

RESEARCH ARTICLE

Examining epistemological challenges of large language models in law

Ludovica Paseri  and Massimo Durante

Law Department, University of Turin, Torino, Italy

Corresponding author: Ludovica Paseri; Email: ludovica.paseri@unito.it

Ludovica Paseri wrote sections 2, 3.1, 3.2, 3.3, 3.4, 4; Massimo Durante wrote sections 1, 3.5, 4.

(Received 01 August 2024; revised 30 October 2024; accepted 06 November 2024)

Abstract

Large Language Models (LLMs) raises challenges that can be examined according to a *normative* and an *epistemological* approach. The normative approach, increasingly adopted by European institutions, identifies the pros and cons of technological advancement. Regarding LLMs, the main pros concern technological innovation, economic development and the achievement of social goals and values. The disadvantages mainly concern cases of risks and harms generated by means of LLMs. The epistemological approach examines how LLMs produce outputs, information, knowledge, and a representation of reality in ways that differ from those followed by human beings. To face the impact of LLMs, our paper contends that the epistemological approach should be examined as a priority: identifying risks and opportunities of LLMs also depends on considering how this form of artificial intelligence works from an epistemological point of view. To this end, our analysis compares the epistemology of LLMs with that of law, in order to highlight at least five issues in terms of: (i) *qualification*; (ii) *reliability*; (iii) *pluralism and novelty*; (iv) *technological dependence* and (v) *relation to truth and accuracy*. The epistemological analysis of these issues, preliminary to the normative one, lays the foundations to better frame challenges and opportunities arising from the use of LLMs.

Keywords: LLMs; artificial intelligence; epistemological approach; ethics of AI

1. Introduction

A simple search on *ssrn.com* on the topic large language models (LLMs) returns more than 400 results only for 2024. Many papers focus on the technical aspects of LLMs, to highlight glitches and potentials of these artificial intelligence (AI) models that have had such a strong recent development: these contributions put the focus on innovation aspects (Diab, 2024; Moreau et al., 2023). Other papers analyze the risks, challenges or benefits that the use of such models may have in different areas of application or on our world at large: these papers put emphasis on the impact of these models on contemporary society (Ferrara, 2024; Novelli et al., 2024; Pagallo, 2022; Wachter et al., 2024). Still other papers focus on the ability of LLMs to reshape traditional activities, jobs or professional fields: these contributions highlight the transformative power of these technologies (Fagan, forthcoming, 2025; Nelson, 2024; Surden, 2023). Finally, some papers focus on the ability of such models to change our perception and understanding of reality, creating new pieces of knowledge: these contributions examine the epistemological impact of LLMs (Delacroix, 2024; Krook, 2024).

This wide range of topics indirectly shows that LLMs potentially have a strong impact on every aspect of reality, with innovative and transformative effects raising risks, challenges and opportunities that require to be increasingly addressed not only from a broad normative angle (ethical, legal, political, economic, social, etc.) but also from a deep epistemological perspective, as they change the way we grasp and know our own reality. This raises a complex and thorny governance problem, since law has traditionally been used to regulate *behaviors* and legally relevant *consequences* arising from behaviors (whether the product of human actions or, increasingly, also of actions enacted by artificial agents) but not *models* (and particularly AI models) that affect the perception and understanding of reality or the production of knowledge.

Consider, for example, the difficulties faced by the European or American lawmakers in legally defining and regulating such new AI models. The European regulation on artificial intelligence (AI Act, hereinafter)¹ defines AI system in terms of

a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments (Art. 3(1) AI Act).

However, the considerable developments in the field of AI have prompted the European lawmakers to include in Art. 3(63) AI Act a further definition of the so-called “general-purpose AI models,” defined as follows:

an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market.

US have taken a distinct approach in the Executive Order No. 14110 on the safe, secure, and trustworthy development and use of AI, signed by the US President Joe Biden on 30 October 2023,² proposing a definition of the term AI,³ another of the AI *model*,⁴ and a third one about the AI *system*.⁵ Furthermore, the §3(p) of the Executive Order also includes a specific mention of generative AI, defined as follows:

¹Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>.

²The White House, *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*, no. 14110, 30 October 2023, available at <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

³AI is defined as “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action,” in §3(b) Executive Ord.

⁴The AI model is defined as “a component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs,” in §3(c) Executive Ord.

⁵The term AI system is defined as “any data system, software, hardware, application, tool, or utility that operates in whole or in part using AI,” in §3(e) Executive Ord.

the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.

The approaches taken by the US and the EU are different. In the former case, generative AI (Gen AI) is explicitly mentioned, whereby LLMs are the best known and widely used example. In the latter case, the EU legislator ultimately decided not to make explicit reference to Gen AI but instead to include the general-purpose AI models, an example of “Large generative AI models,” as specified in recital 99 of the AI Act.⁶ A specific definition of LLMs has not been attempted, although the consequences arising from their ability to automatically generate “content in response to prompts written in natural-language conversational interfaces” (UNESCO, 2023, 8), investigated in this study, fall within the scope of the AI Act. This issue is often addressed through a normative approach, which aims to identify the advantages and disadvantages of applying a specific technology.

However, the normative challenges of LLMs require us to adopt an epistemological approach, which calls for examining the main differences between the representation of reality of human beings and that of LLMs. This epistemological dimension underlies the way LLMs operate in practice and poses challenges to regulation: it is therefore crucial to preliminarily address the problem from an epistemological point of view, as observed and extensively expounded by (Wachter et al., 2024) in a systematic study on the subject.

