klleon

Trends and Future directions of digital human technology

INDEX

Introduction

Introducing the presenter and KLleon.

Introduction

Introducing the presenter and KLleon.

Presentor

Jisu Kang KLleon CRO (Chief Research Officer)

Resume Summary
- CRO, "KLleon"- A Deep Learning-based Video Generation Startup ('18 - present)
- Master's Degree. KAIST, Department of Electrical and Electronics ('11 - '20)
- AI Researcher, "Aizen Global", - *****AI Operating System in Finance ('17 - '18)
- Early Graduation from Sejong Science High School ('09 - '11)

Representative Paper
- "RDIS: Random Drop Imputation with Self-Training for Incomplete Time Series Data,"(2021, AAAI)
- "LEGo: LEGo: Lifelong Learning with Efficient GAN Memory," (2022, ECCV)
- "LFS-GAN: Lifelong Few-shot Image Generation," (2022, NIPS)

Patent
- Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI)
- "얼굴형을 고려하며 딥러닝 네트워크를 이용하는 배경 및 얼굴 합성 방법 및 장치" 출원번호: 10-2021-0073798
- "딥러닝 네트워크를 이용한 배경 및 얼굴 합성 방법 및 장치" 출원번호: 10-2021-0073779
- "딥러닝 네트워크를 이용한 입 모양 생성 방법 및 장치" 출원번호: 10-2021-0142877

Introduction

Introducing the presenter and KLleon.

02   **About KLleon**

Idea

Founding Idea

"What if I could make an online lecture with the face and voice of the person I like?"

Goal

Make Creative Reality

"KLleon dreams of innovation using
AI image generation technology in
the cultural content market."

Mission — Make Digital Communication Sincere

Vision — Let Everyone Become Friends
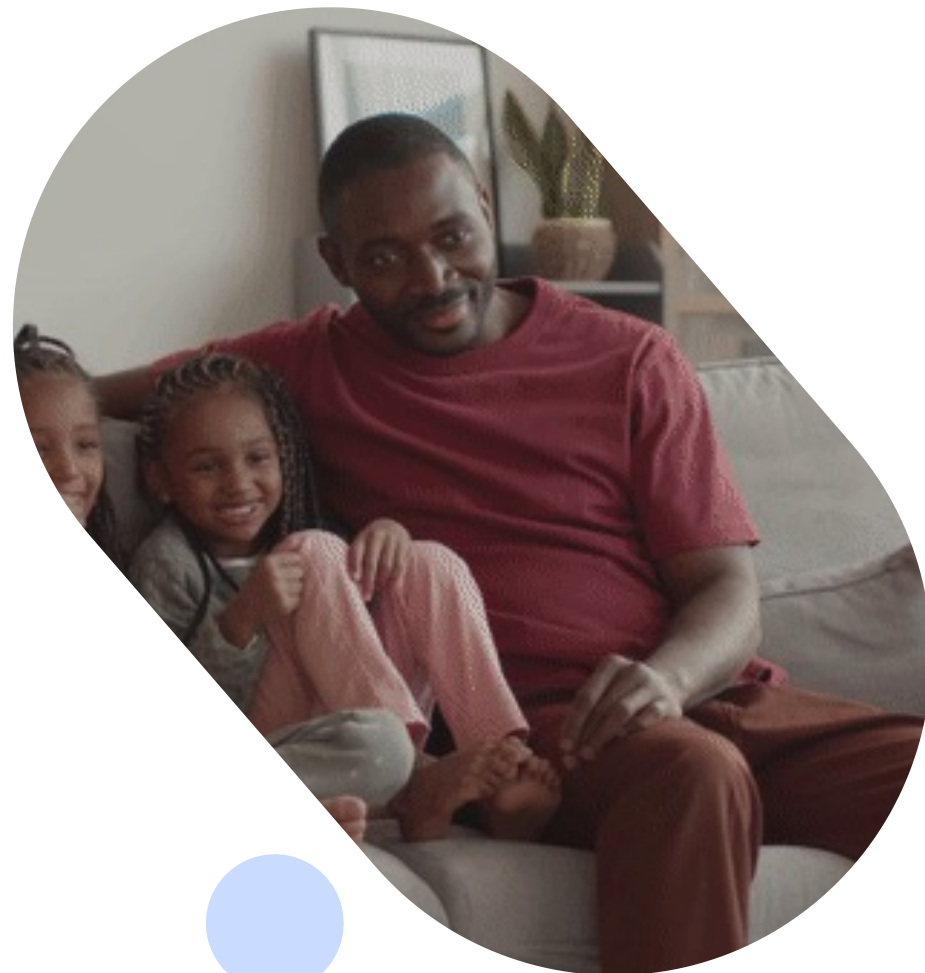Beyond the Limit of Distance

**Key
Word** — Adverturous, Outstanding, Persistent

Services

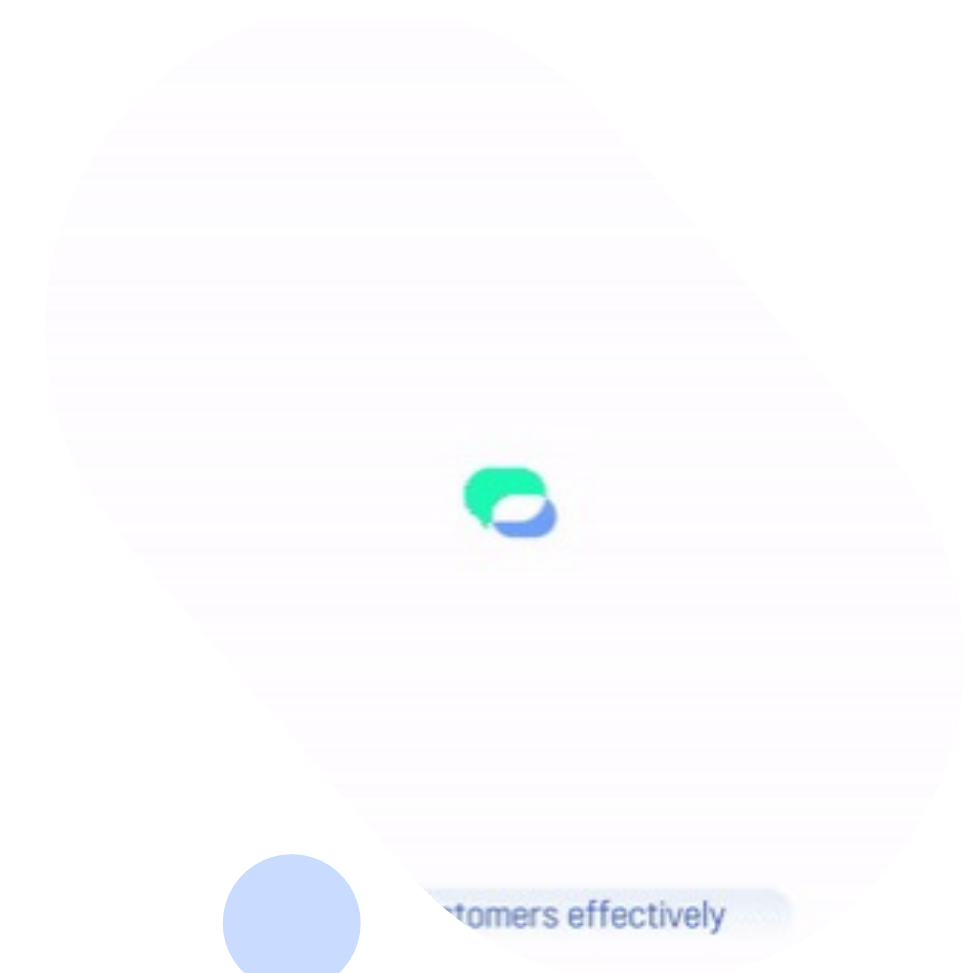# Service for innovation in communication

KLleon intends to innovate communication through three services: KLling/Klone/Kamello.

KLling service is an automatic dubbing solution that translates an actor's script in the video, regenerates the actor's voice, and synchronizes voice and lip.

KLlone service is a digital human creation solution that helps customers create and use digital human with the desired appearance.

KLlone sales service makes customer-tailored sales simple and effective using digital humans.

© KLleon

Digital Human

Innovation

"The digital human can solve temporal and spatial limitations, providing an opportunity to communicate with anyone, anywhere, anytime."

Technologies

Introduces technologies for creating digital humans

Technologies

Introduces technologies for creating digital humans

01   CGX/VFX Based

CGX/VFX

CGX/VFX based Technology

Create a digital human using a 3D engine such as Unreal Engine/Unity.

Unreal Engine is based on C++ and provides a Blueprint function so designers can test for graphic work.

Unity is based on C#, so the entering barrier is low. Unity is light and optimized for multi-platform support such as TV/web/mobile.

