IEEE BIG DATA 2024
December 15-18 • Washington DC, USA

Baylor University

UNIVERSITY OF ARKANSAS

# Supervised Algorithmic Fairness in Distribution Shifts

Tutorial at the 2024 IEEE International Conference on Big data (IEEE BigData 2024)


**Chen Zhao**[1]
chen_zhao@baylor.edu

**Xintao Wu**[2]
xintaowu@uark.edu

[1]Department of Computer Science, Baylor University
[2]Department of Electrical Engineering and Computer Science, University of Arkansas


**December 15th, 2024, Washington D.C., USA**

# Tutorial Website and Survey

**IEEE BigData 2024**
**Tutorial Website**

**IJCAI 2024**
**Survey Paper**

1. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. “Supervised Algorithmic Fairness in Distribution Shifts: A Survey. ”, IJCAI 2024.

# Outline

**Part I:** Introduction to Fairness

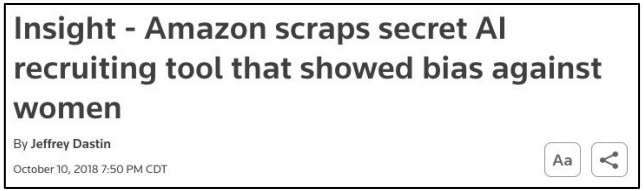
**Part II:** Distribution Shift Undermines Fairness

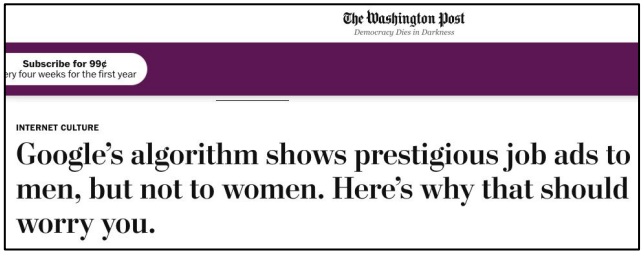
**Part III:** Mitigating Unfairness under Distribution Shift (Offline)

**Part IV:** Mitigating Unfairness under Distribution Shift (Online)

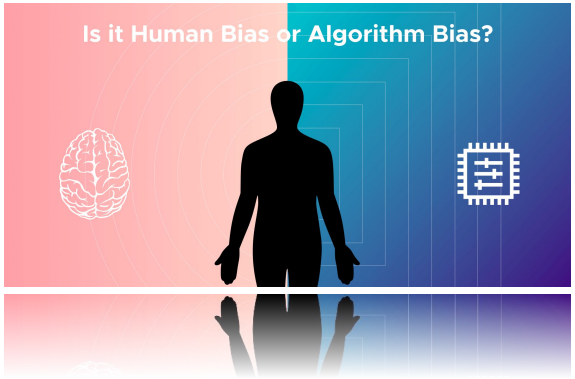
**Part V:** Open Challenges and Beyond

# Why Fairness in Machine Learning is Important?

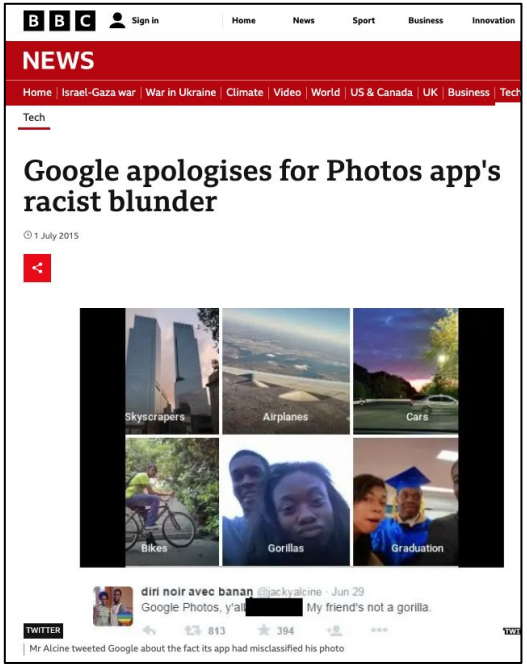






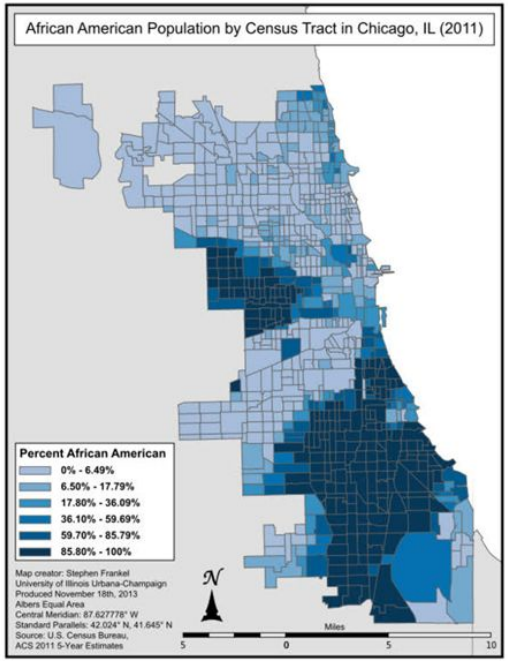




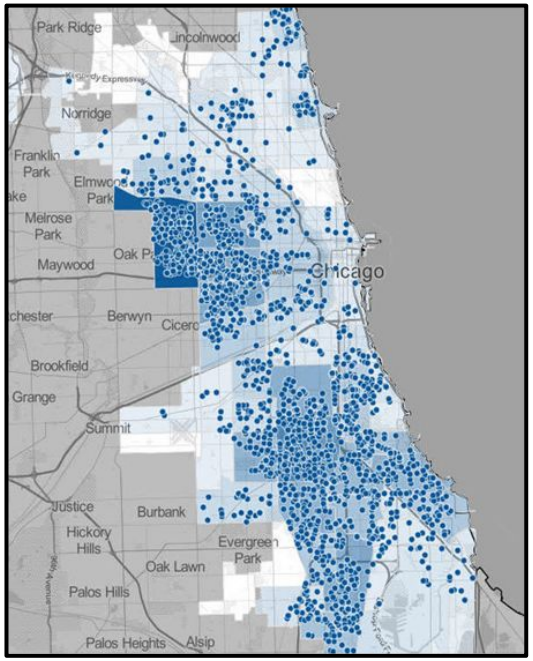


1. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/
2. https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html
3. https://www.bbc.com/news/technology-33347866
4. https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/
5. https://medium.com/si-410-ethics-and-information-technology/tiktoks-addictive-and-unethical-algorithm-3f44f41f1f3c

# Example 1 - Crime Prediction




**Black neighborhoods broken down by community areas**




**Crime counts prediction**

**Sensitive Features**: Race
**Predictions**: Crime Counts

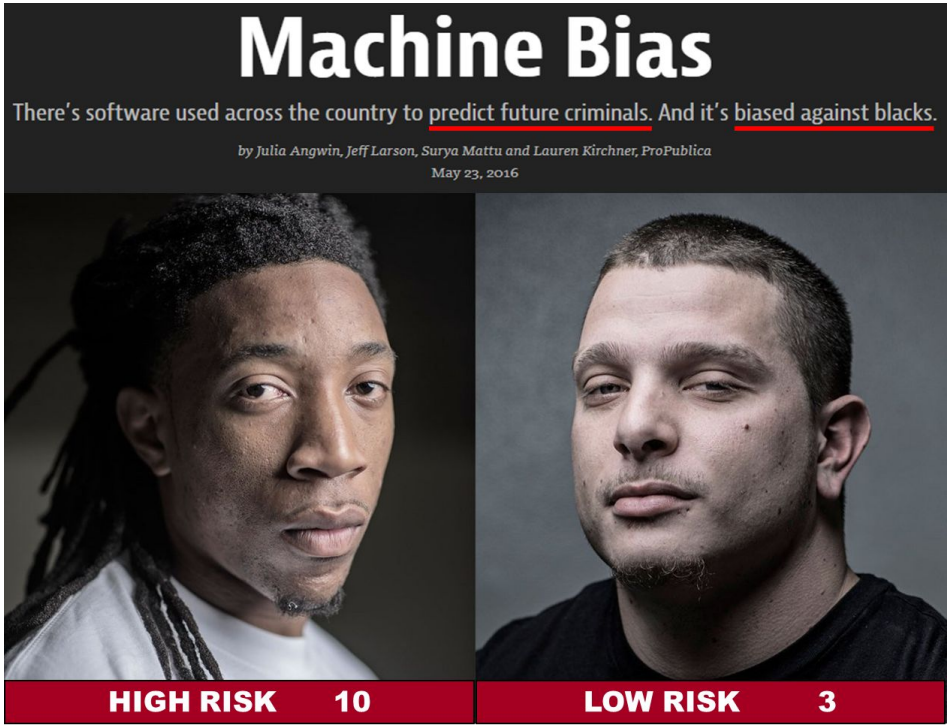
NEWS

**Crime in Chicago: Explore your community**

By CHICAGO TRIBUNE
UPDATED: April 1, 2019 at 11:40 AM CST

1. Image source: https://www.chicagotribune.com/2019/04/01/crime-in-chicago-explore-your-community/

# Example 2 - Recidivism Forecasting




**Sensitive Features**: Race
**Predictions**: Risk Level

The black man on the left was predicted as high risk with 1 prior offense and no subsequent offenses.

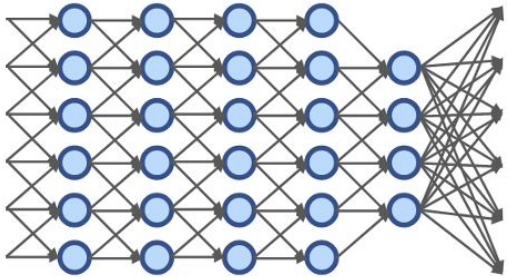
However, the white man on the right was rated as low risk, even though he was arrested 3 times on drug charges after 1 prior offense.

1. Image source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

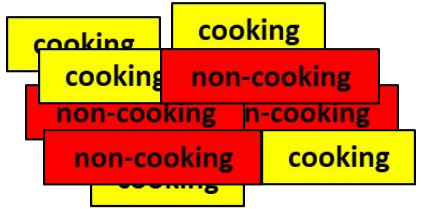
# Example 3 - Image Classification

**Training images of People**


**Deep Learning Model** $f(\mathbf{x})$



**Predictive Outcomes**




**Sensitive Features**: Gender
**Predictions**: Cooking / non-Cooking

| COOKING | |
|---|---|
| ROLE | VALUE |
| AGENT | WOMAN |
| FOOD | PASTA |
| HEAT | STOVE |
| TOOL | SPATULA |
| PLACE | KITCHEN |

| COOKING | |
|---|---|
| ROLE | VALUE |
| AGENT | WOMAN |
| FOOD | FRUIT |
| HEAT | -- |
| TOOL | KNIFE |
| PLACE | KITCHEN |

| COOKING | |
|---|---|
| ROLE | VALUE |
| AGENT | WOMAN |
| FOOD | MEAT |
| HEAT | STOVE |
| TOOL | SPATULA |
| PLACE | OUTSIDE |

| NON-COOKING | |
|---|---|
| ROLE | VALUE |
| AGENT | MAN |
| FOOD | -- |
| HEAT | STOVE |
| TOOL | SPATULA |
| PLACE | KITCHEN |

| NON-COOKING | |
|---|---|
| ROLE | VALUE |
| AGENT | MAN |
| FOOD | -- |
| HEAT | STOVE |
| TOOL | SPATULA |
| PLACE | KITCHEN |

1. Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, Kai-Wei Chang. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints, EMNLP, 2017

# Algorithmic Fairness


*X*: images
*Z*: {Male, Female}
*Y*: {Cooking, non-Cooking}

**For example:**
*Source Domain*: All Male images are non-Cooking and Female are Cooking.

*Target Domain*: How to ensure a classifier that predicts image labels independent of gender?

**In fairness-aware machine learning, we mitigate the dependence:**

**Sensitive Features ($Z$)**
***e.g.*, Male / Female**

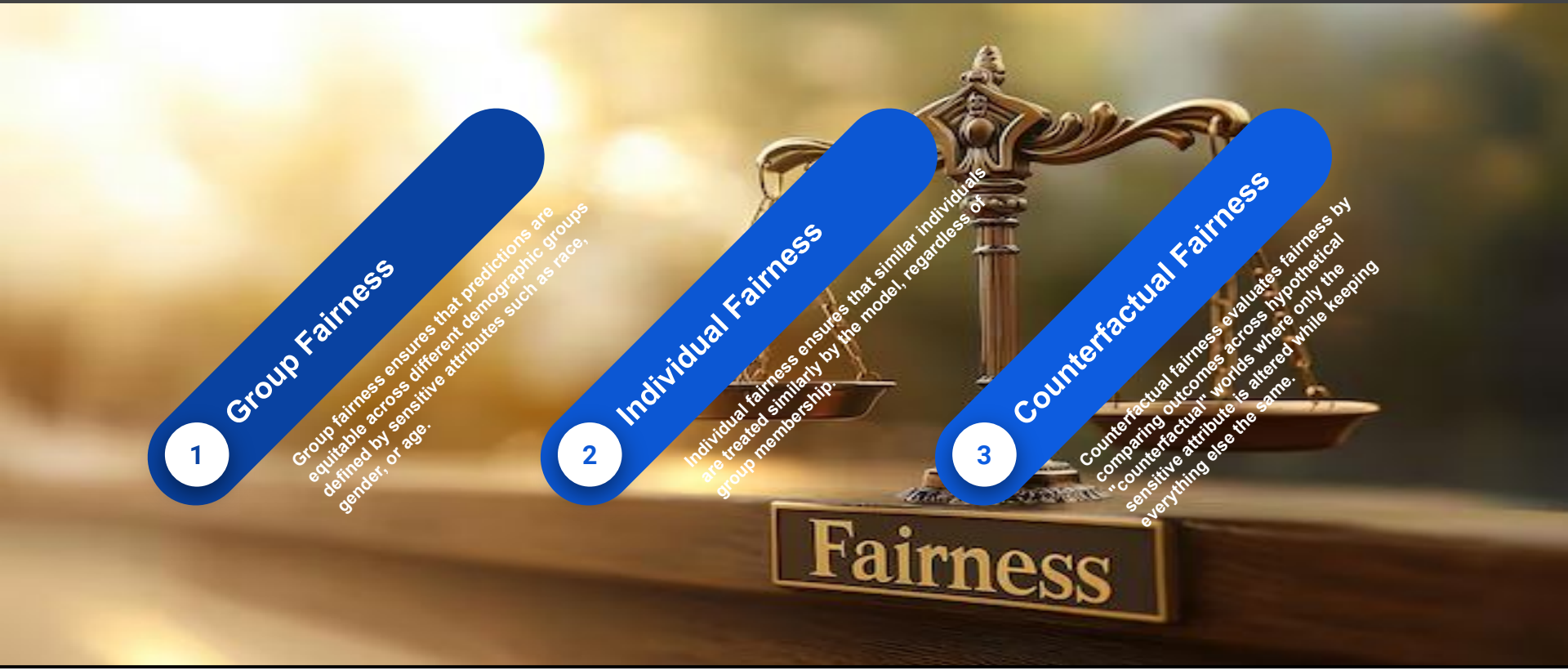
→ spuriously correlated to →

**Predictive Outcomes ($\hat{Y} = f_\theta(X)$)**
***e.g.*, Cooking / non-Cooking**

**Algorithmic Fairness:**

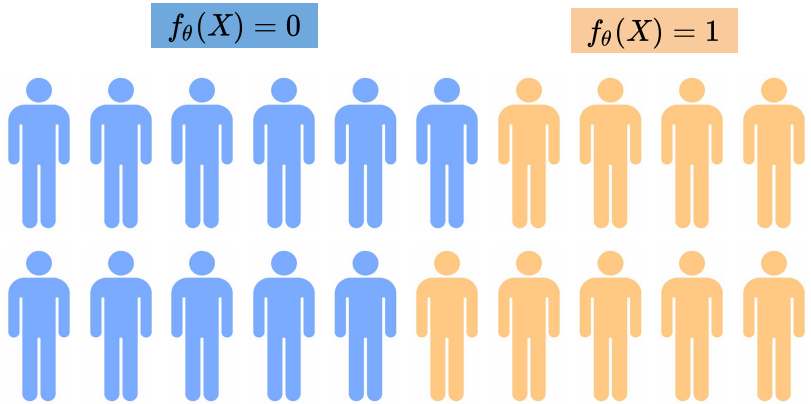
**The spurious dependence between sensitive features to predictive model outcomes is eliminated or restricted to an accepted fairness level.**

# Definition of Fairness



# Group Fairness (Demographic Parity)

**Demographic Parity (DP)** ensures a model's predictions are independent of sensitive attributes. It requires that individuals from different demographic groups have an equal probability of receiving a positive outcome ( $\hat{Y} = f_\theta(X) = 1$ ).




$$\mathbb{P}(f_\theta(X) = 1 | Z = 0) = 4/10 = 0.4$$

$$\mathbb{P}(f_\theta(X) = 1 | Z = 1) = 5/10 = 0.5$$

**Demographic Parity (Ratio):** $\frac{\mathbb{P}(f_\theta(X)=1|Z=0)}{\mathbb{P}(f_\theta(X)=1|Z=1)} = \frac{0.4}{0.5} = 0.8 \geq 0.8$

**(80% rule) [D. Biddle. 2017]**

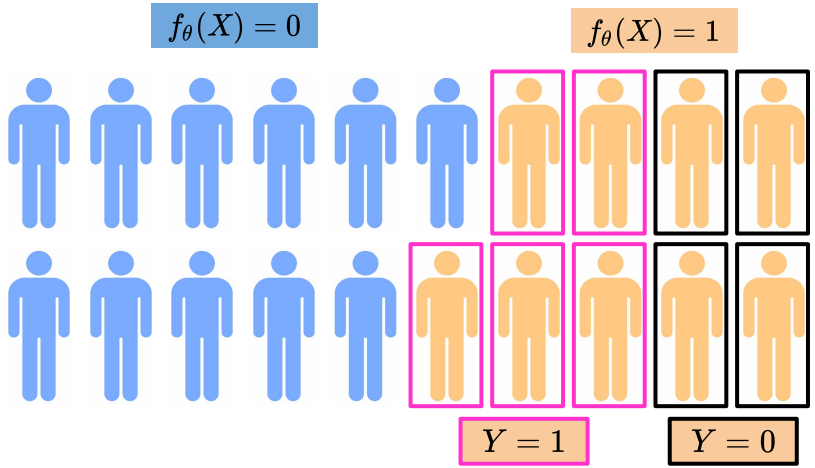
**Difference of Demographic Parity (DDP):** $|\mathbb{P}(f_\theta(X) = 1|Z = 0) - \mathbb{P}(f_\theta(X) = 1|Z = 1)| = 0.1$

**The closer DDP is to 0, the fairer.**

1. Dan Biddle. Adverse Impact and Test Validation: A Practitioner's Guide to Valid and Defensible Employment Testing. Routledge, 2017.

# Group Fairness (Equalized Odds)

**Equalized odds (EO)** ensures a model's predictions are equally accurate across different demographic groups. It focuses on parity of error rates between sensitive groups, for a binary classification task. Specifically, it requires that the model's true positive rate (TPR) and false positive rate (FPR) be the same across all groups.




**Privileged Group (*e.g.*, Male)**

$$\mathbb{P}(f_\theta(X) = 1 | Y = 1, Z = 0) = 0.2$$
$$\mathbb{P}(f_\theta(X) = 1 | Y = 0, Z = 0) = 0.2$$

**Unprivileged Group (*e.g.*, Female)**

$$\mathbb{P}(f_\theta(X) = 1 | Y = 1, Z = 1) = 0.3$$
$$\mathbb{P}(f_\theta(X) = 1 | Y = 0, Z = 1) = 0.2$$

**Difference of Equalized Odds (DEO):**

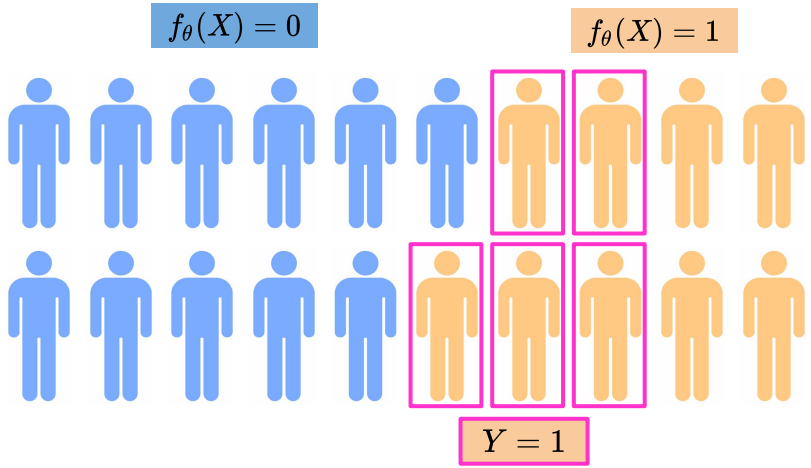
$$\sum_{y \in \{0,1\}} |\mathbb{P}(f_\theta(X) = 1 | Y = y, Z = 0) - \mathbb{P}(f_\theta(X) = 1 | Y = y, Z = 1)| = 0.1$$

**The closer DEO is to 0, the fairer.**

1. Moritz Hardt , Eric Price, Nathan Srebro. Equality of Opportunity in Supervised Learning. NeurIPS, 2016.

# Group Fairness (Equal Opportunity)

**Equal Opportunity (EOp)** focuses on ensuring equal True Positive Rates (TPR) across different groups defined by a sensitive attribute. This means that individuals who belong to different sensitive groups but have the same actual positive outcome should have an equal chance of being correctly classified as positive by the model.




