

# Human Presentator AI-Presentator with Zonos und Memo

By Markus Siegert and Lucas Buchholz

# Contents

1. Realization

2. Introduction

3. Architecture

4. VRAM Usage

5. Zyphra/Zonos

6. Memoavatar Memo

7. Live Demo

8. Discussion

9. End

# Realization

Unlimited FREE Image to Video Generator | Roboneo AI Tutorial (No Watermark, Consistent Chara...



**Realization** → Introduction → Architecture → VRAM Usage

# Introduction

Ideation

Initial concept development and brainstorming phase.

Monolithic implementation

All-in-one application structure with tightly coupled components.

API based implementation

Decoupled architecture with REST APIs for data exchange.

Frontend implementation

Modern reactive frontend consuming API services.



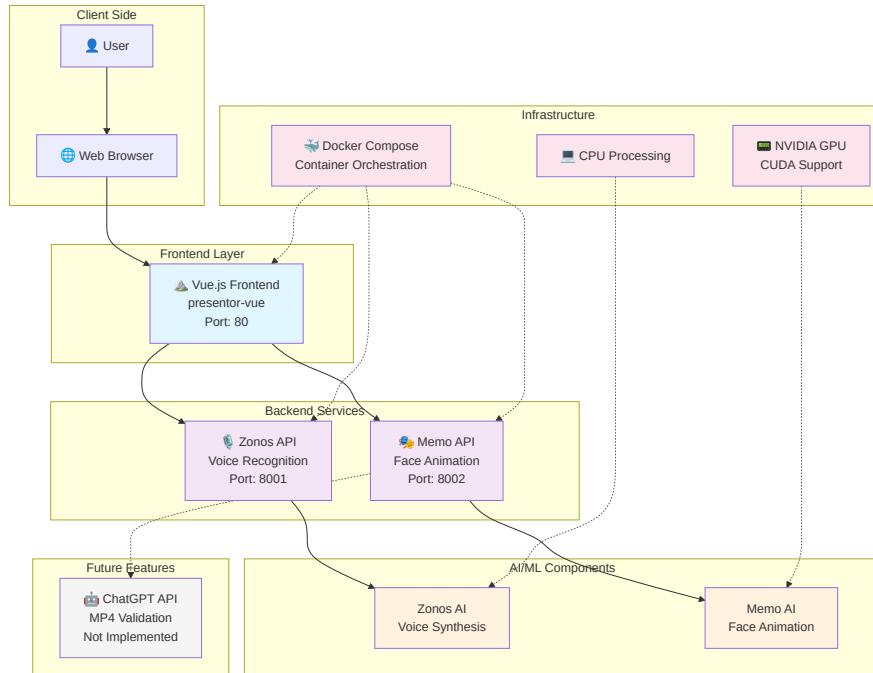
**Planning & Design**

**Single Codebase**

**Backend Services**

**User Interface**

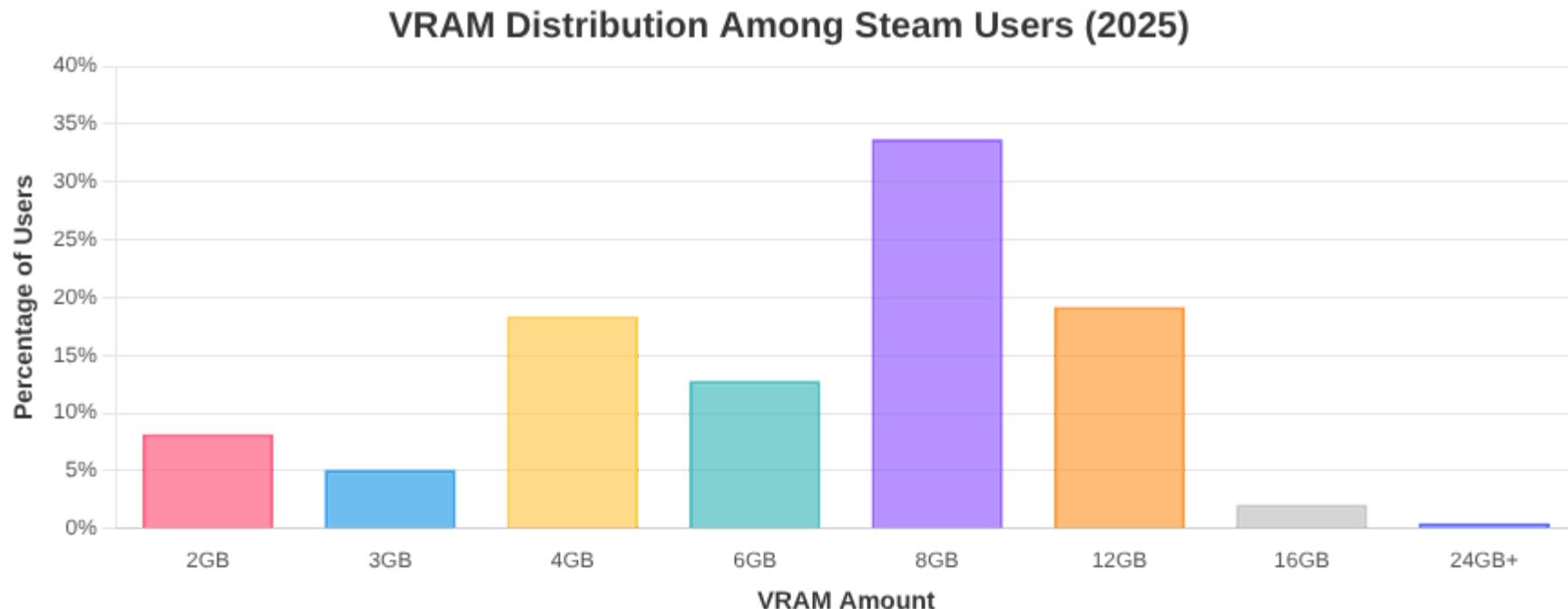
