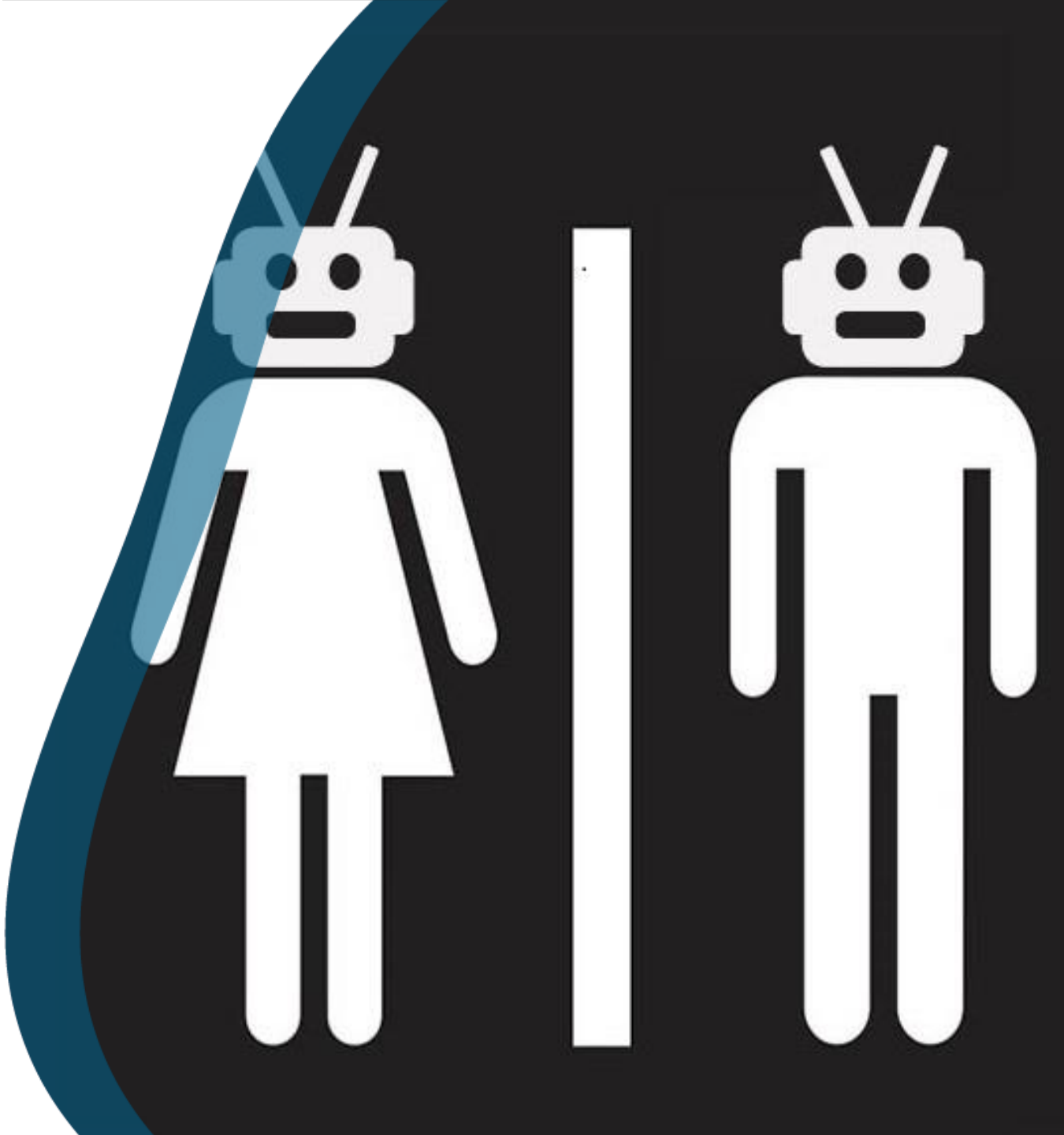
L-Università ta' Malta

What is AI?

And how your bias becomes AI's bias



Dr Claudia Borg
Lecturer
Department of Artificial Intelligence
Faculty of ICT, University of Malta

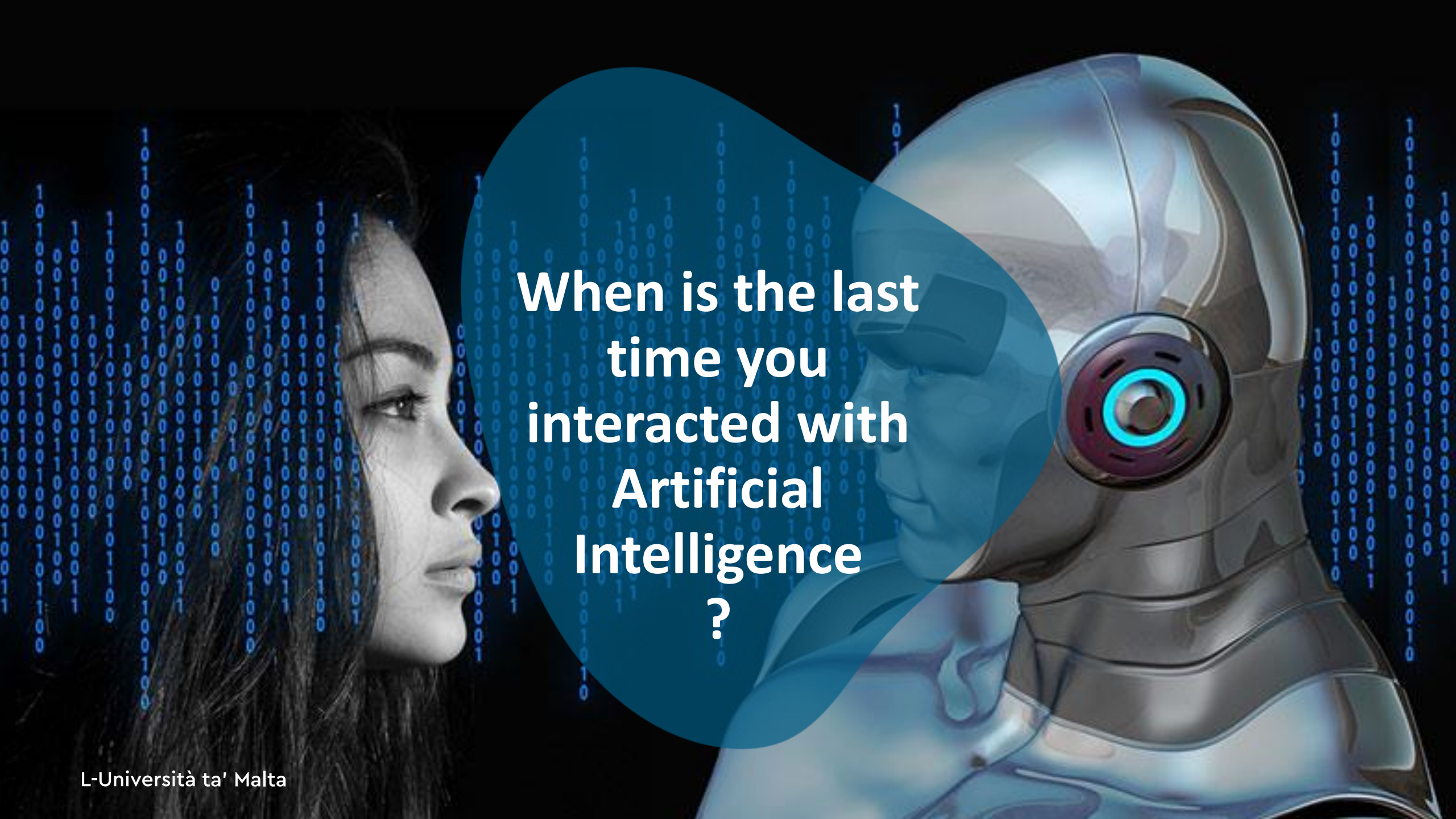
What is Artificial Intelligence?

