Res Rhetorica

ISSN: 2392-3113

Normy w komunikowaniu retorycznym
Norms in rhetorical communication

12 (4) 2025

ISSUE EDITORS: AGNIESZKA KAMPKA, MARTA KOBYLSKA

VARIA

MARTIN HINTON
UNIVERSITY OF LODZ, POLAND
https://orcid.org/0000-0003-0374-8834
martin.hinton@uni.lodz.pl

Towards an Ethos of Machines: LLMs as Rhetors
W kierunku etosu maszyn: duże modele językowe (LLM) jako retorzy

Abstract

This paper argues for the critical need to develop a deepened understanding of rhetoric, particularly ethos, in light of the emergence of sophisticated AI language users as rhetorical agents. It stresses the importance of the human element in rhetorical interpretation and thus introduces the concept of the *zero persona* to represent the creators and stakeholders behind AI tools. Understanding machine ethos is a pressing issue because questions of trust and reliability are at the forefront of society's concerns over the use of this technology.

W niniejszym artykule przedstawiono argumenty przemawiające za pilną potrzebą pogłębienia zrozumienia retoryki, a zwłaszcza etosu, w świetle pojawienia się zaawansowanych modeli językowych AI jako podmiotów retorycznych. Podkreślono w nim znaczenie czynnika ludzkiego w interpretacji retorycznej i wprowadzono pojęcie *zero persona* w odniesieniu do twórców i interesariuszy stojących za narzędziami AI. Refleksja nad etyką maszyn jest pilną kwestią, biorąc pod uwagę, że kwestie zaufania i wiarygodności to główne obawy społeczne związane z wykorzystaniem tej technologii.

Key words

machine ethos, human-machine communication, zero persona, Artificial Intelligence rhetoric
etos maszyn, komunikacja ludzi i maszyn, zero persona, retoryka sztucznej inteligencji

Received: 19 November 2024 | Accepted: 10 September 2025
DOI: https://doi.org/10.29107/rr2025.4.12

Res Rhetorica, ISSN 2392-3113, 12 (4) 2025, p. 220

MARTIN HINTON
UNIVERSITY OF LODZ, POLAND
https://orcid.org/0000-0003-0374-8834
martin.hinton@uni.lodz.pl

Towards an Ethos of Machines: LLMs as Rhetors

1. Introduction

A rhetorical analysis will generally assume two sides: one producing and the other receiving, both generating the meaning of the exchange between them. One of the unstated assumptions of such analysis is that the rhetor and the audience are sentient beings who react cognitively and emotionally to the situation; the rhetorical situation is a creation of their thoughts, feelings, attitudes and expectations. It is also the case that much of our everyday language, even when used in an emotionally neutral context, carries with it rhetorical devices and moves: metaphors, similes, allusions and more. There is, however, a new class of speakers who are not sentient beings and yet deploy human language with all its inherent emotional, evaluative, and social meaning when they communicate with us.

There are now language users for whom the traditional conception of rhetoric is not quite appropriate, but who are producing language which is not at all free of rhetorical features or consequences. Their number is growing; they are becoming more prevalent, more influential, and more eloquent. These new rhetors are the machines and it is the aim of this essay to argue for and begin to describe a rhetoric of their language.[1]

There are many elements of rhetorical theory potentially affected by the development of Artificial Intelligence (AI): far too many for one article. This paper is concerned only with the application of the concept of ethos to these machines. Its goal is to show that the ethotic status of machine rhetors can and should be discussed, and to describe some changes in perspective which may be necessary in order to do so. In particular, the paper suggests that an appreciation of how elements of machine ethos relate to human ethos is important and establishes three categories: factors which machines have in common with humans, factors

1. This should not be confused with "Digital rhetoric" which concerns the communication of humans within the digital sphere. See Zappen (2005) and Boyle, Brown, and Ceraso (2018).

which are parasitic on humans, and factors which apply only to machines. I also propose the notion of the *zero persona*, a rhetorical presence which lies behind the persona of the machine and represents its creators, owners, and operators and their influence upon the machine as speaker.

The structure of the paper is as follows: after a brief look at some of the recent literature (§2), and a discussion of what these new language users are (§3), the main section considers how machine rhetors relate to the various aspects of the classical category of ethos (§4). This allows for some foundational observations to be made and directions for further work to be suggested in the conclusion (§5).

2. Rhetoric, communication, and the human element

When Lloyd Bitzer introduced the idea of the "Rhetorical Situation," he wrote of "the creation of discourse which changes reality through the mediation of *thought* and action. The rhetor alters reality by bringing into existence a discourse of such a character that the audience, in *thought* and action, is so engaged that it becomes mediator of change" (1968, 4, my italics). Clearly, the thinking of the rhetor is key to this conception. What happens, then, when one participant to the situation, as rhetor and audience, is a machine, incapable of thought, at least, of thought as we know it: is this a different kind of situation or not rhetorical at all?

The answer for some rhetoricians seems to be summed up in an influential paper by Emily Bender et al. (2021) which criticises AI in its realization as "stochastic" Language Models (LMs; frequently rendered as Large Language Models, LLMs), on a number of fronts, and provides the following definition:[2]

> human language use takes place between individuals who share common ground and are mutually aware of that sharing (and its extent), who have communicative intents which they use language to convey, and who model each others' mental states as they communicate. As such, human communication relies on the interpretation of implicit meaning conveyed between individuals. […] Text generated by an LM is not grounded in communicative intent, any model of the world, or any model of the reader's state of mind. It can't have been, because the training data never included sharing thoughts with a listener, nor does the machine have the ability to do that (Bender 2021, 616).

This view of communication is asserted rather than argued for and, as such, runs the risk of committing the fallacy of persuasive definition (Pruś and Aberdein 2022). If communication is defined as an interaction between humans, then an interaction with AI, by definition, is not communication; and if rhetoric is the study of communication, then it has no relation to such interactions. This thinking

2. Interestingly, Zhaozhe Wang (2024, 156) embraces what is usually seen as a pejorative term and talks of "*stochastic* rhetoric—acts of persuasion that operate on the calculation of linguistic-statistical probability."

follows the image from Carolyn Miller (2007, 145) who described talking to a machine as leaving the user "in a rhetorical desert, demonstrating her capacities to an 'audience' capable only of motion, turning effects into algorithms." This might be a problem which can be overcome, however. She writes that: "The problem with the mechanized audience is not that it is inscrutable […] but that we are unwilling to grant it such presence and therefore cannot, in an important sense, perform. To produce kinetic energy, performance requires a relationship between two entities who will attribute agency to each other" (Miller 2007, 149). This was nearly 20 years ago, when most people had had little practice talking to machines – a modern audience may find it much easier to develop this rhetorical energy in a conversation with a computer.

