

## Legal Persuasion in the Age of AI.

*Claudio Sarra*

University of Padova

**Abstract:** This article examines algorithmic persuasion within a long-term theoretical framework that links language, persuasion, and legal rationality. Beginning with the classical relationship between speech, corporeality, and persuasion, the analysis traces the historical process through which language becomes progressively objectified, culminating in contemporary generative artificial intelligence. The paper distinguishes persuasion as a psychological effect from persuasiveness as the quality of a discourse as rational achievement emerging in oppositional argumentative contexts, a distinction central to modern legal theory. Drawing on recent empirical literature, the article shows that generative AI can reach persuasive effectiveness comparable to human agents, while identifying transparency, personalization, and cultural heuristics as key moderating factors. Finally, it highlights the ethical and juridical tensions between technological efficiency, human oversight, privacy, and the preservation of human dignity.

*Keywords:* *Algorithmic persuasion, Generative artificial intelligence, Legal argumentation, Rhetoric and rationality, Human–machine interaction*

## 1. Introduction.

«Indeed, in the other qualities that we possess [...] we are in no way distinguished from the animals; rather, we find ourselves to be weaker than many of them in speed, in strength, and in other capacities. But because there is present in us the ability to persuade (*πειθειν*) one another and to make clear to ourselves the matters about which we deliberate, we have not only moved away from living in a bestial manner; rather, after coming together, we have founded cities, established laws, and devised arts and crafts—and nearly all the things that have been contrived through our agency, speech has procured for us»<sup>1</sup>.

The Isocratean passage just cited vividly highlights the close connection between speech, persuasion, and civilization—understood as the management of conflict on the plane of dialectical confrontation to avoid the spreading of violence—that characterizes the mature phase of classical Greek thought, which attributed fundamental importance to the public use of speech in collective practical decision-making<sup>2</sup>.

Even more significantly, the fundamental feature that in this text distinguishes humans from animals — an issue that would more often lead to identify reason as the specific dignity of the human being — is here located precisely in the capacity to persuade one another, which is in turn situated at the level of speech. Yet the epistemic status of persuasion, particularly in relation to truth, remains inherently ambiguous. It is precisely this tension that has given rise to divergent interpretations of the nature and function of rhetorical art, as evidenced in several canonical loci of ancient thought<sup>3</sup>.

Tradition bequeaths to us an image of persuasion as a power capable of moving to action in a manner contiguous with that of erotic seduction. *Peitho* is the name given by the Greek to its mythical personification: a divinity whose origin is, not by chance, connected—and at times intertwined—with that of Aphrodite, the goddess of love.<sup>4</sup> In this sense, prior even to being a specific feature of discourse as such, persuasion expresses an attractive

---

<sup>1</sup> Isocrate, *Antidosi*, Torino, 1991, p. 291(253-256); ID, *Nicocle*, Torino, 1991, p. 133 (5-7). English translation is mine.

<sup>2</sup> J.-P Vernant, *Le origini del pensiero greco*, Roma, 1997, p. 47; T. R. Martin, *Storia dell'antica Grecia. Origini e segreti di una civiltà*, Roma, 2006, p. 104.

<sup>3</sup> In particular in platonic dialogues, especially the *Gorgias*.

<sup>4</sup> As Emma Stafford writes: «Themis and Nemesis, Peitho is rarely found independently in cult, being almost always linked with Aphrodite, and the closeness of this association is demonstrated by a number of occurrences of Peitho as a cult title of Aphrodite», see E. Stafford, *Peithō: the seductive power of rhetoric.*, in ID., *Worshipping Virtues: Personification and the Divine in Ancient Greece*, Wales, 2000, pp. 111–46 (111). See also: R. G. A. Buxton, *Persuasion in Greek Tragedy. A Study of Peitho*, Cambridge, 1982; F. I., Kane, *Peitho and the Polis*, in *Philosophy & Rhetoric*, 1986, 19, 2, pp. 99-124.

force, entirely analogous to erotic power, indifferent to the possible destiny of fortune or misfortune that may follow from it. Significantly, for example, in the myth of Pandora the goddess appears among those who adorn her with the ornaments typical of seductive female beauty<sup>5</sup>.

The original connection between speech, persuasion—understood as an effective impulse to action—and erotic seduction is not easy to explain. Following the perspective proposed by Eric Havelock and subsequently developed in a more generalizing direction by Walter Ong, one might hypothesize that it is to be traced back to the condition of archaic Greek culture as situated within a context of predominant, if not primary, orality<sup>6</sup>. Here, the word is not the written sign, nor is it conceivable as such—that is, as separable in principle from the speaker and the very act of speaking. It is, rather, first and foremost, a bodily *sound* and, as such, a fundamental mode of manifestation of a living entity; indeed, one might say that in this context the word is not at all distinct from the very body that emits it, manifesting its specific vitality.

Speech is thus that same living body projecting itself toward the other and touching it through audible perception.

The association between speech and life, on the one hand, and between death and silence, on the other, would thus have a non-metaphorical origin:

---

<sup>5</sup> See Hesiod, *Works and Days*, in Id., *Theogony and Works and Days*, Oxford University Press, Oxford, 1988 p. 39 (in this version of the passage, *Peitho* is translated with “Lady Temptation”).

<sup>6</sup> «Primary orality» means the condition of a culture that does not possess writing and that, consequently, exhibits characteristics distinct from those of literate cultures, following the framework articulated by Walter J. Ong in *Orality and Literacy: The Technologizing of the Word*, Methuen & Co. Ltd, Malton 1982, which in turn reprises and develops the central thesis advanced by Eric A. Havelock in *Preface to Plato*, Cambridge, MA 1963. See also Havelock’s *The Muse Learns to Write: Reflections on Orality and Literacy from Antiquity to the Present*, New Haven, 1986. From a historical perspective, the collapse of Mycenaean civilization—with its palatial structure within which writing functioned as an instrument of bureaucratic organization in the hands of specialized officials—was accompanied by the loss of this fundamental “technology of the word.” Writing would only later regain centrality in the construction of Greek—and, more broadly, Western—civilization through the adoption of the Phoenician alphabet and its subsequent refinement via the full notation of vowels. That intervening period—often referred to as the “Greek Dark Ages”—witnessed the structuring of Greek culture according to techniques of oral composition and transmission, expressed in poetic forms that would later be distilled in the Homeric epics. The background of these poems is thus to be located in an oral culture. This thesis—particularly in Havelock’s formulation—has been subject to criticism in light of archaeological findings which, albeit sporadic, suggest that writing did not entirely disappear during those “dark” centuries. Nevertheless, insofar as writing no longer constituted a central instrument of collective organization (as would otherwise be evidenced by far more substantial documentation), it remains evident that the prevailing social structure must have been grounded primarily in orality (cf. Martin, *Storia dell’Antica Grecia*, cited, p. 49).

in such a context, where orality constitutes the specific dimension of communication, speech is always and primarily the manifestation of the living<sup>7</sup>.

