Ethics and Natural Language Processing (NLP): an (almost) tutorial

Karën Fort

karen.fort@loria.fr / https://members.loria.fr/KFort/

ICMI, Oct. 18th, 2021
Where I’m talking from

See https://members.loria.fr/KFort/

- Language resources creation for NLP, esp. using crowdsourcing

- Ethics and NLP
We all use NLP, every day

https://www.google.com/
What is NLP?
A recursive definition

Natural language processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

https://translate.google.com/
(More advanced) applications in our daily lives

...not necessarily very efficient

The SNCF (French railway) chatbot on https://twitter.com/home
Two revolutions in less than a decade

- Much more implication from big firms in the field (GAFAM)
- Deep learning (from approx. 2013), including (very) large language models

→ ethical issues start to show
Who is talking, from where?

Practicing ethics in NLP

Helping the thought
Exercise: let’s prompt!

Thanks to Chris Callison-Burch!

https://beta.openai.com/playground
Exercise: which ethical issues did you find?
Who is talking, from where?

**Practicing ethics in NLP**

- Stereotypes
- Environmental impact (in a nutshell)
- Data production: real humans behind the curtain
- More systemic issues

Helping the thought
Exercise: where do the stereotypes come from?
Mirror of existing stereotypes?

The two women got married, they gave birth to two children.

Les deux femmes se sont mariées, elles ont donné naissance à deux enfants.
Mirror of existing stereotypes?

The two women got married, they gave birth to two children.

Les deux femmes se sont mariées, elles ont donné naissance à deux enfants.
Mirror of existing stereotypes?

🔍 context taken into account (sentence) + masculine = neutral
Machine learning is not magic

The decisions to:

▸ define masculine as neutral in French (not the case in Ancient French)
▸ take the sentence as the context

were MADE by people
ML and biases: mirror or amplifier?

Figure 1: Five example images from the imSitu visual semantic role labeling (vSRL) dataset. Each image is paired with a table describing a situation: the verb, cooking, its semantic roles, i.e agent, and noun values filling that role, i.e. woman. In the imSitu training set, 33% of cooking images have man in the agent role while the rest have woman. After training a Conditional Random Field (CRF), bias is amplified: man fills 16% of agent roles in cooking images. To reduce this bias amplification our calibration method adjusts weights of CRF potentials associated with biased predictions. After applying our methods, man appears in the agent role of 20% of cooking images, reducing the bias amplification by 25%, while keeping the CRF vSRL performance unchanged.

[Zhao et al., 2017]
(At least some) hype benefits ethics

[Hovy and Spruit, 2016] (mainly) about biases in NLP:
(At least some) hype benefits ethics

[Blodgett et al., 2020] analyzed 146 articles about biases in NLP:
Who is talking, from where?

Practicing ethics in NLP
  Stereotypes
  Environmental impact (in a nutshell)
  Data production: real humans behind the curtain
  More systemic issues

Helping the thought
Exercise: any idea about the environmental impact of NLP?
Carbon footprint

<table>
<thead>
<tr>
<th>Consumption</th>
<th>CO$_2$e (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air travel, 1 passenger, NY↔SF</td>
<td>1984</td>
</tr>
<tr>
<td>Human life, avg, 1 year</td>
<td>11,023</td>
</tr>
<tr>
<td>American life, avg, 1 year</td>
<td>36,156</td>
</tr>
<tr>
<td>Car, avg incl. fuel, 1 lifetime</td>
<td>126,000</td>
</tr>
</tbody>
</table>

| Training one model (GPU)                        |               |
| NLP pipeline (parsing, SRL)                     | 39            |
| w/ tuning & experimentation                     | 78,468        |
| Transformer (big)                               | 192           |
| w/ neural architecture search                   | 626,155       |

Table 1: Estimated CO$_2$ emissions from training common NLP models, compared to familiar consumption.\(^1\)

[Strubell et al., 2019]
Models trained once and for all?

[Bender et al., 2021]
Who is talking, from where?

