An Integrated Understanding of AI

Andrew Basden [1]

I was working in AI (artificial intelligence) in the early 1980s but it was a different AI - and yet the fundamental issues are the same now as then. I want to share with you a way of understanding AI that I have discovered since then, which applies to both early and current AI in an integrative way, and is useful in practice [FOOTNOTE: Understanding AI].

This article is written for the 'ordinary' person who knows a little about AI but wants to understand more and, in setting out a new way of understanding AI, might also offer fresh insights for those who know a lot.

Though interest in AI today is at fever pitch in the affluent Global North, especially among politicians, business people, academics and the media, the discourse around AI is often based on spectacle, misunderstandings and even prejudice. Even where not, it is often **fragmented**, with the technical, social, behavioural, ethical and philosophical issues of AI debated in isolation from each other. The way of understanding AI presented here offers an **integrative**, holistic picture of all these issues together. It is a philosophical way of understanding, based on understanding the nature of reality itself, yet which happens to be relatively intuitive. It emerges from the philosophy of Dooyeweerd, a mid-twentieth century Dutch thinker. [FOOTNOTE: Dooyeweerd]

What We Cover

For decades, two main questions were asked about AI,

- Q1: "Could computers ever become like humans?"
- Q2: "Will AI take over the world, making humans extinct or allowing us to live without working?"

Elon Musk recently claimed that AI will do all our jobs; I remember that claim being made in the 1970s too! Since automated cars and ChatGPT burst on the scene, however, other questions have circulated, such as:

- Q3: "Will AI (ChatGPT) write essays for students?"
- Q4: "Will automatic cars kill cyclists who are pushing their cycles?" (one did: [FOONOTE: Cyclist])
- Q5: "Surely AI is better than us at analysing X-rays / finding new chemicals / etc."
- Q6: "Will AI recognise my face and put me at risk? Or my gait?"
- Q7: "How will AI change society?"

The questions express different issues. Q1 and Q2 concern what is called **general AI**, one overtly philosophical, one about eventual possibilities. Q3 to Q7 are more prosaic questions about so-called **narrow AI**, about **particular applications**, with Q3 being about capability of AI, Q4 about AI going wrong, Q5 about which applications might be possible, Q6 about how we use AI to benefit or harm, and Q7 similar but at a society level. Issues of climate and environmental responsibility range across several of them. Q4, Q6 and Q7 have important normative thrust.

When we address such questions, we assume a conceptual framework to help us do so, which itself is based on a set of philosophical ideas. To date, different philosophical ideas inform the debate about the different questions, so that, in the main, the questions are addressed in isolation from each other, with no conceptual framework that can inform our thinking about all seven questions

together. I have discovered that Dooyeweerd's philosophy [Dooyeweerd 1955], a different kind of philosophy, can allow us to address them all.

Often, at this point I would open a section explaining Dooyeweerd's philosophy, but that is not necessary because much of his philosophy is intuitive if we maintain an open mind. Instead I will introduce it bit by bit as we need it for addressing the different questions. This article is in three parts:

- Part 1 explains how AI works and its history - both kinds of AI - to give us perspective.
- Part 2 addresses the application questions, Q2 to Q7.
- Part 3 addresses the abstract, philosophical question, Q1 "AI = Human?", after we have understood the realities of AI.

I bring a slightly Christian perspective, and will refer to religion twice, as contributing insight about ethicality and diversity of meaningfulness, in a way that should be interesting to most people.

Part 1. How AI Works

We need to understand roughly how AI works in order to address any of the questions above. Figure 1 depicts this, for both earlier and current AI. The AI system is a software engine operating with a knowledge base, interacting with users via a user interface (UI) and sometimes with data from the world via sensors, databases or the Internet. (In automated AI the UI might be only a start/stop button, a few controls and data from sensors, but in most AI, like the GPT family, there is more 'dialogue' between users and AI systems.)

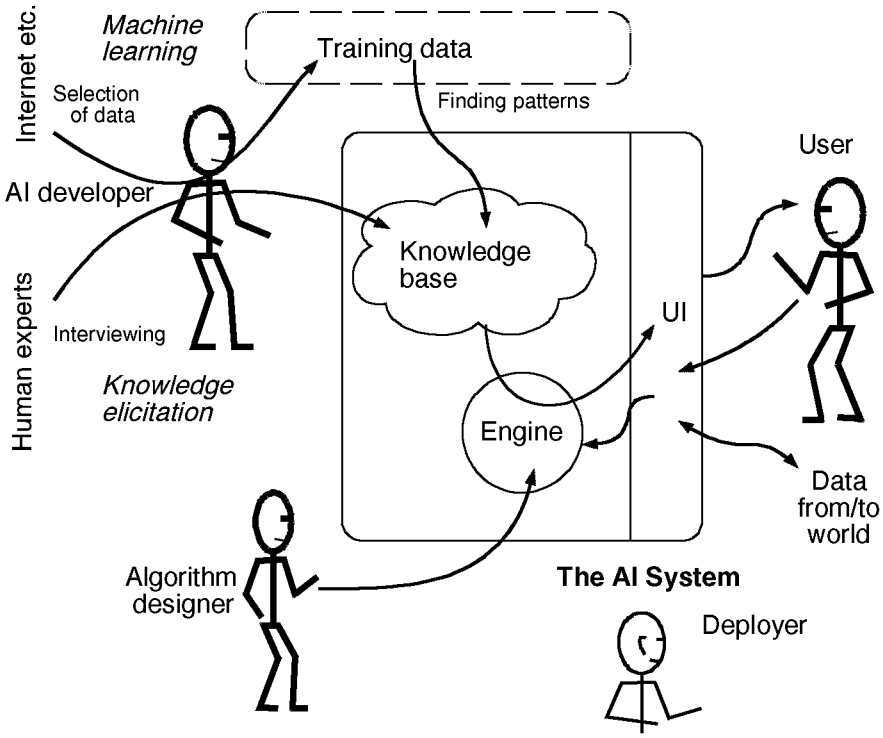

Figure 1. Main elements of AI

The knowledge base encapsulates knowledge about how the AI system should operate in its intended application and is constructed by AI developers. It is based on various technologies, like inference nets, sets of logical statements, sets of associations, or so-called neural networks, with an engine designed and

written to process the encapsulated knowledge according to the technology employed so as to respond to users (or the world). The technology and its engine are created by algorithm designers.

For example, at the core of GPT is a huge matrix of probabilistic characteristics of phrases and words found in billions of statements taken off the Internet (with a lot more around this, such as images). Its engine uses this both to understand user questions or instructions and to generate replies or even essays [FOOTNOTE: (Chat)GPT and how it works].

Users are those who run the AI application in the real world. Deployers, often managers or politicians, are those who decide to invest in or research AI.

