Machine Learning for Time Series

Part 1: Bias

Dr. Emmanuelle Salin
Machine Learning and Data Analytics (MaD) Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg
30.01.2025

Topics Overview

1. Time series fundamentals and definitions (Part 1)
2. Time series fundamentals and definitions (Part 2)
3. Bayesian Inference and Gaussian Processes
4. State space models (Kalman Filters)
5. State space models (Particle Filters)
6. Autoregressive models
7. Data mining on time series
8. Deep Learning (DL) for Time Series (Introduction to DL)
9. DL – Convolutional models (CNNs)
10. DL – Recurrent models (RNNs and LSTMs)
11. DL – Attention-based models (Transformers)
12. DL – From BERT to ChatGPT
13. DL – New Trends in Time Series processing
14. Time series in the real world

Topics Overview

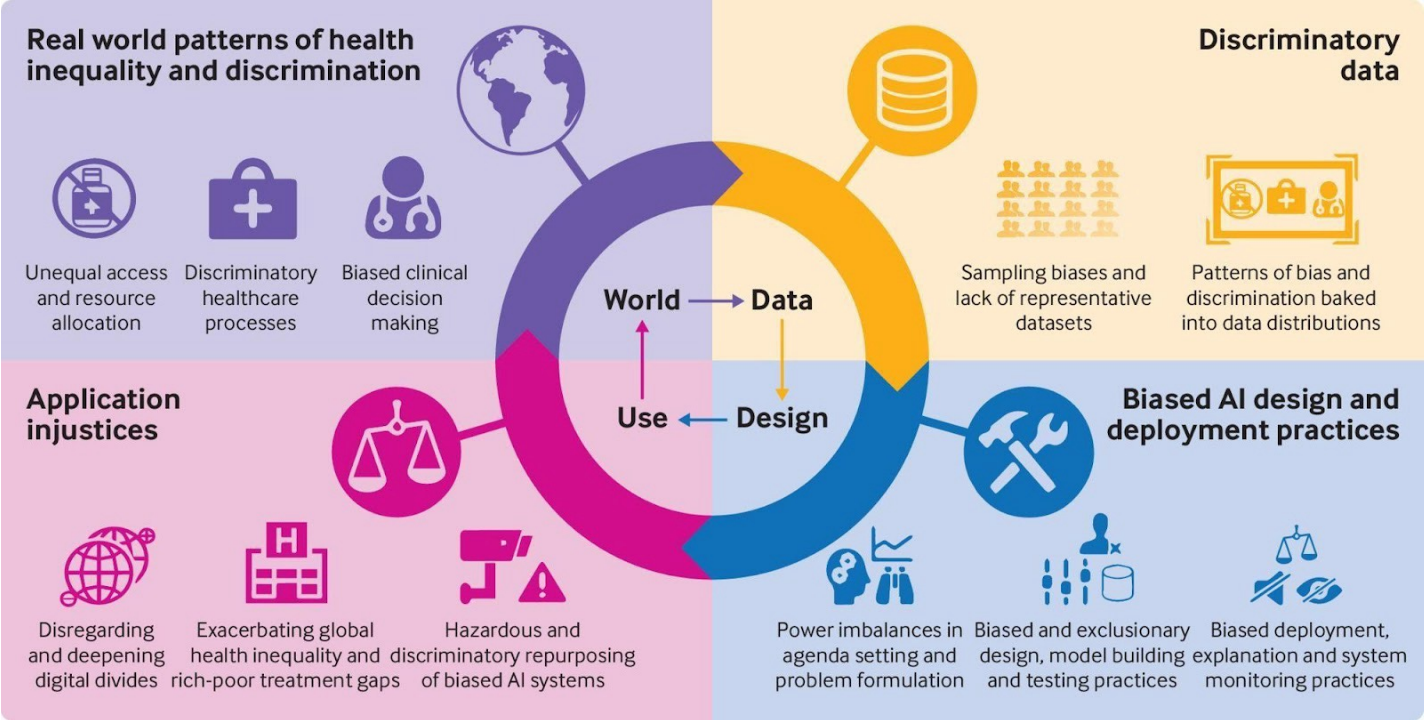
1. Time series fundamentals and definitions (Part 1)
2. Time series fundamentals and definitions (Part 2)
3. Bayesian Inference and Gaussian Processes
4. State space models (Kalman Filters)
5. State space models (Particle Filters)
6. Autoregressive models
7. Data mining on time series
8. Deep Learning (DL) for Time Series (Introduction to DL)
9. DL – Convolutional models (CNNs)
10. DL – Recurrent models (RNNs and LSTMs)
11. DL – Attention-based models (Transformers)
12. DL – From BERT to ChatGPT
13. DL – New Trends in Time Series processing
14. **Time series in the real world**

Introduction to Bias in Time Series Models

What is AI bias?

