outrageously
AMBITIOUS

Module 4: Human & Societal Considerations

Duke
PRATT SCHOOL of ENGINEERING

Module 4 Objectives:

At the conclusion of this module, you should be able to:

1) Differentiate between how AI and humans learn and create predictions
2) Explain approaches to using AI to augment human intelligence
3) Develop strategies to inspire model trust among your users

outrageously
AMBITIOUS

AI and Human Intelligence

Duke
PRATT SCHOOL of ENGINEERING

Artificial general intelligence

- Original conception of AI in 1955 was “artificial general intelligence” (AGI)
  - Ability of an intelligent agent to learn any intellectual task that a human can

> “Machines will be capable, within twenty years, of doing any work a man can do”
> *- Herbert Simon, 1965*

- Common mis-conception even today that AI is equivalent to AGI

Narrow AI

- “Narrow AI” is the ability to accomplish specific pre-learned problem-solving tasks
- All applications today represent Narrow AI
  - Trained models are not easily transferrable to new problems

Golden retriever

Labrador retriever

???

AI vs. human learning

AI is a poor approximation of human learning

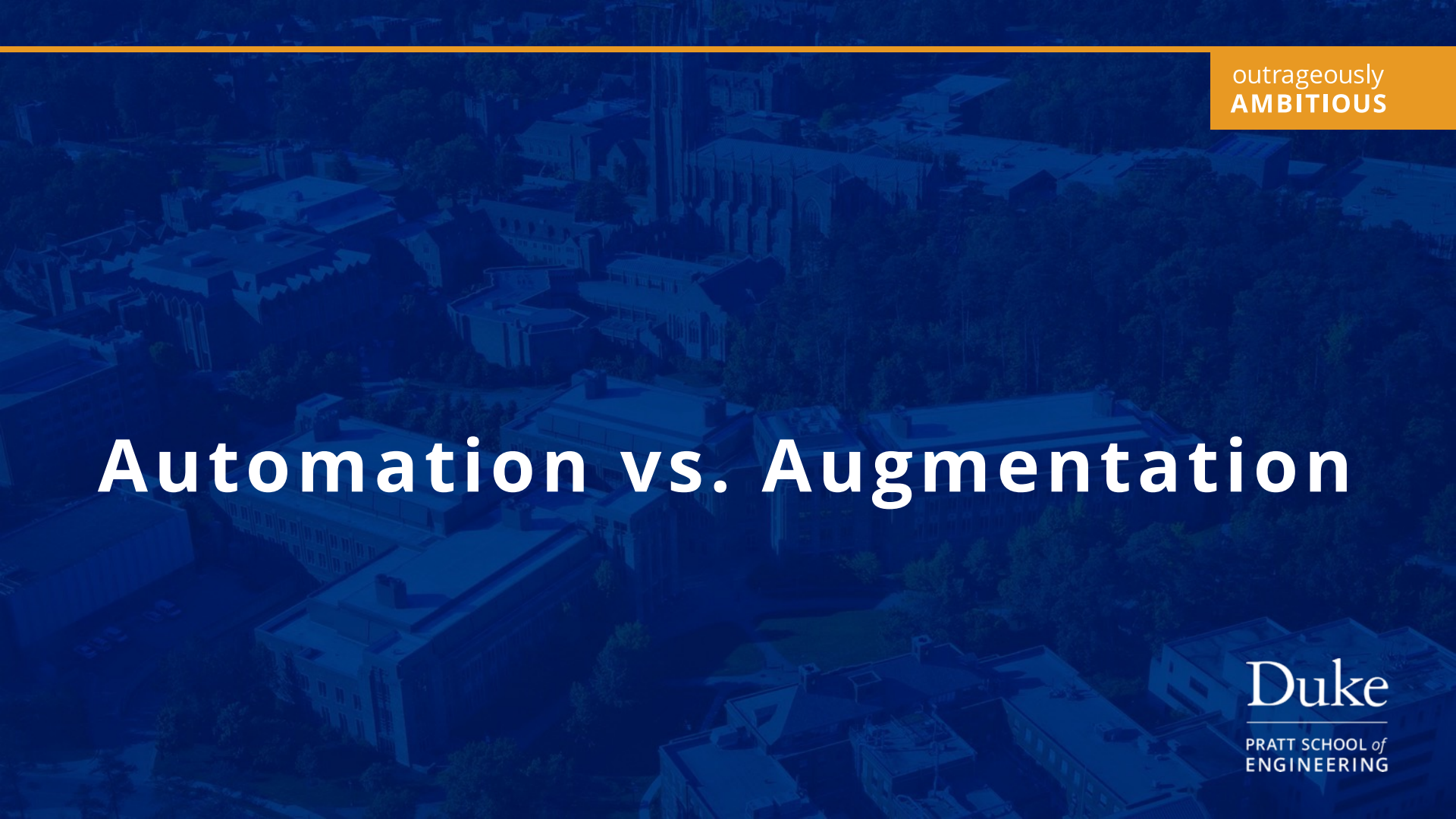
- Uses statistical methods on large amounts of data
- Difficulty understanding context and causation

