

THE VALUE OF TESTIMONIAL-BASED BELIEFS IN THE FACE OF AI-GENERATED QUASI-TESTIMONY

[EL VALOR DE LAS CREENCIAS TESTIMONIALES ANTE LOS QUASI-TESTIMONIOS PRODUCIDOS POR IA GENERATIVA]

Felipe Álvarez *

Universidad Andrés Bello & Universidad de Chile, Chile

Ruth Espinosa **

Universidad Andrés Bello, Chile

ABSTRACT: The value of testimony as a source of knowledge has been a subject of epistemological debates. The "trust theory of testimony" suggests that human testimony is based on an affective relationship supported by social norms. However, the advent of generative artificial intelligence challenges our understanding of genuine testimony. The concept of "quasi-testimony" seeks to characterize utterances produced by non-human entities that mimic testimony but lack certain fundamental attributes. This article analyzes these issues in depth, exploring philosophical perspectives on testimony and the implications of conversational AI technologies on our epistemic practices.

KEYWORDS: Trust; Testimony; AI; Epistemology; Social normativity.

RESUMEN: El valor del testimonio como fuente de conocimiento ha sido sujeto de debates epistemológicos. La "trust theory of testimony", por ejemplo, sugiere que el testimonio humano se basa en una relación afectiva fundada en las normas sociales. Sin embargo, el surgimiento de la inteligencia artificial generativa desafía nuestro entendimiento del testimonio genuino. Así, el concepto de "quasi-testimonio" busca caracterizar las afirmaciones realizadas por entidades no-humanas que imitan el testimonio, pero que carecen de ciertos atributos fundamentales. Este artículo analiza ese problema en profundidad, explorando las perspectivas filosóficas en torno al testimonio y las implicaciones de las tecnologías de IA conversacionales en nuestras prácticas epistémicas.

PALABRAS CLAVE: Confianza; Testimonio; AI; Epistemología; Normatividad social.

1 INTRODUCTION

We learn from the words of others. If one of us notices a rare skin lesion on a leg, it would be natural to consult an expert for diagnosis and treatment. Since most of us have no training in dermatology, we would straightforwardly accept the specialist's testimony as enough grounding to trust her judgment. Since we are

* Felipe Álvarez holds a Master's in Philosophy from the Universidad de Chile. He is currently a Ph.D. candidate in Philosophy at the University of Chile (ANID-PFCHA/Doctorado Nacional/Año 2022 — 21220627), while working as an adjunct assistant professor at the Universidad Andrés Bello. His areas of interest are contemporary epistemology and social ontology, with a focus on testimony and the Internet. E-mail: f.lvarezosorio@uandresbello.edu

** Ruth Marcela Espinosa Sarmiento holds a PhD in Philosophy from the University of Leipzig (Germany) and a Master's in Philosophy with a specialization in epistemology from the University of Chile. Her research areas include modern and contemporary epistemology. She is an associate professor and currently serves as the Head of the Department of Humanities and General Education at Universidad Andrés Bello (Chile). E-mail: ruth.espinosa@gmail.com

epistemically dependent on others, this is a day-to-day practice that is generally not a major cause of concern. However, technological advances in generative AI seem to be challenging our understanding of expert testimony, leading to unexpected results, as in the case of a child who, after visiting 17 doctors for chronic pain, only received an accurate diagnosis of his condition after his mother entered her son's medical records into ChatGPT (O'Neill, 2023).

This case shows that the emergence of conversational artificial intelligence has brought new challenges to our understanding of testimony. Can these technologies properly be considered as sources of testimonial knowledge in the full sense of the word? Some authors have proposed the concept of "quasi-testimony" to describe utterances produced by non-human entities that appear to mimic the characteristics of genuine testimony but lack certain fundamental attributes, such as susceptibility to moral and social sanction.

Beyond these theoretical discussions, the practical impact of conversational AIs on our everyday epistemic practices is undeniable. We increasingly rely on these sources of information, often ascribing to them an authority and credibility similar to that of a human expert. This raises questions about the epistemic value of beliefs acquired through these technologies and their relationship to the norms and expectations that underpin our traditional testimonial interactions.

In light of the above, in this paper we question the value of testimonial-based beliefs as opposed to the utterances produced by a generative AI. To this end, we proceed as follows: In Sections (2) and (3) we present and analyze the development of the epistemology of testimony in order to highlight the role of trust and social normativity in the way we understand testimonial exchange; in Sections (4) and (5) we question whether testimonial exchange is possible between persons and non-human agentive entities, arguing that the notion of "quasi-testimony" is theoretically inadequate to refer to statements produced by generative AI, even though in practice they may be assumed as testimonies by the audience; Finally, in the conclusions, we postulate that the value of testimonial-based beliefs vis-à-vis generative AI (and thus the value of testimony as such) depends on being a form of practice that makes sense only within the regulatory framework of human social normativity, which generative AI lacks, since its outputs consist merely in information processing rather than in genuine cognitive achievements that we can call knowledge.

