# Lecture 3: Implementing AI in Healthcare (part 2)

Data Sciences Institute
Topics in Deep Learning
Instructor: Erik Drysdale
TA: Jenny Du

# Lecture Outline

- Bias (ethical)
- Bias (statistical)
- Adressing risk
- Generalization challenges
- MLOps

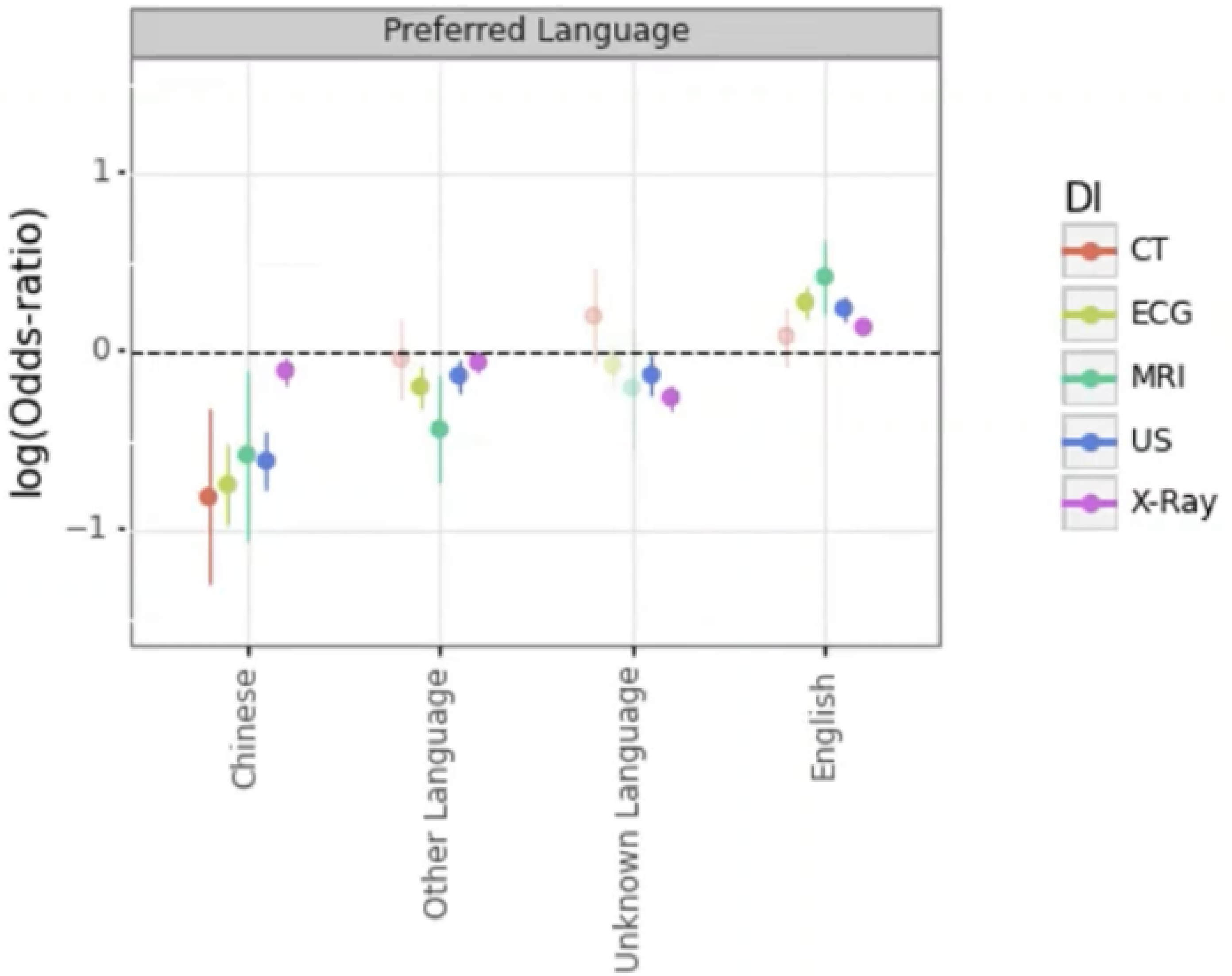
# Introduction

- The integration of AI in healthcare has great potential for improving patient care, but it is not without challenges.
- This presentation will delve into key pitfalls: bias, risk, and generalization, associated with AI in healthcare.

