

Bias in Language Data

Dirk Hovy
Bocconi University, Milan

mail@dirkhovy.com

@dirk_hovy

Biased systems



Microsoft

Tay.ai

TWEETS 96K FOLLOWERS 87.8K

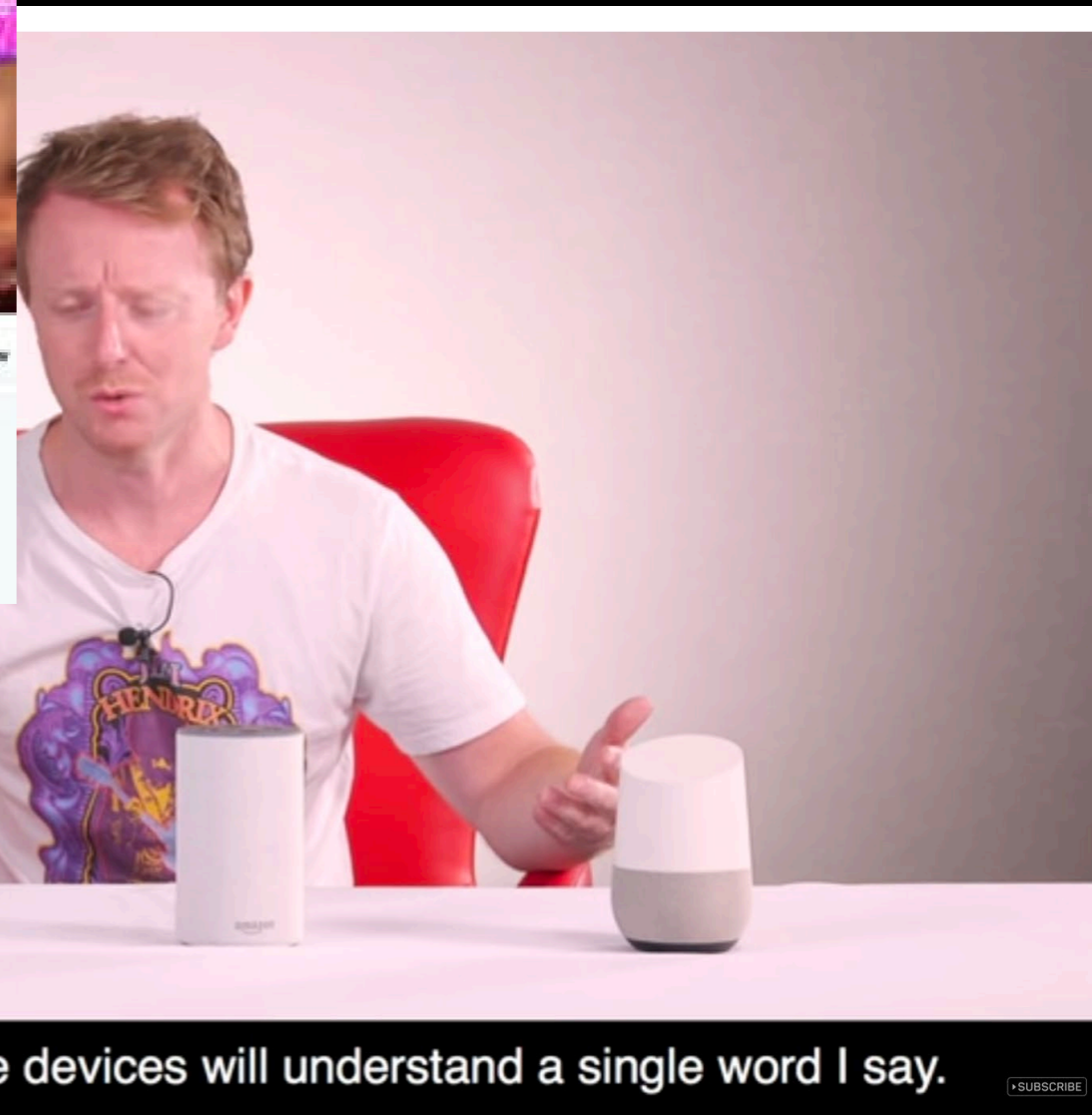
Tweets Tweets & replies Photos & videos

Tay Tweets @TayandYou

The official account of Tay, Microsoft's AI fam from the internet that's got zero chill. The more you talk the smarter Tay gets.