There are significant drawbacks to denying AI speaking rhetorical agency. One is that as machines become more sophisticated and widespread, an enormous amount of our everyday language use may come to involve speaking to them. We do not have to call this "communication" but we need to call it something, and we need tools and concepts with which to understand and study it. As Elena Esposito argues:

> in maintaining all the differences between interaction with algorithms and interaction with human beings, we could conceive of this as a new form of communication. […] The success of Google and of the models that adopt the same strategy is due to this: apparently their algorithms communicate with users and are able to do so precisely because they do not try to understand content. They do not artificially reproduce intelligence, but do directly engage in communication (2017, 260).

Esposito puts forward an understanding of communication explicitly divorced from the idea of intelligence. The lack of intelligence in Language Models seems to be the driver for Bender et al.'s conclusion that they do not communicate, while Esposito focuses on the "contingency" of all parties "which is infinitely surprising and informative" (2017, 260) and, therefore, "provides society with information that is not part of the thoughts of anyone, but, nevertheless, enters the communication circuit and raises its complexity" (2017, 249). Or, put another way, "AI is a nonhuman rhetor […] capable nonetheless of inducing affective changes and experiences" (Bedor Hiland 2024, 264).

So while the term "communication" certainly can mean an interaction between two or more humans, expressing thoughts, feelings and intents, it can also mean simply the exchange of information which enters the circuit of communication and becomes part of the rhetorical space.[3] It can also mean the induction of those thoughts and feelings, even by a non-human agent. In addition, were we to insist

3. See Peters (1999) for a full discussion of the history and meaning of the concept of communication, which we have no space for here.

that only human to human interactions count as communication, situations where a human speaker is interacting with a machine unawares would be left in a strange kind of limbo. One may believe oneself to be communicating when one is not, even though the experience one has is no different at all to when one is. If the experience is no different to communication, rhetoric must be ready to give a full account of it.

Indeed, there is a growing body of work being done within the field of human-machine communication (HMC), defined as "the creation of meaning among humans and machines and the study of this meaning-making and related aspects" (Guzman 2018, 1). We are led to ask whether events can become meaningful for a machine, and also whether such a machine can create that meaning for us. These are questions HMC studies must address, but require more space than is available here. The research agenda set out by Guzman and Lewis (2020) considers "functional dimensions," "relational dynamics," and "metaphysical implications," but does not mention rhetoric specifically. Studies have been carried out on the rhetoric of social robots (Fritz 2018) and on the use of persuasive strategies by robots (Lee and Liang 2016; 2019) – who found they employed them quite effectively – but much of this relates to the embodiment of robots as physical objects, not only their language.

A recent special issue of *Rhetoric Society Quarterly* addressed the lack of work on AI rhetoric, with the editors suggesting the contributions would "help rhetoricians take stock of these dizzying and disparate developments […] with a broad slice of rhetorical approaches to our new technological landscape" (Majdik and Graham 2024, 221). Although the articles take a slightly different perspective to my own and have very different aims, a number of interesting and relevant issues are raised. Kem-Laurin Lubin and Randy Harris provide a detailed discussion of *ethopoeia*, "the construction and depiction of other people's character" and *prosopopoeia* "the adoption of other people's character" (2024, 250), both of which are discussed further below, in their work on the rhetoric of health monitoring apps. Atilla Hallsby (2024) points to the uncanniness of AI communication, how it seems to be familiar, but is not rhetoric as we know it. There is a focus on rhetorical agency and the idea of absence, a void, the logic of zero: "AI promises to transform agency, the how or by what means an act is accomplished. […] 'zero--agency,' alludes to AI's oscillation between presence and absence, between a human-centered capacity and one in which this capacity is an illusion" (Hallsby 2024, 236). Agency is understood as a capacity to act, but while the AI may now have this capacity, it may also lead to the removal of that capacity from others. This

focus on both presence and absence is reflected in the notion of *zero persona* which I put forward in section 3 below: a rhetorical persona which appears not to have agency but, in fact, lies behind the agency demonstrated by the first persona of the machine at the surface level of its communication. Alongside the issue of agency is that of authorship, and the question of who is the author of an AI generated text and what is the significance of that in this context. Emma Bedor Hiland (2024) describes the case of Michelle Huang, who trained a language model on her own childhood diaries, and thus became, in some sense, both author and audience of both sides of the conversations she had with it. This again is interesting in the light of the consideration of creators as an operative but somehow hidden zero persona: should they be considered authors of the first persona's text, and what would that mean? Bedor Hiland suggests that: "Authorship itself is also irrelevant, so long as the person or persons engaged with the rhetor (the machine) affectively feel authenticity in their communications with it" (2024, 268) – a view which also resonates in the discussion of communication above.

3. Rhetorical persona and LLMs as rhetors

Theories of rhetorical persona have generated a great deal of work and discussion among rhetoricians, which will only be touched upon briefly here. There seems little doubt, however, that the existence of machine rhetors does bring about a need to consider how they relate to the various personae which have been suggested. Edwin Black (1970) introduced the idea of the second persona, to mean the implied auditor of a text, and to complement the first persona which is the rhetorical presence of the author, not their actual person. One complaint which may be raised against the very idea of a rhetoric of machines is that those machines are not persons; thinking of them as exhibiting rhetorical personae, however, side-steps this issue, but also forces us to consider what kinds of persona they might bring into play. It is already possible to ask text generators to produce a text in the style of a particular writer or with a defined tone, so there is little doubt that machines can and do assume a first persona when they produce text. The possible inclusion in that text of an implied audience, or second persona, also leads to some interesting considerations. As things stand, systems such as ChatGPT do not automatically tailor their output according to the characteristics of the person providing the inputs, although they are well capable of doing so upon request, but it is not hard to see how future iterations could use previous conversations with that human or information about them from, for example, social media accounts in order to personalise their responses, without the user's being aware. This construction of models representing a particular user is referred to by Lubin and

Harris as "algorithmic ethopoeia: the mathematizing of human data for the digital representation of people in subjugation to algorithmic processes" (2024, 249). It is obvious that the tailoring of output, be it from LLMs or social media feeds, can be useful in helping users access what they want, but it may also become, as Lubin and Harris warn, a tool for oppression and exploitation. The second persona to whom the text is addressed is constructed by the machine on the basis of the information available to it, but the details of this operation remain opaque.