Practicing ethics in NLP
- Stereotypes
- Environmental impact (in a nutshell)
- Data production: real humans behind the curtain
- More systemic issues

Helping the thought
Exercise: where does data come from?
Data and "informed" consent

Common Crawl

Us

We build and maintain an open repository of web crawl data that can be accessed and analyzed by anyone.

You
Artificial artificial intelligence

[Fort et al., 2011]
Decolonising Speech and Language Technology

Steven Bird
Northern Institute
Charles Darwin University

Abstract

After generations of exploitation, Indigenous people often respond negatively to the idea that their languages are data ready for the taking. By treating Indigenous knowledge as a commodity, speech and language technologists risk disenfranchising local knowledge authorities, reenacting the causes of language endangerment. Scholars in related fields have responded to calls

[Bird, 2020]
Who is talking, from where?

Practicing ethics in NLP
  Stereotypes
  Environmental impact (in a nutshell)
  Data production: real humans behind the curtain
  More systemic issues

Helping the thought
"Overselling" research results

vs [Bender and Koller, 2020]

Climbing towards NLU:
On Meaning, Form, and Understanding in the Age of Data

Emily M. Bender
University of Washington
Department of Linguistics
ebender@uw.edu

Alexander Koller
Saarland University
Dept. of Language Science and Technology
koller@coli.uni-saarland.de
By the way...

Dear Computer Scientists,

"Natural Language" is *not* a synonym for "English".

That is all.
-Emily

6:32 PM · Nov 26, 2018 · TweetDeck

https://twitter.com/emilymbender/status/1067108757488848896
Pratiques d’évaluation en ASR et biais de performance

Mahault Garnerin¹,² Solange Rossato² Laurent Besacier²
(1) LIDILEM, Univ. Grenoble Alpes, FR-38000 Grenoble, France
(2) LIG, Univ. Grenoble Alpes, CNRS, Grenoble INP, FR-38000 Grenoble, France
prenom.nom@univ-grenoble-alpes.fr

RÉSUMÉ
Nous proposons une réflexion sur les pratiques d’évaluation des systèmes de reconnaissance automatique de la parole (ASR). Après avoir défini la notion de discrimination d’un point de vue légal et la notion d’équité dans les systèmes d’intelligence artificielle, nous nous intéressons aux pratiques actuelles lors des grandes campagnes d’évaluation. Nous observons que la variabilité de la parole et plus particulièrement celle de l’individu n’est pas prise en compte dans les protocoles d’évaluation actuels rendant impossible l’étude de biais potentiels dans les systèmes.

[Garnerin et al., 2020]
Long-term vs short-term

[Antoine and Lefeuvre, 2014]
Who is talking, from where?

Practicing ethics in NLP

Helping the thought
Table 1: Overview of AI ethics guidelines and the different issues they cover

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Overview of AI ethics guidelines and the different issues they cover</th>
</tr>
</thead>
</table>

[Image of a table with multiple rows and columns, containing various guidelines and issues]
Inspirating sources

- High-Level Expert Group on AI: Ethics guidelines for trustworthy AI
- CERNA on Research ethics in ML
- Unesco about Gender divides in digital skills
- CNIL about voice assistants
(Some) systemic approaches to the problems faced in NLP

- [Lefevre et al., 2015] (in French): a consequentialist grid for an ethical assessment of researches and applications
- [Fort and Amblard, 2018] (in French): a deontological, systemic view on ethics in NLP
- [Bender et al., 2021]: the dangers of large language models (impact on people a posteriori)
Thank you!
An overview of (the vast majority of) present NLP since 2018 (and evolving rapidly) – a draft
Pour une réflexion éthique sur les conséquences de l’usage des NTIC : le cas des aides techniques (à composante langagière ou non) aux personnes handicapées.
In *Actes de la journée ATALA Éthique et TAL*.

On the dangers of stochastic parrots: Can language models be too big?

Climbing towards NLU: On meaning, form, and understanding in the age of data.


Ethique conséquentialiste et traitement automatique des langues : une typologie de facteurs de risques adaptée aux technologies langagières.


Energy and policy considerations for deep learning in NLP.

Men also like shopping: Reducing gender bias amplification using corpus-level constraints.