© KLleon

Unreal

Metahuman Creator

"It used to take months to create a digital human, but now it takes less than an hour."
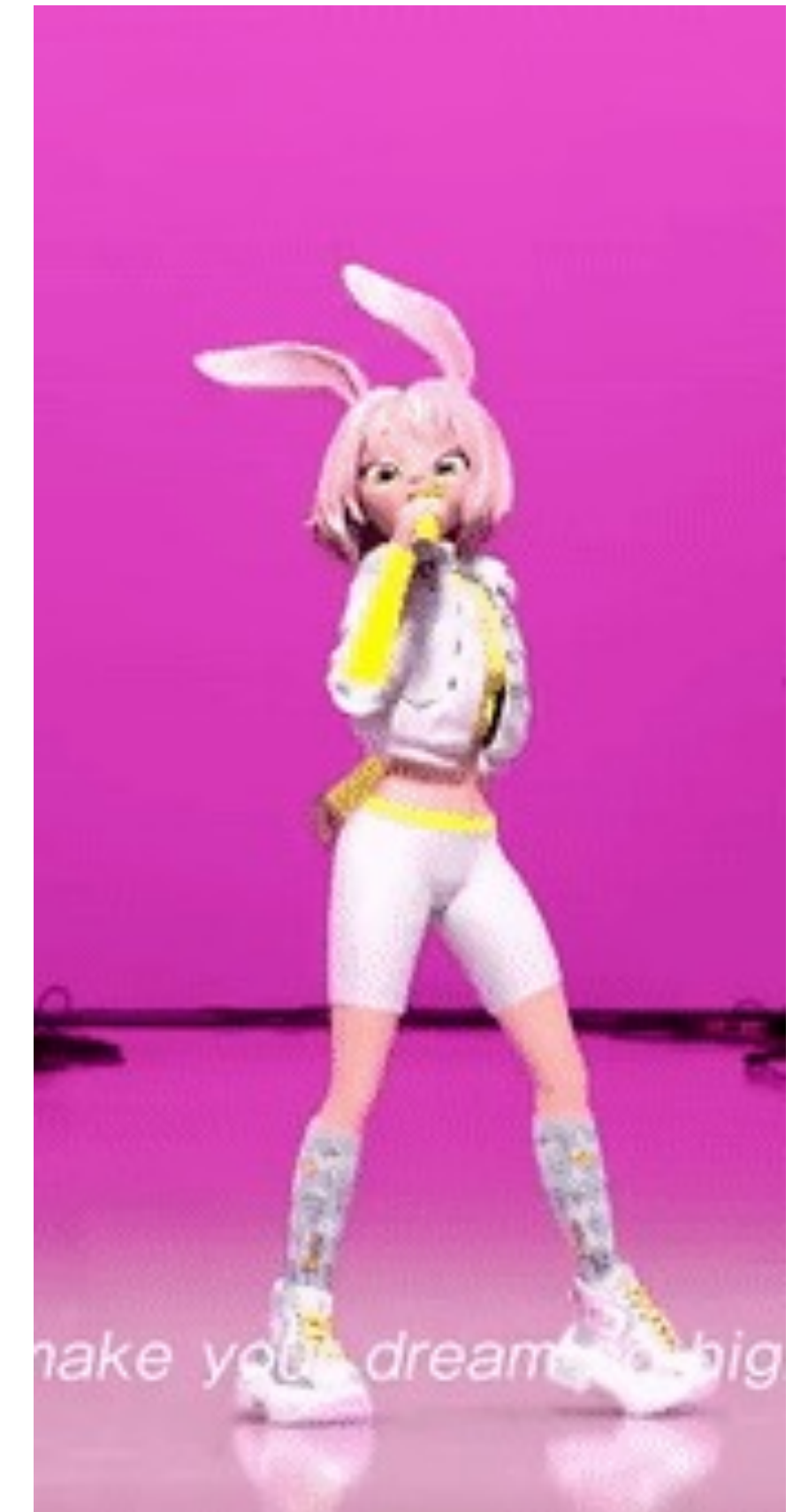

© KLleon

Mocap

Motion Capture

"It is possible to make digital humans imitate human behavior using motion capture equipment."

Example

Digital Human Example

"Various digital humans have been
created using CGX/VFX technology."

Pros/Cons

CGX/VFX

"Advantages and Disadvantages of Creating Digital Humans Using CGX/VFX Technology"

Pros/Cons

Sophisticated digital human creation is possible, but customization is limited using CGX/VFX technology.

Pros — Sophisticated digital human generation and real-time motion generation is possible.

Pros — Full-body digital human creation is possible.

Cons — It usually takes a long time to make. It is also expensive to create the desired digital human.

Cons — In the case of the metahuman creator, customization is difficult.

Technologies

Introduces technologies for creating digital humans

02 AI Based

AI

The Rapid Development of AI Tech

From ALEX Net in 2012 to now, AI technology has been developed rapidly, and AI can now have the creative ability that considered as human only has.

1	2	3	4
2012	**2014**	**2016**	**2018**

Professor Geoffrey Hinton's ALEX Net wins the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with an overwhelming performance.

Generative Adversarial Network (GAN) developed by Professor Ian J. Goodfellow opened the horizon for generative models.

Google DeepMind's AlphaGo catches the world's attention by defeating the world's top professional knight, Sedol Lee.

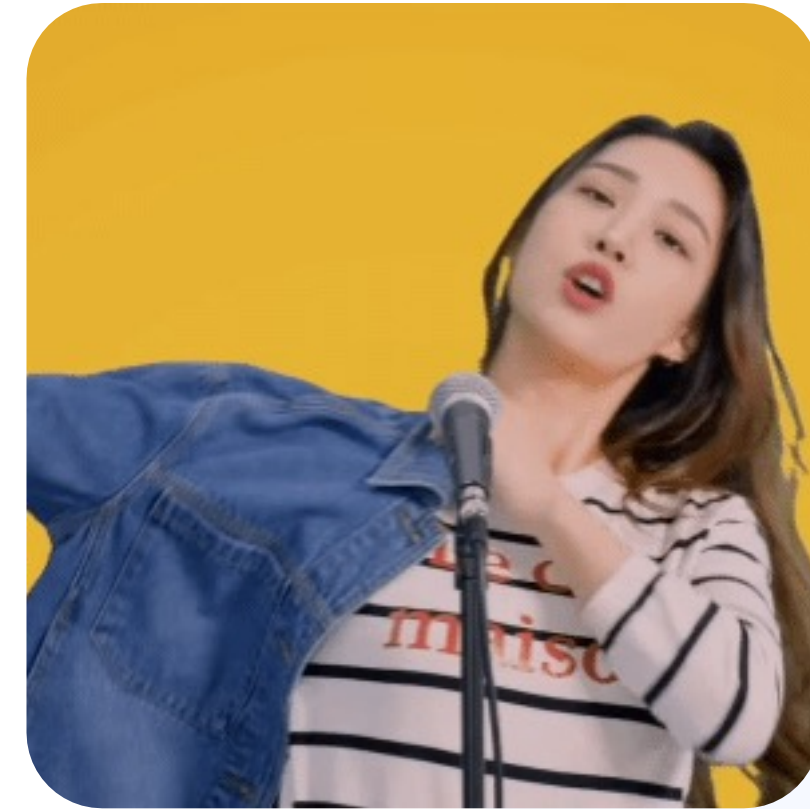Deepfake-based video production has become a hot topic, and various deepfake videos are spreading on the web.

© KLleon

Examples

Various Digital Human

"A variety of digital humans are being created using deepfake technology."

Problem

Limitations

"Deepfake technology requires lots of data and a long training time."

Data and Time Limitations

Using deepfake technology requires data collection and model training.

Data
A large number of target face images (various expressions and angles, 1000+) are required to create a new face.