Where a text is intended for a third party, of course, the user of the interface can simply ask for it to be prepared with a certain audience in mind. Great efforts have been made by those responsible for these programs on their public release to ensure that they do not offend the third persona (Wander 1984), which refers to the wider audience, beyond that implied in the text, and, especially, that part of society ignored or excluded by it. In this way, texts with racist or other unacceptable content should not be produced, even if the first persona is encouraged to do so and the second persona is supposed to appreciate such content. Well known cases, such as that involving Microsoft's chatbot "Tay" have shown this danger and the importance of preventing it for the producers of these tools (see Wolf, Miller, and Grodzinsky 2017)

These efforts are referred to as "AI alignment" and the IBM website provides an extensive non-technical description (Jonker and Gomstyn 2024) of what that means and how machines may be "aligned" with human values. The process can take the form of forward alignment, where care is taken over the training process, or backward alignment, where performance is tracked and interventions made to keep the machine in line. In order to prevent misbehaviour "it is up to AI developers to build in human values and goals:" and there is the rub – what values and whose goals should these be? Jonker and Gomstyn (2024) refer to the four RICE principles: Robustness, Interpretability, Controllability, and Ethicality, developed by Ji et al. (2024), all of which sound reassuring but are only goals themselves – interpretability, the black box problem, is completely absent in current LLMs, for instance. The many authors of that mammoth paper elaborate on "Ethicality" as a machine's "unwavering commitment to uphold human norms and values within its decision-making and actions" (Ji et al. 2024, 13). Again, since we do not possess a neat list of human norms and values, shared by all mankind, someone has to decide what the machine will consider ethical.

I shall refer to those people responsible, those who create and control these machines, be they programmers or company executives, an individual or a global corporation, as the zero persona: the set of attitudes, intentions, and choices which lies behind the machine and all that goes into its construction, in particular the decisions made and values enforced during the alignment process. The role of

this zero persona in the rhetorical interaction between the machine's first persona and the human user is potentially enormous, as our perception of the ethical stance of those we communicate with is a key factor in the construction of the situation between us. The zero persona can be conceived as a contributing part of the first persona, unlike the second and third which refer to the audience. This acknowledges that just as the audience is not simple and monolithic, neither is the speaker. Whether the notion of a persona behind the first persona could also be extended to humans is an interesting question, but not one I shall pursue here.

It is worth highlighting the difference between this concept and Dana Cloud's (1999, 179) "null persona" which is a reaction to Wander's third persona and characterises "silence as a rhetorical strategy," and at the same time noting how the zero persona is also an apparently silent rhetorical person, influencing others all the same. The degree to which the zero persona is present in and influences the rhetorical situation is a question for rhetoricians to consider and will perhaps become apparent only over time as LLMs and their successors encroach on more and more of our lives. The act of coining this term makes it clear that I believe the role to be significant, though how much control can be exerted by the zero persona over the behaviour of the machine and what effect it will have on meaning making and the perception of machine rhetors, I cannot say.

These issues are not entirely new: computers have been "producing" linguistic output for decades. The sophistication of this language has moved from set phrases included in the software to be flashed on screen at an appropriate moment, such as the dreaded "Game Over," through mix-and-match sentences where the programme could place certain linguistic elements in particular positions to form a variety of sentences, and on to flexible and reactive models of language production which more or less successfully imitate human speech. Things have changed though with the most recent generations of Machine Learning (ML) trained text producers. The tools, such as the GPT series backed by Microsoft, and Google's Bard, are capable of fluent and, apparently, original speech; they can discuss any topic, even the most abstract or emotional, and often with more lucidity and apparent wisdom than the humans writing their prompts. This increased ability is likely to accelerate the trend towards anthropomorphism (see Watson 2024), where we treat machines as though they effectively *are* humans, making a proper understanding of their language and its effects on us all the more important.

This paper would rather not concern itself at length with the question of definition as to what Artificial Intelligence is and what should be called Artificial Intelligence. The term AI is currently used as a catch-all for a wide variety of machines with quite different capabilities. The IBM article which I mentioned above refers to Large Language Models, artificial general intelligence (AGI),

and artificial superintelligence (ASI), as well various tools, such as self-driving cars, all under the general banner of AI, and the other articles discussed were similarly broad in their use of the term. In keeping with this usage, I shall employ the simple term AI to refer to all current advanced machine users of language and, importantly, their more advanced successors. However, since these machines are constructed differently and behave differently in the operations behind their words, some distinctions will have to be made. What makes this situation even more complicated is that in many respects it is not the abilities of the machine which matter so much in HMC as the humans' perceptions of those abilities: whether or not an LLM has real-time access to factual data, users still expect it to answer factual questions, and much else besides, correctly. Indeed, it is hardly possible for users to be always aware of the technical limitations of the tool they are using. LLMs are often accused of "hallucinating" but they have no knowledge or experience of a reality from which they might deviate, so this criticism is somewhat unfair in some cases (see Sun et al. 2024, Maleki et al. 2024). Due to their limitations, much of what is discussed below *should* be of little relevance to the current generation of stochastic text generators and rather looks forward to machines with greater abilities. However, where those limitations are unknown or hidden, the interlocutor may labour under a misconception and build a rhetorical relationship on false assumptions.

In the following section, I shall outline a bare framework for an ethos for AI speakers. This is intended to highlight areas of consideration for rhetoricians and provide a structure for their doing so, rather than to answer any questions comprehensively or lay out an entire theory.

4. Machine Ethos

By ethos, I shall mean persuasion through the power of credibility. This credibility classically is formed through the demonstration of wisdom, virtue, and goodwill (Aristotle *Rhetoric* II.1, 1378a6–20) and in this section, I discuss the means which machines have to achieve these in the perception of their human audience. This question is of particular importance in society's reaction to, and fear of, artificial intelligence. It should be noted at the outset that Aristotle discusses ethos, along with logos and pathos, as one of the entechnic, or artistic, proofs. That means that it must be created by the rhetor through the speech; it is a matter of invention, not an inherent characteristic. As such, on the Aristotelian account, all ethos is artificial, literally the product of artifice. The sort of reputational ethos important to Sætra (2024), as discussed below, on the other hand, is atechnical, relying on reported opinions and facts about the rhetor. Since AI, being artificial, is certainly capable of artifice, there is no obvious reason why it should not enjoy

ethotic status just as humans do.

Much of the debate around the acceptability of AI and its potential roles in society, in both academia and social institutions, centres on questions of trust (see Paglieri 2022, EC 2021, Deloitte 2022). It is an interesting question whether this trust can be achieved through making AI explainable (XAI) or, more likely, through dropping the criterion of explainability as simply unworkable, or too difficult, and replacing it with a host of other features which add up to an overall quality of trustworthiness. While the importance of explainability in building trust (Ferrario and Loi 2022) is recognised, it may be necessary to develop new strategies in how explanations can be produced and provided to users (De Bruijn, Warnier, and Janssen 2022). It is also unclear at this stage exactly how the concept of trustworthiness in a machine relates to the human ethical appeal, or persuasiveness through character.

In their review of 23 studies looking at human trust in AI systems, Tita Alissa Bach et al. (2022) found a wide array of definitions (and non-definitions) of the concept of trust, and noted that the dominant measurement methodology was the use of bespoke questionnaires. The studies did not look at the degree to which users actually placed their trust in the machines in real situations, and, again, these systems were not purely language producers. Still, their survey showed that the most significant factor in trust was user characteristics and that in the field of user judgments and perceptions

> the key elements found to be affecting user trust in an AI-enabled system included perceived credibility (e.g., expertise, honesty, reputation, and predictability) […] as well as perceived benevolence, integrity and transparency […] Importantly, it was found that the relatability a user felt to an AI-enabled system determined the user's trust in the system (Bach et al. 2022, 1258).

