

맞다 AI가

듣고 말하는 AI의 완성 VibeVoice 활용 가이드

이동수

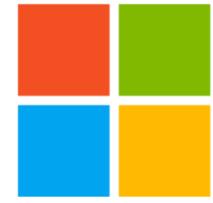


@DS-J-L



@Dongsu Lee



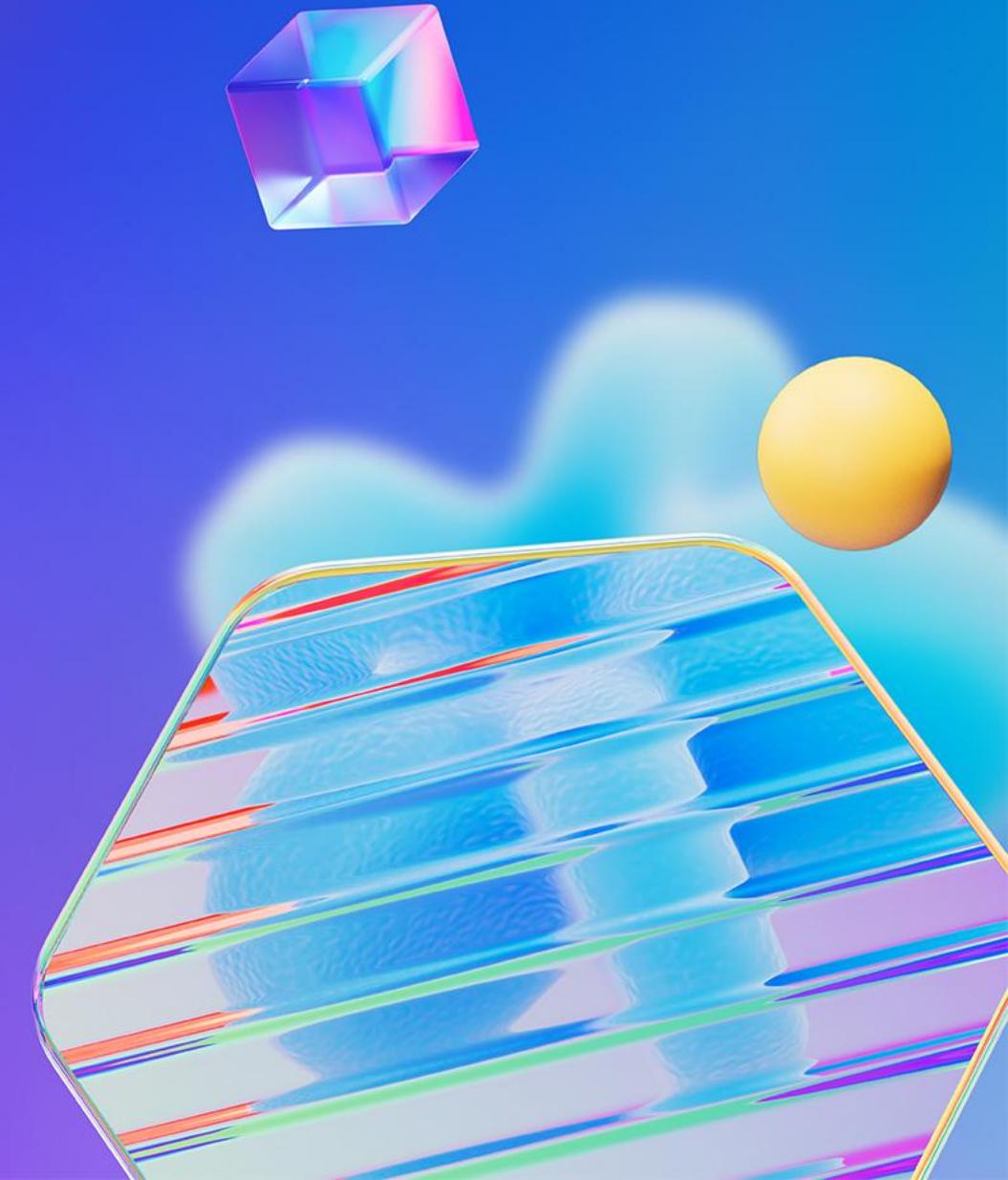


Microsoft

The OceanLightAI logo icon features a stylized sun or moon rising over water, represented by a yellow circle with white and blue wavy lines below it.

oceanLightAI

VibeVoice가 무엇인가요?





