

Tilburg University

Narrating artificial intelligence

Keymolen, Esther

Published in:
Interpreting technology

Publication date:
2021

Document Version
Peer reviewed version

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Keymolen, E. (2021). Narrating artificial intelligence: The story of AlphaGo. In W. Reijers, A. Romele, & M. Coeckelbergh (Eds.), *Interpreting technology: Ricoeur on questions concerning ethics and philosophy of technology* (Philosophy, Technology and Society). Rowman & Littlefield Publishers, Inc..

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Chapter 14

Narrating Artificial Intelligence: The Story of AlphaGo

Esther Keymolen

From self-driving cars to voice assistant Alexa, artificial intelligence (AI) is currently perceived as one of the most promising technologies that is driving innovation. While the field of AI has already been up and running for more than 65 years, recent break-throughs in foundational techniques such as machine learning, the use of neural networks, and natural language processing have put it in the spotlight (Crawford and Whittaker 2016, 2). AI is defined as “the theory and development of computer systems able to perform tasks normally requiring human intelligence” (Jobin, Ienca, and Vayena 2019, 389). The European Commission refers to AI as: “systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.” The commission further adds that: “AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).

What these definitions show is that AI can become an integral part of both the online and offline world, if such a distinction can still be made (The-Online-Initiative 2015). In other words, AI is on its way to becoming an ontological part of the objects and environments we live our lives with and in. And while there is debate on how and to what extend technologies mediate and shape human beings, their relations, and the world they inhabit, there does appear to be some –rare–consensus amongst philosophers that technologies are not merely ‘mute’ instruments. Technology shapes, it matters, and it deserves our attention. This inherently makes AI an utterly important research subject for philosophers of technology.

One important prerequisite for such an endeavor is that philosophers have some kind of access to the technology. Located after the so-called “empirical turn” (Achterhuis 2001), many approaches such as postphenomenology (Ihde 1990), critical constructivism (Feenberg 2009, 2017), and Actor-Network Theory (Latour 1993), have been geared to investigating real-life examples and case-studies. In particular, postphenomenology, with its emphasize on the first-person perspective and the mediating relations opening up everyday life, has increasingly

manifested itself as a case-based, empirically-informed approach (Rosenberger and Verbeek 2015, Aagaard et al. 2018). One of the basic assumptions in postphenomenology is that technologies ‘open up’ the world for human beings in specific, concrete ways and that we can describe these different variations or stabilities in order to understand how human beings perceive of themselves, of others, and of the world around them.

1. How AI Challenges Postphenomenology

At first sight, one could think that this focus on the contextual, every-day use of technology ensures that postphenomenology and AI are a perfect fit. After all, if AI is well on its way to becoming a ubiquitous technology, then investigating how it mediates everyday-life practices seems not only highly relevant but also quite feasible. Nevertheless, AI developments impose certain challenges on the postphenomenological approach which need careful consideration.

First of all, as postphenomenology predominantly focuses on how technologies mediate *everyday life*, the technologies that are central to such analyses are oftentimes mundane objects and/or consumer products. However, not many AI applications already made it into consumer products and the few that exist are not yet highly sophisticated. Oftentimes they are still in a beta-phase, have high error rates, and only possess limited functionality. This obviously does not make them unworthy of investigation. We can still learn a lot from these first-generation commercial AI applications. However, they might also blindside us, as they provide only a limited and maybe even discarded image of what AI can bring about.

This immediately brings us to the second challenge, as the fact that every-day-life AI is not yet highly advanced, does not mean that the progress and developments, that are currently being made behind the scenes, do not warrant philosophical reflection. In other words, don’t we run the risk of ending up with an outdated and toothless philosophy of technology if we only focus on the everyday context of technologies-in-use?

Thirdly, there is the so-called ‘black box’ problem (Pasquale 2015). Even if we would be able to gain access to relevant AI applications, that does not imply that we also understand what we perceive. This can have different causes. Sometimes, the AI application is shielded for proprietary reasons. For instance, when the companies behind the applications do not want to provide too much information which could negatively impact their market position. It might also be that the AI application is opaque due to inherent, technical complexity. Moreover, in general, AI applications operate in intricate networks of human and non-human actors. Behind

a sleek interface resides software, hardware, databases, servers, AI companies, advertisement companies, intelligent services, hackers, etc.

Considering these challenges, it becomes clear that if postphenomenology already in an early stage wants to engage with these upcoming technologies, new ways of gaining access to these technological practices are needed.

It is important to note that postphenomenology is not a static approach and that the scholars in the field certainly cannot be accused of conservatism. Doing justice to its pragmatism roots, there definitely is room to explore and reinterpret the postphenomenological toolbox as it has been initiated by founding father Don Ihde (1990). Over the last couple of years, several researchers have come up with new methodological means to further develop the postphenomenological approach.

Some scholars have combined postphenomenology with Actor-Network Theory to better address the socio-technical context in which technology mediates everyday life (Rosenberger 2018, Keymolen 2017). Others have incorporated elements of the philosophy of Foucault to bring the idea of power more to the foreground (Verbeek 2011). Another very promising direction that currently is being explored is the intensification of the empirical approach, for instance by building on ethnography (Kudina and Verbeek 2019) and investigating the relation between postphenomenology and existing empirical methods (Aagaard et al. 2018).

While some of the listed challenges connected to investigating AI might be mitigated by turning to one of these approaches –e.g. ANT might be a fruitful way for investigating the networks in which AI applications are embedded and empirical research in the form of focus groups or in-depth interviews with experts might be a way of bridging the knowledge gap–, I want to take the opportunity to explore another avenue.

This chapter will aim at broadening the scope of what ‘empirically-informed’ generally comes down to in postphenomenology, by investigating to what extent Ricoeur’s narrative discourse might be useful to interpret and gain access to new technological developments (Ricoeur 1991). Instead of focusing on the first-person experience of every-day life, mediated by technological artefacts, it will take, what Ricoeur would refer to as, a necessary ‘detour’ to explore the stories we tell about AI and how, in these stories, we imagine AI to become part of our everyday life.