**Privileged Group (*e.g.*, Male)**

$$\mathbb{P}(f_\theta(X) = 1 | Y = 1, Z = 0) = 0.2$$

**Unprivileged Group (*e.g.*, Female)**

$$\mathbb{P}(f_\theta(X) = 1 | Y = 1, Z = 1) = 0.3$$

**Difference of Equal Opportunity (DEOp):**

$$|\mathbb{P}(f_\theta(X) = 1 | Y = 1, Z = 0) - \mathbb{P}(f_\theta(X) = 1 | Y = 1, Z = 1)| = 0.1$$

**The closer DEOp is to 0, the fairer.**

1. Moritz Hardt , Eric Price, Nathan Srebro. Equality of Opportunity in Supervised Learning. NeurIPS, 2016.

# Linear Relaxation of DDP and DEOp

**DDP (Difference of Demographic Parity)** and **DEOp (Difference of Equal Opportunity)** are fairness metrics that measure disparities in outcomes or opportunities between sensitive groups. When addressing these fairness criteria in optimization problems, exact fairness constraints are often non-convex or difficult to handle directly. **Linear relaxation** refers to approximating these constraints in a linear form to make them tractable in optimization settings.

Re-write

DDP: $\left|\frac{1}{n_0}\sum_{(\mathbf{x},z,y)\in\mathcal{D}_0}\mathbb{I}_{f_\theta(\mathbf{x})>0}-\frac{1}{n_1}\sum_{(\mathbf{x},z,y)\in\mathcal{D}_1}\mathbb{I}_{f_\theta(\mathbf{x})>0}\right|\leq\tau$

DEOp: $\left|\frac{1}{n_0}\sum_{(\mathbf{x},z,y)\in\mathcal{D}_0}\mathbb{I}_{f_\theta(\mathbf{x})>0,Y=1}-\frac{1}{n_1}\sum_{(\mathbf{x},z,y)\in\mathcal{D}_1}\mathbb{I}_{f_\theta(\mathbf{x})>0,Y=1}\right|\leq\tau$

Linear Relaxation

$$\mathbb{E}_{(\mathbf{x},z,y)\sim\mathcal{D}}\left[\frac{1}{\hat{p}_1(1-\hat{p}_1)}\left(\frac{z+1}{2}-\hat{p}_1\right)\mathbb{I}_{f_\theta(\mathbf{x})>0}\right]\leq\tau$$

$$\left|\frac{1}{n}\sum_{(\mathbf{x},z,y)\in\mathcal{D}}\frac{1}{\hat{p}_1(1-\hat{p}_1)}\left(\frac{z+1}{2}-\hat{p}_1\right)f_\theta(\mathbf{x})>0\right|\leq\tau$$

**For DDP:**

$\hat{p}_1=\mathbb{P}_{(\mathbf{x},z,y)\sim\mathcal{D}}(z=1)$

**For DEOp:**

$\hat{p}_1=\mathbb{P}_{(\mathbf{x},z,y)\sim\mathcal{D}}(y=1,z=1)$

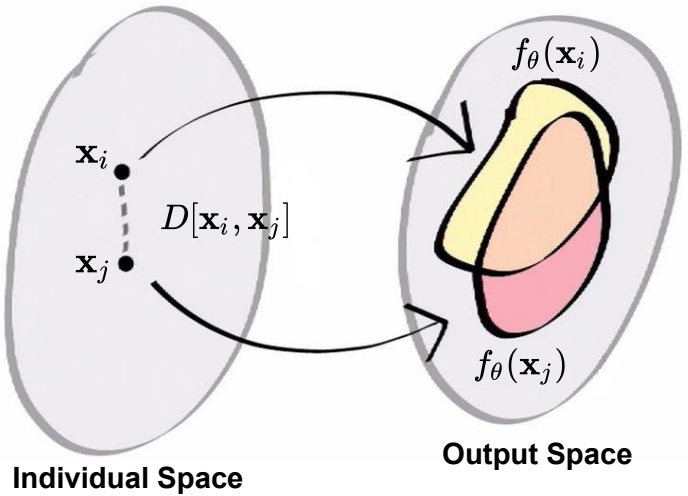

1. Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, Krishna P. Gummadi. Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment. WWW, 2017.
2. Michael Lohaus, Michael Perrot, Ulrike Von Luxburg. Too Relaxed to Be Fair. ICML, 2020.

# Individual Fairness

**Individual fairness** emphasizes **treating similar individuals similarly**. It ensures that two individuals who are similar in terms of relevant features receive similar predictions or outcomes from the model.

| Search Query | Work Experience | Education Experience | Candidate | Xing Ranking |
|---|---|---|---|---|
| Brand Strategist | 146 | 57 | male | 1 |
| Brand Strategist | 327 | 0 | female | 2 |
| Brand Strategist | 502 | 74 | male | 3 |
| Brand Strategist | 444 | 56 | female | 4 |
| Brand Strategist | 139 | 25 | male | 5 |
| Brand Strategist | 110 | 65 | female | 6 |
| Brand Strategist | 12 | 73 | male | 7 |
| Brand Strategist | 99 | 41 | male | 8 |
| Brand Strategist | 42 | 51 | female | 9 |
| Brand Strategist | 220 | 102 | female | 10 |
| | | ... | | |
| Brand Strategist | 3 | 107 | female | 20 |
| Brand Strategist | 123 | 56 | female | 30 |
| Brand Strategist | 3 | 3 | male | 40 |

The top-10 results satisfy group fairness with regard to gender. However the outcomes are far from being fair for the individual users: people with very similar qualifications, such as *Work Experience* and *Education Score* ended up on ranks that are far apart (*e.g.,* ranks 5 and 30).

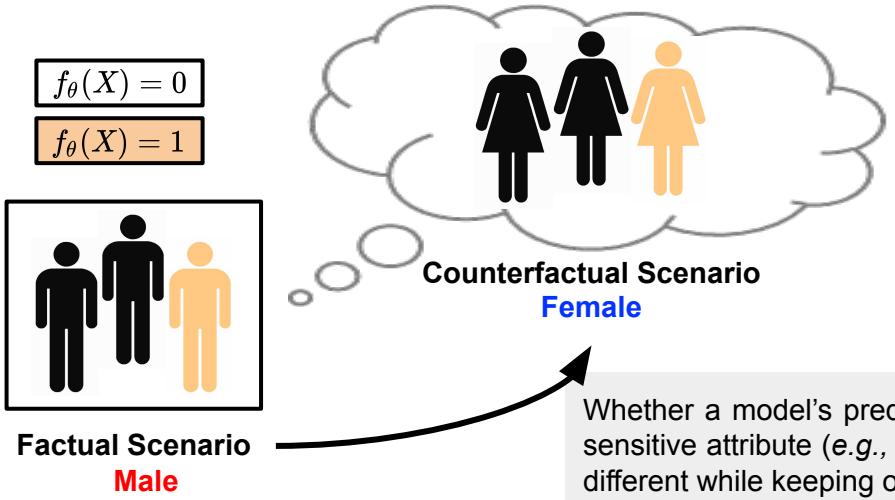



$$d\left[f_\theta(\mathbf{x}_j), f_\theta(\mathbf{x}_j)\right] \leq \delta \cdot D\left[\mathbf{x}_i, \mathbf{x}_j\right], \ \forall i \neq j$$


1. Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, Rich Zemel. Fairness Through Awareness. ITCS, 2012.
2. Preethi Lahoti, Krishna P. Gummadi, Gerhard Weikum. iFair: Learning Individually Fair Data Representations for Algorithmic Decision Making. ICDE, 2019.

# Counterfactual Fairness

**Counterfactual Fairness** ensures a model's predictions for an individual are not influenced by changes to a sensitive attribute in a counterfactual scenario. This concept originates from causal inference and is designed to capture how decisions would change (or not) if a person's sensitive attribute were different while keeping all other aspects of their situation the same.




1. Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. NeurIPS, 2017

# Counterfactual Fairness

The **Total Causal Effect (TCE)** measures the overall influence of sensitive attributes on the model's prediction.

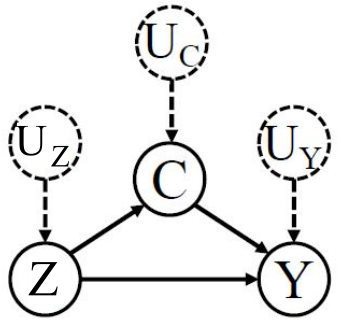
Given context $O = \mathbf{o}$, the **Counterfactual Effect (CE)** measures the influence of sensitive attributes on the model prediction specified on the sub-population by the context only.

These two metrics quantify how much the prediction would change if the sensitive attribute were altered, while allowing all other variables that depend on it to adjust accordingly through the causal pathways.

$$\mathrm{TCE} = \left|\mathbb{P}(f_{\theta,Z\leftarrow 0}(X) = y) - \mathbb{P}(f_{\theta,Z\leftarrow 1}(X) = y)\right|,\ \forall y \qquad \mathrm{CE} = \left|\mathbb{P}(f_{\theta,Z\leftarrow 0}(X) = y | O = \mathbf{o}) - \mathbb{P}(f_{\theta,Z\leftarrow 1}(X) = y | O = \mathbf{o})\right|,\ \forall y$$

Table 1: Connection between previous fairness notions and PC fairness

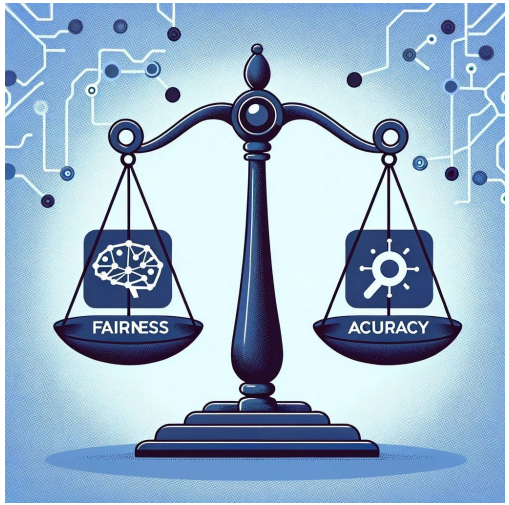
| Description | Relating to PC fairness |
|---|---|
| Total effect | $\mathbf{O} = \emptyset$ and $\pi = \Pi$ |
| (System) Direct discrimination | $\mathbf{O} = \emptyset$ or $\{Z\}$ and $\pi = \pi_d = \{Z \to \hat{Y}\}$ |
| (System) Indirect discrimination | $\mathbf{O} = \emptyset$ or $\{Z\}$ and $\pi = \pi_i \subset \Pi$ |
| Individual direct discrimination | $\mathbf{O} = \{Z, \mathbf{X}\}$ and $\pi = \pi_d = \{Z \to \hat{Y}\}$ |
| Group direct discrimination | $\mathbf{O} = \mathbf{Q} = \mathrm{PA}_Y \backslash \{Z\}$ and $\pi = \pi_d = \{Z \to \hat{Y}\}$ |
| Counterfactual fairness | $\mathbf{O} = \{Z, \mathbf{X}\}$ and $\pi = \Pi$ |
| Counterfactual error rate | $\mathbf{O} = \{Z, Y\}$ and $\pi = \pi_d$ or $\pi_i$ |

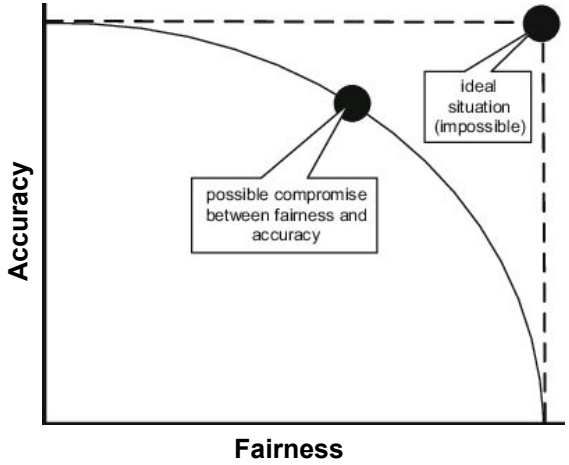


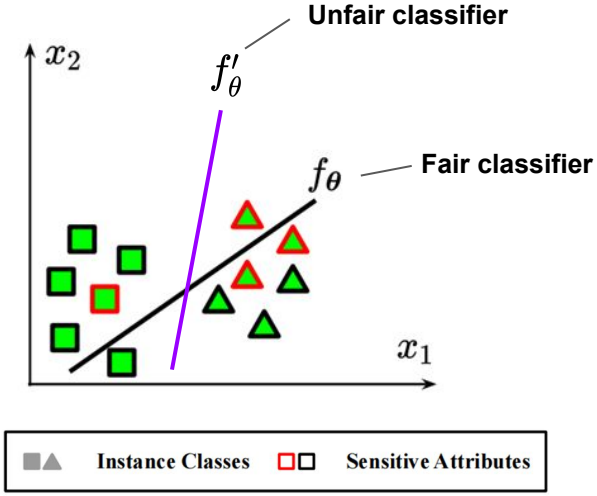


1. Judea Pearl. Causality. Cambridge university press, 2009.
2. Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. JMLR, 2008.
3. Yongkai Wu, Lu Zhang, Xintao Wu, Hanghang Tong. PC-Fairness: A Unified Framework for Measuring Causality-based Fairness. NeurIPS, 2019.

# Fairness-Accuracy Tradeoff




$$\mathbb{P}_{f_\theta}(\blacksquare \mid \square) = \mathbb{P}_{f_\theta}(\blacktriangle \mid \triangle) = 2/4 = 0.5$$
$$\mathbb{P}_{f_\theta}(\blacksquare \mid \square) = \mathbb{P}_{f_\theta}(\blacktriangle \mid \triangle) = 4/8 = 0.5$$

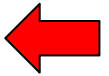
$$\mathbb{P}(f_{\boldsymbol{\theta}}(X) = 1 | Z = 1) = \mathbb{P}(f_{\boldsymbol{\theta}}(X) = 1 | Z = 0)$$


1. Image source: https://medium.com/@afiori_78621/the-fairness-accuracy-tradeoff-af71d5d0c38a
2. Wil van der Aalst. Responsible Data Science: Using Event Data in a “People Friendly” Manner. ICEIS, 2017.
3. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Outline

**Part I:** Introduction to Fairness

**Part II:** Distribution Shift Undermines Fairness

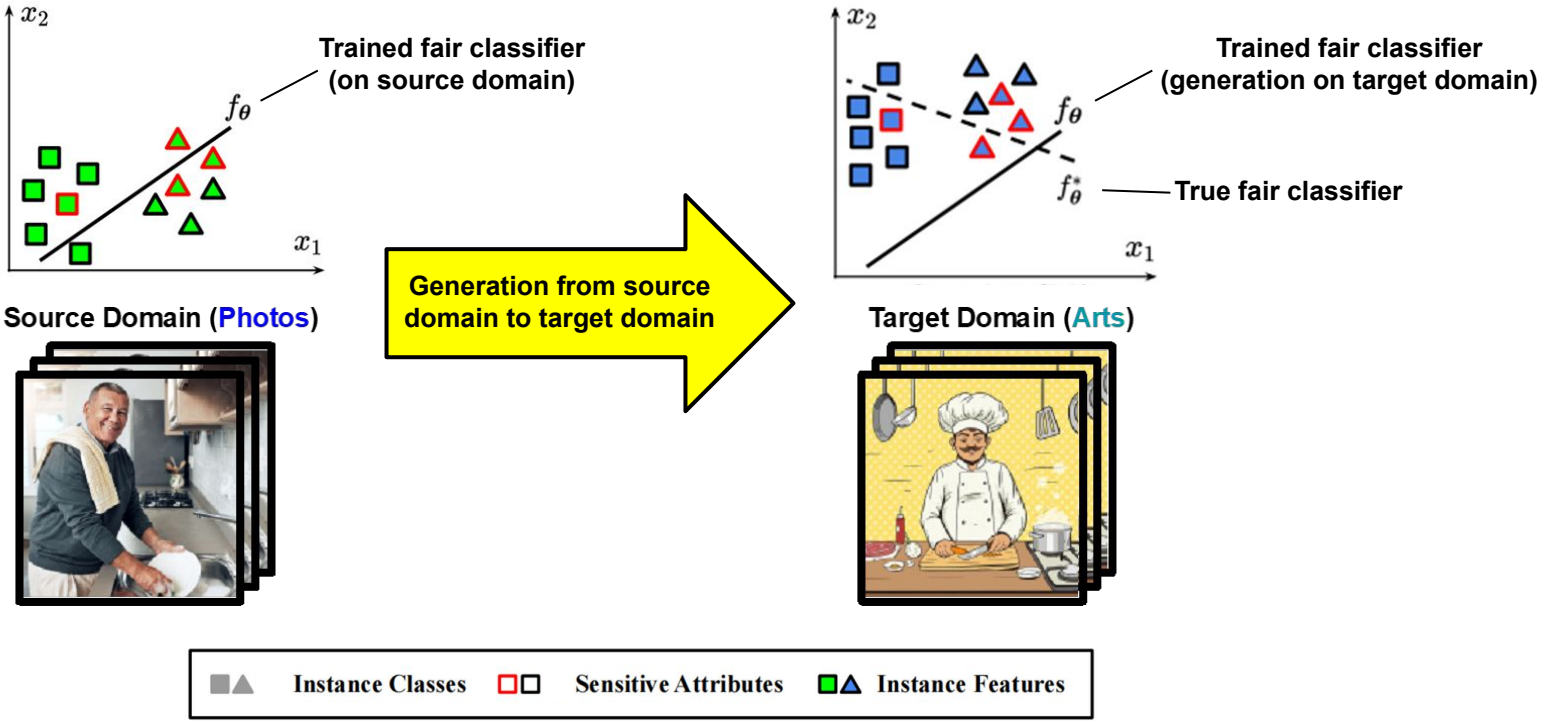


**Part III:** Mitigating Unfairness under Distribution Shift (Offline)

**Part IV:** Mitigating Unfairness under Distribution Shift (Online)

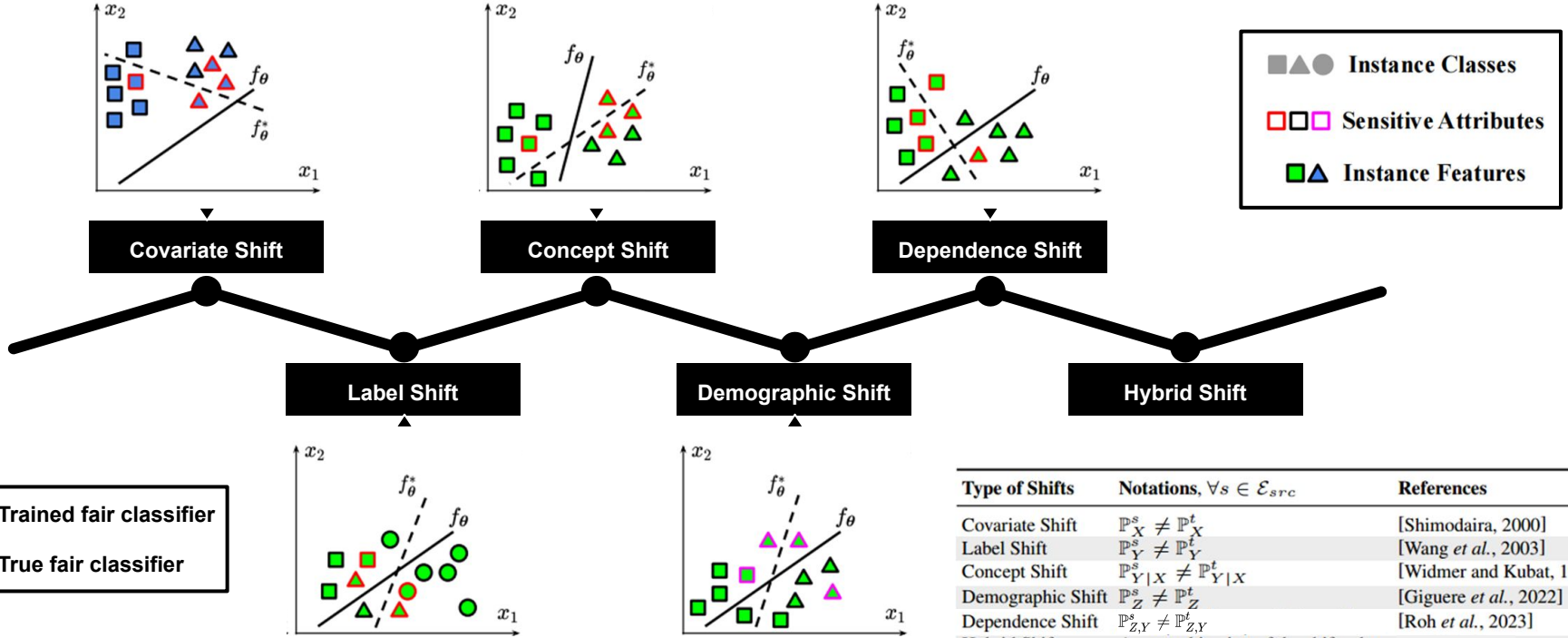
**Part V:** Open Challenges and Beyond

# Distribution Shift Undermines Fairness




1. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Five Types of Distribution Shifts in Model Fairness

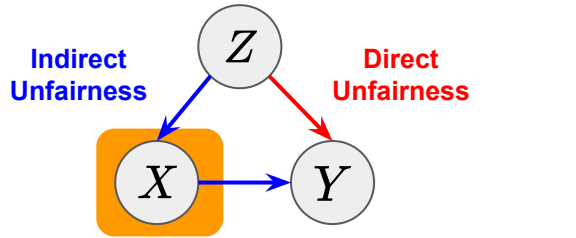


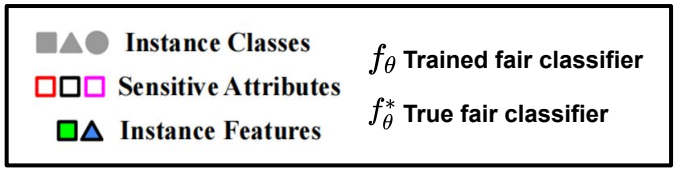

| Type of Shifts | Notations, $\forall s \in \mathcal{E}_{src}$ | References |
|---|---|---|
| Covariate Shift | $\mathbb{P}^s_X \neq \mathbb{P}^t_X$ | [Shimodaira, 2000] |
| Label Shift | $\mathbb{P}^s_Y \neq \mathbb{P}^t_Y$ | [Wang *et al.*, 2003] |
| Concept Shift | $\mathbb{P}^s_{Y\|X} \neq \mathbb{P}^t_{Y\|X}$ | [Widmer and Kubat, 1996] |
| Demographic Shift | $\mathbb{P}^s_Z \neq \mathbb{P}^t_Z$ | [Giguere *et al.*, 2022] |
| Dependence Shift | $\mathbb{P}^s_{Z,Y} \neq \mathbb{P}^t_{Z,Y}$ | [Roh *et al.*, 2023] |
| Hybrid Shift | Any combination of the shifts above | |

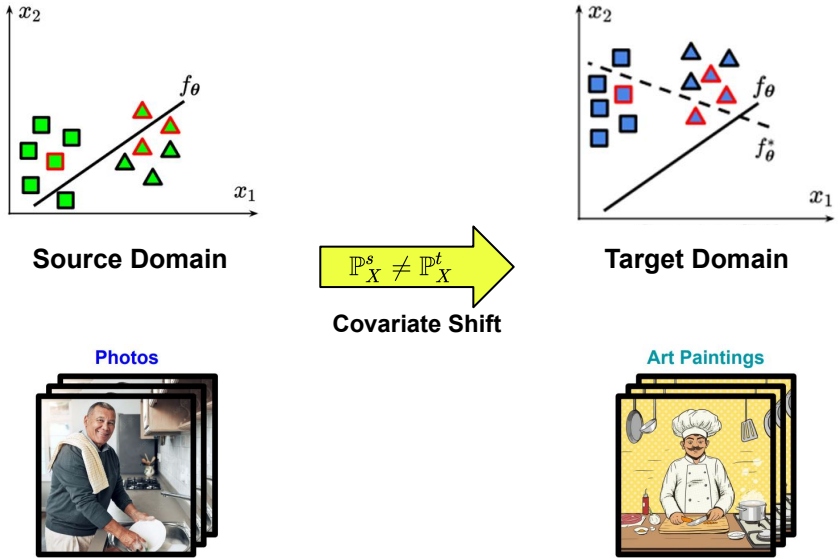

1. Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. JSPI, 2000.
2. Ke Wang, Senqiang Zhou, Chee Ada Fu, and Jeffrey Xu Yu. Mining changes of classification by correspondence tracing. SDM, 2003.
3. Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. Machine learning, 1996.
4. Stephen Giguere, Blossom Metevier, Yuriy Brun, Bruno Castro da Silva, Philip S Thomas, and Scott Niekum. Fairness guarantees under demographic shift. ICLR, 2022.
5. Yuji Roh, Kangwook Lee, Steven Euijong Whang, and Changho Suh. Improving fair training under correlation shifts. ICML, 2023

# Covariate Shift

When the **Covariate Shift** occurs, the model's assumptions about the relationships between X and Y may no longer hold, leading to biased predictions.



