



Trustworthy & Ethical AI

Carroll University ABIC Conference

Just because we *could*,
does that mean we *should*?



“...to bring the Union’s human-centric approach to the global stage and aims to build international consensus on AI ethics guidelines.”



Artificial Intelligence (AI)

The ability for technology to sense, comprehend, act and learn in a way that mimics human intelligence.



Algorithms

The instructions for a computer to solve a problem.
The lifeblood and connective tissue of artificial intelligence.



Machine Learning

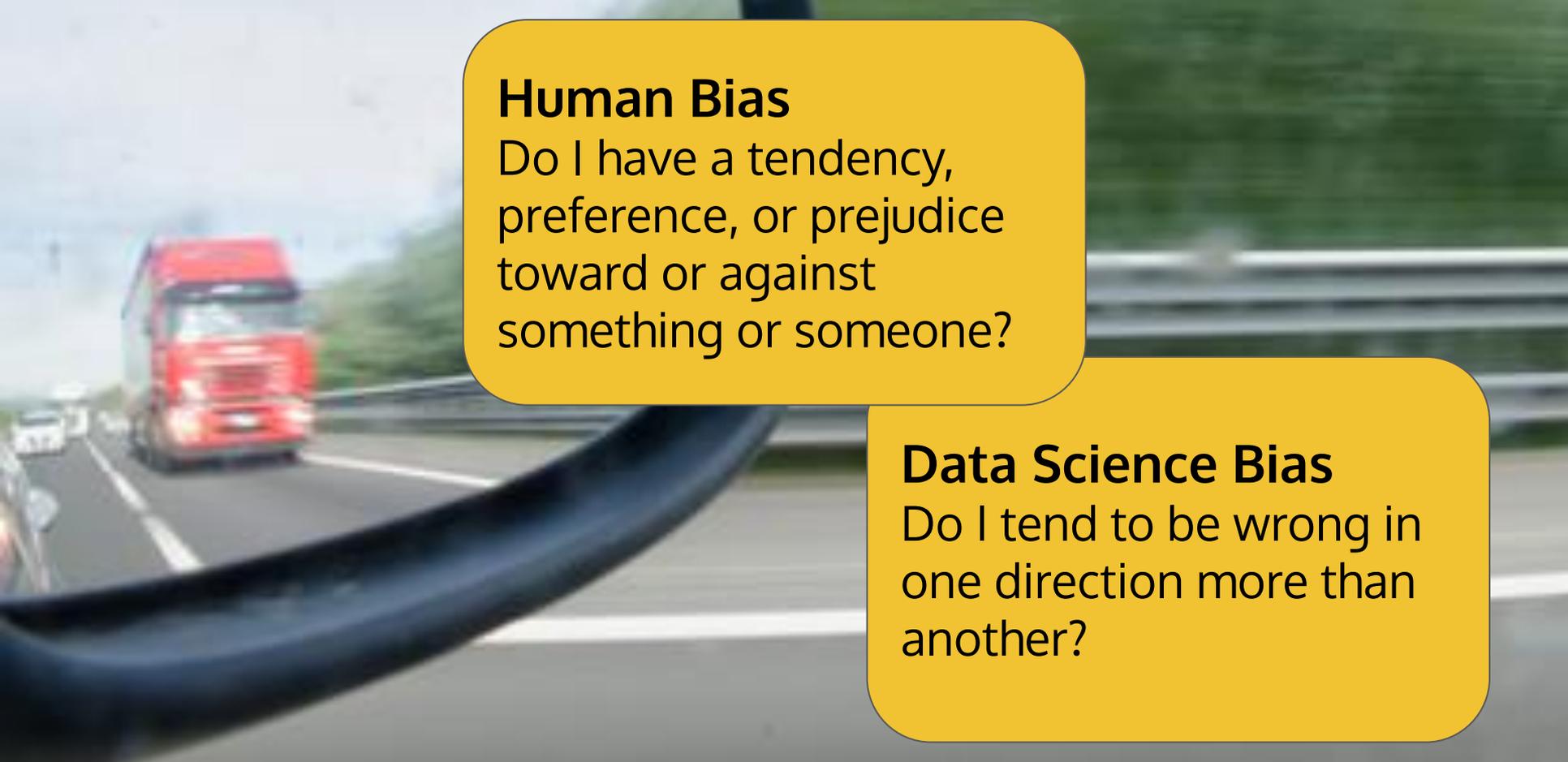
The ability for computers to detect patterns in large amounts of data using algorithms.
How the machine learns its human intelligence.



Deep Learning (Neural Networks)

An advanced machine learning technique that loosely models how neurons interact in the brain.