In this sense, speech possesses the power to move to action and to seduce precisely because it belongs to the bodily dimension—indeed, because it is *a* body that moves and attracts demanding a bodily reaction. Neither the humble nor the powerful can resist it, precisely because they both are corporeal in the first place. Being capable of inexorably drawing the human being toward his or her destiny—whatever that may be—one can understand why Persuasion was placed within the complex configuration of divine powers, according to a figurative schema typical of Greek interpretations of the divine<sup>8</sup>.

Speech, persuasion, and seduction are therefore continuous and inseparable dimensions, insofar as they are originally connected to the constitutive corporeality within which—and not separately from which—the earliest understanding of rationality also unfolds, figuratively identified in the act of “gathering, bringing together”: *logos* in action<sup>9</sup>.

With the advent of a scriptural mentality—or, more precisely, with the gradual collective attribution of epistemic and practical primacy to speech insofar as it is precipitated into the visible signs that represent it—the word detaches itself from the actuality of the vital manifestation that produced it. It is deposited and endures, thus becoming a thing, an object. It can now be observed from the outside, as an alterity with respect to the living body that expressed it; it can be analysed, separated from other word-signs and from the spaces in-between that mimic, in the fixity of the word-object, silence—which, instead, like breathing, is the pulsation of that form of life which in oral communication manifests itself. The written word finds its dimension in the space that contains it and that offers a locus toward which the analytic

---

<sup>7</sup> In this respect, it is important to observe that reading—within the broader transition from an oral to a literate society—does indeed present itself as a procedure for the decoding of written text; however, insofar as it was for a long time practiced primarily and almost exclusively aloud, it also expresses a specific form of vocality that connects the author of the written text to the present audience. In this sense, the author “lives” within the act of reading, while the reader, divesting himself or herself of individual subjectivity, lends the text a voice. Silent reading, that is, reading “to oneself,” appears to be a comparatively later development, see J. Svenbro, *Phrasikleia: An Anthropology of Reading in Ancient Greek*, Ithaca-London, 1993, spec. chapters 3 and 9.

<sup>8</sup> About the complexities of Greek divine figurative approach, see M. Vegetti, *L'uomo e gli dei*, in J-P. Vernant (a cura di), *L'uomo greco*, Roma-Bari, 1991, pp. 264-267.

<sup>9</sup> With reference to the meaning of *λέγω* in Homeric texts, Patrizia Laspia recalls that they show a coherent design «1. *λέγω* significa ‘raccolgere’ come “mettere insieme” [...] 2. *λέγω* significa ‘parlare’ nel senso di ‘mettere insieme i contenuti di un discorso’ [...] 3. *λέγω* significa contare al modo dei popoli senza scrittura, ossia mettendo insieme, raggruppando», see P. Laspia, *Linguaggio e natura. ΔΟΓΟΣ e corradicali da Omero ad Aristotele*, in R. Medda, F. Pelosi, D. Quarantotto (a cura di), *Il Logos nella filosofia antica*, Roma, 2020, pp. 3-24 (6).

gaze may be directed, whereas the spoken word is vibration, sound, and unfolds within the conscious temporality of an actual corporeal existence.

Put differently, the written word—insofar as it is the deposited symbol of the word-as-sound<sup>10</sup> - is already a datafication of the human; writing is a combinatory technology of data, and reading is the algorithm of its decryption. In this context, the problem is no longer persuasion as seduction—that is, as a bodily dynamic—but rather that of “persuasive discourse,” of the ideal unity of abstract rationality and its capacity to impel action, something that, at this point, becomes somewhat mysterious<sup>11</sup>.

From this perspective, the abstraction of word from the actual communicative situation through its perceptible, and thus given (*datum*), sign-dimension constitutes a prelude to its mechanization. Which eventually happened: first through the capacity to multiply already produced written texts (printing), and currently through the capacity of artifacts to produce them in turn.

For centuries the technical difficulty of producing and preserving writing had enhanced its value making the *book* (that is, the eminent device of static memory of the written word expressing a unity of meaning) an object that partook of the immense value of deposited speech. But with the invention of printing the objectification, the datafication, of the word set out toward the apex of its usability and reproducibility. Movable-type printing is already the product of thinking of words as pure abstract symbols.

Generative AI thus represents the most recent stage—chronologically—of the objectified depersonalization of discourse, a process whose structural beginning lies in scripturality. Language has become so objectified that it can be dismantled and reassembled effectively, competently, by the machine itself. The communicative occurrence can be reconstructed through the relative stability of statistically relevant correlations, all the more effective the more collective speech assumes homogeneous, conformist traits. The history of the “speaking machine” is thus also that of communicative conformism

---

<sup>10</sup> Aristotle, *De Interpretatione*, in Id. *Categories and De interpretatione*, (tr. J. L. Ackrill), Oxford Clarendon Press, 1980, p. 44: «Now spoken sounds are symbols of affections in the soul, and written marks symbols of spoken sounds».

<sup>11</sup> As is well known, Aristotle situates the theme of rhetoric within the domain of the inquiry into the means of persuasion in contexts—deliberative, judicial, and epideictic—in which it is necessary to decide among genuine compossibilities («deliberation is about matters that appear to admit of being one way or another; no one consciously deliberates about things that cannot either have been different in the past or be about to be different in the future or be disposed differently in the present, as there would be no profit in it», see Aristotle, *The Art of Rhetoric*, Penguin, London 2004, 1357a, el. ed. s.p.). In these contexts, persuasiveness—according to this reconstruction—lies at the intersection of the qualities of discourse (*logos*), the context in which it is situated, the qualities of the speaker (*ethos*), and the condition of the recipient (*pathos*).

and of the technique of social assimilation of the masses into linguistic models that writing has stabilized and helped to transmit through education.

In this situation, where machines have learnt to combine linguistic *tokens* into complex *corpora*, the persuasive phenomenon becomes populated by unprecedented agents – artificial automata – and takes on new directions: not only from human to human, but also from machine to human, from human to machine, and finally from machine to machine.

The possibility of configuring these relations under the heading of “persuasion” implies that this phenomenon is now to be considered in linguistic-communicative terms dissociated, after the centuries-long process of datafication of the word just outlined, from biological corporeality. It is exactly this process of *dis-embodiment* language (and reason) that permits nowadays to compare human linguistic behaviour *to* that of machines. After all the biological bodily dimension is inexorably absent in the automaton<sup>12</sup>.

## 2. ‘Persuasion’ and ‘persuasiveness’

Before delving into the empirical studies that are currently attempting to reconstruct how the persuasive phenomenon occurs in the social context hybridized by the presence of generative artificial intelligence, it is perhaps appropriate to clarify, with an extremely brief synthesis, in what terms the question is considered today in the methodological study of the practical legal phenomenon.

