Explainable Artificial Intelligence through the lense of Recommender Systems
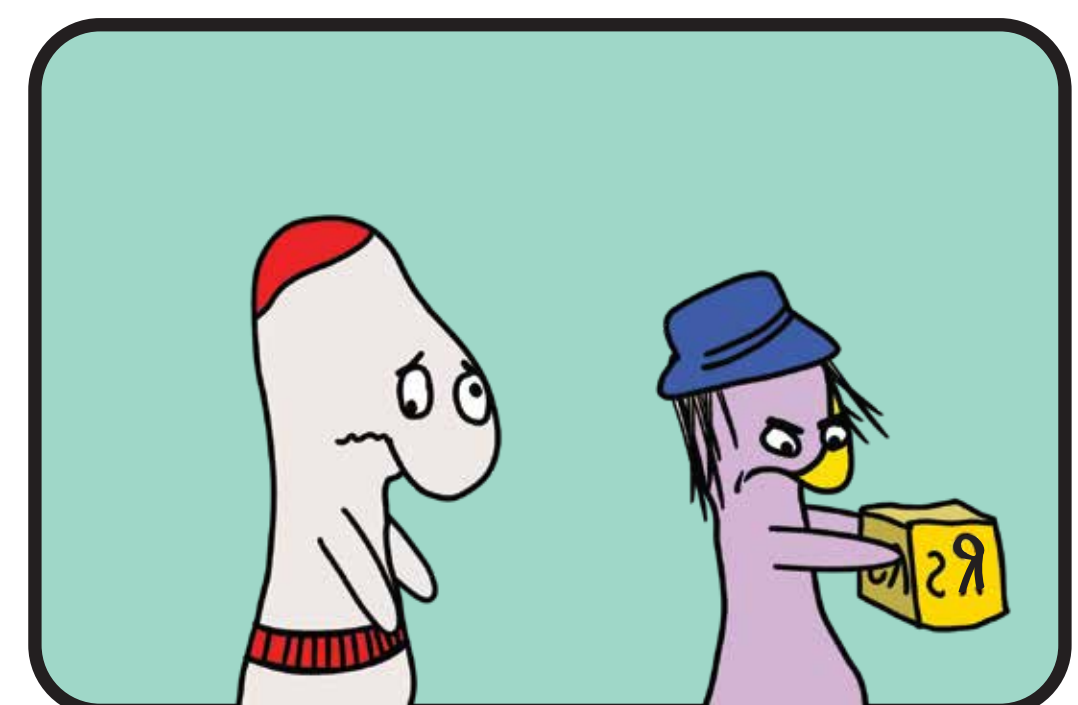