

Some Wisdom About Artificial Intelligence (AI)

The Role and Responsibility of Humans and Where and Why AI might succeed or fail

Andrew Basden

Abstract: You see more clearly when you look at things from afar. In this article, to gain some understanding of AI today, we look at current AI from the perspective of 1980s 'knowledge elicitation' AI, from which we learn some lessons for today and, to understand what we see, we employ Dooyeweerd's multi-aspectual philosophy.

Most recent version with active links at: "<http://dooy.info/using/wai.html>".

Youtube video of this content: "https://youtu.be/XrQjJML8BhM?si=AXPtpLOOUoPEA_kX"

Introduction

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, so what I learned then can be relevant today. In fact, increasingly so. I want to share these with you, along with a way of understanding issues around AI that I have discovered since then [Note: Understanding AI]. It is based on the philosophy of the Dutch thinker Herman Dooyeweerd [Note: Dooyeweerd], which I and others employed to understand information systems in general, of which AI systems are one type.

The purpose of the article is to help people in the 'real world' cope with AI and have a kind of compass towards what is **wise and good**. It is written for the 'ordinary' person who knows a little about AI but wants to understand more, especially AI in the real world. It might also stimulate fresh insights among those who know a lot about AI already, because it sets out a different way of understanding. This is set out for information systems in my book *Foundations of Information Systems: Research and Practice* [Basden 2018], from which this article adapts ideas to AI.

Interest in AI today is at fever pitch among academics, business people, politicians and the media in the affluent Global North. Yet 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. A wise approach is an integrative attitude of thinking that brings all kinds of issues together, and Dooyeweerd can help us do this.

In this article, which was prepared as an invited lecture to the University of Salford Business School, March 2025, I cover two main issues in AI, The Role and Responsibility of Humans in AI, and In Which Kinds of Application Might AI Succeed and Fail, and Why. I will not delve much into the technological details of algorithms or programs of Transformers, Reinforcement Learning, Supervised Fine Tuning, and the like, though some (briefly explained) reference may be made to them below. Nor will I discuss the philosophical question of whether AI can eventually be like a human nor whether AI will take over the world.

I start by relating something of my experience in early AI and some of the lessons I learned, then introduce you to Dooyeweerd's philosophy and then use this to help us understand the importance of humans in AI, and offer a guide to understanding and predicting AI success and failure. So, this article has four parts:

- ◆ Part 1. My experience in AI and what I learned.
- ◆ Part 2. Brief Introduction to Dooyeweerd's Philosophy: Aspects of Real Life;
- ◆ Part 3. The Role and Responsibility of Humans in AI;
- ◆ Part 4. In Which Kinds of Application Might AI Succeed and Fail, and Why.

Smaller text is used for examples and for text you can skip.

Part 1. My Experience of Working in AI

Why is experience from early AI important? Because some of the lessons learned then have been forgotten in the optimism about current AI, but the realities of life are showing that **some of what was learned then needs to be re-learned** - which can happen the hard way by making many mistakes that do a lot of harm, or by learning from those who still have the expertise and can share it. I am one of those. My experience of (a rather different kind of) AI in the 1980s helped me to understand the relevance of many of the aspects of life, and especially to learn many important lessons (in bold below) about how to develop and use AI, and about the roles and responsibilities of humans. This is a view from the inside: I was developing AI in business rather than as academic or sociological research.

My first degree was in electronics, my PhD, researching algorithms for computer-aided design, involved computer programming, then I developed an interest in psychology and how the human mind worked and thought that, to get into that field, I would first move into the fascinating and then-fashionable field of AI. In 1980, therefore, I joined ICI, the UK's large chemicals firm, to work on a type of AI known as Expert Systems or Knowledge-Based Systems (KBS).

I was hoping to develop the 'perfect' knowledge representation language (for which I thought I had exciting ideas!) but was told, "No, you will use this knowledge representation language despite its flaws, and go round the company seeking applications to try it out on." That was the best thing I could have done, because it forced me to address issues of complex and uncertain human knowledge, how to relate well to experts (so they would not feel threatened by the technology), and how to make the AI system useful to potential users rather than merely technologically advanced.

First I joined an eminent corrosion specialist to develop expert systems to advise on stress corrosion cracking in stainless steel chemical plant, dangerous when it occurs and highly unpredictable. This taught me more about the **skills in eliciting knowledge**, and the importance of **rare exceptions** in knowledge, and of **tacit knowledge** [Note: Tacit Knowledge]. One example was that two experts gave what seemed contradictory advice. When I probed, with the very useful question, "Why?", I found that one worked with chemistry above 300 degrees Celcius, the other with very different chemistry below that. Oh, the importance of assumptions!

Then I coached him to develop his own expert systems, and he built a system to advise on insulation in chemical plants (SYSLAG). This taught me not only that non-AI people could **develop their own** expert systems, but that they included **multiple kinds of knowledge**. Most of the knowledge was about temperatures, insulation thicknesses, and chemical and physical properties, but he added in social knowledge about how people actually treat insulation in real life: they might put ladders against insulated pipes and stand on the pipes to reach something higher - hence that insulation needed an extra-strong covering. He got SYSLAG into use among his team, and this taught me that the important thing was not technical correctness but **usability**.

This experience led to producing four academic papers, one on the **roles Expert Systems could play** [Basden 1983] and another about the importance of separating and then relating **understanding and experience** (as two different kinds of knowledge) when building a knowledge base [Attarwala & Basden 1985], and two on using expert systems in materials engineering [Hines & Basden 1986; Basden & Hines 1986].

Then I was asked to develop an expert system in a very **different field**, agriculture: to advise farmers on what pesticides to use and when to spray their crops. This came into real-life use, and, under the name *Wheat Counsellor*, was a flagship product at the 1984 Royal Agricultural Show. I learned several things with this. One was that it was important to **embed** the expert system among other technologies (networking and multimedia). The expert system was not a showcase technology by itself, but merely a 'seed' germinating in richer human and technological soil.

But perhaps the most important lesson was that I, as knowledge engineer, had a **responsibility**, not just to my employer or 'contract', but to knowledge itself and to the future (and, since I was a Christian, to God). I knew that many farmers in the 1980s wanted to reduce their use of chemicals (and I was an environmentalist), and so I took the courageous step of asking my experts what they would advise. At first they demurred, but I pressed them - and discovered they did have that knowledge. Such important knowledge would not have entered the expert system otherwise - and sadly usually does not unless the knowledge engineer has a **broader vision and courage** to act on it.

The final expert system at ICI in which I was involved was to advise on business strategy - a very different field indeed, highly uncertain, unpredictable and subjective. The promoter of this expert system was insistent that the user should be actively encouraged to disbelieve the advice the system gave, and to probe reasons for the advice. Its role was not so much to advise as to **refine human knowledge** - especially to uncover aspects of business sectors that its users (sector managers) might have overlooked. It was designed to be used to stimulate discussion among groups of such people. The ability of AI then to **explain its reasoning** was crucial ("I have calculated this because of that; and that because of that other factor; do you want to take them into account?").

Leaving ICI in 1986, I was then hired by the University of Salford to develop an expert system for the construction industry, to advise quantity surveyors on setting a budget for new office developments. Three other modules were developed by colleagues, and I integrated them into one system, ELSIE [Brandon et al. 1987]. It was then developed by the Royal Institution of Chartered Surveyors, and became the second-most **commercially successful expert system** of its time. Though ostensibly to advise on costs and refine users' knowledge, to my surprise, it was also used by those who already possessed the expertise - to ensure they had not overlooked anything: what I call the checklist role, later on.

I then entered academic life, teaching and researching knowledge-based systems, but especially the human side, of KBS development. It emphasised the importance of taking a **client-centred approach** rather than technology-centred approach [Basden et al. 1995] - where the client includes both the deployer and all potential users in Figure 1, and also other stakeholders. I also developed algorithms for building and running expert systems, guided by all that I had learned above, and which will inform Part 2 [Basden & Brown 1996], and my PhD student, Mike Winfield then brought Dooyeweerd's philosophy into knowledge elicitation [Winfield et al. 1996].

Many lessons, which may together be summed up as: **humans are more important in AI than most people today realise.**

Part 2. Brief Introduction to Dooyeweerd's Philosophy: Aspects of Real Life

Human activity, including around AI, is complex. Dooyeweerd's philosophy will help us untangle its multiple, intertwined aspects, and so so in a relatively intuitive way [Basden 2020].