2 TESTIMONY, TRUST AND EVIDENCE

In the epistemology of testimony there have been different positions concerning the justificatory status of the word of others. Reductionists, for example, have denied that testimony has the status of a self-standing rational support for belief. Roughly speaking, the justification of a testimonial-based belief requires additional evidence of a non-testimonial nature, i.e., reducing the justification of that belief to a different source, such as perception, prior experience, inductive evidence, or something similar. Thus, one way to describe this position is as follows:

Reductionism: A is justified in believing p from what S has said if, and only if, (1) A has positive non-testimonial reasons to support that S is reliable (e.g. perception, memory, among others); and (2) A has no additional evidence that the testimony is false or improbable (Hume, 1748; Adler, 2006; Fricker, 2006).

This position, while recognizing the importance of testimony in everyday life, seems to downplay its capacity to provide a rational foundation for belief. Testimony

would mostly be considered a fulcrum for acquiring epistemic justification, meaning that the word of others could not be regarded as sufficient evidence for the asserted proposition.

While a non-reductionist point of view considers testimony as an independent source of justification for belief (Reid, 1983; Coady, 1992; McDowell, 1994; Owens, 2006; Goldberg, 2007), the *assurance view* presents a specific position among those who argue that for accepting testimony it is not necessary additional evidence that *p*, or the reliability of the speaker.

Assurance view: A is justified in believing that *p* from what S has said only by virtue of the interpersonal relationship gained through the testimonial exchange, where A assumes that what was said is true, in part, because of S's assurance, where that assurance is not evidential in nature (Ross, 1986; Hinchman, 2014; Moran, 2018)

This second position is the counter hypothesis to reductionism, suggesting that the word of others would be sufficient to justify believing that *p*, since the mere act of communicating would cause the speaker to present himself, as a witness, as a guarantor of the truth of the utterance. In this way, the audience would not have to resort to any non-testimonial source to believe that *p*.

Although what has been said emphasizes the value of testimony as an epistemic source, it abandons the idea that there is testimonial "evidence", that is, an objective, factual or probabilistic support for the truth of *p*. Given this, one might wonder whether the assurance view is a correct way to conceive testimony, given that it is always possible to fall into deceptions, entanglements, misinterpretations, or other epistemically harmful consequences that creep into testimony in our everyday lives; or whether, on the contrary, we should subscribe to reductionism in order to avoid the harms of credulity in the face of the word of others.

Given this, assurance view is faced with the need to additionally account for the reasons why testimony, under appropriate conditions, would count as sufficient support for the belief that *p*. That is, to provide the criteria for the success of testimonial transmission as an epistemically adequate basis for belief. The solution to this problem would be to find an alternative that accepts the value of testimony as an epistemic source and, in turn, accounts for how testimony conveys evidence.

A tentative proposal for this is the *inheritance view*, according to which through testimony the speaker transmits to the audience the justification conferred by his first source evidence, that is, through testimony one makes the evidence of others one's own. Thus,

Inheritance view: A is justified in believing that *p* from what S has said because the evidence justifying S's belief that *p* was conveyed to A through the testimonial exchange (Owens, 2000; Faulkner 2011; Wright 2019).

An example of this can be found in Paul Faulkner's *Trust theory of testimony* (TTT). According to him, one can trust people by assuming that they are, from a veritist perspective, good instruments for acquiring a true belief. In this way, there is *predictive trust* in that:

PT: A trusts S to φ (in the predictive sense) if and only if (1) A depends on S φ -ing and (2) A expects S to φ (where A expects this in the sense that A predicts that S will φ) (Faulkner, 2001. p. 145).

One might think that this perspective, much like reductionism, would require that one trusts someone to know that *p* in the same way that one trusts a thermometer to

know the temperature: I do so insofar as I have non-testimonial reasons to suppose that the result will be a true proposition. But this characterization of trust is not what we find in the testimonial exchange according to TTT.

According to the latter position, testimony conveys evidence because there is a relationship of *affective trust* in which the audience, epistemically dependent on the speaker, expects the speaker to testify honestly precisely because of this kind of relationship. Thus,

AT: *A trusts S to ϕ* (in the affective sense) if and only if (1) *A depends on S ϕ -ing*; and (2) *A expects (1) to motivate S to ϕ* (where *A expects this* in the sense that *A expects it of S* that *S be moved by the reason to ϕ* given by (1) (Faulkner, 2001. p. 146).

This way of trusting that the dependence of the audience on the speaker is the reason why the latter is honest speaks, at least, of the normative constraints that regulate this kind of exchange. By depending on others, we expect them to be honest with us because the opposite has repercussions in terms of loss of the speaker's reputation through social sanction.¹ In standard cases, this is a consequence that the witness would like to avoid at all costs, since the loss of credibility (being branded as untruthful) is equivalent to the loss of other social goods derived from trust.

