

# AI and the Black-Box Problem

By: Ethan Greene



**Who Believes AI Has a Place In Society?**

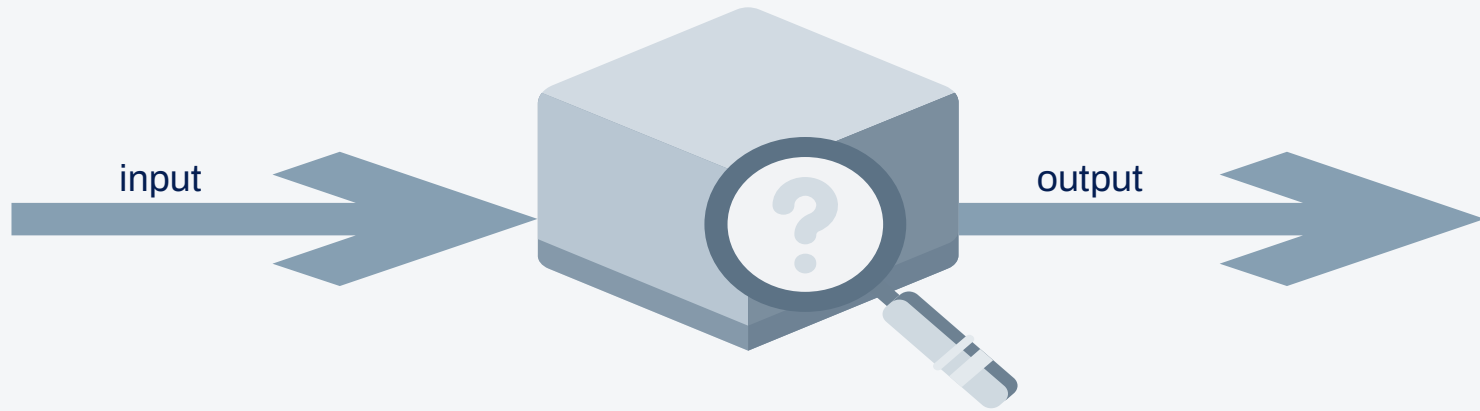
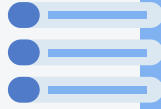
# AI's Demand and Growth

- AI Market Size to grow to \$407 billion USD by the year 2027
- AI will have a 21% increase on the United States GDP by the year 2030
- 1 in 10 cars will be self driving by the year 2030
- 97% of business owners say that chat GPT will help their business

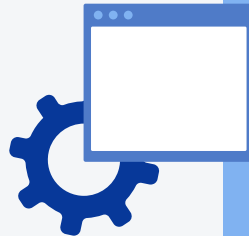
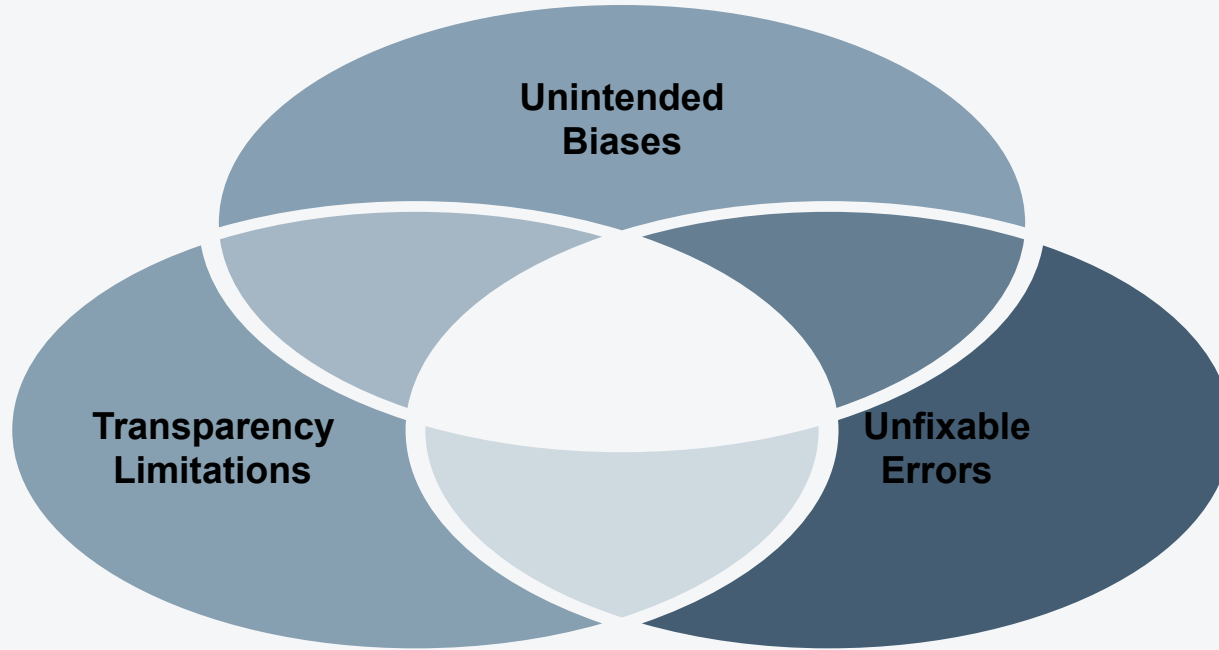


