First impression assessment of digital human applicant images generated with posture prompts and text prompts

Wataru Ganaha, Shota Hioki, Masashi Nishiyama

Tottori University (Japan)

INTERNATIONAL CONFERENCE ON DIGITAL HUMAN MODELING AND APPLICATIONS IN HEALTH, SAFETY, ERGONOMICS AND RISK MANAGEMENT (DHM), HCII2025

鳥取大学 Tottori University

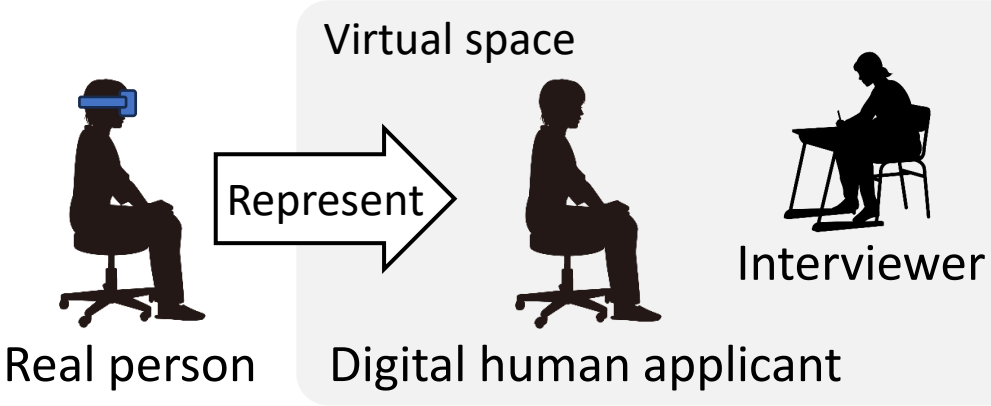
Introduction (1/2)

The use of digital humans in virtual interviews has attracted attention in recent years.

- Digital human: an entity in a virtual space that realistically imitates a real person's appearance and behavior.
- The use of digital humans is designed to help interactions in virtual interviews feel natural.

Like interviews in real space, interviews in virtual space generally involve an **applicant** and an **interviewer**.

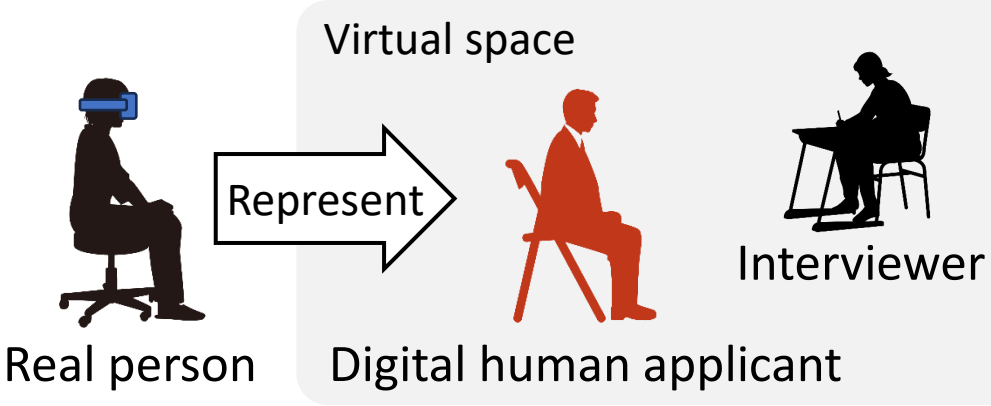
[J. Autism Dev. Disord, 2014]
[Res. Autism Spectr. Disord, 2020]



Recent analytical study [CSCW Companion, 2023]

Interviewer interacted in virtual space with digital human applicants differing in physical characteristics.