Definition:

- Machine learning bias refers to biased results due to human biases that **skew** the original **training data or AI algorithm**, leading to **distorted outputs** and **potentially harmful outcomes**.



Real-world Example of Bias in Time-series Problem

Predictive policing

- Predictive policing uses machine learning algorithms to **analyze time-series crime data** to anticipate future criminal activity locations and times.

  - Characteristics: Temporal patterns of crimes, seasonality, and trend detection.

  - Types of data used: Calls for service, arrest records, and historical crime reports.

Hypothetical Example: Metropolis City's Predictive Policing

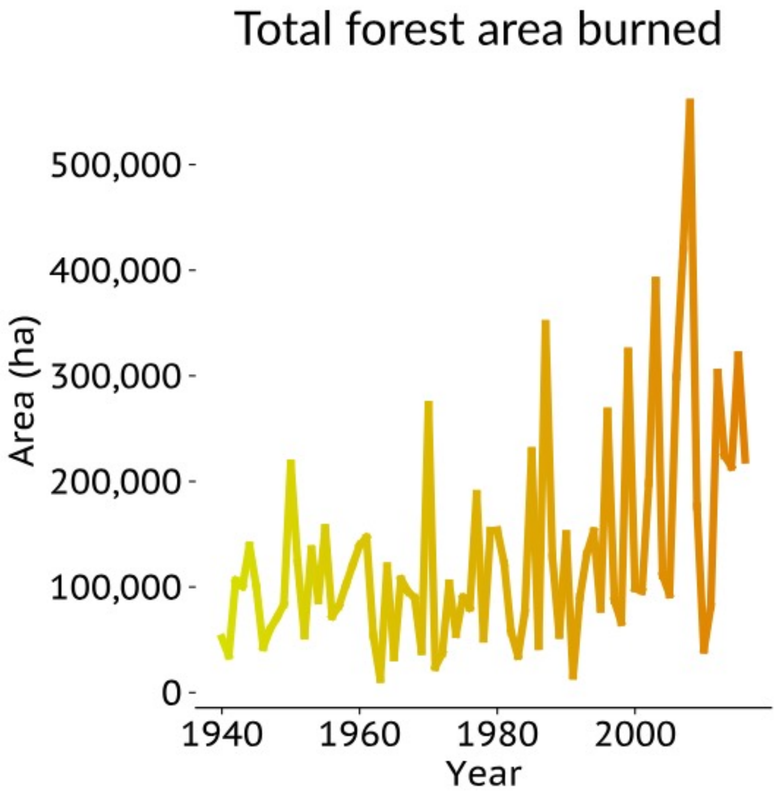
- Metropolis City has adopted a predictive policing system to help allocate its police resources more efficiently. The city has a diverse population, with various neighborhoods differing significantly in socioeconomic status.

Types of Bias in Time Series Data

Historical Bias

Historical Bias: Data used to train an AI system no longer reflects the current **reality**.

- Predictive policing systems rely heavily on historical crime data to make predictions. If law enforcement practices **in the past were biased**—intentionally or unintentionally—towards certain communities, the data generated will reflect these biases.
  - *Example: The model is trained on data from a period of discriminatory policing practices, leading to higher crime predictions in minority neighborhoods due to past over-policing.*