- any device that **perceives** its environment …

  and takes **actions** …

  that maximize its chance of successfully achieving its **goals**

Our Perception of AI

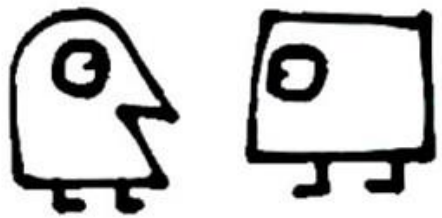

When is the last time you interacted with Artificial Intelligence ?

72
112 68
100

3 stages of AI

Narrow AI

Dedicated to assist with or take over specific tasks



General AI

Takes knowledge from one domain, transfers to other domain

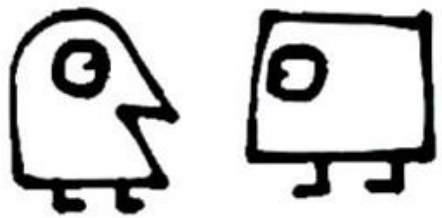
Super AI

Machines that are an order of magnitude smarter than humans

3 stages of AI

Narrow AI

Dedicated to assist with or take over specific tasks



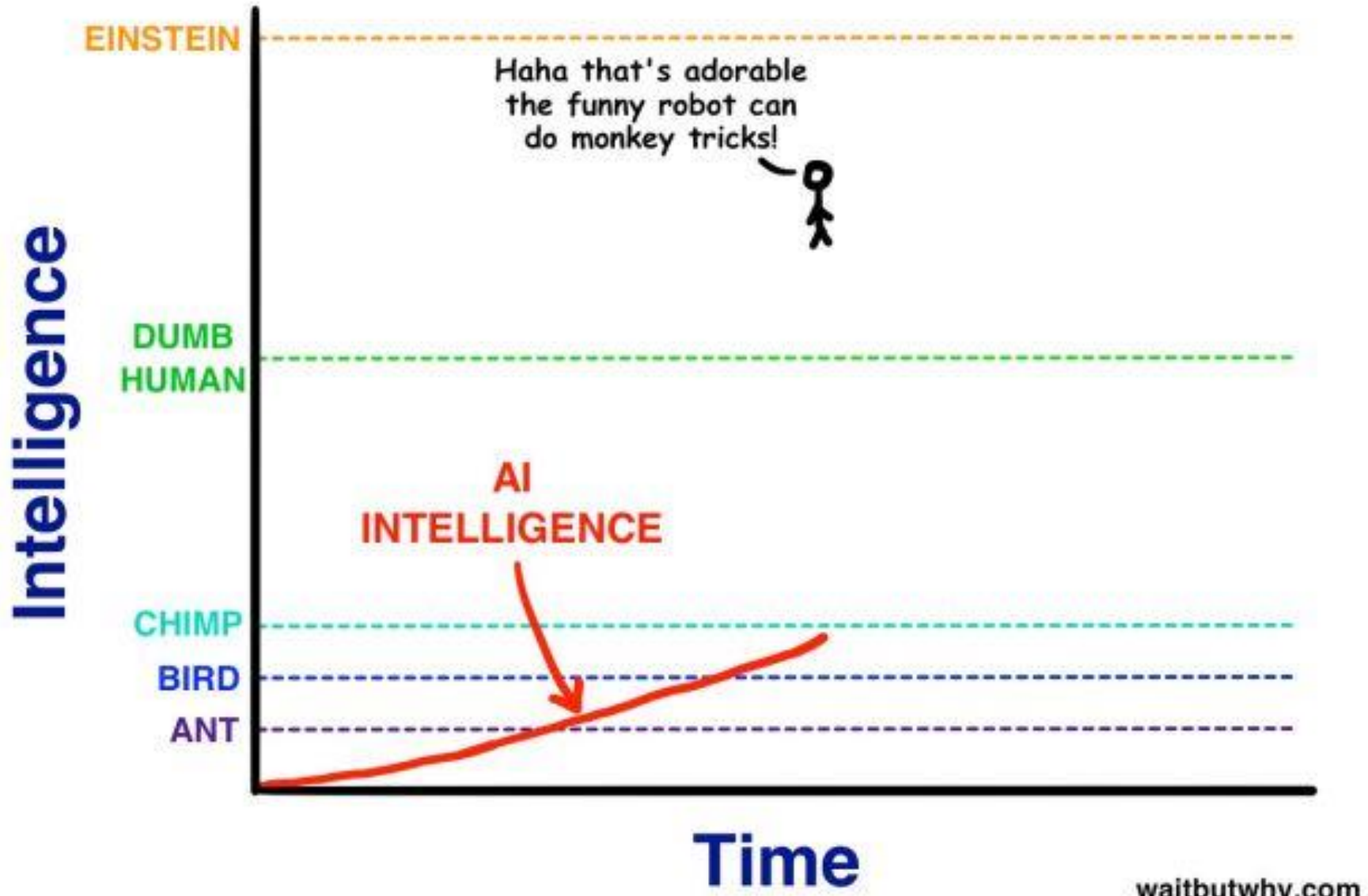
General AI

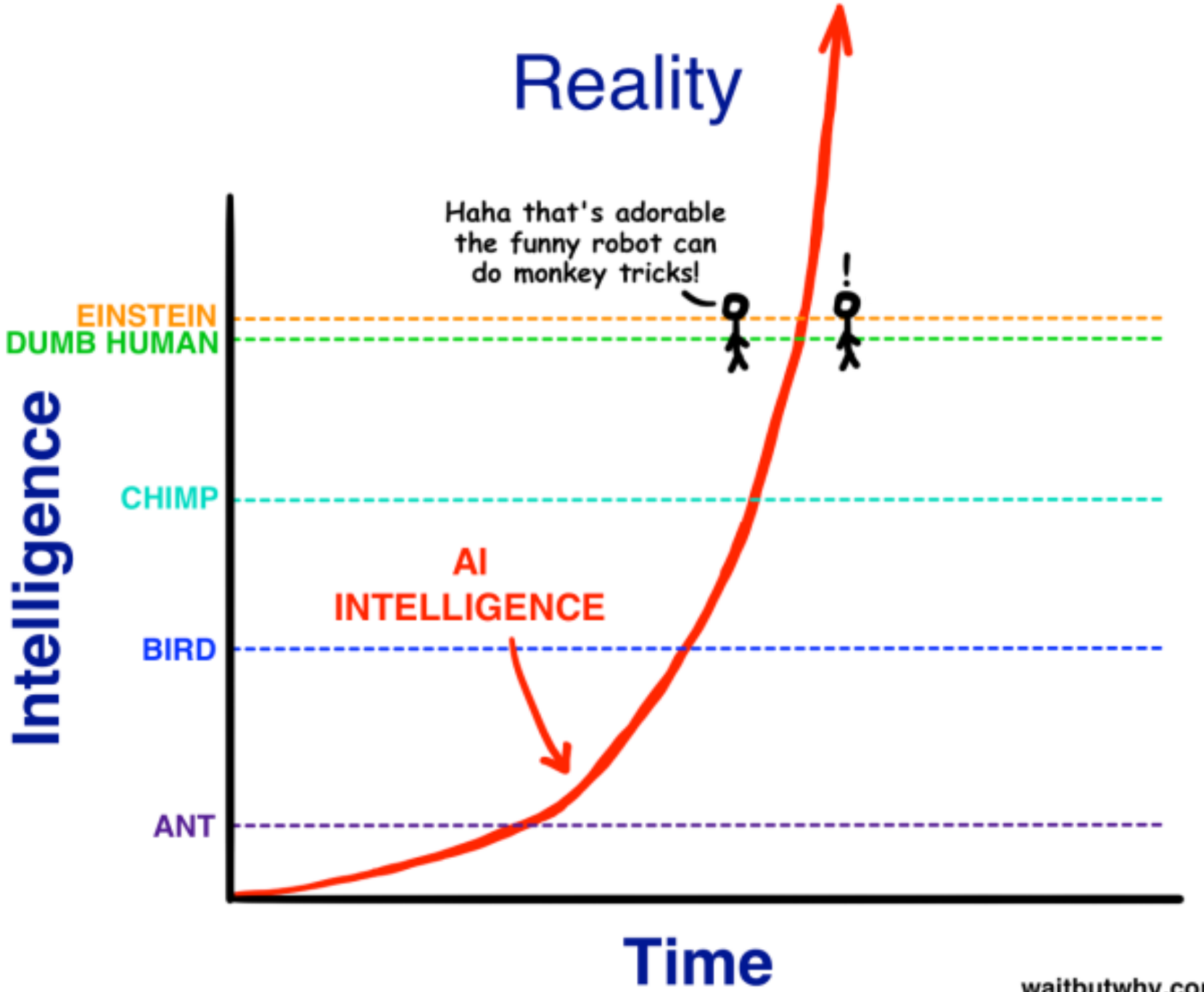
Takes knowledge from one domain, transfers to other domain

Super AI

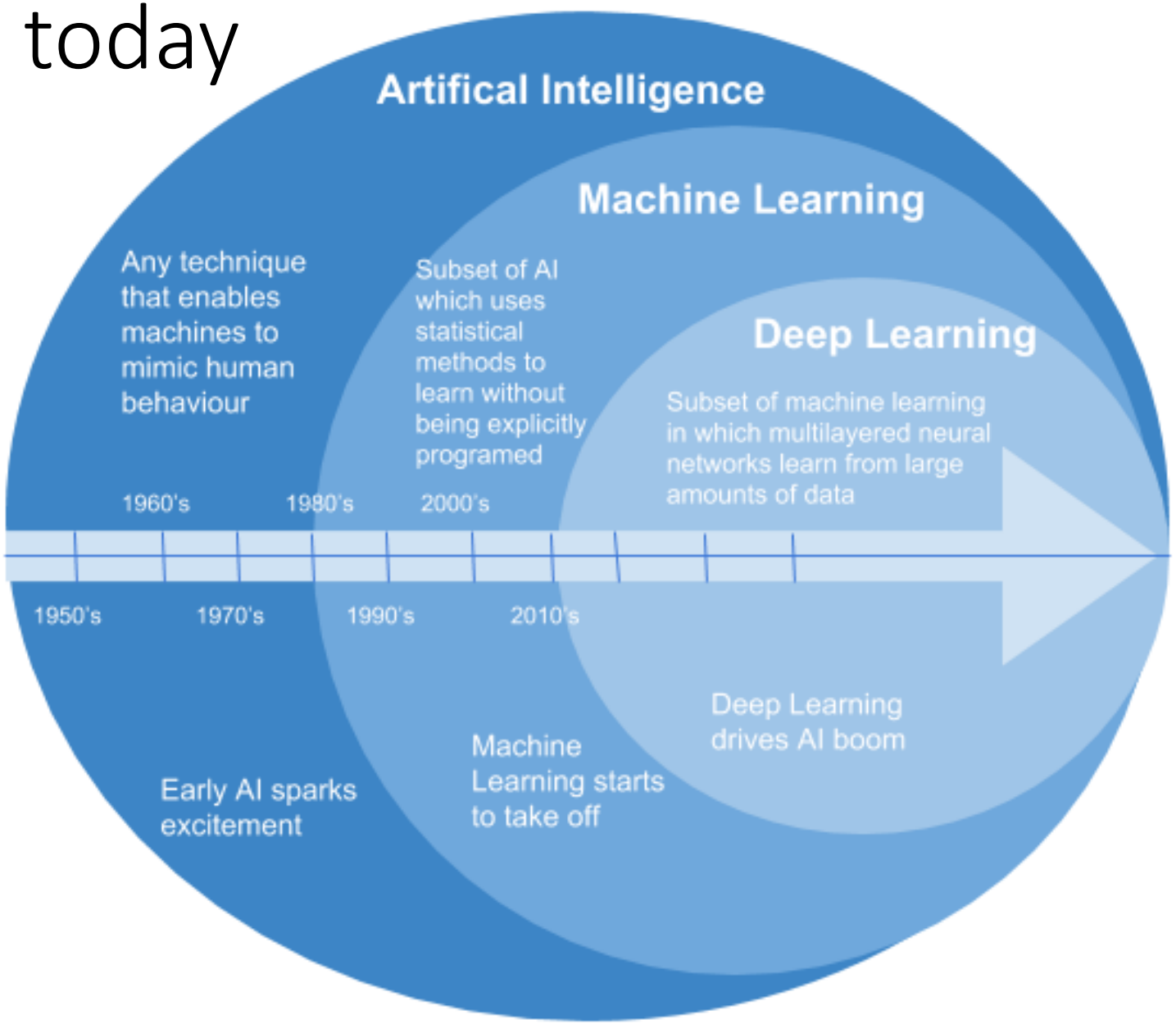
Machines that are an order of magnitude smarter than humans

Our Distorted View of Intelligence




Reality
Haha that's adorable the funny robot can do monkey tricks!
!
EINSTEIN
DUMB HUMAN
CHIMP
AI INTELLIGENCE
BIRD
ANT
Intelligence
Time
waitbutwhy.com