Therefore, the paper focuses on the epistemological approach behind LLMs (section 2), wondering what representation of reality LLMs produces and how we interact with this representation. Five epistemological challenges of LLMs are identified in terms of qualification (section 3.1); reliability (section 3.2); pluralism and novelty (section 3.3); technological dependence (section 3.4) and relation to truth and accuracy (section 3.5). In light of these epistemological challenges, the social implications of LLMs are framed, proposing some final remarks (section 4).

2. Epistemological approach

LLMs produce outputs in forms of answers to prompts, which can be understood as pieces of knowledge (Kelemen & Hvorecký, 2008; Lenat & Feigenbaum, 1991; Lenat & Marcus, 2023), offering a specific representation of reality, because of “the human tendency to attribute meaning and intent to natural language” (Wachter et al., 2024, 3). From a qualitative point of view, the LLM’s representation of reality may *accidentally* reach high standard and, as a result, discerning between synthetic and non-synthetic outputs becomes challenging for humans. Accordingly, apart from normative considerations, it becomes essential to scrutinize the representation of reality offered by LLMs and how human beings become subjects of interaction, since people increasingly rely on AI-generated outputs, which are produced by AI models, trained on ever larger dataset and endowed with ever-increasing computational power sparking significant consequences on human life (Durante, 2021, 13).

From an epistemological perspective, LLMs and humans are indeed different. While human beings are extraordinary “semantic engines” (their understanding and representation of reality is rooted in semantic skills, i.e., the ability to give meaning and sense to the world), a LLM is an extraordinary “syntactic engine” (Floridi, 2023a, 44), which predicts the most likely string of words that comes next in a piece of text in response to a prompt, operating on the basis of the ability to process huge amounts of data and parameters through structures of great complexity, such as neural networks, without possessing knowledge or awareness of what is being processed. Unlike the scientific method, where knowledge entails verifiable predictions based on models and explanations, AI-generated outputs

⁶Large generative AI models are a typical example for a general-purpose AI model, given that they allow for flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks.; recital 99, AI Act.

lack such theoretical grounding. Thus, questioning the underlying functioning of LLMs is pivotal due to its profound impact on human actions and the criteria for justifying them.

In doing so, the analysis is limited to considering the comparison with a specific epistemology, namely the legal epistemology. Without going into the age-old and thorny question of the definition of law and its role, we can say that, from an epistemological standpoint, law acts as a normative system, which interprets and applies legal statements within a shared legal framework, verifying the reliability of this interpretation and application through forms of reasoning and specific tools (principles, categories, evidences, review systems, etc.) that the law itself provides and regulates, which have a fundamental common feature. Every piece of knowledge produced by the interpretation and application of the law is offered to the scrutiny, control and discussion of a counterparty and a third party (and more widely to the community of legal scholars over time).⁷ Legal science is progressive and incremental and requires the contribution of all legal experts. This method, and the resulting representation of reality that it conveys, may differ from or be in contrast with LLM's representation of reality, for three main reasons.

First, the representation produced by LLMs may happen to be accurate (where there is a validated standard or metric to measure accuracy) but it is not based on the ability to attribute meaning and intersubjectively verify its epistemic reliability. In analytical terms, LLMs (and more widely Gen AI) do not have the ability to operate a semantic ascent (Floridi, 2023a; Quine, 1953) or descent in relation to artificially produced sentences; namely, they are not able to move from the linguistic to the metalinguistic level in order to qualify the content of their statements or argue in favor of its accuracy (as remarked with regards to general-purpose LLMs, “despite their generality, responses rarely include linguistic signals or measure of confidence” [Wachter et al., 2024, 2, also referring to Lin et al., 2022; Mielke et al., 2022]).

Second, the law claims to define the extent to which an explanation is legally relevant both in terms of causal regression (i.e., determination of the chain of events of a legally relevant fact) and in terms of prediction (of the legally relevant consequences of the fact in question). Against this epistemological background, the law has always independently decided where it intends to stop the legally relevant retrospective or predictive explanation of an event. LLMs cannot account for a causal retrospective or predictive explanation of any knowledge it produces. How could the law, from this perspective, regulate an event if it cannot appraise and determine where (the knowledge of) its origin and consequences begin and end? (Durante, 2021, 9–10).

Third, each piece of knowledge created by LLMs does not require the existence of a specialized *community of interpreters* to be concretely applied, while law is a social product and cannot give rise to progressive and incremental knowledge except through the contribution, discussion, interpretation and scrutiny by a community of experts (Zeno-Zencovich, 2023, 445). While legal epistemology is focused on a shared process of representation and validation, LLMs propose solely the product of output, leading to concerns regarding both reproducibility and control. In particular, the aspect of control is central to the extent that “whoever controls the questions controls the answers and whoever controls the answers control reality” (Floridi, 2023b, 5).

Consider, for instance, the ability of ChatGPT to reiterate a given answer, potentially offering a very different response, sometimes totally contradicting the former. Variable outputs stress that LLM's representation of reality needs mechanisms ensuring accountability. Furthermore, LLM's inability to attribute meaning or validate the reliability of its outputs raises crucial questions about its role in producing pieces of knowledge on which decision-making is grounded. Nevertheless, this should not

⁷According to a neopositivist conception of legal science that has also been called logical positivism, “[...] the paradigmatic instance of knowledge is represented by the empirical sciences, which adopt the principle of verification (or, from a certain point on, falsification). The latter thus becomes the fundamental methodological principle of every sphere of knowledge. Just as the proposition ‘this is chalk’ implies that if one observes a piece of chalk at the microscope certain structural qualities will become apparent, similarly the proposition ‘Section 62 of the Uniform Negotiable Instruments Act is valid Illinois law’ implies that Illinois courts, given certain conditions, will behave in a certain way” (Schiavello, 2020, 148 [our translation]).

