# Audio-Visual Deepfake Datasets

| Dataset Name | Link / Reference | Number of Clips | Type (Real/Fake) | |
|---|---|---|---|---|
| FakeAVCeleb | Google Sites+1 | Info not specified | Real & Fake | Classic audio+video deep |
| AV-Deepfake1M / 1M++ | arXiv+2Hugging Face+2 | Large-scale | Real & Fake | Content-driven A/V manip |
| LAV-DF | GitHub+1 | Info not specified | Real & Fake | Designed for temporal fo |
| AVLips | arXiv+1 | Info not specified | Real & Fake | Lip-sync set spotlighting |
| LRS2 / LRS3 | Oxford Robotics Institute+1 | Large-scale | Real (not deepfakes) | Clean lip-reading corpora |

# Baseline AV Models & Pretrained LLMs

## A. Audio-Visual (AV) Models

| Model Name | Purpose | Pretrained? | Link | Notes |
|---|---|---|---|---|
| SyncNet | Audio-Visual sync scoring | Yes | Paper | Computes lip-sync scores between audio and vide |
| AV-HuBERT | VSR embeddings / lip-reading | Yes | GitHub | Extracts viseme features and transcripts from vide |

*Optional models for reference:* | Wav2Lip | Lip-sync generation baseline | Yes | GitHub | Can generate lip movements from audio; useful for stress-testing pipeline |

## B. Pretrained LLMs

| LLM Name | Purpose | Size | Pretrained? | Link | Notes |
|---|---|---|---|---|---|
| Local 7–8B Instruct | Semantic / prosody scoring | 7–8B | Yes | [Repo/Link] | Runs locally; evaluates transcript |
| LLaMA 7B | Optional semantic scoring | 7B | Yes | Hugging Face | Instruction-tuned; local inference |
| GPT-4 / GPT-3.5 | Optional cloud API scoring | 175B | Yes | OpenAI | Cloud API; can score semantic/pr |