Source: CALFIRE

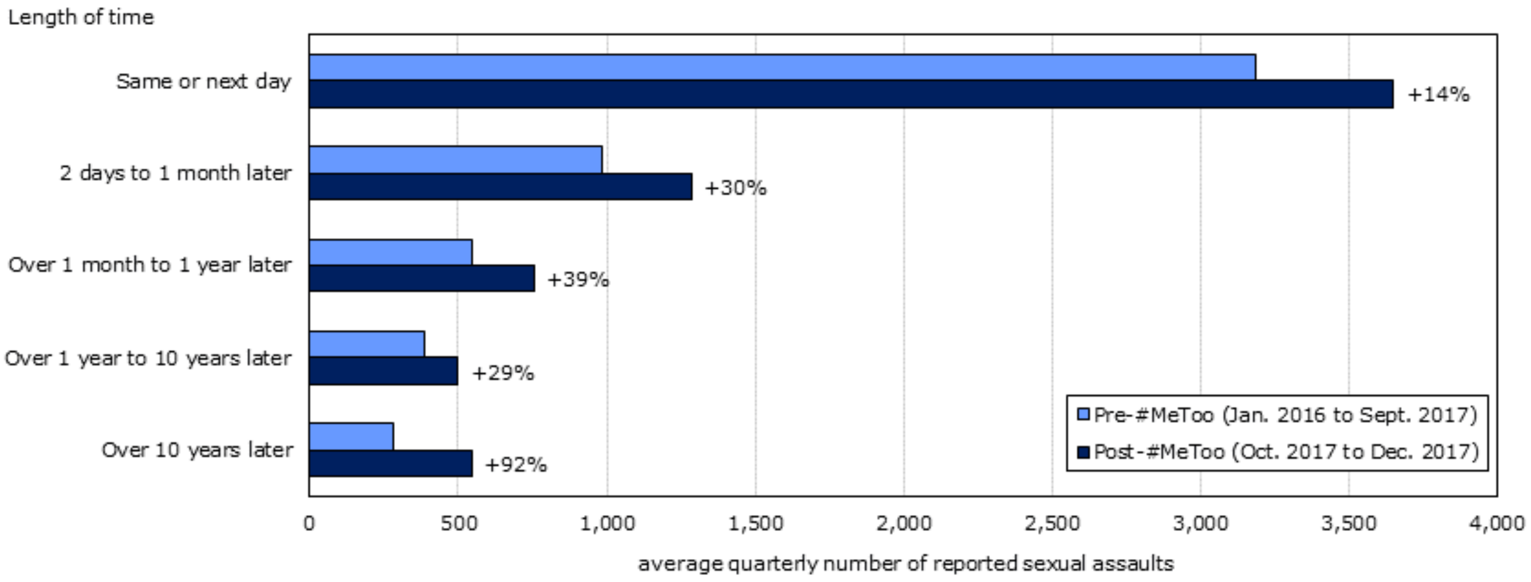
Historical bias also applies to **climate change**: climate related data from before the XXIst century does not reflect current reality.

Types of Bias in Time Series Data

Data Representativeness

Data Representativeness:

- The data used might not well **represent** the entire population or the range of crime types.
  - *Example: Certain crimes, like property crimes, are more likely to be reported and logged, whereas others, like domestic violence, might be underreported.*



The #MeToo movement saw a significant increase in report of sexual assault to law enforcement in Canada.

