Investigating Mutual Gaze in Real-Time LLM-Driven Avatar Interaction on 2D Screens

HEIKO DREWES, LMU Munich, Germany
ABDULLATIF DINC, LMU Munich, Germany
FLORIAN ALT, LMU Munich, Germany

With the growing presence of AI-driven avatars in everyday applications, understanding how users perceive and respond to such agents becomes increasingly important. A key open question is whether gaze awareness, an avatar's ability to react to the gaze of a user, contributes to more natural and socially engaging interaction, particularly on conventional 2D screens. We present a real-time system that integrates live speech recognition, LLM-generated responses, and gaze feedback to enable naturalistic unscripted conversations with a virtual avatar. We conducted a within-subjects study (N = 17) comparing a gaze-aware avatar with a nonreactive baseline. The results of the eye-tracking show that gaze awareness significantly altered user eye movement patterns, increasing mutual gaze episodes, and generating subconscious gaze regulation. Our findings demonstrate that realistic gaze behavior affects user interaction on a subconscious level, even outside immersive environments. We discuss implications for designing future conversational AI systems that foster natural social presence.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**;

Additional Key Words and Phrases: gaze awareness, eye contact, avatar interaction, avatar design space, gaze response model

ACM Reference Format:
Heiko Drewes, Abdullatif Dinc, and Florian Alt. 2026. Investigating Mutual Gaze in Real-Time LLM-Driven Avatar Interaction on 2D Screens. *Proc. ACM Hum.-Comput. Interact.* 10, 3, Article ETRA005 (May 2026), 16 pages. https://doi.org/10.1145/3806019

1 Introduction

The rapid development of artificial intelligence (AI) technologies has enabled avatars to converse in ways increasingly similar to those of a human dialog. Text-to-speech, speech-to-text, and LLM-based dialog systems have reached a level of quality where virtual characters can engage users in fluid context-aware conversation [Allbert et al. 2025; Mahmood et al. 2025]. Such avatars are already entering diverse domains, from customer service chatbots and virtual tutors to healthcare assistants and companionship applications, where they represent the visible, social front-end of AI systems. Despite these advances, most avatars still lack one of the most fundamental aspects of human interaction: realistic eye contact. Eye contact signals attention, empathy, and trust, and its absence can make even sophisticated avatars appear detached or uncanny [Kullmann et al. 2025].

Previous research demonstrated that gaze-aware avatars can increase perceived social presence and naturalness of interaction [Bee et al. 2010b; Garau et al. 2001; Wang and Gratch 2010]. Yet, most of this work focused on highly controlled or immersive settings, such as virtual reality (VR).

Authors' Contact Information: Heiko Drewes, LMU Munich, Munich, Germany, heiko.drewes@ifi.lmu.de; Abdullatif Dinc, LMU Munich, Munich, Germany, dinc.abdullatif@campus.lmu.de; Florian Alt, LMU Munich, Munich, Germany, florian.alt@ifi.lmu.de.

This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 2573-0142/2026/5-ARTETRA005
https://doi.org/10.1145/3806019

At the same time, looking at the sales numbers for 2D and stereoscopic displays, 2D screens still dominate everyday interaction. As there is current research on the generation of 2D avatars [Pham et al. 2024], we expect a considerable part of AI interactions will occur on conventional 2D screens in the next years. Here, eye contact is harder to reproduce due to the Mona Lisa effect – the illusion that the gaze of a face on a picture follows the viewer regardless of the viewing angle [Boyarskaya et al. 2015]. Moreover, 2D interfaces typically lack integrated eye trackers, and webcam-based eye trackers are not accurate enough (yet) to detect eye contact. This raises the question whether the benefits of gaze-aware avatars are worth the cost of additional hardware. Finally, the emergence of LLMs introduces a new level of conversational realism, allowing unscripted interaction with AI agents. Yet, it remains unclear how users respond to gaze-aware behavior when the dialog itself is live and unpredictable.

