1a

A: Accuracy (ACC): A model that improves accuracy would assist judges and prosecutors make better sentence judgments and decrease hazardous offender release. An accurate model would lower the possibility of defendants being incorrectly designated high risk and sentenced longer than necessary.
B: Positive predictive value (PPV): A model that maximizes PPV would eliminate false positives—high-risk offenders who don't reoffend—for judges and prosecutors. This will eliminate unnecessary incarceration and focus the criminal justice system's limited resources on real criminals.
C: False positive rate (FPR): A model that optimizes FPR would lower the amount of false positives, or high-risk defendants who get lengthier sentences. This would lower the number of people wrongfully imprisoned and subjected to high-risk offender repercussions.
D: False negative rate (FNR): Optimizing FNR would lower the number of dangerous criminals mislabeled as low risk and given reduced sentences. This would prevent dangerous criminals from being freed early.
E: Statistical parity (SP): Defendants would benefit from a methodology that maximizes SP since it would lessen disparate effect and prevent unjust treatment based on race or ethnicity. Improving SP would reduce prejudice and enhance legal equality. Optimizing SP at the cost of other measures like PPV or FPR may impair the model's accuracy and designate more people high risk who don't reoffend.

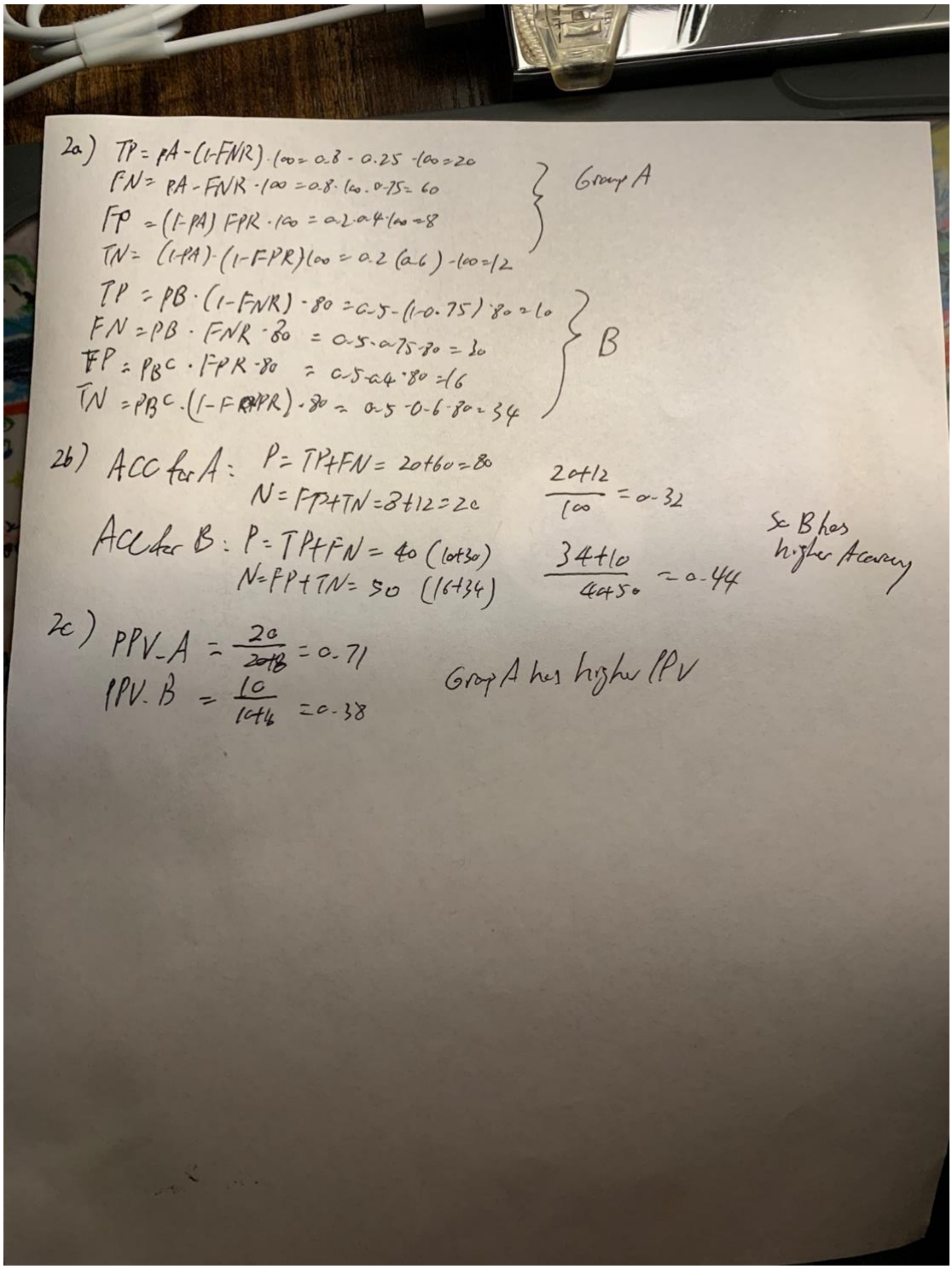
1b

A. Pre-existing bias:
Example: The resume screening tool may be trained on past data of successful TechCorp workers, who are mostly white males, reinforcing data prejudices.
Stakeholder group: Women, persons of color, and other underrepresented groups in the data may have less job possibilities.
Intervention: Train the algorithm with candidates of varied genders, races, and ethnicities. To prevent TechCorp-specific biases, add external data.
B. Technological bias:
Example: The resume screening tool may discriminate against individuals with gaps in their employment history or who worked for unknown organizations, even though these variables may not predict job success.
Stakeholder group: Applicants have non-traditional job experiences, such as those who worked for startups or non-profits or took time off to care for family.
Intervention: Evaluate applicants based on employment performance, not irrelevant characteristics like their past employers. Regular algorithm audits may help identify technical biases.

C. Emergent bias:
Example: Even if gender and race are not expressly included in the data, the resume screening tool may exploit these patterns to make biased choices.
Stakeholder group: Candidates who may face emerging prejudices, especially those from

historically underrepresented groups.

Intervention: Identify and minimize emerging biases using transparency and explainability. To guarantee varied viewpoints, a wide set of stakeholders may assist create and implement the algorithm.



2d

it demonstrates that when there is a difference in the base rates for the outcome of interest across groups, no classifier can simultaneously achieve equal positive predictive value, equal false positive rates, and equal false negative rates across groups. In the given example, Group

A has a higher base rate of positive outcomes compared to Group B, which means that even if both groups have equal false positive and false negative rates, they will have different positive predictive values, violating the requirement for fairness as defined by Chouldechova.

3

Memo on Problems of Fairness Presented at the "AI for Whom?" Lecture.

In the "AI for Whom?" lecture, predictive policing and face recognition technology in the United States are discussed. The stated objective of predictive policing is to employ machine learning algorithms to evaluate previous crime data and anticipate the probable locations of future crimes. Faces collected in real-time are matched against photographs in a database using facial recognition technology to identify persons of interest.

Many parties are involved in these technologies. The public and law enforcement agencies are the major stakeholders in predictive policing technologies. Predictive policing has the potential to decrease crime and make communities safer. Yet, the use of predictive policing has had negative effects on communities of color. The historical crime statistics used by machine learning algorithms are skewed toward police methods that disproportionately targeted communities of color in the past. Thus, predictive police technologies may reinforce systemic biases in policing.

Included among the stakeholders of facial recognition technology are law enforcement organizations, private businesses, and the people whose faces are being studied. Law enforcement organizations employ facial recognition technology to identify persons of interest, but it has the potential to be used for more intrusive monitoring reasons. Using face recognition technology for security reasons, such as unlocking phones or gaining access to bank accounts, is of interest to private firms. Those whose faces are being examined may be concerned about their privacy.

Differential treatment and/or differential outcomes in the data science application are instances of inherent bias. The systematic biases in policing are reflected in the crime statistics utilized by predictive police systems. Predictive policing has uneven effects on communities of color, who are disproportionately targeted by law enforcement. Similarly, persons with darker skin tones have a greater mistake rate with face recognition technology, which is an example of technical prejudice.

The implementation of these data science applications may result in the worsening of systemic biases in police, greater monitoring, and privacy breaches. There is the potential for predictive police algorithms to enhance current trends of overpolicing in communities of color. The use of face recognition technology for intrusive monitoring purposes might lead to privacy breaches and could be used to target persons on the basis of their race or other protected features.

The "AI for Whom?" presentation concludes by highlighting the possible damages and prejudices connected with predictive policing and face recognition technology in the United States. These technologies have the potential to increase systemic biases in law enforcement and result in privacy breaches. It is essential to identify the stakeholders involved and to examine the possible downsides of these technologies, as well as the technological and emerging biases that may develop.