

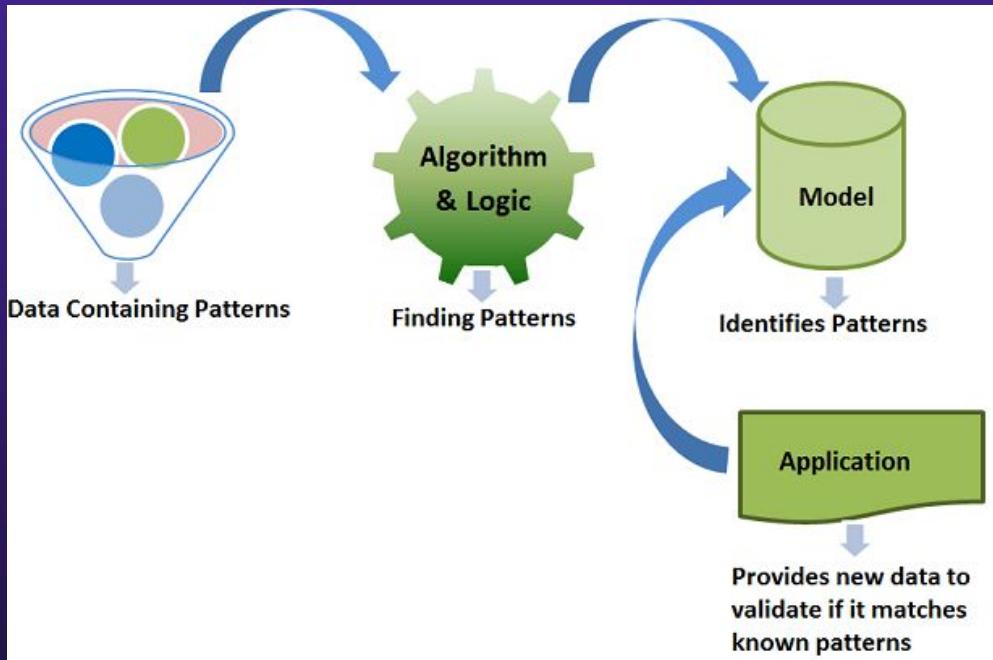


Avoiding bias in machine learning algorithms

© Kat Wallen, CodeClan 2022

Introduction

ML algorithms are **powerful tools**



Ethical challenges

- ★ Reinforces societal biases¹:
Gender²
Racial³
Sexuality
Ethnicity
- ★ Privacy
- ★ Informed Consent
- ★ Discrimination
- ★ Exploitation of vulnerable people



Potential solutions



★ Human oversight

- AI must remain **human-centric**
- commissions/groups
- > diversity

★ Careful selection of training data

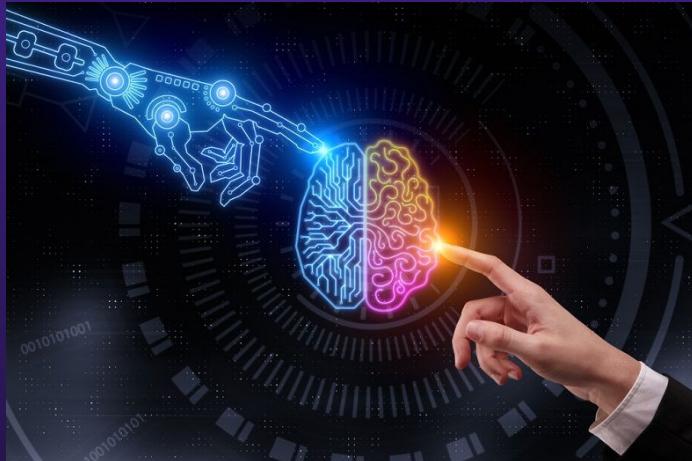
- ↓sampling bias
- assess input & output for bias

★ Legislation/regulation



Final thoughts

- ★ ML algorithms and AI are only going to become **more** ubiquitous & more sophisticated



- ★ Integration of humans and AI - **stronger** and **safer** than working separately
- ★ No simple or easy solution!

References

1. Bolukbasi, T et al. (2016) Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (29)
2. Dastin, R. (2018) Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, retrieved 28/07-2022
3. Kostick-Quenet, K. M. et al. (2022) Mitigating Racial Bias in Machine Learning. *Journal of Law, Medicine & Ethics* Volume 50 (1), pp. 92-100.