# Architecture



Architecture → VRAM Usage → Zyphra/Zonos → Memoavatar Memo

# VRAM Usage

Based on Steam Hardware Survey 2025



VRAM Usage → Zephyra/Zonos → Memoavatar Memo → Live Demo

# Zyphra/Zonos

- Open-weight TTS model
- Phoneme, Pitch and Speaking Rate are adjustable
- Voice cloning from 30s–5min voice samples

```
model = Zonos.from_pretrained("Zyphra/Zonos-v0.1-transformer", device=device.ty

wav, sampling_rate = torchaudio.load(temp_input_path)
speaker = model.make_speaker_embedding(wav, sampling_rate)

cond_dict = make_cond_dict(text=text, speaker=speaker, language=language)
conditioning = model.prepare_conditioning(cond_dict)

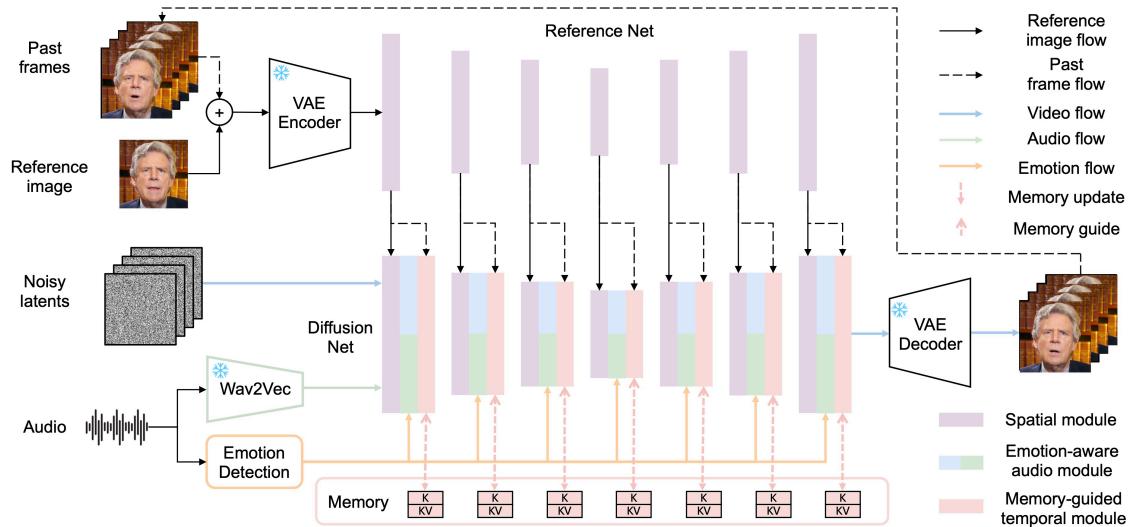
codes = model.generate(conditioning)

wavs = model.autoencoder.decode(codes).cpu()
torchaudio.save(output_file_path, wavs[0], model.autoencoder.sampling_rate)
```



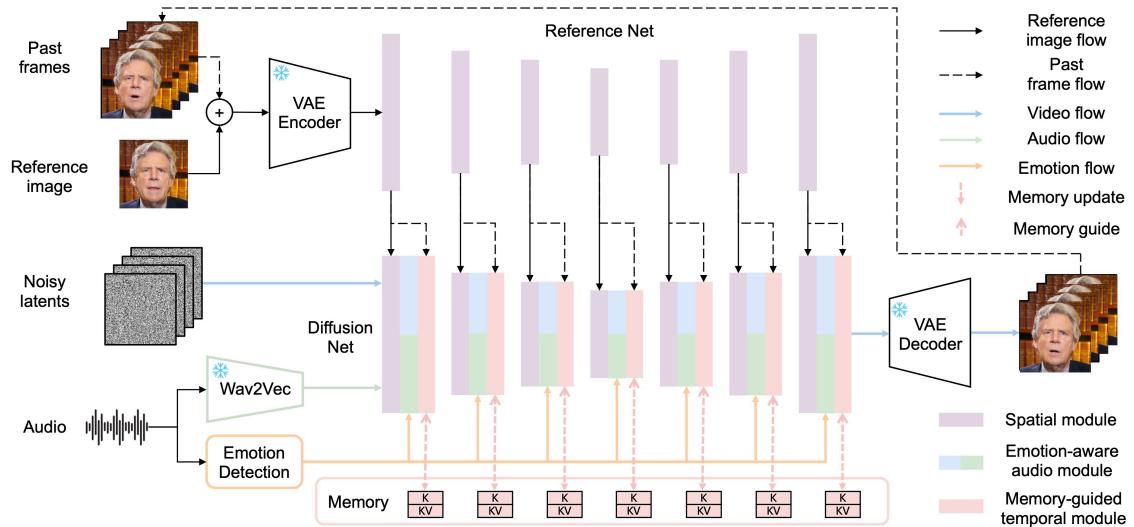
[Zyphra/Zonos](#) → Memoavatar Memo → Live Demo → Discussion

