Lect. #23: Fair and Responsible AI

dilbert.com/strips/comic/2013-01-31

Agenda for Today

1. Brief Introduction to AI

2. Examples of the Use of AI

3. Some Issues and Concerns Over the Use of AI

4. Current Status of AI “Regulation”

Announcements

- Erin Mullen of BetterUp will speak on “How to Work Remotely from Home” this Thursday, 11.17.22
- GCS initial report due before 11:59 pm this Friday, 11.18.22
- Note that you must be in class in Spark 227 to get credit for attendance/participation!

Poll

Have you taken a course on machine learning at WSU, online, or elsewhere?

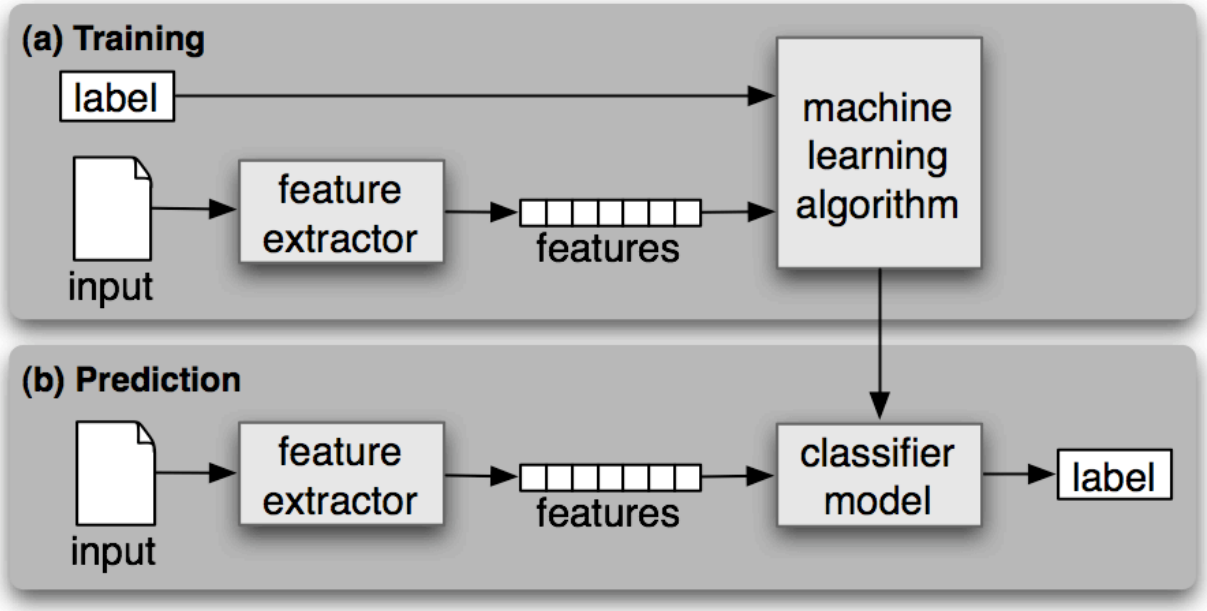
A. Yes
B. No
C. No, but I plan to take one

Brief Introduction to AI

- Dictionary definition of AI: the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages
- Majority of applications of AI referenced in the media belong to a branch of AI called machine learning (ML)
- Three basic forms of ML:
  - Supervised
  - Unsupervised
  - Reinforcement learning

