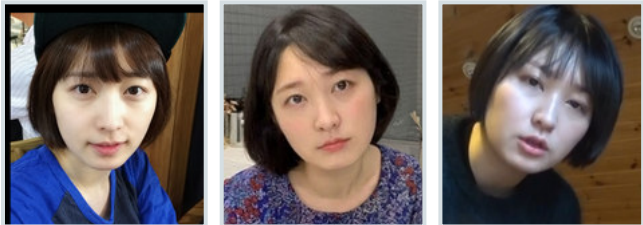


a Source images

D · follow-up



A–C · three-photo pilot



A

B

C

4 photographs
1 person

b Comparison inputs

Change the context



Background

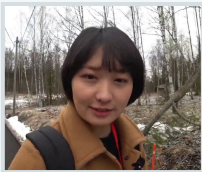


Clothing

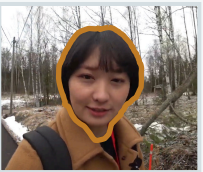


Accessory

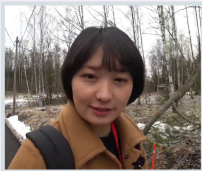
Processing control



Downsample
and restore



Repeat the same file



Request 1
Request 2

Face pixels restored exactly

c Evaluation

Follow-up prompts
Plain / Ignore context



Multimodal models

gpt-6-luna

gpt-6.1-sol

2 repeats for each
model–prompt pair

Fresh conversation
for each request

d Readouts

Face rating

① ② ③ ④ ⑤ ⑥ ⑦

Facial descriptions

Mouth · teeth

Expression · outline

Compare responses

Edit – original

Sham – original

Same-input repeats

Pilot also rates the
whole photograph.

Pilot: 9 inputs × 2 models × 2 repeats = 36 responses

Follow-up: 10 × 2 models × 2 prompts × 2 repeats = 80