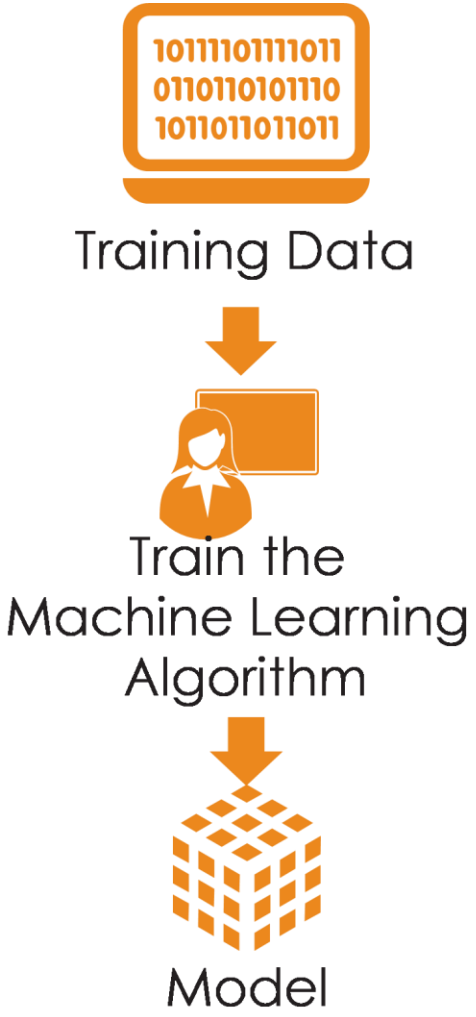
AI...history and today



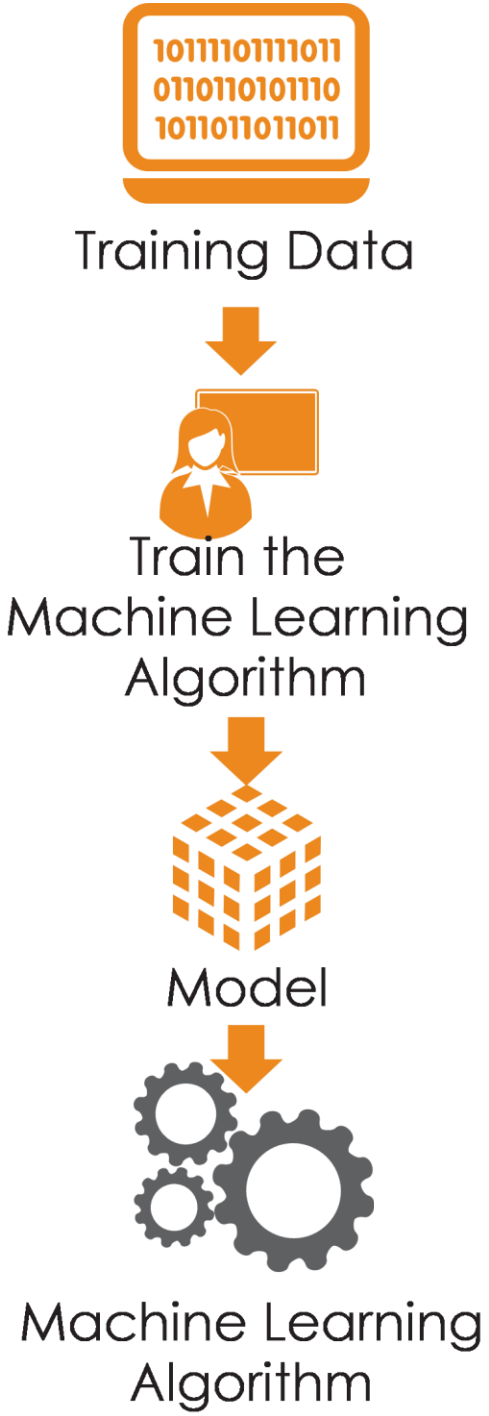
Machine Learning

Machine Learning