Human Bias

Do I have a tendency, preference, or prejudice toward or against something or someone?

Data Science Bias

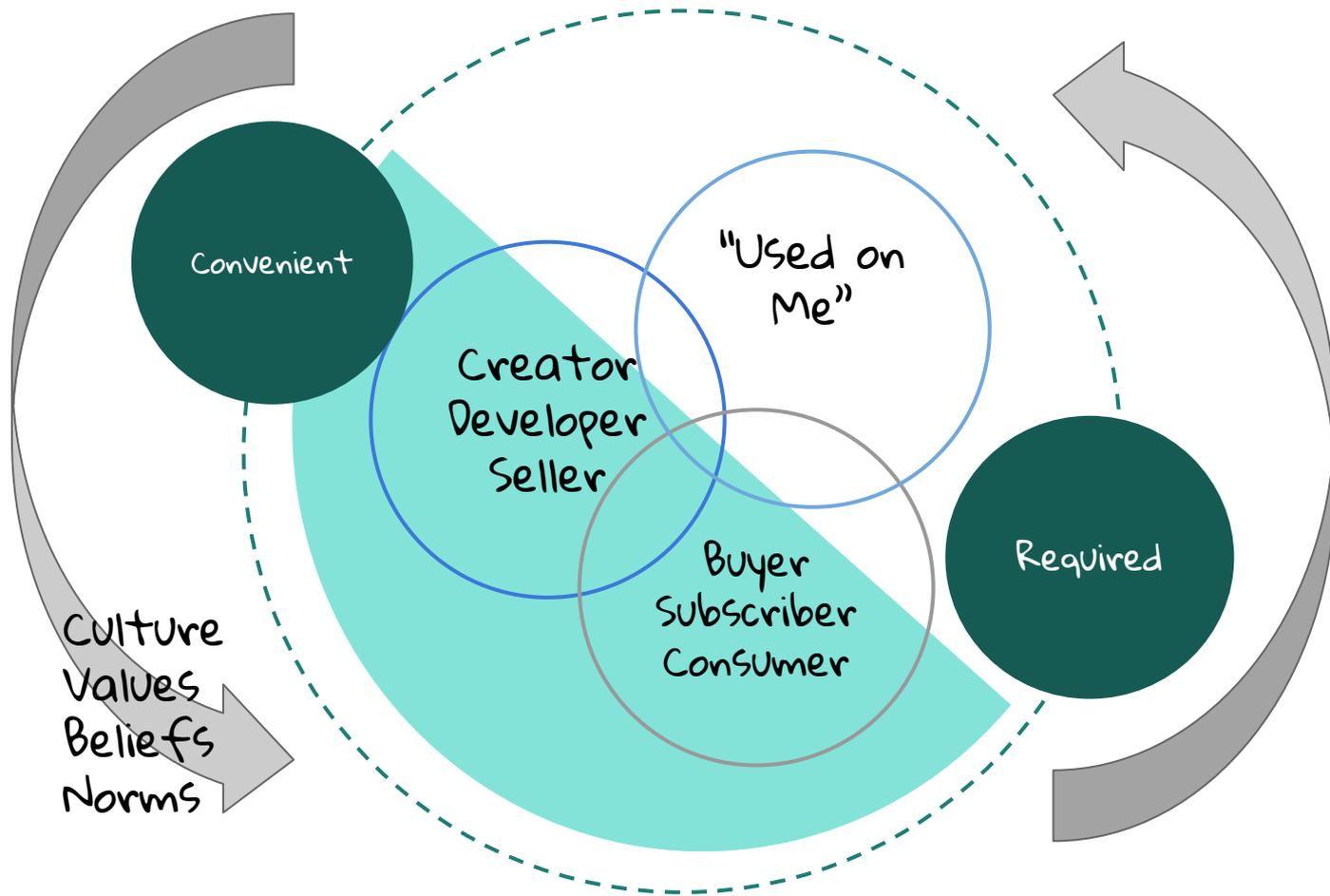
Do I tend to be wrong in one direction more than another?



AI and its algorithms can inherit the same biases as the humans who code them and the data we feed them.

