

Solutions

Each answer shows the steps, then the **result**.

2.1

- it learns from **data** or follows rules that already contain human bias, or the data leaves some groups out

2.2

- **group C**
- it is only **5 percent** of the training photos, so the program had few examples to learn from

2.3 any two:

- collect **representative** data from every group
- **test** accuracy for each group separately
- include **diverse** people in the development team
- **review** the rules and factors the program uses

2.4

- a program makes the same biased decision **many times**, for **many** people, very quickly

3.1 B

- groups rare in the training data are served worse

3.2 B

- comparing results for each group reveals where the program serves some worse

3.3 B

- past refusals in the data are repeated by the program

3.4

(a)

- if past classes had few students from some group (e.g. girls, or students from some schools), the data contains **few examples** of them succeeding
- so the program **recommends them less often**, repeating past patterns rather than measuring ability

(b) e.g.:

- able students are **not invited** to the class and lose the chance to learn advanced maths

(c) any two:

- **check** the recommendations for each group of students and compare them
- use **current** measures of ability, e.g. a test, instead of similarity to past students
- let **teachers review** and override recommendations