Privacy-enhancing language-based machine learning for detecting effective language use in job interviews

Clement Canel, Dnyaneshwari Rakshe, and Ratnakshi Gore

Problem Statement

- Can we train ML models to detect how well someone explains their answer in a job interview—while reducing bias from who the speaker is?

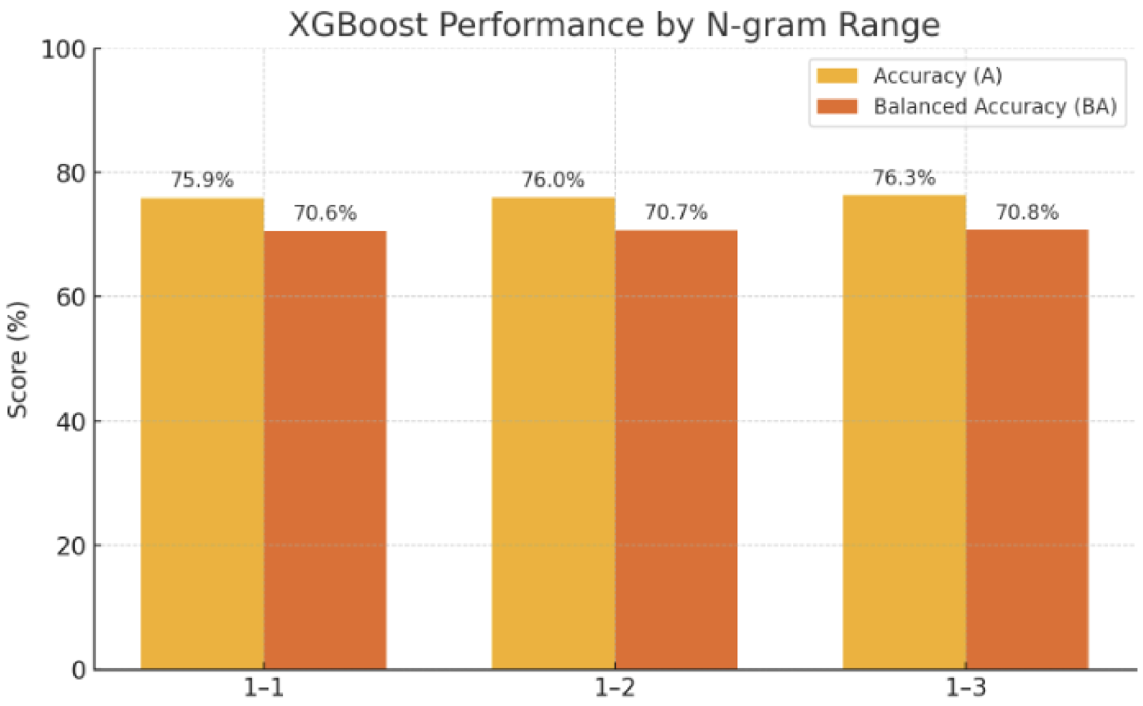
- We aim to classify responses (e.g., succinct, over-explained) and explore **privacy-enhancing techniques** to make feedback **fairer**.