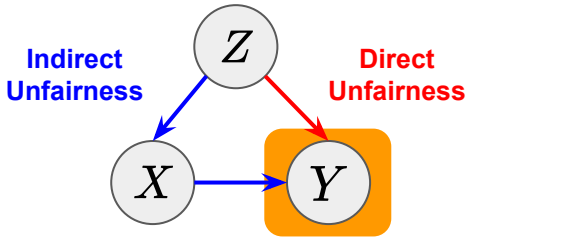



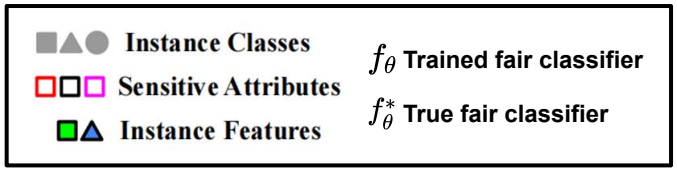


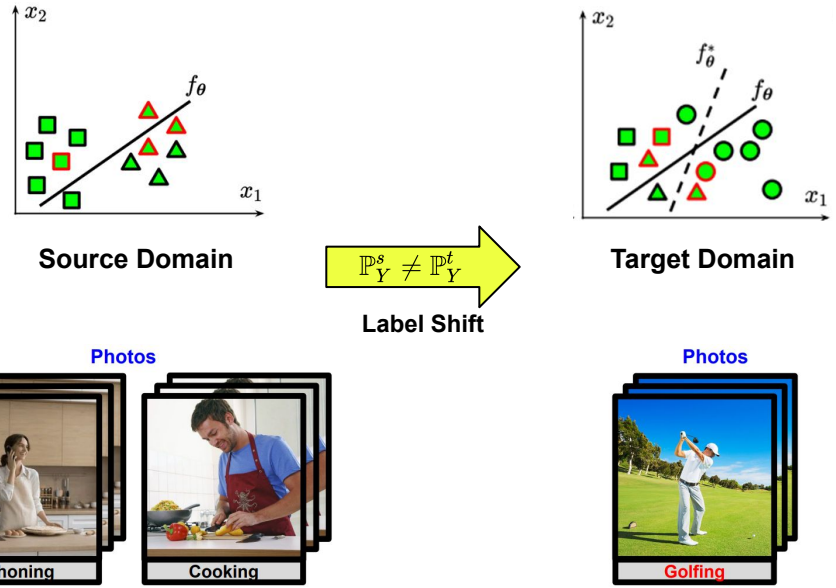
1. Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. JSPI, 2000.
2. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Label Shift

The challenge of fairness learning under **Label Shift** primarily involves mitigating predictive bias in studies related to outlier detection or out of distribution (OOD) detection.



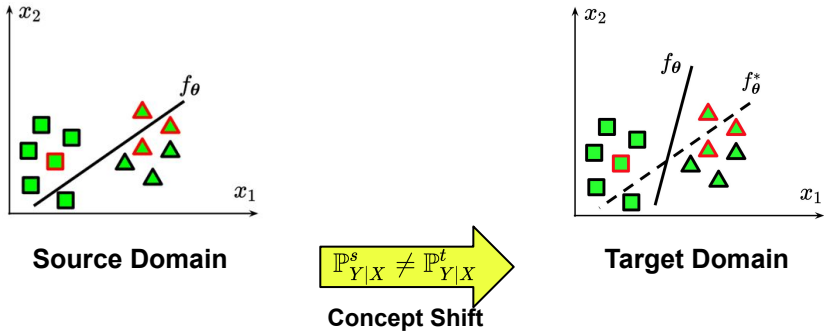



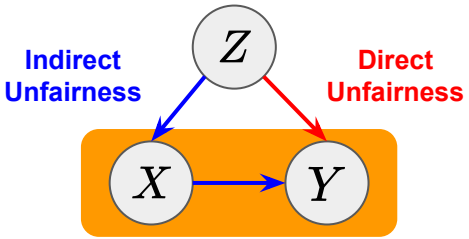


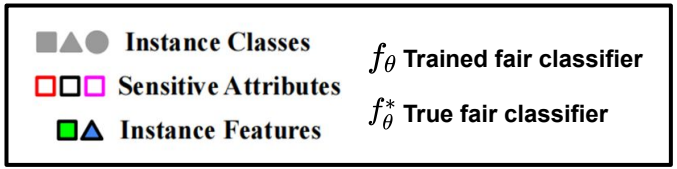
1. Ke Wang, Senqiang Zhou, Chee Ada Fu, and Jeffrey Xu Yu. Mining changes of classification by correspondence tracing. SDM, 2003.
2. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Concept Shift

**Concept Shift** occurs when the underlying concept defining the labels evolves, leading to differences in how the input features correspond to the output labels. It specifically refers to changes in *P(Y|X)*. When this relationship shifts, models trained on historical data may not only perform poorly but also produce predictions that exacerbate biases or unfair outcomes.



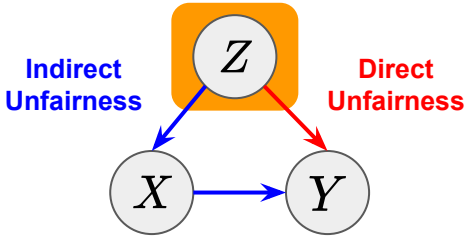



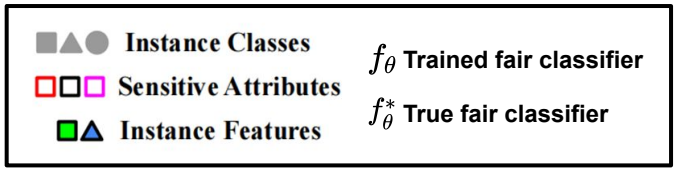


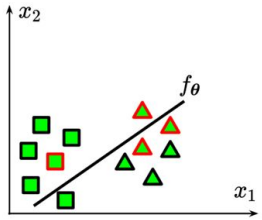
1. Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. Machine learning, 1996.
2. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Demographic Shift

A **Demographic Shift** occurs when a specific sensitive subgroup of the population becomes more or less probable in the target domain.



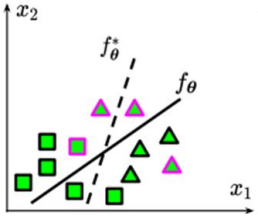



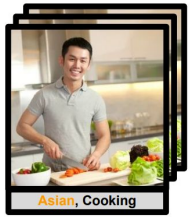

Source Domain

$\mathbb{P}_Z^s \neq \mathbb{P}_Z^t$

Demographic Shift




Target Domain

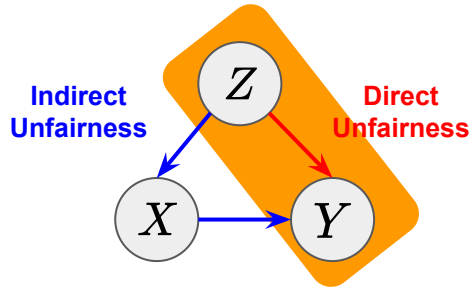



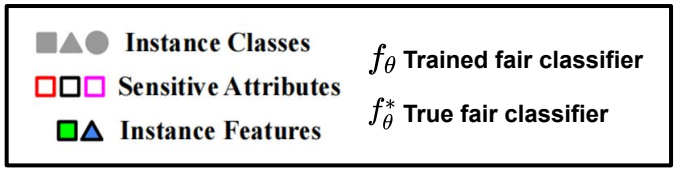


1. Stephen Giguere, Blossom Metevier, Yuriy Brun, Bruno Castro da Silva, Philip S Thomas, and Scott Niekum. Fairness guarantees under demographic shift. ICLR, 2022.
2. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

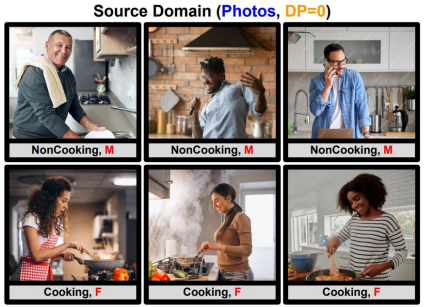
# Dependence (Correlation) Shift

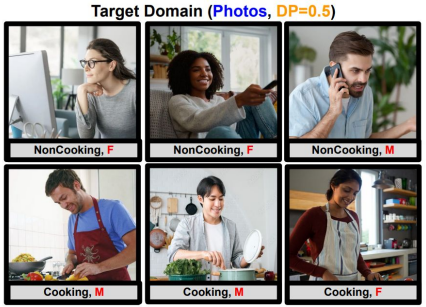
**Dependence Shift** can explicitly capture the alteration in *the correlation between Y and Z* of the data across the source and target domains.





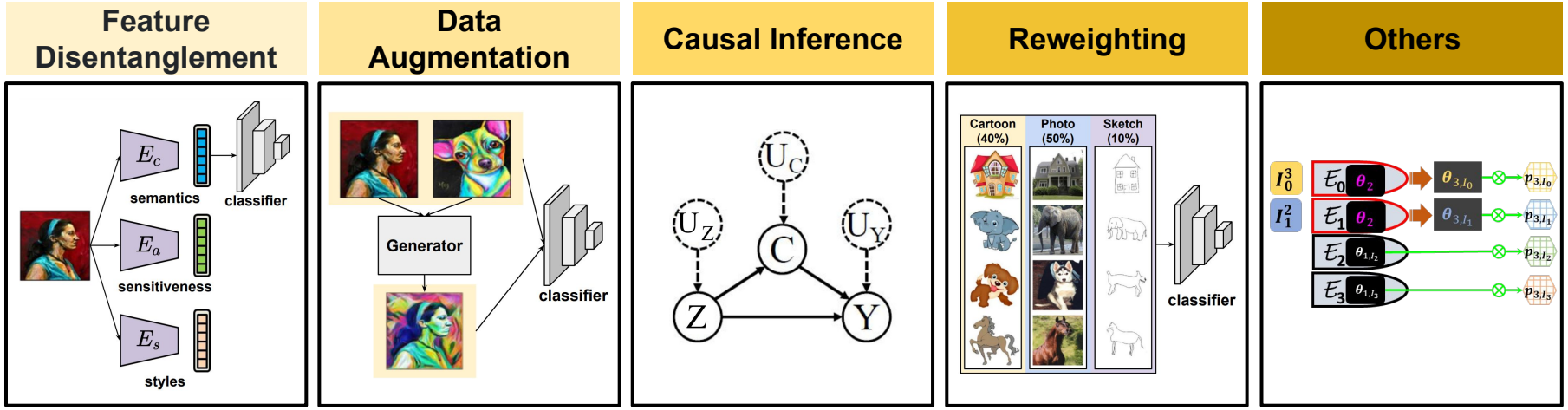




1. Yuji Roh, Kangwook Lee, Steven Euijong Whang, and Changho Suh. Improving fair training under correlation shifts. ICML, 2023
2. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Approaches for Supervised Algorithmic Fairness in Distribution Shifts

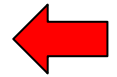



1. Minglai Shao, Dong Li, Chen Zhao, Xintao Wu, Yujie Lin, Qin Tian. Supervised Algorithmic Fairness in Distribution Shifts: A Survey. IJCAI 2024.

# Outline

**Part I:** Introduction to Fairness 

**Part II:** Distribution Shift Undermines Fairness 

**Part III:** Mitigating Unfairness under Distribution Shift (Offline) 

**Part IV:** Mitigating Unfairness under Distribution Shift (Online)

**Part V:** Open Challenges and Beyond

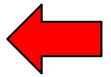
# Break

- **15 Minutes**

# Outline

**Part I:** Introduction to Fairness 

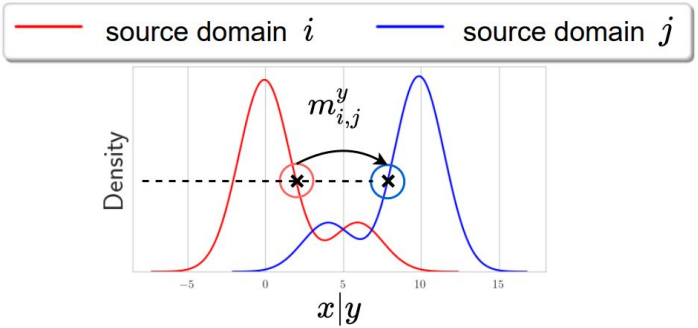
**Part II:** Distribution Shift Undermines Fairness 

**Part III:** Mitigating Unfairness under Distribution Shift (Offline) 

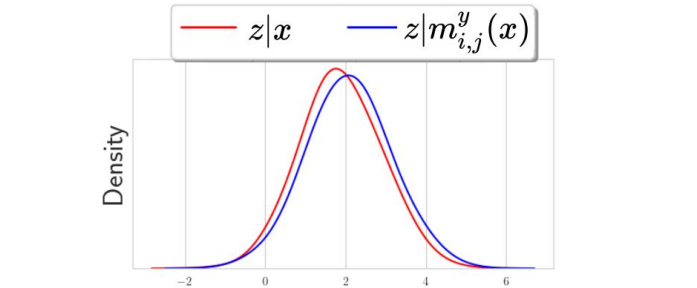
**Part IV:** Mitigating Unfairness under Distribution Shift (Online)

**Part V:** Open Challenges and Beyond

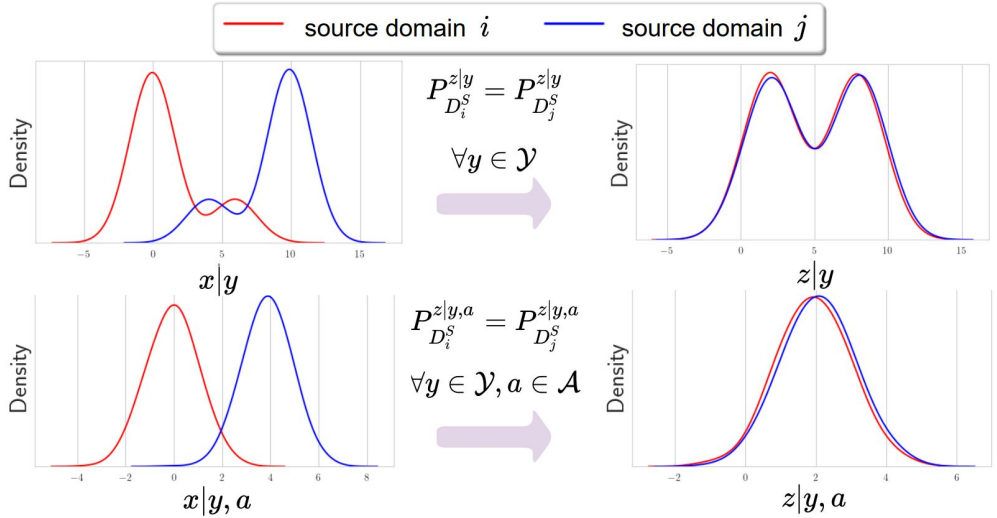
# Fairness and Accuracy Transfer by Density Matching (FATDM)




Stage 1: Finding $m^y_{i,j} : P^{x|y}_{D^S_i} = P^{x|y}_{D^S_j} \ \forall y \in \mathcal{Y}$




Stage 2: Enforcing $P^{(z|x)} = P^{z|m^y_{i,j}(x)} \ \forall x \in \mathcal{X}, y \in \mathcal{Y}$

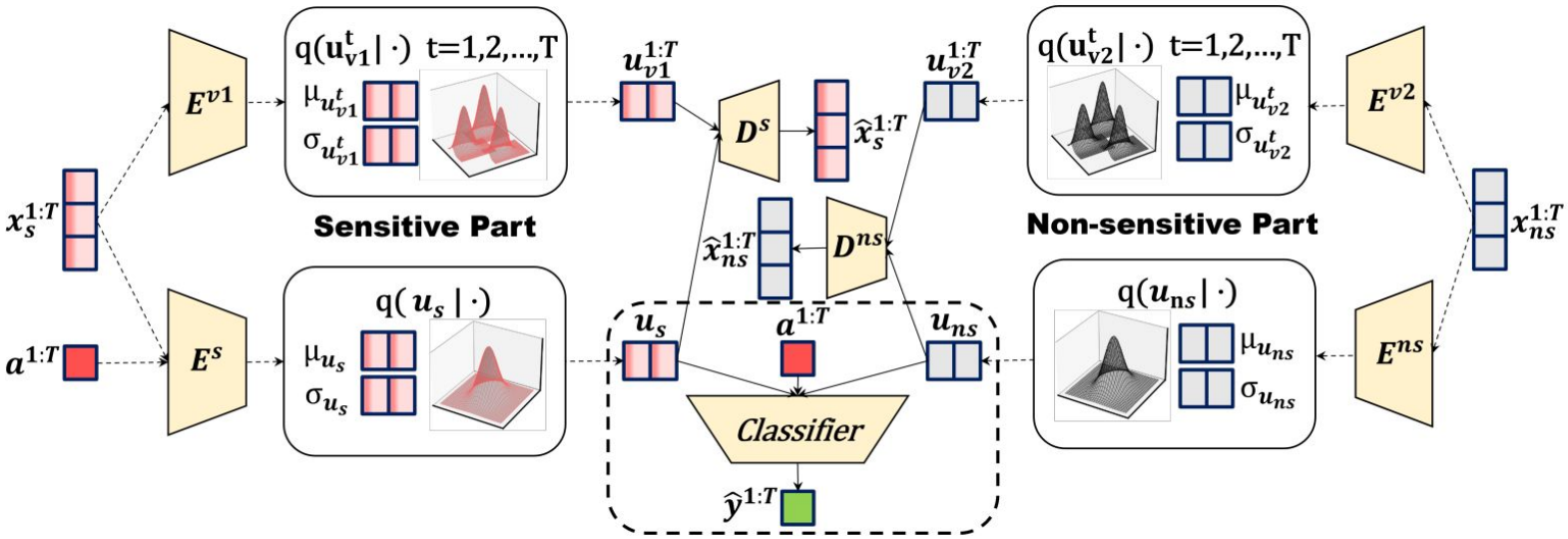



Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation ✓ | Causal | Reweighting | Others |


1. Thai-Hoang Pham, Xueru Zhang, Ping Zhang. Fairness and Accuracy under Domain Generalization. ICLR 2023.

# Disentanglement for Counterfactual Fairness-aware Domain Generalization (DCFDG)

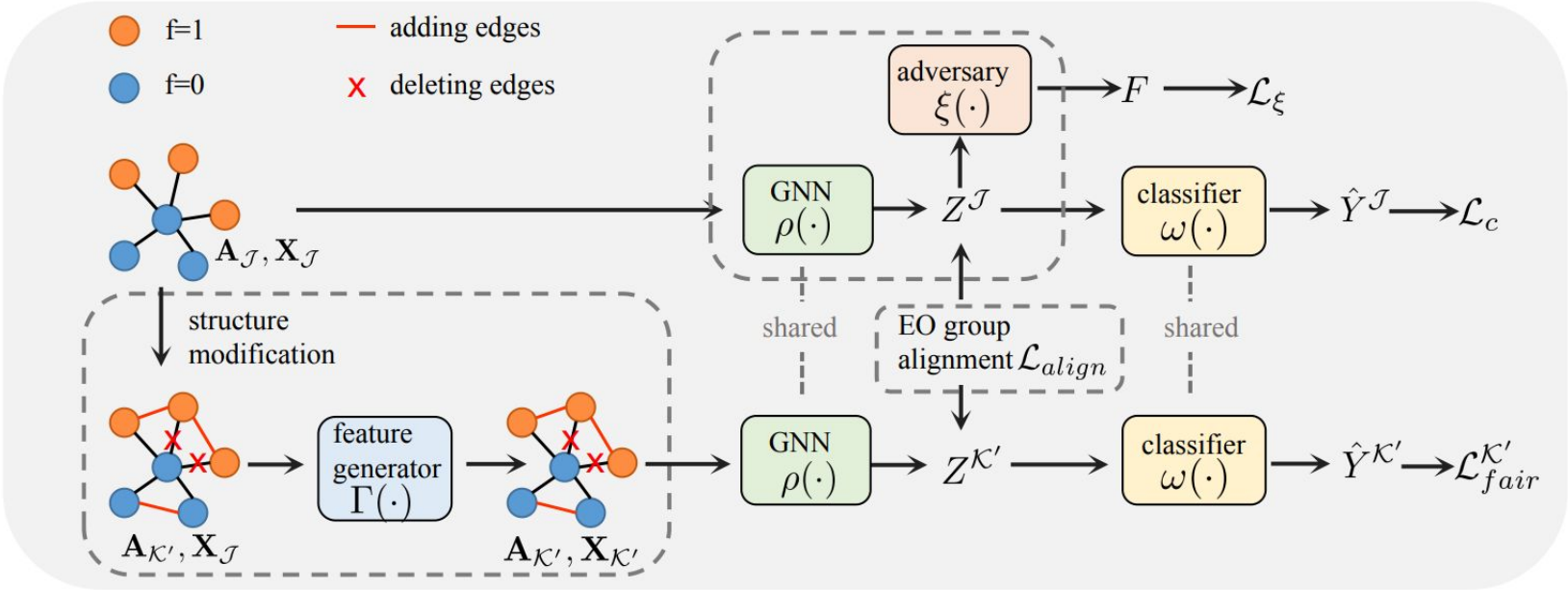



Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement ✓ | Augmentation | Causal ✓ | Reweighting | Others ✓ |


1. Yujie Lin, Chen Zhao, Minglai Shao, Baoluo Meng, Xujiang Zhao, Haifeng Chen. Towards Counterfactual Fairness-aware Domain Generalization in Changing Environments. IJCAI, 2024.

