Mechatronic Engineering program

Basics of AI and Deep Learning: 14: Engineering ethics Broad AI perspective

Ziemowit Dworakowski AGH University of Krakow

1



2

Case study II

ZD.

Margaret is a *data science* specialist in a software-for-surveillance company. She notices a mistake in data acquisition routines rendering the system to be insensitive to people dressed in blue.

R

She notifies her Team Leader of the problem. He says is too late for any changes, system goes for production. The client will not know system's weakness.

Margaret should: A: Listen to her team leader (he takes responsibility after all) B: Escalate the problem (become a "whistleblower")

If escalate, then how?
If listen, then to what extent?









Case study VI Sam co-developed a key algorithm for processing sensor data in a scientific institute. A paper is later published on the suspiciously similar topic by his mentor and team – without his name. He is told he was "just an intern" and that the final version was heavily

modified. He should:



A: Let it go – its normal in academia or industry, and he really was "just an intern". Maybe his contribution was not as valuable as he thought.

ZD.

B: Raise the issue – risking his future recommendation and possibly getting a label of "problematic" in future endeavours...

7



20.

"Come gather 'round people wherever you roam And admit that the waters around you have grown And accept it that soon you'll be drenched to the bone If your time to you is worth savin' Then you better start swimmin' or you'll sink like a stone For the times they are a-changin'..." - Bob Dylan

Broad AI picture and perspectives of development ... (claims here might be strongly subjective and biased...)

















AI just pretends it thinks ...

Its just prediction of a next word. Yes, there is a layered structure, self-organization and building inference, but each of this steps just predicts words over and over again. There is no conscience, no real knowledge and understanding.

> ...but if you fake a skill well enough... ...you end up having this skill in the end.

Right now LLMs are completely indistinguishable in-conversation from people. If not for system prompts, you wouldn't be able to prove they are not sentient...

There are still some tasks (formal logic, analogical reasoning, spatial navigation) in which people are better

...but how long?

16



















