TALKING AND THINKING WITH AI: HOW AI CHATBOTS RESTRUCTURE EPISTEMIC IDENTITY AND VIRTUE

Mateja Centa Strahovnik,
Vojko Strahovnik

Introduction[1]

The integration of large language models (LLMs) or AI-driven chatbots into everyday communication and knowledge practices has important implications for how we think and understand ourselves as epistemic agents. The "Just Google it" tack is rapidly getting replaced by "Just ask ChatGPT." This paper investigates how AI chatbots reshape epistemic identity and virtue, focusing on the effects of human-AI interaction. One of the proposals is, thus, that if one wants to fully understand and accommodate the epistemic dimensions of human-AI interactions, one has to employ the notion of epistemic identity. Epistemic identity is the aspect of a person's identity that shapes and is shaped by their ways of forming beliefs, engaging in epistemic practices (such as reasoning, judgment, and discourse), their core epistemic beliefs, and their values and goals regarding knowledge and knowledge-related practices. The use of AI chatbots can ultimately impact one's epistemic identity, including epistemic virtues such as curiosity, intellectual humility, and open-mindedness on the one hand, and raise

[1] This publication was produced within the framework of the research project J6-60105 *Theology and Digitalization: Anthropological and Ethical Challenges* (supported by ARIS—the Slovenian Research and Innovation Agency) and the research project *Epistemic Democracy in a Digital Era* (EDDD; HRZZ-IP-2024-05-4113, funded by the Croatian Science Foundation).

https://doi.org/10.35469/poligrafi.2025.510 | Poligrafi, no. 119/120, vol. 30, 2025, pp. 75–95

questions about epistemic agency and responsibility on the other. This paper emphasizes the importance of reflective practices that maintain human autonomy, foster epistemic responsibility, and promote virtuous collaboration with AI systems. Ultimately, this paper advocates for a partial reimagining of epistemic identity and epistemic collaboration in the age of AI, seeking a balance between the advantages of AI and the significance and uniqueness of human intellectual endeavors.

The evolution from "Just Google it" to "Just ask ChatGPT" represents more than a mere change in the technological preferences of users; it signposts a more fundamental transformation in how humans conceptualize knowledge acquisition and related epistemic practices. This shift marks the transition from information retrieval and evaluation to conversational knowledge generation, fundamentally altering the process of technology-mediated inquiry itself. When we "Google something," we engage (in the good case scenario) in actively searching for, evaluating, and synthesizing information. The user must formulate the search term(s) or query, navigate through multiple sources, assess credibility, and construct knowledge and understanding from disparate pieces of information. Although the process is technology-mediated, it still allows for aspects of epistemic vigilance as the general and critical disposition necessary for responsible knowledge acquisition to be utilized. Dan Sperber conceptualizes epistemic vigilance as a set of (human) cognitive mechanisms that help us evaluate the reliability and trustworthiness of the information we receive from others. Epistemic vigilance operates on many levels and includes varied aspects, e.g. we assess (i) the competence, knowledgeableness, and benevolence of whoever is providing the information, (ii) the plausibility and consistency of the information given what we already know, and (iii) the context and manner of communication—that is, why this is person telling us this, and whether their way of communicating suggests honesty or deception.[2] This, of course, does not mean that the online Google searching

[2] Sperber and his coauthors also argue that although epistemic vigilance is an epistemic virtue, it must be moderated by our other epistemic and practical concerns, to avoid being too vigilant and consequently miss out on valuable information and become overly suspicious of others. Also, epistemic vigilance does not need to always be conscious or deliberate, since it is also effective when it takes place through intuitive and automatic processes. (Dan Sperber et al.,

allows the full potential of epistemic vigilance to be in play, since the entire setting is highly impoverished. Many of the traditional cues that our epistemic vigilance utilizes are either absent or very compromised, e.g., there is no direct social interaction with cues such as someone's face or tone of voice, or the reputation and track record might be missing (given that anonymity and pseudonymity are widespread online). Furthermore, some of the cues themselves are easily manipulated, and some superficial cues could easily mislead (e.g., paying undue attention to the really professional and well-made design of the webpage). Social media sites and algorithms add another layer of complexity through their tendency to create echo chambers and filter bubbles that can reinforce our existing beliefs while limiting exposure to corrective information.[3] And finally, the extent of information available online greatly surpasses offline and might overload our vigilance capacities. Putting aside these limitations, the "Google it" approach preserves some aspects of the investigative dimension of human inquiry.

In contrast to this, when someone "asks ChatGPT," they enter into what appears to be a dialogue with a seemingly (extremely) knowledgeable interlocutor. In most cases, the AI chatbot presents information in the form of complete and coherent responses that mimic human expertise. This might give rise to a certain sense of epistemic immediacy, i.e., the instantaneous provision of seemingly authoritative knowledge. This is not to say that one is prohibited from utilizing as many tools of epistemic vigilance as possible, even in this context. But the inherent danger is obvious. This is the shift being made from the model of "distributed cognition" and humans actively orchestrating various information sources, toward a model of "delegated cognition" where huge and highly capable AI systems perform much of the cognitive or epistemic work. Such a shift also includes an impact on epistemic habits, expectations, and practices, and ultimately, on our epistemic virtues and epistemic identity. Furthermore, such AI chatbots typically provide information (as a sort of decontextualized knowledge) stripped of

---

"Epistemic Vigilance," *Mind & Language* 25, no. 4 (2010): 359–393, https://doi.org/10.1111/j.1468-0017.2010.01394.x).