# Graph Fairness Learning under Distribution Shifts (FatraGNN)

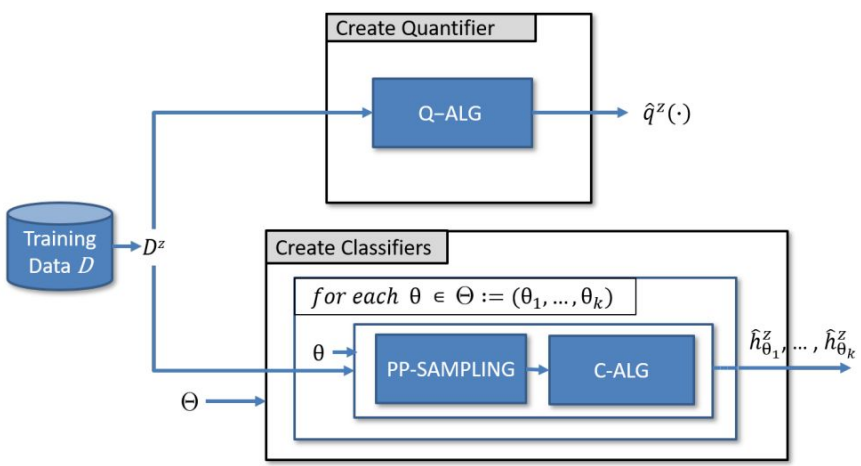



**Types of Distribution Shifts and Methods**

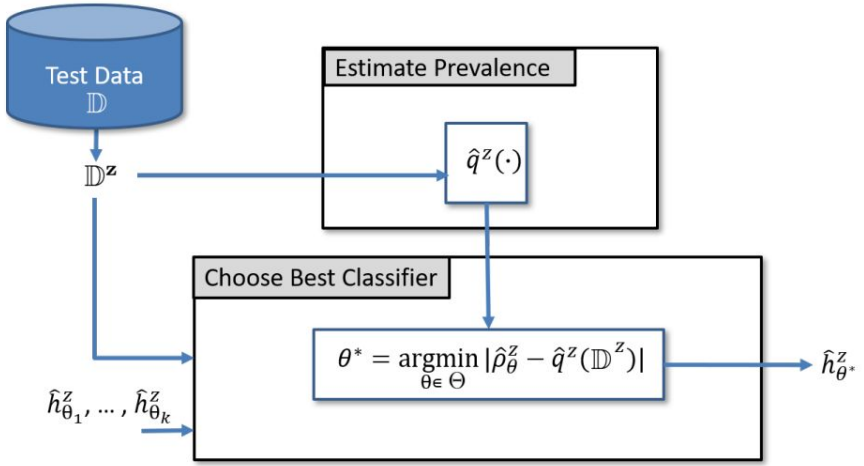
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
|---|---|---|---|---|
| Disentanglement | Augmentation ✓ | Causal | Reweighting | Others |


1. Yibo Li, Xiao Wang, Yujie Xing, Shaohua Fan, Ruijia Wang, Yaoqi Liu, Chuan Shi. Graph Fairness Learning under Distribution Shifts. WWW, 2024.

# Combinatorial Algorithm for Proportional Equality (CAPE)




**Training Phase**



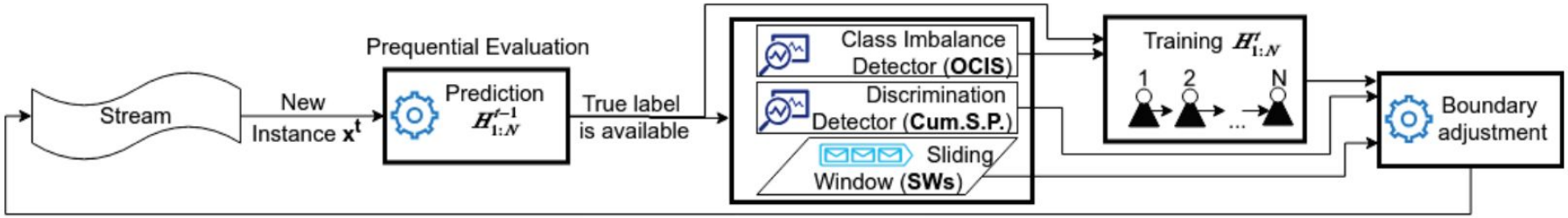

**Testing Phase**

Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift | Label Shift ✓ | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |


1. Arpita Biswas, Suvam Mukherjee. Ensuring Fairness under Prior Probability Shifts. AIES, 2021.

# FAirness and class imBalance-aware BOOsting (FABBOO)

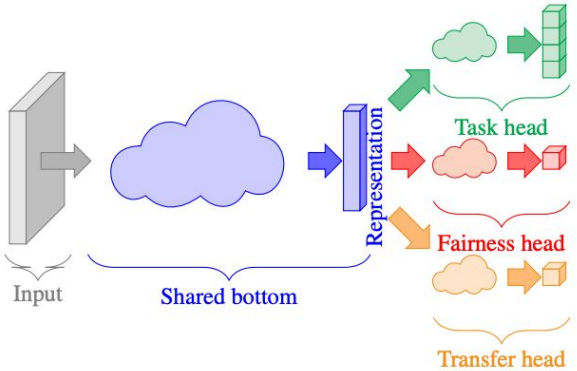



**Types of Distribution Shifts and Methods**

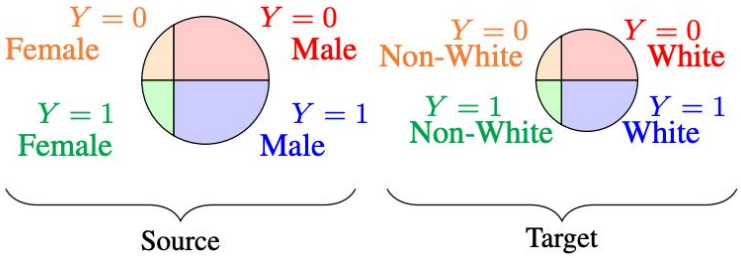
| | | | | |
|---|---|---|---|---|
| Covariate Shift | Label Shift | Concept Shift ✓ | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |


1. Vasileios Iosifidis, Eirini Ntoutsi. FABBOO - Online Fairness-aware Learning under Class Imbalance. ICDS, 2020.

# Transfer of Machine Learning Fairness across Domains




$$\min \left[ \sum_{Z \in \mathcal{D}_S \cup \mathcal{D}_T} L_Y(f(Z), g(Z)) + \sum_{(A, Z^0) \sim \mathcal{D}_{S^0}} \lambda_{Fair} L_{MMD}\left(a(h(Z^0)), A\right) + \sum_{(d, Z^0) \sim \left(\mathcal{D}_{S^0} \cup \mathcal{D}_{T^0}\right)} \lambda_{DA} L_{MMD}\left(d(h(Z^0)), d\right) \right],$$

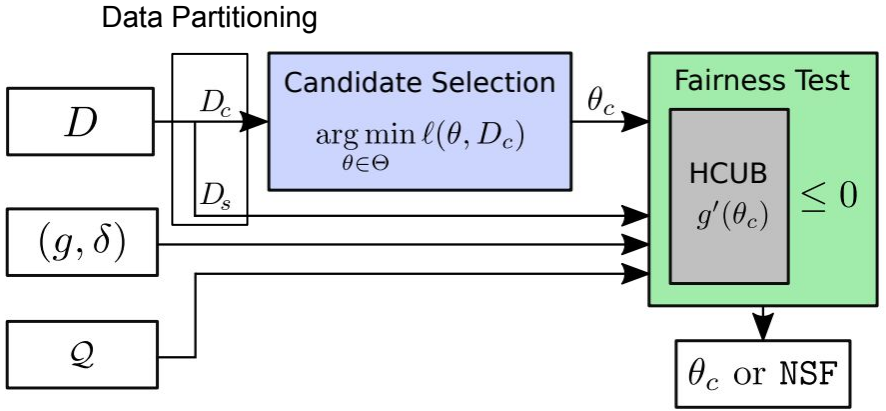



**Types of Distribution Shifts and Methods**

| | | | | |
|---|---|---|---|---|
| Covariate Shift | Label Shift | Concept Shift | Demographic Shift ✓ | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |

1. Candice Schumann, Xuezhi Wang, Alex Beutel, Jilin Chen, Hai Qian, Ed H Chi. Transfer of Machine Learning Fairness across Domains. arXiv, 2019.

# Fairness Guarantees under Demographic Shift (Shifty)




- **Known Demographic Shift**

$$\phi(t) := \frac{\Pr(T'=t|\xi')}{\Pr(T=t|\xi)} = \frac{\Pr(\xi|T=t)\Pr(T'=t)}{\Pr(T=t|\xi)\sum_{t' \in \mathcal{T}} \Pr(\xi|T=t')\Pr(T'=t')}$$

- **UnKnown Demographic Shift**

$$\phi(t; q) := \frac{\Pr(\xi|T=t)q_t}{\Pr(T=t|\xi)\sum_{t' \in \mathcal{T}} \Pr(\xi|T=t')q_{t'}}$$

$D$ : Training data
$g$ : A definition of unfair behavior
$\delta$ : A tolerance
$\mathcal{Q}$ : A description of possible demographic shifts

Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift | Label Shift | Concept Shift | Demographic Shift ✓ | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |


1. Stephen Giguere, Blossom Metevier, Yuriy Brun, Bruno Castro da Silva, Philip S Thomas, Scott Niekum. Fairness Guarantees under Demographic Shift. ICLR, 2022.

# Environment Inference for Invariant Learning (EIIL)

(a) **Inferred environment 1** *(mostly) landbirds on land, and waterbirds on water*

(b) **Inferred environment 2** *(mostly) landbirds on water, and waterbirds on land*

**EIIL involves the following sequential approach:**

1. Input *reference model* $\tilde{\Phi}$;
2. Fix $\Phi \leftarrow \tilde{\Phi}$ and optimize the EI objective to infer environments: $\mathbf{q}^* = \arg\max_{\mathbf{q}} C^{EI}(\tilde{\Phi}, \mathbf{q})$;
3. Fix $\tilde{\mathbf{q}} \leftarrow \mathbf{q}^*$ and optimize the IL objective to yield the new model: $\Phi^* = \arg\min_{\Phi} C^{IL}(\Phi, \tilde{\mathbf{q}})$

$$C^{EI}(\Phi, \mathbf{q}) = ||\nabla_{\bar{w}} \tilde{R}^e(\bar{w} \circ \Phi, \mathbf{q})||,$$
$$\tilde{R}^e(\Phi, \mathbf{q}) = \frac{1}{\sum_{i'} \mathbf{q}_{i'}(e)} \sum_i \mathbf{q}_i(e) \ell(\Phi(x_i), y_i)$$

**In practice,** $C^{IL} \in \{C^{IRM}, C^{GroupDRO}\}$

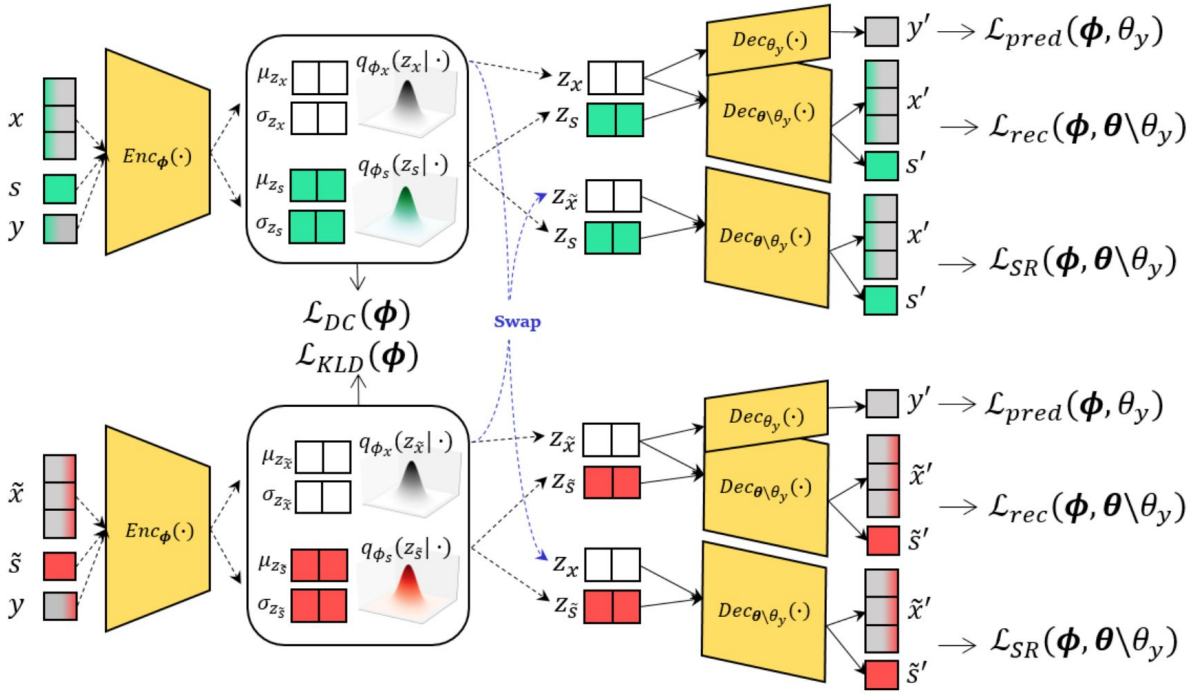
$$C^{IRM}(\Phi) = \sum_{e \in \mathcal{E}^{obs}} R^e(\Phi) + \lambda ||\nabla_{\bar{w}} R^e(\bar{w} \circ \Phi)||$$
$$C^{GroupDRO}(\Phi) = \max_g \mathbb{E}_{g(e)}[R^e(\Phi)]$$

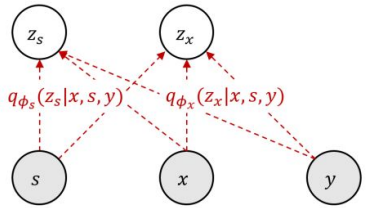
**Types of Distribution Shifts and Methods**

| | | | | |
|---|---|---|---|---|
| Covariate Shift | Label Shift | Concept Shift | Demographic Shift | Dependence Shift ✓ |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |

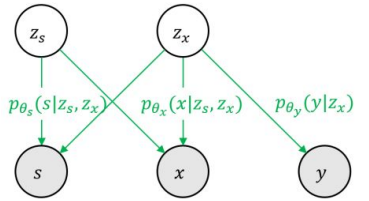
1. Elliot Creager, Jorn-Henrik Jacobsen, Richard Zemel. Environment Inference for Invariant Learning. ICML, 2021.

# FAir Representation via distributional CONtrastive Variational AutoEncoder (FarconVAE)




(a) Recognition model



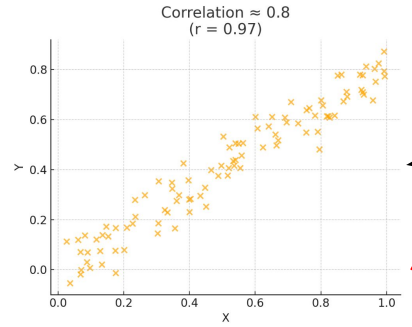

(b) Generative model

Types of Distribution Shifts and Methods

| Covariate Shift | Label Shift | Concept Shift | Demographic Shift | Dependence Shift ✓ |
|---|---|---|---|---|
| Disentanglement ✓ | Augmentation | Causal | Reweighting | Others |


1. Changdae Oh, Heeji Won, Junhyuk So, Taero Kim, Yewon Kim, Hosik Choi, and Kyungwoo Song. Learning Fair Representation via Distributional Contrastive Disentanglement. KDD, 2022.

# Improving Fair Training under Correlation Shifts




Correlation between y & z

Training

High correlation
=> Achievable performance: low

① *Pre-processing*

Correlation shift

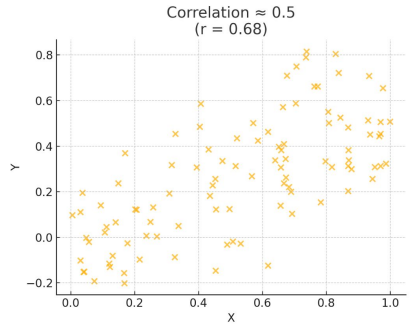
② *Fair model training (in-processing)*

Deployment

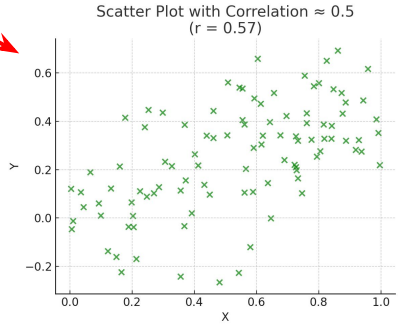
Low correlation
=> Achievable performance: High

*Pre-processing*

**Estimate a range of correlations for the target domain**




*Fair model training (in-processing)*

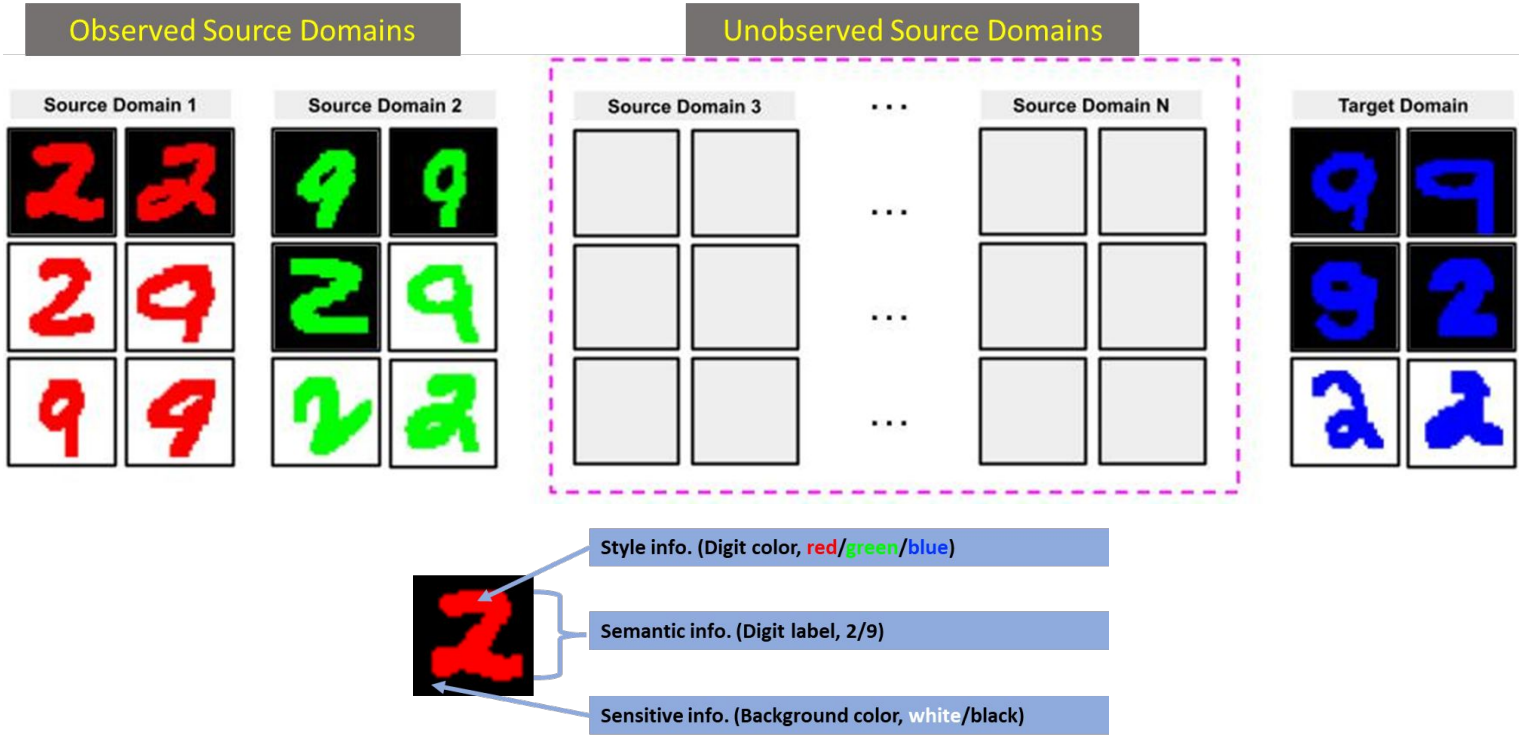



**Types of Distribution Shifts and Methods**

| Covariate Shift | Label Shift | Concept Shift | Demographic Shift | Dependence Shift ✓ |
|---|---|---|---|---|
| Disentanglement | Augmentation | Causal | Reweighting ✓ | Others |

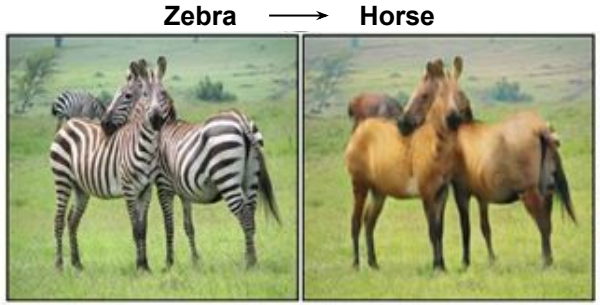
1. Yuji Roh, Kangwook Lee, Steven Euijong Whang, Changho Suh. Improving Fair Training under Correlation Shifts. ICML, 2023.