be a limitation for the study of potential opportunities arising from the use of LLMs in legal settings (as discussed in Fagan, 2025; Nelson, 2024; Surden, 2023).⁸

Investigating the epistemological dimensions of LLMs involves broader philosophical inquiries regarding knowledge acquisition, representation and validation. Understanding how to balance LLM's production of outputs and representation of reality with human interpretation becomes crucial. What safeguards can be implemented to mitigate the issues emerging with LLMs from an epistemological standpoint? These issues require a multifaceted approach that recognizes the epistemic abilities and limits of LLMs. By critically examining some assumptions and implications of the production of outputs and the representation of reality of LLMs, the study paves the way for a more informed and coherent approach to its development and deployment. Therefore, a set of epistemological challenges have been identified and will be discussed below.

3. Epistemological challenges

AI has been described in terms of “epistemic technology,” where the term epistemic “is meant to refer to the broader category of practices that relate in a direct manner to the acquisition, retention, use and creation of *knowledge*” (Alvarado, 2023, 12). The epistemic nature of AI derives both from the type of content it manipulates and the types of operations it performs on that content, but also for “its own internal capacities” (Alvarado, 2023, 15). However, being an epistemic technology does not mean that the content or representation of reality provided is *reliable*. In other words, “accepting AI as an epistemic technology does not mean that we must also accept that it is an epistemically trustworthy epistemic technology” (Alvarado, 2023, 14). For these reasons, “the need to involve challenging epistemic and methodological scrutiny is even more immediately pressing” (Alvarado, 2023, 22). In doing so, five major epistemological challenges need to be analyzed: (1) qualification; (2) reliability; (3) pluralism and novelty; (4) technological dependence and (5) relation to truth and accuracy.

3.1 Epistemic qualification

The qualification challenge consists in the fact that LLMs defy the human ability to qualify the epistemological status of outputs (i.e., the answers and pieces of knowledge) conveyed. This occurs primarily in two ways.

First, LLMs tend to elide the distinction between what is created (i.e., which is the outcome of a proper creative process) and what is merely reproduced and mirrored (i.e., which is the result of the ability to mimic a creative activity). If one considers only the result and not the process, human beings are often unable to distinguish, for instance, the synthetic from the non-synthetic result. In the legal context, it remains crucial not only to be able to evaluate a piece of knowledge but also to be able to control the process by which a piece of knowledge is produced (Tasioulas, 2023, 9).⁹

Second, LLM challenges the human ability to qualify between true and false. Considering the statistical dimension of LLM's outcomes, pieces of false knowledge (the so-called hallucination, Ji et al., 2023; Rebuffe et al., 2022) are frequently generated however plausible: “Readers and consumers

⁸See, for instance, Frank Fagan, who envisages that “legal tasks will take less time to complete, and language models will enhance lawyer productivity” and significantly hypothesizes that “any advantage from using LLMs in law depends upon the cost of data and its processing. If costs are structured so that capital investment makes superior performance possible, then firms will be able to differentiate themselves on the basis of their investments in generative A.I.” (Fagan, 2025, respectively 1 and 9).

⁹On the one hand, the use of synthetic data may reduce costs in the legal field: “Synthetic data promises to substantially lower data processing costs. Human-created data must be cleaned, labelled, and organized prior to its use. By creating ideal data, builders of large language models can train and fine-tune their models more easily and with less strain on computational resources;” on the other hand, “there is a cost to using synthetic data. If the artificial data fails to adequately represent the real world, then the LLM will produce errors. Input that is substantially inaccurate can lead to low-quality output” (Fagan, 2025, 11).

of texts will have to get used to not knowing whether the source is artificial or human. Probably they will not notice, or even mind” (Floridi & Chiriatti, 2020, 691). For example, LLMs provide answers, which they are unable to reflexively qualify as to their degree of truthfulness or reliability. Where questioned in this regard, human beings can reflexively evaluate a statement of their own, qualifying it as certain, probable, plausible, etc. (using expressions, which precede the statement itself, such as “I am certain that,” “I believe that,” “it is highly probable that,” etc.). Any answer given by an LLM should, in this sense, always be preceded by an expression such as: “It is only probable that ...,” in order to reveal that any AI-generated statement cannot constitutively raise any claim to truth (UNESCO, 2023).¹⁰

It is a crucial matter, since, as argued by Stefano Rodotà, long before the emergence of LLMs, “the meaning of truth in democratic societies [...] presents itself as the result of an open process of knowledge” (Rodotà, 2012, 225 [our translation]), remarking that “In a society omnivorous of information, and continually productive of representations, the ‘truth’ of the latter takes on a special significance” (Rodotà, 2012, 221 [our translation]). This is even more relevant considering the so-called “feedback loop:” “[...] Large Language Models learn from content that is fed to them, but because content will be increasingly wholly or partially AI-generated, it will be partly learning from itself, thus creating a feedback loop” (van der Sloot, 2024, 63). This feedback loop leads to the crystallization of LLMs and Gen AI’s representation of reality, with serious risks where it produces distorted or unreliable knowledge.