[3] Thi C. Nguyen, "Echo Chambers and Epistemic Bubbles," *Episteme* 17, n. 2 (2020): 141–161, https://doi.org/10.1017/epi.2018.32.

its original context, i.e., the debates, methodologies, and assumptions that shaped its production, which might be retained to a considerable degree in the "just Google it" paradigm. While some might brush aside such a gloomy predicament as an overreaction in light of recent developments related to AI chatbots, one should nonetheless be careful and pay attention to empirical findings. Some recent behavioral studies indicate that users spend (significantly) less time cross-referencing sources when using conversational AI compared to traditional search engines. Also, the latter type of system is perceived as more useful, enjoyable, and offers higher levels of satisfaction.[4]

Michael Lynch proposes that we can frame the developments related to our epistemic agency and epistemic environment in the digital age in terms of the difference between pollution and corruption. "Corruption is not the same as pollution. Pollution is something that happens to a system; corruption is something that happens within a system. […] (I)nformation culture is corrupt when the rules of evidence and reliability that some of its participants allegedly adhere to—their epistemic principles, in other words—are not the ones they more frequently employ. This phenomenon might be what some people mean when they talk about living in a "post-truth" culture. Of course, we don't literally live in a world where nothing is true. Truth exists as much as it ever has. What has happened is that our information culture has become so corrupt as to tolerate and encourage self-deceptive attitudes toward truth and evidence. It encourages us to care more about our convictions than about truth, but to tell ourselves we are doing otherwise."[5] What makes the case of AI chatbots specific is that such corruption is covert and thus very hard to detect.

This paper delves deeper into some of the aspects related to this shift, emphasizing its inherent risks and some possible mitigation measures. The paper proceeds in the following way. Section 2 elaborates on the notions of epistemic identity and epistemic virtue in more detail and

---

[4] Ruiyun Rayna Xu, Yue Feng and Hailiang Chen, "ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience," *ArXiv*, https://doi.org/10.48550/arXiv.2307.01135.

[5] Michael P. Lynch, *Know-it-All Society: Truth and Arrogance in Political Culture* (New York, NY: WW Norton, 2019), 36.

highlights their interconnectedness. Section 3 explores some of the possible impacts of the use of AI systems on our epistemic identity and epistemic virtues, and critically evaluates them. It also addresses the question of how we should conceptualize, from the epistemic aspect, the relevant AI systems. At the end (subsection 3.1), the impacts of the use of AI systems and human-AI interaction on epistemic virtues are highlighted. Section 4 concludes the paper with a proposal for restructuring the conceptualization of human epistemic identity in light of the emergence of generative AI systems and widespread human-AI interactions. The key aspect of the proposal is to conceive or envisage them as epistemic partners.

Epistemic Virtue and Epistemic Identity

Epistemology recently underwent an important shift to virtue-based theorizing. As part of this turn, the principal focus shifted from the properties of one's beliefs to the characteristics of the epistemic agent, in particular to epistemic processes and character traits or virtues. The simplest way to define epistemic virtues is to frame them as the epistemic (intellectual or cognitive) qualities (abilities, dispositions, learned habits, personality traits, etc.) of individuals that help them pursue their epistemic goals (true beliefs about the world, knowledge, understanding, wisdom, etc.). Epistemic virtues are thus dispositions that make someone good at acquiring knowledge, forming true beliefs, etc. They include things like intellectual honesty, curiosity, open-mindedness, critical thinking, humility about one's epistemic limitations, epistemic perseverance, etc. Consequently, epistemic virtuousness is the state of possessing and exercising epistemic virtues, which means being intellectually virtuous in how one seeks information, evaluates evidence, and forms beliefs. It requires consistently exhibiting responsible intellectual character—that is, being disposed to think, inquire, and believe in ways that are conducive to gaining knowledge and avoiding falsehood. Epistemic virtuousness includes fulfilling one's epistemic responsibilities. Intellectual virtues like curiosity, honesty, and thoroughness are not merely cognitive skills but involve deep motivational commitments

to truth and understanding.[6] It is also worth pointing out that the focus on epistemic virtues and epistemic virtuousness also highlights the importance of epistemic agency and the agent's epistemic responsibility

In order to capture the comprehensive impact of human-AI interactions, one can go a step further than this and introduce the term epistemic identity, which is thus far fairly underdeveloped and not utilized much in the literature. Furthermore, such a notion can be very useful when considering human-AI interaction from the epistemic point of view. We propose delineating it along the following lines. Epistemic identity can be initially understood as the aspect of one's identity that is intimately intertwined with one's manner and scope of forming beliefs and other epistemic practices (such as thinking and reasoning, judgment, assessment of credibility, engagement in discourse, etc.), one's beliefs about the essence of knowledge, one's associated epistemic standards, and one's epistemic sensibility. A similar but somewhat narrower use of the term 'epistemic identity' is pervasive in education studies and psychology, and is closely related to the learning styles of students and their stances toward knowledge and knowledge-acquiring practices. E.g., one can find talk of epistemic identity associated with talk about epistemic positions, e.g., "evaluative or reflective," "absolutist," "personal," and "rule-based" epistemic positions.[7] Epistemic identity can also mean treating certain beliefs as central or part of one's identity. As such, it can influence the information we attend to, remember, and seek, as well as how we reason. Finally, epistemic identity includes what one sees as a worthy goal of epistemic endeavors or practices.[8] In this

---

[6] Linda Zagzebski, *Virtues of the Mind* (Cambridge, MA: Cambridge University Press, 1996); Vojko Strahovnik, "Uvod v vrlinsko epistemologijo [Introduction to Virtue Epistemology]," *Analiza,* 8, no. 3 (2004): 101–118.