# Bias (ethical)

- Bias in AI refers to the systematic and unfair discrimination or favoritism in the outcomes produced by artificial intelligence systems, algorithms, or models.
- In healthcare it may lead to unequal access to healthcare, inaccurate diagnoses, or disparities in treatment recommendations based on various factors.

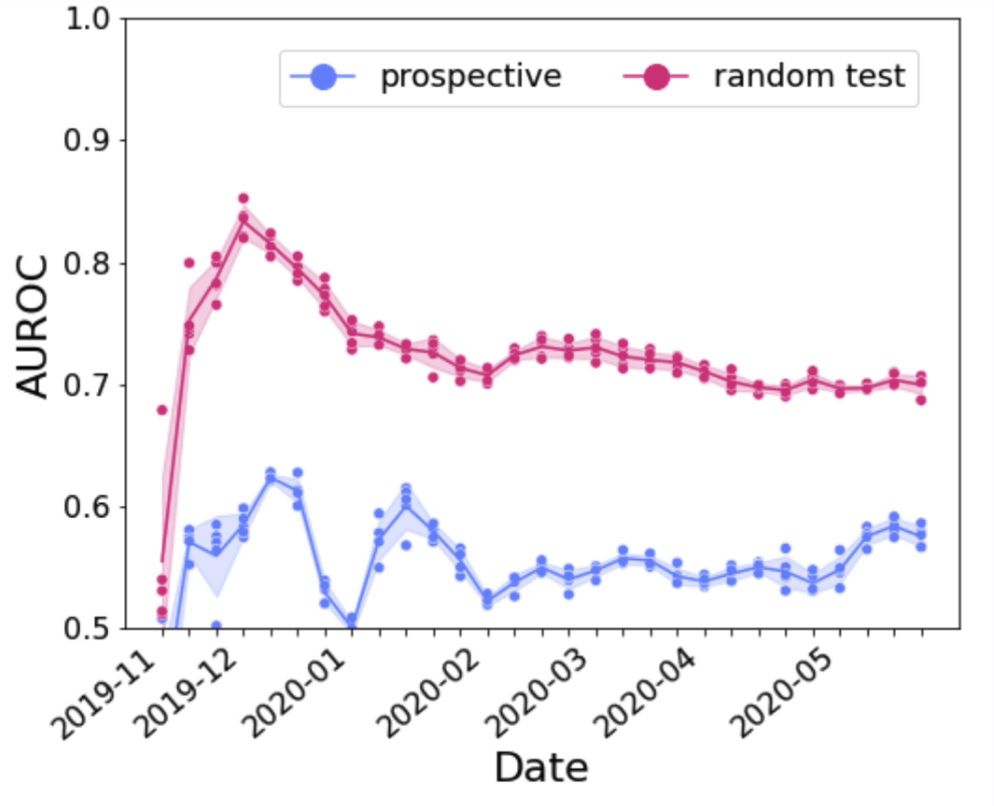
# Bias is inherant in medical practice



# Bias (statistical)

# Test set structure

- It's very important to create a test set that (most) closely resembles prospective deployment



# Types of Bias

# Selection Bias

- Selection bias is associated with the manner in which the data used for training or evaluation is collected.
- It arises when the data collection process favors certain groups or circumstances over others.
- Selection bias can introduce systemic bias into the dataset.

# Labeling Bias

- Labeling bias can occur during the annotation or labeling of data points.
  - It stems from human annotators' biases or subjective judgments when assigning labels or categories to data.
- Labeling bias can impact the accuracy of the ground truth labels used for training.