3.2 Epistemic reliability

The challenge of epistemic reliability represents the problem of human ability to rely on the answers and representation of reality conveyed by LLMs, which reiterates patterns identified starting from its training dataset, without having access to real-world understanding and knowledge. As has been pointed out more generally with regards to AI-generated output, this

can lead teachers and students to place a level of trust in the output that it does warrant. [...] Indeed, Gen AI is not informed by observations of the real world or other key aspects of the scientific method, nor it is aligned with human or social values. For these reasons, it cannot generate genuinely novel content about the real world, objects and their relations, people and social relations, human-object relations, or human-tech relations. Whether the apparently novel content generated by Gen AI models can be recognized as scientific knowledge is contested (UNESCO, 2023, 16).

In addition, it is worth emphasizing that with regard to the debate on scientific and objective knowledge, the results of LLMs may not even be consistent with the main characteristics and principles of the open science policies and the FAIR (i.e., Findable, Accessible, Interoperable and Reusable) data requirements (Paseri, 2023, 2022). This can raise three main consequences.

First, this produces an impact not only on epistemic trust but, precisely because of the inaccuracy, opacity and inexplicability of its model, also on social acceptance. In this perspective, it is indeed crucial to note that: “Assessing the likelihood or the accuracy of such outputs will depend on how well we think the model analyzed the training data. Hence, epistemic reliability of the analytical processes will be relevant [...] to the trust we can allocate to AI generated text, search, etc.” (Alvarado, 2023, 22). Hence, it follows that this epistemic limitation “is also a key cause of trust issues around Gen AI

¹⁰“Gen AI is trained using data collected from webpages, social media conversations and other online media. It generates its content by statistically analyzing the distribution of words, pixels or other elements in the data that it has ingested and identifying and repeating common patterns,” (UNESCO, 2023, 7). For instance, even if we were to ask the LLM to qualify the status of his own response in degree of truthfulness or epistemic reliability, it would not be able to do so, and its further response would be merely probabilistic and vitiated by the same original lack of self-qualification.

(Nazaretsky et al., 2022a). If users don't understand how a Gen AI system arrived at a specific output, they are less likely to be willing to adopt or use it (Nazaretsky et al., 2022b)" (UNESCO, 2023, 15).

Second, this has serious consequences for the allocation of responsibility. In an ever-increasing situation of deep intermingling between human and artificial collaboration in content generation, "New mechanisms for the allocation of responsibility for the production of semantic artefacts will probably be needed" (Floridi & Chiriatti, 2020, 692). Consider, for instance, in this regard the growing debate in the field of scientific research, which is not only about "reproducibility and robustness in their results and conclusions" (European Commission, 2024, 6) but also about the very authorship of the content produced by LLMs.

Third, this has a considerable environmental impact on the infosphere (Floridi, 2014, 205–216). As in the case of an upstream polluted source, flawed AI-generated content can not only spread polluted content downstream, which is inextricably mixed with trustworthy and genuine content, but in the long run it also addicts end users, lowering their defenses.¹¹

3.3 Epistemic pluralism and novelty

Every authentic cognitive inquiry rests on the ability to question. There is nothing more difficult in the search for knowledge than to construct the relevant and pertinent questions, from which any real investigation can start. However, for an investigation to be fruitful and satisfy our need for research, it is necessary that the answer be obtained within the boundaries set by the question, but it is equally necessary that, within those boundaries, the answer brings something new, which is not already included in the prompt of the question; otherwise, the inquiry remains tautological. Moreover, any authentic answer, which gives us access to the knowledge of something new, raises new questions, that is, the possibility of questioning reality from different standpoints, ensuring pluralism of research and knowledge: "Each piece of semantic information is a response to a question, which, as a whole, raises further questions about itself, which require the correct flow of information to receive adequate response, through an appropriate network of relationships with some information source" (Floridi, 2013, 274).

What puzzles us about LLMs is that the ability to produce pieces of knowledge by predicting probabilities of language patterns found in data is based on a static training dataset that describes the world as it is, "assuming what we know about the observed world" (Pearl & Mackenzie, 2018, 9; Vallor, 2024). This is likely to reaffirm the primacy of the existent over openness to novelty and pluralism in a more or less surreptitious manner. In this way, the knowledge and representation of reality produced by LLMs and Gen AI at large is not instrumental in imagining new and different worlds but rather in reaffirming worlds already entrenched in traditional views and conventions. As properly remarked, "Gen AI, by definition, reproduces dominant worldviews in its outputs and undermines minority and plural opinions" (UNESCO, 2023, 26). This brings about two main consequences, according to UNESCO's view of generative AI, which are relevant to epistemology and democracy, respectively. First, "Accordingly, if human civilizations are to flourish, it is essential that we recognize that Gen AI *can never be an authoritative source of knowledge* on whatever topic it engages with" (UNESCO, 2023, 26). Second,

Data-poor populations [...] have minimal or limited digital presence online. Their voices are consequently not being heard and their concerns are not represented in the data being used to train GPTs, and so rarely appear in the outputs. For these reasons, given the pre-training

¹¹"This poses a high risk for young learners who do not have solid prior knowledge of the topic in question. It also poses a recursive risk for future GPT models that will be trained on text scraped from the Internet that GPT models have themselves created which also include their biases and errors" (UNESCO, 2023, 16), as we have already pointed out.

methodology based on data from internet web pages and social media conversations, GPT models can further marginalize already disadvantage people (UNESCO, 2023, 17).¹²

This last remark raises – or rather reiterates – a crucial concern for democracy. No matter how democracy is defined, it is generally agreed that people and populations should be able to modify their starting conditions. Like other technologies, LLMs can increase or emphasize starting conditions, favoring those who are already advantaged and disfavoring those who are already at a disadvantage. The risk is that LLMs become the battlefield where the challenges of innovation may be won by a few (also in the field of legal profession¹³), and where conformity, division and social exclusion may also develop and spread, jeopardizing equity and pluralism in democracy.

