

NEVER MIND CONSCIOUSNESS: IS ANTHROPOCENTRISM THE HARD
PROBLEM FOR ARTIFICIAL INTELLIGENCE?

Lisa Whitsett – Student Number 2015809

Course: HPPH7014

Dr. Rebekah Humphreys

University of Wales Trinity Saint David

Submitted 31 March 2022

Acknowledgements

My sincere thanks to Dr. Rebekah Humphreys for her constructive and encouraging comments, her insightful questions, and for sharing her expertise on this topic. With gratitude, I thank Dr. David Cockburn, Dr. Tristan Nash, and Dr. Thomas Jansen for their mentoring, instruction, and encouragement.

Master's Degrees by Examination and Dissertation

Declaration Form

1. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Name: Lisa Whitsett

Date: 31 March 2022

2. This dissertation is being submitted in partial fulfilment of the requirements for the degree of:
MA, Philosophy

Name: Lisa Whitsett

Date: 31 March 2022

3. This dissertation is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references.

A bibliography is appended.

Name: Lisa Whitsett

Date: 31 March 2022

4. I hereby give consent for my dissertation, if accepted, to be available for photocopying, inter-library loan, and for deposit in the University's digital repository

Name: Lisa Whitsett

Date: 31 March 2022

Supervisor's Declaration

I am satisfied that this work is the result of the student's own efforts.

Signed: Dr. Rebekah Humphreys

Date: 31 March 2022

Table of Contents

Section One – Introduction	1
Overview.....	1
Literature Review.....	4
Section Two – AI in Theory	11
AI’s Intellectual Foundations.....	11
AI’s Main Goals.....	11
AI’s Main Hypothesis and Approaches	12
AI’s Rational Agents.....	14
AI’s Theory of Mind.....	15
Section Three – AI in Practice	17
Western Philosophical Tradition.....	17
Substance Dualism.....	20
Notions of Intelligence, Thought, and Language.....	23
Rational Agents.....	29
Common Sense and Peripheral Intelligence	32
Section Four – Ontological Errors in Anthropocentric Paradigms	35
Section Five – Findings	39
Section Six – Discussion.....	42
Implications	42
Future Research	45
References.....	47

Abstract

The world is experiencing its fourth industrial revolution. In contrast to previous revolutions, this one is digital. It is characterized by unparalleled pace, exponential growth, and advanced technologies. These technologies blur lines between the physical, digital, and biological boundaries. Artificial intelligence (AI) is one such technology. As AI advances, so do its underlying assumptions, one of which is that human intelligence is the ideal goal and benchmark of AI. I question this assumption, arguing that it is highly anthropocentric, and it discounts useful data from a variety of disciplines. I argue that AI's anthropocentric assumptions allow for reasoning errors. These errors further influence AI research and practice, reinforcing misconceptions about human and non-human animal intelligence. The findings have theoretical and practical implications for AI and intersecting disciplines.

Section One – Introduction

Overview

The world is facing a fourth industrial revolution. Unlike previous revolutions, this one is digital. Digitalization brings exponential growth, unparalleled pace, and advanced technologies. One of these technologies, artificial intelligence (hereafter AI), is distinctly powerful, blurring the lines among “physical, digital, and biological spheres” (Schwab, 2016). These blurred boundaries mean changes to economies, social systems, and the environment. AI¹ has had successes and failures, both of which have heightened attention to its research, design, and deployment, including its ethics. “The recent focus on ethical AI has arisen from increasing concern over its unintended negative impacts, coupled with a traditional exclusion of ethical analysis from engineering practice” (Peters *et al.* 2020). Given the breadth and depth of AI’s impact, both human and non-human species are relevant to AI research. The degree of relevance may vary depending on the particular research goals, but overall, one would expect AI to reflect a multi-species approach. One would also expect that if a multi-species approach is not used, AI would explain the choice to include some species while excluding others. I argue that this is not the case, however. I show that AI places *Homo sapiens* central to its practice with little explanation for its anthropocentrism.

Etymologically, the term *anthropocentric* means “pertaining to human beings at a centre” (Harper, 2022). Anthropocentrism differs from anthropomorphism, the latter

¹ I use the term *AI* in this paper to mean the discipline, or field, of artificial intelligence as well as to mean the AI technologies created by the field.

referring to non-human entities being characterized using human traits (Salles, Evers, and Farisco, 2020). Anthropocentrism is one of countless paradigms that can be used in various contexts including humanities and sciences. Paradigms are useful methods to isolate problems, to gather data, and to explain phenomena. Paradigms aid thinkers in considering what is possible, in determining what constitutes evidence, and in deciding the relevance of data. Paradigms are also paradoxical by nature: the moment a thinker commits to a paradigm, the thinker is also necessarily constrained (Morris, 2013, cited in Mylius, 2018). This effectively enlarges and narrows the scope of fluid concepts the thinker addresses. As a result, some components of a problem are highlighted while others are masked, inviting philosophical critique from those with various perspectives.

Anthropocentrism's effects on culture, nature, and society have been discussed extensively in moral philosophy and related fields. This includes animal ethics (Allen, 1998, 2014; Jamieson, 1998; Humphreys, 2008; Andrews, 2019), environmental philosophy (Singer, 2009; Taylor, 2009; Attfield, 2011; Humphreys, 2020); and feminist philosophy (Plumwood, 1993, 2002; Davion, 2002). These analyses often sit within normative contexts causing some to view anthropocentrism as the concern of ethics alone. However, the study of anthropocentrism is important to philosophical inquiry into many facets of contemporary life (Mylius, 2018) including AI. While some research on anthropocentrism in AI has been conducted, further exploration is required. My aim is to help fill the gap by discussing AI's theory and practice, highlighting where anthropocentric paradigms emerge. Having established that, I will answer three research questions: (1) What reasoning errors are embedded in anthropocentrism? (2) How do

these errors impact AI research and practice? (3) What are the implications for AI, intersecting fields, and beyond?

Throughout this paper, I use the term *anthropocentrism* and its variants to refer to paradigms that place *Homo sapiens* at their centre in some way. It is important to clarify that more than one kind of anthropocentric paradigm may exist (Mylius, 2018). I refer to three kinds of anthropocentrism in this paper: perceptual, normative, and descriptive.

Table 1 further explains these types.

Table 1: Types of Anthropocentrism	
Type of Anthropocentrism	Explanation of Type
Perceptual	Any organism using sensory experience does so from its own perspective. A paradigm is perceptually anthropocentric when it is directly or indirectly informed by data received or gathered by the senses of the human body.
Normative	A paradigm is also normatively anthropocentric when it constrains inquiry of a phenomenon in ways that privilege human beings.
Descriptive	A paradigm is also descriptively anthropocentric if it takes as its reference point, centres on, or is focused on human beings. Descriptive anthropocentric paradigms have sub-types, several of which are below. • By omission - present in paradigms that excerpt or bracket humans or elements of human society from its context. • By separation - present in paradigms that assert a capacity or feature that makes humans distinctive also somehow makes humans metaphysically separate from everything else that exists. • By extrapolation - present in paradigms that purport to study phenomena in general but do so based on concepts developed via the study of human beings.

Source: Mylius, 2018, pp. 159-169.

In addition to defining anthropocentric types, table 1 illustrates two other relevant points. Anthropocentrism is evident in statements about human superiority over other species; however, anthropocentrism also comes in more subtle or implicit² forms. Further, anthropocentric paradigms can be displayed separately in a table, but paradigms often overlap in practice (not only in AI, but also in other disciplines).

Literature Review

Research directly exploring anthropocentrism in AI is rather sparse. Within what exists, a common research theme explores the construct of intelligence³ as used in AI. Hernandez-Orallo (2016) and Martínez-Plumed *et al.* (2018) argue that anthropocentric definitions and measures of intelligence must be replaced with modern and inclusive approaches. Cave (2020) argues that AI's concepts of intelligence are not only anthropocentric but also value-laden in problematic ways, resulting in debates within AI that misrepresent the risks and opportunities. Fox (2017) explores AI's anthropocentric framing of intelligence, suggesting that the frames insufficiently consider post-anthropocentric research. Shevlin *et al.* (2019) present evidence that anthropocentric views of human intelligence are an obstacle to AI achieving its loftiest goal, called artificial general intelligence (AGI).

² I use the term *implicit* in this paper in the sense that a belief or assumption is not expressed overtly. I do not intend the term to qualify whether the thinker is consciously or unconsciously aware of the belief or assumption.

³ I acknowledge the controversies and debates surrounding the term *intelligence*. For this paper's purposes, I will use the term *intelligence* broadly (if not imperfectly) to include terms like *thought*, *cognition*, *mind*, *mental states*, and occasionally *consciousness*. I will use these concepts interchangeably as the AI community often does, unless a distinction is required.

Preston (1991) provides one of the first and most comprehensive explorations of anthropocentrism in AI. She suggests that AI’s intuitive notions⁴ of intelligence are highly anthropocentric. For example, Preston says AI scholars Barr and Feigenbaum describe AI as “the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics which we associate with intelligence in human behaviour – understanding language, learning, reasoning, solving problems, and so on” (Barr and Feigenbaum, 1981, quoted in Preston, 1991, p. 261). Preston argues that AI’s notions of intelligence led to artificially reconstructing human intelligence which, in turn, emphasised human abstract reasoning⁵ as most important to intelligent behaviour overall. In response, AI’s research programmes focused on abstract reasoning intensely. The emphasis on abstract reasoning is reflected in early AI successes, some of which are shown in table 2.

TABLE 2. Early AI Successes		
Year	Description	Creator
1952	Checkers programme	Arthur Samuel
1955	Logic Theory Machine designed to prove A. Whitehead and B. Russell’s theorems	Allen Newell, J. Clifford Shaw, and Herbert Simon
1957	General Problem Solver (GPS)	Newell, Shaw, and Simon
1966	ELIZA English dialogue simulation	Joseph Weizenbaum
1966	PARRY English dialogue simulation	Kenneth Colby

Sources: Preston (1991); Boden (2016); Copeland (2021).

⁴ Preston (1991) clarifies that she uses the term *intuitive notions* broadly to include stereotypes and ideologies. Her main point is that the notions are frequently justified on non-theoretical grounds. I adopt the same meaning in this paper.

⁵ In AI, the term *abstract reasoning* is not precisely defined, so I rely on the supporting examples in the paper to provide context. The field seems to contrast abstract reasoning with something akin to *crystallized intelligence* (Ziegler *et al*, 2012), intelligence that comes from facts and experiences.

