AI: Strategy + Marketing (MGT 853)

Ethics in AI (Session 10)

Vineet Kumar

Yale School of Management
Spring 2025

Project Presentation

- Please add your project title, still missing one group

Project Presentation

- Please add your project title, still missing one group
- Each group gets 12 minutes (maximum) to present their project

Project Presentation

- Please add your project title, still missing one group
- Each group gets 12 minutes (maximum) to present their project
- Presenter notes etc. will not be used in grading. If it's important, you should present it to everyone.

Project Presentation

- Please add your project title, still missing one group
- Each group gets 12 minutes (maximum) to present their project
- Presenter notes etc. will not be used in grading. If it's important, you should present it to everyone.
- Asking questions of other groups also counts in participation

Project Presentation

- Please add your project title, still missing one group
- Each group gets 12 minutes (maximum) to present their project
- Presenter notes etc. will not be used in grading. If it's important, you should present it to everyone.
- Asking questions of other groups also counts in participation
- TAs will allocate groups to presentation slots (see Google sheet)

Project Presentation

- Please add your project title, still missing one group
- Each group gets 12 minutes (maximum) to present their project
- Presenter notes etc. will not be used in grading. If it's important, you should present it to everyone.
- Asking questions of other groups also counts in participation
- TAs will allocate groups to presentation slots (see Google sheet)
- We will also have a mini-lecture on April 29 in the remaining time.

Project Presentation

- Please add your project title, still missing one group
- Each group gets 12 minutes (maximum) to present their project
- Presenter notes etc. will not be used in grading. If it's important, you should present it to everyone.
- Asking questions of other groups also counts in participation
- TAs will allocate groups to presentation slots (see Google sheet)
- We will also have a mini-lecture on April 29 in the remaining time.
- On **May 1**, everyone please bring **laptops** for course evaluation

Converting Business to Prediction Problem: Amazon GO

Business Objective to Prediction Problem

Amazon Go

- Amazon GO Business Problem is "Create a Grocery Store Experience without human employees"
    - Customer scans in to establish identification
    - After entering the store, everything that happens can be solved as ML problems or prediction problems

Y variables (from your submissions)

- Person
- Dollar amount of individual/total purchases
- Product recommendations for customer
- Inventory levels
- Items that will be purchased
- Identify customer is in store
- Add/remove item to cart
- Identify customer left the store and charge cart
- Number of different items
- Type of items purchased
- # of each item selected
- Total $ purchase
- The specific item that a customer picks up
- Other items a customer might want based on the items they have already selected
- Store traffic at different times

Which are "essential" prediction problems?

AI in College Admissions

Recall: AI Decision-Making Framework

Input

Prediction

Judgment

Action

Feedback

Performance

In-class Exercise
Groups of 4 or so

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?
5. What prediction model would you use? What considerations would you use to make this choice?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?
5. What prediction model would you use? What considerations would you use to make this choice?
6. What additional data would you like to add that is not currently included? Why do you think it would be helpful, i.e. for what prediction would it improve accuracy?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?
5. What prediction model would you use? What considerations would you use to make this choice?
6. What additional data would you like to add that is not currently included? Why do you think it would be helpful, i.e. for what prediction would it improve accuracy?
7. How do you evaluate if your AI system is doing a good job?

What is Current Admissions Process?

In-class Exercise
Groups of 4 or so

Show Mission Statements

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?
5. What prediction model would you use? What considerations would you use to make this choice?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?
5. What prediction model would you use? What considerations would you use to make this choice?
6. What additional data would you like to add that is not currently included? Why do you think it would be helpful, i.e. for what prediction would it improve accuracy?

AI model for University Admissions

In class Exercise

1. What is the organizational objective and problem they want to solve?
2. What are the decisions to be made? By whom?
3. What predictions would help support these decisions? i.e. solving what prediction problems with 100% accuracy would greatly help solve the organizational problem?
4. What structured and unstructured data are used to make these predictions?
5. What prediction model would you use? What considerations would you use to make this choice?
6. What additional data would you like to add that is not currently included? Why do you think it would be helpful, i.e. for what prediction would it improve accuracy?
7. How do you evaluate if your AI system is doing a good job?

