# Defining and Contextualizing Fairness

## Your score: 12/12

Go back

1. When implementing a loan approval algorithm where historical data shows different default rates across demographic groups, which approach most effectively addresses the inherent trade-offs identified by fairness impossibility theorems?

A. ◯ Implement all fairness constraints simultaneously with equal weighting to demonstrate comprehensive commitment to fairness.

B. ◯ Create a Pareto frontier visualizing the trade-offs between calibration, equal opportunity, and demographic parity; then select an operating point based on explicit analysis of historical lending patterns, regulatory requirements, and business constraints.

C. ◯ Always prioritize demographic parity over other fairness metrics since it's the only metric that guarantees equal outcomes across groups.

D. ◯ Rely exclusively on the metric that's easiest to implement technically, since impossibility theorems prove that all fairness approaches have equivalent limitations.

2. A company is developing an algorithmic lending system that will operate across multiple jurisdictions, including the United States and European Union. Which

approach to protected attributes most accurately reflects the current legal consensus for ensuring compliance?

A. ◯ The system should be "blind" to all protected attributes, removing them from the data to prevent explicit discrimination.

B. ◯ The system should consider protected attributes exclusively during a post-processing fairness evaluation phase but not during model training.

C. ◯ The system should collect protected attribute data for testing disparate impact but may need different implementations across jurisdictions due to conflicting requirements regarding the use of protected attributes in decision-making.

D. ◯ The system should standardize on demographic parity across all protected attributes as the mathematical fairness definition that best satisfies all relevant legal frameworks.

3. What is the most accurate characterization of the relationship between different mathematical fairness definitions based on current impossibility theorems?

A. ◯ With sufficient data and computational resources, all desirable fairness definitions can be simultaneously satisfied through advanced multi-objective optimization techniques.

B. ◯ Mathematical impossibility results prove that certain combinations of fairness criteria cannot be simultaneously satisfied in most real-world scenarios, requiring explicit trade-off decisions.

C. ◯ The incompatibility between fairness definitions is primarily a practical implementation issue that can be resolved through better algorithm

design rather than a fundamental mathematical limitation.

D. ◯ Fairness definitions are only incompatible when protected attributes are highly correlated with legitimate predictive features; in low-correlation scenarios, all fairness criteria can be simultaneously satisfied.

4. In developing a loan approval system, stakeholders disagree about appropriate fairness metrics. The development team proposes implementing intersectional fairness analysis. Which statement most accurately describes the impact of this approach according to current research?

A. ◯ Intersectional analysis will resolve stakeholder disagreements by identifying a universal fairness definition that protects all demographic subgroups simultaneously.

B. ◯ Intersectional analysis will increase fairness for all groups by ensuring that the model satisfies demographic parity across all possible subgroup combinations.

C. ◯ Intersectional analysis will reveal potentially hidden fairness disparities at demographic intersections that single-attribute analysis might miss, while still requiring explicit trade-off decisions between competing fairness definitions.

D. ◯ Intersectional analysis will demonstrably reduce bias by removing all protected attributes and their proxies from the model to ensure colorblind fairness for all groups.

5. In a legal challenge alleging algorithmic discrimination in an employment screening tool, the defending company demonstrates that: (1) the algorithm's

features are strongly predictive of job performance based on validated studies, (2) all features serve necessary business functions, and (3) no alternative algorithm configuration could achieve similar performance with less disparate impact. According to current U.S. anti-discrimination law, what would likely be the outcome of this analysis?

A. ◯ The company would lose the case because any statistically significant disparate impact constitutes illegal discrimination regardless of business necessity.

B. ◯ The company would win the case because the business necessity defense and absence of less discriminatory alternatives generally satisfy the legal requirements under disparate impact doctrine, even if some disparities remain.

C. ◯ The company would win only if they prove they did not know the algorithm would create disparities, demonstrating absence of discriminatory intent.

D. ◯ The company would lose because algorithmic systems are held to a strict liability standard that prohibits any disparate outcomes regardless of business justification.

6. A university admissions office is building a predictive model to identify applicants likely to succeed academically. Historical admissions data shows different acceptance rates and academic performance patterns across demographic groups. Which statement most accurately represents the implications of fairness impossibility theorems for this scenario?

A. ◯ The university should abandon algorithmic approaches entirely since fairness impossibility theorems prove that all prediction systems will

inevitably discriminate.

B. ◯ The university should use demographic parity as the only fairness metric because it's mathematically guaranteed to ensure equal representation regardless of qualifications.

C. ◯ The university cannot simultaneously achieve calibration (equal meaning of predicted success probabilities across groups), balance for successful students (similar scores for those who succeed), and balance for unsuccessful students (similar scores for those who don't succeed) unless the base rates of academic success are identical across groups.

