



Legal Obligations for AI Explainability: Open Questions on Technical Methods



2nd July 2026

The 4th World Conference on eXplainable Artificial Intelligence

Madeleine Waller, Paul Waller, Karen Yeung



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



The
Alan Turing
Institute



Thorney Isle Research



UK Research
and Innovation



UNIVERSITY of
BRADFORD



UNIVERSITY OF
BIRMINGHAM

25 October 2021

Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms

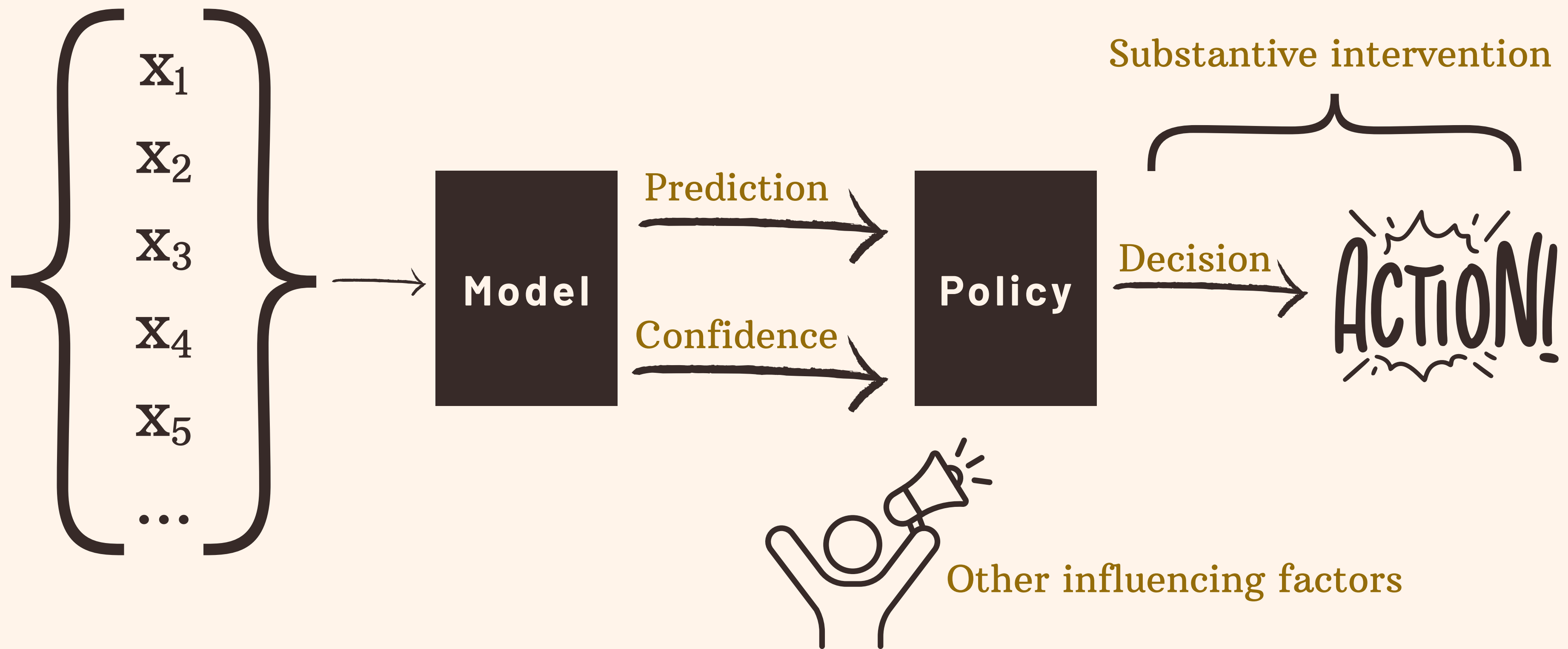
Warnings AI tools used by government on UK public are 'racist and biased'

Transparency campaigners welcome government move to publish details of system algorithms

Facial recognition use by South Wales Police ruled unlawful

🕒 11 August 2020





AI Explainability

The degree to which a system or a set of practices and tools support a person's ability to:

Explain and communicate the rationale underlying the behaviour of the system

Demonstrate that the processes behind its design, development, and deployment have been undertaken in ways that ensure its sustainability, safety, fairness, and accountability across its particular contexts of use and application

Legal obligations for ADMS

GDPR Article 15

‘provide certain kinds of information to the data subject, including, but not limited to, the purpose of the personal data collected and processed, **meaningful information about the logic** in processing their personal data, and the significance and the envisaged consequences of such processing for the data subject’

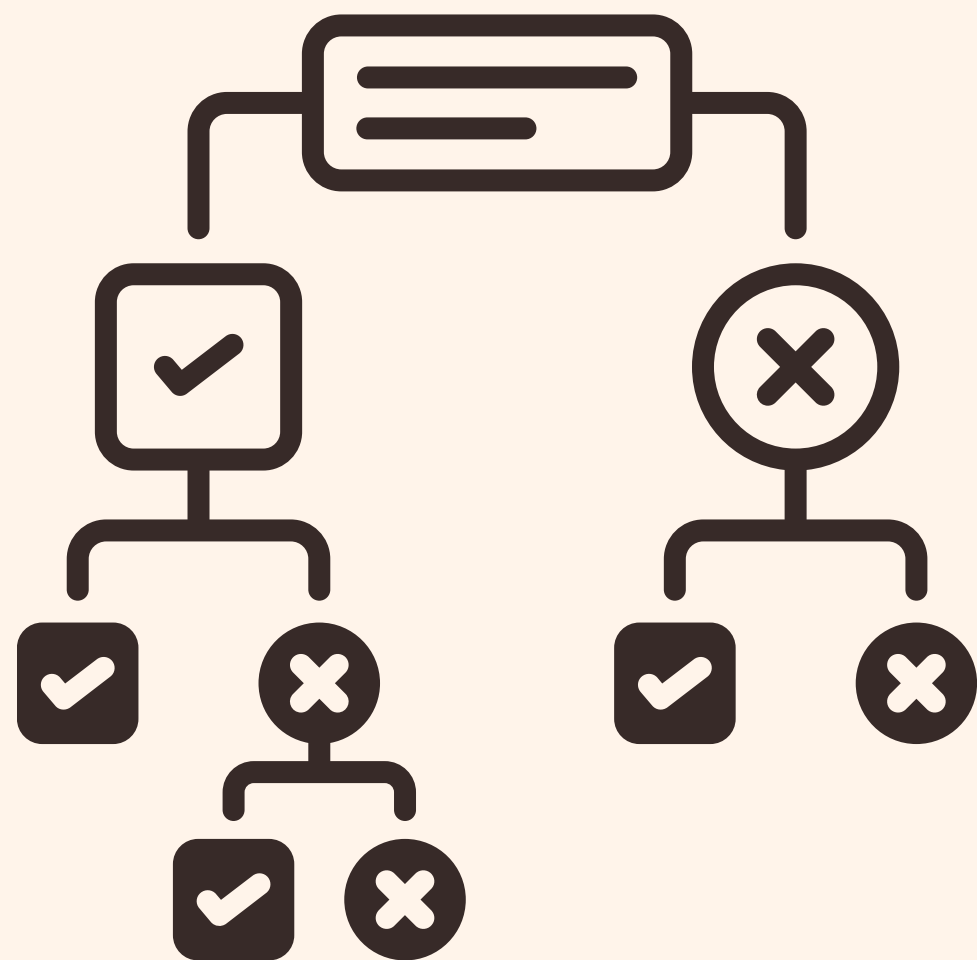
EU AI ACT Article 50

‘the natural persons concerned are informed that they are interacting with an AI system’