ASR

Realtime TTS

TTS



ASR

Realtime TTS

TTS



ASR

Realtime TTS

TTS



ASR

Realtime TTS

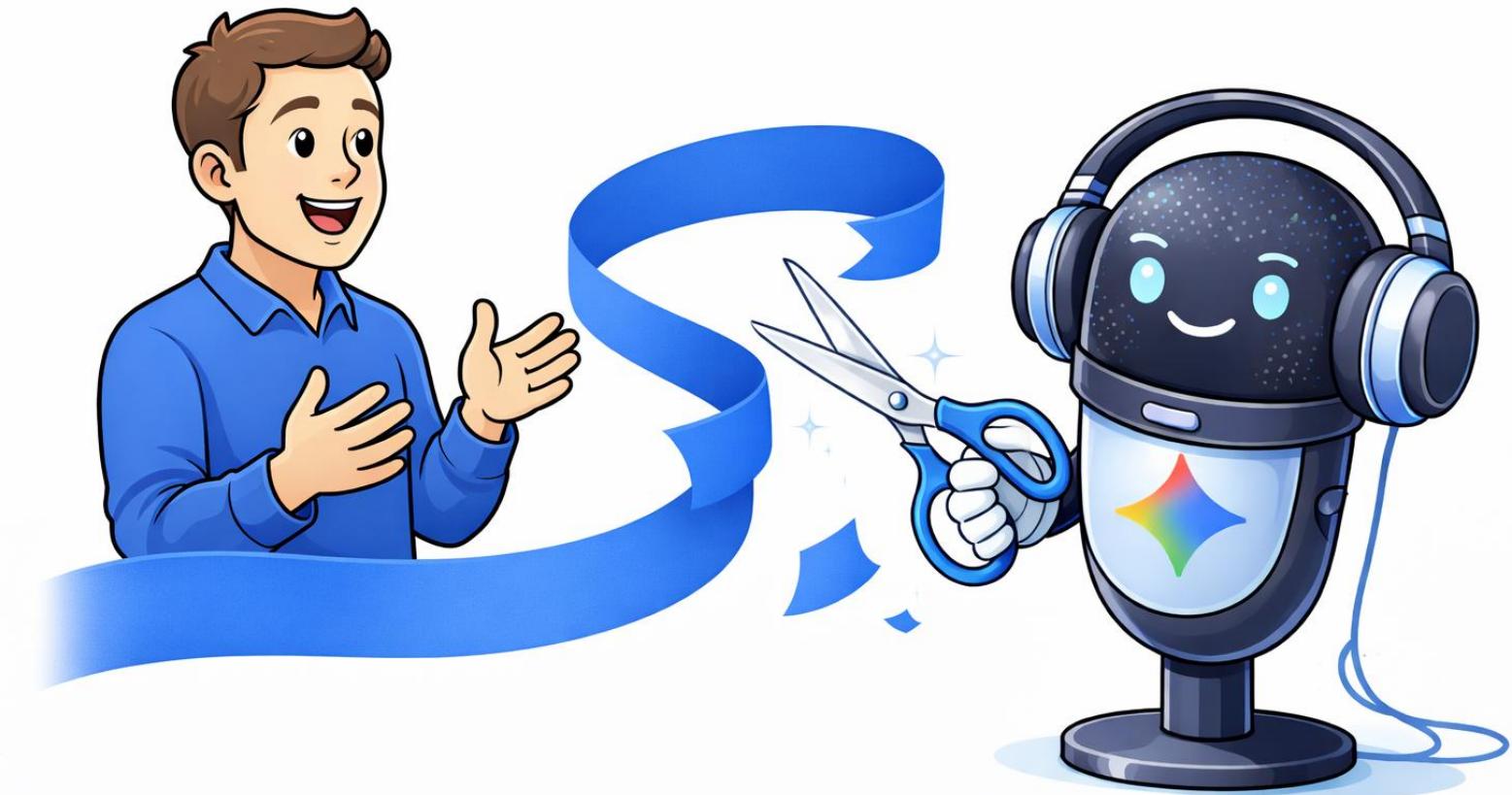
TTS

VibeVoice ASR





Gemini Pro



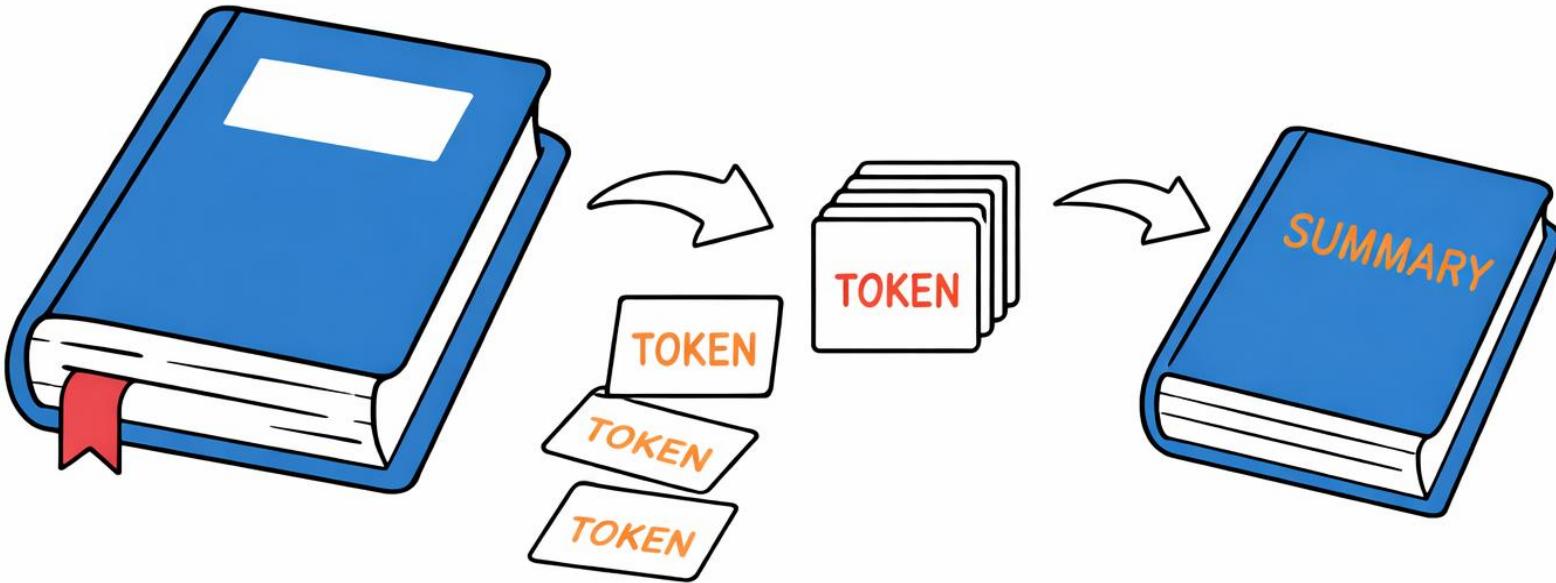
음성이 너무 길면 30초 씩 자른다



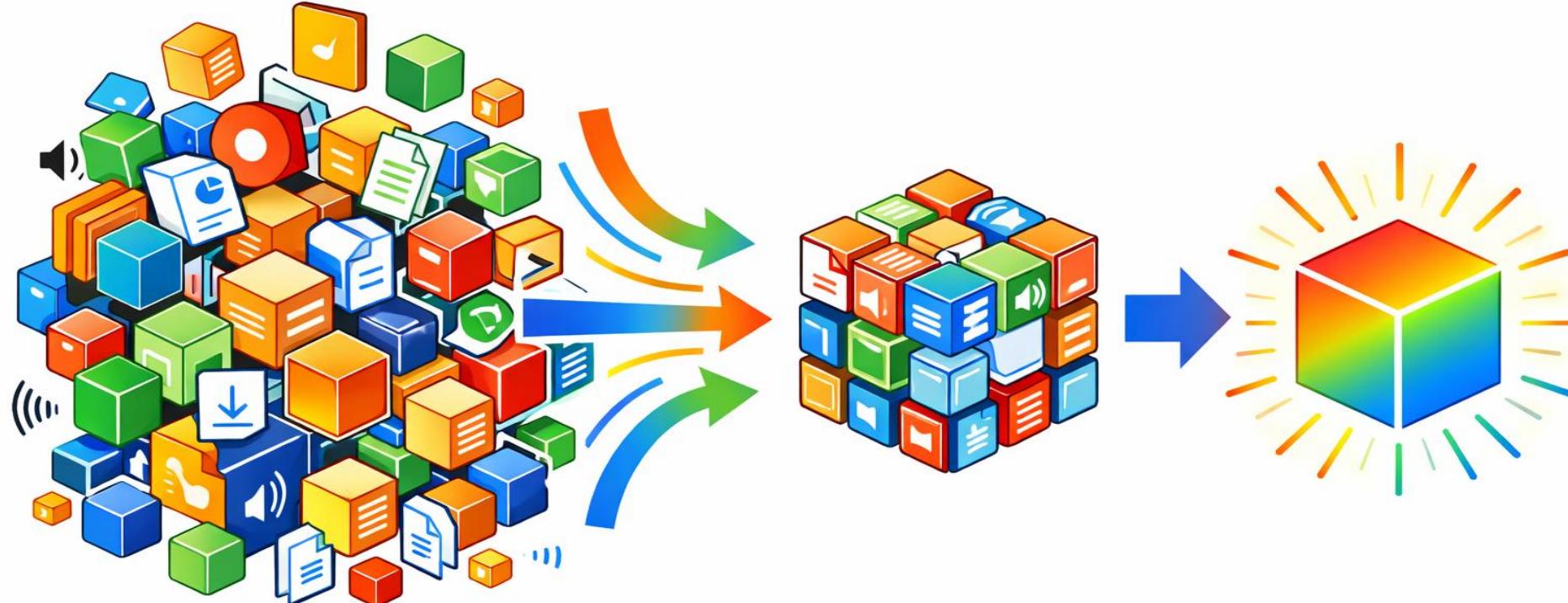
30초 씩 자르면 이전 내용을 기억 못한다



자르지 않고 한 시간 분량을 들을 수 있다