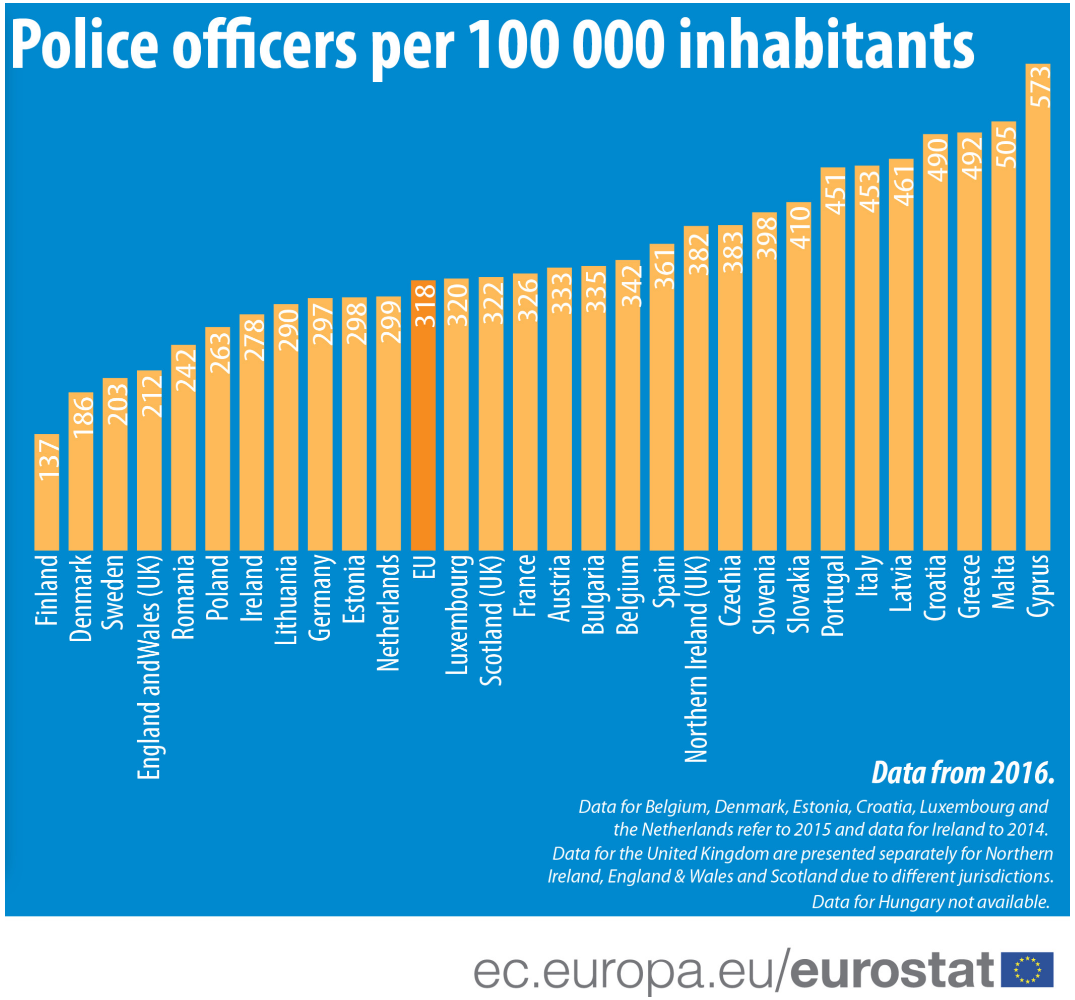
https://www150.statcan.gc.ca/n1/pub/85-002-x/2018001/article/54979-eng.htm

Types of Bias in Time Series Data

Selection Bias

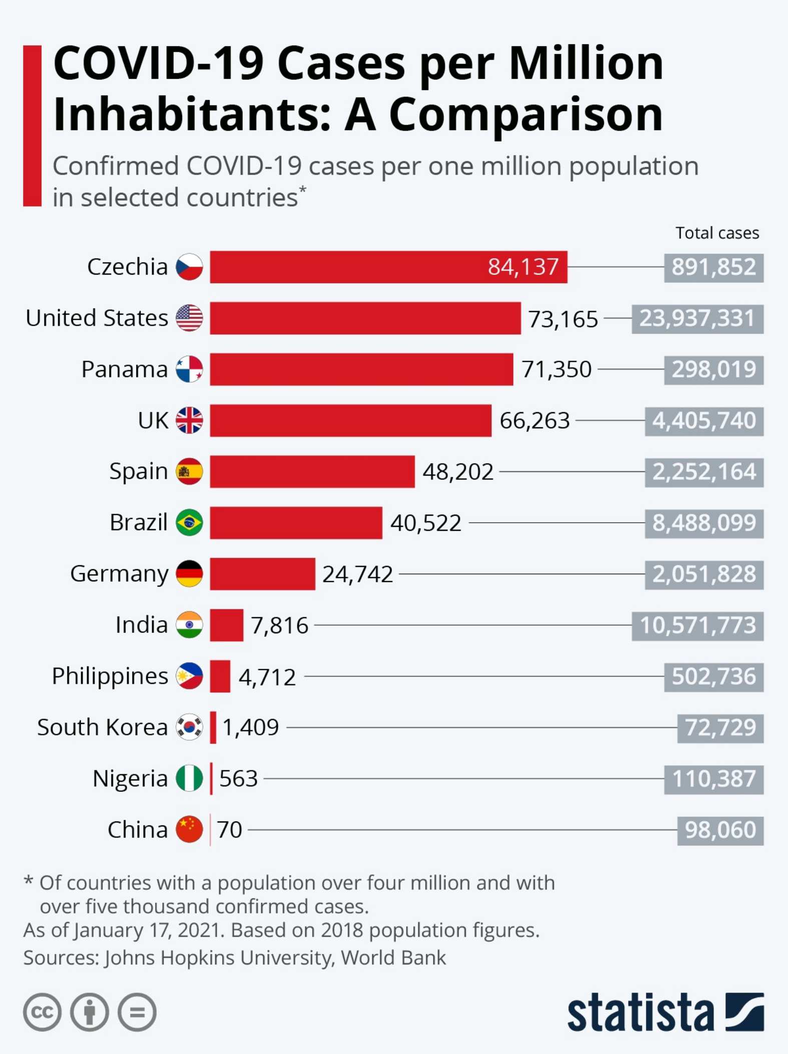
Selection Bias:

- The **selection of the data** for the training dataset of the model is biased, potentially due to unequal distribution of law enforcement resources.
  - *Example: The dataset is skewed towards urban districts with more reported crimes, or towards countries with better integration of data collection in their processes.*



Measurement Bias:

- This might occur if the variables used to **measure** the training data are biased. For instance, measurement of crime risk can be flawed.
  - *Crime data have been recorded inconsistently across different precincts. Some precincts have a history of classifying petty theft and shoplifting under general theft, while others have more specific categorizations. This inconsistency leads to a skewed understanding of crime types across Metropolis City.*

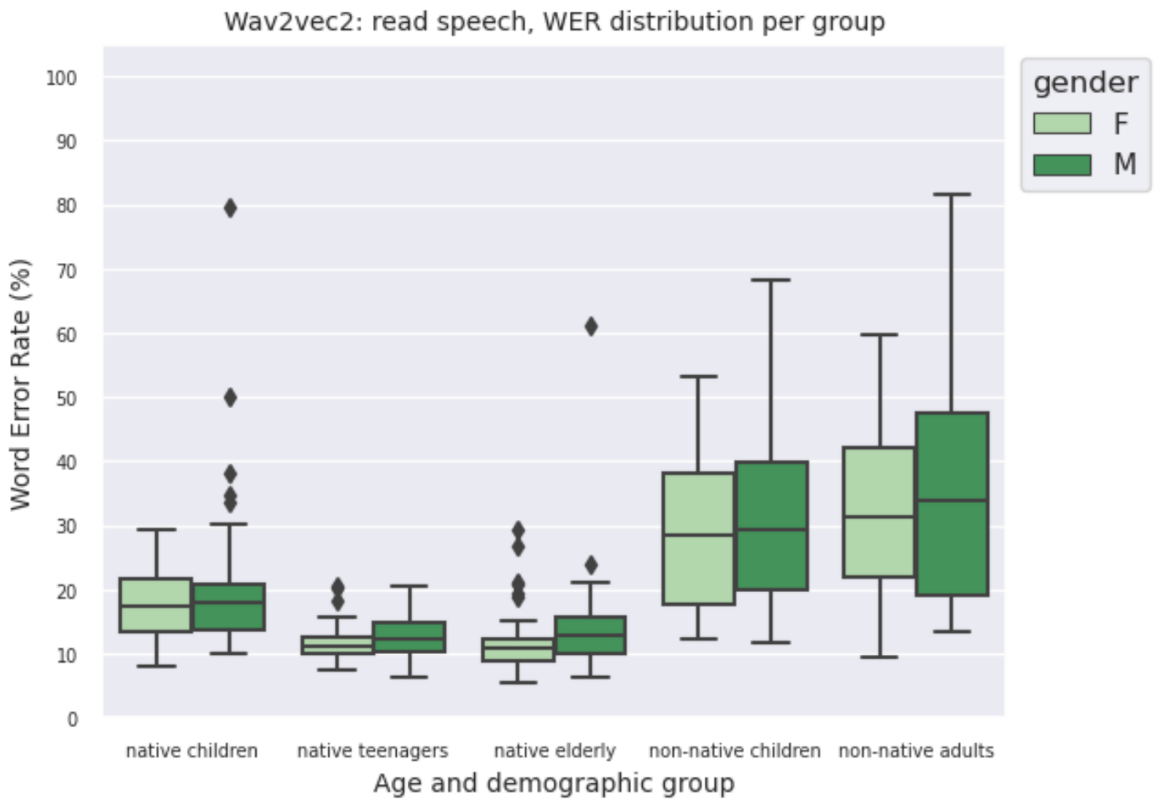


Types of Bias in Time Series Data

Contextual and Socioeconomic Bias

Contextual and Socioeconomic Bias:

- The algorithmic models may not account for the **socioeconomic and contextual factors** that influence crime rates, such as poverty, unemployment, and education levels. These factors can vary widely across different communities and are crucial for understanding crime differently.
  - *High-crime predictions are made for economically disadvantaged areas without considering underlying factors that contribute to crime, such as lack of access to jobs or inadequate housing.*

