

# Tell me Habibi, is it **Real** or **Fake**?

Kartik Kuckreja<sup>†</sup>, Parul Gupta<sup>∧</sup>, Injy Hamed<sup>†</sup>, Tamar Solorio<sup>†</sup>, Muhammad Haris Khan<sup>†</sup>, Abhinav Dhall<sup>∧</sup>

*ICLR-2026*



<sup>†</sup> Mohammad Bin Zayed University of AI, <sup>∧</sup> Monash University, Australia

Arabic is the fifth most spoken language globally, often blended with English in everyday conversations through code-switching.



While **Generative AI** bridges cultures and languages, it also fuels the rise of **deepfakes**.



Although there have been some attempts at multilingual deepfake datasets [1,2], **the amount of Arabic data is bleak,** while **code-switching is completely ignored!**

# Introducing **ArEnAV**

First **Arabic-English Code-Switching Audio-Visual Deepfake** Dataset

Containing 387k videos, over 765 hours of video content

Contains **Arabic only**, **English-Arabic Code-Switching** and **Diglossic Code-Switching Content** (code-switching between MSA\* and various Arabic Dialects (Egyptian, Levantine, etc))

# Data Generation Pipeline



الطقس ليس سيئاً ("The weather is not bad.")

العيال فازوا ("The boys won.")

نحنصنع hope ("We create hope.")

الطقس سيئاً... الطقس ("The weather is ... bad.")

العيال يلسوا ("The boys sat.")

نحنصنع fun ("We create fun.")

Sync-Net    Yolo-v5-Face    Whisper

8,800 videos  
48,000 clips

### Transcript Manipulation

###System Message###  
You are a controlled text-perturbation bot. Here is the transcript of an audio. Please use the provided operations to modify the transcript to change its sentiment...

###Input###  
"input": {"language": "ar", "change\_mode": "meaning\_only", "tokens": ["", "الطقس", "ليس", "سيئاً", ""]}

###Output###  
"operations": [{"operation": "delete", "old\_word": "ليس", "new\_word": null, "index": 1}, ...  
"operations": [{"operation": "replace", "original\_dialect": "gulf", "new\_dialect": "emirati"}]...  
{ "operation": "replace", "old\_word": "hope", "new\_word": "fun", "index": 2 }

### Fake Clip

Insert    Replace    Delete

Option 1: Diff2Lip    Option 2: Latent-Sync

Ref Frames

Loudness Norm

Full speech fake sentence → Whisper → Find and Crop → Cropped speech fake word

Option 1: Voice Cloning    Option 2: TTS + Voice Conversion

Denoisier → Clean Voice + BG Noise

Fake Transcripts

Video Manipulation

Audio Manipulation

# Social and Scientific Impact

## Social

1. Improves protection for *Arabic-speaking and code-switching populations*
2. Increases visibility and inclusion of non-English languages in AI research
3. Supports safer deployment of media technologies in multilingual regions
4. Encourages broader participation from researchers in underrepresented communities

## Scientific

1. Enables controlled study of how language mixing alters multimodal forgery signals
2. Allows evaluation of model reliance on semantic vs acoustic-visual cues
3. Creates ground truth for failure analysis of AV models under distribution shift
4. Opens new research questions at the intersection of sociolinguistics and AI security



Thank You for watching!



Checkout **ArEnAV**

Dataset Link:



GitHub Link:

