Teaching Ethics in Natural Language Processing: DIY

Karën Fort

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

ERCIM, Oct. 18th, 2022
Where I’m talking from
See https://members.loria.fr/KFort/

► Language resources creation for NLP, esp. using crowdsourcing

► Ethics and NLP
Survey in the community in 2015: ethics in your curricula

Is there any class on ethics in the curricula in which you participate? [approx. 200 responses]

- No: 67.5%
- Yes: 19%
- I don't want to answer: 12.5%
- NA: 1%

Karën Fort, Alain Couillault. Yes, We Care! Results of the Ethics and Natural Language Processing Surveys. LREC 2016.
DIY Ethics: teaching experience

- from 2018
- in Grenoble, Tours, Nancy, Panthéon-Assas and Malta
- different contexts:
  - an introduction to a real class on the subject ("pédagogie du faire")
  - a one-shot seminar
DIY Ethics: organization

- 2.5 to 3h
- groups of 3 to 4
- one newspaper article per group (carefully chosen subjects) in their language (FR/EN)
- 40 min. of reading + discussion + preparation
- 5 min. of presentation to everyone, without slides (everybody participates) + 5 min. questions
- no evaluation

They think (autonomy), they explain (peer-learning), they debate
The Internet Is Enabling a New Kind of Poorly Paid Hell

For some Americans, sub-minimum-wage online tasks are the only work available.

By Alana Semuels

Machine Learning in the wild

Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter

Attempt to engage millennials with artificial intelligence backfires hours after launch, with TayTweets account citing Hitler and supporting Donald Trump

Data minimization and informed consent

Data Broker Is Selling Location Data of People Who Visit Abortion Clinics

It costs just over $160 to get a week's worth of data on where people who visited Planned Parenthood came from, and where they went afterwards.

By Joseph Cox

Carbon footprint

https://www.technologyreview.com/2019/06/06/239031/
training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/
Long term consequences

Alexa, is this a good idea?

Amazon uses kid’s dead grandma in morbid demo of Alexa audio deepfake

Amazon taps emotional woes of pandemic, grief to push developing Alexa feature.

SCHARON HARDING - 6/23/2022, 7:14 PM

Confirmation bias (and so much more)

How Wrongful Arrests Based on AI Derailed 3 Men's Lives

Robert Williams, Michael Oliver, and Nijeer Parks were misidentified by facial recognition software. The impact cast a long shadow.

https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/
Red line?

If AI Is Predicting Your Future, Are You Still Free?

Part of being human is being able to defy the odds. Algorithmic prophecies undermine that.

https://www.wired.com/story/algorithmic-prophecies-undermine-free-will/
Opening the ethical chakras: not just consequentialism

What is good?

Kant: Act in line with duties
Aristotle: Strive for excellence
Bentham: Do the math
Arendt: Think

7:54 AM · 8 juin 2022 · Twitter for iPhone
A question of time

Shows that:

- ethics is about time
- in **one hour** we can start the reflection
- as specialists, we can go further than the journalist
The teacher’s role: if necessary

- widening their perspective:
  - the world is not only about white (male) human beings living in rich countries/regions, with Internet access
The teacher’s role: if necessary

- widening their perspective:
  - the world is not only about white (male) human beings living in rich countries/regions, with Internet access

- deepening their analysis:
  - what happened technically?
  - can we avoid this? How?
  - who made the decisions?
The teacher’s role: if necessary

▶ widening their perspective:
  ▶ the world is not only about white (male) human beings living in rich countries/regions, with Internet access

▶ deepening their analysis:
  ▶ what happened technically?
  ▶ can we avoid this? How?
  ▶ who made the decisions?

▶ making them feel how important (ie responsible) they are
Explore the ethical issues of computer science research through dystopia.

The workshop “Ethics in computer science: write your own dystopia” is intended for PhD students, students and researchers in computer science, natural language processing and philosophy. It aims to create a space for reflection and exchange to explore together the issues of research ethics through science fiction. It will take place on November 3 and 4, 2022 at the Château du Montet, in Vandoeuvre-lès-Nancy.

Next Step: doctoral training at LORIA (Nov. 3rd & 4th)

Day 1

9h – 12h: Introduction covering the philosophical bases (Karën Fort), a case study on smart office setting (Sarah Carter) and the role of dystopia and science fiction to help thinking about ethical issues (Ilaria Tiddi).
Practical activities: creating the teams, brainstorming on research topics and their potential ethical issues

12h – 13h30: Lunch

13h30 –17h: Practical activities: reflecting on ethical issues, preparing a backbone scenario from a chosen situation/story

17h30 – 18h: Pitching

Next Step: doctoral training at LORIA (Nov. 3rd & 4th)

Day 2 Writing a dystopia

A professional writer will guide your team to create your own dystopian short story. Starting from your theme, you will identify your story’s moral debate, imagine your characters and explore dramatic situations. Then you will learn to structure your storyline in order to build momentum and meaning, before actually writing the final text and ultimately sharing it with the rest of the group.

9h – 12h:
- Introduction to creative writing & dystopia
- Characters’ moral debate
- Dystopian spiral

12h – 13h30: Lunch

13h30 – 17h:
- Unfolding the narrative
- Writing the story

17h – 18h30:
Sharing the stories (12mn/group including 5mn questions)

Expected outcome

▶ 20-25 PhD students trained, more to come
▶ 6 or 7 productions, hopefully publishable (leaflets, short stories, comics?)
▶ training to be "exported" to other labs (Inria?)
Thank you for your attention and use our Bingo!

## NLP Ethics Excuse Bingo

<table>
<thead>
<tr>
<th>If I don’t, someone else will</th>
<th>Who are you to decide?</th>
<th>Ethics is relative to culture</th>
<th>There are positive uses too</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethics review is censorship</td>
<td>Science is neutral</td>
<td>Well, you flew to this conference</td>
<td>People are biased too</td>
</tr>
<tr>
<td>Negative outcomes are not predictable</td>
<td>Workers there are happy for $0.05</td>
<td>There are no alternatives</td>
<td>Don’t slow down progress</td>
</tr>
<tr>
<td>You want to go back to candles?</td>
<td>Ethics review is US imperialism</td>
<td>The data was publicly accessible</td>
<td>Stop being political</td>
</tr>
</tbody>
</table>

CC-BY-SA

Emily M. Bender & Karënn Fort 2022

https://twitter.com/KarnFort1/status/1532699418067193857
Fort, K. and Couillault, A. (2016). Yes, We Care! Results of the Ethics and Natural Language Processing Surveys.