[7] Marie L. Caltabiano, Raoul J. Adam and Rebecca Denham, "Epistemic Identity and Undergraduate Students' Understandings of Psychology," *International Journal of Education, Psychology and Counselling* 4, no. 30 (2019): 299–314, https://gaexcellence.com/ijepc/article/view/3045; Elise J. West, "Perry's Legacy: Models of Epistemological Development," *Journal of Adult Development* 11 (2004): 61–70, https://doi.org/10.1023/B:JADE.0000024540.12150.69.

[8] Besides linking epistemic identity to individuals, there is also a sense that ties this notion to groups or communities. One such framing understands such epistemic identity as partially constituted by epistemic practices as "socially organised and interactionally accomplished ways that members of a group propose, communicate, assess, and legitimise knowledge

sense, it pertains to a central aspect of one's epistemic agency. More systematically, epistemic identity can be defined as

> a structure that consists of three main elements, that is (i) basic epistemic beliefs, which can be either identity-constituting or identity-affecting, (ii) epistemic practices, dispositions, virtues, habits, capacities, etc., and (iii) one's epistemic values (in the broader sense, which includes such things as epistemic goals or aims, priorities, ideals, norms, standards, sensitivity, etc.).

Virtues and virtuousness are thus an integral part of one's epistemic identity. In the remainder of this section, we will address some open questions pertaining to the proposed notion of epistemic identity. Nick Byrd relates epistemic identity to "the phenomenon of treating certain beliefs as part of one's identity," which can lead to a situation in which "I might prioritize my epistemic identity over other epistemic goods" and such an identity can subsequently "influence what we seek, what we attend to, what we perceive, and thereby what we remember, whom we listen to, and how we reason."[9] It should be emphasized that Byrd does not limit the scope of the relevant identity-determining beliefs to epistemic beliefs; for him, an epistemic identity can be determined or shaped by any sort of belief that has consequences for our other beliefs and epistemic practices.[10]

One open question regarding such a view is whether such beliefs are constitutive of epistemic identity or merely beliefs that influence epistemic identity. We propose understanding epistemic identity in the former, narrower way. Epistemic identity thus comprises those beliefs that are themselves epistemic beliefs (e.g., "In order to know something, my

---

claims". Gregory J. Kelly and Peter Licona, "Epistemic Practices and Science Education," in *History, Philosophy and Science Teaching: New Perspectives*, ed. Michael R. Matthews (Cham: Springer Verlag, 2017), 139–165.

[9] Nick Byrd, "Bounded Reflectivism and Epistemic Identity," *Metaphilosophy* 53, no. 1 (2021): 57–58, https://doi.org/10.1111/meta.12534.

[10] This is clear from the examples that Byrd uses to demonstrate this, such as the following. "Suppose that I identify with a religion. If you criticise some aspect of my religion, then I might reflectively defend my religious beliefs rather than dispassionately submit to the best arguments and evidence. In short, I might prioritise my epistemic identity over other epistemic goods. Or suppose that you identify with a particular political party—one that explicitly codifies its ideological commitments in a party platform that is recited in its public speeches, advertisements, and so on. In other words, you identify not only with the party but also with its values and beliefs. In this case your political identity is an epistemic identity." (*Ibid.*, 57–58)

beliefs must be accompanied by a feeling of certainty." or "One should not trust one's senses on all occasions, since they might be misleading.") or beliefs that significantly affect other aspects of epistemic identity (e.g., beliefs like "I always go with my gut feelings when it comes to evaluating other people's motives."). Holding a particular religious or political belief strongly does not necessarily mean that this affects one's epistemic identity *per se* (although they are undoubtedly a part of the overall identity). Epistemic identity pertains more centrally to how we know (and believe) than what we know (and believe). We also need to take into account that someone might hold particular epistemic beliefs about themselves that are not true or not well supported. For example, one might think of oneself as a truly unbiased and fair evaluator or grader without this being the case. In such cases, these beliefs constitute one's self-ascribed epistemic identity. This self-ascribed epistemic identity might thus differ from actual epistemic identity, though it is important to note that both can significantly affect other beliefs, motives, practices, and values.