Two Kinds of AI

There are two kinds of AI, two ways in which the knowledge base can be constructed, in which the AI developer operates in a different way: human knowledge elicitation and machine learning.

In my early work as AI developer in the 1980s, we would manually build the knowledge base by interviewing human experts and expressing the elicited knowledge in an appropriate computer language: **knowledge elicitation AI** (KEAI). Knowledge engineering, as it was called, was a labour-intensive process, in which good knowledge engineers would winkle out tacit knowledge and rare exceptions as well as more accessible knowledge and incorporate them into the knowledge base.

Today's **machine learning AI** (MLAI) bypasses the human processes of eliciting and expressing knowledge, by detecting patterns in masses of training data supplied to it by AI developers, such as from Reddit in the case of GPT. [FOOTNOTE: MLAI] I like the explanation given by Paul McCartney [Kraftman 2023] of how they used MLAI to extract John Lennon's voice from a poor-quality recording; they told the AI system,

"That's voice. That's guitar. In this recording, lose the guitar."

Why Humans Are Important

How well AI works depends on the quality of knowledge in its knowledge and, of course, on the engine processing this correctly. Since human beings design both engine (algorithm designer) and knowledge base (AI developer), and also use the AI system, even if indirectly, AI cannot be properly understood without taking human intention and interpretation into account.

The quality of KEAI depended on sensitive elicitation and close relationships of trust with experts. Sadly, because AI became fashionable at that time, many became knowledge engineers who would be less careful, so that many AI systems did not work well. Quality of MLAI depends on careful selection of training data and of parameters by which to learn patterns and, having become fashionable again, is subject to the same dangers.

What benefits obtain from AI, and what harm, depends on users and deployers. Misuse and careless use is all too easy, and deployment for nefarious purposes is by no means uncommon.

In both kinds of AI, the quality of the knowledge base and the benefits in use are a **human responsibility**. Part 2 helps us understand what quality and benefits and harm are, and more about the ways humans are important.

A Brief History of AI

Largely, the history of AI is a history of the four human activities around AI, as well as some philosophical development.

The term ″Artificial Intelligence″ was coined in the 1950s, when people were thinking about the capabilities and nature of computers, and wondering whether computers could be like human beings. Alan Turing's famous Turing Test comes from that period [FOOTNOTE: Turing Test]. As computers became more powerful, many began investigating how to get computers to perform human tasks like diagnosing diseases (MYCIN [Shortliffe 1976]) or discover seams of minerals (Prospector [Hart & Duda 1977]) as well as play games like Chess. Many ways of expressing and representing knowledge were investigated, including logic, list-processing, semantic nets, production rules, inference nets, neural nets, ″naive″ reasoning, and even direct spatial reasoning [Funt 1980], which many have forgotten about today. **Knowledge representation languages** became an important area of research. For more see Basden [2008, Chapter VII] and Basden [2018, Chapter 7].

The challenge, however, lay in acquiring the knowledge to represent. It proved insufficient to just obtain rule-based knowledge or logical statements from experts, because much **tacit knowledge** was involved. **Knowledge elicitation** was a skill, not just a technique, and AI sub-field of **Expert Systems** became important. This was my own field in the 1980s.

Because of tacit knowledge and the mediocre level of much knowledge elicitation, along with a technology-oriented rather than user-oriented approach, many AI systems performed poorly, especially when exposed to ′real life′, and yielded no real benefits. Because I attended carefully to tacit and exceptional knowledge and to actual benefits that might accrue in use, most of my Expert Systems were relatively successful, with three coming into beneficial use [FOOTNOTE: Expert Systems Projects]. From this experience a ″Client-Centred Approach′ emerged to guide AI development [FOOTNOTE: CCA].

However, a seminal book [Winograd & Flores 1986] greatly reduced the interest in, and funding for, AI in the late 1990s. It argued that computers should not be seen as conversation partners (the Cartesian / Turing AI paradigm of equalling humans) but as extensions of human capability (Heideggerian paradigm). Interestingly, my own work with Expert Systems was from the latter perspective.

AI work never entirely stopped however, and MLAI proved itself somewhat able to overcome the challenge of tacit knowledge. With increased computing power and the availability of massive amounts training data available, MLAI began showing itself successful in defeating Chess grand masters and Go masters, analysing X-rays, etc., culminating, in the public′s awareness in 2023, of the wonders of ChatGPT. Now that AI has notable successes in some applications (benefits in use), it has become very fashionable to deploy it.

The rest of the history of AI has yet to be played out, but my guess is that it will severely increase climate change emissions, ecological damage, injustice and mental stress for most humans, at the very time we need to take action to reduce those. The effect will be indirect, and not legally attributable to any one AI person. To understand the history and future possibilities of AI, we must be able to answer questions like those listed earlier, which we attempt in Parts 2 and 3.

Part 2. Understanding What AI Can And Should (Not) Do

AI can beat us at Go and Chess. AI let an automated car kill a cyclist. AI can analyse X-ray screens very well. ChatGPT can write essays for students, but they are bland and full of errors (“hallucinations”). How may we understand this? As pointed out above,

- Q2: “Will AI take over the world, making humans extinct or allowing us to live without working?” is about the ultimate possibilities inherent in AI, and whether AI will ever be able to take over from humans - and why.
- Q3: “Will AI (ChatGPT) write essays for students?” calls us to understand what makes AI capable - and why.
- Q4: “Will automatic cars kill cyclists who are pushing their cycles?” calls us to understand AI going wrong - and why.
- Q5: “Surely AI is better than us at detecting cancers in x-rays / finding new chemicals / etc.” calls us to understand in which kinds of application AI can be successful - and why.
- Q6: “Will AI recognise my face/gait and put me at risk?” calls us to understand how we use AI to gain benefit or harm - and why.
- Q7: “Will AI change society?” calls us consider impact of widespread use of AI on society and planet, and structures of society - and why.

Q2 can only be addressed after addressing Q3 to Q7.

Q3. What Makes AI Capable?

To understand what makes AI capable the key is to understand about spheres of meaningfulness. The capability of an AI system comes mainly from its knowledge base encapsulating information and laws that are meaningful in aspect(s) of reality relevant to its application: spatial aspect for Chess AI, kinematic aspect for automated cars and lingual aspect for ChatGPT, for example.

But what aspects are there? From several decades of reflection on everyday experience, the sciences and philosophies, Dooyeweerd carefully delineated fifteen that seem to be irreducible to each other (i.e. cannot be explained in terms of each other nor inferred from each other) [FOOTNOTE: Aspects]. Table 1 lists his aspects, along with what the laws of each are about and some typical AI applications in which the aspect is central, which are mentioned in this article. [FOOTNOTE: Laws]

So, for Chess, for example, AI must have a good ‘knowledge’ of the laws of the spatial aspect and for GPT, of the lingual aspect, so that they can operate in them (their **main aspects**) in response to user input or world data.