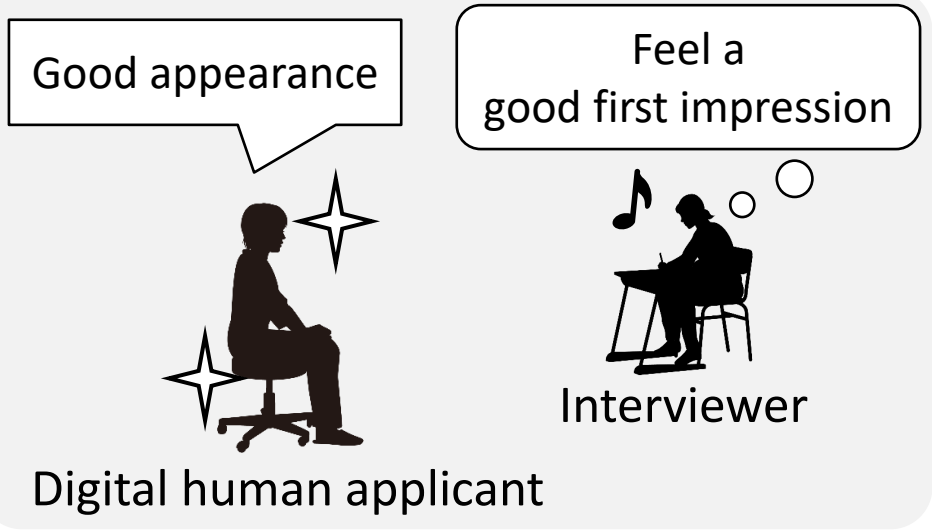
As in this study, we examined a case in which the appearance of a digital human applicant differed from that of the real person it represented.



Introduction (2/2)

An interviewer's **first impression** of an applicant is often an essential factor influencing their decision-making.

The **appearance of a digital human applicant** may significantly influence the interviewer's **first impression**.

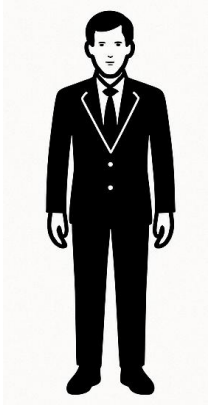


Digital human generation techniques:

- Typically, designer-crafted 3D models are used.
- Recently, services like MetaHuman and Avatar Cloud Engine create highly realistic digital humans.

https://www.unrealengine.com/ja/metahuman https://developer.nvidia.com/ace

In virtual space interviews, it is often preferable for the digital human applicant to resemble a real-world human in **whole-body** appearance.



However, current services mainly focus on facial and hair realism, generating **upper-body** representations.

Generating digital human applicants

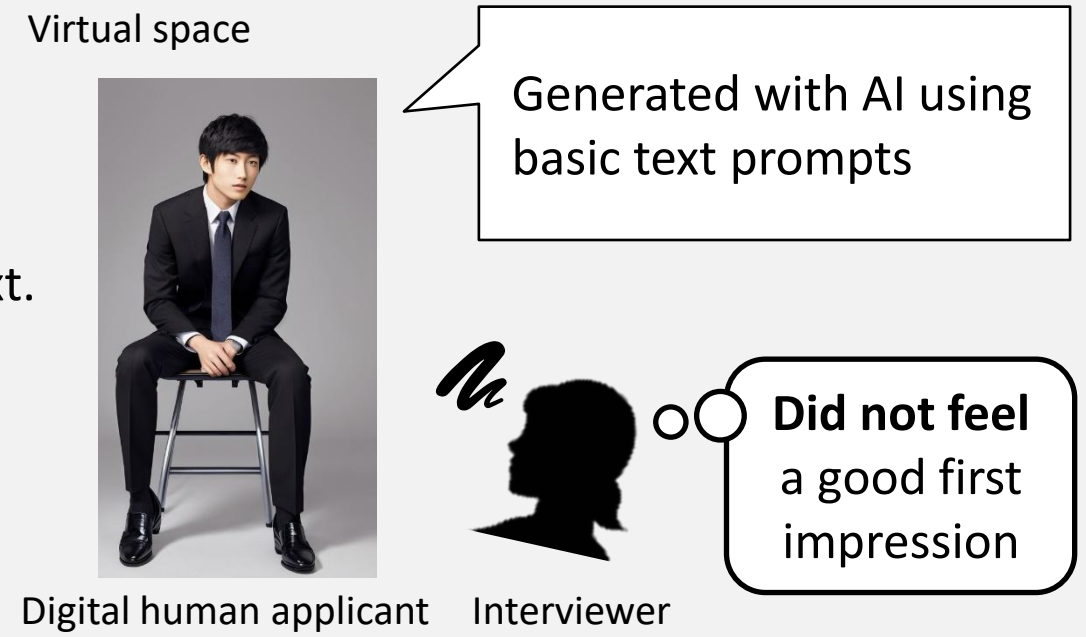
We considered the use of AI-based image generation such as Stable Diffusion [CVPR, 2022] to generate the whole-body appearance of digital human applicants.

