

# Different Horses for Different Courses: Comparing Bias Mitigation Algorithms in ML

Prakhar Ganesh<sup>1,2\*</sup>, Usman Gohar<sup>3\*</sup>, Lu Cheng<sup>4</sup>, Golnoosh Farnadi<sup>1,2</sup>

<sup>1</sup>Mila, <sup>2</sup>McGill University, <sup>3</sup>Iowa State University, <sup>4</sup>University of Illinois Chicago

\*Equal contribution

# Meet the racers!



# Meet the racers!



**Neighsayer**

# Meet the racers!



**Neighsayer**



# Meet the racers!



**Neighsayer**



**Gavin**

# Meet the racers!

**Wins most races.**



**Neighsayer**



**Gavin**

# Meet the racers!

**Wins most races.**



**Neighsayer**

**Wins longer races.**



**Gavin**

***“The race track acts as a lens that selectively highlights certain strengths and obscures others.”***

***“The race track acts as a lens that selectively highlights certain strengths and obscures others.”***

**Races in ML? Benchmarking frameworks!**

*“The race track acts as a lens that selectively highlights certain strengths and obscures others.”*

**Races in ML? Benchmarking frameworks!**

*“The **benchmarking context** acts as a lens that selectively highlights certain strengths and obscures others.”*

# Regulation for Fair ML

## EU AI Act: first regulation on artificial intelligence

The use of artificial intelligence in the EU will be regulated by the AI Act, the world's first comprehensive AI law. Find out how it will protect you.

<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

## BLUEPRINT FOR AN AI BILL OF RIGHTS

MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE

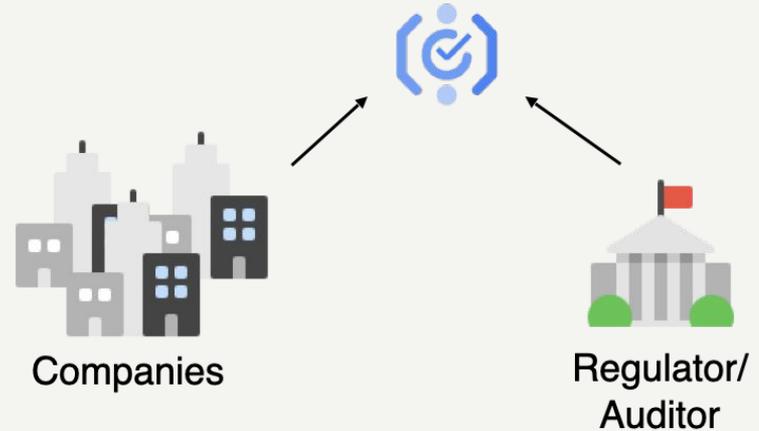
OSTP

<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

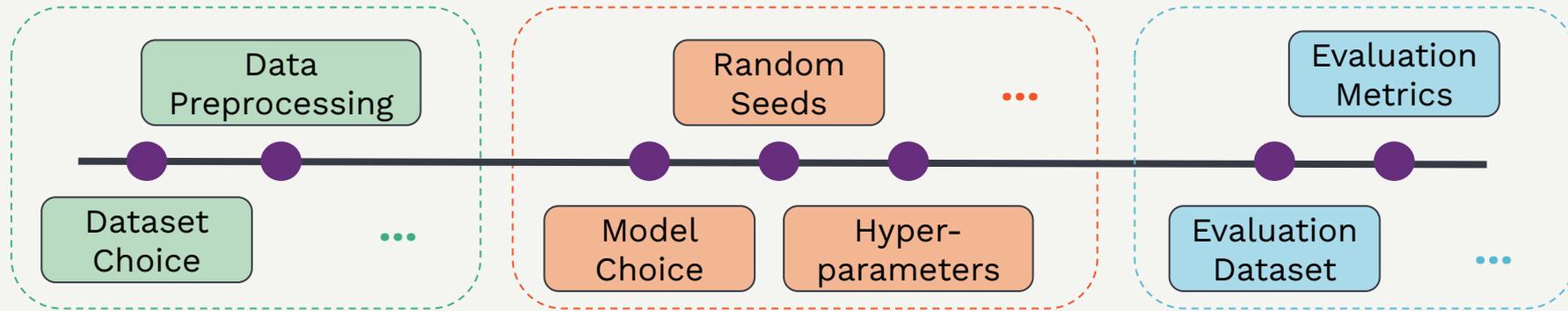
## AI Watch: Global regulatory tracker

<https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker#home>

Compliance?



