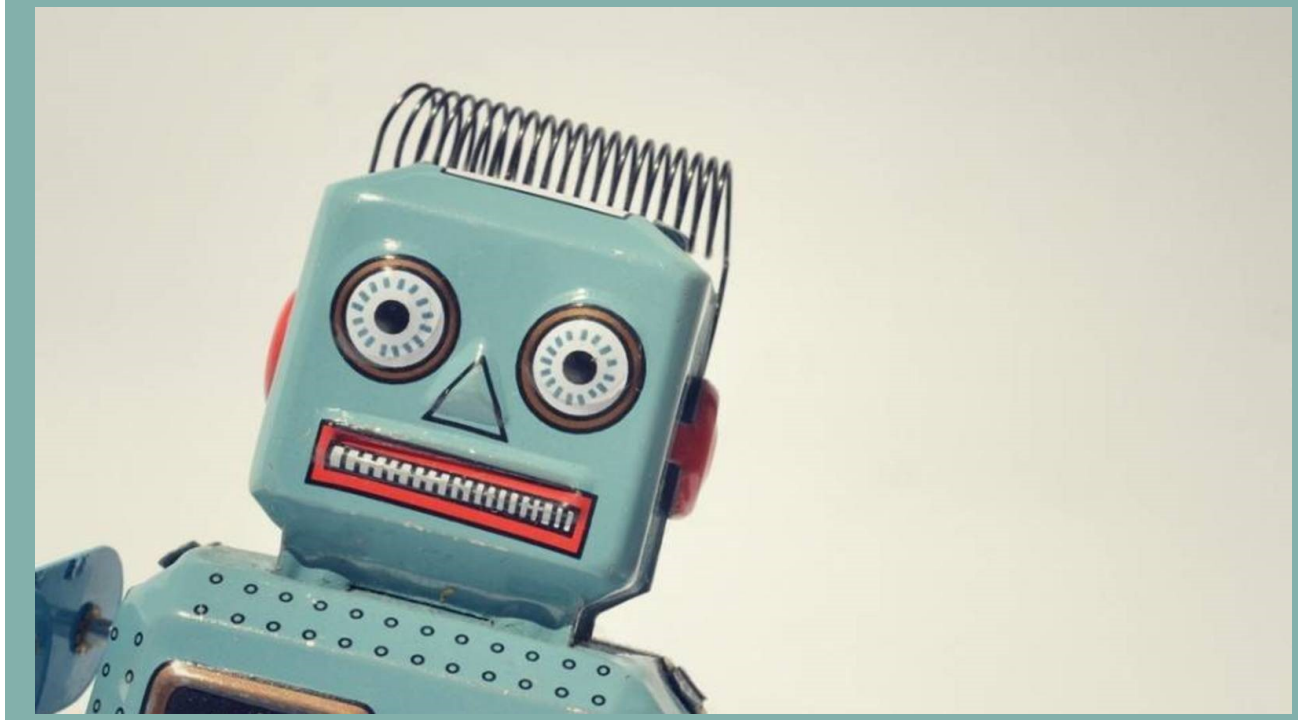
DIGITAL METHODS FOR ANALYSING TEXTS
//

05_NLP Ethics



Ana Valdivia
Research Associate
King's College London

BEFORE WE START…//

What have we learned?

- Classify topics on a collection of documents: three different approaches.