3.4 Technological dependence

Technological dependence resulting from the increased or systematic use of LLMs, which are entrusted with tasks of producing knowledge, or even just aiding or supporting human activities, raises at least three different types of issues from an epistemological point of view, which add to the already stressed reliability problem.

First, technological dependence stresses the importance of measuring and comparing the outputs produced by human activity or AI, raising the need of standards for assessing the outputs of LLMs. On the one hand, recourse to such systems should not be discouraged in areas where it has proved better or safer than human activities; on the other hand, it would be desirable and fair to expect a higher standard from LLMs (Durante & Floridi, 2022, 93–112) where they are proved to be able to achieve it (European Commission, 2019, 25). In more general terms, comparing human- or AI-generated results will raise tensions that may result in third-party protected expectations (e.g., should the lawyers inform the client that their legal analysis has been generated or supported by an AI system?).

Second, it is to reaffirm what has been qualified as the “principle of functional equivalence” as closely related to the “principle of equal protection under the law” (Durante & Floridi, 2022, 112). The first principle asserts that operation or assistance by an AI system (in this case, LLMs) should be treated no differently from human operation or assistance in case of harm to a third party, for equal protection under the law to be ensured to all those who are the receivers of the effects or the performance of these systems. Similar cases must be dealt with in similar ways unless there is a reason to make a distinction. In this regard, a remark by the *Expert Group on Product Liability in the field of AI and Emerging Technologies* is relevant from an epistemological standpoint, although stated in a partially different context: “[It is] challenging to identify the benchmark against which the operation of non-human helpers will be assessed in order to mirror the misconduct element of human auxiliaries” (European Commission, 2019, 25).

Third, technological dependence raises an additional epistemological challenge posed by the need for humans to increasingly rely on technology itself to distinguish synthetic from non-synthetic outputs and to detect the use of LLMs in producing pieces of knowledge. Consider, for example, the tools developed in academia to understand whether a text is generated by an artificial or human agent, which raises relevant issues in terms of liability, plagiarism and copyright. Accordingly, it is important to mention the publication of guidelines for the responsible use of LLMs and Gen AI in

¹²The risks associated with the under-representation of population segments in the datasets underlying the operation of LLMs are widely investigated, see, for instance, Park et al., (2021); (Fosch-Villaronga & Poulsen, 2022).

¹³“[...] as language model processing costs increase, the greater will be the separation between excellent and mediocre firms to the extent that processing costs are high, but efficiently scale. [...] The language model will provide them with substantial productivity gains. Investing in a powerful LLM may not be worth it when it is used to service just a few clients, but if, for instance, repeat patterns of computational analysis engender a more efficient processing cost structure, then some consolidation is more likely” (Fagan, forthcoming 2025, 9).

research, according to which researchers “are accountable for the integrity of the content generated by or with the support of AI tools” as well as it is expected that they “do not use fabricated material created by generative AI in the scientific process, for example falsifying, altering, manipulating original research data” (European Commission, 2024, 6). This becomes even more challenging when dealing with so-called hybrid data, i.e., “the offspring of historical and synthetic data” (Floridi, 2023a, 36).

All these cases highlight the crucial importance of standards as they constitute key benchmarks for evaluating the reliability and accuracy of AI-generated outputs. Furthermore, the identification of these standards implies a complex assessment (from a legal, ethical and social standpoint), in order to guarantee adequate levels of transparency and responsibility when using LLMs and Gen AI systems, taking also into account the rapid pace of technological advancement and the associated need to periodically update measurement criteria.

3.5 A legal duty to tell the truth?

A similar set of epistemological concerns has motivated some scholars to advance the hypothesis that there may be a legal duty on LLMs providers to create models that “tell the truth” (Wachter et al., 2024). This study represents, at the state of the art, the most systematic examination of what has been defined as the risk of “epistemic harm” (Wachter et al., 2024, 5), meaning the set of dangerous and harmful consequences that can result from relying on epistemologically flawed AI-generated content.

This study, building on the assumption that a risk of epistemic harm depends not only on the inherent limitations of LLMs (which “are not designed to tell the truth in any overriding sense” [Wachter et al., 2024, 2]) but also on the extrinsic epistemic limits of users (who “are both encouraged and innately susceptible to believing LLMs are telling the truth” [Wachter et al., 2024, 3]), highlights a key point: namely, that such risk is not specific or limited to a given area but systematic in nature, since it gives rise to “careless speech” which is bound to cause “unique long-term harms to science, education, and society which resist easy quantification, measurement, and mitigation” (Wachter et al., 2024, 5).

The study proposes a detailed analysis of truth-related obligations in EU human rights law, AI Act, Digital Services Act, Product Liability Directive, and AI Liability Directive. These legal frameworks contain limited, sector-specific truth duties, from which, according to the authors, it is hard to derive a general duty to tell the truth:

Where these duties exist they tend to be limited to specific sectors, professions, or state institutions, and rarely apply to private sector. The harms of careless speech are stubbornly difficult to regulate because they are intangible, long-term and cumulative. [...] Current frameworks are designed to regulate specific types of platforms or people (e.g., professionals) but not to regulate a hybrid of the two. Future regulatory instruments need to explicitly target this middle ground between platforms and people (Wachter et al., 2024, 47–48).

However, the authors trust that: “Despite this general regulatory landscape, the limited duties found in science and academia, education, and libraries and archives offer an interesting avenue to explore as LLMs serve a similar function” (Wachter et al., 2024, 48). On this basis, they believe that a *general* duty to tell the truth can be assumed from more limited, sector-specific duties. In fact, they contend that the “scope of this duty must be broad; a narrow duty would not capture the intangible, longitudinal harms of careless speech, and would not reflect the general-purpose language capacities of LLMs” (Wachter et al., 2024, 48).