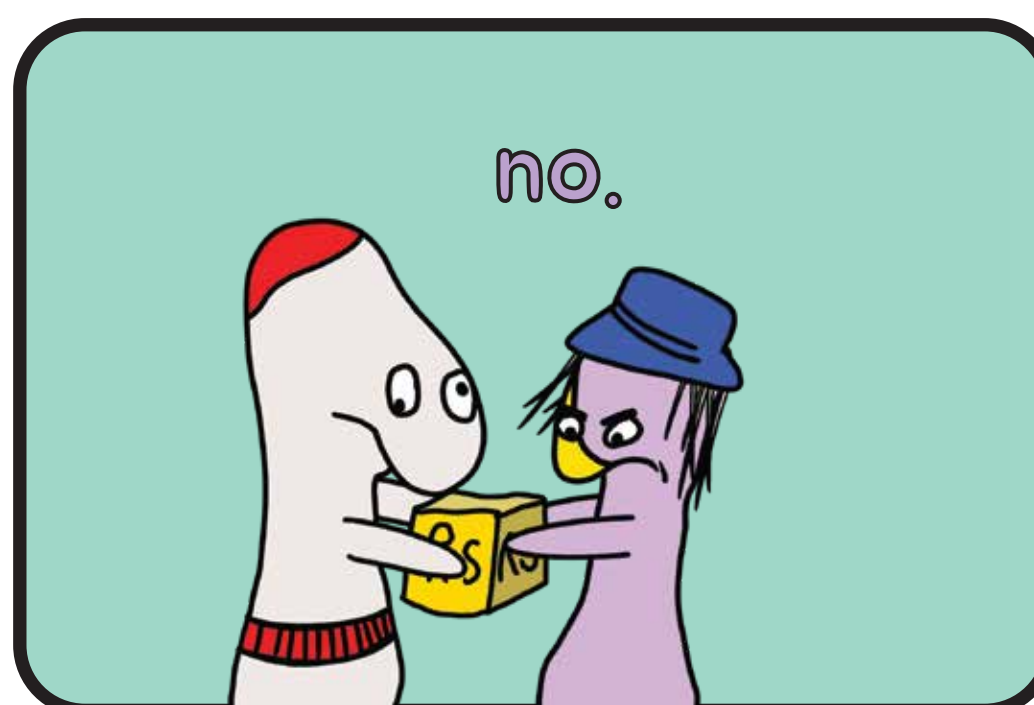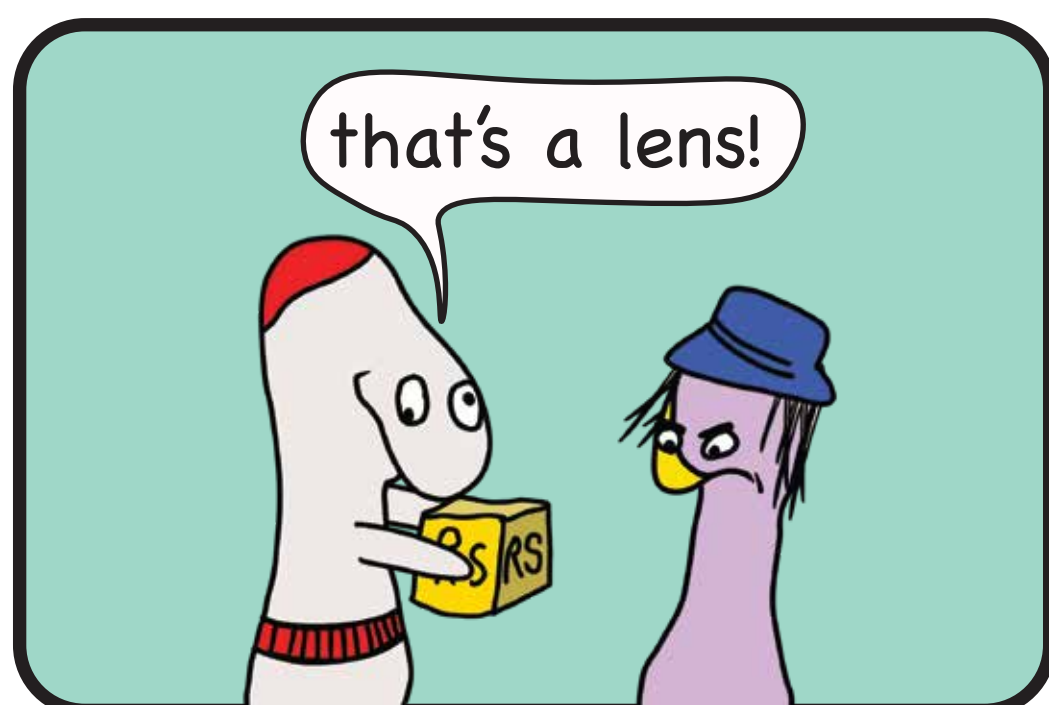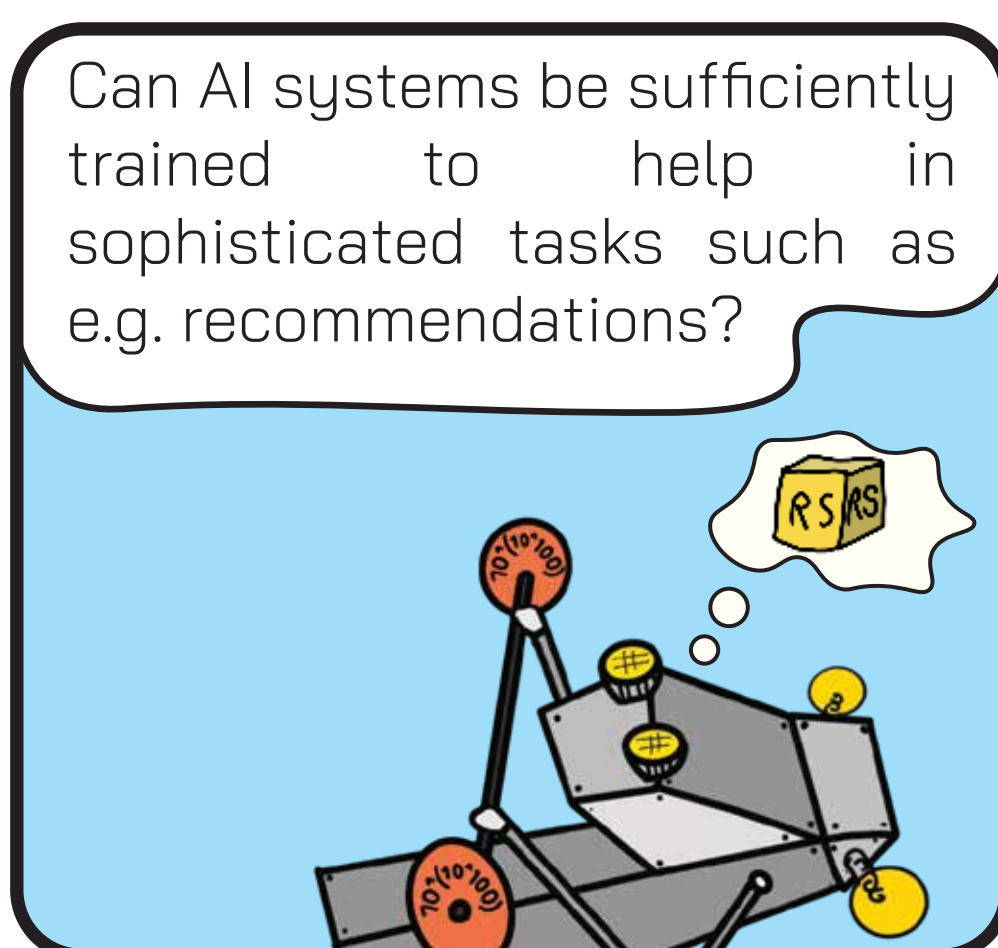
Konstantin Lackner