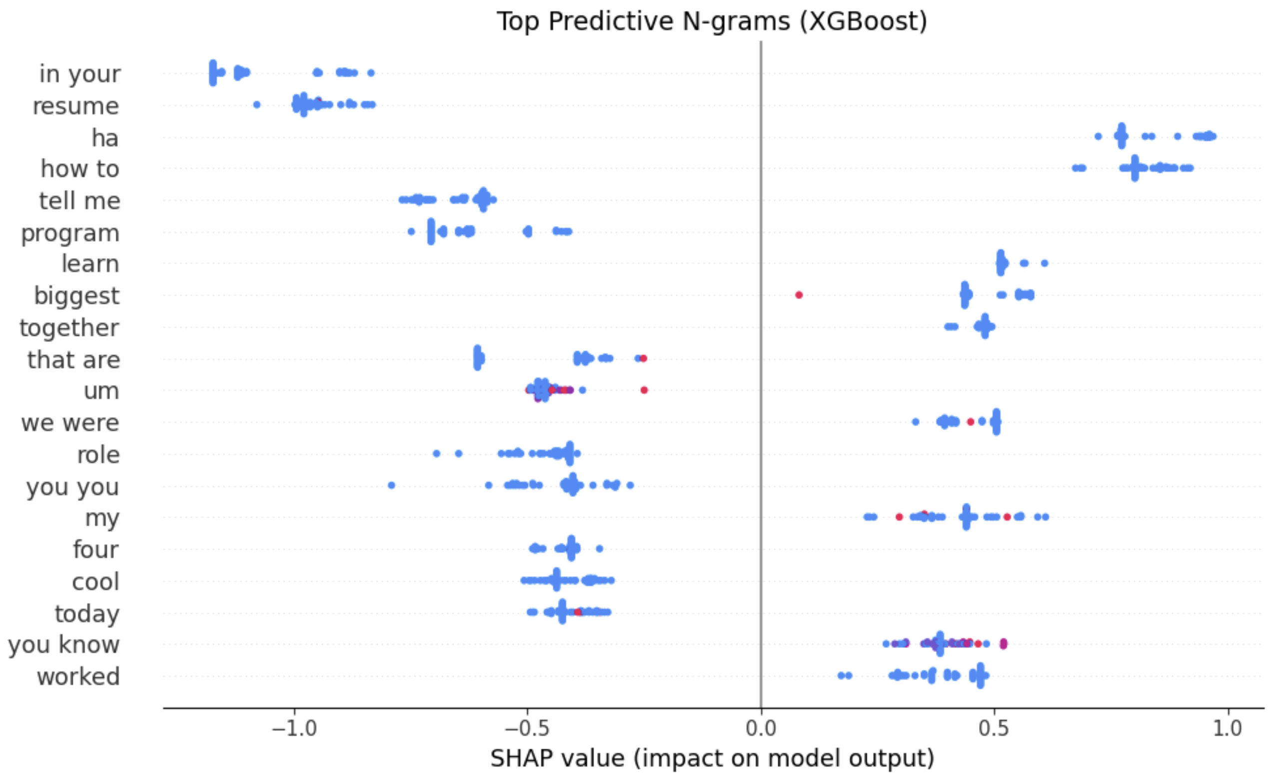
Methodology - Data Overview

- **Dataset: VetTrain transcripts and Behavioral annotations**
  - Participants: 38 (with anonymized IDs)
  - Data: Interview transcripts (questions and responses)
  - Labels: Four classes based on response explanation quality
  - Classes for degree of explanation:
    - Under-explained: 23 samples
    - Succinct: 107 samples
    - Comprehensive: 122 samples
    - Over-explained: 34 samples
- **Preprocessing**
  - Steps: Removed speaker labels, extracted Q&A pairs
  - Parsed transcripts and aligned behavioral annotations

Feature Engineering

Feature	Extracted Features	Purpose
Linguistic	TF-IDF, CountVectorizer	Style, vocabulary, n-grams
Structural	Duration, Lag, Word Count	Speech pace, verbosity
Semantic	Sentiment, Subjectivity	Emotional tone
Topic Similarity	TopicID, TopicProb, Q-A similarity score	Relevance of answer to question