There are several other attempts that try to elaborate on what determines epistemic identity, how it is formed, and what purpose it serves. Loren Demerath speaks about epistemological identity theory, which she sees as explaining "how individuals enhance their knowledge of self and the world by creating and maintaining identities" and this "theory reconceptualizes commitment to an identity as the degree to which that identity organizes and clarifies one's experience of the world and him/herself."[11] She further claims that we are continually constructing or reconstructing our identity; i.e. we strive to preserve, improve, or alter our self-perception and the epistemological identity theory contends that we act in this way for epistemological reasons: to give us a feeling of meaning(fullness) by making us feel as though we understand our world, ourselves, and the ties that connect them. We create our identities in order to maintain that sense of meaning and significance. Identity loss could result in emotional and cognitive disruption—a loss

---

[11] Loren Demerath, "Epistemological Identity Theory: Reconceptualising Commitment as Self-Knowledge," *Sociological Spectrum* 26, no. 5 (2006): 491, https://doi.org/10.1080/02732170600786208.

of epistemic certainty and security regarding who we are. This latter aspect relates more to the function or role of epistemic identity.[12]

Lastly, one can also mention another important element of epistemic identity. One's epistemic identity also includes what one sees as a worthy goal of one's epistemic practice, e.g., truth, forming and maintaining beliefs based on sufficient evidence, forming and maintaining beliefs that must be in some way significant, relevant, or useful, knowledge, understanding, believing in accordance with the intellectual virtues, etc. Standards and norms that we follow in epistemic practices or deem important for them are also part of the evaluative dimension. Although there are different proposals on how to frame and understand epistemic virtue and epistemic identity, the general idea sketched above should be clear enough for our purposes.

Aliens in the Space of Reasons: AI Chatbots and Human-AI Interactions

Why is the aspect of epistemic identity important in human-AI relationships or, to put it more neutrally, in our use of AI systems? We do not particularly highlight or question the impact on our epistemic identity when getting a soda from a vending machine, driving a car, using a public transportation system, or using a calculator. What makes AI and AI implementations special in this regard? Furthermore, are there any specific features of epistemic identity that make it more malleable in light of human-AI interaction (as opposed to our other forms of identity)? Initially, one could raise the following considerations. For some other aspects of the human-AI relationship, e.g., AI's moral responsibility or the possibility of a genuine emotional or personal relationship, the influencing force of AI can be more easily dismissed on the basis that an AI system merely responds in a manner that can be interpreted as emotional engagement (e.g. providing care, companionship, friendship, or romantic involvement in a reciprocal way)[13] or that

---

[12] *Ibid.*, 492.

[13] Anne Zimmerman, Joel Janhonen and Emily Beer, "Human/AI relationships: challenges, downsides, and impacts on human/human relationships," *AI Ethics* 4 (2024), 1555–1567,

it makes no sense at all to ascribe moral responsibility to AI-based systems, e.g. self-driving cars. (This is not to say that someone else cannot be morally responsible, such as the programmers or the company that produces and sells the car.) Now, the domain of epistemic identity is not so clear-cut. Although it can of course be questioned whether an AI system like an LLM model really engages in reasoning, offers us explanation or other epistemic activities, it is at least sensible to look into the human-AI relationship from this angle.

One starting question could be whether AI systems, like ChatGPT, DeepSeek, or Claude, are themselves epistemic agents. What does it mean to be an epistemic agent? A conventional answer is that an epistemic agent is an agent with the ability to perform knowledge-related activities, such as gathering, processing, analyzing, evaluating, and utilizing information. This includes forming beliefs, justifying them, revising them in light of new information or evidence, and utilizing them in further tasks. Also included is the ability to communicate knowledge and beliefs to others, as well as to understand and interpret the knowledge communicated by others. Lastly, agency usually includes a certain degree of autonomy and intentionality, related to knowledge-seeking behavior in the case of epistemic agency. *Prima facie,* AI systems such as LLM-based chatbots do not meet the conditions for epistemic agency since they lack, at least, the required mental states such as beliefs and intentions. But AI chatbots might well emulate all this in a way that nudges our perception of such AI systems toward treating them as agents. It is also worth noting that the subjective mental models, including the range of attitudes or stances that we, humans form and adopt toward such systems, can be primed in ways that inhibit their recognition as epistemic agents.[14] Advertising by AI developers cer-

---

https://doi.org/10.1007/s43681-023-00348-8.

[14] Bojan Žalec, "Ali je umetna inteligenca inteligenca v pravem pomenu besede?: vprašanje psihičnih značilnosti in splošnosti [Is artificial intelligence an intelligence in the true sense of the word?: the issue of mental characteristics and generality]," *Bogoslovni vestnik* 83, no. 3 (2023): 816–822, https://doi.org/10.34291/BV2023/04/Zalec; Pat Pataranutaporn et al., "Influencing human-AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness," *Nature Machine Intelligence* 5 (2023), 1076–1086, https://doi.org/10.1038/s42256-023-00720-7.

tainly uses this aspect when trying to sell AI system as products under names like "assistant," "companion," "tutor," "copilot," etc.