# Memoavatar Memo



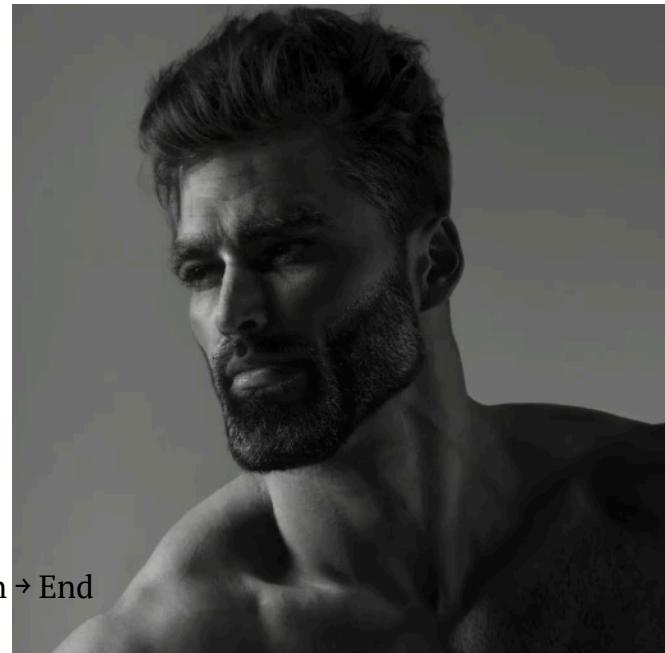
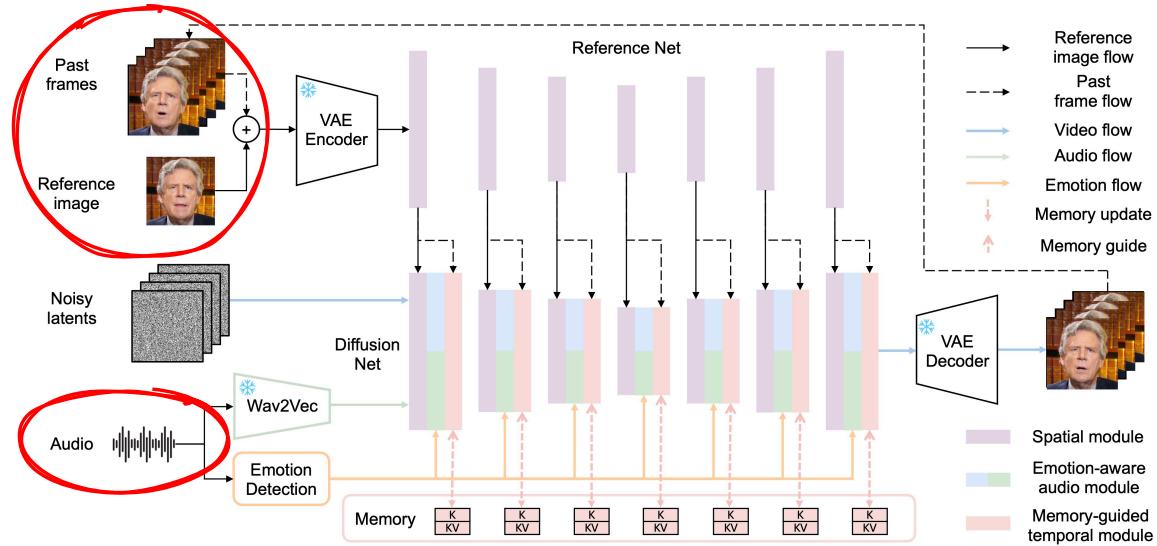
**Memoavatar Memo → Live Demo → Discussion → End**

# Memoavatar Memo



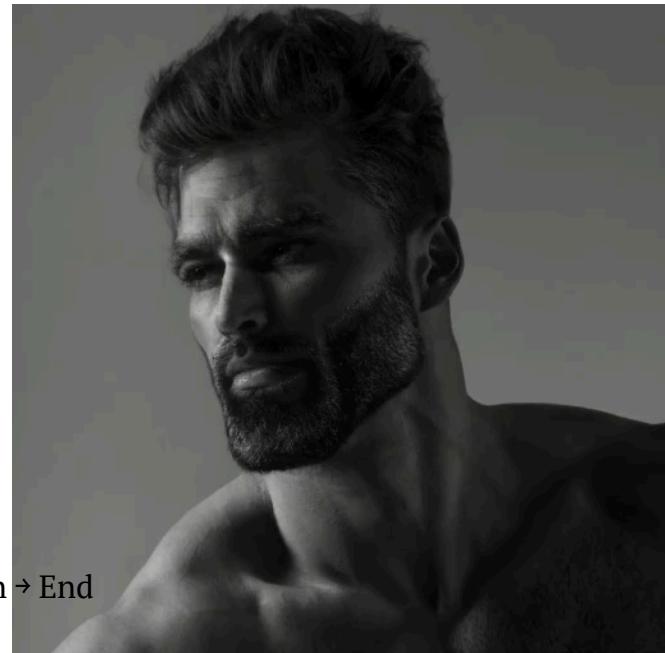
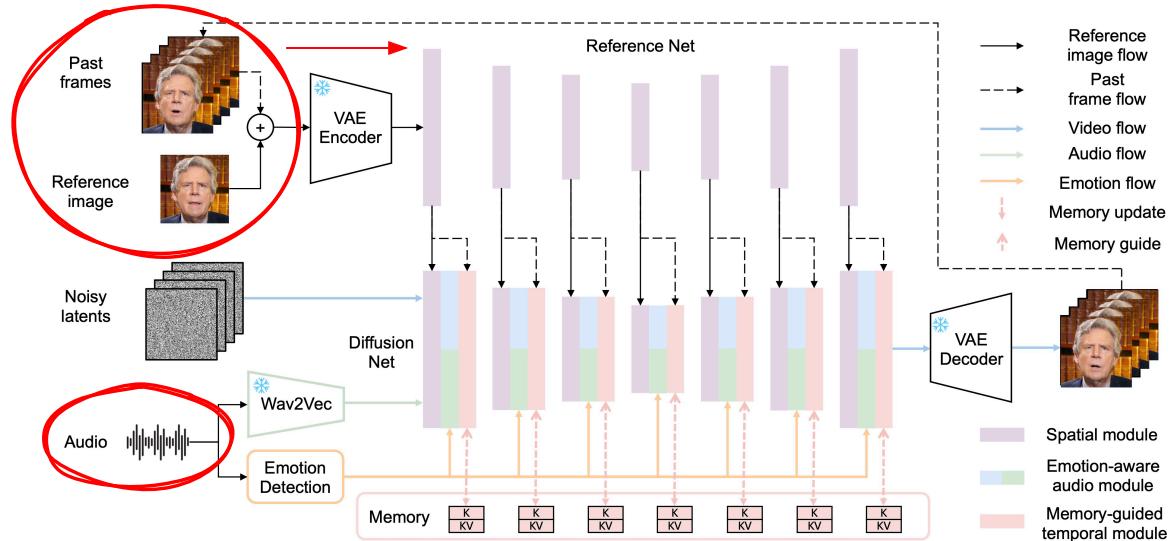
**Memoavatar Memo → Live Demo → Discussion → End**

