AI’s Dual Nature: Empowerment / Enslavement

67- 04/09/2023 **The application of AI is spreading massively across all sectors, from healthcare to entertainment, encompassing politics, defense, security, agriculture, environment, provision of goods and services, transportation, information, and finance. Penetrating ever more deeply into our private lives, using AI enhances and even surpasses our capacities to an unprecedented extent… At the risk of making us so dependent that we could lose control over our future. This raises the question of AI use : as a fantastic emancipatory lever for humanity or as a threat of servitude (voluntary or unconscious). To find out more, Selfpower Community turned to Professor Joanna Bryson. The meeting took place at her office in the Hertie School (Berlin Center), a few floors above the bustling Friedrichstraße. Her insights aid in discernment, a responsibility incumbent upon every scientist who identifies as an engaged actor in civil society.**

Leading expert in artificial intelligence (AI), Pr. Joanna Bryson delves into critical issues concerning the technology potentials and human values. As an activist she pleads for an AI use respectful of human dignity, rights, fundamental freedoms, and democratic values.

Is AI a powerful tool for individual and collective empowerment or a a Trojan horse concealing a new form of enslavement ?

Joanna Bryson : I think it already is kind of a Trojan horse. We already are seeing that people don’t like the recommender algorithms that get them to spend too much time on social media. So it’s already a bit that way. But at the same time it’s exaggerated. The battles we’re having to fight now is to get people to understand what is being lost through Twitter by

the undermining of Twitter. We no longer have this forum where academics and diplomats were having or following real conversations, where the general population could check if somebody is real and can be taken seriously.

So we've lost all that because of deliberate actions by the current owners of Twitter. And yet many people are only saying bad things happen on social media because the recommender algorithms addict you and you start having a negative self image. It can be true and legitimate to worry about that and to protect against it. But again, before Musk had bought it, Twitter had invested hugely in having the kinds of moderation that reduced those effects. I'm not saying those effects aren't real, but that they're moderated by the fact that people actually were learning so much. Twitter users, but not Facebook users, were actually well informed about things like, for example, Covid, or how to handle the government edicts or what was coming out in health care. So people tended to be much better informed if they were Twitter users.

We do not make AI to be good or bad, but what we do with it can be good or bad. And sometimes the tools that we're good can suddenly be flipped and made bad just because someone had an insight and the other way around. There's no reason to think you can't do it the other way around.

There's nothing intrinsic in AI about whether it helps with dignity or it doesn't. That's us. We make those decisions, right?

Definition
AI, by convention, describes non living artefacts that demonstrate capacities to perceive contexts for action,
to act, and to associate contexts to actions…
in order to act intelligently ie « Doing the right thing at the right time ».
(Bryson, & Winfield, 2017 ; Byson, 2019)

Yes, but it’s harder to decipher the logic behind its model, whereas, for example, it’s easy to understand the positions taken by an opinion magazine, or the message of an advertisement : you can quickly see where it wants to take you…

Joanna Bryson : I agree with you digital technology can be used to create very obscure systems but at the same time, we’re also getting much more powerful at penetrating and expressing complexity than we ever have before. That’s why weather forecasts are getting better. That’s why we understand more about the universe. There’s been a lot of people with this false narrative that AI is necessarily opaque. No, that’s not true. It’s like anything else that we could force people to be clearer about. How did they build it? How did they test it? When were they sure they could release it? How did they make these decisions?

It’s not that you’re ever going to know when there’s a neural network with trillions of weights what each weight does. But you know, that’s true about people, too. We have trillions of connections in our brains that we’re not obliged to know what each of those connections does, right? But we’re obliged to be able to regulate each other. People figure out what works and what doesn’t, and then everybody figures out best practice to follow. For example, everyone at this point must understand that machine learning reflects the biases of the society that the data was built on. Right? And so everyone goes and checks that now before they release a product.

And speaking of countering power, there’s some really interesting new work happening like these investment corporations that own like entire portfolios of other companies, they also are trying to make sure that there’s a relatively stable thing going on in the future. And so there’s these new kinds of regulation and pressures.

Establishing rules of transparency and traceability gives society the power to control. Is it enough ?

Joanna Bryson : Of course not. I'm now working on trying to make sure that we use AI to enhance the value of workers so that they get higher wages rather than using it to make workers more similar to each other, which will in general tend to decrease wages because they become more exchangeable. For example, there's a great paper about ChatGPT some researchers at MIT did. It's a study where they they looked different group of people who wrote text for a living like working in advertising or whatever. And then they taught them how to use ChatGPT and they all loved it. They said it made them faster. And what they found was that for three quarters of the people, their writing improved using ChatGPT. When the researchers came back few weeks after, 70 or 80% of people, still used because they found it was helpful, but not the top writers..

But the reverse is also true. You know, that Chinese wages have been going up too fast. So China sees that as a problem, the government does, so they want to use AI to make jobs easier and to reduce salaries.

I would like to see us find ways to use this technology to help everybody be more productive. So, it's more of an expression of who we are as individuals, not just an expression of like this is how the job is. There'll probably be both, though.

AI influence rests in the hands of those who wield it, emphasizing our collective responsibility to shape its trajectory. The discourse on AI's future must continue, framed by awareness, regulation, and a commitment to its beneficial potential.

How do you explain the way ChatGPT presents its answers in an

assertive and peremptory tone,
without giving its sources or giving false ones ?
Joanna Bryson : The whole point about the way that ChatGPT is working is it's mining data. Maybe you'll get lucky if you look for something
about which there is already consensus. However, it is important to know that it will still always come up with something that
looks like language whether or not there is a known answer, because that's what it's been trained on. It's been trained
on competent, confident language, the kinds of people that write newspaper articles, for example The machines are tools that we
buy and most machines that we deal with are things that our corporations have thought of. Whatever the algorithm is, you could put in a different corpus and you would get different answers too, but it would be a similar level of quality in terms of sound.

Decryptage

As AI's role remains ambiguous, Professor Bryson shares a nuanced perspective. The whole questions are about how do we use AI to help us ? As European consumers we have to make sure that product law is being applied in order to prevent or mitigate the risk of surveillance, manipulation, coercion, exploitation, autocracy.

Ideally, we should be aware of the potentials and limitations inherent in its design. This presupposes that the tech industry invests in transparency and provides access to the logic and construction parameters of algorithms, allowing a comprehensive understanding of their ins and outs. Open access to these explanations is the only way to maintain control over the results delivered – a matter of democracy.

Interview by Marie-Georges Fayn

Joanna Bryson

Pr. Joanna Bryson is professor of ethics and technology at the Hertie School in Berlin, Germany. She was one of the inaugurala expert members of the Global Partnership on AI (GPAI), an international initiative to support the responsible development and use of AI. She has been consulted by various governments, organizations, and media outlets on AI-related issues, such as the European Commission, the UK Parliament, the OECD, the UN, the BBC, the Guardian, etc.

Bibliography

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186.

Bryson, J. J. (2022) One Day, AI Will Seem as Human as Anyone. What Then? Wired – Ideas, https://www.wired.com/story/lamda-sentience-psychology-ethics-policy/

Bryson, J. J. (2019). The past decade and future of AI's impact on society. *Towards a new enlightenment*, 150-185.

Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, *50*(5), 116-119.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283*.

Theocharis, Y., Cardenal, A., Jin, S., Aalberg, T., Hopmann, D. N., Strömbäck, J., … & Štětka, V. (2021). Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries. *new media & society*, 14614448211045666.