원본 데이터를 토큰으로 바꿔 요약본을 만든다



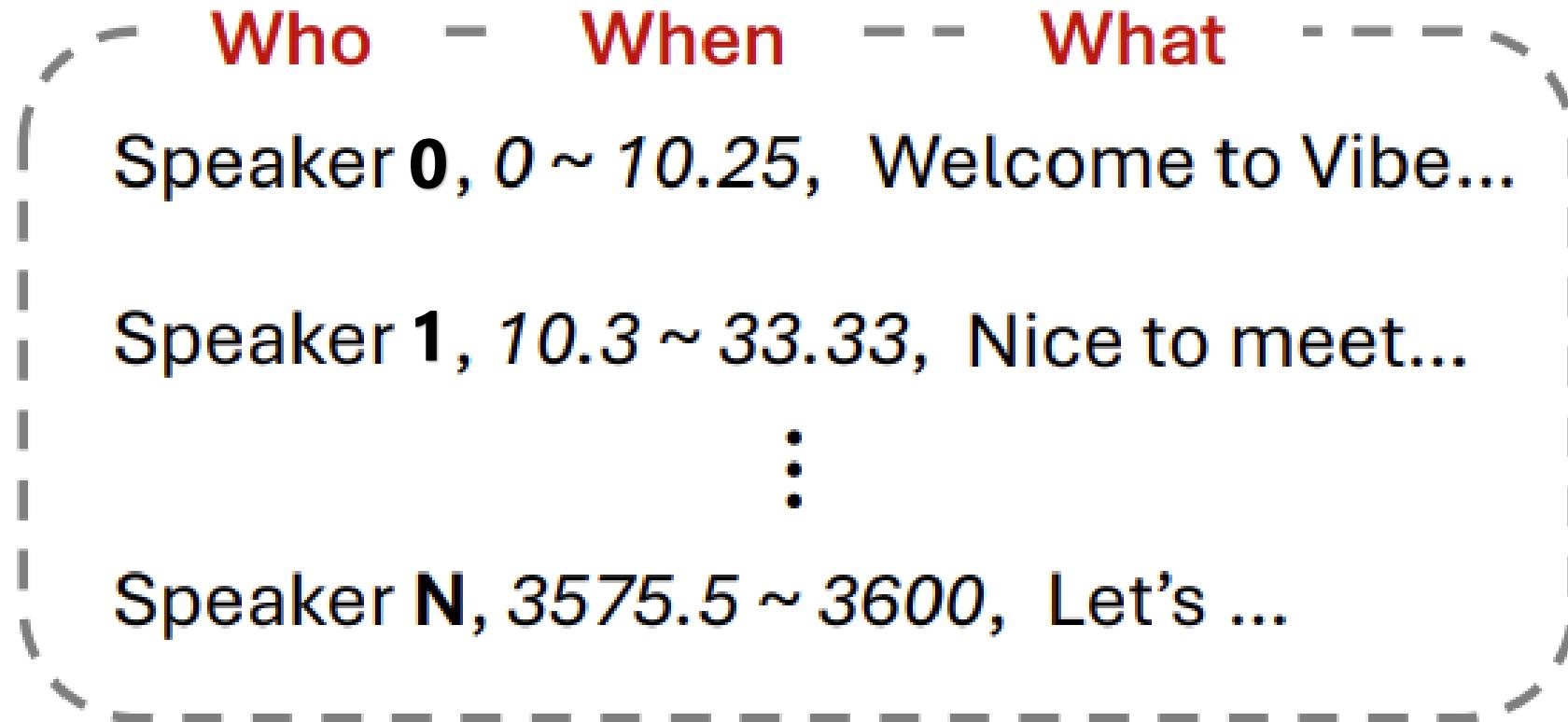
데이터 24,000개 → 압축 3200배 → 7.5 토론

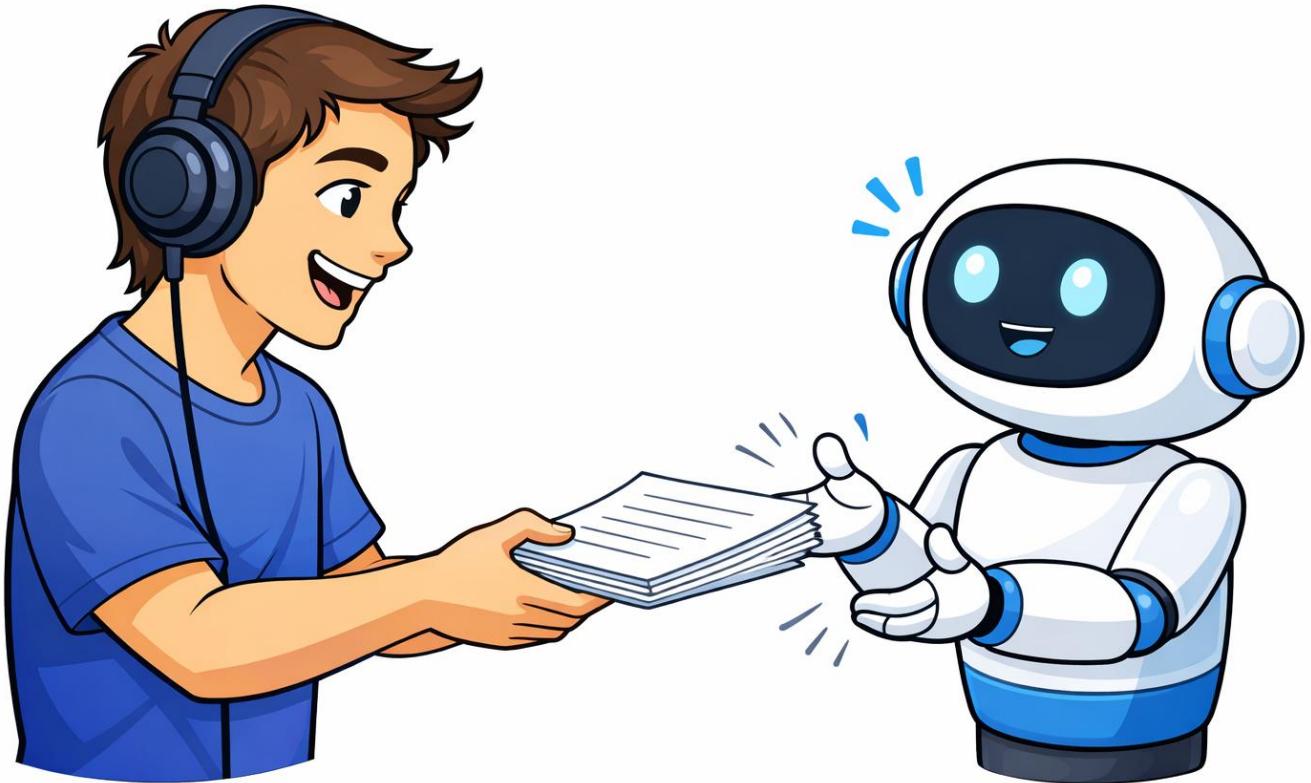
모델명	입력 데이터	초당 생성 데이터	압축률
Whisper (OpenAI)	16,000 Hz	50 per sec	320
Wav2Vec 2.0 (Meta)	16,000 Hz	50 per sec	320
Gemini 2.5/3.0Pro (Google)	24,000 Hz	32 per sec	750
VibeVoice (Microsoft)	24,000 Hz	7.5 per sec	3,200

모델명	입력 데이터 16,000 Hz	초당 생성 데이터 50 units/sec	압축률
Whisper (OpenAI)	$50 \times 3,600 = 180,000$	50 units/sec	320
Wav2Vec 2.0 (Meta)	$7.5 \times 3600 = 27,000$	50 units/sec	320
Gemini 2.5/3.0Pro (Google)	24,000 Hz	32 tokens/sec	750
VibeVoice (Microsoft)	24,000 Hz	7.5 tokens/sec	3,200