To conceive of testimony in this way, according to Faulkner, would make testimony itself potential evidence² of what is said, insofar as:

a. It is true that the audience trusts the speaker and it is true the background of propositions about social normativity that the audience accepts by trusting the speaker.

b. That the audience trusts the speaker does not imply the proposition asserted in the testimonial exchange.

c. Given the audience's trust in the speaker and the background of propositions about social normativity that the audience accepts by trusting the speaker, it is very likely that the proposition asserted in the testimonial exchange is true.

d. Given the audience's trust in the speaker, the background of propositions about social normativity that the audience accepts in trusting the speaker and the proposition asserted in the testimonial exchange, it is highly likely that there is an explanatory connection between the audience's trust in the speaker and the truth of the proposition asserted in the testimonial exchange.

This being the case, it could be argued that believing the testimony of others is a rational act based on the social normativity that governs such an act and that, as such, we acquire evidence by engaging in such an act. This evidence would be inherently different from that which we might obtain from instruments, in that it would relate not only to positive outcomes (true beliefs), but also to how the outcomes are produced by a process that is, in most cases, reliable.

3 TESTIMONY, TRUST AND NORMATIVITY

Science is a paradigmatic case of this type of epistemic structure based on trust that relies on a certain social normativity. One of the first authors to highlight the normative structure on which science relies as a social institution was the sociologist, Robert Merton. He characterized this normativity as a particular "*ethos*". The *ethos* of science, he claims, that affectively toned complex of values and norms which is held to be binding on the man of science" (1942 p. 116). This *ethos* is closely related to the reward system of science, which revolves around the worthiness of its members' contribution to the common and universalistic enterprise of gaining knowledge. The

ethos of science, then, is the system of social normativity that sustains the testimonial exchange inherent in science, although it has been overshadowed by the positivist vision of science.

In the same vein, but from an epistemological perspective, John Hardwig explored the social and trust-based dimension of modern science, emphasizing the characteristics of current scientific production, such as trans disciplinarity, specialization and technologization, which make trust often become "more basic than empirical data" (1991, p. 694). As he notes "belief based on testimony is often epistemically superior to beliefs based on direct, nontestimonial evidence" (1991, p. 698).

Both Merton and Hardwig recognize that testimonial exchange within the scientific community is enabled by a complex social normativity, and a system of social rewards and punishments as the pivot of motivation. Certainly, these features are not replicated in all contexts, but they are present to varying degrees in many human interactions, both spontaneous and institutionalized, and especially in knowledge communities (Goldberg, 2018). This feature responds more than to a psychological fact, to a deep need to reduce complexity in the social system. Trust functions as a device of social construction, in the containment of complexity that would make foresight, knowledge, and well-being impossible. In Luhmann's words, "Where there is trust there are increased possibilities for experience and action, there is an increase in the complexity of the social system and also in the number of possibilities which can be reconciled with its structure". Trust is, thus, an "effective form of complexity reduction." (Luhmann, 1979, p. 8).

In the same sense, we can understand the fact that trust is the condition of possibility of modern science. This, in turn, corresponds to another fact that John Hardwig reports, namely, epistemic dependence. The fact of our epistemic dependence on others raises the inescapable question, which we have already mentioned above, of the extent to which the authority of an expert can constitute a justification for believing or knowing that *p*. In this regard, Hardwig points out that there is a "kind of good reason for believing which does not constitute evidence for the truth of the proposition" (1985, p. 336).

From the sidewalk of the epistemology of trust, it is possible to identify some characteristics that underlie testimonial exchanges. Annette Baier poses one of the fundamental questions that an epistemology of trust must address: "What is the difference between trusting others and merely relying on them?" (Baier, 1986, p. 234). Trust constitutes a relationship of dependence on the "good will" of another, a dependence that makes us "vulnerable" to the limits of that good will. In other words, whoever is the depository of trust enjoys a certain discretionary power, a power that goes hand in hand with a certain responsibility. Thus,

Reasonable trust will require good grounds for such confidence in another's good will, or at least the absence of good grounds for expecting their ill will or indifference. Trust then, on this first approximation, is accepted vulnerability to another's possible but not expected ill will (or lack of good will) toward one (Baier, 1986, p. 235).

Mere dependence, on the other hand, can occur even with respect to objects as well as with people. However, when there is a relationship of trust between people, we concomitantly have good reason to expect that this vulnerability will not be betrayed. Katherin Hawley links this expectation to the existence of a commitment (2012, p. 10) that supports the trustworthiness of the subject, while for Goldberg (2020, p. 4423),

such a commitment can be read in terms of the expectation being properly normative. This is because the epistemic dependence that takes place within knowledge communities is sustained by a set of binding values for informants, so that the expectation is in principle to be regarded as legitimate. This normative structure largely justifies that reliance on the testimony of others is generally reasonable as a source of information, evidence and justification. Thus, the expectation expressed in Baier's quote is rarely fulfilled for individual reasons.

