1. ElevenLabs

- **Name of the organization:** ElevenLabs
- **Is it open-sourced:** No
- **The Architecture:** Proprietary neural network architecture
- The data source: Private dataset
- **Is multilingual:** Yes
- **Supported languages:** English, Chinese, Spanish, Hindi, French, German, Japanese, Arabic, Polish, Tagalog, Tamil
- **Related papers:** NA
- **Link to test the Model:** https://elevenlabs.io/

2. MetaVoice

- **Name of the organization:** MetaVoice
- **Is it open-sourced:** Yes
- **The Architecture:** Deep learning models with emphasis on emotional variety. Using a causal GPT to predict the first two hierarchies of EnCodec tokens. Then diffused up to the waveform level, with post-processing applied to clean up the audio.
- **Data source:** Private dataset
- **Multilingual:** Yes
- **Supported languages:** English, French, Spanish, German, Polish
- **Related papers:** NA
- **Repo Link:** https://github.com/metavoiceio/metavoice-src
- **Features:** Prioritizes generating speech that closely resembles natural human speech patterns, including emotional tone and rhythm, Provides the ability to clone existing voices with minimal reference audio (especially for American and British accents) opens up possibilities for creating personalized TTS experiences, Trained on a massive dataset of speech data (100,000 hours) allows MetaVoice-1B to capture a wide range of speech variations and nuances.
- **Drawbacks:** Difficult to assess its accuracy in terms of word-for-word replication. This might be because the model prioritizes naturalness over strict accuracy. The model has a limit on the amount of text it can process at once (around 2048 tokens). This might necessitate splitting longer texts into smaller chunks for processing.
- **Others:** Details about the dataset format, instruction about fine tuning the model along with parameters is given on the repo but not tested.

3. OpenVoice V2

- **Name of the organization:** MyShell AI
- **Open-Sourced:** Yes
- **Architecture:** Transformer-based architecture

- **Data Source:** LJSpeech dataset and custom datasets for voice cloning
- **Multilingual:** Limited (English as primary, zero-shot cloning for some languages)
- **Supported languages:** British English, American English, Indian English, Australian English, Spanish, French, Chinese, Japanese, Korean
- **Related papers:** https://arxiv.org/abs/2204.01134
- **Github Repo: myshell-ai/OpenVoice: Instant voice cloning by MyShell. (github.com)**
- **HuggingFace:** https://huggingface.co/myshell-ai/OpenVoiceV2#quick-use
- **Features:**Emotional Nuance Control, Free for commercial use,
- **Drawbacks:** Limited Naturalness, Limited Language Support,

4. WhisperSpeech

- **Name of the organization:** Coqui AI (formerly ML Kit)
- **Open-Sourced:** Yes
- **Architecture:** The general architecture is similar to AudioLM, SPEAR TTS from Google and MusicGen from Meta. We avoided the NIH syndrome and built it on top of powerful Open Source models: Whisper from OpenAI to generate semantic tokens and perform transcription, EnCodec from Meta for acoustic modeling and Vocos from Character Inc as the high-quality vocoder.
- **Data Source:** LJSpeech dataset, English LibreLight database
- **Multilingual:** yes
- **Supported languages:** English like languages
- **Related papers:**NA
- **Github Repo:** https://github.com/collabora/WhisperSpeech
- **HuggingFace:** https://huggingface.co/spaces/collabora/WhisperSpeech
- **Features:** Voice Cloning,

5. Play.HT 2.0

- **Name of the organization:** Play.ht
- **Is it open-sourced:** No
- **The Architecture:** deep learning architecture
- **The data source:** Private dataset
- **Is multilingual:** Yes
- **Supported languages:** English
- **Related papers:** NA
- **Github Repo: playht/pyht: PlayHT Python SDK -- Text-to-Speech Audio Streaming (github.com)**
- **Link to test the Model:** (https://play.ht/)

6. StyleTTS 2

- **Name of the organization:** NVIDIA Research
- **Open-Sourced:** Yes
- **Architecture:** Mel-spectrogram inversion using WaveNet vocoder
- **Data Source:** LJSpeech dataset
- **Multilingual:** Limited (primarily English)
- **Supported languages:** English (primarily)
- **Related papers:** https://arxiv.org/abs/2306.07691
- **Github Repo:** https://github.com/yl4579/StyleTTS2
- **Features:** Have potential to produce human level speech quality, Optimized for fast inferencing, Can adapt to a new speakers voice with minimal data(at least 30 minute of audio)

7. XTTSv2

- **Name of the organization:** Baidu AI (Research project)
- **Is it open-sourced:** Yes
- **The Architecture:** Transformer-based architecture with WaveNet vocoder
- **The data source:** LJSpeech dataset
- **Is multilingual:** Limited (primarily English)
- **Supported languages:** English (primarily)
- Related papers::https://arxiv.org/abs/2406.04904
- **Github Repo:** **coqui-ai/TTS: 🐸💬 - a deep learning toolkit for Text-to-Speech, battle-tested in research and production (github.com)**
- **Link to test the Model:** https://huggingface.co/coqui/XTTS-v2

8. MeloTTS

- Name of the organization: Ruochen Wang et al. (Research project)
- Is it open-sourced: Yes
- The Architecture: Transformer-based architecture with WaveNet vocoder
- The data source: LJSpeech dataset
- Is multilingual: Yes
- Supported languages: English (American), English (British), English (Indian), English (Australian), Spanish, French, Chinese (mix EN), Japanese, Korean
- Related papers:NA
- Link to test the Model: https://github.com/myshell-ai/MeloTTS

9. GPT-SoVITS

- Name of the organization: NVIDIA (Research project)
- Is it open-sourced: Yes
- The Architecture: Combines Generative Pre-training Transformer (GPT) with Spectrogram Vocoder using Information Theoretic Loss (SoVITS)

- The data source: LJSpeech dataset
- Is multilingual: Yes
- Supported languages: English, Chinese, Japanese
- Related papers: NA
- Link to test the Model: https://github.com/RVC-Boss/GPT-SoVITS , https://huggingface.co/lj1995/GPT-SoVITS

10. Parler TTS

- Name of the organization: DeepMind (Research project)
- Is it open-sourced: Yes
- The Architecture: Transformer-based architecture with WaveNet vocoder
- The data source: LJSpeech dataset
- Is multilingual: No
- Supported languages: English in various accents
- Related papers: https://www.text-description-to-speech.com/
- Link to test the Model: https://huggingface.co/parler-tts

11. Vokan TTS

- Name of the organization: Resemble AI
- Is it open-sourced: No
- The Architecture: Proprietary deep learning architecture
- The data source: Private dataset
- Is multilingual: Yes
- Supported languages: English, Hindi, Chinese, Korean, Japanese, Polish, German, Spanish, Russian, Romanian, more european languages
- Related papers: NA
- Link to test the Model:https://huggingface.co/ShoukanLabs/Vokan

12. VoiceCraft 2

- Name of the organization: VoiceCraft
- Is it open-sourced: Yes
- The Architecture: token infilling neural codec language model
- The data source: Private dataset
- Is multilingual: Yes
- Supported languages: English, and european languages
- Related papers: https://jasonppy.github.io/assets/pdfs/VoiceCraft.pdf
- Link to test the Model https://github.com/jasonppy/VoiceCraft

13. Pheme

- Name of the organization: PolyAI
- Is it open-sourced: Yes
- The Architecture:Neural seq to seq
- The data source: Private dataset
- Is multilingual: Yes
- Supported languages: Not mentioned
- Related papers: https://arxiv.org/pdf/2401.02839
- Link to test the Model https://github.com/PolyAI-LDN/pheme