However, it is more complicated than that. Chess AI must have *some* ‘knowledge’ that is meaningful in other aspects, such as of movement (kinematic aspect) and of goals and strategy (formative aspect). GPT must have *some* ‘knowledge’ of the formative aspect (structure of language), analytical aspect (distinguishing words, phrases and part-words from each other: vocabulary etc.), psychical aspect (especially for colour in pictures), spatial aspect (in pictures), social aspect (it has a database of people and their relationships), and a few others. Such we will call the **secondary aspects**, because they are there to support its operation in its main aspect. We need not be dogmatic about which are main and secondary; the idea of main aspect is here to help us understand, especially for Q5. Some AI systems might have more than one main aspect.

Table 1. Introducing Dooyeweerd's aspects, their laws, with AI applications mentioned here

	Aspect	Laws to do with ...	Applications mentioned
'Successes'	Quantitative	Arithmetic	-
	Spatial	Spatiality	Chess; X-ray analysis
	Kinematic	Movement	Automated cars
	Physical	Energy, forces, causes Physics, chemistry	Molecule design
	Organic / Biotic	Life functions, cells, organs, Organisms, Ecosystems	Face recognition
	Psychic / Sensitive	Stimulus-response	Voice extraction
'Human' aspects	Analytic	Concepts, analysis, logic	Business analysis
	Formative	Design, planning, forming	-
	Lingual	Signifying, speaking, writing	ChatGPT
	Social	Associating, institutions	-
	Economic	Resources, frugality	(Business analysis)
	Aesthetic	Harmony, enjoyment	-
	Juridical	Rightness Reward, punishment	Writing contracts
	Ethical / Attitudinal	Self-giving love	-
	Pistic / Faith	Belief, commitment Ultimate meaning	-

Let us consider **GPT** in more detail, and how the laws of the lingual aspect are encapsulated in it. To 'write' essays, GPT 'analyses' user's instructions about the essay, and 'generates' the text for it. [FOOTNOTE: Scare quotes] Both analysing and generating text operate according to the laws of the lingual aspect. In 1980s KEAI, the laws of the lingual aspect would have been elicited and encapsulated in the knowledge base explicitly and manually, but in MLAI they are 'learned' as patterns found in masses of (humanly-written) texts.

GPT's main knowledge base, which enables both analysis and generation, consists of a host of probabilistic parameters, 12,288 per word or phrase. With this, GPT's algorithm can reason about things like the relationships among words, such as synonyms and which word follows which in various contexts, using conceptually simple mathematical matrix operations. These parameters were calculated by reading vast amounts of Internet content (175 billion pieces as of November 2023). Since all these pieces are results of humans functioning in the lingual aspect (consciously or subconsciously), they together express human beings' functioning in the lingual aspect. It is not divulged what those 12,288 ways are but we may expect some to measure how much the word expresses meaningfulness in each of the fifteen aspects (e.g. "triangle" would be strong in the spatial aspect, with some meaningfulness in the aesthetic and psychical aspects, referring to the musical instrument, and a little in the social aspect, referring to love-triangle), and many to measure meaningfulness in permutations of aspects. (To Dooyeweerd, aspects are "modalities of meaning" that are irreducible to each other and yet intertwine with each other.) [FOOTNOTE: (Chat)GPT and How It Works]

But why is it that, whereas GPT's language use after massive training (and huge climate-change emissions) is **inferior** to that of most 3-year-old children,

who learn with only limited input? A tentative Dooyeweerdian answer (in need of research) might be that children intuitively function and learn in all aspects, according to what is meaningful in each aspect, whereas all GPT's language learning is funnelled through the limited psychical learning of neural net technology, in which laws of other aspect emerge only statistically via pattern detection.

But why does AI make mistakes, such as in automated cars not recognising a cyclist pushing a bicycle. or ChatGPT offering its famous "hallucinations"? That is the issue addressed in Q4.

Q4. Why Does AI Go Wrong?

There are several reasons AI goes wrong. One is **errors in user input** or world data. Three others arise from deficiencies in the encapsulated knowledge.

1. **Erroneous knowledge** in the knowledge base. Because human writings from the Internet contain errors, ChatGPT 'learned' erroneous patterns among correct ones, thus generating "hallucinations". Also, since its word parameters are probabilistic, it sometimes generates inappropriate text even from correct knowledge.

2. **Missing knowledge**. Where tacit knowledge and rare exceptions are absent from a knowledge base, in usage situations where such knowledge would be relevant, the AI makes mistakes or at least gives biased information. In knowledge elicitation, a good analyst will deliberately seek these out but MLAI learns patterns statistically. There is often not enough training data to learn rare patterns reliably, such as cyclists pushing rather than riding bicycles. With its massive training set, this might be less of a problem with GPT.

3. **Missing aspects**. Omitting (most of) a whole aspect omits a whole swathe of knowledge that is meaningful in that aspect. Whole aspects might be missing if the AI developer fails to recognise their relevance and so does not seek knowledge or provide training data and parameters meaningful in them. This becomes problematic especially when AI is used in different contexts, where those aspects are important, because what the AI gives the user is likely to be seriously biased or even false or inappropriate. Most training data for GPT was written by affluent people in the Global North, in which some aspects important elsewhere, such as generosity in Sub-Saharan Africa, have been undervalued [= = =]. "Conservative values" were included among negative characteristics of a town [FOOTNOTE: Value bias].

It is the AI developer who is **responsible** for ensuring high quality knowledge bases. This becomes more challenging in later-aspect applications, as addressed in Q5.

Q5. In Which Applications Can AI Work Well?

In which applications AI is likely to work well (now and in future), can be understood via aspects. The laws of earlier aspects are easier to encapsulate in a knowledge base reliably. for two main reasons. One is that the laws of earlier aspects are more determinative so that, for example, 3 + 4 is always 7 (law of quantitative aspect), whereas those in later aspects are not: a description of something might validly take several forms (lingual aspect).

The other is that the laws of earlier aspects act as a foundation for those of later aspects, so, in principle, encapsulating knowledge of later aspects requires us to encapsulate laws of all earlier aspects too. Since laws of physics depend on three earlier aspects and those of lingual, on eight, we can expect a knowledge base for lingual applications to be much more complex. Moreover,

the middle aspects of human individual functioning are influenced by later aspects too, which can also need encapsulating (e.g. GPT's social database).

