



“ethics in AI” relate to those two humans, but the other two humans, AI developer and algorithm designer, also have important responsibility and discussion of “ethics in AI” should include them too.

It is mainly these two with which this chapters concerns itself. First, we discuss the two types of AI and a major challenge faced by both, then we discuss the types of application in which AI is likely to fail or succeed, then the responsibilities of the algorithm designer and the AI developer. To do this, we employ Dooyeweerd’s philosophy and our approach will, as far as possible, apply to both types.

## Two Types of AI

Today, AI systems are built by machine learning (MLAI: machine learning AI); in the 1980s AI systems (which includes “knowledge based systems” or “expert systems”) were built by **knowledge elicitation** (KEAI: knowledge elicitation AI), the lower route into the knowledge base in Figure 1. My own experience in AI was developed within KEAI [Basden 1983; Attarwala & Basden 1985; Basden, Brandon & Watson 1995]. In KEAI, analysts (“knowledge engineers”) interview experts to elicit knowledge that is relevant to the intended application, which knowledge is then coded into the knowledge base manually - for example, the essayist’s expertise on what constitutes a good essay and how to write one and what errors to avoid. Knowledge elicitation was an expert activity, in which developing a good social accord with experts was of supreme importance if one wanted to elicit high quality knowledge. It was time-consuming and needed to be learned by apprenticeship. Partly because KEAI was seen by management as expensive, from the 1990s onwards, many hoped that this human process could be bypassed by machine learning.

It is useful to regard **machine learning AI** (MLAI) as KEAI with the human element of knowledge elicitation bypassed by automation. To achieve this, a massive amount of data is analysed by algorithms to detect relevant patterns - for example patterns about good essay writing. This is the learning or training phase of MLAI. ChatGPT was trained using nearly 200 billion pieces of information drawn from the Internet. (It is assumed that the reader understands roughly how MLAI is trained using training data, and here we introduce details; for a good explanation see Chapter 2 of this book or Wolfram [2023].)

For an AI system to work well, its knowledge base should ideally be

- ◆ accurate,
- ◆ comprehensive,
- ◆ unbiased and
- ◆ future-proofed,

(the latter to cope with exceptional or unexpected situations especially those that have never happened before) rather than inaccurate, partial, biased or backward-looking.

One challenge in KEAI is **tacit knowledge** [Polanyi 1958], knowledge that the expert has but cannot, or does not usually, tell us. Explicit knowledge, which the experts can readily tell us, is the kind of knowledge that is taught on courses or written about, but is usually only a part of the knowledge that is actually brought to bear in practice. The rest is tacit knowledge, of which there are several kinds: muscular tacit knowledge such of how to ride a bicycle, social tacit knowledge of how to operate well in social settings, skills that were learned 30 years ago that have become embedded in the expert’s operations, knowledge of the unusual or exceptional conditions, and so on. If tacit knowledge is omitted from the knowledge base then the AI system, working well in most circumstances, will fail

without warning in others, so cannot be relied upon. During the earlier period we found many ways to make some kinds of tacit knowledge explicit, especially by asking “Why?” and “When not?” etc., which jog the expert’s memory on skills they learned decades ago, but with muscular and sensory knowledge this is not possible. See Chapter 13 for more on tacit knowledge.

What this means in practice is that KEAI is best when designed to *work with* the human user rather to *replace* them, because some tacit knowledge is supplied by the user, and the explicated knowledge encapsulated in the KEAI knowledge base becomes useful to ensure that the user does not overlook important questions [Basden 1983]. This is what made the expert systems in which I was involved successful: we designed them to form human-machine wholes in a way that made interaction very easy and natural so that the good quality explicated knowledge in AI system works smoothly with the tacit and contextual knowledge of the users [Basden, Brandon, Basden & Watson 1995].

In **MLAI**, some tacit knowledge can be learned by detecting patterns from such data insofar as the data has arisen from real life experience. This includes muscular or sensory tacit knowledge insofar as their effect is detectable in spatial or physical or behavioural patterns. That is one of MLAI’s advantages over KEAI, especially for things like Xray analysis or self-driving cars - though the vast amounts of training data required to capture it can be challenging to obtain.

