Copyright Notice

deeplearning.ai

AI and Society

Introduction

Andrew Ng

AI and society

- AI and hype
- Limitations of AI

  -Bias

  -Adversarial attacks

  AI and Ethics
- AI, developing economies, and jobs
- Conclusion

deeplearning.ai

AI and Society

A realistic view of AI

Andrew Ng

Goldilocks rule for AI

- Too optimistic: Sentient / super-intelligent AI killer robots coming soon
- Too pessimistic: AI cannot do everything, so an AI winter is coming
- Just right: AI can't do everything, but will transform industries

Limitations of AI

- Performance limitations
- Explainability is hard (but sometimes doable)

Right-sided Pneumothorax (collapsed lung)

[Rajpurkar et al. (2018). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.]
[Wang et al. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. IEEE CVPR]
[Images source: NIH Clinical Center Image dataset: https://nihcc.app.box.com/v/ChestXray-NIHCC]

Limitations of AI

- Biased AI through biased data
- Adversarial attacks on AI

deeplearning.ai

AI and Society

Discrimination / Bias

Andrew Ng

AI learning unhealthy stereotypes

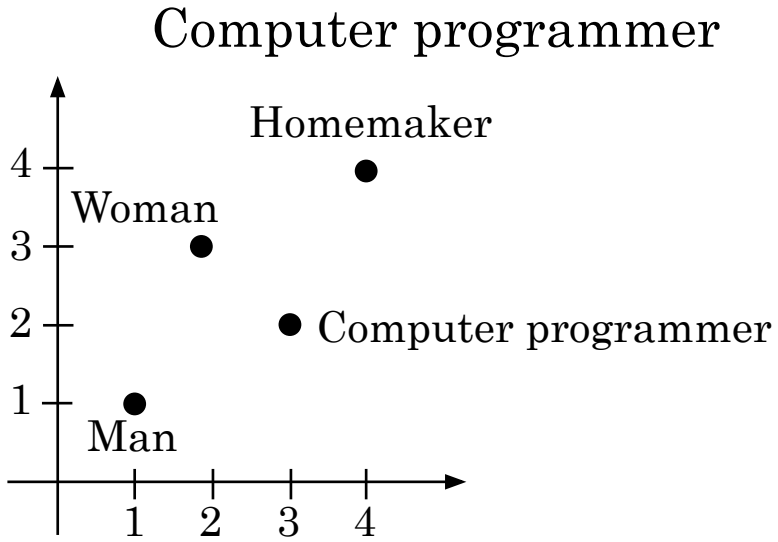
- Man : Woman as Father : Mother
- Man : Woman as King : Queen
- Man : Computer programmer as Woman : Homemaker

Man: (1,1)

Computer programmer: (3,2)

Woman: (2,3)

Homemaker: (4,4)



Bolukbasi et al. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings.

Why bias matters

- Hiring tool that discriminated against women
- Facial recognition matching dark skinned individuals to criminal mugshots
- Bank loan approvals
- Toxic effect of reinforcing unhealthy stereotypes

Combating bias

- Technical solutions:
  - E.g., “zero out” the bias in words
  - Use less biased and/or more inclusive data
- Transparency and/or auditing processes
- Diverse workforce
  - Creates less biased applications

deeplearning.ai

AI and Society

Adversarial attacks on AI

Andrew Ng

Adversarial attacks on AI

Hummingbird

Minor perturbation

Hammer

Hare

Minor perturbation

Desk

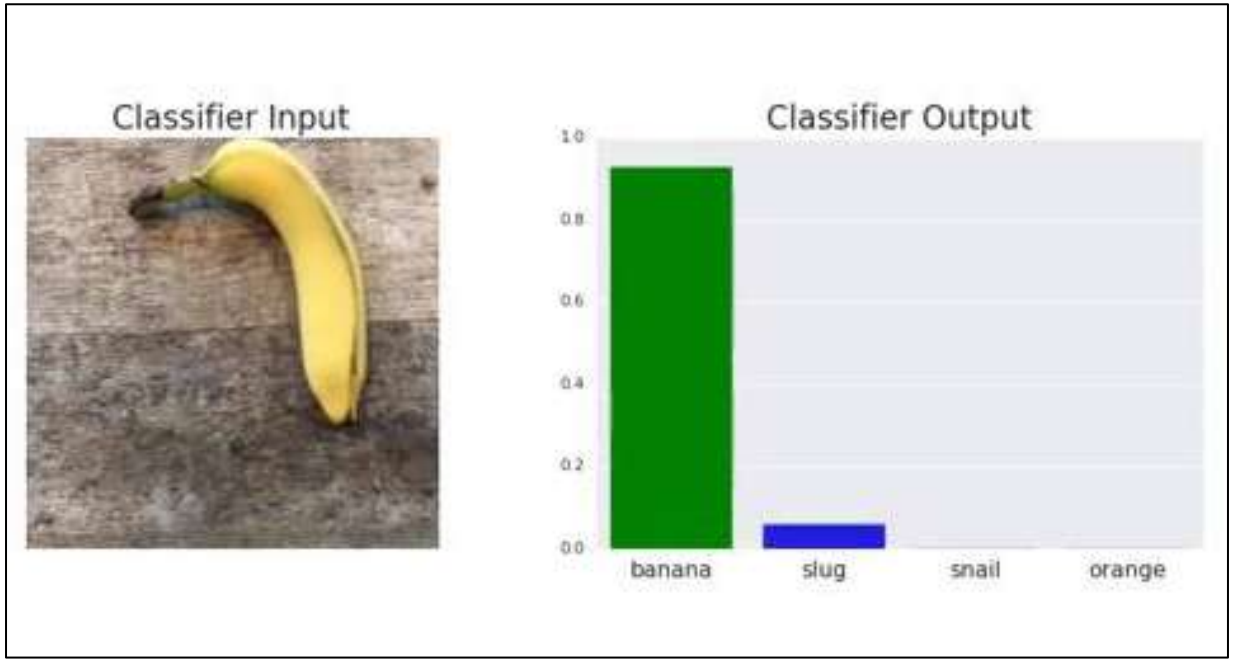
Physical attacks

"Milla Jovovich"