Time
Trainning is required to create a new face, and it takes more than 12 hours to learn.
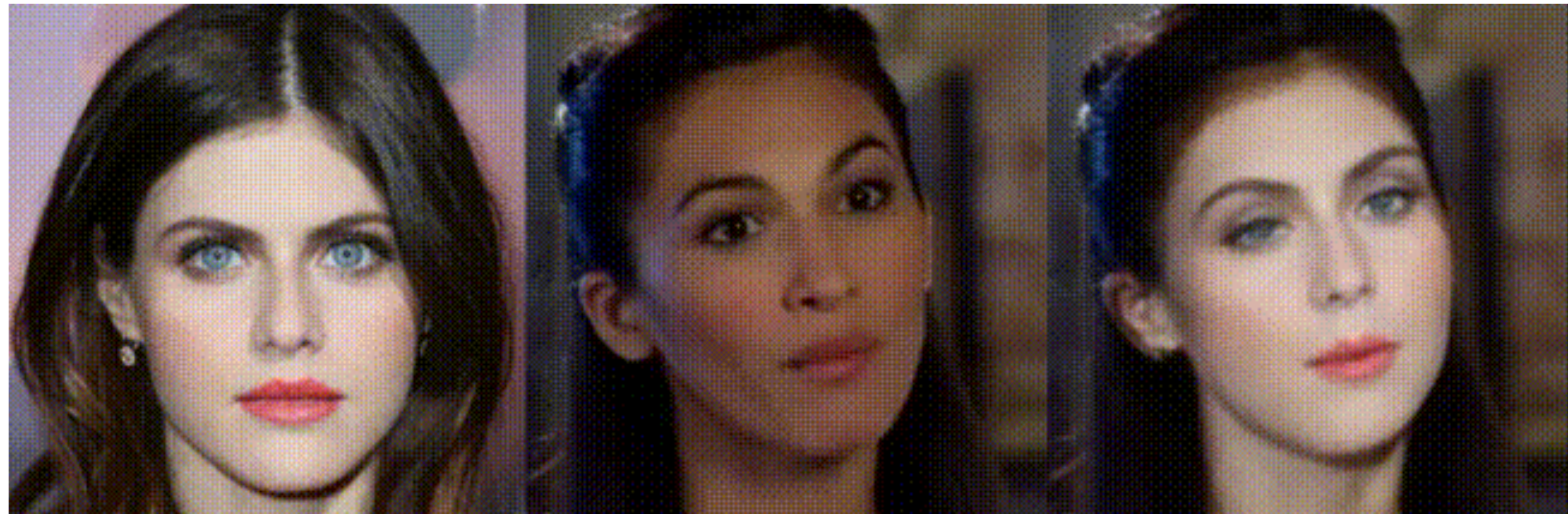
Tool
To use deepfake technology, you need to be able to handle the tools for production.

Reenactment
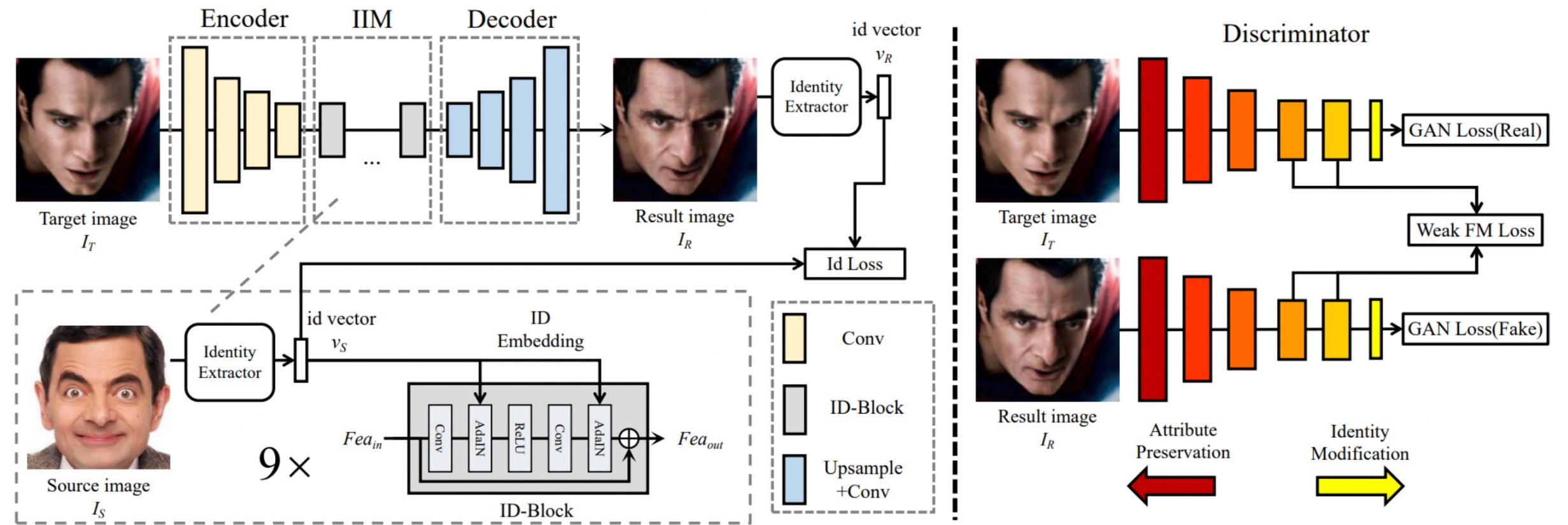
Only one image...

"Unlike deepfake technology, which took a long time to collect data, you can now create the face you want with just one picture."
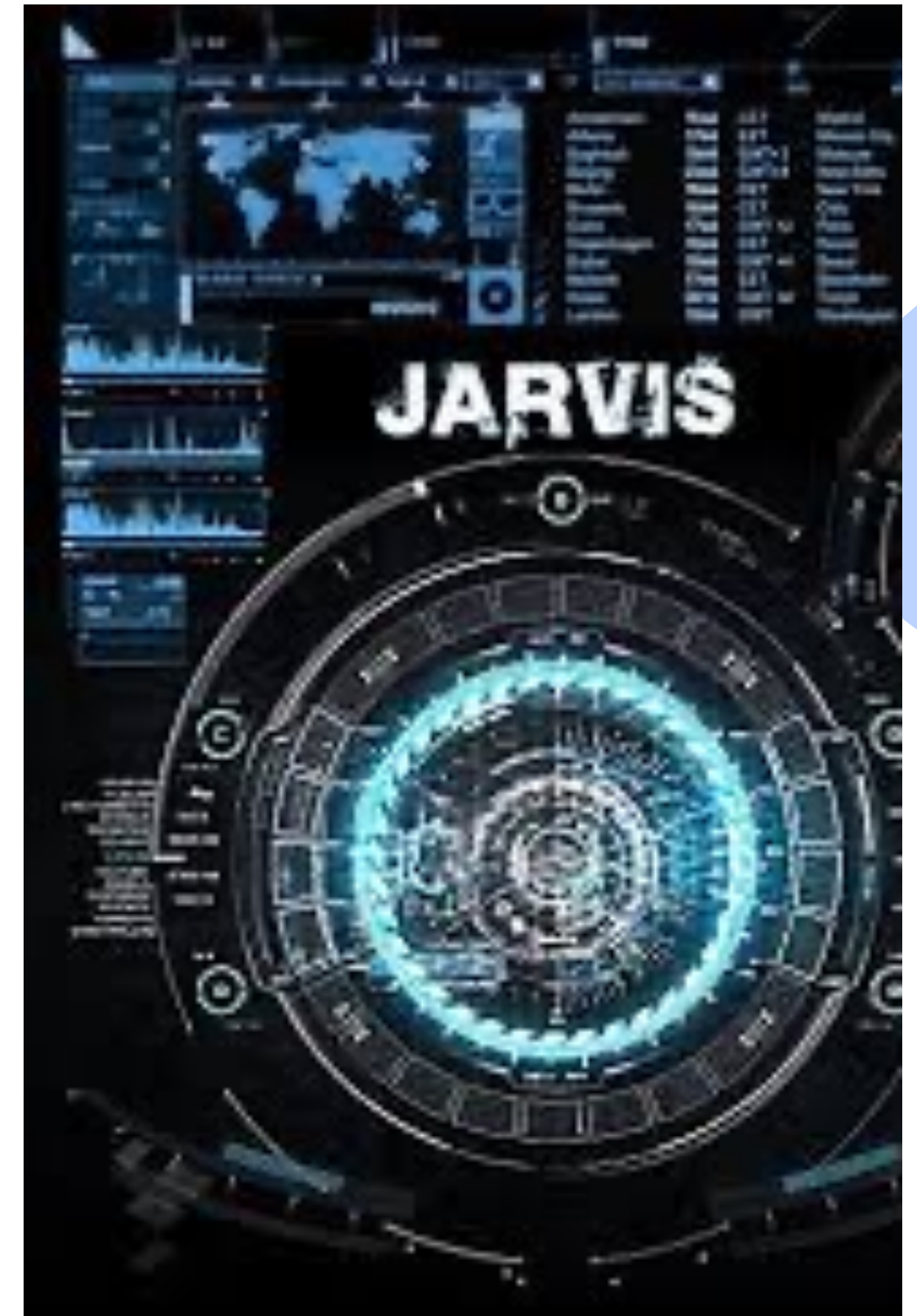
How

Like a painter

"Painters can imagine a person's profile and various expressions just by looking at one picture because painters have been drawing many faces of several people."

Question

Digital Human?