When using AI-based image generators, **text prompts** are generally employed.

Basic text prompts:
Used to define appearance, clothing, posture, and context.

Additional text prompts:
Used to adjust generated images.

However, controlling posture with text prompts alone is difficult and time-consuming.



We aimed to generate posture-controlled images by providing **posture prompts** using ControlNet. [ICCV, 2023]

To our knowledge, no studies have explored whether combining posture and text prompts can improve the first impression of a digital human applicant in an interview setting.

Purpose

We investigate whether interviewers perceived a good first impression from digital human applicants generated by the combination of posture prompts and additional text prompts.

Virtual space

Digital human applicant

Generated with AI combining
- Basic text prompts
- **Posture prompt**
- **Additional text prompts**

Felt
a good first impression

Interviewer
(Participant acting as interviewer)

Hypothesis:

An interviewer perceives a good first impression from an applicant when it is generated by a combination of posture prompts that control the applicant's posture and additional text prompts that modify the generated results.

We generated digital human images with varying prompts and conducted a subjective assessment where participants rated their first impressions as good or bad.

Experimental design for subjective assessment

In preparing the stimulus images, we set the conditions for combining prompts to generate digital humans.


C1:
w/o posture prompts
w/o additional text prompts


C2:
w/ posture prompts and
w/o additional text prompts


C3:
w/o posture prompts
w/ additional text prompts


C4:
w/ posture prompts
w/ additional text prompts

- AI-based image generator: Stable Diffusion with BeautifulRealisticAsians
  [CVPR, 2022] https://civitai.com/models/25494/beautiful-realistic-asians
- The total number of stimulus images presented to each participant:
  6 images × 2 genders × 4 conditions = 48 images.

Prompts for generating stimulus images

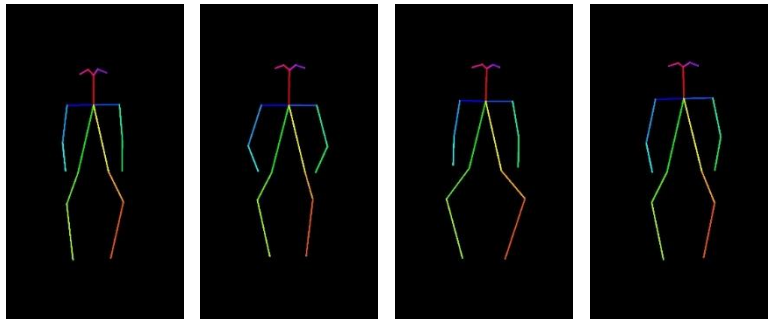
Text prompts for Stable Diffusion

Basic text prompts (Male applicant)	full-body, best-quality, realistic, a-20-year-old-Japanese-man, wearing-black-business-suit, derby-tie, seated-posture-on-a-folding-chair, no-background
Basic text prompts (Female applicant)	full-body, best-quality, realistic, a-20-year-old-Japanese-woman, wearing-black-business-suit, seated-posture-on-a-folding-chair, no-background
Additional text prompts	smile, hands-on-lap, look-at-viewer

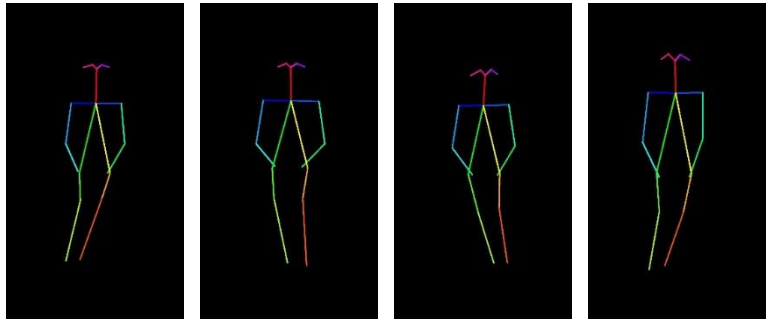
- Basic prompts included words describing appearance, clothing, posture, and context.
- Additional prompts specified a smile, hands on knees, and a forward gaze to reduce variability in expression and posture.

