BIAS IN EVALUATING CODE CONTRIBUTIONS

IDENTIFYING **DIVERSITY** IN ONLINE GROUPS AND CREATING **AWARENESS**

Nadri et al., On the Relationship Between the Developer's Perceptible race and ethnicity and the Evaluation of Contributions in GitHub. **TSE'21**

Rodriguez-Perez et al., Perceived diversity in software engineering: a systematic literature review. **EMSE'21**

Nadri et al., I*nsights into Non-merged Pull Requests in GitHub: Is There Evidence of Bias Based on Perceptible Race?*. **IEEE Software'21**

Reza Nadri

Gema Rodríguez-Pérez

Meiyappan Nagappan

OPEN SOURCE (**OS**) PROJECTS ARE TYPICALLY THE RESULT OF **COLLECTIVE EFFORTS** PERFORMED BY DEVELOPERS **AROUND THE WORLD**

OS THINKS OF ITSELF AS A **MERITOCRACY** WHERE **THE QUALITY** OF THE CONTRIBUTION IS THE KEY

BUT, **DIVERSITY** ASPECTS MIGHT AFFECT THE **ACCEPTANCE** OR **REJECTION** OF THE CONTRIBUTIONS

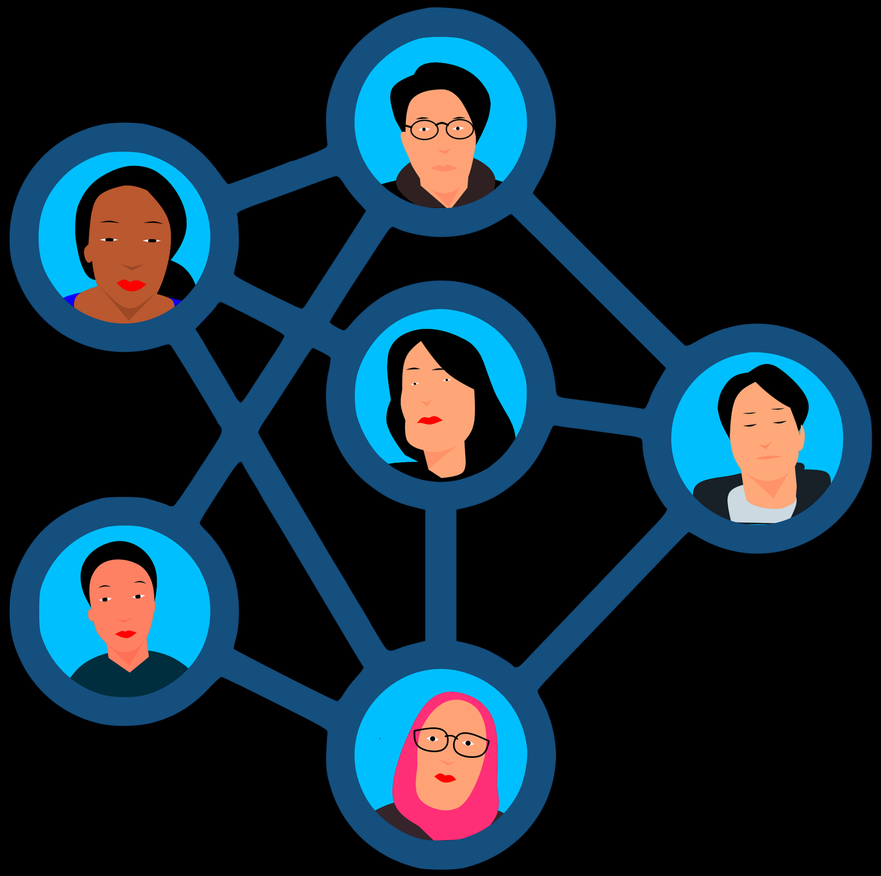
Among developers who have identifiable gender outside a project, women have lower acceptance [1]

[1] Terrell, et al. “Gender differences and bias in open source: Pull request acceptance of women versus men,”

Statistically significant differences in pull request acceptance rate between developers from different countries [2]

[2] Rastogi, et al. “Relationship between geographical location and evaluation of developer contributions in github”

MOREOVER A GITHUB'S SURVEY REVEALED THAT[1]:



Developers are aware of the ethnicity of their team members in GitHub

30% of developers have felt sometimes negative experiences due to diversity in terms of national origin, language, and ideology.