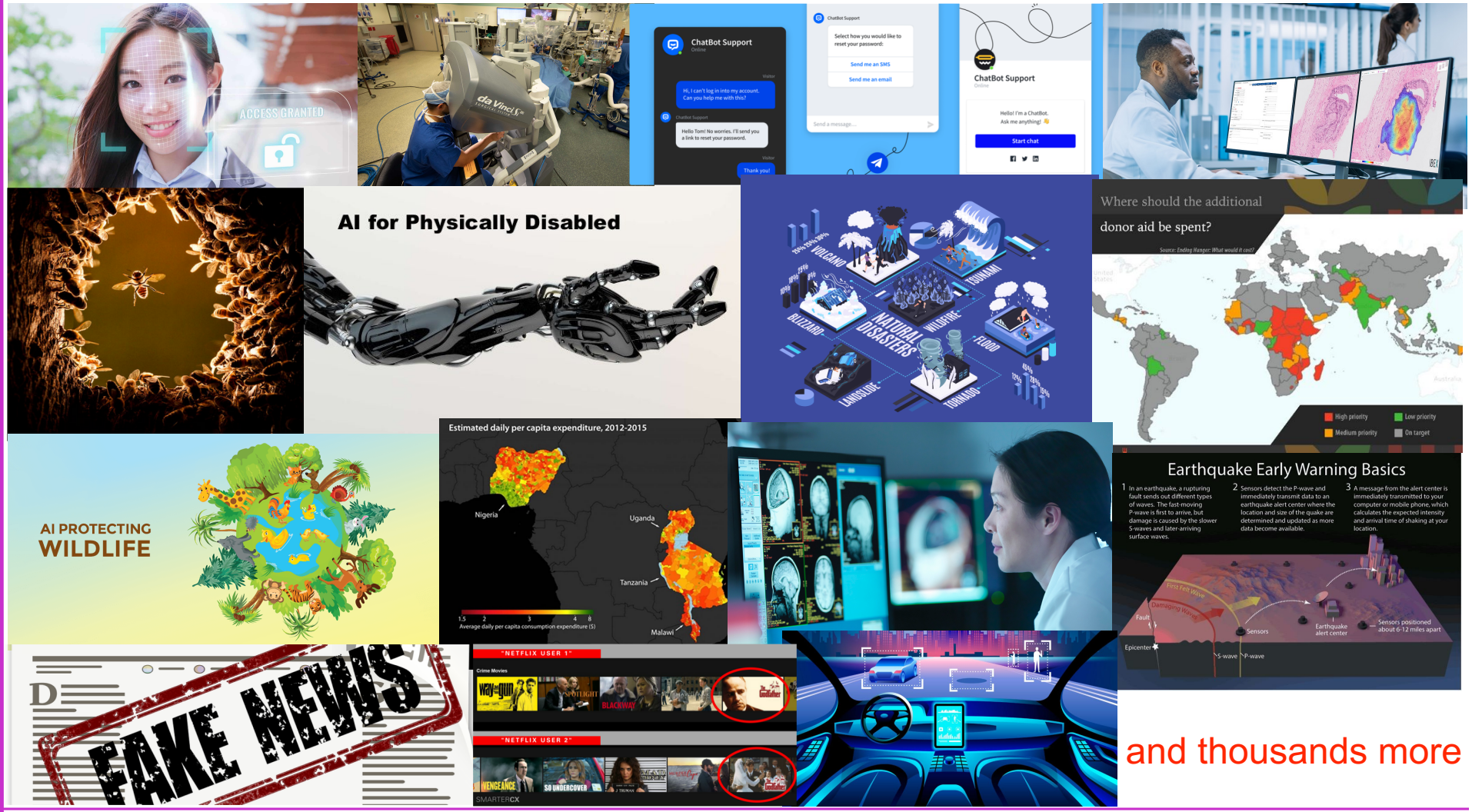
Brief Introduction to AI (cont)

- Most AI/ML applications referenced in the media use supervised ML (deep learning, NLP, etc.)
- Supervised ML uses labeled training data



www.nltk.org/book/ch06.html

Examples of the Use of AI



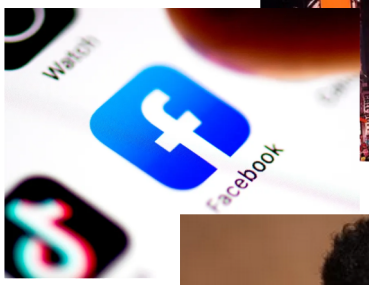
and thousands more

Breakout Discussion (5 min)

- In the previous slide, over a dozen AI applications were given. In your breakout discussion, use your imagination to come up with ways these applications could be abused.
- Choose someone to present your discussion results.