With the recent advances in Neural Network (NNs) and Deep Learning (DL) systems (ChatGPT, Stable Diffusion, ...) a research field that is gaining interest is XAI. 'XAI' stands for eXplainable Artificial Intelligence, in other words the push for systems that aren't just proficient at generating content (be it text, images, audio or whatever) but also capable of backtracing why they created it precisely in the way they did.

WOW!

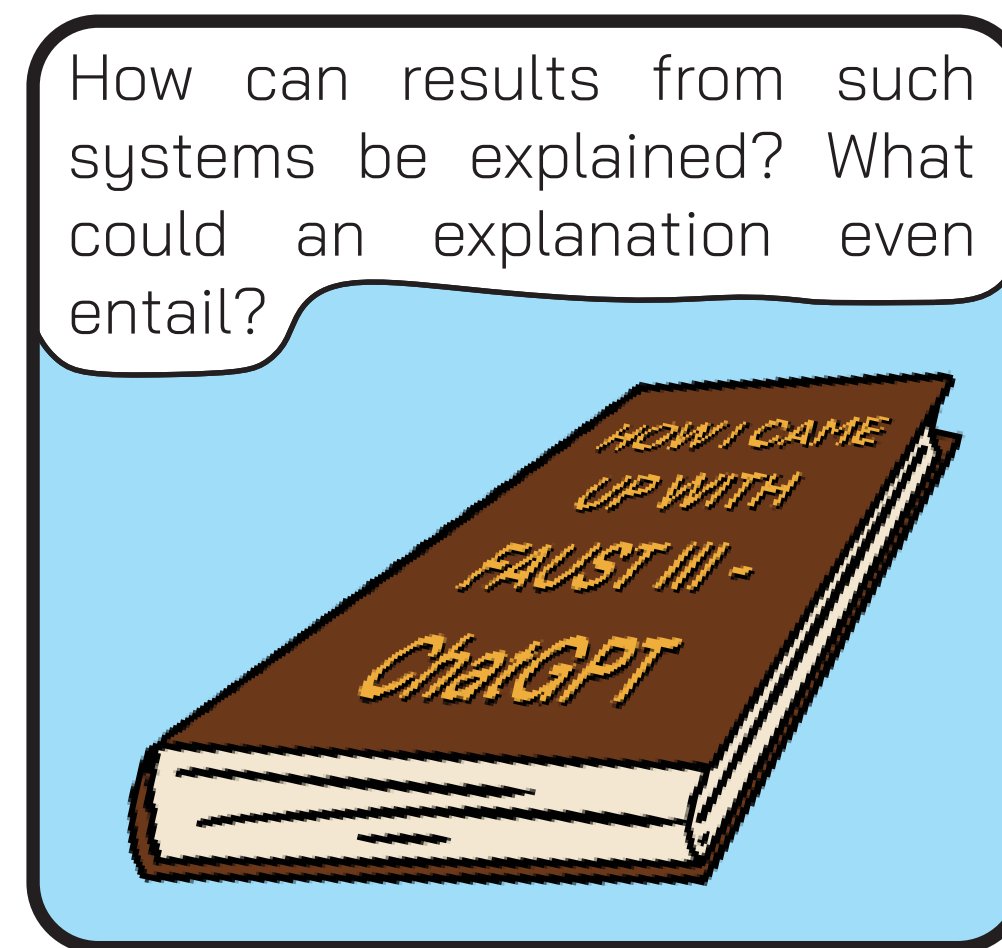how did you come up with that???
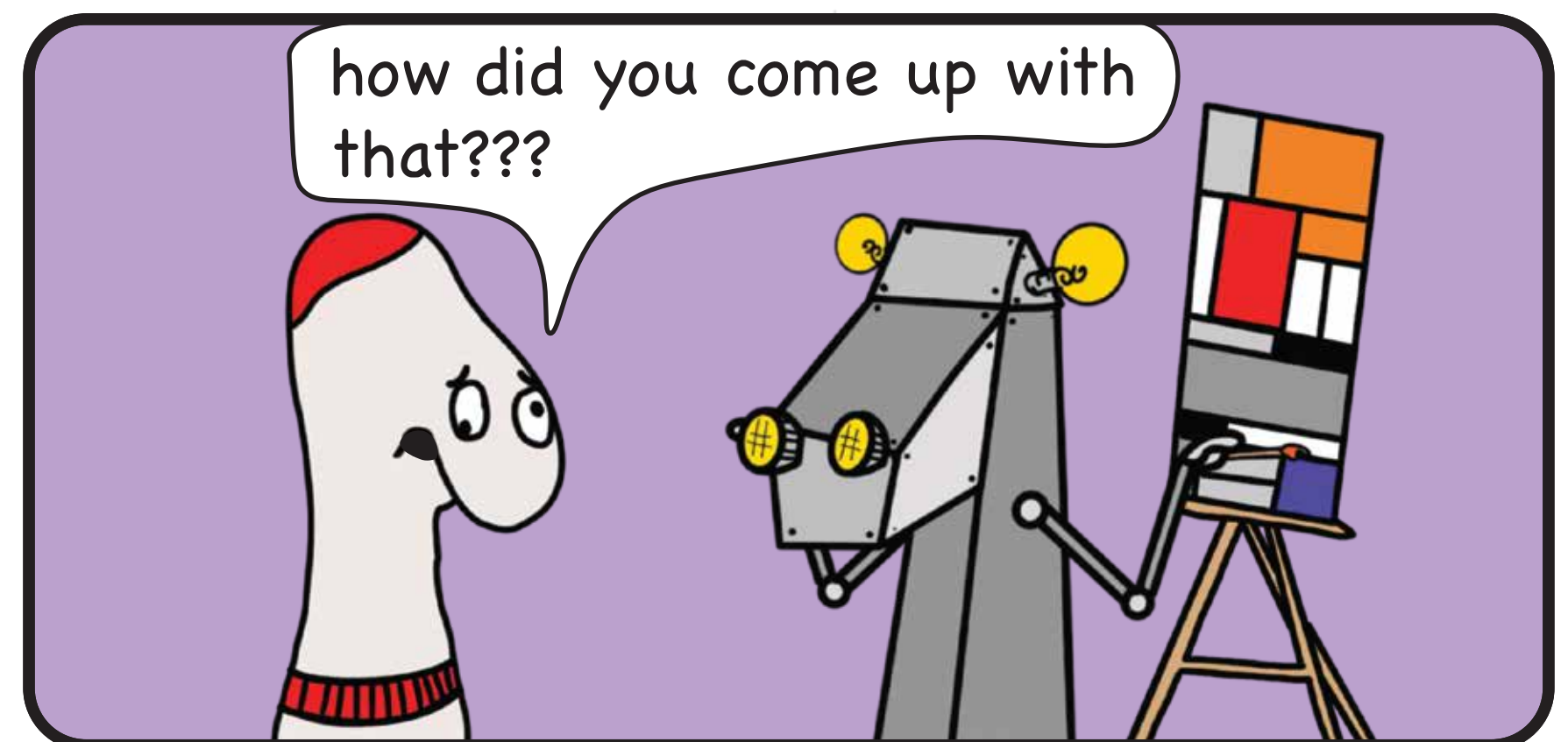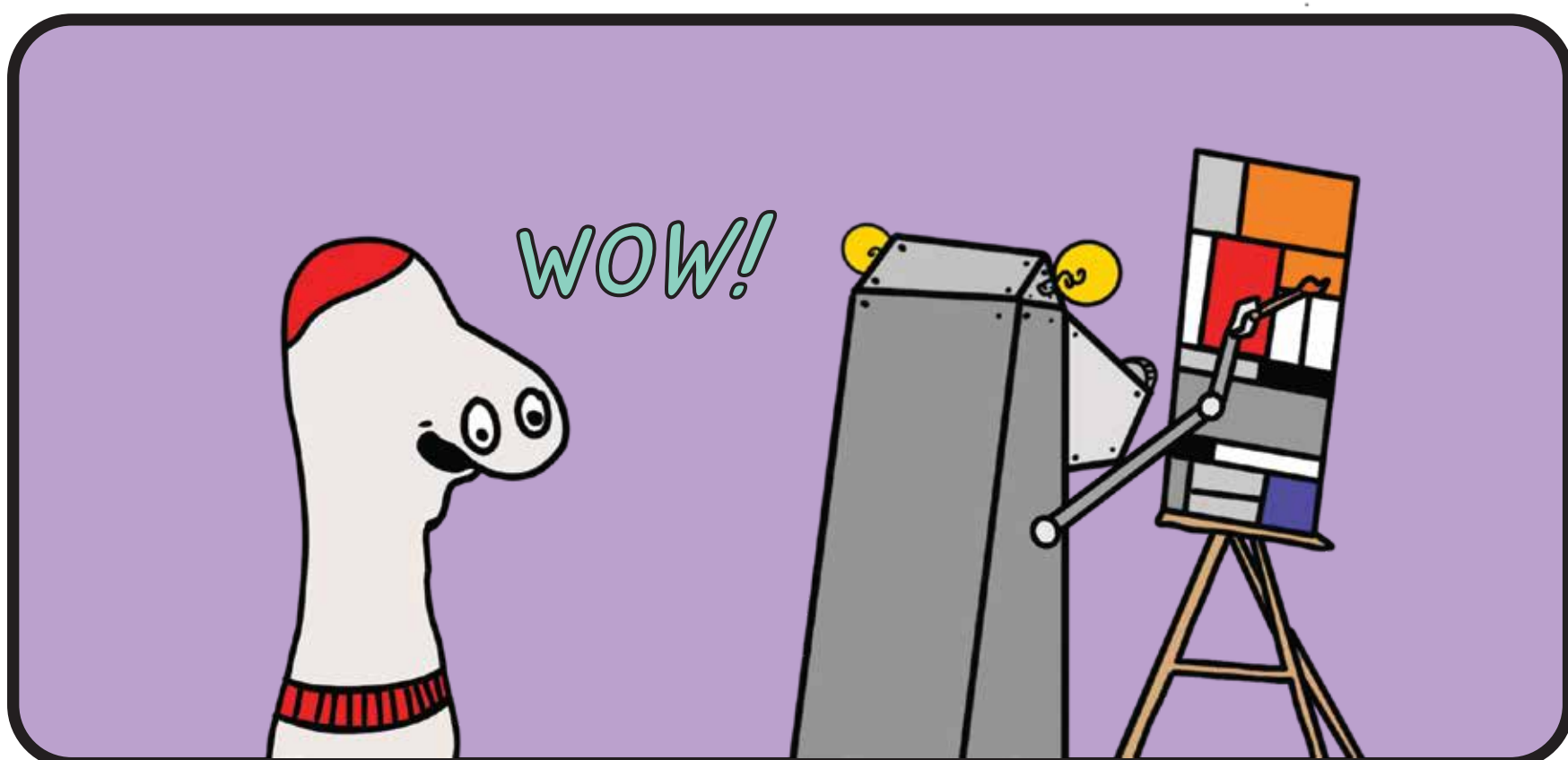
With NNs and DL permeating different fields of research, I argue that one of said fields being permeated - Recommender Systems - is a perfect lens through which one can assess the issues with these new systems. And while some people might not agree with this assertion, this view opens up a set of intriguing questions.

