Crash Course

ML Concepts

**Advanced ML models** 

▶ Neural networks (75 min)

▶ Large language models (LLMs) (45

Production ML systems (70 min)

**■** Introduction (5 min)

**■** Types of bias (5 min)

■ Identifying bias (10 min)

**■** Evaluating for bias (5 min)

**■** Demographic parity (10 min)

**■** Counterfactual fairness (10

**≡** Equality of opportunity (10 min)

■ Mitigating bias (5 min)

Automated machine learning (30 min)

► Embeddings (45 min)

Real-world ML

▼ Fairness (110 min)

min)

min)

Home

**∓** Filter

Benefits and Drawbacks

Was this helpful? 🖒 🗇

**Accepted Candidates** 

Send feedback

Sign in

Info

Chat

API

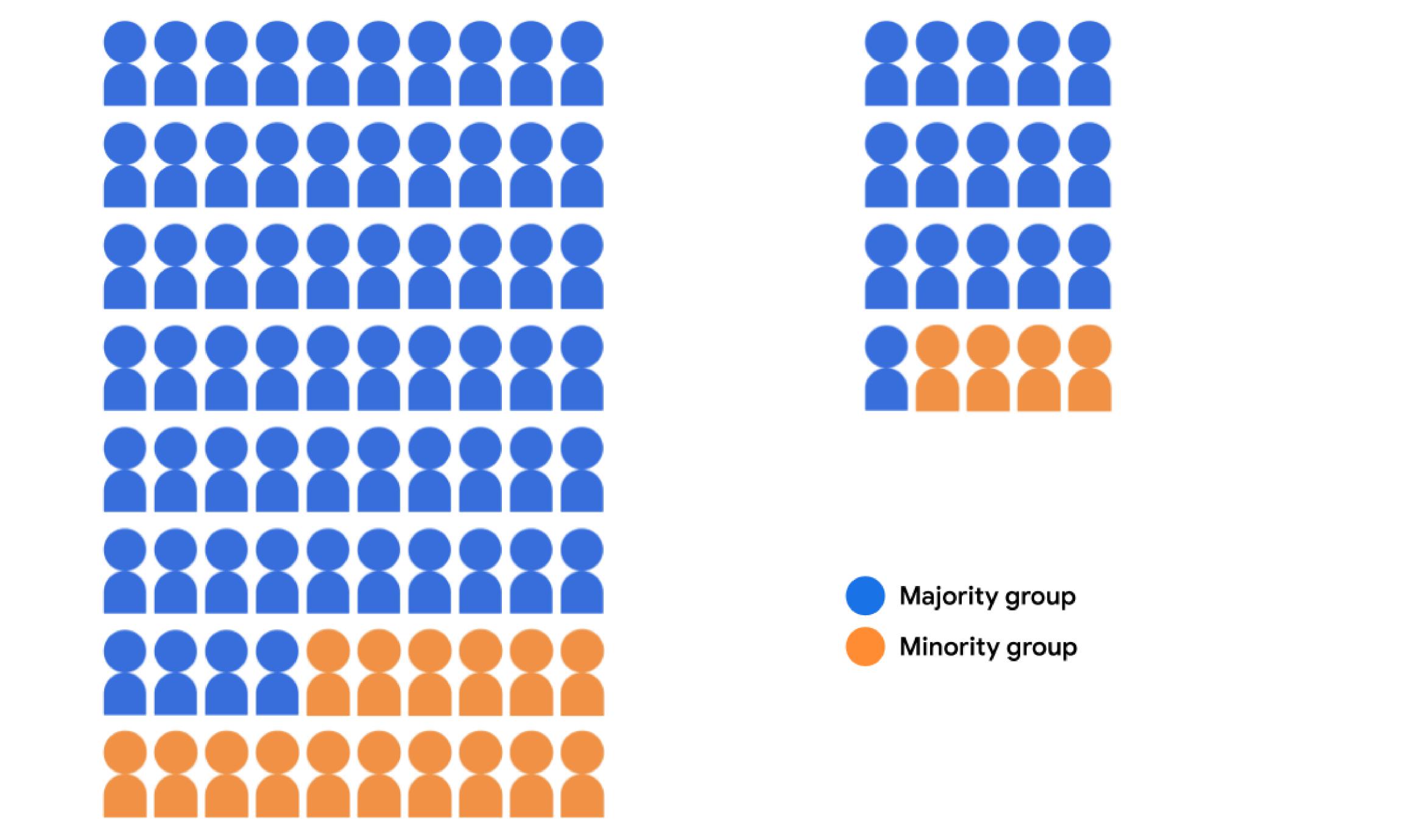
Home > Products > Machine Learning > ML Concepts > Crash Course Fairness: Demographic parity 🗖 -