Where MLAI fails, compared with good KEAI, is in several ways. One is that even vast amounts of training data are **not enough** to capture all tacit or exceptional knowledge, such as that cyclists sometimes walk with their cycles rather than ride them - which is why one driverless car killed a cyclist [Note: Driverless car kills cyclist], whereas with KEAI such possibilities can be uncovered by asking What-else questions. A second, major deficiency is that the knowledge encapsulated is **not transparent** so MLAI becomes a ‘black box’, unable to explain how and why it came to the answer or decision it did, because the knowledge encapsulated in MLAI is all mixed together and cannot be separated. In KEAI, because its pieces of knowledge retain their distinctness, such explanations can, in principle, be given. There is much discussion about these problems with MLAI, and some of what was learned during KEAI era is being unearthed, but too little. So AI providers are having to do ever increasing amounts of knowledge elicitation to avoid LLMs giving illegal or immoral advice, such as pornographic or violent.

A third disadvantage is that MLAI is **backward-looking** because it is trained on information about past experience, whereas KEAI can be forward-looking if the knowledge engineer elicits knowledge about future possibilities that have not yet occurred. A final disadvantage, widely recognised, is the massive power consumption needed for training on vast datasets, and the consequent increase in **greenhouse gas emissions**. Is the fashionable, automatic, promoted, deployment of MLAI ethical? Such a question is discussed elsewhere.

## **In Which Kinds of Application Can AI be Successful?**

As I argue in Basden [2008, 197], any computer program may be seen as encapsulating in computer code the laws by which the program is to operate. The philosophy of Dooyeweerd offers a multi-aspectual view on reality, distinguishing fifteen different spheres of law or aspects, which are irreducible to each other and yet inter-dependent; see Chapters 3 and 17. With these, we can understand AI more clearly and build AI modules more systematically.

For example, the knowledge base of a Chess-playing program encapsulates laws of the spatial aspect (for the Chess board), of the kinematic (moves), and of the formative aspect (laws of planning and achieving, and of forming strategies). It might

also encapsulate laws of the aesthetic aspect (playfulness and harmony of the whole), and maybe a few others. There is little need to encapsulate laws of the physical or biotic aspects.

In a computer game, the laws can depart from reality but for any program to be used in real life, including AI systems, the laws it encapsulates must be faithful to those of reality - in every aspect that is relevant to the application of the AI.

If the laws that are encapsulated are not accurate or not comprehensive enough, then the AI system will fail to perform well in some circumstances. It is not necessary to encapsulate all the laws of the kinematic aspect in Chess, for example, since movement there is limited to small, restricted movements forwards, backwards and sideways, without laws about speed etc.

In automated driving, at least the laws of the following aspects are likely to be needed:

- ◆ Kinematic aspect - the primary aspect: all its laws;
- ◆ Spatial aspect - especially of roads and distances: almost all its laws;
- ◆ Quantitative aspect - for things like speed comparisons, amount of fuel, number of human occupants: all its laws;
- ◆ Physical aspect: comprehensive encapsulation of some of its laws, such as of friction, fuel or battery consumption, fuel as explosive, joint wear, momentum, etc. but little about quantum physics or relativity or planetary physics;
- ◆ Biotic aspect: laws about life and limb, but not about reproduction nor about ecology;
- ◆ Sensory aspect: the laws of sight and visual recognition and proximity detection, and also of the need to be seen and heard, and also some knowledge about reaction times etc.;
- ◆ Analytical aspect: laws of being able to distinguish important features from unimportant, and to reason about them;
- ◆ Formative aspect: laws about goal setting and achievement, for example options of routes to the destination;
- ◆ Lingual aspect: to understand road signs, etc. - but not to write essays!
- ◆ Social aspect: perhaps knowledge about the difference between driving to work, to a party, to family, etc. - but not to socialise;
- ◆ Juridical aspect: the country's laws about driving on the nation's roads, and also some of the laws of natural justice;
- ◆ And so on.

