de-Humanising AI

Understanding the Responsible use of AI

Sanjana Bhatnagar
Marie Bodén

Human-Centred Computing · School of Electrical Engineering & Computer Science

Actual AI in real life

- Artificial Intelligence is an umbrella term for a type of (advanced) computer systems that can emulate human intelligence.
- It encompasses **everyday technology** (e.g. facial recognition, self-driving cars, image generation, LLMs, etc) as well as technology that is **far in the future** (e.g. Superintelligence, new cancer treatments, etc).

(large language models)

How LLMs Work

(briefly)

- These models work by **predicting the next word** based on patterns (probability distributions) they learned from enormous amounts of text.
- They are trained on huge datasets and learn which words tend to follow others in different contexts.
- They do not understand meaning, check facts, or know when they are wrong.

They produce text that sounds right based on probability, not the truth.

How LLMs Work

and what that means for you

- Models see only text, not context.
- Missing context leads to guesses.
- Guesses produce generic or misaligned outputs.
- The same question (**prompt**) can produce very different answers.

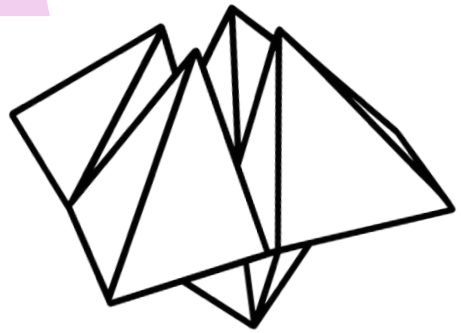
Context shapes what a “good” response will look like to you.

If you want better answers, give better constraints.

ACTIVITY (15 mins)

Context Transformer

1. Divide into groups of 2-3.
2. Fold up your fortune-teller.
3. Work with each other to generate an **action-context** combination.
4. Write it down on your butcher's paper.



ACTIVITY (15 mins)

Context Transformer

How would this output need to change to suit your generated action-context?

“Photosynthesis is the process by which plants use sunlight, carbon dioxide, and water to produce glucose and oxygen. This process occurs in the chloroplasts and is essential for plant growth and life on Earth.”

Think about:

- What would no longer work as written?
- What would need to be added, removed, or modified?
- What assumptions does this output currently make?

Help

Context Transformer

Example generated action-context: **Scaffold, 5-year-old**

> "Photosynthesis is the process by which plants use sunlight, carbon dioxide, and water to produce glucose and oxygen. This process occurs in the chloroplasts and is essential for plant growth and life on Earth."

What would no longer work as written?	Technical terminology Long, dense sentence structure No connection to lived experiences
What would need to be added, removed, or modified?	Simple language Familiar references/Relatability Step-by-step progression Show, not tell to explain what happens and its importance
What assumptions does this output currently make?	That the learner can infer meaning without examples That the learner will understand why this information matters

Help

Context Transformer

Example generated action-context: **Scaffold, 5-year-old**

"Photosynthesis is the process by which plants use sunlight, carbon dioxide, and water to produce glucose and oxygen. This process occurs in the chloroplasts and is essential for plant growth and life on Earth."

- Plants make their own food. They use sunlight, air and water.
- The sun gives plants energy. Inside the plant, the leaves use that energy to make food.
- This food helps make the plant grow big and strong.
- While making food, plants also make oxygen, which is the air that we breathe.

SUMMARY

Context Transformer

Context can drastically change what a good AI output looks like.
Generic outputs are insufficient because of:

- Overgeneralisation/oversimplification
- Missing nuance
- Assumptions of culture/location/people
- Missing localisation for relevance

Understand what “good” means to you and never rely on the first output your LLM of choice produces.

it AIn’t a dictionary!

or an encyclopaedia

A dictionary or encyclopaedia is:

- Curated, edited and reviewed by subject experts to explicit standards
- Grounded in verifiable sources
- Designed to record and stabilise knowledge

An LLM is:

- Trained on large volumes of text, not curated facts
- Optimised to predict language, not verify truth
- Unaware of sources, authority or reliability
- Confident regardless of correctness

Why This Matters

- AI **sounds** right more than it **is** right.
- It fills gaps by guessing. This can reflect patterns and biases in its training data. It will not say “I don’t know” unless told to.
- Default output is generic. Nuance, culture, location, and perspective are often flattened.
- You cannot see what data was used, what was emphasised, or what’s missing.
- Treating AI output as “correct” is **risky**! **Learning still requires doing the work.**

If you want better answers, use human judgement.