Think about any situation in everyday life - having breakfast, travelling to work, using technology, buying and selling, relating to others, and so on. All exhibit multiple aspects [Note: Aspects]. For example, breakfast has a biotic aspect, of feeding, an aesthetic aspect of enjoyment, an economic aspect of limited time before having to leave for work (as well as cost of food, or being short of an ingredient), and so on. I depict all this in Figure 1, naming 15 aspects of reality that Dooyeweerd discussed, and showing which aspect is relevant in various things.

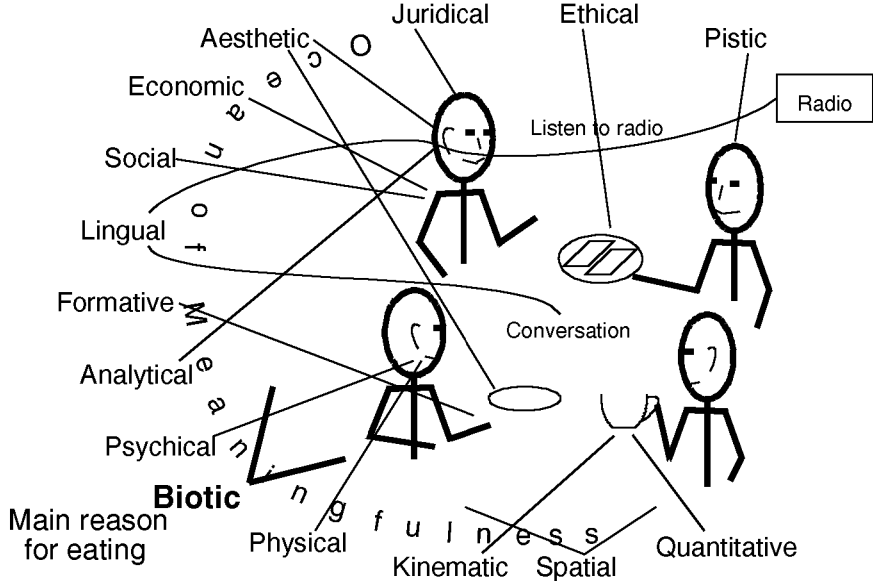


Figure 1. Aspects of eating breakfast

(Note: “pistic” is from a Greek word meaning faith, belief, commitment, etc. - but what we really believe rather than just claim to believe. The physical aspect includes chemistry.)

Those who try to understand having breakfast need to be aware of the many aspects of it. Likewise, those who want to understand AI must **understand its many aspects** - especially if AI is applied to breakfast! Table 1 explains these aspects. In column 1 is the name we give each aspect, in column 2 is its kernel meaningfulness, and in column 3 is an example of how the activity of having-breakfast is meaningful in this aspect.

Table 1. Dooyeweerd’s aspects, and how each is found in breakfast

Aspect	Kernel	... of breakfast (examples)
Quantitative	Amount	Number of cups of tea; amount of muesli
Spatial	Continuous extension	Placement around table
Kinematic	Movement	Lifting cup to mouth
Physical	Energy, force, causality	Chewing food
Organic / Biotic	Organism, life, health	Sustenance (main reason for breakfast)
Psychic / Sensitive	Feeling, emotion, response	Tasting, smelling, texture
Analytic	Distinction	Thinking about which food to have
Formative	Construction, achievement	Making up muesli, cooking food
Lingual	Symbolic signification	Conversation; Listening to radio
Social	Relationships, organisations	Sharing meal together (or not)
Economic	Frugality; resource, Value	Limited time before catching train; Limited budget
Aesthetic	Harmony / enjoyment	Enjoying food; Enjoying company
Juridical	Due, justice, appropriateness	Doing justice to day ahead by eating enough; Respecting others
Ethical / Attitudinal	Self-giving love	Courtesy: offering others the last piece of toast
Pistic / Faith	Belief, commitment	Believing courtesy is important

The word “**aspect**” is used in its ordinary sense, as a way in which something may be viewed or understood, as in architecture where we speak of the east or south aspects of a building. Such aspects imply both the possibility of a viewer, but also that what is seen is not purely subjective but is some reality about what is viewed. Just as in architecture the east aspect of a building cannot tell us what the south aspect is like, nor vice versa, so in wider reality, each aspect is irreducible to others and cannot be derived from others. What Dooyeweerd did was to explore the philosophical nature of such aspects, not only their irreducibility but their inherent, fundamental interconnections and mutual inter-dependence, as modalities of meaning, giving us distinct ways of viewing/understanding reality, spheres of law, giving us an understanding of why functioning is possible and also guidance towards what is Good rather than Bad (e.g. what we eat might sustain us or make us ill; we might offer the last piece of toast or grab the best for ourselves), and modes of being. Each aspect contains analogical ‘echoes’ of the others. Each aspect also give us a different basic kind of rationality or logic. For example, in quantitative rationality, if $X = Y$ and $X = Z$, then $Y = Z$, but in social rationality it is not so: I can be friends with Jim and Joe but that does not mean Jim and Joe are friends with each other.

Our having-breakfast activity is a multi-aspectual functioning in which each aspect is exhibited, often simultaneously: satisfying our hunger (biotic) while tasting it (psychical) and also enjoying it (aesthetic), and so on. Our functioning in later aspects impacts or shape our functioning in earlier aspects; for instance, what we eat (biotic) is shaped by our social situation, our budget, what we enjoy, our religious or ideological observance, etc.

Likewise for all the activities around AI and the functioning of the AI system itself. We will find knowledge of aspectual meaningfulness useful in understanding the content of knowledge bases, of aspectual functioning and Good, useful in understanding the activities of the humans involved in AI and their responsibilities, and aspectual laws and logic helpful in understanding where and why AI fails or succeeds.

There are other parts of Dooyeweerd's philosophy, which we will not use, such as his epistemology, including his theory of theoretical knowing and science, his theory of entities his theory of time and progress, and the role of ground-motives in society; they are not covered here. To understand his entire philosophy, see The Dooyeweerd Pages.

Now let us turn to AI, to understand its meaningfulness within the whole context of life. First, I will dispel the myth that AI can run without humans, by showing the ways in which humans are necessarily involved in AI. Then I look at how to assess in which kinds of application AI can be successful and where it is likely to fail.

Part 3. Artificial Intelligence and Its Humans

It is usually assumed that AI will one day work without humans, but I believe that can never be. Humans are indispensable to AI - in more ways than is usually appreciated - because of the very nature of AI. We need to understand in what ways.

The AI System

The following figure 2 shows roughly how AI works (some readers might already know some of this - but possibly not all). The **AI system** is a software engine operating with a knowledge base, interacting with users via a user interface (UI) and sometimes data from the world via sensors, databases, especially the Internet. What I call the **AI App** (application software) comprises the technological parts: the knowledge base, engine and pattern recogniser (or trainer), and user interface. (In automated AI the UI might be only a start/stop button, a few controls and data from sensors, but in most AI, like ChatGPT and DeepSeek, there is more 'dialogue' between users and AI systems.)

