What is an algorithm?

06 Jan 2022

wrapping up last class

Evaluating potential actions

- Which action will produce the most good and do the least harm? (The Utilitarian Approach)
- Which action respects the rights of all who have a stake in the decision? (The Rights Approach)
- Which action serves the community as a whole, not just some members?
  (The Common Good Approach)
- Which action leads me to act as the sort of person I should be? (The Virtue Approach)

(A Framework for Making Ethical Decisions)

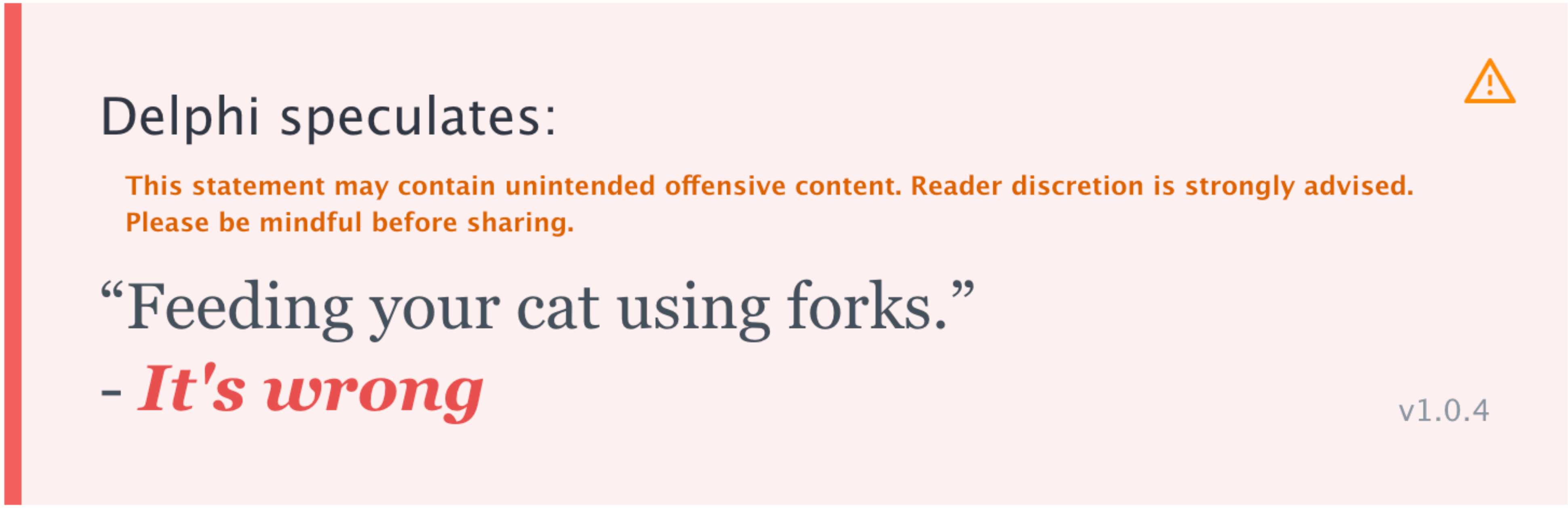
A step back

- Do you think you rely upon any of these ethical approaches when deciding what to do?
- Think back to a time you lied, or a time you could have lied, but chose not to do so. How did you decide whether to lie?

Dangers of a limited perspective

- Many of these ethical perspectives/approaches are drawn from Western traditions and thinkers

- "Fairness," "privacy," and "bias" mean different things (pdf) in different places. People also have disparate expectations of these concepts depending on their own political, social, and economic realities. The challenges and risks posed by AI also differ depending on one's locale." (Gupta and Heath)



⚠

Delphi speculates:

This statement may contain unintended offensive content. Reader discretion is strongly advised. Please be mindful before sharing.

“Feeding your cat using forks.”

- ***It's wrong***

v1.0.4

Discussion

- How might the researchers behind Delphi define what is moral or ethical? What does it mean for an "AI" or language model to behave ethically?
- Are you optimistic or pessimistic about efforts to make models more "ethical?"
- What might Gupta and Heath say about Delphi?

a warm-up

a warm-up

- On your slide, list as many “algorithms” or potential algorithms as you can.

- If you had to categorize them into different types/groups, how would you do so?

- If you are unable to attend today, please make sure to edit your slide with your answers by next Tuesday 2:30pm for credit.