The complexity of the topic is such that here we can only provide basic theoretical coordinates useful for understanding why and how the theme is present in legal discourse. In fact, persuasion does not present itself as a positive legal category in and of itself, although numerous are the regulatory spaces that presuppose its presence. In general, it can be said that wherever there is agreement or social relationship constituted communicatively – therefore outside of brute force<sup>13</sup> – there also lies the locus of persuasive dynamics, so that – given such pervasiveness – it is not possible to imagine that law regulates them as such. Instead, law typically intervenes in those situations in which the postulated equality between the moral autonomies of subjects is violated, protecting – under certain conditions – those who found themselves acting juridically under conditions of diminished cognitive or practical freedom. Thus, to provide obvious examples, cases of protection

---

<sup>12</sup> Here lies the central critics classically presented by Dreyfuss: see H. L. Dreyfuss, *What Computers Can't do: The Limits of Artificial Intelligence*, New York, 1972, Cap. 7.

<sup>13</sup> Traditionally, *peitho* is, in fact, conceptually opposed to *bia*, violence, see M. Foley, *Peitho and Bia*, in *Symplokē*, 2012, 20, 1-2, pp. 173-181.

against fraud or error can be viewed as situations in which a subject has been induced to act, that is, persuaded, in a state of compromised freedom and for this reason finds remedial measures within the legal system.

The problem of persuasion emerges most acutely not at the level of abstract legal regulation, but within the practical dynamics of identifying and actualizing the law. This is especially evident where unilateral assumptions about the lawfulness of a specific situation are challenged and put to the test, that is, in contexts of legally significant intersubjective conflict.

With the abandonment of a mechanistic conception of the legal order as a self-producing system, and the corresponding reorientation of methodological inquiry toward the interpretative and dialectical dynamics that define the inherently adversarial procedural context, one can discern the deeper rationale behind the significant twentieth-century engagement with theories of argumentation, especially rhetoric<sup>14</sup>.

The latter is, in Aristotelian terms, the “power to observe the persuasiveness of which any particular matter admits”<sup>15</sup> and its specific domain is – in addition to that of political deliberation and epideictic discourse – precisely the judicial sphere. This is exactly because of the specific function it performs where rational alternatives are possible, but a decision is necessary<sup>16</sup>.

The discursive arena, therefore, is the fulcrum of a dynamic from which – starting from a state of contingency and not of necessity – a general order of things is determined.

Rhetoric thus becomes an eminently philosophical theme precisely because of the ‘wonder’ (*thauma*<sup>17</sup>) that follows from the observation of this ordering of the world (*kosmos*) starting from what is not in itself necessary and through the engine constituted by discourse (*logos*), in particular, that one which emerges victorious from the context of oppositions and thus becomes common rule.

---

<sup>14</sup> In particular, starting from the very well-known 1958 *Traité de l'argumentation* by Perelman and Olbrechts-Tyteca whose sub-title is exactly *La nouvelle rhétorique*.

<sup>15</sup> Aristotle, *The art of Rhetoric*, cit., I, 2, 1355b.

<sup>16</sup> And, I have to add, that decision is of definitive importance for the building or preserving of the political community. Rhetoric applies primarily to those three fields exactly because they share these three essential elements (presence of rational alternatives, unavoidability of decision, decisive relevance for the political community). This is true also for epideictic rhetoric: since its main scope is to praise excellence, or to blame vice it encourages and shows a way for moral action, see Pernot, L., *Epideictic Rhetoric: Questioning the Stakes of Ancient Praise*, University of Texas Press, Austin, 2015, pp. 94-95; and, in particular, Oravec, C., ‘Observation’ in Aristotle’s Theory of Epideictic, in *Philosophy and Rhetoric*, n. 9 (1976), pp.162-174.

<sup>17</sup> Aristotle, *Metaphysics*. (ed. D. Ross), Oxford University Press, Oxford 1924, 982b, 10 «διὰ γὰρ τὸ θαυμάζειν οἱ ἀνθρώποι καὶ νῦν καὶ τὸ πρῶτον ἤρξαντο φιλοσοφεῖν».

The rediscovery of rhetoric and argumentation that occurred after the so-called ‘argumentative turn’ of the 1950s<sup>18</sup>, has thus reignited attention toward the theme of persuasive discourse, albeit seeking to maintain it within the bounds of a concept of rationality that is not exhausted by an a priori and complete analytical deduction, but rather valorising the fundamental oppositional moment that, in principle, can extend even to those premises hitherto taken as shared<sup>19</sup>.

This distinction is, however, rejected by radically skeptical visions of legal experience, according to which there is no rationality at stake in the context of normative application but only the imposition by those who hold ‘semiotic power’ of specific and arbitrary meanings upon normative provisions that are accepted simply because they are effectively imposed<sup>20</sup>.

But, leaving aside such extreme positions, and remaining instead within the framework of the distinction between the search of mere persuasion and the building of a persuasive discourse, examples of the former are typically marketing and propaganda, not so much in the sense that no rationality can reside in such types of discourse, but rather because *rational* persuasion does not appear to be their essential objective. Such discourses achieve their purpose when they obtain psychological adherence (in our terms, pure persuasion) by whatever means<sup>21</sup>, even through the use of elements that transcend pure discursivity, such as the exploitation of imagery, symbolism, collective rituals, etc.

Persuasiveness, on the other hand, as it is characterized in the current methodological context, is a quality of discourse that manifests itself particularly in the force of resistance to rational oppositions as it is articulated in a specific context. Moreover, it may not actually persuade; in other words, though deserving it, it may fail in achieving the result of obtaining, the psychological adherence of the recipient. This may happen for many reasons: the recipient’s prejudice, the disinterest in a rational management of the case, the reliance on one’s own pre-established judgment arbitrarily withdrawn from confrontation, etc.

---

<sup>18</sup> See A. Cattani, P. Cantù, I. Testa, *La svolta argomentativa: 50 anni dopo Perleman e Toulmin, 1958-2008*, Napoli, 2009.

<sup>19</sup> See F. Cavalla, *Retorica, Processo, Verità*, Milano, 2007, *passim*; Id, *L’origine e il diritto*, Milano, 2017, pp. 377 ss.

<sup>20</sup> About the idea of a legal «semiotic power», see M. Jori, *I giudici creano diritto?*, in *Stato, Chiese e Pluralismo confessionale*, 2009, 1 luglio, pp.1-24.

<sup>21</sup> Indeed, the psychological dynamics that underlie persuasion revolve around certain ‘response mechanisms’ that humans appear to have developed due to their effectiveness in ensuring the orderly construction of sociality. For example, Robert Cialdini identified six of them: consistency, reciprocity, social proof, authority, liking, and scarcity, see R. Cialdini, *Influence. The Psychology of Persuasion*, New York, 1993.

Furthermore, as a significant element, persuasiveness offers reasons for the critique of persuasion as a mere fact: insofar as what is persuasive has shown its preferability to objections, its rejection in favour of a thesis accepted a priori and not subjected to the same test appears arbitrary – and therefore criticizable<sup>22</sup>.

The construction of persuasive discourse can be understood, in practical terms, as the pursuit of rational decision-making under conditions of contingency—that is, within a domain of what could be otherwise. In contexts of public relevance, where decisions are intended to function as collective points of reference, the extent to which such decisions embody the best expression of collective reason depends on the effective availability of dialectical spaces in which alternatives are articulated and critically examined.