[1] Vasilescu, et al. "Perceptions of diversity on github: A user survey," in 2015 International Workshop on Cooperative and Human Aspects of Software Engineering.

INTEGRATORS' PERCEPTION OF RACE AND ETHNICITY OF SUBMITTERS MIGHT ACTIVATES WHATEVER BIASES THE INTEGRATORS HAVE ABOUT PEOPLE WITH THE SAME RACE AND ETHNICITY OF SUBMITTERS

LET'S LOOK INTO THE PR FROM THE HISPANIC CONTRIBUTOR

GOOD!, A NEW PR FROM A SOUTHERN ASIAN CONTRIBUTOR

A LARGE SCALE STUDY ABOUT THE PERCEPTION OF **RACE AND ETHNICITY** IN **PULL REQUEST EVALUATION**

+10

Non-Trivial

46,191

Engineered Software Projects

GitHub API

Filtering

2,507,591

Merged and rejected Pull requests

Stanford Named Entity Recognizer

NamePrism

Name-Prism

493,170

Deriving race and ethnicity from name

57.2% OF GITHUB DEVELOPERS HAVE A PERCEPTIBLE RACE AND ETHNICITY, BUT 42.8% DEVELOPERS ARE CLASSIFIED AS UNKNOWN

NamePrism

74M names

118 countries

F1 0.79

493,170

Deriving race and ethnicity from name

AIAN *	API **	Black	Hispanic	White	2PRace
0.0	0.01	0.48	0.0	0.51	0.0
0.0	0.00	0.00	0.97	0.03	0.0
…	…	…	…	…	…