We explore whether gaze awareness in an AI-driven avatar on a 2D screen affects the user's perception and eye movement behavior. We developed a system that integrates live speech recognition, LLM-generated responses, and real-time gaze feedback, allowing participants to interact with the avatar in a natural conversation. In a within-subjects study (N = 17), participants interacted with a gaze-aware and a non-gaze-aware avatar while answering quiz questions. During the interaction, we recorded both the avatar's and the user's gaze. Our results show that gaze awareness significantly influences users' eye movement patterns even though most participants did not notice this consciously.

For the research community, our work highlights that gaze remains a powerful social cue in human–AI communication [Klein 2025; Koller et al. 2023], even outside immersive settings. It also provides a practical foundation for studying such effects: our real-time LLM-integrated system demonstrates how to synchronize gaze data, speech, and dialog to capture natural behavior. This methodological approach can support future research on adaptive avatars, conversational agents, and affective computing systems.

2 Related Work

2.1 Avatars and Social Presence

Avatars are digital representations of people or agents that mediate social interaction in virtual environments. Garau et al. [2001] showed that the type of gaze behavior of a humanoid avatar significantly influences communication quality. They found that eye movements realistically adapted to the conversation partner and context lead to more positive user ratings than random or uncoordinated movements. Similarly, Nowak and Biocca [2003] demonstrated that a human-like appearance increases social presence – the feeling of encountering a "real other".

2.2 Gaze Behavior in Avatars

Eye movements are among the most powerful non-verbal signals in human communication. Early work by Lee et al. [2002] developed statistical models for when a virtual avatar should look at its conversation partner or avert its gaze. Al Moubayed et al. [2012] later examined the *Mona Lisa effect* – the illusion that a face on a flat screen appears to look back at the viewer regardless of viewpoint – and proposed technical adjustments to mitigate it.

Recently, the generation of eye movements moved from rule-based to data-driven approaches. Canales et al. [2023] trained ML models to synthesize realistic gaze and head movements based on recorded human conversations. Dembinsky et al. [2024] further applied generative adversarial networks (GANs) to create lifelike gaze patterns. These methods demonstrate that plausible eye behavior enhances perceived realism and engagement. In contrast, unnatural gaze, such as rigid staring or complete lack of eye contact, can lead to discomfort and the *uncanny valley* effect [Izumi et al. 2025].

Research on reactive avatars for single users reinforced the importance of gaze. Garau et al. [2001] found that avatars with gaze behavior tailored to a conversation significantly increase communication quality compared to random gaze. Kipp and Gebhard [2008] introduced *IGaze*, an immersive system in which the avatar's reactive eye and head movements followed different strategies. Bee et al. [2010a,b] confirmed that such responsiveness increases social presence and that users unconsciously treat gaze-aware avatars as genuine partners. AnimeGaze provides technical solutions to minimize the Mona Lisa effect and achieve convincing mutual gaze from any viewing angle [Izumi et al. 2025].

2.3 Social-Psychological Foundations of Eye Contact

Social psychology provides a theoretical basis for designing believable gaze behavior in avatars. Argyle and Dean [1965]'s theory of *intimacy balance* proposes that people regulate eye contact and interpersonal distance to maintain comfortable levels of familiarity. Too much eye contact can feel intrusive, while too little reduces engagement. Kendon [1967] examined how speakers use gaze during turn-taking: they typically avert their eyes when speaking freely and reestablish contact to signal the end of a turn. Goffman [1963] described *civil inattention* as the brief eye contact that strangers make to acknowledge each other's presence before deliberately looking away to avoid interaction. Rogers et al. [2018] found evidence of within-participant consistency for the duration of fixations to the eye region. Maran et al. [2020] showed that more social gazing occurs in groups. These findings reveal that human gaze behavior follows subtle social rules that operate largely below conscious awareness. Garau et al. [2003] showed that avatars emulating such behavior - looking away when "thinking" and seeking eye contact when "listening" - are perceived as more natural.