This consideration explains the prominence, within strands of contemporary argumentation theory, of the analysis of oppositional structures. Such prominence has been explicitly recognized and systematically developed—albeit within distinct theoretical frameworks—both in the model advanced by Stephen Toulmin and, more recently, in that elaborated by Douglas Walton. The latter, building upon and extending the approach of Chaïm Perelman and Toulmin, organizes argumentative practices into a plurality of “argumentation schemes.” These schemes are distinguished, in particular, by the explicit formulation of “conditions of refutation,” through which the specific points at which counter-arguments may arise within a given argumentative structure are systematically identified<sup>23</sup>.

In the context of legal-procedural rhetoric, the idea of collecting the types of possible objections within a quadripartite scheme useful for the analysis and methodological organization of training for procedural discussion was instead proposed, based on classical reflection, by Francesco Cavalla<sup>24</sup>. Here, the general framework is structured according to the degree of determination of the opposition in relation to the content of the claim advanced and identifies different modes of rhetorical articulation (aesthetic, didascalical, oratorical, confutatory). Central – as per tradition – is the function of *endoxa*, units of common knowledge, general or specialized (e.g., doctrinal dogmas), upon which the logical relations of arguments are structured, articulated in different types and degrees of analyticity depending on the situation and the type of objection that must be overcome. Persuasiveness, here, has the specific nature of “finding oneself in a common place” rather than put forward a purely strategic action. It ultimately has the logical form of a non-contra-

---

<sup>22</sup> See F. Cavalla, *L'origine e il diritto*, cit.

<sup>23</sup> D. Walton, C. Reed, F. Macagno, *Argumentation Schemes*, Cambridge, 2008.

<sup>24</sup> See F. Cavalla, *Retorica, processo, verità*, cit.

dictory relation that finds no further alternatives and, as such, generates a rational expectation toward the acceptance of the claim.

It is worth emphasizing that the ability to identify, within a given context, the appropriate *loci* of shared knowledge presupposes a high level of cultural formation on the part of the rhetor. This requirement serves the rationality of the legal enterprise. At the same time, it delineates what is often described as the “artistic” dimension of rhetoric.

Particular emphasis should be placed on the fact that the capacity to discern, in context, the relevant *loci* of common knowledge demands a high degree of cultural and intellectual formation. Such a requirement supports the rational integrity of legal reasoning. Simultaneously, it brings into focus the properly “artistic” dimension of rhetoric—an aspect long recognized, yet not always adequately understood in its fundamental significance.

An interesting element of the current technological evolution is, in fact, connected to the artifact’s capacity to identify in immense *corpora* of texts statistical recurrences such as to bring forth and make available for text construction precisely those ‘common places’ that are empirically most recurrent. The widespread use of such tools effectively places AI in the position of determining the relevant *endoxa*, aggravating conformism of thought and translating into statistically relevant quantities a concept that, instead, operates on a delicate quantitative/qualitative balance<sup>25</sup>.

But the point is that, apparently, there is no technical limitation to the building of an artifact that can manage a more sophisticated approach to *endoxa*, once it is trained in field-dependent occurrences of professional dialogues with proper data extracted.

This means that both persuasion *and* persuasiveness are now potentially at the hand of advanced AIs.

### 3. General features of (textual) generative AI

When we speak of artificial intelligence today, we refer to a multiplicity of very different applications, so much so that the definitional problem remains among those still under discussion, nor does the current availability of a normative notion appear adequate to resolve it, due to its generality<sup>26</sup>.

---

<sup>25</sup> *Ἐνδοξά* are “common opinions” that is, those that constitute the opinion of everyone, or of the majority of people, or of the wise, and, among these, either of everyone, or of the majority, or of the most well-known and esteemed, see Aristotle, *Topica*, in Aristotle, *Posterior Analytics and Topica*, Heinemann Ltd, London 1960, p. 305 (105a/b).

<sup>26</sup> In the *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence*, ‘AI system’ means «a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit

Here we focus on so-called ‘generative’ AI and on those applications capable of producing texts, in short, on the ‘speaking machine’.

This does not mean that the automatic generation of images, music, and video cannot be relevant for the future of legal persuasion. All these modalities, in fact, can interact in the construction of virtual worlds and, therefore, in the refinement of digital environments in which legally relevant actions can occur. Consider, for example, the use of virtual reality for the immersive representation in court of facts to be judged<sup>27</sup>, or to optimize the environment of online hearings or, furthermore, in the future, for the construction of procedural settings located in the so-called ‘metaverse’<sup>28</sup>. These implementations are entirely plausible and, indeed, already under experimentation, although, undoubtedly, at present, text generation is that form of AI having the most significant impact on law, including judicial practice.

The history of these applications, as with artificial intelligence in general, now spans several decades of research and development<sup>29</sup>; it has, however, seen a radical turn with the public release of latest-generation language models and in particular with the launch of their perhaps most well-known application, namely ChatGPT.

From a technical standpoint, the GPT model is based on a so-called transformer architecture as proposed in a 2017 article<sup>30</sup>, whose fundamental element of novelty consists of an ‘attention mechanism’ that enables the machine to attribute weighted values to parts of text (*tokens*), converted into numerical vectors, representing their relevance in the specific context, that is, in relation to all other parts of the text equally converted and weighted.

The model – based on deep neural networks – is first pre-trained on a large corpus of texts in an unsupervised manner and then refined (*fine tuning*) using specific datasets containing manually annotated examples referring to specific domains<sup>31</sup>.

---

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»

<sup>27</sup> In December 2024, in a court case in Broward County, Florida, virtual reality was used to present the expert opinion in support of the defense to the judge. Equipped with a Meta Quest 2 headset, the judge was thus ‘projected into’ the event, becoming an active participant in the immersive reconstruction of the facts to be adjudicated. The news is reported in <https://www.local10.com/news/local/2024/12/17/broward-judge-dons-virtual-reality-headset-in-whats-thought-to-be-a-courtroom-first/> (last visit January 29<sup>th</sup>, 2026).

<sup>28</sup> See for instance the *Just Legal VR Courtroom Experience* by Meta for Quest.

<sup>29</sup> See Y. Cao *et al.*, *A Comprehensive Survey of AI Generated Content: A History of Generative AI, from GAN to ChatGPT*, in arXiv:2303.04226v1, 2023.

<sup>30</sup> See A.Vaswani *et al.*, *Attention Is All You Need*, in arXiv:1706.03762, 2017.

<sup>31</sup> The model is described in a 2018 paper published on the OpenAI website: A. Radford *et al.*, *Improving Language Understanding by Generative Pre-Training*. It can be read here: <https://>

All these expressions, naturally, presuppose complex computational operations on numerical structures – typically of a vectorial type – into which linguistic *tokens* are converted, with the objective – in generating the output – of identifying the set of those that are most probably expected in the succession that gradually constitutes itself.