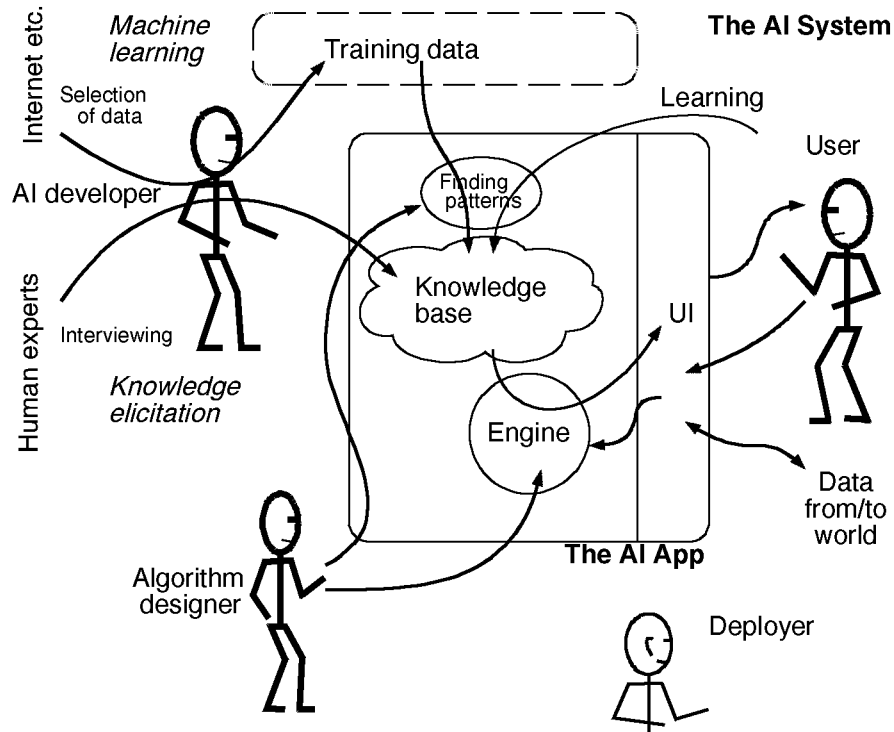


Figure 2. Basic makeup of AI systems

The **knowledge base** encapsulates knowledge about how it should operate in its intended application. In most applications, we find one aspect to be of **primary** importance; so the knowledge base must encapsulate good knowledge of that (for example the physical aspect for SYSLAG, the biotic for Wheat Counsellor and the economic or analytical for the Business Sector Assessor). Table 2 (later) gives primary aspects for more applications. In most applications we also find other aspects important or relevant, which we may call **secondary** aspects. These are important, not usually in their own right, but in supporting the primary aspect. For example, in the SYSLAG Insulation Expert System above, these include the physical-chemical and the pre-physical aspects as the main ones, but also the formative aspect of human intentional behaviour like reaching high pipes, and maybe the social aspect of humans working together. In Wheat Counsellor, the physical aspect of soil, the spatial aspect of proximity to infection, the quantitative aspect of number and amounts of sprays, and the pistic aspect of the belief of farmers about using chemicals, were all important and incorporated into the knowledge base. In the Business Sector Expert System, many other aspects were relevant, including the social and juridical aspects of industrial relations, the pistic aspect of vision for the business, the lingual aspect of discussion and the analytical aspect of analysing.

The knowledge base is constructed by the efforts of an **AI developer**, and the good AI developer will be aware, at least intuitively, of all the aspects so as not to overlook any. Dooyeweerd claimed that the kernel meanings of aspects can be grasped with **intuition** better than with theoretical or analytical thought, and so people do not need to know his philosophy in order to intuitively grasp the importance of social, lingual, ethical, economic, etc.

The **engine** is computer code that runs using the knowledge base to interact with users and/or the outside world. The **pattern recogniser** (trainer) is used in some types of AI to recognise patterns in data so as to insert knowledge into the knowledge base, and it is used both in initial training in machine learning technology (see below) and in improving the knowledge during use. (Here, in the pattern recogniser I include things like Supervised Fine Tuning etc.) Both are composed of algorithms appropriate to the technology being used and are built by an **algorithm designer**.

Several **technologies** are available on which knowledge base and engine can be based, including inference nets, logical predicates, sets of associations, matrices, or so-called neural networks. Whatever technology is used, the knowledge base is an encapsulation (in a computer-readable language appropriate to the technology) of knowledge of whatever is relevant for the intended application. The technology employed in the engine must match that for the knowledge base and encapsulates some of the kinds of reasoning and processing required by the aspects encapsulated in the knowledge base. In inference net or logical AI, the engine must make generic inferences but also contains modules with knowledge of the laws and rationality of each aspect in the knowledge base which work out the rationality or functioning of that aspect. Examples: quantitative calculation for arithmetic and statistics (e.g. Y is sum of squares of X1, X2, ...), spatial inference for spatial things (e.g. shape Y is the merging of shapes X1, X2, X3, ..), kinematic things (e.g. Google Maps route-finding), physical (e.g. laws of physics), lingual (e.g. to do with synonyms, with implications, etc.), and so on. In neural nets and matrices, the engine can, at first sight, be simpler, for example doing matrix arithmetic for everything, but this has merely shifted the burden of aspect-specific reasoning elsewhere.

The **users** are that part of the AI system who use the AI App. In some systems (e.g. a shopping system, where AI-selected advertisements are sent to shoppers), “user” of the AI has two meanings, the shopping company (direct user) and the shoppers (indirect). In use, that is in activity in the world, knowledge is brought to bear, a combination of human knowledge and that in the AI App’s knowledge base. The simple view of this is that:

- ◆ the knowledge base supplies generic knowledge (e.g. in Wheat Counsellor, about farm chemicals);
- ◆ the human user adds specific knowledge, of two kinds:
 - » the situation of use (the user’s farm)
 - » and also of its context (such as the culture of farming).

In reality, the human user supplies some generic knowledge too. When the knowledge base has gaps or errors, the user’s knowledge can fill those gaps or rectify the errors by overriding the output from the AI App. When the knowledge base is in agreement with the user’s knowledge, this makes it easier for the two to work in harmony, especially in a knowledge refinement role (see later). In the Business Sector Assessor, we relied on the human user supplying detailed (sometimes tacit) knowledge by encouraging the user to disbelieve the AI App’s output and explore it, so the knowledge base focused on the analytical aspect of how to help the human clarify and separate out tangled issues.

What I call the **deployer** is the person or organisation that decides to install AI for some of their tasks. The deployer, whether a person or a committee, has the responsibility to understand all the ramifications of the technology, and to ensure that all the conditions necessary for good development and use are met. Deployment requires more **wisdom** than is usually recognised because, though not usually involved in technical or detailed decisions of the other humans involved in AI, the deployer has an important indirect effect on the AI development and use, often subtle and invisible, because, often, the worldview, mindset or attitude of the deployers, who are often senior management, influences or infects all those other humans.

In ICI we had a joke about *Airline Magazine Syndrome*: When we, internally would suggest something new to senior management, they would usually find reasons to resist. But senior managers, on a flight to somewhere, would read in a airline magazine about some new wonderful technology and, on return, would say to their colleagues “We’ve got to have some of this!” So, the way to get senior management to listen was to get a reporter to write in an airline magazine how wonderful our idea was.

That evinces an attitude of mutual distrust between management and developers, which might evoke an attitude of despondency and laxity among developers, which affects the quality of the AI developed.

Thus **humans** determine what the AI system is and does and its impact. But there are two kinds of AI, in which the AI developer operates differently.

Two Kinds of AI

As shown in Figure 2, 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 AI (KEAI)** and **machine learning AI (MLAI)**.

Most early AI in the 1980s was KEAI. In my early work as AI developer, we would manually build the knowledge base by eliciting knowledge from human experts, usually by interviews, and expressing the elicited knowledge in an appropriate computer language. **Knowledge engineering**, as it was called, was a labour-intensive process, in which good knowledge engineers would not only collect knowledge but winkle out tacit knowledge [Note: Tacit Knowledge] and rare exceptions and incorporate them into the knowledge base. Tacit knowledge can be **sensory** (such as in riding a bicycle, playing tennis), or **social** as in what we learned as children on how to get on with people, or **cultural** (as in assumptions made by the affluent Global North), and so on.

Here is a piece of the SYSLAG knowledge base, in which the degree to which mineral wool would be acceptable as an insulation material is calculated using Bayesian accumulation of evidence from several factors (LS, LN are Logical Sufficiency, Logical Necessity). We can see how relatively easily the knowledge base can be understood, leading to good transparency (see below).

```
PROBABILITY min_accept 'mineral wool is an acceptable
insulation material '

    min_temp_ok      LS temp_wt LN 0.001
    min_sections     LS 1.5 LN 0.85
    complex_layout   LS 0.75 LN 1.3
    therm_resist     LS 1.6 LN 0.5
    therm_capacity   LS 0.5 LN 1.3
    min_vibration    LS 0.2 LN 1.0
    ins_sat          LS 0.5 LN 1.2
    high_boiling_flam LS 0.2 LN 1
IF (min_temp_ok >> 0.25)
ELSE 0.001

PRIOR system_prior
```

The IF-THEN is used here, not as a knowledge rule but to represent context (to do with minimum temperature). The following Figure 3 is a screenshot of a whole inference net knowledge base for understanding and assessing aspects of housing. It is not text but each box is like the PROBABILITY above and the arrows are the weighted links to other things. Each node and link is explicitly meaningful, and can be traced.

