GNN and LLM Insights: Multimodal Cues and Gender Disparities in Video Conversations

Dimosthenis Stefanidis, George Pallis, Marios D. Dikaiakos, Nicos Nicolaou

{dstefa02, pallis, mdd}@ucy.ac.cy, Nicos.Nicolaou@wbs.ac.uk

Computer Science Department, University of Cyprus

Motivation 1

- **Interpersonal conversations shape major decisions:** from investor funding pitches to job interviews; yet understanding what truly influences these outcomes is challenging.
- **Multimodal signals** like facial expressions, voice tone, and spoken words interact dynamically, making it complex to detect their exact relation.
- **Most of the existing research** study these signals in isolation, missing critical interactions and temporal dynamics.
- **Crucially, biases (e.g., gender biases) may hide** in these subtle multimodal interactions, potentially disadvantaging specific groups.

Contributions 2

Research Question

"Do multimodal cues (visual, vocal, and verbal) affect investment decisions differently for men vs women?"

- **End-to-end pipeline** that turns raw unstructured video conversations into 30 multimodal cues + rich controls
- **Graph Neural Network (GNN)** tailored to conversational chronology; CAPTUM + GNNExplainer for transparency
- **Comprehensive bias audit** across models.

Variables 3

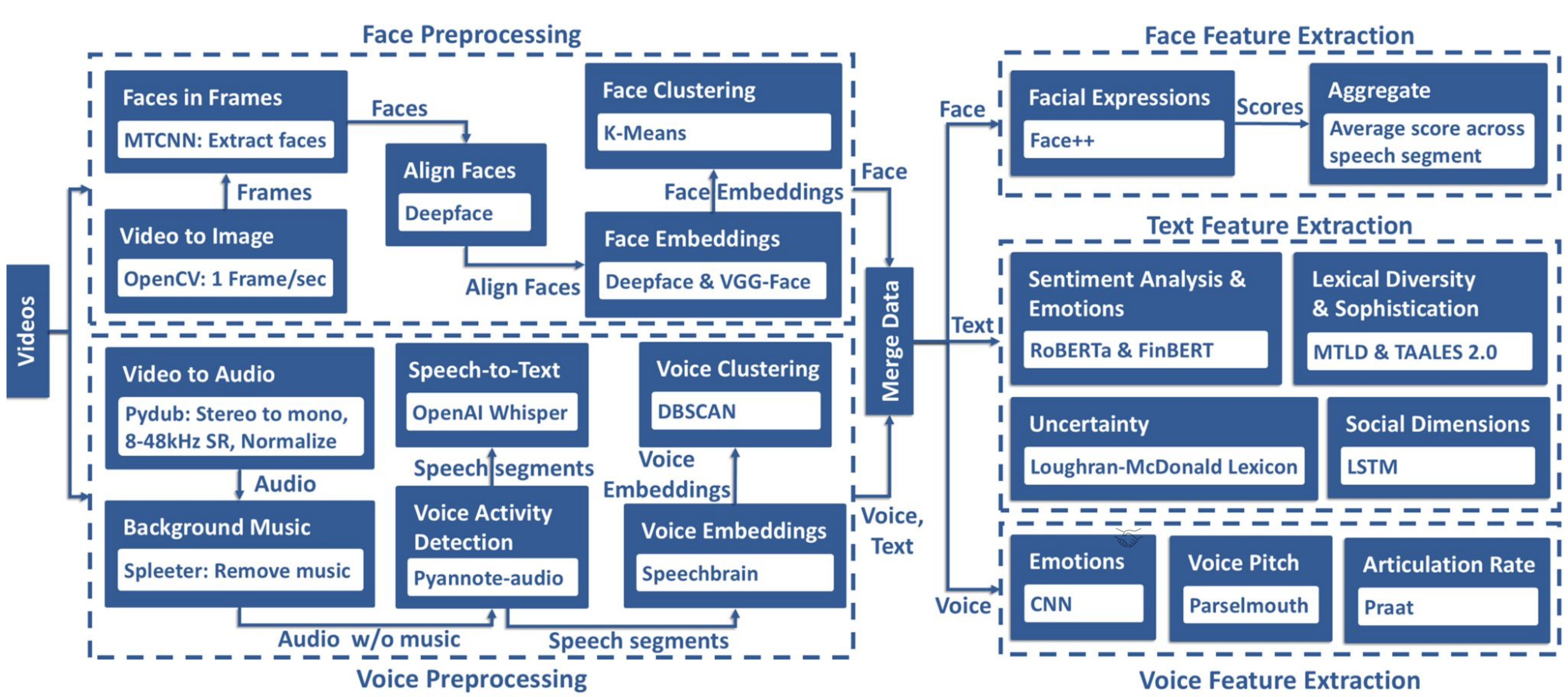
Independent Variables

- **Audio:** Voice Emotions, Voice Pitch, Voice Articulation Rate
- **Image:** Facial Expressions
- **Text:** Sentiment Analysis & Text Emotions, Social Dimensions in Conversation, Lexical Diversity & Sophistication, Uncertainty

