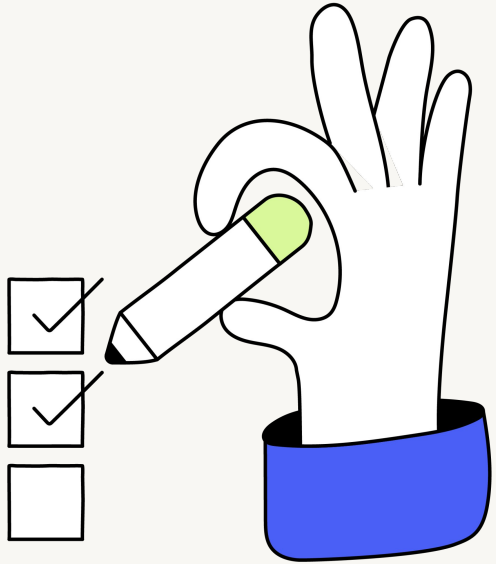
Module 8

Workshop 2: Fairness in facial emotion recognition

Icebreaker

Share one algorithmic-bias headline that surprised you lately.

Drop your answers in the chat.



Module 8

Workshop 2: Fairness in facial emotion recognition

Agenda

- **Review:** Recap key concepts from async unit 3.
- **Demo:** A guided walkthrough of how to:
  - Derive a skin-tone proxy (ITA)
  - Train a baseline CNN
  - Compute fairness and visualise error rates
  - Apply light-weight mitigations for fairness and accuracy
- **Practice:** Hands-on evaluating & mitigating skin-tone bias in a facial-emotion classifier.
- **Closing:** Key takeaways and next steps.

Learning objectives

- **Derive and bin a skin-tone proxy** from face crops into six demographic groups for downstream fairness analysis.
- **Train and evaluate** a CNN baseline using class-weighted loss and regularisation to establish raw performance benchmarks.
- **Compute and visualise fairness-aware metrics** (e.g. per-bin FNR, max-gap) on disaggregated slices—skipping under-represented bins—and interpret disparate error patterns.
- **Apply lightweight mitigations** (post-hoc temperature scaling; tone-aware instance re-weighting) and critically assess their impacts on the fairness–accuracy trade-off and business governance KPIs.

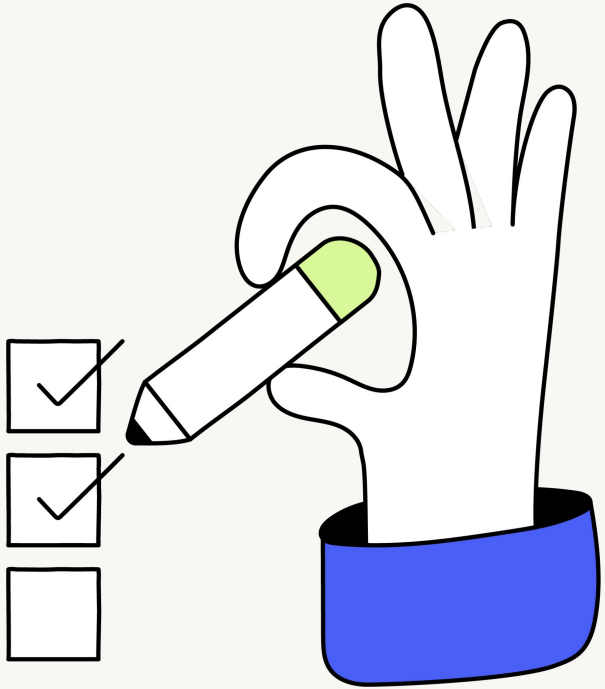
Async review: Unit 3

Go to My Multiverse, Module 8
Workshop 2

Navigate to the **“Async review”** page

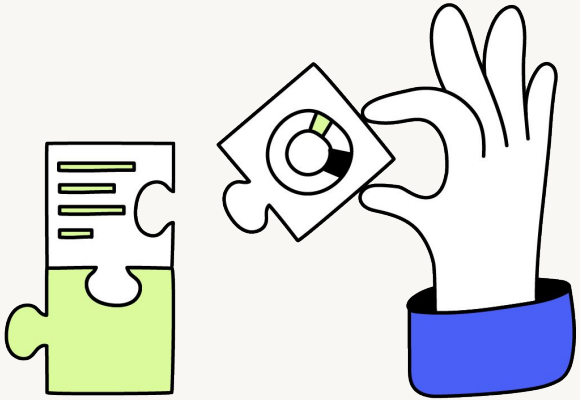
Review the summary of Unit 3, then test your knowledge with the quiz:

Bias and Variance Tradeoff checkpoint



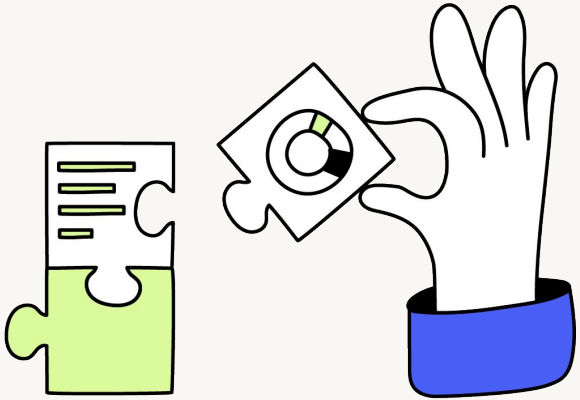
Fundamentals of bias & variance

- **Define statistical bias** as the systematic error from under-fitting, where overly simple models miss key patterns.
- **Define statistical variance as error** from over-reacting to training-set noise, causing models to generalise poorly.
- **Explain the bias–variance trade-off curve**, showing how total error splits into bias, variance, and irreducible noise as complexity changes.
- **Evaluate bias and variance** impacts on model reliability, interpretability, and business alignment.



Mitigation & auditing strategies

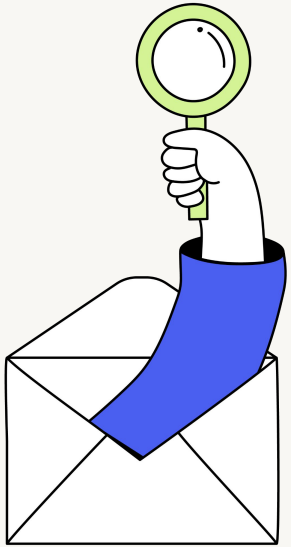
- **Mitigate bias** by increasing model capacity or dataset size, and control variance through regularisation or ensembling.
- **Assess mitigation effectiveness** by comparing fairness improvements against any loss in predictive accuracy.
- **Implement detection techniques** such as cross-validation, regularisation, data augmentation—to identify and correct under- and over-fitting.
- **Conduct bias–variance–fairness audits** on real data to recommend governance-aligned, evidence-based adjustments.



Skills application demo: Fairness in facial emotion recognition demo

Skills application demo

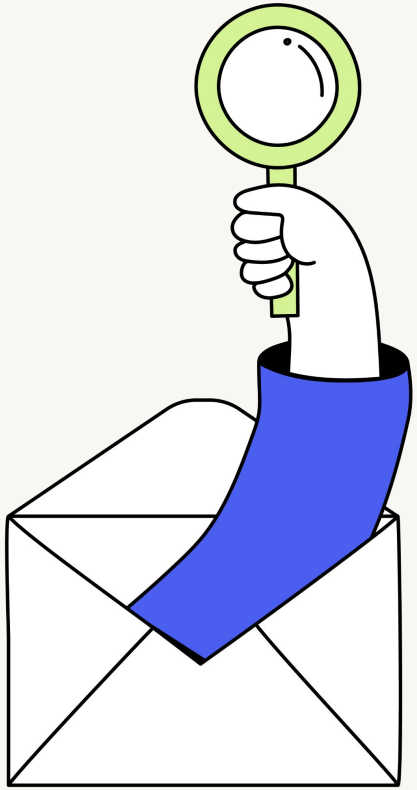
- **Objective:** Apply an end-to-end ensemble classification pipeline to predict London house price categories, reinforcing practical skills in data preparation, model training, tuning, and evaluation for robust, generalisable results.
- **What you'll do:**
  - Load, explore, clean, and impute the London House Price dataset.
  - Train Random Forest, Gradient Boosting, and XGBoost models with GridSearchCV and RandomizedSearchCV.
  - Evaluate via confusion matrices, accuracy, precision, and recall; interpret key trade-offs.
  - Reinforce strategies for model robustness and generalisation.



Skills application: Fairness in facial emotion recognition

Skills application options

- Work in breakout rooms in small groups.
- Use the Jupyter Notebook and provided dataset.
- The challenge is designed to be solved individually, but collaboration is encouraged.
- Find all instructions and resources on the Skills application page.

