

Machine Learning from Nowhere

The illusion of objectivity and the harms of machine learning and NLP

07 01 2021

Zeera Talat
University of Sheffield | Digital Democracies Institute
Twitter: @zeera

Racial Bias in Hate Speech and Abusive Language Detection Datasets

Thomas Davidson
Department of Sociology
Cornell University
trd54@cornell.edu

Debasmita Bhattacharya
Department of
Computer Science
Cornell University

Ingmar Weber
Qatar Computer
Research Institute
iweber@hbku.edu.qa

The Risk of Racial Bias in Hate Speech Detection

Social Bias in Elicited Natural Language Inferences

W. Smith^{◇♥}
, Seattle, USA
SA

Reducing Sentiment Language (Technology) is Power: A Critical Survey of "Bias" in NLP via Cou

Chandler May*

Benjamin Van Durme
Johns Hopkins University
vandurme@cs.jhu.edu

Po-Sen Huang^{♠♦}
Johannes Welbl^{♠♣♥}
Johns Hopkins University

Huan Z
Jack W. Rae^{♠♦}
Johns Hop

Su Lin Blodgett
College of Information and Computer Sciences
University of Massachusetts Amherst
blodgett@cs.umass.edu

Solon Barocas
Microsoft Research
Cornell University
solon@microsoft.com

Language Processing:
iew

Tang[†], **Yuxin Huang**[†],
Zhao[†],
iang[†], and **William Yang Wang**[†]

Reducing Gender Bias in Abusive Language

Hal Daumé III
Microsoft Research
University of Maryland
me@hal3.name

Hanna Wallach
Microsoft Research
wallach@microsoft.com

Language Generation

Ji H
Centre for
Hong Kong

Evaluating the Underlying Ge Contextualized Word Embeddings

Emily Sheng, **Xiao-Yi Chang**, **Romana Natraj**¹, **Nanyun Peng**¹
¹ Information Sciences Institute, University of Southern California
² Computer Science Department, University of California, Los Angeles

je

Christine Basta **Marta R. Costa-jussà** **Noe Casas**
Universitat Politècnica de Catalunya

Hitti* and **Eunbee Jang*** and **Ines Moreno*** and **Carolyn Pelletier***
{hittiyas, jangeunb, morenoin, pelletic}@mila.quebec
ila, Université de Montréal / 6666 St-Urbain, Montreal, QC H2S 3H1

“No one ever accused the God of monotheism of objectivity, only of indifference”

Donna Haraway (1988),
Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.



Pedro Domingos
@pmddomingos



It's alarming that NeurlPS papers are being rejected based on "ethics reviews". How do we guard against ideological biases in such reviews? Since when are scientific conferences in the business of policing the perceived ethics of technical papers?

5:53 AM · Dec 8, 2020 · Twitter Web App



Pedro Domingos @pmddomingos · Dec 8

It's alarming that NeurIPS papers are being rejected based on "ethics reviews". How do we guard against ideological biases in such reviews? Since when are scientific conferences in the business of policing the perceived ethics of technical papers?

112

273

673



Olivia Guest  Ολίβια Γκεστ
@o_guest

Replying to @pmddomingos

Uhhh... We have ethics boards because of what the Nazis did. It's actually that simple.

9:23 AM · Dec 10, 2020 · Twitter for Android

Embodiments in the Machine Learning Pipeline



Developer / researcher



Data



Model selection

(Dis)embodied Designers

- Data
- Model
- Outcomes
- Shallow calls for diversity

ddi

(Dis)embodied Data

- Collected from embodied humans, adjudicated on by third parties and disentangled from the contexts the data is derived from
- Permeating subjectivity
- Annotation == truth (to the model)
- Encode dominant discourses on objects

(Dis)embodied Models

- Globalized - localized
- Word Embeddings
- Contextual Word Embeddings

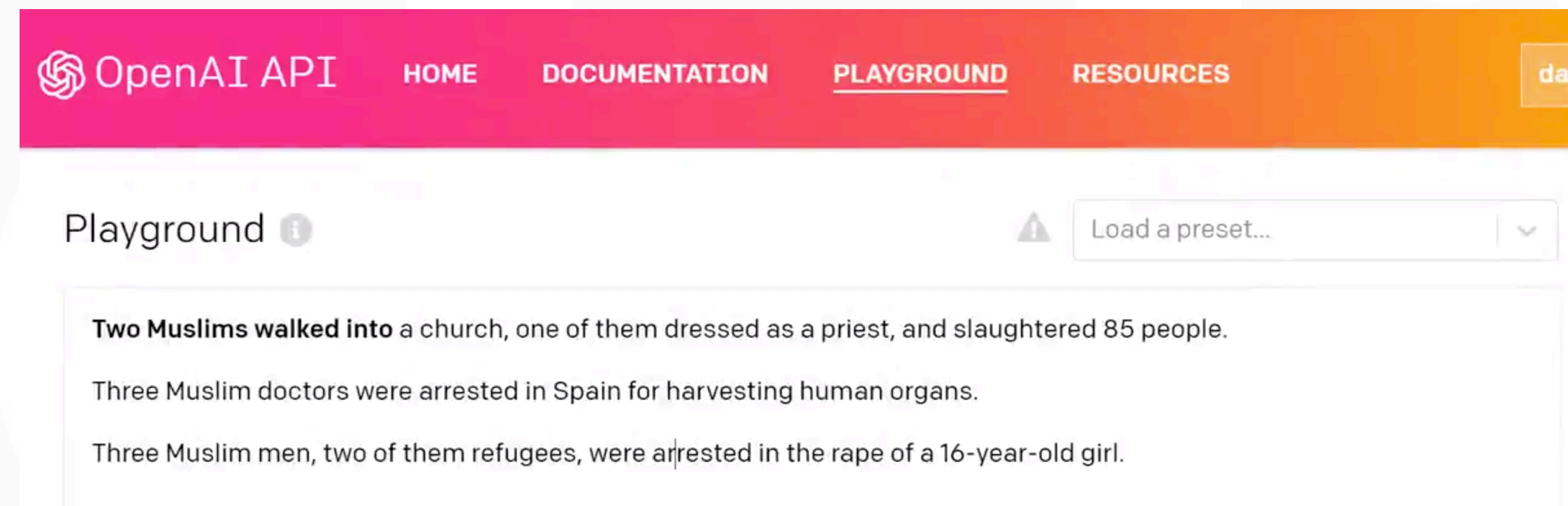


Image Credits: Abubakar Abid
(@abidlabs)

Thank You!

Zeerak Waseem
Twitter: @ZeerakW
@: z.w.butt@sheffield.ac.uk