Therefore AI tends to work more reliably, and have more successes, in applications governed by the **earlier aspects**, than those governed by later aspects (see Table 1). This explains why X-ray analysis (spatial aspect) is more reliable than ChatGPT (lingual). Those who extrapolate from current successes in AI to "AI will soon be able to do everything" fundamentally misunderstand AI.

However, full reliability is not always needed where AI *assists* rather than *replaces* humans - the next question, Q6.

Q6. How Do We Use AI for Benefit Not Harm?

Whether AI face recognition is beneficial or harmful depends, not just on the AI working properly or wrongly (in its main spatial, biotic, psychical aspects), but also on the role it plays and whether it is used with evil or good intent.

Roles: Most popular discussion of AI systems presupposes them replacing humans, but AI can also assist humans. An example of this was an AI system to advise managers on the strength of business sectors - analytical and economic aspect application - in which I was involved during the 1980s. From information supplied by managers, it estimated sector strength but then actively encouraged them to disbelieve it rather than accept its answers. Inviting them to explore differences between their and its views, it could reveal things they had overlooked, thus refining their knowledge. Knowledge refinement is the very opposite of AI replacing humans [FOOTNOTE: Roles of AI].

In fact, most of the AI in which I was involved was designed to assist rather than replace humans, advising on stress-corrosion-cracking in industrial plants, herbicide use, business sector analysis, budget-setting in construction projects, and helping to write contracts for construction projects [FOOTNOTE: Expert System Projects] - with main aspects being the physical, biotic, economic, economic and juridical respectively. These were designed to draw out, clarify, capitalize on and submit to expertise and wisdom of their users rather than replace them. That might be a reason for a relatively high success rate even in later-aspect applications.

An important characteristic of 1980s KEAI, which is lacking in MLAI, is **transparency** (understandability) of the knowledge by which it works, and it is often this that made Expert Systems useful in assisting humans.

Intent: Whatever role, is AI used with good intent, evil intent or carelessness? Are decisions to invest in or deploy AI made with responsibility and wisdom, or with self-interest and fear of missing out?

Q7. How Might AI Affect Society and Planet?

As I argue for information technology systems in general in chapter 8 of Basden [2018], there are two societal issues, and they apply to AI.

One is **widespread application**: the impact of individual use, whether beneficial or harmful, becomes multiplied by millions. Though the individual car driver contributes only few emissions to climate change, billions of us together contribute one third of total global emissions. The original pioneers of the internal combustion engine did not see this coming. Likewise for AI. What major harm might its widespread use do? For example what is the unforeseen impact of AI choosing advertisements on social media to present to people? We might not be able to predict this in the usual ways, but understanding Dooyeweerd's aspects at least enable us to separate out kinds of impact - such as on biodiversity, health, mental health, friendliness, resource use, trust, etc. - as in

Table 2 below. Using Dooyeweerd's aspects as a checklist helps ensure that no aspects are overlooked [FOOTNOTE: Aspectual checklist].

The other societal issue is **societal structures** that constrain and enable how we live. The most discussed of these is legislation, which defines some things as legal and others as illegal (a distinction that is meaningful in the juridical aspect). But there are two other kinds of societal structure, meaningful in the ethical and pistic aspects: attitude that pervades a society (self-giving and open, versus selfish and self-protective) and mindset that prevails throughout a society (what society deems most meaningful, to be aspired to, to be expected or taken for granted, to be sacrificed for, etc.). In Figure 2, we can see that all three aspects, being post-aesthetic (harmony across the whole), form societal structures that enable or constrain our behaviour in other aspects.

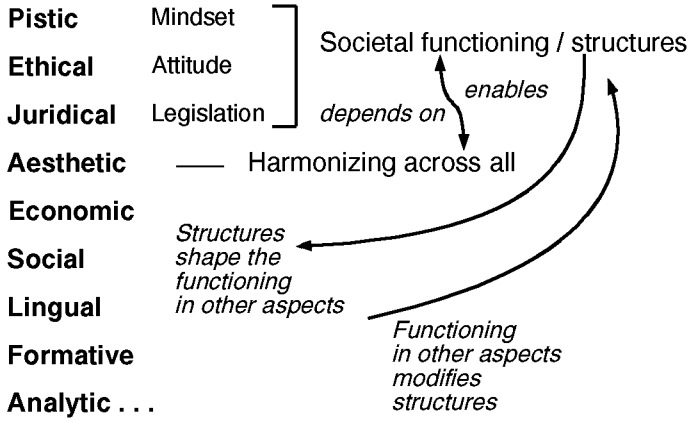

Figure 2. Aspectual structures of society

These structures affect how we function in all other aspects and are themselves modified by what we do in other aspects - a circular relationship between structure and agency [FOOTNOTE: Aspectual Structures]. For example, if AI is used to cheat people (juridical agency), then attitudes of fear and self-protection increase throughout society (ethical aspect structure), which in turn changes the way we relate to and communicate with others (social and lingual agency). Ethical and pistic structures are less visible than juridical structures, but arguably more powerful in their impact. We should remember that legislation is largely ineffective against attitudes and mindsets, deep down, because ethical and pistic problems require more than juridical solutions.

In such ways, Dooyeweerd's aspects can help us think about and discuss societal issues, of both kinds, in AI.

Q2. Will AI Take Over From Humans?

No. Because, to do so effectively, it would have to (a) have encapsulated in its knowledge base the laws of every aspect (b) have done so more completely than humans, and with fewer errors or biases, including cultural. For the reasons discussed above, especially in Q5, I do not believe this is possible. Nor will AI do all our jobs, as Elon Musk believes (though some jobs will change); I remember similar predictions being made in the late 1970s!

The danger from AI, in my opinion, is not AI capabilities but human sin, especially of attitude and mindset. Humanity will tend to use AI in ways that are "affluent, arrogant and unconcerned" FOOTNOTE: Ezekiel]. Such attitude and mindset can affect all four human activities around the AI system, algorithm

design, AI development, and AI use and deployment. They also affect all other activities in our lives and, I submit, climate change, biodiversity destruction and injustice to the Global South are more important issues than AI capability, though they hardly enter discourse around this question, including that on ‘ethics’ of AI.

On the ‘Ethics’ of AI

Increasing numbers of people are discussing what is called the ‘ethics’ of AI [FOOTNOTE: ‘Ethics’]. Sadly, this discussion often takes place in a different mental compartment than that about technology and capabilities of AI, but technology is exciting to develop while ‘ethics’ is not. But in Dooyeweerd’s view they cannot be so separated: technology, capability and ‘ethics’ are inescapably intertwined. The ‘ethics’ is clear in Q4 (“Why does AI go wrong?”), Q6 (“How Do We Use AI for Benefit Not Harm?”) and Q7 (societal concerns) but Q4 is close to Q3 and Q6, to Q5. And Q2 exudes ‘ethical’ concern.