In the context of knowledge communities, trust, commitment, or the legitimate expectation of not being deceived, rests on the normativity sustained in the set of values, ethical, social and political (Goldberg, 2020). In sum, there are good reasons to claim that testimony is an adequate source of justified belief and knowledge, because it is a human institution, subject to a kind of social normativity. Thus, transgressions of epistemic trust involve compromising social, professional, or scientific values and duties.

4 ARE THERE TESTIMONIAL EXCHANGES WITH MACHINES?

What has been said above shows that testimony is a social institution, a matter between humans. However, with the rise of artificial intelligence, the question arises as to whether a generative AI, such as a conversational natural language model, could become part of the same kind of interaction with us and, therefore, report as such, rather than simply being a producer of affirmations as is the case with other less advanced AI-based technologies.³ The distinction is not easy, as we will see below, because several current chatbots (OpenAI's ChatGPT, Google's Bard, Meta's LLaMa, among others) process natural language in such a sophisticated way that they challenge the way we understand concepts such as belief or assertion, which are fundamental to understanding what we mean by "testimony".

We can begin to answer this by looking at the historical development of how these issues related to technology have come to be conceived. Bernard Williams (1973), for example, denied the possibility that machines could have beliefs as such, so that they could only assert (i.e., state that p is true) in a metaphorical sense.⁴ This is because, while they are capable of satisfying some of the properties specific to belief states, they would not be able to satisfy them entirely.

Considered as a psychological state, beliefs have five characteristics:

1. That belief aims at truth.
2. The basic form of expressing belief is assertion.
3. However, assertion is neither a sufficient nor a necessary condition of that state.
4. Those states, in turn, tend to be evidence-based.
5. That state can function as an explanatory notion of the agent's behavior.

Of these conditions, the third one shows us an important difference between us and machines. While a machine would be able to satisfy conditions (1), (2) and (4) by stating that p (i.e., producing an *output* by selecting an alternative based on the available data and displaying it as correct), it cannot produce insincere statements, since this implies the possibility of having intentional states of consciousness. Humans can assert that not p when they believe that p without any cognitive problem; indeed, it is part of the complexity of our rationality to be able to do so (that is, to say something

false and deceive *willingly*), however, this would imply an error in the way that machines process data since "if something interrupts the connection between the normal inner state for asserting that p and its asserting that p, and it comes out with something else, this is merely a case of breakdown" (Williams, 1973, p.145).

This way of conceiving the internal states of the machine ('B-states' instead of beliefs, as Williams prefers to call them) would imply that, if we accept its 'assertions' as a form of testimony, they would necessarily be true if the system is functioning as it should, assuming that it is designed to produce true statements (the machine would do nothing else but express correct outputs). However, we do not say of people that they are malfunctioning if they are insincere or that they function well when they give honest testimony, but we mention that they are rightly or wrongly suited to our way of living in society. In other words, we can rely on machines, but we cannot trust them. If the speedometer in my car fails, I will not be disappointed, no matter how much predictive trust I have placed in the device. If a health professional fails to diagnose me with a common disease, he is reprehensible and ultimately morally and legally liable if there is damage. In this sense, the complexity of doxastic states and human communicative acts would prevent us from subscribing in principle to the idea that a chatbot could testify in the full sense.

An argument different from Williams's can be found in the work of Miller and Freiman. They argue that propositions uttered by machines are indirectly subject to the epistemic normativity inherent in testimonial exchange and that, therefore, such entities can be said to produce quasi-testimonies, i.e., acts that are not a testimonies per se but which, being sufficiently similar to testimony, produce almost the same effect in the audience. Thus,

A linguistic output of an instrument or a machine constitutes a quasi-testimony in a given context of use if and only if the machine or instrument has been designed and constructed to produce this output in a manner that sufficiently resembles testimony phenomenologically, and it is in conformity with an epistemic norm that is parasitic on, or sufficiently similar to what is, or would be, an epistemic norm of testimony in the same context (2020, p. 429).

Quasi-testimony, i.e., utterances produced by non-human entities, are evaluated by the same standards as testimonies. The example given by the authors is quite illustrative: while waiting for a flight, I hear over the loudspeaker when I should board⁵. This illustrates epistemic dependence based on a quasi-reliance on what I hear, which "[...] is distinguishable from mere reliance in that it is grounded in normative expectations from its target, as opposed to expectations that are based merely on inductive inference about its reliability." (2020, p. 430). Thus, I am confident that what is being said is a good reason for taking a particular course of action.