"Can a virtual human created by changing only the face to the desired look be called an actual digital human?"
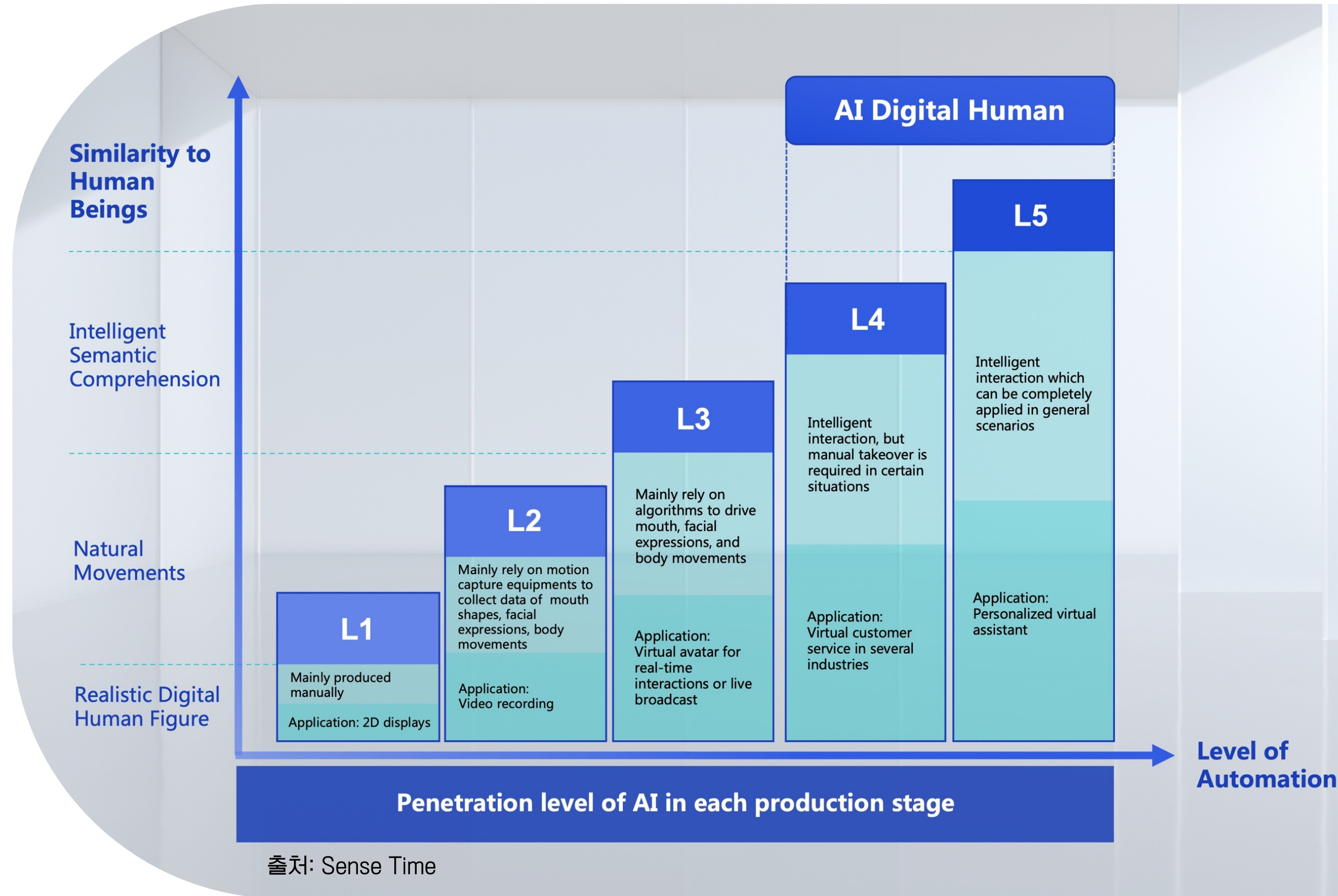
Question

Digital Human?

"There are many steps left to become a complete digital human."

Similarity to Human Beings

AI Digital Human

Intelligent Semantic Comprehension

Natural Movements

Realistic Digital Human Figure

L1
Mainly produced manually

Application: 2D displays

L2
Mainly rely on motion capture equipments to collect data of mouth shapes, facial expressions, body movements

Application: Video recording

L3
Mainly rely on algorithms to drive mouth, facial expressions, and body movements

Application: Virtual avatar for real-time interactions or live broadcast

L4
Intelligent interaction, but manual takeover is required in certain situations

Application: Virtual customer service in several industries

L5
Intelligent interaction which can be completely applied in general scenarios

Application: Personalized virtual assistant

Level of Automation

Penetration level of AI in each production stage

출처: Sense Time

© KLleon

KLleon Tech

Introducing KLleon's technology to create a digital human.

KLleon Tech

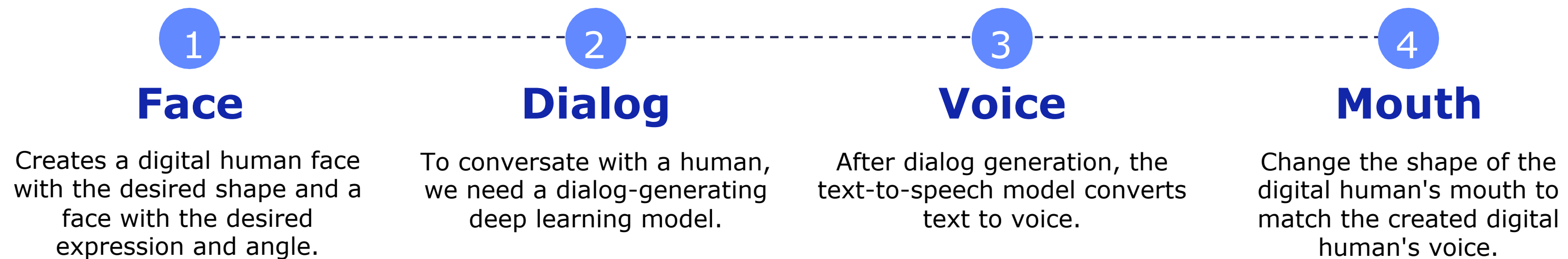**Introducing KLleon's technology
to create a digital human.**

01   **Overall
Process**

Process

Digital Human Creation Process

"Various deep learning models are needed to create a digital human, such as face generation, dialogue generation, voice generation, and lip-sync generation."

1 **2** **3** **4**

Face

Creates a digital human face with the desired shape and a face with the desired expression and angle.

Dialog

To conversate with a human, we need a dialog-generating deep learning model.

Voice

After dialog generation, the text-to-speech model converts text to voice.

Mouth

Change the shape of the digital human's mouth to match the created digital human's voice.

© KLleon

KLleon Tech

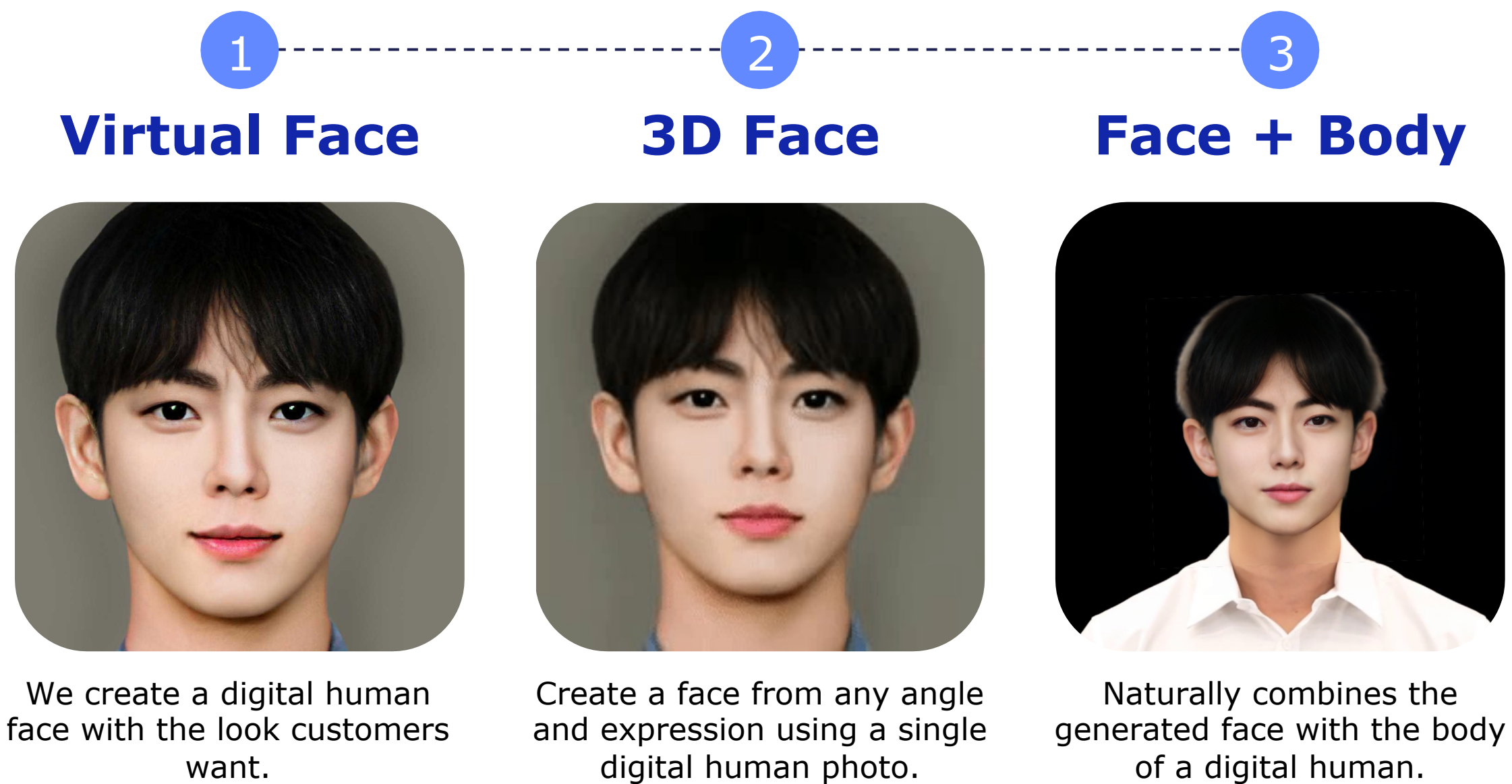**Introducing KLleon's technology
to create a digital human.**