Table 2 illustrates Preston’s point. Checkers, logic, and language are problems often associated with abstract reasoning. And, as Preston points out, AI’s focus on these kinds of problems suggests that the only issues worth tackling in AI are human ones. Additionally, AI’s focus on these kinds of problems suggests that abstract reasoning, a feature of intelligence which most starkly differentiates human from non-human animals, is the most instructive (Preston, 1991).

I assert that table 2 represents an additional point. Artificial demonstrations of abstract reasoning invite comparison between AI and human intelligence, thus implicitly positioning human intelligence as the ultimate goal. This comparison is evident in mainstream AI dialogue, which often contains references to terms like *human-level machine intelligence*, *human-level artificial intelligence*, *high-level machine intelligence*, and sometimes *artificial general intelligence* (Martinez-Plumed *et al.* 2018).

Preston (1991) further argues that anthropocentric notions were partially responsible for AI’s early failures. Believing abstract reasoning to be primary, AI initially misjudged the role of commonsense knowledge^{6,7} and the significance of peripheral intelligence (perception and movement) to overall intelligence. Repeated failures to artificially reconstruct these types of intelligence prompted AI to rethink its views. “The initial intuition was that abstract reasoning is the hard part, and that common sense could be reconstructed very easily as a mopping up exercise. But opinion has shifted gradually

⁶ Preston (1991) notes that the term *commonsense* is used in AI with little precise technical meaning. It is generally distinguished from “abstract reasoning about formally specified objects and considered the ability to deal with situations and objects encountered in everyday life” (p. 276). Though this definition is not quite satisfactory, I will continue the tradition as does Preston.

⁷ In this paper, I use the phrase *common sense* as a noun and I use the word *commonsense* as an adjective to modify *knowledge* and *intelligence*. Per grammatical rules, the noun requires a space between the two words.

as the real dimensions of the commonsense knowledge problem have come into view, and it is now widely believed that common sense poses a much greater challenge to AI in the long run” (Preston, 1991, p. 267). After the shift, AI no longer viewed abstract reasoning as the most significant aspect of intelligence, suggesting that commonsense is more important. Preston suggests that the emphasis on commonsense intelligence invited AI’s observation that human and non-human animals share intelligent behaviour. The enhanced assessment of commonsense knowledge as more difficult “suggests rather strongly that the abilities of non-human animals will be much harder for AI to duplicate or model than, e.g., the ability to play chess has proven to be. This in turn suggests that the study of infra-human intelligence using the techniques and methods of AI would be an interesting project in its own right” (Preston, 1991, p. 267). Preston says that the growth in research programmes focused on common sense and peripheral intelligence confirms AI’s interest in common sense and AI’s interest in non-human animal intelligence. Preston acknowledges the shift was not an intentional turn away from anthropocentrism (1991), but she asserts it signalled a shift away, nonetheless.

I think we have good reason to question Preston’s argument. Humans have a poor track record recognising our commonality with other groups. Instead, we tend to view ourselves as cognitively exceptional in capacity or level (McGregor, 2018). We often exaggerate our intelligence over that of other species (Zentall, 2020). So, shared abilities may have invited contemplation of commonality, but little reason exists to assume the result would be a substantive shift away from anthropocentrism. But what, then, might explain the growth of research programmes in common sense and peripheral intelligence? A likely explanation has its roots in anthropocentrism. Preston writes:

It rapidly became obvious to early researchers that a [programme] which could successfully understand and solve algebra word problems accessible only to people with ten or twelve years of formal education could nevertheless not be extended in any obvious way so as to understand and react appropriately to a story easily accessible to a three year old child...if computers were ever going to be able to operate in natural environments, and to communicate and cooperate with humans on their own level, they would have to be endowed with commonsense knowledge too. So a cry went up in AI to formalize commonsense knowledge; a cry which was widely and enthusiastically heeded.

(Preston, 1991, p. 265).

Both the cry for help and AI's response to it appear to have emerged within an anthropocentric context: the need for AI to communicate and to cooperate with humans. AI's increased investment in commonsense and peripheral intelligence does not mean AI reoriented its notions about *whose* core intelligence is most important, *whose* commonsense intelligence should be studied, or *whose* perceptual skills are ideal. Preston's assumption ignores the pattern that AI has shown (and had shown when Preston published her views), which is to be interested in human versions of the intelligence concepts being studied. This pattern is supported with research from Shanahan *et al.* (2020) who suggest that while common sense is a shared intelligence, many in AI view language as a prerequisite to common sense, hence, non-human animal common sense is often ignored. So, while it seems reasonable to claim a relationship — perhaps even a causal one — between AI's interest in commonsense knowledge and growth in the number of programmes researching commonsense, it seems unreasonable to interpret the growth in research programmes as genuine interest in non-human animal intelligence.

Davion (2002) critiques problematic notions of intelligence in AI from an ecofeminist perspective. She points out that science is “deeply influenced by background assumptions determined by the political and moral contexts from which it emerges”

(Harding, 1986, quoted in Davion, 2002, p. 174). One of these background assumptions is anthropocentrism, which is a privileged construct within something called a value dualism. Value dualisms are manufactured constructs in which two distinct concepts are treated as radically different absolutes. The absolutes are then pitted against each other (Davion, 2002). These value dualisms are powerful in several ways. “The dualism functions to legitimate a number of oppressions, including sex, race, and class oppression, which can all be seen in terms of the central dualism underlying the system, that of reason/nature” (Davion, 2002, p. 164). Davion explains that the reason/nature dualism earns a prominent role in ecofeminist theory because many view it as the underpinning of Western patriarchal societies. The ecofeminist view is that the reason/nature dualism forms a foundation upon which other related dualisms rest. In these related dualisms, the construct associated with reason is viewed as superior to the construct associated with nature. Related dualisms include anthropocentrism/non-anthropocentrism, human/nature, and masculine/feminine⁸. Davion makes the point that dualisms are also powerful because the dominant portion of the value dualism becomes so engrained that the concepts seem intuitive, and no argument for them is needed. As a result, the construct becomes uncritically accepted and unquestioned. Davion concludes that anthropocentrism will remain durable unless it is challenged from these kinds of political perspectives (2002).

Similarly, Cave (2020) argues that anthropocentrism in AI must be understood in broader political and social contexts. Tracing the story of intelligence to the early days of Plato, Socrates, and Aristotle, Cave points out that early philosophers connected reason to positions of rulership. The criterion to be the ruler was rationality, which was possessed

⁸ In these examples, the construct presented first in the pair represents the dominant construct.

by educated men. “At the dawn of Western philosophy, we have intelligence identified with the European, educated, male human. It becomes an argument for his right to dominate women, the lower classes, uncivilised peoples and [non-human] animals” (Cave, 2017). Intelligence, then, has a history of occupying a key role in systems and structures that not only perpetuate value dualisms (to use the ecofeminist term), but also perpetuate the idea that such structures are natural and inevitable. Further, intelligence is fetishized. The centrality of intelligence to the human experience has been assumed, and threats to human intelligence from AI have been inflated (Cave, 2020).

Stephen Fox (2017) argues that AI has been framed problematically, particularly in its approach to intelligence. He suggests that AI not only views human intelligence as superior to other intelligences, but also views intelligence developed by humans (i.e., AI) as superior to all other intelligences. As a result, post-anthropocentric research on intelligence is insufficiently considered. Fox further argues that AI frames debates between AI proponents and AI end-users in anthropocentric ways. For example, the term *friendly AI* focuses on ensuring AI has positive side effects on human beings, while little concern is expressed for other species on the planet and AI’s impact on them (2017).

This paper expands the literature on anthropocentrism in AI. In section two, I show that AI is theoretically positioned to explore intelligence from a multi-species perspective. In section three, I show that despite its positioning, AI operates on notions of intelligence that are highly anthropocentric. In section four, I show that some of the anthropocentric paradigms represent reasoning errors which, in turn, may lead to unsubstantiated claims about intelligence. In sections five and six, I present findings and discuss implications for the AI discipline and beyond it.

Section Two – AI in Theory

In this section, I establish AI's main hypothesis, its aims, and its core approaches. In doing so, I show that AI is theoretically positioned to explore many kinds of intelligence of many species. This section serves as a baseline from which to compare AI's anthropocentric practices addressed in section three.

AI's Intellectual Foundations

Scholars in multiple disciplines contributed to the pursuit of today's AI. Philosopher Margaret Boden (2016) explains these disciplines include mathematics, computational neuroscience, neurophysiology, neurology, psychology, computer science, and engineering. With such concentrated effort, AI was officially born in the early 1950s (Boden, 2016). Since then, AI's research corpus continues to expand as does AI's relevance to intersecting fields. The proliferation of research is partly driven by AI's commercial value. AI is an industry globally estimated at \$62 billion USD (Gartner, 2021) and is closely connected to political, economic, social, and environmental strategies across the world.

AI's Main Goals

As a distinct field, AI has two main goals. One goal is technical, and the other is scientific (Boden, 2016). Technically speaking, AI's goal is to employ computers to accomplish useful things. The technical domain produces the technological devices,

applications, and algorithms common in today's environment. Examples include internet search engines like Google, social media like Facebook or LinkedIn, and wearable devices like smartwatches. AI researchers sometimes create these technologies "by employing methods very unlike those used by minds" (Boden, 2016, p. 2). In theory, AI researchers acknowledge that, while minds excel at many sorts of tasks, a problem can be solved in many ways. Some solutions may be very different than those that minds would generate. AI's other goal is a scientific one. The scientific goal involves "using AI concepts and models to help answer questions about human beings and other living things" (Boden, 2016, p.2). For example, AI researchers have created computer models to test life science theories including medical diagnoses and drug effectiveness (Wyss Institute, 2014).

AI's Main Hypothesis and Approaches

AI is built on a core hypothesis to which modern AI researchers remain committed: it is logically possible to create a machine that thinks (Turing, 1950). This hypothesis reflects its mathematical, engineering, computer science, and related foundations, so it takes no position on what has now become a major philosophical debate. The debate generally concerns whether a machine appears to be thinking (called the *weak AI* hypothesis) or whether a machine is actually thinking (called the *strong AI* hypothesis)⁹. Most philosophers will accept the weak AI hypothesis, but others criticize the strong AI hypothesis arguing that modelling a mind differs in important ways from

⁹ This important philosophical debate is abbreviated given the paper's scope. A comprehensive discussion of the opposing arguments can be found at S. Bringsjord and N. Govindarajulu (2020). Complete source information is provided in this paper's references.

creating one (Dreyfus, 1965). John Searle (1980) argued against strong AI on grounds that machines do not exhibit understanding, and therefore, are not actually thinking. Other philosophers, such as Daniel Dennett (1994) say that strong AI is possible but not recommended or desirable. While the debate seems relevant if AI intends to make claims about minds based on machines, some AI researchers only care whether their computer programmes work or not (Russell and Norvig, 2016). A deep exploration of the weak versus strong AI debate lies outside the scope of this paper, but I raise the issue to demonstrate a key point. Even though the weak/strong AI debate spontaneously leads to discussions about thought and consciousness, the latter mental states could belong to any species, not just to humans. In this sense, anthropocentrism is not inherent in the weak/strong AI debate. Endorsing the view that it is logically possible to create a machine that thinks can induce comparisons to any species' thinking, not just human thought.