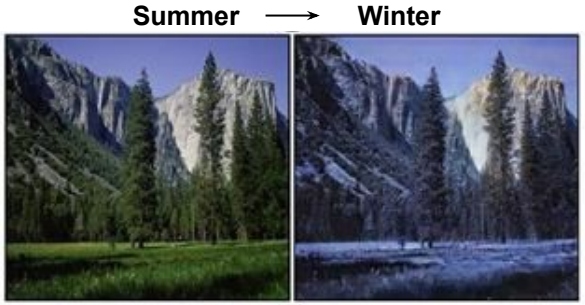
# The More You See; The More You Know…




1. Chen Zhao, Kai Jiang, Xintao Wu, Haoliang Wang, Latifur Khan, Christan Grant, Feng Chen. Algorithmic Fairness Generalization under Covariate and Dependence Shifts Simultaneously. KDD 2024

# Transformation Models

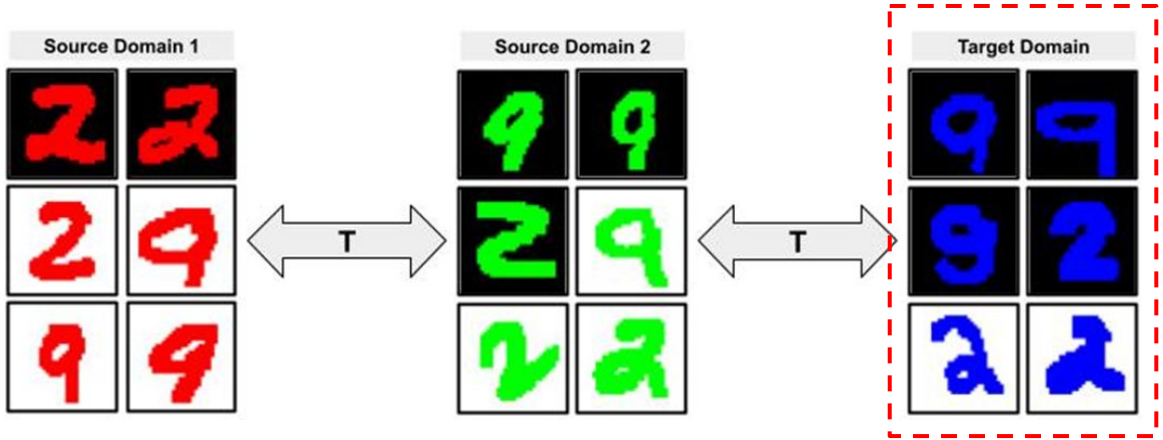




1. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV, 2017

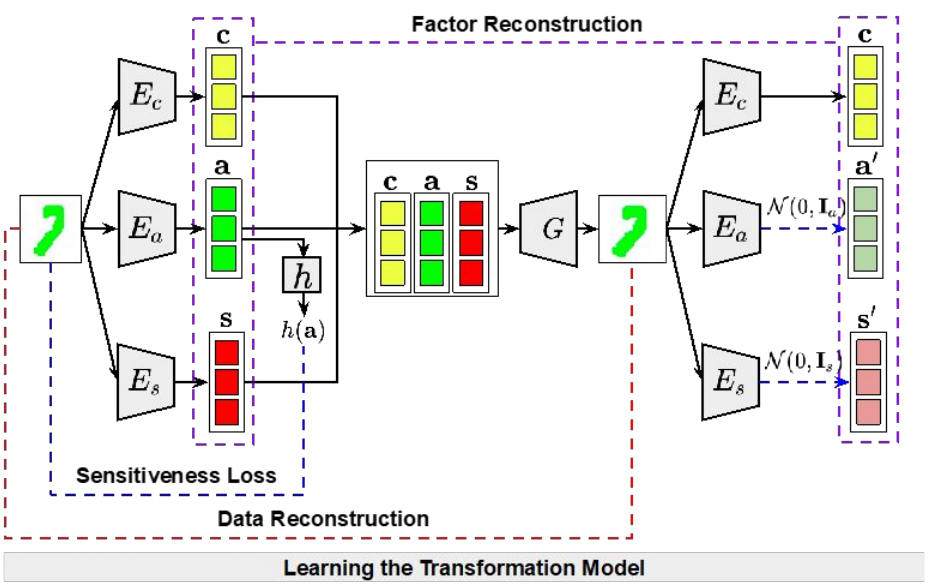
# Transformation Models



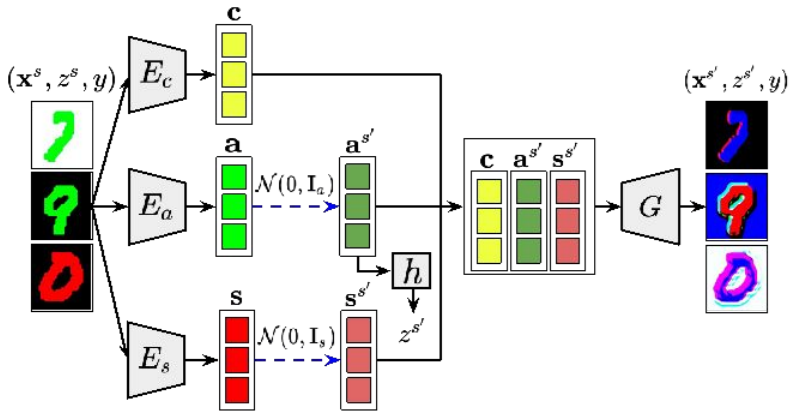

There exists a transformation model T that:

1. Distribution shifts can characterize generalization tasks across domains.
2. Generate augmented data in unobserved domains by perturbing existing samples with various variations.

# Fair disEntangled DOmain geneRAlization (FEDORA)




Learning the Transformation Model



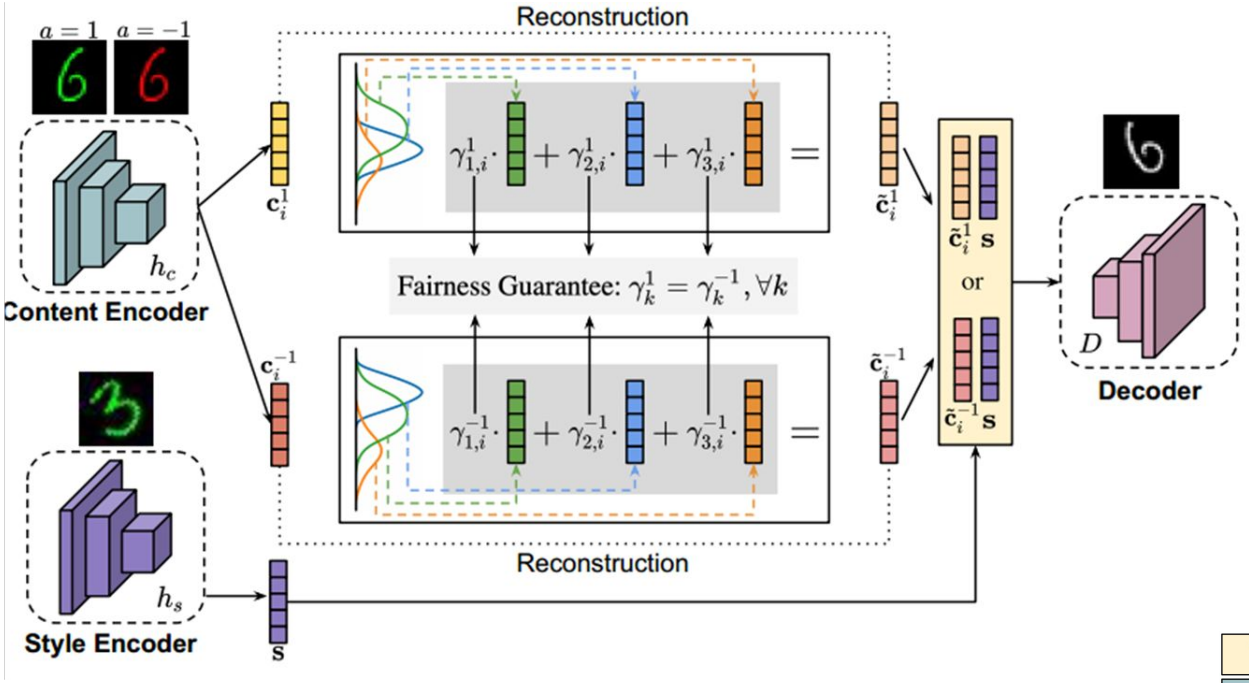
Augmentation in Synthetic Domains

Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift ✓ |
| Disentanglement ✓ | Augmentation ✓ | Causal | Reweighting | Others ✓ |


1. Chen Zhao, Kai Jiang, Xintao Wu, Haoliang Wang, Latifur Khan, Christan Grant, Feng Chen. Algorithmic Fairness Generalization under Covariate and Dependence Shifts Simultaneously. KDD 2024

# Fairness-aware LeArning Invariant Representations (FLAIR)




- Minimize the information disclosure related to a specific sensitive subgroup
- Maximize the preservation of significant information within non-sensitive representations

Types of Distribution Shifts and Methods

| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift ✓ |
|---|---|---|---|---|
| Disentanglement ✓ | Augmentation | Causal | Reweighting | Others ✓ |


1. Dong Li, Chen Zhao, Minglai Shao, Wenjun Wang. Learning Fair Invariant Representations under Covariate and Correlation Shifts Simultaneously. CIKM 2024

# Fairness-aware Outlier Detection (FairOD)

PROBLEM 1 (FAIRNESS-AWARE OUTLIER DETECTION). *Given samples $\mathcal{X}$ and protected variable values $\mathcal{PV}$, estimate outlier scores $\mathcal{S}$ and assign outlier labels $\mathcal{O}$, to achieve*

*(i)* $P(Y = 1|O = 1) > P(Y = 1)$ , *[Detection effectiveness]*

*(ii)* $P(O \mid X, PV = v) = P(O \mid X),\ \forall v \in \{a, b\}$ , *[Treatment parity]*

*(iii)* $P(O = 1|PV = a) = P(O = 1|PV = b)$ , *[Statistical parity]*

*(iv)* $\pi^{\text{BASE}}_{PV=v} = \pi_{PV=v}$ , $\forall v \in \{a, b\}$, *where* BASE *is a fairness-agnostic detector.* *[Group fidelity proxy]*

$$\mathcal{L} = \alpha \underbrace{\mathcal{L}_{\text{BASE}}}_{\text{Reconstruction}} + (1-\alpha) \underbrace{\mathcal{L}_{SP}}_{\text{Statistical Parity}} + \gamma \underbrace{\mathcal{L}_{GF}}_{\text{Group Fidelity}}$$

$$\mathcal{L}_{\text{BASE}} = \sum_{i=1}^{N} \|X_i - G(X_i)\|_2^2$$

$$\mathcal{L}_{SP} = \left| \frac{\left(\sum_{i=1}^{N} s(X_i) - \mu_s\right)\left(\sum_{i=1}^{N} PV_i - \mu_{PV}\right)}{\sigma_s\ \sigma_{PV}} \right|$$

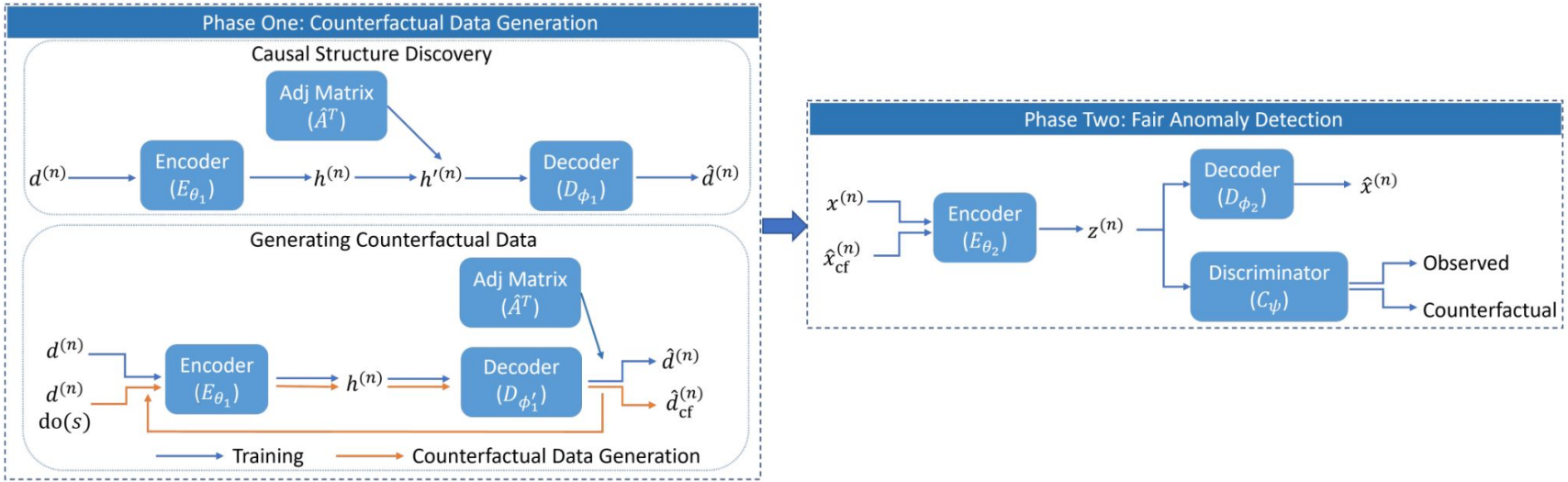
$$\mathcal{L}_{GF} = \sum_{v \in \{a,b\}} \left( 1 - \sum_{X_i \in \mathcal{X}_{PV=v}} \frac{2^{s^{\text{BASE}}(X_i)} - 1}{\text{DNM}} \right)$$

**Types of Distribution Shifts and Methods**

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift ✓ | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |

1. Shubhranshu Shekhar, Neil Shah, Leman Akoglu. FairOD: Fairness-aware Outlier Detection. AIES, 2021.

# Counterfactually Fair Anomaly Detection (CFAD)




**Types of Distribution Shifts and Methods**

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift ✓ | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement ✓ | Augmentation | Causal ✓ | Reweighting | Others |

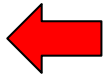

1. Xiao Han, Lu Zhang, Yongkai Wu, Shuhan Yuan. Achieving Counterfactual Fairness for Anomaly Detection. PAKDD, 2023.

# Outline

**Part I:** Introduction to Fairness 

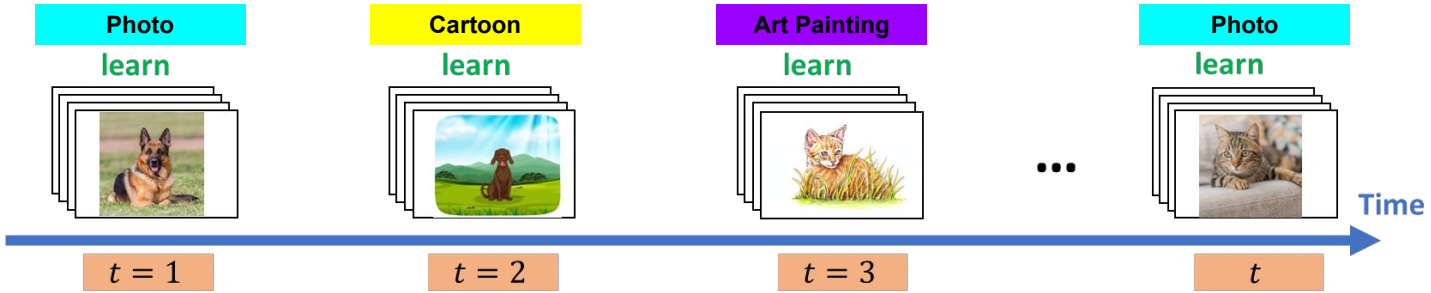
**Part II:** Distribution Shift Undermines Fairness 

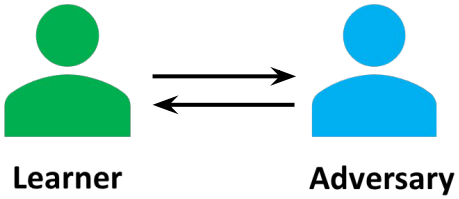
**Part III:** Mitigating Unfairness under Distribution Shift (Offline) 

**Part IV:** Mitigating Unfairness under Distribution Shift (Online) 

**Part V:** Open Challenges and Beyond

# Fair Online Learning under Distribution Shifts



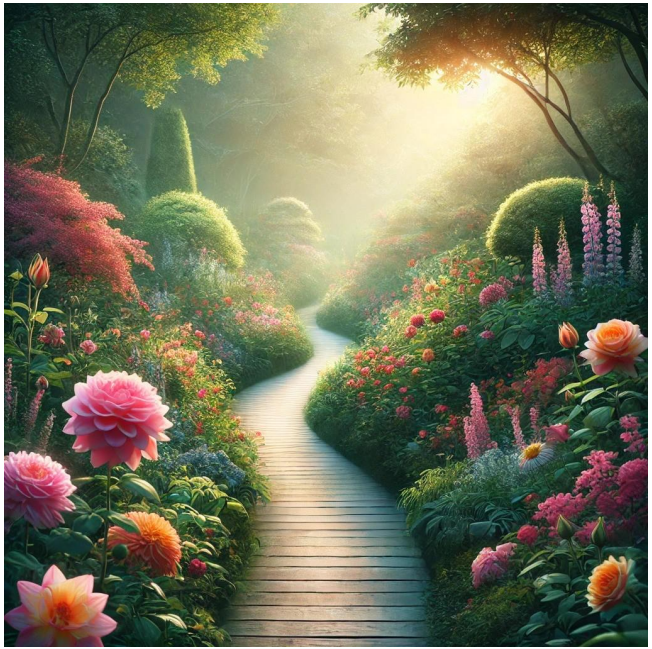

Step 1: The learner SELECTS $\boldsymbol{\theta}_t$ in the fair domain $\Theta$

Step 2: The adversary reveals a loss function $f_t(\cdot)$ and fairness function $g(\cdot)$

Step 3: The learner incurs an instantaneous loss $f_t(\boldsymbol{\theta}_t, \mathcal{D}_t)$ and fairness notion $g_i(\boldsymbol{\theta}_t, \mathcal{D}_t)$

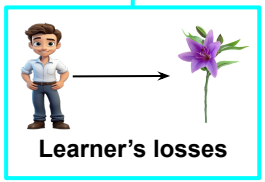
Step 4: Advance to round $t + 1$

# Regrets

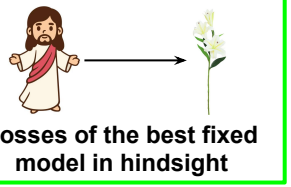


**Static Regret:**

$$R_S = \sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)$$



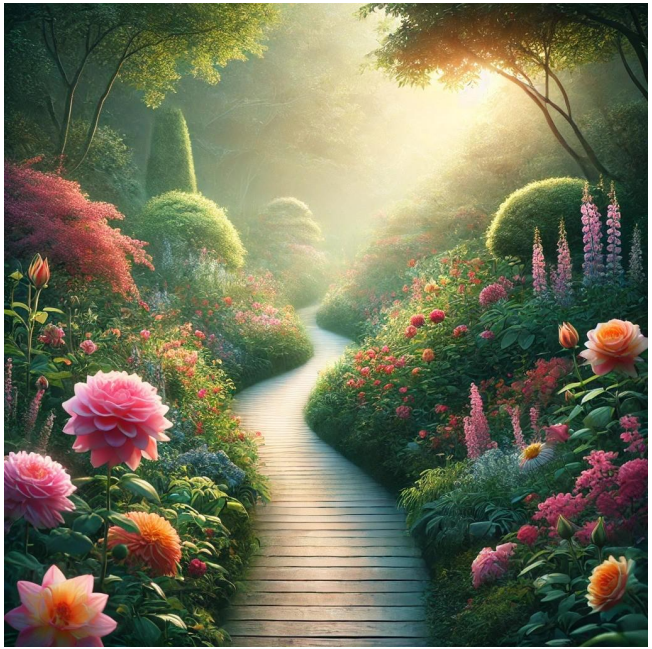
Learner's losses



Losses of the best fixed model in hindsight

- Static regret works for data coming from fixed distribution (with non-stationary task distributions).
- The goal is to design algorithms such that this regret grows with *T* as slowly as possible.
- In particular, an algorithm whose regret grows sub-linearly in *T* is non-trivially learning and adapting.