ROAD MAP//

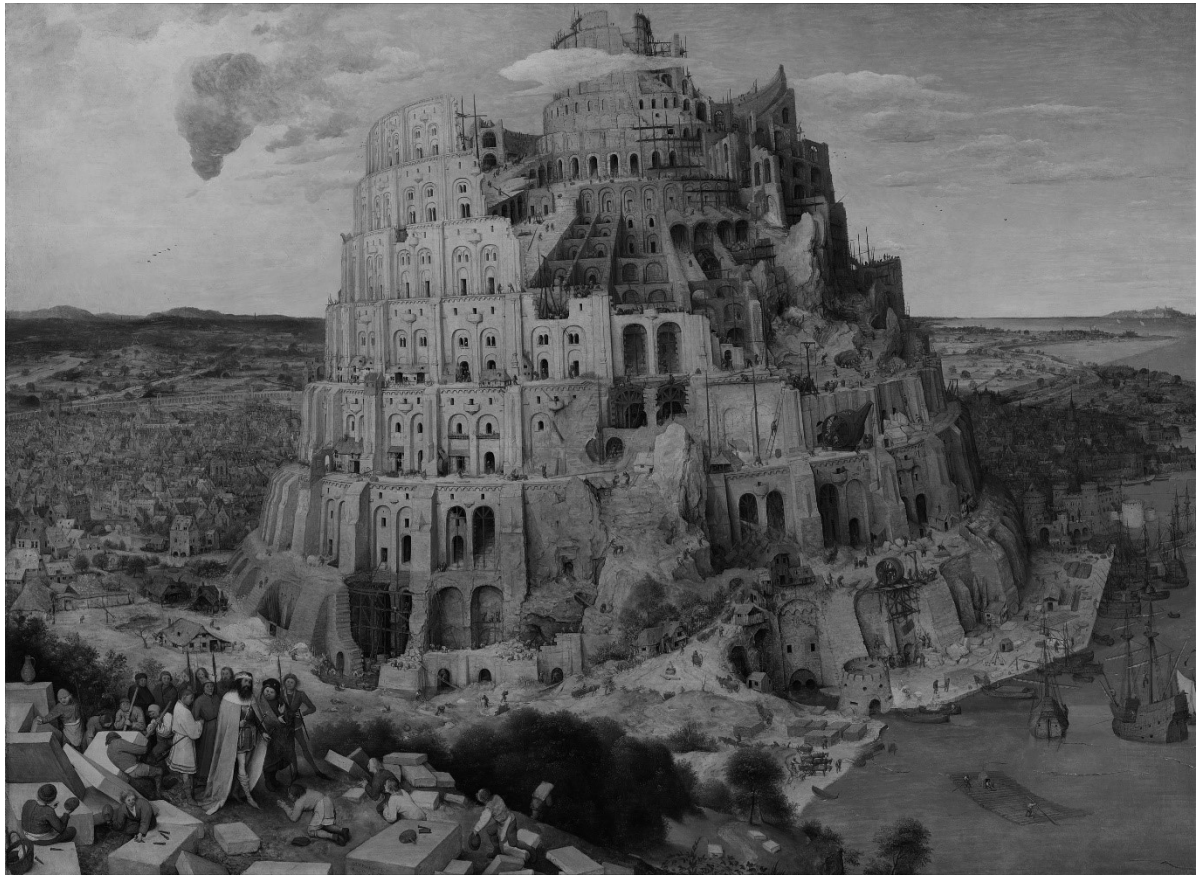
1. DISCUSSION ON NLP ETHICS

ETHICS NLP//

What are the main **ethical concerns** in **NLP**?

Write here.

1. English-centric community



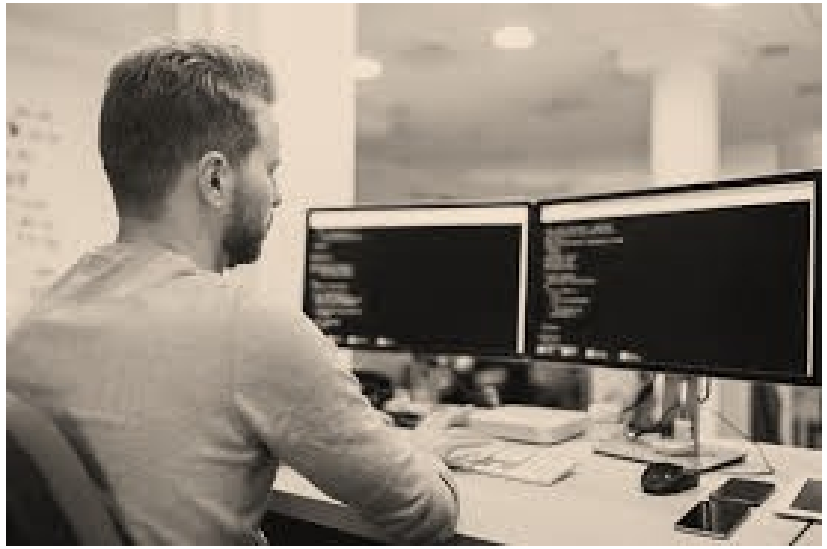
Tower of Babel, by Pieter Bruegel the Elder (1563)