# ML Pipeline



# FFB: Fair Fairness Benchmark

# FFB: Fair Fairness Benchmark

***“Fairness methods can be hindered by variations in dataset preprocessing and the use of different backbones. These discrepancies can lead to inconsistent and unfair comparison”***

# FFB: Fair Fairness Benchmark

***“Fairness methods can be hindered by variations in dataset preprocessing and the use of different backbones. These discrepancies can lead to inconsistent and unfair comparison”***

**Recommended Solution: Using a uniform evaluation setup**

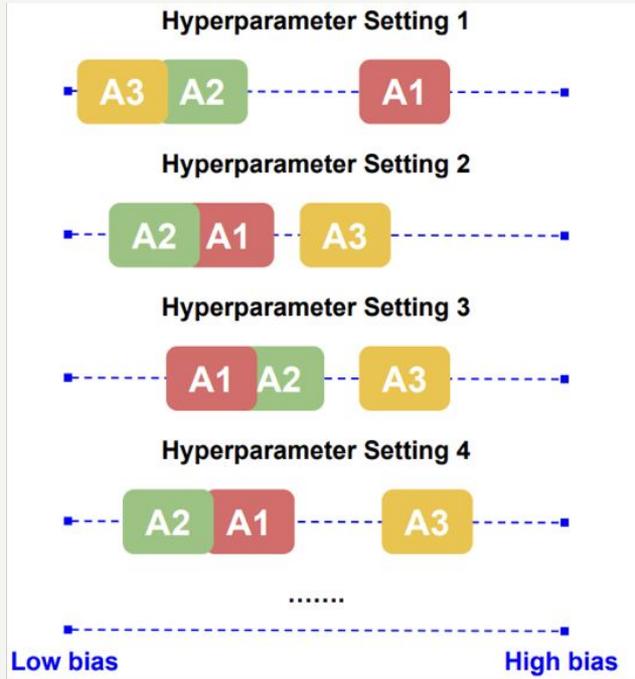
# FFB: Fair Fairness Benchmark

***“Fairness methods can be hindered by variations in dataset preprocessing and the use of different backbones. These discrepancies can lead to inconsistent and unfair comparison”***

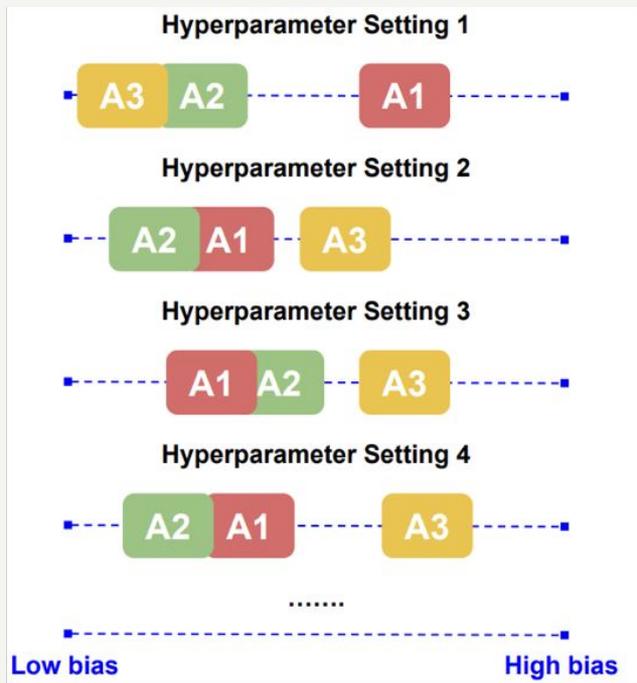
**Recommended Solution: Using a uniform evaluation setup**