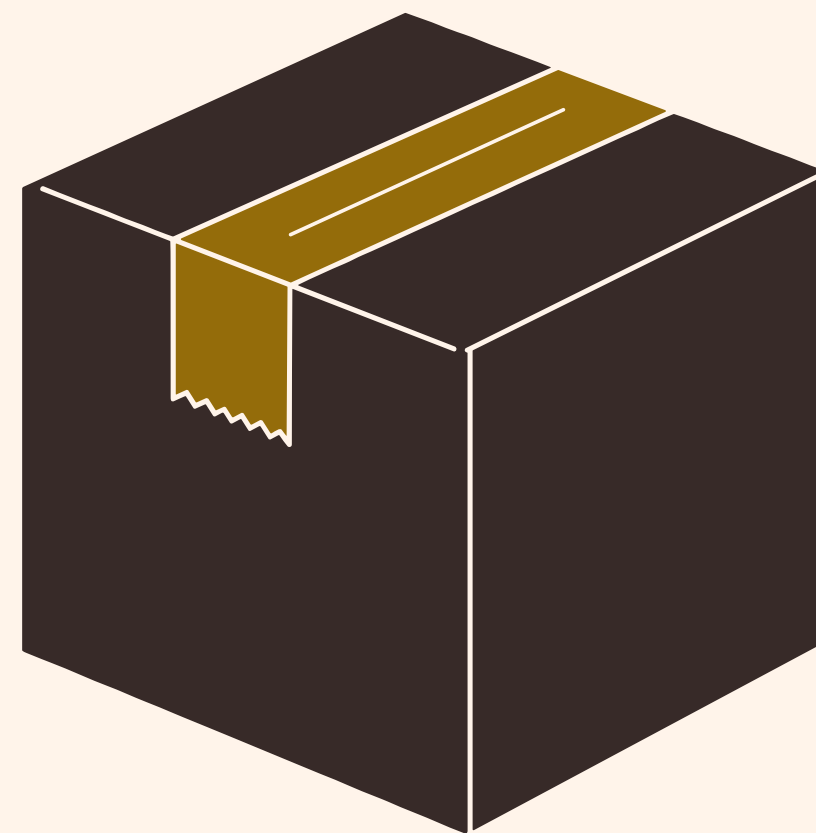
EU AI ACT Article 86

‘clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken’

XAI Methods



Interpretable



Black-box

Functional Explanations

- X_1
- X_2
- X_3
- X_4
- X_5
- ...



Prediction

Confidence



Substantive intervention

Decision

ACTION!



Other influencing factors



Why did you go over the speed limit?

- I pressed the accelerator harder
- I was in a hurry
- I was rushing to the hospital after learning that my daughter has been taken to the Emergency Room following a school accident

Is this a justification?