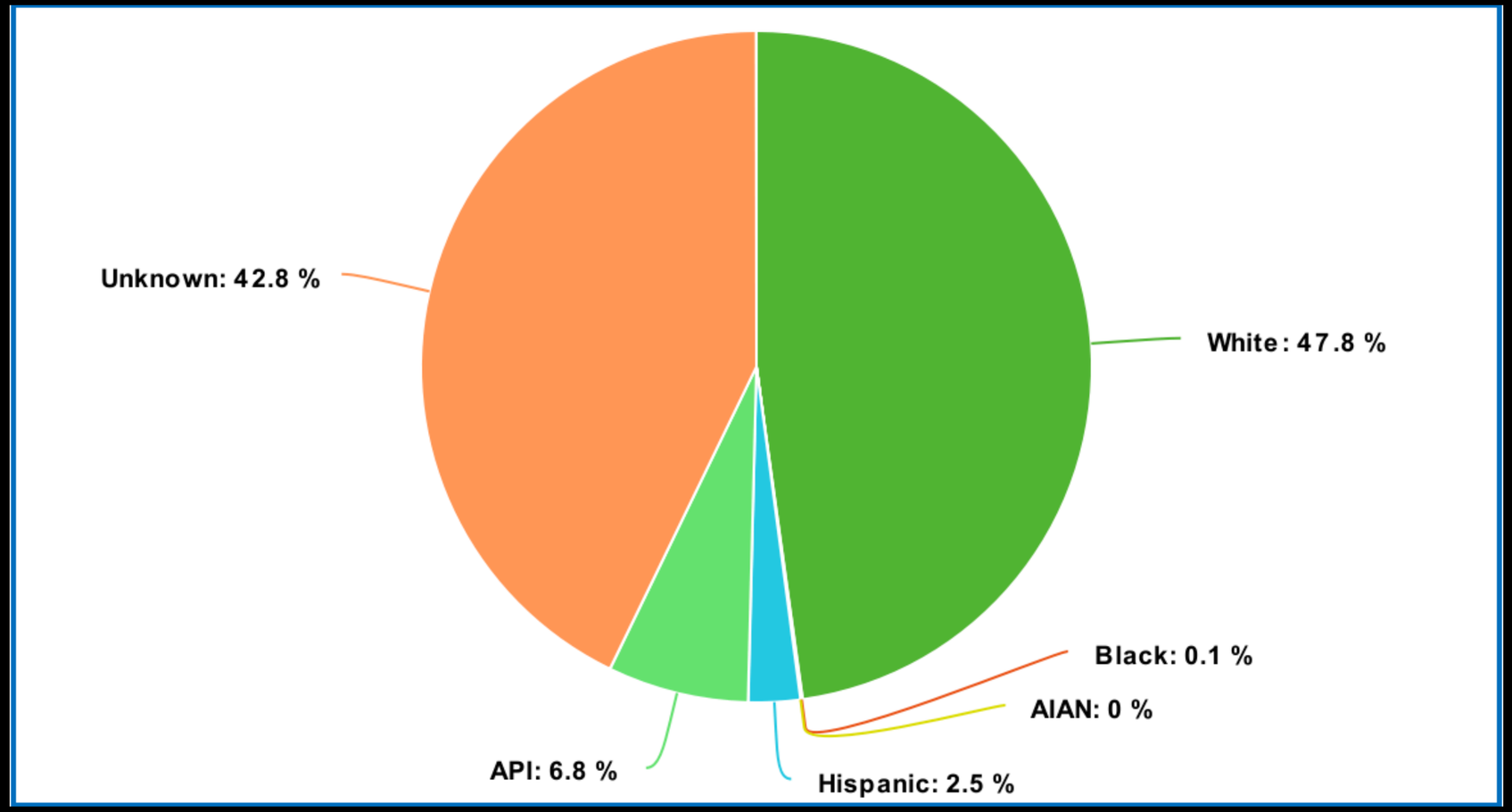
Unknown

Hispanic

* AIAN= American Indian and Alaska Native

** API= Asian Pacific Islander

LESS THAN 10% OF CONTRIBUTIONS SUBMITTED IN OS PROJECTS ARE FROM PERCEPTIBLE NON-WHITE DEVELOPERS



DOES **PERCEPTIBLE RACE AND ETHNICITY** HAVE A **RELATIONSHIP** WITH THEIR **CONTRIBUTIONS**?

Independent

API Black Hispanic White

…

+

Control

Files changed in a PR

Commits in a PR

Developer's Experience

Developer's Tenure

[…]

Mixed-Effects Regression Model

Dependent

THE **RELATIONSHIP** OF **HISPANIC AND API** DEVELOPERS ON PULL REQUEST ACCEPTANCE **IS LOWER** WHEN COMPARED TO PERCEPTIBLE WHITE DEVELOPERS