On this page Exercise: Check your intuition

Rejected Candidates

One method we can use to evaluate our model's predictions for fairness is to compare the admissions rate for the majority group and the minority group. If the two admissions rates are equal, then the model's predictions exhibit demographic parity: a student's chance of being admitted to the university doesn't vary by demographic group. Suppose the admissions model accepts 16 candidates from the majority group and 4 candidates from the minority group. The model's

decisions satisfy demographic parity, as the acceptance rate for both majority and minority candidates is 20%.



groups have an acceptance rate of 20%. The following table quantifies the numbers supporting the rejected and accepted candidates in Figure 2.

Figure 2. Candidate pool split into two groups: Rejected Candidates and Accepted Candidates. Both the majority (blue) and minority (orange)

	Majority group	Minority group	
Accepted	16	4	
Rejected	64	16	
Acceptance Rate	20%	20%	

## your guidance on the benefits and drawbacks of this approach. Can you identify one pro and one con of evaluating the model's predictions using demographic parity?

Exercise: Check your intuition

Can you identify any problems with the feature data?

Continue reading the next section, Benefits and Drawbacks, for our summary of the key pros and cons of using demographic parity.

The university admissions committee is considering using demographic parity as the fairness evaluation metric for their model, and needs

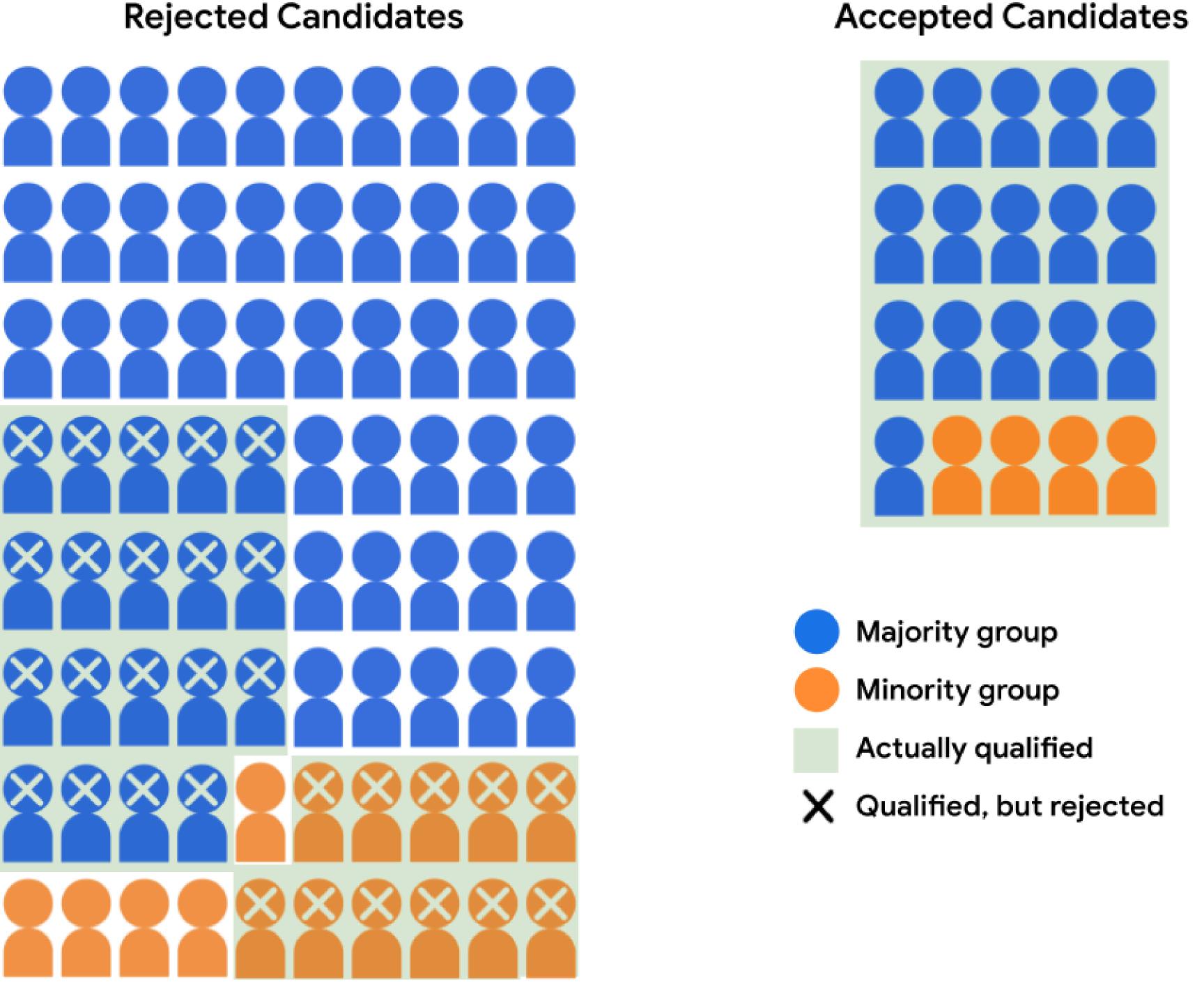