**So What's The Problem?**

# The Black-Box Problem



# Why Does It Matter



# Unintended Biases

- Racial discrimination
  - Facial recognition software found to fail at a higher rate for people of color compared to white males.
- Age discrimination
  - IBM's AI was trained on data created on data from younger aged individuals showing it wasn't as effective for some older generations
- Gender discrimination
  - Amazon Hiring Process downgrades applications including "female-indicating words" such as: "Women's" in "Women's chess club captain"

# Limited Transparency & Unfixable Errors

- Issues developing trust with these systems
  - A doctor using a diagnostic software he doesn't understand
- Accountabilities issues
  - Who's at fault when these systems fail?
- Healthcare
  - Getting denied healthcare without explanations could cause legal ramifications
- Inability to correct faulty outputs



# So What's The Solution



- Visualization Tools
- Auditing and Testing
- Explainable AI

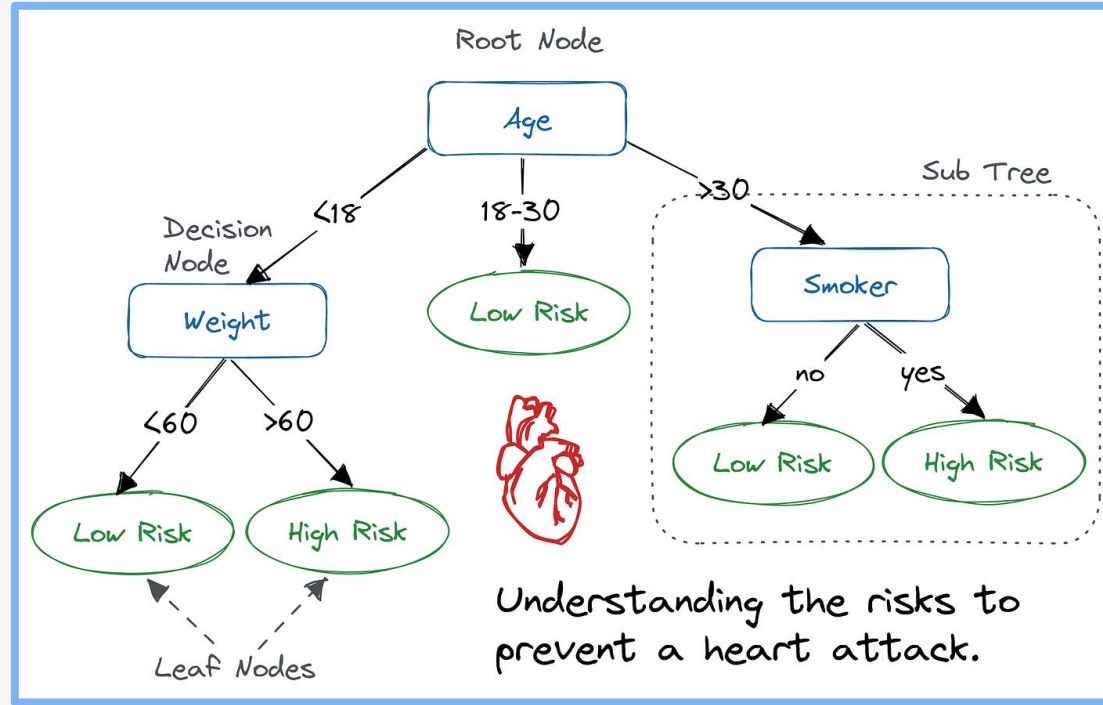


# Visualization Tools

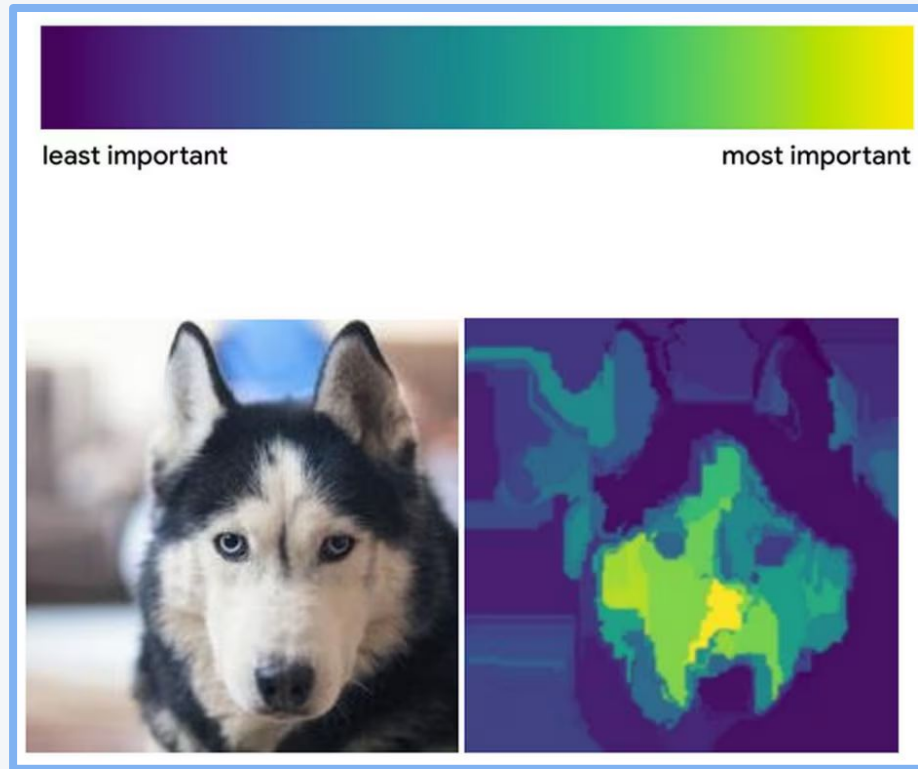
- Decision Trees
- Heatmaps
- Partial Dependence plots
