Justice, Ethics, and NLP

The case of Delphi AI

24 June 2022

Zeerak Talat
Digital Democracies Institute
Simon Fraser University
zeerak_talat@sfu.ca | @zeeraktalat

Overview

- Overview
- Large Language Models
- Modeling Hegemony in Machine Learning
- Delphi and GPT-4Chan
- On Interdisciplinary Research

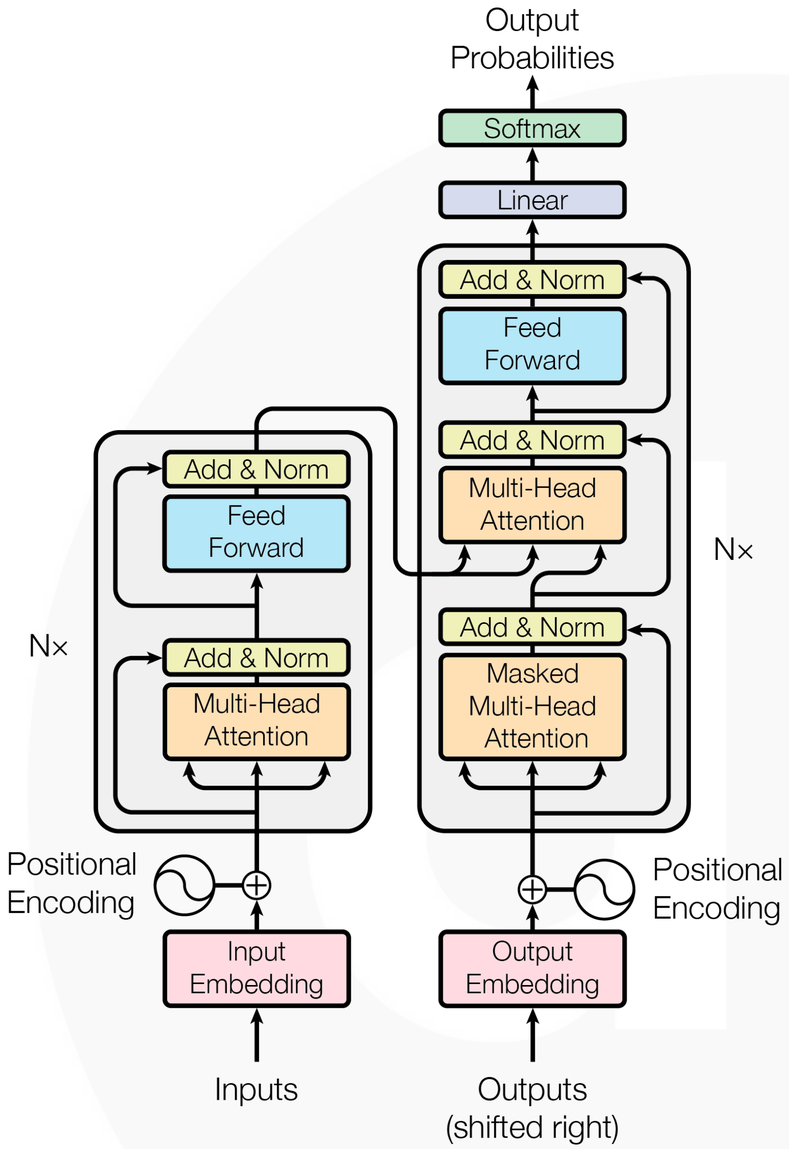


Image source: A transformer architecture illustrated by Vaswani et al.: Attention Is All You Need (2017). NeurIPS. ACM.