누가? 언제? 어떤 내용을 말했는지 정확히 기록

Rich Transcription

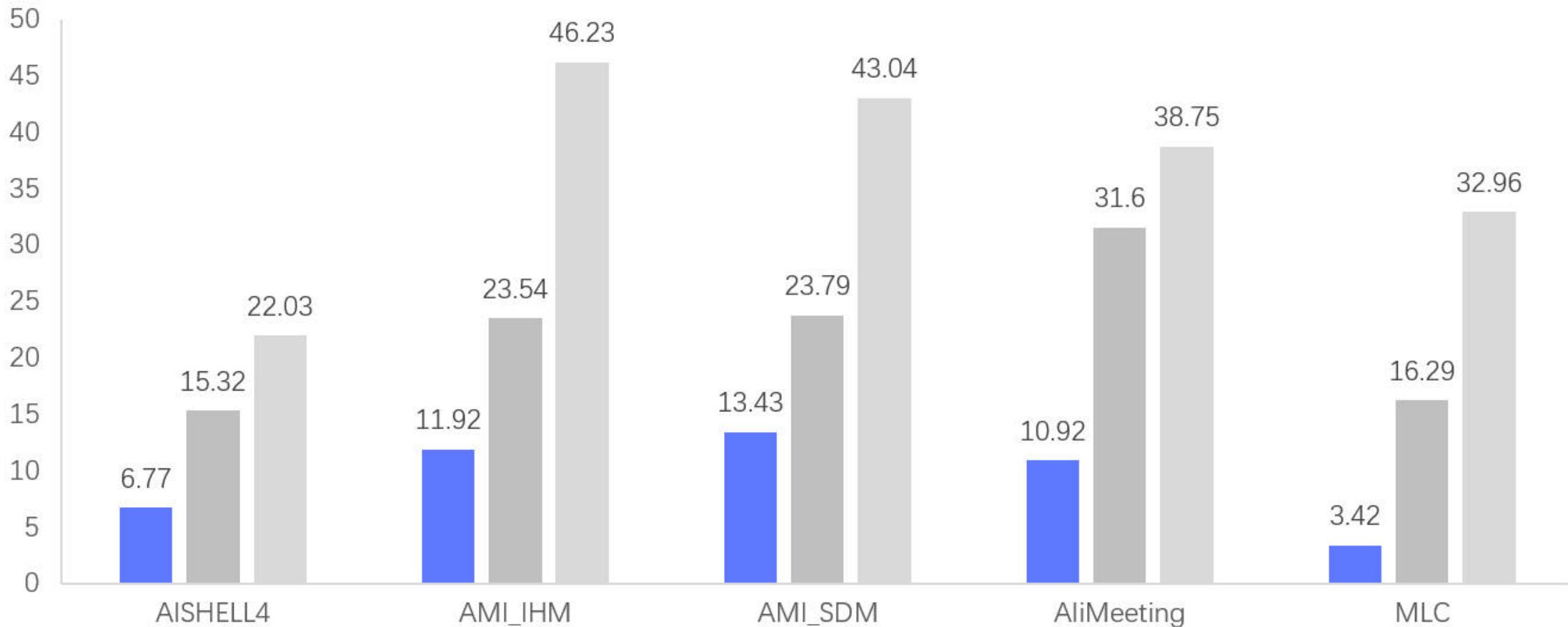




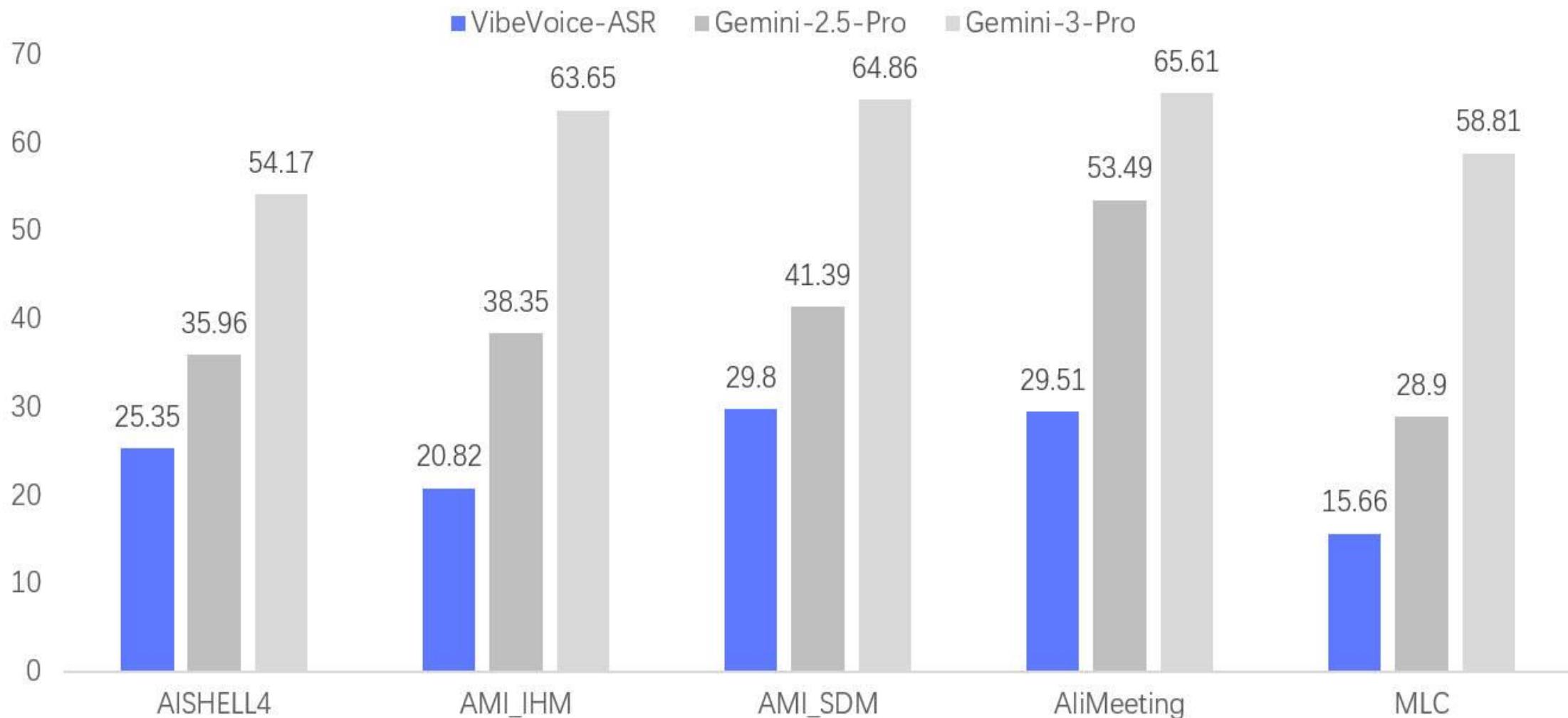
컨텍스트를 주입해줄 수 있다

DER (Diarization Error Rate) ↓

■ VibeVoice-ASR ■ Gemini-2.5-Pro ■ Gemini-3-Pro