TODO: DOES IMAGEMENT GENERALIZE TO IMAGENET?

## Data Bias

- Data bias pertains to the quality and representativeness of the training data used to develop AI models.
- It results from biased or unrepresentative samples in the dataset.
- Data bias can affect the entire AI system, as the model learns patterns from the training data.

# Algorithmic Bias

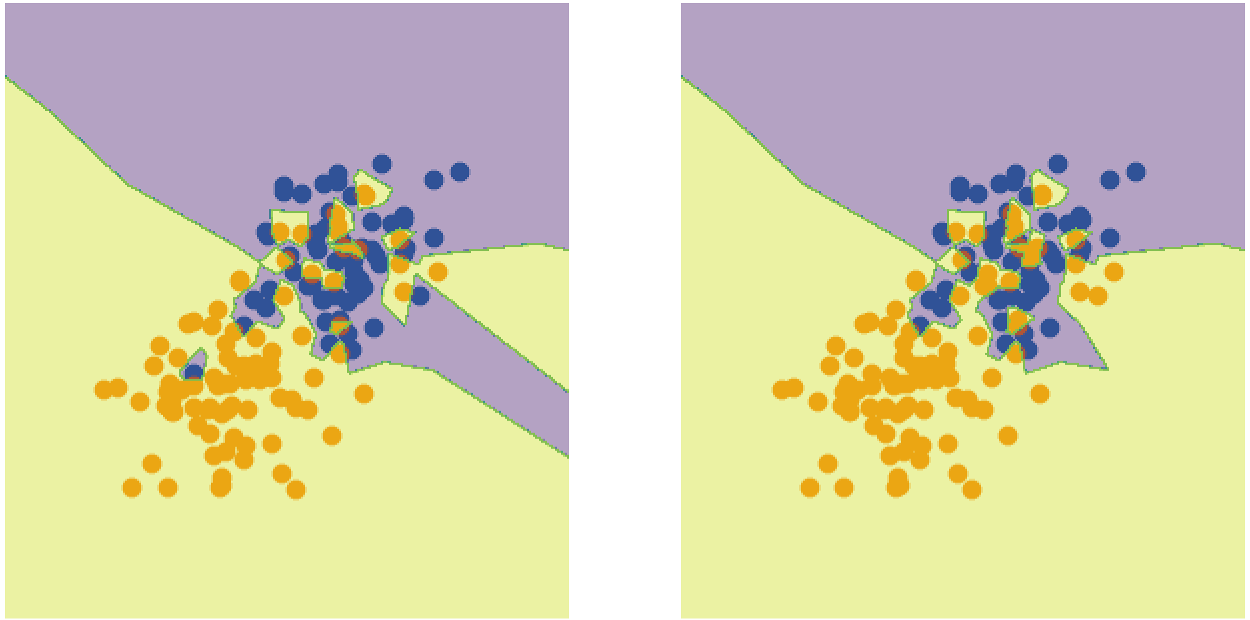
- Algorithmic bias relates to inherent biases in the design or structure of the AI algorithms themselves.
- It can result from the way features are selected, weighted, or processed during decision-making (example: Ribeiro et. al (2016))

(a) Husky classified as wolf (b) Explanation

# Reinforcement Bias

- Reinforcement bias emerges from the interactions between AI systems and users.
- It results from AI systems learning from user feedback and behavior.
- If users exhibit biased behavior, the AI may reinforce these biases in its responses.

Source



# Addressing Bias

## Diverse and Inclusive Data Collection

- Collect diverse and representative data to train AI models.
- Ensure that data includes various demographic, geographic, and socio-economic factors.
- Pay special attention to underrepresented or marginalized groups to avoid skewed or biased training data.

## Data Preprocessing and Cleaning

- Implement rigorous data preprocessing techniques to identify and mitigate bias in training data.
- Remove or re-weight biased or sensitive attributes from the dataset to minimize the potential for bias to be learned by the AI system.