(Notice how readily parameters come to mind when thinking with aspects.)

MLAI must be trained in all the relevant laws of all relevant aspects. However, it is wise to go beyond that which seems relevant during development, for two reasons. One is that, in use, the AI system is likely to meet **exceptional circumstances**, which is it not always easy to anticipate - such as the bike-pushing cyclist [Note: Cyclist]. The other is that **extensions of use** beyond the original intention should be anticipated, such as driving off-road or along farm tracks where there are animals rather than people. Or, if a variant of Chess were devised where speed matters, the Ai Chess player would need to encapsulate the speed laws of the kinematic aspect. Human knowledge elicitation (to anticipate such possibilities) cannot be entirely bypassed.

By and large, the laws of earlier aspects are simpler, more limited and more determinative [De Raadt 1991], than those of later aspects and hence much easier to discover and encode. They are simpler because the laws of any aspect depend foundationally on (and thus incorporate) the laws of all earlier aspects. For example, the laws of the physical aspect depend on those of the kinematic,

spatial and quantitative, the laws of the social depend on all these plus all aspects up to the lingual. They are more determinative in that functioning enabled and guided by the laws of early aspects is less variable and more predictable than that enabled by later aspects.

Because of unpredictability, **generative AI** sometimes makes up answers that might be statistically appropriate but often are actually wrong - what we call "hallucinations". Generative AI follows laws of the formative aspect.

**Large Language Models** (LLMs) like ChatGPT are much more complex than self-driving cars, because their main aspect is the lingual. That they use generative AI reflects the dependency of the lingual on the formative aspect. They they use massive matrices holding the parameters of tokens (pieces of signified meaning, word-concepts) reflects its dependency on the analytical. See Wolfram [2023] for full description. Its dependency on the psychical and spatial aspects is found in their ability to generate paintings. Their lingual aspect itself comprises three levels: syntax, semantics and pragmatics, which are meaningful in the formative, lingual and post-lingual aspects respectively. It turns out that LLMs only poorly and partially encapsulates laws of pragmatics, even though it processes such tacit knowledge in its training data [Hu et al. 2023; Sravanthi et al. 2024] - which is one important reason for LLM failure.

What this means is that MLAI applications meaningful in later aspects require much more training than those in earlier aspects. It also means that AI is likely to work better in applications meaningful primarily in early aspects such as Chess (spatial), self-driving cars (primarily kinematic), or generating chemicals (physical) or proteins (physical-biotic) analysing x-rays (spatial-biotic). In early-aspect applications it has proved possible to replace humans with AI, but not so successfully in later aspects. The essays that ChatGPT generates do not gain students good marks!

In fact, the above reasoning suggests that in such later-aspect applications AI might never be able to fully or mimic replace humans where the outcome matters. Instead, they can at best **assist** humans. To gain good marks, the student using ChatGPT should take its ideas for essays and then carefully check and expand those ideas, then write the essay. Many recognise that AI should work with, rather than replace, humans. In the 1980s we found this to be true of KEAI too and it was important to identify roles in which AI can be truly beneficial, such as refining human knowledge or acting as a checklist to help human experts ensure they did not overlook anything [Basden 1983]; Brandon et al. 1988]. One of my colleagues demanded that the AI system be designed to actively discourage people from believing its advice and, instead, explore its reasoning. The challenge then is to design the user interface of the AI system to make this assisting humans seamless. KEAI has an advantage here of the transparency of its knowledge; MLAI lacks this.

So one cannot extrapolate MLAI's recent successes in early-aspect applications to later aspects, because of their increasingly complex laws. Especially **Artificial General Intelligence** is unlikely ever to be successful because it aims to replace humans in every aspect and thus the laws encapsulated for every aspect must be complete, accurate, unbiased and future-proof - which is unlikely to ever be the case.

But back to the real world. Let us consider the responsibility of the AI developer and the algorithm designer.