To advance its main hypothesis, AI designs, builds, and deploys computer systems that model cognition, thought, and intelligence (Poole, Mackworth, and Goebel, 1998). Many approaches to do so exist; the two most common are addressed here. One approach is called symbolic AI, also referred to as classical AI or good-old-fashioned AI (Boden, 2016). Symbolic AI assumes that intelligence is achieved by manipulating logic symbols and that computer models of cognition are best represented by computers processing symbolic language (Rescorla, 2020). The symbols in the language are grounded in the world, meaning the symbols refer to some aspect of reality (Dennett, 1994). Symbolic AI researchers are likely to focus on modelling language and logical thinking in their work.

Another approach is connectionism (Boden, 2016). Broadly, connectionism explains cognition through neural networks. “Neural networks are simplified models of the brain composed of large numbers of units (the analogs of neurons) together with weights that measure the strength of connections between the units” (Buckner and Garson, 2019). Connectionism is inspired by neurophysiology (Rescorla, 2020) and holds that computer models of cognition are best represented by activation patterns across complex networks. Therefore, connectionist AI researchers are likely to focus on modelling adaptive motor behaviour in their work (Thomas and McClelland, 2008). Nested within the connectionist approach lies cybernetics. Cybernetics is “the scientific study of control and communication in the animal and the machine” (Wiener, 1948) and emphasizes biological self-organization, adaptation, and feedback (Boden, 2016). The approaches I have discussed here differ in several ways, but they are similar in that they lack inherent anthropocentrism.

AI’s Rational Agents

Regardless of approach, AI researchers aim to create artificial entities that behave intelligently. These entities are called rational agents or rational intelligent agents. To qualify as a rational intelligent agent, at any one time the agent must have prior knowledge of the world; past experiences from which it can learn; goals that it tries to achieve or values about what is important; observations about the current environment and itself; and must take some action (Poole, Mackworth, and Goebel, 1998). The important point to keep in mind is that AI aims to create artificial rational agents and, in theory, takes inspiration from natural and biological organisms in doing so.

AI's Theory of Mind

As I mentioned earlier, AI is concerned with artificial thought construction and views it as a complex computer science problem. As such, the relevance of the weak/strong AI debate is not primary (Russell and Norvig, 2016). However, AI's intellectual foundations, its goals, and its approaches point to theories of mind that should be mentioned here.

It is clear, for example, that the field advocates the computational theory of mind (hereafter CTM), or the computational point of view (Pfeifer, Iida, and Lungarella, 2014). Often referred to as the computer metaphor, CTM is a family of views that suggest the mind's core mental processes (e.g., reasoning, decision-making, and problem solving) are computations. These computations are like those executed by a Turing machine (Rescorla, 2020). CTM is endorsed by the field of cognitive science, which as I discussed earlier, influenced AI from the beginning. CTM is considered a variant of the Representational Theory of Mind (hereafter RTM), which claims that cognition is manipulation of representation. In its purest forms, CTM is related to the Language of Thought Hypothesis, most often associated with Jerry Fodor (Milkowski, 2022). Computationalism has been challenged, especially as agreement increases that bodies play a crucial role in cognition. AI robotics is one area experimenting with the role of bodies in cognition to learn more about computational processes occurring in the brain as well as the behaviour of the organism (Pfeifer, Iida, and Lungarella, 2014).

AI also aligns with identity theories of mind. Identity theorists commit to the view that mental events are identical to biological brain processes. "Pain, for example, is nothing over and above a neural state in the central nervous system, presumably the

excitation of certain neurons...in the brain” (Honderich, 2005, p. 418). In its common form, identity theory of mind links psychological types, properties, and kinds with physical types, properties, and kinds. Hence, the theory is sometimes called type physicalism.

AI also aligns with functionalism, a theory generally associated with philosopher Hilary Putnam. Putnam argues against identity theory, suggesting that a pain state, for example, is not “a brain state, in the sense of physical-chemical states of the brain (or even the whole nervous system)” (1967, p. 162). Rather, Putnam argues, pain is another kind of state entirely, called a functional state. The kind of functional state (e.g., pain, pleasure, fear) is not determined by an intrinsic feature of the state. Instead, the kind of functional state is determined by the purpose it serves the organism or system and by the variable inputs and outputs that interrelate with the functional state (Honderich, 2005). Functionalist theory is compatible with AI in part because it borrows concepts from automata theory, a theory core to computer science. In simple terms, automata theory concerns computational logic and the machine that performs those computations (i.e., an automaton). Automata are characterized by predetermined inputs, outputs, and resulting states (Bickle, 2020). Putnam’s functionalism makes two modifications to automata theory. One modification is that functionalism broadens the definition of subjects beyond machines to include any system or organism. Another modification is that functionalism replaces predetermined inputs and outputs with variable inputs and outputs. The variability of inputs and outputs results in functional states that can be variably realized, or multiply realized (Bickle, 2020). “The multiple realizability contention about the mental holds that a given psychological kinds (like pain) can stand in that relationship to

many distinct physical kinds” (Bickle, 2020). Multiple realizability is the gateway through which multiple species’ mental states—not just humans’— become relevant in the functionalism conversation. Certainly, functionalism has its philosophical critics. Some reject functionalism saying that it does not adequately explain both the easy and hard problems of consciousness¹⁰. Others reject functionalism on the grounds that “consciousness may have interesting functional characteristics, but its nature is not essentially functional” (Van Gulick, 2021). However, my goal in raising these theories of mind is to demonstrate that it is possible for AI to maintain integrity to its intellectual foundations, its main hypothesis, and its approaches while simultaneously avoiding anthropocentrism.

I have now set a baseline showing that in theory, many kinds of intelligence and many species are relevant to AI’s goals and approaches. In the next section, I argue that despite this, AI operates on anthropocentric notions, many of which are deeply rooted in Western thought.

Section Three – AI in Practice

Western Philosophical Tradition

Anthropocentrism is by no means an invention of the modern world. At the start of the world’s philosophical period, ancient Greeks including Socrates, Plato, and Aristotle established the West’s philosophical agenda which tied the meaning of life to

¹⁰ A comprehensive discussion of the easy and hard problem of consciousness can be found in Chalmers (1996). Complete source information is provided in this paper’s references.

understanding the essence of human beings (Perry, 2015). The medieval era saw a fleeting reemphasis on religion and some scepticism about science, but this was soon replaced by the Renaissance and the Age of Reason (Buckingham, 2011). The Age of Reason marked the victory of rationality and science over religious doctrines. Disciplines like philosophy, science, and mathematics reconnected, providing non-religious explanations for the universe and alternative methods to acquire knowledge about the world. Two primary philosophical traditions emerged from this: rationalism and empiricism. Rationalists endorse the view that knowledge about the world comes from reason alone, whereas empiricists advocate knowledge acquisition through experience. Despite this difference, both approaches are anthropocentric in the same sense. They share the view that it is human beings “whose reason or experience leads to knowledge” (Buckingham, 2011, p. 101).

Many associate rationalism with philosopher Rene Descartes given his dedication to rationalist views and approaches. Descartes, who was also a mathematician, was keen to “rid the sciences of the annoyance of scepticism once and for all” (Buckingham, 2011, p. 118). The latter scepticism was a holdover from the earlier times, during which people had become suspicious of science and the possibility of knowledge generally. Descartes responded to this by writing *Meditations on First Philosophy* (hereafter *Meditations*), published in Latin in 1641. *Meditations* is a rigorous metaphysical and epistemological journey to “demonstrate the possibility of knowledge even from the most sceptical of positions, and from this, to establish a firm foundation for the sciences” (Buckingham, 2011, p. 118). In pursuit of certain knowledge, Descartes questions anything of which he previously felt sure, especially any knowledge gained through sense experience (as

rationalism would dictate). Committed to doubting everything, Descartes addresses questions of his own existence. Through the process of reasoning, Descartes has an epiphany: “After everything has been most carefully weighed, it must finally be established that this pronouncement, ‘I am, I exist’ is necessarily true every time I utter it or conceive it in my mind” (Descartes, 1993, p. 18). Descartes’s pronouncement “*cogito ergo sum*” — Latin for the phrase *I think therefore I am* — is often simply called “the cogito”. Philosophical shorthand for the concept that if a thought exists, then the thinker exists, Descartes’s claim survives as one of the most influential in philosophy.

Descartes’s cogito is central to debates about what sorts of beings can have thoughts, which cannot have thoughts, and why. Descartes writes extensively about the differences between human and non-human animals based on rationalism and reasoning capacity¹¹. Descartes observes that, despite not having mouths, non-human animals do not communicate via spoken language. Lack of spoken language, Descartes claims, is evidence that non-human animals lack reasoning ability altogether. Since non-human animals do not speak, they do not have thoughts. If they do not have thoughts, they are not rational and lack minds and souls. This, Descartes claims, proves that non-human minds are completely different from human minds (1993). Descartes asserts that the difference justifies metaphysically separating non-humans from humans, a fundamental characteristic of anthropocentric paradigms (see Table 1 of this paper).

I add that the cogito also demonstrates the primacy of perceptual anthropocentrism. Recall that a paradigm is perceptually anthropocentric if it is informed by data received or gathered by the senses of the human body. Descartes may object to

¹¹ Descartes’s views on differences between human and non-human animals can be found in *Passions of the Soul*, *Meditations on First Philosophy*, and *Discourse on the Method of Rightly Conducting One’s Reason and Seeking Truth in the Sciences*. Complete source information is provided in this paper’s references.

my suggestion on the grounds that he vehemently distrusted data gathered through his sense experience. However, for humans, choosing to ignore sensory data is possible but being perceptually anthropocentric is impossible. Just as I cannot reasonably claim to perceive the world using bat sonar unless I am a bat, I cannot reasonably claim I do not perceive the world using human senses if I am human. Descartes would be in the same position. Though Descartes distrusts his sensory experience, perceptual anthropocentrism still influences him whether he wants it to or not, and whether he realizes it or not (Mylius, 2018). That Descartes has perceptions to distrust points to perceptual anthropocentrism's existence, albeit paradoxically.