02   **Face
Generation**

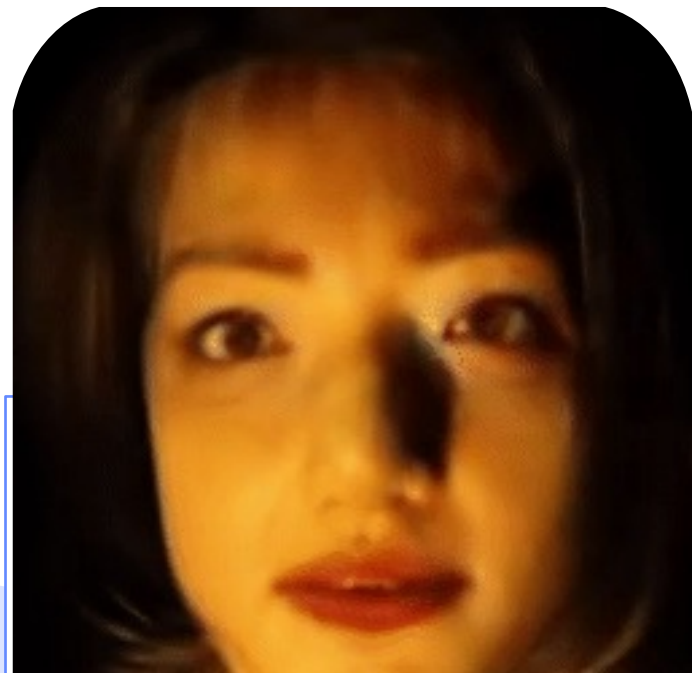**Face**

Face Generation Process

"Three processes are required to generate a digital human face: creating a virtual face, transforming to a 3D face, and combining the face and body."

1	2	3
Virtual Face	**3D Face**	**Face + Body**

We create a digital human face with the look customers want.

Create a face from any angle and expression using a single digital human photo.

Naturally combines the generated face with the body of a digital human.

© KLleon

Compare

Face Generation Quality

"Our digital human face generation performance is very sophisticated and fast compared to other companies."

- SOTA face generation performance

- Composite performance is degraded when lighting is complex
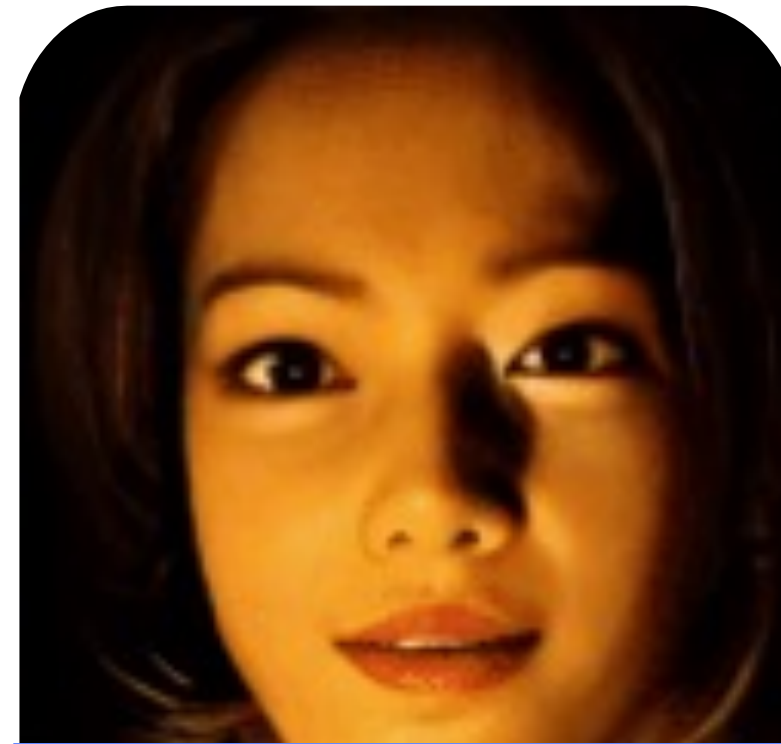
- KLleon face generation performance
- Natural and sophisticated facial generation is possible even in challenging environments.
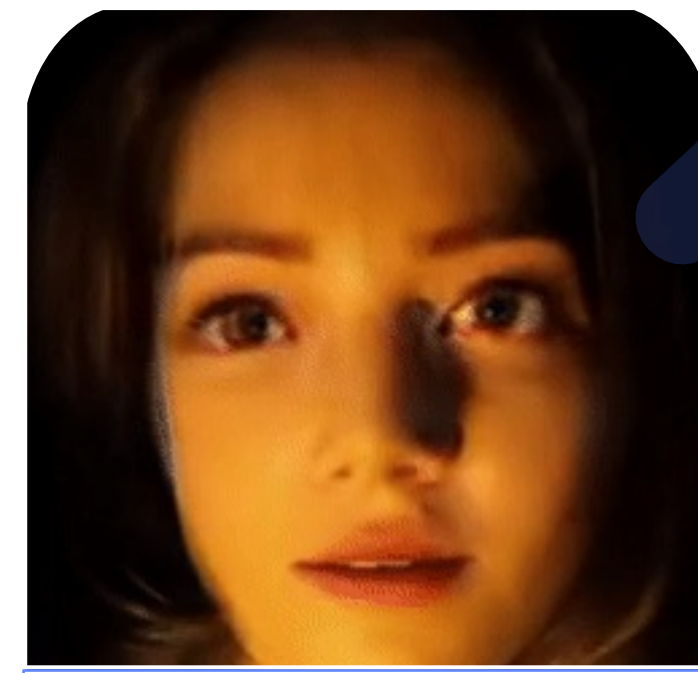
- SOTA face generation performance

- Composite performance is degraded when lighting is complex

Compare

Face Generation Quality

"Our digital human face generation performance is very sophisticated and fast compared to other corporations."

	Resoultion	FPS	Accuracy	Quality
Company A	256x256	24FPS	48%	1.29
Company B	256x256	17FPS	44%	2.36
Company C	256x256	30FPS	68%	0.72
KLleon	**1024x1024**	**123FPS**	**70%**	**1.66**

Quantitative comparison between KLleon and other SOTA technologies

© KLleon

KLleon Tech

**Introducing KLleon's technology
to create a digital human.**

03   TTS &
NLP

NLP

Dialog Generation

"Dialog generation model used for communication between humans and digital humans"

Speech Signal

Hypothesis
are there any action movies to
see this weekend

Automatic Speech Recognition (ASR)

Language Understanding (LU)
- Domain Identification
- User Intent Detection
- Slot Filling

Text Input
Are there any action movies to see this weekend?

Semantic Frame
request_movie
genre=action, date=this weekend

Natural Language Generation (NLG)

Dialogue Management (DM)
- Dialogue State Tracking (DST)
- Dialogue Policy

Text response
Where are you located?

System Action/Policy
request_location

Task Oriented Dialogue System

© KLleon

TTS

Text-to-Speech

"KLleon's TTS model converts the text produced by the dialogue generation model into the voice of a digital human. KLleon's TTS model requires only 30 seconds of voice to generate desired voice."

Voice Model

. . . . . .

