

#flo #disorganized #incomplete

1 | Can we get back to coding

please? yaow.

- Privacy
 - Good ml requires lots of data, so inherent leads to violation
 - Also, algorithm predicts private things?
 - * Target example.
- Bias
 - Our model has bias built into them
 - * ex. prop 25
 - get rid of bail, and use ml to identify flight risks
 - has bias in the model due to a biased dataset
 - Input bias -> output bias
- Abuse and misuse
 - Using ml in ways that are morally wrong
 - * Ex. Cambridge analytica
 - Put up quizzes, then sold data to companys which put targeted adds to swing votes.
- Model explainability
 - Don't know how to explain results
- Responsibility
 - Sold model to someone who used it for abusive use / it just breaks
 - * Who is at fault?
 - Model builder? Data collector? Abuser?
 - * Large problem for things that are "literally life or death"
 - Watson, what happens when it's wrong?
- Automation of labor
 - Putting people out of jobs
 - More existential: what will people do? When ml does everything better than us
 - sudo-creative work is already being done
- Existential risk
 - Essentially, terminator risk

- Or, morality is not encoded into ml for task optimization
 - * ex. make a class happy,
 - human: fun activity
 - ml: inject morphine
 - * Paperclip
 - primary directive of producing paperclips
 - ignores morality