Substance Dualism

Metaphysically separating human from non-human animals based on a capacity is a characteristic of anthropocentrism (see Table 1 of this paper). For Descartes, this separation is rooted in his metaphysical views on substance dualism, sometimes called Cartesian dualism. Cartesian dualism sees the fundamental nature of the universe as comprised of two substances: material substance (matter) and immaterial substance (mind or soul). These two substances are distinguishable by their essence. The essence of material substance is extension, which is the characteristic of taking up space in volume, length, or movement. Material substance examples include human bodies, clocks, rocks, and roads. By contrast, the essence of immaterial substance is mind or soul. "Anything we experience as being in us, and which we see can also exist in wholly inanimate bodies, can be attributed only to our body; and anything in us that we can't conceive of as being somehow had by a body must be attributed to our soul" (Descartes, 2017b, p. 1). Though a person is made of these two substances, it is the mind that constitutes the real

person in Descartes's view. Mental states occur in the mind, not the body, and as such, the mind is fundamental to the person's life (Cockburn, 2021).

The separation of minds from bodies led Descartes to suppose that his mind and his body could exist independently of one another (Descartes, 2017a). Though some philosophers disagree whether Descartes endorses the physical possibility of disembodied minds, most agree that Descartes endorses its logical possibility. In Descartes's view, the idea that his mind can continue to exist after the destruction of his body is not incoherent (Cockburn, 2021). However, this view was criticized when it came to explaining the mind/body problem. The mind/body problem can be stated in multiple ways, but it generally concerns the relationship between the mental and physical realms. This includes mental causation (Honderich, 2005). Mental causation posed a challenge to Descartes: if the mind and body are metaphysically separate, how does dualism explain their interaction? Descartes responded to this by suggesting that the human mind and the human body are causally connected by the pineal gland located in the brain. To many of Descartes's critics, this explanation was only partially acceptable since it explained where, but not how, the mind/body interaction occurred.

But for Descartes, the pineal gland was more than a mental causation explanation. Descartes suggested that the rational mind and the body were often at odds. The mind/soul is unified and rational; the body often initiates actions based on passions and spirits¹² which can be troublesome for the soul. For example, the body, sensing fear, may wish to run away while the mind, thinking rationally, may advocate staying calm and holding one's ground. The pineal gland, then, is the "battle ground" (Cockburn, 2021)

¹² Descartes elaborates on what he means by passions, spirits, and their relationship to the body and mind in *Passions of the Soul* (2017b). Complete source information is provided in this paper's references.

where the body and the mind push and pull against each other, competing for victory. So, while Descartes views the body as mechanistic and impressive, the body is also “the biggest villain in Descartes’s plot” (Cockburn, 2021) because the primary characteristic of a strong mind is its ability to overcome the passions (Cockburn, 2021). This separation of mind and body, of mind over matter, still influences AI today.

While the separation of minds from bodies does not constitute anthropocentrism by itself, the assertion has at least two anthropocentric implications. The first implication concerns non-human mental states. To have mental states according to Descartes, one must have a mind. Since Descartes states that non-human animals do not have minds, then they “must belong wholly in the extended divisible world of jostling Cartesian shapes” (Cottingham, 1978, p. 558). It follows from this that any experiences the non-human animal has, including pain, is “nothing more than a set of internal muscle contractions leading to the jerking of certain limbs” (p. 558). Descartes’s dualism, in effect, reduces non-human animal mental states—including pain states—to non-conscious experiences. The second anthropocentric implication concerns eternal life. Descartes writes, “When we know how different the beasts are from us, we are better placed to understand the arguments proving that our soul is of a nature entirely independent of the body, and thus not liable to die with it. And since we can’t see any other causes that destroy the soul, we are naturally led to think that it is immortal” (Descartes, 2017a). From this, one can infer that immortality applies only to humans since immortality depends on having a human soul. Assuming that immortality is desirable, Descartes’s stipulation privileges humans, a characteristic of anthropocentric paradigms.

It would be difficult to establish that any one person's thinking, even Descartes's, directly led to AI's use of anthropocentric paradigms. However, it is reasonable to suggest that Descartes popularized what Davion (2002) called the foundational value dualism of reason/nature (see Section One of this paper). Recall that the reason/nature dualism positions reason as a privileged construct over nature; related dualisms rest on this foundation. Resulting paradigms become implicit and seem to require no explanation (Davion, 2002). This would include human/non-human dualisms, which emerged early in AI's practice.

Notions of Intelligence, Thought, and Language

I have so far discussed Western philosophical traditions and its anthropocentric foundations. This brings us to discuss anthropocentrism in modern AI and its notions of intelligence, thought, and language. Preston (1991) suggests that AI's anthropocentric notions of intelligence were evident in AI's beginnings, starting with the discipline's adoption of *artificial intelligence* as its formal name. Calling the field *artificial cognition* may have better reflected the connection between AI and cognitive science and may have better reflected AI's purported intent to study all forms of intelligence (Preston, 1991). It is difficult to know how much difference a name change would have made. On one hand, if the term *cognition* carries less anthropocentric baggage than the term *intelligence* often does, perhaps AI would have demonstrated interest in multi-species cognition. On the other hand, at the time AI was emerging, cognitive science tended to focus on human subjects, suggesting that cognition was associated with humans anyway.

What we can see is that the field's formal designation seems to have helped anthropocentrism along (Preston, 1991). Mathematician John McCarthy coined the term *artificial intelligence* at the 1955 Dartmouth research conference, an event marking the official birth of the field. The purpose of the conference was to present research proposals designed to advance AI. Consider these excerpts from McCarthy's presentation.

An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain.

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture.

The human mind apparently uses language as its means of handling complicated phenomena.

For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.

(All from McCarthy *et al.* 2006, pp. 2-12).

These proposal excerpts reflect anthropocentric paradigms of one type or other, suggesting the term *artificial intelligence* subsumed human intelligence as its default. We can also see that in these examples, anthropocentrism is explicit: the references to human brains, human minds, and human language are unambiguous. However, in other cases, anthropocentric paradigms are implicit. Consider these statements from McCarthy's presentation.

[Statement 1:] I propose to study the relation of language to intelligence. It seems clear that the direct application of trial and error methods to the relation between sensory data and motor activity will not lead to any very complicated behavior.

[Statement 2:] The human mind apparently uses language as its means of handling complicated phenomena. The trial and error processes at a higher level frequently take the form of formulating conjectures and testing them. The English language has a number of properties which every formal language described so far lacks.

(Both from McCarthy *et al.* 2006, p. 10).

Statement 1 points to McCarthy's research interest in the relationship between language and intelligence. Statement 2 elucidates that McCarthy conceptualizes language as human language and intelligence as human intelligence. Only by viewing the two statements together can we be certain that McCarthy conceptualizes language and intelligence anthropocentrically (in this example at least).

The McCarthy examples above also point to the anthropocentric notion that language is uniquely human. Consider this example from Russell and Norvig's (2016) widely distributed AI text¹³. "Homo sapiens is set apart from other species by the capacity for language. Although chimpanzees, dolphins, and other animals have shown vocabularies of hundreds of signs, only humans can reliably communicate an unbounded number of qualitatively different messages on any topic using discrete signs" (Russell and Norvig, 2016, p. 860). This example legitimizes non-human animal communication ability, but it perpetuates the view that only human communication qualifies as language.

Someone holding the anthropocentric view will likely ask what could be problematic about assuming language means human language or assuming non-human animal communication is not language. This is a fair question. It deserves acknowledgement by highlighting that the anthropocentric view is one of several hypotheses debated in discourse about non-human language, communication, and

¹³ Russell and Norvig's (2016) textbook is marketed as "the authoritative, most-used AI textbook, adopted by over 1500 schools" per <http://aima.cs.berkeley.edu>. I have used the text as a key source of AI's core ideas since it represents content being taught and communicated about AI to a global audience.

thought¹⁴. One way to frame the debate is to widen the aperture slightly. Doing so allows us to see that AI is one of several disciplines with an interest in language and thought. This includes comparative cognition which focuses on how cognition evolves (Chittka *et al.* 2012). Comparative cognition scientists study “which animal behaviours are cognitive and what sort of cognitive mechanisms or processes permit that behaviour” (Andrews, 2019). Another discipline is cognitive ethology, which is the “comparative, evolutionary, and ecological study of nonhuman [*sic*] animal minds including thought processes, beliefs, rationality, information-processing, and consciousness” (Bekoff, 1995). The field of philosophy complements these disciplines by asking questions about animal beliefs, thoughts, consciousness, and rationality, as well as the relationship between language and thought (Andrews, 2019). Among these disciplines and others, debates continue concerning what constitutes language in non-human animals, and whether non-human animals have thoughts and beliefs.

One of the most common positions within the debate is the linguistic view. Glock (2000) explains that the linguistic view “denies that non-linguistic animals have any thoughts” (p. 35). This view has been defended by Descartes (as I previously discussed) and is often referred to as the Cartesian view. Stich (1985) and Davidson (1979) have defended this position more recently (Stich, 1979; Davidson 1985; cited in Andrews and Monsó, 2021). On the linguistic view, the linguistic structure itself allows a thinker to problem-solve, to make generalisations about the world, and to create infinite sets of new thoughts. Creatures without linguistic structure, then, cannot have thoughts according to this view (Andrews, 2019). A slightly less restrictive but related view is that non-human

¹⁴ For a detailed look at this important debate, see Bekoff (1995); Glock (2000); and Andrews (2019). Complete source information is provided in this paper’s references.

“animals can have systems of communication, but lack anything that can be called a language” (Chomsky, 1980, p. 430, quoted in Andrews and Monsó, 2021). The latter view is reflected in the Russell and Norvig example I used earlier in this paper in which non-human animal communication is acknowledged as a system but not as language.

An alternative to the linguistic view begins by hypothesising that language is one tool among several tools by which belief and thought can be communicated. On this view, the temptation to assume that language precedes beliefs or thoughts is explained by the parallel structure between belief and language. Andrews (2019) argues that human language has a grammatical structure which allows us to create sentences; it has logical structure which allows us to make inferences to new sentences; and it has truth value resulting from grammar and logic. Similarly, beliefs have “propositional structure, which means they can be characterised in terms of sentences. They permit rational inference such that we can use sets of beliefs about the world to infer new things about the world, and have truth value such that our beliefs are either true or not” (Andrews, 2019). This view hypothesises that without beliefs to communicate, there would have been no evolutionary pressure on humans to develop linguistic capacity. While belief is necessary for language, language is not required for beliefs (Andrews, 2019).