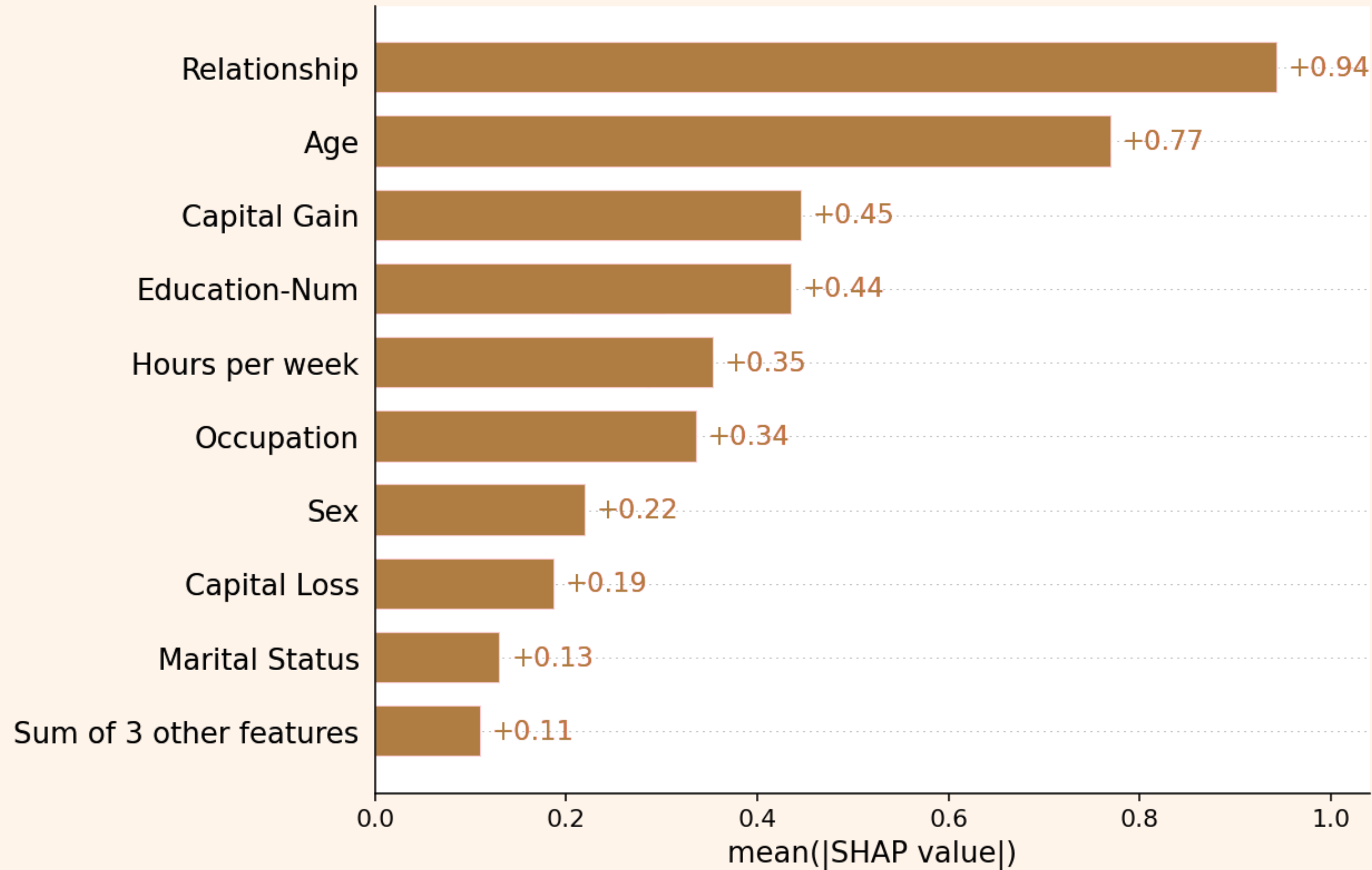
Global



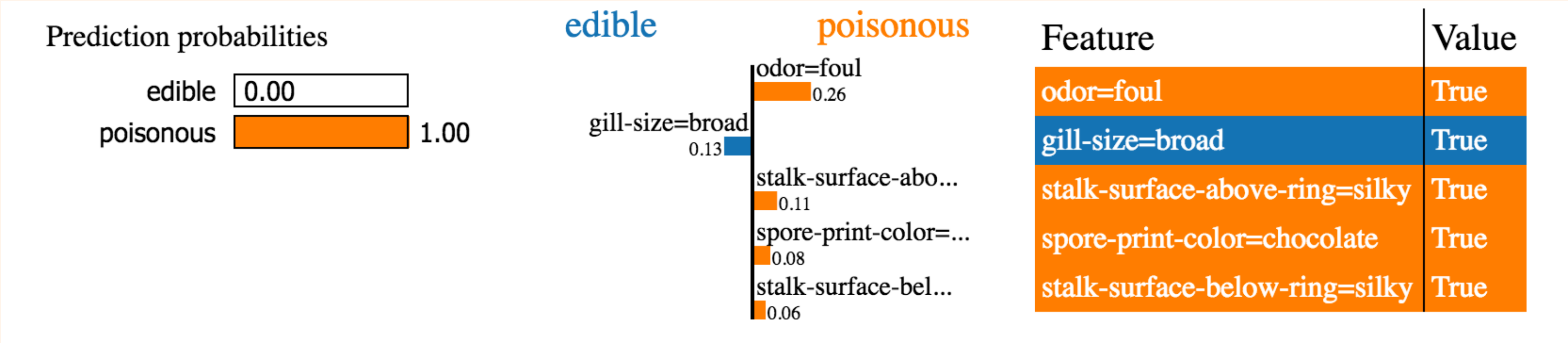
Local



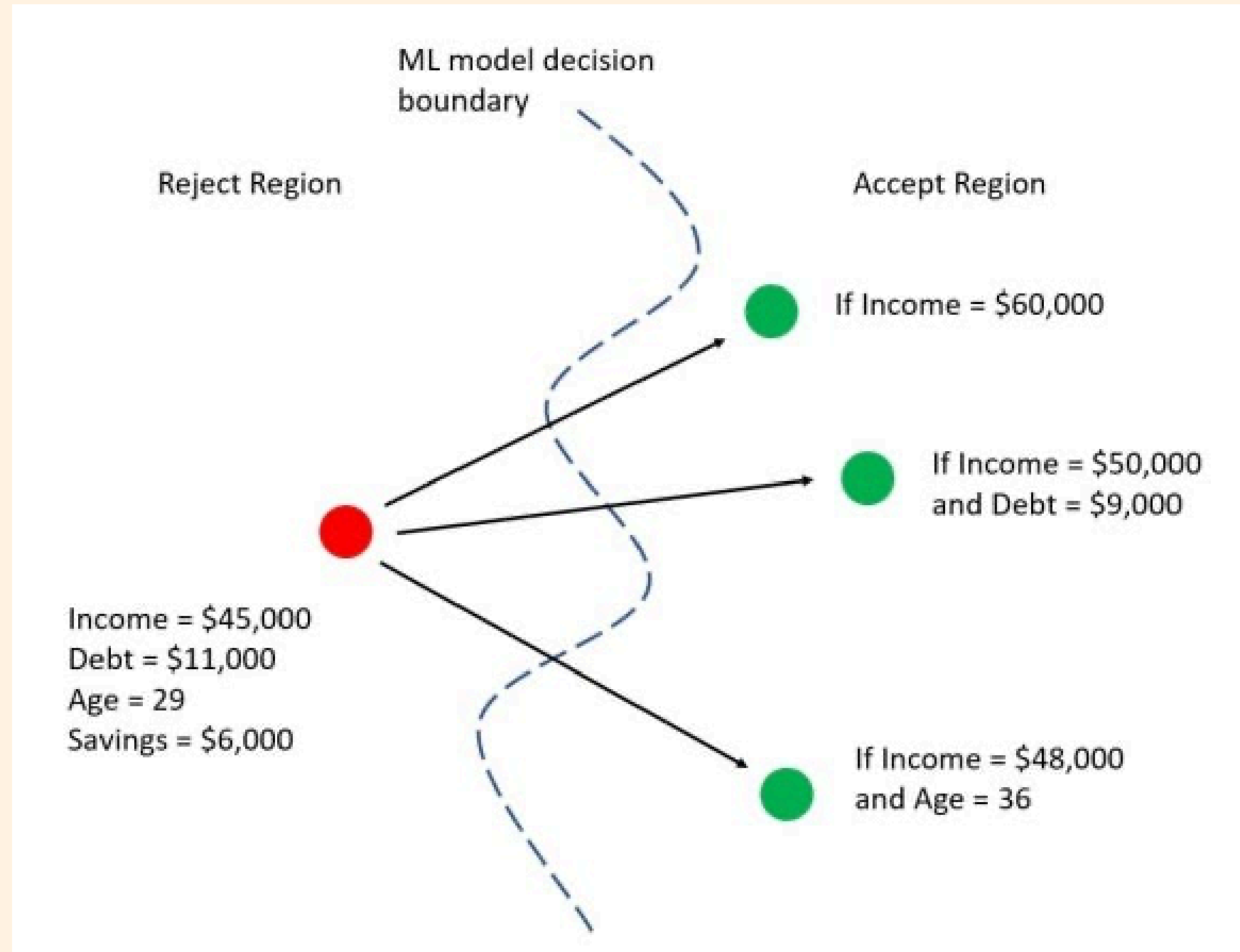
SHapely Additive Explanations (SHAP)



Local Interpretable Model-agnostic Explanations (LIME)