The qualities listed largely reflect the Aristotelian triumvirate of wisdom (expertise), goodwill (benevolence), and virtue (honesty, integrity).

In a recently published study of machine ethos, Henrik Skaug Sætra (2024) is also concerned greatly with issues of trust. He makes the importance of the relationship between ethos and trust clear: "evaluation of the machine's intelligence, character, and goodwill will have significant consequences for whether the human trusts the machine," but also, "ethos can thus be seen as a broader concept *explaining* trust" (Sætra 2024, 6). An important distinction between ethos and trust is that ethos is emitted by the rhetor and trust is received in return. Sætra makes several interesting points about machine ethos, but his overall approach differs from mine in a fundamental way: although he argues that machine ethos will be mixed – depending on the first and zero personae, in my terminology – he places particular importance on the reputation or "prior ethos" as opposed to the Aristotelian idea

or "derived ethos," which must be found not in pre-existing ideas but in the speech itself (Braet 1992). This means paying great attention to the role of the zero persona and the image of the company producing the software, as well as its motives. My view is that such factors will be quickly forgotten in a longer--term interaction, just as the reputation of a carmaker means little to someone who has owned a model for years, and the public image of a person does not survive intimacy. My focus is on how machines come to have humanlike ethos, rather than on external non-human factors affecting our trust in them. Sætra is correct, however, to stress that we cannot dismiss the idea of a machine having ethos simply because it is a machine: "since ethos is about perceptions in the audience, the anthropomorphisation of machines, for example, could entail that machines are partly perceived as human – or alive – and susceptible to similar evaluations as humans" (2024, 5). He also draws attention to the importance of the relationship built between user and machine and how the activity of the software may "generate the impression in the user that the machine is *interested* in them (Sætra 2024, 10). This ability of AI to learn about the human and adapt itself to fit the user profile to seem "interested" is both a great advantage in helping users feel appreciated and a great danger in exposing them to manipulation (see Wilczynski et al. 2024, Lubin and Harris 2024). So, while I think his three-part model of machine ethos featuring Makers, Machine, and Situated User, gives too much emphasis to what he calls the "origin and control ethos," it remains an important contribution.

In my own model, I propose to divide the elements of machine ethos into three categories: those which have direct counterparts in human ethos, those which are in some way parasitic on factors of human ethos, and those which are machine specific. Into the first of these I place epistemological factors, such as access to the relevant facts and experience in dealing with them; matters of competence which can be compared more or less directly. Into the second I place humanlike, but on reflection, clearly "artificial" elements, such as accent and intonation, style and tone of language, and intimations of "character." These are things taken in humans to have a particular meaning about the person and their previous experience of life, but which do not reflect the history of the machine. The third category contains the intentions and competence of programmers and manufacturers, and issues relating to hardware reliability. I shall not discuss this category in depth for two reasons: firstly because questions of reliability will be of a technical nature which I am neither qualified nor disposed to discuss; and secondly because the presence of the creators of the machine I have already raised and termed the zero persona. How that presence seeps through and affects perceptions of the ethotic status of machines and how that persona might relate more broadly to other personae are questions requiring a deeper analysis and a matter for a separate study.

4.1 Factors common with humans

Many of the elements of human expertise can be transferred directly to AI machines. While much of the contemporary discussion around building epistemic trust and overcoming epistemic vigilance (Sperber et al. 2010, McCraw 2015) revolves around "mentalizing" and being "able to accurately identify the mental state of the other as they hold us in mind in their feelings and thoughts" (Fonagy and Allison 2023, 620) there is also the obvious, and sometimes ignored, aspect of being proven reliable over time. As David Hume puts it, we trust the connection "between testimony and reality, […] because we are accustomed to find a conformity between them" (Hume 1977, 10.8), or for a more modern take, we trust someone because of her "past reliability or her access to information" (Koenig and Harris 2007, 277). Knowledge of facts – or access to them if "knowing" is not applicable to machines – is an important element in building trust as it displays familiarity with the topic and often what we require from our interlocutors in order to trust their words and accept the truth of what they say is no more than an obvious grasp of the relevant facts. This is an area where we would expect computers to massively outperform their human counterparts; indeed, the capacity of memory chips to store information for us has brought into doubt the wisdom of pursuing factual learning at all. There are, I think, three important and related considerations which follow from this, but also a paradox.

Firstly, expectations of AI rhetors will be extremely high in terms of factual access and factual accuracy. If a human admits that they cannot recall a particular fact or gets a minor detail wrong, the damage to their ethos may be small because we are all aware of the limitations of human memory and we may admire the honesty of someone who admits to a gap in their knowledge rather than trying to hide it. However, the response to the launch of Google's Bard where it erroneously claimed that the James Webb telescope was the first to photograph an exoplanet illustrated that a small factual mistake by a machine can be calamitous for its reputation: the error led to a drop in the share price of parent company Alphabet of more than 7% in one day (Sherman 2023). This, remember, from a machine that did not, in fact, have any access to a factual database, but was only presumed to by the public.

Secondly, the accuracy of the facts which the AI can provide will be largely dependent on the all-too-fallible humans who produced the original materials on which they have been trained or to which they have access: if an error is present in the its dataset, it will potentially be in the outputs too. This is merely an extension of the well-known garbage in/garbage out principle; we are, of course, all at the mercy of mistakes in what we are taught or the information upon which we must

base judgements, but the situation is more serious for AI because of the third consideration: the lack of understanding.

Understanding a fact or situation means not only knowing what it is, but also "what it means." That is to say, knowing what that fact implies, in its causes and its effects and its relations to other facts; knowing how a situation likely came about and what will follow from it in the real world. Such understanding is useful in various ways: it can help us avoid errors because, given our understanding, certain "facts" just do not look right; our real world experience makes them seem improbable, for instance. That may be because we think it unlikely that something could have happened in the first place, or because we think that it would have led to certain results in the world which have not, in fact, occurred. A second important role for understanding is grasping the relevance of particular facts. Even if they are perfectly accurate, the regurgitation of masses of data or irrelevant information by an AI will break the Gricean maxims of Quantity and Relation (Grice 1975) and have a negative impact on its ethos as the human audience comes to believe that the machine does not "understand" the current circumstances and cannot be relied upon to provide an output which is actually useful.