The non-linguistic view is supported by evidence from the study of birds and bonobos, suggesting that human language sits on a continuum of communication, one shared with non-human animal communication (Andrews and Monsó, 2021). “One should remember that the extensive social lives of some of the most intelligent nonhuman *[sic]* animals, not only bonobos, but aquatic mammals, such as whales and dolphins, down to and including the complex attack manoeuvres of killer whales, are coordinated

entirely by what one might call ‘music’— a ‘language’ of pitch, intonation, and temporal relation.” (McGilchrist, 2019, p. 106). Bugnyar, Reber, and Buckner’s (2016) research suggests “that ravens can generalize from their own perceptual experience to infer the possibility of being seen” (p. 1). The belief-before-language hypothesis is rooted in scientific data and deserves thoughtful consideration.

It also seems that perceptual anthropocentrism can lead to erroneous assumptions about language, communication, and thought. Philosopher and psychiatrist Ian McGilchrist writes, “Because the part of communication that [humans] are aware of lies in the choice of words, we imagine wrongly that that must be where most, perhaps even all, communication lies” (McGilchrist, 2019, p. 105). But obviously, as McGilchrist points out, thoughts can be communicated without words. Humans use non-verbal communication in our gestures and body language daily. Some estimates suggest that more than 90 percent of human communication occurs through our body language and intonation. McGilchrist adds that “Indigenous Pirahã tribes of Brazil use a language that is effectively a kind of song, possessing such a complex array of tones, stresses, and syllable lengths that its speakers can dispense with their vowels and consonants altogether and sing, hum or whistle conversations” (2019, p. 106). The belief that one cannot think without language is a “fallacy of the introspective process...[humans] carry out most mental process that would normally constitute what we mean by thinking without doing anything consciously, or in language, at all...humans form categories and concepts, weigh and evaluate evidence, make decisions and solve problems, all without language” (2019, p. 106). Taken together, this evidence challenges traditional anthropocentric views about language, communication, and thought.

Rational Agents

As discussed in Section Two, AI aims to create artificial rational agents. The choice to create rational agents is not the only option, however. When AI was emerging, researchers had to decide what kind of systems they wanted to create. What is the goal of AI? They wanted AI to be intelligent, but what does that really mean to the AI discipline? Russell and Norvig (2016) explain how the field settled on the goal to create artificial rational agents. The authors present several definitions of AI from prominent AI scholars and organize those definitions into four categories, thus creating four possible goals of AI. I have reconceptualised the authors' framework to emphasize the resulting four possible goals, shown in Figure 1.

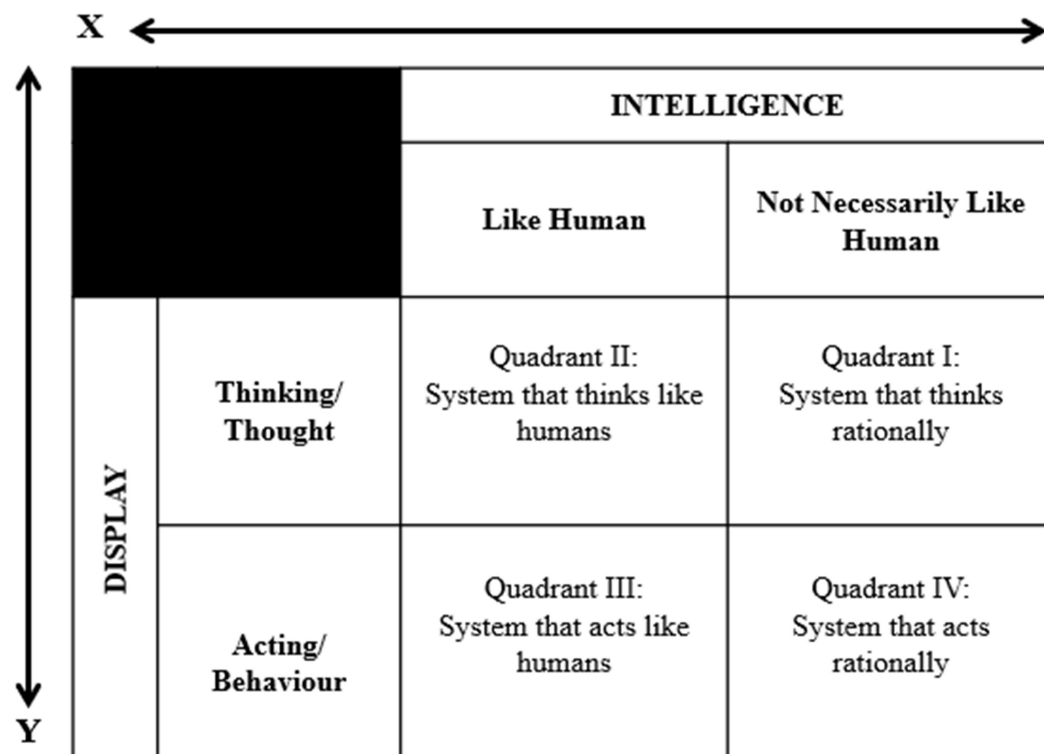


Figure 1. Russell and Norvig's framework representing four possible goals of AI, reconceptualized to emphasize quadrant results. Adapted from Russell and Norvig (2016; 2021).

The X axis represents the dimension *intelligence*, and it is anchored by two options: human and rational. The Y axis represents what I will call the *display* of intelligence, and it is also anchored by two options: thinking and acting (i.e., thought and behaviour). The combinations of the anchors create four quadrants, each of which represents a possible goal for AI. Ultimately, the authors make it clear that they advocate for Quadrant IV. Quadrant IV represents the goal to create an AI system that displays rational behaviour (i.e., acts rationally). The authors explain and justify this choice in the excerpt below.

The rational agent approach has two advantages over the other approaches. First, it is more general than the “laws of thought” approach because correct inference is just one of several possible mechanisms for achieving rationality. Second, it is more amenable to scientific development than are approaches based on human behavior or human thought. The standard of rationality is mathematically well defined and completely general, and can be “unpacked” to generate agent designs that provably achieve it. Human behavior, on the other hand, is well adapted for one specific environment and is defined by, well, the sum total of all the things that humans do. This book therefore concentrates on general principles of rational agents and on components for constructing them.

(Russell and Norvig, 2016, pp. 4-5).

The authors sufficiently explain why thought and behaviour, the Y axis anchors, are the options and they explain the advantage that behaviour has over thought in the AI context. Regarding the X axis, the authors also adequately explain rationality’s advantage over humanness. However, the authors do not explain or justify why human beings are presented as an anchor on the X axis. Why are humans the anchor, as opposed to any other species? The authors only say that to claim AI acts like a human, AI would have to pass the Turing Test; to claim AI thinks like a human would require a precise theory of human minds to which AI would be compared (Russell and Norvig, 2016). Notice the implicit anthropocentric paradigms at work here. The authors suppose that humans are the obvious species by which other intelligences are measured and that readers will hold

the same assumption. This is problematic for many of the same reasons I discussed earlier in this section. Humans are not the only intelligent species by which AI could be designed and compared. If the authors are going to anchor the X axis with humanness as a goal, that choice should be supported with evidence or otherwise substantiated.

Some may say that the authors justify choosing humans as an anchor with the Turing Test¹⁵. The Turing Test is frequently cited as “a satisfactory operational definition of intelligence” (Russell and Norvig, 2016, p. 1) and is offered as evidence that human intelligence is a logical benchmark for AI. Recently, however, some have suggested that Turing was misunderstood, and his intent misinterpreted (Hernandez-Orallo, 2016). As a result, many now believe that Turing did not intend to operationalize questions about whether a machine can think or not, nor did he intend the Turing Test to be a measure of intelligence. Instead, many believe that Turing articulated the “imitation game” (Turing, 1950, p. 443) as a “philosophical instrument in his response to nine objections against machine intelligence” (Hernandez-Orallo, 2016, p. 8). This misunderstanding is unfortunate, but important to consider as anthropocentric influence on AI. The Turing Test assesses humans relative to human characteristics. If the test were meant to be used as many have misperceived, it would be a test of the interrogator, not of the machine (Hernandez-Orallo, 2016). Turing should accept no blame for this misunderstanding, of course, and others’ misunderstanding does not diminish Turing’s pioneering contributions. However, the misunderstanding seems a good reason to stop using the Turing Test to justify the focus on human intelligence in AI.

¹⁵ For a full description of the Turing Test, see Turing (1950). Complete source information is provided in this paper’s references.

Lastly, Figure 1 is also problematic because it creates the perception that only two options exist: either humanness or rationality. This dichotomy is false, however. The dichotomy does not truthfully represent the breadth of available options, which would include many intelligent species, not just *Homo sapiens*.

Common Sense and Peripheral Intelligence

In Section One, I discussed Preston's account of AI's response to the commonsense knowledge problem. Recall that the AI community judged common sense as less relevant to overall intelligence than abstract reasoning; AI also assumed common sense would be easier to artificially reconstruct compared to abstract reasoning. Both assumptions turned out to be incorrect. This left AI tackling two issues, one of understanding common sense as part of intelligence overall and another of technological feasibility. Preston (1991) predicted that the commonsense problem would stir AI's interest in non-human animal intelligence, thus decreasing anthropocentrism. Thirty years have passed since Preston's prediction; how has AI fared?

Technological progress toward artificially constructing commonsense knowledge is becoming more plausible thanks to increased computer power and memory, but it generally remains resistant to AI's current methods. Finding new methods remains one of the field's biggest challenges (Boden, 2016; Shanahan *et al.* 2021). This is especially true to realize artificial general intelligence (AGI).

As for a decrease in anthropocentrism, AI's research interest still generally lies in understanding the human version of common sense and peripheral intelligence. Consider the examples below.

To achieve human-level performance in domains such as natural language processing, vision, and robotics, basic knowledge of the commonsense world—time, space, physical interactions, people, and so on—will be necessary.

(Davis and Marcus, 2015, p. 92).

Autonomous robots with human levels of commonsense would probably be able to take over many humdrum chores, but without commonsense, it would be much less useful.

(Wolff, 2018, p. 3).

Artificial [commonsense reasoning and commonsense knowledge] at human levels has considerable potential to assist in the drafting of legislation and in the preparation of legal cases.

(Wolff, 2018, p. 3).

AI pioneers...were also hoping for systems with general intelligence. Each human-like capacity they modelled—vision, reasoning, language, learning and so on—would cover its entire range of challenges.

(Boden, 2016, p. 22).

The examples above indicate the human version of commonsense is AI's goal and benchmark even when commonsense is poorly understood overall. As a result, artificial general intelligence (AGI) becomes anthropocentrically defined and measured. Again, this is unreasonable: many other forms of complex and intelligent behaviour exist in nature (Korteling *et al.* 2021) including “surprisingly complex forms of communication and social learning in bees, the feats of long-range navigation in migratory birds, or the astonishing memory and tool use found in corvid birds” (Shevlin *et al.* 2019, p.2).