Only 30 sec voice

→

Few-Shot TTS

Result

KLleon Tech
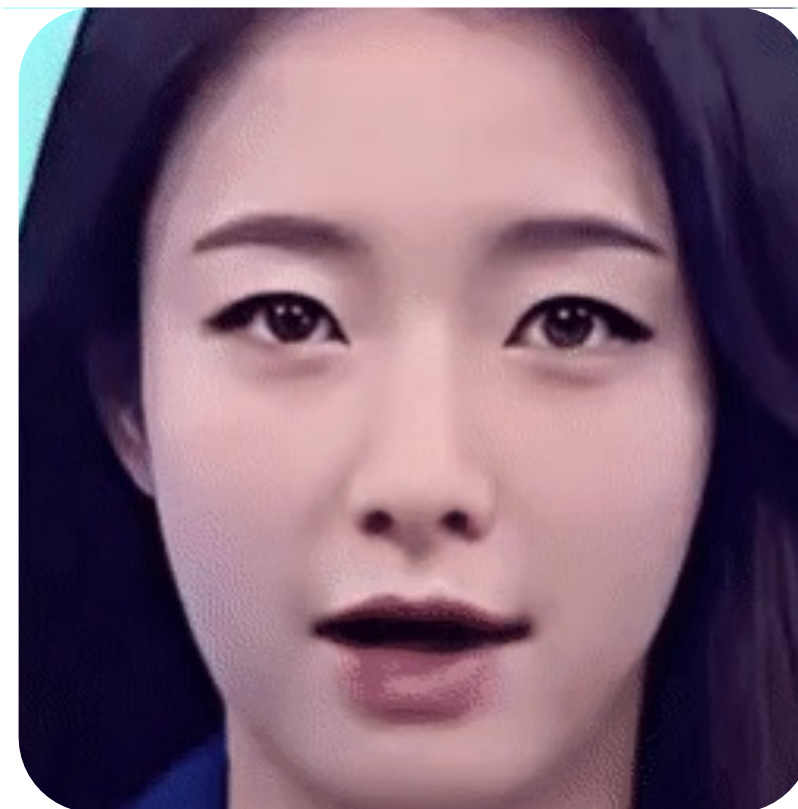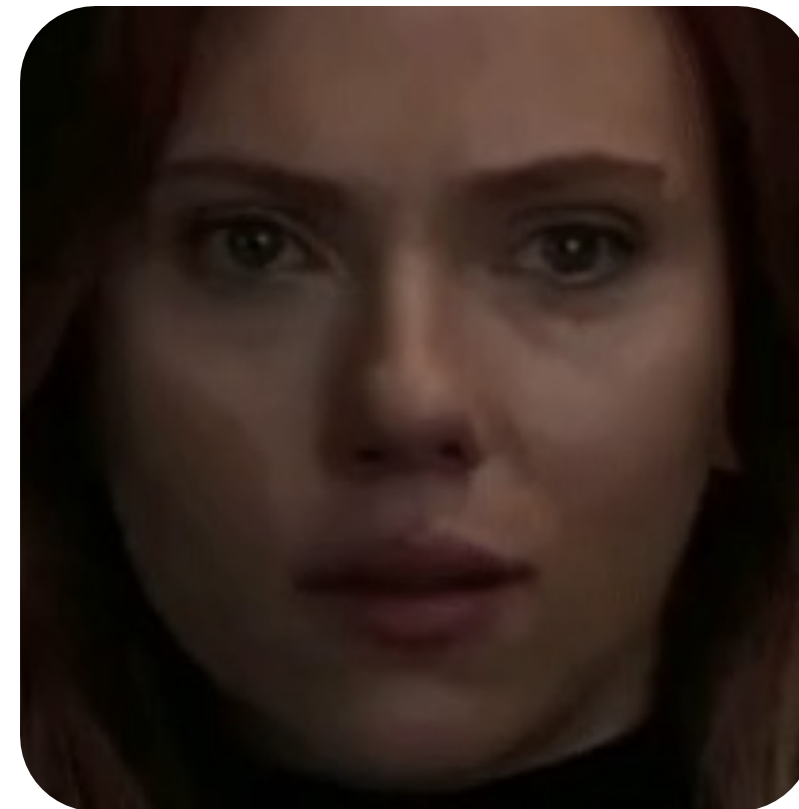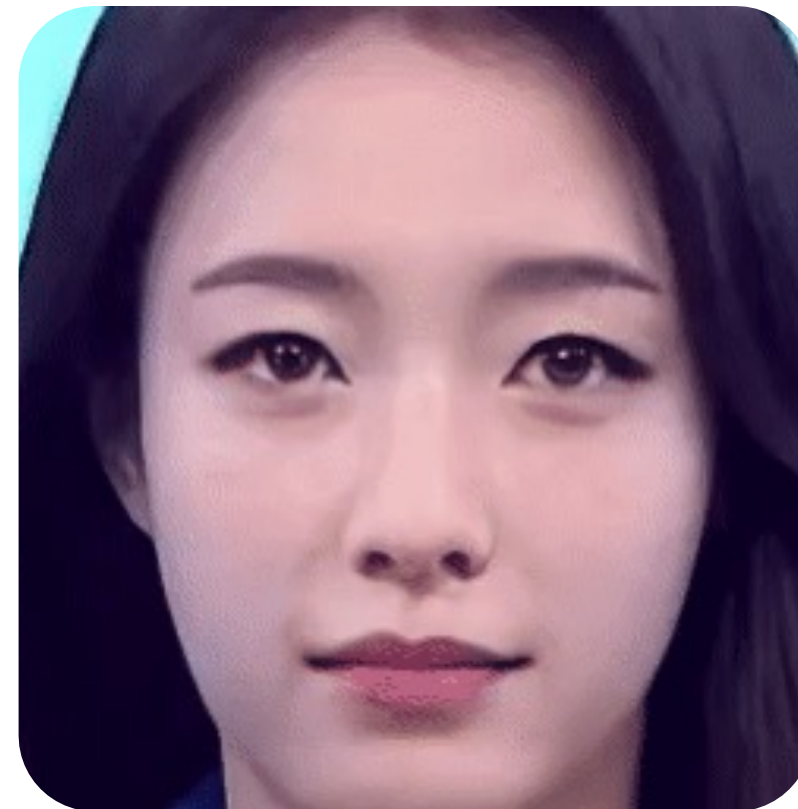
**Introducing KLleon's technology
to create a digital human.**

04   # Lip Sync
Generation

Lip Sync

Lip Generation

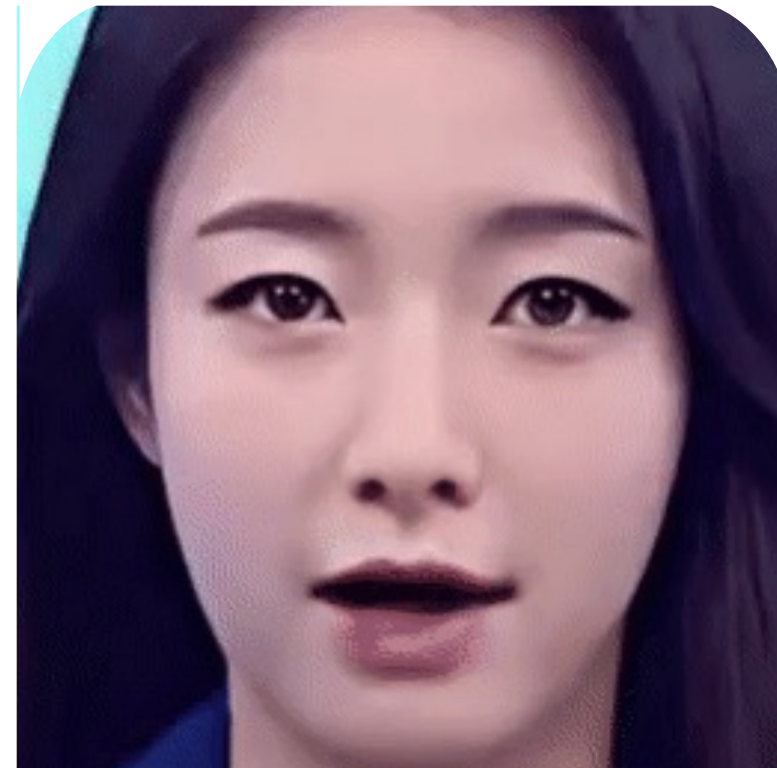"We need to synchronize the voice and the mouth of the digital human."
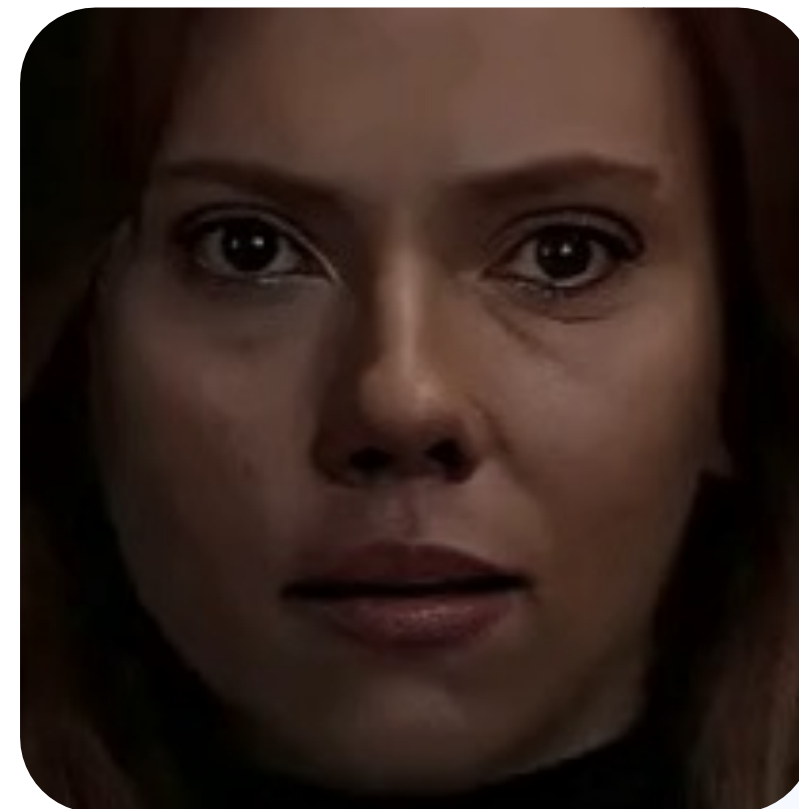
Compare
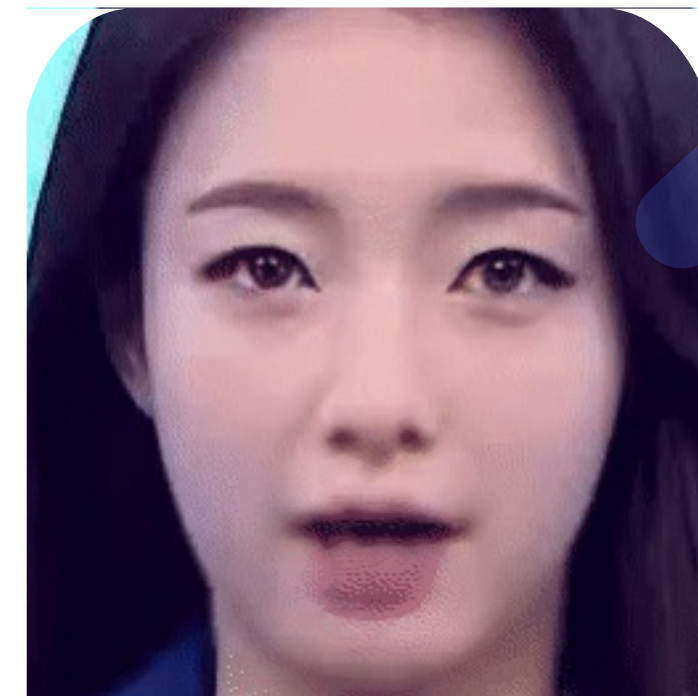
Lip Sync Generation Performance

