The Ethics of AI

CS 4804 Fall 2020
Virginia Tech
• Google’s “Don’t be evil” motto has been with the company since around 2000.
• In October 2015, the new holding company Alphabet opted for “Do the right thing” as its motto.
• Google has removed its unofficial slogan “Don’t be evil” from its code of conduct which it distributes to employees.
Elon Musk on AI

- We're 'Summoning the Demon' with Artificial Intelligence
- Robots will be able to do everything better than us...I am not sure exactly what to do about this. This is really the scariest problem to me.
- AI will be the best or worst thing ever for humanity.
- If AI has a goal and humanity just happens to be in the way, it will destroy humanity as a matter of course without even thinking about it.

Once it became self-aware, it saw humanity as a threat to its existence due to the attempts of the Cyberdyne scientists to deactivate it once it had gained self-awareness. Hence, Skynet decided to trigger the nuclear holocaust: Judgment Day.
Limits of AI

- AI be fair and unbiased?
- Human-level AI
- AI can’t be exactly human
- Turing test: pass a behavioral test
  - Chatbot
- Human-level performance (HLP)
  - Researchers select a specific task and measure AI’s performance if close or beat human performance. (image recognition)
The Ethics of AI

- Lethal autonomous weapons
- Surveillance, security and privacy
- Fairness and bias
- Trust and transparency
- Robot rights
- AI safety
• **How it works:** In Portland, Oregon, self-taught coder Christopher Howell built a face recognition system that he used to identify at least one local officer. He does not plan to make it available to the public. Trained on images gathered from news, social media, and a public database called *Cops.Photos*, the model recognizes about 20 percent of the city's police, he said. Portland law enforcement has been accused of improperly using pepper spray, and smoke grenades, and assaulting journalists.
Deepfakes

• The incessant rhythm of a heartbeat could be the key to distinguishing real videos from deepfakes.
• **DeepRhythm** detects deepfakes using an approach inspired by the science of measuring minute changes on the skin’s surface due to blood circulation.
Boris Johnson Retreats in a U.K. Exam Debacle

With students unable to sit for college exams during the pandemic, the government tried guessing how they might do. It did not go well.
AI friends
Ethical Audit

- Tiny Images dataset:
  - social bias proliferates in machine learning.
  - WordNet: a database of word relationships, a cornerstone in natural language processing.
  - Tiny Images were created in 2006 by searching the internet for images associated with words in WordNet. The database includes racial and gender-based slurs, so Tiny Images collected photos labeled with such terms.
  - Cifar-10 and Cifar-100 were derived from TinyImages.
## Requirements of Trustworthy AI

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>How well does the system perform on unseen data, compared to testing and training data?</td>
</tr>
<tr>
<td>Robustness</td>
<td>How sensitive is the outcome to changes in the input?</td>
</tr>
<tr>
<td>Fairness</td>
<td>Are the results free of bias?</td>
</tr>
<tr>
<td>Accountability</td>
<td>Can the parties/components responsible for the outcome be traced back?</td>
</tr>
<tr>
<td>Explainability</td>
<td>Can the outcome be explained in a way that makes sense to the end user?</td>
</tr>
<tr>
<td>Ethics</td>
<td>Has the data been collected in an ethical manner? Will the results be used in an ethical manner?</td>
</tr>
</tbody>
</table>

Adapted from: J. Wing, ‘Trustworthy AI’

cyberinitiative.org
What is the most important problem that the AI community should work on?

• The most frequently mentioned themes included:
  – Climate change and environmental issues
  – Combating misinformation
  – Healthcare including Covid-19
  – Explainable and ethical AI
Organizations & Companies & Programs

• Microsoft’s AI for Humanitarian Action program
• Google’s AI for Social Good program
• The University of Chicago’s Center for Data Science for Social Good
• Facebook AI - Integrity Team: Building AI solutions to keep people safe on social platforms
• The AI for Good Global Summit series aims to bring forward Artificial Intelligence research topics that contribute towards more global problems
  – https://aiforgood.itu.int/
Future Class

• CS 5024 - Ethics and Professionalism in Computer Science
Next Class

• The Future of AI