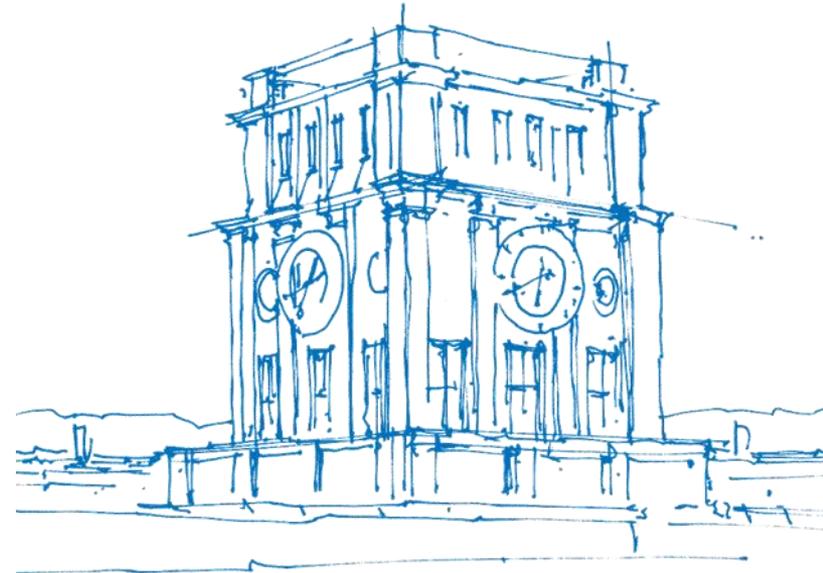# Seminar Ethics in NLP Intro Session

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

TUM Uhrenturm

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 9th, June 23rd, July 7th, July 21st
- 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

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

Irene Solaiman et al. 2019: "Release Strategies and the Social Impacts of Language Models".
https://arxiv.org/abs/1908.09203

Topic 4: Natural language understanding

Emily M. Bender and Alexander Koller 2020: "Climbing towards NLU:
On Meaning, Form, and Understanding in the Age of Data". https://aclanthology.org/2020.acl-main.463.pdf

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."
https://ide.mit.edu/sites/default/files/publications/2017%20IDE%20Research%20Brief%20False%20News.pdf

Topics

Topic 7: Explainable NLP

Marina Danilevsky et al. 2020. "A Survey of the State of Explainable AI for Natural Language Processing" https://arxiv.org/abs/2010.00711

Topic 8: Privacy and anonymity in NLP applications

Chen Qu et al. 2021. "Natural Language Understanding with Privacy-Preserving BERT" https://arxiv.org/abs/2104.07504

Topics

Topic 9: NLP for Social Good

Zhijing Jin et al. 2021. "How Good Is NLP? A Sober Look at NLP Tasks through the Lens of Social Impact." https://arxiv.org/abs/2106.02359

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

Lipton, Zachary C., and Jacob Steinhardt. 2019. "Troubling trends in machine learning scholarship."
https://dl.acm.org/doi/pdf/10.1145/3317287.3328534

Topic 12: Environmental impact of NLP

Emma Strubell, Ananya Ganesh and Andrew McCallum. 2019. Energy and Policy Considerations for
Deep Learning in NLP. https://www.aclweb.org/anthology/P19-1355.pdf

Contents

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

Next Steps: Distributing Topics

Questions?