## Benefits and Drawbacks

The key benefit of demographic parity for our admissions example is that it ensures that both majority and minority groups are represented in the admitted class of students in the same proportions as they are in the candidate pool. That is, if the applicant pool is comprised of 80% majority-group candidates and 20% minority-group candidates, demographic parity guarantees that the group of admittees will also be 80% majority-group students and 20% minority-group students.

number of students classified as "qualified" vs. "unqualified") into account when evaluating how the 20 admissions slots should be allocated. Let's take another look at the composition of our candidate pool from above. However, this time we'll classify candidates not only by

However, demographic parity has one significant drawback: it does not take the distribution of predictions for each demographic group (the

demographic group, but also by whether the model has scored each candidate as "qualified" or "unqualified":



minority groups are shaded in green, and the qualified students who were rejected are marked with an X. Of the 35 qualified majority-group students, 16 were accepted. Of the 15 qualified minority-group students, 4 were accepted.

Figure 3. The same breakdown of Rejected and Accepted candidates as in Figure 2, but this time all qualified students in both the majority and

The following table quantifies the numbers supporting the rejected and accepted candidates in Figure 3.

	Majority group	Majority group		Minority group	
	Accepted	Rejected	Accepted	Rejected	
Qualified	16	19	4	11	
Unqualified	0	45	0	5	

Majority acceptance rate =  $\frac{\text{qualified majority students accepted}}{\text{total qualified majority students}} = \frac{16}{35} = 46\%$ 

students is 46%, whereas the acceptance rate for qualified minority students is only 27%.

details, see the Google Developers Site Policies. Java is a registered trademark of Oracle and/or its affiliates.