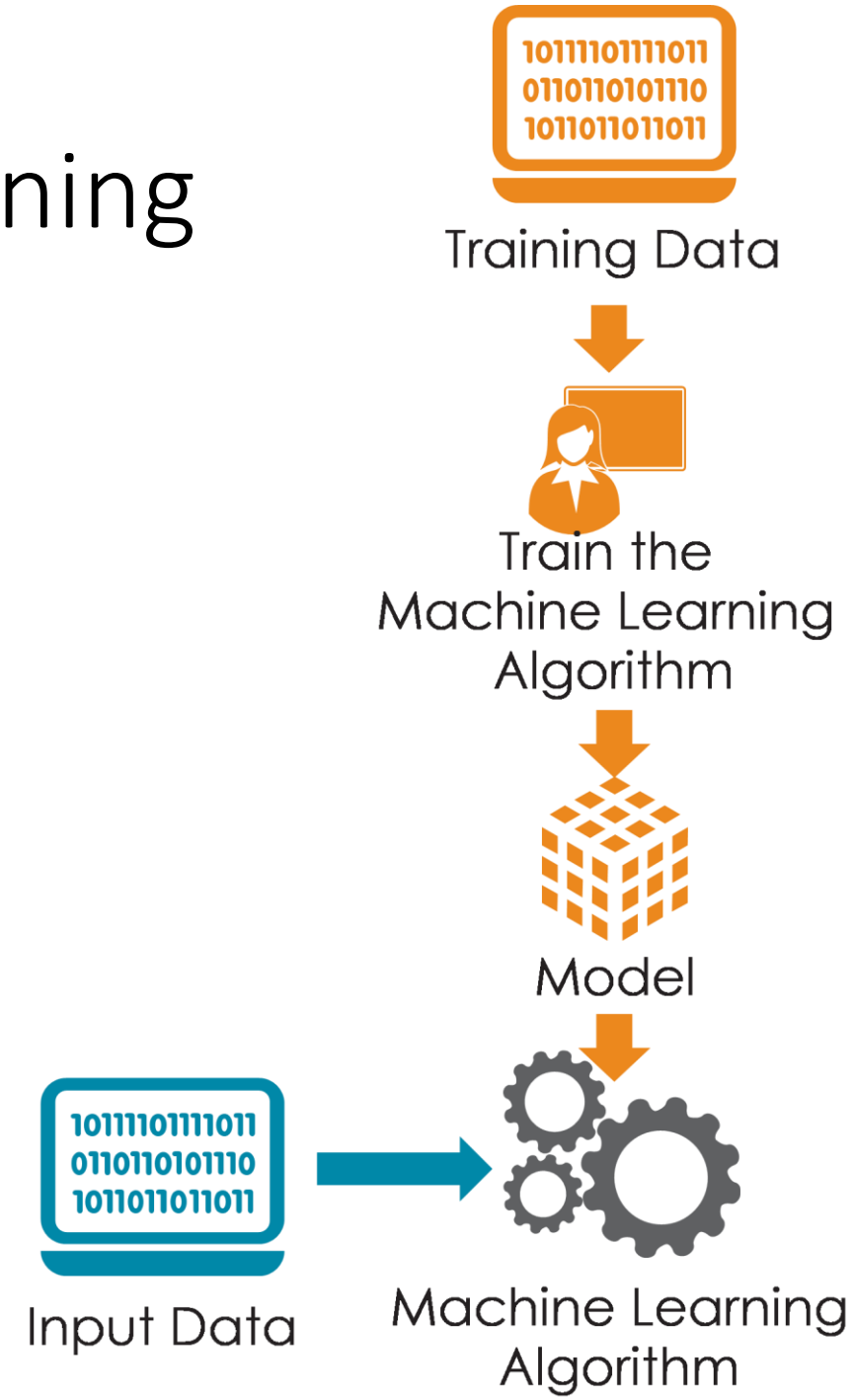
Machine Learning



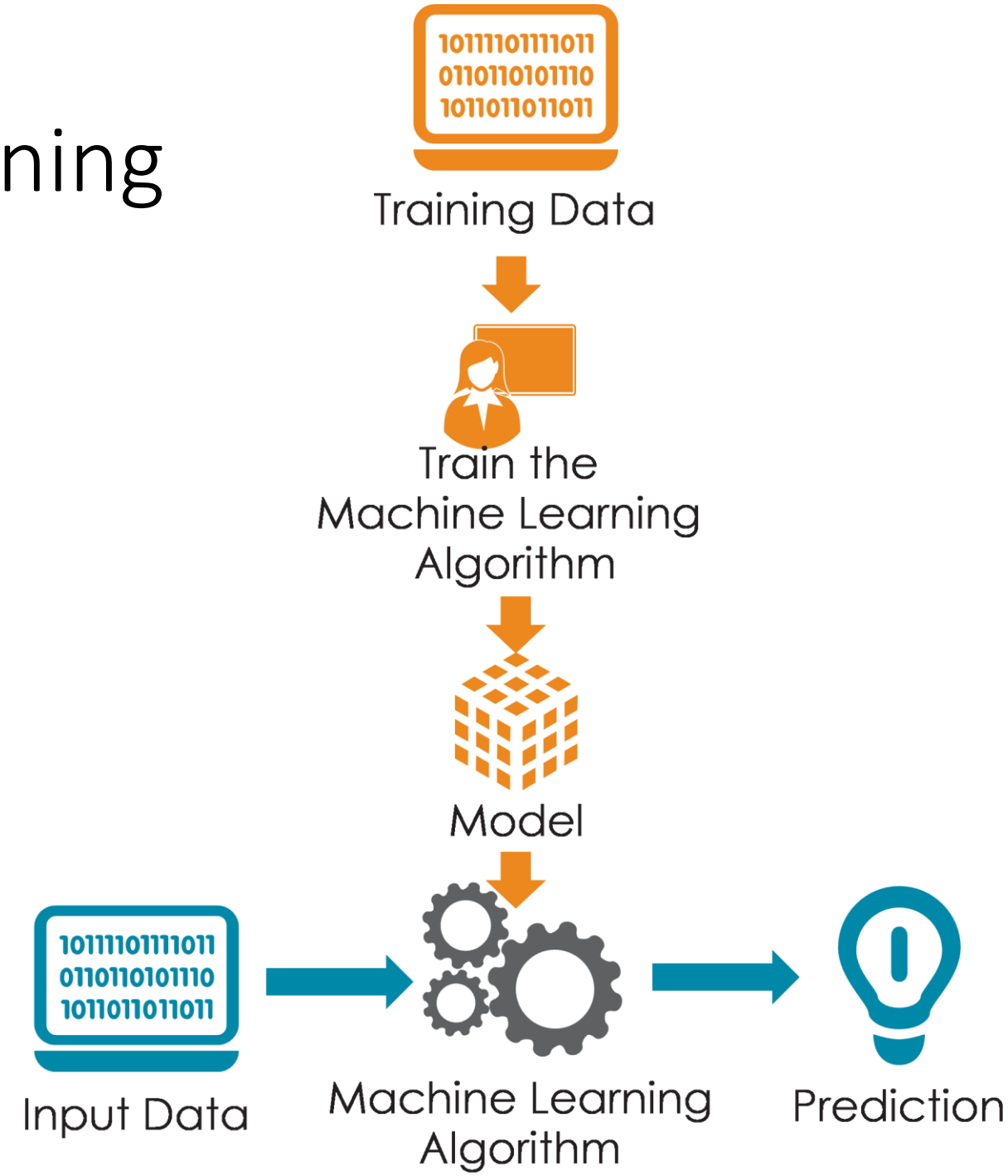
Machine Learning



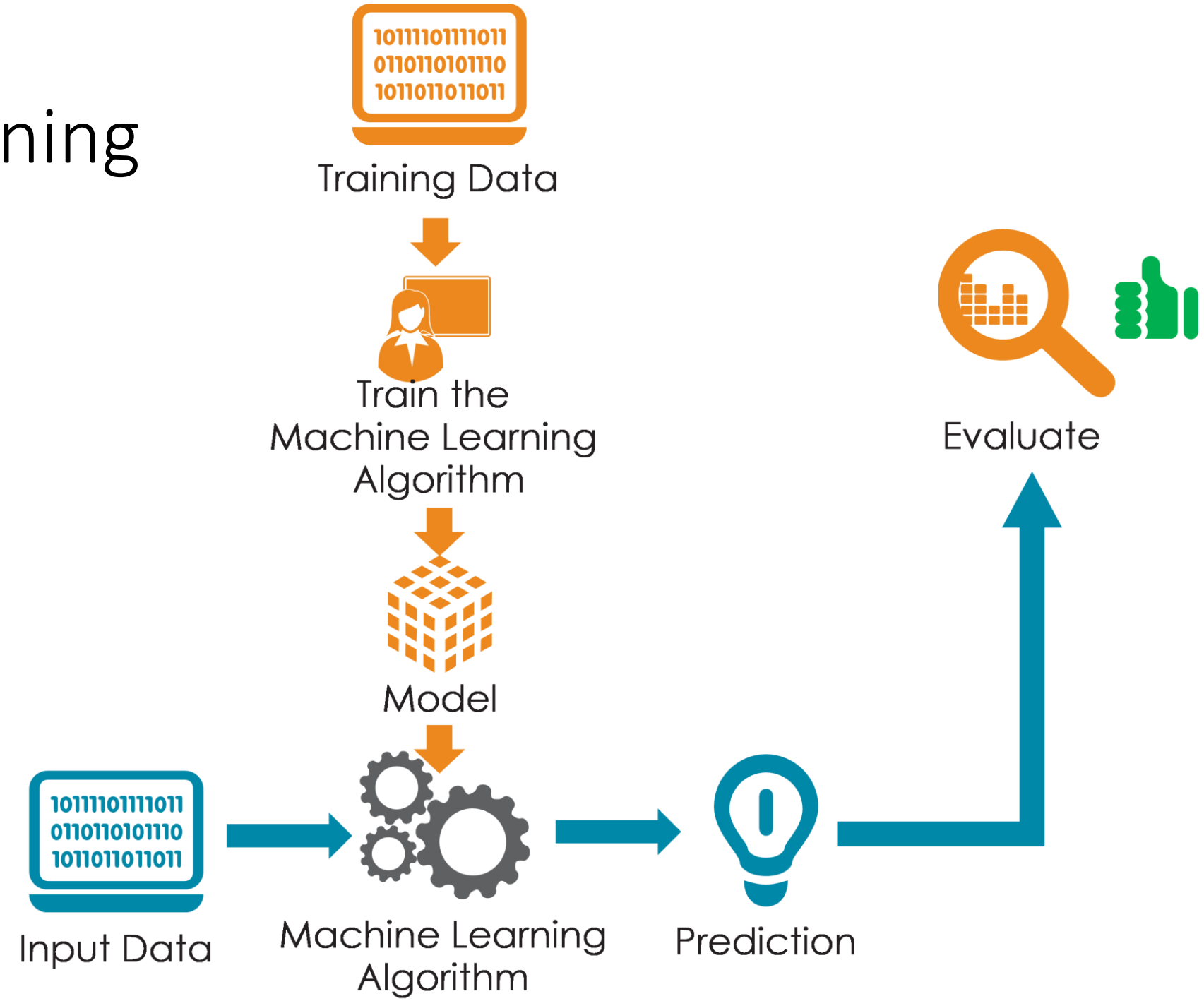
Machine Learning