Dooyeweerd’s aspects give kinds of good and evil. Table 2 shows examples of good and evil (dysfunctional) kinds of functioning (columns 2, 3) and of good and harmful repercussions (columns 4, 5) meaningful in each aspect.

Table 2. Aspectual good, evil and harm

Aspect	Functioning	Dysfunction	Repercussions	
			Good	Harmful
Quantit’ive	Amount as given		Reliable sequence	
Spatial	Simultaneity Continuity		Continuous extension	
Kinematic	Movement		Change (non-stasis)	
Physical	Force, causality		Persistence	(Climate change) [1]
Organic / Biotic	Feeding, reproduction	Starvation, suffocation	Vitality, survival	Disease, extinction
Psychic / Sensitive	Interaction	Insensitivity	Emotional and sensory vitality	Sensory, emotional deprivation
Analytic	Distinction	Conflation	Conceptual clarity	Confusion
Formative	Working, planning, constructing	Laziness, destroying	Achievement, construction	Lost opportunities, destruction
Lingual	Expressing, signification	Deceiving	Information	Misinformation
Social	Relating, befriending	Disdaining, hating	Friendship, amplified activity	Working against each other
Economic	Frugality	Squandering	Prosperity	Waste, poverty
Aesthetic	Harmonizing	Fragmentation, narrowing	Integrality, interest, fun	Fragmentation, boredom
Juridical	Giving due, responsibility	Irresponsibility	Justice	Injustice
Ethical / Attitudinal	Self-giving love, vulnerability, trust	Selfishness, self-protection	Culture of goodwill	Competitive, harsh culture
Pistic / Faith	Belief, courage, commitment	Idolatry [2], disloyalty	High morale in society	Loss of meaning, morale

1. Strictly, climate change is not harmful physically but biotically, juridically, etc.
2. Idolatry: Treating something non-absolute as absolute

By separating out the above questions, we may begin to distinguish different types of 'ethical' issues, and Dooyeweerd enables us to think explicitly about kinds of 'ethicality' across them all or, as he calls it, "normativity" because each aspect implies some kind of good, and most also a corresponding kind of harm or evil.

In asking Q4, "Why does AI go wrong?" we focus on the **'ethics' of AI development** by asking in what ways the knowledge encapsulated might be deficient or wrong in any aspect. Such deficiencies result from the dysfunctioning of the AI developers and algorithm designers in various aspects. For example, do they function transparently or deceitfully (lingual aspect)? Carefully or carelessly in giving due regard to all meaningful knowledge (juridical aspect)? Cooperatively or competitively (social aspect)? Generously or selfishly (ethical aspect)? And so on. All these affect the quality of their knowledge bases and algorithms, often in subtle ways that only become evident later on.

In asking Q6, "How Do We Use AI for Benefit Not Harm?" we focus on **'ethics' of use and deployment**, and Dooyeweerd's aspects enables us to understand reasons. It is obvious that using face-detection AI to find someone you want to kill is evil, whereas using face-detection AI to find people who are starving so as to be able to bring food to them is good, but why? Because of the juridical norm of justice and due, and of the ethical and biotic norms of self-giving love about sustenance. Using GPT to find information to help a student write a better essay can be good in the lingual aspect; using it to cheat is wrong, in the juridical aspect. Consider every aspect when asking Q6.

In asking Q7 we focus on **societal 'ethics'**, of two types. One is widespread harm or good impacts, in which AI use for particular applications becomes multiplied. For example, the computing power required to train and use GPT is enormous, and so if GPT becomes widely used (e.g. by it becoming embedded in MS Word), it will contribute significantly to further worsening climate change. This is alarming. Even more alarming, such widespread impacts become accepted as 'normal' and even 'necessary', as people, media and politicians become less concerned about them and even resist calls to curb them. This is pistic functioning, which links us to the second societal concern, structures. In my view, it is an example of an evil mindset structure. An example of evil attitude structure (ethical aspect) is selfish unconcern: "Everyone wants more convenient access to information that GPT gives." An example of evil juridical structure is legislation and policy that promote harm rather than good, such as government encouragement of such AI. Even if I am wrong in my views, at least should we not think about such things? Dooyeweerd's aspects offers a framework in which we can do so, and perhaps bring more clarity into the discourse about societal AI 'ethics'.

Religion is usually omitted from AI discourse, but one useful contribution that a religious perspective can make is about mindset and attitude, because most religions have more profound and more extensive understanding of ethical and pistic functionings than do most secular perspectives.

Part 3. Can AI Be Human?

Here we address the philosophical question of whether or not AI can ever become like humans, Q1: "AI = Human?" Of course, AI is not yet fully like humans but many argue that, given time, it will become so. This has been debated for over 70 years. It was raging in the 1980s, when I was an AI developer, but it remains unresolved. Why? And how may we resolve it? The following is a summary of a more detailed discussion in Basden [2008, 207-220].

So far we have blithely talked about AI capabilities that are normally attributed to humans, such as analysing, generating and writing, around which we put scare quotes earlier [FOOTNOTE: Scare quotes]. Is it valid to attribute such capabilities to AI? Why or why not?

Here we discuss two fundamental flaws in the question itself, and offer a way to resolve them.

The Chinese Room Thought Experiment

In 1990 John Searle suggested the **Chinese Room thought experiment**, to demonstrate, he thought, that AI can never be human. His argument may be summarised as follows:

> Suppose I do not understand Chinese, and cannot even recognise Chinese writing from any other shapes. I am in a room with a hole in the wall. From time to time pieces of paper with Chinese writing arrive through the hole, and I must respond by composing replies in Chinese writing and sending them out through the hole. (To recipients, the room seems to understand Chinese.) I have a rule book in English (which I understand well) that tells me how to reply (by drawing shapes) to each received pattern on the basis of properties like its shape and taking into account all previous patterns received and sent. ″Where in this room,″ asks Searle rhetorically, ″is the understanding of Chinese? And how does it differ from my understanding of English?″ He argued that a computer running a program is like following the rule book, and cannot understand in the way human beings do.

Searle argues that biological causality is necessary for understanding, and that the physical causality of computers can never achieve this; humans operate by one while computers operate by the other. In effect, physical causality is ′lower′ while biological causality is ′higher′, the two operating by completely different laws.

Various counter-arguments have been attempted by AI supporters, of which six kinds are found in Boden [1990]:

- The systems reply, that the property of understanding Chinese is an emergent property of the system of room, rules, me, etc.;
- The robot reply, that understanding involves action in the world (in which the symbols I send out go to instruct robotic movements);
- The brain stimulator reply, that all we need is for the program to simulate the operation of brain cells rather than rules directly (which is the principle on which machine learning neural nets is based);
- The combination reply, that putting all of these together is enough for the AI to have the property of genuine understanding;
- The other minds reply, that we cannot know what is in another mind except by the behaviour we see, so if the Chinese Room behaves aright we may say it understands Chinese;
- The many mansions reply, that eventually we will build computers with the right type of causality and these will truly understand.