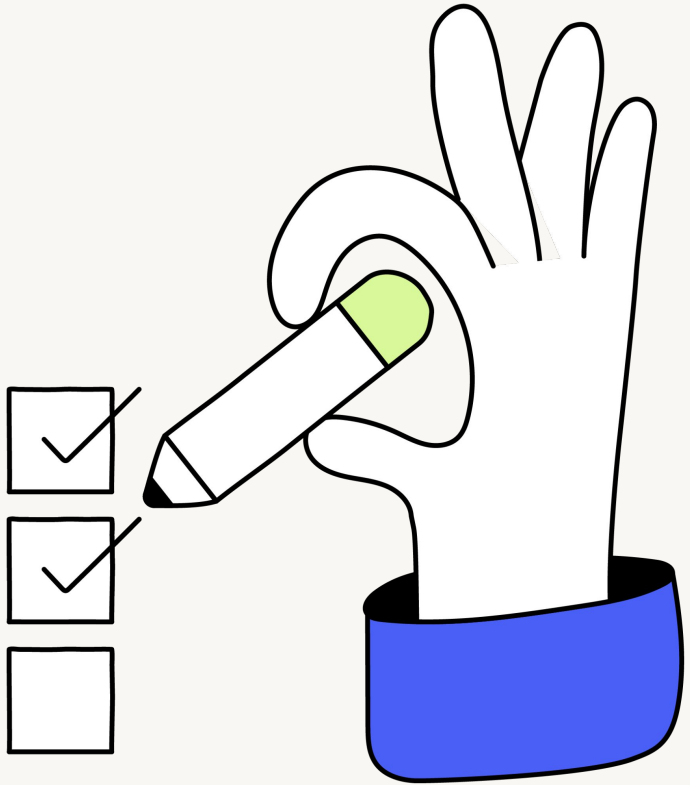


Access the skills application

Go to My Multiverse, Module 8 Workshop 2

Go to the Skills application page and follow the instructions provided

Be prepared to discuss your learnings when we come back together



Skills application debrief

Skills application debrief

Which fairness-audit and mitigation techniques did you apply?

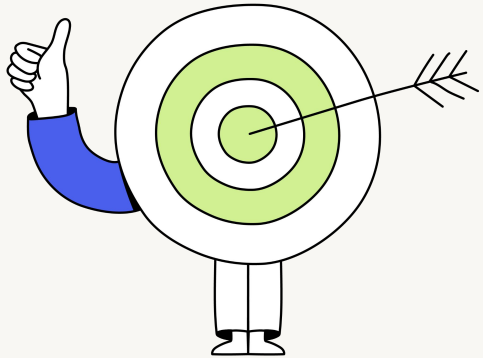
How did they affect per-bin FNR gaps and overall model accuracy?

Key takeaways

1. **Rigorous slicing is your foundation:**Derive and bin your ITA skin-tone proxy, clean and encode demographics, and ensure you handle under-represented cohorts before any modelling.

2. **Calibration & re-weighting amplify fairness:** Temperature scaling and tone-aware instance re-weighting each tackle bias differently; tune both to tighten FNR gaps across groups.

3. **Metrics fuel insight:** Per-slice FNR, max-gap, and confusion matrices aren't vanity stats—they spotlight exactly where your model misfires by demographic.



Applying key takeaways to your organisation

Applying key takeaways

1. Where does messy or incomplete data crop up in your projects? How are you currently exploring, imputing, and encoding it to build more reliable models?

2. Which classification challenges in your team could benefit from ensemble methods? How might hyperparameter tuning with GridSearchCV or RandomizedSearchCV unlock performance gains?

3. How are you interpreting model outcomes today? In what ways could confusion matrices, precision, recall, and accuracy sharpen your insights and drive better business decisions?

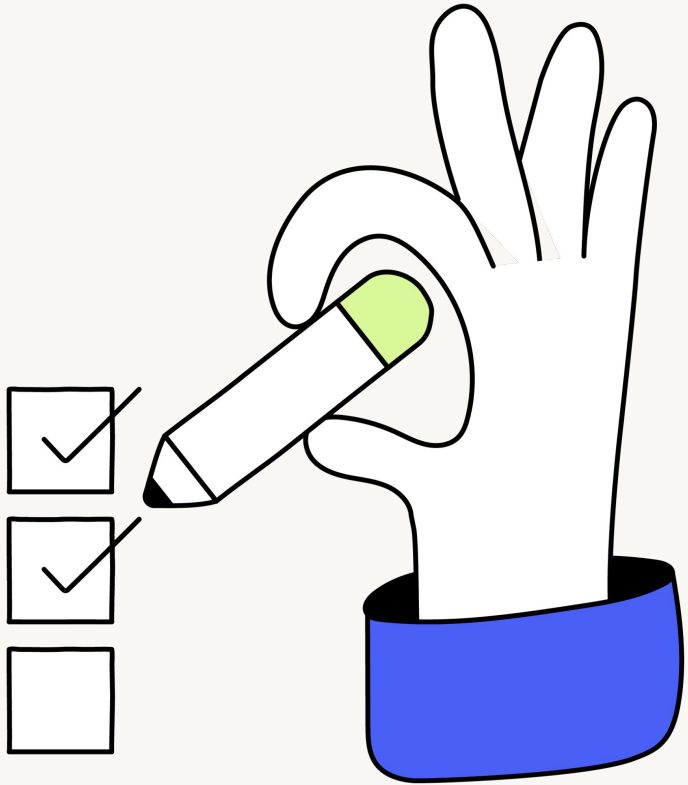
Share your plan to apply new skills

Go to My Multiverse, Module 8
Workshop 1

On the Apply new skills to your role page, find the discussion activity

Take 5 minutes to share how you will apply this skill in your role.

Take time this week to read what others have written and get new ideas!



Closing out

- **Next workshop:** Wrap Up
- **Have questions?**
  - Chat with Atlas
  - Reach out to your Cohort Coach
  - Sign up for Tutoring
- **Log your OTJ hours**
- **Take the Workshop Survey** on My Multiverse

multiverse