A similar aspect is hinted at by Bert Heinrichs and Sebastian Knell in the paper in which they introduce the talk about artificial intelligent agents (AIA) as "aliens in the space of reasons."[15] The space of reasons can be delineated as the domain of epistemic activity that is governed by normativity, i.e., the domain in which concepts, beliefs, and justifications operate in relation to knowledge as the final goal.[16] Participation in the space of reasons entails being subject to normative constraints, e.g., bound by what counts as a reason for belief or action. Being in the space of reasons is not merely a matter of being causally responsive to the environment; it is to be accountable to norms of justification and assertion. Heinrichs and Knell argue that AIAs, e.g., widely used AI recommendation systems or AI chatbots, have already seemingly entered the space of reasons since they make knowledge claims and attempt to justify, explain, and elaborate on them. But this is a mere appearance; AIAs do not quite play by the rules of the space of reasons. And thus the "alien" talk captures how AI systems appear to participate in rational discourse and knowledge claims but operate according to completely different principles than human reasoners. In particular, they don't engage in the same kind of reason-giving, responsibility-taking practices that characterize genuine participation.

A more modest view, but a view that nonetheless captures the special significance of AI systems for considering epistemic identity and virtue, is that AI systems are different to other technologies since they are very closely related to human epistemic agency and epistemic practices. Such a move substantially aligns with Ramon Alvarado's proposal that what makes AI so special is that it is, first and foremost, an example of

---

[15] Bert Heinrichs and Sebastian Knell, "Aliens in the Space of Reasons? On the Interaction Between Humans and Artificial Intelligent Agents," *Philosophy & Technology* 34 (2021), 1569–1580, https://doi.org/10.1007/s13347-021-00475-2.

[16] The notion of the "space of reasons" emerges from the thought of Wilfrid Sellars and is linked to his characterisation of knowledge, i.e., to his claim that "in characterising an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says" (Wilfrid Sellars, *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1956), 169.)

epistemic technology and, furthermore, a type of epistemic technology in a very strong or prominent sense. AI's defining characteristic lies in its essence as an epistemic technology, which separates it from other technologies that we use or interact with. Alvarado highlights that AI is a case of epistemic technology because (i) it is primarily employed in epistemic contexts (such as inquiry), (ii) it primarily deals with epistemic content (e.g., propositions, models, and data), and (iii) it carries out epistemic operations or tasks on epistemic content (such as analysis, predictions, inferences, recognitions, etc.).[17] And it is not merely AI's design or what it can do, but AI is an epistemic technology in terms of what it is used for as well.[18]Even if we go with the more modest proposal of thinking of AI systems as pieces of epistemic technology, one can meaningfully question the impact of such a technology on human epistemic agents and their epistemic identity through technological mediation. Peter-Paul Verbeek convincingly demonstrated that technologies do not merely provide neutral tools for pre-existing human purposes but actively shape human experience and agency.[19] Consequently, and in the context of this paper, AI systems do not simply help us access knowledge, but they significantly alter what counts as knowledge, how knowledge is validated, and what kinds of epistemic agents we become. Furthermore, one can speak about a new type of epistemic dependence.[20] AI systems as epistemic technology introduce epistemic dependence that is neither fully social or interpersonal (since AI systems lack certain aspects that would make them epistemic agents) nor fully individual (since AI systems embody collective human knowledge, thought in complex ways, which differs from social knowledge). All this culminates in the technological mediation of human epistemic identity operating through human-technology relations,[21] where AI

---

17 Ramon Alvarado, "AI as Epistemic Technology," *Science and Engineering Ethics* 29, no. 32 (2023), 19, https://doi.org/10.1007/s11948-023-00451-3.

18 *Ibid*., 15.

19 Peter-Paul Verbeek, *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. (University Park: Pennsylvania State University Press, 2005).

20 John Hardwig, "Epistemic dependence," *Journal of Philosophy* 82, no. 7 (1985), 335–349, https://doi.org/10.2307/2026523.

21 Don Ihde, *Technology and the Lifeworld: From Garden to Earth* (Bloomington: Indiana University Press, 1990).

systems assume various roles, e.g., the role of an extension of our cognitive capacities, interpreter of complex information, quasi-agent, or represent a set of background conditions that subtly shapes our epistemic environment. One of the central questions is, therefore, how all this impacts (enhance, diminish, etc.) human epistemic agency, including the question of whether we are witnessing the emergence of genuinely augmented epistemic agents, or we are observing a gradual atrophy of essential intellectual capacities. Before returning to the question about epistemic identity, let us consider the issues mentioned from the perspective of epistemic virtues, in particular epistemic curiosity, humility, and open-mindedness.

AI and Epistemic Virtues

Epistemic curiosity might be framed as a disposition to seek knowledge and understanding, as well as being willing to pursue questions even when they lead to complexity or uncertainty. Curiosity involves both the motivation to learn and the persistence to pursue epistemic inquiry. AI chatbots might impact epistemic curiosity by reducing exploratory questioning and research in light of instant responses to questions. At least, such AI systems (*via* immediateness and seeming comprehensiveness) alter the temporal structure of inquiry. Traditional inquiry involves what may be labeled "epistemic friction"—the necessary delays, dead ends, and unexpected discoveries (creating space for reflection and the development of investigative skills and epistemic virtues). Such friction serves an important function. The AI chatbots often respond in a confident, complete, and authoritative manner (even when "hallucinating"), already synthesizing multiple perspectives[22] and creating a podium for thinking that understanding has been achieved through the transfer of information alone. While AI chatbots might give us access to an extremely broad array of topics, highlighting them

[22] Such AI responses may play into or satisfy what Daniel Kahneman identifies as our cognitive need for closure, i.e., the need to attain precise knowledge instead of confusion and ambiguity on a subject and the desire to make a decision as soon as possible (Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011)).