This narrative approach is not meant to replace the more ‘traditional’ postphenomenological analysis or the empirical approaches described above. Rather, it should be seen as a new addition to the continuous expanding methodological toolbox of

postphenomenology. In particular it could serve well as a strategy to engage with the increasing number of stories that surface on the development of AI. AI is presented to us in movies, TV series, books, performances, and –as will be the case in this chapter– documentaries. These stories generally fall out of the scope of the other approaches as they are not perceived as ‘empirical,’ ‘phenomenological,’ or ‘focused on everyday-life.’ This chapter will show that engaging in a hermeneutical interpretation of these stories is however not only valuable in itself, but can also better equip us to undertake a postphenomenological analysis.

Central to this chapter will be the documentary “AlphaGo” (Kohs 2017), which narrates one of the greatest breakthroughs in the field of AI. AlphaGo is the name of an algorithm that learned how to play the very complex and ancient Chinese board game Go and was developed by DeepMind, an AI company later acquired by Google. By engaging with this story, we should become able to get a better understanding of state-of-the-art AI and read ourselves into a context where we interact with its transformative power, before it is even part of everyday life (Kaplan 2006). Moreover, as the documentary’s goal is to share information on AI in an accessible manner, it might prove to be a good way to open at least some parts of the AI black box.

This chapter will first provide an overview of some of the key concepts of Ricoeur’s narrative discourse. This will be followed by a short methodology section and a synopsis of the documentary. Next, the actual analysis of the documentary will take place, by making use of Ricoeur’s key concepts and highlighting the relevance for a postphenomenological analysis. The chapter will end with a concluding section, summarizing the most important outcomes.

2. Ricoeur’s Narrative Discourse

One key starting point in Ricoeur’s work—which it actually shares with postphenomenology—is that human life cannot be directly known or understood, but that it is always mediated. And for Ricoeur, it is particularly mediated through *language*. However, as spoken language can be short-lived, Ricoeur predominantly turns to written texts or analogue forms that capture language, such as, in our case, a documentary. By focusing on ‘captured’ language instead of spoken language, how the text becomes interpreted –its meaning– can become detached from the intentions of the actor who initiated the words.

Although Ricoeur did not come up with a systematic, general theory of interpretation, in his work he did develop a variety of fruitful concepts and techniques that can enable a rich, hermeneutic interpretation of texts. Particularly, his narrative discourse which focuses on the

understanding of actions appearing in a certain place and timespan mediated by texts, (both actions and texts understood in a broad sense) is relevant to the analysis of AlphaGo.

Key Concepts

How the reader reads herself into the texts, interprets its structure and genre can all be seen as a form of “appropriation” (Pellauer 2016). By making the text ‘ours,’ we are not just interpreting text, we are trying to make sense of the world (*referentiality*), to relate to others (*communicability*), and to get to know ourselves (*self-understanding*) (Ricoeur 1991, 27). This subjectivity also entails that there is no such thing as a finite interpretation. Interpretations need to be checked and contrasted with other interpretations and over time, interpretations might lose their explanatory force, making room for new ones (also see: Gadamer 1972).

A Story

“A story,” Ricoeur (1981, 239) claims,

describes a sequence of actions and experiences of a certain number of characters, whether real or imaginary. These characters are represented in situations which change or to the changes of which they react. These changes, in turn reveal hidden aspects of the situation and the characters, give rise to a new predicament which calls for thought or action or both. The response to this predicament brings the story to its conclusion.

What ‘makes’ a story is however not the mere succession of events leading to an acceptable conclusion. A story is just as much about the expectations one has while engaging with texts (both as a narrator and reader), actively bringing together bits and pieces to make a meaningful whole. A story is characterized by an active coming together of a chronological and a non-chronological dimension. The former, Ricoeur refers to as the *episodic dimension*, the latter is the *configurational dimension*. Both in *the art of narrating a story* as well as in *the art of following a story* these dimensions play an important role and actually are intrinsically intertwined (Ricoeur 1981, 241). After all, just as even the simplest narrative is more than merely the succeeding episodes leading up to a conclusion, demanding some act of interpretation, of finding a closure, also constructing meaningful totalities out of a story cannot be done if one would completely abolish its narrative structure.

This two-dimensional distinction actually resonates – to a certain extent – with the idea of multistability, one of the core, postphenomenological concepts that refers to the ontological openness of artefacts to different interpretations and their ability to furnish different relations.

Just as a story is more than its chapters, scenes, or episodes, a technology is also more than the sum of its screws and bolts, of its software and hardware. Just as people have to read themselves into the text to interpret and give meaning, people co-shape the mediating relations with their devices. Just as a text can have a different meaning to different people, so can technology.

Genre

It goes without saying that not all stories are the same: a theatre piece is not the same as a book, a comic is different from a novel. Also, a documentary as AlphaGo is situated in a specific genre domain. It is non-fiction, which generally is understood as ‘representing reality.’ However, that it refers to reality does not make it a mere copy of reality either. Every non-fiction is also in a certain way fiction. The documentary in itself is an audio-visual artefact adhering to independent rules and structures which are inherent to the genre (therefore, fiction), however, it is simultaneously also referring to something outside of itself, to actions taking place in the real world (therefore, non-fiction). In other words, both fiction and non-fiction as *productive imaginations and representations* prescribe a new reading of reality, making the life-world appear under specific circumstances.

The fact that we are able to pinpoint the genre of a text is because it is embedded in a tradition that is located between *innovation* and *sedimentation*. We recognize characteristics which allow us to categorize a piece of work as a documentary, thriller, or tragedy. However, this is never an exhaustive description. It always remains a singular work, which allows for new relations with the tradition (Ricoeur 1991, 23-24). It is the variation between the two poles of innovation and sedimentation, that in the case of the documentary will result in a productive representation.

Emplotment

A plot is a key part of a story. First, it “serves to make *one* story out of the multiple incidents” (Ricoeur 1991, 21). Secondly, as a plot also gathers all kinds of heterogeneous components, from actors who perform actions to those who suffer them, from well-planned interactions to unintended consequences, it is both “concordant and discordant” in nature (Ricoeur 1991, 21). Finally, the plot mediates the episodic and configurational dimension of the story. It is where the competition between succession and configuration becomes explicit.