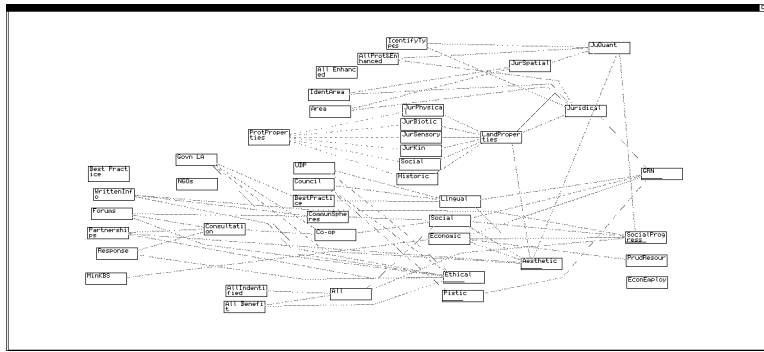


Figure 3. Inference net knowledge base about environmental protection.

Today's machine learning AI (MLAI) bypasses the human processes of eliciting, probing, expressing and representing knowledge, by **training** the knowledge base, by detecting patterns in masses of training data supplied to it by AI developers, such as from Reddit in the case of ChatGPT [Note: Training]. [Note: 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 App two bits of training, one bit of use,

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

A diagram of a neural net would be similar to the inference net, except that the nodes would be arranged in layers, and each would be connected to all in the next and the previous layers, and neither nodes nor links would have any specific meaning, but each would be a kind of mix of the meaning of all the training data.

The training of MLAI is mainly by detecting patterns in training data, but with extra stages (which are discussed later). MLAI training seemingly bypasses the necessity to craft modules that encapsulate the laws of each aspect by reducing them all to a single aspect. This seems to hold out great promise of cheaper AI by reducing labour costs, in that knowledge elicitation can be challenging and labour-intensive, but as we see later that promise is often not kept. Instead of labour, MLAI requires huge amounts of energy in training - and therefore generates huge amounts of climate change gases.

Example: Microsoft had a noble goal of becoming climate-negative (not just net zero) but then they bought OpenAI, makers of ChatGPT, and their climate footprint jumped 29% since 2020! [Donnelly 2024]

If we come to depend heavily on MLAI training, we will lock ourselves into ever-increasing power consumption and climate-change emissions - at the very time that we need to drastically reduce the ecological footprints of the Global North? **Where is the wisdom in that?** Part of the structural problem here is the problem is that the market is distorted by massive subsidies for energy production and taxes on human labour, thus incentivizing a shift towards MLAI training, and part is that MLAI is an unquestioned fashionable technology. **Should we not return to KEAI** for many applications, especially those for which Large Language Models are used? Especially since MLAI does not work so well in such applications (see later).

Why Humans Are Important

It is widely assumed that in MLAI less human input is required (and some look forward to a time when none is required). My experience tells me the first is often wrong, and my philosophy tells me the second is and will always be wrong. **Human input is and will always be essential and important**, and increasingly needed, maybe until we get to the point of making human input central, as in KEAI.

My experience served to emphasise to me the **four main roles humans play** in AI - of each of which I gained some experience while in industry or the surveying profession and continued research thereafter. The four roles are found not only in the 1980s knowledge elicitation AI but even today in machine learning AI, though they might take on a different flavour.

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

Figure 4 shows the impacts that AI and its use can have - five different kinds of impact. In most discussion of AI use, only one is discussed: the Good that AI can do.

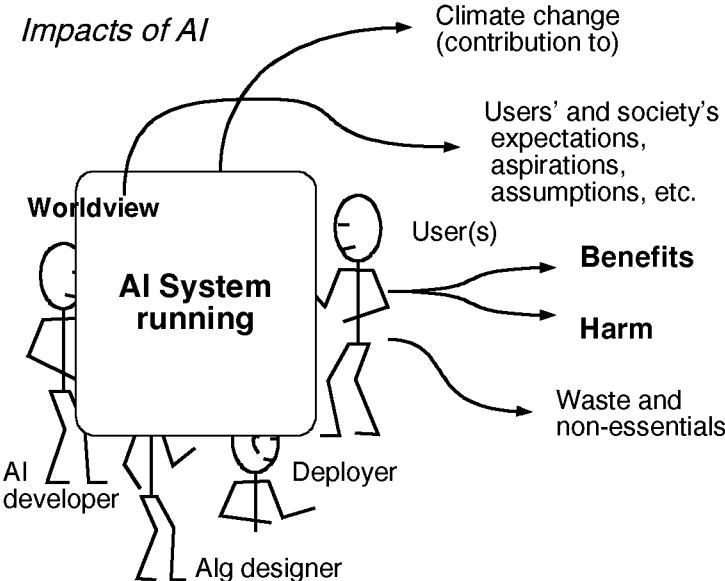


Figure 4. Impacts of AI and its use.

The quality of **KEAI** depends on **sensitive elicitation** and close relationships of trust with experts. Sadly, because AI became fashionable, many became knowledge engineers who would be less careful, so that many AI Apps ended up with poor quality knowledge bases and did not work well. That was one reason AI fell out of fashion during the 1990s.

The equivalent in **MLAI** is that its quality depends on **careful selection** of training data and of the parameters by which to learn patterns, but that is not all. There are three basic strategies for training MLAI: Supervised Learning, with labeled data (of which that is an example), Unsupervised Learning, which finds patterns in unlabelled data (e.g. customer segmentation), and Reinforcement Learning, by trial and error (often using game-playing bots). One might think that only supervised learning requires human input, but in fact the other two do too. While humans tell supervised learning what the labels are and in unsupervised and reinforcement learning the AI works this out for itself, humans must still tell unsupervised and reinforcement learning **what is meaningful** (such as "customer" and what goes along with being a customer, or what constitutes "error"). It is

not that AI cannot generate its own labels, rules etc. from data (because even a rule may be derived from correlation between patterns), but that AI cannot, without humans, discern what is meaningful in making up such rules and hence the basis on which the machine should do so.

Meaningfulness is something that most discussions of the ultimate potential miss, but it is where Dooyeweerd's philosophy can be very helpful because it is a philosophy with meaningfulness at its very centre. Dooyeweerd's aspects can help us clarify the constellations of meaningfulness that we might wish to tell the learning subsystem. [Note: Training]

In Large Language Models extra work is needed to make them more accurate and acceptable. Supervised Fine Tuning (SFT) and Reinforcement Learning with Human Feedback (RLHF), which occur after initial training to make the knowledge base work properly, both involve humans. As Wolfe [2023] puts it for SFT,

"The results of SFT are heavily dependent upon the dataset that we curate. If this dataset contains a diverse set of examples that accurately capture all relevant alignment criteria and characterize the language model's expected output, then SFT is a great approach. However, how can we guarantee that the dataset used for SFT comprehensively captures all of the behaviors that we want to encourage during the alignment process? This can only be guaranteed through careful manual inspection of data, which is i) not scalable and ii) usually expensive. ... As such, SFT, despite its simplicity, requires the curation of a high-quality dataset, which can be difficult."

Notice the word "curation", indicating no mere collecting of data but very careful, imaginative and creative crafting of meaningful sets of criteria.

In addition, other, special-purpose modules might need to be added, which also involves humans to determine what they are and to craft and implement them. For example, the designers of ChatGPT realised they needed to add a social database, of who knows whom etc., and then needed a bolt-on module to prevent GPT returning pornography and violence as answers - both of which involved humans to supply knowledge. [Note: Preventing Porn]

The philosophical reasons why I believe that never will AI need zero human input is because of the need to 'tell' the machine what is meaningful, and which modalities of meaning (aspects) are relevant. Even in MLAI that seems to learn the parameters by itself, it does so from others that have had human input; human input of meaningfulness will always occur somewhere down the line, including in the programming of the engine and pattern recognition algorithms.

Another issue to do with meaningfulness: Harm as well as benefits. Look at any introduction of AI applications, and you will see varied possible **benefits**, but seldom mention of **Harm**, less still on wasteful uses and **non-essentials**, seldom on the huge power consumption and **climate emissions** from AI, and even less on the impact on our **worldview**, mindset, **lifestyle**, etc. Harm can be direct from things like wrong advice or indirect via changing our lifestyles, expectations or aspirations. The AI App (its knowledge base and algorithms) embodies the worldview of the algorithm designer and the AI developer: their way of seeing the world, and what is important to them - usually technology, their own reputation and the economic success of their company, but seldom environmental issues for which we are all responsible (which is why I emphasised the challenge to the farm chemical industry in my story above). Example: Microsoft's seeming downplaying of its commitment to carbon-negativity by opportunistically purchasing OpenAI. It seems that its commitment to carbon-negative was very weak. This is a matter of mindset and attitude.

