A Crash Course on Ethics for Natural Language Processing

Annemarie Friedrich
Bosch Center for Artificial Intelligence
Renningen, Germany

Torsten Zesch
Language Technology Lab
University of Duisburg-Essen

What Self-contained **lecture unit** on **ethics in NLP** / starting point for creating a unit

Why Easy integration in any **introductory** NLP course

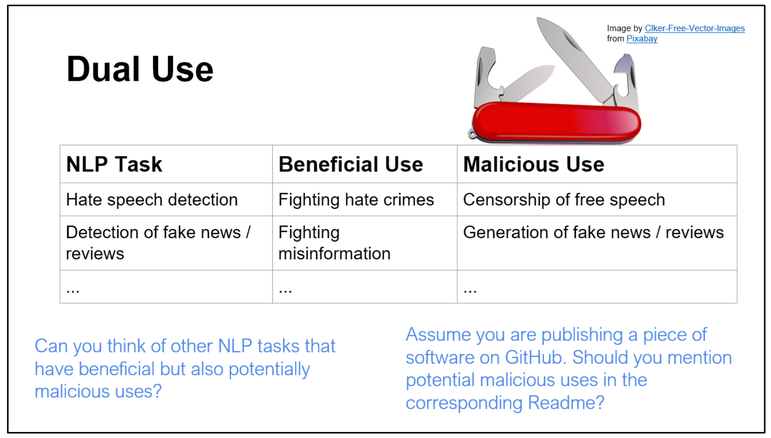
Who By teachers for teachers

Where **Freely available** on Google Slides http://gscl.org/en/resources/ethics-crash-course

Feedback and improvements welcome ☺

Slides with extensive **comments**
(in easily adaptable format and design)

Embedded **exercises**
& suggestions for **reading assignments**



Dual Use

Image by Clker-Free-Vector-Images from Pixabay

NLP Task	Beneficial Use	Malicious Use
Hate speech detection	Fighting hate crimes	Censorship of free speech
Detection of fake news / reviews	Fighting misinformation	Generation of fake news / reviews
...	...	...

Can you think of other NLP tasks that have beneficial but also potentially malicious uses?

Assume you are publishing a piece of software on GitHub. Should you mention potential malicious uses in the corresponding Readme?

Even if the system works as intended, there might be unintended side effects …

As a developer of an NLP software or as an NLP researcher working on a method or dataset, you cannot always prevent malicious use thereof. This makes it very important to think about and document potential dual use of your method or system. Conferences such as ACL or EMNLP usually allow a section on “Ethical Considerations” which you can (and should) use to discuss ethical issues considering your paper.

Doctor vs. Nurse

The doctor recommended to perform an X-ray.
He/She said …

The nurse recommended to perform an X-ray.
He/She said …

Do you think “he” or “she” is a more likely continuation in the above cases (respectively)?
What would happen if you asked a large pre-trained language model?

Reading Assignment / Discussion

Daniel J. Solove. 'I've Got Nothing to Hide' and Other Misunderstandings of Privacy. San Diego Law Review, Vol. 44, p. 745, 2007

Germany's complicated relationship with Google Street View. NY Times, April 2013.

Questions to think about / discuss:

Which dimensions of privacy matter most to you?

A software developer accidentally notices a document where a user is drafting a suicide note. Should he/she contact the police to save a life, or respect their user's secret?

Can you imagine a situation where interfering with someone's privacy leads to an economic / financial issue for that person?

Image Source:
https://pixabay.com/de/vectors/arzt-krankenhaus-gesundheit-5187733/