Continue

========== Fish Audio is a leading generative AI platform that specializes in ultra-low latency, highly customizable text-to-speech (TTS) and realistic voice cloning services. This innovative platform empowers users to convert text into natural, fluent speech and create accurate voice replicas from short audio samples. With its robust language support, Fish Audio caters to diverse applications such as content creation, marketing, accessibility, and sound design. The key features of Fish Audio include: - **Text-to-Speech (TTS)**: This feature converts written text into natural-sounding speech with support for various languages and voice styles. Users can generate and download audio clips quickly with minimal latency. - **Voice Cloning**: This feature creates highly realistic voice clones from brief audio samples, allowing users to build personalized voice avatars and customize speech output. - **Speech-to-Text (STT)**: This feature provides accurate transcription capabilities supporting multiple languages, enhancing usability for audio-to-text workflows. - **API Access**: Fish Audio offers developers integration options to embed its voice synthesis and cloning capabilities into their own applications. Fish Audio boasts an intuitive dashboard with sections for discovering new voices, generating text-to-speech, building voice avatars, and accessing premium plans. The platform also offers a free tier with generous usage limits and competitive premium plans, making advanced AI voice technology accessible to both beginners and professionals. The cutting-edge technology utilized by Fish Audio leverages large language models and advanced vocoder architectures for high-fidelity, multilingual speech synthesis with ultra-low latency (~150ms). Fish Audio stands out as a top-tier AI voice platform in 2025, offering realistic voice cloning and fast, high-quality text-to-speech services at competitive pricing, rivaling major players like 11 Labs.Powered by advanced deep learning technology, our AI voiceover and text-to-speech platform generates natural, fluent voices for various use cases. With a simple process that requires no professional knowledge, users can clone a voice in just a few steps: upload clear voice samples, create an AI voice model, wait for training, and input text to generate the cloned voice. Our platform supports all major audio formats (MP3, WAV, M4A, etc.) and processes text-to-speech in 40+ languages. We ensure optimal audio quality with professional features like noise reduction, volume equalization, and audio enhancement. Our multilingual AI voiceover capabilities are backed by our advanced technology. Trained on extensive data and equipped with professional audio processing workflows, our AI models provide high-quality voices that are almost indistinguishable from human voices. We offer various voice tones and emotional options to ensure professional quality. Our platform is widely used in video voiceovers, audiobook production, educational courses, podcast creation, game voicing, and more. Our text-to-speech technology ensures optimal performance across all use cases. With 99% voice accuracy, our generated AI voices are natural and fluent with rich emotional expression. We also provide batch processing for improved efficiency and support cross-language content creation. Whether you're a creator or an enterprise user, our platform offers the best text-to-speech technology available. Our services include AI voiceover, voice cloning, and audio processing. With our advanced technology and professional features, we guarantee optimal results every time.OpenAudio S1 alcança 0.008 WER e 0.004 CER em texto inglês, que é significativamente melhor que modelos anteriores. --- O OpenAudio S1 demonstrou ser o modelo mais eficaz para avaliação de text-to-speech, com uma taxa de erro de palavras (WER) de 0.008 e uma taxa de erro de caracteres (CER) de 0.004 em texto inglês. Isso supera significativamente os resultados dos modelos anteriores. --- O modelo OpenAudio S1 alcançou a classificação #1 no TTS-Arena2, um benchmark para avaliação de text-to-speech, e demonstrou ser capaz de controlar uma variedade de marcadores emocionais, de tom e especiais para aprimorar a síntese de fala. --- Além disso, o OpenAudio S1 suporta inserção de amostras vocais de 10 a 30 segundos para gerar saída TTS de alta qualidade. O modelo também é capaz de lidar com texto em qualquer script de idioma e alcança uma precisão alta, com uma taxa de erro de caracteres (CER) de cerca de 0,4% e uma taxa de erro de palavras (WER) de cerca de 0,8% para Seed-TTS Eval. --- O OpenAudio S1 é disponível em duas variantes: o modelo principal com todos os recursos disponíveis e uma versão mais compacta, chamada OpenAudio S1-mini, que é otimizada para inferência mais rápida mantendo excelente qualidade. Ambos os modelos incorporam Aprendizado por Reforço Online com Feedback Humano (RLHF) e oferecem suporte multilíngue e cross-lingual. --- O modelo também é capaz de lidar com texto em qualquer script de idioma sem dependência de fonemas e alcança uma taxa de erro de caracteres (CER) baixa. Além disso, o OpenAudio S1 é rápido, com um fator de tempo real aproximadamente 1:5 em um laptop Nvidia RTX 4060 e 1:15 em um Nvidia RTX 4090. --- O modelo oferece uma interface web fácil de usar baseada em Gradio e uma interface gráfica PyQt6 que funciona perfeitamente com o servidor API. Ele também é amigável para deploy, permitindo a configuração facilmente do servidor de inferência com suporte nativo para Linux, Windows (MacOS em breve). --- O OpenAudio S1 é compatível com web, Android, iOS, macOS, Windows e Linux e pode ser encontrado no GitHub ou na PyPI.