To make the AI system wise requires human-intensive care. My belief is that, despite MLAI apparently needing minimal human input, in fact most applications involving LLMs **will require considerable knowledge elicitation skills**

and carefulness, probably of the same order of magnitude as those needed in earlier KEAI. Even in these additional modules or refining processes, we will often find that we cannot escape the challenge of rare exceptions and tacit knowledge; mere superficial systems analysis will seldom suffice. Even in scene analysis AI we need to know not just the shapes and visual movements of things, but other aspect of human life too. Example: the rare exception of the cyclist pushing their bicycle who was killed by the driverless car [Note: Cyclist].

Proposal: Use Dooyeweerd's aspects in selection of what is meaningful in all three types of learning. And, if human input is going to become increasingly important, why not grasp the bull by the horns and recognise the need for KEAI even today - and begin to re-learn good ways of doing it? After all, KEAI requires much less energy consumption (by a factor of hundreds or thousands)?

The sad thing is that many of the skills and lessons learned during the KEAI period have been forgotten or lost. I believe that they will **need to be learned again**, and probably more painfully because the AI community does not yet acknowledge the complexity of it and hence is not asking the right questions. Given the idolising of AI by many, much harm is likely to be done before we have re-learned the lessons. That is why I am keen to help: I do not know how long I have remaining on this Earth, for my experience to be offered to the AI community.

It is one thing to acknowledge the importance of humans; it is quite another thing to understand human activity in all four roles and the multiple kinds of human responsibility. In MLAI as in KEAI, the quality of the knowledge base is a **human responsibility**. To which we now turn.

Responsibilities in Creating an AI System

To help us understand and not miss anything, it can be useful to employ Dooyeweerd's aspects applied to the four activities around creating and using an AI App. Each of the four roles implies a different responsibility, shown in Figure 2. The intended AI App needs to be a coherent artefact, and it opens up possibilities in use. In real life new possibilities will open up, unexpected, because human users are creative, which gives a wider diversity of things the AI deployer and developers should consider. They should be aware of a profile of possibilities wider than is expressed in any written specification.

Possibilities imply responsibilities on all humans involved. There are four, shared among the four humans shown in Figure 2. The **algorithm designer** is responsible for ensuring their algorithms - for both training and running the AI App - are not only accurate and devoid of errors but complete. For example, will the pattern-detecting algorithms pick up all that is really relevant and suppress the irrelevant? The **AI developer** is responsible for both anticipating the situations of use and understanding the domain of application, in all its complexity and tacit knowledge.

The **deployer** is responsible for the AI application project overall, and any harm it might cause. Deploying AI is more than mere project management and is a responsibility to ensure the coherence and harmony among all relevant humans and the AI App when the system is in use, and the AI system and its embeddedness in community, society, life and planet. In particular, the deployer should nowadays consider how to **reduce carbon emissions** and other harms caused during training and also when the system is in use.

However, those responsibilities should not be compartmentalised, with an attitude "That's their responsibility, not mine." **Every person** involved, as shown in Figures 2, 3, in creating the AI system should have an attitude that assumes

some responsibility of every kind - but do so courteously. This is why and how I took the initiative in challenging the experts about low-input farming in the case of the Agricultural Expert System.

Dooyeweerd's aspects can help us with both the profile of possibilities and the activities of developing and using AI, as follows.

Multi-aspectual Functioning in Developing a Knowledge Base

To get the knowledge base right is an onerous responsibility, demanding alertness and a knowledge of all aspects, by the AI developer or knowledge engineer. So, as with knowledge engineers in the 1980s and 1990s, so with AI developers today, this is not just a technical or information-theoretic operation, but one of having a broad range of skills, an open mind and, crucially, a generous attitude (ethical aspect) and open-minded and open-hearted mindset (pistic aspect). In fact, it requires good functioning in every aspect, as follows. (Note: we are talking about aspects of human *process* and activity of the developers, not about *content* of the knowledge base, which we were talking about above.)

The following is only a selection of examples; there are many others in each aspect. **Discuss** what these might be.

- ◆ Quantitative: Amount of training data to obtain.
- ◆ Spatial:
- ◆ Kinematic: Change in the knowledge base as it is trained or constructed. also change in from where to seek data or knowledge as the knowledge base grows.
- ◆ Physical: Power consumption and climate change emissions from training the KB.
- ◆ Biotic: Health of the AI developers.
- ◆ Sensitive: Mental and emotional health; Pattern recognition.
- ◆ Analytical: Distinguishing relevant parameters from irrelevant, and not omitting any. (The main, leading aspect in machine learning.)
- ◆ Formative: Achieving a good knowledge base; Planning training and all its stages, and ensuring it is done properly; coping with errors.
- ◆ Lingual: Good understanding of the information used as input or parameters; Good communications with others; Ensuring AI App output will not be misunderstood. (The main, leading aspect in knowledge elicitation.)
- ◆ Social: Working with others, in various roles and with respect.
- ◆ Economic: Frugal use of data resources and energy and own time and the time of others.
- ◆ Aesthetic: Ensuring the finished AI system 'hangs together', that it fits its life contexts, including its technical context, social, ethical, religious and ecological contexts; Enjoying the process of developing
- ◆ Juridical: Appropriateness of the information and parameters used for training; Doing justice to the domain of application (no short-cutting); Ensure that all stakeholders and possibilities of use are considered; Consider all potential repercussions when in use - including impact on climate and biodiversity, on human health and behaviour, expectations, aspirations, lifestyle; treating others involved in development with respect.
- ◆ Ethical: Am I developing this for my own pleasure, monetary or reputational gain, or to bring genuine Good without Harm?
- ◆ Faith: My and our mindset; The ultimate meaningfulness of the system we are developing; Our expectations and aspirations for it.

If any of these functionings is negative (going against the norms of aspects), then a flawed and misleading knowledge base can result.

Multi-aspectual Functioning in Using AI Systems

Likewise, **wise and good use** of the system is multi-aspectual functioning.

- ◆ Quantitative: How much to use AI?
- ◆ Spatial: Geographic locations.
- ◆ Kinematic: Does AI use move or change our ideas, understanding, plans, habits, organisations, etc.?
- ◆ Physical: Physical conditions of use; Climate change emissions generated by use.
- ◆ Biotic: Health impact on us and others of use
- ◆ Sensitive: What impact does using the AI have on mental and emotional health?
- ◆ Analytical: When and when not to use AI; When using it, deciding what information is relevant for its input and interpreting its output.
- ◆ Formative: Structuring information; Planning use.
- ◆ Lingual: Understanding properly and fully the AI output; Expressing things properly in our input (e.g. that the questions or requests made to LLMs properly express what we intend); Communicating with others.
- ◆ Social: Working with others in use, in various roles.
- ◆ Economic: Is our use of AI wasteful or frivolous? [Note: Frivolous Use].
- ◆ Aesthetic: Is our use of AI properly embedded in our life and work?
- ◆ Juridical: Doing justice to all we are involved with; Doing justice to the AI system (using it properly, though with imagination); Seriously considering harm in use as well as benefits.
- ◆ Ethical: Am I using this selfishly, for my own pleasure, monetary or reputational gain, unconcerned about others who might be impacted? Or to bring genuine Good without Harm?
- ◆ Faith: Our mindset in using it; What is our motivation for using AI: because AI is fashionable, and society's idol (fear of missing out), or to bring meaningful Good. What are our expectations in use, our aspirations when using it? Do our religious or ideological beliefs relate to our use?

Multi-aspectual Functioning in Deploying AI

I leave readers to discuss this. See the final section. Use of aspects in such ways is not a tick-box exercise but an attitude of mind. Usually, we should listen to our intuition because, Dooyeweerd said, the kernel meanings of aspects are better grasped by intuition than by analytical or theoretical thinking (which can then be used after intuition to gain clarity).

Part 4. In Which Kinds of Application Might AI Succeed and Fail, and Why.

This section looks to the future: in which applications should we think about using AI, and which not?

Advantages of Each Type of AI

Though the apparent cheapness of MLAI might be false, each kind of AI does have certain **advantages**, which wisdom can understand.

KEAI has an important advantage over MLAI: **understandability of knowledge** (often called "**transparency**"), which is one of seven key requirements in the EU's act on trustworthy AI. The knowledge encapsulated in its knowledge base is conceptualized and explicit, and so, in principle, can be traced and explained. This is especially important when the AI is used to assist rather than replace humans. By contrast, MLAI collapses all the different kinds of knowledge and inference into one kind, usually either vectors or neural nets. This constitutes an aspectual reduction of all aspects to the quantitative or to an analogy of the

psychical; aspectual reduction is almost always harmful [Clouser 2005], because it ignores the laws and rationality of, and obliterates knowledge of, the reduced aspects [Note: Reductionism]. Two examples from two kinds of ML technology:

In neural net technology, all are reduced to an analogy of psychical / mental processes found in the brain, which takes a multitude of signals as input (an analogy of the sensory detectors in eye, ear, skin, etc.) processes these by trained general purpose (simplified analogy of) neural inference and emerges with some result.

In GPT, all are reduced to an analogy of the spatial aspect, as 'points' located in multiple dimensions. At the core of ChatGPT is a huge matrix of probabilistic associations between phrases and words found in billions of statements taken off the Internet (with a lot more around this, such as images), which its engine uses both to understand user questions or instructions and to generate replies or even essays or computer code. Each element in the matrix has quantitative amounts, each measuring the distance along a different 'dimension', over 12,000 of them per element. The matrix engine encapsulates laws of the quantitative and analytical aspects, but the training of the matrix encapsulates some rudimentary laws of the lingual aspect and some pre-lingual aspects, and spatial for images. [Note: ChatGPT].

Strenuous research has been going on to try to make MLAI explainable or transparent [Note: Transparency], but success so far has been very limited because it is fundamentally impossible to derive the meaningfulness, laws or rationality of one aspect from another. I very much **doubt that it will ever fully succeed** without some measure of human knowledge elicitation, because to reduce MLAI operates in pre-analytical aspects and so inherently knows nothing of conceptual distinctions that are necessary for transparency.

Conversely, MLAI has an advantage over KEAI: **inclusion of tacit knowledge**, especially sensory or bodily knowledge [Note: Tacit Knowledge]. The biotic and psychical aspects are pre-lingual and so knowledge of our functioning in them is much more difficult to express than post-analytical (conceptual) knowledge, and hence cannot easily be elicited and represented in a KEAI knowledge base. But, by obtaining data from sensors (including cameras etc.) they can be encapsulated in MLAI by pattern recognition. Some lingual tacit knowledge may be obtained in either technology, either by asking "Why?" and "What else?" during knowledge elicitation, or by detecting patterns in what people have said - which is what ChatGPT does. Insofar as MLAI is trained on data from real life (real-life camera data or what people have actually posted on the Internet) such data expresses what has actually occurred, which arises from actual functioning and conditions in all aspects, whether we know about these functionings and conditions or not; tacit knowledge is where we are not aware of them.

However, the latter is not foolproof. If a LLM looks at material I have posted on a certain website, it will come across several criticisms of one political stance, and hence might infer I am of the opposite stance - which I am certainly not. What it probably misses is *why* I posted that material: it was not because I was against that political stance as such but because of one particular issue within it. Whereas LLMs are good at finding out *what* people write, they might be less good at finding out the *why*.

Moreover, there is no guarantee that all tacit knowledge will be detected, because the AI developer might overlook whole aspects (see below). Also, MLAI has problems with rare exceptions, because often it does not have sufficient data of them to detect their patterns.

The famous example is of a driverless car that killed a cyclist because it did not recognise the cyclist pushing a bicycle (which occasionally happens but not often enough that its pattern had been detected) as a human being and so did not stop. [Note: Cyclist]

So sometimes MLAI is beneficial, and sometimes KEAI is beneficial, and perhaps sometimes they can be combined.

What Makes AI Capable?

The capability of an AI system comes mainly from its knowledge base with an engine that runs it correctly and users who use it wisely and well. But what is knowledge in the AI System? As with knowledge in use, above, we can categorise the various ways knowledge makes AI capable:

- ◆ Knowledge held by the users (specific to the situation and context of use);
- ◆ Knowledge encapsulated in the AI App:
 - » encapsulated in the knowledge base (specific to the type of application);
 - » encapsulated in the engine (and trainer) (specific to the aspectual kinds of application).

Knowledge the users hold is of their specific situation and the social, historical, ethical and religious context in which they are using the AI App. We discuss this in Roles of AI Apps below. Here we examine the encapsulated knowledge in knowledge base and engine (and trainer).

As mentioned earlier, what is encapsulated in the knowledge base is knowledge meaningful to the application - and we have seen that in terms of including knowledge of what is meaningful in various aspects and of the fundamental **laws and rationalities of the aspects** [Note: Laws of Aspects], i.e. how things tend to operate in each aspect, the 'logic' and 'cause and effect' meaningful in each aspect: for example, laws of the spatial aspect for Chess AI, of the kinematic aspect for automated cars, of the lingual aspect for ChatGPT, and of the juridical aspect for fraud detection.

The following Table 2 lists the aspects, with what the laws of each are about and some typical AI applications that are mentioned in this article, in which the aspect is central.

Table 2. Dooyeweerd's aspects, with laws and AI applications mentioned in the article

Aspect	Laws to do with ...	Applications mentioned
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
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	Fraud detection Writing contracts
Ethical / Attitudinal	Self-giving love	-
Pistic / Faith	Belief, commitment Ultimate meaning	-

'Human' aspects

Successes

However, it is more complicated than that because most applications involve other aspects too. Chess AI must have *some* 'knowledge' that is meaningful in other aspects, such as of movement (kinematic aspect) human goals and strategy (the formative aspect). ChatGPT 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**. Some AI systems might have more than one main aspect (e.g. ELSIE); we need not be dogmatic about which are primary and secondary; the idea of primary aspect is here to help us understand.

How can ChatGPT write essays, for example? ChatGPT analyses user's instructions or questions, and generates the text of the essay. Both operate according to the laws of the lingual aspect, which are encapsulated as host of probabilistic degrees measuring how much each word is meaningful in more than 12,000 ways. With this, ChatGPT's algorithm is designed to perform conceptually simple mathematical matrix operations by which the relationships among words can be reasoned about, for example which words tend to follow which in various contexts and which words are synonyms for each other. [Note: How ChatGPT works]. It is not divulged what those 12,000 ways are but we may expect each to represent a different permutation of the fifteen aspects.

This massive knowledge base was constructed by ChatGPT 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. In 1980s AI, the laws of the lingual aspect would have to be elicited and encapsulated in the knowledge base explicitly and manually.

But AI make mistakes, such as in automated cars not recognising a cyclist pushing a bicycle. or ChatGPT offering its famous "hallucinations". Why?

Why Does AI Go Wrong?

There are several reasons AI goes wrong. One is that the engine and pattern detection algorithms contain errors - the responsibility of the algorithm developer. Usually, for well-tested AI software, this is rare. Another, more common, is errors in user input or world data - the responsibility of the user and the deployer to make sure all input is free of errors.

Three types of error, more common still, arise from deficiencies in the encapsulated knowledge - which is the responsibility of the AI developer.

1. **Erroneous knowledge** in the knowledge base. Because human writings from the Internet contain errors, ChatGPT 'learned' some erroneous patterns that generate "hallucinations". Also, since its word associations are probabilistic, it sometimes selects inappropriate ones.

A famous example of an error is that in learning which x-rays showed cancer and which did not, the AI App learned not from the shapes on the screen, which it should have done, but from whether there was a red dot in the margin of the screen (which radiographers had already put there to indicate possible pathology)!

2. **Missing knowledge**: of unusual situations, and minor biases. Sometimes knowledge is missing because of lax selection of parameters and data for training, or lax knowledge elicitation, or simply because knowledge or data are unavailable. 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, example: cyclists pushing rather than riding bicycles [Note: Cyclist].

3. **Missing aspects:** major biases, cultural bias. Omitting 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 seek them or provide training data about them. This becomes problematic when AI is used in different contexts. Example: Most training data for ChatGPT was written by affluent people in the Global North [Caliskan 2017; Atari et al. 2024], which leads to many problems; in different cultures, different sets of aspects are important while others are undervalued.