2.4 AI-Driven Avatars and Mutual Gaze

Recent advances in AI have led to a convergence of language models and embodied avatars. These systems can interact with users not only via text or speech, but also through gestures, facial expressions, and gaze. Despite impressive verbal fluency, current AI models remain limited in their ability to convey authentic eye contact. Izumi et al. [2025] noted that while large language models excel at conversation, "AI is still unable to look users in the eye". To address this, their AnimeGaze system introduced a hardware setup in which a camera is positioned behind a semi-transparent display, aligning the virtual eyes with the optical path to the user, thereby enabling genuine mutual gaze.

At the software level, Cassell et al. [2000] pioneered the concept of *embodied conversational agents*, integrating speech with non-verbal behavior models as early as 2000. Building on this idea, Park et al. [2023] presented *generative agents* that simulate daily routines and social interactions within virtual environments. Fink et al. [2024] reviewed educational and social applications of AI avatars, emphasizing their potential to evoke positive emotions and engagement through synchronized speech, expression, and motion.

3 Research Approach

3.1 Research Questions

The goal of this research is to understand how *gaze awareness* in an AI-driven avatar influences the users' perception and eye movement behavior during natural interaction on a 2D screen.

Characteristics of an avatar include visual and behavioral aspects such as gender, voice, appearance, body, lip, and eye movements, and mimics [Rashik et al. 2024], leading to many possible combinations. In addition, the *context* of interaction, both spatial and social, can shape how users experience the avatar. Spatial context refers to viewing distance and display size, while social context includes the purpose of the interaction (e.g., professional assistant, conversational partner, or virtual companion) and whether bystanders are present. Each of these factors contributes to users' gaze behavior, making it essential to isolate the role of gaze awareness itself. Our research focuses on disentangling the specific influence of gaze awareness from other properties of an avatar that may affect user perception.

Avatars may also vary in the types of user behavior they recognize and respond to, such as mimics, body movements, voice tone, or gaze direction. Technically, such forms of awareness are now feasible through facial recognition, emotion detection, and eye tracking. However, it remains an open question how avatars should react to these signals. Ma et al. [2022] raised this question about emotional responses in voice assistants and Bee et al. [2010b] examined gaze behavior. The high dimensionality of this design space makes it difficult to identify which aspects most strongly influence perceived naturalness and social presence. With all these uncertainties and without reliable effect-size estimates or variability measures from the literature that transfer to our specific, real-time, LLM-driven 2D avatar setting, we designed an exploratory study.

Our work takes a focused approach to one dimension of avatar awareness — *gaze awareness* — and investigates its influence while controlling other factors. Specifically, we ask:

> **RQ1:** How does gaze awareness of an avatar on a 2D screen influence users' eye movement behavior during interaction?
>
> **RQ2:** How does gaze awareness affect the users' satisfaction with the interaction and the users' subjective perception of the avatar's interest, awareness of presence, naturalness, and functionality?

We conducted a comparative within-subject study in which participants interacted with a gaze-aware and a non-gaze-aware version of the same avatar. The study is designed to capture both objective measures (eye tracking data) and subjective impressions (questionnaire responses).

3.2 Avatar and Interface

For a natural conversational experience, the avatar was designed to resemble a human without entering the *uncanny valley* [Mori et al. 2012]. Extremely realistic avatars may appear unsettling, while overly stylized or cartoon-like characters can reduce perceived authenticity. We used an established avatar model from previous research [Izumi et al. 2025], implemented via *ChatVRM*, an open-source browser-based project for VRM avatars[1]. This choice ensures visual comparability with existing work and enables reproducibility. Participants could select either a male or a female avatar with a corresponding gender-specific synthetic voice. Providing only one gendered avatar would likely introduce avoidable gender-related discomfort or discussion, while an androgynous avatar would not align with our goal of a human-like conversational setting. There is a debate on the gender of voice assistants [West et al. 2019], and although it is an interesting topic, it was not the focus of this study. The avatar displayed subtle idle body movements and lip synchronization during speech, but no facial mimics, to avoid introducing uncontrolled emotional cues. Idle eye movements were maintained to preserve realism [Canales et al. 2023].

