Exam

- Longest possible (you may not need all of it)
- Open materials
Exam

• design a system to do X

• here is a system, how/why can it fail?

• here is an output of a system, which system is it?
  • our system made a mistake. explain why you think it happens / how this system is implemented

• linguistic structures and annotations

• suggest features for a problem

• compute values according to METHODX

• fix this algorithm

• terms and concepts

• is X a good solution to Y? why? why not?
my research agenda

(iextract slides)
Another Problem with Existing NLP Systems

what languages are we processing?
Reading between the Lines

–

The Hidden Biases of NLP

Dirk Hovy
Computer Science Department
University of Copenhagen

Bocconi University, Milan

mail@dirkhovy.com
@dirk_hovy
Note: Prof. Reut Tsarfaty is interested in and active in Hebrew NLP research
And now to something completely different (?)
how to make a racist AI?

(notebook)
How is this related to previous issue?
Another angle

[Yanai's slides]
Learning about the World / biases in text

(slides)
Ethics

think about what you do.
Ethics

should this system be built?

(see Yulia's slides)
ETHICS

- Same decision can be both good and bad.
- Let's discuss:
  - Predict health risk
  - Ad targeting, recommendation, search ordering
  - Automatic filtering of harmful content
Ethics

• Who is going to use your system and why?

• Who may get harmed from your system?
  • Intentionally
  • Unintentionally

• Who will be excluded?
Ethics

algorithms are opinions embedded in code
ML models automate the status quo

- Cathy O'neil
Ethics

algorithms are opinions embedded in code

ML models automate the status quo

- Cathy O'neil
Ethics algorithms are opinions embedded in code
ML models automate the status quo

Cathy O'neil

WEAPONS OF MATH DESTRUCTION
How Big Data Increases Inequality and Threatens Democracy

CATHY O’NEIL

‘Wise, fierce and desperately necessary’
JORDAN ELLENBERG
Ethics

• What if your system is 100% accurate?

• What if its 90% accurate?

• How are the mistakes distributed?
Ethics

• What biases do you encode into your system?
  • Intentionally
  • Unintentionally
Ethics

• Biases in train data. Can be in:
  • $P(x)$
  • $P(y)$
  • $P(y|x)$
  • $P(x|y)$
Ethics

• Biases in model
  • Which features? feedback loops?

• Biases in Interpretation
To conclude the course

• We saw that language is central to communication and information

• We saw that language is complex

• We saw algorithms that work on language data

• We saw where things work and where they can fail

• We've seen many low-level blocks, and some applications
To conclude the course

• We hope it was interesting
• We hope you acquired useful skills
• We hope you will make good use of them
• We hope you think of the implications of your work.