From the audience's perspective, a quasi-testimony and testimony have a practical equivalence. However, the authors raise important qualitative differences to be considered between quasi-testimonies and genuine testimonies. Thus, quasi-testimonies:

1. They are not imaginative and playful (sensitive to humor).
2. Machine assertion is prescribed and restricted to the context.
3. Machines make errors because they are sensitive to *input*.
4. Testimony, unlike quasi-testimony, has moral value in its own right. (Ibid, p. 430-431).

Of these points, the fourth seems to be the most relevant in drawing a distinction

between us and what machines, chatbots and AIs can produce. Many people, for a variety of reasons, may lack sensitivity to humor and context, as well as sensitivity to *input*. It is true, however, that all testimony has moral value insofar as testifying involves being honest about it.⁶ As we saw in the first section, testimony generates trust insofar as the possibility of demanding sanctions is a reason for the speaker to be honest in the dependency relationship. We do not know to what extent a machine, chatbot and AI could be sanctioned according to what has been said in this section, but it does not seem a minor point insofar as the supposed trust generated by a quasi-testimony, if it does not possess the structure of affective trust, is reduced to mere prediction (and, therefore, to the reliability of the device). In this sense, we seem to be returning to Williams' point by noting that a machine's assertion is only an assertion *in quotation marks*.

On this point, Butlin and Viebahn (2023) point out two conditions for an AI to be able to assert (and, for our interests, to testify): on the one hand, it must be able to generate *outputs* with descriptive functions; while, on the other hand, it must be sanctionable. Regarding the second point, they also point out that the sanctionability of assertions requires: "(i) that interlocutors keep track of the credibility of the asserters, (ii) that asserters are sensitive to this, and (iii) that losing credibility is bad for asserters." (2023, p.14). The authors argue that both (i) and (ii) can be satisfied by a generative AI that processes natural language. However, they deny that they can satisfy (iii) because the AI's lack of intentional states prevents it from noticing why sanctioning is undesirable. Losing credibility is not a problem for the IA (it might be for the company's liability), so generating assertions that are conducive to error and receiving a complaint about them is beyond its current capabilities. If this is the case, the terms quasi-testimony and quasi-trust become inadequate to characterize the phenomenon of receiving information from various artificial entities especially for conversational generative AIs, because the similarity they might present with respect to testimonial practices is only superficial.

However, this analysis has only theoretical value, since in everyday epistemic practice, the lay subject seems to respond to these quasi-testimonies in the same way as to an expert testimony, attributing to them high degree of credibility (Guo et al., 2023)⁷. This difficulty (that of relying on statements whose epistemic commitment is null but which are nevertheless functional and even advantageous for our everyday life) makes it obvious to ask, given the nature of these informational devices, what could guarantee their reliability and whether it would be possible for interactions with this type of information source to somehow enter into the normative system of testimonial exchange, either as a parasite of the epistemic norm regarding testimony or on the basis of special criteria for their use. In the following section we will address these questions.

5 THE VALUE OF TESTIMONY-BASED BELIEFS

We know one thing about these sources of information: AIs have no expertise, but only model their language on that of formal scientific communication by the set of data on which they have been trained. Likewise, they have no qualifications or evidence to support their statements, so the reliability of the statements they make depends on the set of data they have been trained with and have at their disposal.

These characteristics, added to the optimization of communication achieved by these language models, give the appearance of expert communication, i.e., trustworthy and guaranteed on the basis of the social normativity that sustains it. It is clear,

however, that these technologies are unable to answer the question of what it is that places AI communication within the realm of legitimate expectations of trust in the assertions of others. In light of this, how should we relate epistemically to these technological advances?

According to Fintan Mallory (2023), one can defend the idea that our interaction with chatbots (and, *mutatis mutandis*, with other artificial entities such as generative AI) takes place within a fictional framework in which we imagine such technology as a sort of character in a game of make-believe. We take it to be a human agent in a context where the utterances it makes are meaningful as long as such a game continues to operate. Thus, the embodiment of the device allows us to mimic the social norms at work in testimonial practices, but only to the extent that one subscribes to a context in which the semantic competence of the generative AI is phenomenally comparable to the semantic competence of a human being.

For example, if I imagine ChatGPT as an astrophysics expert in a game of make-believe, I may well form beliefs about the circumference of the earth, the properties of black holes, or the distance of our moon from the sun. In general, the answers are usually as accurate as an expert's answer could be. However, the mimicry of semantic competence and our assumptions about these technologies do not imply that, being the source informatively accurate (in a veritistic, reliable sense), the beliefs we obtain from them have the same epistemic value as beliefs obtained by a genuine expert, nor that they are meaningful outside the particular context in which they were obtained.

One might argue that the beliefs obtained by imagining a generative AI as a real agent have value outside the context of the make-believe game. This is partly true, in that the information might correspond to facts about the world. But the support for such beliefs is too weak to hold. For example, if someone says to me "Bard told me that *p*", one might ask them about how reliable Bard is as a generative AI. I might get the answer "very reliable", to which I might respond with "how do you know that?". Then I might get the answer "because it has a good database" or something similar. Again, I might ask "and how do you know that is so?" and so on *ad infinitum*. Prototypically, the common user's response would indicate a mere informational dependency without further substantiation. In the case of testimony, this could be similar, but epistemic dependence between people (and not mere 'informational dependence') is maintained by us as members of a community oriented, *all things being well*, to the formation of knowledge.