tcpWER (Time-Constrained Permutation WER)↓



VibeVoice ASR

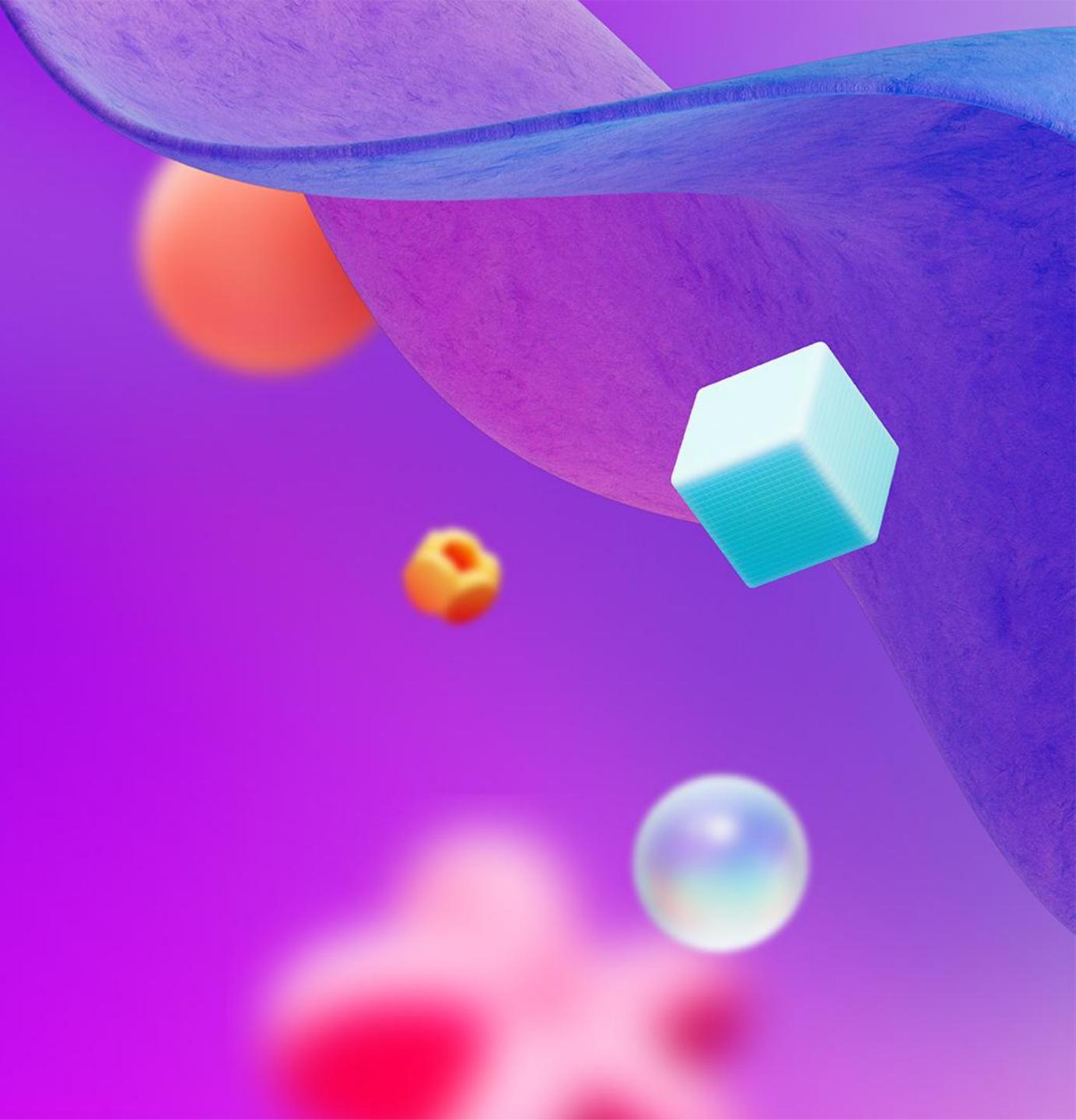
Gemini 3.0

Gemini 2.5

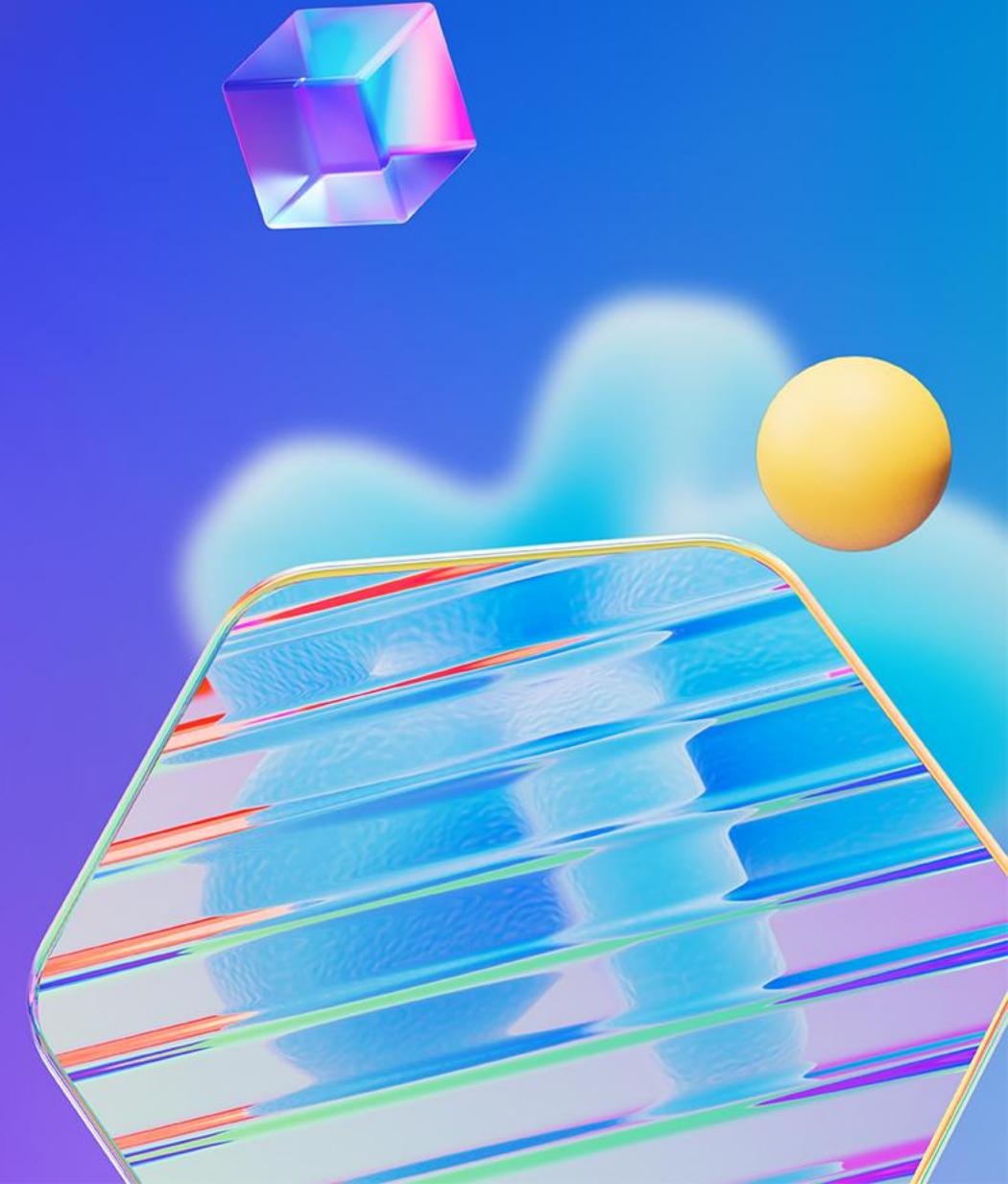


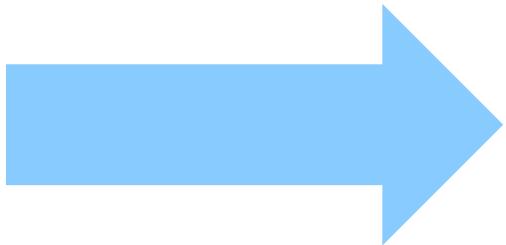
64가지의 테스트 항목 중 13개 제외 모두 1위!