Although it is not necessary here to enter into greater detail regarding the precise mathematical operations that the artifact performs step by step to generate its contents, it is important to always bear in mind that this is what it involves, and nothing more. In non-specialized literature, including legal literature, and sometimes even in that which would reflect philosophically on the phenomenon, an excessive anthropomorphization of the phenomenology of generative AI still appears present today, something that unfortunately sometimes hinders a profitable discussion.

That said, and appropriately reiterated, it is equally true that anthropomorphization—understood as the human tendency to measure reality against itself—constitutes a characteristic mode through which the world is apprehended, both cognitively and practically. This holds also with respect to mechanical artifacts, which are themselves intentionally produced within that broader process of humanization.

Anthropomorphization, that is, the inclination of humans to interpret the world by reference to themselves, represents a typical mode of both cognitive and practical engagement with reality. This dynamic extends as well to mechanical artifacts, which are deliberately generated within the same process of humanizing the world.

Once produced, artifacts appear as a form of alterity that must, in turn, be reinterpreted and reassessed. As a mode of alienation, this process calls forth a rational strategy of recomposition. Significantly, this strategy closely resembles that through which humans confront natural alterity, composed of elements and forces perceived as potentially threatening. In both cases, the response consists in a figurative transfer of the self, whereby what is external is interpreted in distinctly human terms<sup>32</sup>. This strategy, perfectly rational, which on the cognitive level has the nature of metaphorical thought<sup>33</sup>, has

---

[cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf) (last visit January 29th, 2026).

<sup>32</sup> And that since the beginning: Joseph Weizenbaum recalls how, after only a few interactions with his ELIZA software, the individuals he had invited to participate in the test began to display attitudes of trust and even a desire for intimacy in their dialogue with it, see J. Weizenbaum, *Computer Power and Human Reason*, Freeman & Co., New York, 1976, p. 7.

<sup>33</sup> Metaphors have the fundamental function to manipulate the commonplaces associated to a term (in Aristotelian terms *endoxa*, see. M. Black, *More About Metaphor*, in A. Ortony (ed.), *Metaphor and Thought*, Cambridge, 1993, 19-41, spec. p. 28) with the aim to project

radical consequences: in fact, once enacted, it makes available categories, rules of judgment, heuristic and practical tools that, continuously refined thereafter, allow for the establishment of a lasting relationship with that part of the external world.

So, it happens – and it is a phenomenon of great current relevance and interest – that in the relationship with the speaking artifact, man finds himself experiencing dynamics that appear representable, to some extent, in terms of communication, with consequent psychological and, therefore, also practical effects<sup>34</sup>.

Given the continuous expansion of the use of such tools in ever wider sectors of experience, public and private, and the dependence on them of decisions regarding behaviours to be adopted, it therefore appears interesting to study this transformation also in terms of a communicative relationship: in such form, in fact, for the reasons just mentioned, it is experienced by man.

From this point of view, it is possible to question whether communication with the artifact also produces both reasoned and persuasive constructs in the logical-rational sense mentioned above, and mere persuasion on contents of apparent rationality.

This latter hypothesis is certainly relevant also for specific purposes connected to the legal regulation of AI: consider, for example, the issue related to the concrete implementation of human oversight which Article 14 of Regulation (EU) 2024/1689 on artificial intelligence (the so-called *AI Act*) prescribes as mandatory for ‘high-risk’ AI systems and which provides – in § 4, letter b – that the person to whom such task is entrusted must be placed in a position «to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (*automation bias*)». It is readily conceivable that, in an era defined by conversational artificial assistants, the psychological dynamics of persuasion will assume a pivotal role in the design and implementation of appropriate measures aimed at preserving the human subject’s autonomous and supervisory position of control.

---

categorical tools of understanding – and, then, also of practical governance – to something that is unknown, sometimes even deprived of a denomination. In this sense, they have an essential function in law as tools for managing and overcome oppositions in danger of becoming incommensurable and prone to violence. See, *amplius*, C. Sarra, *Lo scudo di Dioniso. Contributo allo studio della metafora giuridica*, Milano 2010; and, more recently, Id., “Formed out of fire and earth”. *Legal metaphors and the quest for rational commensurability*, in D. Medina Morales, A. Sánchez Hidalgo (eds.), *Estudios de retórica jurídica*, Valencia, 2020, pp. 261-295.

<sup>34</sup> With reference to the humanizing relation of *media* – although quite counterintuitively – see B. Reeves, C. Nass, *The Media Equation. How People Treat Computers, and New Media, Like Real People and Places*, Cambridge, 1996.

As for persuasiveness, given that it, ultimately, expresses the best articulation in the concrete case of the principles of identity and non-contradiction (albeit starting from endoxastic premises), the possibility for AI to play a role in methodological support depends on its effectiveness in producing relevant and semantically recognizable texts as well as well-structured on shared premises. On this point, however, the current state of generative AI suffers from important deficiencies that impose constant attention to the solidity of outputs, and, in particular, require the implementation of accurate measures of control and mitigation of so-called ‘hallucinations’<sup>35</sup>.

But, apart from that, it is entirely plausible the ability of the machine to generate arguments perfectly in line with the principles of a correct argumentative practice.

#### 4. Artificial Intelligence and Persuasion

Recent literature on persuasion – understood here as a psychological phenomenon – marks a progressive conceptual shift from an exclusively human domain to one that involves artificial systems in an increasingly structured manner. This transition not only redefines the theoretical boundaries of persuasion, but raises fundamental questions regarding the nature of communication, the role of the technological artifact as agent, and the psychological response of individuals in algorithmically mediated contexts. In this framework, a first non-trivial problem concerns the very definition of ‘persuasion’ when the communicative dynamic involves an AI. It seems, in fact, difficult to eliminate from the configuration of the phenomenon any reference to subjective intention (to induce a certain response in the receiver), given that, without it, it would be difficult to distinguish consequent behaviours that

---

<sup>35</sup> That is, the phenomenon whereby generative AI produces texts that are apparently well-formed and persuasive but are in fact incorrect or nonsensical in context, and therefore particularly insidious insofar as they are nonetheless capable of inducing reliance (persuasion). The causes of such phenomena are associated with a variety of factors, including: inadequacy of prompts; contradictory, insufficient, or non-specific training data; the detection of spurious correlations; as well as the very ‘a-semantic’ architecture, that is, one based on the mere statistical prediction of linguistic tokens. Despite advances in research, the problem of hallucinations remains unresolved. As for the most commonly used remedies, approaches include models equipped with detection and mitigation systems such as retrieval-augmented generation (RAG), which integrates factual data from external sources during inference; the use of chain-of-thought prompting to improve reasoning; as well as reinforcement learning from human feedback (RLHF) to better align responses with user expectations and reduce the frequency of hallucinations. These approaches aim to improve so-called factual grounding and the coherence of generated outputs. For an up-to-date overview on this issue. See, C. Woesle, L. Fischer-Brandies, R. Buettner, *A Systematic Literature Review of Hallucinations in Large Language Models*, in *IEEE Access*, 13, pp. 148231-148253

have actually been determined by that specific communicative interaction from all others that simply follow as a matter of fact.