## **How Should MLAI Systems Be Trained? The Responsibility of AI Developers**

The **responsibility of the AI developers** is to ensure that the knowledge base is 'honest' and 'faithful' to the reality of the application, that is, it holds accurate and complete knowledge that is unbiased and future-proof. That is the ideal but it is usually impossible, so the AI developer should understand such limitations and design the AI system around them, giving careful attention not only to the knowledge base but also to how it will assist users in various roles.

KEAI requires highly sensitive and skilful interviewing of experts, understanding what they tell, and actively seeking other knowledge in every aspect relevant to the domain. In my experience, the good knowledge engineer would proactively challenge the experts about exceptions etc. and would benefit from apparent disagreements among experts by exploring the root of disagreements, rather than merely accepting some kind of 'democratic' average.

For example, in my work with two corrosion experts, when I asked why their knowledge differed, it was revealed that one worked with chemical reactions above 300 degrees Celcius, and the other below, and the chemistry changes there. This opened up a lot of avenues for new knowledge. I had to have at least some understanding of the physical aspect to do this. In another application, in agriculture, I proactively asked what advice would be given to farmers who wanted to reduce their chemical input, but had to persevere through a barrier presupposition of ever-increasing chemical use - and was motivated to do so by my Christian beliefs about our responsibility to ecology. Again, this opened up new avenues for knowledge elicitation. It also reduced bias to industrial agriculture and also made the final AI system more flexible and trusted when it went into use. [Basden 2025]

It is less easy for the MLAI developer to take such actions, however. There are typically seven stages in training MLAI: identifying the parameters on which to train the knowledge base, collecting data, pre-processing, selecting an appropriate model, actual training, evaluation and refining. Contrary to popular assumptions that machine learning requires minimal human input, that is true only of running the training software; human input is vital in all other stages for a good AI system that will actually work well. KEAI methods are needed in each of them. For example refining includes finding ways to identify content as pornographic or violent, and to correct cultural bias.

Dooyeweerd's aspects can be very useful in each stage by offering a taxonomy of meaningfulness and of kinds of normativity; for example, the age of a person is meaningful in the biotic, psychical and social aspects at least. The suite of aspects can be used almost as a type of checklist and cue to stimulate better thinking by the developers, to separate out aspects, and then ask "What else might be meaningful in this aspect?" I will illustrate this with just the first two stages.

**Stage 1, parameter selection** (also in other stages). Of ChatGPT's 12288 parameters (GPT4 has many more; see Chapter 13 of this book), I have not yet been able to discover how they were chosen, and suspect that many of them were themselves selected by some AI system built for that purpose but which would have begun by humans choosing a number of parameters as a starting point. On which parameters the knowledge base is trained is crucially important, because, if any are missing, such as about road signs in self-driving cars, then the AI system will not work well. Many of them are not obvious and must be discovered carefully by processes similar to the old knowledge elicitation. Any AI generation of parameters from the starting set would magnify errors and omissions in the original set, especially errors of cultural bias and the ignoring of aspects. So, the AI developer has a responsibility to check and double-check the set of parameters finally employed to train the knowledge base.

So it is useful to have a clear idea of what kinds of parameter are meaningful and important in the application, and thus MLAI parameter selection can learn

from KEAI. In 1995, Mike Winfield (also [Winfield et. a. 1996]) devised a method of obtaining a good set of parameters using Dooyeweerd’s aspects, including those often overlooked: MAKE, **Multi-aspectual Knowledge Elicitation** . An expert in the application area is interviewed, but first has Dooyeweerd’s aspects explained briefly, to use as a checklist for discussion. The analyst begins by asking “Can you identify a couple of aspects that are important to you?” and then “Please tell me some parameters that are meaningful in these aspects,” then cycling back. Ideally, aspects and parameters are plotted on a large piece of paper, as shown in Figure 3.