VibeVoice ASR: Demo

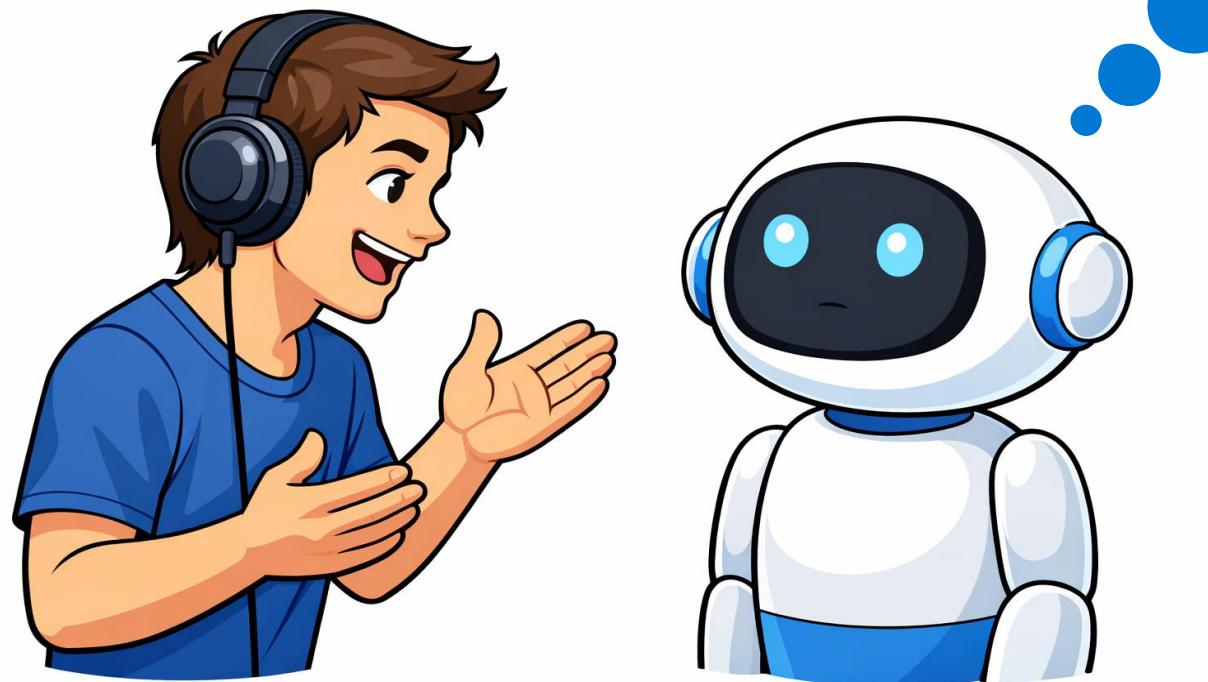


VibeVoice Realtime TTS

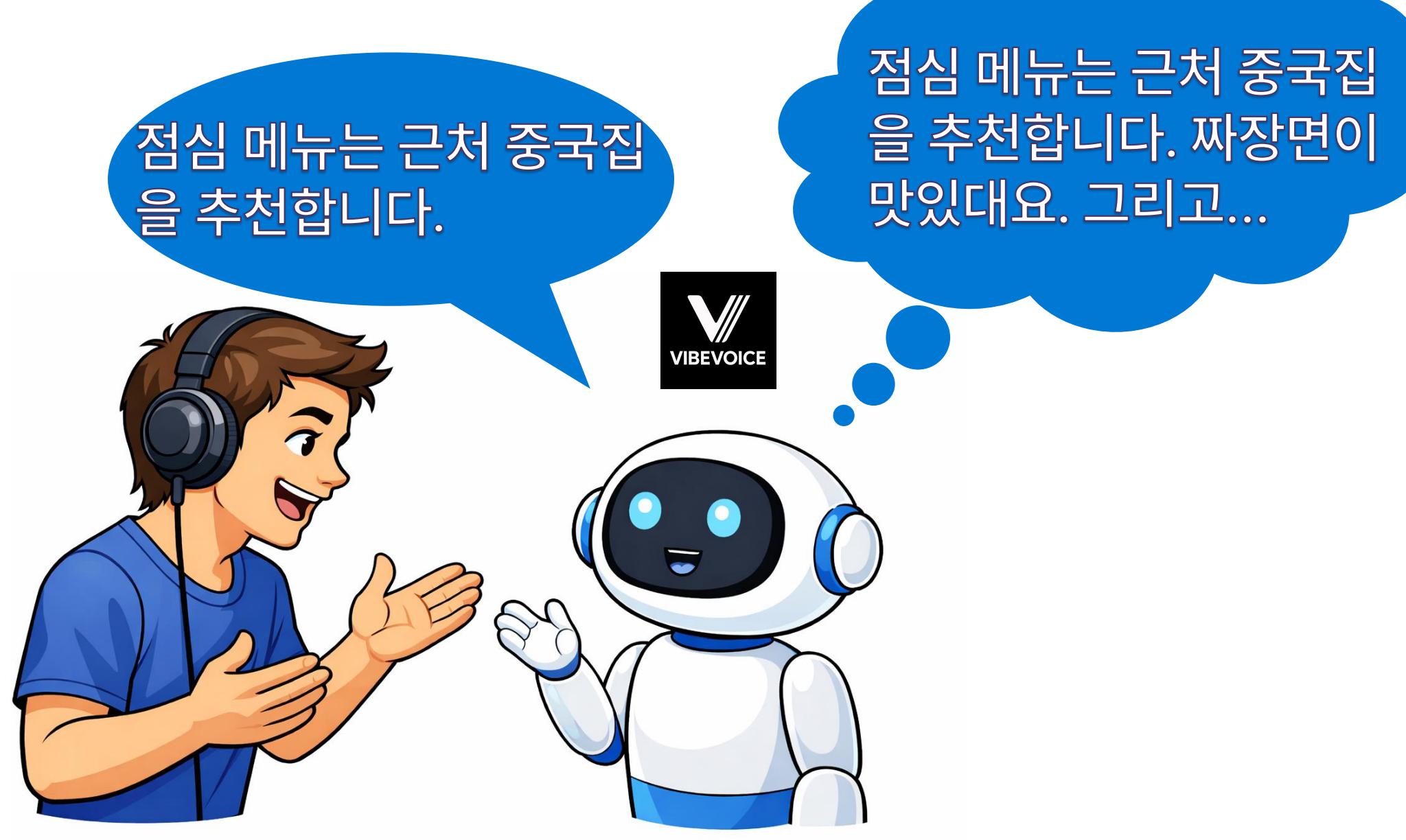




텍스트를 끝까지 생성 이후 음성 생성



점심 메뉴는 근처 중국집
을 추천합니다. 짜장면이
맛있대요. 그리고...



점심 메뉴는 근처 중국집
을 추천합니다.

점심 메뉴는 근처 중국집
을 추천합니다. 짜장면이
맛있대요. 그리고...