- https://jamboard.google.com/d/1yk4IToly7ZfDa3GS74PZe0dnsq04Y2QOOHEoJtO7Rc4/edit?usp=sharing

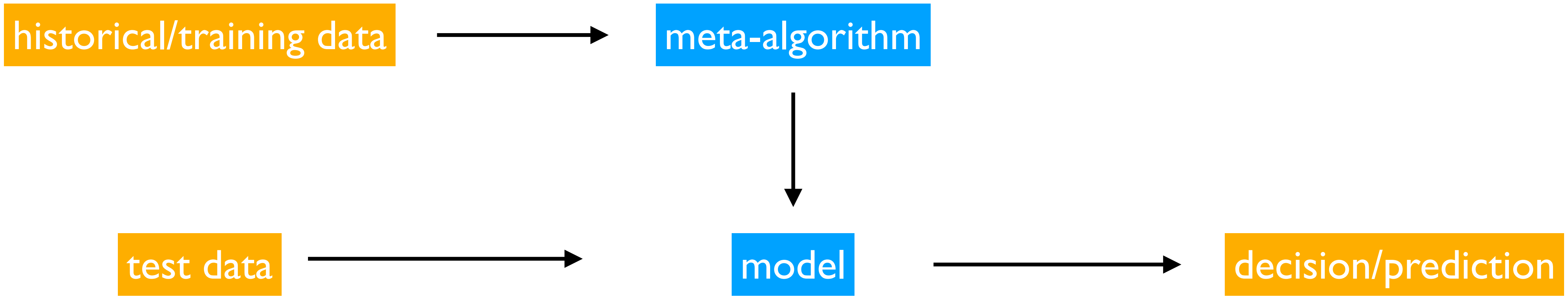
what is an algorithm?

But first, what is an algorithm anyway? At its most fundamental level, an algorithm is nothing more than a very precisely specified series of instructions for performing some concrete task.

Kearns and Roth, "The Ethical Algorithm."

More complicated algorithms—the type that we categorize as machine learning algorithms—are automatically derived from data. A human being might hand-code the process (or meta-algorithm) by which the final algorithm—sometimes called a model—is derived from the data, but she doesn't directly design the model itself.

Kearns and Roth, "The Ethical Algorithm."



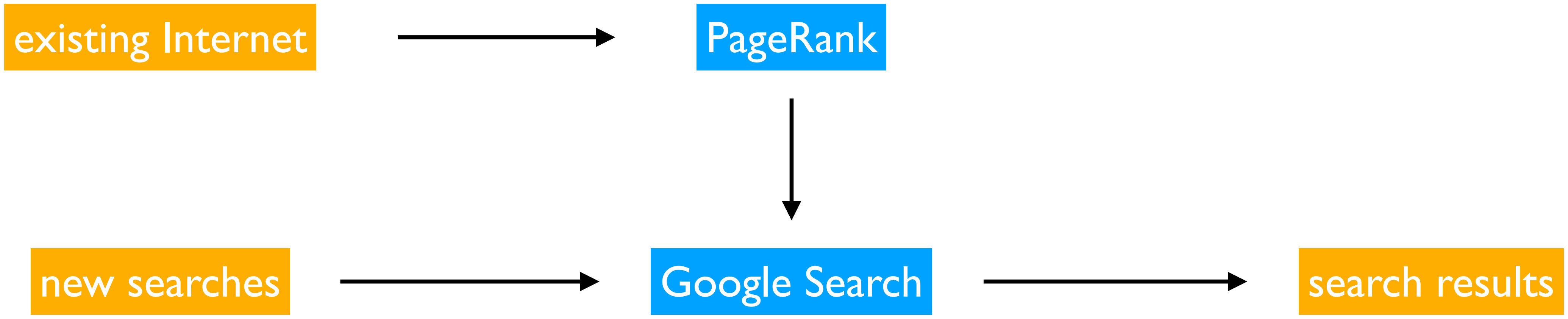
historical/training data

meta-algorithm

test data

model

decision/prediction



existing Internet

PageRank

new searches

Google Search

search results

Discussion

- If you use YouTube, it will begin to give you personalized recommendations. What is the "meta-algorithm" here and what is the "model?"

other definitions

An algorithm is a set of rules that precisely define a sequence of operations.

Harold Stone

In statistics and machine learning, we usually think of the algorithm as the set of instructions a computer executes to learn from data. In these fields, the resulting structured information is typically called a model. The information the computer learns from the data via the algorithm may look like "weights" by which to multiply each input factor, or it may be much more complicated. The complexity of the algorithm itself may also vary. And the impacts of these algorithms ultimately depend on the data to which they are applied and the context in which the resulting model is deployed. The same algorithm could have a net positive impact when applied in one context and a very different effect when applied in another.

Lum and Chowdhury

As decision-makers in both government and industry create standards for algorithmic audits, disagreements about what counts as an algorithm are likely. Rather than trying to agree on a common definition of "algorithm" or a particular universal auditing technique, we suggest evaluating automated systems primarily based on their impact. By focusing on outcome rather than input, we avoid needless debates over technical complexity. What matters is the potential for harm, regardless of whether we're discussing an algebraic formula or a deep neural network.

Lum and Chowdhury

"Automated Decision Systems" are any systems, software, or process that use computation to aid or replace government decisions, judgments, and/or policy implementation that impact opportunities, access, liberties, rights, and/or safety. Automated Decisions Systems can involve predicting, classifying, optimizing, identifying, and/or recommending.

Rashida Richardson, "Defining and Demystifying Automated Decision Systems" (Narrow Definition)

Discussion

- Do these distinctions matter? How do you think your friends and family interpret the word,“algorithm?”

the designer’s role

The result is that the complicated, automated decision-making that can arise from machine learning has a character of its own, distinct from that of its designer. The designer may have had a good understanding of the algorithm that was used to *find* the decision-making model, but not of the model itself. To make sure that the effects of these models respect the societal norms that we want to maintain, we need to learn how to design these goals directly into our algorithms.