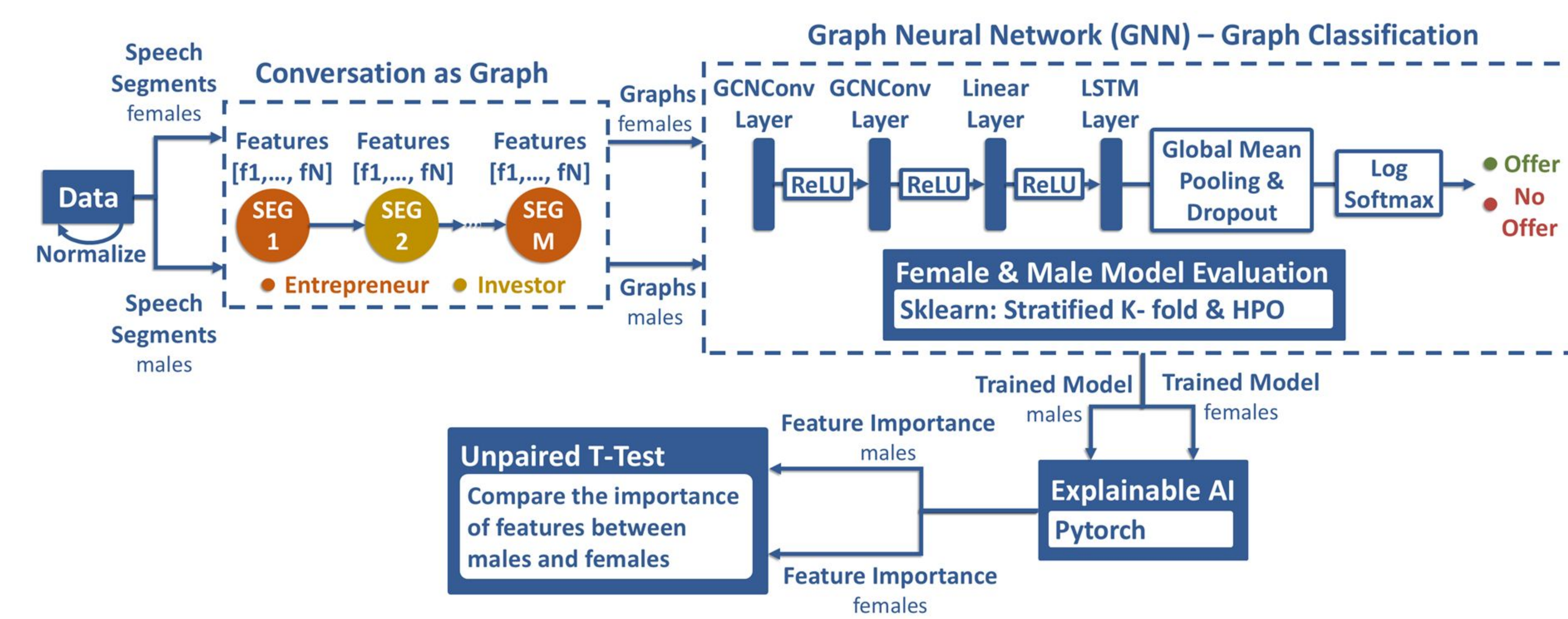
Dependent Variable

- **Receive funding:** whether a company has been offered funding or not, with values 1 and 0 respectively (binary)