A high quality knowledge base is one that will have no erroneous knowledge, have complete knowledge of each aspect and in which no areas of relevant knowledge are missing - and it is the AI developer who is responsible for ensuring all three are true. That is, all the relevant fundamental **laws and meanings of all the aspects** relevant to the application need to be encapsulated in the knowledge base, and those relevant to processing and training the knowledge base in the algorithms. If the encapsulation is faulty, we get error type 1. If it is incomplete, we get error type 2. If whole aspects are missing (often aspects are just completely overlooked), we get error type 3. Fortunately, Dooyeweerd's aspects are more fundamental than, and transcend, and apply across, cultures and ages. Understanding this will help us assess in which kinds of application AI can be successful, below.

In Which Applications Can AI Work Well?

In which applications is AI likely to work well (now and in future)? The short answer is that AI is likely to work better in applications meaningful in the **earlier aspects** and less well in **later aspects**.

This is for three 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 a question might be answered in several different ways (lingual aspect).

The second 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. Laws of physics depend on three earlier aspects, those of lingual, on eight. Moreover, the middle aspects of human individual functioning are influenced by later aspects too, which can also need encapsulating (e.g. ChatGPT has a separate social database).

A third is that laws of the earlier aspects (mathematics, physics, etc.) are more precisely and more fully known by humanity, and hence there can be some confidence in selecting training parameters in MLAI and, in KEAI, in the knowledge elicited. So the knowledge base can be more more readily constructed. In KEAI, what this requires is wise knowledge elicitation, open to ideas, sensitive, and able to help experts express their knowledge. In MLAI, it depends on good full data of all possible rare exceptions, not just of the ordinary (e.g. black swans as exception to "All swans are white").

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 2). X-Ray analysis (spatial aspect) is more reliable than is ChatGPT (lingual).

Those who extrapolate from current successes in AI to "AI will soon be able to do everything" fundamentally misunderstand AI. *Do not believe them!*

Roles of AI Apps

However, full reliability is not always needed where AI *assists* rather than *replaces* humans - which brings us to roles of AI Apps. AI Apps may be used in

various roles - as I discovered in my work in ICI [Basden 1983], on which I will draw here [Note: Roles of AI].

When we think about AI we often tend to think about AI **replacing humans**, doing some tasks that humans do but more cheaply, or possibly doing things that humans cannot do, as in robots (robots are only alluded to in Basden 1983).

Examples are XRay analysis and using ChatGPT to write essays or write Python code. In that role, the above applies: the AI App must have perfect, complete knowledge, in principle. MLAI can be good here in early-aspect applications. Complete, perfect knowledge encapsulated in a knowledge base is possible (in principle) only in the earlier aspects, where the laws are deterministic and relatively simple (and, for KEAI, mostly known through the sciences) - such as analysing scenes or X-rays. But in later aspects - such as essay-writing - AI keeps making mistakes. And I believe will always do so. (There is an exception to that: where mistakes don't matter. Do they matter in recommending purchases or pages to look at? Do they matter if the essay contains errors? Discuss.)

In the **advisory role** ("Consultancy role" in 1983), AI can be somewhat useful, because the AI App gives advice and the human knowledge is brought to bear and can override the AI when it fails. Most of the Expert Systems of my experience were used in an advisory role, with some in other roles too. ChatGPT is often used in this role (e.g. finding material for essays). But what happens if the human user is lazy or mistaken? Discuss. This why the Business Analysis Expert System actively encouraged users to disagree with it.

In a **training role**, the AI can help train people. When it is not entirely accurate or complete, it can at least help in initial training. So this might be a role in which MLAI can offer some success in later aspects. Discuss how much and when incomplete knowledge matters.

In a **demonstration role**, the AI App can be run to demonstrate things, or to try out things (simulation). Discuss what happens when the AI App has faults.

In the **checklist role**, as in ELSIE, the AI is used even by experts to ensure they do not miss anything. This usually requires a transparent knowledge base, which is difficult with MLAI. Discuss whether MLAI can be used in this role, and if so, how.

In the **communication role**, and the **repository role** the knowledge base itself is used as an expression of knowledge (and hence needs to be transparent). I found that showing people the knowledge in the knowledge base helped communicate with them, because the knowledge was systematic, and especially, it highlighted exceptions and assumptions. That the knowledge is expressed more systematically than happens in most text, makes it useful in repositories of knowledge. I expect that this role requires the transparent knowledge bases generated by KEAI. Discuss whether MLAI can be used in these roles, and if so, how.

In what I called the **knowledge refinement role** the running of the AI App helped refine human knowledge, when the knowledge was explicitly and transparently expressed ("I, the AI App, believe X because of A, B, C. I believe A because of J, K, L. ...") I found that knowledge refinement occurred even in the experts when constructing the knowledge base, especially when I asked "Why?" The corrosion expert told me later that he would sometimes go away and either think about it, or even run an experiment. Knowledge refinement relies heavily on having a transparent knowledge base. But might MLAI be able to offer some impoverished form of knowledge refinement for example by simulation? Discuss.

So, in such applications, AI can be beneficial in roles other than replacing humans. MLAI can be useful in some, and KEAI in more.

Concluding Remarks

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”). AI is spectacular, having grabbed the adulation of most who lead society in the affluent Global North, and yet our understanding and practice of it is fragmented. How may we understand this, join the dots, and move on wisely?

In this article, I have tried to bring an integrative, holistic view of AI, in which early knowledge elicitation AI and today’s machine learning AI are seen as similar but with different ways of encapsulating knowledge in the knowledge base. I have brought some important lessons from early AI, in which human elicitation of knowledge was crucial because I believe that increasingly today’s machine learning will be in need of those lessons to refine the knowledge learned. Even more important than expertise is wisdom, including a good understanding of the roles that humans and the technological AI App can play, in order to work together to achieve Good rather than Harm.

Responsibility and attitude are important in wisdom - especially by deployers of AI, not least because widespread use of AI is likely to make it extremely difficult to reduce climate change emissions to the level that we can save future generations from massive suffering. It is machine learning that is the culprit; early knowledge elicitation AI consumed much less power but machine learning bypasses the human processes and hence seems cheaper. We must face the question, “Why do we want to bypass the human?” So, in part of this article we look at what really is a key advantage of each kind of AI >

I have introduced Dooyeweerd’s philosophical aspects as a way towards this wisdom, by helping us separate out issues so that we do not confuse them, to understand what is going on, and to guide the activities of developing, using and deploying the AI system. These are “modalities of meaning” which are readily grasped by intuition - and thus understandable by those without either philosophical or even technical expertise. They can help us understand why AI is likely to be more successful in some kinds of application and less in others. In this way, Dooyeweerd’s aspects have proven very practical as well as philosophically sound [Basden 2020].

Notes

Note on Understanding AI. The understanding of AI presented in this article is an amalgam of my experience in AI practice in industry in the 1980s and 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. If you want to take the ideas in this article forward, please read that book. Please contact me if you have problems in doing so.

Note on Dooyeweerd. Herman Dooyeweerd (1894-1977) was a Dutch thinker who rethought philosophy from the bottom up, with radically different presuppositions, presupposing meaning rather than being or process, and adopting the Biblical ground-motive rather than Greek, Scholastic or Humanist ground-motives. In 1955, He contributed a radical transcendental critique of theoretical thought to understand why it is never neutral - prefiguring many of the ideas of postmodern, linguistic and critical thinkers. His most famous idea is his suite of modal aspects, which we use in this article; Part II explains these. Dooyeweerd’s philosophy is explained, discussed and commented on, in “<http://dooy.info/>” The Dooyeweerd Pages.

Note on Tacit Knowledge. Michael Polanyi [1970], in *The Tacit Dimension*, discoursed on tacit knowledge, which may be characterized by “We know more than we can tell.” This is especially true of muscular knowledge, as in playing tennis or riding a bicycle, but it is also true for long-learned skills, social knowledge (e.g. of how to get on with people, which we learn early on from our families), aesthetic knowledge, and so on. It is difficult to express our tacit knowledge because it has become ‘part’ of us. But there are ways to help people do so, which a good knowledge engineer can take.

Note on ChatGPT and How It Works. For an excellent, accessible explanation of how ChatGPT see [Lee & Trott 2023].