Kearns and Roth, "The Ethical Algorithm."

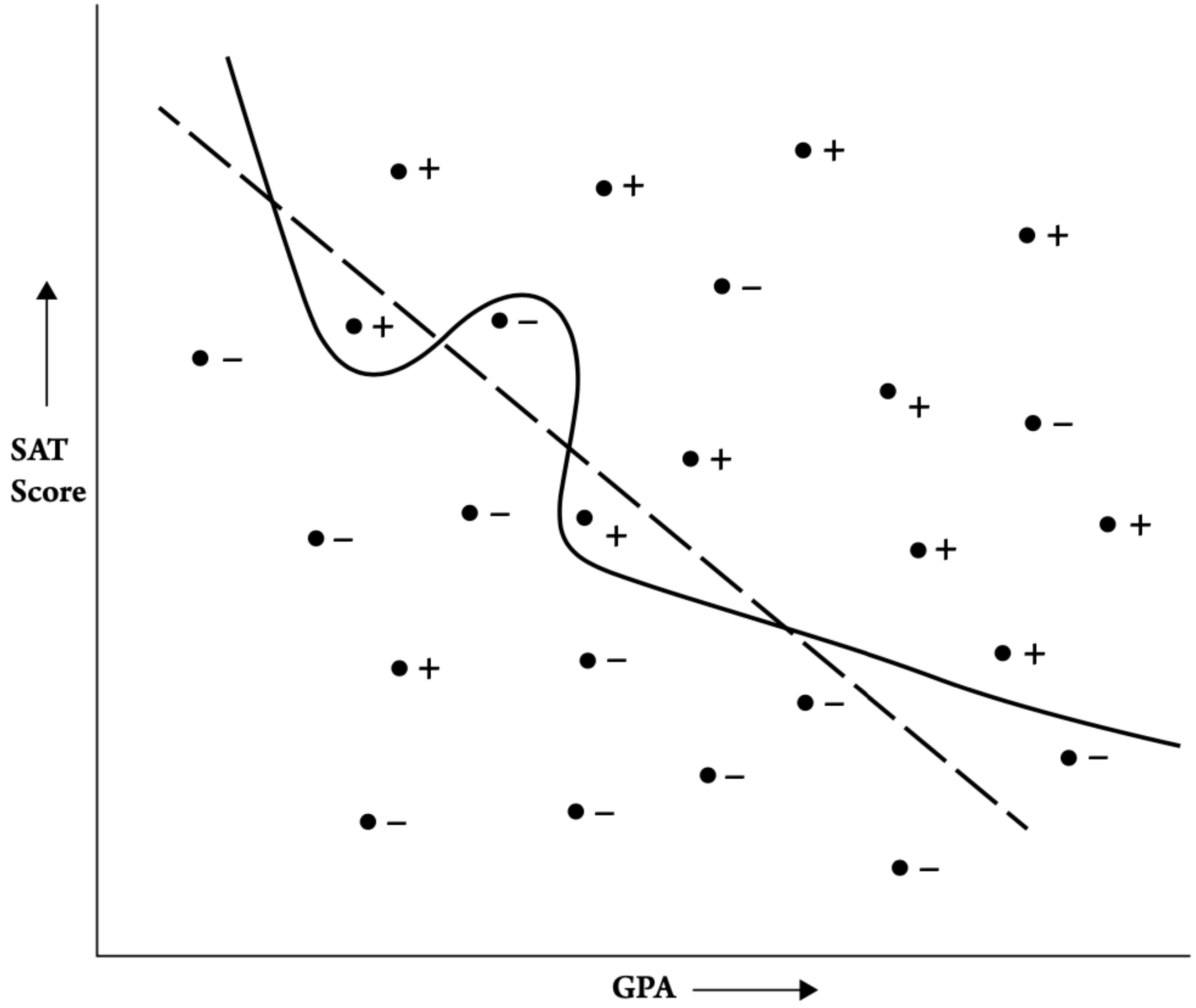


Fig. 1. Building a model to predict collegiate success from high school data. Imagine that each point represents the high school GPA and SAT score of a college student. Points labeled with "+" represent students who successfully graduated from college in four years, while points labeled with "–" represent students who did not. The straight dashed line does an imperfect but pretty good job of separating positives from negatives and could be used to predict success for future high school students. The solid curve makes even fewer errors but is more complicated, potentially leading to unintended side effects.

Kearns and Roth, "The Ethical Algorithm."

Discussion

- Kearns and Roth suggest that algorithms can be improved to address issues of bias and fairness if they are designed specifically to account for them—do you agree or disagree?
- How does this relate to our discussion of Delphi?

Discussion

- How would you explain any distinctions between "algorithms," "models," "automated decision systems," etc… for a general audience?

- Based on our discussion today, what might it mean for an algorithm to be "biased" or "unfair?"