



deeplearning.ai

Andrew Ng

AI and Society

Introduction

AI and society

- AI and hype
- Limitations of AI

- Bias

AI and Ethics

- Adversarial attacks

- AI, developing economies, and jobs
- Conclusion



deeplearning.ai

Andrew Ng

AI and Society

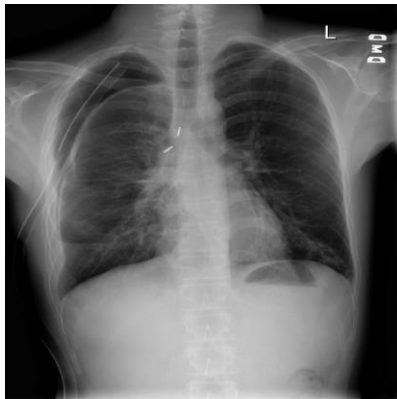
A realistic view of AI

Goldilocks rule for AI

- Too optimistic: AI will solve all of humanity's problems and create utopia
- Too pessimistic: Super intelligent/sentient killer AI can lead to human extinction
- Just right: AI is a powerful tool, but has limitations. We can mitigate its potential harms and use it to create tremendous value.

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

Andrew Ng

AI and Society

Discrimination / Bias

AI learning unhealthy stereotypes

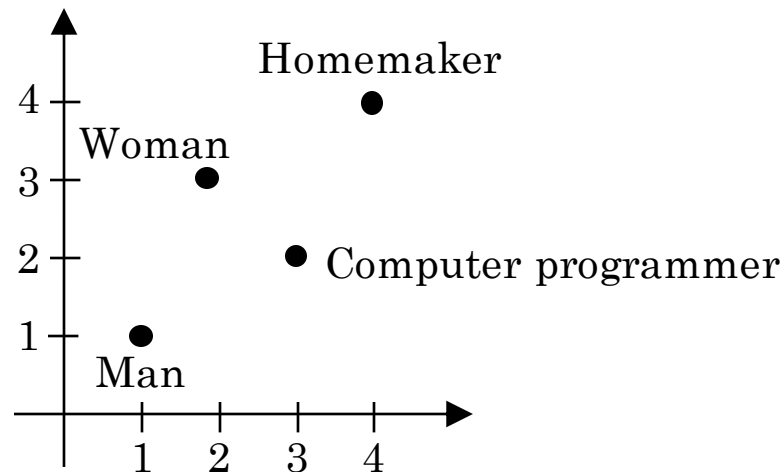
- **Man** : **Woman** as **Father** : **Mother**
- **Man** : **Woman** as **King** : **Queen**
- Man : Computer programmer as Woman : Homemaker

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

Andrew Ng

AI and Society

Adversarial attacks
on AI

Adversarial attacks on AI



Hummingbird

Minor perturbation →



Hammer



Hare

Minor perturbation →



Desk

Physical attacks



“Milla Jovovich”

