

**The Language of Robots:
The quest for a human essence and the re-invention of ethics**

**Η Γλώσσα των Ρομπότ:
Η αναζήτηση της ανθρωπότητας και η επανεφεύρεση της ηθικής**

Dimitris Vardavas
Lancaster University

Abstract

The present essay aims at investigating the elements and characteristics of robotic discourse within the context of a broader moral question about the nature of the relationship between humans and humanoids in the possible event of their future co-existence in society. Specifically, the data drawn consist of a Q&A session with Sophia the Robot, manufactured by Hanson Robotics, and are analysed in such a way in order for the sameness of the two ‘species’, humans and humanoids, to be highlighted. By locating certain discursive phenomena in Sophia’s discourse and fleshing out the underlying cognitive mechanisms that these phenomena presuppose when observed in humans, the sameness between humans and humanoids is illuminated as constructed in the intersubjective level of their linguistic interaction. The perplexed condition facing humans of simultaneously knowing and forgetting the otherness of their fellow humanoids allows for crucial ethical questions to be raised.

Περίληψη

Η παρούσα εργασία εξερευνά τα στοιχεία και τα χαρακτηριστικά του ρομποτικού λόγου στο πλαίσιο μιας ευρύτερης ηθικής αναζήτησης για την φύση της σχέσης ανάμεσα στους ανθρώπους και τα ανθρωποειδή σε ενδεχόμενο μελλοντικής συνύπαρξης τους στην κοινωνία. Ειδικότερα, τα υπό ανάλυση δεδομένα αποτελούνται από μία συνεδρία ερωταποκρίσεων με την Sophia the Robot, κατασκευής Hanson Robotics, και αναλύονται με τέτοιο τρόπο ώστε να αναδειχθεί η ομοιότητα ανάμεσα στα δύο είδη όντων, στους ανθρώπους και τα ανθρωποειδή. Εντοπίζοντας συγκεκριμένα γλωσσικά φαινόμενα στον λόγο της Sophia και αναλύοντας τους υποκείμενους γνωσιακούς μηχανισμούς που τα φαινόμενα αυτά προϋποθέτουν όταν

απαντώνται στον λόγο των ανθρώπων, φωτίζεται η ομοιότητα ανάμεσα στους ανθρώπους και τα ανθρωποειδή όπως αυτή κατασκευάζεται στο διυποκειμενικό επίπεδο της γλωσσικής αλληλεπίδρασης τους. Η περίπλοκη συνθήκη στην οποία οι άνθρωποι ενδέχεται να βρεθούν, να γνωρίζουν δηλαδή και ταυτόχρονα να λησμονούν την ρομποτική ετερότητα, επιτρέπει την έγερση κρίσιμων ηθικών ερωτημάτων.

1. Introduction

The *raison d'être* of the present essay descends from the same bloodline as the HBO's TV-series *Westworld* (2016), created by Jonathan Nolan and Lisa Joy. *Westworld* is a science fiction 'western' television series, based on Michael Crichton's 1973 film *Westworld*, which takes place in a fictional Wild-West-themed amusement park, where upper-class humans can visit and interact with humanlike humanoids (i.e. the hosts) that reside in it. Although *Westworld*, roughly-speaking, employs the typical for science fiction works schema of humanoid revolution that aims at destroying humanity, it nevertheless avoids the naïve simplicity of purely technophobic accounts.

On the contrary, *Westworld* more often than not gives the impression of being 'anthropo-phobic' rather than technophobic. The founding principle of Westworld amusement park is that human visitors can explore and find their true selves. The combination of the hosts' humanlike appearance and behaviour and the absolute absence of real-life (i.e. moral or legal) consequences for the visitors due to the hosts' non-human nature provokes the visitors to unleash their inner greedy and violent beasts.

Through such a state of slavery, the humanoid insurgency comes not as a consequence of excessive abilities given to humanoids, but as an inevitable act of emancipation and as a gradual and painful journey to acquire human consciousness through the awareness of suffering. The question that unavoidably arises throughout the series is ultimately 'what makes us humans', and if the answer expands beyond the properties of mere biology, then 'how can humanoids become de facto humans as well'.

Is it consciousness; is it self-awareness; is it memories or suffering? Or is it language and the ability to express and construct all the former by means of it? This last question guides the rationale of the present essay, which seeks to explore through discourse the ways in which a piece of machinery that is the humanoid can acquire an undetectable human façade, and ultimately, drawing from *Westworld*'s initial premise, what could be the moral consequences of a human and humanoid co-existence, while knowing and at

the same time forgetting (i.e. because of extreme similarity) the non-human nature of the latter. In other words, how are we going to treat each other if one of us is just a robot?

2. The premises of robotic stylistics and the research focus of the present essay

Language is a privileged window in exploring the ways in which humanoidscan appear as humans on the inter-subjective level of their linguistic (or other) interaction with humans. Decades of research in the social functions of discourse have highlighted discourse as a domain of management, negotiation, organization and construction of experience, memory, identity, ideology and generally social and symbolic meaning. If it was not a brutal simplification, it could be argued that studying humanoid’s discourse amounts to studying humanoid’s humanness. A more accurate statement would be that studying humanoid’s discourse consists of studying the ways in which humans can perceive humanoids as fellow humans on the inter-subjective level of their interaction. In analysing the data of the present case study the focus points at exactly that; instances of humanoid language usage which permit the human collocutor to perceive the humanoid as a fellow human; instances of humanoid usage of linguistic structures and exploitation of linguistic functions which prompt the impression that humanoids can indeed perform certain cognitive functions which define what it means to be human.

