

Case Studies

Scenario 4: Fairness in AI-Assisted Screening for Diabetic Retinopathy

Background: An AI tool is being developed to assist ophthalmologists in screening for diabetic retinopathy from images of the retina acquired using a retinal scanner. Diabetic retinopathy is a condition that can lead to blindness if not detected and treated early. The condition is more prevalent in certain populations, such as Hispanic and African American communities. Current screening relies on experts reviewing the retinal images, but this process misses about 15% of positive cases, leading to preventable vision loss.

Development and Training: The AI tool will be trained using retinal images as input data and expert diagnoses as output labels. Training and evaluation will be based on data acquired from a network of 10 hospitals across the USA. The goal is to make the tool available globally to improve screening for diabetic retinopathy.

Questions to consider:

- What dangers/risks of the use of AI for this problem can you identify at this stage?
- How would you go about addressing these?
- What fairness metric(s) do you think might be appropriate when assessing the AI tool for potential bias?

REMINDER - DEFINITIONS OF FAIRNESS

- *False Negative Rate (FNR): the rate at which positive cases are missed by the classifier*
- *Equalised odds - equal true positive rate (TPR) & false positive rate (FPR) for each protected group*
- *Equal opportunity - only equalise either FPR or FNR, not both*