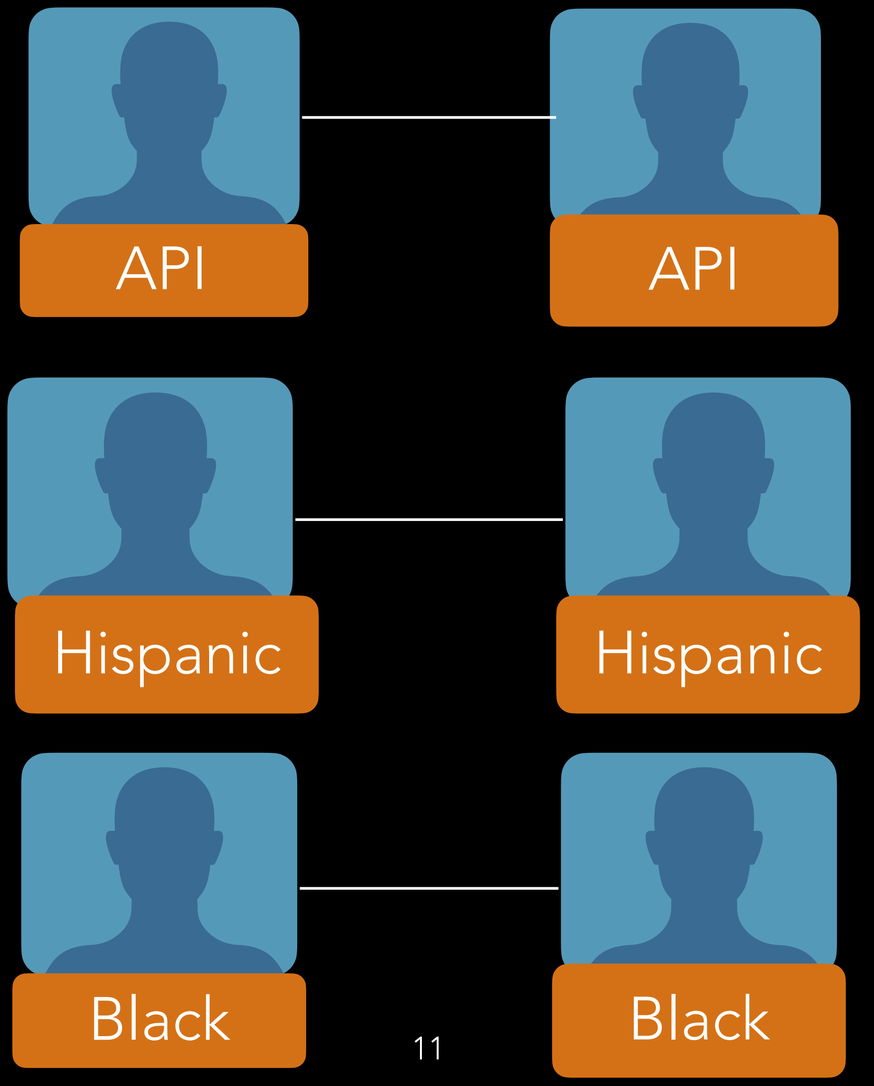
Submitter

Perceptible **Hispanic** developers have **6%** lower odds of getting their pull request accepted

Submitter

Perceptible **API** developers have **10%** lower odds of getting their pull request accepted

NON-WHITE DEVELOPERS ARE MORE LIKELY TO GET THEIR PR ACCEPTED WHEN THE INTEGRATOR IS FROM THEIR SAME RACE AND ETHNICITY (COMPARED TO PERCEPTIBLE WHITE DEVELOPERS)



NON-WHITE DEVELOPERS ARE MORE LIKELY TO GET THEIR PR ACCEPTED WHEN THE INTEGRATOR IS FROM THEIR SAME RACE AND ETHNICITY (COMPARED TO PERCEPTIBLE WHITE DEVELOPERS)

Submitter

Integrator

Perceptible **Hispanic** submitters have **75%** higher odds of getting their pull request accepted when the integrator is estimated as Hispanic

NON-WHITE DEVELOPERS ARE MORE LIKELY TO GET THEIR PR ACCEPTED WHEN THE INTEGRATOR IS FROM THEIR SAME RACE AND ETHNICITY (COMPARED TO PERCEPTIBLE WHITE DEVELOPERS)

Submitter

Integrator

Perceptible **API** developers have **36%** higher odds of getting their pull request accepted when the integrator is estimated as API

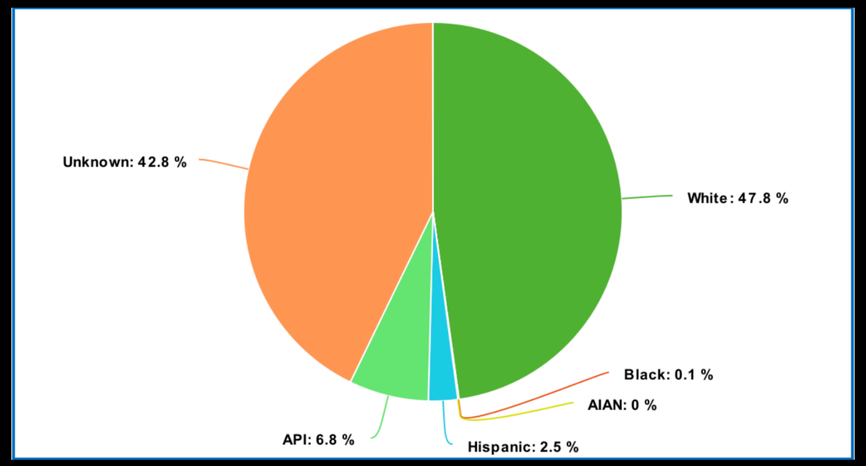
NON-WHITE DEVELOPERS ARE MORE LIKELY TO GET THEIR PR ACCEPTED WHEN THE INTEGRATOR IS FROM THEIR SAME RACE AND ETHNICITY (COMPARED TO PERCEPTIBLE WHITE DEVELOPERS)

