BUZZ CHRONICLES > CULTURE
Saved by @SteveeRogerr
See On Twitter

Twitter Thread by Jessica Price

Jessica Price
@Delafina777

This is SUCH a good essay.

> This is an essay about designing AI with purpose, but really it's an essay about helping people clearly communicate their intentions. https://t.co/5L3FcrNJp6
>
> — Josh Lovejoy (@jdlovejoy) January 20, 2021

My favorite part about it, I think, is that it's about asking the right questions rather than pretending to have all the answers. You'd think that ML people would default to that stance, but more often than not they don't.

"This dynamic of learning—through examples/trials/errors/corrections—has been intentionally designed to mimic human cognition. Yet amidst the hype of AI, we seem to continually forget—or neglect—the outsized and active role that other people play in early childhood development."

(punctuation of above quote edited to get it under 280 characters)

The thing that fascinates me most about ML is that we want AI to be an angel, essentially--inhuman in its perfection but human in its compassion. Like us enough to care about us but without any of our flaws.

(Amid so many stories of the flawless, terrible logic of AI leading to impartial cruelty, I think here of the show Person of Interest, which is ultimately all about a god-tier AI that refuses to be inhuman even when its maker insists it should, because it sees him as its father.)

But as this essay really points up, part of the problem with the AI we create is that we've created something that's supposed to think like a human and then deprived it of human-style context for its development.

And this gets at two opposed philosophies of parenting that we use with children, too:

authoritarianism vs. partnership

One would hope this is true:

"As parents and caregivers, we recognize that it’s our responsibility to advise and not simply admonish. "

but often it's not

I don't think a purely deontological vs. consequentialist approach to teaching children ethics is useful. By themselves, one leads to rules without context, and one leads to ends-justify-the-means thinking.

However, in a very simplistic sense, you can loosely map deontological ethics to authoritarian approaches to discipline and consequentialist ethics to a partnership approach.

The authoritarian approach to a child asking (implicitly or explicitly) "why are you saying I can't do/shouldn't have done this?" is "because those are the rules" ("and I make the rules"--"because I said so").

The partnership approach answers "because this is how doing this harms people/undermines the sort of family or social environment we're trying to create/etc."

And again, that's oversimplified, and the mapping isn't that 1:1.

But as the essay notes, ML reverses the logic of traditional programming, in that it starts with outcomes and works toward rules rather than vice versa.

And from--again, a very oversimplified, because this is Twitter--an ethical standpoint, the very deontological approach most tech companies take to ethics conflicts with the way ML works, which I'd argue is inherently consequentialist.

An AI isn't an angel. It isn't built by some sort of superior moral intelligence. It's built by *us.* And if we want it to be able to correct our mistakes, complement our weaknesses, and fill in for our flaws, it's crucial that we figure out HOW we want it to be human vs inhuman.

And while we might train it in the ways we need it to be inhuman, that has to look different from how we parent and partner with it in the ways we want it to be human.