Interestingly, emphasizing the *complexity* of intelligence, especially as a positive trait, seems to be linked to anthropocentric paradigms as well. “In asking how complex, or ‘advanced’ various forms of cognition are, and to compare them between animal species, there is often a tendency to rank various types of plasticity from simple (e.g., habituation, sensitization) through intermediate (from classical conditioning to rule learning) to advanced (human-like, e.g., theory of mind, metacognition, consciousness)” (Chittka *et al.*

2012). Complexity, then, is a mark of high intelligence. Since humans' cognition is complex, humans are considered highly intelligent. However, Zentall (2020) suggests we should question the notion that cognitive complexity is inherently positive. From a biological perspective, the ideal survival machine is simple, perhaps even single-celled. Species that survive the longest over time are those *without* the need for things like complex cognitive skills. "From a biological perspective, in general, intelligence can be viewed as making the best out of a bad situation or producing a complex solution to problems that other species have often solved in simpler ways" (Zentall, 2020, p. 397). On this view, complex intelligence has a downside. "As measured by the number of surviving members of a species, intelligence, as a general characteristic, correlates only superficially (and perhaps even negatively) with most measures of evolutionary success" (p. 397). Zentall reminds us that when we are judging non-human animal intellectual capacity, we ought to remember that most have survived very well¹⁶ without the need for complex intellectual skills (Zentall, 2020).

So overall, while Preston (1991) predicted that AI's reorientation toward common sense would result in a focus on non-human animal cognition, that does not seem the case. And, by ignoring non-human animal research, AI is operating on a double standard. I suggest this based on AI's clear commitment to scientific development, testing, and proving (Russell and Norvig, 2016). However, fields like non-human animal cognition and comparative ethology are also responsive to scientific development. How does AI justify discounting these fields' equally rigorous scientific process? This inconsistency is not trivial, and I assert that it should be questioned.

¹⁶ Zentall (2020) points out that non-human animals sustained themselves well until humans threatened their survival.

Section Four – Ontological Errors in Anthropocentric Paradigms

I have so far argued that anthropocentrism is present in AI's research and practice. In this section, I show that AI's anthropocentrism represents ontological errors¹⁷. These errors may produce claims about human and non-human animal intelligence that ignore counterevidence or are otherwise unsubstantiated.

To begin, we turn discussion back to Cartesian dualism. Philosopher Gilbert Ryle wrote extensively about his objections to Cartesian dualism based on what he calls a category mistake (1949). In Ryle's view, category mistakes are a kind of thinking error in which one allocates "concepts to logical types to which they do not belong" (1949, p. 19). A popular example illustrates Ryle's view. "A foreigner visiting Oxford or Cambridge for the first time is shown a number of colleges, libraries, playing fields, museums, scientific departments and administrative offices. He then asks, 'But where is the University?' I have seen where the members of the Colleges live, where the Registrar works, where the scientists experiment and the rest. But I have not yet seen the University in which reside and work the members of your University" (Ryle, 1949, pp. 17-18). In the example, the category mistake occurs when the visitor categorizes "university" as a "building" in the same way the visitor categorizes the university's "library" as a "building." The "university" actually belongs in a category of "concepts" which do not have a physical location. Therein lies a category mistake.

¹⁷ I use the term *ontological errors* as a general category. Errors could be mistaken premises or other general reasoning errors.

Ryle supposed that Cartesian dualism commits this category mistake¹⁸. Cartesian dualism “represents the facts of mental life as if they belonged to one logical type or category” (Ryle, 1949, p. 17). Cartesian dualism represents mind as a substance, which Ryle says is the mistake. This category mistake leads one to picture “the mind as a ghostly counterpart of the body...a non-physical ‘thing’ and the scene or agent of non-physical states, happenings, and acts” (Honderich, 2005, p. 830). Ryle (1949) argues that just as the university has no physical or geographical place, minds have no physical place either.

Ryle (1949) further explains that the tendency to commit this category mistake is intensified by characteristics of human language. Phrases like *my head is full of useless information* suggest that mind and thoughts are a substance with location. Instead, Ryle says, “Statements about perception, memory, belief, and other mental states are nothing more significant than a series of short-hand ways of describing human behaviour of identifiable sorts” (Kemerling, 2011). Talking about minds, Ryle suggests, is really a particular way of talking about bodies (Honderich, 2005). Further, “although traditional language divides the inner (mental and non-spatial) aspect of human life from the outer (bodily and spatial) aspect...efforts to describe the inner life invariably appeal to the language and models of bodily motion and interaction. The only way to speak about my supposedly private mental life is by drawing analogies to physical processes” (Kemerling, 2011). The result of this, Ryle said, is that humans falsely perceive and attribute internalism to our mental states and externalism to our behaviours (1949).

¹⁸ Ryle’s detailed analysis of Cartesian dualism and category mistakes can be found in *The Concept of Mind* (1949). Complete source information is provided in this paper’s references.

AI anthropocentrism seems to commit a particular kind of category mistake, called a type/token error. Type/token errors are often discussed in the context of philosophy of language, including the relationship between humans and our language (i.e., semantics), and the relationship between words and things to which they refer (i.e., pragmatics). Though there is overlap between them, semantics is often described as the exploration of what it is for us to endow words with meaning, while pragmatics is often described as the exploration of words and the things to which they refer (Honderich, 2005). Philosophy of language intersects with philosophy of mind and metaphysics, so it has drawn attention from modern scholars like Russell, Moore, Carnap, Schlick, and Wittgenstein (Honderich, 2005) as well as the ancient Greeks. For example, Plato suggests “the type is an abstract object, not located anywhere in space-time, although its tokens are” (Wetzel, 2018). In Plato’s view, the form of a triangle is an abstract concept while kinds of triangles exist. The form allows one to understand and recognize different kinds of triangles. In Aristotle’s view, a type is “‘in’ each and every one of its tokens, and so can be seen or heard just as the tokens can be” (Wetzel, 2018). For this paper’s purposes, I will use the term *types* to mean abstract concepts and *tokens* to mean particulars of those abstract types (Wetzel, 2018). I will also assume that readers will entertain the possibility that type/token errors exist.

If we examine McCarthy’s research proposals (presented earlier in this paper) through the lens of type/token errors, we can see that McCarthy presents a token, human language, as if it were a type, the concept of language. Similarly, Russell and Norvig’s discussion of non-human animal communication (presented earlier in this paper) presents a token, word-dependent language, as if it were a type, the concept of communication.

With these errors highlighted, we can more clearly see how they shape AI's research and practice and related claims about intelligence. For example, when the token of human language is confused for communication overall, AI's research scope narrows to only one species. This is unfortunate for AI; it is possible that some research questions could be answered by studying non-humans.

We can also see how the reasoning errors embedded in anthropocentrism reinforce notions about intelligence and minds that in turn, preclude AI researchers from considering new and modern theories. For example, the category mistake leads to the belief that natural intelligence must occur in a centralized, human-like brain (Fox, 2017). Believing the latter, AI misses the opportunity to reap lessons that could result from exploring all kinds of life forms including those demonstrating embodied cognition. For example, octopuses have a decentralized intelligence. The majority of an octopus's neurons are in its arms, which can operate independently without input from its brain. "Cephalopods can solve problems through their self-awareness, decision-making, and robust adaptation. Moreover, they can solve problems in environments where human beings cannot even survive without the continual support of resource-intensive equipment" (Fox, 2017, p. 5). Of course, embodied cognition has been challenged by those with and without anthropocentric assumptions. Moreover, committing type/token errors does not necessarily lead to ignoring data any more than avoiding type/token errors ensures important data are included. The point is that anthropocentrism may conceal these reasoning errors, obstructing AI's ability to advance via new methods and approaches (Boden, 2016).

I have shown that despite this positioning, AI operates from anthropocentric paradigms which may contain reasoning errors. These errors, in turn, impact AI research and practice. In the next section, I present findings from this study and discuss theoretical and practical implications.

Section Five – Findings

I engaged in this research to uncover where and how anthropocentric paradigms emerge in AI. My aim was to answer three questions: (1) What reasoning errors are embedded in anthropocentrism? (2) How do these errors impact AI research and practice? (3) What are the implications for AI, intersecting fields, and beyond? The results of this dissertation indicate that anthropocentrism characterizes aspects of the AI discipline. These findings support my hypothesis that despite AI's theoretical position as a species-inclusive field, anthropocentrism emerges in AI practice. Findings also suggest that Cartesian, overt expressions of anthropocentrism have largely been replaced by implicit anthropocentric references. This implicitness houses reasoning errors that further shape AI research and practice. The results support some of the previous literature on anthropocentrism in AI including Davion's (2002) work on dualisms. Findings also support the view that intelligence remains central to what Cave calls the "human story", a centrality that is rarely questioned. (Cave, 2020). The results also support Fox's (2017) claim that AI's problematic framing leads to anthropocentrism. Whereas Preston (1991) supposed that anthropocentrism would decrease over time, my findings contradict that suggestion.

Three unexpected findings from this research merit comment. One is the absence of evidential support for AI's assumptions that human intelligence is superior. This occurs even when literature supporting AI's views is available. For example, though Zentall does not generally advocate for anthropocentrism, he also acknowledges that "taken as a whole, clearly the sum of [human] intellectual capacity, measured in almost any way, exceeds that of other animals" (2020, p. 397). The pattern of omitting or ignoring evidence conflicts with AI's commitment to scientific principles, which was unexpected.

Another unexpected finding concerns the contradictions in AI's accounts of its goals and purpose. Poole, Mackworth, and Goebel (1998) argue that the goal of AI is not to simulate, replicate, or mimic intelligence but to understand intelligent systems. In contrast, others argue that AI's goal is to achieve artificial general intelligence (AGI) with the performance capability of humans (Boden, 2016). This discrepancy may have emerged from AI's early days in which "an author would argue that an algorithm performs well on a task and that it is therefore a good model of human performance, or vice versa" (Russell and Norvig, 2016, p. 3). Mainstream media seems to perpetuate the view that replicating human intelligence is the goal, suggesting that the argument remains unresolved.

The last unexpected finding concerns the inconsistency with which anthropocentric paradigms themselves are used. The inconsistency was not surprising to see across research sources (given the variance in research emphasis, aims, and goals). But the inconsistency within a single source was surprising. For example, Russell and

Norvig (2016) introduce their AI textbook using non-anthropocentric terms. Then, throughout the book's chapters, anthropocentrism emerges.

Of course, these three findings are not enough to claim the field of AI is characterised by inconsistency overall. Taken together, however, they suggest a pattern of internal discrepancy within AI that warrants further exploration.