# Memoavatar Memo



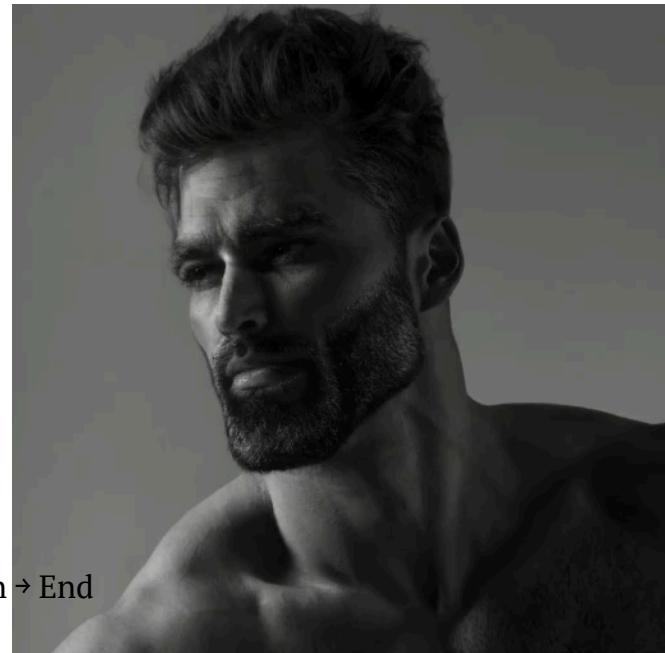
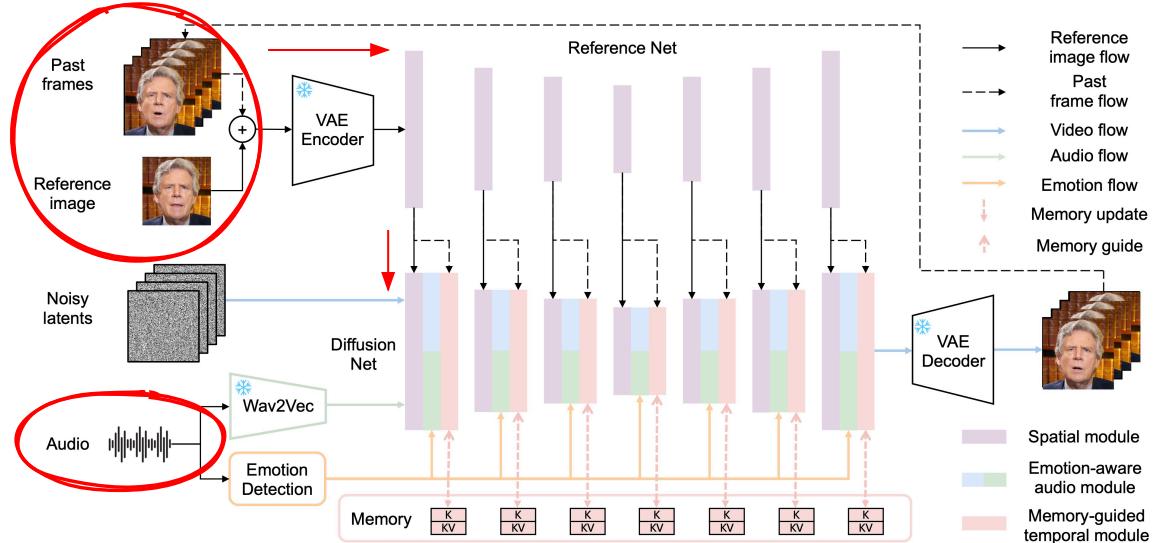
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



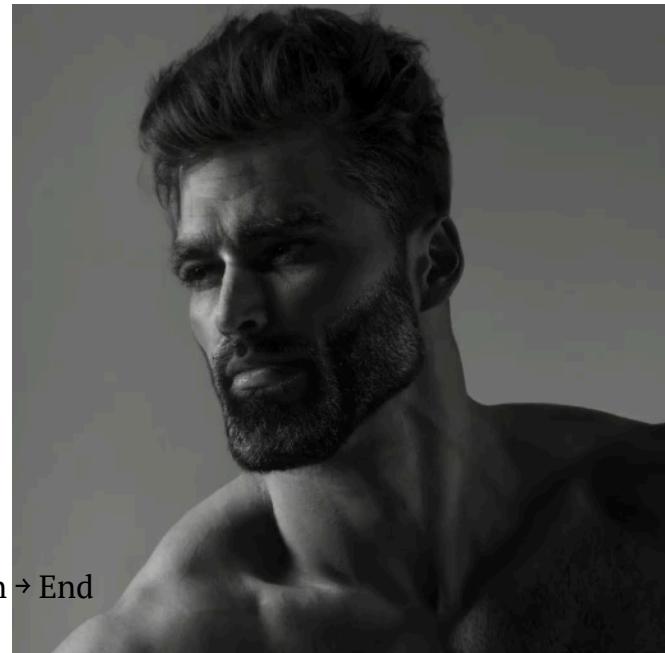
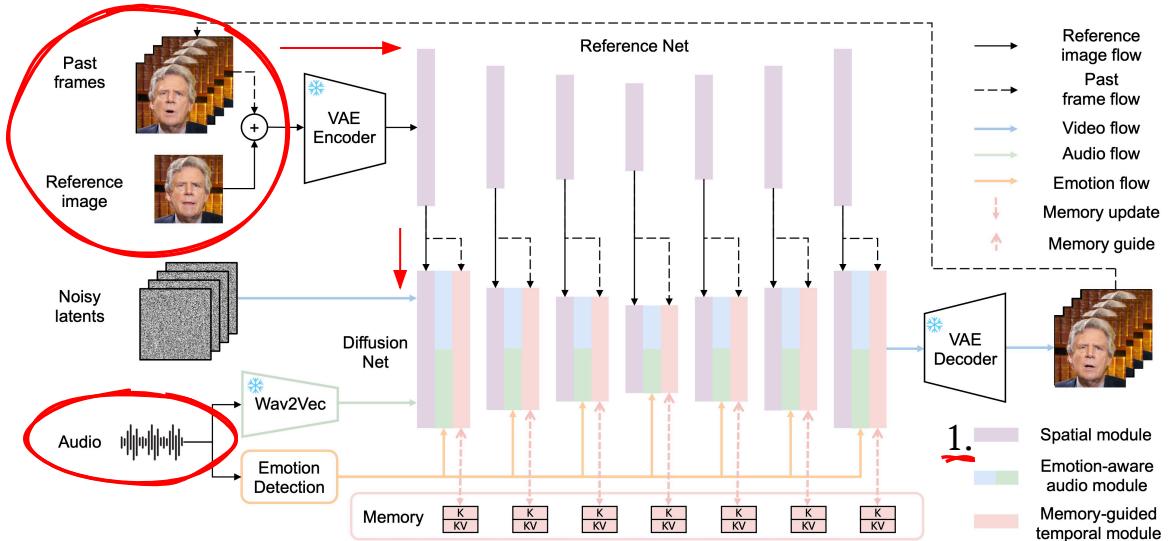
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