Because we are, through culture and tradition, familiar with different types of plots, we can also learn from them. We can relate a plot to certain virtues or “forms of excellence” (23). Ricoeur speaks of “phronetic understanding” to emphasize its practical character (23). It is not

so much about understanding the techniques or rules which a plot follows. Rather, the significance of a text lies in the “*intersection of the world of the text and the world of the reader*” (26). It relates more to lived experience than to theoretical knowledge.

It is not merely in the text, but also in the reader, in the act of reading that the *emplotment* takes place. It is where hermeneutics and phenomenology come together and “where narrative and life can be reconciled with one another” (26), as “reading is itself already a way of living in the fictive universe of the work.” Or in other words, “stories are recounted but they are also *lived in the mode of the imaginary*” (27).

A text is not something closed in upon itself, it is the projection of a new universe distinct from that in which we live. To appropriate a work through reading is to unfold the world horizon implicit in it which includes the actions, the characters and the events of the story told. As a result, the reader belongs at once to the work’s horizon of experience in imagination and to that of his or her own real action (Ricoeur 1991, 26).

3. Ricoeur and AlphaGo

AlphaGo is a 2017 documentary directed by Greg Kohs (Kohs 2017). In this movie, viewers follow the developments of the AI application AlphaGo: a computer program developed by the AI start-up DeepMind (owned by Google) to play the ancient board game called Go.

In the AI community, the board game Go has long been considered as one of the biggest challenges for AI to master. While the rules of the game might be rather straightforward, –two players take turns placing either black or white stones on a 19 by 19 grid board with the aim to capture the opponent’s stones or surround empty space to make points of territory (Silver and Hassabis 2016) – the game itself is: “mind-bogglingly complex– far more complex than chess. A game of 150 moves (approximately average for a game of Go) can involve 10^{360} possible configurations” (Granter, Beck, and Papke Jr 2017, 619). Playing Go, there are more possibilities “than there are atoms in the universe,” the DeepMind blog emphasizes (Silver and Hassabis 2016). Because of this infinite possibilities, Go is –much more than chess– associated with creative and imaginative thinking, something that is thought of as being difficult, if not impossible, for AI to achieve (Bory 2019, 631).

In order to nevertheless successfully try to reduce this complexity, AlphaGo utilizes two deep neural networks: a policy network to provide the probability of certain moves and a so-called value network that delivers a position evaluation (Silver et al. 2017).

First, AlphaGo engages in supervised learning to train the policy network. It uses human expert games to learn how Go is played. Next, by reinforcement learning, the algorithm also plays millions of games against an instantiation of itself to improve its value network. (Li and Du 2018, 78). Based on the combination of these trained networks, AlphaGo searches for the proper moves by anticipating possible future states (a so-called lookahead search) and predicts the most-likely successful state.

AlphaGo looks ahead by playing out the remainder of the game in its imagination, many times over - a technique known as Monte-Carlo tree search. But unlike previous Monte-Carlo programs, AlphaGo uses deep neural networks to guide its search. During each simulated game, the policy network suggests intelligent moves to play, while the value network astutely evaluates the position that is reached. Finally, AlphaGo chooses the move that is most successful in simulation. (Silver and Hassabis 2016)

After training on 160,000 recorded games of professional human Go players and then playing more than 30 million games against versions of itself, AlphaGo was ready for its first match against a human professional player. In October 2015, we witness AlphaGo beat Fan Hui, a European Go champion with 5-0. After recovering from this defeat, Hui joins the DeepMind team to help further train AlphaGo. Subsequently, in March 2016, a new historical match is organized: AlphaGo will play Go against Lee Sedol, one of the world's best Go players. Now, this match, again consisting out of five games, receives wide media attention and ends –much to the surprise of both the Go and the AI community– in a victory for AlphaGo. The AI wins with 4-1.

The Art of Reading

For my analysis of the documentary, I proceeded as follows: I watched the documentary itself four times. The first time was one year ago. At that time, I did not anticipate writing about the documentary. This viewing was therefore, to a certain extent, pre-theoretical. Obviously, it is not possible to put all knowledge and experience between brackets, but while watching, I did not actively investigate how the content of the documentary resonated with my background knowledge. This first viewing of the documentary was, nevertheless, important as it ‘set the scene’ for the analysis. It allowed me to familiarize myself with the overarching narrative and storyline, to gain a first understanding of the plot and the actors involved. While watching but also after the documentary ended, I remember experiencing feelings of excitement, wonder,

empathy, and also resistance and sadness. The documentary touched upon topics which were meaningful to me and sparked my interest; it became clear that there was something at stake.

The second viewing had the explicit goal of identifying themes, ideas, starting points to engage in a hermeneutic analysis of the documentary. I made notes, paused the documentary frequently, played back some parts, etc. The third viewing took the form of multiple close-reading sessions of the transcript of the documentary. In this way I was able to further elaborate my preliminary analysis of the second viewing. The fourth viewing was again a total viewing of the documentary in order to loosely validate and further fine-tune my analysis that came out of the previous rounds.

In addition to the multiple rounds of viewing (and reading), I also engaged with critical texts, comments, and analyses of the documentary and academic literature on AI. This took place after the second viewing and enabled me to develop a richer interpretation of the documentary.

Emplotment in AlphaGo: Game as a Structuring Narrative

Following Ricoeur, we know that the way in which a story is structured, is not neutral. It pre-sorts our understanding and interpretation of a specific subject. The dominant narrative of the documentary is without any doubt shaped by the game Go itself. By using the game as an overarching structure, actors immediately are assigned specific and recognizable roles: they are participants, opponents, winners and losers, judges, commenters, and supporters. This game narrative also allows for a clear delineation of the action space by introducing specific rules, attributes, and expectations.