***But is that really fair?***

# Towards Context Dependent Evaluation

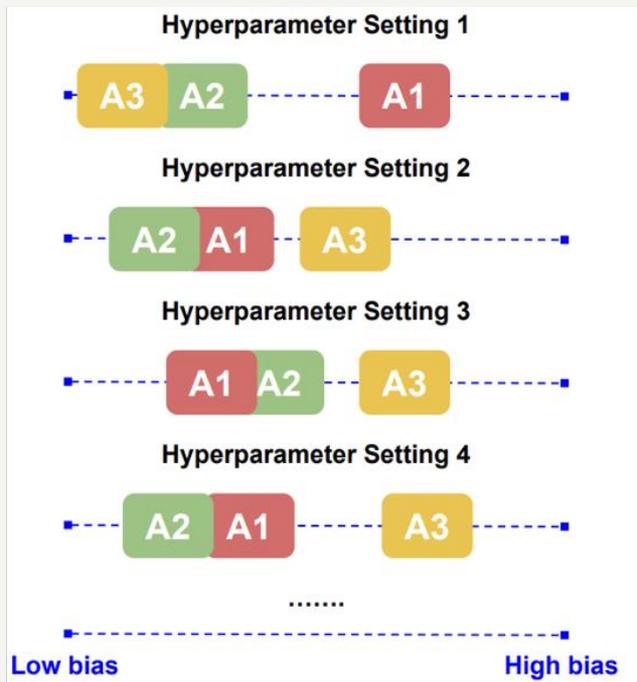


# Towards Context Dependent Evaluation



What is the best method if the developer is constrained to hyperparameter setting 3?

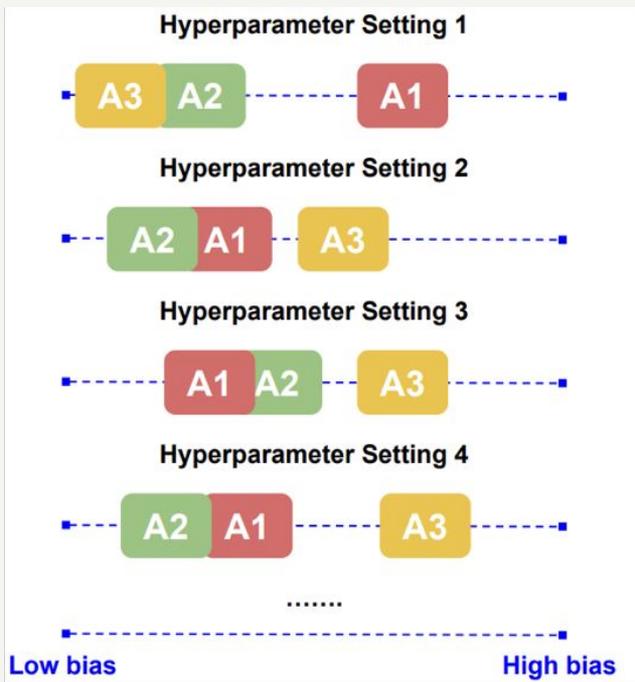
# Towards Context Dependent Evaluation



What is the best method if the developer is constrained to hyperparameter setting 3?

A1

# Towards Context Dependent Evaluation

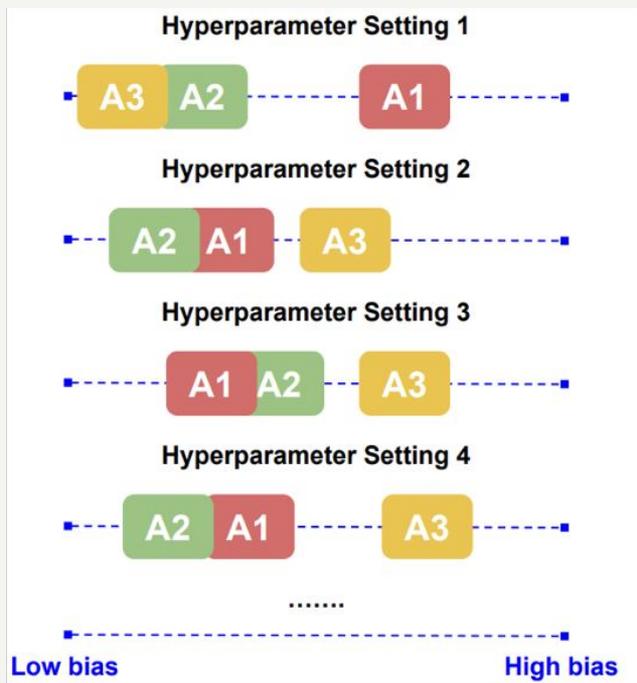


What is the best method if the developer is constrained to hyperparameter setting 3?

A1

What is the best method if the developer can use any hyperparameter?