[1]https://github.com/pixiv/ChatVRM

Fig. 1. **Avatars used in the study.** Participants could choose between a female (left) and a male (right) avatar implemented via ChatVRM. Both avatars included subtle idle body motion and lip synchronization but no facial mimic to ensure experimental control.

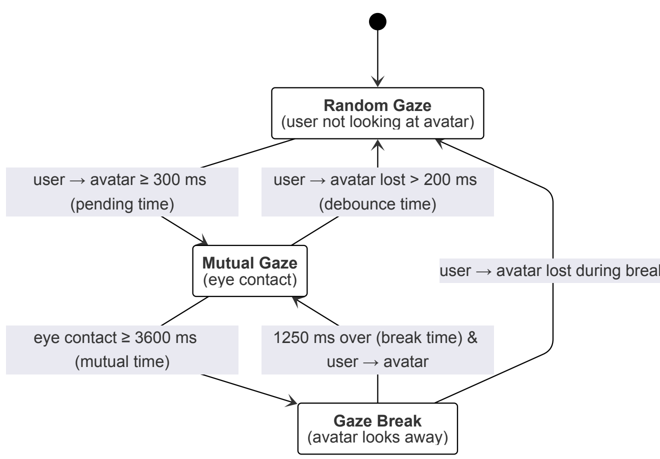

Fig. 2. **Finite-state model controlling the avatar's gaze-aware behavior.** The model alternates between random gaze, mutual gaze, and gaze-break states based on real-time eye-tracking data from the user.

3.3 Gaze Awareness Model

Gaze awareness enables the avatar to sense the user's gaze direction and adapt its own eye movements accordingly. Human eye contact is governed by complex social norms. Direct gaze typically signals attention or engagement, but prolonged eye contact can also imply dominance or intimacy [Eibl-Eibesfeldt 1997]. Mutual eye-contact even impacts mate choice [Hoffmann et al. 2024].

We implemented a finite-state model that controls the avatar's eye behavior based on user gaze input. The model comprises three states: *random gaze*, *mutual gaze*, and *gaze break*. In the random gaze state, the avatar performs small, naturalistic saccades unrelated to the user's gaze. When the system detects that the user is looking at the avatar, it transitions to the mutual gaze state, establishing virtual eye contact. After a short interval, the model triggers a gaze break in which the avatar averts its eyes to mimic typical human gaze behavior. This cyclic process produces short, responsive gaze interactions that appear natural without being intrusive. Figure 2 illustrates the state transitions.

3.4 Context of Interaction

The spatial context refers to the visual composition of the scene, including the size of the avatar and the distance to the user. The avatar was shown from the waist up, providing a balanced view between close-up facial detail and body posture. To avoid distraction, we used a static background free of salient and moving elements, or readable text (see Figure 3). We chose a non-blank, static background as a conservative design choice to avoid potential attentional anchoring on the avatar. The background was the same for both avatars.

Fig. 3. **Avatar and visual context.** The female avatar is shown from the waist up against a static background without salient details to minimize visual distraction while maintaining ecological validity.

The social context describes the purpose and content of the interaction. To maintain a structured, yet conversational setting, the avatar assumed the role of a quiz master asking general knowledge questions from an established general knowledge standard by Nelson and Narens [1980] and covers various areas of knowledge. This setup introduced a light cognitive load on the participants, diverting their attention away from the avatar. In contrast to previous work, that often relied on more controlled setups and post-hoc ratings (including gaze-perception tasks) [Garau et al. 2001; Moubayed et al. 2012; Nowak and Biocca 2003], this task created a more natural turn-taking dynamic.