D. ◯ The university should prioritize individual fairness exclusively since group fairness metrics are mathematically impossible to satisfy.

7. Which of the following most accurately characterizes how historical discrimination patterns persist across technological transitions according to current scholarship?

A. ◯ Historical biases typically diminish as technologies advance, with newer computational systems inherently reducing discrimination through increased precision and objectivity

B. ◯ Discrimination patterns transform but persist across technological transitions, often becoming encoded in new technologies through problem formulations, data practices, and optimization choices that reflect existing social hierarchies

C. ◯ Historical patterns are primarily relevant to legacy systems but largely eliminated in modern AI through technical advances like feature selection algorithms and regularization techniques

D. ◯ Technological transitions create disruptions that typically reset discrimination patterns, with biases in new technologies emerging independently rather than inheriting historical patterns

8. When developing a hiring algorithm that must balance multiple stakeholder perspectives on fairness, which approach most accurately reflects the current technical consensus?

A. ◯ Implement demographic parity constraints to ensure equal representation across all protected groups, as this is the most broadly accepted fairness definition.

B. ◯ Optimize for maximum prediction accuracy first, then apply post-processing adjustments to correct for any observed disparities across protected groups.

C. ◯ Conduct a structured analysis of context-specific factors—including historical discrimination patterns, stakeholder perspectives, and domain requirements—before selecting and prioritizing among potentially conflicting fairness definitions.

D. ◯ Select the fairness definition with the strongest legal precedent in employment law to minimize compliance risks, even if it does not address all stakeholder concerns.

9. In the context of developing a hiring algorithm where historical data shows different representation across demographic groups, what is the most accurate technical characterization of the tension between individual fairness and demographic parity?

A. ◯ Individual fairness and demographic parity are mathematically compatible as long as the similarity metric is carefully designed to incorporate historical context.

B. ◯ Individual fairness fundamentally requires treating similar candidates similarly based on merit, while demographic parity may require selecting candidates with different qualifications to ensure equal representation, creating an inherent tension when qualification distributions differ across groups due to historical factors.

C. ◯ Individual fairness is always the preferred metric because it avoids the statistical infeasibility issues that make demographic parity impossible to implement.

D. ◯ Demographic parity is always preferable because individual fairness metrics are too subjective to implement effectively.

10. In developing a resume screening algorithm for technical roles, a team discovers that the mathematical encoding of educational background creates larger vector distances between international universities and elite U.S. institutions than between elite and non-elite U.S. institutions, despite similar educational quality. Applying the concept of codification of social categories, which approach most appropriately addresses this issue?

A. ◯ Remove educational institution features entirely from the model to eliminate any potential bias based on educational background.

B. ◯ Apply dimensionality reduction techniques like PCA to educational features to create a more compact representation that minimizes differences.

C. ◯ Develop an encoding approach informed by educational quality metrics rather than institutional prestige, validate this encoding through outcomes analysis across demographic groups, and incorporate uncertainty measures for institutions with limited representation in the data.

D. ◯ Standardize all educational institutions to binary values indicating only whether the candidate has the required degree level, regardless of institution.

11. A healthcare company is developing an AI system to prioritize patients for specialized care, trained on historical treatment data. Which approach most effectively applies historical pattern analysis to this scenario according to current research?

A. ◯ Conduct a technical analysis of model accuracy across demographic groups while avoiding historical considerations that might introduce subjective biases into the assessment process.

B. ◯ Research whether the training data includes diverse demographic representation and ensure protected attributes are removed from model inputs to prevent historical biases from influencing predictions.

C. ◯ Conduct a focused historical analysis of how medical technologies and classification systems have historically classified and measured disease across demographic groups, examining how these patterns might manifest in the current system through problem formulation, feature selection, and outcome definition.

D. ◯ Apply industry-standard fairness metrics like demographic parity to ensure the system meets established benchmarks, as these metrics

already incorporate relevant historical considerations.

12. A data scientist is developing a natural language processing system for criminal justice risk assessment and notes that the training data contains offense descriptions using different terminology patterns for defendants from different racial backgrounds, with some actions described as "aggressive" versus "assertive" in ways that correlate with race. Which approach best applies the concept of classification politics to address this issue?

A. ◯ Remove all subjective terminology from the dataset, restricting the model to objective numerical features like prior offense counts.

B. ◯ Apply uniform terminology transformations that standardize all language regardless of context to ensure computational consistency.

C. ◯ Examine the historical context of these terminological differences, analyze the implicit classification systems they represent, and either reframe descriptions to eliminate bias-laden terms or develop representations that explicitly account for these documented patterns.

D. ◯ Average the language patterns across groups to create a balanced representation that incorporates all perspectives equally.

Go back