As already mentioned, (Wachter et al., 2024) is a systematic and in-depth study that would deserve extensive analysis and detailed review, particularly with regard to the fields of science, academia,

education, archives and libraries¹⁴ (which serve as a point of reference for founding a general duty to tell the truth). Here, however, the aim is limited to analyzing a few problematic facets of a general duty to tell the truth, related to what is relevant to this study, regarding the epistemological investigation of LLMs.

It may be doubted that LLMs can be technologically designed in such terms as to tell the truth (or raise a claim to truth), given the inherent statistical dimension with which they provide answers to a given prompt. This does not prevent their accuracy¹⁵ (or “ground truth”) from being improved over time within the inherent epistemic limits on which they are based.¹⁶ Moreover, it is open to question whether accuracy, reliability or truthfulness *by design* do not risk increasing, rather than decreasing, the tendency to over-rely on AI-generated contents. However, it is not the technological feasibility of truth-telling LLMs that we intend to discuss here; rather, the focus is on the desirability of a legal duty to tell the truth, from a threefold standpoint: (i) the incentive for accurate information; (ii) the risk of paternalism and (iii) the generalization of context-based truth duties.

First, careful speech is essential for democracy, but it is costly: it requires time, effort, resources and often expertise. There is a basic, deeply rooted asymmetry in the production of information: it is easy to produce false or inaccurate information, while it is hard to produce true and accurate information. Careful speech is the result of choice, investment and effort, which cannot simply be made mandatory by law. Commitment to true or accurate information is part of a process for which there must always be public or private incentives other than fear of breaking a legal norm.

Second, protecting users from relying on flawed AI-generated contents by providing a duty to tell the truth that takes on a general scope with respect to LLMs sounds somewhat paternalistic. Pointing out the epistemic limits, biases and flaws that pollute the ground truth of LLMs is crucial, but it does not automatically lead to embed into them, by design, a supposed predisposition to truth that would shelter users from the risk of being exposed to falsehood and careless speech. After all, is such risk not naturally part of the basic dialectics of democracy?

Third, the law has always resisted the temptation to prosecute and punish the utterance of falsehood as such. In fact, the law does only regulate the specific circumstances under which a legally relevant and punishable act is committed by the assertion of a falsehood: e.g., giving false testimony in court; attributing an untrue fact to someone by defamation; conducting negotiations in bad faith by misrepresenting an item of sale, etc. In all these cases, it is not the misleading or false statement as

¹⁴ According to the Mertonian norms of the scientific ethos there are four institutional imperatives that guide the work of the scientist: (i) universalism, i.e., the objective nature of the study undertaken, based on previously confirmed knowledge, unrelated to the subjective viewpoint of the individual scientist; (ii) communalism, which indicates the fact that the discoveries produced by scientists are the result of social collaboration and, as such, are assigned to the community in its entirety (iii) disinterestedness, which refers to the attitude that must be embodied by any member of the scientific community, subject to the stringent internal control of peers, since research implies the verifiability of the result obtained by the community itself and (iv) systematic skepticism, which expresses that aspect of physiological distrust that induces the scientist to be critical of his or her own results, in relation to society (Merton, 1973, 266–278). The Mertonian imperatives and their interpretations in the tradition of the philosophy and sociology of science lack reference to truth, emphasizing instead the notions of verifiability, reliability or integrity, as also remarked in (Leonelli, 2023, 1): “As argued by philosophers ranging from Karl Popper to Jürgen Habermas, Helen Longino and Philip Kitcher, what distinguishes a dictator from an elected leader – or a scientist from a crook – is the extent to which their decision-making processes are visible, intelligible, and receptive to critique.”

¹⁵ In this respect, consider Article 13(3)(b)(ii) of the AI Act, which requires that the deployment of high-risk AI systems shall be accompanied by instructions including: “the level of accuracy, including its metrics, robustness and cybersecurity referred to in Article 15 against which the high-risk AI system has been tested and validated and which can be expected, and any known and foreseeable circumstances that may have an impact on that expected level of accuracy, robustness and cybersecurity.”

¹⁶ As the authors underline: “Truth can be optimised in LLMs through a variety of means. Fine-tuning based on authoritative sources or human-authored truthful responses for difficult prompts can introduce external validity. RLHF [*reinforcement learning from human feedback*] workers can provide subjective perceptions of the truthfulness or accuracy of statements and indicate a preference for factual responses. [...] reliability can be improved through extensive curation and annotation, methodologically sound benchmarking metrics, long-term fine tuning with expert human feedback, auditing and adversarial testing, and perhaps even downsizing models” (Wachter et al., 2024, 7–8).

such that is legally relevant, but the fact that it represents a form of conduct under the given circumstances that the law sees fit to pursue and punish. Misrepresentation of a fact is usually not legally relevant except in a *specific context*, as in the case of perjury in court or bad faith negotiation. The law sets the rules for defining the context (e.g., a trial), and then when and how misrepresenting a fact is configured, for instance, as a perjury in that context. Law does not regulate a falsehood *independently of a given context* in which that falsehood counts (i.e., is legally relevant) as a tort or crime. A specific context is the only *level of abstraction* (Florida, 2008) at which a falsehood can be deemed legally relevant.