Fails to see stop sign

Banana

Physical attacks



[Sharif et al. (2016). Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition.]
[Eykholt et al. (2018). Physical Adversarial Examples for Object Detectors.]
[Brown et al. (2018). Adversarial Patch.]

Adversarial defenses

- Defenses do exist, but incur some cost
- Similar to spam vs. anti-spam, we may be in an arms race for some applications

deeplearning.ai

AI and Society

Adverse uses of AI

Andrew Ng

Adverse uses of AI

- DeepFakes
  - Synthesize video of people doing things they never did
- Undermining of democracy and privacy
  - Oppressive monitoring of individuals
- Generating fake comments
- Spam vs. anti-spam and fraud vs. anti-fraud

deeplearning.ai

AI and Society

AI and developing economies

Andrew Ng

Developing economies

"leapfrog"

-Mobile phones

-Mobile payments

-Online education

How developing economies can build AI

- US and China are leading, but all AI communities are still immature
- Focus on AI to strengthen a country's vertical industries
- Public-private partnerships to accelerate development

deeplearning.ai

AI and Society

AI and jobs

Andrew Ng

AI's impact on jobs worldwide

Jobs replaced by 2030

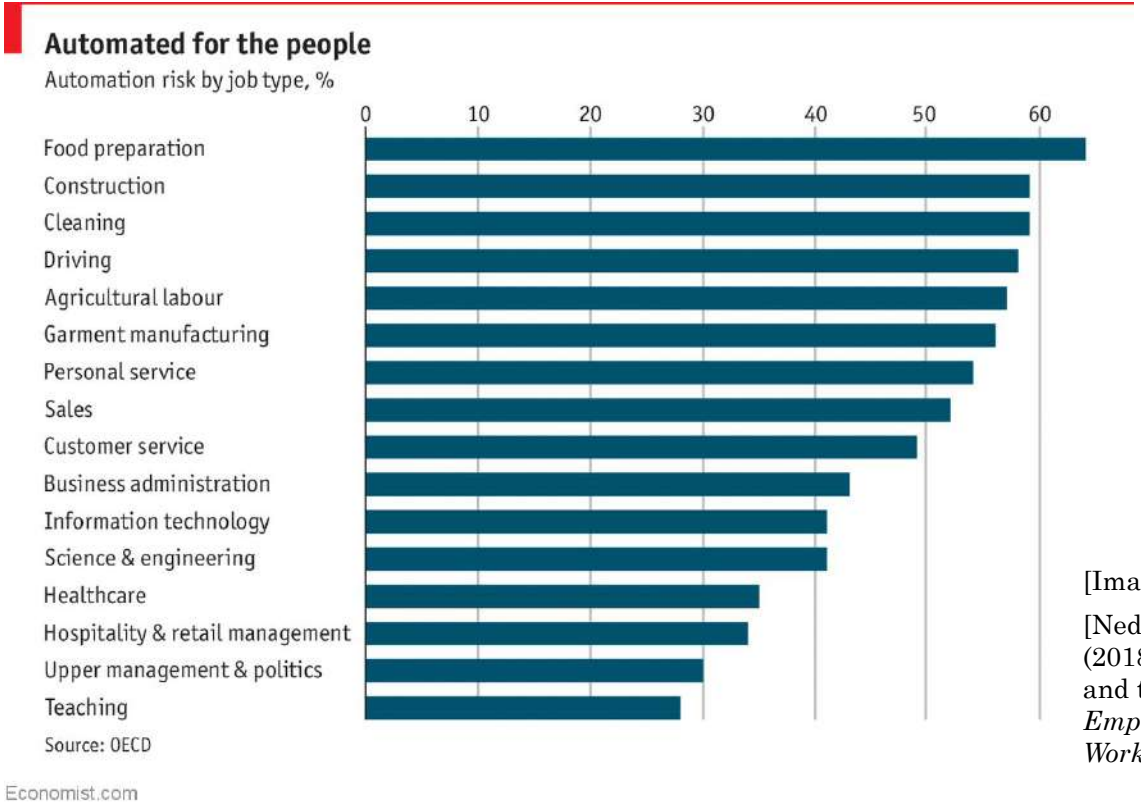
400-800 mil

Jobs created by 2030

555-890 mil

[Source: McKinsey Global Institute.]

AI's impact on jobs worldwide



[Image credit: Economist.com]

[Nedelkoska, L. and G. Quintini. (2018). Automation, skills use and training. *OECD Social, Employment and Migration Working Papers*, No. 202.]

Some solutions

- Conditional basic income: provide a safety net but incentivize learning
- Lifelong learning
- Political solutions

deeplearning.ai

AI for Everyone

Conclusion

Andrew Ng

What you've learned

- What is AI?
- Building AI projects
- Building AI in your company
- AI and society

Keep learning!

- Online courses, books, blogs, ...
- deeplearning.ai mailing list

Thank you!