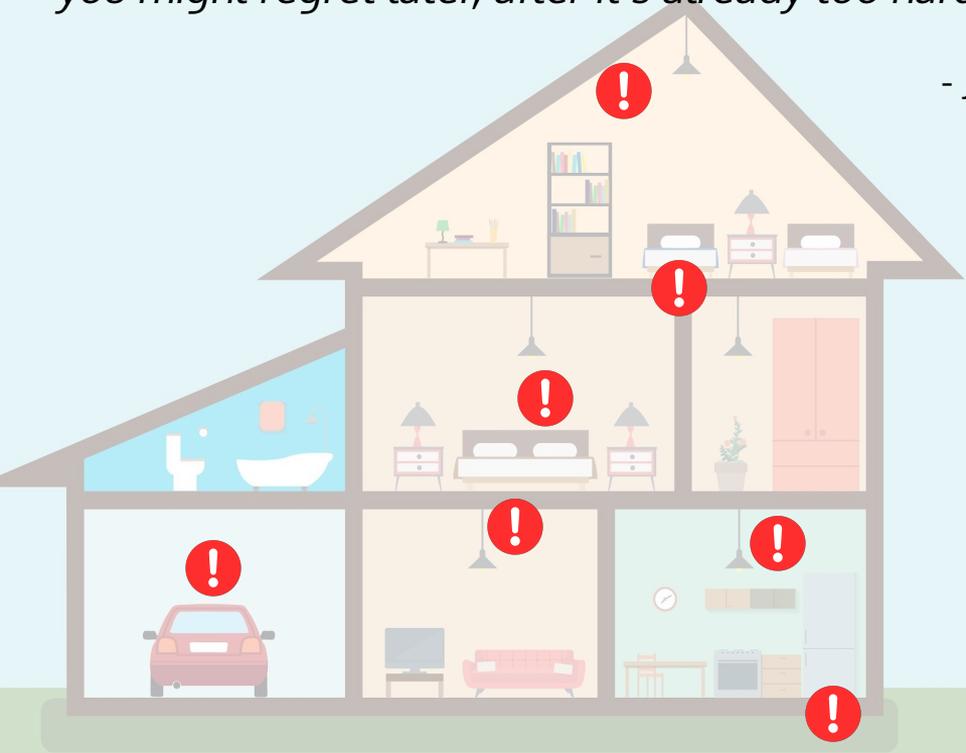
Because AI can have blind spots too.™



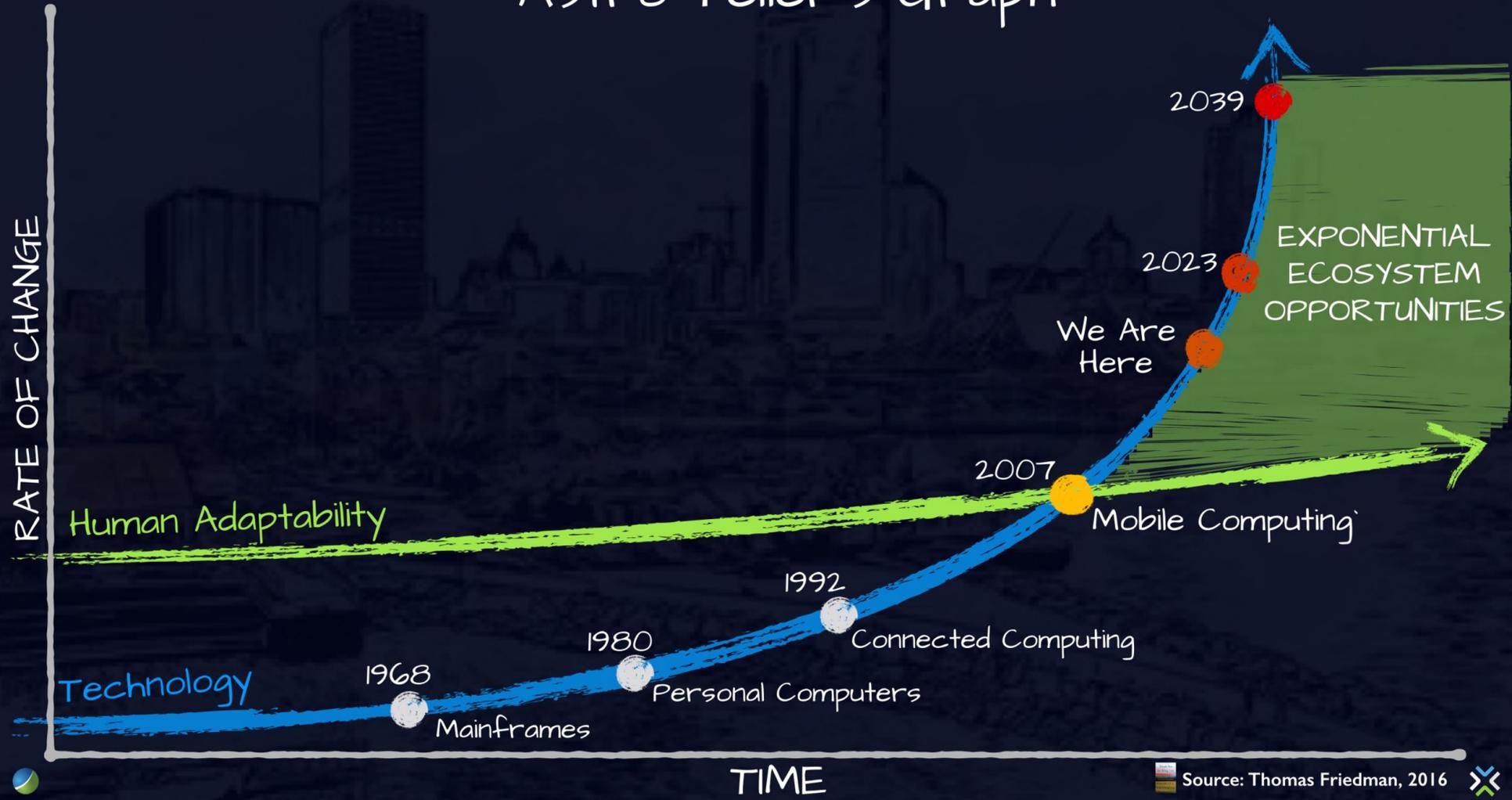


"I think of machine learning kind of as asbestos...It turns out that it's all over the place, even though at no point did you explicitly install it, and it has possibly some latent bad effects that you might regret later, after it's already too hard to get it all out."

- Jonathan Zittrain, Professor, Harvard Law School, 2019.