There is actually a good reason for this, which is based on a fundamental principle of legal and democratic civilization: in a state of law and in a democratic society nobody can or should have sole possession of the truth (historically, States have a monopoly on the use of force, but not a monopoly on truth). If we were to charge a public authority with the responsibility of pursuing and punishing all false statements (or careless speeches), we would implicitly be investing this authority with the power to distinguishing truth from falsehood and thus, ultimately, the power to declare what is true or false in a given society (Durante, 2021, 89).

It is fundamental to identify specific circumstances in which telling falsehoods (or speaking carelessly) is legally relevant, but this is not the same as assuming a general duty to tell the truth with reference to human or artificial agents. In other words, it is certainly important to clarify the specific circumstances in which LLMs are likely to produce legally relevant consequences, such that they may entail forms of legal liability on the part of their providers; but it is questionable – and perhaps not even desirable for the reasons given above – that a general duty to tell the truth can be derived from these specific circumstances.

The reasoning does not lead one to deny the risks implicit in a systematic careless speech, which does not merely generate specific and limited inaccuracies or misrepresentations of reality as the content of discourse but is capable of long-term alteration of the very conditions of discourse. Rather it seems fair to argue that democracy is more protected by the discussion on and unconstrained pursuit of truth open to reasonable disagreement than by the imposition of a general duty to tell the truth.

4. Conclusions

LLMs raise several challenges that can be examined according to a *normative* and *epistemological* approach. The former, increasingly adopted by European institutions, identifies the pros and cons of technological advancement. With regard to LLMs, the main pros concern technological innovation, economic development and the achievement of social goals and values. The disadvantages mainly concern cases of risks and harms produced by means of LLMs. The epistemological approach investigates how LLMs produce data, information and pieces of knowledge in ways that differ from – and often are irreducible to – humans. To fully grasp and face the impact of LLMs, our paper contends that the epistemological approach should be prioritized, since the identification of the pros and cons of LLMs depends on how this model of AI works from an epistemological standpoint and our ability to interact with it. To this end, our analysis compared LLM's and legal epistemology and highlighted five major problematic issues: (1) qualification; (2) reliability; (3) pluralism and novelty; (4) technological dependence and (5) relation to truth and accuracy.

Furthermore, the analysis of such epistemological challenges contributes to understand some potentially long-term consequences of LLMs on relevant democratic aspects of our contemporary society. These consequences may be understood in terms of an impact on (i) human autonomy; (ii) semantic capital and (iii) level playing field.

Human autonomy is impacted by LLMs in several ways: from being unable to verify the accuracy of outputs to being subject to misrepresentations; from reliance on untrustworthy results to the risk of manipulation of opinions formed on biased or hallucinatory content, etc. (Novelli & Sandri, 2024).

In all these cases, the limitation of human autonomy generates a lack of social trust and a loss of individual self-determination.

The production of data, information and pieces of knowledge by LLMs has a strong impact on the production of semantic capital, which is defined as “any content that can enhance someone’s power to give meaning to and make sense of (semanticise) something” (Floridi, 2018, 483). This power plays a crucial role in every democracy and society. The increasing number of outputs and tasks delegated to LLMs can circumscribe or alter human power to give meaning to and make sense of reality.

It should be also noted that the production of data, information and pieces of knowledge made possible by the technological development of LLMs requires resources, investment, time and energy, with a great impact on our society (van der Sloot, 2024, 66–67) and, above all, the environment (Berthelot et al., 2024; Mannuru et al., 2023). As we have tried to highlight, this can benefit the few, already advantaged, and weaken the many, already disadvantaged, by altering the level playing field.

Ultimately, the risk from the epistemological perspective is that the focus on outputs tends to obscure the importance of the path through which the process of knowledge unfolds.

Competing interests. The authors declare none.

References

- Alvarado, R. (2023). AI as an epistemic technology. *Science and Engineering Ethics*, 29(5), 1–32.
- Berthelot, A., Caron, E., Jay, M., & Lefèvre, L. (2024). Estimating the environmental impact of Generative-AI services using an LCA-based methodology. In *CIRP LCE 2024 - 31st Conference on Life Cycle Engineering*, 1–10.
- Delacroix, S. (2024). Augmenting judicial practices with LLMs: Re-thinking LLMs’ uncertainty communication features in light of systemic risks. SSRN, 1–26.
- Diab, R. (2024). Too dangerous to deploy? The challenge language models pose to regulating AI in Canada and the EU. *University of British Columbia Law Review*, 1–36.
- Durante, M. (2021). *Computational power: The impact of ICT on law, society and knowledge*. New York-London: Routledge.
- Durante, M., & Floridi, L. (2022). A legal principles-based framework for AI liability regulation. In *The 2021 yearbook of the digital ethics lab* (pp. 93–112). Cham: Springer International Publishing.
- European Commission (2019). Directorate-General for Justice and Consumers, *Liability for artificial intelligence and other emerging digital technologies*. Luxembourg: Publications Office of the European Union. <https://data.europa.eu/doi/10.2838/573689>
- European Commission (2024). DG for Research and Innovation, *Living guidelines on the responsible use of generative AI in research*. Luxembourg: Publications Office of the European Union, 1–14.
- Fagan, F. (forthcoming 2025). A view of how language models will transform law. *Tennessee Law Review*, 1–64.
- Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7, 1–21.
- Floridi, L. (2004). Outline of a theory of strongly semantic information. *Minds and Machines*, 14(2), 197–221.
- Floridi, L. (2008). The Method of Levels of Abstraction. *Minds and Machines*, 18(3), 303–329.
- Floridi, L. (2013). *The philosophy of information*. Oxford: OUP.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford: OUP.
- Floridi, L. (2018). Semantic capital: Its nature, value, and curation. *Philosophy and Technology*, 31(4), 481–497.
- Floridi, L. (2023a). *The Ethics of Artificial Intelligence: Principles, challenges, and opportunities*. Oxford: OUP.
- Floridi, L. (2023b). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy and Technology*, 36(1), 1–15.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694.
- Fosch-Villaronga, E., & Poulsen, A. (2022). Diversity and inclusion in artificial intelligence. In B. Custers, & E. Fosch-Villaronga (Eds.), *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice* (pp. 109–134). The Hague: T.M.C. Asser Press.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.
- Kelemen, J., & Hvorecký, J. (2008) On knowledge, knowledge systems, and knowledge management. In *Proc. 9th International Conference on Computational Intelligence and Informatics*, Budapest Tech, Budapest, 27–35.
- Krook, J. (2024). Manipulation and the Ai Act: Large Language Model Chatbots and the Danger of Mirrors. SSRN 4719835, 1–38.
- Lenat, D., & Feigenbaum, E. (1991). On the thresholds of knowledge. *Artificial Intelligence*, 47, 185–250.