  - Shows the relationship between a feature and it's predicted output which allows users to know how much changing that feature will change the output



# Decision Trees



# Heat Maps



# Auditing and Testing

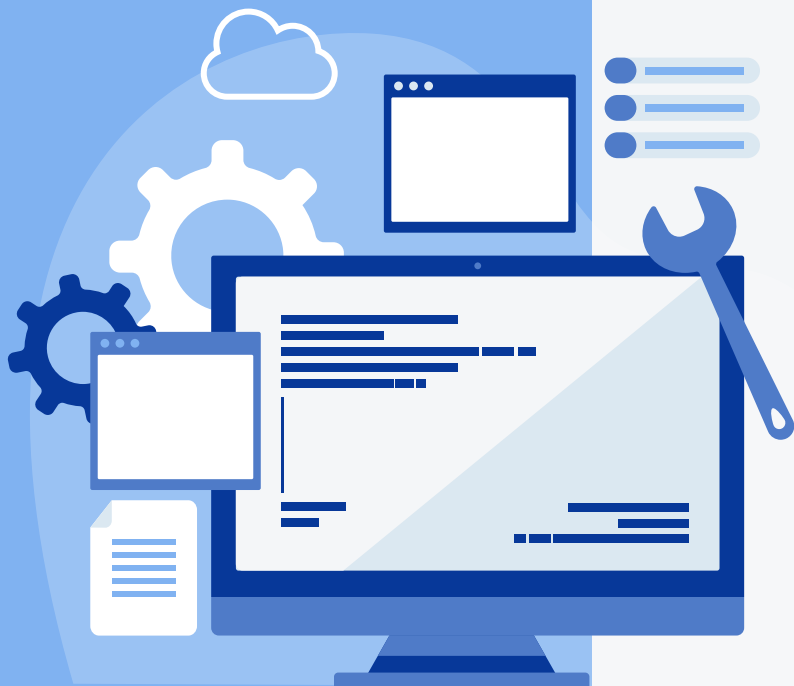
- Scheduled audits on AI systems to help mitigate biases that may be present in decision making
  - European Union's
    - AI act
  - United States':
    - Algorithmic Accountability Act
- Performing testing on these systems with specified minimum allowable accuracy levels before they are allowed to be published or verified

# Explainable AI

- Outputs can be paired with reason and support
- Outputs worded in ways that users can understand
- Accurate depiction as to how AI came up with this solution
- Allows AI to develop knowledge limits and work within the bounds from which it was created
- Enhance AI's credibility with society
- Helps us to continue to develop AI

# Conclusion

- AI has benefits
  - Data analysis
  - Military advancements
  - Cost efficient alternatives
- Do these outweigh the unexplainable biases, errors, and limitations on visibility?



# Thank You

questions?