XG Boost model for Speaker Classification

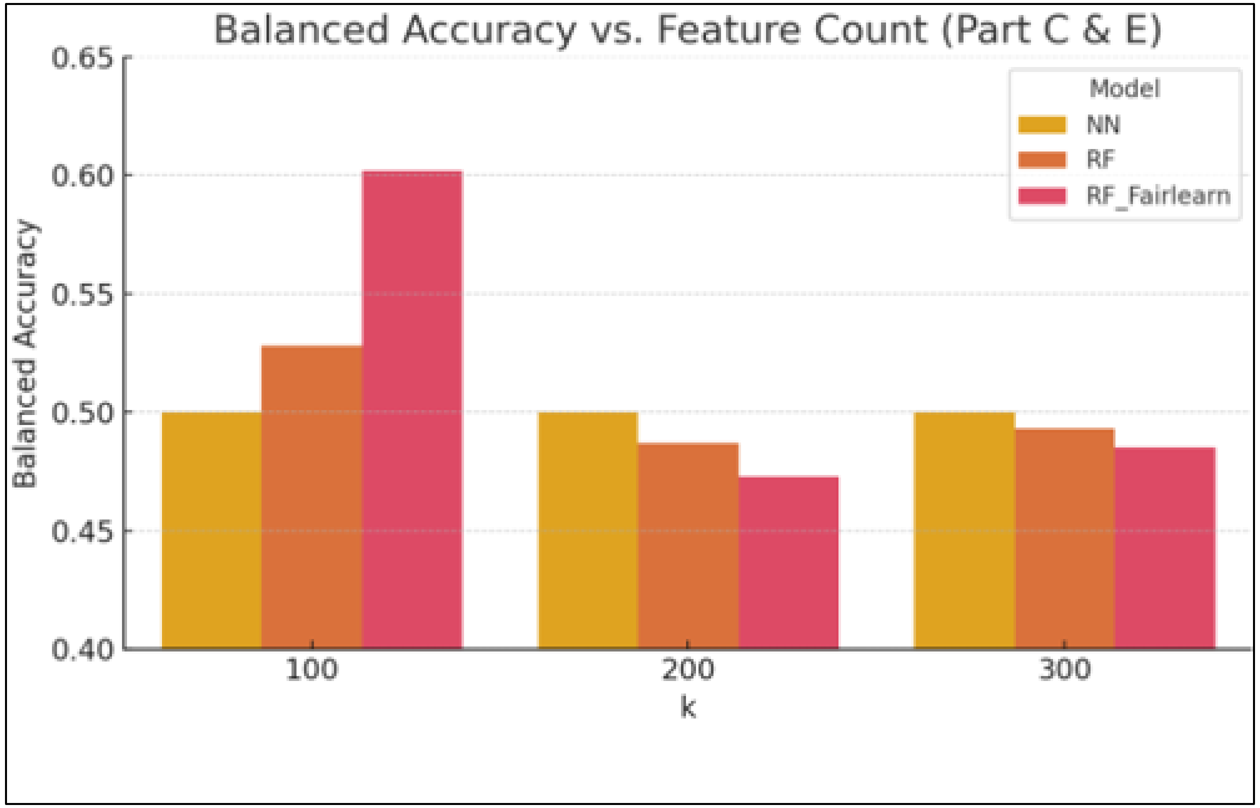
- Used **CountVectorizer** to extract unigram, bigram, and trigram features.
- Particularly strong at handling structured data and sparse features like text n-grams.
- It builds a series of trees that split based on presence or absence of these n-grams.
- Each decision tree splits based on thresholds of n-gram frequencies

FNN model for Speaker Classification

- Architecture**:** 3-layer neural network with ReLU activations and Dropout for regularization.
- Used **TF-IDF vectorization** with **unigrams and bigrams**.
- Features ranked using **chi-squared statistical test.**
- These features help distinguish speaker identity even without explicit speaker tags.
- Achieved high performance with:
  Accuracy : 0.78
  Balanced Accuracy : 0.73

```
Top 200 features by χ²:
['about' 'about time' 'about you' 'about your' 'alright' 'alright alright'
 'alright cool' 'alright so' 'and was' 'any' 'any experience'
 'any questions' 'anything you' 'appreciate' 'appreciate you'
 'appreciate your' 'are' 'are you' 'ask' 'ask that' 'ask you' 'at your'
 'awesome' 'background' 'bit about' 'can you' 'cool' 'cool cool' 'cool so'
 'cool um' 'curious' 'curious like' 'description' 'design build' 'did you'
 'do you' 'doing well' 'essentially' 'excellent' 'fantastic' 'fit for'
 'for you' 'for your' 'good' 'good good' 'good so' 'good um' 'gosh'
 'gotcha' 'great' 'great so' 'great that' 'great well' 'ha' 'ha ha' 'haha'
 'has there' 'haskell' 'have any' 'have for' 'have good' 'have you'
 'high level' 'how did' 'how do' 'how would' 'however' 'huh' 'impactful'
 'impactful leadership' 'in terms' 'in your' 'interview' 'interviewing'
 'is there' 'it sounds' 'it was' 'job description' 'last question'
 'like your' 'me about' 'me little' 'meet you' 'method' 'mission' 'mm'
 'my' 'name' 'name how' 'obviously' 'obviously you' 'of curious' 'oh' 'ok'
 'ok cool' 'ok ok' 'ok um' 'okay' 'okay great' 'okay so' 'once'
 'other questions' 'perfect' 'question' 'questions' 'questions do'
 'questions for' 're interviewing' 'resume' 'role' 'role that' 'say is'
 'says' 'sir' 'so how' 'so is' 'so tell' 'so um' 'so what'
 'software development' 'sorry' 'sounds like' 'speak' 'spent' 'star'
 'tell' 'tell me' 'terms' 'terms of' 'that awesome' 'that great'
 'that job' 'that was' 'that with' 'that you' 'the' 'them'
 'there anything' 'they' 'thing you' 'this job' 'to ask' 'today' 'uh' 'um'
 'umm' 'very cool' 'was' 'wasn' 'we are' 'we had' 'well' 'well good'
 'west point' 'what' 'what are' 'what questions' 'what would' 'what you'
 'when you' 'would you' 'yeah' 'yeah yeah' 'yep' 'yes' 'you' 'you and'
 'you ask' 'you did' 'you enjoy' 'you feel' 'you for' 'you go' 'you had'
 'you handle' 'you have' 'you like' 'you might' 'you said' 'you say'
 'you tell' 'you the' 'you think' 'you to' 'you uh' 'you ve' 'you were'
 'you you' 'your' 'your background' 'your career' 'your education'
 'your experience' 'your favorite' 'your leadership' 'your resume'
 'your service' 'your team' 'your time' 'your work']
```