Submitter

Integrator

Perceptible **Black** developers have **9** times higher odds of getting their pull request accepted when the integrator is estimated as Black

LESS THAN 10% OF CONTRIBUTIONS SUBMITTED IN OS PROJECTS ARE FROM PERCEPTIBLE NON-WHITE DEVELOPERS



8

THE RELATIONSHIP OF HISPANIC AND API DEVELOPERS ON PULL REQUEST ACCEPTANCE IS LOWER WHEN COMPARED TO PERCEPTIBLE WHITE DEVELOPERS

Submitter

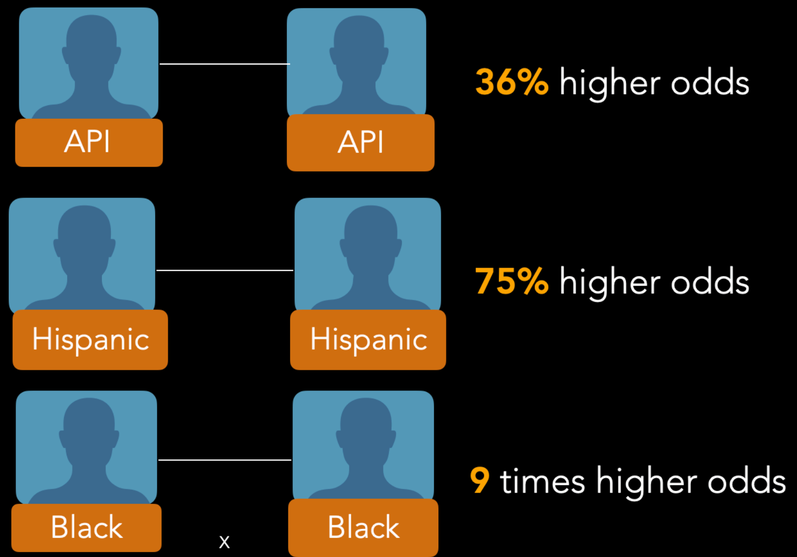
Perceptible **Hispanic** developers have **6%** lower odds of getting their pull request accepted

Submitter

Perceptible **API** developers have **10%** lower odds of getting their pull request accepted