that's a lens!

no.

One is originating from a rather technical, implementation perspective asking both:

Can AI systems be sufficiently trained to help in sophisticated tasks such as e.g. recommendations?

Clearly this does not work very well yet - Is it, however, just a question of training with absurd amounts of data? How well can you train systems and at which point are we just chasing rainbows? For what can and should NNs be used?

How can results from such systems be explained? What could an explanation even entail?

HOW I CAME UP WITH FAUST III - ChatGPT

Can resuts and the embeddings responsible for them be backtraced to areas in a system that were active in their generation? Can we deconstruct our recommendations using NN approaches?

(...One is a technical, implementation perspective...) the other is the HCI perspective:

Without explainability... How do we control what is shown to whom?
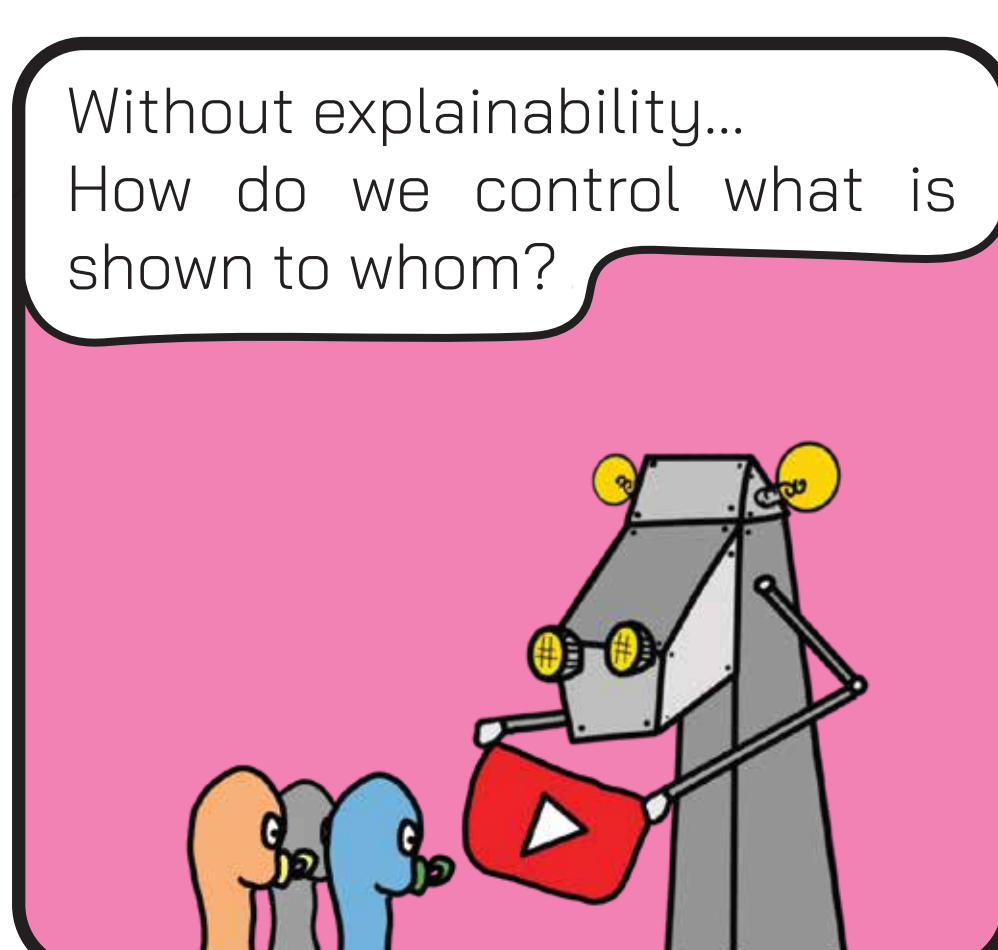
What political content is shown to whom on Social Media? We do not necessarily need to govern these decisions. However, the systems are implicitly deciding anyway, we might as well understand how.