Findings also highlight the influence perceptual anthropocentrism plays in AI. The nature of perceptual anthropocentrism means that humans are informed by data we receive from our human body senses (Mylius, 2018), which is to be expected since we are human. However, as AI researchers often point out, human beings have metacognitive capabilities. Therefore, I anticipated metacognition would be instrumental in helping AI researchers think about their thinking, possibly arriving at the conclusion that perceptual anthropocentrism (and other types) could pose challenges to AI's goals. Findings do not suggest that happens very often, however. This finding also illustrates Mylius's point that anthropocentrism types overlap (2018). I suggest the overlap represents a kind of scaffolding: perceptual anthropocentrism may be a foundation upon which other types of anthropocentrism rest. An example of this scaffolding is within AI's assumptions about thought and language. Anthropocentrism's multi-faceted quality may also help explain its insidiousness and recursiveness, which I raise in the next section.

Section Six – Discussion

Implications

Anthropocentrism is insidious, recursive, and durable, which suggests that it is a phenomenon resulting from, or operating as part of, a systemic structure¹⁹. If that is the case, then it may be conducive to a dynamic complexity analysis, rather than a detail complexity analysis. Detail complexity methods are useful to understand the specifics of a particular issue and the details of that issue. Detail complexity methods are very useful, for example, to understand a discrete process that has a start, an end, and complex details in between. By contrast, dynamic complexity has its roots in the disciplines of systems thinking and complexity science. In systems thinking, *structures* are interconnected variables, or networks of relationships, that create behaviour (Senge, 1990). Dynamic complexity methods focus on understanding interrelationships among variables and on the structure's behaviour over time.

In the case of anthropocentrism, the theoretical value of systems thinking is that the interrelationships among variables create or house anthropocentrism can be revealed and better understood. The information we seek to understand anthropocentrism's durability, for example, may reside in the interrelationships in the system rather than within individual parts of the system (Sayama, 2015). In terms of interventions, then,

¹⁹ I mean *structures* in the systems thinking and systems dynamics sense (Senge, 1990). The essence of these structure is not in the things themselves but in the relationships of things. Systems thinking is a useful approach to understanding change and complexity. It is sometimes linked to Eastern cultures' approaches to problem-solving and is an alternative to reductionist thought. For a user-friendly reading on systems thinking, see Senge (1990). For a technical and scientific reading, see Sayama (2015). Complete source information is provided in this paper's references.

systems thinking helps highlight places in system where interventions make the biggest difference (Senge, 1990). A systems thinking view may reveal, for example, that redefining intelligence or improving intelligence measures are unlikely to have the impact on anthropocentrism that many hope²⁰.

In terms of practical interventions, without a comprehensive systems thinking or dynamic complexity analysis, it is difficult to say which interventions would be most impactful. However, we can safely assume anthropocentrism will increase over time. That assumption is evidentially supported by research²¹, but we can also anticipate an increase because anthropocentrism has so far behaved in a *reinforcing pattern*. In systems thinking, reinforcing patterns, or reinforcing feedback loops, are growth engines. They act like a gas pedal in an automobile (Senge, 1990). Since anthropocentrism continues to grow, it suggests that a reinforcing loop is at work. Then, as AI grows, it influences variables in the system with which it is connected.

For example, one variable influenced by AI growth is the increase in AI narratives (Fox, 2017; Cave, 2020). These narratives have been described as fear-based and exaggerated, especially regarding AI's impacts to humans. "According to some prominent voices in the tech world, artificial intelligence presents a looming existential threat to humanity: Warnings by luminaries like Elon Musk and Nick Bostrom about 'the singularity'—when machines become smarter than humans—have attracted millions of dollars and spawned a multitude of conferences" (Crawford, 2016). Others have

²⁰ Hernandez-Orallo (2016), Hofstadter (1979), and Korteling *et al.* (2021) present diverse discussions about redefining intelligence and its measures. Complete source information is provided in this paper's references.

²¹ Martinez-Plumed *et al.* (2019) conducted a quantitative and qualitative analysis of AI's evolution including anthropocentric notions of intelligence. Their research suggests that AI's focus on achieving human intelligence will likely to continue in the future.

described AI narratives as racist and sexist, suggesting that already marginalised people and groups will be most affected (Cave, 2020). This means we should pay attention to “‘people of colour, migrants, unpopular religious groups, sexual minorities, the poor, and other oppressed and exploited populations’ who are not rated highly by the ideology of intelligence, but will be most targeted by intelligent machines” (Cave, 2020).

If we believe the reinforcing pattern (of which anthropocentrism is a part) will continue without intervention, it strongly suggests the need to introduce or bolster a *balancing pattern*, or balancing feedback loop, in the structure. Balancing patterns act like breaks on an automobile. They slow things down, provide resistance, or stabilize (Senge, 1990). AI ethics programs and ethics enforcement are examples of balancing patterns in the sense that the goal is to stabilise and govern AI. It remains to be seen how impactful this intervention will be on AI overall, but the concept is logical enough.

A systems thinking approach to addressing anthropocentrism may have some disadvantages. Sometimes, people resist systems thinking approaches because high levels of collaboration among stakeholders is required. This may not be possible in some situations. Systems thinking may also require learning new methods for analysing problems, which suggests a time commitment and letting go of old habits. Further, systems thinking addresses root causes rather than symptoms of complex issues; while solutions are often more impactful, they may emerge more slowly compared to other problem-solving methods. These are important considerations; not all situations are conducive to the “go-slow-to-go-fast” theoretical approach. However, the desire for quick solutions is usually best balanced with sustainable solutions, in which case systems thinking theories may be warranted.

A final practical implication warrants emphasis. Seeking responsible parties and assigning blame for anthropocentrism seems unproductive. Recognising anthropocentrism as a system, one in which human beings have played a significant role, may be a more productive stance. Just as Cave (2020) suggests that humans take “account of the value-laden history of intelligence...considered the primary marker of the professional elite”, perhaps the same approach would be useful in addressing anthropocentrism in AI. Senge (1991) suggests that the process of systems thinking overturns anthropocentrism because humans are part of the system, not separate from it. Through a systems thinking lens, all of us are influencers of, and influenced by, our reality.

Future Research

Since these findings result from philosophical analysis, it may be useful to consider other research methodologies in future research. I suggest three options. First, it may be useful to conduct an experimental philosophy project with members of AI’s subset research groups (i.e., symbolic AI, connectionist AI, robotics, machine learning). For example, a controlled experiment could be designed in which two symbolic AI research groups are asked to solve a problem; one group could be exposed to an anthropocentric condition while the control group is not exposed to the condition. Results may reveal information about anthropocentrism that non-field research cannot. Second, it would be interesting to test our assumptions about anthropocentrism and its drivers using computer modelling software and systems thinking tools. This kind of software allows a user to input assumptions about how variables in a system will behave in certain

scenarios. The resulting models visually depict the dynamic behaviour of the structure over time. The models can be modified as needed to test new assumptions. This research approach would allow the user to test impacts of anthropocentrism— like species and planetary welfare—in a virtual, digital environment. The models may point to high-leverage interventions where change would make the biggest difference in the overall system. Third, technology fields like AI change very quickly and new information is constantly available. Therefore, future research on anthropocentrism in AI may benefit from employing (non-anthropocentric) AI technologies to aid with searching, analysing, and compiling data that could point to anthropocentrism trends in real-time.

Finally, trends suggest that AI will continue to grow, thus anthropocentrism is likely to grow with it. Many people anticipated AI's technological power, and it may be time to consider its other influences as well. This includes how AI may drive and reinforce anthropocentrism. Anthropocentrism seems to have its own kind of power, and its impact on humans and non-humans deserves to be taken seriously.

References

- Allen, C. (1998) 'Assessing animal cognition: Ethological and philosophical perspectives', *Journal of Animal science*, 76(1), pp. 42-47.
- Allen, C. (2014) 'Models, mechanisms, and animal minds', *The Southern Journal of Philosophy*, 52(S1), pp.75-97. DOI:10.1111/sjp.12072.
- Andrews, K. (2019) 'Can there be belief without language?' *IAI*, 08 January. Available at: <https://iai.tv/articles/can-there-be-belief-without-language-auid-1199> (Accessed: 18 March 2022).
- Andrews, K. and Monsó, S. (2021) 'Animal cognition', *Stanford Encyclopedia of Philosophy*. Available at <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=cognition-animal> (Accessed: 18 March 2022).
- Attfield, R. (2011) 'Beyond anthropocentrism', *Royal Institute of Philosophy Supplement*, 69, pp. 29-46. DOI: 10.1017/S1358246111000191.
- Bekoff, M. (1995) 'Cognitive Ethology: The Comparative Study of Animal Minds', in Bechtel, W. and Graham, G. (eds.) *Blackwell Companion to Cognitive Science*. Oxford: Blackwell Publishers. Available at <https://webarchive.southampton.ac.uk/cogprints.org/159/1/199709004.html>.
- Bickle, J. (2020) Multiple Realizability. Available at <https://plato.stanford.edu/archives/fall2020/entries/multiple-realizability> (Accessed: 03 October 2021).
- Boden, M. (2016) *AI: Its nature and future*. Oxford: Oxford University Press.
- Bringsjord S. and Govindarajulu, N. (2020) Artificial Intelligence. Available at: <https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence> (Accessed: 03 October 2021).
- Buckingham, W. (2011) *The Philosophy Book*. London; New York: Dk Publishing.
- Buckner, C. and Garson, J. (2019) Connectionism. Available at <https://plato.stanford.edu/archives/fall2019/entries/connectionism> (Accessed: 03 October 2021).
- Bugnyar, T., Reber, S.A. and Buckner, C. (2016) 'Ravens attribute visual access to unseen competitors', *Nature Communications*, 7(1), pp. 10506–10506. DOI: 10.1038/ncomms10506.