# Towards Context Dependent Evaluation



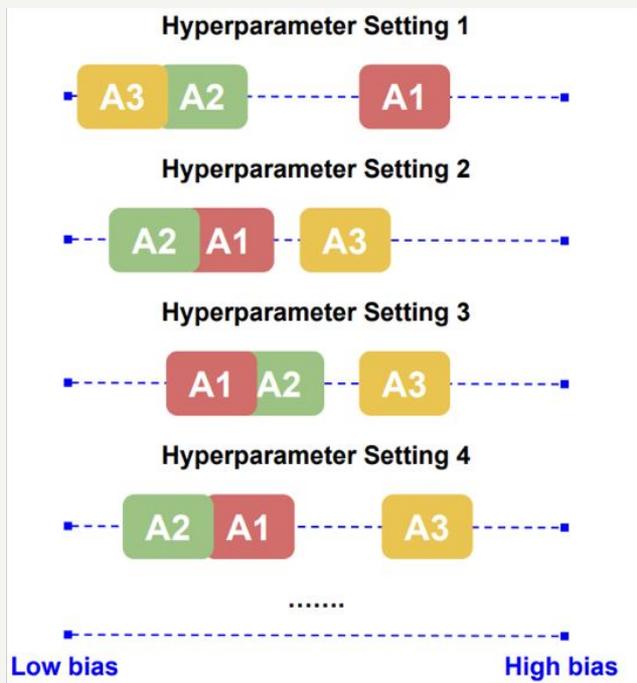
What is the best method if the developer is constrained to hyperparameter setting 3?

A1

What is the best method if the developer can use any hyperparameter?

A3

# Towards Context Dependent Evaluation



What is the best method if the developer is constrained to hyperparameter setting 3?

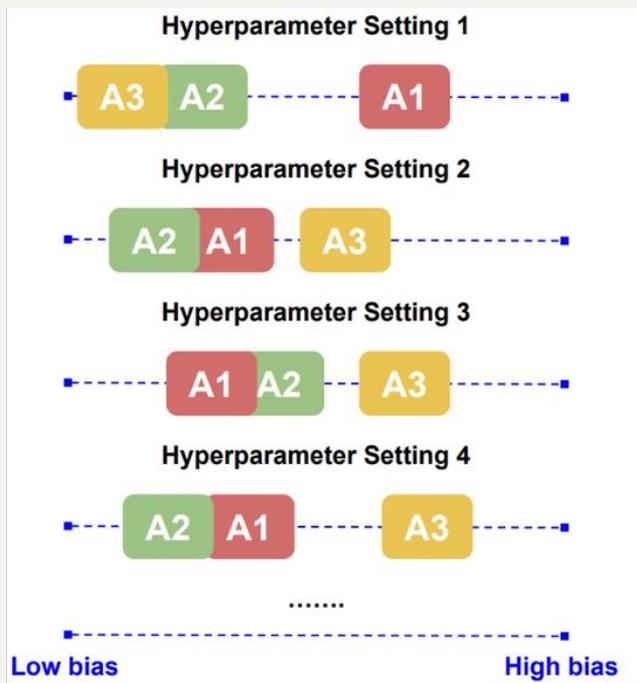
A1

What is the best method if the developer does not know the hyperparameter setting in advance, and thus wants the most robust technique?

What is the best method if the developer can use any hyperparameter?

A3

# Towards Context Dependent Evaluation



What is the best method if the developer is constrained to hyperparameter setting 3?

A1

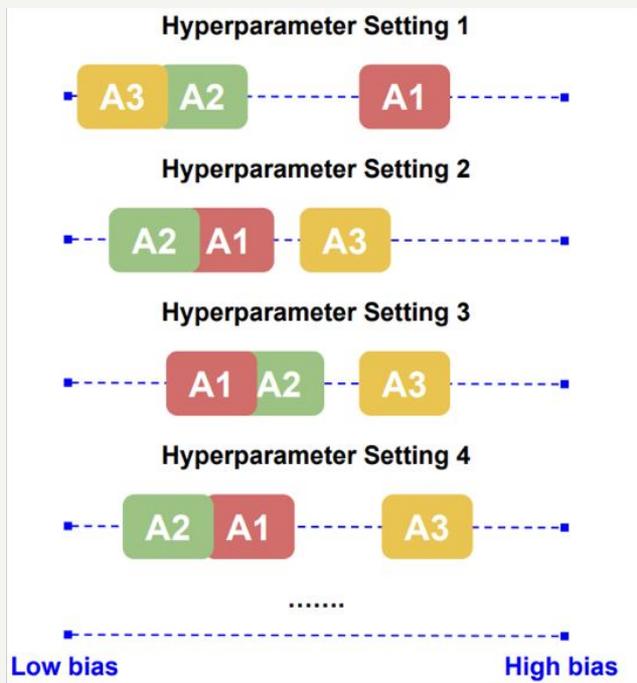
What is the best method if the developer does not know the hyperparameter setting in advance, and thus wants the most robust technique?

