Ethics in NLP

Karën Fort

karen.fort@loria.fr / https://members.loria.fr/KFort
Toward a systemic approach
   Advertizing vs publishing
   Artificial artificial intelligence
   Environmental impact (in a nutshell)
   Conflicts of interests

"All your data are belong to us"

What about guidelines?

To finish
Very few systemic approaches to the problem

- [Lefeuvre 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)
Toward a systemic approach

Advertizing vs publishing

Artificial artificial intelligence

Environmental impact (in a nutshell)

Conflicts of interests

“All your data are belong to us”

What about guidelines?

To finish
"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
Toward a systemic approach

- Advertizing vs publishing
- Artificial artificial intelligence
- Environmental impact (in a nutshell)
- Conflicts of interests

"All your data are belong to us"

What about guidelines?

To finish
Data production: real humans behind the curtain

[Fort et al., 2011]
Data and "informed" consent
Toward a systemic approach

- Advertizing vs publishing
- Artificial artificial intelligence
- Environmental impact (in a nutshell)
- Conflicts of interests

"All your data are belong to us"

What about guidelines?

To finish
Carbon footprint

<table>
<thead>
<tr>
<th>Consumption</th>
<th>CO₂e (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>

<table>
<thead>
<tr>
<th>Training one model (GPU)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP pipeline (parsing, SRL)</td>
<td>39</td>
</tr>
<tr>
<td>w/ tuning &amp; experimentation</td>
<td>78,468</td>
</tr>
<tr>
<td>Transformer (big)</td>
<td>192</td>
</tr>
<tr>
<td>w/ neural architecture search</td>
<td>626,155</td>
</tr>
</tbody>
</table>

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

[Strubell et al., 2019]

Note: this concerns only 1 source out of four [Bannour et al., 2021] ⇒ largely under-estimated
Models trained once and for all?
from a presentation of [Bender et al., 2021]

[Bender et al., 2021]
Water consumption

Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models

Pengfei Li, Jiayi Yang, Mohammad A. Islam, Shaolei Ren

The growing carbon footprint of artificial intelligence (AI) models, especially large ones such as GPT-3 and GPT-4, has been undergoing public scrutiny. Unfortunately, however, the equally important and enormous water footprint of AI models has remained under the radar. For example, training GPT-3 in Microsoft’s state-of-the-art U.S. data centers can directly consume 700,000 liters of clean freshwater (enough for producing 370 BMW cars or 320 Tesla electric vehicles) and the water consumption would have been tripled if training were done in Microsoft’s Asian data centers, but such information has been kept as a secret. This is extremely concerning, as freshwater scarcity has become one of the most pressing challenges shared by all of us in the wake of the rapidly growing population, depleting water resources, and aging water infrastructures. To respond to the global water challenges, AI models can, and also should, take social responsibility and lead by example by addressing their own water footprint. In this paper, we provide a principled methodology to estimate fine-grained water footprint of AI models, and also discuss the unique spatial-temporal diversities of AI models’ runtime water efficiency. Finally, we highlight the necessity of holistically addressing water footprint along with carbon footprint to enable truly sustainable AI.
Toward a systemic approach

- Advertizing vs publishing
- Artificial artificial intelligence
- Environmental impact (in a nutshell)
- Conflicts of interests

"All your data are belong to us"

What about guidelines?

To finish
BigTech’s presence in NLP [Abdalla et al., 2023]
Toward a systemic approach

"All your data are belong to us"
Data in NLP
Definition
What Happens to Data?
Back to Consent

What about guidelines?

To finish
Toward a systemic approach

”All your data are belong to us”

Data in NLP
Definition
What Happens to Data?
Back to Consent

What about guidelines?

To finish
Today’s NLP

Massive Language data

Language model

Used by

Language data

Annotated corpus

Reference corpus

Evaluation

Application

Manual annotation

Used by

Produce

Language data

Used by

Massive Language data

Used by

Produce

Language data

Manual annotation
Why it’s important!

Ben Hamner @benhamner · Oct 9
Programming: 10% writing code. 90% figuring out why it doesn’t work

Analyzing data and ML: 1% writing code. 9% figuring out why code doesn’t work. 90% figuring out what’s wrong with the data
Toward a systemic approach

"All your data are belong to us"
Data in NLP
Definition
What Happens to Data?
Back to Consent

What about guidelines?

To finish
data  noun, plural in form but singular or plural in construction, often attributive

da·ta  |  \'dā-tə, dā-  also  dā- \n
Definition of data

1  : factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation

   // the data is plentiful and easily available
   — H. A. Gleason, Jr.

