

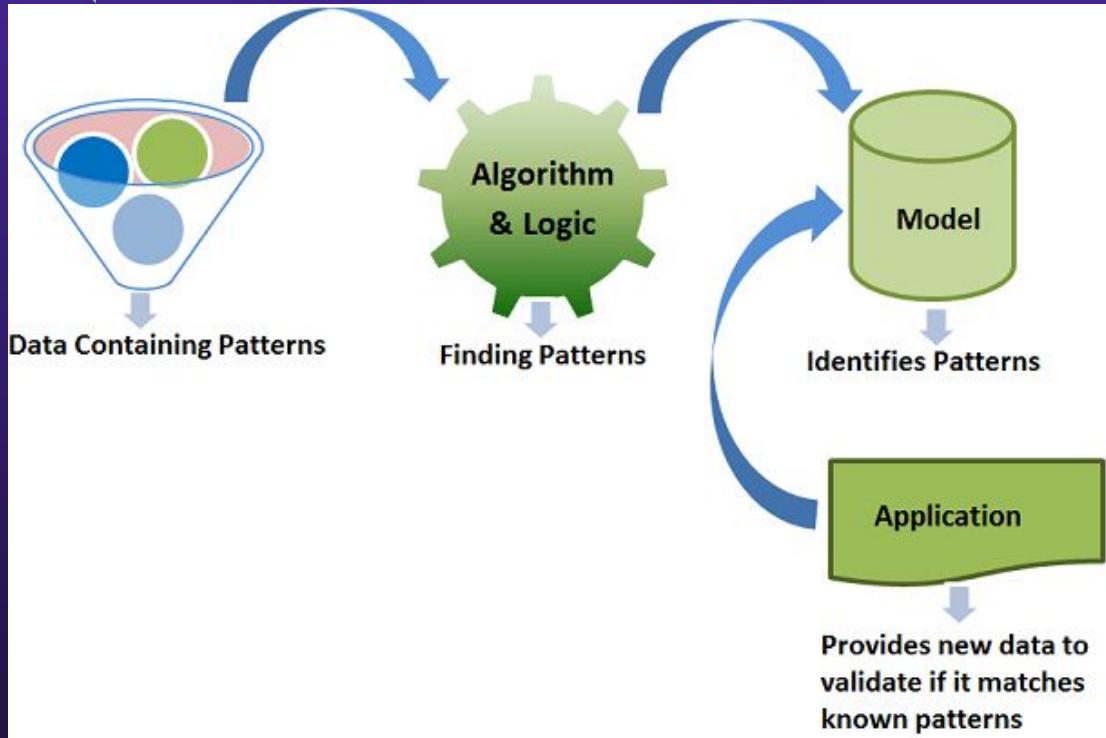


Avoiding bias in machine learning algorithms

© Kat Wallen, CodeClan 2022

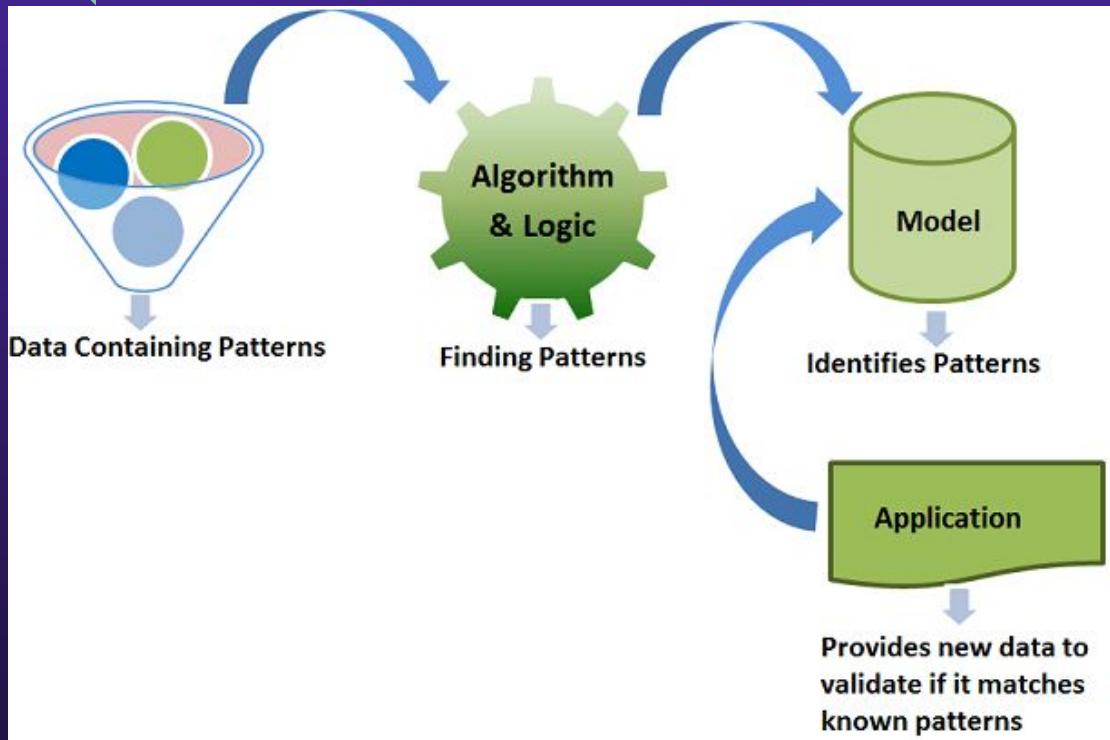
Machine Learning algorithms

★ Powerful tools



Machine Learning algorithms

★ Powerful tools



Ethical challenges

- ★ Reinforce societal biases¹:
Gender²
Racial³
Sexuality
Ethnicity⁴



Ethical challenges

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



Ethical challenges

- ★ Reinforces societal biases¹:
Gender²
Racial³
Sexuality
Ethnicity⁴
- ★ Privacy
- ★ Informed Consent
- ★ Discrimination
- ★ Exploitation of vulnerable people





Solutions?

- ★ **Careful selection of training data**
 - ↓sampling bias
 - assess input & output for bias

Solutions?



- ★ **Careful selection of training data**
 - ↓sampling bias
 - assess input & output for bias

Solutions?