Choosing for the game as organizing structure of the documentary is telling on different levels. First of all, game and play are often seen as a key aspect of human life (Frissen et al. 2015). As Huizinga (1955, 173) in his classic book *Homo Ludens* states: human culture “arises in and as play, and never leaves it.” Following Huizinga, play is characterized by the fact that it is an activity, human beings freely engage in. It is different from their ordinary, everyday life, as it is not spurred by the need to fulfill basic, everyday needs such as finding food or shelter. Rather, one engages in play for the sake of play itself. Taking place within a specific timeframe and on a specific location, with its own rules and order, people who play are absorbed by the game, they are completely dedicated to it, often experiencing profound feelings of excitement, joy, fear, and tension (also see Gadamer 1972). All in all, to play a game is to enter a “sacred sphere” (Huizinga 1955, 9).

Huizinga explains that many forms of living –from developing rituals, poetry, to religion and warfare– are built up on patterns of play. As play is such a fundamental form of social life, it is decisive that in the documentary, it is in the context of a game, human and AI ‘meet’ and ‘get to know’ each other. Following Huizinga, this element of play will, however, not just vanish when the game is over. It will put a long-lasting, decisive mark on future human-AI interactions and the societal practices that will be established within the AI domain.

While Huizinga emphasizes the distinction between everyday life and play –he refers to the “magical circle” people step into when playing– I do not perceive this distinction to be an impermeable hiatus. Rather, from a first-person perspective, we can be immersed in play, while also becoming aware of that ‘immersiveness.’ In other words, we can be in-play while also being aware of how this play relates to the world around us. It is our double, eccentric position (cf. Plessner 1975), our reflexivity that gives room to emplotment in a game. It is where the act of play and life come together, where the role we play and the life we live touch each other and a new light becomes shed on our perceptions, beliefs, and actions. At that moment, game truly mediates life. In the documentary, Frank Lantz, Director NYU Game Centre explains:

Go is putting you in a place, where you’re always at the very farthest reaches of your capacity. There’s a reason that people have been playing Go for thousands and thousands of years, right? It’s not just that they want to understand Go. They want to understand what understanding is. And maybe that is truly what it means to be human. (Kohs 2016, 00:01:05)

Secondly, play and game are not merely something closely connected to human life, they are also inherently intertwined with technology, particularly with AI. In order to build AI, one needs input of the real world, so-called training data. The real world, however, is a messy place and as a result training data can be biased, multi-interpretable, or too complex to distill meaningful patterns from. This can make it hard to train AI in a successful way. By focusing on games, one takes a very specific, well-defined snippet of real life. On the one hand a game is rule-based, making it relatively easy to familiarize AI with the action-space in which it has to perform. On the other hand, a game like Go gives access to a specific form of human intelligence, bringing together logic and creativity, both seen as essential to –in the end– develop strong AI. The game as a specific representation of the world is therefore instrumental in building AI.

Thirdly, there is also a commercial interest in choosing for the game narrative. As a game is distinct from everyday life, it provides a safe space to experience the power of AI. AI can ‘win’ without immediately threatening the familiar world as it does not *directly* have a tangible impact on that world. By presenting AI’s progress to the world as a game, DeepMind is able to: “gain trust in its new products, turning the old imaginary of a new intelligence as a potential Frankenstein’s monster into the narrative of togetherness, in which AI is an essential partner for the progress of humankind” (Bory 2019, 639).

4. Sedimentation and Innovation in AlphaGo

Now that we have established that ‘game’ is the overarching structure of the documentary’s narrative, we now move to the interpretation of the documentary itself. Following Ricoeur we know that in the art of reading, we are always maneuvering between the poles of sedimentation and innovation. Our unique and meaningful understanding of a story cannot materialize without the background knowledge guiding our reading. In reading we are not just focused on the story itself but on the web of familiar concepts and traditions in which the story is embedded. Such a first important element of sedimentation can be found in the historical sequence of games between man and machine, in which the documentary implicitly is embedded. In 1997, there was the chess game between IBM’s Deep Blue and the then world champion Kasparov. Deep Blue won. In 2011, IBM’s Watson defeated two champions –Ken Jennings and Brad Rutter– in the questions-based game Jeopardy. And in 2016, there is the match of AlphaGo vs. Lee Sedol.

Man vs. Machine Games

All these games between technology and human beings are to display, prove even, that in a domain, which is generally thought of as being dominated by human superiority, technology can outclass the human intellect. This antagonistic setting brings forth the frame of the human player and the technological player as rivals, opponents, enemies. With every lost game, beliefs about human capacities, human uniqueness, and the place of the human in the world are put under pressure. With Ricoeur, we can wonder how much of this framing will impact our future perceptions and interactions with AI.

In addition to this dominant, antagonistic framing, also some other shared characteristics can be identified, particularly in the AlphaGo and the Deep Blue match. On both occasions there was: a leading Western tech company involved, a non-Western world

champion as human opponent, and both games attracted a lot of media attention. (Bory 2019, 631). This media attention plays an important role in the AlphaGo documentary. The recurring press conferences and TV and internet commentators which appear in the documentary serve as an important source of interpretation. As many people might still be new to the game of Go, the media literally tell them what they are looking at.

Research into the overlap and differences between Chinese and American coverage of the event indicates that overall, when AlphaGo was framed as human by, for instance, describing it by making use of human qualities such as ‘intuitive’ or ‘creative,’ this often brought along questions on what it actually means to be human. Covering and co-experiencing the unexpected performance of the AI opened up a conceptual space to reflect on the porous boundaries between human and machine. This fundamental discussion took place both in the American and the Chinese media (Curran, Sun, and Hong 2019).

A difference between the American and Chinese media interpretation, however, was that the Chinese press made use of the ‘non-threat’ frame much more than the American press. While the Chinese press acknowledged the impact of AlphaGo’s performance on the assumptions of what it means to be human and what it means to play Go, they overall did not assess this as a danger. The researchers suggest that this might be traced back to different cultural backgrounds –or macroperceptions in postphenomenological terms–. The Chinese public might be more familiar with Go, as the game is part of their education. They might also be more open to AI because of a less apocalyptic, religious tradition –such as Christianity in the Western world– and a different history of science fiction (Curran, Sun, and Hong 2019, 5-6).