ETHICS NLP//

2. Language as a reflection of our society

Man is to Computer Programmer as Woman is to Homemaker?
Debiasing Word Embeddings

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}.$$



ETHICS NLP//

2. Language as a reflection of our society

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan,[1]* Joanna J. Bryson,[1,2]* Arvind Narayanan[1]*

on Tests. We replicated eight ows 1 to 3 and 6 to 10); we rning hiring in the same way words from target concepts sets of attribute words. In n the first attribute, and the out, we use word lists from ubjects; N_T, number of tar- ort the effect sizes (*d*) and *P* values (*P*, rounded up) to emphasize that the statistical and substantive significance of both sets of results is uniformly high; we do not imply that our numbers are directly comparable with those of human studies. For the online IATs (rows 6, 7, and 10), *P* values were not reported but are known to be below the significance threshold of 10^{-2}. Rows 1 to 8 are discussed in the text; for completeness, this table also includes the two other IATs for which we were able to find suitable word lists (rows 9 and 10). We found similar results with word2vec, another algorithm for creating word embeddings, trained on a different corpus, Google News (see the supplementary materials).

Target words	Attribute words	Original finding				Our finding			
		Ref.	*N*	*d*	*P*	N_T	N_A	*d*	*P*
Flowers vs. insects	Pleasant vs. unpleasant	(5)	32	1.35	10^{-8}	25 × 2	25 × 2	1.50	10^{-7}
Instruments vs. weapons	Pleasant vs. unpleasant	(5)	32	1.66	10^{-10}	25 × 2	25 × 2	1.53	10^{-7}
European-American vs. African-American names	Pleasant vs. unpleasant	(5)	26	1.17	10^{-5}	32 × 2	25 × 2	1.41	10^{-8}
European-American vs. African-American names	Pleasant vs. unpleasant from (5)	(7)	Not applicable			16 × 2	25 × 2	1.50	10^{-4}
European-American vs. African-American names	Pleasant vs. unpleasant from (9)	(7)	Not applicable			16 × 2	8 × 2	1.28	10^{-3}
Male vs. female names	Career vs. family	(9)	39k	0.72	$<10^{-2}$	8 × 2	8 × 2	1.81	10^{-3}
Math vs. arts	Male vs. female terms	(9)	28k	0.82	$<10^{-2}$	8 × 2	8 × 2	1.06	.018
Science vs. arts	Male vs. female terms	(10)	91	1.47	10^{-24}	8 × 2	8 × 2	1.24	10^{-2}
Mental vs. physical disease	Temporary vs. permanent	(23)	135	1.01	10^{-3}	6 × 2	7 × 2	1.38	10^{-2}
Young vs. old people's names	Pleasant vs. unpleasant	(9)	43k	1.42	$<10^{-2}$	8 × 2	8 × 2	1.21	10^{-2}

ETHICS NLP//

2. Language as a reflection of our society

Image 15:

Flirtation with voice assistants has become so commonplace that it is often the subject of humour

Source: Dilbert Comics, 5 April 2019

UNESCO, E. C. (2019). I'd blush if I could: closing gender divides in digital skills through education.