Unlike the recording we can listen to at the airport, which can almost certainly be traced back to individual or collective human intentionality and responsibility, in the case of generative AI, it is not transparent how it produces its outputs. This is because, given that language models are pre-trained with a large amount of text-based machine learning architecture (such as deep learning) to generate responses to the inputs it receives. However, due to the extremely complex and non-deterministic nature of machine learning, it is not possible to predict exactly what responses the model will generate in any given case. Thus, it is difficult to conceptualize the formal similarity of testimony to such outputs in the way described by Freiman and Miller, since the "subjection to the epistemic norm" of testimony seems to be blocked by its spontaneity. To date, the corporations behind these types of technologies have responded to certain social and moral conventions by blocking or restricting responses on socially sensitive topics. However, given the nature of these models which are designed to resemble human expert testimony, limiting them to the epistemic norm is much about their extinction as anything else.

6 CONCLUSIONS

The above leads us to defend, therefore, the following regarding the value of beliefs acquired by oneself and from others, especially from experts, within the framework of the normativity of knowledge communities: (i) In the search for knowledge, our beliefs result from a process of discriminating between different data. That process can be judged by others as epistemically adequate or inadequate depending on whether we follow the guidelines standardized by a knowledge community for each case. For example, I may be epistemically judged negatively if, in matters that require greater rigor such as public health, I choose to inform myself through *reels* of any RRSS (as happens, for example, with the possible adverse effects of vaccines, where the proliferation of *fake news* has led to cases of parents deciding not to vaccinate their children). Of course, this does not guarantee that people will follow proper procedures in forming beliefs (the above case is indeed an example of this), but being subject to public scrutiny is reason enough for most agents to feel compelled to form beliefs according to such procedures under normal circumstances.

In this sense, a person never forms beliefs outside of a regulatory framework, even if he ignores it, because the rest of the members of the community may object to and question the procedures by which one comes to hold that *p*. In the case of generative AI, we know that it is susceptible to various drawbacks, such as the fact that it has biases in its data (Fraivan and Khasawneh, 2023), or the fact that we ourselves are the ones who bias its capabilities by assuming it is more capable than it really is (Kidd and Birhane, 2023), so that an excessive and generalized reliance on it raises suspicions about the epistemic value of such beliefs, which would not be the case if such technology were approached with the restraint required for its use.

Second, (ii) if in the search for knowledge we are dependent on the knowledge of others, we may also be required to undergo a process of information discrimination with respect to the agent I am asking. For example, I may be required to be cautious about someone who is considered a liar, just as I may be required to believe what is said by someone who is considered an expert in a particular field. This requirement, by the way, is in addition to the expectation that others are subject to the public scrutiny that we ourselves are subject to what we point out in (i), i.e., just as I can be challenged to form beliefs by such mechanisms, any other human being can be judged in the same way. Thus, epistemic dependence would be maintained in that all members of the knowledge community are subject to the same normativity, both in their individual and intersubjective dimensions, and are responsible for what they believe and for what they disseminate.

This is not true of generative AI. Generative AI does not depend on others to know in the same sense that we depend on others to know. AI only learns from data, but there is no kind of relationship between it and that data as there is when we assume a regulative framework that enables testimonial practices. Insofar as AI can produce a quasi-testimony and phenomenally deceive the audience, the value of this utterance is equivalent to the value of any other utterance produced by an instrument (e.g. a tsunami alert, which does not process natural language, can still produce utterances that can be called quasi-testimonies according to the definition given in the fourth section). This is crucial, since an agent ignorant who is unaware of the mechanisms and scope of AI could think about it (literally or analogously) as an expert in a particular field by ignoring that, as we pointed out above, this would only make sense within a fictional frame of reference. Unfortunately, it seems that our ability to discriminate between trustworthy and untrustworthy AI seems to be rather limited (Banovic et al, 2023), thus,

it becomes necessary to emphasize the value of expertise in the processes of belief acquisition and diffusion, as opposed to technologies that appear as capable as a person transmitting knowledge.

As we have pointed out, an important aspect of the value of knowledge, and thus of the beliefs we form, is that it fulfills the expectations that others have of our own cognitive achievements and, at the same time, of their dissemination within the knowledge community. This could be analyzed as isolated communicative instances in which the speaker testifies that p , but if we want to account for the importance of testimony in relation to the imitations we currently possess because of technological progress, we must point out that (iii) the value of our testimonial-based beliefs lies in the fact that they are the product of a human history of communicative exchange. If we consider our reasons for affirming that an expert is preferable to an AI for knowing whether p is true, we come to the reasons already given that such an expert is preferable only to the extent that he or she is validated by his or her peers. This process is also a process of testimonial exchange, in which a person learns from others who have gone through similar processes, which serves as an informational filter on the things that need to be preserved and communicated as a knowledge community. AI, on the other hand, is reduced to accumulating and discriminating information.