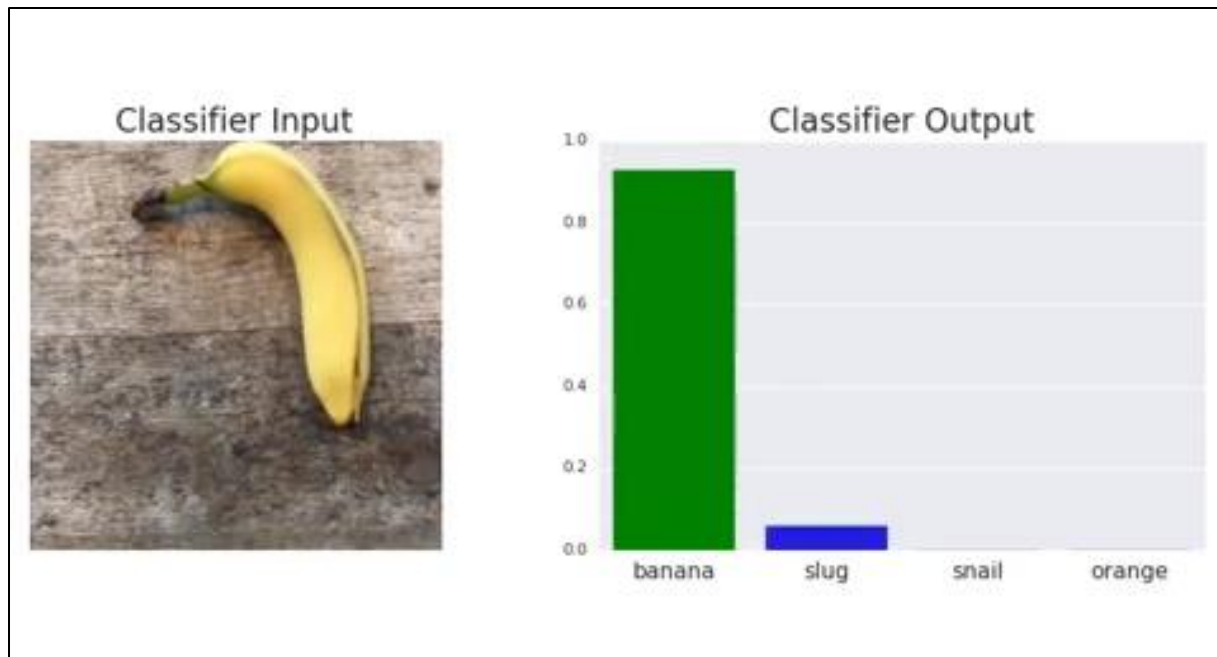


Fails to see stop sign



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

Andrew Ng

AI and Society

Adverse uses of AI

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

Andrew Ng

AI and Society

AI and developing economies

Developing economies

“leapfrog”

-Mobile phones

-Mobile payments

-Online education

How developing economies can build AI

- The US, China, UK, Canada, and many others have vibrant AI communities. But all countries still have a lot of development ahead
- Focus on AI to strengthen a country's vertical industries
- Public-private partnerships to accelerate development



deeplearning.ai

Andrew Ng

AI and Society

AI and jobs

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

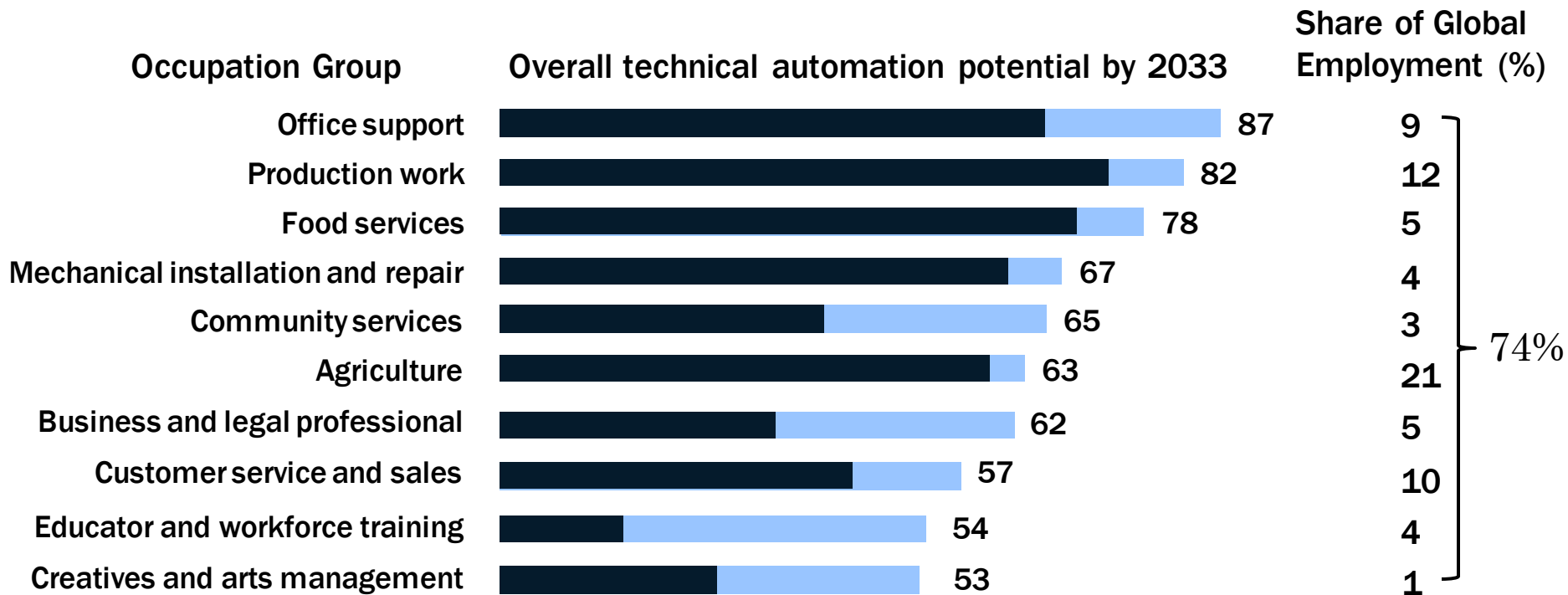


Figure adapted from “The Economic potential of generative AI: The next productivity frontier”, McKinsey Digital, June 2023

With Generative AI Without Generative AI

Some solutions

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



deeplearning.ai

Andrew Ng

AI for Everyone

Conclusion

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!