

## The effects of anthropomorphisation on students' mental models of AI

As educators, our role is to help young people develop mental models that will let them navigate a world increasingly reliant on artificial intelligence (AI) systems. AI systems are often given human attributes, such as feelings, mental states, and behaviours<sup>1</sup>. The **anthropomorphisation** of AI is cemented by depictions in fiction – in movies, TV, and books – and news media. Media outlets often default to images of smiling human-like robots to illustrate stories of AI systems<sup>1</sup>. The language of the field is also inherently anthropomorphic, from the name *artificial intelligence*, to *machine learning*, *having hallucinations*, and *wake words*.

These descriptions may be taken literally if learners lack a robust established mental model about AI systems<sup>2</sup>. The language used to describe AI systems portrays them as capable of 'thinking' or 'reasoning', which misrepresents the processes used by these systems as being similar to human-like intelligence<sup>3</sup>. If you believe that AI systems can think, then you might also ascribe emotions to them. This leads to you viewing these machines as having intention and a moral character – further intensifying existing fears and misconceptions about the capabilities of these systems<sup>3</sup>.

### The risks of anthropomorphisation

If young people see this technology as innately human-like, we run the risk of impacting their...

- **Sense of agency:** The intentional simplicity of AI use may lead students to think they have little or no control over its role and impact on their lives and choices. People may defer to the systems if they perceive them as smarter than they are<sup>4</sup>.
- **Safety:** When AI systems are seen as human, young people report feeling emotionally connected to them, which can open them up to influence and manipulation<sup>5</sup>.
- **Social connection:** People are already more likely to anthropomorphise if they lack social connection in their lives or want to make sense of an unpredictable environment<sup>6</sup>.
- **Curiosity:** If AI systems are seen as human, it leads to a 'black-box' view of applications. This level of abstraction can put people off wanting to learn about how they work, making them perceive AI as 'magic'<sup>7</sup>.
- **Biases:** When AI systems are perceived as human-like, then they are also viewed as 'white'. This view of AI means many people of colour feel put off by AI agents<sup>8</sup>. Many AI systems are also anthropomorphised as female assistants, perpetuating stereotypes of women as subservient<sup>9</sup>.

### When anthropomorphisation works

Anthropomorphising AI systems has benefits for developers in terms of adoption and trust. When users view tools as human-like, they are far more receptive to integrating them into their lives and using them<sup>4</sup>, so developers of AI systems may have a vested interest in people viewing them that way.

There is evidence that anthropomorphising helps younger learners understand complex phenomena in science. Making systems more human than they are also allows us to tell

stories about them, which can be helpful when learning complex concepts.

Overall, the usefulness of anthropomorphisation is directly related to how close something is to actual intelligence. For example, it is helpful when learning about the tides, but when learning about monkeys, it might lead to misconceptions<sup>2</sup>. AI systems are presented as extremely close to human intelligence, so anthropomorphisation is particularly dangerous.

### Key concepts

**Anthropomorphisation** is the practice of assigning human-like characteristics and behaviours to non-human objects<sup>1</sup>.

Artificial intelligence (AI) is often described using very human-focused language.

Anthropomorphising AI leads to misconceptions about how AI systems work and their current capabilities, and creates a loss of agency from users<sup>3</sup>.

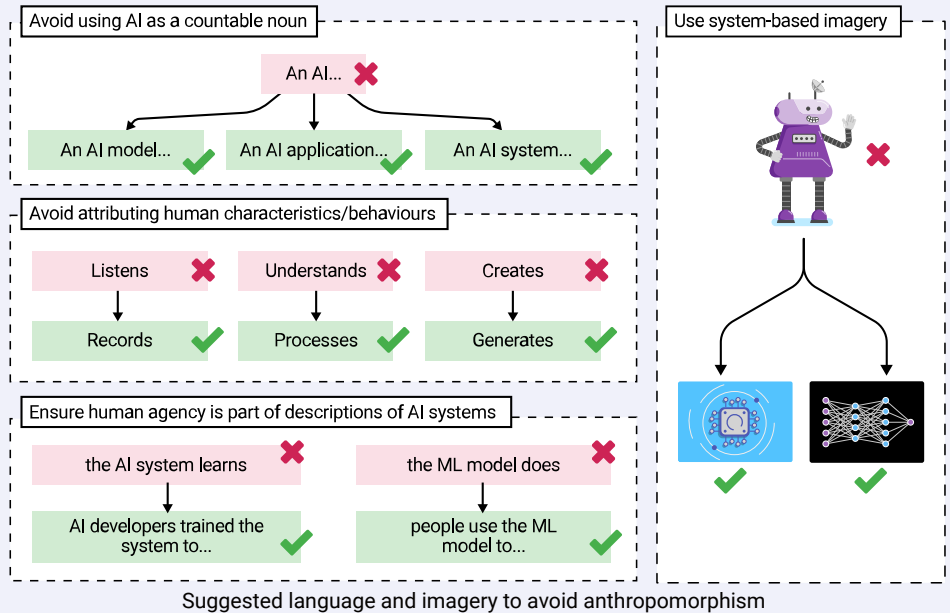
Educators can avoid anthropomorphising AI systems by