with explanations etc., and thus satisfying our curiosity, the final epistemic goal would nonetheless be the development of what John Dewey called "reflective thinking"—the sustained effort to understand not just what is the case, but why it is the case and how it relates to broader patterns of knowledge, including a disposition to suspend judgment and maintain a healthy skepticism when apt.[23]

Intellectual humility pertains to the apt recognition of one's epistemic limitations, coupled with appropriate confidence in areas of genuine knowledge. It includes calibrating one's confidence to one's actual epistemic position and remaining open to correction while acknowledging the apt epistemic status of others. Intellectual humility thus implies that we have an adequate or realistic and non-haughty look at our intellectual capabilities, strengths, and weaknesses, and that we exhibit an ability to properly assess and evaluate different ideas and views in a way that includes respect for those who do not agree with us, etc. It therefore includes intrapersonal and interpersonal dimensions. It enables us to establish a proper relationship with ourselves as epistemic agents, which *inter alia* includes being open to new facts and insights, an ability to integrate new knowledge into our existing knowledge, an ability to assess the relevance of this knowledge, etc. At the same time, it puts us into a cognitive space with others in a way that allows non-haughty, non-condescending, and solidary participation in the common pursuit of truth and in public discourse. Understood in this way, we can distinguish intellectual humility as an epistemic virtue.[24] One difficulty arising from human-AI epistemic interactions is the question of knowing what these systems know and don't know. Traditional epistemic humility assumes that we can, at least in principle, evaluate the reliability of different knowledge sources. But algorithmic systems operate through forms of "intelligence" that seem fundamentally different from human cognition. When engaging with AI systems that

---

23 John Dewey, *How We Think* (Lexington, MA: D.C. Heath & Co., 1910).

24 Vojko Strahovnik, "Humility, religions, and dialogue," *Poligrafi* 22, no. 87/88 (2017): 3-21, https://ojs.zrs-kp.si/index.php/poligrafi/issue/view/32; Mateja Centa Strahovnik and Vojko Strahovnik, "Epistemic virtues and interreligious dialogue: a case for humility," *Annales - Series historia et sociologia* 30, no. 3 (2020): 395–404, https://doi.org/10.19233/ASHS.2020.25.

demonstrably outperform human reasoning in specific domains, humility may inadvertently foster epistemic deference rather than critical engagement. Addressing this requires epistemic vigilance, not merely humility, in order to challenge and interrogate the hidden biases encoded within AI systems.

Open-mindedness is a virtue that centers on the willingness to consider alternative viewpoints and revise beliefs in light of new evidence. Human-AI epistemic interactions significantly affect and alter open-mindedness. The first issue concerns possible over-reliance on AI-generated answers or suggestions, replacing critical evaluation with passive acceptance. Second, AI systems can systematically narrow the range of perspectives and evidence that humans encounter. Third, AI systems can undermine people's confidence in their own capacity to evaluate evidence and consider alternative viewpoints. Lastly, AI systems can make it difficult for humans to distinguish between their own thinking and AI-influenced thoughts, and furthermore detect patterns of what type of influences individuals are most prone to.[25]

In addition to the impact on particular epistemic virtues (not necessarily on their conceptualization, but e.g., domain and manner of operation), over-reliance on AI-generated knowledge poses several distinct epistemic risks that go beyond simple concerns about accuracy. Even when AI systems provide correct information, excessive dependence can undermine the development and maintenance of crucial intellectual capacities.[26] This is not to deny the epistemic usefulness of such AI systems. Some recent studies have highlighted their effectiveness in the learning process. E.g., Greg Kestin and his collaborators developed an AI tutor and have established that "when students interact with our AI tutor, at home, on their own, they learn significantly more than when

---

[25] Rosenberg, Louis, "The Manipulation Problem: Conversational AI as a Threat to Epistemic Agency," *arXiv* (2023), https://doi.org/10.48550/arXiv.2306.11748.

[26] Nataliya Kosmyna et al., "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task," *arXiv* (2025), https://arxiv.org/abs/2506.08872.

they engage with the same content during an in-class active learning lesson, while spending less time on the task."[27]

Conclusion: From Knowing About and From AI to Knowing with AI

Before considering the reconceptualization of human-AI epistemic interactions, we will briefly consider some of the impacts of human-AI interactions on epistemic agency and epistemic responsibility. Epistemic agency refers to an individual's capacity to actively direct their own cognitive processes, evaluate information critically, and take responsibility for their beliefs and knowledge claims. This concept sits at the intersection of epistemology, philosophy of mind, and moral philosophy, as it concerns both the mechanics of belief formation and the normative dimensions of intellectual responsibility. First, as AI systems become more capable and prevalent, human users may experience automation bias as the tendency to over-rely on automated systems and under-utilize human judgment.[28] Coupled with the reduced critical evaluation of AI-provided information, this can have detrimental effects on our epistemic identity and knowledge. Second, there is a tendency to defer to AI systems without careful evaluation of their reliability, but due to their perceived technical sophistication. Third, cognitive offloading,[29] as the delegation of tasks to AI systems, can alter the characteristics of epistemic agency itself. In the context of human-AI interaction, what is called for is reflective engagement—the maintenance of a critical distance even while benefiting from AI capabilities, which in turn might involve meta-epistemic skills or virtues to evaluate not just information but the processes by which information is transmitted and validated. Lastly, as part of effective human-AI collaboration, humans need to retain a sort of strategic epistemic control, which is responsibility for

---

[27] Kestin, Greg, et al. "AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting," *Scientific Reports* 15 (2025), 17458, https://doi.org/10.1038/s41598-025-97652-6.