This kind of findings emphasizes the importance of connecting macroperception and microperception in postphenomenological research in AI. While the dominant frame of man vs. machine may significantly impact the mediating relations that can be developed with AI tools and services, the macro-context in which these relations are constructed, should not be overlooked. An overarching framing like the man vs. machine perspective can still have different meanings and normative connotations due to different macroperceptions. While Ihde has always emphasized that microperception and macroperception are inherently intertwined, he does seem to suggest that they should be analyzed in different ways: postphenomenology is best suited to look into the microperception, whereas the macroperception asks for a cultural hermeneutics approach (Ihde 1990, 30). In interpreting the AlphaGo documentary, it becomes clear that a phenomenological analysis can highly benefit from such a hermeneutical approach.

This raises the question if a postphenomenological analysis should not be executed in tandem with a hermeneutic analysis.

Next to the elements of sedimentation, there are also elements of *innovation* to be found in the documentary. While AlphaGo is embedded in this canonical sequence of man vs. machine games, it also differs from its predecessors. The documentary reinterprets the man vs machine framing in three distinct, innovative ways (Bory 2019, 632). First, whereas Deep Blue was a hardware-based machine with enormous calculating force, AlphaGo was presented as a body-less, self-learning intelligence. Secondly, where the functioning of Deep Blue was hidden from its opponent and the public, the tech team behind AlphaGo went to great lengths to describe the way the AI operates. Thirdly, where Deep Blue was predominantly described as mimicking human intelligence, AlphaGo is characterized as going “beyond its human guide,” coming up with “something new, and creative, and different” (Kohs 2017, 00:50:48).

The Tragedy of AlphaGo

Another important element which helps us to interpret the documentary is its *tragic-like emplotment*. Resonating with the old Greek tragedies –as another form of sedimentation–, we can recognize Lee Sedol as a tragic hero. He enters the game freely, cheered along by the whole Go community, together with the media who watch and comment on this historic game, much as if they take on the role of the choir in old Greek tragedies. Sedol is –overly– confident that he is going to win, but then fate strikes.

Also, the antagonistic setting of man vs. machine confrontations in which the Alphago-Sedol game is embedded, resonates with the Greek tragedies of grand conflicts. With every lost game, he comes to realize that it was *hubris* which led him to accept the challenge. He ignored the signs: the previous man-machine games which were all won by the machine (including the most recent game of AlphaGo against European champion Fan Hui), the overall disruptive developments in the AI domain, and the excellent reputation of the DeepMind AI team, specifically. After entering the magical circle of the game, there is, however, no way back. He is captured by the rules of the game and has to play until the end. In AlphaGo vs. Sedol, fate and freedom come together, enforcing the game’s tragical character. Sedol reflects on his daunting position: “I can’t believe this is happening. Regardless of your opponent’s level, to be defeated not by three-zero, but five-zero? Losing to AlphaGo by five-zero would really hurt my pride” (Kohs 2016, 1:03:06).

After every game, Sedol has to face the press. The first time, he stated to be surprised by the high playing level of AlphaGo, the second time he said to be speechless, the third time he felt he had to apologize to the audience. Sedol:

I think I have to express my apologies first. If I had been able to play better or smarter, the results might have been different. I think I disappointed too many of you this time. I want to apologize for being so powerless. I've never felt...this much pressure, this much weight. I think I was too weak too overcome it. (Kohs 2016, 1:02:05)

Experiencing self-doubt and guilt, the responsibility of playing on behalf of 'humanity' becomes an enormous burden for Sedol to bare. However, nobody really doubts his efforts. The fact that Sedol actually manages to win the fourth match, only underlines his determinacy. As De Mul describes the tragic hero, he stresses that: "When they go down in their struggle, this is generally not the result of a lack of determination, but rather of the inhuman grandness of their efforts. They lose themselves in the catastrophic course of events" (De Mul 2014, 45).

Phronetic Understanding

As viewers of the documentary, we can empathize with Sedol as we all have first-hand experiences of being captured between freedom and necessity, between making a choice and not being able to escape its consequences. Aristotle states that watching a tragedy unfold before your eyes invokes feelings of compassion and fear with the audience, liberating "the intense emotions that are called forth by the tragic events" (De Mul 2014, 43).

There is something to be learned from the suffering of Sedol. Or as Ricoeur states it, we can move from tragic wisdom to practical wisdom (Ricoeur 1994). The one-dimensional perspective on for instance "the human" or "the machine" or "intelligence" when confronted with the complexity of life can lead to sometimes insoluble conflicts. Deciding on the best action in these situations is not so much about coming up with a univocal solution, as it is about developing a judgment that acknowledges and responds to the tragic character of the conflict (also see: Nussbaum 2002).

A tragedy, such as the AlphaGo game does not only forces us to face the limitations of our human condition, it also gives us the opportunity of undergoing what De Mul calls an "aesthetic experience" (De Mul 2014, 50). Actually, on two occasions in the documentary, an outspoken aesthetic experience occurs. The first aesthetic experience is move 37 in the second game, made by AlphaGo, the second one is move 78 in the fourth game by Sedol.

During the second match, while Sedol takes a break to smoke, AlphaGo plays move 37. Almost all professional commentators agree that this is a move, no human would play. AlphaGo actually agrees with this judgement as the AI states that there is a 1 in 10,000 probability that this move would have been played by a human player. When Sedol comes back into the room and sees move 37, the camera focuses on his face and we witness his confusion. Sedol comments:

I thought AlphaGo was based on probability calculation and that it was merely a machine. But when I saw this move, I changed my mind. Surely, AlphaGo is creative. This move was really creative and beautiful (Kohs 2016, 00:52:17) ... This move made me think about Go in a new light. What does creativity mean in Go? It was a really meaningful move (Kohs 2016, 00:53:35).

On several occasions in the documentary, the word “beautiful” is employed to describe move 37. However, one can argue that beautiful is a too limited expression of what happens here. Move 37 is not something we merely desire or that sparks our admiration. It is both a stunning and devastating move (cf. Yu 2016). Stunning, because it opens up the possibility of understanding and playing Go in a completely new way. Devastating, because creativity and intuition seemingly no longer solely belong to human beings. Move 37 sparks *sublime* beauty (on tragedy and the sublime, see: De Mul 2014, 56-59). We can feel the agony of Lee Sadol, while we simultaneously also share in his wonder for this extraordinary move. A move, which one the one hand is *non-human*, in the sense that a human player would probably never had played it and on the other hand, is utterly *human*, as it displays a capacity for creativity and intuition, we until now had only ascribed to humans.