3.5 Technical Implementation

The system architecture integrates multiple components for real-time and multi-modal interaction (see Figure 4). A microphone records the participant's speech, which is transcribed using a speech-to-text (STT) module and passed to the LLM for dialog generation. The LLM output is then synthesized to speech (TTS) and rendered by the avatar in synchrony with lip movements.

A Flask server[2] coordinates communication between components, including the avatar front end (running in an Electron shell[3]), the OpenAI real-time API[4], and a ZeroMQ[5] bridge connected to a Tobii 4C eye tracker, which continuously reports gaze coordinates, used to update the avatar's gaze state in real time.

This implementation allows controlled experiments on gaze behavior while preserving the natural flow of conversation. The architecture can be reproduced or extended for future studies investigating multi-modal AI interaction.

[2]https://flask.palletsprojects.com
[3]https://www.electronjs.org
[4]https://openai.com/index/introducing-the-realtime-api
[5]https://zeromq.org/

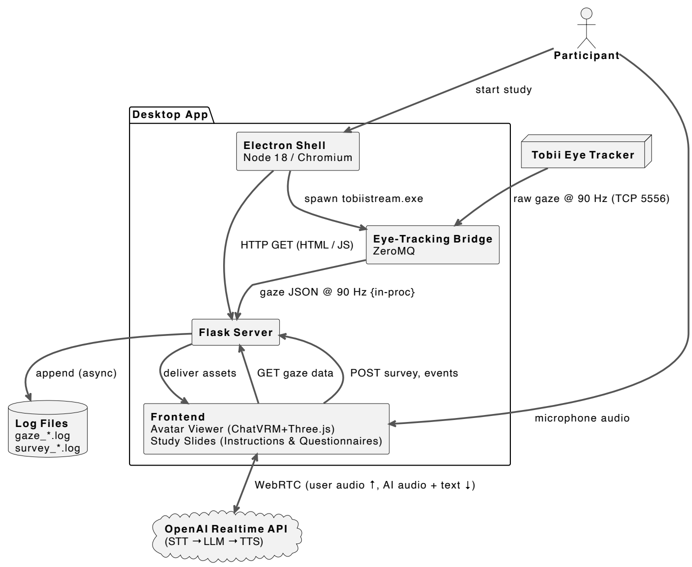

Fig. 4. **System architecture of the real-time, gaze-aware avatar.** The system combines speech-to-text (STT), large language model (LLM) dialog generation, and text-to-speech (TTS) output with continuous eye-tracking feedback. A central Flask server coordinates communication between modules, ensuring low-latency interaction.

4 User Study

4.1 Study Design

The study employed a within-subject comparative design with two experimental conditions: *(1) a gaze-aware avatar* and *(2) a non-gaze-aware avatar* serving as baseline. The within-subjects design minimized inter-individual differences and allowed participants to directly compare both conditions. As both rounds showed the same avatar and background, the condition order was counterbalanced among the participants to mitigate the effects of learning or fatigue.

The **independent variable** was *gaze awareness* (two levels: gaze-aware vs. non-gaze-aware). The **dependent variables** comprised subjective and objective measures:

- **Subjective measures:** Participants' perceptions of avatar interest, awareness, naturalness, and general satisfaction, as well as their comparative preference for one version.
- **Objective measures:** Eye-tracking metrics describing user and avatar gaze behavior, including proportion of gaze time, mean gaze episode duration, episode rate, mutual gaze frequency, and gaze aversion patterns.

Participants interacted with the avatar in two quiz rounds, one per condition, while their gaze and the avatar's gaze direction were recorded. After each round, participants evaluated the avatar by completing a short questionnaire and a comparative questionnaire after both rounds.

4.2 Apparatus

All sessions were conducted on a 16-inch laptop with a Tobii 4C eye tracker mounted below the display. Participants sat approximately 60–70 cm from the screen in a dimly lit room to ensure reliable eye tracking performance. The audio was played through the laptop speakers and a microphone recorded the participants' speech. The system used the architecture described in Section 3.5.