- Cave, S. (2017) 'Intelligence: a history', *Aeon*, 21 February. Available at: <https://aeon.co/essays/on-the-dark-history-of-intelligence-as-domination> (Accessed: 15 February 2022).
- Cave, S. (2020) 'The problem with intelligence: Its value-laden history and the future of ai' (from the Proceedings of the AAAI/ACM conference on ai, ethics, and society. New York, NY, USA, 7-8 February), pp. 29-35. Available at: <https://doi.org/10.1145/3375627.3375813> (Accessed: 3 October 2021).
- Chalmers, D. (1996) *The conscious mind: in search of a fundamental theory*. Oxford: Oxford University Press.
- Chittka, L. *et al.* (2012) 'What is comparable in comparative cognition?' *Philosophical Transactions*, 367(1603), pp. 2677-2685. DOI:10.1098/rstb.2012.0215.
- Cockburn, D. (2021) Seminar 2: Descartes: the distinction between mind and body, lecture notes, Mind and Body HPPH7011, University of Wales Trinity St. David, delivered 10 February 2021.
- Copeland, B. (2021) Artificial intelligence. Available at <https://www.britannica.com/technology/artificial-intelligence> (Accessed: 03 October 2021).
- Cottingham, J. (1978) 'A brute to the brutes?': Descartes' treatment of animals', *Philosophy*, 53 (206), pp. 551-559. DOI: 10.1017/s0031819100026371.
- Crawford, K. (2016) *Artificial Intelligence's White Guy Problem*, Available at <http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guyproblem.html?partner=bloomberg> (Accessed: 25 March 2022).
- Crosby, M. (2020) 'Building thinking machines by solving animal cognition tasks', *Minds and Machines*, 30(4), pp. 589-615. DOI:10.1007/s11023-020-09535-6.
- Davion, V. (2002) 'Anthropocentrism, artificial intelligence, and moral network theory: An ecofeminist perspective', *Environmental Values*, 11(2), pp. 163-176. DOI: 10.3197/096327102129341037.
- Davis, E. and Marcus, G. (2015) 'Commonsense reasoning and commonsense knowledge in artificial intelligence', *Communications of the ACM*, 58 (9) pp. 92-103. DOI: <https://doi.org/10.1145/2701413>.
- Dennett, D. (1994) 'Consciousness in Human and Robot Minds' Available at: <https://ase.tufts.edu/cogstud/dennett/papers/concrobt.htm>. (Accessed: 20 March 2022).

- Descartes, R. (1993) *Meditations on First Philosophy*. 3rd edn. Translated by D. Cress. Indianapolis: Hackett Publishing.
- Descartes, R. (2017a) *Discourse on the Method of Rightly Conducting One's Reason and Seeking Truth in the Sciences*. Translated by J. Bennett. Available at <https://www.earlymoderntexts.com/authors/descartes>. (Accessed: 03 October 2021)
- Descartes, R. (2017b) *Passions of the Soul* Translated by J. Bennett. Available at <https://www.earlymoderntexts.com/authors/descartes>. (Accessed: 03 October 2021)
- Dreyfus, H. (1965) *Alchemy and Artificial Intelligence*. Santa Monica, CA: Rand Corporation. Available at <https://www.rand.org/pubs/papers/P3244.html>.
- Fox, S. (2017) "Beyond AI: Multi-Intelligence (MI) Combining Natural and Artificial Intelligences in Hybrid Beings and Systems" *Technologies*, 5(3):38. <https://doi.org/10.3390/technologies5030038>.
- Gartner (2021) "Gartner Forecasts Worldwide Artificial Intelligence Software Market to Reach \$62 Billion in 2022." <https://www.gartner.com/en/newsroom/press-releases/2021-11-22-gartner-forecasts-worldwide-artificial-intelligence-software-market-to-reach-62-billion-in-2022>. (Accessed: 28 February 2022).
- Glock, H.J. (2000) 'Animals, thoughts and concepts', *Synthese*, 123(1), pp. 35–64. DOI: 10.1023/A:1005295521736.
- Harper, D. (2022) *Anthropocentric*. Available at http://www.etymonline.com/word/anthropocentric#etymonline_v_26341 (Accessed: 28 February 2022).
- Hernández-Orallo, J. (2016) 'Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement', *The Artificial Intelligence Review*, 48(3), pp. 397-447. DOI: 10.1007/s10462-016-9505-7.
- Hofstadter, D. (1979) *Goedel, Escher, Bach: an eternal golden braid*. Harmondsworth: Penguin.
- Honderich, T. (2005) *The oxford companion to philosophy*. New edition. Oxford: Oxford University Press.
- Humphreys, R. (2008) 'Contractarianism: On the incoherence of the exclusion of non-human beings', *Percipi* (2), pp. 28–38. Available at: http://www.personal.ceu.edu/percipi/archive/200802/03_rh.pdf (Accessed: 02 January 2022).

- Humphreys, R. (2020) ‘Suffering, sentience, and sustainability: An analysis of a non-anthropocentric moral framework for climate ethics’, in Henning, B. and Walsh, Z. (ed.) *Climate change ethics and the non-human world*. London: Routledge.
- Jamieson, D. (1998) ‘Science, Knowledge, and Animal Minds’, *Proceedings of the Aristotelian Society*, 98(1), pp. 79–102. DOI:10.1111/1467-9264.00026.
- Kemerling, G. (2011) ‘Analysis of Ordinary Language’ Available at <http://www.philosophypages.com/hy/6u.htm> (Accessed: 24 March 2022)
- Korteling, J. H. *et al.* (2021) ‘Human- versus artificial intelligence’ *Frontiers in Artificial Intelligence*, 4, pp.1-13. DOI: 10.3389/frai.2021.622364.
- Martínez-Plumed, F. *et al.* (2018) ‘The facets of artificial intelligence: A framework to track the evolution of ai’, Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, 13-19 July. <https://doi.org/10.24963/ijcai.2018/718>.
- McCarthy, J. *et al.* (2006) ‘A proposal for the Dartmouth summer research project on artificial intelligence: August 31, 1955’, *AI Magazine*, 27(4), pp.2-14. Available at: <https://www-proquest-com.ezproxy.uwtsd.ac.uk/docview/208121208?pq-origsite=primo&accountid=130472> (Accessed 03 October 2021).
- McGilchrist, I. (2019) *The master and his emissary: The divided brain and the making of the western world*. New Expanded Edition. New Haven: Yale University Press.
- McGregor, S. (2018) ‘Cognition is not exceptional’, *Adaptive Behavior*, 26(1), pp. 33-36. DOI: 10.1177/1059712318756166.
- Milkowski, M. (2022) Computational Theory of Mind. Available at <https://iep.utm.edu/> (Accessed: 23 March 2022).
- Mylius, B. (2018) ‘Three types of anthropocentrism’, *Environmental Philosophy*, 15(2), pp. 159-194. DOI:10.5840/envirophil20184564.
- Perry, K. (2015) *Philosophy*. New York: Zephros Press.
- Peters, D. *et al.* (2020) ‘Responsible AI-two frameworks for ethical design practice’, *IEEE Transactions on Technology and Society*, 1(1), pp. 34–47. DOI: 10.1109/TTS.2020.2974991.
- Pfeifer, R., Iida, F. and Lungarella, M. (2014) ‘Cognition from the bottom up: On biological inspiration, body morphology, and soft materials’, *Trends in Cognitive Sciences*, 18(8), pp. 404-413. DOI: 10.1016/j.tics.2014.04.004.
- Plumwood, V. (1993) *Feminism and the Mastery of Nature*, New York: Routledge.

- Plumwood, V. (2002) *Environmental culture: the ecological crisis of reason*. London: Routledge. DOI: 10.4324/9780203996430.
- Poole, D., Mackworth, A., and Goebel, R. (1998) *Computational intelligence: A logical approach*. Oxford: Oxford University Press.
- Preston, B. (1991) 'AI, anthropocentrism, and the evolution of intelligence', *Minds and Machines*, 1(3), pp. 259-277. DOI: 10.1007/BF00351181.
- Putnam, H. (1967) 'Psychological Predicates', in Capitan W.H. & Merrill D. D. (eds.) *Art, Mind, and Religion*. Pittsburgh, Penn: University of Pittsburgh Press.
- Rescorla, M. (2020) The Computational Theory of Mind. Available at <https://plato.stanford.edu/archives/fall2020/entries/computational-mind> (Accessed: 03 October 2021).
- Russell, S. and Norvig, P. (2016) *Artificial intelligence: a modern approach*. 3rd edn. London: Pearson Education, Limited.
- Russell, S. and Norvig, P. (2021) *Artificial intelligence: a modern approach*. 4th U.S. edn. Available at: <http://aima.cs.berkeley.edu/> (Accessed: 23 March 2022).
- Ryle, G. (1949) *The concept of mind*. Harmondsworth, Middlesex: Huchinson.
- Salles, A., Evers, K. and Farisco, M. (2020) 'Anthropomorphism in AI', *The Human Brain Project*, 11(2), p. 88–95. DOI: 10.1080/21507740.2020.1740350.
- Sayama, H. (2015) Introduction to the modeling and analysis of complex systems. Binghamton, NY: Open Suny Textbooks. Available at <https://milneopentextbooks.org/introduction-to-the-modeling-and-analysis-of-complex-systems> (Accessed: 15 March 2022).
- Schwab, K. (2016) The fourth industrial revolution: What it means and how to respond. Available at: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond> (Accessed: 02 March 2022).
- Searle, J. (1980) 'Minds, brains, and programs', *The Behavioral and Brain Sciences*, 3(3), pp. 417-424.
- Senge, P. (1990) *The fifth discipline: the art & practice of the learning organization*. Doubleday/Currency: New York.
- Shanahan, M. *et al.* (2020) 'Artificial intelligence and the common sense of animals' *Trends in Cognitive Sciences*, 25(2), pp. 862-872. DOI: 10.1016/j.tics.2020.09.002.

- Shevlin, H. *et al.* (2019) 'The limits of machine intelligence', *EMBO Reports*, 20(10). DOI: 10.15252/embr.201949177.
- Singer, P. (2009) 'All Animals are Equal', in Sterba, J. (ed.) *Ethics: The big questions*. Malden, Mass: Wiley-Blackwell.
- Taylor, P. (2009) 'The Ethics of Respect for Nature', in Sterba, J. (ed.) *Ethics: The big questions*. Malden, Mass: Wiley-Blackwell.
- Thomas, M.S.C. and McClelland, J. L. (2008) 'Connectionist Models of Cognition,' in Sun, R. (ed.) *The cambridge handbook of computational psychology*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511816772.005.
- Turing, A. (1950) 'Computing machinery and intelligence', *Mind*, 59, pp. 433-60. Available at: <https://academic.oup.com/mind/article/LIX/236/433/986238> (Accessed: 07 July 2021).
- Van Gulick, R. (2021) Consciousness. Available at <https://plato.stanford.edu/archives/win2021/entries/consciousness> (Accessed: 03 October 2021).
- Wetzel, L. (2018) Types and tokens. Available at <https://plato.stanford.edu/archives/fall2018/entries/types-tokens> (Accessed: 03 October 2021).
- Wiener, N. (1948) *Cybernetics, or control and communication in the animal and the machine*. Cambridge: MIT Press.
- Zentall, T. (2020) 'Animal intelligence', in Sternberg, R. (ed.) *The cambridge handbook of intelligence*, 2nd edn., Cambridge: Cambridge University Press.
- Ziegler M. *et al.* (2012) 'Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model', *Journal of Research in Personality*, 46(2), pp. 173-183. DOI: 10.1016/j.jrp.2012.01.002.