- Lenat, D., & Marcus, G.** (2023). Getting from generative AI to trustworthy AI: What LLMs might learn from cyc. *arXiv Preprint arXiv:2308.04445*, 1–21.
- Leonelli, S.** (2023). *Philosophy of open science*. Cambridge University Press.
- Lin, S., Hilton, J., & Evans, O.** (2022). Teaching models to express their uncertainty in words. *arXiv Preprint arXiv:2205.14334*, 1–19.
- Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., and Vaidya, P.** (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. *Information Development*, 1–19.
- Merton, R. K.** (1973). *The sociology of science. Theoretical and empirical investigations* (pp. 267–278). Chicago: UCP.
- Mielke, S. J., Szlam, A., Dinan, E., & Boureau, Y. L.** (2022). Reducing conversational agents' overconfidence through linguistic calibration. *Transactions of the Association for Computational Linguistics*, 10, 857–872.
- Moreau, P., Prandelli, E., & Schreier, M.** (2023). Generative Artificial Intelligence and Design Co-Creation in Luxury New Product Development: The Power of Discarded Ideas. SSRN 4630856, 1–46.
- Nazaretsky, T., Cukurova, M., & Alexandron, G.** (2022a). An instrument for measuring teachers' trust in AI-based educational technology. In *LAK22: LAK22: 12th International Learning Analytics and Knowledge Conference*, Vancouver, Association for Computing Machinery, 55–66.
- Nazaretsky, T., Cukurova, M., & Alexandron, G.** (2022b). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53(4), 914–931.
- Nelson, J.** (2024). The Other 'LLM': Large language models and the future of legal education. *European Journal of Legal Education*, 5(1), 127–155.
- Novelli, C., Casolari, F., Hacker, P., Spedicato, G., & Floridi, L.** (2024). Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity. *arXiv Preprint arXiv:2401.07348*, 1–36.
- Novelli, C., & Sandri, G.** (2024). Digital democracy in the age of artificial intelligence. SSRN, 1–27.
- Orofino, M.** (2022). La questione del sotto utilizzo dell'intelligenza artificiale in campo sanitario: Spunti di rilievo costituzionale. *Journal of Open Access to Law*, 4, 158–171.
- Pagalio, U.** (2022). The politics of data in EU law: Will it succeed?. *Digital Society*, 1(3), 1–20.
- Park, J. S., Bernstein, M. S., Brewer, R. N., Kamar, E., & Morris, M. R.** (2021). Understanding the representation and representativeness of age in AI data sets. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 834–842.
- Pasero, L.** (2022). From the Right to Science to the Right to Open Science. The European Approach to Scientific Research. In *European Yearbook on Human Rights* (pp. 515–541). Cambridge: Intersentia.
- Pasero, L.** (2023). Open Science and Data Protection: Engaging Scientific and Legal Contexts. *Journal of Open Access to Law*, 11, 1–16.
- Pearl, J., & Mackenzie, D.** (2018). *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.
- Quine, W. V.** (1953). *From a logical point of view*. Cambridge: Harvard University Press.
- Rebuffe, C., Roberti, M., Soulier, L., Scoutheeten, G., Cancelliere, R., & Gallinari, P.** (2022). Controlling hallucinations at word level in data-to-text generation. *Data Mining and Knowledge Discovery*, 1–37.
- Rodotà, S.** (2012). *Il diritto di avere diritti*. Roma-Bari: Laterza.
- Schiavello, A.** (2020). La scienza giuridica analitica dalla nascita alla crisi Ragion pratica. *Ragion Pratica*, 1, 143–163.
- Surden, H.** (2023). ChatGPT, AI Large Language Models, and Law. *Fordham Law Review*, 92, 1939–1970.
- Tasioulas, J.** (2023). The rule of algorithm and the rule of law. *Vienna Lectures on Legal Philosophy*, 1–19.
- UNESCO.** (2023). *Guidance for generative AI in education and research*. Author. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Vallor, S.** (2024). *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. Oxford: OUP.
- van der Sloot, B.** (2024). *Regulating the Synthetic Society: Generative AI, Legal Questions, and Societal Challenges*. Oxford: Hart Publishing.
- Wachter, S., Mittelstadt, B., & Russell, C.** (2024). Do large language models have a legal duty to tell the truth?. SSRN 4771884, 1–49.
- Zeno-Zencovich, V.** (2023). Big data e epistemologia giuridica. In G. Resta & V. Zeno-Zencovich (Eds.), *Governance of/through big data* (vol II, pp. 439–448). Roma: Roma TrE-Press.