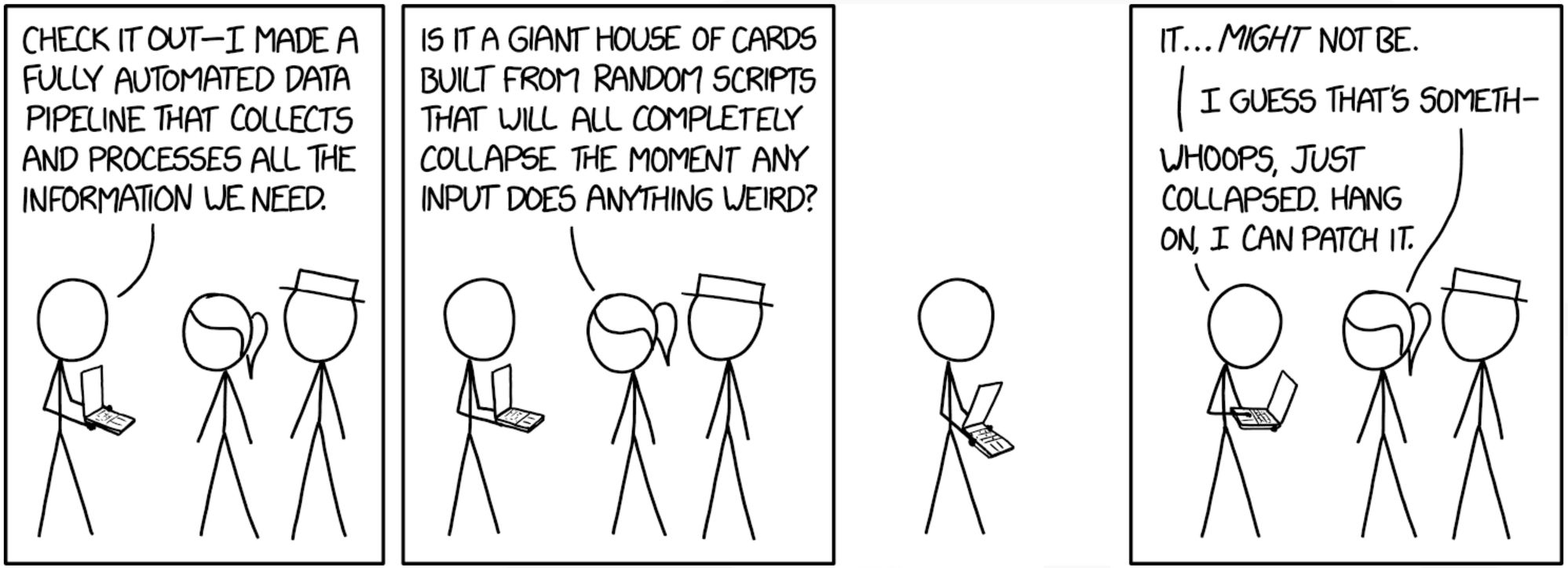


Image source: Data Pipeline. XKCD. https://xkcd.com/2054/