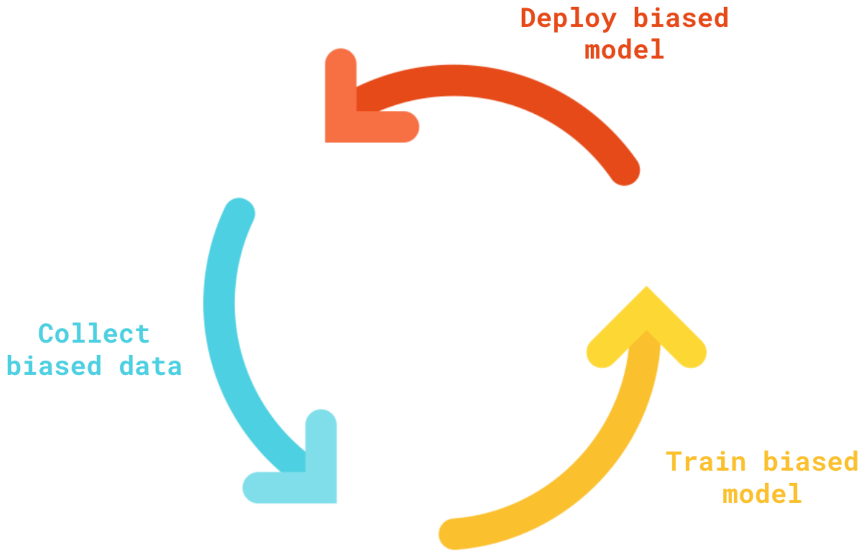


Types of Bias in Time Series Data

Feedback Loop

Feedback Loop:

- Predictive policing can create a feedback loop. Increased police presence in predicted hotspots leads to more recorded offenses, **reinforcing the model's biased predictions** and perpetuating over-policing in those areas.



https://towardsdatascience.com/algorithm-fairness-sources-of-bias-7082e5b78a2c

Ethical Concerns

Fairness, Transparency and Accountability

Fairness in Decision-Making:

- **Reinforcement of Existing Biases**: If the data is biased, the predictions will likely inherit and reinforce these biases, disproportionately targeting marginalized communities.
- **Discriminatory Outcomes**: Where certain groups are unfairly targeted or subjected to increased scrutiny based on biased predictions.

Transparency and Accountability:

- Many predictive policing systems use complex machine learning algorithms that can be difficult to interpret, making it challenging for stakeholders to **understand how decisions are made** and to hold systems accountable for unfair outcomes.

Privacy Concerns:

- **Data Collection and Use**: Predictive policing involves the collection and analysis of **large amounts of data**, which can raise privacy concerns, particularly if individuals' personal information is collected without their consent or used beyond its original purpose.

- **Surveillance**: The use of predictive policing tools can lead to **increased surveillance** of certain communities, infringing on individuals' rights to **privacy and freedom** from unwarranted police attention.

Prohibition of Predictive Policing

Concerns and Criticisms:

- **Data Protection and Privacy:** Extensive data collection raises concerns about legal bases and privacy intrusions**.**
- **Algorithmic Discrimination:** Risk of exacerbating biases and structural injustices within the justice system.
- **Presumption of Innocence:** Predictive identification may prematurely label individuals as potential criminals, conflicting with the presumption of innocence.

Prohibition Under the AI Act (EU):

- Initially considered high-risk AI, now classified as prohibited under Article 5(1)(d) due to concerns outlined in Recital 42, emphasizing the need to protect individual rights and prevent unjust practices.

Strategies to Mitigate Bias and Address Ethical Concerns

Data Sources and Auditing

Diverse and Comprehensive Data Sources:

- Use a **broad spectrum of data sources** beyond traditional crime reports to include community input, socioeconomic factors, and other relevant data that might illuminate the root causes of crime.
- Incorporate qualitative data such as community surveys and interviews to provide **context** that quantitative data alone might miss.

Regular Auditing and Monitoring:

- Regularly **audit algorithms for bias and fairness** by examining the outputs and their impact across different communities and demographics.
- Implement **feedback mechanisms** to continually assess the performance and impact of predictive models on policing practices.

Strategies to Mitigate Bias and Address Ethical Concerns

Transparency and Bias Reduction

Transparent Model Development:

- Develop algorithms with **transparency** in mind, allowing stakeholders to understand how decisions are being made and which **factors influence predictions**.
- Ensure that models are **interpretable** so that law enforcement personnel can **understand** and scrutinize the factors leading to predictions, rather than treating them as **black-box solutions**.