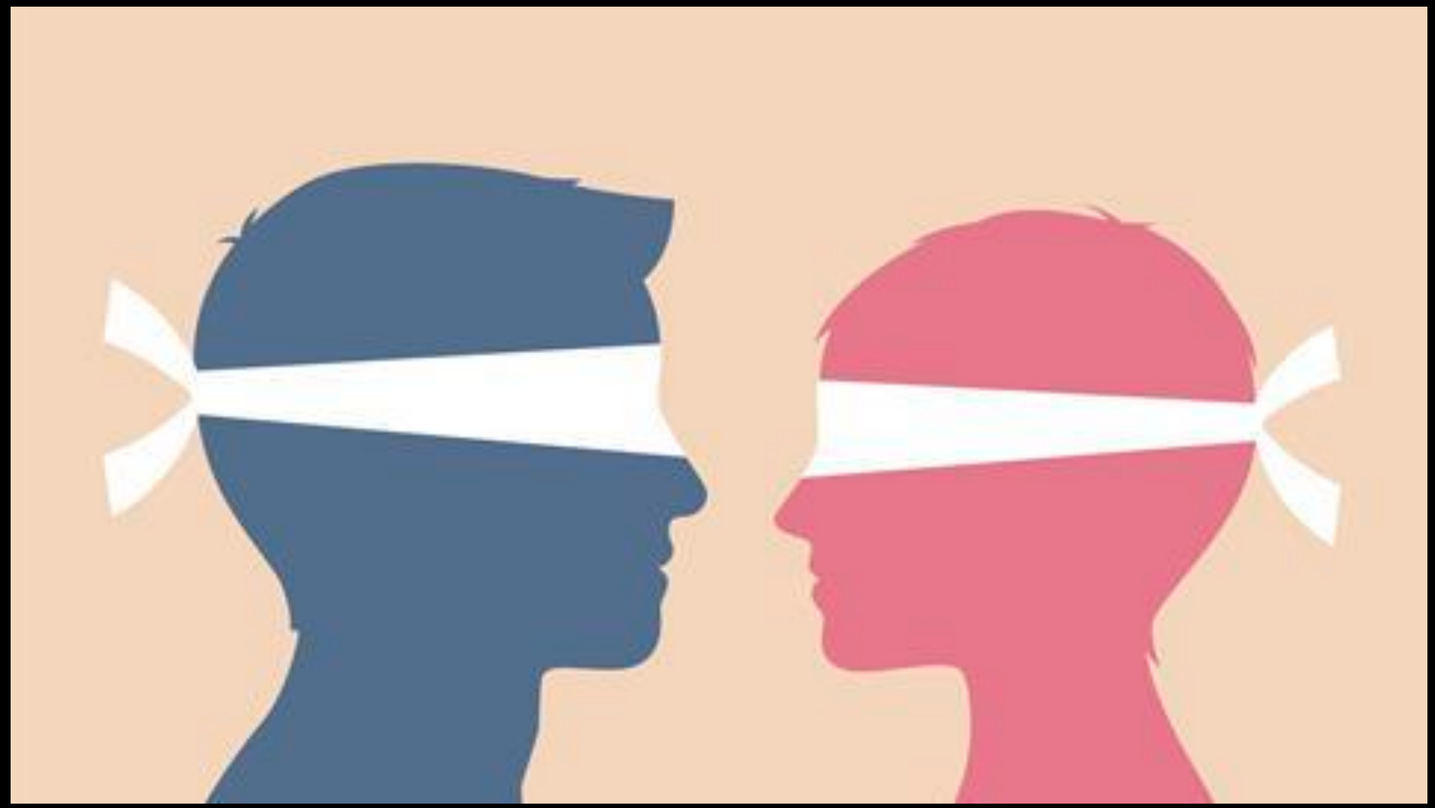
10

NON-WHITE DEVELOPERS ARE MORE LIKELY TO GET THEIR PR ACCEPTED WHEN THE INTEGRATOR IS FROM THEIR SAME RACE AND ETHNICITY (COMPARED TO PERCEPTIBLE WHITE DEVELOPERS)



- REPRESENTATION IS DISPROPORTIONAL TO POPULATION
- UNCONSCIOUS BIAS EXISTS

SO WHAT NOW? AUTHOR BLINDED EVALUATION?



MAYBE?

INSTEAD WE CAN ACTIVELY SUPPORT A DIVERSE GROUP OF CONTRIBUTORS.

INSTEAD WE CAN ACTIVELY SUPPORT A DIVERSE GROUP OF CONTRIBUTORS.

Nadri et al., On the Relationship Between the Developer's Perceptible race and ethnicity and the Evaluation of Contributions in GitHub. **TSE'21**

Nadri et al., *Insights into Non-merged Pull Requests in GitHub: Is There Evidence of Bias Based on Perceptible Race?*. **IEEE Software'21**

Rodriguez-Perez et al., Perceived diversity in software engineering: a systematic literature review. **EMSE'21**

Reza Nadri

gema.rodriguezperez@ubc.ca

mei.nagappan@uwaterloo.ca

@MeiNagappan