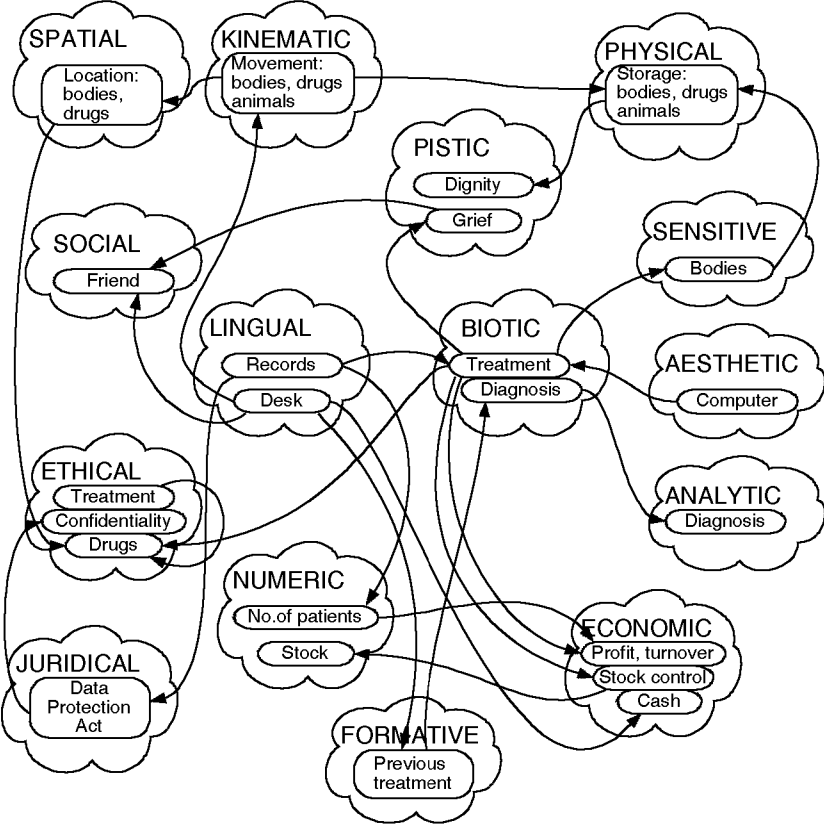


Figure 3. Simple MAKE diagram generated for veterinary practice.

As more parameters are identified other aspects become relevant and are added to the diagram, along with more parameters. Lines indicate relationships among parameters and even aspects. After the flow of parameters slows, it is appropriate to ask, “Are there any other aspects that we have not covered, which are important?” and, if so, these are added in similar manner. In his research, Winfield all interviewees identified almost every aspect as relevant. He found that frequently tacit knowledge would be made explicit and that interviewees greatly valued the process.

Imagine a much more comprehensive diagram of this kind. It would most likely provide a good starting set of parameters from which to begin parameter selection and, in later stages, it can remind the developer in which areas the AI system needs to be evaluated or refined.

**Stage 2, collection of training data.** Training data must be selected and cleaned up, and this is likewise a challenging, expert task, because it determines the quality of the knowledge base (accuracy, comprehensiveness, lack of bias and forward-lookingness). The training of LLMs from the Internet is particularly

challenged (see Brown et al. [2020] and Chapter 13). Internet data contains many errors and gaps. Some of the gaps are just lack of detail or of the latest knowledge, such as when this is unavailable for copyright reasons.

Bias, especially **cultural bias**, can result from too little attention being given to various aspects. For example, most of the Internet text was written by affluent, educated people of the Global North who, because of their secularist, economic worldview, overplay the analytical, formative and economic aspects of science, criticality, technology and money, and downplay the ethical and pistic aspects of things like generosity and religious faith that are important in other cultures.

Errors, gaps and bias are widely recognised and discussed but **backward-lookingness** is less discussed. It occurs because MLAI is always trained on data that expresses past experience, so MLAI systems learn from the past and might not cope well when future conditions differ. By contrast, good KEAI can be future-looking if the knowledge engineer has taken the trouble to ask the experts about possibilities, especially unusual ones.

Reference to Dooyeweerd's aspects, perhaps using MAKE, can help in several stages of the training process.