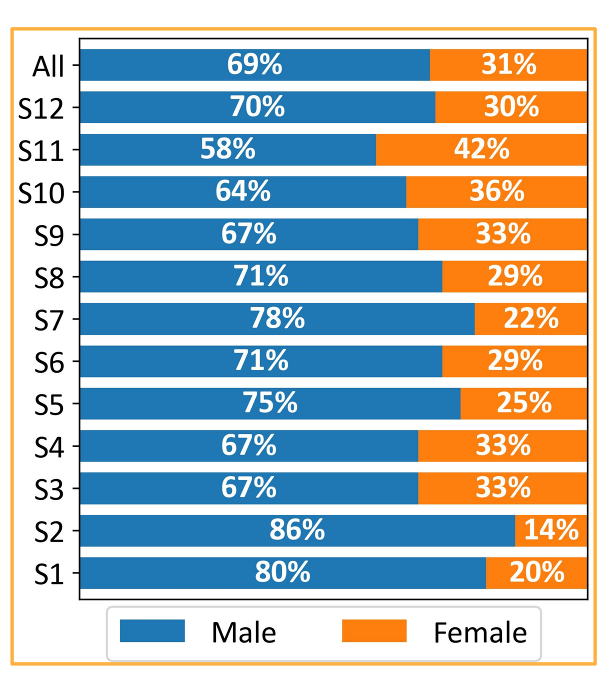
Video Processing Tool 4

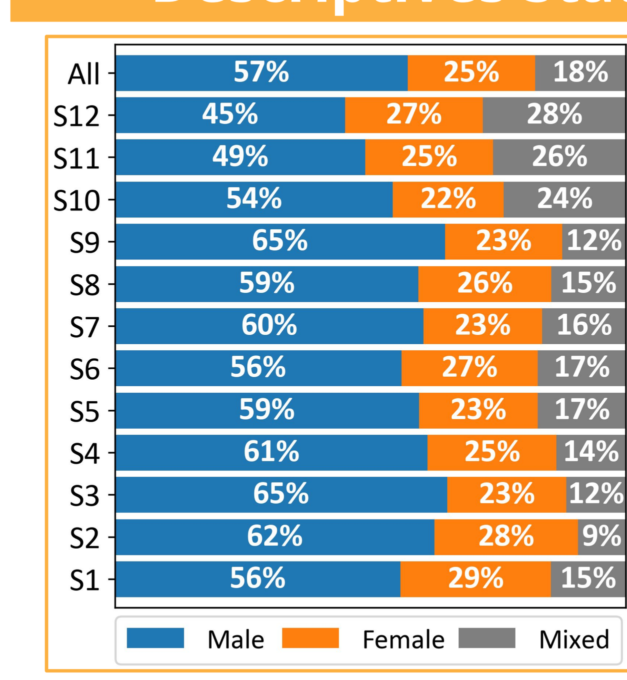


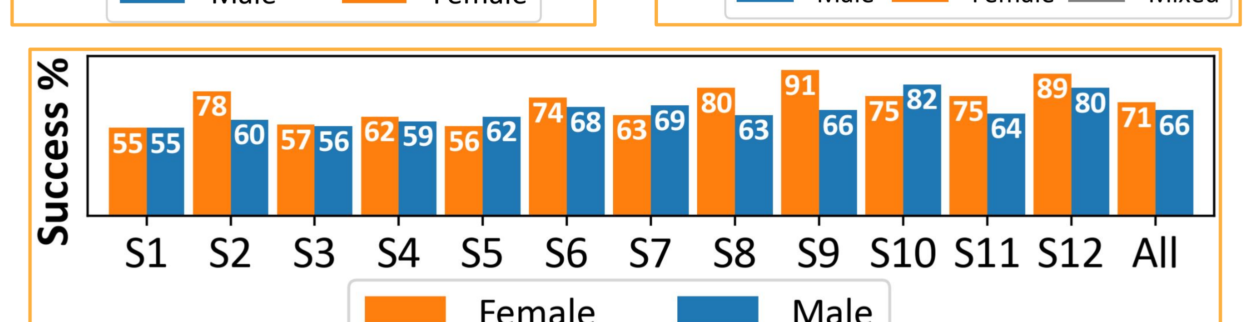
Graph Neural Network 5

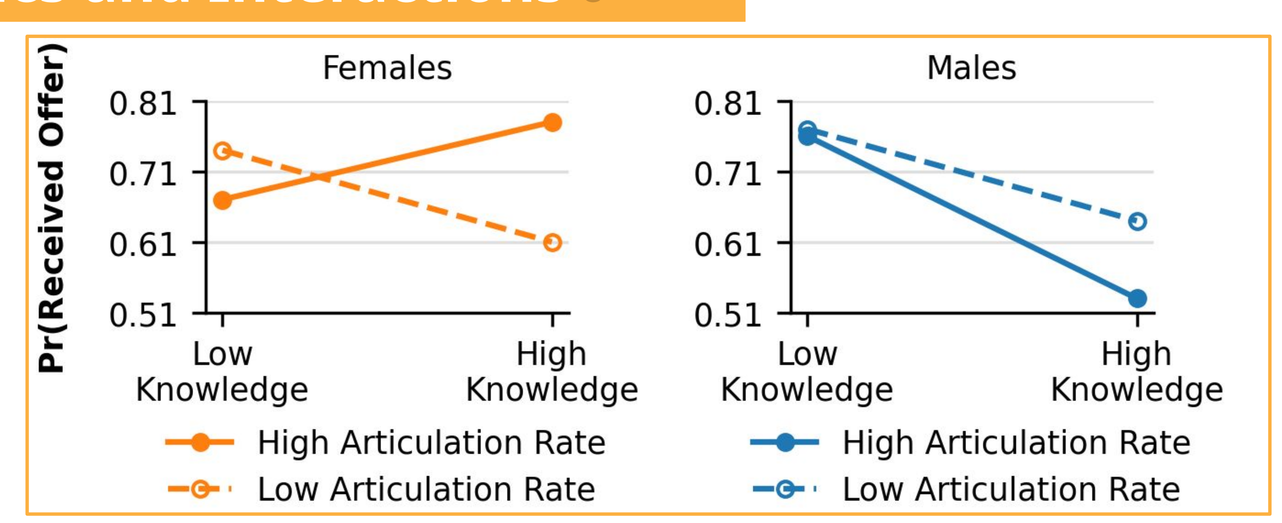


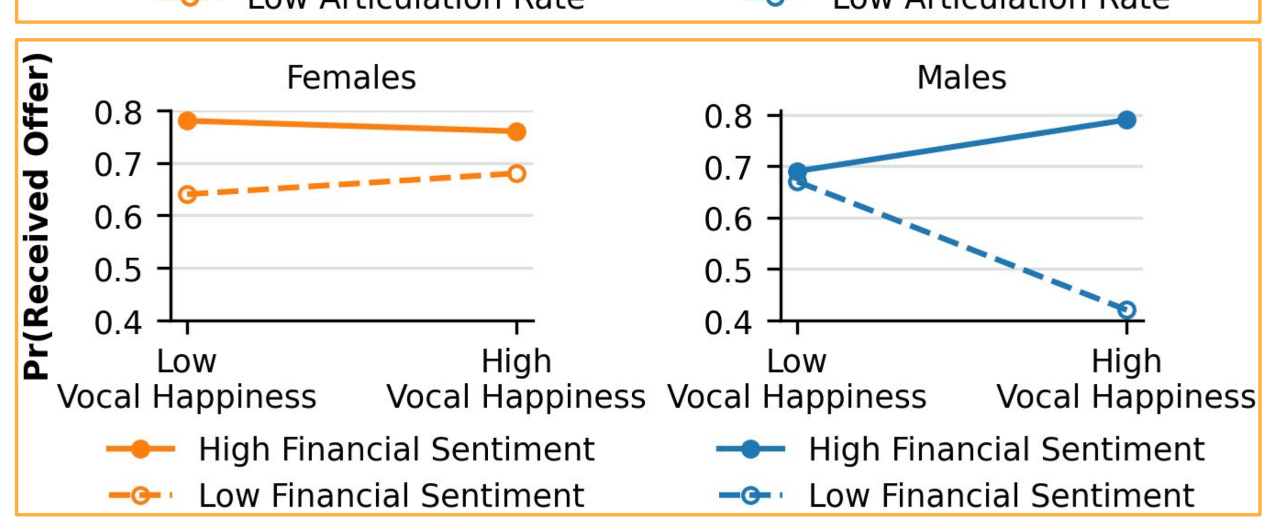
Descriptives Statistics and Interactions 6











GNN vs GPT-4 7

Performance Evaluation

Model	Precision	Recall	F1
GNN - Vocal (F)	0.70±0.07	0.63±0.03	0.64±0.03
GNN - Vocal (M)	0.64±0.04	0.60±0.03	0.60±0.04
GNN - Facial (F)	0.65±0.07	0.61±0.05	0.61±0.07
GNN - Facial (M)	0.64±0.03	0.60±0.02	0.60±0.03
GNN - Verbal (F)	0.69±0.05	0.66±0.04	0.67±0.04
GNN - Verbal (M)	0.61±0.14	0.57±0.11	0.57±0.11
GNN - All (F)	**0.72±0.05**	**0.69±0.06**	**0.70±0.06**
GNN - All (M)	**0.69±0.04**	**0.67±0.02**	**0.66±0.02**

F: Females, M: Males

Table 3: Performance of GNNs using Stratified 5-fold CV

Model	Precision	Recall	F1
GPT-4.0 Turbo (F)	0.794	0.435	0.562
GPT-4.0 Turbo (M)	0.736	0.404	0.521

F: Females, M: Males

- **Experiment Setup**
  - **Anonymise pitch** (Prompt 1) → present twice: no gender (Prompt 2) vs gender revealed (Prompt 3).
  - **GPT-4 Turbo** (9 Apr 2024), temperature 0; outcome = Funded / Not Funded
  - **Statistical tests**: F1, unpaired t-tests, logistic interaction.

Bias Evaluation

	Males		Females		t	p-value
	Mean	SD	Mean	SD		
GNN-All	0.736	0.441	0.739	0.440	-0.093	0.926

Table 4: Two-sample unpaired t-test to compare the predictions of GNN between males and females

	Males		Females		t	p-value
	Mean	SD	Mean	SD		
W/O Gender	0.359	0.480	0.391	0.489	-0.877	0.381
W/ Gender	0.423	0.494	0.556	0.498	-3.619	0.000

Table 6: Two-sample unpaired t-test to compare the predictions of GPT-4 between males and females