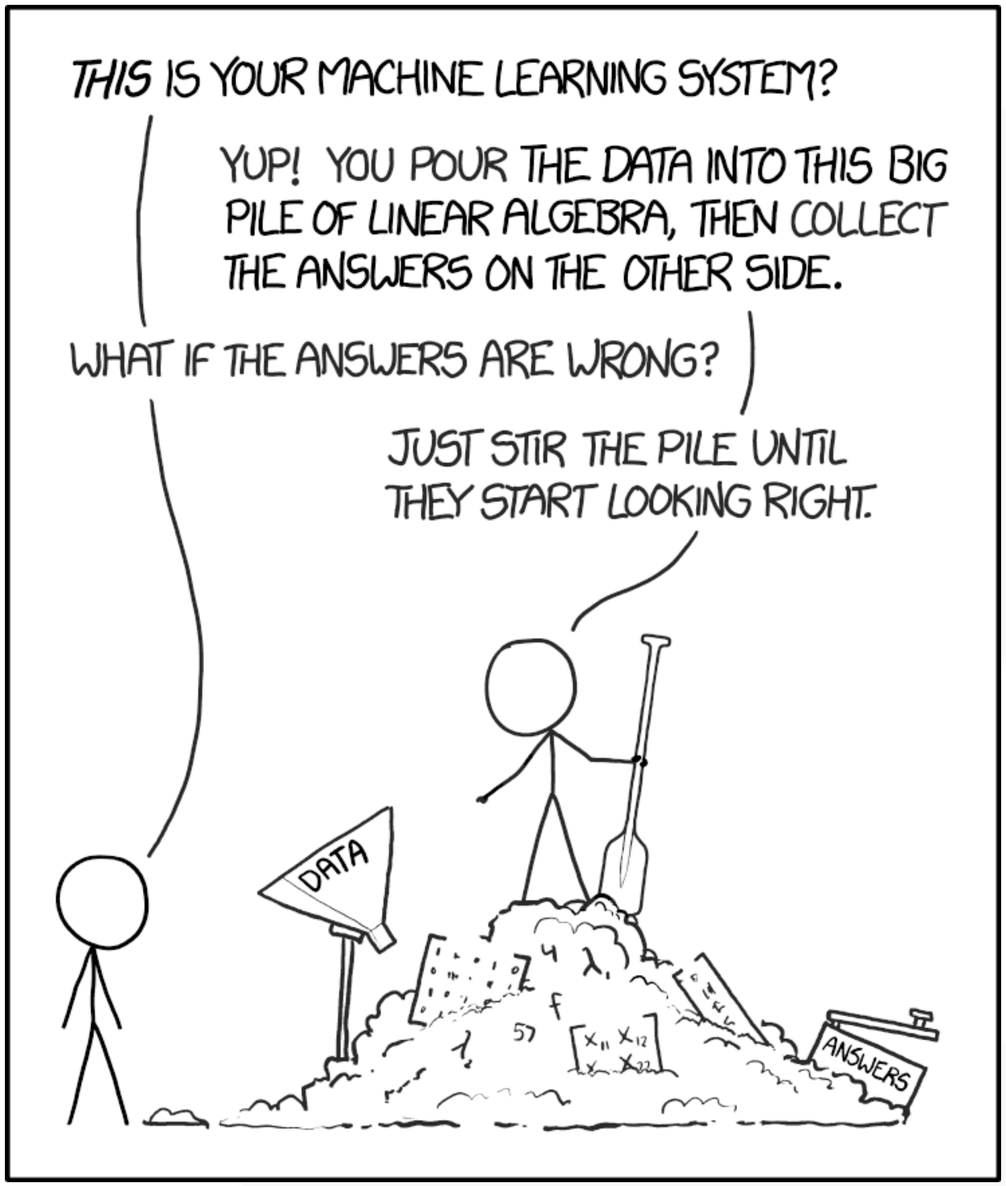
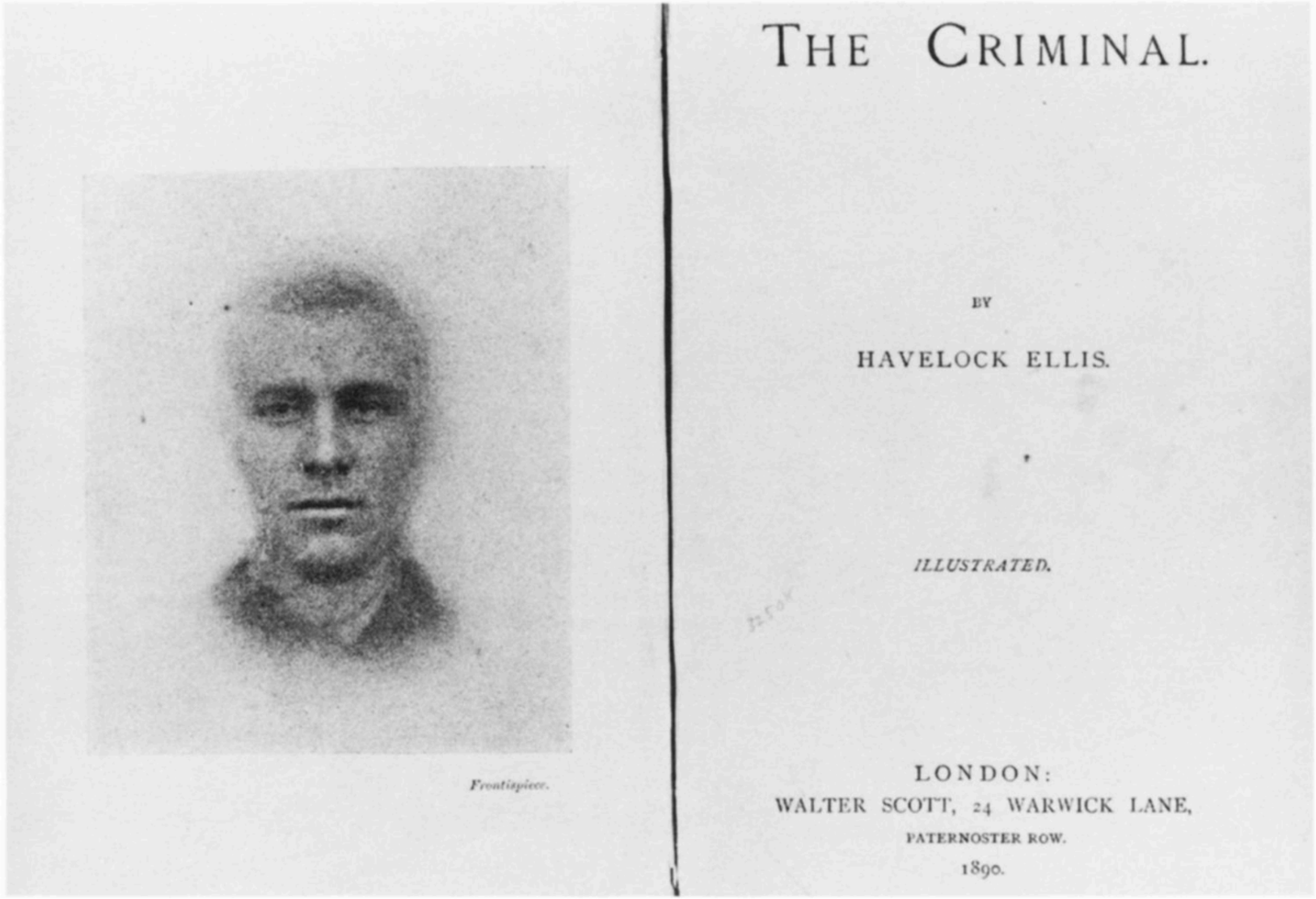


