Seminar Ethics in NLP Intro Session
Master-Seminar – E4N: Ethics in Natural Language Processing (IN2107, IN4428)

Tobias Eder, M.A. M.Sc.
Prof. Dr. Georg Groh

Research Group Social Computing, Department of Informatics,
Technical University of Munich

05.05.2022
Contents

• Ethics in NLP
• This Seminar (Dates, Expectations, Guidelines)
• Topics
• Next Steps
Why Ethics in NLP?

- Ethics have become an integral part in the NLP community
- Both in a broader context and Ethics in AI
- Increased real world usage of NLP systems
- Problems with these systems are often not obvious
Contents

• Ethics in NLP
• This Seminar (Dates, Expectations, Guidelines)
• Topics
• Next Steps
Seminar Schedule

• Today: Topic selection
• Four presentation / discussion sessions
• Seminar paper due by August 1st
Presentation sessions

Seminar dates:

- June 9\textsuperscript{th}, June 23\textsuperscript{rd}, July 7\textsuperscript{th}, July 21\textsuperscript{st}
- Time: 2pm to 5pm
- Room 03.11.018 FMI
- 3 presentations per meeting
- Everyone needs to have done the basic reading for each topic (see slide and website)
Presentations

• 30-minute length presentation on your chosen topic
• Technical and ethical aspect for each topic
• Make people understand the technical side…
• … where the ethical problems arise …
• … and what suggestions there are to deal with them
• 15-minute discussion on the topic after each presentation
• Be a moderator for one discussion session that’s not your own
Seminar paper

Due date by August 1st (Monday) at 1pm

• ~8 pages long (template see webpage)
• On your presentation topic in more detail
• Should include a longer (~2 page) ethical discussion section towards the end
• Can include points that were raised during your presentation's discussion

Ethics in NLP Intro
Contents

- Ethics in NLP
- This Seminar (Dates, Expectations, Guidelines)
- Topics
- Next Steps
Topics

Topic 1: Data sources and labeling
Emily M. Bender and Batya Friedman. 2018. Data statements for NLP: Toward mitigating system bias and enabling better science. https://www.mitpressjournals.org/doi/pdf/10.1162/tacl_a_00041

Topic 2: Bias in models and debiasing
Yoberg Mor Geva and Jonathan Berant. 2019 "Are We Modeling the Task or the Annotator? An Investigation of Annotator Bias in Natural Language Understanding Datasets"
https://arxiv.org/abs/1908.07898
Topics

Topic 3: Language models and automated text generation

Topic 4: Natural language understanding
Topics

Topic 5: Hate Speech detection
Sean MacAvaney et al. 2019. "Hate speech detection: Challenges and solutions”
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0221152

Topic 6: NLP for (and against) disinformation
Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The spread of true and false news online.”
Topics

**Topic 7: Explainable NLP**

**Topic 8: Privacy and anonymity in NLP applications**
Topics

Topic 9: NLP for Social Good

Topic 10: Accountability for ML Algorithms
Hal Daume III "Should the ML and NLP Communities have a Code of Ethics?" https://nlpers.blogspot.com/2016/12/should-nlp-and-ml-communities-have-code.html + Professional Code of Conducts
Topics

Topic 11: Empiricism in ML Research

Topic 12: Environmental impact of NLP
Contents

- Ethics in NLP
- This Seminar (Dates, Expectations, Guidelines)
- Topics
- Next Steps
Next Steps: Distributing Topics
Questions?