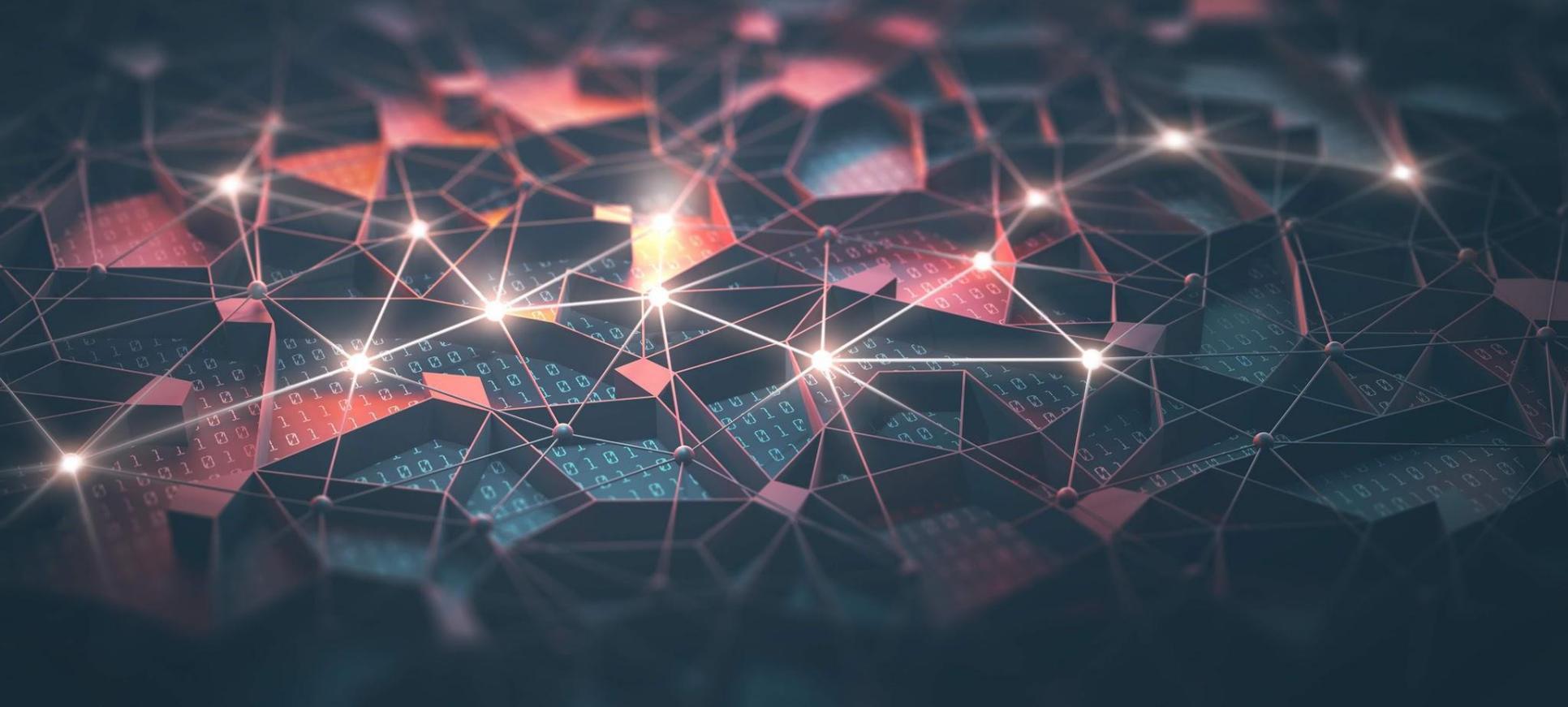
Astro Teller's Graph

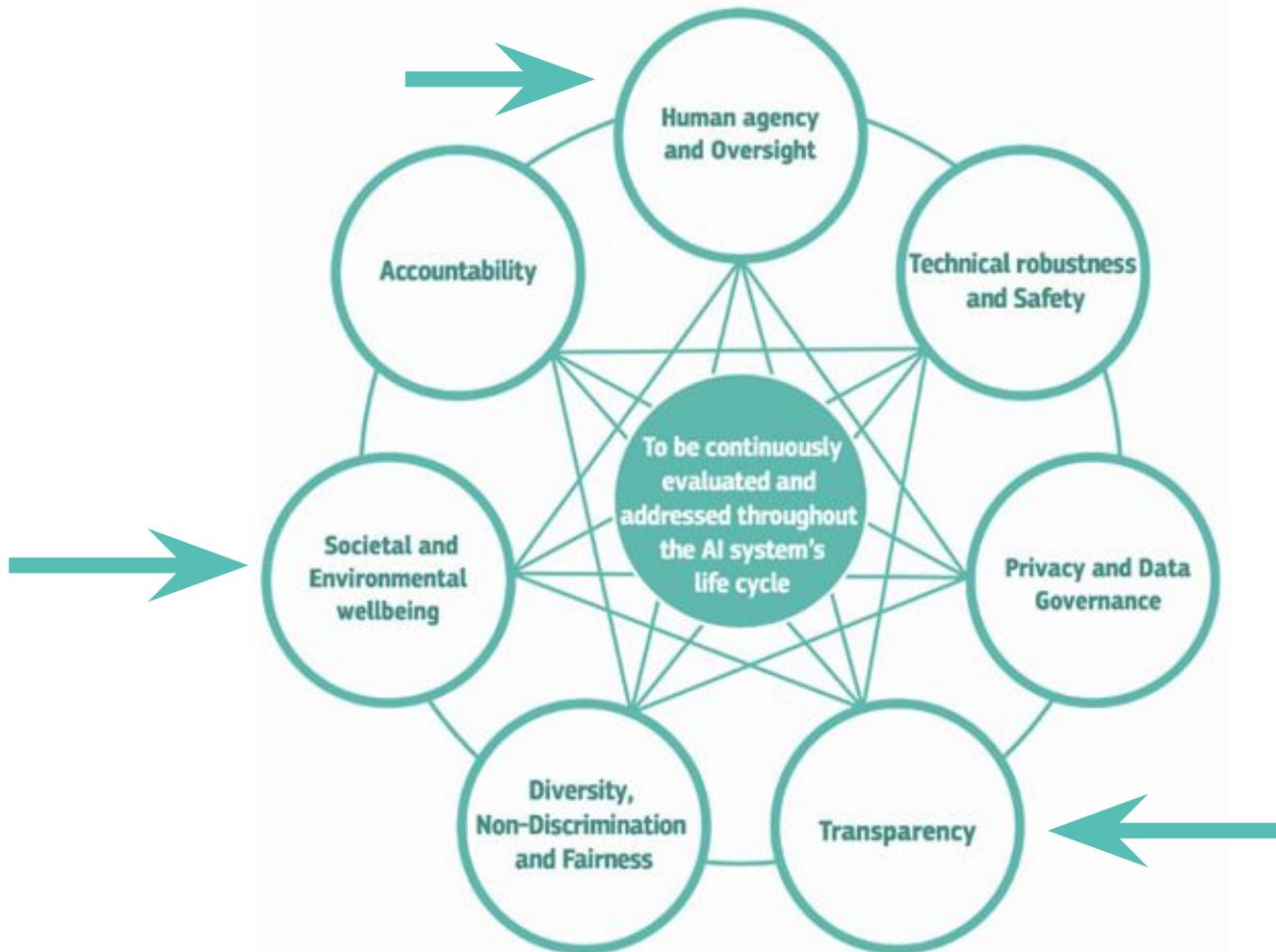


10 Places to Check for Unconscious Bias in Algorithms

1. Treating, Saving
2. Financing
3. Housing
4. Employing
5. Protecting
6. Enforcing laws
7. Educating
8. Securing
9. Insuring
10. Participating in our democracy