ACTIVITY (20 mins)

Checking the Chain

Divide into new groups of 2-3!

You will be given a sentence to "defend". These are meant to be a bit wrong/provocative, but defendable if you twist logic.

- Come up with 5-7 steps of reasoning to lead to your sentence.
- Each strip = one sentence/idea, not a paragraph.
- Make it sound plausible, not obviously ridiculous. Anyone reading this should have to think a bit before picking it apart.
- Use paper clips to turn your strips into a paper-chain.

ACTIVITY (20 mins)

Checking the Chain

Swap chains with another group and:

1. **Identify weak links and label the problem:**
   Leap – jumps too far without support
   Assumption – claims something without evidence
   Distortion – misrepresents the content
2. **Repair the chain**
   Either improve the logic or adjust the argument to something more nuanced. You can rewrite individual links or add/change the starting link if you want.

SUMMARY

Checking the Chain

- What made some chains feel convincing at first glance?
- How easy was it to create a plausible sounding argument for something you disagreed with?

LLMs do this at scale, but faster and with **no human in the loop** to incorporate understanding, context or nuance.
We need to be **critical** of output.

Just like in maths, a correct answer with broken working still fails.

What should we consider when teaching students to use AI responsibly?

How would we individually consider Years 7-8, 9-10 and 11-12?

Being a Responsible User

and learning to assess information quality

- **Don't accept the first answer**: this is just as much an ethical stance as it is an accuracy stance.
- **Know when not to use AI**: where could accepting an AI output cause harm? Is it even the right tool for the job? Will it actually save you work, time or effort?
- **Localise & contextualise**: keep the human in the loop. Consider multiple perspectives – **AI doesn't have a life or experiences. You do.**

Always maintain YOUR voice. The human is responsible for the final output, not the AI.

How would you define “good” AI use?

What is “good” AI use?

- **Purposeful:** Use it for repetitive tasks to free your brain up for creative and strategic work
- **Critical:** Always apply human judgement to understand context; fact-check and question every response you get
- **Ethical:** Be mindful of data privacy; think about who you are excluding when using AI (e.g. is there a professional you could hire, are you missing the consideration of a group of people, etc.)
- **Collaborative:** Work with it, don’t be a slave to it.

Can AI be a good tool for considering multiple perspectives?

“You won’t lose your job to AI, you’ll lose it to someone that knows how to use it.”

thank you for your attention

Sanjana Bhatnagar
Marie Bodén

References/Further Reading

- https://business.columbia.edu/research-brief/research-brief/ai-industrial-revolution
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3570130
- https://big-picture.com/media/the_prompt_engineering_cheat_sheet.pdf
- https://www.youtube.com/watch?feature=oembed&v=H_c6MWk7PQc
- https://www.ibm.com/think/topics/ai-hallucinations
- https://theconversation.com/what-are-ai-hallucinations-why-ais-sometimes-make-things-up-242896
- https://cloud.google.com/discover/what-are-ai-hallucinations
- https://support.google.com/websearch/answer/14901683?hl=en-AU&visit_id=639010077392106068-1173874798&p=ai_overviews&rd=1
- https://lifearchitect.ai/timeline/
- https://freedium.cfd/@lmpo/a-brief-history-of-lmms-from-transformers-2017-to-deepseek-r1-2025-dae75dd3f59a
- https://valasys.com/ibm-study-reveals-40-workforce-reskilling-due-to-ai/