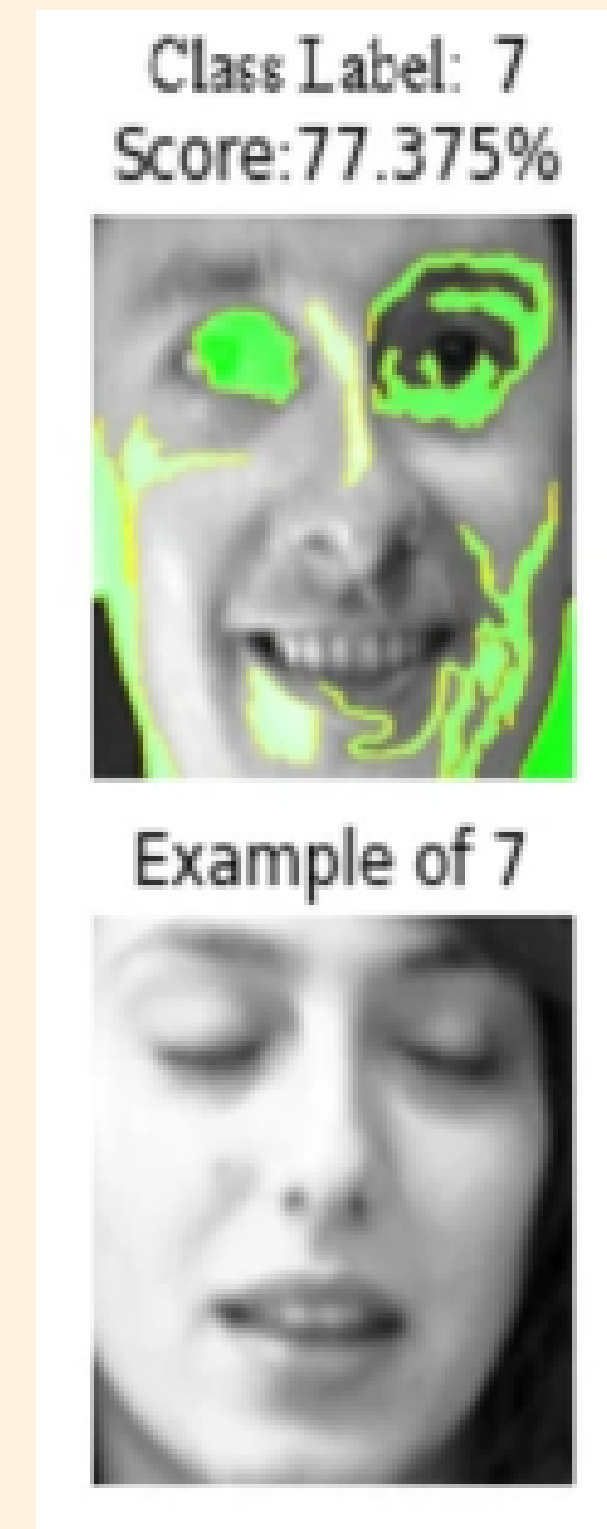
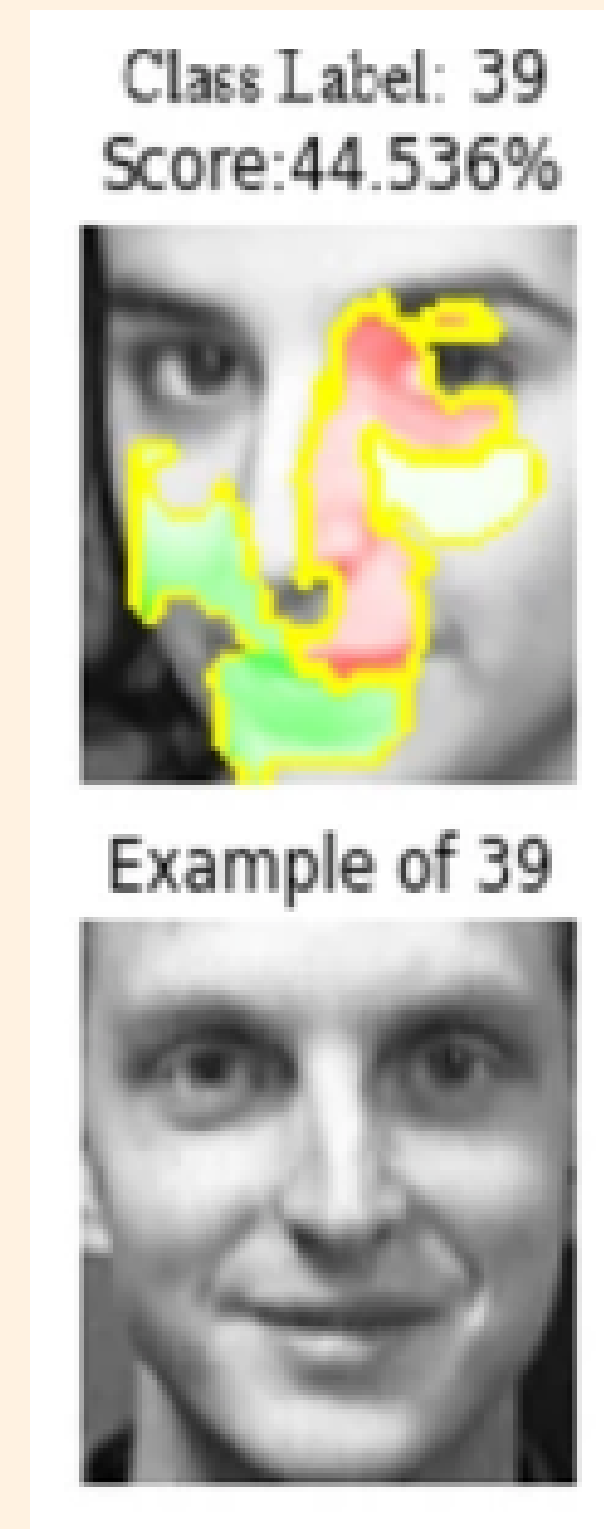
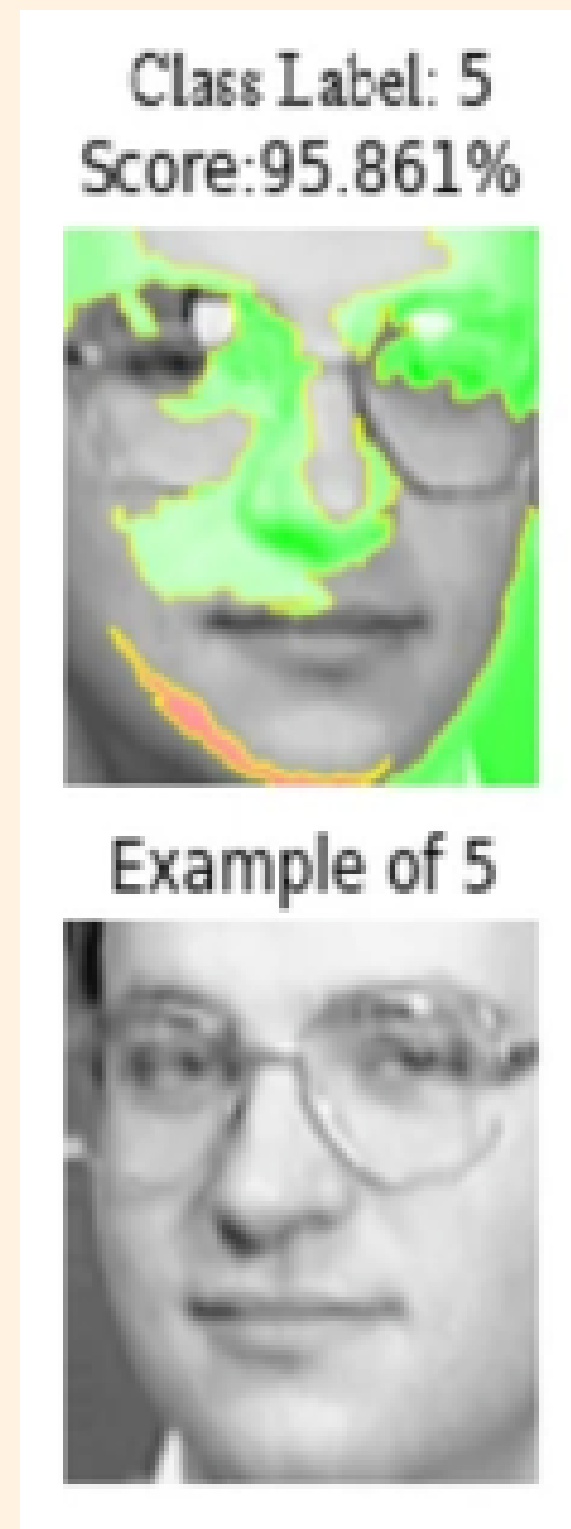
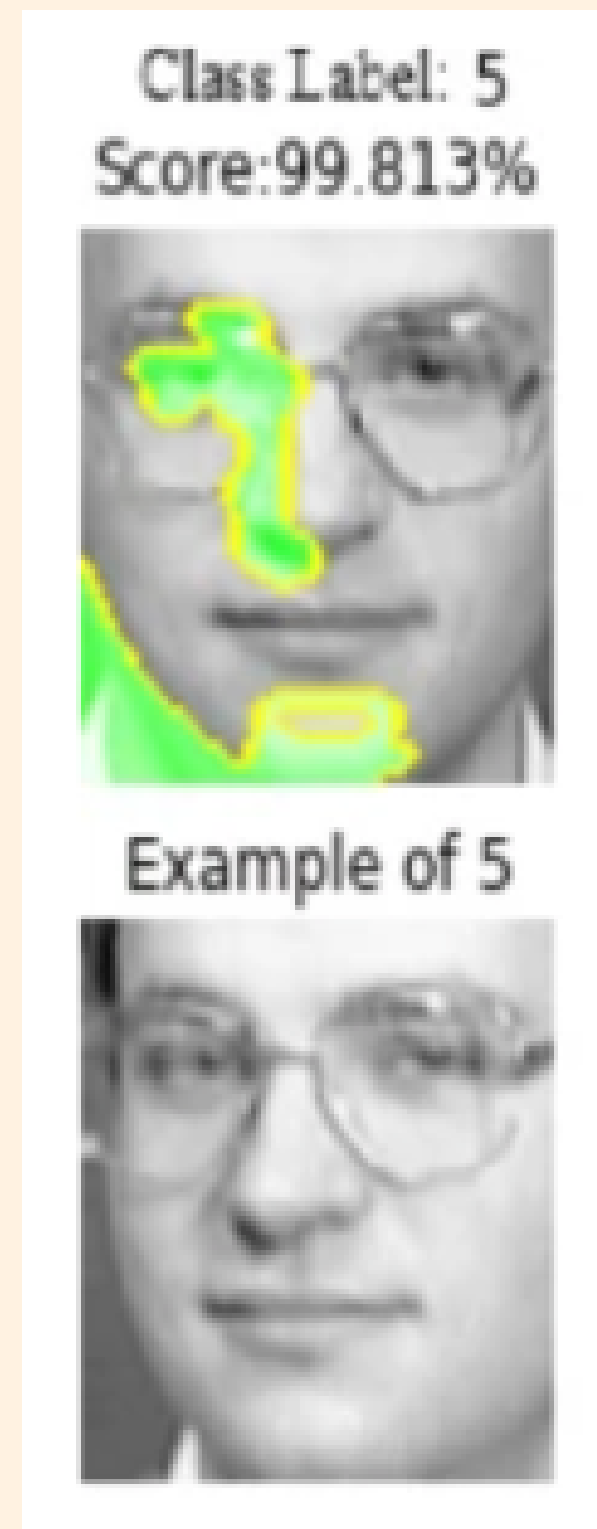


Counterfactual Explanations



You will be accepted if:

- your income increases to \$60,000,
- your income increases to \$50,000 while debt decreases to \$9,000 or
- you're 7 years older with a slightly higher income.

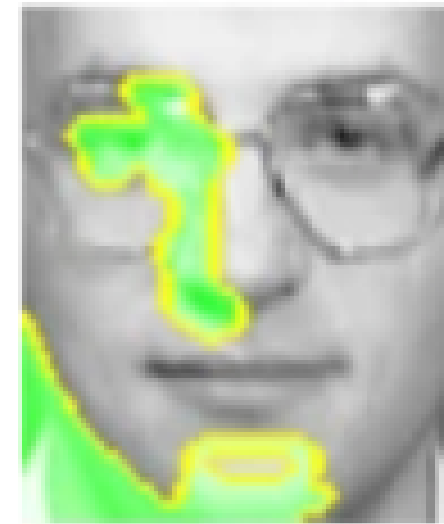


Rajpal, A. et al. (2023). “XAI-FR: Explainable AI-Based Face Recognition Using Deep Neural Networks.” *Wireless Pers Commun* 129, pages 663–680. Available at: <https://doi.org/10.1007/s11277-022-10127-z> (Accessed 22.5.24).