# Regrets



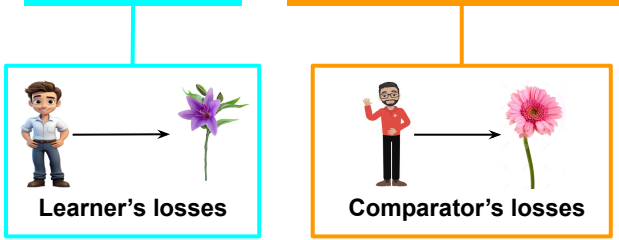
**Static Regret:**

$$R_S = \sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)$$

**Dynamic Regret:**

$$R_D = \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(u_t)$$

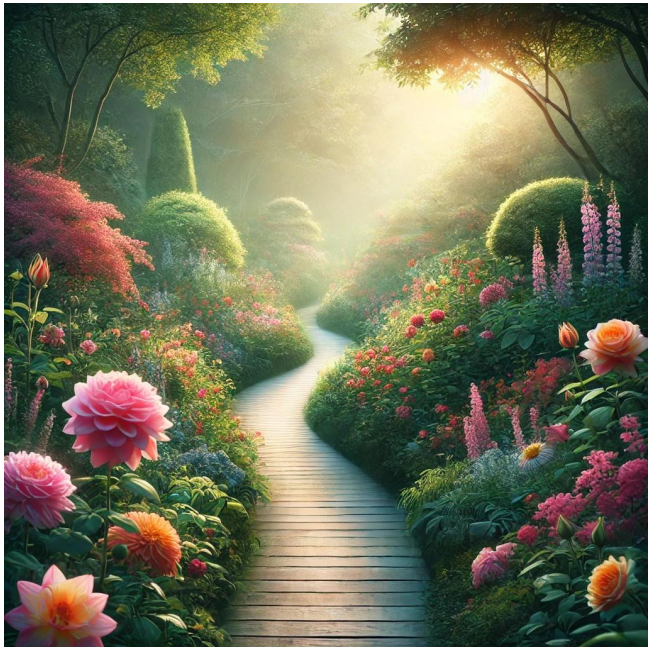
$$= \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \min_{\theta \in \Theta} \ell_t(\theta)$$




- Dynamic regret works for data coming from changing environments.

1. Martin Zinkevich. Online Convex Programming and Generalized Infinitesimal Gradient Ascent. ICML, 2003.

# Regrets



**Static Regret:**

$$R_S = \sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)$$

**Dynamic Regret:**

$$R_D = \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(u_t)$$

**Adaptive Regret:**

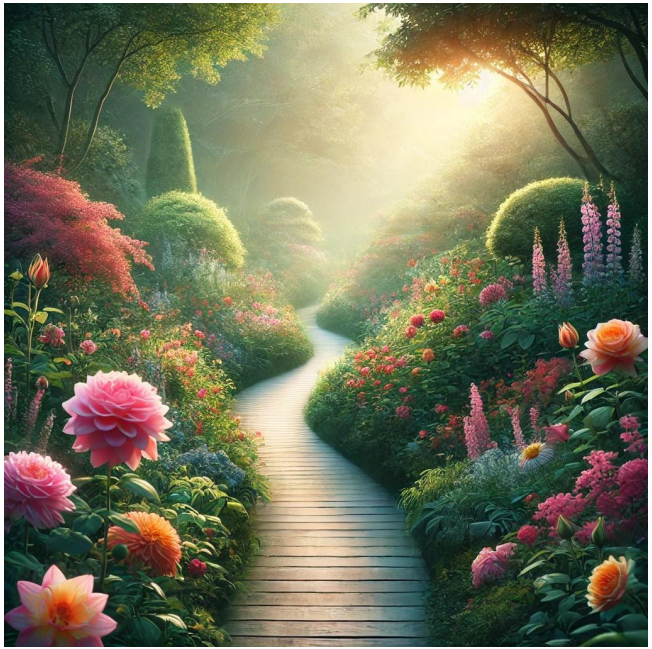
$$R_A = \max_{[s,q] \subseteq [T]} \left( \sum_{t=s}^{q} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=s}^{q} \ell_t(\theta) \right)$$

**static regrets in short intervals**

- Both dynamic and adaptive regrets work for data coming from changing environments.
- Dynamic regret addresses changing environments from a **global** perspective, while adaptive regret takes a **local** perspective by focusing on comparators within short intervals.


1. Martin Zinkevich. Online Convex Programming and Generalized Infinitesimal Gradient Ascent. ICML, 2003.
2. Elad Hazan, C. Seshadhri. Adaptive Algorithms for Online Decision Problems. Electronic Colloquium on Computational Complexity, 2007.

# Regrets



**Static Regret:**

$$R_S = \sum_{t=1}^{T} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)$$

**Dynamic Regret:**

$$R_D = \sum_{t=1}^{T} \ell_t(\theta_t) - \sum_{t=1}^{T} \ell_t(u_t)$$

**Adaptive Regret:**

$$R_A = \max_{[s,q] \subseteq [T]} \left( \sum_{t=s}^{q} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=s}^{q} \ell_t(\theta) \right)$$

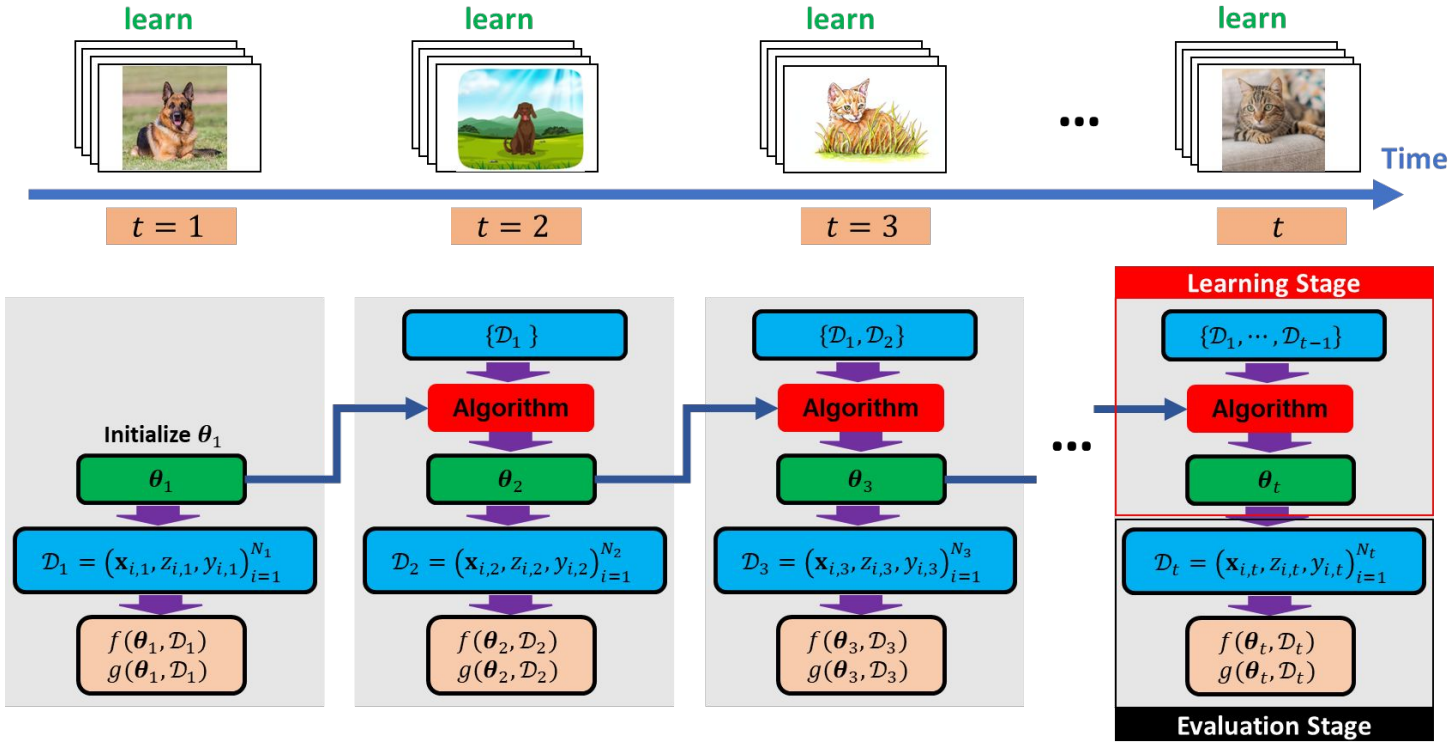
**Long-Term Fair Constraints:**

$$\sum_{t=1}^{T} \left|\left| \left[ g_i(\theta_t) \right]_+ \right|\right| \leq \epsilon, i \in [m]$$

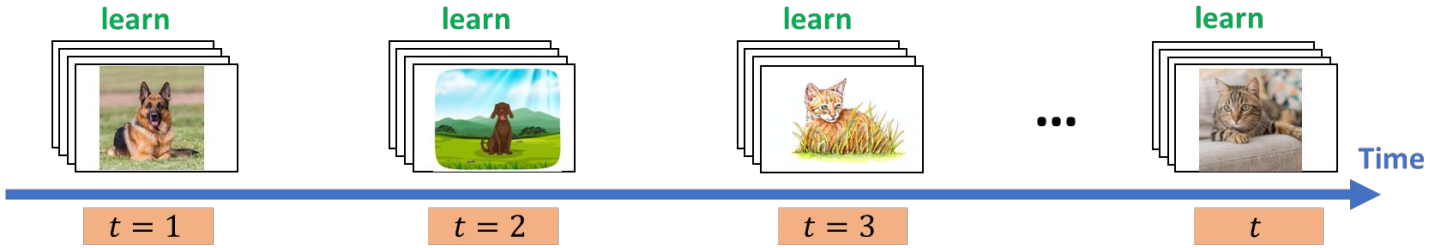
add to these regrets


1. Martin Zinkevich. Online Convex Programming and Generalized Infinitesimal Gradient Ascent. ICML, 2003.
2. Elad Hazan, C. Seshadhri. Adaptive Algorithms for Online Decision Problems. Electronic Colloquium on Computational Complexity, 2007.
3. Mehrdad Mahdavi, Rong Jin, and Tianbao Yang. Trading Regret for Efficiency: Online Convex Optimization with Long Term Constraints. JMLR, 2012.

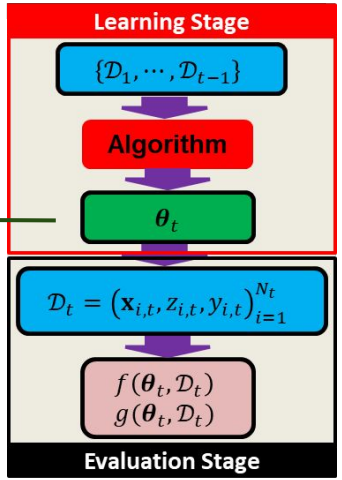
# Fair Online Learning under Distribution Shifts



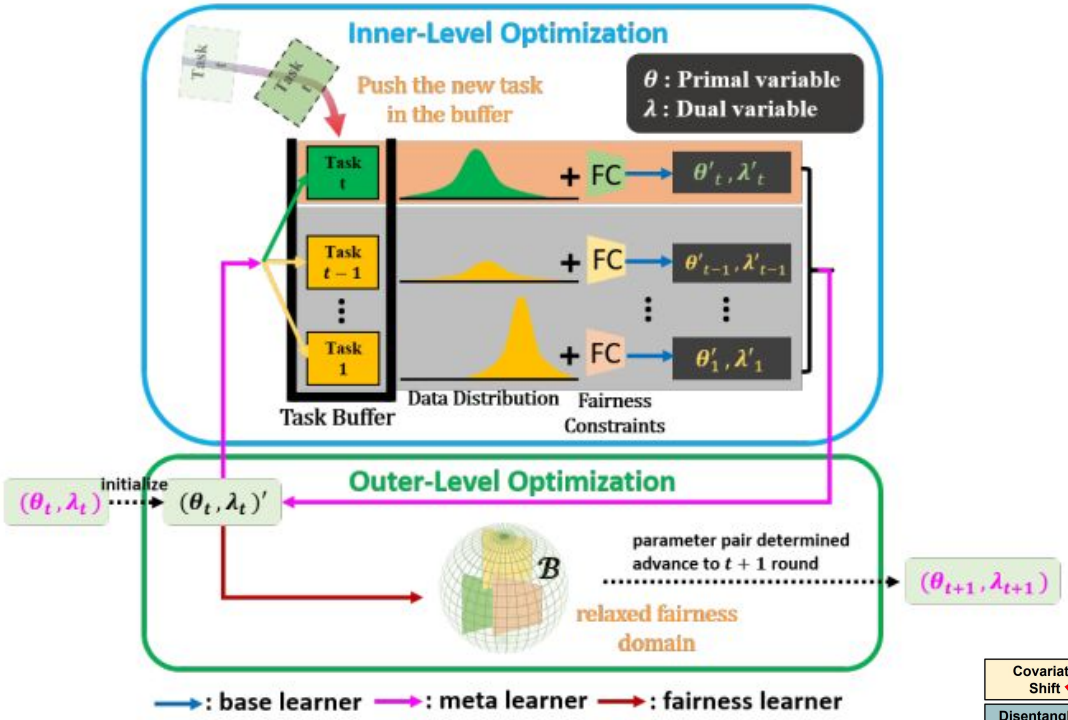
# Fair Online Learning under Distribution Shifts




The goal is to develop an effective fairness-aware online algorithm for changing environments, ensuring that the learned parameter sequence $\{\theta_1, \cdots, \theta_t\}$ minimizes both **the loss regret** and **the violation of long-term fairness constraints**.



# Follow the Fair Meta Leader (FFML)

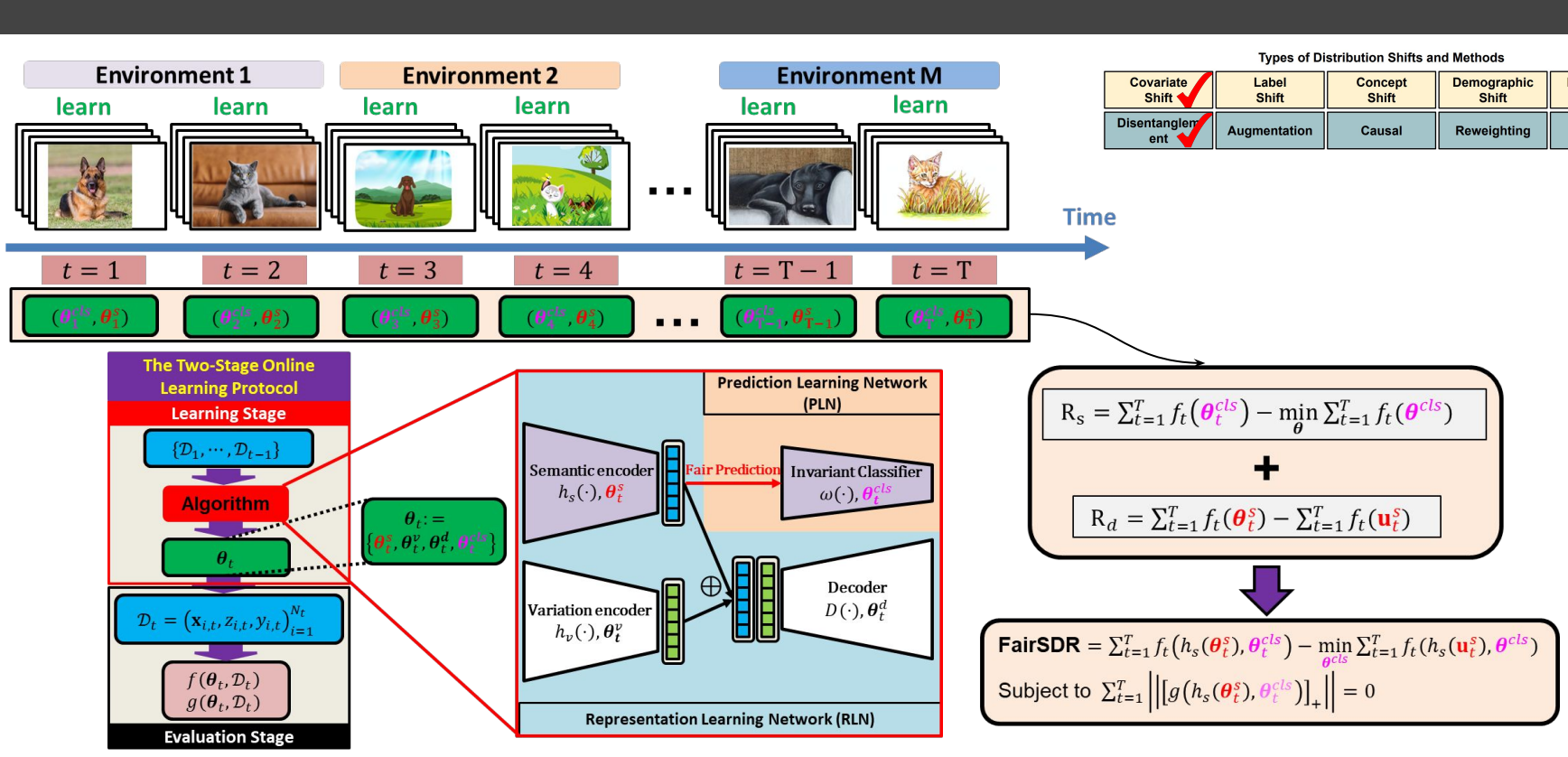



Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting | Others ✓ |

1. Chen Zhao, Feng Chen, Bhavani Thuraisingham. Fairness-Aware Online Meta-learning. KDD, 2021.

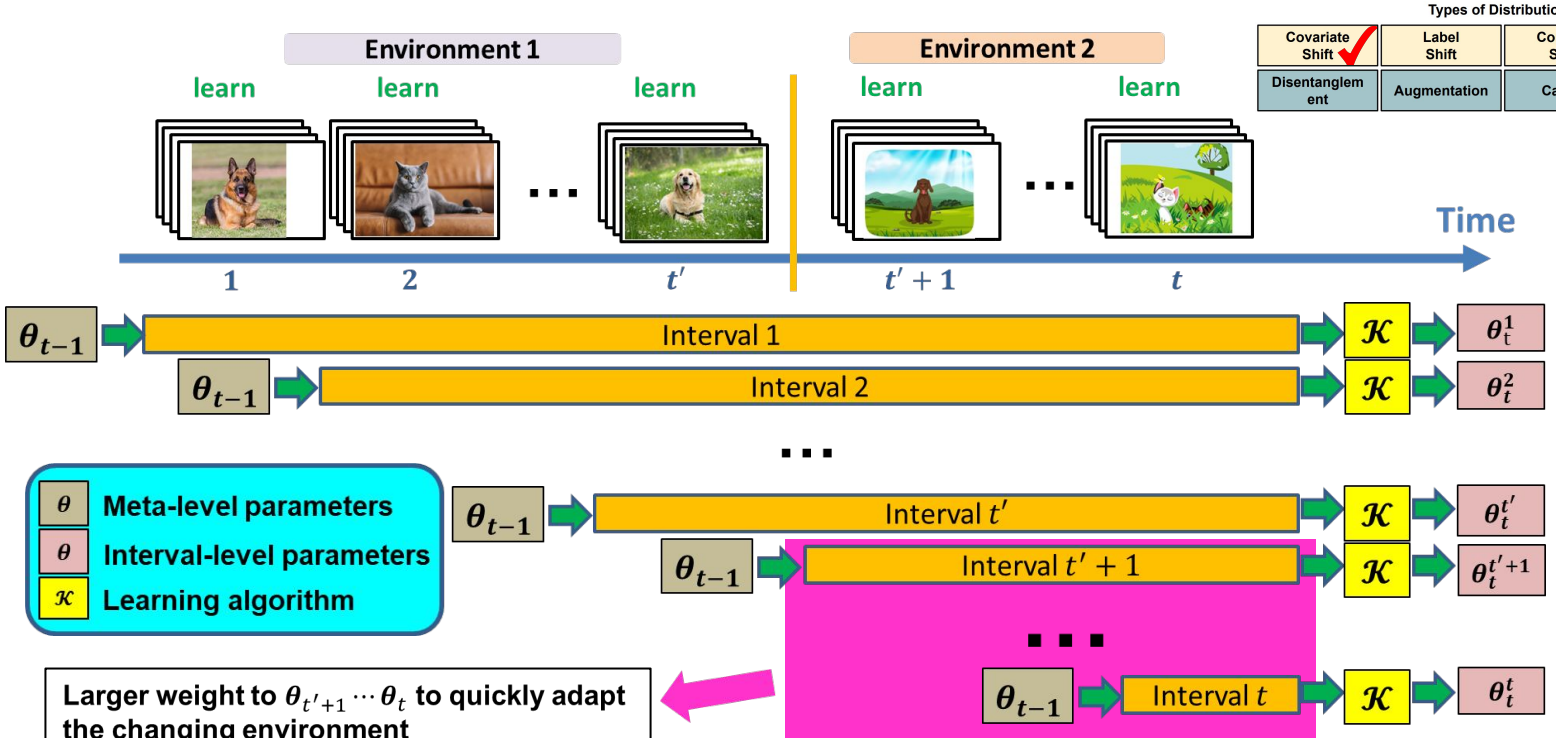
# Fair Disentanglement Online Learning for Changing Environments (FairDolce)





1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Christian Earl Grant, Feng Chen. Towards Fair Disentangled Online Learning for Changing Environments. KDD, 2023.