Bias-Reduction Algorithms:

- Employ machine learning techniques specifically designed to **reduce algorithmic bias**, such as fairness constraints during model training or adversarial debiasing.
- Explore **counterfactual fairness approaches** to ensure that changes in sensitive attributes (like race or socioeconomic status) do not unfairly impact predictions.

Strategies to Mitigate Bias and Address Ethical Concerns

Ethical Guidelines and Personnel Training:

- Develop ethical **guidelines** and **provide training** for law enforcement on the responsible use of technology, emphasizing awareness of potential biases and their impacts.

Scenario-Based Testing:

- Conduct scenario-based simulations to evaluate how algorithms behave under different conditions and to understand potential biases in predictions.

Adversarial Debiasing

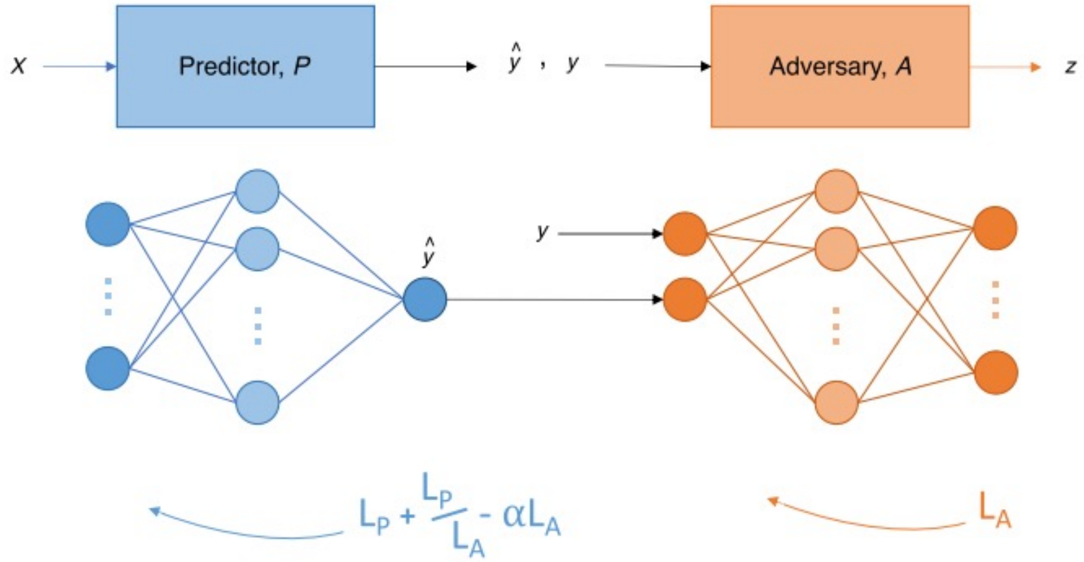
Using **adversarial training**, a process where two models (or components) are trained **simultaneously** with **opposing** objectives, can reduce bias in the predictions of a primary model.

- **Primary Model (Predictor):** The main machine learning model, trained to perform a task such as classification or regression. Initially, this model may produce biased predictions based on the training data.
- **Adversarial Model:** A secondary model, often called the adversary or discriminator, is trained simultaneously with the primary model. The goal of the adversarial model is to predict or identify the sensitive attributes (such as race, gender, etc.) from the outputs of the primary model or its intermediate representations.

Adversarial Debiasing

Training Process:

- During training, the **primary** model is optimized to achieve **high performance on its main task** (e.g., classification), while the **adversarial** model tries to accurately **predict the sensitive attribute** from the primary model's outputs.
- The primary model is encouraged to produce **representations** that make it **difficult** for the adversarial model to correctly **identify the sensitive attribute**. It forces the primary model to limit the information about sensitive attributes used in its predictions.



Yang, J., Soltan, A.A.S., Eyre, D.W. *et al.* An adversarial training framework for mitigating algorithmic biases in clinical machine learning. *npj Digit. Med.* **6**, 55 (2023). https://doi.org/10.1038/s41746-023-00805-y

Bias: Conclusion

- **Real world applications can be subject to bias from multiple sources**

- **This can lead to ethical concerns for sensitive applications**

- **Strategies can be implemented to mitigate this bias**