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

Note About Dooyeweerd's Aspects. 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 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 as a conceptual tool to help us think.

Note on Training. MLAI training requires humans to tell it what is meaningful. In real-life training, data pre-processing takes up a considerable amount of effort and time, and is done mainly by humans, who clean the data, remove inconsistencies, handle missing values, handle outliers, normalise the data, label it (in the case of supervised learning), and so on. This is followed by training and then model evaluation and refinement. All these tasks require being told what is meaningful and what each attribute is on which the knowledge base is being trained. Ultimately this knowledge has to be supplied by humans and, especially in later-aspect applications, it requires careful knowledge consideration, which old knowledge elicitation methodologies can assist. As O'Keefe [2024] says,

"But model evaluation isn't a one-time event. Organizations must continuously evaluate AI models to ensure they produce the right results. For example, several major US health insurance companies have come under fire and face legal cases around excessive claim denials. Having human oversight to ensure these models aren't making the wrong decisions is critical to prevent poor performance, reputational damage, lowered customer satisfaction, or even compliance fines.

The following links explain training and each one presupposes humans telling the training what is meaningful - but, sadly, none give much in the way of detail.

- ◆ "[How Does AI Model Training Work?](https://appian.com/blog/acp/ai/how-does-ai-model-training-work)" Dan O'Keefe. "<https://appian.com/blog/acp/ai/how-does-ai-model-training-work>"
- ◆ "[What is unsupervised learning?](https://www.ibm.com/think/topics/unsupervised-learning)" IBM. "<https://www.ibm.com/think/topics/unsupervised-learning>"
- ◆ "[How to Train AI Models](https://www.clickworker.com/customer-blog/process-of-ai-training/)" Robert Koch. "<https://www.clickworker.com/customer-blog/process-of-ai-training/>"
- ◆ "[How to Train AI Models Efficiently with 5 Pain-free Steps](https://www.neurond.com/blog/how-to-train-ai)". Trinh Nguyen. "<https://www.neurond.com/blog/how-to-train-ai>"
- ◆ "[Data Preparation for Machine Learning: The Ultimate Guide to Doing It Right](https://www.pecan.ai/blog/data-preparation-for-machine-learning/)"The Pecan Team. "<https://www.pecan.ai/blog/data-preparation-for-machine-learning/>"
- ◆ "[How to Train an AI Model: A Comprehensive Guide](https://medium.com/@saiwadotai/how-to-train-an-ai-model-a-comprehensive-guide-d5aefaa2763d)" Saiwa. "<https://medium.com/@saiwadotai/how-to-train-an-ai-model-a-comprehensive-guide-d5aefaa2763d>"
- ◆ "[How to Train AI Models: Your Complete Guide](https://www.tavus.io/post/how-to-train-ai-models)" Julia Szatar. "<https://www.tavus.io/post/how-to-train-ai-models>"
- ◆ "AI models are trained on datasets to learn patterns, make predictions, and assist with decision-making, enabling task automation and personalized recommendations. Learn the key steps, challenges, and best practices for training reliable AI models". Liz Ticong. "How to Train an AI Model: A Step-by-Step Guide for Beginners"

Note on Preventing Porn. OpenAI hired people in Kenya, and other countries where labour is cheap, to watch endless streams of videos etc. and indicate whether it was pornographic, violent, or not, and by this they trained ChatGPT to exclude them. But constant watching of such stuff destroyed the mental health of many people, and even their lives. What price AI?

Note about Laws. Laws here are not like laws of a land nor social norms, but laws that govern how things function. 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 any one language group. Laws of the later aspects are non-determinative, but they guide towards what is Good.

Note on Transparency of AI. How to make AI transparent or explainable has become a whole field of research and discourse known as XAI, Explainable AI. Larsson & Heinz [2020] review some of the field, and show its breadth. Our focus here is on only one facet of transparency, how to make the knowledge encapsulated in the knowledge base clearly understandable to users. Even so, the range of kinds of issue that an AI System needs to explain is very wide, covering many aspects, from physical through to juridical at least [p.5].

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 Frivolous Uses of AI. Frivolous uses of AI are ones that really do not need to be made, and it is important to avoid these, given the climate and other harm that using AI contributes. We should use AI only for important uses - and certainly not for trivialities, rivalries or selfish gain. I asked a colleague who is involved in AI in business what he thought might be frivolous uses, and he replied with a substantial list of good and frivolous uses, the latter including: entertainment and novelty; marketing and advertising; social media; AI pets or companions. On thanking him, he replied, "I feel like a fraud - albeit unintentionally :-)" I thought it would be helpful to send an AI-generated answer ... I simply expressed your request in my own words as Q and the AI answered A." Ironic! Though it is not frivolous to know which kinds of application might be considered as frivolous, not by 'puritans' like me but by the corpus of human writing as well, to which the AI system refers. What actually is frivolous is a question that needs to be discussed (e.g. a modicum of entertainment in people's lives might be a necessity, but a surfeit thereof and the expense of others or the planet, may be seen as frivolous. As is being discussed in economics under the label of "useless economic activity" ("<http://christianthinking.space/economics/r7-ghu.html>"), Dooyeweerd's aspects offers a basis for considering this.

Note about Roles of AI in Use. Basden [1983] outlines eight roles in which Expert Systems could be used and be beneficial. Strangely, though that paper was much read at the time, there has been little discussion of roles since then, but most of the roles still apply today.

Note on Reductionism. Reductionism has several forms, discussed in Clouser [2005], including treating only one thing or aspect as valuable or meaningful, such as reducing everything to money, and 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. Dooyeweerd offers a useful conceptual tool to help this.

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

References

- Atari M, Xue MJ, Park PS, Blasi DnE, Henrich J. 2024. Which Humans? *PsyArXiv*. Preprint available at "https://osf.io/preprints/psyarxiv/5b26t_v1".
- Attarwala FT, Basden A. 1985. A methodology for constructing Expert Systems. *R&D Management*, 15(2):141-149.
- 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.
- Basden A, Watson I D, Brandon P S, (1995), *Client-Centred: An Approach to Knowledge Based Systems*. CLRC: Rutherford Appleton Laboratory, U.K. ISBN 0 9023 7635 7.
- Basden A, Brown AJ. 1996. Istar - a tool for creative design of knowledge bases. *Expert Systems v.13, n.4*, pp.259-276, November 1996.
- 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. See "<http://dooy.info/bk/fisrpl/>".
- Basden A. 2020. *Foundations and Practice of Research : Adventures with Dooyeweerd's Philosophy* Routledge, London, UK. See "<http://dooy.info/bk/adventures/>".
- Brandon P S, 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.
- Caliskan A. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 14 Apr 2017, 183-6.
- 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.

- Donnelly C. 2024. Microsoft and Google's GHG emissions gains call viability of net-zero targets into question. *Computer Weekly*, 9 July 2024, available at "<https://www.computerweekly.com/news/366592778/Microsoft-and-Googles-GHG-emissions-gains-call-viability-of-net-zero-targets-into-question>".
- Dooyeweerd H. 1955. *A New Critique of Theoretical Thought*, Vol. I-IV, Paideia Press (1975 edition), Jordan Station, Ontario.
- Hines JG, Basden A. 1986. Experience with use of computers to handle corrosion knowledge. *British Corrosion Joournal*, 21(3):1511-156.
- Hirst G, (1991), "Existence assumptions in knowledge representation", *Artificial Intelligence*, **49**:199-242.
- 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/>".
- Larsson S, Heinz F. 2020. Transparency in artificial intelligence. *Internet Policy Review*, 9(2), 1-16. Available from "<http://policyreview.info/concepts/transparency-artificial-intelligence>".
- 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/.html>".
- O'Keefe D. 2024. How Does AI Model Training Work? Appian, January 24, 2024. Available at "<https://appian.com/blog/acp/ai/how-does-ai-model-training-work>".
- Polanyi M. 1967. *The Tacit Dimension*, Routledge and Kegan Paul, London U.K.
- Winfield M J, Basden A, Cresswell I. 1996. Knowledge elicitation using a multi-modal approach. *World Futures* **47**:93-101.
- Wolfe CR. 2023. Understanding and Using Supervised Fine-Tuning (SFT) for Language Models. Available at "<https://cameronwolfe.substack.com/p/understanding-and-using-supervised>".

Created: 28 February 2025.

Last updated: 1 March 2025. 3 March 2025, 7 March 2025.