Posture prompt for ControlNet

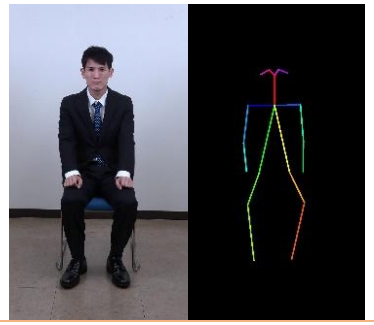
Posture prompt (Male applicant)

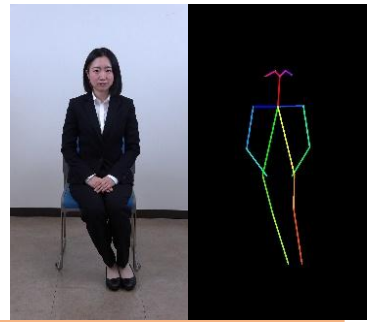


Posture prompt (Female applicant)



Posture was estimated from the images and used as prompts.



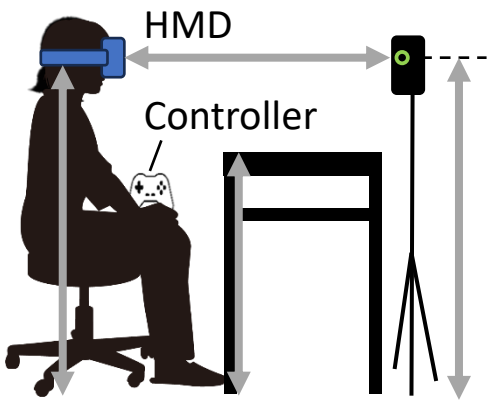


OpenPose was used for estimate.
[PAMI, 2021]

Setting

15 Japanese students (14 male and 1 female, average age of 23.9±2.8 years) participated.

We used an HMD (VIVE Pro Eye) to construct a virtual space simulating an interview.



Question:

Q1: Do you feel a **good** first impression of the applicant?
Q2: Do you feel a **bad** first impression of the applicant?

Participants responded on a 5-point scale
(1: no, 2: likely no, 3: neutral, 4: likely yes, 5: yes).

Experimental procedures:

P1: The questions and the method of answering were explained to the participants.

P2: The reference image was displayed for 2 seconds.

P3: The stimulus image was displayed for 9 seconds.

P4: The participants were asked to answer questions Q1 and Q2.

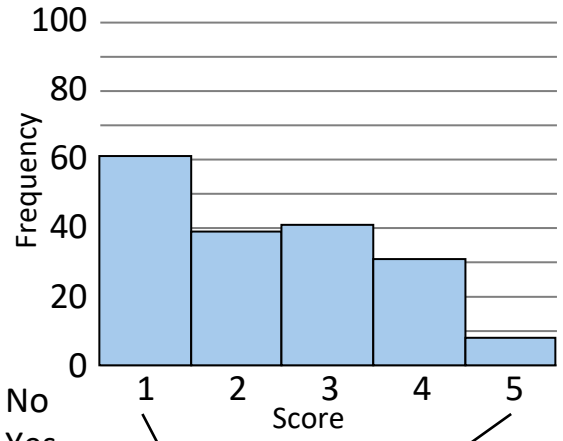
P5: The procedures P2 to P4 were repeated until the participants completed their responses for all 48 stimulus images.

[2x]

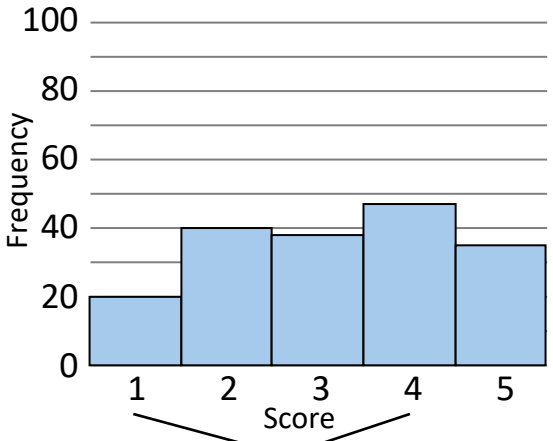
Histograms of Q1 subjective scores (Good first impression)

The high Q1 subjective score indicates a **good** first impression.