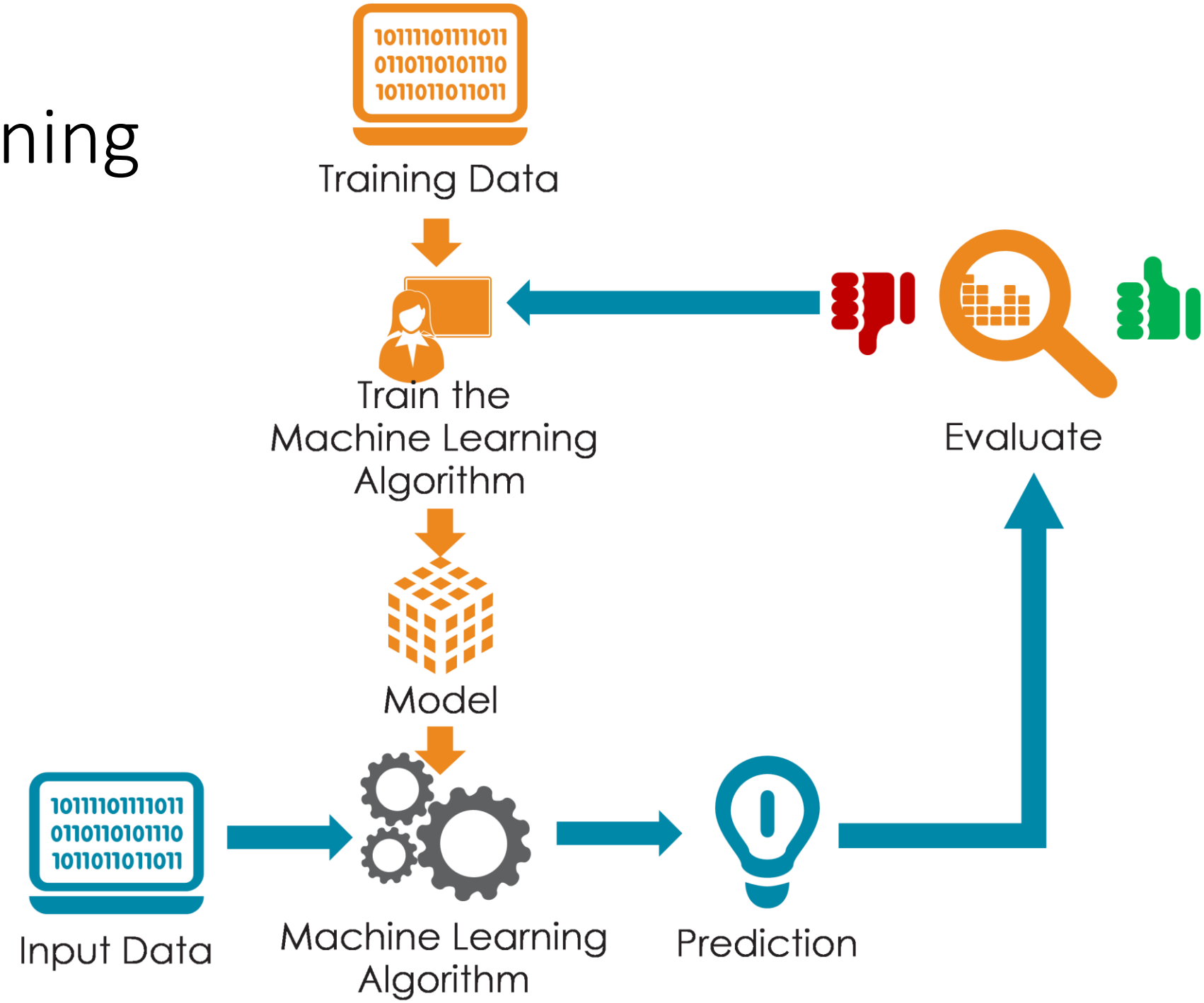
Machine Learning



Machine Learning



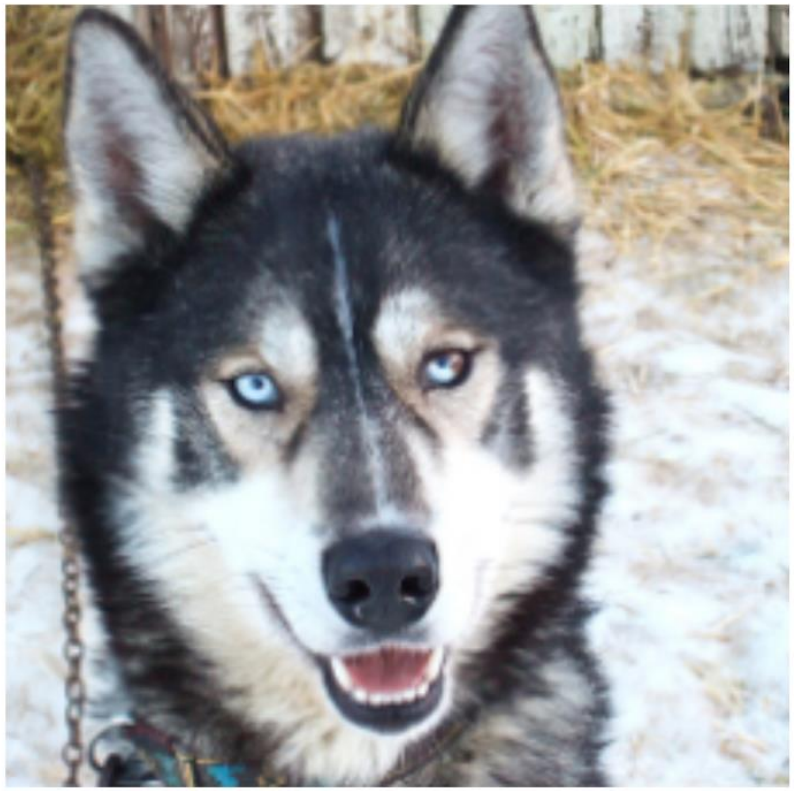
Machine Learning



Husky or wolf?