In the fourth match, with move 78, Sedol is able to play a wedge move at the center of the board, making the game so complex that AlphaGo is no longer able to evaluate its position correctly. The AI starts to play strange and even silly moves and resigns after a while. Sedol wins the match. At the press conference, move 78 is called a “god’s move.” Asking Sedol, what he was thinking when he played that move, he replies: “At that point in the game, move 78 was the only move I could see. There was no other placement. It was the only option for me, so I put it there.” AlphaGo, however, estimated –once more– that only 1 out of 10,000 humans would actually play that move. In a sense, move 78 came to mirror move 37. Twin moves, but not identical twins as they were both extraordinary, both beautiful, but not both sublime.

With move 78, as one DeepMind scientist explains, we are left: “a little bit in awe of the human brain’s power, in particular, Lee’s amazing ability to cause AlphaGo problems and find something seemingly out of nothing. And so, we really want to understand what had happened” (Kohs 2016, 1:14:47). In other words, at least for the DeepMind crew, move 78 was not merely beautiful, it also quickly became another opportunity to further improve their AI, commodifying move 78. People watching the victory, on the other hand, were completely overwhelmed by excitement, resulting in an intense and all-consuming joyful experience. However, therefore also their experience did not have a sublime character, as it lacked the ambiguity of going through conflicting emotions as a whole, which is a typical aspect of the sublime. Rather, these feelings of joy and excitement, at least for a short period of time, smothered the feelings of fear and loss that had been growing with every lost game. Sedol (Kohs 2016, 1:12:19) explains:

I heard people shouting in joy when it was clear that AlphaGo had lost the game. I think it is clear why. People felt helplessness and fear. It seemed like we humans are so weak and fragile. And this victory meant...we could still hold our own. As time goes on, it will probably be very difficult to beat AI. But winning this one time, I felt like it was enough. One time was enough. I heard from so many people saying they were running out in the street. They were so happy. They were chanting, they were celebrating.

By interpreting the tragic-like emplotment of the documentary, we can identify some basic characteristics of human-AI interactions which could be useful for a postphenomenological analysis. First, hubris leads human beings to mistakenly believe that they can completely control AI and that it is merely an instrument to be played with or to have at your disposal. This hubris is not merely in Sedol, who takes on the challenge of competing with AlphaGo, but also in the computer scientists of DeepMind, who are convinced that they contain AlphaGo and can explain its actions. In other words, hubris might blind people for the mediating workings of the technology. Secondly, this loss of control is not merely a bad thing. As we saw, there is also beauty in a technology which accomplishes things we thought of as belonging solely to the human domain. The complexity of experiencing the sublime in our interactions with technology demands from postphenomenology a sensitivity for those conflicting feelings that exist in tandem. It is not just that different people can have different mediating relations with an AI application, but within a specific mediation, different, conflicting meaning-given movements can co-exist.

5. Transfiguration: When Documentary and Life Intersect

In this section the art of narrating a story and of following a story come together. In the previous sections, we have established the game-like structure of the documentary, how it moves between the poles of sedimentation and innovation and appeals to our phronetic understanding. Now we are able to explore how this story mediates our relation to ourselves, others, and the world around us.

Self-understanding

The documentary explains that Go is been thought of not just as a game, but as having “a very deep philosophy. The Go board reflects the individual who’s playing. The truth is gonna show itself on the board. You won’t be able to hide it” (Kohs 2016, 00:16:45). Or as Fan Hui explains: “Because all the things I learned in my life is [sic] with Go. It looks like a mirror. I see Go, I also see myself. For me, Go is real life” (Kohs 2016, 00:04:20).

Playing Go is also a form of self-expression, a way of distinguishing oneself from others. Sedol: “It has become a type of creation of mine. I want my style of Go to be something different, something new, my own thing. Something that no one has thought of before” (Kohs 2016, 00:18:25). So, if Go is a form of self-expression and a way to get to know oneself, it is interesting to investigate if AlphaGo mediates this relation with the self in a new way.

In the first part of the documentary, we get a glimpse of such mediation through Fan Hui who just lost, as a professional Go player, from a computer program for the first time in history. During the game, we see him go through a kaleidoscope of emotions: surprise, irritation, fear, resignation. After the final game, he comments: “I feel something very... strange. I lose with [sic] a program. And I don’t understand myself anymore” (Kohs 2016, 00:10:16). He leaves and it takes more than an hour for him to return to the DeepMind offices. To the question if he is okay, he replies to experience mixed feelings. He is not happy to have lost the game, but he is happy to be part of this historic moment. After going through the emotions of losing the game, of losing himself, he seemingly has found a new part of his identity: the first human Go professional to have lost from a computer program, making him part of an historical event. Later in the documentary, he will join the DeepMind team to help them improve AlphaGo.

Lee Sedol too comes to see himself in a new light through his interaction with AlphaGo. Notwithstanding, or actually maybe because of the four games that he lost, he declares: “I feel

thankful and feel like I have found the reason I play Go. I realize it was really a good choice, learning to play Go” (Kohs 2016, 01:25:36).

All in all, both Go players in their interaction with AlphaGo, go through a *personal transition*. Losing from AlphaGo invokes some sort of existential anxiety at first, but also provides room for a new way of understanding themselves. They change through their interaction with AlphaGo. By focusing on these first-person experiences, we can nuance the earlier discussed antagonistic framing inherent in the man vs. machine narrative. The human Go players are not diminished but transformed by their interaction with the machine.

Communicability

In the documentary, AlphaGo becomes ‘the other.’ From a postphenomenological point of view, we could speak of an –augmented– alterity relation. The Go players interact with AlphaGo as another actant (Latour 1993).