"Our lip sync generation performance is very sophisticated and fast compared to other companies."
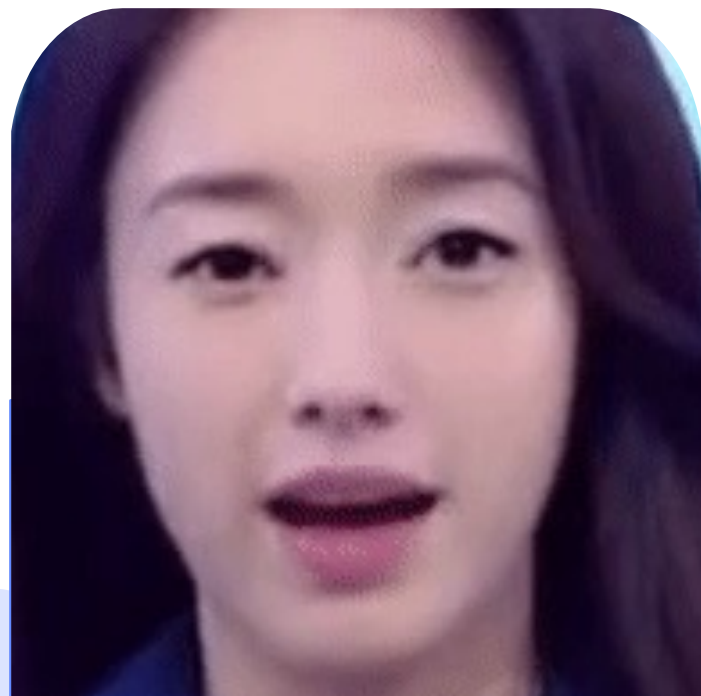
- SOTA lip-sync generation performance

- Poor quality, unnatural facial movements

- KLleon lip-sync generation performance

- High Resolution and natural facial movements

- SOTA lip-sync generation performance

- Low Resolution

Compare

Lip Sync Generation Performance

"Our lip sync generation performance is very sophisticated and fast compared to other companies."

	Resoultion	FPS	Accuracy	Required Data
Company A	256x256	Train required	7.2	More than 100
Company B	256x256	11FPS	6.4	1 image
Company C	98x98	24FPS	6.9	1 image
KLleon	**512x512**	**37FPS**	**7.0**	**1 image**

Quantitative comparison between KLleon and other SOTA technologies

© KLleon

Future Tech

Introducing the future technology direction for creating digital humans.

Future Tech

Introducing the future technology direction for creating digital humans.

01 **Road Map**

Future

Technology direction for complete communication

"We suggest the direction of technological development necessary for complete communication with digital humans."

1　　　　**2**　　　　**3**

Body　　**Emotion**　　**Memory**

"Body generation technology is essential to generate not only the face but also the body's movement."

"For complete communication, it is necessary to understand human emotions and be able to communicate accordingly."

"To fully communicate, a digital human needs to remember conversations with a human and learn new conversations to improve continually."

Future Tech

Introducing the future technology direction for creating digital humans.
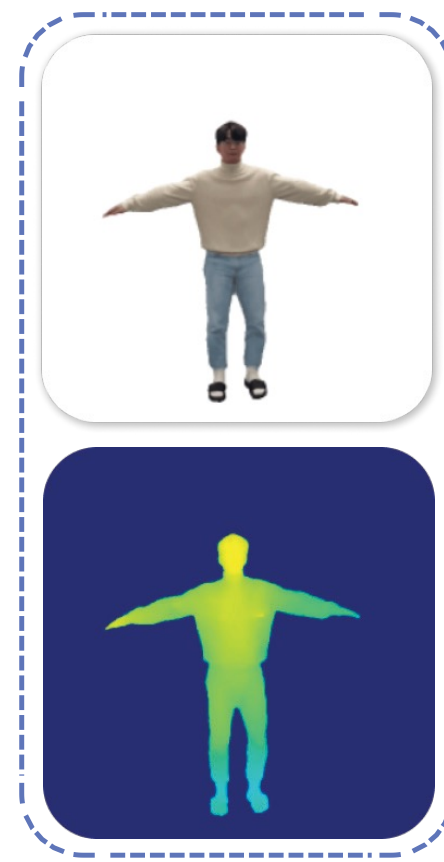
02   **Body Generation**

Body

Body Movement

"Complete digital human generation is possible when not only face but also body movement is possible."

Deep Learning

Motion

Body Generation

Future Tech

Introducing the future technology direction for creating digital humans.

03   **Conversation with Emotion**

Emotion

Emotional Interaction

"Digital humans need to understand and communicate with human's emotions."

Everyting match emotions

For emotion-based conversations, movements, voices, and conversations must match emotions.

Multi-modal
Not only human words but also actions contain emotions. Facial expressions, voices, and conversation content must all be integrated to understand a human's emotions.

NLP
The digital human need to create conversations that match the feelings of human.

Voice
The digital human need to synthesize a voice that matches the emotion of the spoken text.

Video
The digital human should automatically generate facial expressions and actions to fit the conversation rather than imitate the actors' facial expressions and movements.