- ◆ **To improve completeness**, identify which aspects tend to be overlooked, and then seek data meaningful in those aspects.
- ◆ **To overcome cultural bias**, which is often a functioning in later aspects like the ethical and pistic-faith, obtain data meaningful in those aspects that have come from other cultures. In a secularist culture, one would expect the faith aspect to be less present, in a rationalist culture, the ethical aspect, in a money-dominated culture, the aesthetic and biotic aspects (e.g. biodiversity), and so on, so seek data from non-secularist, non-rationalist, non money-dominated cultures. A useful **measure of bias** would be to count the number of pieces of knowledge meaningful in each aspect. I cannot find any evidence that creators of LLMs have done anything like this.
- ◆ To understand **future possibilities**, ask experts questions like "What else?" and "When not?" relevant to each aspect.

Good quality knowledge is of little use, however, if the underlying algorithms are deficient.

## The Algorithm Designers and Their Responsibility

The **algorithm designer** [Note: Algorithms] has the primary responsibility for ensuring that the way knowledge can be encoded, stored in the knowledge base and activated by the engine are fully appropriate to all the aspects that are relevant in the application. For the quantitative aspect, this is usually easy because computers already do arithmetic and matrix manipulations.

With the spatial aspect it is more complex. Conventionally, two-dimensional spatial figures like routes or shape boundaries, such as the U-shaped woodland in Figure 2(a), would often be stored in the knowledge base as a list of (x,y) coordinate pairs.

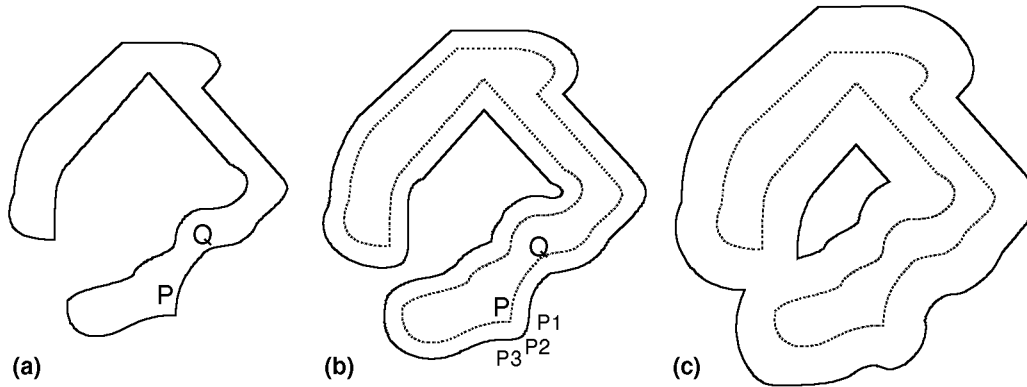


Figure 2. Complex shape: (a) original, (b) expansion, (c) further expansion to become a shape with hole; see Basden 2008; 2018.

Suppose that the AI system needs to deal with expansions of the boundary as in Fig 2(b). (For example, in the area around the woodland, birds nest on the ground and predators emerge from the wood each night up to 100 m to prey on nests, and the system is to predict which nests are under threat.) That is still supportable using a list of (x,y) pairs. But suppose the expansion is larger, as in Fig 2(c), then we get a shape with a hole, which is no longer supportable with one list; it needs two, or more depending if there are more holes. But many algorithm developers would not think of this at first. Spatially important applications need appropriate algorithms that fully incorporate spatial laws.

For later aspects, such as the lingual of LLMs similar mistakes occur. By choosing to represent word meanings by sets of numeric parameters in gigantic matrices, which very conveniently may be manipulated using matrix arithmetic (conceptually simple but power-hungry) they are largely **reducing the lingual to the quantitative** aspect and thereby the richness of language to a few oversimplified word operations. This exacerbates LLM errors introduced by deficient training (see below), and even more so since the numbers are merely probabilistic, so can be wrong in specific cases.