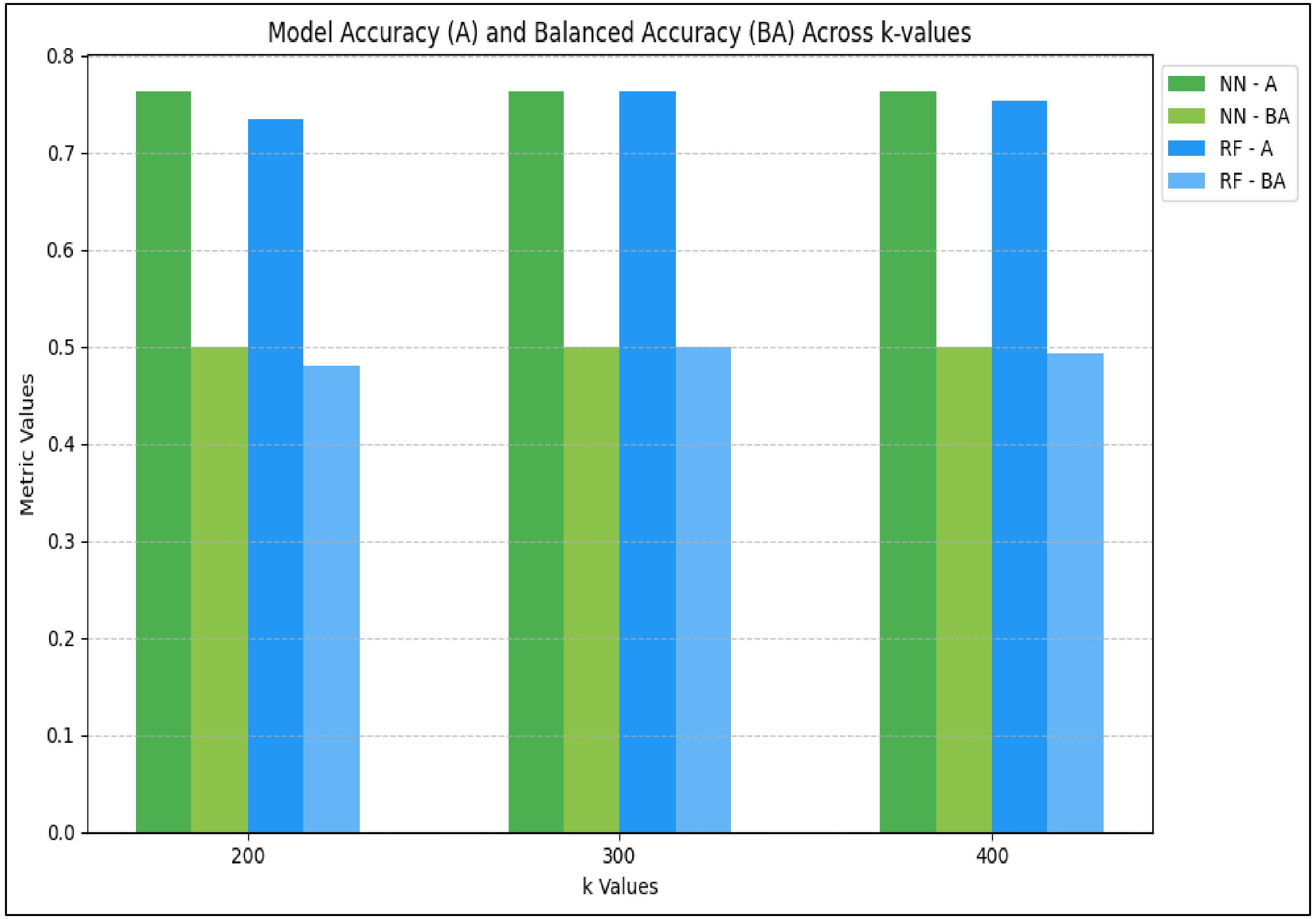
Model Performance Comparison



- **Classes Considered:**
  - **Succinct**
  - **Under-explained**

K-values: 100, 200, 300
Tree-Based Model: Random forest
Deep Learning Model: FNN
Toolbox for in-processing: Fairlearn