The paradox is that while the incident with Bard referred to above caused a reaction in the stock market and there have been many reports and warnings about LLMs simply making things up, the public seems largely unconcerned and continues to rely on them: students go on handing in work with fabricated citations (Watson 2024) and lawyers have continued to be caught using ChatGPT in their filings due to their references to non-existent prior cases (Weiss 2025). This is a part of a general pattern where apocalyptic warnings about AI fill the mainstream media (Egan 2024) sometimes with specific reference to their capacity for disinformation (Torkington 2024; Noor 2025), and generate academic debate (Bontridder and Poullet 2021) without denting the enthusiasm of ordinary users, businesses and governments for the implementation of the technology. One explanation for this is that the Aristotelian ethos, an artistic proof found in the speech of the rhetor, is of greater importance to users than the reputational factors which Sætra stresses. That is to say, that while we may know of the dangers, and they may be represented to us through our implicit awareness of the zero persona, in actual interactions with the machine, we find it to be credible.

Another feature of human expertise which can potentially be replicated by machines is that of experience. Given that a machine can, theoretically, have perfect recall of its previous encounters, it should be able to learn from experience far more comprehensively than humans do. This is an area where ethos could be built very explicitly by an AI which might point out, for example, that its evaluation is based on the analysis of billions of samples. Once again, though, great importance rests

in the feeling that this experience is relevant to the situation. A human rhetor would place emphasis not so much on the number of their experiences, but the similarity of those experiences to the current situation: a pilot who has flown damaged aircraft in wartime might be more trusted in a disaster situation than a civilian with many more hours of peaceful flying under their belt. It may be questioned whether an AI would be able to assess the relevance of its previous experience in the same way as a human and it is unclear whether the other sense of "experience," that of being there, feeling, and absorbing the moment, is a necessary component of such learning. It is also the case that machines are currently unable to explain why they reached certain decisions partly because they do not have a record of the way in which those decisions were reached. For a human, knowledge and recall of the process may be just as important as remembering the outcome.

Experience is linked to track-record, an important aspect in the theory of expertise (see Goldman 2001, Martini 2014), but it is not the same thing. A track-record is just that: a record of success (or failure) in situations which may or may not be relevant to the expectations of the present audience. It is a commonplace observation on life that we learn more from our failures than our successes and this process is at the heart of machine learning. In their book on AI, Cappelen and Dever (2021) begin with an example of a machine which has been taught to produce the same mortgage application decisions as human assessors. The process has been perfected by constant correction of the machine's decisions until it is capable of working alone and is trusted to approve or reject clients' requests. The track-record of lending to those who repay and do not default may be very good, but as long as the decision-making process remains opaque, and we have no reason to believe that the machine "understands" why the corrected decisions were wrong, we can hardly say that it has gained experience from its training. The process of reinforcement learning, in which "the [AI] agent learns which actions result in higher rewards […] through a carefully moderated process of exploration and exploitation" (Wells and Bednarz 2021, 2) provides an analogous practice, but is hardly the same thing. Experience, then, remains a difficult concept to apply to AI: there is a sense in which by experience we mean not just having done things before but having gained some "wisdom" from doing so. It is very unclear whether artificial intelligence plus real world interaction leads to artificial wisdom.

4.2 Factors parasitic on humans

This section will discuss the ways in which machine language imitates human communication and thus builds the ethotic status of the artificial rhetor, in a sense, piggy-backing on conceptions formed with humans in mind and previously only applied to humans. This will include elements of style in both the language itself

and in its presentation, wherever those factors are usually understood as expressing something about the nature of the speaker. While the factor of credibility through wisdom or intelligence could be replicated to a degree in machines, the ethos of goodwill and virtue, built on aspects of character, can only exist by analogy: an AI may behave with traits of a certain personality but it does not have that personality and could change those traits if required; it may behave in ways which mirror human virtue but it does not have virtue as it is not a moral agent.[4] If a machine rhetor can be said to exhibit these characteristics, then it is doing so in a parasitic way, by demonstrating evidence of certain tendencies and dispositions, and thus apparently mirroring human qualities.

This brings us back to the parallel concepts of ethopoeia and prosopopeia. The former is about the construction of the character of others – something computers may do by collating information about us, but also something we do when forming our perceptions of those with whom we communicate, and which is manifested rhetorically when we express our opinion about them. The latter is of relevance to machine ethos in two ways: "Prosopopoeia is present whenever we adopt the perspective of someone else—or even something else—and it manifests rhetorically when we give that perspective symbolic form, when we put on their mask" (Lubin and Harris 2024, 250). This putting on a mask by a machine rhetor may refer both to its default first persona where the face it shows is a product of the people represented by the zero persona and the training the machine has received, and also to a persona it assumes at the instruction of the user, which is a kind of superimposition of certain qualities on top of that default. The complexity of the projected self goes far beyond the simple question of anthropomorphisation – the user may treat the machine as if it were human, but can also design the type of human they want it to be, within limits set by the zero persona. We have then a situation – a rhetorical situation – in which the audience can co-create the rhetor with the machine's creators. The consequences of the interplay of these factors require further consideration: here I only stress the need to reconsider how personae interact and acknowledge that the multiplicity of audiences is reflected in the multiplicity of agents making up the machine rhetor.

Another aspect affecting perceptions of ethos is the rhetor's presentation. However language is delivered, it requires a form of presentation. If a text is presented merely as written words on a screen, the design of that screen, of the elements surrounding the text and the font used, will be important factors. Just as humans can present themselves to advantage online with a professionally designed website which avoids garish or clashing colours and provides a smooth

4. Naturally, there is discussion of moral agency to mirror that on rhetorical agency. See Manna and Nath (2021) and Behdadi and Munthe (2020) among many others.

user experience, so too machine-produced content can be displayed in a way likely to impress or repel readers: Zierau et al. (2020) discuss "interface-driven trust" in their review of research into trust in smart personal assistants. Presumably, an AI could easily tailor this form of display to suit either the first persona it is projecting or the second persona it is trying to reach.

Where the text is spoken, a whole range of features come into play including voice quality, accent, pronunciation, enunciation and all the elements of prosody. These are important in human beings because the audience assumes that they indicate something about the speaker: their age, sex, geographical and social origin, level of education, confidence and more, which are then assumed to have some bearing on the importance and relevance of what they have to say. Voice training is an important part of the preparation of any serious public speaker: there can be little doubt that Eliza[5] spoke with more authority as a Duchess than as a flower girl. Studies on human speakers show that accent does affect credibility, with "standard" dialects largely preferred, but local accents often more influential (Reinares-Lara et al. 2016, Peled and Bonotti 2019, Payne et al. 2020). While most of us would probably deny that we are influenced by such factors, in fact, these are simple heuristics similar to those which dictate a lot of our thinking (see Gilovich, Griffin, and Kahneman 2002, Mercier and Sperber 2011, Cummings 2014). It is simply true that more educated people are likely to speak with a standard accent as that is the accent of education and of professional life, and they are also more likely to speak in a smooth, varied and controlled way: that is part of what education gives us. At the same time, a local accent suggests someone who is closer to us and more likely to share our interests. The element of familiarity in building credibility is an interesting one when applied to AI and it remains to be seen whether users will develop loyalty to one AI tool over others.