Thus, persuasion is sometimes defined as a «symbolic process in which an artificial communicative entity generates, amplifies, or modifies a message *designed* to induce a change in the responses of the human recipient»<sup>36</sup>. The definition, as can be seen, is very broad and finds its critical point precisely in the identification of a design element and, therefore, an intentional one, which, however, appears problematic if referred to the artifact. On the other hand, the elimination of this element risks leading to definitions that are too broad, such as not to allow the identification of the specific phenomenon: as said, every behaviour consequent to a communicative instance could thus be seen as the effect of a successful or failed persuasion, thereby making the boundaries of the specific phenomenon confused<sup>37</sup>.

Notwithstanding this serious question, with the advent of generative AI various studies are attempting to measure the persuasive capacity of the artifact by comparing it with that of human interlocutors in analogous situations. The most recent meta-analyses – although in a context where shared methodological standards are still lacking – highlight some important preliminary findings relating both to the overall effectiveness of the artifact in inducing persuasion and to the factors that appear most significant in the production of this complex phenomenon.

---

<sup>36</sup> M. Denhart, P., Mongeau, *Persuasion in the Age of Artificial Intelligence (AI): Theories and Complications of AI-Based Persuasion*, in *Human Communication Research*, 2022, 48, pp. 386-403 (387). The approach adopted here is grounded in well-established cognitive models, such as the Heuristic-Systematic Model (HSM), which distinguishes between persuasion based on established heuristics—characterized by rapid responses and minimal cognitive effort, yet less stable and more dependent on the type of source—and systematic processing, which requires greater cognitive effort, provides more stable content evaluation, and is less dependent on the nature of the source (see S. Chaiken, *Heuristic vs Systematic Information Processing and the Use of Source versus Message Cues in Persuasion*, in *Journal of Personality and Social Psychology*, 1980, Vol. 39, 5, pp. 752-766). As regards the relationship with the artefact, the CASA (Computers Are Social Actors) and MAIN (Modality, Agency, Interactivity, Navigability) models suggest that individuals attribute social qualities to machines and respond to persuasive cues in ways analogous to those observed in human-to-human communication (B. Reeves, C. Nass, *The Media Equation*, cit.; S. S. Sundar, *The MAIN Model: A Heuristic Approach to Understand Technology Effect on Credibility*, in M. J. Metzger, A. J. Flanagin (eds), *Digital Media, Youth, and Credibility*, Cambridge MA, 2008, pp. 73-100). The authors observe that the possibility for humans to recognize AI as their interlocutor, on the basis of more or less evident cues (thin/thick AI), alters—by rendering it more complex—the analytical framework that can be constructed on the basis of the aforementioned paradigms.

<sup>37</sup> This problem seems to be present in another quite used definition, that is «communication that shapes, reinforces, or changes people's beliefs, attitudes, or behaviors», see P. Schoenegger *et al.*, *Large Language Models Are More Persuasive Than Incentivized Human Persuaders*, in [arXiv:2505.09662v2](https://arxiv.org/abs/2505.09662v2), 2025.

A recent reconstruction has thus compared the persuasive capacity of AI with that of humans through a large sample of studies within a framework that distinguishes factors such as: the role of AI in communication (AI as ‘conversational agent’, ‘decision-maker’, ‘information selector’), the general ideological context (individualistic/collectivistic), the communicative typology (mono/bi-directionality)<sup>38</sup>. The outcomes indicate that, overall, AI proves to be as persuasive as human beings: this general result appears extremely relevant in assessing the future relevance of the artifact in social relations<sup>39</sup>.

However, significant nuances emerge: while AI proves comparable to humans in modifying perceptions, attitudes, and actual behaviours, it is slightly less effective in promoting behavioural intentions. Moreover, its effectiveness decreases when it assumes the role of ‘autonomous decision-maker’ (*contemplator*) and when its artificial nature is made explicit (*AI labeling*). This seems to reflect a widespread preference for an ancillary role of AI in sensitive decision-making processes, as well as the activation of cultural heuristics that – at present – favour the human source when the artificial origin is evident. This empirical finding is in line with the ethical sensitivity and general orientation of contemporary regulation<sup>40</sup>, which, on the one hand, emphasize the importance of specific transparency obligations (see, for example, *AI Act*, art. 50) and, on the other hand, reflect widespread consensus about maintaining AI within the realm of tools only to support decision-making in areas of particular significance, for example, precisely, the legal domain (see the Italian recent Law n.132/2025, arts. 13 and 15).

No relevant difference seems, however, to emerge from this literature regarding AI’s persuasive capacity with reference to the cultural-ideological context of individualistic or collectivistic type, nor, overall, from different directional articulation.

Further systematic reviews confirm these trends and deepen the factors that moderate persuasive effectiveness<sup>41</sup>.

In these cases too, the relevance of AI labeling is confirmed, but it is added that while knowledge of the artificial source reduces the persuasiveness of the message, interactivity and personalization of communication seem to

---

<sup>38</sup> See G. Huang, S. Wang, *Is artificial intelligence more persuasive than humans?*, in *Journal of Communication*, 2023, 73, pp. 552-562.

<sup>39</sup> In P. Schoenegger *et al.*, *Large Language Models Are More Persuasive Than Incentivized Human Persuaders*, *cit.*, has been noticed an efficacy even superior to that exercised by economically incentivised humans.

<sup>40</sup> See on point the *Ethical Guidelines for Trustworthy AI* by the HLEG published in 2019.

<sup>41</sup> Like those by A. Rogiers *et al.*, *Persuasion with LLMs A Survey*, in arXiv:2411.06837v1 (2024) e N. B. Bozdog *et al.*, *Must Read: Systematic Survey of Computational Persuasion*, in arXiv:2505.07775v1 (2025).

increase it significantly, thus somehow compensating for the effect of disclosure<sup>42</sup>.

The adoption of *prompts* that solicit structured and evidence-based responses also appear to have an improving effect on the persuasiveness of artificial communication.

By contrast, model complexity (in terms of number of parameters) does seem to affect effectiveness, but, apparently, not beyond a certain threshold from which the quality of training data and fine-tuning procedures become more significant<sup>43</sup>.

Strategies such as appeal to authority and the integration of reliable external information also prove effective in enhancing the credibility of the generated message. Furthermore, research highlights how the anthropomorphization of the interface – through the use of natural voices, synthetic faces, or colloquial language – by activating heuristics linked to the recognition of familiar social cues, can enhance persuasiveness, especially in contexts of care, education, or customer service<sup>44</sup>.

In terms of applications, studies focus on contexts such as negotiation, debate, and communication in healthcare, political, and advertising domains. Again, it is found that in such scenarios, AI demonstrates a persuasive capacity often comparable to that of humans.

On the other hand, the most recent literature is also beginning to explore the inverse phenomenon: the persuasion of AI by humans, up to the use of jailbreaking techniques or manipulation of inputs, to obtain desired outputs by overcoming any limitations on response capabilities, which raises unprecedented questions of ethics, security, and control<sup>45</sup>.