Model Performance Comparison



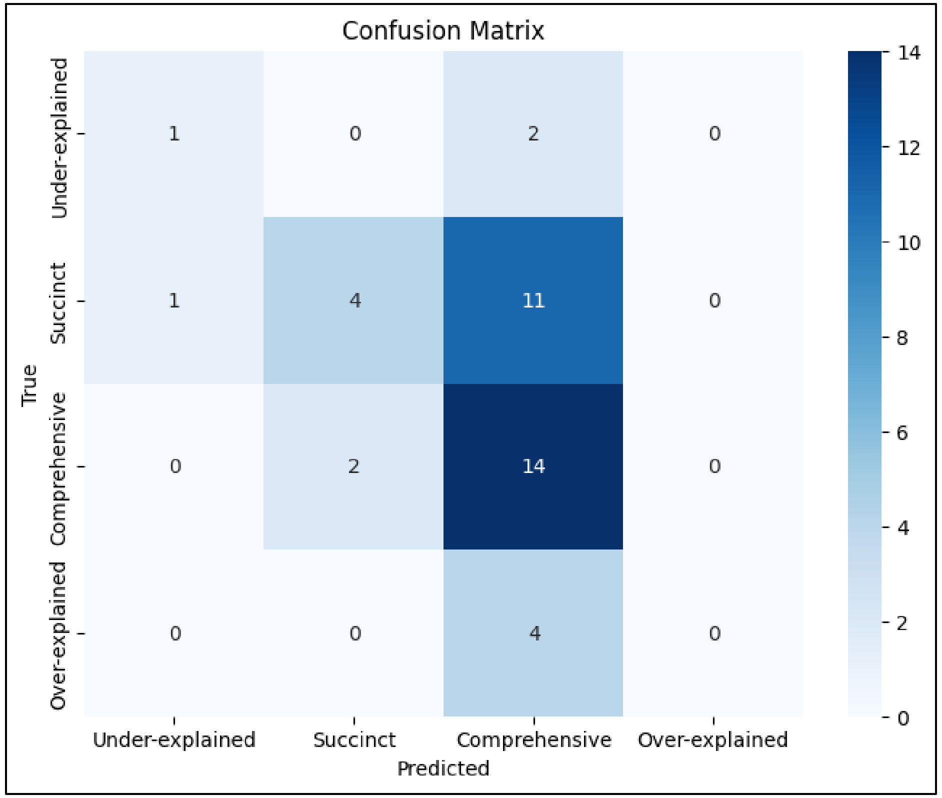
- **Classes Considered:**
    - **Comprehensive**
    - **Over-explained**

K-values: 200, 300, 400
Tree-Based Model: Random forest
Deep Learning Model: FNN

Observations:

- **Accuracy (A):** 0.76
- **Balanced Accuracy (BA):** 0.50

Experimenting with Transformers



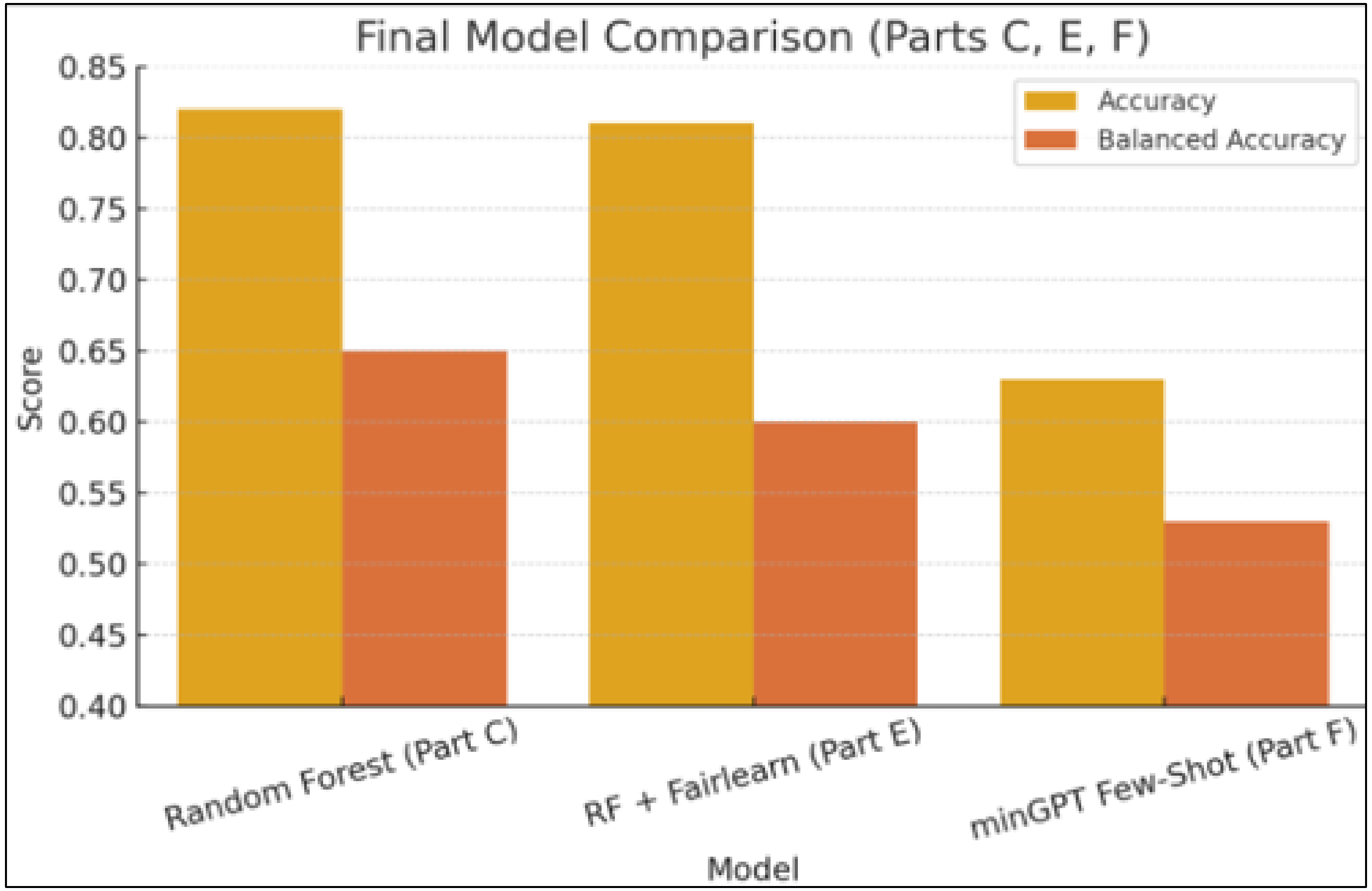
Transformer used: **minGPT**

- Most true labels are predicted as 'Comprehensive', especially those originally labeled as 'Succinct' and 'Over-explained'.
- 'Comprehensive' class has the highest true positive count (14), showing the model leans heavily toward this class.
- No instances of 'Over-explained' were predicted correctly.
- Indicates model bias and need for class rebalancing techniques like oversampling or class weights.

Limitations of minGPT:

- Lightweight architecture
- Shallow context understanding
- Outdated training techniques
- Not-Domain Tuned

Performance Comparison



Random Forest vs. RF+Fairlearn vs minGPT

Conclusion

- ML models successfully classify explanation quality but bias persists.
- Fairlearn mitigates some imbalances but further tuning is required.
- Transformers skew heavily toward ‘Comprehensive’ responses.
- Next steps: Enhanced fairness strategies, improved class balance, domain-specific fine-tuning.

Any Questions?

THANK YOU!