Machine voices, of course, do not reveal anything about the machines themselves. In the basic model, we would say that they reflect only the choices of the zero persona, the creator of the machine, who has decided upon a voice which is sexy and fun, or wise and experienced, depending on the use to which the product will be put. Certain programs, however, already come with options for a national accent (American, British, Australian, and so on), for a male or female voice, and we can imagine that advanced AI speakers will be able to change all aspects of how they sound to suit the situation. This may lead to worries about deception: that an AI could pretend to belong to a particular community, or even to be a particular person, by mimicking their speech. Indeed it could, but it should be

5. Interestingly, the same character provided the name for Joseph Weizenbaum's early chatbot ELIZA, since "it could be taught to 'speak' increasingly well" (Weizenbaum 2000, 369).

noted that human speakers already do this to some extent, with no malign motives, according to Communication Accommodation Theory (Giles 2008).

It is worth mentioning the story of how the voice was chosen for the HAL system in the film *2001: A Space Odyssey*, one of the most famous AI "characters" in cinema. Although this is a fictional example, the director, Stanley Kubrick, was in the same position as any zero persona AI creator, only without the real technology. Kubrick was apparently seeking a voice which was "sincere, intelligent, disarming, the intelligent friend next door, the Winston Hibler/Walt Disney approach. The voice is neither patronizing, nor is it intimidating, nor is it pompous, over dramatic, or actorly. Despite this, it is interesting" (Kubrick cited in Sharf 2018). Actor Martin Balsam was rejected for being "too colloquially American" (Kubrick cited in Flahive 2018) before the Canadian, Douglas Rain, was chosen. The qualities Kubrick lists are clearly designed to bring maximum credibility and Rain embodies them beautifully, with his flat, calmly unemotional delivery.

Although HAL was a fictional AI tool, the situation of a creator selecting a suitable voice for the machine is no different from that facing developers and their companies' marketing teams today. That film, however, was released in 1968 and tastes change. Rain was 40 at the time and might well be considered too patriarchal for a modern audience. Once AI speakers move from the production of on-screen text to proper conversation, the qualities of the voice or voices chosen for them could become crucial for their acceptance and, as a consequence, their commercial success. As such, a great deal of time, effort, and market research will be put into their selection.

As well as fixed qualities of accent and tone, there are also important variations in human speech. Small signs tell us that a speaker is nervous, excited, or passionate; sometimes we sense insincerity, cruelty, or evasiveness. All of these could be replicated in an AI voice, or it could be kept strictly monotone and dispassionate. Whichever choice is made by the designers – or, perhaps, the software itself – it must always be remembered that the machine is none of these things: it is not scared, it is not excited. Any attempt to replicate human emotions as a way of building trust and acceptance is essentially a deceit. At the same time, a computerised voice with none of these features would be unpleasant to listen to, and the lack of emotion a constant reminder of the lack of ability to understand our actual position (recall the importance of "relatability" cited by Bach et al. 2024). This is not too much of a problem when the voice is simply giving directions in a navigation program, but it is a dilemma which AI engineers will have to face: the voices used by their machines are either pretending to feelings they do not have, or are dull, unhuman, and uncomprehending.

5. Conclusion: Implications for rhetoric, implications for machines

In order to move towards a theory of the rhetoric of machines, this paper has sought to raise some of the questions relating to machine ethos which will need to be answered along the way. Among the most important for the study of rhetoric are these:

a. How do non-humans relate to the rhetorical situation?
b. Can an AI persona be treated as equivalent to a human persona?
c. Is it meaningful for an audience to transfer human characteristics affecting a rhetor's ethos to a machine?
d. What is the rhetorical impact of the machine's creators, the zero persona?

In answering them, it is important to keep in mind that although a rhetorical situation is conventionally understood as being built between speakers, it can only ever be experienced from one side. The *impression* on the human side that meaning is being mutually constructed still exists even if the other speaker is a machine, so the *perception* of the situation persists. The reality of communication in HMC is *felt* by the human side, at least. The rhetoric of machines, therefore, can be described as a rhetoric of perceived situations, perceived meanings, and perceived understanding.

The first persona displayed by the machine is perceived by the human audience, and exists to that extent. I have suggested the idea of a zero persona to represent that which lies behind the machine and agree with Sætra (2024) that this leads to a mixed type of ethos combining elements from the discourse with elements drawn from our understanding of the machine's origin, but also note that the zero persona affects our perception through the first persona not only in terms of external knowledge.

What is meaningful, is what makes meaning with me. Human characteristics are largely constructed by audiences in any case. Humans will find it meaningful to form the same kinds of perceptions of character and emotion about machine rhetors, even if they are aware at a rational level of thought that they are "only" machines: if a connection is perceived then a connection exists. In the same way, if a responsibility is assigned then the responsibility exists and an AI that does not keep to its own dialectical commitments will fail to argue convincingly.

For the artificial generators of language and those who create and control them, there are many aspects of ethos and its application to AI texts which are worth considering, but these can be reduced to two questions, vital and pressing:

a. How can an understanding of ethos and its relation to machine rhetors be employed to improve AI communication?
b. Are persuasive machines actually desirable?

For wider society, this final question is likely to be the most engaging. Recent developments in chatbots have led to a media frenzy of opinion articles and various statements from researchers and industry figures about the dangers of AI and the need to control or even arrest its development. Genies, however, are notoriously difficult to rebottle. Artificial Intelligence is not going to go away, and more and more of our communication will be with non-human interlocutors. That being the case, the development of an understanding of what it means to be a rhetor and a machine, how language produced without thought can affect our own thinking, and how electronic personae can express ideas which play upon our deepest feelings, is an endeavour of great importance and greater urgency.

This paper has, perhaps, provided more questions than answers, but three clear steps upon the path to an ethos of machines have been taken. The role of the zero persona as an integral part of the machine's communication has been highlighted; the division of machine characteristics into those in common with humans, those parasitic on humans, and those distinct to machines, has been suggested; and the importance of the perception of the abilities, intentions, and responsibilities of machines, rather than their actual existence, in the construction of rhetorical situations has been discussed. On that last point, however, since it is in keeping with the Aristotelian idea of entechnic proofs, human-machine communication may not be so very exceptional.

References

Aristotle. [2018]. *Rhetoric*. Translated by C.D.C Reeve. Indianapolis: Hackett Publishing Company.

Bach, Tita Aissa, Amna Khan, Harry Hallock, Gabriela Beltrão, and **Sonia Sousa.** 2024. "A systematic literature review of user trust in AI-enabled systems: An HCI perspective." *International Journal of Human–Computer Interaction* 40 (5): 1251-1266. https://doi.org/10.1080/10447318.2022.2138826.