Police Stop and Search



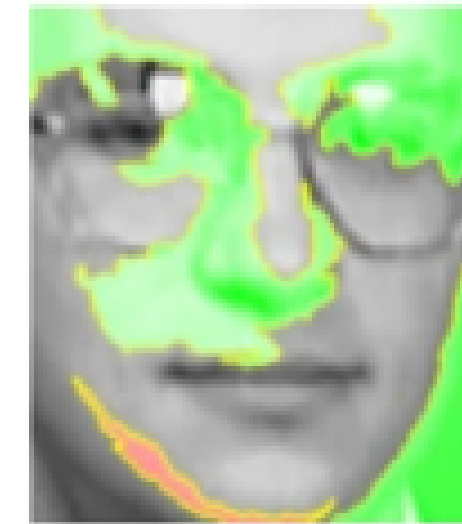
Class Label: 5
Score:99.813%



Example of 5



Class Label: 5
Score:95.861%



Example of 5



Which algorithms are used to provide inputs to the decision, why were these chosen, and why are they considered appropriate for use in this specific context?

Prediction

What factors informed the decision and why are they considered appropriate for the context and impact of the substantive interventions?

Why is the policy considered appropriate and lawful, including its objectives, specification, communication, supervision and implementation?

Contextual factors

DECISION!

Human-in-the-loop
vs
Fully automated

Prediction

Confidence

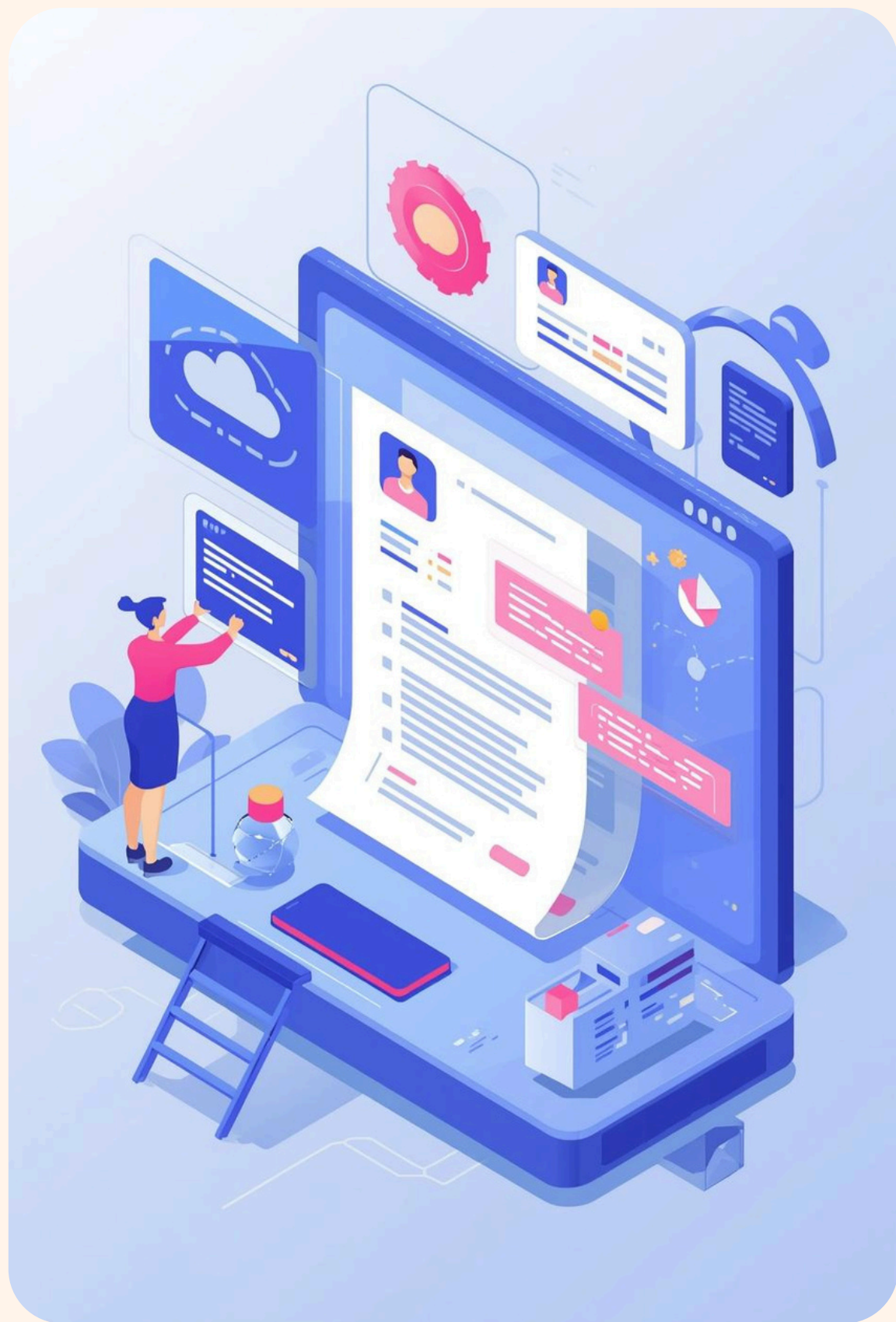
Substantive intervention

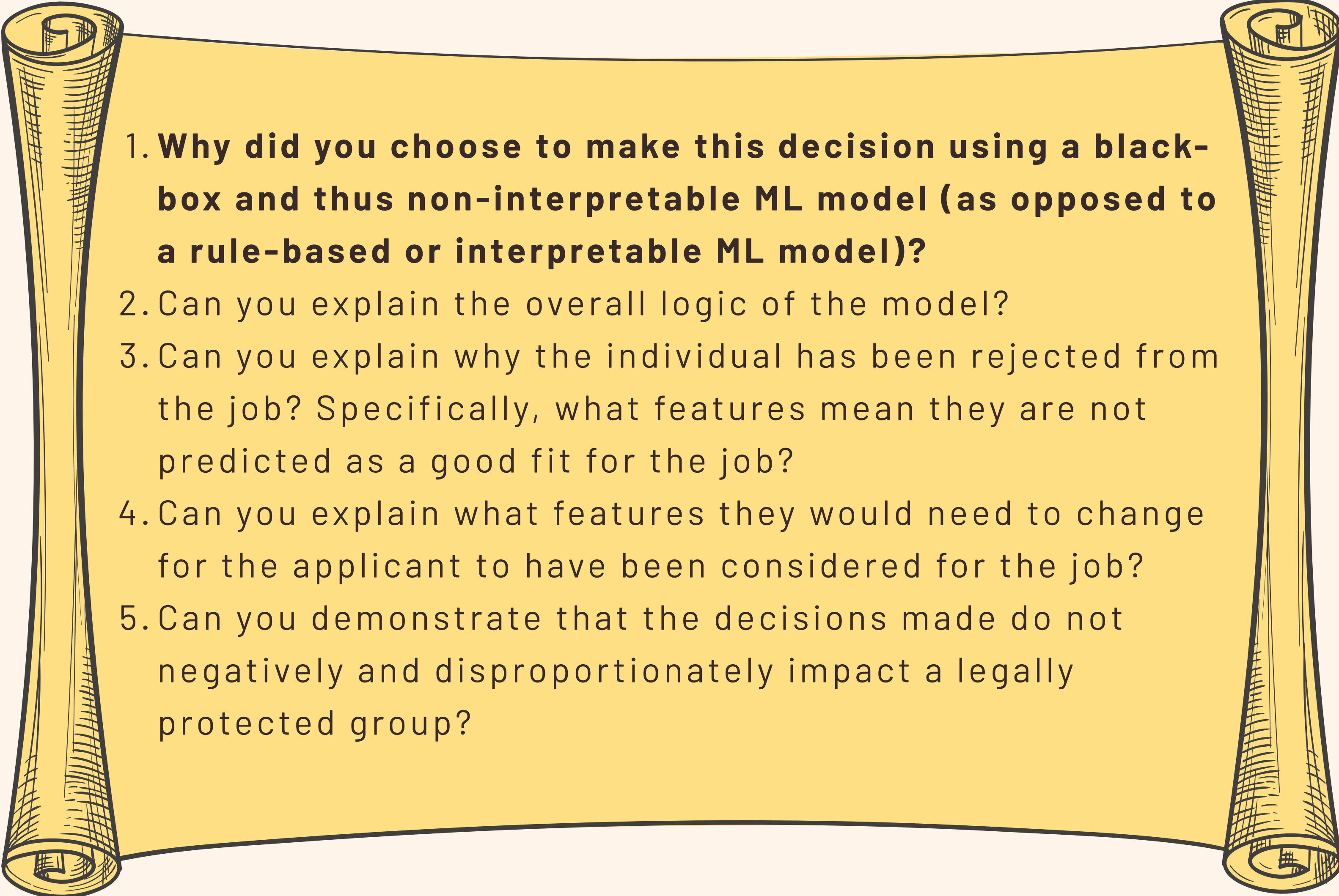
Decision

ACTION!