Let's tabulate the acceptance rates for qualified students for both groups:

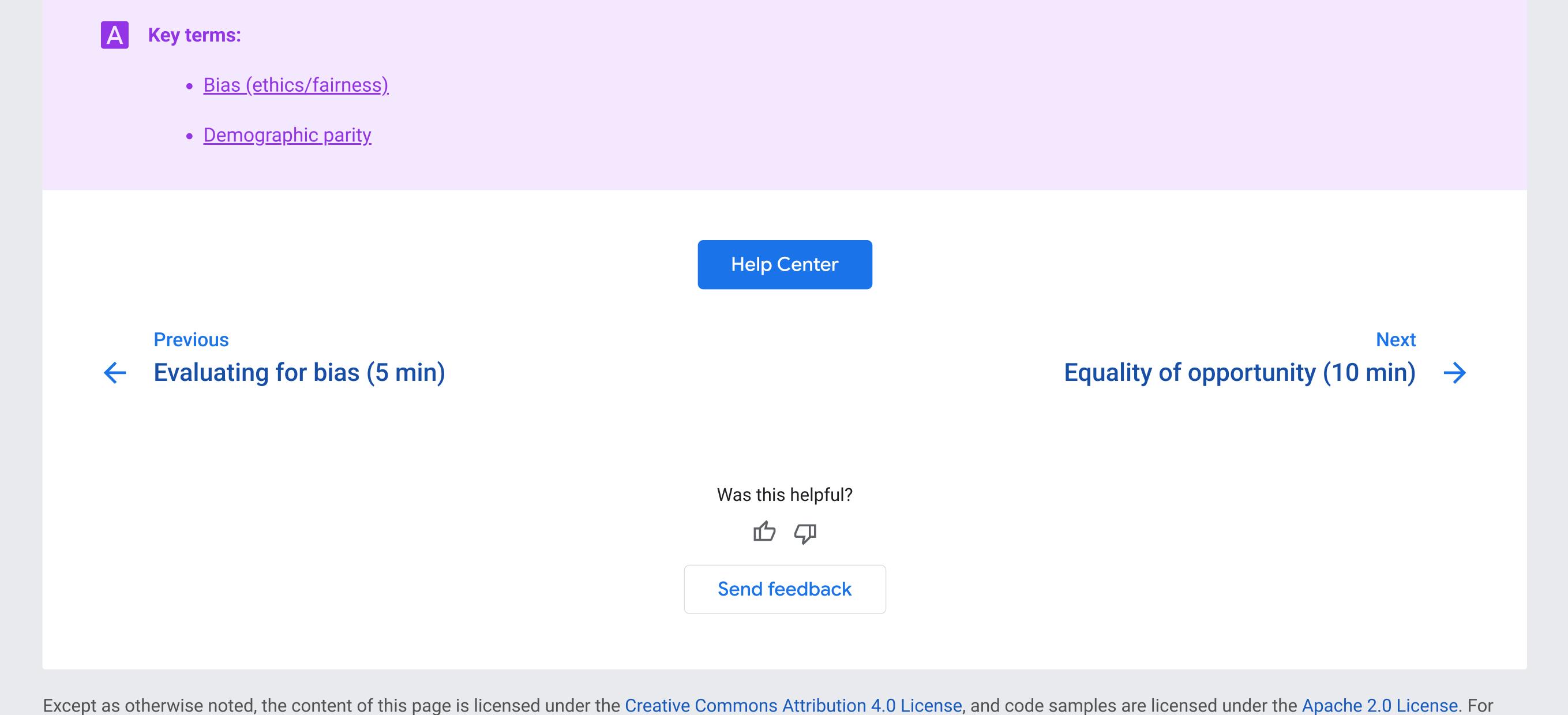
 $\text{Minority acceptance rate} = \frac{\text{qualified minority students accepted}}{\text{total qualified minority students}} = \frac{4}{15} = 27\%$ 

In cases like this, where the distribution of a preferred label (such as "qualified") varies greatly for both groups, demographic parity may not be the optimal metric to choose for evaluating fairness. In the next section, we'll look at an alternative fairness metric, equality of opportunity, which takes these differences into consideration.

Even though both groups have an overall acceptance rate of 20%, satisfying demographic parity, the acceptance rate for qualified majority

the model is designed to achieve and the social context in which the model's predictions will be used.

Note: There may be additional benefits/drawbacks of demographic parity not discussed here that are also worth considering, depending on both the goals



Last updated 2024-10-09 UTC.

Terms | Privacy