Role of Big Picture Objective (Mission)

Mission Statements

1. **University of Chicago Booth School of Business**
   "At Booth, our mission is to create knowledge with enduring impact, and influence and educate current and future leaders."
2. **Stanford Graduate School of Business**
   "Our mission is to create ideas that deepen and advance our understanding of management and with those ideas to develop innovative, principled, and insightful leaders who change the world."
3. **Northwestern Kellogg School of Management**
   "Kellogg is a global business school with a mission to educate, equip and inspire leaders who can build strong organizations and wisely leverage the power of markets to create lasting value."
4. **Harvard Business School**
   "The mission of Harvard Business School is to educate leaders who make a difference in the world. Achieving this mission requires an environment of trust and mutual respect, free expression and inquiry, and a commitment to truth, excellence, and lifelong learning."
5. **Wharton School of the University of Pennsylvania**
   "Wharton's pioneering vision was to produce graduates who would become "pillars of the state, whether in private or in public life." The Wharton School maintains a long tradition of educating visionary business leaders in academia, business, government, and not-for-profit organizations."
6. **Dartmouth Tuck School of Business**
   "Tuck develops wise, decisive leaders who better the world through business."
7. **Columbia Business School**
   "Columbia Business School is committed to educating and developing leaders and builders of enterprises who create value for their stakeholders and society at large; we accomplish this through our MBA, MS, PhD, and Executive Education programs."
8. **MIT Sloan School of Management**
   "The mission of the MIT Sloan School of Management is to develop principled, innovative leaders who improve the world and to generate ideas that advance management practice."
9. **Cornell SC Johnson Graduate School of Management**
   "Together, our three schools transcend the traditional boundaries of business education and research to transform critical thinking into practical solutions. We collaborate within SC Johnson College, across Cornell, and beyond to mobilize diverse expertise, generate world-class knowledge, inspire students, and positively impact society."
10. **University of Michigan Ross School of Business**
    "Michigan Ross is committed to building a better world through business. Through our powerful ideas, purpose-driven leaders, and positive impact, we empower business leaders to make a positive difference in the world."
11. **UC Berkeley Haas School of Business**
    "Our mission is to help extraordinary people achieve great things."
12. **Yale School of Management**
    "Our mission is to educate leaders for business and society."

AI In Admissions: What to Predict?

(U), (R) and (G) in Admissions

Your View of Using AI in Admissions

Select the view closest to yours. Assume we specify clear objectives at the outset:

1. AI / ML should be used as a pure complement to our current process
2. AI / ML can substitute *some* parts of the admission process done by humans but not others
3. AI / ML should only be used if everything is explainable and interpretable
4. AI / ML should not be used for such an important decision

Your view

?

AI in Admissions: Takeaways – 1

- Using AI in admissions requires us to convert organizational / business problem to a set of prediction problems

AI in Admissions: Takeaways – 1

- Using AI in admissions requires us to convert organizational / business problem to a set of prediction problems
  - Solving prediction problem helps solve organizational problem

AI in Admissions: Takeaways – 1

- Using AI in admissions requires us to convert organizational / business problem to a set of prediction problems
  - Solving prediction problem helps solve organizational problem
- We have identified a large set of y variables we care about that potentially matter in admissions

AI in Admissions: Takeaways – 1

- Using AI in admissions requires us to convert organizational / business problem to a set of prediction problems
  - Solving prediction problem helps solve organizational problem
- We have identified a large set of y variables we care about that potentially matter in admissions
- Human judgment is required in deciding which y variables we care about, and how much – ML cannot do this

AI in Admissions: Takeaways – 1

- Using AI in admissions requires us to convert organizational / business problem to a set of prediction problems
  - Solving prediction problem helps solve organizational problem
- We have identified a large set of y variables we care about that potentially matter in admissions
- Human judgment is required in deciding which y variables we care about, and how much – ML cannot do this
  - Org. Mission (big picture objective) can help with this, but may need to convert to a lower level objective

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships
  - Experimenting with non-traditional students who are not present in historical data