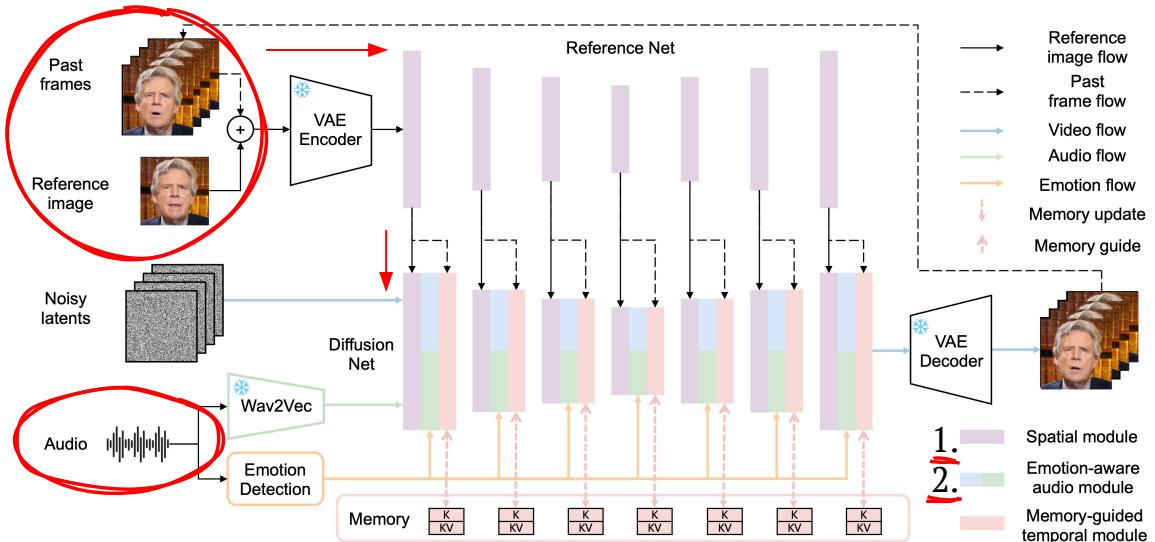
**Memoavatar Memo → Live Demo → Discussion → End**

# Memoavatar Memo



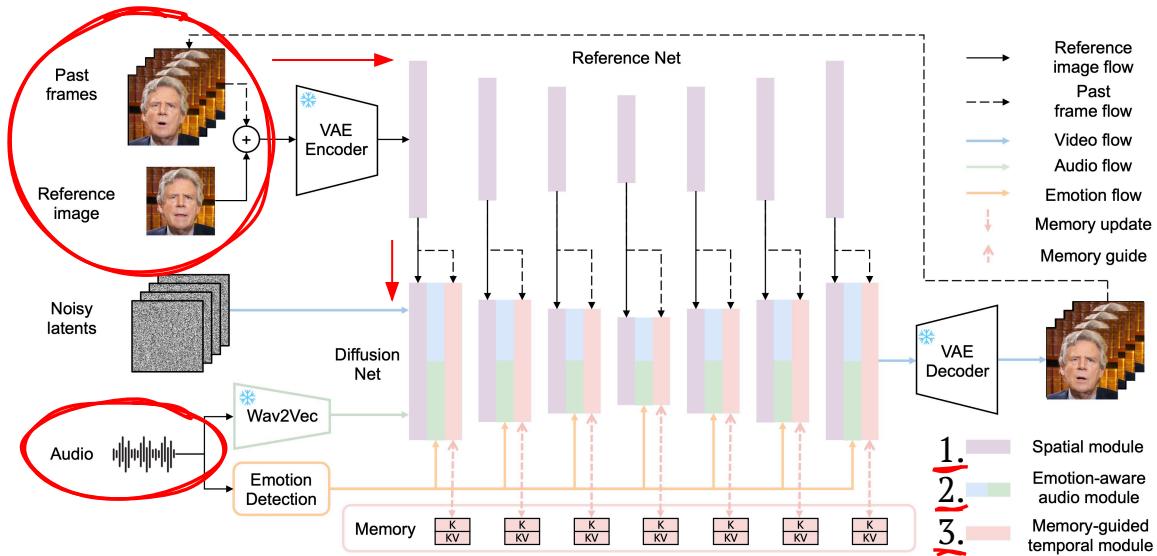
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