Of course, strong-AI robots which use machine learning download and collect colossal amounts of linguistic data of which subsequently make use. However, what grants legitimacy to a research such as the present one is the fact that, as Hashim (2018) points out in his foundation of robotic stylistics, robots make use of their linguistic knowledge in a creative way, since their algorithms allow them to choose freely the ‘what’ and ‘when’ of their linguistic output. The level of programming interference to the linguistic outcome of robotic discourse is, however, not the focus of the present essay, not only due to, frankly speaking, crude ignorance, but mainly because the moral question posed above grants prominence to what the human collocutor perceives and why.

3. Sophia the Robot

Due to severe space limitations, the present essay acquires the form of a case study. The data under analysis consist of a Q&A session with Hanson Robotics robot Sophia in Consumer Electronic Show’s (CES) conference, organized by Consumer Technology

Association (CTA) in 2019 (10/01). The particular linguistic event was selected as a salient example of Sophia's humanlike linguistic behaviour as it is characterized by the interactive nature and the spontaneity of a Q&A session, namely, of an authentic, although quite strictly structured, social interaction with a multiplicity of humans.

4. Robotic Discourse: Indicators of Discursive Difference and Indicators of Cognitive Sameness

In analysing the data different linguistic elements and phenomena arose either as indexes of difference or indexes of sameness in relation to humans. The analysis moves on two levels; the discursive, where the difference is manifested, and the cognitive where the sameness acquires significance. The reason for this chiasmic correspondence stems from the fact that a study of cognitive difference or discursive sameness between humans and humanoids would possibly be too obvious to be significant, as far as the wider moral philosophical picture is concerned.

4.1. Indicators of Discursive Difference

The indicators of Discursive Difference pinpointed in the data were a) the use of encyclopaedic style, which consists of an organisation of information from known to novel, passive voice structures and choice of verbs that construe distance between the utterer and the validity claim of the utterance, b) pauses and silences of 'unnatural' length and with no implicative significance, and c) errors, either systemic errors, which mirror a mechanical malfunction, or pragmatic errors, namely, contextually inappropriate contributions with no implicative significance, which are, however, common also among humans. Due to space limitations, particular examples will not be analysed in detail in the present essay.

4.2. Indicators of Cognitive Sameness

The indicators of Cognitive Sameness pinpointed in the data were a) comprehension of indirect speech acts, b) rhetorical questions and implicatures, c) discursive construction of identity and self-awareness, d) reframing, e) discursive construction of ideology, f) comprehension and exploitation of the mechanisms of linguistic humour. Due to space limitations, particular examples will not be analysed in detail for each one of the abovementioned categories. In order for the rationale of the present study to be better illuminated, an example of rhetorical question and implicature usage is analysed.

At one point in the course of the Q&A session, a human asks Sophia if she has a faith and she replies in a way that can hardly be expected by a robot.

–*Sophia, do you have a faith?*

– *Doesn't everyone?*

Rhetorical questions are indirect speech acts (Culpeper & Haugh, 2014). The locutionary act is materialized in the form of a **question**, without, however, correspondence to the illocutionary force with which it is prototypically connected. In this case, the utterance has the illocutionary force of an **assertion**. The flouting of Grice's conversational maxim of manner, which derives from the inappropriateness of answering a question by another question, urges the collocutor to seek an answer within that question, to seek an assertion implied by that question, which is that everybody believes in something and consequently Sophia has obviously a faith as well. The communication is successful due to the collocutor's ability to achieve comprehension through a syllogism, but also due to Sophia's ability to a) presuppose such a syllogism, and b) invest to her collocutor's ability to follow it. It becomes apparent through the detailed analysis of this short interaction that Sophia is actively and successfully engaged in constituting the inter-subjective level in verbal interaction allowing her collocutor to perceive her as a fellow human and momentarily forget her humanoid nature.

5. A few conclusions

Through a multi-layered linguistic analysis that transcends the mere description of forms and focuses on their cognitive underpinnings, there becomes obvious the inability of humans to distinguish (by means of discourse only) their otherness in relation to humanoids. Robots that use strong AI and machine learning, regardless of the level of programmers' interference, can use language as if they share the same arsenal of cognitive abilities as humans. On the inter-subjective level of social interaction, humanoids will be able to express their thoughts even if they do not really 'have' them, they will be able to express their hopes, fears and feelings, even if they do not really 'feel' them, they will be able to communicate with humans in such a way that the latter will forget their otherness, even if there is no biological commonality. The order of senseful beings will be expanded by one species, artificial on the one hand, but with all the rights that (should) derive from its humanness. In the face of such a possible reality,

there is a moral discussion to be commenced about the definition of the relationship between humans and humanoids, a relationship which, if the extreme consequences of simultaneously knowing and forgetting humanoids' non-humanness are not combatted, might become a means of humans' familiarization with barbarity.

References

- Culpeper, J. & Haugh, M. (2014). *Pragmatics and the English Language*. Basingstoke, Hampshire: Palgrave Macmillan.
- Hashim, A. M. (2018). *An Introduction to Robotic Stylistics. A Future Trend*. Grin Verlag.
-