What immediately stands out in this interaction between players and AlphaGo, is the lack of intersubjectivity. Yes, AlphaGo plays moves which are creative and beautiful, but it does not play Go with its opponent. Playing is not just following the rules of the game –which AlphaGo does perfectly–, more importantly, it is a coming together of actors who co-experience the event by looking at each other, reading emotions, smelling, hearing, talking. These parts of the game are lost when AlphaGo plays. Fan Hui explains that normally when playing Go, he can feel “many, many things. But with AlphaGo, you can feel nothing. So, when you can feel nothing when you play, you have more and more questions about yourself” (Kohs 2016, 00:36:39). Here, we see an immediate connection between the second-person and first-person experience. As the former lacks intersubjectivity, the subjective experience becomes much more encompassing, overwhelming even.

In the documentary, we see Lee Sedol looking at Aja Huang, who is sitting opposite of him, placing the stones on the board on behalf of AlphaGo. As a player it comes naturally to look at one’s opponent’s face. However, Aja Huang is not Sedol’s opponent. Huang too feels how unusual this interaction is: “I can actually feel the spirit and courtesy of a great Go player like Lee Sedol...It was the first time he faced a strange opponent, I think. It is non-human, has no emotion, it’s cold. But he stayed very calm. And I can feel his mental strength” (Kohs 2016, 00:33:08). The limited role of Huang and the absence of AlphaGo in the interaction makes that after a couple of games, Sedol brings along friends to go through the moves once more when the game is over. This is something Go-players usually do together, but with AlphaGo has

become impossible. All in all, the *lack of intersubjectivity* makes *loneliness* the fundamental mood of this interaction.

This need for understanding the other and building a meaningful connection is also reflected in the active search of the audience to find ways to interpret AlphaGo. They are looking for ‘tells’ of AlphaGo which might give away something of its *modus operandi* and many journalists interview DeepMind scientists to explain to them how AlphaGo ‘thinks.’

It is striking how far the DeepMind crew goes to demystify AlphaGo in the documentary. There are quite a few scenes shot in the control room where we witness the scientists follow, comment, and interpret the moves of AlphaGo. Seemingly, the muteness of the artificial intelligence is balanced with the human networks in which AlphaGo is embedded.

Although, DeepMind wants to stay far away from anthropomorphizing AI, they also stress its humanness, even without its ability for intersubjectivity. One of the DeepMind-scientists states:

AlphaGo is human-created, and I think that’s the ultimate sign of human ingenuity and cleverness. Everything that AlphaGo does, it does, because a human has either created the data that it learns from, created the learning algorithm that learns from that data, created the search algorithm. All of these things have come from humans. So, really, this is a human endeavor (Kohs 2016, 00:33:08).

Referentiality

Finally, the documentary also tells us something about the world and how it changes because of AlphaGo. First of all, it puts the world of Go itself in a different light. When people start to realize that AlphaGo is not a ‘normal,’ mediocre program and that what at first sight might look like a wrong move, a mistake, actually turns out to be a game-changing move, they acknowledge Go has entered a new paradigm because of AlphaGo. “The lessons that AlphaGo is teaching us are going to influence how Go is played for the next thousand years,” one commentator claims (Kohs 2016, 01:20:05). Sedol: “What surprised me the most was... that AlphaGo showed us that moves humans may have thought are creative, were actually conventional” (Kohs 2016 01:21:03).

However, that Go has shown people that they might be less creative than they initially thought, did not make them stop play. On the contrary, one of the effects of the AlphaGo game and the documentary was a rush on Go boards and stones.

The documentary wants to show us also another world, which reaches far beyond the game of Go. Particularly at the end of the documentary the narrative of the man vs. machine framing is abandoned to sketch a future where AI and human beings empower each other. Here the mutual shaping vision, which is key to postphenomenology and other approaches after the empirical turn, becomes apparent. Reporter Cade Metz reflecting on the game makes the link between Sedol's transformation and how this might happen to all of us:

At least in a broad sense, move 37 begat move 78, begat a new attitude in Lee Sedol, a new way of seeing the game. He improved through the game. His humanness was expanded after playing this inanimate creation. And the hope is that, that machine, and in particular, the technology behind it, can have the same effect with all of us (Kohs 2016, 01:24:42).

6. Conclusion

The aim of this chapter was to investigate in which way the documentary AlphaGo can give us access to AI technology which is not yet part of everyday life, so we can come to understand AI's mediating qualities in a timely manner and inform a postphenomenological analysis. By adopting key elements of Ricoeur's narrative discourse, we engaged in a step-by-step analysis of the documentary, bringing forth some valuable findings.

Amongst others, we established the importance of actively connecting micro and macroperception. While theoretically interlinked, in practice they are often addressed separately. The cultural embeddedness of micro-perceptions can however lead to different and even conflicting human-AI-world relations, as the framing of US and Chinese media outlets show. Moreover, this macro-perception is also important to understand the tragic emplotment that resonates in the documentary, referring back to the narrative of the ancient Greek tragedies. Finally, the macro-perception is also important to grasp the power relations that play a role in the set-up of the documentary. By choosing a game-setting to display their AI's strength, DeepMind (Google) creates a safe action-space where AlphaGo never becomes too threatening.

When it comes to the first-person experience, we saw that the initial hubris of the human actors led them to believe they could control the AI, blinding them for the mediating workings of AlphaGo. Or, loosely referring to Ihde (1990): they wanted the transformation of the AI without its mediation. This loss of control can, however, go together with a feeling of awe. The complex, layered experience of the sublime is something we might expect to perceive more

often in our future interactions with AI. This asks of postphenomenology to invest in crafting a new sensitivity for those conflicting feelings coming together in such an experience.

As all mediating technologies bring some aspects of reality to the fore and hide others in the background, we learn from our analysis that one of the crucial aspects that gets lost in AI-interactions is intersubjectivity, resulting in feelings of strangeness and loneliness. That people look for ‘tells,’ look for togetherness with AlphaGo is not mere anthropomorphizing, it is a fundamental need of people to interpret and to be interpreted. Recent investigations in explainable AI, in developing methods to explain what happens in the black box, predominantly focus on explanations that enable engineers to make the system more robust or to comply with legal requirements. However, our analysis indicates that explainability might also be *crucial to live a flourishing life together with AI*. What kind of explanation is needed, or to put it differently, what kind of *mediation* is desired from an existential point of view, might be a valuable question postphenomenology could actually contribute to if it makes room for new understandings of what it means to operate as an ‘empirically-informed’ approach. In our analysis of AlphaGo, Ricoeur’s narrative discourse has proven to be a valuable way to interpret and gain access to AI.