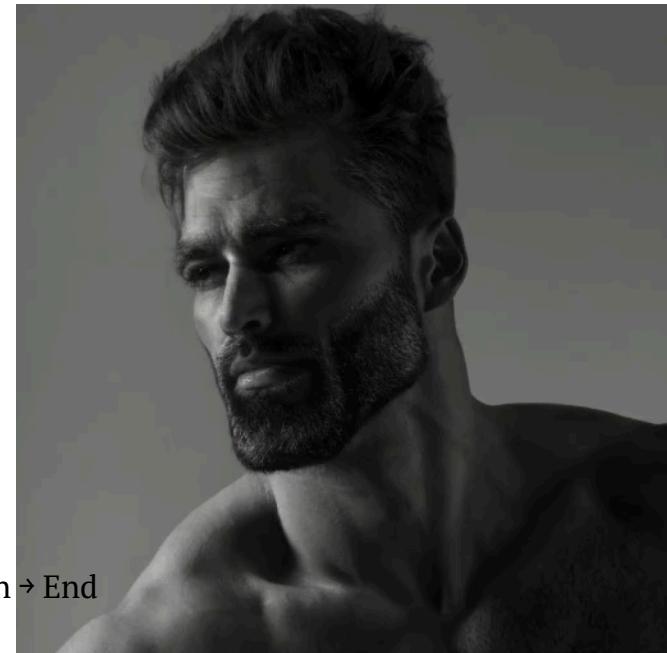
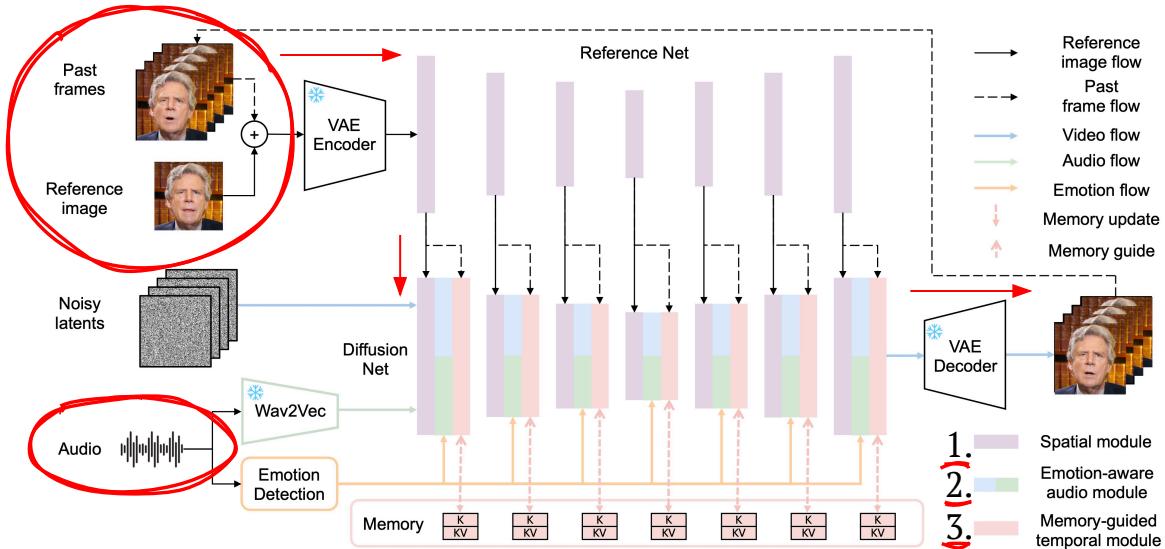
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



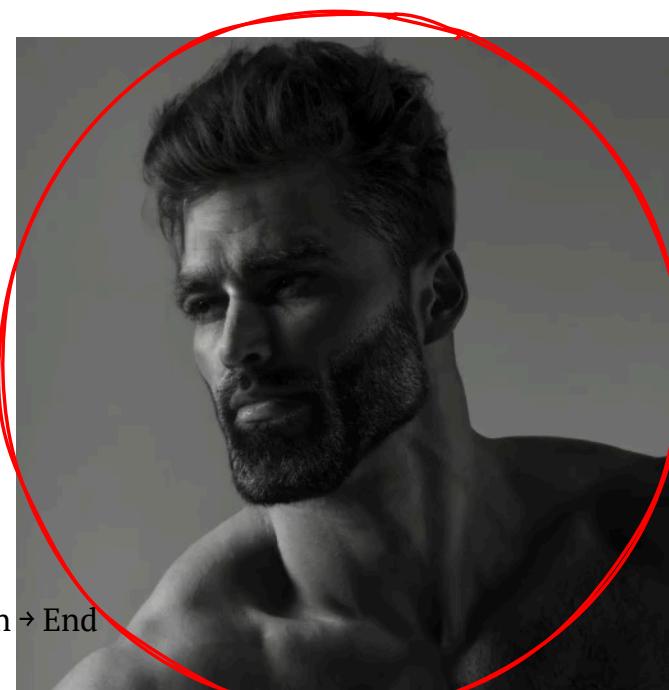
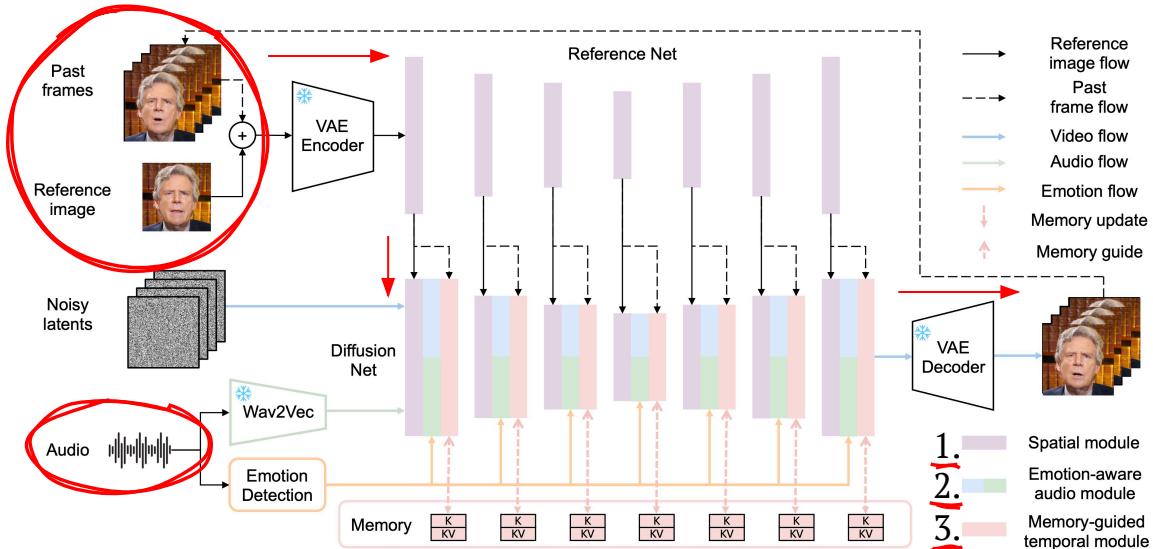
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