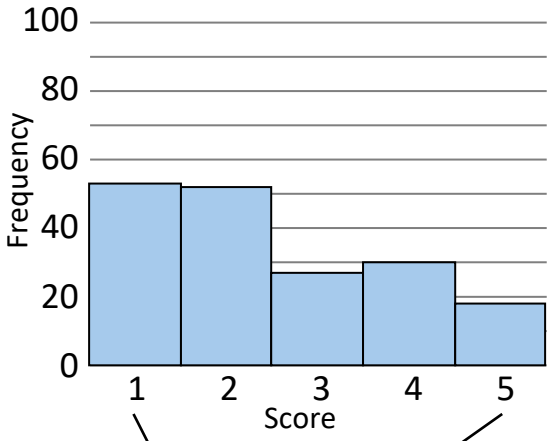
C1: w/o posture prompts
w/o additional text prompts



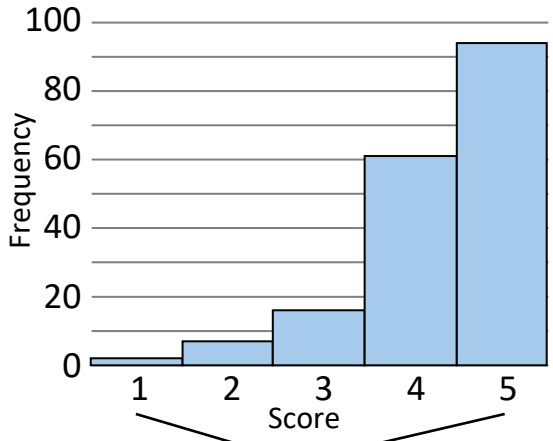
C2: w/ posture prompts and
w/o additional text prompts



C3: w/o posture prompts
w/ additional text prompts



C4: w/ posture prompts
w/ additional text prompts



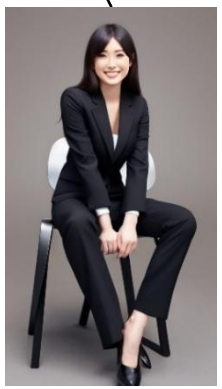
1 = No
5 = Yes

Frequent

Rare

Frequent

Rare



Frequent

Rare

Frequent

Rare

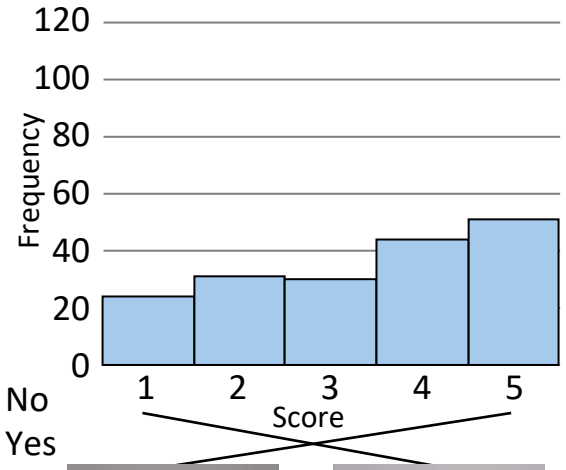
Images belonging to the group of each score.

We can see that images belonging to the high Q1 subjective score makes a **good** first impression.

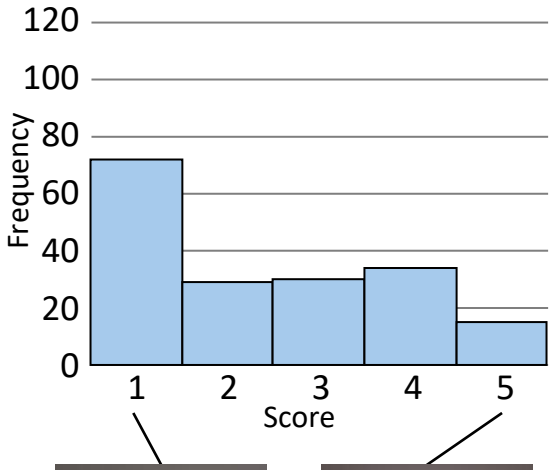
Histograms of Q2 subjective scores (Bad first impression)

The high Q2 subjective score indicates a **bad** first impression.