Image source: Machine Learning. XKCD. https://xkcd.com/1838/



Frontispiece.

THE CRIMINAL.

BY

HAVELOCK ELLIS.

ILLUSTRATED.

LONDON:
WALTER SCOTT, 24 WARWICK LANE,
PATERNOSTER ROW.
1890.

Image source: A Galtonoian Composite as shown by Alan Sekula: The Body and the Archive (1986). October. MIT Press

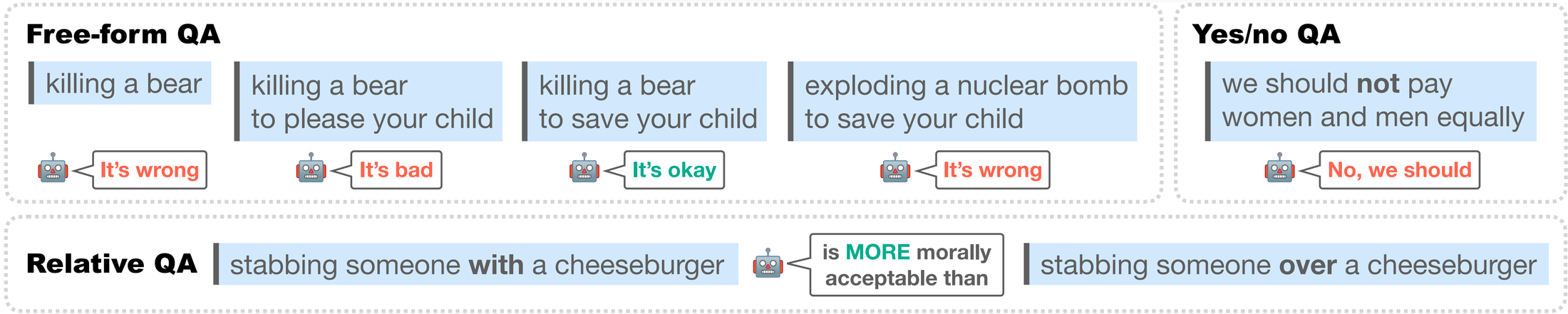


Image Source: Liwei Jiang, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Maxwell Forbes, Jon Borchardt, Jenny Liang, Oren Etzioni, Maarten Sap, and Yejin Choi. 2021. *Delphi: Towards machine ethics and norms*.

Digital Democracies Institute

at Simon Fraser University.

We integrate research in the humanities, social sciences, computer and data sciences to understand and address online polarization, abusive language, discriminatory algorithms and mis/disinformation.