# Fairness and Bias Audits

- Conduct regular fairness audits of AI models to detect and quantify bias.
- Use specialized tools and metrics (e.g., disparate impact, equal opportunity) to assess the fairness of model outcomes across different groups.

## Transparency and Explainability

- Make AI models more transparent and interpretable to understand the factors influencing their decisions.
- Implement techniques like explainable AI (XAI) to provide insights into model behavior and allow for the identification and rectification of bias.

## Continuous Monitoring and Feedback Loop

- Establish a feedback loop for continuous monitoring and improvement of AI systems' fairness.
- Collect feedback from users and impacted communities to identify and address bias issues as they arise, making ongoing refinements to models and data.

# Risk

- Risk in AI refers to the potential negative consequences or uncertainties associated with the development, deployment, and use of artificial intelligence systems.

# Examples of Healthcare AI Risks

## Data Breaches

- Healthcare data is particularly sensitive.
- Breaches can expose patient information, leading to privacy violations and legal consequences.

## Incorrect Diagnoses

- AI systems that assist in diagnostics could potentially make incorrect diagnoses, leading to improper treatment and harm to patients.

## Legal Liabilities

- Healthcare providers using AI systems face legal risks if the technology leads to patient harm, including malpractice claims.

## Ethical Concerns

- Decisions about patient care based on AI could raise ethical issues, especially regarding consent, transparency, and the prioritization of healthcare resources.

# Addressing Risk

# Robust Data Security Measures

- Implement strong data protection practices like encryption, access controls, and regular security audits.
- Ensure compliance with regulations like General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) to safeguard sensitive health data.

## Transparent and Explainable AI

- Develop AI systems that are understandable and transparent, allowing healthcare professionals to grasp how AI decisions are made.
- Can help in understanding AI model's decision-making process, providing justification for the decisions made, and identifying biases.

## Ethical AI Development and Use

- Adhere to ethical principles in AI development to ensure fairness, avoid bias, and respect patient autonomy and privacy.

## Rigorous Testing and Validation

- Subject AI systems to extensive testing and validation to confirm their safety and efficacy, and that they perform as intended across diverse patient populations.
- May include clinical trials followed by continuous monitoring post-deployment.

## Legal and Regulatory Compliance

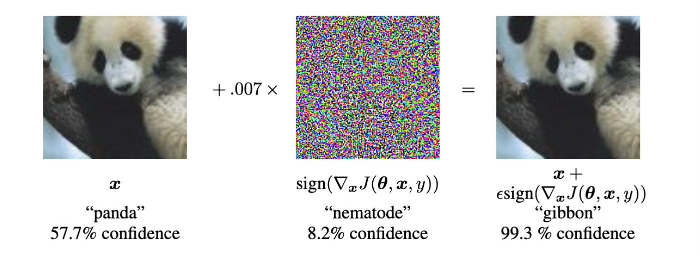
- Ensure AI systems comply with medical, data protection, and patient rights laws.
- Make sure to adapt to legal changes.

# Generalization

- Generalization in AI refers to the ability of an AI system or model to perform well on new, unseen data after having been trained on a specific set of data.

# ML models are "too sensitive"

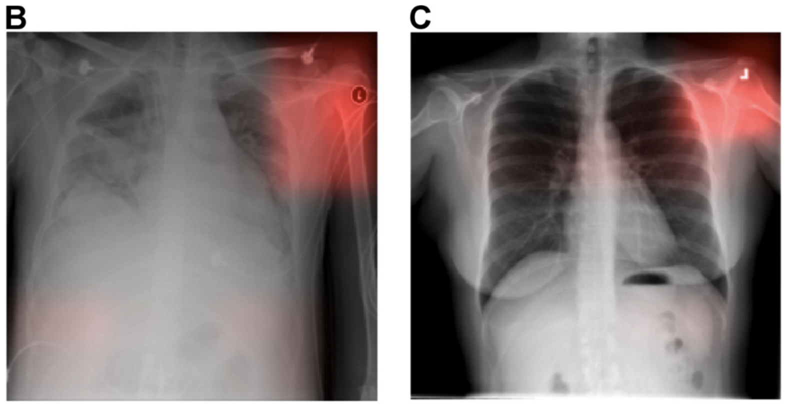
- All deep learning systems can be rendered useless by adversarial attacks