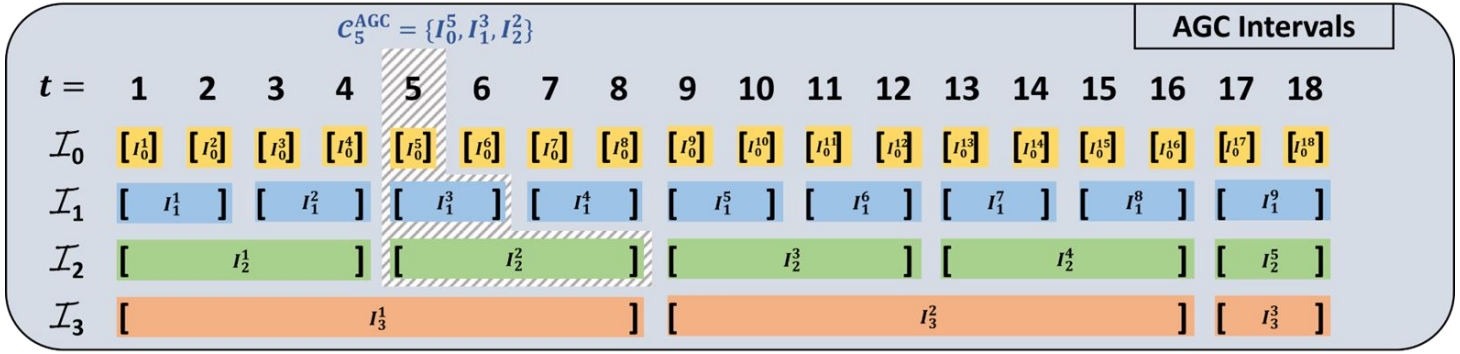
# Setting of Dynamic Intervals (DI)






1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Dynamic Environment Responsive Online Meta-Learning with Fairness Awareness. TKDD, 2024.

# Setting of Adaptive Geometric Covering (AGC) Intervals



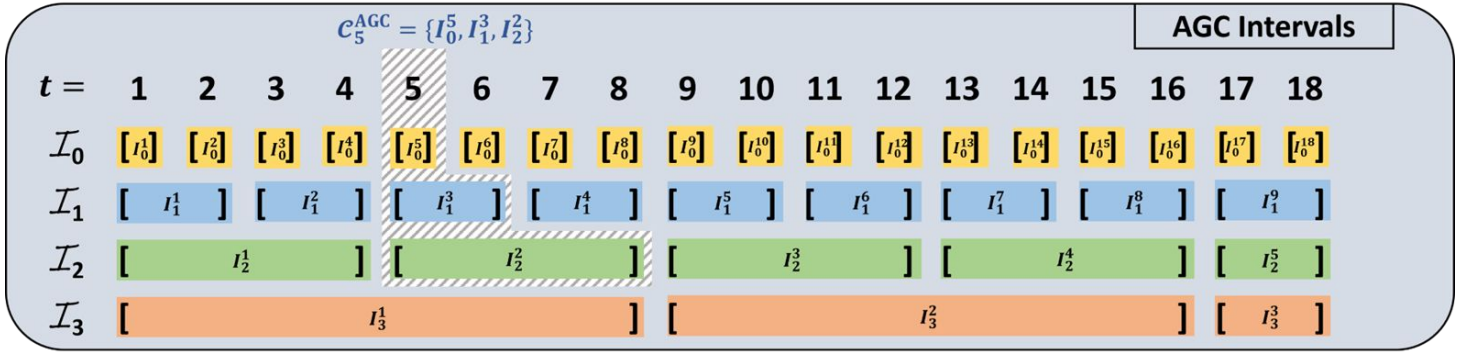

- AGC divides intervals into $N$ sets of intervals (*i.e.*, $\{\mathcal{I}_k\}_{k=0}^{N-1}$), where $N = \lfloor \log_2^T \rfloor$.
- Each $\mathcal{I}_k$ contains equal length intervals, where $\mathcal{I}_k = \left\{ I_k^i \middle| [(i-1)\cdot 2^k + 1,\ \min\{T, i\cdot 2^k\}] : i \in \mathbb{N} \right\}$.
- At each time $t$, a target set $\mathcal{C}_t$ is selected, where $\mathcal{C}_t$ includes all intervals starting with $t$. It mathematically defined as $\mathcal{C}_t = \{I | I \in \mathcal{I}, t \in I, (t-1) \notin I\}$.
- AGC requires $T$ needs to be known and fixed in advance.

Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting ✓ | Others |

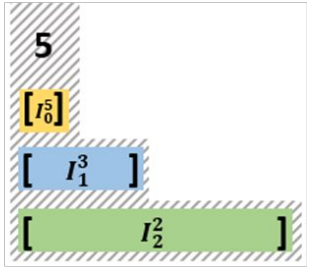

1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Adaptive Fairness-Aware Online Meta-Learning for Changing Environments. KDD, 2022.

# Setting of Adaptive Geometric Covering (AGC) Intervals




For example, when $t = 5$

$\mathcal{C}_5 = \{[I_0^5], [I_1^3], [I_2^2]\}$

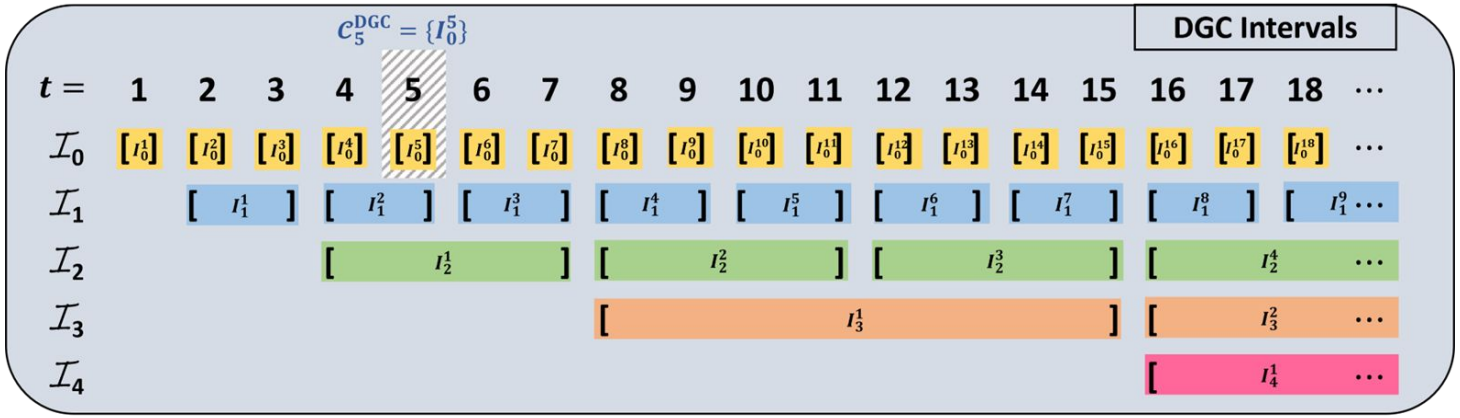



Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting ✓ | Others |


1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Adaptive Fairness-Aware Online Meta-Learning for Changing Environments. KDD, 2022.

# Setting of Dynamic Geometric Covering (DGC) Intervals



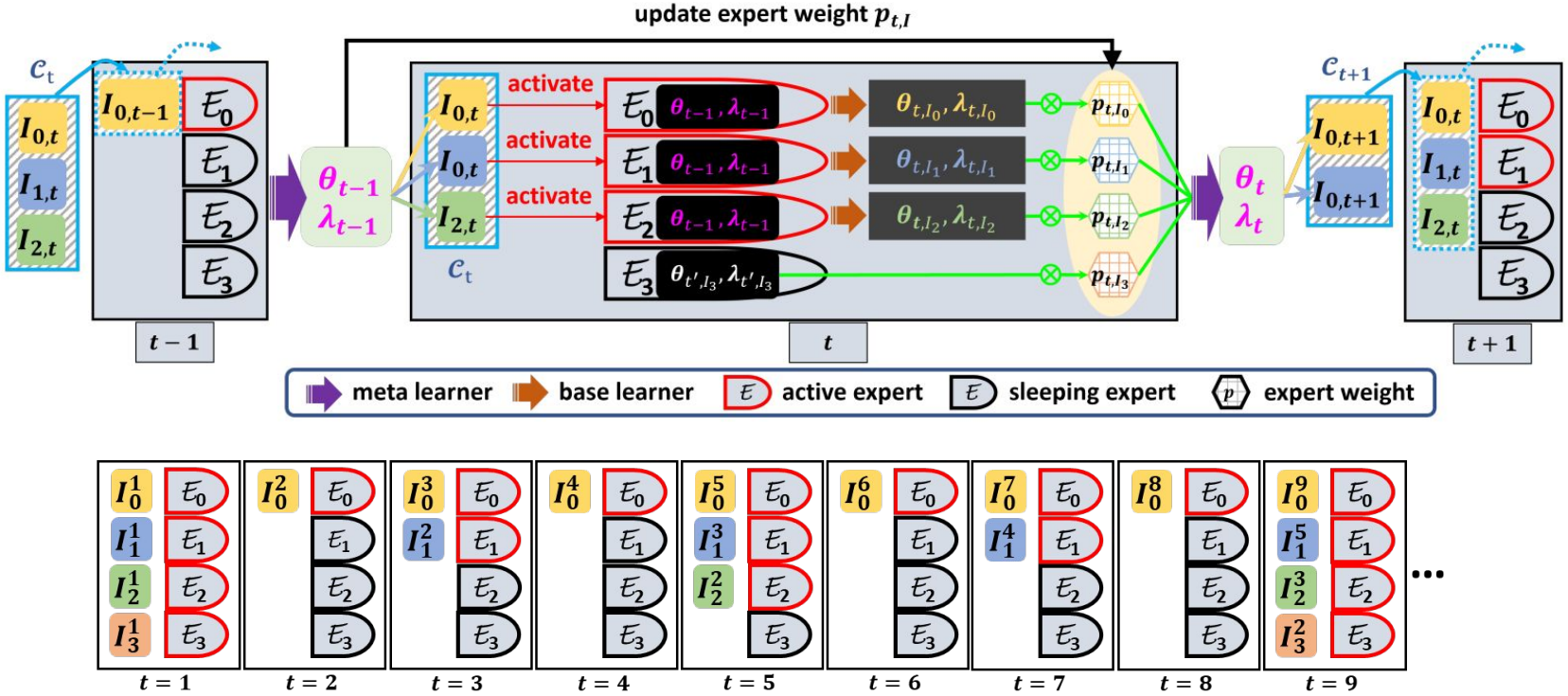

- DGC **dynamically** divides intervals into $\lfloor \log_2^t \rfloor - 1$ sets of intervals.
- Each $\mathcal{I}_k$ contains equal length intervals, where $\mathcal{I}_k = \left\{ I_k^i \middle| [i \cdot 2^k,\ (i+1) \cdot 2^k - 1] : i \in \mathbb{N} \right\}$.
- At each time $t$, a target set $\mathcal{C}_t$ is selected, where $\mathcal{C}_t$ includes all intervals starting with $t$. It mathematically defined as $\mathcal{C}_t = \{I | I \in \mathcal{I}, t \in I, (t-1) \notin I\}$.

Types of Distribution Shifts and Methods

| | | | | |
|---|---|---|---|---|
| Covariate Shift ✓ | Label Shift | Concept Shift | Demographic Shift | Dependence Shift |
| Disentanglement | Augmentation | Causal | Reweighting ✓ | Others |

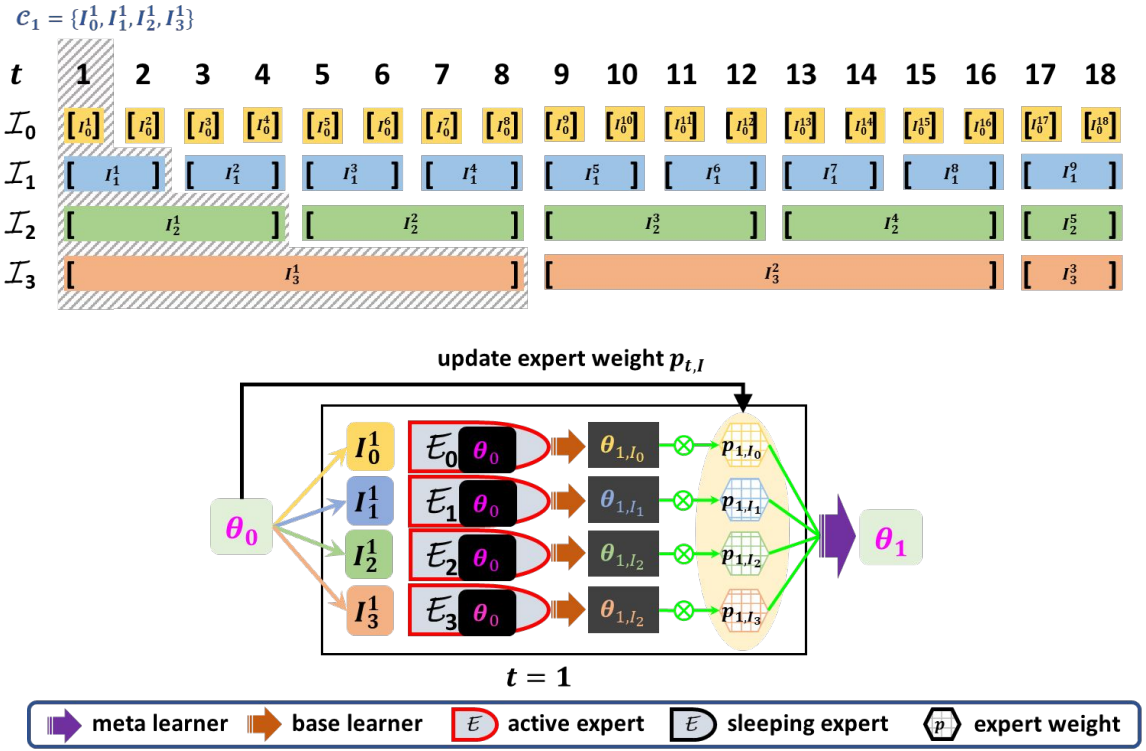

1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Dynamic Environment Responsive Online Meta-Learning with Fairness Awareness. TKDD, 2024.

# Learning Experts for AGC and DGC



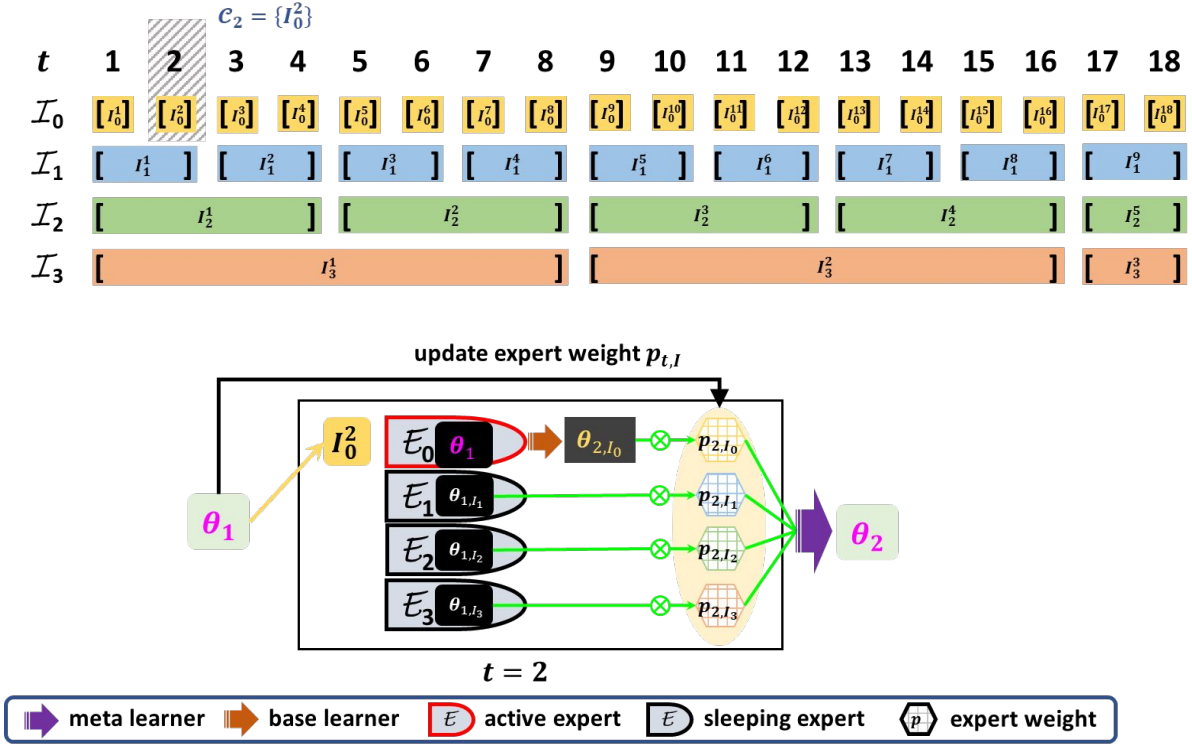

1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Adaptive Fairness-Aware Online Meta-Learning for Changing Environments. KDD, 2022.

# Learning with Experts (When t=1)



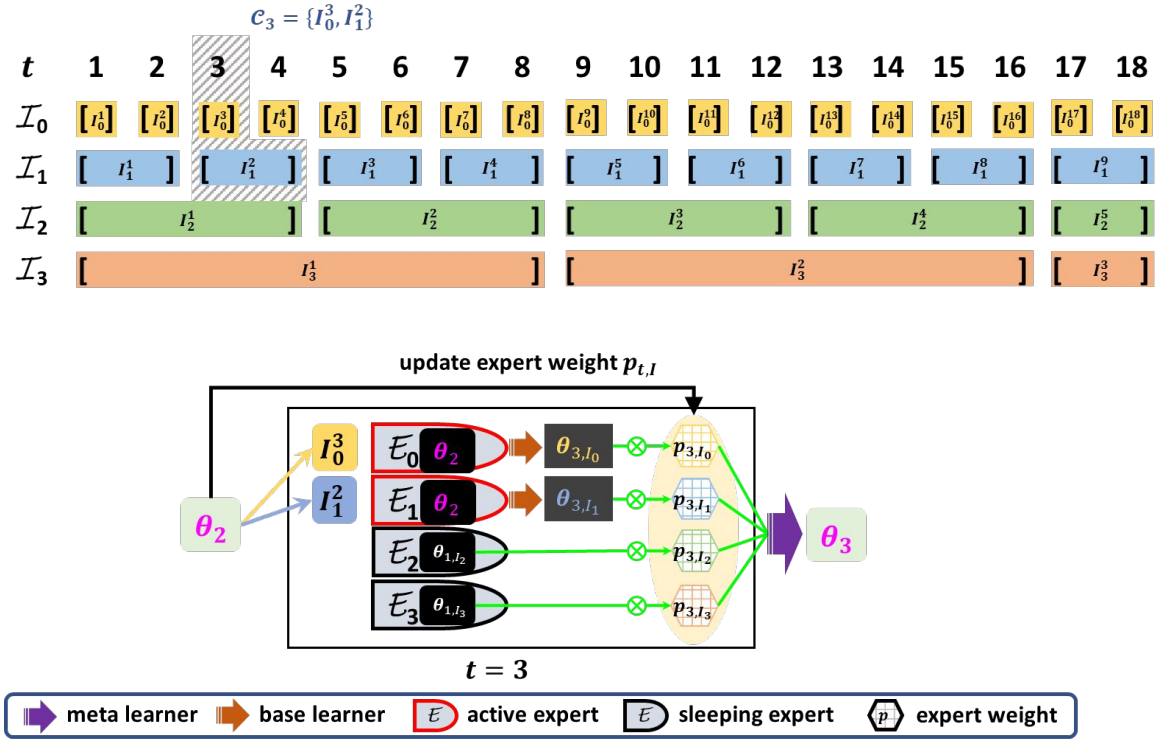

1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Adaptive Fairness-Aware Online Meta-Learning for Changing Environments. KDD, 2022.

# Learning with Experts (When t=2)




1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Adaptive Fairness-Aware Online Meta-Learning for Changing Environments. KDD, 2022.

# Learning with Experts (When t=3)




1. Chen Zhao, Feng Mi, Xintao Wu, Kai Jiang, Latifur Khan, Feng Chen. Adaptive Fairness-Aware Online Meta-Learning for Changing Environments. KDD, 2022.