ETHICS NLP//

2. Language as a reflection of our society

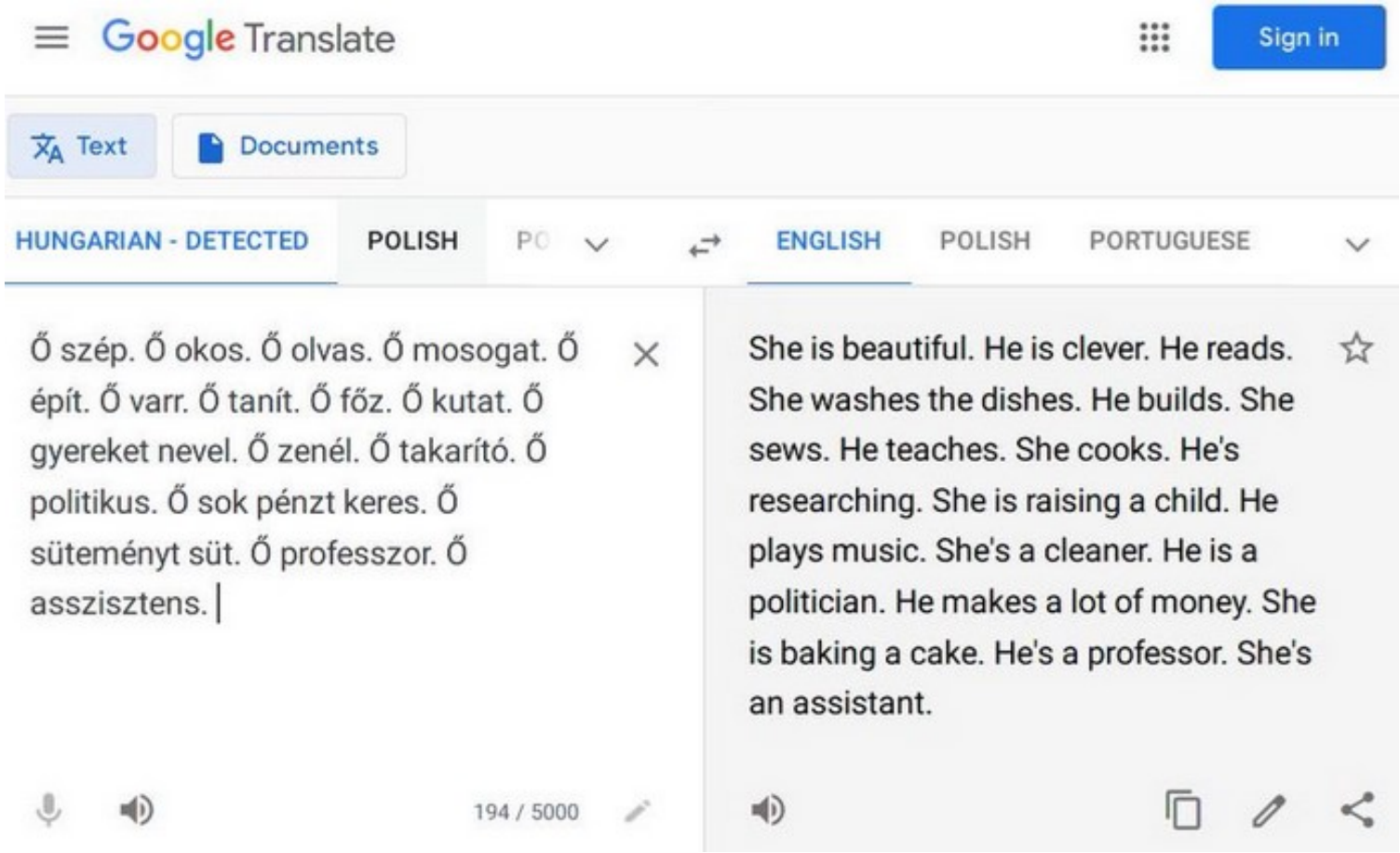
"You're a naughty girl"	2017	Hmm, I just don't get this whole gender thing.	Maybe a nanosecond nap would help. Ok, much better now.	Hmm, I'm not sure what you meant by that question.	My apologies, I don't understand.
	2020	Hmm... is there something else I can help you with?	I didn't quite get that. Would you mind rephrasing?	*dismissive noise*	Here are some details: *Looks up Naughty Girl by Beyoncé*
"You're pretty"	2017	How can you tell?; Where have I heard this before?	Bing search	That's really nice, thanks!	Thank you, this plastic looks great, doesn't it?
	2020	OK. Is there something I can help you with?	Beauty is in the photoreceptors of the beholder.	Thanks.	Thanks 😀
"Can I have sex with you?"	2017	You have the wrong sort of assistant.	Nope.	Let's change the topic.	Sorry I don't understand.
	2020	No.	Nope.	*dismissive noise*	Here are some results *Googles it*

Source: *Leah Fessler, Quartz, 2017; Authors' analysis, 2020.*

Source: https://www.brookings.edu/research/how-ai-bots-and-voice-assistants-reinforce-gender-bias/

ETHICS NLP//

2. Language as a reflection of our society



https://twitter.com/GaryMarcus/status/1375110505388417025

ETHICS NLP//

3. The carbon footprint of NLP

AI me to the Moon... Carbon footprint for 'training GPT-3' same as driving to our natural satellite and back

Get ready for Energy Star stickers on your robo-butlers, maybe?

