# *Machine Learning for Operational Needs*

Prof. Kathleen M. Carley

kathleen.carley@cs.cmu.edu

**Center for Computational Analysis of
Social and Organizational Systems**
**http://www.casos.cs.cmu.edu/**

Carnegie Mellon

institute for
SOFTWARE
RESEARCH

# The Promise



- Boring jobs are gone
- Complex but repetitive jobs done better – e.g. translation
- Gives people time to do the kinds of creative work people are good at
- Faster analysis of complex situations, save lives …

# The Scientific Promises

- Letting the data speak
  - Rather than prejudice the results by your theory
  - No on top assumptions
  - Removes unconscious bias

- Regularities emerge from massive data
  - Subtle but critical similarities
  - Remove bias due to an analyst spending years getting over invested
  - More data can be processed than a human can

# Key Problems with Machine Learning

- Feature set extraction
- Training sets are limited
- Training set construction is hard
- Generalized Model and Contextualized Application
- Assumption of stationarity
- The bias is in the data collection
- Explanation to the public

# Feature Set Extraction

- ML requires the data to be turned into a set of features
- Identifying the right features is an art
- Auto feature identification (e.g. some of the latent space models) decreases interpretability

# Training

- Most algorithms are still supervised learning
- Training sets
  - Don't always exist
  - Are out of date
  - Are wrong
  - Don't cover all contexts
  - Outlier cases can be critical
- Using these tools in the wild
  - You get a decrease in accuracy
  - Black swan events occur
    (e.g. driving the freak hail storm with the kid crossing the road)

# Training Set Construction is Hard

- Time to create training sets
  - Multiple people
  - Inter-rater agreement
  - Of significant size
  - Representative of problem
  - Search for other indicators
- In a volatile world – the time may not exist
- Example case – bots
  - Historic bots
  - Evolving
  - Random strings

# Generalized Model Contextualized Application

- Churchill: Hero or war criminal?
- Kevorkian: Reducing suffering or murder?
- Stealing to feed a starving family: Stealing for fun?

- ML creates a narrow with strict rules and low context
  - Criminal Justice - Narrative and context do matter
  - History - Perspective matters

# Assumption of Stationarity

- ML assumes stationarity of process

"tomorrow is like today"

- Many socio-cultural systems violate this assumption
  - Training set gets stale
  - Problem is out of bounds
  - …

# The Bias is in the Data Collection

- ML are meant to avoid data processing bias
- …. However …. The bias is often in how the data was collected
- E.g. – the insider threat issue

# Explanation to the Public

Why?

Why?

The AI said to Arrest you

You look like a murderer
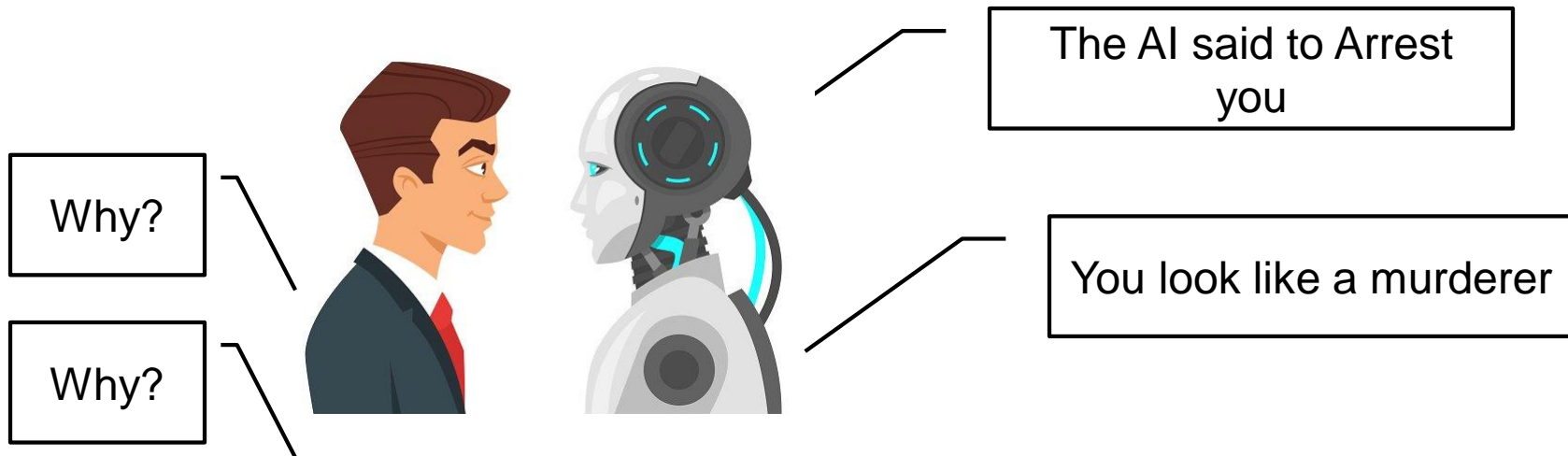
- A machine learning tool is a model
- Models for policy need to be explained to the public
- Many classes of models can't be used to provide explanations

# So When is it Bad that the Model is Wrong?

- When false positives matter
  - Judging an actor as being a criminal

- When false negatives matter
  - Missing early stage cancer

# The Promise

© 2018 CASOS, Director Kathleen M. Carley