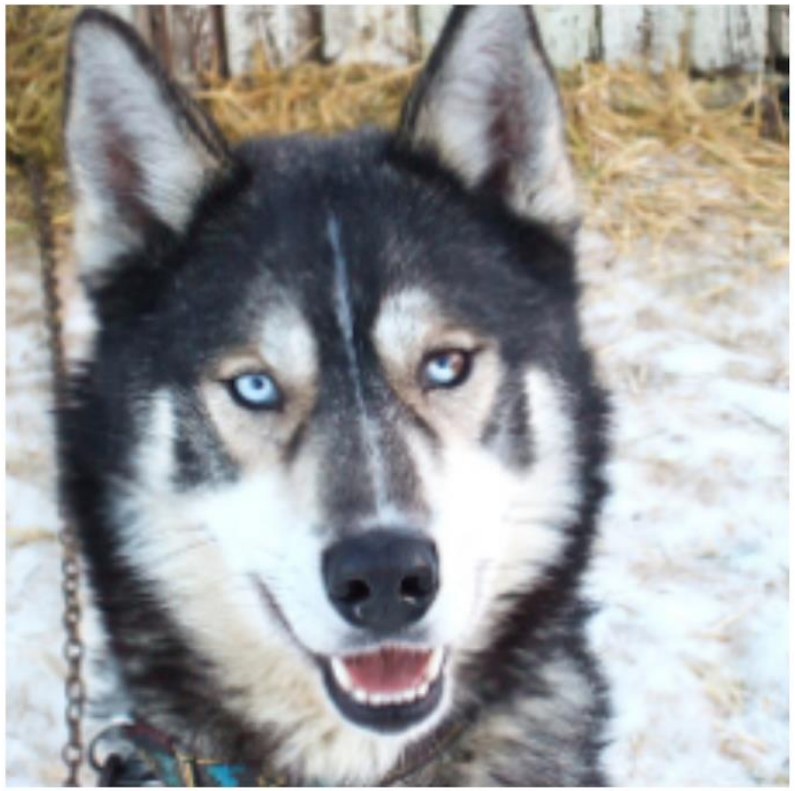
Husky or wolf?

Husky or wolf?



(a) Husky classified as wolf

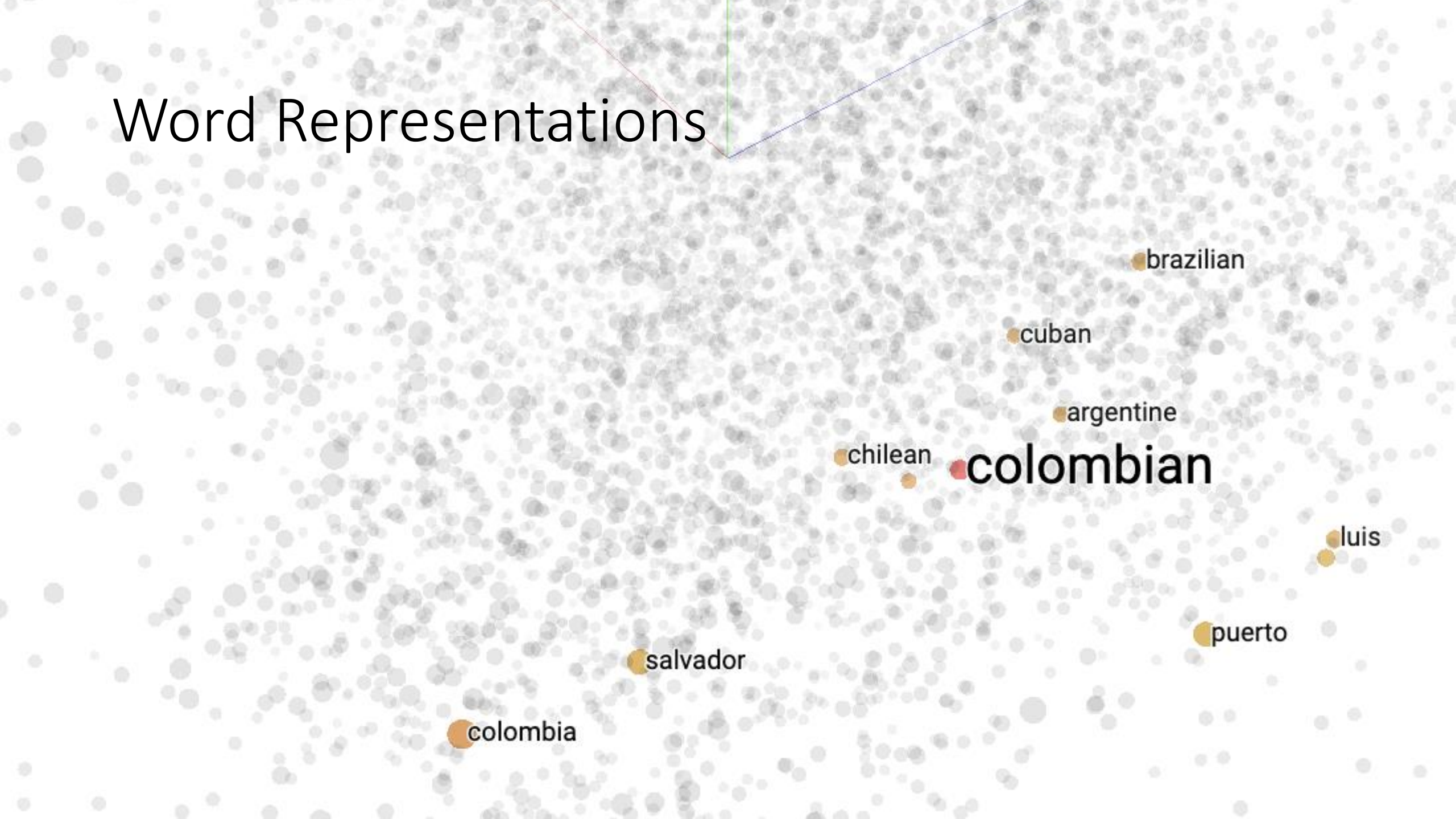
~~Husky or wolf?~~ Snow!



(a) Husky classified as wolf

(b) Explanation

Word Representations



Word Representations reflect cultural bias

- Ask “Paris : France :: Tokyo : x”
  - x = Japan
- Ask “father : doctor :: mother : x”
  - x = nurse
- Ask “man : computer programmer :: woman : x”
  - x = homemaker

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." In *Advances in Neural Information Processing Systems*, pp. 4349-4357. 2016.

Word Representations reflect cultural bias

- Psychological findings on US participants:
- African-American names are associated with unpleasant words (more than European-American names)
- Male names associated more with math, female names with arts
- Old people's names with unpleasant words, young people with pleasant words.