Without explainability... How do we account for fairness in our systems?

i choose you... again!
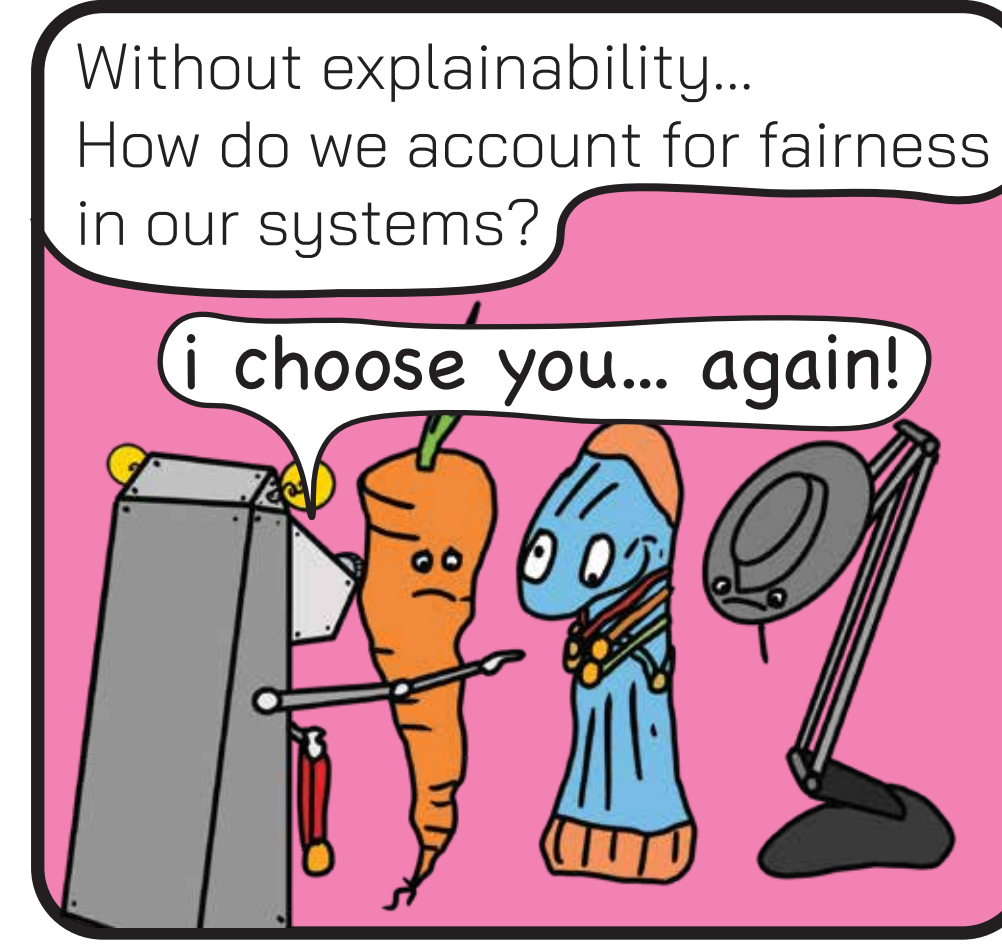
With no way to describe what the system does and how, it becomes increasingly more difficult to ensure fairness or even describe what that fairness may be on a not just purely statistical level but conceptionally.