[28] Lisanne Bainbridge, "Ironies of automation," *Automatica* 19, no. 6 (1983): 775–779, https://doi.org/10.1016/0005-1098(83)90046-8.

[29] Andy Clark and David Chalmers, "The extended mind," *Analysis* 58, no. 1 (1998): 7–19.

setting goals, evaluating outputs, and making final judgments about knowledge claims. AI systems could contribute to this and inform human epistemic practices without replacing them. Epistemic responsibility refers to the obligation to form, maintain, accept, and convey beliefs in a responsible manner, including seeking sufficient evidence, avoiding benightedness and ignorance, being open to revision when presented with contrary evidence, etc. The question of responsibility for AI-mediated knowledge claims presents novel challenges that the existing frameworks of epistemic responsibility are not fully equipped to address. The standard view of epistemic responsibility assumes that knowledge claims can be traced to identifiable human agents who can be held accountable for the evidence and reasoning supporting their assertions, to whom the knowledge assertion could be attributed, and who can elaborate on it further (answerability).[30] AI systems complicate this picture by introducing what we might call opacity and distributed agency problems that make responsibility attribution vague and difficult. The challenge is that AI systems are neither fully autonomous agents (they lack genuine understanding, intentionality, and thus the authorship status)[31] nor simple tools (they exhibit complex behaviors that can surprise). This liminal status requires updating the models of epistemic responsibility attribution.

One possible framing of the changes in the epistemic environment brought about by AI chatbots is the move to conceptualize them as human epistemic partners instead of mere epistemic tools. Why would such a move be sensible and advantageous? Epistemic tools could be framed as technologies that are designed in a way that cognitively empower the human mind while humans still maintain agency throughout the knowledge-production process. One of the most straightforward examples of such tools are calculators that boost computational ability but at the same time necessitate a human understanding of mathematical

---

[30] Cf. David Shoemaker, "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics* 121, no. 3 (2011): 602–632, https://doi.org/10.1086/659003.

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

concepts, proof construction, and problem-solving strategies. The same thing can be said about libraries, databases, and search engines, which are not only places that contain information, but they are also places where aspects of the synthesis, evaluation, and application are performed. Still, they remain in the role of epistemic tools.

Epistemic tools typically exhibit the following characteristics:

- Transparency: Their operations are relatively understandable to users.
- Controllability: Users maintain significant control over how they are employed.
- Complementarity: They enhance rather than replace human cognitive abilities.
- Selectivity: They require human judgment to select, understand, interpret, and apply their outputs.

Epistemic partners, on the other hand, are systems that seem to be directly involved in the process of knowledge creation and transmission, so they not only enable human epistemic agency but actually change its character. AI systems, especially LLM, are frequently epistemic partners, as they not only supply information but can also do the work of understanding, combining, and creatively thinking that is similar to human epistemic practice.

Epistemic partners are characterized by the following core features:

- Opacity: Their internal operations are largely opaque to their users.
- Autonomy: They operate according to their own internal logic rather than direct human control.
- Substitutability: They can replace rather than merely augment human epistemic endeavors.
- Generativity: They produce novel outputs rather than simply retrieving and transmitting existing information.

A couple of further elaborations. First, one might raise the concern that the use of the term 'partner' suggests substantive agency, responsibility, and (epistemic) trust on both sides. But the term partner is used in a broader sense here, as in when we use the term for speaking about, e.g., a sparring partner. Alternatively, one can speak about quasi-partnership as pertaining to the status and nature of interaction

as elaborated above. Second, note that the partner-tool distinction is not absolute; the same technology can function as either, depending on how it is used. E.g., AI systems can function as tools when used to generate ideas that are then critically evaluated, or as partners when their outputs are accepted without substantial human contribution. This framework thus suggests that AI systems have evolved beyond the traditional tool relationship into something approaching genuine epistemic partnership; they are entities that bring their own capabilities and contributions to the collaborative pursuit of knowledge, even if they lack consciousness or intentionality in the human sense.

To conclude, contemporary AI systems occupy an ambiguous position. They are marketed and often experienced as partners—entities capable of understanding questions, providing explanations, and engaging in dialogue. However, they lack the genuine understanding, intentionality, and responsibility that characterize human epistemic partners. This creates what we might call "pseudo-partnership," or "quasi-partnership"—a relationship that has the surface features of epistemic collaboration without its deeper structures of mutual understanding and shared responsibility. Nonetheless, the partnership status of AI systems might help us navigate the AI chatbot-populated epistemic environment better.

Bibliography

Alvarado, Ramon. "AI as Epistemic Technology." *Science and Engineering Ethics* 29, no. 32 (2023). https://doi.org/10.1007/s11948-023-00451-3.

Bainbridge, Lisanne. "Ironies of automation." *Automatica* 19, no. 6 (1983): 775–779. https://doi.org/10.1016/0005-1098(83)90046-8.