"One of the fascinating things about the search for AI is that is has been so hard to predict which parts would be easy or hard… it turns out to be much easier to simulate the reasoning of a highly trained adult expert than to mimic the ordinary learning of every baby"

- Alison Gopnik, UC Berkeley

Human vs. AI prediction

Human prediction	AI prediction
Limited memory and processing ability	Ability to process vast amounts of data
Decisions influenced by many factors – emotions, biases, physical state	Consistent decisions not influenced by external factors
Lean on mental heuristics to reach decisions that fit with our concept of the world	Makes decisions based solely on data (which can contain biases)
Can apply commonsense reasoning to novel situations	Cannot reason beyond the data it was trained on


outrageously
AMBITIOUS
Automation vs. Augmentation
Duke
PRATT SCHOOL of
ENGINEERING

Automation vs. augmentation

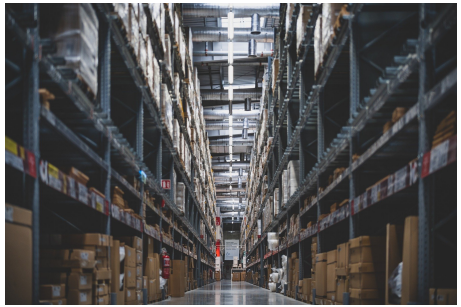
- AI can easily learn routine tasks which follow consistent patterns
- Other tasks require understanding of context, or ability to reason
- Use cases for AI fall into two categories:
  - **Automation** – replacing humans
  - **Augmentation** – supporting humans

Automation

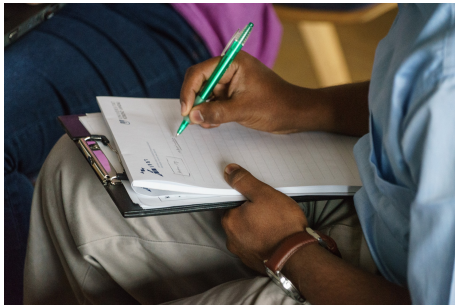
- AI expands the scope of what can be automated, including tasks involving computer vision or NLP
- McKinsey study of 800 occupations found that 60% have more than 30% of activities which could be automated[1]
  - Physical activities in structured environments
  - Data collection or processing

1) https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for

Automation



Warehouse logistics



Transcription & translation

Factory automation & QC

Customer support

Impact on jobs

- McKinsey estimates that 15% of the global workforce (~400 million) could be displaced by automation by 2030[1]
- Some job categories will go away while others will rapidly grow
  - Will accelerate shift in required workforce skills
- Labor productivity expected to rise from 0.5% in 2010-14 to average 2%

1) https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for

Augmentation

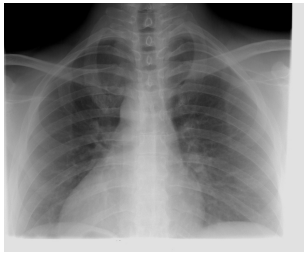
- An alternative use of AI is to complement humans rather than replace them
- Advantages of human-computer collaboration:
  - Complimentary skillsets
  - Human control over process

Cyborg chess

Forms of AI augmentation

Triage

Insurance underwriting



Radiology

Decision support

Investing

Medical diagnosis


outrageously
AMBITIOUS
Inspiring Model Trust
Duke
PRATT SCHOOL of
ENGINEERING

Inspiring Model Trust

"All models are wrong, but some are useful"
- George E.P. Box

- Your model will often be wrong
- How do you get users to trust it regardless?
- If users don't trust your model, they won't use it

Inspiring Model Trust

- Communicating performance
- Presenting confidence
- Providing explanations
- Acknowledging limitations
- Human-in-the-loop

Communicating performance

Outcome Metrics

- Refers to the desired business impact for your users
- Stated in terms of the expected impact (which is often $)

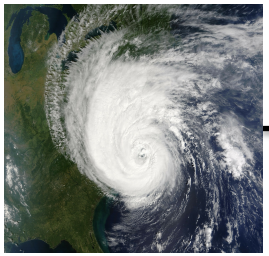
Output Metrics

- Refers to the desired output from the model
- Measured in terms of a model performance metric

How to use metrics to inspire trust:

1) Communicate outcomes for previous users
2) Track and communicate model metrics for current user