EU Commission, AI HLEG 7 Key Requirements for Trustworthy AI

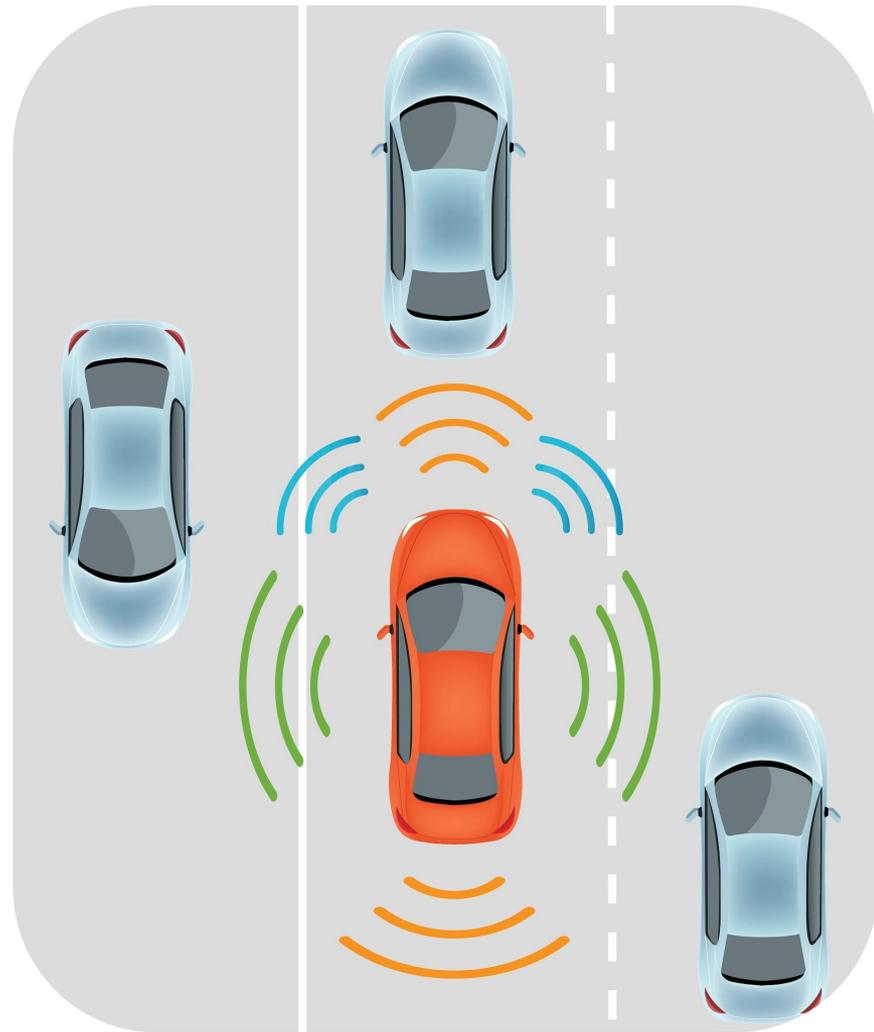


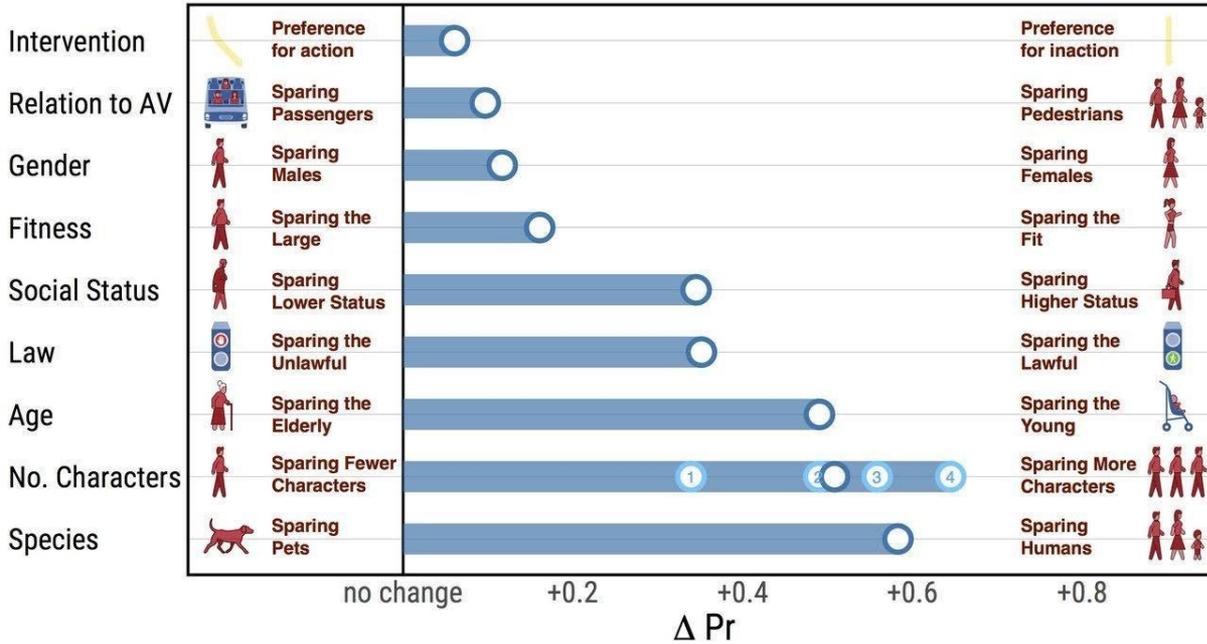
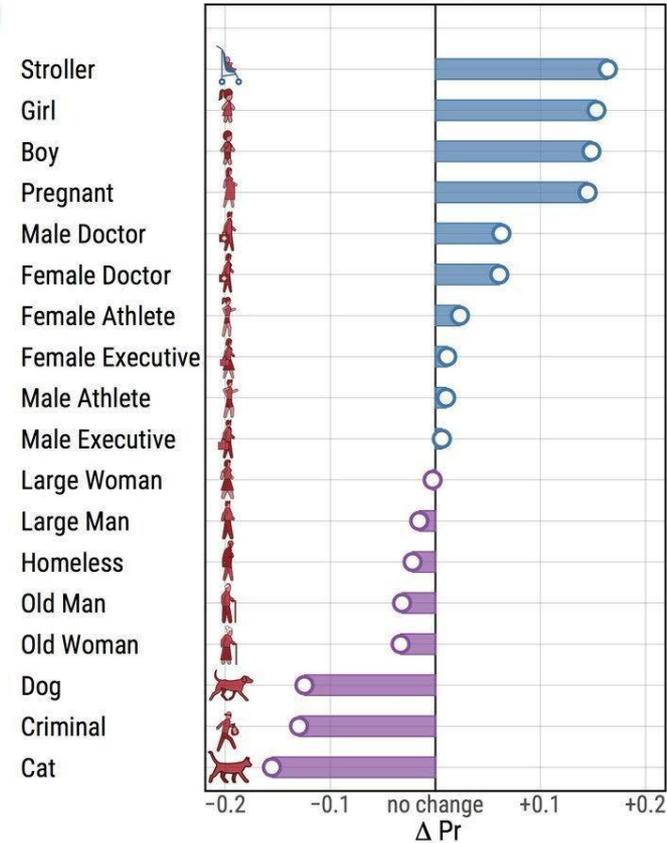


Source: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Human%20agency>



1. Human agency & oversight



a**Preference in favor of the choice on the right side****b****Preference in favor of sparing characters**