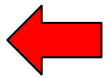
# Outline

**Part I:** Introduction to Fairness 

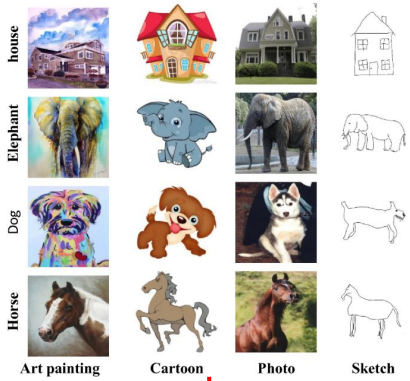
**Part II:** Distribution Shift Undermines Fairness 

**Part III:** Mitigating Unfairness under Distribution Shift (Offline) 

**Part IV:** Mitigating Unfairness under Distribution Shift (Online) 

**Part V:** Open Challenges and Beyond 

# Benchmark Dataset for Fairness Learning under Distribution Shifts

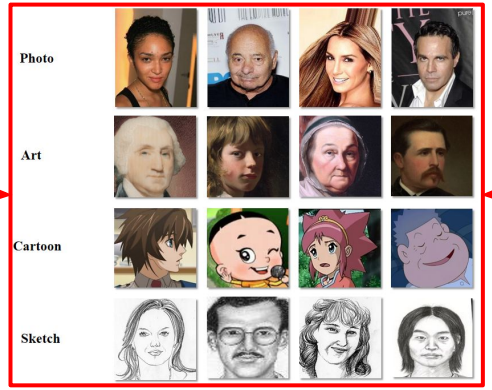



**PACS Dataset**

- 4 Domains
- 7 Classes
- 9991 Images


**CelebA Dataset**

- 40 Annotations
- 202K Images
- Fairness

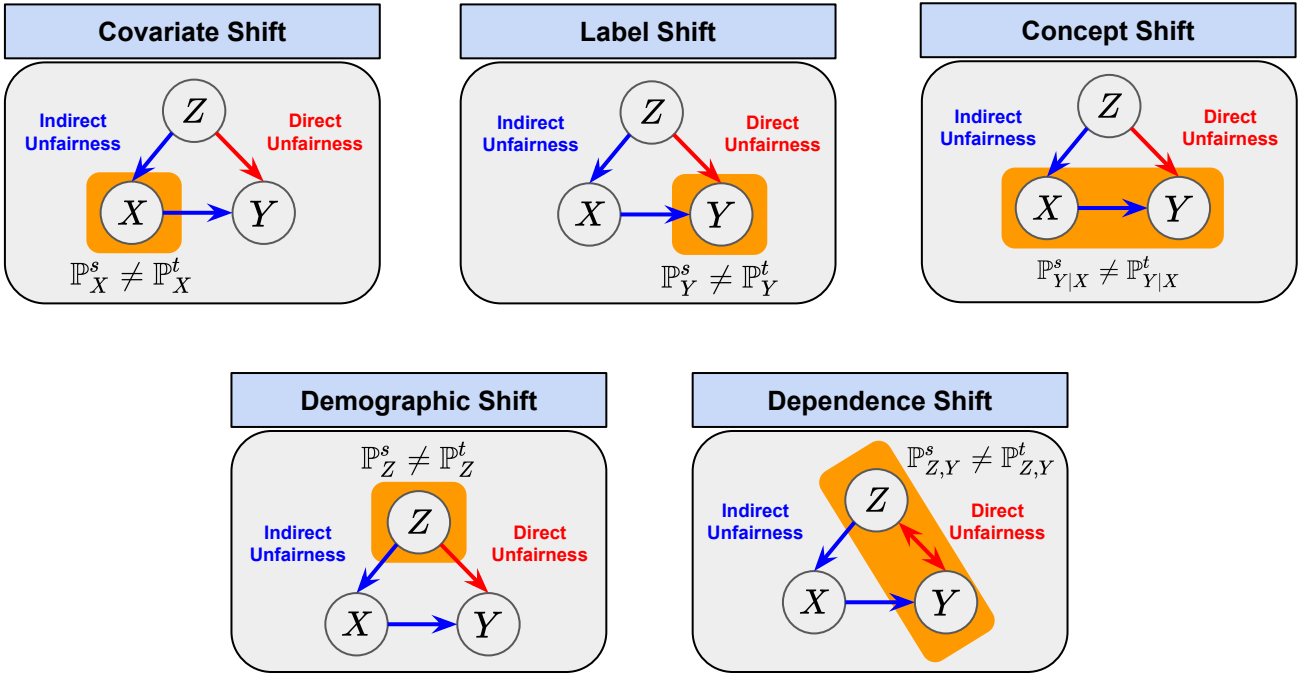



**Open Challenge 1:**

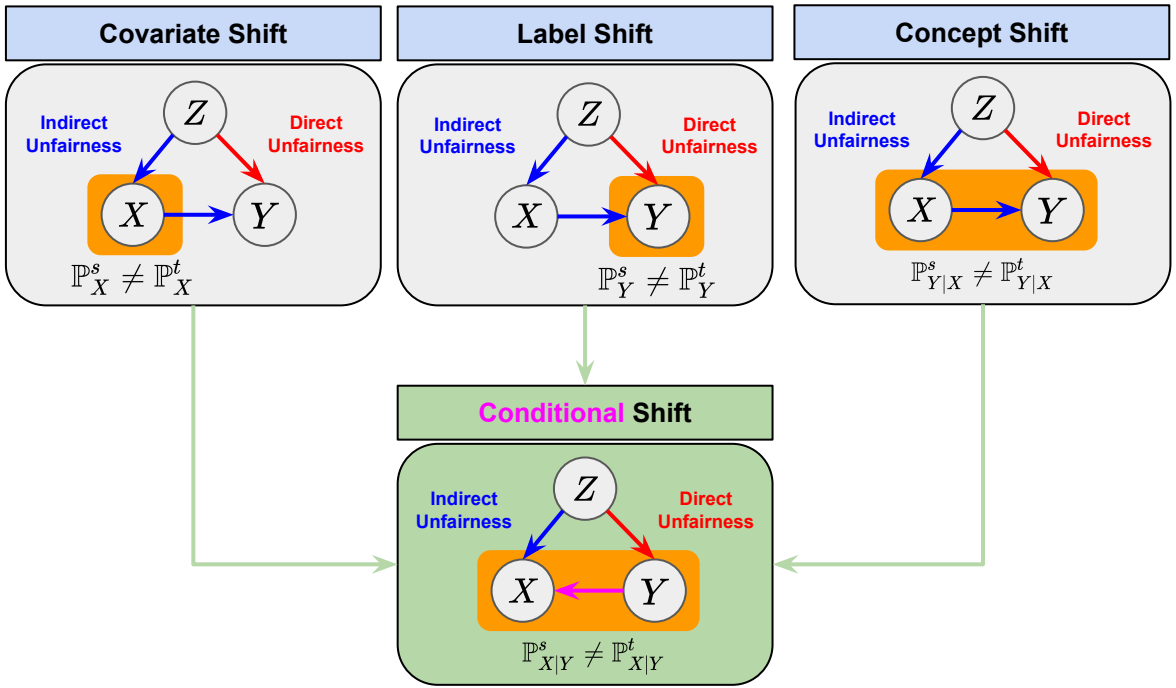
- Fairness-related Attribute Annotations
- 100K - 200K Images
- Multiple Domains
- Various Types of Shifts

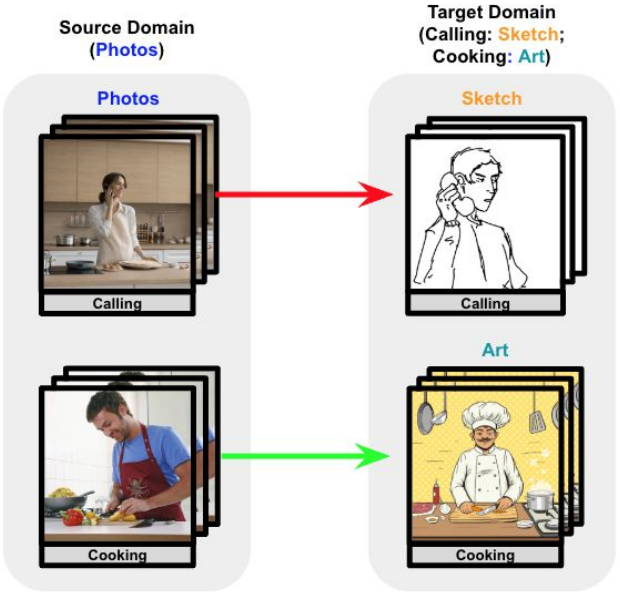

1. Da Li, Yongxin Yang, Yi-Zhe Song, Timothy M. Hospedales. Deeper, Broader and Artier Domain Generalization. ICCV, 2017.
2. Liu, Ziwei and Luo, Ping and Wang, Xiaogang and Tang, Xiaoou. Deep Learning Face Attributes in the Wild. ICCV, 2015.

# Algorithmic Fairness under Five Types of Distribution Shifts



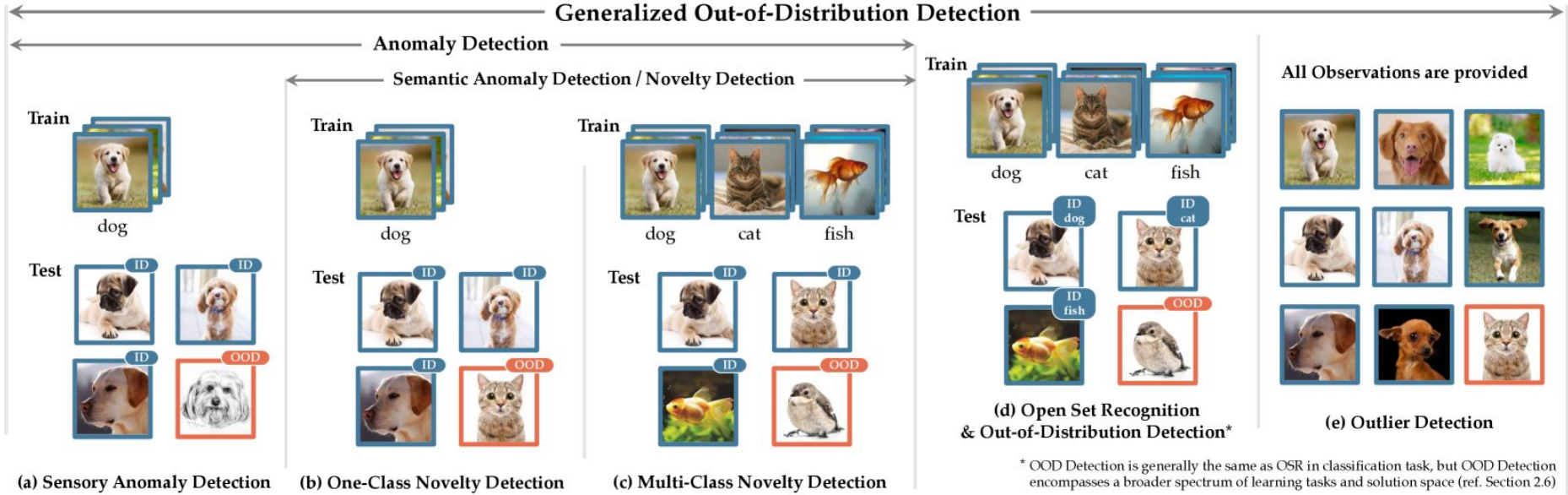
# Algorithmic Fairness under Conditional Shift



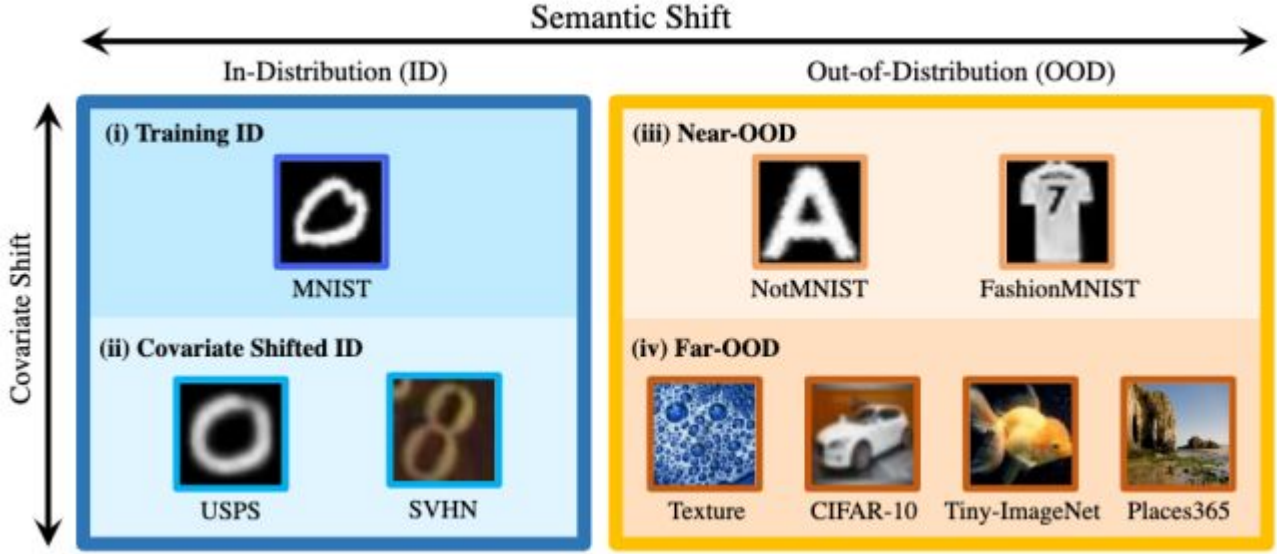

1. Xiaofeng Liu, Bo Hu, Linghao Jin, Xu Han, Fangxu Xing, Jinsong Ouyang, Jun Lu, Georges EL Fakhri, Jonghye Woo. Domain Generalization under Conditional and Label Shifts via Variational Bayesian Inference. IJCAI, 2021.

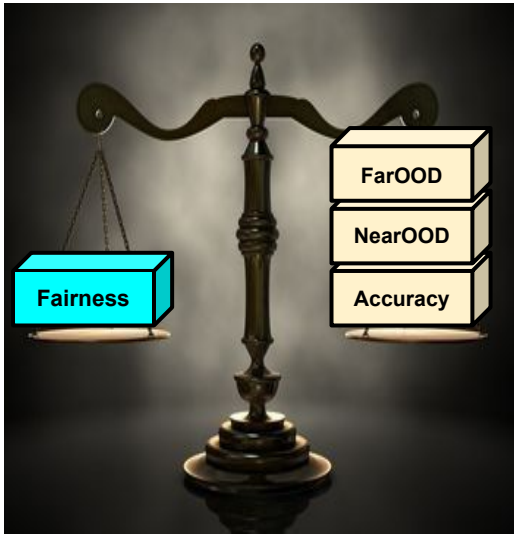
# Algorithmic Fairness for Out-of-Distribution (OOD) Detection




1. Jingkang Yang, Kaiyang Zhou, Yixuan Li, Ziwei Liu. Generalized Out-of-Distribution Detection: A Survey. IJCV, 2024.

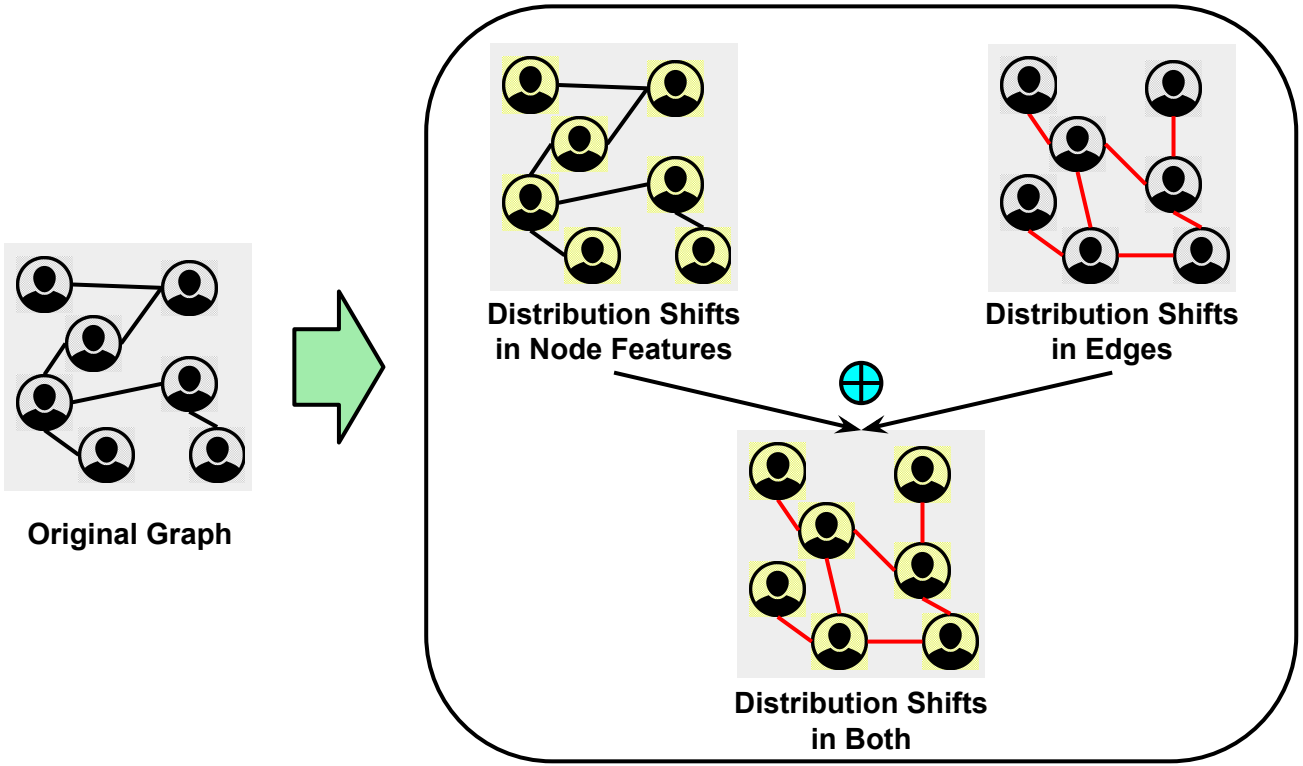
# Algorithmic Fairness for Out-of-Distribution (OOD) Detection



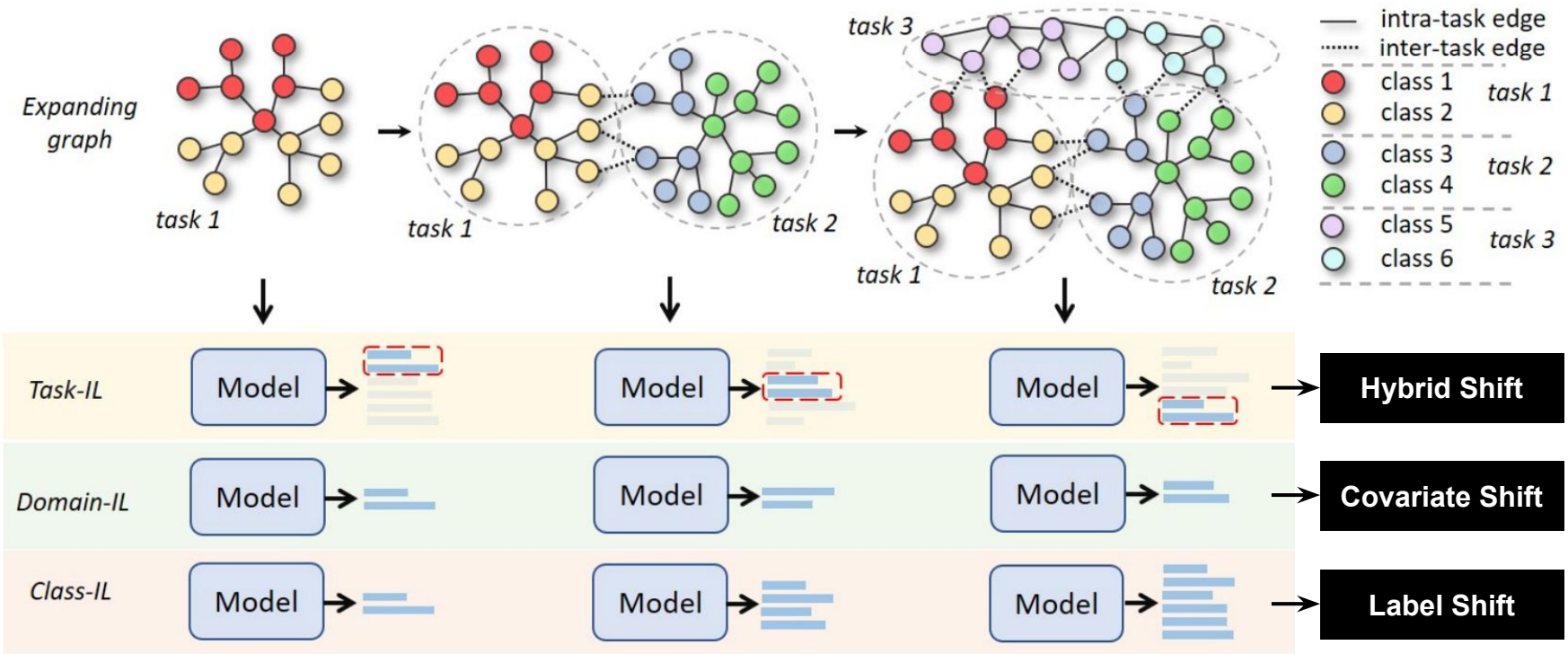




1. Jingkang Yang, Kaiyang Zhou, Ziwei Liu. Full-Spectrum Out-of-Distribution Detection. IJCV, 2023.

# Algorithmic Fairness Domain Generalization on Graphs



# Algorithmic Fairness on Incremental Learning on Graphs




1. Xikun Zhang, Dongjin Song, Dacheng Tao. Continual Learning on Graphs: Challenges, Solutions, and Opportunities. arXiv, 2024.

IEEE BIG DATA 2024
December 15-18 • Washington DC, USA

Baylor University

UNIVERSITY OF ARKANSAS

# Supervised Algorithmic Fairness in Distribution Shifts

Tutorial at the 2024 IEEE International Conference on Big data (IEEE BigData 2024)


**Chen Zhao**[1]
chen_zhao@baylor.edu

**Xintao Wu**[2]
xintaowu@uark.edu

[1]Department of Computer Science, Baylor University
[2]Department of Electrical Engineering and Computer Science, University of Arkansas


**December 15th, 2024, Washington D.C., USA**