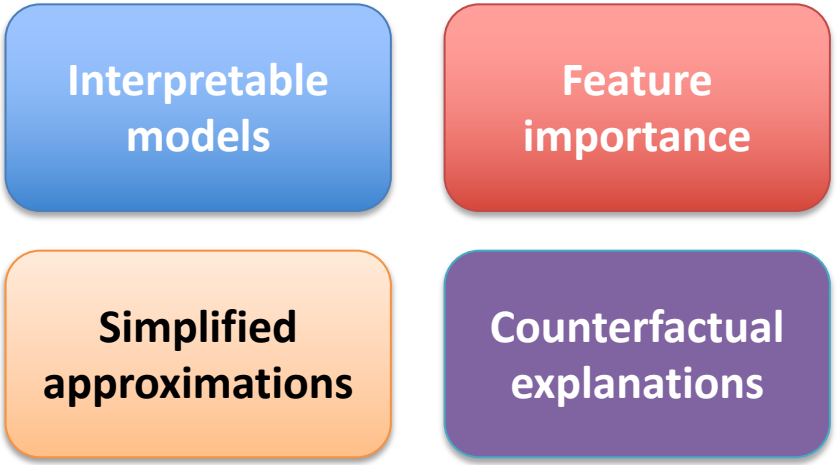
Presenting confidence



versus

Providing explanations

- When the model is off, transparency into model output helps users understand why it differs from reality



Acknowledging limitations

- Sometimes the best option is to acknowledge the limitations of the model
- Rather than giving the user a wrong answer, we give them no answer
- Instead, we can suggest an alternative solution for them

Human-in-the-loop

- Human quality control of model outputs can flag issues before they impact users
- Particularly important early in model rollout
- Opportunity to correct model or provide alternative support to customer
- Should use carefully – not try to guess when model is right or wrong


outrageously
AMBITIOUS
Change Management
Duke
PRATT SCHOOL of
ENGINEERING

AI’s impact on workflows

- Many AI systems create disruption to users’ existing workflows through automation

- Change management is key to successful adoption

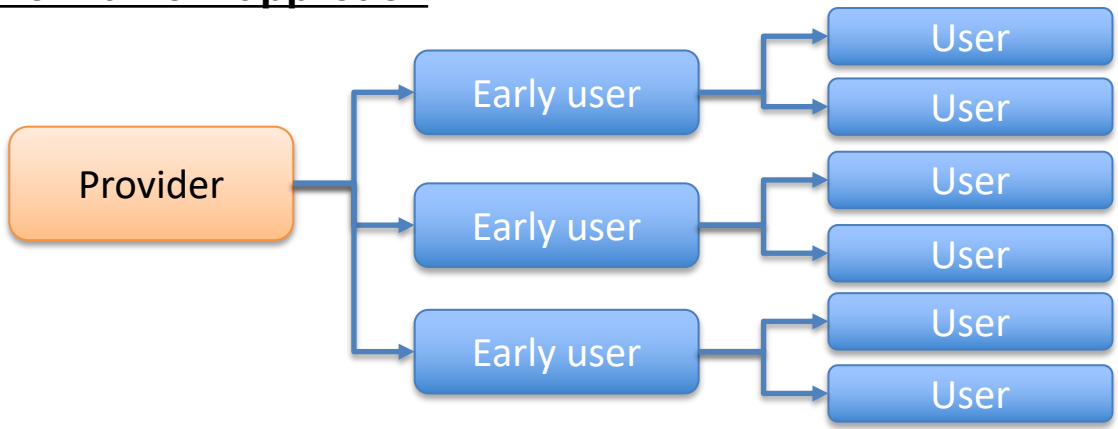
User onboarding

- Proactive user education is the key to successful change management
- Communicate the value that the AI system will create
- Focus on how it will enhance their daily lives – “what’s in it for me”
- Anticipate concerns or fears

Deployment & training

- Step-by-step deployment roadmap beginning with early adopters
- Gather feedback from early users
- Training can build confidence

“Train the Trainer” approach



Monitoring adoption

- Build in methods for monitoring product usage
- Compare adoption relative to initial expectations
- Low usage may indicate need for more education and/or training


outrageously
AMBITIOUS
Wrap-up
Duke
PRATT SCHOOL of
ENGINEERING

Wrap-Up

- AI learns and makes predictions very differently from humans
- Augmentation approaches such as triage and decision support can take advantage of complimentary skillsets
- Intentional focus on building model trust and proper onboarding can ensure adoption


outrageously
AMBITIOUS
Course Wrap-up
Duke
PRATT SCHOOL of
ENGINEERING

What we've covered

- Designing human-centered AI products
- How to protect user privacy
- Ensuring fairness, accountability & transparency in AI
- Human – AI collaboration and encouraging adoption

Final thoughts

- AI is designed to learn from the past
- Propagates that learned past forward to the future
- But what if we want a different future?
  - Who should design it?
  - What values should shape it?