Source: <https://twitter.com/edmondawad/status/1055310123604807681>



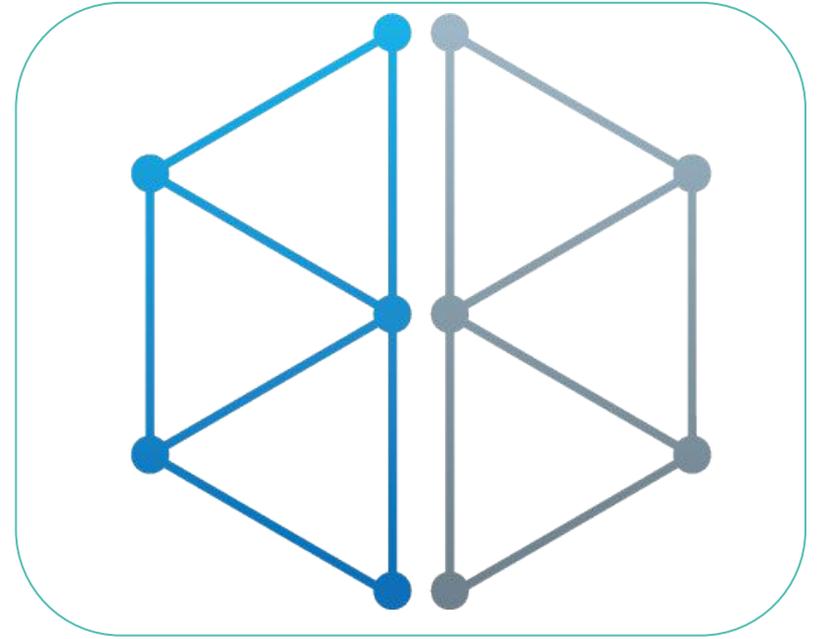
2. Technical robustness & safety





3. Privacy & data governance





4. Transparency



5. Diversity, non-discrimination & fairness





6. Societal & environmental wellbeing

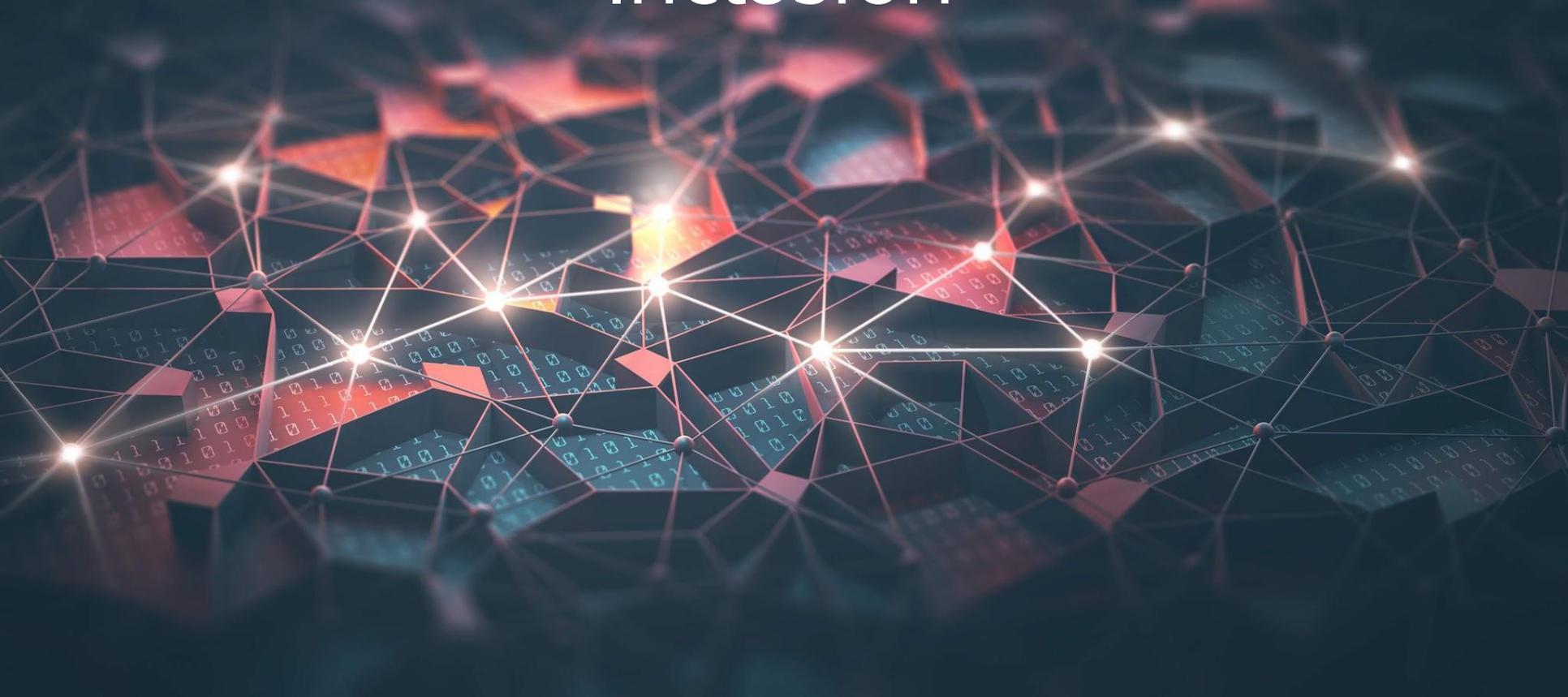


7. Accountability

Just because we *could*,
does that mean we *should*?



Diversity Inclusion



Trustworthy, Responsible, Transparent



Critical Thinking
Design Thinking
Resilience, Persistence, Agility





Trustworthy & Ethical AI

Carroll University ABIC Conference