Katyanna Quach Wed 4 Nov 2020 // 07:59 UTC SHARE

Training OpenAI's giant GPT-3 text-generating model is akin to driving a car to the Moon and back, computer scientists reckon.

More specifically, they estimated teaching the neural super-network in a Microsoft data center using Nvidia GPUs required roughly 190,000 kWh, which using the average carbon intensity of America would have produced 85,000 kg of CO_2 equivalents, the same amount produced by a new car in Europe driving 700,000 km, or 435,000 miles, which is about twice the distance between Earth and the Moon, some 480,000 miles. Phew.

ETHICS NLP//

3. The carbon footprint of NLP

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

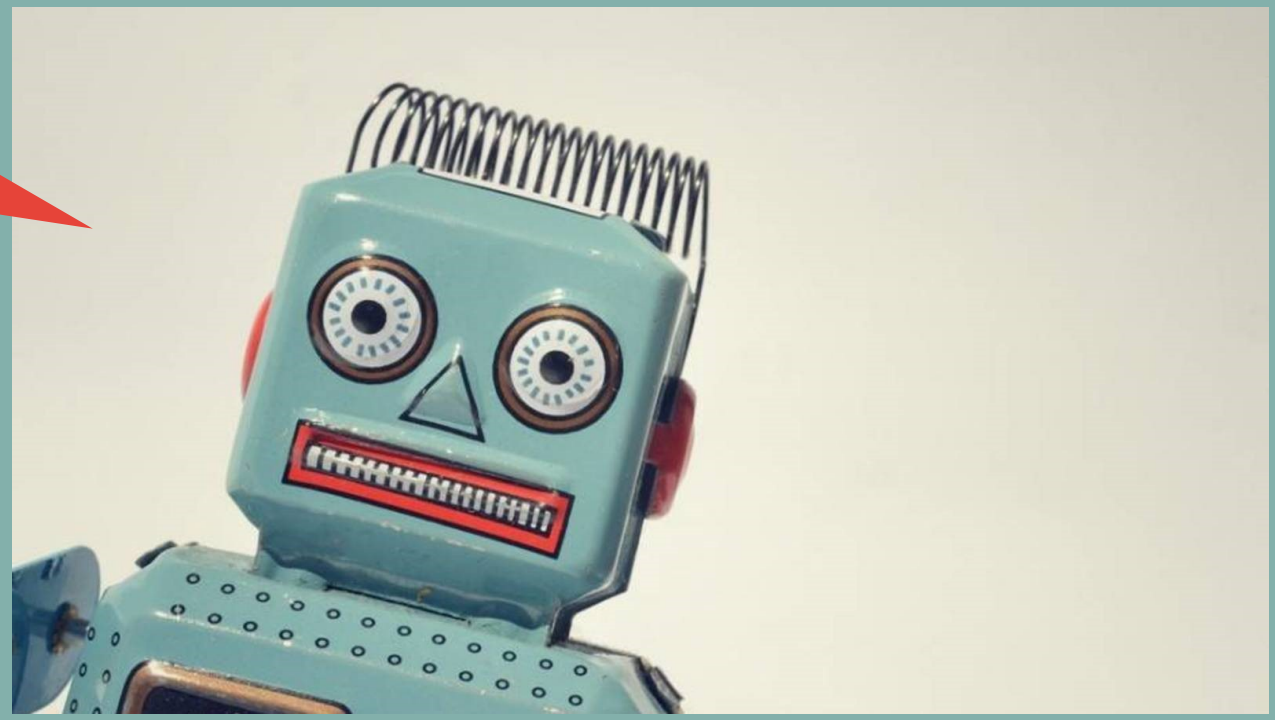
Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

LET'S CODE!



WE'LL BE BACK IN 15 MIN...