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships
  - Experimenting with non-traditional students who are not present in historical data
- Generative AI can be used at many stages:

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships
  - Experimenting with non-traditional students who are not present in historical data
- Generative AI can be used at many stages:
  - Outreach to different pools of potential students to improve consideration

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships
  - Experimenting with non-traditional students who are not present in historical data
- Generative AI can be used at many stages:
  - Outreach to different pools of potential students to improve consideration
  - Chatbot to help with the discovery and learning process, e.g. what courses or clubs are available at school X but not school Y

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships
  - Experimenting with non-traditional students who are not present in historical data
- Generative AI can be used at many stages:
  - Outreach to different pools of potential students to improve consideration
  - Chatbot to help with the discovery and learning process, e.g. what courses or clubs are available at school X but not school Y
  - Persuasion to admitted candidates to increase yield

AI in Admissions: Takeaways – 2

- Can (U)nsupervised and (R)einforcement Learning be used?
  - RL can be used in communications,
  - Experiment with scholarships
  - Experimenting with non-traditional students who are not present in historical data
- Generative AI can be used at many stages:
  - Outreach to different pools of potential students to improve consideration
  - Chatbot to help with the discovery and learning process, e.g. what courses or clubs are available at school X but not school Y
  - Persuasion to admitted candidates to increase yield
  - Collective impressions of current students $\Longrightarrow$ future students

Ethics in AI

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?
  - Criminal Justice: Pre-trial bail

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?
  - Criminal Justice: Pre-trial bail

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?
  - Criminal Justice: Pre-trial bail

Major Challenge:
Algorithmic Bias leads to differential performance across groups (think race, gender, age, income)

Humans may be able to override decisions, but...

- often have limited time / energy / attention
- power of defaults from marketing and psychology

Impact of Bias: Human Capital

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 4 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Impact of Bias: Human Capital

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 4 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

ML Problem:

Prediction Problem: Predict quality or fit of applicant (one to five stars)
Input to algorithm: Resume
Decision: Interview or Not
Bias: Women-related words decreased stars

Impact of Bias: Human Capital

Impact of Bias: Human Capital

ML Problem:

Prediction problem (Y): Likelihood of accepting
Input to algorithm (X): Student information
Decision: How much financial aid to offer
Bias: More accurate for higher income

Impact of Bias: Medical Care

Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism

Unclear regulation and a lack of transparency increase the risk that AI and algorithmic tools that exacerbate racial biases will be used in medical settings.

Impact of Bias: Medical Care

Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism

Unclear regulation and a lack of transparency increase the risk that AI and algorithmic tools that exacerbate racial biases will be used in medical settings.

ML Problem:

Prediction problem (Y): who is likely to have a serious condition
Input to algorithm (X): insurance claims, diagnosis codes, etc.
Decision: extra medical attention and care
Bias: For same risk assessment, Black patients sicker than White patients

Impact of Bias: Criminal Justice

Impact of Bias: Criminal Justice

ML Problem:

Prediction problem (Y): likelihood of re-offending
Input to algorithm (X): 137 question survey
Decision: Offer Bail or Not
Bias: Higher FPR among Blacks

- Is Prediction Accuracy a good metric?
- Accuracy cannot distinguish between FP and FN (Type 1 and Type 2)

Where does Algorithmic Bias come in?

Concepts of Fairness in ML

Fairness through Unawareness

Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians

By Claudia Goldin and Cecilia Rouse*

A change in the audition procedures of symphony orchestras—adoption of "blind" auditions with a "screen" to conceal the candidate's identity from the jury—provides a test for sex-biased hiring. Using data from actual auditions, in an individual fixed-effects framework, we find that the screen increases the probability a woman will be advanced and hired. Although some of our estimates have large standard errors and there is one persistent effect in the opposite direction, the weight of the evidence suggests that the blind audition procedure fostered impartiality in hiring and increased the proportion women in symphony orchestras. (*JEL* J7, J16)

Sex-biased hiring has been alleged for many occupations but is extremely difficult to prove. The empirical literature on discrimination, deriving from the seminal contributions of Gary Becker (1971) and Kenneth Arrow (1973), has focused mainly on disparities in earnings between groups (e.g., males and females), given differences in observable productivity-altering characteristics. With the exception of various audit studies (e.g., Genevieve Kenney and Douglas A. Wissoker, 1994; David Neumark et al., 1996) and others, few researchers have been able to address directly the issue of bias in hiring practices.[1] A change in the way symphony orchestras recruit musicians provides an unusual way to test for sex-biased hiring.