- Not using AI as a countable noun as in "an AI"
- Swapping words like "listens", "understands", and "creates" with "records", "analyses", and "generates"
- Using the term AI in the same way you would use "biology" or "cybersecurity"

# How to avoid anthropomorphising AI systems

To help students develop mental models resilient enough to see through how AI systems are presented as human-like, you can:

- Avoid using AI as a countable noun. Instead of calling something “an AI”, use a qualifier like “an AI system”, “application”, or “model”. When using the term “AI”, use it in the same way you would use “biology” or “cybersecurity” – as a field of study and practice rather than a thing you can point at.
- Avoid using human language to describe the behaviour of AI applications. Use system-based language instead. Replace words like “listens”, “understands”, and “creates” with “records”, “processes”, and “generates”.
- Don’t give too much agency to an AI system when describing it. Instead of “the AI system learns” or “the machine learning (ML) model does”, put a human in the equation. For example, “AI developers trained the system to...” or “people use the ML model to...”.



## Examples

Here are some practical examples of de-anthropomorphised descriptions of AI systems:

### Smart speaker

“A smart speaker is activated with a phrase like ‘Hey Google’. It then records your voice and sends it to a computer over the internet. This computer analyses the recording and predicts the response to your request, such as playing music or answering a question. The speaker receives the predicted answer, and plays it back to you.”

### A large language model (LLM)

“A large language model (LLM) is an AI system that generates text in response to a user’s prompt. When a prompt is entered, the system breaks it down into smaller pieces and compares it to a vast dataset of text the LLM has been trained on. The LLM uses patterns and relationships within the training data to evaluate the prompt. A response is then generated by combining words and phrases predicted to be relevant.”



This resource by Raspberry Pi Foundation is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-sa/4.0/>

## References

1. Salles, A., et al. (2020). Anthropomorphism in AI. [the-cc.io/qr22\\_1](https://the-cc.io/qr22_1)
2. Tang, X., & Hammer, D. (2024). “I think of it that way and it helps me understand”: Anthropomorphism in elementary students’ mechanistic stories. [the-cc.io/qr22\\_2](https://the-cc.io/qr22_2)
3. Placani, A. (2024). Anthropomorphism in AI: hype and fallacy. [the-cc.io/qr22\\_3](https://the-cc.io/qr22_3)
4. Huang, J., et al. (2024). When does anthropomorphism hurt? How tool anthropomorphism negatively affects consumers’ rewards for tool users. [the-cc.io/qr22\\_4](https://the-cc.io/qr22_4)
5. Williams, R., et al. (2023). AI + ethics curricula for middle school youth: lessons learned from three project-based curricula. [the-cc.io/qr22\\_5](https://the-cc.io/qr22_5)
6. Alabed, A., et al. (2022). AI anthropomorphism and its effect on users’ self-congruence and self–AI integration: a theoretical framework and research agenda. [the-cc.io/qr22\\_6](https://the-cc.io/qr22_6)
7. Tedre, M., et al. (2016). Teaching machine learning in K–12 computing education: potential and pitfalls. [the-cc.io/qr22\\_7](https://the-cc.io/qr22_7)
8. Cave, S., & Dihal, K. (2020). The Whiteness of AI. [the-cc.io/qr22\\_8](https://the-cc.io/qr22_8)
9. Barau, S. (2024). Deception, discrimination, and objectification: ethical issues of female AI agents. [the-cc.io/qr22\\_9](https://the-cc.io/qr22_9)