A2

What is the best method if the developer can use any hyperparameter?

A3

# Towards Context Dependent Evaluation



What is the best method if the developer is constrained to hyperparameter setting 3?

A1

What is the best method if the developer does not know the hyperparameter setting in advance, and thus wants the most robust technique?

A2

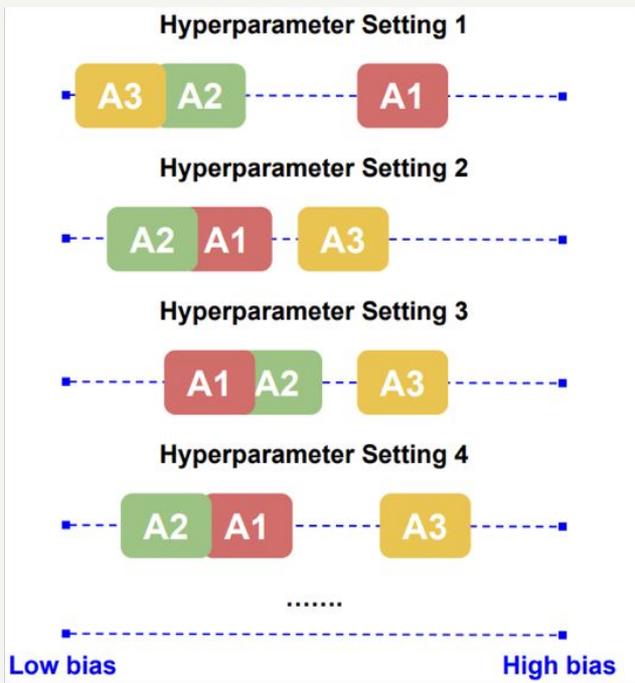
What is the best method if the developer can use any hyperparameter?

A3

What is the best method if ...

?

# Towards Context Dependent Evaluation



What is the best method if the developer is constrained to hyperparameter setting 3?

A1

What is the best method if the developer does not know the hyperparameter setting in advance, and thus wants the most robust technique?

A2

What is the best method if the developer can use any hyperparameter?

A3

What is the best method if ...

?

No method is the "best" method in every context!

# Empirical Results: Some Takeaways

# Empirical Results: Some Takeaways

- **No method provides the best fairness-utility tradeoff for every single set of hyperparameters or dataset.**  
**In fact, we find some techniques that outperform all others in one setting but do not even converge in a different setting.**

# Empirical Results: Some Takeaways

- **No method provides the best fairness-utility tradeoff for every single set of hyperparameters or dataset.**  
**In fact, we find some techniques that outperform all others in one setting but do not even converge in a different setting.**
  
- **Given the opportunity to perform hyperparameter optimization, most mitigation algorithms can provide competitive score. This gives us the opportunity to look beyond the fairness-utility tradeoff and improve other meta-objectives!**

**Current fairness evaluation practices are broken, creating  
*'perceived superiority'* of certain mitigation techniques over others.**

**Current fairness evaluation practices are broken, creating ‘*perceived superiority*’ of certain mitigation techniques over others.**

**High fairness variance exacerbates the problem, echoing the call for context-aware evaluations and more holistic frameworks.**

**Current fairness evaluation practices are broken, creating *'perceived superiority'* of certain mitigation techniques over others.**

**High fairness variance exacerbates the problem, echoing the call for context-aware evaluations and more holistic frameworks.**

**Come talk to us @ Poster Session**

**Paper**



**\* Prakhar Ganesh**  
Mila, McGill University  
prakhar.ganesh@mila.quebec



**\* Usman Gohar**  
Iowa State University  
ugohar@iastate.edu



**Lu Cheng**  
University of Illinois Chicago  
lucheng@uic.edu



**Golnoosh Farnadi**  
Mila, McGill University  
farnadig@mila.quebec