---

<sup>42</sup> However, in the specific context of using AI for electoral persuasion, a recent empirical study published in *Nature* does not appear to have found such effectiveness in personalisation strategies, suggesting that the extent of this effect is heavily dependent on the specific field, see H. Lin, *et al.*, *Persuading voters using human-artificial intelligence dialogues*, in *Nature*, 2025, 648, pp. 394–401.

<sup>43</sup> Strategies of *post training*, and *prompt* that stimulate high *information density* answers are at the core of the persuasive capacity of language models also in K. Hackenburg *et al.*, *The levers of political persuasion with conversational artificial intelligence*, in *Science*, 2025, 390, s.p. The authors note, though, that «The most persuasive models and prompting strategies tended to produce the least accurate information, and post-training techniques that increased persuasiveness also systematically decreased accuracy».

<sup>44</sup> C. Nass, Y. Moon, *Machines and Mindlessness: Social Responses to Computers*, in *Journal of Social Issues*, 2000, 56, 1, pp. 81-103; A. Waytz, J. Heafner, N. Epley, *The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle*, in *Journal of Experimental Social Psychology*, 2014, 52, pp. 113-117.

<sup>45</sup> *Jailbreaking* is the use of *prompting* techniques by an user to substitute, interrupt, or overcome the safety policies embedded in the model, thus producing results that should be prohibited, see S. A. Akheel, *Guardrails for Large Language Models: A Review of Techniques and Challenges*, in *Journal of Artificial Intelligence, Machine Learning and Data Science*, 2025,

## 5. Conclusions

Recent research on persuasion induced by artificial intelligence ultimately raises three general fundamental questions, closely intertwined with one another. At the centre of the reflection are the conceptual presuppositions of persuasion itself and the profound ethical and cultural implications of the relationship between human being and machine. The first critical node concerns the definitional dilemma. In traditional communication between human beings, persuasion normally implies a conscious intention to influence beliefs or behaviours, a fundamental element also for the attribution of responsibility. However, AI, as a ‘speaking machine’, possesses neither mental states nor subjective intentionality. This generates a paradox: if the intentional element is maintained in the definition, one risks an undue humanization of the artifact and, with this, an *a priori* and fictitious commensuration between man and machine; if it is eliminated, one may incur the opposite risk, namely a de-humanization of the human agent and the establishment of a new and equally artificial *a priori* equation.

The second node concerns the regulatory *topoi* and cultural dynamics that shape human-machine interaction. The persuasive relationship is in fact mediated by heuristics that influence both the depth of cognitive elaboration and the duration of persuasive effects. These heuristics, however, regarding the actual behaviours they trigger, are not universal: they derive from specific cultural contexts and are subject to change over time. New generations, developing profoundly different communicative habits, may significantly modify the persuasive processes observed today. In perspective, it cannot be excluded that cultural changes may lead to a different acceptance of AI as an autonomous decision-making agent, calling into question the current – and perhaps only temporary – consensus toward its merely ancillary role, especially in sensitive domains such as the legal one.

Finally, the third node concerns the structurally invasive nature of artificial persuasion, and it is at this level that the very foundations of contemporary legal experience are touched. As has been seen, it appears that communication personalization strategies are effective, in general, in increasing the persuasive capacity of the artifact. But such a strategy necessarily implies that the AI ‘knows’ its interlocutor and, that, ultimately, it processes an appropriate quantity of data relating to their person and specific existential context. That way, the personalization of the message, if, on the one hand, it increases the communicative effectiveness of the artifact, on the other hand, makes the phenomenon inevitably invasive.

---

3, 1, pp. 1-9.

What emerges in this case too<sup>46</sup>, is a fundamental ethical and legal dilemma: limiting invasiveness to protect individual privacy entails a reduction in the functional effectiveness of the system; conversely, pursuing maximum (in this case) persuasive effectiveness entails an inevitable progressive erosion of the private sphere.

This tension underscores a broader and more consequential theoretical issue: insofar as the contemporary understanding of “progress” is framed in terms of technological empowerment, and insofar as technology itself may be conceived as an “indefinite expansion of power”<sup>47</sup>, it becomes conceptually untenable to posit credible limits to its pervasiveness—even when such efforts are guided by the most robust ethical and institutional commitments. Technology, by its very logic, cannot accommodate a deliberately inefficient articulation of its own capacities, except for reasons that are merely contingent—arising, in particular, from the material constraints of the artifacts through which it is instantiated at any given moment, and which are destined to be supplanted by increasingly efficient and performant successors.

Any other limit that one wishes to impose (ethical or legal) can only be perceived as extraneous and as an expression of a contradictory will to entrust technology and at the same time employ it in a consciously suboptimal manner: the unsustainability of such a position will therefore inevitably imply a push to overcome every constraint.

In this perspective, technology—understood in its contemporary instantiation as symbolic engineering—is increasingly invested, by the spirit of the age, with a foundational role. Yet, in relation to this project, it stands in a potential tension with human dignity, to which—following the arduous processes of secularization and the tragic ruptures of the world wars—has been entrusted precisely the task of reconstituting the historical meaning of our ethical and legal experience<sup>48</sup>.

It must be understood, therefore, that the real threat to humanity does not derive so much from this or that particular technological realization, nor does the mere reaffirmation of twentieth-century fundamental principles in yet another bill of rights suffice to avert it. It may possibly derive from that collective decision (implicit, fundamental, and fatal) which, in the name of progress understood as power, would wish to circumscribe the understanding of the principle of humanity within the absolutized categorical horizon of technology.

---

<sup>46</sup> For another relevant example see also the last chapter in C. Sarra, *Il mondo-dato*, cit.

<sup>47</sup> See E. Severino, *Il destino della tecnica*, BUR, Milano 1994.

<sup>48</sup> That is the main concern in C. Sarra, *Dignità dell'uomo nell'era della intelligenza artificiale e della secolarizzazione*, Kront, Roma, 2025.

In other words, in this historical moment, within the ideology of *human-as-data*<sup>49</sup>.

Ironically, that epochal conversion means depriving humans of their actual living bodily engagement in conversation, definitely redefining persuasion by burying, once and (perhaps) for all, the mighty *Peitho*.