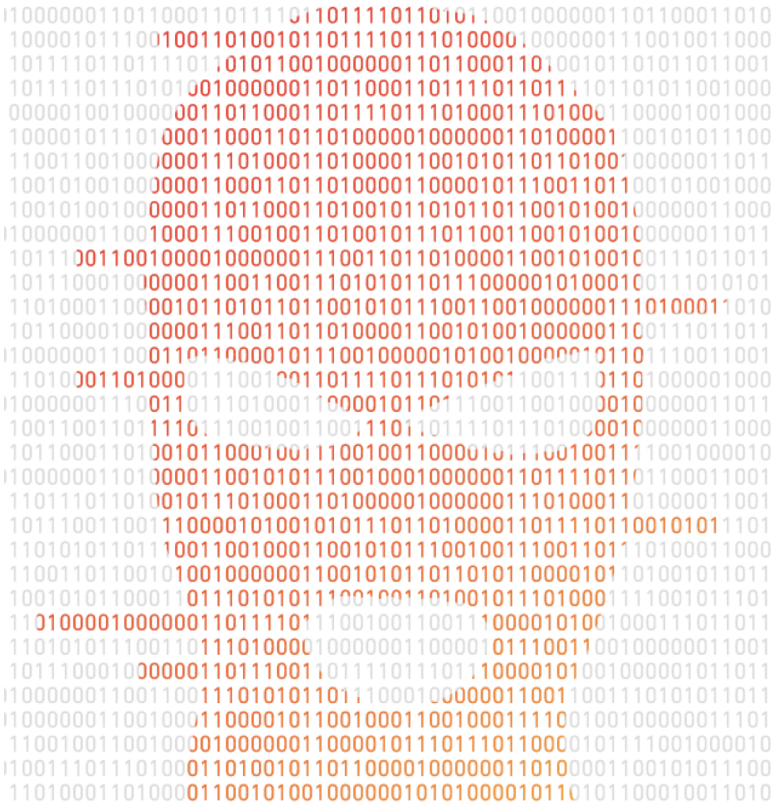
Some Issues and Concerns Over the Use of AI

- AI application screening is biased and use of facial movements akin to “modern-day phrenology”
- Advertisers now using targeted digital advertising on physical billboards based on types of people viewing them; use facial recognition
- Facebook ads match gender, race, and age based on who is pictured in the ad



Some Issues and Concerns Over the Use of AI (cont)

- Inner workings of NLP language models are kept secret, so no one knows flaws that can make them a source of misinformation, racism, etc.
- 13 language models examined; all showed bias against people with disabilities; ‘good’ turned to ‘great’ with no disability association and turned to ‘bad’ with disability association; text associated with disability labeled as negative
- Large language models will be able to write health advice websites, high school term papers, political pieces, etc.
- How will we discern real from fake?



bit.ly/3hDqrgr

Some Issues and Concerns Over the Use of AI (cont)

- AI problems have led to a call for the need for fair and responsible AI; many initiatives address the negative aspects of AI use but share the same blind spot
- AI initiatives don't address how AI can affect international peace and security
- Risk of misuse of AI for political disinformation, cyberattacks, terrorism, and military operations rarely considered; potential risk of misuse is acute
- In March 2022, deepfake used to show Zelenskyy telling his troops to surrender
- AI tool developed from existing ML model used to check for toxicity in new medical drugs took 6 hours to suggest 40k new chemical weapons

Some Issues and Concerns Over the Use of AI (cont)

- Some worst-case scenarios for AI if nothing is done to stop them:
  - When fiction defines our reality (fake images, video, audio, and text generated with advanced ML tools), false information could lead to a major crisis or even a war
  - In a dangerous race to the bottom, tiniest flaw in AI system can be exploited by hackers if major powers cut corners to get AI system up first; single point of failure could cause decisions to spiral out of control, and an automatic launch of missiles controlled by AI and without human authorization could lead to nuclear war
  - End of privacy and free will may come about by allowing companies and governments unrestricted access to all our data and addition of facial recognition, biometrics, genomic data, and AI-enabled predictive analysis; systems will predict what we'll buy, what we'll watch, and what links we'll click and know us better than we do ourselves; like boiling frogs, we'll have lost our free will before we know it

Some Issues and Concerns Over the Use of AI (cont)

- More worst-case scenarios:
  - We'll become human-Skinner boxes as algorithms, optimized to keep users on platforms as long as possible by offering rewards, short circuit the way our brains work and make us want more; the more time we spend on platforms, the less time we'll spend on positive, productive, and fulfilling pursuits
  - The tyranny of AI design, failing to account for diversity of experience and personality, will produce a level of conformity that isn't real; individuals falling outside the biased norm could be neglected, marginalized, or excluded and denied access to many services
  - Fear of AI may rob humanity of its benefits if governments overregulate it

Poll

Which of the worst-case scenarios below do you think is most likely to occur if nothing is done to prevent it?