Byrd, Nick. "Bounded Reflectivism and Epistemic Identity." *Metaphilosophy* 53, no. 1 (2022): 53–69. https://doi.org/10.1111/meta.12534.

Caltabiano, L. Marie, Raoul J. Adam, and Rebecca Denham. "Epistemic Identity and Undergraduate Students' Understandings of Psychology." *International Journal of Education, Psychology and Counseling* 4, no. 30 (2019): 299–314. https://gaexcellence.com/ijepc/article/view/3045.

Centa Strahovnik, Mateja, and Vojko Strahovnik. "Epistemic virtues and interreligious dialogue: a case for humility." *Annales - Series historia et sociologia* 30, no. 3 (2020): 395–404. https://doi.org/10.19233/ASHS.2020.25.

Clark, Andy, and David Chalmers. "The Extended Mind." *Analysis* 58, no. 1 (1988): 7–19.

Demerath, Loren. "Epistemological Identity Theory: Reconceptualizing Commitment as Self-Knowledge." *Sociological Spectrum* 26, no. 5 (2006): 491–517. https://doi.org/10.1080/02732170600786208.

Dewey, John. *How We Think*. Lexington, MA: D.C. Heath & Co., 1910.

Hardwig, John. "Epistemic dependence." *Journal of Philosophy* 82, no. 7 (1985), 335–349. https://doi.org/10.2307/2026523.

Heinrichs, Bert, and Sebastian Knell. "Aliens in the Space of Reasons? On the Interaction Between Humans and Artificial Intelligent Agents." *Philosophy & Technology* 34 (2021): 1569–1580. https://doi.org/10.1007/s13347-021-00475-2.

Ihde, Don. *Technology and the Lifeworld: From Garden to Earth.* Bloomington: Indiana University Press, 1990.

Kahneman, Daniel. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux, 2011.

Kelly, Gregory J., and Peter Licona. "Epistemic Practices and Science Education." In *History, Philosophy and Science Teaching: New Perspectives*, edited by Michael R. Matthews, 139–165. Cham: Springer Verlag, 2018.

Kestin, Greg, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. "AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting." *Scientific Reports* 15 (2025), 17458. https://doi.org/10.1038/s41598-025-97652-6

Kosmyna, Nataliya, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task." *arXiv* (2025). https://arxiv.org/abs/2506.08872.

Lynch, P. Michael. *Know-it-All Society: Truth and Arrogance in Political Culture*. New York, NY: WW Norton, 2019.

Nguyen, C. Thi. "Echo Chambers and Epistemic Bubbles." *Episteme* 17, no. 2 (2020): 141–161. https://doi.org/10.1017/epi.2018.32.

Pataranutaporn, Pat, Ruby Liu, Ed Finn, and Pattie Maes. "Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy, and effectiveness." *Nature Machine Intelligence* 5 (2023): 1076–1086. https://doi.org/10.1038/s42256-023-00720-7.

Rosenberg, Louis. "The Manipulation Problem: Conversational AI as a Threat to Epistemic Agency." *arXiv* (2023). https://doi.org/10.48550/arXiv.2306.11748.

Sellars, Wilfrid. *Science, Perception, and Reality*. London: Routledge and Kegan Paul, 1956.

Shoemaker, David. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121, no. 3 (2011): 602–632. https://doi.org/10.1086/659003.

Sperber, Dan, Fabrice Clement, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. "Epistemic Vigilance." *Mind & Language* 25, no. 4 (2010): 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x.

Strahovnik, Vojko. "Uvod v vrlinsko epistemologijo [Introduction to Virtue Epistemology]." *Analiza* 8, no. 3 (2004), 101–118.

Strahovnik, Vojko. "Humility, religions, and dialogue." *Poligrafi* 22, no. 87/88 (2017): 3–21. https://ojs.zrs-kp.si/index.php/poligrafi/issue/view/32.

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

Verbeek, Peter-Paul. *What Things Do: Philosophical Reflections on Technology, Agency, and Design.* University Park: Pennsylvania State University Press, 2005.

West, J. Elise. "Perry's Legacy: Models of Epistemological Development." *Journal of Adult Development* 11, no. 2 (2004). https://doi.org/10.1023/B:JADE.0000024540.12150.69.

Xu, Ruiyun Rayna, Yue Feng, and Hailiang Chen. "ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience." *arXiv*:2307.01135. https://doi.org/10.48550/arXiv.2307.01135.

Zagzebski, Linda. *Virtues of the Mind.* Cambridge, MA: Cambridge University Press, 1996.

Zimmerman, Anne, Joel Janhonen, and Emily Beer. "Human/AI relationships: challenges, downsides, and impacts on human/human relationships." *AI Ethics* 4 (2023): 1555–1567. https://doi.org/10.1007/s43681-023-00348-8.

Žalec, Bojan. "Ali je umetna inteligenca inteligenca v pravem pomenu besede?: vprašanje psihičnih značilnosti in splošnosti [Is artificial intelligence an intelligence in the true sense of the word?: the issue of mental characteristics and generality]." *Bogoslovni vestnik* 83, no. 3 (2023), 813–823. https://doi.org/10.34291/BV2023/04/Zalec.