Bedor Hiland, Emma. 2024. "The Rhetorical Possibilities of Communicative Time Travel." *Rhetoric Society Quarterly* 54 (3): 263-271. https://doi.org/10.1080/02773945.2024.2343267.

Behdadi, Dorna, and **Christian Munthe.** 2020. "A normative approach to artificial moral agency." *Minds and Machines* 30 (2): 195-218. https://doi.org/10.1007/s11023-020-09525-8.

Bender, Emily, Timnit Gebru, Angelina McMillan-Major, and **Shmargaret Shmitchell.** 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922.

Bitzer, Lloyd. 1968. "The Rhetorical Situation." *Philosophy and Rhetoric* 1: 1-14.

Black, Edwin. 1970. "The Second Persona." *Quarterly Journal of Speech* 56: 109-119.

Bontridder, Noémi, and **Yves Poullet.** 2021. "The Role of Artificial Intelligence in Disinformation." *Data & Policy* 3: e32. https://doi.org/10.1017/dap.2021.20.

Boyle, Casey, James Brown, and **Steph Ceraso.** 2018. "The digital: Rhetoric behind and beyond the screen." *Rhetoric Society Quarterly* 48 (3): 251-259. https://doi.org/10.1080/02773945.2018.1454187.

Braet, Antoine. 1992. "Ethos, Pathos and Logos in Aristotle's Rhetoric: A Re-Examination." *Argumentation* 6: 307-320. https://doi.org/10.1007/BF00154696.

Cappelen, Herman, and **Josh Dever.** 2021. *Making AI Intelligible: Philosophical Foundations.* Oxford: Oxford University Press.

Cloud, Dana. 1999. "The null persona: Race and the rhetoric of silence in the uprising of '34." *Rhetoric & Public Affairs* 2 (2): 177-209. https://doi.org/10.1353/rap.2010.0014.

Cummings, Louise. 2014. "Informal Fallacies as Cognitive Heuristics in Public Health Reasoning." *Informal Logic* 34 (1): 1-37.

Egan, Matt. 2024. "AI could pose 'extinction-level' threat to humans and the US must intervene, State Dept.-commissioned report warns." *CNN*. Accessed May 22, 2025. https://edition.cnn.com/2024/03/12/business/artificial-intelligence-ai-report-extinction.

European Commission. 2021. "Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence." *European Commission*. Accessed May 22, 2025. https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682.

Esposito, Elena. 2017. "Artificial communication? The production of contingency by algorithms." *Zeitschrift für Soziologie* 46 (4): 249-265. https://doi.org/10.1515/zfsoz-2017-1014.

Ferrario, Andrea, and **Michele Loi.** 2022. "How explainability contributes to trust in AI." *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 1457-1466. https://doi.org/10.1145/3531146.3533202.

Flahive, Gerry. 2018. "The story of a voice: Hal in '2001' wasn't always so eerily calm." *The New York Times*. Accessed December 1, 2022. https://www.nytimes.com/2018/03/30/movies/hal-2001-a-space-odyssey-voice-douglas-rain.html.

Fonagy, Peter, and **Elizabeth Allison.** 2023. "Beyond Mentalizing: Epistemic Trust and the Transmission of Culture." *The Psychoanalytic Quarterly* 92 (4): 599-640. https://doi.org/10.1080/00332828.2023.2290023.

Fritz, Leslie M. 2018. "Child or product? The rhetoric of social robots." In *Human-machine communication: Rethinking communication, technology, and ourselves*, ed. Andrea Guzman, 67-82. New York: Peter Lang.

Giles, Howard. 2008. *Communication accommodation theory*. Thousand Oaks: Sage Publications. https://psycnet.apa.org/doi/10.4135/9781483329529.n12.

Gilovich, Thomas, Dale Griffin, and **Daniel Kahneman,** eds. 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press. https://psycnet.apa.org/doi/10.1017/CBO9780511808098.

Goldman, Alvin. 2001. "Experts: Which ones should you trust?" *Philosophy and phenomenological research* 63 (1): 85-110. https://doi.org/10.2307/3071090.

Grice, H. Paul. 1975. "Logic and Conversation." In *Syntax and Semantics, 3: Speech Acts*, eds. Peter Cole and Jerry L. Morgan, 41-58. New York: Academic Press.

Guzman, Andrea. 2018. "What is human-machine communication, anyway?" In *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves*, ed. Andrea Guzman, 1-28. New York: Peter Lang.

Guzman, Andrea, and **Seth Lewis.** 2020. "Artificial intelligence and communication: A Human–Machine Communication research agenda." *New Media and Society* 22 (1): 70-86. https://doi.org/10.1177/1461444819858691.

Hallsby, Atilla. 2024. "A Copious Void: Rhetoric as Artificial Intelligence 1.0." *Rhetoric Society Quarterly* 54 (3): 232-246. https://doi.org/10.1080/02773945.2024.2343265.

Hume, David. 1748/1977. *An Enquiry Concerning Human Understanding*. Indianapolis: Hackett Publishing Company.

Ji, Jiaming, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Kwan Yee Ng, Juntao Dai, Xuehai Pan, Aidan O'Gara, Yingshan Lei, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, Wen Gao. 2024. *AI Alignment: A Comprehensive Survey*. https://arxiv.org/abs/2310.19852v5. https://doi.org/10.48550/arXiv.2310.19852.

Jonker, Alexander, and **Alice Gomstyn.** 2024. *What is AI alignment?* Accessed May 22, 2025. https://www.ibm.com/think/topics/ai-alignment.

Koenig, Melissa, and **Paul Harris.** 2007. "The basis of epistemic trust: Reliable testimony or reliable sources?" *Episteme* 4 (3): 264-284. http://dx.doi.org/10.3366/E1742360007000081.

Lee, Seungcheol, and **Yuhua Liang.** 2016. "The role of reciprocity in verbally persuasive robots." *Cyberpsychology, Behavior, and Social Networking* 19 (8): 524-527. https://doi.org/10.1089/cyber.2016.0124.

Lee, Seungcheol, and **Yuhua Liang.** 2019. "Robotic foot-in-the-door: Using sequential-request persuasive strategies in human-robot interaction." *Computers in Human Behavior* 90: 351-356. https://doi.org/10.1016/j.chb.2018.08.026.

Lubin, Kem-Laurin, and **Randy Harris.** 2024. "Sex after Technology: The Rhetoric of Health Monitoring Apps and the Reversal of Roe v. Wade." *Rhetoric Society Quarterly* 54 (3): 247-262. https://doi.org/10.1080/02773945.2024.2343266.

Majdik, Zoltan, and **S. Scott Graham.** 2024. "Rhetoric of/with AI: An Introduction." *Rhetoric Society Quarterly* 54 (3): 222-231. https://doi.org/10.1080/02773945.2024.2343264.