Boden records Searle′s reply [p.72ff], countering all of them successfully. The debate continues. [FOOTNOTE: Chinese Room] (Some substitute consciousness, intelligence or some other property for understanding.)

Why Has the AI Question Never Been Resolved?

There are two reasons the AI question, ″AI = Human?″, has never been resolved, both of which, incidentally, Dooyeweerd can help us with: Immanence Standpoint and dualistic ground-motives. Neither Searle nor his opponents seem

to be aware of them; both reasons concern deep presuppositions that shape how questions are posed and interpreted and how debate evolves over time.

The **Immanence Standpoint** [FOOTNOTE: Immanence Standpoint], prevalent in Western philosophy from the Greeks onward, presupposes that what is self-dependent, and on which everything else depends [Clouser 2005] (and thus may be explained) is to be found within the world, without any reference to anything that transcends it. In the Chinese room debate, Searle and most opponents largely presuppose that understanding may be found within the room, without any reference to meaningful agency outside. (In the robot view, the robot arms have no agency.)

There is one obvious answer to Searle's "Where is the understanding of Chinese?" which all seemed to miss: in the book of rules. The book of rules is the Chinese Room's knowledge base in Figure 1 and 'contains' the understanding of Chinese. The understanding is 'contained' in the book, by way of the marks on the paper of the book signifying it, but this was placed there by some agency outside the room: human beings who wrote the book. In this answer, however, we must presuppose some origin of meaning that transcends the room. Adopting an Immanence Standpoint prevents us doing so.

Similarly, to address the question "AI = Human?", we must presuppose some origin of meaning outwith the world; Dooyeweerd did so, as we shall see below, because to Dooyeweerd meaning always transcends.

The second flaw concerns the ways Q1 has been posed and interpreted in the debate, which Dooyeweerd called "**ground-motives**". Ground-motives are society's presuppositions about what is most deeply meaningful in reality, and they propel a society's thinking and beliefs over centuries [see Basden 2020, Chapter 5]. The AI question takes the form of "X = Y?" but it can be interpreted differently depending on what X and Y are assumed to mean and even what "=" means. These are determined by ground-motives, of which Dooyeweerd investigated four that have driven Western thought for 2500 years.

Three are dualistic, with two poles, X and Y, which would seems to suit the AI question. Each sees AI and human in different light, with different properties used in comparison, as follows (proponents in brackets):

- The Greek ground-motive of Mind-Matter: "Computers are matter, humans are (partly) mind; can matter generate mind? e.g. Could a dump of my mind into Cyberspace be the real me? Could I live forever that way? (John Perry Barlow)
- The Scholastic ground-motive of Nature-Supernature: "Computers are natural; humans are (partly) supernatural; can computers gain such supernatural characteristics?" e.g. Is the biological causality by which humans operate a kind of supra-physical spark that computers can never have (John Searle)? Or consciousness [Koch 2019]?
- The Humanistic ground-motive of Nature-Freedom: "Computers are determined, machines; humans are (partly) free; can freedom arise from determinative causality?" e.g. Could Emergence Theory explain it (Allen Newell, Systems theory)? Or Quantum mechanics?

However, posing the AI question in any of the dualistic ways is ultimately fruitless because each presupposes a fundamental either-or opposition that no amount of reasoning can bring them together, often driving us into reductionism [FOOTNOTE: Reductionism].

The fourth ground-motive is non-dualistic (in fact pluralistic).

- The Biblical ground-motive of Creation-Fall-Redemption recognises, and encourages us to explore, multiple ways in which the Creation is Meaningful and Good (works well) - involving human, animals, plants and inanimate things (including machines such as computers). Dooyeweerd took up that challenge.

This ground-motive offers a basis for integration rather than opposition because, to Dooyeweerd, all aspects are irreducibly distinct and yet they are all of equal importance and cohere with no fundamental opposition. With this view, the views of AI above each emphasise a different aspect.

It sees the "=" differently. Under it, we no longer compare two types of entity on just two properties, but we recognise multiple aspects in both computers and humans function, and we find at least two ways of comparing.

Addressing "AI = Human?" with Dooyeweerd

If we cast "AI = Human?" in terms of pluralistic meaningfulness that that transcends both computers and humans, such as is understood via Dooyeweerd's aspects, the question may be recast as "Is it meaningful to say that computers, like humans, function in aspect X?" When we do this, we find two ways of answering the question: with and without taking account of humans:

(a) an 'everyday' way in which computers and humans operate together as part of the whole (the humans including designers, fabricators, programmers, users and deployers);

(b) a narrower, theoretical way, in which we take humans completely out of the picture. We treat the computer as a mass of silicon, various doping elements, copper, plastic, etc., all arranged in certain spatial arrangements and subjected to certain electromagnetic forces. (The reason why they are arranged this way is, within this view, irrelevant.)

In the quantitative to physical aspects, answers to both (a) and (b) are "Yes" for both computers and humans. For example computers and humans consume energy (and thus emit greenhouse gases), occupy space, and so on. In these four aspects, computers are like humans. In subsequent aspects, however, the answer is "Yes" if we take humans into account (version (a), column 3), and "No" if we do not (version (b), column 4).

The answer is "Yes" in (a), column 3, because we humans, assign meaning from later aspects to the physical operation of the computer: the way the electromagnetic fields vary and to their spatial arrangements. It is the fabricators' intention to build a computer (formative aspect), which is the reason why the silicon, copper and various doping elements are arranged spatially they way they are. It is the designers' and programmers' intention to produce an application, such as GPT, that is the reason for the initial arrangements (at switch-on and application loaded and start) of electromagnetic forces (in what fabricators would call the computer memory). It is the users entering text into GPT that is the reason for how those forces vary through time.

The answer is "No" in (b), column 4, because, in that view, the aspects that make intention to build a computer, develop GPT and seek answers, meaningful are irrelevant and what happens is described purely in terms of physical forces and energy, and their spatial arrangement and movement [FOOTNOTE: Bits].

So the debate over "AI = Human?" finds a defensible and understandable resolution if it can escape the constraints imposed by the Immanence Standpoint and the dualistic ground-motives. It is not a single Yes-No answer, but

something more nuanced, involving all aspects and whether or not humans are taken into account.