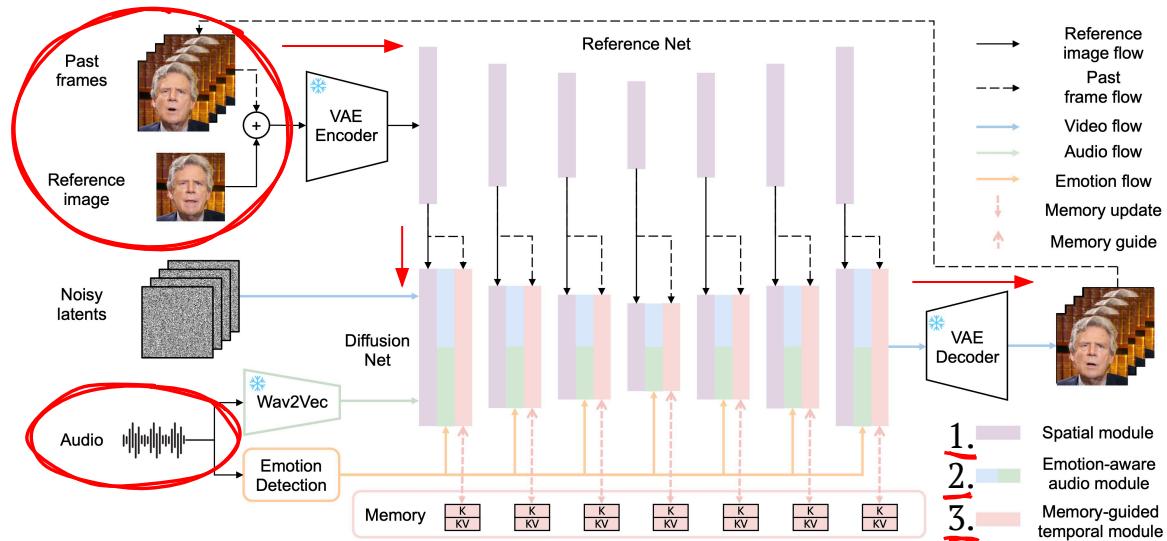
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



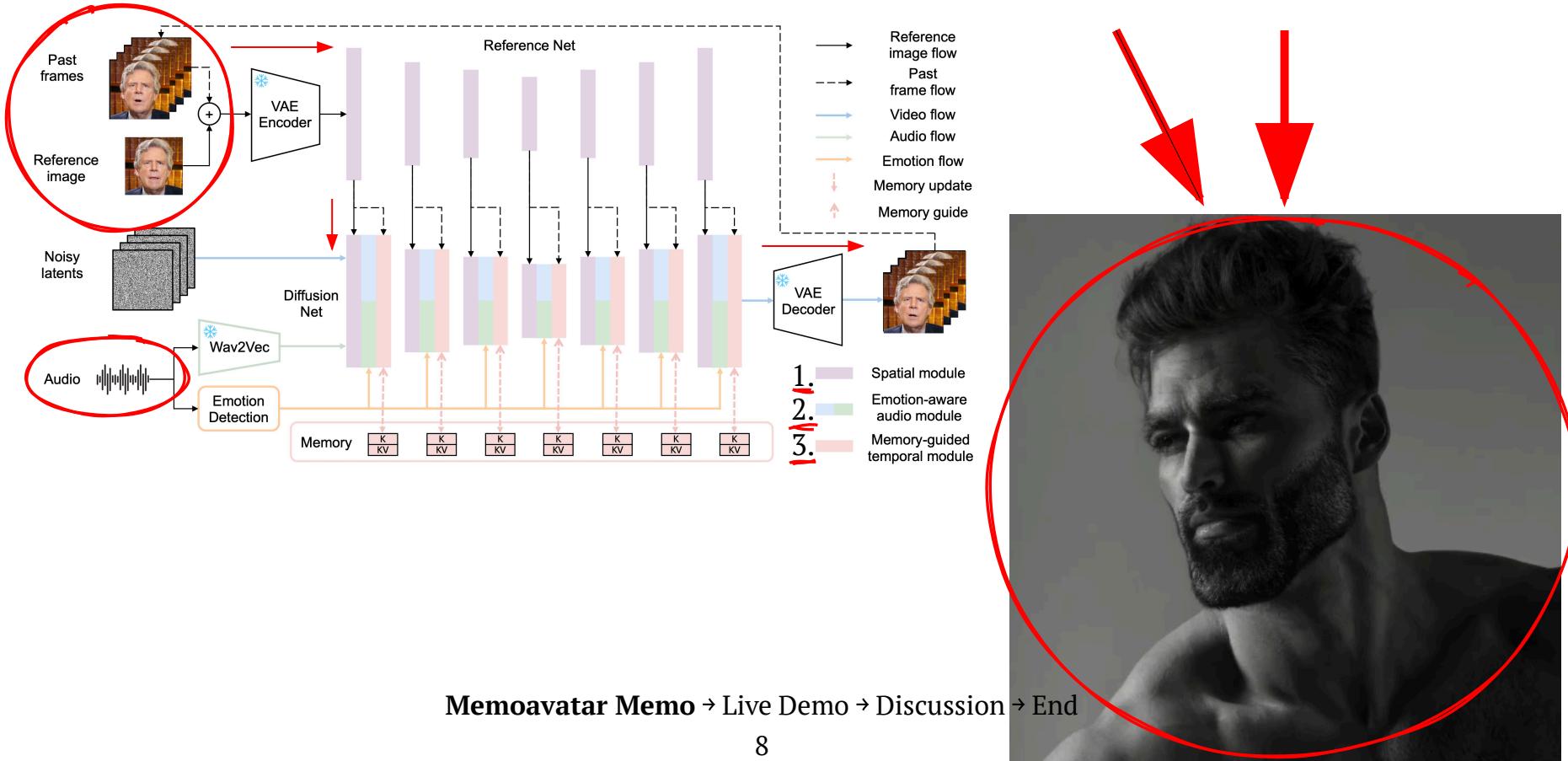
Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo

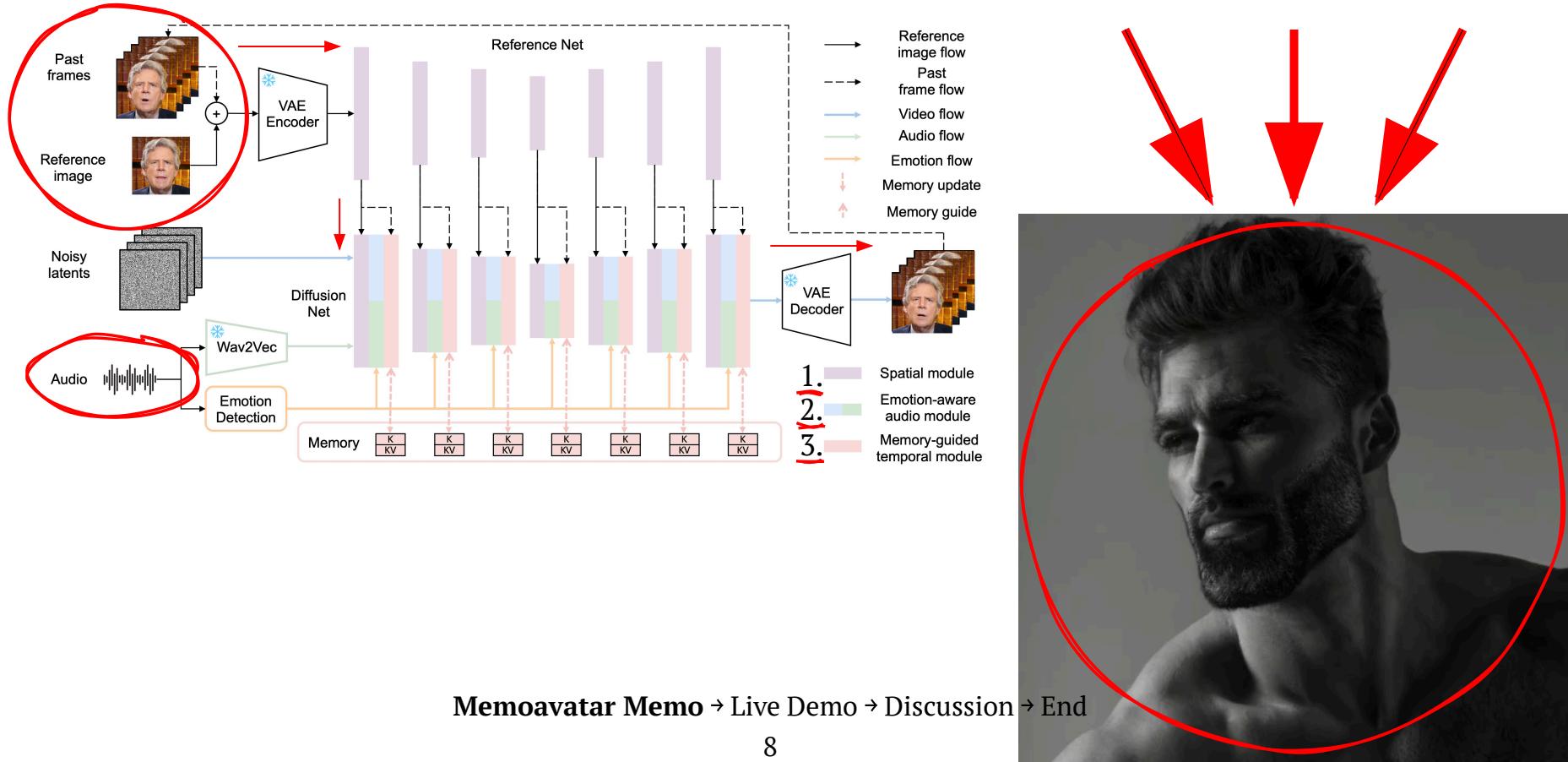


Memoavatar Memo → Live Demo → Discussion → End

# Memoavatar Memo



# Memoavatar Memo



## Live Demo

Look1!

Live Demo → Discussion → End

# Discussion

- State of the Art Technologies
  - Zyphra/Zonos 2025
  - Memoavatar/Memo 2024
- Not dependent on commercial sources, but high computing capacity is a prerequisite
- Training with own voices and generation with character picture
- In Comparison Open-Source AI models are undoubtedly improving and reaching peak performance.

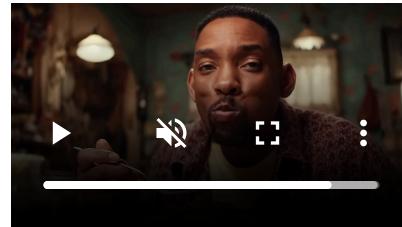
**Discussion → End**

# Discussion

- State of the Art Technologies
  - Zyphra/Zonos 2025
  - Memoavatar/Memo 2024
- Not dependent on commercial sources, but high computing capacity is a prerequisite
- Training with own voices and generation with character picture
- In Comparison Open-Source AI models are undoubtedly improving and reaching peak performance.



Stable Diffusion: Will Smith eating Spaghetti (March 2023)



Google Veo 3: Will Smith eating Spaghetti (May 2025)

Discussion → End

**End**

End

Thanks!

End