Accountability
and
how to deal with errors

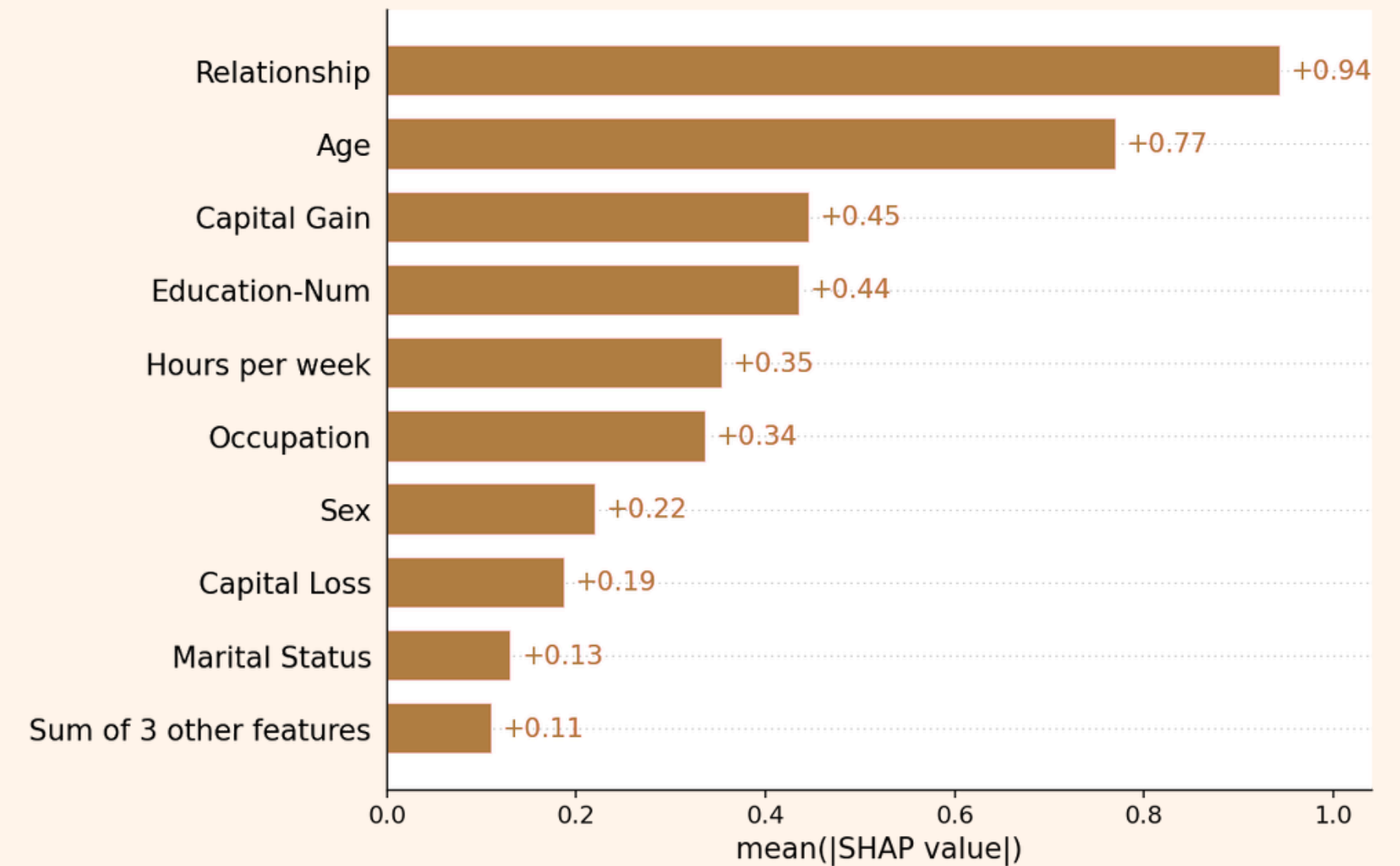
Recruitment Example



- 
1. **Why did you choose to make this decision using a black-box and thus non-interpretable ML model (as opposed to a rule-based or interpretable ML model)?**
 2. Can you explain the overall logic of the model?
 3. Can you explain why the individual has been rejected from the job? Specifically, what features mean they are not predicted as a good fit for the job?
 4. Can you explain what features they would need to change for the applicant to have been considered for the job?
 5. Can you demonstrate that the decisions made do not negatively and disproportionately impact a legally protected group?

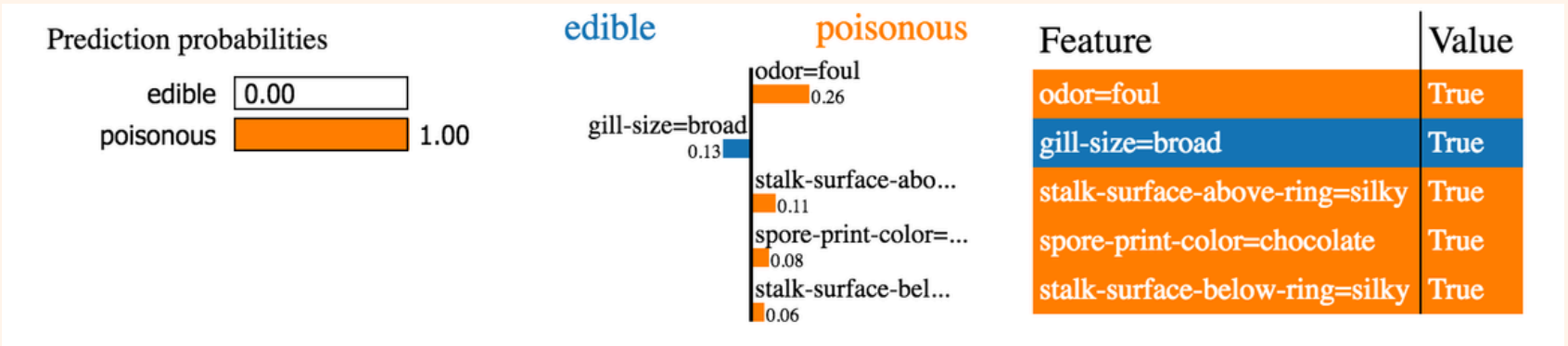
2. Can you explain the overall logic of the model?

SHapely Additive Explanations (SHAP)



