# Explain: Algorithmic Bias

Sophie Wang
Prof. Sendhil Mullainathan
BUSN 32200: Artificial Intelligence
6 March 2022

CHICAGO BOOTH

# Does AI really have “intelligence”?

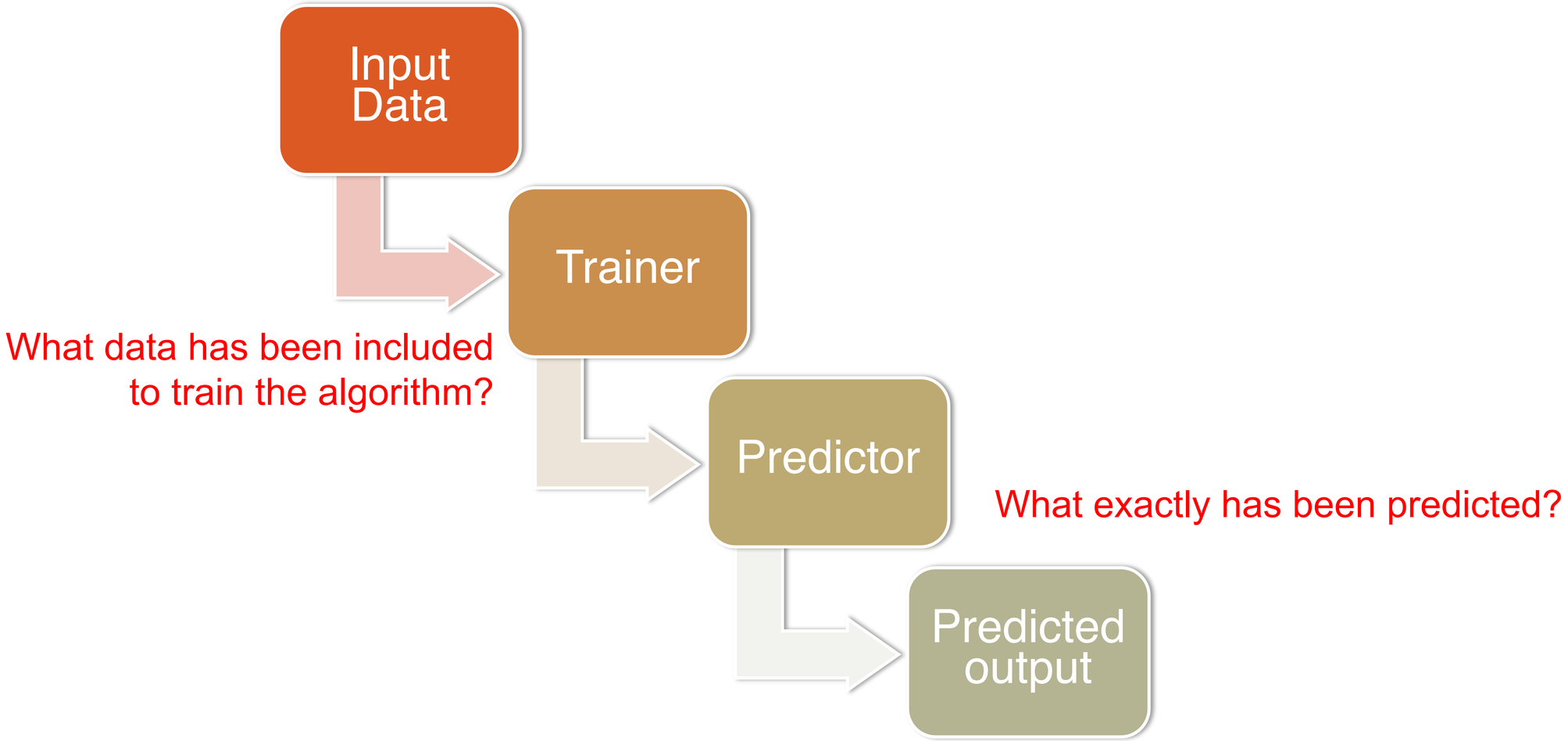
- **Background knowledge:**

  Four Components of a supervised learning application:
  - Input, desired output
  - Data
  - Trainer (training algorithm needed)
  - Predictor (deployment algorithm needed)

- **Conclusion**: The only learning source of AI: data

- **Problems?**

# Two Problems Lead to Bias



# Algorithmic bias type 1: *Non-diverse training data*

- The two algorithms are aimed at the right target
  - but they fail to represent underserved groups.

- Causes biases in AI

  *- Example from Healthcare: A medical algorithm that was trained on primarily White patients performed poorly in Black patients.*[1]

[1] *Michael W. Sjoding, et al. "Racial bias in pulse oximetry measurement." New England Journal of Medicine 383, no. 25 (2020): 2477-2478.*

# Algorithmic bias type 2:
# *ideal label VS. predicted label*

- Recall the previous example from laws
  - oftentimes “proxies” cannot capture the reality
  - *Example from Laws: A prison system’s minimum height and weight requirement for hiring that discriminates against female applicants.*[1]

- A <u>mismatch</u> between the ideal label and predicted label

[1] *Robert P. Bartlett et al. “Algorithmic Discrimination and Input Accountability under the Civil Rights Acts (August 1, 2020).”*

Drew by Joost Swarte

# Who is easier to be fixed:
# *Biased humans or biased algorithms?*

***Recommended Reading**:*
*"Biased Algorithms Are Easier to Fix Than Biased People,"*
*Sendhil Mullainathan, New York Times, 12/6/19*

- Think about this open question for a minute…
- AI systems are incredibly narrow in what they can do
- They can do certain tasks, but human-style generalizations do not apply.[1]
- Solutions to fix algorithmic bias
  - Be careful at the blueprint stage: when and how will the algorithms be used?
  - Automated human check: how far is the predicted label from the ideal label?
  - Retrain the model on a more accurate label

[1] "The Seven Deadly Sins of AI Predictions," Rodney Brooksarchive, *MIT Technology Review, 10/6/17*