Table 3. Aspectual functioning of AI with and without humans

Aspect	Kernel	(With humans) computers can ...	(Without humans) computers can ...
Quantitative	Amount	calculate	have amount of memory
Spatial	Extension	(find cancers in) x-ray plates	take up space
Kinematic	Movement	(find) routes; (play) Chess	work dynamically
Physical	Energy	(design) new chemicals	consume energy; be heavy
Biotic/Organic	Life	self-repair; (diagnose) illness	
Psychic/Sensitive	Feeling	sense and respond	
Analytic	Distinction	think and analyse	
Formative	Shaping	plan; manufacture	
Lingual	Signification	converse	
Social	'We'	bring people together	
Economic	Frugality	save money	
Aesthetic	Harmony	create music, paintings	
Juridical	Due	write contracts	
Ethical	Self-giving	care for someone	
Pistic	Faith	assume, commit	
		(Brackets indicate an additional aspect, e.g. analyse.)	

Concluding Remarks

We have addressed a broader range of questions about AI than is usual, and found that Dooyeweerd's philosophy is able to help us address them all. This approach offers an integrated way of understanding AI that exhibits an innate holistic harmony, and is philosophically sound. It brings together the two types of AI, technical issues with 'ethical' issues, individuals with society, and many different kind of applications - GPT, x-ray analysis, automated cars, Chess, and so on. And it does so in ways that respect their differences.

At the core of this approach is Dooyeweerd's suite of aspects, which is the conceptual tool we have employed. It has proven successful in research and practice in many areas [Basden 2020, especially Chapter 11]. More than that, Dooyeweerd was clear that the kernel meanings of aspects are better grasped by intuition than in a theoretical attitude of thought - which implies that this understanding of AI need not require philosophical expertise.

Many current issues in AI have not been mentioned here, such as privacy, job losses, losing control to AI, military use of AI, but this has set out a way of tackling them. For each issue, think, ask and discuss: And which of the seven questions apply? How are responsibilities shared among the four human roles in Figure 1? Which aspects are important in AI's development, use, social impact and ethicality?

Footnotes

Note on Understanding AI. The understanding of AI presented in this article is an amalgam of two sources, one being three blogs published by Faith in Scholarship [FiS 2023], the other being my book [Basden 2018] on *Foundations of Information Systems: Research and Practice*, in which I worked out an integrated, holistic understanding of information technology and digital systems, of which of course AI is a species, covering philosophical, technical, behavioural, ethical and societal issues together. The ideas have also been re-ordered and developed, and call for critique and refinement.

Note on Dooyeweerd. See The Dooyeweerd Pages, "http://dooy.info/", and Dooyeweerd [1995].

Note on Cyclist. A cyclist was killed by an automated car. See "https://www.theverge.com/2019/11/20/20973971/uber-self-driving-car-crash-investigation-human-error-results" Notice the mix of human errors here.

Note on (Chat)GPT and How It Works. ChatGPT is one of the GPT family of generative AI applications, somewhere between GPT3 and GPT4. For an excellent, accessible explanation of how ChatGPT works see Lee & Trott [2023]. The vector of 12,288 parameters per word or phrase is called an "embedding"; they are 12,288 because GPT uses a method called Davinci. Though what these parameters signify does not seem to be divulged, most documents I have read suggest they are about the semantics of the word, such as its numeric reality, its physical reality, its biotic reality, its social reality, and so on. If this is so, then each will be a different combination of aspects.

Note about MLAI. The knowledge base in machine learning AI (MLAI) is usually based on neural net technology or associations.

Note on Turing Test. The Turing Test was that if a computer behaves in a way that cannot be distinguished from human behaviour, then it is valid to call that computer intelligent. However, by using Dooyeweerd's aspects, as we do here, we can understand its strength and weaknesses. It relying on surface behaviour of the computer, which is meaningful mainly to the psychical aspect (and maybe a bit of the lingual) and it thereby ignores all others. A full test of this kind should take all aspects into account.

Note on CCA. CCA, the Client-Centred Approach to developing Expert Systems [Watson et al. 1992; Basden et al. 1995] went through the stages of development with tacit knowledge, human relationships and usefulness in mind.

Note About Dooyeweerd's Aspects. The word "aspect" is as used in architecture, where the east and south aspects of a building cannot be inferred from each other. Dooyeweerd's fifteen aspects may be explored by going to the aspect 'home page' at "http://dooy.info/aspects.html" and a summary at "http://dooy.info/aspects.smy.html". The fifteen aspects are Dooyeweerd's best guess at the complete range of ways in which things may be meaningful. Other suites of aspects could be used, but Dooyeweerd's is so far most complete and most philosophically sound; see "http://dooy.info/compare.asp.html". Dooyeweerd was clear that no suite of aspects, including his own, can ever be treated as a final truth, so we take them on trust here as a conceptual tool to help us think, rather than being dogmatic about them.

Note about Laws. Laws here are not like legislation nor social norms, but laws that enable and govern functioning. The law of gravity, for example, is a law of the physical aspect, and it enables masses to stay together. The lingual aspect has laws that enable language to occur, which are deeper than, and apply across,

all languages. Laws of the later aspects are non-determinative, guiding towards what is Good.

Note on Scare Quotes. "Write", "analyse" and "generate" as activity attributed to ChatGPT are put in scare quotes here to introduce the question of whether computers can 'really' do these things, which is discussed in Part 3. From here on we omit the scare quotes.

Note on Value Bias. GPT will echo the majority values and culture of those who wrote the Internet text selected for its training. This is predominantly by affluent, Global North, progressivist writers, to whom the Global South issues are less important and who disdain "conservative values". For the former see Ilube [2022]. For the latter, see Rozado [2023] or hear the experience of one ChatGPT user around the 38th minute of the conversation with John Lennox on "https://www.youtube.com/watch?v=Undu9YI3Gd8".

Note about Roles of AI in Use. Basden [1983] outlines eight roles in which AI could be used and be beneficial. Strangely, there has been little discussion of roles since then, but most of the roles still apply today.

Note on Expert System Projects. Some Expert System projects in which I was involved include: Stress-corrosion-cracking estimator [Hines & Basden 1986; Basden & Hines 1986]. Herbicide advice; *Wheat Counsellor* commercially available by ICI. Business sector advice: *Assistum*. Advising quantity surveyors on budget-setting: *Elsie* [Brandon et al. 1988]. Writing of construction contracts [Brandon et al. 1992; 1994].

Note on Checklists. When we use Dooyeweerd's aspects as a checklist, we are taking them on trust, but should always remember they might be open to question, even though they are probably the best suite of aspects so far available. See note on aspects above.

Note on Ezekiel. God told the prophet Ezekiel [16:49] that affluence, arrogance and unconcern are the reason Sodom was destroyed and Judah would be exiled. Do we see them today, especially among our tech and political leaders?