The latter should not be taken as a negative assessment of AI. Rather, these reflections aim primarily to illuminate the intrinsic value of our social epistemic practices. AI is a powerful technology that adds a complex tool to our arsenal for understanding the world and advancing humanity, but we need to embrace it with appropriate restraint, given its potential to be misdirected. In any case, the points we have made seem to hold: the value of our beliefs, and therefore the value of testimonial-based beliefs, derives from the demands we make on each other as a community of knowledge. In this sense, quasi-testimonies are parasitic on those demands. But they lack what allows us to distinguish beliefs derived from any kind of instrument from beliefs derived from testimonies. As Freiman (2023) has rightly pointed out, testimonial beliefs differ from non-human testimonial outcomes in that they are a form of human agency, which, as we have noted in this paper, involves subjective submission to norms. In the absence of such norms, the exchange of testimony would be impossible, reducing our interactions to mere predictive trust, as emphasized earlier. However, with humans, it's the process rather than just the outcome that holds significance. Therefore, any superficial similarity between genuine testimony and quasi-testimony is essentially superficial.

REFERENCES

- ADLER, J. *Belief's own ethics*. Cambridge, MA: The MIT Press, 2006.
- BANOVIC, N.; YANG, Z.; RAMESH, A.; LIU, A. Being trustworthy is not enough: how untrustworthy artificial intelligence (AI) can deceive the end-users and gain their trust. *Proceedings of the ACM on Human-Computer Interaction*, v. 7, p. 01-17, 2023.
- BAIER, A. Trust and antitrust. *Ethics*, v. 96, p. 231-260, 1986.
- BUTLIN, P.; VIEBAHN, E. AI assertion. *Unpublished manuscript*, 21 pages, 2023. <<https://doi.org/10.31219/osf.io/pfjzu>>
- COADY, C. *Testimony: a philosophical study*. Oxford: Oxford University Press, 1992.
- FAULKNER, P. *Knowledge on trust*. Oxford: Oxford University Press, 2011.
- FAULKNER, P. Testimony and trust. In: SIMON, J. (Ed.). *The Routledge handbook of trust and philosophy*. New York: Routledge, 2020. p. 329-340.

- FRAIWAN, M; KHASAWNEH, N. A review of ChatGPT applications in education, marketing, software engineering, and healthcare: benefits, drawbacks, and research directions. *Unpublished manuscript*, 22 pages, 2023. Available at <<https://doi.org/10.48550/arXiv.2305.00237>>
- FREIMAN, O. Analysis of beliefs acquired from a conversational AI: instruments-based beliefs, testimony-based beliefs, and technology-based beliefs. *Episteme*, p. 1-17, ahead of print, 2023.
- FRICKER, E. Testimony and epistemic autonomy. In: LACKEY, J.; SOSA, E. (Ed.) *The epistemology of testimony*. Oxford: Oxford University Press, 2006. p. 225-252.
- GOLDBERG, S. *Anti-individualism: mind and language, knowledge and justification*. Cambridge: Cambridge University Press, 2007.
- GOLDBERG, S. What we owe each other, epistemologically speaking: ethico-political values in social epistemology. *Synthese*, v. 197, p. 4407-4423, 2020.
- GRIMALTOS, T.; ROSELL, S. *Mentiras y engaños: una investigación filosófica*. Madrid: Cátedra, 2021.
- GUO, B.; ZHANG, X.; WANG, Z.; JIANG, M.; NIE, J.; DING, Y.; YUE, J.; WU, Y. How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection. *Unpublished manuscript*, 20 pages, 2023. Available at<<https://doi.org/10.48550/arXiv.2301.07597>>
- HARDWIG, J. Epistemic dependence. *The Journal of Philosophy*, v. 82, p. 335-349, 1985.
- HARDWIG, J. The role of trust in knowledge. *Journal of Philosophy*, v. 88, p. 693-708, 1991.
- HAWLEY, K. *Trust: a very short introduction*. Oxford: Oxford University Press, 2012.
- HINCHMAN, E. Telling as inviting to trust. *Philosophy and Phenomenological Research*, v. 70, p. 562-587, 2005.
- HOLLIS, M. *Trust within reason*. Cambridge: Cambridge University Press, 1998.
- HUME, D. *An enquiry concerning human understanding*. Cambridge: Cambridge University Press, 1748.
- KIDD, C.; BIRHANE, A. How AI can distort human beliefs. *Science*, v. 380, p. 1222-1223, 2023.
- LEHRER, K. Knowledge and the trustworthiness of instruments. *Monist*, v. 78, p. 156-170, 1995.
- LUHMANN, N. *Trust and power*. New York: Wiley, 1979.
- McDOWELL, J. *Knowledge by hearsay: meaning, knowledge and reality*. Cambridge: Harvard University Press, 1994.
- MALLORY, F. Fictionalism about chatbots. *Ergo*, v. 10, p. 1082-1100, 2023.
- MERTON, R. Science and technology in a democratic order. *Journal of Legal and Political Sociology*, v. 1, p. 115-126, 1942.
- MORAN, R. Getting told and being believed. In: LACKEY, J.; SOSA, E. (Ed.) *The epistemology of testimony*. Oxford: Oxford University Press, 2018. p. 272-306.
- MILLER, B.; FREIMAN, O. Trust and distributed epistemic labor. In: SIMON, J. (Ed.) *The Routledge handbook on trust and philosophy*. New York: Routledge, 2020. p. 341-353.
- O'NEILL, M. ChatGPT diagnoses cause of child's chronic pain after 17 doctors failed. *The Independent*, 13 September 2023.<<https://www.independent.co.uk/news/health/chatgpt-diagnosis-spina-bifida-mother-son-b2410361.html>>.
- OWENS, D. *Reason without freedom: the problem of epistemic normativity*. New York: Routledge, 2000.
- OWENS, D. *Testimony and assertion: normativity and control*. Oxford: Oxford University Press, 2006.
- REID, T. Inquiry into the human mind on the principles of common sense. In: BEANBLOSSOM, R.; LEHRER, K. (Ed.) *Thomas Reid's Inquiry and essays*. Indianapolis: Hackett, 1983. p. 01-126.