   // comprehensive data on economic growth have been published
   — N. H. Jacoby

2  : information in digital form that can be transmitted or processed

3  : information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful
Art. 4 GDPR
Definitions

For the purposes of this Regulation:

(1) ‘personal data’ means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

https://gdpr-info.eu/art-4-gdpr/
Sensitive Data specifically protected?

Art. 9 GDPR

Processing of special categories of personal data

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation shall be prohibited.

https://gdpr-info.eu/art-9-gdpr/
Sensitive Data: exceptions

2. Paragraph 1 shall not apply if one of the following applies:

   (a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;

   (b) processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorised by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject;

   (c) processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent;

   https://gdpr-info.eu/art-9-gdpr/
Sensitive Data: exceptions again

(d) processing is carried out in the course of its legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or to former members of the body or to persons who have regular contact with it in connection with its purposes and that the personal data are not disclosed outside that body without the consent of the data subjects;

(e) processing relates to personal data which are manifestly made public by the data subject;

(f) processing is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity;

(g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;

https://gdpr-info.eu/art-9-gdpr/
Sensitive Data: exceptions again again

(h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;

(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy;

https://gdpr-info.eu/art-9-gdpr/
processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

https://gdpr-info.eu/art-9-gdpr/
Toward a systemic approach

"All your data are belong to us"
Data in NLP
Definition
What Happens to Data?
Back to Consent

What about guidelines?

To finish
Data Lifecycle

Create

Manage

Distribute

Retrieve

Archive

01 Design | Capture | Tag

02 Review | Annotate | Approve

03 Share | Deliver

04 Search | Find | Reuse

05 Preserve | Expire

Digital Asset Lifecycle

Haztowichp - CC BY-SA
Toward a systemic approach

"All your data are belong to us"

Data in NLP
Definition
What Happens to Data?
Back to Consent

What about guidelines?

To finish
What Consent Means (or not), by ©Boulet
The Nuremberg Code (1947) states that consent can be voluntary only if:

▶ participants are **able** to consent
▶ they are **free from coercion**
▶ they **comprehend** the risks and benefits involved
1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.

2. If the data subject’s consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

https://gdpr-info.eu/art-7-gdpr/
Art. 7 GDPR: Conditions for consent (2/2)

3. 1 The data subject shall have the right to withdraw his or her consent at any time. 2 The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. 3 Prior to giving consent, the data subject shall be informed thereof. 4 It shall be as easy to withdraw as to give consent.

4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

https://gdpr-info.eu/art-7-gdpr/
Consequences in Practice

There is no consent if no decision is made:

▶ opt in vs opt out
▶ importance of the default settings
▶ possibility to withdraw one’s consent at anytime
Toward a systemic approach

"All your data are belong to us"

What about guidelines?
   Beware of guidelines

To finish
Toward a systemic approach

"All your data are belong to us"

What about guidelines?
   Beware of guidelines

To finish
Table 1  Overview of AI ethics guidelines and the different issues they cover

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy protection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Security</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Source: Hagendorff, 2020]
"Currently, AI ethics is failing in many cases. Ethics lacks a reinforcement mechanism. Deviations from the various codes of ethics have no consequences. And in cases where ethics is integrated into institutions, it mainly serves as a marketing strategy. Furthermore, empirical experiments show that reading ethics guidelines has no significant influence on the decision-making of software developers.” [Hagendorff, 2020]
Guidelines and checklists are attractive:

- simple
- illusion of exhaustiveness

But they are far from enough:

”Neither the risk analysis informed by engineering practice, nor the socially informed engineering practice can be replaced by the other.” [Gurses et al., 2011]
Making the Most of Guidelines

1. start thinking/discussing without them
2. use them as a complement
3. do not limit your thinking because you checked all the list in the grid
Some guidelines I recommend

1. AI HLEG Ethics guidelines for trustworthy AI (EN or FR or …)
2. The consequentialist grid of analysis [Lefeuvre et al., 2015] (FR)
3. CERNA Machine learning ethics report (FR and EN)
4. CCNE Chatbots ethics report (FR)
Toward a systemic approach

"All your data are belong to us"

What about guidelines?

To finish

WYHTR: What You Have To Remember
▶ data is everywhere in NLP
▶ data lifecycle and ethical hotspots
▶ consent, consent, consent
Reading List

Please participate!

ACL ethics committee reading list