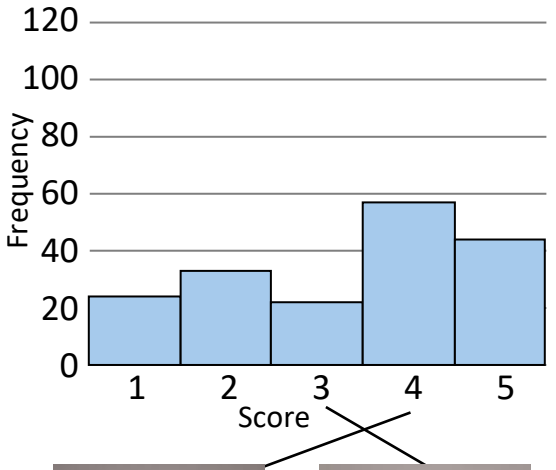
C1: w/o posture prompts
w/o additional text prompts



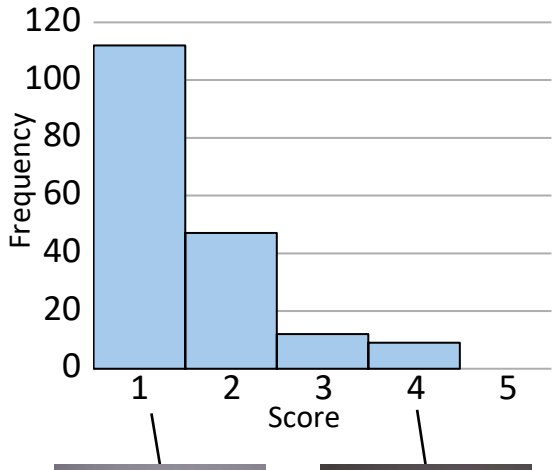
C2: w/ posture prompts and
w/o additional text prompts



C3: w/o posture prompts
w/ additional text prompts



C4: w/ posture prompts
w/ additional text prompts



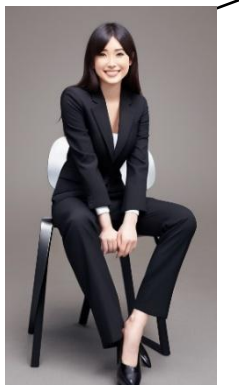
1 = No
5 = Yes

Frequent

Rare

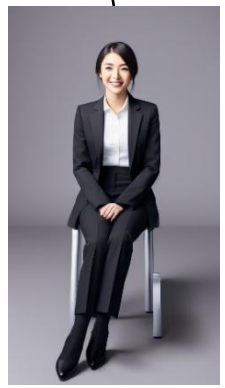
Frequent

Rare



Frequent

Rare



Frequent

Rare

Images belonging to the group of each score.

We can see that images belonging to the high Q2 subjective score makes a **bad** first impression.

Two-way ANOVA

We applied the aligned rank transform to the subjective scores from questions Q1 and Q2, and conducted a two-way ANOVA with repeated measures. [HFCS, 2011]

Question	Factor	F value	p value	Main effect	Interaction
Q1 (**Good** first impression)	Posture prompts	255.69	$< .001$	**Present**	-
	Additional text prompts	43.38	$< .001$	**Present**	-
	Posture prompts × Additional text prompts	35.48	$< .001$	-	**Present**
Q2 (**Bad** first impression)	Posture prompts	247.25	$< .001$	**Present**	-
	Additional text prompts	15.41	$< .001$	**Present**	-
	Posture prompts × Additional text prompts	13.75	$< .001$	-	**Present**

Interaction effects were observed for both Q1 and Q2, indicating that the combination of posture prompts and additional text prompts significantly changed the interviewers' first impression of the digital human applicants.

The simple main effect

We conducted the Wilcoxon signed-rank test as a test for simple main effects.

Posture prompts

Question	Factor	p value	Simple main effect
Q1 (**Good** first impression)	w/o additional text prompts	$< .05$	**Present**
	w/ additional text prompts	$< .05$	**Present**
Q2 (**Bad** first impression)	w/o additional text prompts	$< .05$	**Present**
	w/ additional text prompts	$< .05$	**Present**