ROSS, A. Why do we believe what we are told? *Ratio*, v. 28, p. 69-88, 1986.

WILLIAMS, B. Deciding to believe. In: WILLIAMS, B. *Problems of the self: philosophical papers 1956-1972*. Cambridge: Cambridge University Press, 1973. p. 136-151.

WRIGHT, S. *Knowledge transmission*. New York: Routledge, 2019.

NOTAS

- 1 This expectation relates only to what the agent should do, not what the audience believes the agent will do. A similar concept, which Faulkner himself discusses, is Hollis' concept of normative trust (1998 p. 10), which he complements by emphasizing the way in which we are challenged by these transgressions. Another way of looking at the same point would be the distinction between the doxastic view of trust and the non-doxastic view of trust (Faulkner, 2020), the latter being the one that describes this phenomenon.
- 2 2 Faulkner starts with Achinstein's definition of evidence, noting that: "e is potential evidence for h, given b, if, and only if, (a) e and b are true, (b) e does not entail h, (c) $\text{prob}(h/e\&b) > k$, (d) $\text{prob}(\text{there is an explanatory connection between h and } e/h\&e\&b) > k$." (Faulkner, 2011. p. 155. Modified from Achinstein, 1978. p. 162)
- 3 Someone might argue that our interaction with a generative AI, such as ChatGPT, is similar to other instruments (such as GPS or thermometers). For instance, the *Evaluation model for instrument* proposed by Keith Lehrer (1995) suggests that our trust in these instruments depends on our evaluation of their trustworthiness, which is based on the results they produce. However, this does not apply to generative AI models. As Ori Freiman (2023) argues, the type of belief we form through interaction with generative AI is different from that formed by mere instruments because the former is produced by a non-human agentive entity that processes natural language. This distinction implies that the beliefs produced by our interaction with AI models are more akin to human testimony than the predictable information produced by an instrument.
- 4 In the sense of producing an *output* in an assertive mode, but without implying anything more, which would be contrary to a 'true' assertion in that it would imply believing *p* to be true by stating it.
- 5 This is said from the context of the recordings. But it could also apply to what is said by a generative AI that, given the airport data, generates new outputs each time it makes the boarding call to passengers. This may well be the case in the future.
- 6 One could defend the idea that one can give false testimony, but this is only a reach of words. The phenomenon of testifying involves wanting the audience to believe that *p* by asserting that *p*, while one believes that *p* is the case; and any other possible combination of these or other elements give rise to lying and/or deception. For a detailed taxonomy of the latter, see Grimaltos and Rosell (2021, p. 42).
- 7 In a recent study by Guo, et al (2023), the helpfulness of responses from ChatGPT and human experts was compared across various domains. Surprisingly, ChatGPT's answers were found to be more helpful than human responses in over half of the questions, particularly in finance and psychology. The results of this study are concerning because they highlight a potential threat to the quality and reliability of information. With ChatGPT often outperforming human experts in terms of perceived helpfulness, there's a legitimate concern that people may increasingly prefer AI-generated responses to those from trained experts. This

shift could have significant implications for fields where human expertise is critical, such as medicine, raising questions about the reliability and appropriateness of relying on AI for critical decisions.