Multi-modal

Emotion Recognition

"Since human emotions are complex, we need to analyze the voice, facial expression, and conversation content in an integrated way to recognize emotions."
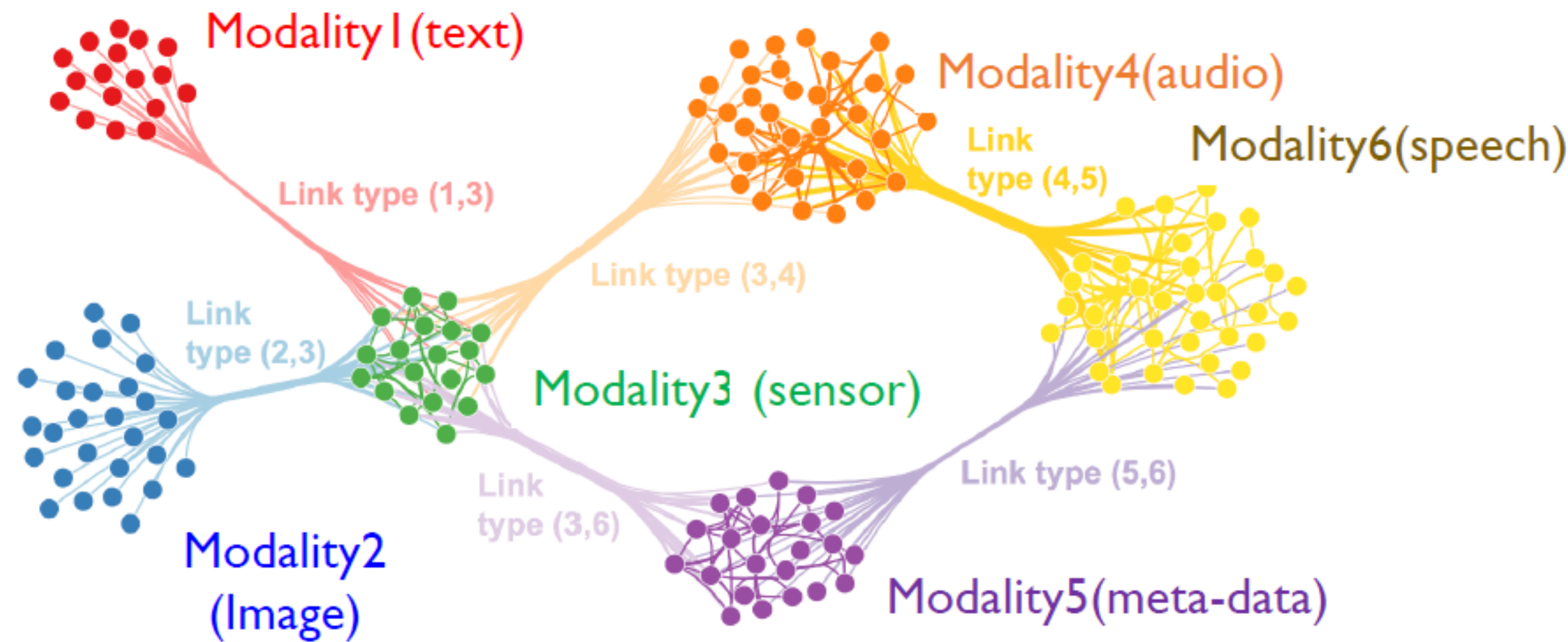
NLP

Emotion based Conversation

"The digital human should create text that matches the emotions of the human."

Hi there! I'm a chatbot and I'm here to answer your questions!

You can ask me all sorts of questions about our services. But I can't make a registration for you (just not yet!).

But first things first, how are you doing? 😄

I'm good! How about yourself?

Awesome! I'm great as well. I mean just look at that weather outside! 🌤

Hi! I'm a chatbot and I'm here to answer your questions.

You can ask me all sorts of questions about our services. However, I can't make a registration for you (not yet!).

That's okay.

Where can I find more information about your products?

You can find information about our products by clicking on this link: products

Chatbot with **high extraversion**:
- focus on social talk (e.g. chit-chat)
- high use of exclamation marks
- usage of emojis

Chatbot with **low extraversion**:
- focus on task at-hand (e.g. no chit-chat)
- moderate use of exclamation marks
- no emojis

© KLleon

Voice

Voices that match emotions

"It is necessary to generate a voice that matches the emotions in the text spoken by digital humans."

Video

Actions

"The digital human must create facial expressions and actions that match the emotions in the digital human's text."

3D Skeleton Model driven by TTS   Our Result - Environmental Broadcaster

Future Tech

Introducing the future technology direction for creating digital humans.

Conversation with Memory

Memory

Memory based Conversations

"The digital human must remember previous conversations and be able to communicate based on memory."

Remember and learn conversations

For complete conversation, the digital human need to remember and learn from existing conversations.

Memory   Remember the conversations that the digital human had with a human.

Lifelong   The digital human should be continuously learnable based on the conversation content.

© KLleon

Memory

Dialog Memory

"For digital humans to empathize and become friendly through conversations, they need to be able to store and remember existing conversations."

Blender Bot 2.0

Static ML

Adaptive ML
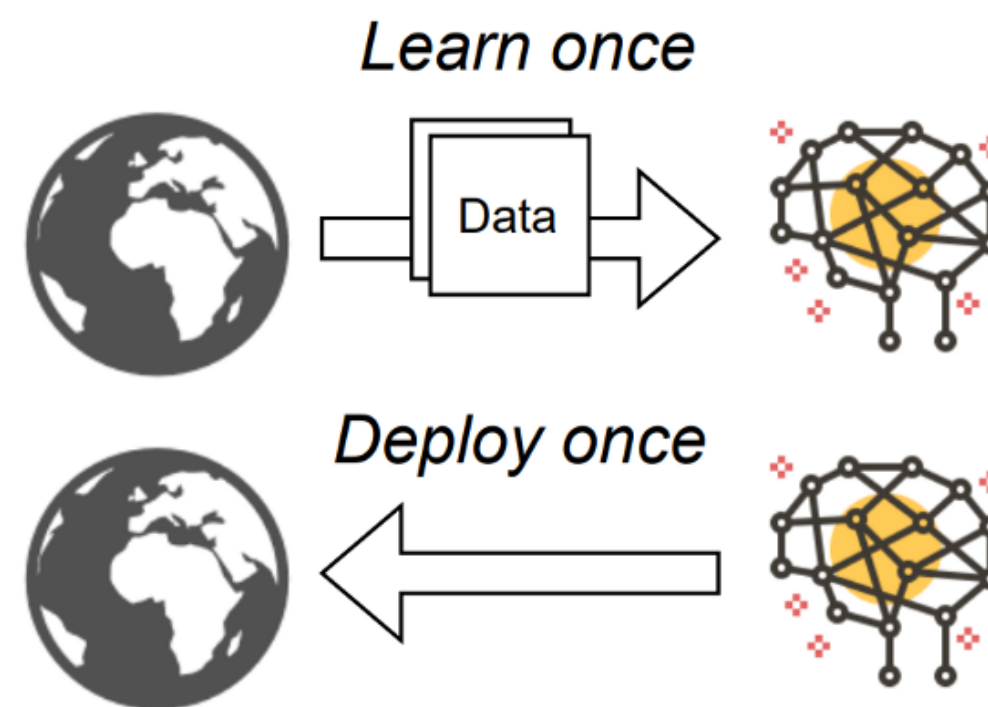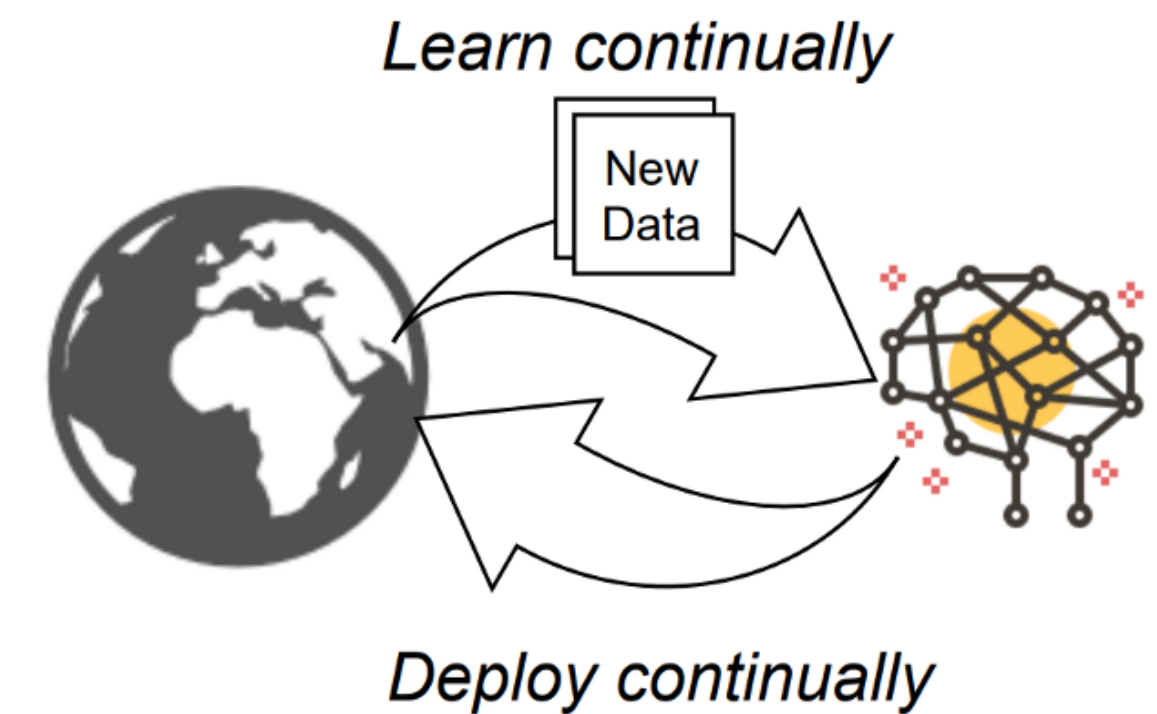
Lifelong

Continuous Evolution

"Digital humans need to be able to learn through conversation and continue to evolve."

Learn once

Data

Deploy once

Learn continually

New Data

Deploy continually

Thank You

Jisu.kang@klleon.io

klleon