Additional text prompts

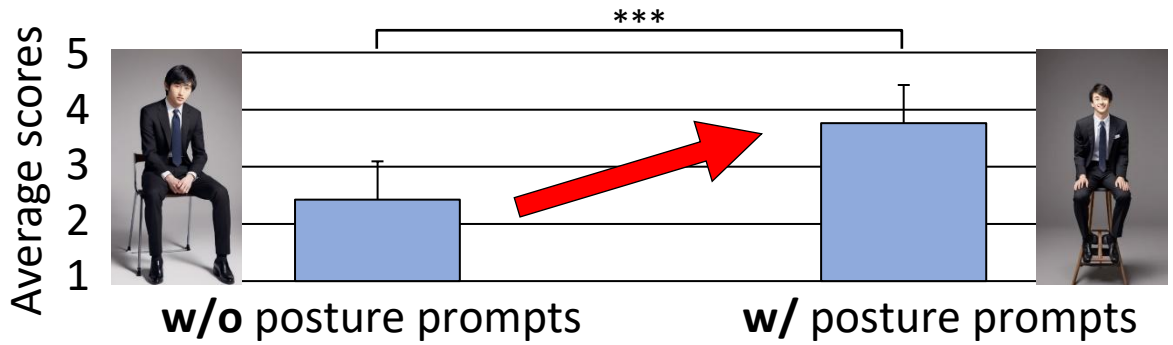
Question	Factor	p value	Simple main effect
Q1 (**Good** first impression)	w/o posture prompts	$\geq .05$	Absent
	w/ posture prompts	$< .05$	**Present**
Q2 (**Bad** first impression)	w/o posture prompts	$\geq .05$	Absent
	w/ posture prompts	$< .05$	**Present**

For Q1, a simple main effect was observed with the posture prompt, but not without it. The same pattern was observed in Q2.