As argued in Basden [2008] chapter VII and Basden [2018] Chapter 7), each aspect needs a different set of algorithms to encapsulate its rich laws. Algorithm designers should no longer be content with reducing the richness of any aspect to algorithms meaningful in an aspect that they find conceptually simple. The algorithms should be **appropriate to the knowledge**, not reductionist. The algorithm designer is not merely a lowly backroom-worker but must be expert enough in every relevant aspect to respect its richness, including the unusual exceptions that can occur, and find ways of incorporating them appropriately in the software of the knowledge base and engine. **Dooyeweerd's aspects** makes it possible to recognise these issues, frame them and explore them.

## Conclusions

It is being found in practice that MLAI, though impressive, fails without a lot of extra human effort to refine it. Much of this effort needs to be treated as skilled knowledge elicitation. MLAI needs to learn KEAI methods and skills, not just (as seems to happen now) bolt on modules in reaction to problems like pornography and extreme violence that happen to arise. To sum up this chapter:

1. The "ethics of AI" concerns not only the widely-discussed harm or good that results from using AI or from its impact on society, which are the responsibility of the user and deployer of AI, but also the extent to which AI systems mislead or give incorrect information or action, which is the responsibility of the AI

developer, and the extent to which the algorithms that run the AI system are appropriate to the aspects in which the kind of application is meaningful, which is the responsibility of the algorithm designer.

2. Today's machine-learning AI can learn from earlier knowledge-elicitation AI because in both cases the accuracy, completeness, unbiasedness and future-proofing of the knowledge base crucially affects how the AI system operates.

3. AI will tend to be more successful and reliable when the main aspects in which its application is meaningful are the earlier ones. It will tend to be less reliable for later aspects. Yet even these can still be useful if designed to assist rather than replace humans.

4. The process of developing an AI system and training its knowledge base can be greatly assisted by reference to Dooyeweerd's aspects, because aspects are fundamentally distinct ways in which things can be meaningful.

MLAI needs to learn from KEAI and the considerable experience amassed three decades ago, such as recounted in Basden, Brandon & Watson [1995] needs to be rediscovered but few of us are now left to pass it on.

The world currently finds LLMs fashionable and much discourse about AI is geared to LLMs and their problems and apparent benefits; tomorrow it might find something else fashionable. The aspectual approach here will still apply because it lays down principles relevant to any kind of application; one just needs to shift aspect, from the lingual core of LLMs to the core aspect of whatever is the next kind of AI.

## Notes

**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 Algorithms.** Algorithms include both the data structures and the procedures that work on them.

## References

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. 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/fispr/>".

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. 2025. Some Wisdom About Artificial Intelligence (AI): The Role and Responsibility of Humans and Where and Why AI might succeed or fail. Part I. Available at: "<http://dooy.info/using/wai.html>" and "<http://dooy.info/using/wai.pdf>".

Brandon PS, Basden A, Hamilton I, Stockley J. 1988. *Expert Systems: The*

*Strategic Planning of Construction Projects*. The Royal Institution of Chartered Surveyors, London, UK.

De Raadt JDR. 1991. *Information and Managerial Wisdom*. Paradigm Publications, Idaho, USA.

Hu J, Floyd S, Jouravlev O, Fedorenko E, Gibson E. 2023. *Proc. 61st Annual Meeting of the Association for Computational Linguistics*, Vol I: long papers, 4194-4213.

Polanyi M. 1958. *Personal Knowledge. Towards a Post-Critical Philosophy*. London.

Sravanthi SL, Doshi M, Kalyan TP, Murthy R, Dabre R, Bhattacharyya P. 2024. PUB: A pragmatics understanding benchmark for assessing LLMs pragmatics capabilities. *Findings of the Association for Computational Linguistics, ACL 2024*, 12075-12097.

Winfield M J, Basden A, Cresswell I. 1996. Knowledge elicitation using a multi-modal approach. *World Futures* **47**:93-101.

Wolfram S. 2023. What Is ChatGPT Doing ... and Why Does It Work?. Available at "<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>"

-----