

# 'Intelligence' without understanding

## How a lack of common sense is holding back AI

### 1 Introduction

**Artificial Intelligence** (AI) is becoming a bigger part of our lives, being used for both specialised and everyday mundane tasks. However, unlike humans, AI lacks any real understanding and often makes errors as a result. One of the key reasons why these AI systems are prone to error is because of their lack of a specific type of understanding: **common sense**.

### 2 Common Sense

Despite being an intuitive concept to most people, the term 'common sense' is difficult to define in a concrete and universally accepted way (Shiele and Shiele, 2024). However, a key feature of common sense is that it is inherently **social** and shared (Vico, 1999; Whiting and Watts, 2024).

We gain common sense by **learning from others** (peers, teachers, parents) and through **lived experience**, such as observation, imitation and generalisation. We typically presume that others share awareness and understanding of this stock of knowledge, leading us to regard common sense as being simply '**obvious**'.

This **shared knowledge and understanding** allows us to navigate everyday life and make decisions, particularly in situations of uncertainty, by enabling us to draw inferences about why certain events occur and to adjust our plans and actions accordingly.

### 3 Why AI lacks understanding

AI systems do not experience the world the way that humans do and often struggle with **unforeseen circumstances** and those not in its training data. For instance, a robot vacuum obstructed by a curtain may fail to complete its task; however, it would likely be unable to detect exactly what went wrong, or fix the issue itself. In contrast, a human could recognise the obstruction, remove it and resume the task effortlessly.

This demonstrates the profound **fragility** and **brittleness** of these systems (Brachman and Levesque, 2022; Mitchell, 2021, p.2; McCarthy, 1984).

Humans have **contextual models** and possess the ability to understand what is **relevant** (or irrelevant) to a given situation. This is something that AI systems lack. There is no meaning attached to the words or images they may incorporate or generate, in part because the system does not have the relevant context or awareness of normative principles (both of which can be captured by common sense) and therefore does not have an explicit grasp of **abnormal** conditions.

While AI systems are capable of highly complex tasks, such as defeating chess grandmasters (Silver *et al.*, 2016) or predicting protein folding (Heaven, 2020), they often struggle with tasks that are (to humans) foundational and effortless, including **basic perception and mobility** skills that even toddlers can perform (Moravec, 1988).

### ALMI: A Case Study



Without common sense, a robot may fail to recognise and adhere to social cues, cultural practices and empathetic considerations that are second nature to human caregivers.

The TiaGo robot (featured on the left), developed by PAL Robotics, is designed for use in a variety of settings. For instance, the ALMI project employs a TiaGo robot as an assistive care solution for supporting older adults. However, for the robot to operate safely and effectively in such environments, it must be able to interpret complex human actions, speech and expressions and act in ways that align with human expectations and norms.

For instance, if an elderly individual appears to be distracted, tired or confused, handing them a sharp or hot object could pose a significant safety risk. While a human caregiver may intuitively avoid such actions as a common sense response to the given situation, a robot's inability to interpret and appropriately respond to such contextual cues could result in user distress or harm.

### 4 Implications

Lack of understanding leads to error. AI '**hallucinates**' when the system presents false, fabricated and often misleading information as accurate. The system itself does not recognise that it is making an error and may even assure the human that it is correct as they cannot reliably tell fact from fiction. This is problematic because the errors carry **no meaning** or significance to the system itself, even when their consequences may be harmful or even fatal to humans.

These three news headlines report cases of AI hallucinations. In the first, an AI suggested that adding bleach would make a recipe more 'aromatic'. The second features an AI falsely accusing a real professor of sexual harassment with fabricated evidence. In the third case, ChatGPT generated fake citations that were used by lawyers in court without having first verified their validity.

### 5 Looking to the Future / Potential Solutions

It is this seemingly obvious understanding of 'the way the world works' that is crucial for successful functioning and AI systems, lacking this capacity, often make **unexplainable** and **unpredictable** errors.

1. It is likely that we cannot prevent AI errors altogether, but with the endowment of common sense we may be capable of **changing the type of errors** that these systems make, such that they are less random and nonsensical than current errors like 'hallucinations'.
2. Being given some capacity for common sense could help AI to identify errors as and when they are made which could make it easier to **avoid or reduce the effects** of harmful consequences.
3. If AI systems possessed this type of understanding, they would be better equipped to avoid or **mitigate similar errors in future** - recognising the factors that contributed to preventing the errors or, at the very least, knowing when to seek human assistance.

### References

- Brachman, R.J., Levesque, H.J. (2022). 'Toward a New Science of Common Sense'. *Proceedings of AAAI-36*.  
 Heaven, D. (2020). 'DeepMind's protein-folding AI has solved a 50-year-old grand challenge of biology'. *MIT Tech Review*.  
 McCarthy, J. (1984). 'Some Expert Systems Need Common Sense'. *Annals of the New York Academy of Sciences*.  
 Milmo, D. (2023). 'Two US lawyers fined for submitting fake court citations from ChatGPT'. *The Guardian*.  
 Mitchell, M. (2021). 'Why AI is Harder Than We Think'. *Association for Computing Machinery*.  
 Moravec, H. (1988). 'Mind Children: The Future of Robot and Human Intelligence'. *Harvard University Press*.  
 Novac, M. (2023). 'Supermarket AI Gives Horrifying Recipes For Poison Sandwiches And Deadly Chlorine Gas'. *Forbes*.  
 Shiele, B. and Shiele, A. (2024). 'The A-sociability of AI: Knowledge, social interactions and the dynamics of common sense'.  
 Silver, D., et al. (2016). 'Mastering the game of Go with deep neural networks and tree search'. *Nature*.  
 Verma, P. and Oremus, W. (2023). 'ChatGPT invented a sexual harassment scandal and named a real law prof as the accused'. *The Washington Post*.  
 Vico, G. (1999). 'New Science'. Translated from the Italian by D. Marsh. *Penguin Classics*.  
 Whiting, M.E., and Watts, D.J. (2024). 'A framework for quantifying individual and collective common sense'.

Isobel Standen

Department of Philosophy and  
Institute for Safe Autonomy,  
University of York

Supervisors

Professor Radu Calinescu  
Professor Alan Thomas



These errors are also problematic as they are often **subtle** and **sporadic**. More and more frequently, AI users are deferring to and relying on the system rather than their own common sense and better judgement. However, as shown by the news headlines on the left, there can be severe consequences to humans placing **too much trust** in AI systems.