- ★ **Careful selection of training data**
 - ↓sampling bias
 - assess input & output for bias

- ★ **Legislation/regulation**

Solutions?



★ Careful selection of training data

- ↓sampling bias
- assess input & output for bias

★ Legislation/regulation

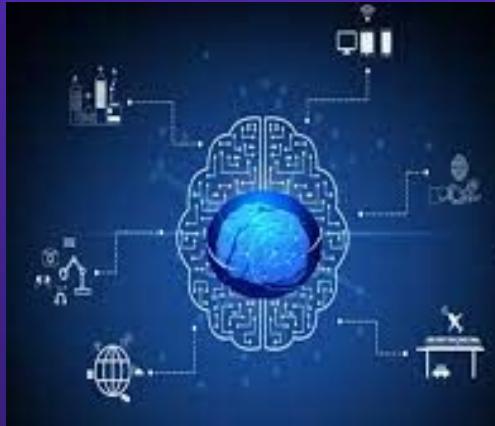


★ Human oversight

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

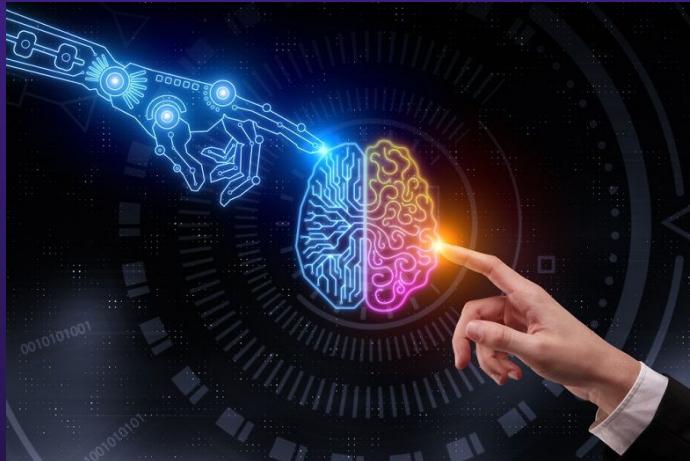
Final thoughts

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



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 Thompson*. Retrieved online July 28th 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.

4. Noseworthy, PA et al. (2020) Assessing and Mitigating Bias in Medical Artificial Intelligence: The Effects of Race and Ethnicity on a Deep Learning Model for ECG Analysis. *Circ Arrhythm Electrophysiol*. 13(3).