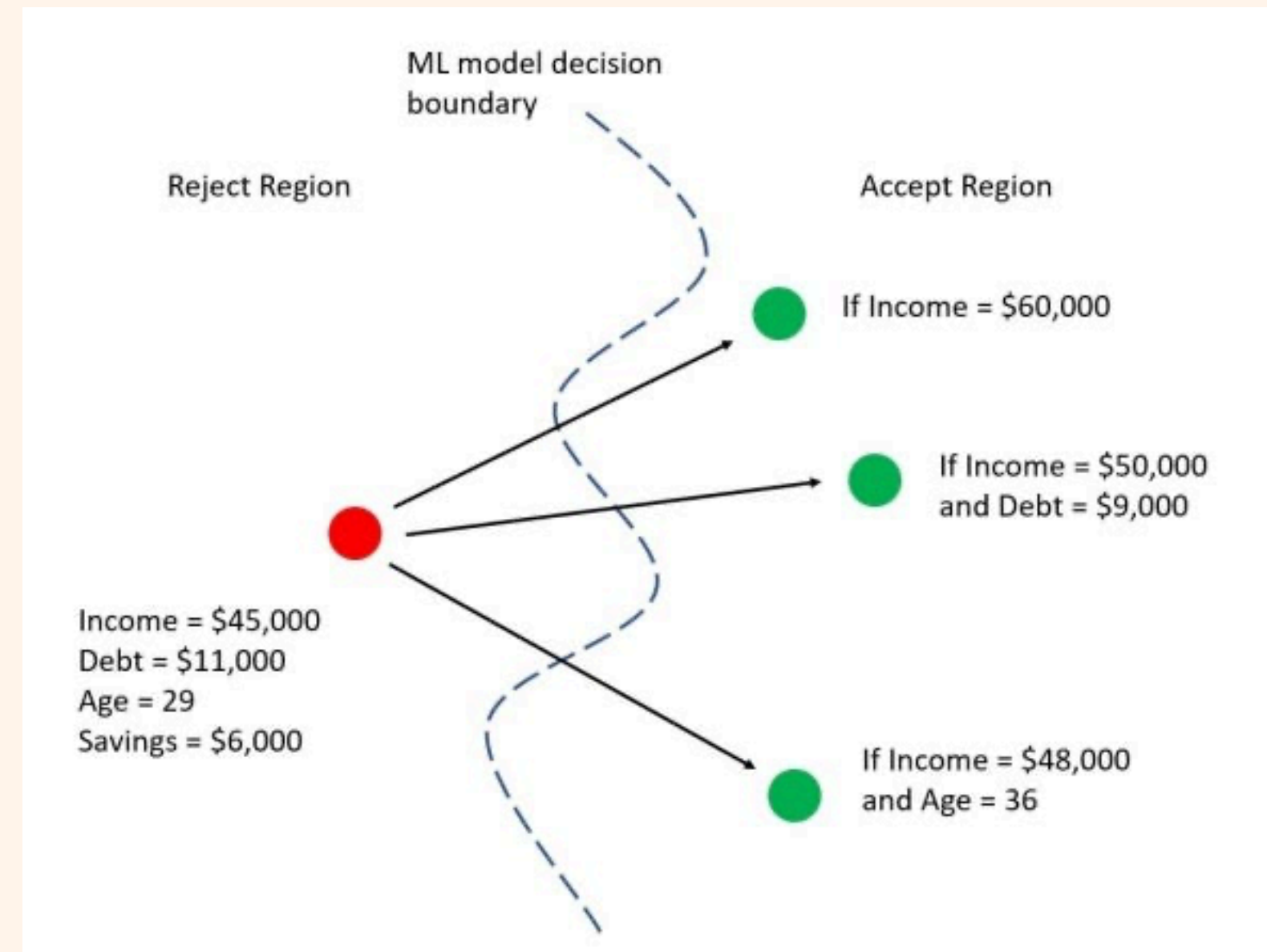
3. Can you explain why the individual has been rejected from the job? Specifically, what features mean they are not predicted as a good fit for the job?

Local Interpretable Model-agnostic Explanations (LIME)



4. Can you explain what features they would need to change for the applicant to have been considered for the job?

Counterfactual Explanations



**5. Can you demonstrate
that the decisions
made do not negatively
and disproportionately
impact a legally
protected group?**

functional explanation $\neq$ legal justification

- Whether XAI methods can 'help' to satisfy legal obligations depends on the nature and scope of the applicable legal obligations
- XAI could help understand the underlying logic of a fully automated system and thus help demonstrate compliance with e.g., Article 15 of GDPR

Open Questions

WHAT?

explanations are required

WHAT?

explanations are required

types of ADMS are explanations most likely to support

WHAT?

explanations are required

types of ADMS are explanations most likely to support

motivates their creation and use, if XAI methods are not legally sufficient

WHAT ?

explanations are required

types of ADMS are explanations most likely to support

motivates their creation and use, if XAI methods are not legally sufficient

HOW ?

does sectoral context affect the usefulness of XAI methods

WHAT ?

explanations are required

types of ADMS are explanations most likely to support

motivates their creation and use, if XAI methods are not legally sufficient

HOW ?

does sectoral context affect the usefulness of XAI methods

can legal obligations inform the design of technical explainability approaches

WHAT ?

explanations are required

types of ADMS are explanations most likely to support

motivates their creation and use, if XAI methods are not legally sufficient

HOW ?

does sectoral context affect the usefulness of XAI methods

can legal obligations inform the design of technical explainability approaches

can explanations be evaluated across legal and technical criteria

WHO?

is legally entitled to receive explanations

WHO ?

is legally entitled to receive explanations

WHERE ?

can XAI methods have been practically useful for legal compliance

WHO ?

is legally entitled to receive explanations

WHERE ?

can XAI methods have been practically useful for legal compliance
can we definitively say that XAI methods do not satisfy legal obligations

WHO ?

is legally entitled to receive explanations

WHERE ?

can XAI methods have been practically useful for legal compliance
can we definitively say that XAI methods do not satisfy legal obligations
can explanations be evaluated across legal and technical criteria

Thank you!

Any questions?

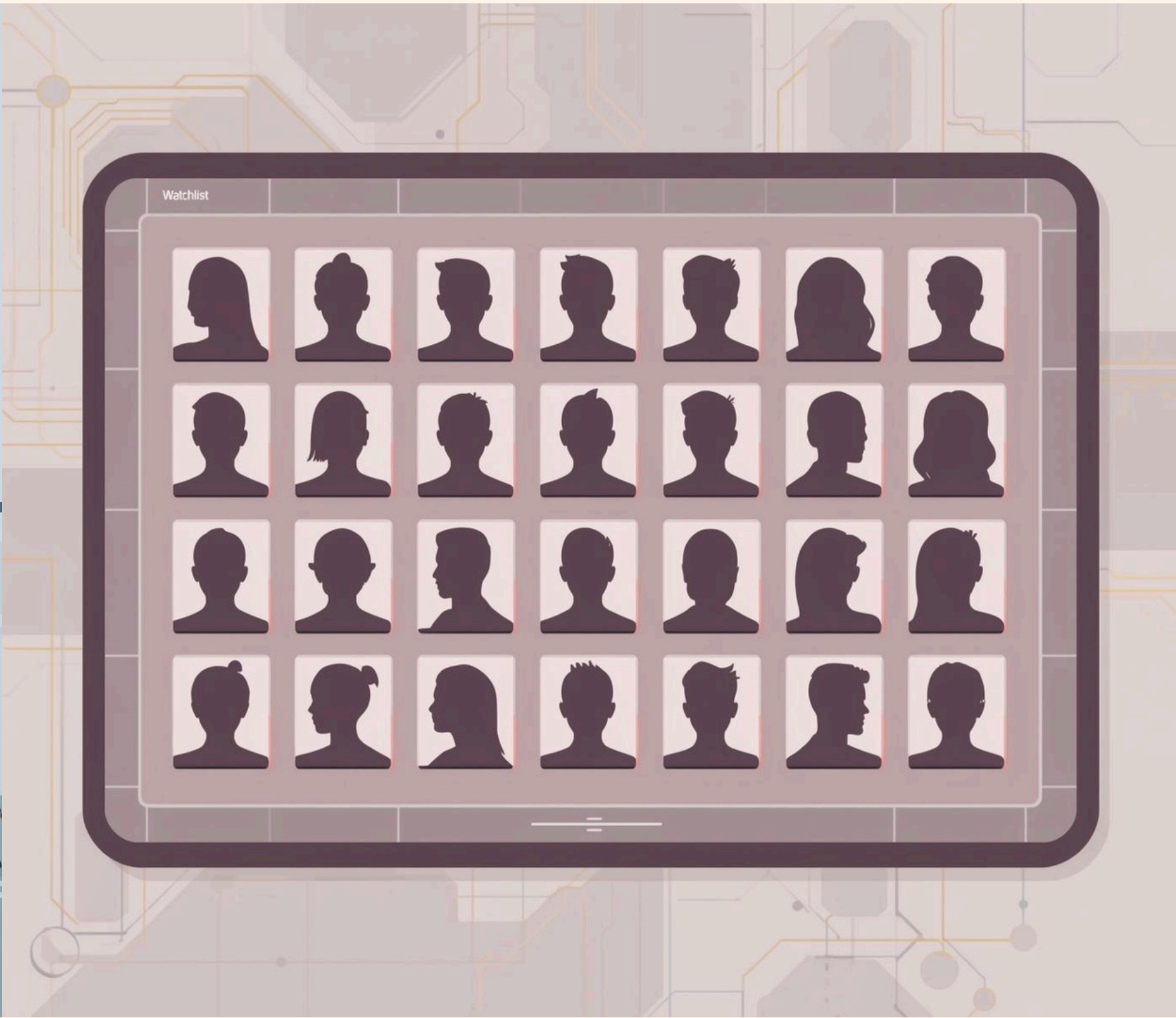


Keep in touch:

madeleine.waller@uni-tuebingen.de

 [/maddiewaller/](https://www.linkedin.com/company/maddiewaller/)