References

- Aagaard, Jesper, Jan Kyrre Berg Friis, Jessica Sorenson, Oliver Tafdrup, and Cathrine Hasse, eds. 2018. *Postphenomenological Methodologies. New Ways in Mediating Techno-Human Relationships*. Lanham: Lexington Books.
- Achterhuis, Hans. 2001. *American Philosophy of Technology: The Empirical Turn*. Indianapolis: Indiana University Press.
- Bory, Paolo. 2019. "Deep new: The Shifting Narratives of Artificial Intelligence from Deep Blue to AlphaGo." *Convergence* 25 (4):627-642.
- Crawford, Kate, and Meredith Whittaker. 2016. *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. New York: AI Now.
- Curran, Nathaniel Ming, Jingyi Sun, and Joo-Wha Hong. 2019. "Anthropomorphizing AlphaGo: a content analysis of the framing of Google DeepMind's AlphaGo in the Chinese and American press." *AI & SOCIETY*:1-9.
- De Mul, Jos. 2014. *Destiny Domesticated: The Rebirth of Tragedy out of the Spirit of Technology*. Albany: SUNY Press.
- Feenberg, Andrew. 2009. "Democratic rationalization: technology, power and freedom." In *Readings in the Philosophy of Technology*, edited by D.M. Kaplan, 139-155. Lanham: Rowman & Littlefield Publishers.
- Feenberg, Andrew. 2017. *Technosystem. The Social Life of Reason*. Cambridge: Harvard University Press.
- Frissen, Valerie, Sybille Lammes, Michiel De Lange, Jos De Mul, and Joost Raessens. 2015. "Homo ludens 2.0: Play, Media, and Identity." In *Playful Identities: The Ludification of Digital Media Cultures*, edited by Valerie Frissen, Sybille Lammes, Michiel De Lange, Jos De Mul, and Joost Raessens, 9-50. Amsterdam: Amsterdam University Press.
- Gadamer, Hans-Georg. 1972. *Wahrheit und Methode*. Tübingen: Mohr.
- Granter, Scott R., Andrew H. Beck, and David J. Papke, Jr. 2017. "AlphaGo, Deep Learning, and the Future of the Human Microscopist." *Archives of pathology & laboratory medicine* 141 (5):619-621.
- Huizinga, Johan. 1955. *Homo Ludens: A Study of the Play-element in Culture*. Boston: Beacon Press.

- Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press.
- Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9):389-399.
- Kaplan, David M. 2006. "Paul Ricoeur and the Philosophy of Technology." *Journal of French and Francophone Philosophy* 16 (1/2):42-56.
- Keymolen, Esther. 2017. "Trust in the Networked Era: When Phones Become Hotel Keys." *Techné: Research in Philosophy and Technology* 22 (1):51-75.
- Kohs, Greg, Gary Krieg, Josh Rosen, and Kevin Proudfoot. 2017. *AlphaGo*. United States: Moxie Pictures.
- Kudina, Olya, and Peter-Paul Verbeek. 2019. "Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy." *Science, Technology, & Human Values* 44 (2):291-314.
- Latour, Bruno. 1993. *We Have Never Been Modern*. Harvard: Harvard University Press.
- Li, Fangxing, and Yan Du. 2018. "From AlphaGo to power system AI: What engineers can learn from solving the most complex board game." *IEEE Power and Energy Magazine* 16 (2):76-84.
- Nussbaum, Martha C. 2002. "Ricoeur on Tragedy: Teleology, Deontology, and Phronesis." In *Paul Ricoeur and Contemporary Moral Thought*, edited by John Wall, William Schweiker, W. David Hall and David Hall, 264-276. New York: Routledge.
- Pasquale, Frank. 2015. *The Blackbox Society. The Secret Algorithms That Control Money and Information*. Harvard: Harvard University Press.
- Pellauer, David and Bernard Dauenhauer. 2016. "Paul Ricoeur." In *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2016/entries/ricoeur/>
- Ricoeur, Paul. 1981. *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. Edited and translated by John B. Thompson. Cambridge: Cambridge University Press.
- Ricoeur, Paul. 1991. "Life in Quest of Narrative." In *On Paul Ricoeur: Narrative and Interpretation*, edited by David Wood, 20-33. London: Routledge.
- Ricoeur, Paul. 1994. *Oneself as Another*. Chicago: University of Chicago Press.
- Rosenberger, Robert, and Peter-Paul Verbeek. 2015. "A Field Guide to Postphenomenology." In *Postphenomenological Investigations. Essays on Human-technology Relations*, edited by Robert Rosenberger and Peter-Paul Verbeek, 9-42. Lanham: Lexington Books.

- Rosenberger, Robert. 2018. "Why it Takes Both Postphenomenology and STS to Account for Technological Mediation. The Case of LOVE park." In *Postphenomenological Methodologies. New Ways in Mediation Techno-human Relationships*, edited by Jesper Aagaard, Jan Kyrre Berg Friis, Jessica Sorenson, Oliver Tafdrup and Cathrine Hasse, 171-198. Lanham: Lexington Books.
- Silver, David, and Demis Hassabis. 2016. "AlphaGo: Mastering the ancient game of Go with Machine Learning." *Google AI Blog*, July 2, 2016.
<https://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html>
- Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, and Adrian Bolton. 2017. "Mastering the Game of Go Without Human Knowledge." *Nature* 550 (7676): 354-359.
- The-Online-Initiative. 2015. "Background Document: Rethinking Public Spaces in the Digital Transition." In *The Onlife Manifesto*, edited by Luciano Floridi, 41-48. Cham: Springer.
- Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: The University of Chicago Press.
- Yu, Haofeng. 2016. "From Deep Blue to DeepMind: What AlphaGo Tells Us." *Predictive Analytics and Futurism* (13): 42-45.