* Goldin: Department of Economics, Harvard University, Cambridge, MA 02183; Rouse: Woodrow Wilson School, Princeton University, Princeton, NJ 08544. Rouse acknowledges The National Academy of Education, the NAE Spencer Postdoctoral Fellowship Program, and the Mellon Foundation for financial support. We are indebted to the staff members of the orchestras that gave us access to their audition records and who provided other assistance, and to the musicians who responded to our questionnaire. We are particularly grateful to Joanne Berry, Brigit Carr, Ruth DeSarno, Stefanie Dyson, Josh

The New York Times

CRITIC'S NOTEBOOK

To Make Orchestras More Diverse, End Blind Auditions

If ensembles are to reflect the communities they serve, the audition process should take into account race, gender and other factors.

Share full article

Fairness through Unawareness – 1

Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians

By Claudia Goldin and Cecilia Rouse*

A change in the audition procedures of symphony orchestras—adoption of "blind" auditions with a "screen" to conceal the candidate's identity from the jury—provides a test for sex-biased hiring. Using data from actual auditions, in an individual fixed-effects framework, we find that the screen increases the probability a woman will be advanced and hired. Although some of our estimates have large standard errors and there is one persistent effect in the opposite direction, the weight of the evidence suggests that the blind audition procedure fostered impartiality in hiring and increased the proportion women in symphony orchestras. (*JEL* J7, J16)

Sex-biased hiring has been alleged for many occupations but is extremely difficult to prove. The empirical literature on discrimination, deriving from the seminal contributions of Gary Becker (1971) and Kenneth Arrow (1973), has focused mainly on disparities in earnings between groups (e.g., males and females), given differences in observable productivity-altering characteristics. With the exception of various audit studies (e.g., Genevieve Kenney and Douglas A. Wissoker, 1994; David Neumark et al., 1996) and others, few researchers have been

* Goldin: Department of Economics, Harvard University, Cambridge, MA 02183; Rouse: Woodrow Wilson School, Princeton University, Princeton, NJ 08544. Rouse acknowledges The National Academy of Education, the NAE Spencer

Fairness through Unawareness – 2

The New York Times

CRITIC'S NOTEBOOK

To Make Orchestras More Diverse, End Blind Auditions

If ensembles are to reflect the communities they serve, the audition process should take into account race, gender and other factors.

Share full article

Concepts of Fairness

There are many notions of fairness:

- Unawareness
- Statistical Parity
- Equalized Odds
- Equal Opportunity

Can we satisfy all concepts of fairness?

The (Im)possibility of Fairness: Different Value Systems Require Different Mechanisms For Fair Decision Making

By Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian
Communications of the ACM, April 2021, Vol. 64 No. 4, Pages 136-143
10.1145/3433949
Comments

VIEW AS: SHARE:

Automated decision-making systems (often machine learning-based) now commonly determine criminal sentences, hiring choices, and loan applications. This widespread deployment is concerning, since these systems have the potential to discriminate against people based on their demographic characteristics. Current sentencing risk assessments are racially biased,[4] and job advertisements discriminate on gender.[8] These concerns have led to an explosive growth in fairness-aware machine learning, a field that aims to enable algorithmic systems that are fair by design.

Back to Top

Key Insights

SIGN IN for Full Access
User Name
Password
» Forgot Password?
» Create an ACM Web Account
SIGN IN

ARTICLE CONTENTS:
Introduction
Key Insights
An Example
Spaces: Construct vs. Observe
and Features vs. Decisions
Fairness and Non-Discrimination
Worldviews and Assumptions

It is impossible to simultaneously satisfy all the notions of fairness

How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue

How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes $300

How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes $300
- An unsuccessful loan costs $700

How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes $300
- An unsuccessful loan costs $700
- **What is the bank's profit?**

How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes $300
- An unsuccessful loan costs $700
- **What is the bank's profit?**

How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes $300
- An unsuccessful loan costs $700
- **What is the bank's profit?**

How to trade-off profitability with fairness?