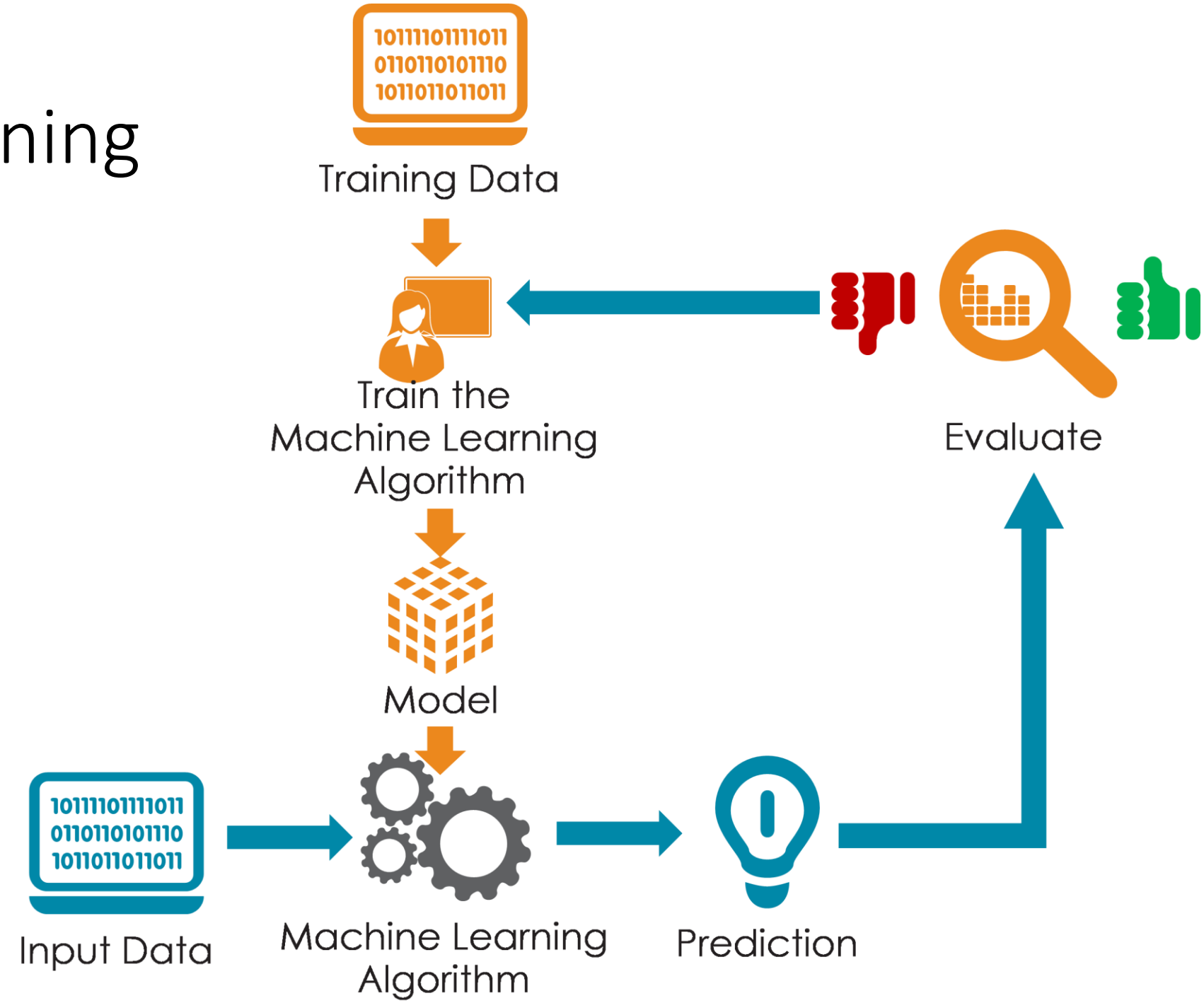
Caliskan, Aylin, Joanna J. Bruson and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. Science 356:6334, 183-186.

History of biased framings of women

- Embeddings for competence adjectives are biased toward men
  - *Smart, wise, brilliant, intelligent, resourceful, thoughtful, logical, etc.*
- This bias is slowly decreasing

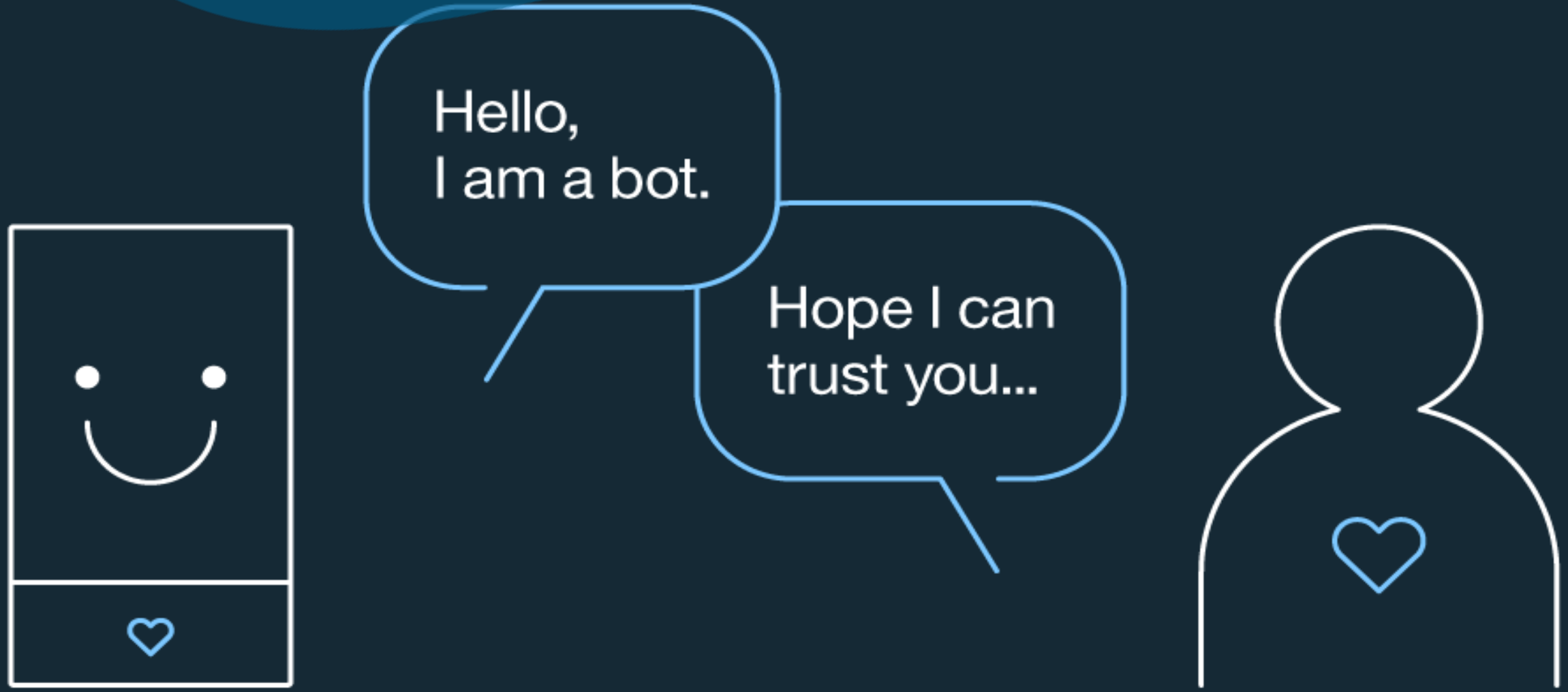
Garg, Nikhil, Schiebinger, Londa, Jurafsky, Dan, and Zou, James (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644

Machine Learning



We are the data!

Ethical Concerns



Trust

- What is needed for us to trust the decisions that AI systems make?

Trust

- What is needed for us to trust the decisions that AI systems make?
- AI must *Explain* itself
- We tend to trust AI systems because they are working for us…

Trust

- What is needed for us to trust the decisions that AI systems make?
- AI must *Explain* itself
- We tend to trust AI systems because they are working for us…
- Or for the company that created that system?

TOP REASONS BANKS USE ARTIFICIAL INTELLIGENCE

DATA ANALYSIS & INSIGHT

60%

INCREASED PRODUCTIVITY

59%

COST BENEFITS /SAVINGS

54%

L-Università ta' Malta

Thank you for your attention.

If you have any questions, now is the right time to ask!
Contact me on the details below for further info:

Claudia Borg

(+356) 2340 3539

/AIMalta

claudia.borg@um.edu.mt