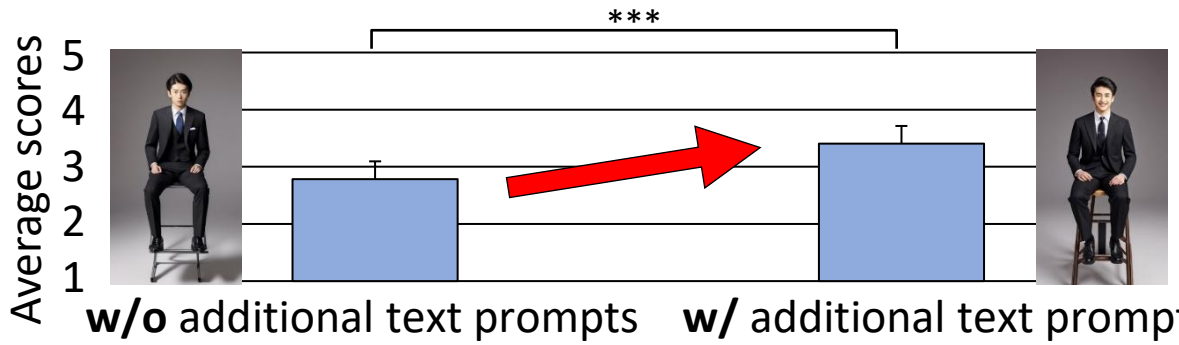
Average subjective scores

Wilcoxon signed-rank test

*** : Simple main effect ($p < .001$)

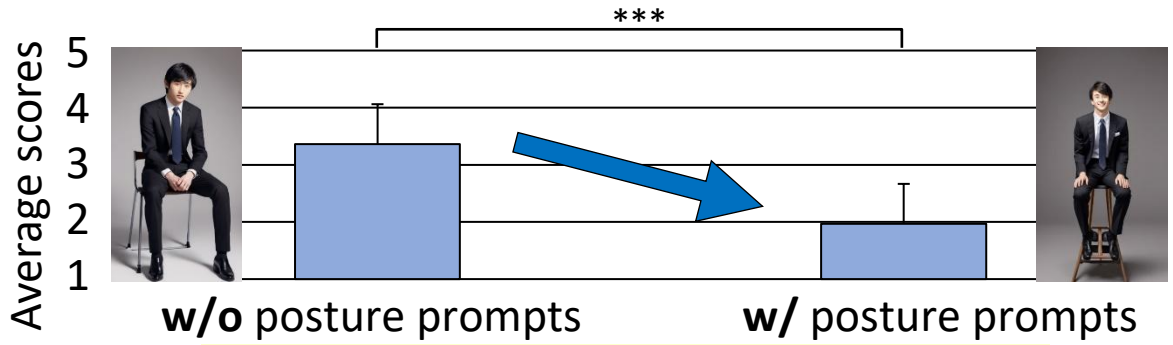


Higher score with prompts

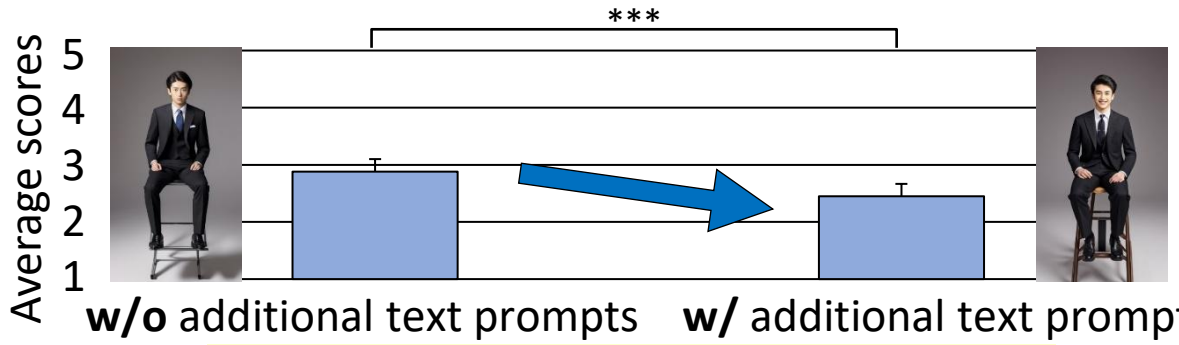


Higher score with prompts

Q1: Good first impression



Lower score with prompts



Lower score with prompts

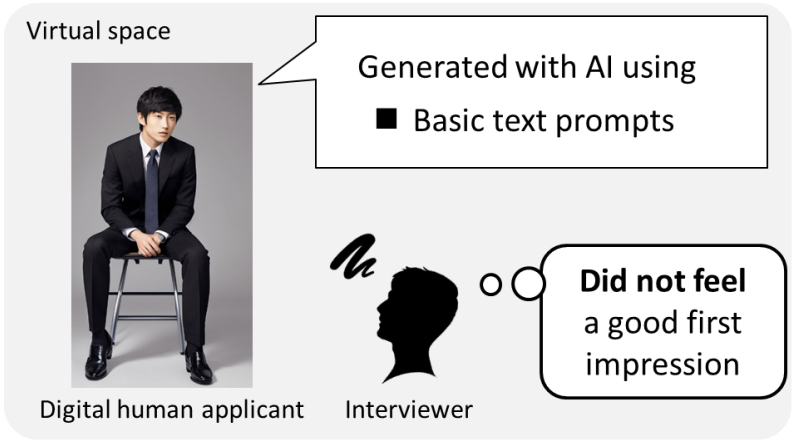
Q2: Bad first impression

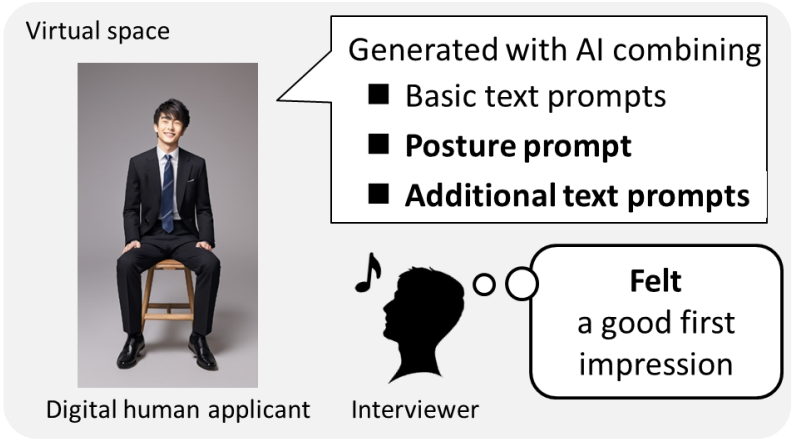
When both prompts were used, the average subjective score for Q1 was higher and the average subjective score for Q2 was lower. This suggests that using posture prompts and additional text prompts improved interviewers' first impressions of the digital human applicants.

Conclusions

Purpose:

Evaluate the first impressions of digital human applicants generated using different prompt combinations.





Results:

- **Combining posture and additional text prompts significantly improved first impressions** of digital human applicants.
- **Posture prompts alone were more effective** than additional text prompts alone in enhancing perceived impressions.
- **The absence of posture control often led to poor impressions**, even when additional text prompts were used.
- **Statistical analyses (two-way ANOVA and Wilcoxon test)** confirmed that the interaction of prompts had a strong effect on perceived first impressions.

Future work:

Identify which specific postures and textual expressions most strongly influence perceived first impressions.