첫 음성 출력까지 2~3초에서 0.3초로 단축

기존 다른 모델

텍스트를 생성하는 시간

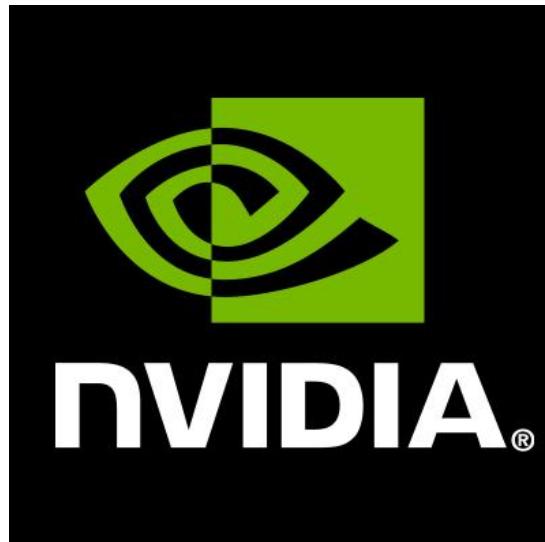
음성을 생성하는 시간



텍스트를 생성하는 시간

음성을 생성하는 시간

Interleaved Diffusion



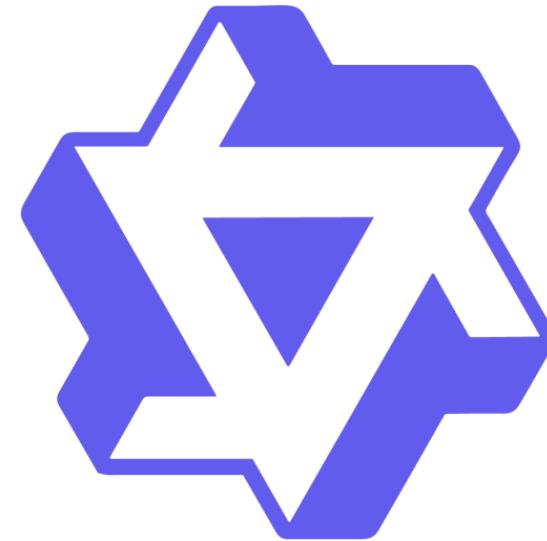
7B

>

0.5B

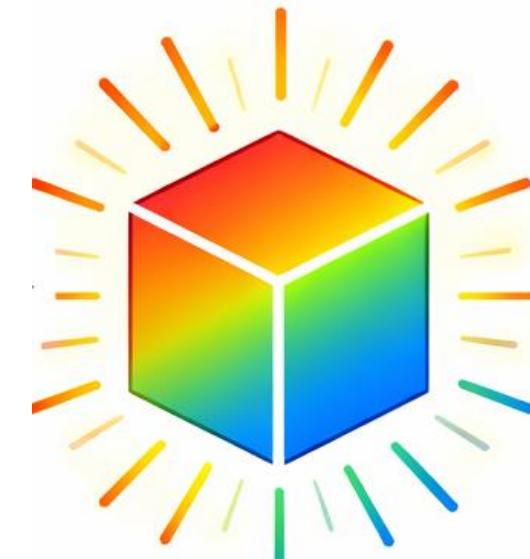
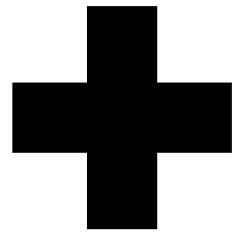
<

7B





0.5B



7.5 토큰

VibeVoice Realtime TTS : Demo

VibeVoice TTS





문장을 자연스럽게 읽고 흉내는 가능!



유튜브 @NadimExplainsAI

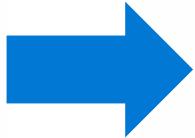
바이브 보이스 TTS로 생성한 대화



짧은 샘플로도 목소리 복제 가능



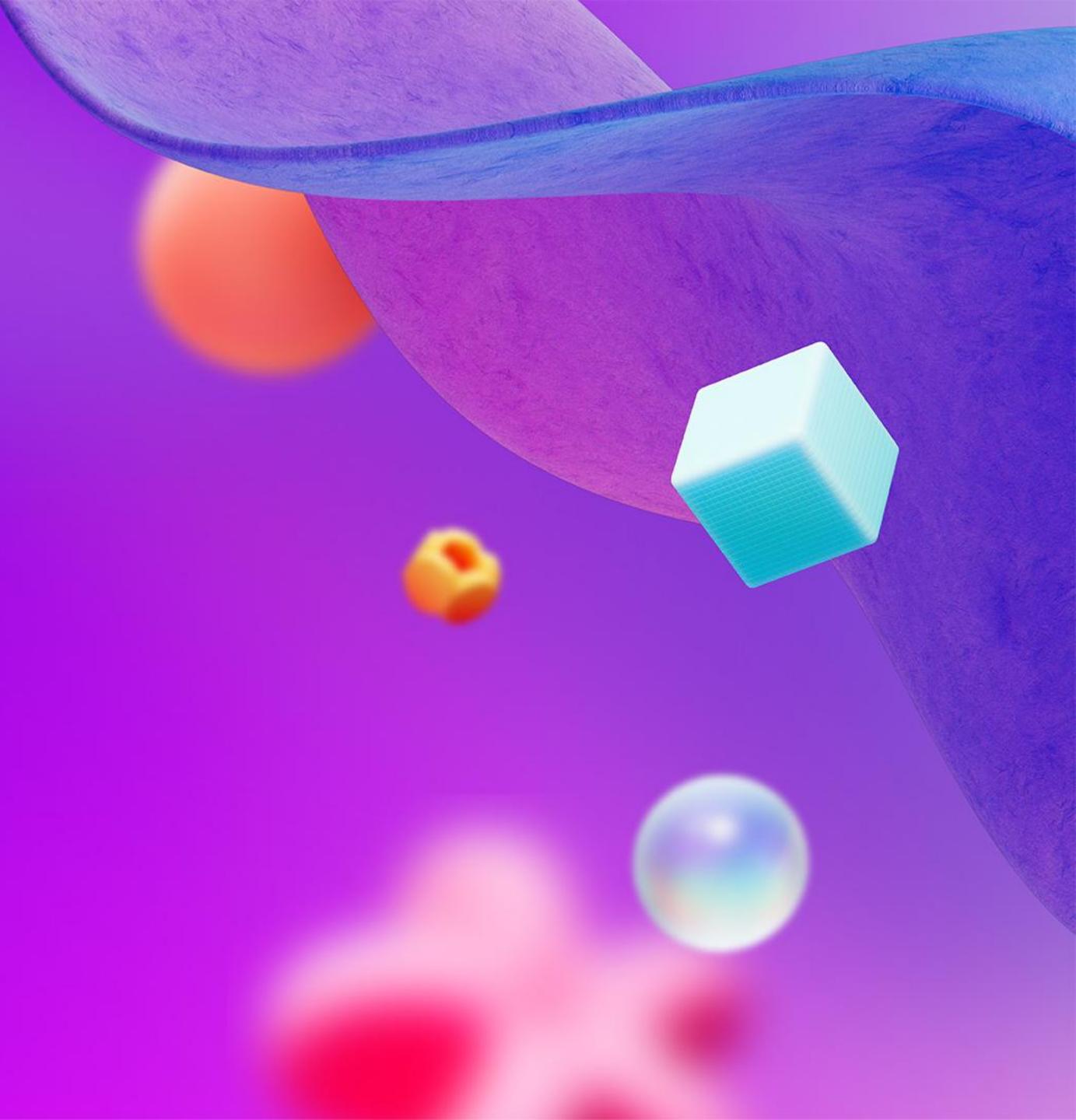
혼자 말하는 것이 아닌 다화자 대화



각자 대본을 읽는 것이 아닌 실제 대화 같은 느낌



VibeVoice TTS





보이스피싱



딥페이크



가짜 뉴스

책임감 있는 개발자가 되려면?

<https://bit.ly/ms-responsible>



VibeVoice 실제로 사용해보고 싶다면?

<https://bit.ly/vibevoice>



맞다 AI가

감사합니다

듣고 말하는 AI의 완성
VibeVoice 활용가이드

경북대학교 이동수