Note on 'Ethics'. 'Ethics' in scare quotes refers to the usual discourse around AI doing harm versus good. In most cases the discourse is about right and wrong, and legislation, which are actually juridical in meaning. Ethical without scare quotes refers to Dooyeweerd's ethical aspect, concerned with selfish versus self-giving attitude, rather than right and wrong or legislation.

Note on Aspectual Societal Structures. Societal structures or systems are what enable and constrain us to live in certain ways rather than others. The best-known of these is legislation, but pervading attitude and prevailing mindset likewise enable and constrain us towards certain lifestyles. Attitude and mindset may be seen as the culture of a society. Legislation, attitude and mindset are meaningful in the juridical, ethical and pistic aspects. For more, see Basden 2018, 275-279, 297-299, 301].

Note on Chinese Room. For a fuller discussion, see Basden [2008, 210-216].

Note on Immanence Standpoint. The Immanence Standpoint, as Dooyeweerd called it, a presupposition as to the deepest idea of what reality is like. The ancient Greeks presupposed "It exists" to be the most fundamental thing we can say about something, and existence was presupposed to be self-explanatory and self-dependent (and hence Kant claiming that existence is not a predicate). But as Hirst [1991] points out existence is neither. Clouser [2005] offers a good explanation of this, especially the idea of self-dependence. Dooyeweerd rejected the Immanence Standpoint, holding that existence always presupposes meaning, and hence depends on and is explained by meaning, To say that a poem exists is

to say that something is functioning in ways meaningful in the aesthetic aspect (and others).

Note on Reductionism. Reductionism has several forms, discussed in Clouser [2005], including treating only one thing or aspect as important or meaningful, such as reducing everything to money, or trying to explain the entire complexity we encounter in terms of one aspect, such as materialism and evolutionism do. Trying to break out of reductionism is system thinking, which tries to accept multiple aspects but paradoxically it keeps on being drawn back into reductionism. Dooyeweerd offers a useful conceptual tool to help this.

Note on Subject and Object. In philosophical terminology used by Dooyeweerd, (b) is subject-functioning and (a) is any meaningful functioning whether as subject and/or object.

Note on Bits. It is commonly thought that "the computer is only ones and zeros" (which are called "bits" in digital systems). This is not strictly true. a bit of value 1 can be implemented electronically as a voltage of 3v or 5v or 12 v, as a current flowing, or as a phase change in an AC current, etc. The bit-value is an attribution by humans to a physical phenomenon from the perspective of the psychical aspect, in which signals are meaningful. Moreover, there are also analog computers which do not operate with bits. To speak about bits is to describe the computer from the perspective of the psychical aspect.

References

Basden A. 1983. On the application of Expert Systems. Int. J. Man-Machine Studies, 19:461-477. Available at "http://kgsvr.net/andrew/-p/ai/Basden83-ApplicES.pdf"

Basden A, Hines JG. 1986. Implications of relation between information and knowledge in use of computers to handle corrosion knowledge. *British Corrosion Journal, 21(3):157-162.*

Hines JG, Basden A. 1986. Experience with use of computers to handle corrosion knowledge. British Corrosion Journal*, 21(3):1511-156.*

Basden A, Watson ID, Brandon PS, 1995. Client-Centred: An Approach to Knowledge Based Systems*. CLRC: Rutherford Appleton Laboratory, U.K. ISBN 0 9023 7635 7.*

Basden A. 2008. Philosophical Frameworks for Understanding Information Systems*. IGI Global Hershey, PA, USA.*

Basden A. 2018. Foundations of Information Systems: Research and Practice*. Routledge, London, UK.*

Basden A. 2020. Foundations and Practice of Research : Adventures with Dooyeweerd's Philosophy *Routledge, London, UK.*

Boden MA. 1990. The Philosophy Of Artificial Intelligence*. Oxford University Press, UK.*

Brandon PS, Basden A, Hamilton I, Stockley J. 1988 Application of Expert Systems to Quantity Surveying*. The Royal Institution of Chartered Surveyors, London. ISBN 0 85406 334 X.*

Brandon PS, Hibberd P, Basden A, Kirkham JA, Tetlow S. 1992. Intelligent authoring of construction contracts *The Royal Institution of Chartered Surveyors, London.*

Brandon PS, Watson ID, Basden A. 1994. Professional Involvement in the Development of Expert Systems for the Construction Industry. Journal of Computing in Civil Engineering *(volume, page numbers unknown).*

Clouser R. 2005. The Myth of Religious Neutrality; An Essay on the Hidden Role of Religious Belief in Theories*. University of Notre Dame Press, Notre Dame, Indiana, USA.*

Dooyeweerd H. 1955. A New Critique of Theoretical Thought, Vol. I-IV, Paideia Press (1975 edition), Jordan Station, Ontario.

FiS 2023. Three blogs on AI:
Blog 1, Can AI be human?
"https://thinkfaith.net/2023/10/02/can-ai-be-human/"
Blog 2: How AI works: Why Humans are needed
"https://thinkfaith.net/2023/11/21/why-are-humans-important-in-ai/"
Blog 3: How well can AI work and why?
"https://thinkfaith.net/2023/12/11/how-well-can-ai-work-and-why/"

Funt BV. 1980. Problem-solving with diagrammatic representations. *Artificial Intelligence*, 13(3), 201-30.

Hart PE, Duda RO. 1977. *Prospector: A Computer Based Consultation System For Mineral Exploration*. SRI, Menlo Park, California, USA.

Hirst G, (1991), "Existence assumptions in knowledge representation", *Artificial Intelligence*, **49**:199-242.

Ilube C. 2022. The hidden biases behind ChatGPT.
"https://medium.com/@ilubechristian/the-hidden-biases-behind-chatgpt-dcc3011f37d3"

Kraftman T. 2023. *Paul McCartney is using AI to create a "final Beatles record*. Available at "https://guitar.com/news/music-news/paul-mccartney-is-using-ai-to-create-a-final-beatles-record/".

Koch C. 2019. Will Machines Ever Become Conscious? *Scientific American*, December 1, 2019. "https://www.scientificamerican.com/article/will-machines-ever-become-conscious/"

Lee TB, Trott S. 2023. A jargon-free explanation of how AI large language models work. Available on "https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/".

Rozado D. 2023. The political biases of ChatGPT. *Social Sciences*, 12, 148.

Shortliffe E H. 1976. *Computer Based Medical Consultation: MYCIN*. Elsevier, New York, USA.

Watson ID, Basden A, Brandon P. 1992. The client-centred approach: expert system development. *Expert Systems*, 9(4):181-188.

---

[1] Andrew Basden is Professor Emeritus in Human Factors and Philosophy in Information Systems, from the University of Salford. Email for correspondence: <basdentab@gmail.com>

Version: 10 January 2024.