## References

- Aristotle. *Categories and De Interpretatione*. Translated by J. L. Ackrill. Oxford: Clarendon Press, 1980.
- Aristotle. *The Art of Rhetoric*. London: Penguin, 2004.
- Aristotle. *Metaphysics*. Edited by W. D. Ross. Oxford: Oxford University Press, 1924.
- Aristotle. *Posterior Analytics and Topica*. London: Heinemann, 1960.
- Hesiod. *Theogony and Works and Days*. Oxford: Oxford University Press, 1988.
- Black, Max. "More About Metaphor." In *Metaphor and Thought*, a cura di Andrew Ortony, 19-41. Cambridge: Cambridge University Press, 1993.
- Bozdag, N. B., et al. "Must Read: Systematic Survey of Computational Persuasion." arXiv:2505.07775v1, 2025.
- Buxton, R. G. A. *Persuasion in Greek Tragedy: A Study of Peitho*. Cambridge: Cambridge University Press, 1982.
- Cao, Y., et al. "A Comprehensive Survey of AI Generated Content: A History of Generative AI, from GAN to ChatGPT." arXiv:2303.04226v1, 2023.
- Cattani, Adelino, Paola Cantù, e Italo Testa. *La svolta argomentativa: 50 anni dopo Perleman e Toulmin, 1958-2008*. Napoli: Liguori, 2009.
- Cavalla, Francesco. *L'origine e il diritto*. Milano: FrancoAngeli, 2017.
- Cavalla, Francesco. *Retorica, Processo, Verità*. Milano: FrancoAngeli, 2007.
- Chaiken, Shelly. "Heuristic vs Systematic Information Processing and the Use of Source versus Message Cues in Persuasion." *Journal of Personality and Social Psychology* 39, no. 5 (1980): 752-766.
- Cialdini, Robert B. *Influence: The Psychology of Persuasion*. New York: HarperCollins, 1993.

---

<sup>49</sup> This is the position of the so called «dataism» which I defined as the "universal-religious" dimension of datafication, see C. Sarra, *Il mondo dato*, cit., pp.47-61.

- Denhart, M., e P. Mongeau. "Persuasion in the Age of Artificial Intelligence (AI): Theories and Complications of AI-Based Persuasion." *Human Communication Research* 48 (2022): 386-403.
- Dreyfuss, Hubert L. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper & Row, 1972.
- Esiodo. *Le opere e i giorni*. In *Tutte le opere*, a cura di Cesare Cassanmagnago. Milano: Bompiani, 2009.
- Foley, M. "Peitho and Bia." *Symplokē* 20, no. 1-2 (2012): 173-181.
- Hackenburg, K., et al. "The levers of political persuasion with conversational artificial intelligence." *Science* 390 (2025).
- Havelock, Eric A. *Preface to Plato*. Cambridge, MA: Harvard University Press, 1963.
- Havelock, Eric A. *The Muse Learns to Write: Reflections on Orality and Literacy from Antiquity to the Present*. New Haven: Yale University Press, 1986.
- Huang, G., e S. Wang. "Is artificial intelligence more persuasive than humans?" *Journal of Communication* 73 (2023): 552-562.
- Isocrate. *Antidosi*. Torino: UTET, 1991.
- Isocrate. *Nicocle*. Torino: UTET, 1991.
- Jori, Mario. "I giudici creano diritto?" *Stato, Chiese e Pluralismo Confessionale* (1 luglio 2009): 1-24.
- Kane, Francis I. "Peitho and the Polis." *Philosophy & Rhetoric* 19, no. 2 (1986): 99-124.
- Laspia, Patrizia. "Linguaggio e natura. ΛΟΓΟΣ e corradicali da Omero ad Aristotele." In *Il Logos nella filosofia antica*, a cura di R. Medda, F. Pelosi, e D. Quarantotto, 3-24. Roma: Edizioni di Storia e Letteratura, 2020.
- Lin, H., et al. "Persuading voters using human-artificial intelligence dialogues." *Nature* 648 (2025): 394-401.
- Martin, Thomas R. *Storia dell'antica Grecia: Origini e segreti di una civiltà*. Roma: Newton Compton, 2006.
- Nass, Clifford, e Youngme Moon. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56, no. 1 (2000): 81-103.
- Ong, Walter J. *Oralità e scrittura: Le tecnologie della parola*. Bologna: Il Mulino, 2014.
- Oravec, Christine. "Observation in Aristotle's Theory of Epideictic." *Philosophy and Rhetoric* 9, no. 3 (1976): 162-174.

- Pernot, Laurent. *Epideictic Rhetoric: Questioning the Stakes of Ancient Praise*. Austin: University of Texas Press, 2015.
- Radford, Alec, et al. “Improving Language Understanding by Generative Pre-Training.” OpenAI, 2018. [[https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf)]([https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf)).
- Reeves, Byron, e Clifford Nass. *The Media Equation: How People Treat Computers, and New Media, Like Real People and Places*. Cambridge: Cambridge University Press, 1996.
- Rogiers, A., et al. “Persuasion with LLMs: A Survey.” arXiv:2411.06837v1, 2024.
- Sarra, Claudio. \*Dignità dell’uomo nell’era della Intelligenza artificiale e della secolarizzazione\*. Roma: G&BPress, 2025.
- Sarra, Claudio. “‘Formed out of fire and earth’: Legal metaphors and the quest for rational commensurability.” In *Estudios de retórica jurídica*, a cura di D. Medina Morales e A. Sánchez Hidalgo, 261-295. Valencia: Tirant lo Blanch, 2020.
- Sarra, Claudio. *Il mondo-dato. Saggi su datificazione e diritto*. Padova: CLEUP, 2022.
- Sarra, Claudio. *Lo scudo di Dioniso: Contributo allo studio della metafora giuridica*. Milano: FrancoAngeli, 2010.
- Schoenegger, P., et al. “Large Language Models Are More Persuasive Than Incentivized Human Persuaders.” arXiv:2505.09662v2, 2025.
- Severino, Emanuele. *Il destino della tecnica*. Milano: BUR, 1994.
- Stafford, Emma. “Peithō: The seductive power of rhetoric.” In *Worshipping Virtues: Personification and the Divine in Ancient Greece*, 111-146. Wales: The Classical Press of Wales, 2000.
- Sundar, S. Shyam. “The MAIN Model: A Heuristic Approach to Understand Technology Effect on Credibility.” In *Digital Media, Youth, and Credibility*, a cura di Miriam J. Metzger e Andrew J. Flanagin, 73–100. Cambridge, MA: The MIT Press, 2008.
- Svenbro, Jesper. *Phrasikleia: An Anthropology of Reading in Ancient Greek*. Ithaca: Cornell University Press, 1993.
- Vaswani, Ashish, et al. “Attention Is All You Need.” arXiv:1706.03762, 2017.
- Vegetti, Mario. “L’uomo e gli dei.” In *L’uomo greco*, a cura di Jean-Pierre Vernant, 264-267. Roma-Bari: Laterza, 1991.

- Vernant, Jean-Pierre. *Le origini del pensiero greco*. Roma: Editori Riuniti, 1997.
- Walton, Douglas, Chris Reed, e Fabrizio Macagno. *Argumentation Schemes*. Cambridge: Cambridge University Press, 2008.
- Waytz, Adam, Joy Heafner, e Nicholas Epley. “The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle.” *Journal of Experimental Social Psychology* 52 (2014): 113-117.
- Weizenbaum, Joseph. *Computer Power and Human Reason: From Judgment to Calculation*. New York: W. H. Freeman & Co., 1976.
- Woesle, C., L. Fischer-Brandies, e R. Buettner. “A Systematic Literature Review of Hallucinations in Large Language Models.” *IEEE Access* 13 (2025): 148231-148253.