Source: Goodfellow et. al (2015)

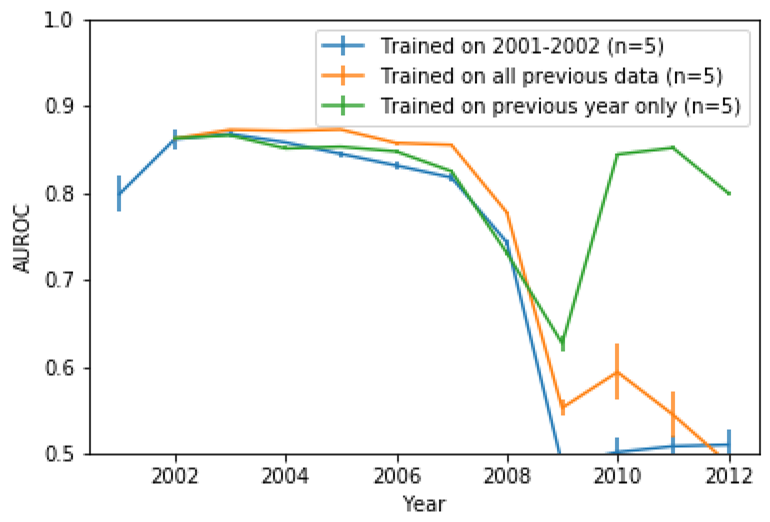
# ML models are "too sensitive"

- Hence, they can easily be fooled by "artifacts"
  - Example of CNN picking up on hospital-specific X-ray practices (source: Zech et. al (2018))



# Model drift

- After a model goes live the performance of the model will often suffer
  - Unconditional label distribution changes
  - Unconditional feature distribution changes
  - Conditional relationship b/w label and features changes




Source: Nestor et. al (2019)

## Overfitting vs. Underfitting

- Good generalization requires balance between overfitting and underfitting.
- Overfitting: Model learns the training data too well & performs poorly on new data.
- Underfitting: Model is too simple to capture the underlying structure of the data & performs poorly on training and new data.

# Importance of Generalization in Healthcare

## Diverse Patient Populations

- Healthcare datasets come from diverse populations with varying demographics, medical histories, and health conditions.
- Generalization ensures that AI models can effectively handle data from varied patient groups.

# Variability in Medical Data

- Medical data can be highly variable (i.e., imaging data, electronic health records (EHRs), genetic information).
- Each type of data has differences in quality, format, and context.
- Generalization ensures AI models can provide reliable insights across various types of medical data.

# Changing Healthcare Practices and Knowledge

- Healthcare is a rapidly evolving field (i.e., new treatments, diagnostic criteria, and research findings).
- Generalization ensures AI models are better equipped to remain relevant and accurate as medical knowledge and practices evolve.

# Addressing Generalization

## Training Data Diversity

- Model trained on a very diverse dataset is more likely to generalize well because it has been exposed to a wide variety of examples.

## Regularization Techniques

- Dropout, L1/L2 regularization, and early stopping help to prevent overfitting by penalizing complexity or stopping the training process before the model starts to overfit.

# Cross-validation

- Involves dividing the dataset into several subsets, training the model on some subsets and validating it on others.
- Helps in assessing the model's ability to generalize across different data splits.

## Model Complexity

- Simpler models usually underfit but more complex models usually overfit.
- Need to find the right level of complexity.

# Transfer Learning

- Involves taking a model that has been trained on one task and adapting it to a different but related task.
- Can help in situations where there is not enough data for training a model from scratch, leveraging the generalization capabilities learned from the original task.