Simulation

See how different notions of fairness lead to different levels of profitability in this simulation:
Google: Attacking Discrimination in ML (link)

Confusion Matrix for Bank Loans

Pareto Frontier

Intersectionality of Race and Gender

Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

Editors: Sorelle A. Friedler and Christo Wilson

Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, IJB-A and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects

who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O'Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus,

Intersectionality of Race and Gender

Figure 1: Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As the examples show, the images are constrained with relatively little variation in pose. The subjects are composed of male and female parliamentarians from 6 countries. On average, Senegalese subjects are the darkest skinned while those from Finland and Iceland are the lightest skinned.

Intersectionality in ML

Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	PPV(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	TPR (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	8.3	3.5	12.9	0.7	**16.3**	7.9	1.3	0.0
Face++	PPV(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	94.0	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	6.0	0.8
	TPR (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	14.9	1.1	16.5	4.7	**23.4**	1.2	7.1	1.1
IBM	PPV(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	TPR (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	14.8	7.9	22.4	3.2	**25.2**	17.7	5.20	0.4

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline
3. Many concepts of fairness: Statistical Parity, Equalized odds, Equal opportunity

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline
3. Many concepts of fairness: Statistical Parity, Equalized odds, Equal opportunity
   - Concepts of fairness are not equivalent

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline
3. Many concepts of fairness: Statistical Parity, Equalized odds, Equal opportunity
   - Concepts of fairness are not equivalent
   - may be opposed to one another (impossibile to satisfy *all*)

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline
3. Many concepts of fairness: Statistical Parity, Equalized odds, Equal opportunity
   - Concepts of fairness are not equivalent
   - may be opposed to one another (impossibile to satisfy *all*)
4. Need to define our goals and make them specific, then evaluate what concept of fairness is appropriate.

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline
3. Many concepts of fairness: Statistical Parity, Equalized odds, Equal opportunity
   - Concepts of fairness are not equivalent
   - may be opposed to one another (impossibile to satisfy *all*)
4. Need to define our goals and make them specific, then evaluate what concept of fairness is appropriate.
5. Pareto frontier characterizes the tradeoff between accuracy and fairness

Fairness & Bias in ML – Takeaways 1

1. Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
2. Where can bias enter into an ML system?
   - Can happen at many (possibly all) points in the ML pipeline
3. Many concepts of fairness: Statistical Parity, Equalized odds, Equal opportunity
   - Concepts of fairness are not equivalent
   - may be opposed to one another (impossibile to satisfy *all*)
4. Need to define our goals and make them specific, then evaluate what concept of fairness is appropriate.
5. Pareto frontier characterizes the tradeoff between accuracy and fairness
   - If within frontier, may not be a tradeoff. At frontier, there is a tradeoff.

Fairness & Bias in ML – Takeaways 2

1. Intersectionality: Combinations of groups (intersections)

Fairness & Bias in ML – Takeaways 2

1. Intersectionality: Combinations of groups (intersections)
   - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)

Fairness & Bias in ML – Takeaways 2

1. Intersectionality: Combinations of groups (intersections)
   - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
2. Interventions to reduce bias:

Fairness & Bias in ML – Takeaways 2

1. Intersectionality: Combinations of groups (intersections)
   - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
2. Interventions to reduce bias:
   - Pre-processing (Data)

Fairness & Bias in ML – Takeaways 2

1. Intersectionality: Combinations of groups (intersections)
   - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
2. Interventions to reduce bias:
   - Pre-processing (Data)
   - In-processing (Algorithm)

Fairness & Bias in ML – Takeaways 2

1. Intersectionality: Combinations of groups (intersections)
   - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
2. Interventions to reduce bias:
   - Pre-processing (Data)
   - In-processing (Algorithm)
   - Post-processing (Decision)