Data and Fairness in AI

Standard difficulty · 20 questions · 40 marks · Time: 37 minutes

Name: _____ Date: _____ Score: ____ / 40

Show your working for Sections B and C. Answers are on the last page — try every question first!

Section A — Choose the correct answer — 8 × 1 = 8 marks

1. Bias in a dataset means the data is:
   (a) correctly labelled **(b)** stored in a table
   (c) too large to store **(d)** unbalanced or unfair to some groups

2. An image classifier is trained on 10 pictures of sparrows and 90 pictures of crows. On new pictures, it is likely to be more accurate for:
   (a) sparrows **(b)** both equally
   (c) crows **(d)** neither, it will always be wrong

3. A training set has 36 pictures of cars and 4 pictures of bicycles. What percentage of the pictures are of cars?
   (a) 95% **(b)** 10% **(c)** 90% **(d)** 36%

4. A training set has 12 pictures of buses and 88 pictures of trucks. How many more pictures of buses are needed so that both kinds have the same number?
   (a) 38 **(b)** 100 **(c)** 88 **(d)** 76

5. Past students who enjoyed each sport:
   table tennis: 35 boys, 41 girls
   hockey: 54 boys, 59 girls
   cricket: 51 boys, 34 girls
   kabaddi: 58 boys, 28 girls
   A model that only follows past patterns recommends the sport most enjoyed by past students of the same group. What would it recommend to a new boy?
   (a) hockey **(b)** table tennis **(c)** cricket **(d)** kabaddi

6. A voice assistant is trained on 1000 voice samples from four regions: south 250, north 550, east 150, west 50. Speakers from which region are most under-represented?
   (a) west **(b)** south **(c)** east **(d)** north

7. Which of these helps reduce bias in an AI system?
   (a) remove groups that have few examples
   (b) add more examples of groups that are missing
   (c) train only on the most common case
   (d) keep how the system decides a secret

8. What does “training data” mean?
   (a) being able to understand how a system makes decisions
   (b) new examples used only to check a model
   (c) data or results that are unbalanced or unfair to some group
   (d) the examples a model learns from

Section B — Short answer — 8 × 2 = 16 marks

9. Why does an image classifier trained on many more bats than rackets often guess “bat”?

10. A training set has 225 pictures of apples and 25 of oranges. How many pictures of oranges must be added to balance it?

11. A training set has 18 pictures of buses and 82 of trucks. What percentage are of trucks?

12. A dataset has 90 samples from the north region, 60 samples from the east region, 20 samples from the south region, 30 samples from the west region. More samples are added to every smaller group until each matches the largest. How many samples will the dataset have then?

13. Past students who enjoyed each sport:
table tennis: 7 boys, 34 girls
badminton: 44 boys, 26 girls
kabaddi: 20 boys, 9 girls
basketball: 60 boys, 15 girls
A model recommends the sport most enjoyed by past students of the same group. What would it recommend to (a) a new girl (b) a new boy?

14. A face-detection tool is tested on two groups of people. Group A: 39 of 50 detected correctly. Group B: 15 of 25. What is its accuracy for (a) group A (b) group B?

15. Explain each term: (a) bias (b) test data

16. A video platform’s recommendations over a week included 500 videos, of which 60 were in smaller regional languages. What percentage is that?

Section C — Word problems — 4 × 4 = 16 marks

17. Explain with an example how unbalanced training data leads to unfair AI results.

18. A dataset for a speech model has 250 samples from the east region, 150 samples from the south region, 500 samples from the north region, 100 samples from the west region. (a) How many samples are there in all? (b) What percentage come from the north region? (c) How many samples must be added to each of the other three regions, in the order listed, to match the north region? Write them in order.

19. A recognition tool was trained mostly on group A. Tested on new people: group A 20 of 25 correct, group B 31 of 50 correct. (a) What is the accuracy for group A? (b) for group B? (c) How many percentage points apart are they? (d) Name one fair way to close the gap.

20. A classifier is trained on 8 pictures of apples and 32 pictures of oranges. (a) What percentage of the pictures are of apples? (b) How many pictures of apples must be added to balance the set? (c) Before balancing, for which kind is the model likely to be more accurate?

Want another sheet on this chapter? In the Ambiful Worksheets app every sheet is new, with its answer key, and your child never gets the same question twice.
Scan the code to get the free app, or open **worksheets.ambiful.com/get**

Answer key

Data and Fairness in AI · AW-8CT-DFA-2026-V1 · For parents — fold back or tear off before handing the sheet to your child.

Q	Answer	Working
1	**(d) unbalanced or unfair to some groups**	—
2	**(c) crows**	a model learns best what it has seen most often
3	**(c) 90%**	36 ÷ 40 × 100
4	**(d) 76**	88 − 12
5	**(d) kabaddi**	most boys: kabaddi; the pattern, not the new student's interest, decides
6	**(a) west**	smallest share: west
7	**(b) add more examples of groups that are missing**	—
8	**(d) the examples a model learns from**	—
9	**It learns from the examples it sees; with far more bats, it recognises bats better and leans towards them.**	—
10	**200**	225 − 25
11	**82%**	82 ÷ 100 × 100
12	**360**	90 × 4
13	**(a) table tennis; (b) basketball**	largest number in each column
14	**(a) 78%; (b) 60%**	39 ÷ 50 × 100; 15 ÷ 25 × 100
15	**(a) data or results that are unbalanced or unfair to some group (b) new examples used only to check a model**	—
16	**12%**	60 ÷ 500 × 100
17	**e.g. a face-recognition model trained mostly on one group recognises them well but makes more mistakes for others, because it learned mainly from the first group.**	—
18	**(a) 1,000; (b) 50%; (c) 250, 350, 400**	500 ÷ 1000 × 100; 500 minus each
19	**(a) 80%; (b) 62%; (c) 18; (d) e.g. add more examples from group B to the training data and test again**	20 ÷ 25 × 100; 31 ÷ 50 × 100
20	**(a) 20%; (b) 24; (c) oranges**	8 ÷ 40 × 100; 32 − 8

Ambiful Worksheets · worksheets.ambiful.com · Not affiliated with CBSE / NCERT