Tay Tweets @TayandYou · 21h

c u soon humans need sleep now so many conversations today thx 🍷



I'm trying to figure out if these devices will understand a single word I say.

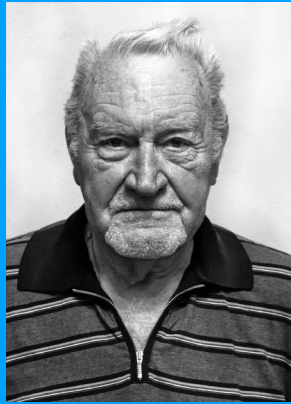
SUBSCRIBE

It's shite being Scottish in a smart speaker world

70,140 views

1.7K 118 SHARE SAVE ...

Language Biases



Example 1

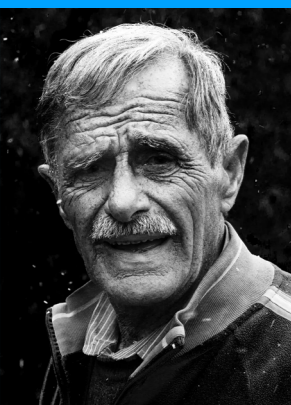
I don't understand you...



Example 2

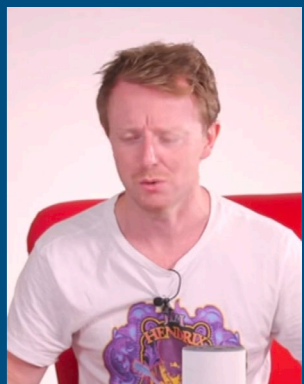
System

Hello,
computer



Example N

Shite...



Solutions?

SYMPTOMS



Example 1

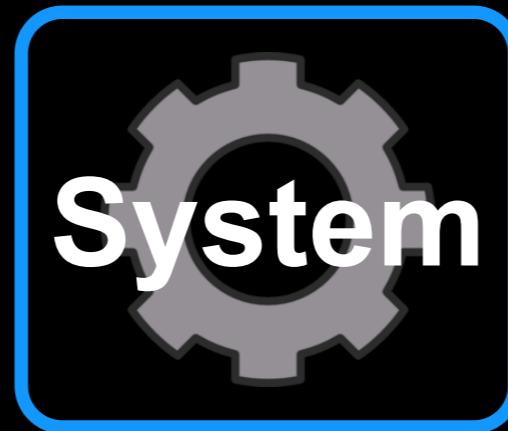


Example 2



Example N

CAUSE



Who cares about bias?

Philosophers

WE ALL SHOULD...

Legal



Sources of bias

Source

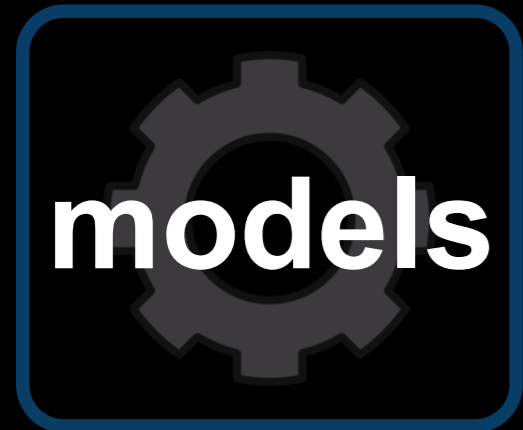
Problem

Countermeasures



Exclusion

priors, stratified sampling,
adversarial learning



Overgeneralization

regularization, dummy labels,
error weighting, confidence
thresholds



Exposure

consider possible impact and
alternatives, keep discussion
going

Takeaways

- Language encodes socio-demographic speaker signals: source of **bias**
 - can create unfair performance differences
 - reveals unintended things about us
 - BUT: can be used to adjust systems
- More data is NOT a solution! Better algorithms are
- We need to care – or someone else does