LIME

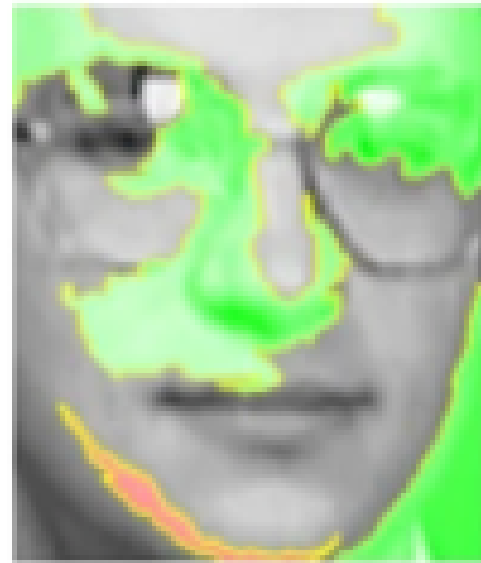
Class Label: 5
Score:99.813%



Example of 5



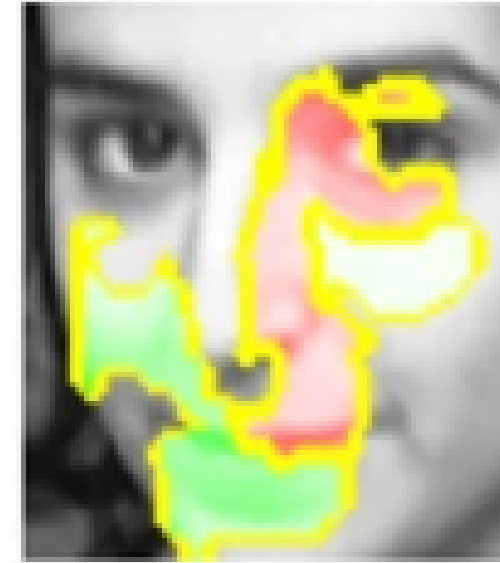
Class Label: 5
Score:95.861%



Example of 5



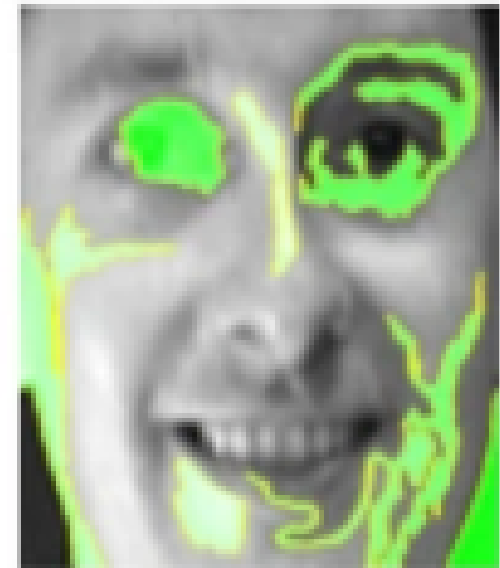
Class Label: 39
Score:44.536%



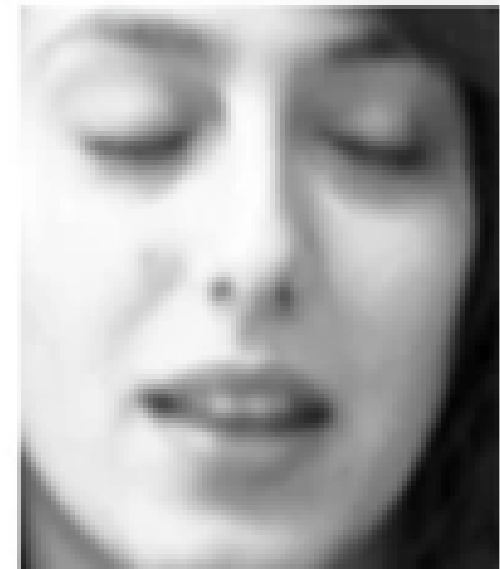
Example of 39

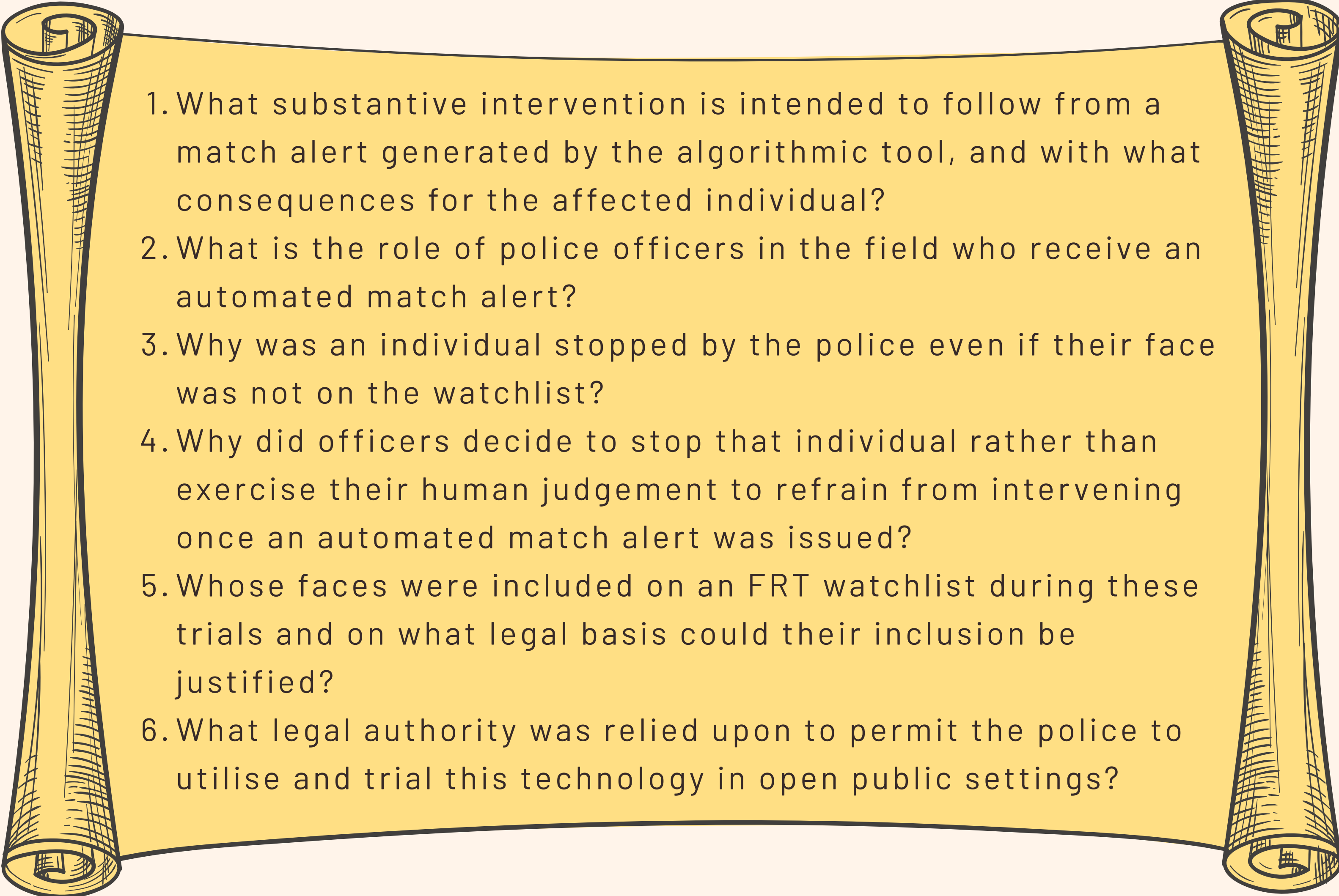


Class Label: 7
Score:77.375%



Example of 7



- 
1. What substantive intervention is intended to follow from a match alert generated by the algorithmic tool, and with what consequences for the affected individual?
 2. What is the role of police officers in the field who receive an automated match alert?
 3. Why was an individual stopped by the police even if their face was not on the watchlist?
 4. Why did officers decide to stop that individual rather than exercise their human judgement to refrain from intervening once an automated match alert was issued?
 5. Whose faces were included on an FRT watchlist during these trials and on what legal basis could their inclusion be justified?
 6. What legal authority was relied upon to permit the police to utilise and trial this technology in open public settings?