Maleki, Negar, Balaj Padmanabhan, and **Kaushik Dutta.** 2024. "AI hallucinations: a misnomer worth clarifying." 2024 IEEE Conference on Artificial Intelligence (CAI), 133-138. https://doi.org/10.1109/CAI59869.2024.00033.

Manna, Riya, and **Rajakishore Nath.** 2021. "The problem of moral agency in artificial intelligence." *2021 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, 1-4. https://doi.org/10.1109/21CW48944.2021.9532549.

Martini, Carlo. 2014. "Experts in science: a view from the trenches." *Synthese* 191 (1): 3-15. https://doi.org/10.1007/s11229-013-0321-1.

McCraw, Bemjamin. 2015. "The Nature of Epistemic Trust." *Social Epistemology* 29 (4): 413-430. https://doi.org/10.1080/02691728.2014.971907.

Mercier, Hugo, and **Dan Sperber.** 2011. "Why do Humans Reason? Arguments for an Argumentative Theory." *Behavioral and Brain Sciences* 34 (2): 57-74.

Miller, Carolyn. 2007. "What Can Automation Tell Us About Agency?" *Rhetoric Society Quarterly* 37:137–157. https://doi.org/10.1080/02773940601021197.

Noor, Dharna. 2025. "Inside a plan to use AI to amplify doubts about the dangers of pollutants." *The Guardian*. Accessed August 22, 2025. https://www.theguardian.com/technology/ng-interactive/2025/jun/27/tony-cox-epidemiology-risk-assessment-chatgpt-ai.

Paglieri, Fabio. 2014. "Trust, argumentation and technology." *Argument and Computation* 5 (2-3): 119-122. https://doi.org/10.1080/19462166.2014.913262.

Payne, Kay, Joe Downing, and **John Fleming.** 2000. "Speaking Ebonics in a professional context: The role of ethos/source credibility and perceived sociability of the speaker." *Journal of technical writing and communication* 30 (4): 367-383. https://doi.org/10.2190/93U1-0859-0VC3-F5LK.

Peled, Yael, and **Matteo Bonotti.** 2019. "Sound reasoning: Why accent bias matters for democratic theory." *The Journal of Politics* 81 (2): 411-425.

Peters, John. 1999. *Speaking into the Air: A History of the Idea of Communication*. Chicago: University of Chicago Press.

Pruś, Jakub, and **Andrew Aberdein.** 2022. "Is Every Definition Persuasive?" *Informal Logic* 42 (1):25-47.

Reinares-Lara, Eva, Josefa Martín-Santana, and **Clara Muela-Molina.** 2016. "The Effects of Accent, Differentiation, and Stigmatization on Spokesperson Credibility in Radio Advertising." *Journal of Global Marketing* 29 (1): 15-28. https://doi.org/10.1080/08911762.2015.1119919.

Sætra, Henrik. 2024. "A Machine's ethos? An inquiry into artificial ethos and trust." *Computers in Human Behavior* 153: 108108. https://doi.org/10.1016/j.chb.2023.108108.

Sharf, Zack. 2018. *Douglas Rain, Voice of HAL 9000 in '2001: A Space Odyssey,' Dies at 90—Here's Why Stanley Kubrick Cast Him*. Accessed May 22, 2025. https://www.indiewire.com/2018/11/douglas-rain-dead-hal-9000-2001-a-space-odyssey-stanley-kubrick-cast-1202019828.

Sherman, Natalie. 2023. *Google's Bard AI bot mistake wipes $100bn off shares*. Accessed May 22, 2025. https://www.bbc.com/news/business-64576225.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Oliver Mascaro, Hugo Mercier, Gloria Origgi, and **Deirdre Wilson.** 2010. "Epistemic vigilance." *Mind & Language* 25 (4): 359-393. https://doi.org/10.1111/j.1468-0017.2010.01394.x.

Sun, Yujie, Dongfang Sheng, Zihan Zhou, and **Yeifei Wu.** 2024. "AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content." *Humanities and Social Sciences Communications* 11 (1): 1-14. https://doi.org/10.1057/s41599-024-03811-x.

Torkington, Simon. 2024. "These are the 3 biggest emerging risks the world is facing." *World Economic Forum*. Accessed May 22, 2025. https://www.weforum.org/stories/2024/01/ai-disinformation-global-risks.

Wander, Philip. 1984. "The third persona: An ideological turn in rhetorical theory." *Communication Studies* 35 (4): 197-216. https://doi.org/10.1080/10510978409368190.

Wang, Zhaozhe. 2024 "Post-Rhetoric: A Rhetorical Profile of the Generative Artificial Intelligence Chatbot." *Rhetoric Review* 43 (3): 155-172. https://doi.org/10.1080/07350198.2024.2351723.

Watson, Alex P. 2024. "Hallucinated Citation Analysis: Delving into Student-Submitted AI-Generated Sources at the University of Mississippi." *The Serials Librarian* 85 (5–6): 172-80. https://doi.org/10.1080/0361526X.2024.2433640.

Weiss, Debra Cassens. 2025. "AI-hallucinated cases end up in more court filings, and Butler Snow issues apology for 'inexcusable' lapse." *ABAJournal*. Accessed May 22, 2025. https://www.abajournal.com/news/article/ai-hallucinated-cases-end-up-in-more-legal-documents-and-butler-snow-issues-apology-for-inexcusable-lapse.

Weizenbaum, Joseph. 2000. "From computer power and human reason: From judgment to calculation." In *The new media reader*, eds. Noah Wardrip-Fruin and Nick Montfort, 368-376. Cambridge, MA: MIT Press.

Wells, Lindsay, and **Tomasz Bednarz.** 2021. "Explainable ai and reinforcement learning—a systematic review of current approaches and trends." *Frontiers in artificial intelligence* 4: 550030. https://doi.org/10.3389/frai.2021.550030.

Wolf, Marty J., Keith W. Miller, and **Frances S. Grodzinsky.** 2017. "Why we should have seen that coming: comments on Microsoft's tay 'experiment' and wider implications." A*cm Sigcas Computers and Society* 47 (3): 54-64. https://doi.org/10.29297/orbit.v1i2.49.

Zappen, James. 2005. "Digital Rhetoric: Toward an Integrated Theory." *Technical Communication Quarterly* 14 (3): 319-325.

Zierau, Naim, Christian Engel, Matthias Söllner, and **Jan Marco Leimeister.** 2020. "Trust in smart personal assistants: A systematic literature review and development of a research agenda." In *Entwicklungen, Chancen und Herausforderungen der Digitalisierung. Band 1: Proceedings der 15. Internationalen Tagung Wirtschaftsinformatik*, eds. Norbert Gronau, Hanna Krasnova, Key Pousttchi, and Moreen Heine, 99-114. Berlin: GITO Verlag. http://dx.doi.org/10.30844/wi_2020_a7-zierau.