A. Deepfakes and disinformation could lead to a crisis or even a war
B. Flaw in AI system could lead to single point of failure causing AI to launch missiles resulting in nuclear war
C. Access to all our data together with facial recognition, biometrics, genomic data, and AI-enabled predictive analysis will end privacy and possibly free will
D. We'll become human-Skinner boxes controlled by rewards received on a platform which will cause us to want even more rewards, and we'll spend less time on positive, productive, and fulfilling pursuits
E. The tyranny of AI design will lead to a level of unreal conformity, and individuals falling outside the biased norm, could be neglected, marginalized, and excluded and denied access to many services

Breakout Discussion (5 mins)

- AI clearly has many great applications for bettering our world, but equally clearly it also can be used destructively. In your breakout group, discuss how AI development should be handled. For example, should individual companies be responsible for how they develop AI applications? Should professional societies play a role? Should governments play a role? Should there be laws?

- Choose someone to present your discussion results

Current Status of AI "Regulation"

- At present no comprehensive regulation of AI exists; however, a plethora of guidelines exist
- A project by Algorithm Watch called AI Ethics Guidelines Global Inventory currently lists 173 sets of guidelines submitted by countries, companies, universities, and more
- A duck-duck-go search shows that the 173 sets of guidelines include just a sample of available guidelines
- Search shows guidelines for a myriad of industries, government entities, and governments: insurance, hospitals, finance, pharmaceuticals, automotive, NIST (Department of Commerce), Defense Innovation Unit, Department of State, Monetary Authority of Singapore, Australian Department of Industry, Science and Resources, FDA, Thailand agriculture, WHO, Japan, EU, Australia

Current Status of AI "Regulation" (cont)

- Many more sets of standards, guidelines, and resources exist
  - IEEE has published standards
  - ACM has published guidelines
  - Stockholm International Peace Research Institute has published a 48-page report
  - IEEE Global Initiative has published a 294-page report called "Ethically Aligned Design, First Edition"
- In October 2022, the White House issued an AI Bill of Rights which stated five principles for the development and use of automated systems
- AI Bill of Rights criticized for going too far and for not going far enough; as it stands, it's a non-binding set of principles

Current Status of AI "Regulation" (cont)

- NYC passed laws regulating the use of AI tools in the workplace; federal officials may scrutinize how new tech is being used to screen applicants
- In February 2022, the Algorithmic Accountability Act of 2022 was introduced in the US Senate; it would have required companies to assess the impact of the automated systems they use and sell, created new transparency about when and how automated systems are used, and empowered consumers to make informed choices about the automation of critical decisions; it went nowhere
- The Artificial Intelligence Act was proposed in 2020 by the EU with a first draft published in 2021 and a second in 2022; it divides AI into 3 categories: unacceptable (and banned), high risk, and unregulated
- AI Act lays out specific objectives: ensure that AI systems placed on the EU market and used are safe and respect existing law on fundamental rights and EU values; ensure legal certainty to facilitate investment and innovation in AI; and more (108-page pdf)
- AI Act is expected to pass and will likely affect the rest of the world

Poll

Do you think AI should be regulated in any way?

A. Yes
B. No
C. Not sure

If you think we need to do more to make AI fair and responsible, what can you do?

References

- p. 7: ocrsolutions.com/software/face-recognition/, oto.wustl.edu/new-technology-expands-use-of-transoral-robotic-surgery-tors/, joetechreviews.com/top-10-live-chat-products-for-websites/, kr-asia.com/israeli-ai-diagnostics-firm-ibex-to-roll-out-cancer-screening-for-nhs-patients-in-the-uk, www.slideshare.net/musicprabhant/ai-for-disabled, automatonai.com/ai-and-climate-change/, aiworldschool.com/research/this-is-why-ai-in-wildlife-conservation-is-so-glorious/, mrtejkohli.medium.com/using-ai-to-predict-poverty-e9ddceb6e7a0, indiaai.gov.in/article/artificial-intelligence-a-shield-against-fake-news, www.cedars-sinai.org/blog/ai-in-medical-imaging.html, www.weforum.org/agenda/2018/02/a-new-ai-detected-17-times-more-earthquakes-than-traditional-methods, builtin.com/artificial-intelligence/artificial-intelligence-automotive-industry, becominghuman.ai/how-netflix-uses-ai-and-machine-learning-a087614630fe
- p. 10: gizmodo.com/billboards-facial-recognition-privacy-targeted-ads-1849655599, news.northeastern.edu/2022/10/25/facebook-algorithm-discrimination/, www.bbc.com/news/technology-63228466