4.3 Participants

Participants were recruited through university mailing lists and personal contacts. They volunteered without financial compensation and provided informed consent in accordance with institutional ethics guidelines. Participants received written and verbal information on the study, data privacy, and their right to withdraw at any time.

4.4 Procedure

Each session lasted approximately 25 minutes and followed four phases: *introduction and setup*, *round 1*, *round 2*, and *wrap-up*. Participants completed a short demographic form and calibrated the Tobii 4C eye tracker using a standard five-point calibration. Participants were briefed that the quiz would not assess their knowledge, but would serve to study natural interaction with an avatar. The eye-tracker calibration made it apparent that we were recording eye tracking data. However, we did not tell the participants that we manipulated the avatar's gaze. After calibration, they selected either the male or the female avatar based on personal preference.

In each round, the avatar acted as a quiz master, asking general knowledge questions generated by the LLM-based dialog system. The participants answered verbally, and the avatar responded with synthesized speech and corresponding gaze behavior depending on the assigned condition (gaze-aware or baseline).

All participants rated five aspects, interest (Q1), presence (Q2), naturalness (Q3), satisfaction (Q4), and functionality (Q5), of each interaction on a 5-point Likert scale ('1=totally disagree' to '5=totally agree') immediately after each round.

(1) The avatar appeared interested in me.
(2) The avatar appeared to be aware of my presence.
(3) The avatar's gaze behavior appeared natural.
(4) I was satisfied with the interaction overall.
(5) The conversation between me and the avatar worked properly.

The within-subject design allows for comparative questions. After both rounds, we asked the participants: *(1) Which version did you find more natural?* and *(2) Which version would you prefer to use in the future?*

4.5 Limitations

The study was conducted in a controlled laboratory setting to ensure stable eye tracking and consistent system performance. Although this environment may not capture all nuances of everyday interaction, it was essential to isolate the effects of gaze awareness under reproducible conditions.

The avatar's behavior was intentionally simplified – without facial expressions, gestures, or emotional prosody – to control confounding influences. This reduction focused the experiment on gaze behavior, ensuring that observed effects could be attributed specifically to gaze awareness rather than to multimodal expressivity.

Minor timing variations may occur in the real-time speech synthesis pipeline; however, latency measurements and participant feedback indicated that conversational flow remained natural. These controlled conditions represent standard trade-offs in HCI experimentation and do not affect the validity of the reported findings.

The study was conducted with only 17 participants. Hence, the results should be interpreted with care. There are many deliberate design decisions in the study, and their effects are not yet fully understood. Interaction with avatars is an ongoing area of research, and the presented work provides valuable directions for future research.

5 Results

5.1 Participants

Seventeen volunteers (11 male, 6 female) participated in the study. Their ages ranged from 18 to 54 years ($M = 30.5$, $SD = 9.3$), approximately half between 26 and 32 years; two participants were older than 50 years. Two participants reported using corrective lenses. Five of the six female participants chose the female avatar and eight of the eleven male participants chose the male avatar. There was a clear preference for an avatar of the own gender. However, such preferences may depend on cultures and this finding should not be generalized.

5.2 Eye-Tracking Results

Eye-tracking data were preprocessed using Python scripts and aggregated per participant and condition. Mutual gaze episodes were detected based on spatial and temporal overlap between the user and avatar gaze vectors ($8.0^{\circ} x\, 5.0^{\circ}$ for the user and $3.9^{\circ} \times 1.7^{\circ}$ for the avatar). For each condition, we computed:

- **Proportion of gaze time:** fraction of total interaction time spent looking at the avatar.
- **Mean episode duration:** average lengths of continuous gaze episodes.
- **Episode rate:** number of gaze episodes per second.
- **Mutual gaze time:** proportion of time when user and avatar looked at each other.
- **Aversion metrics:** rate and duration of instances in which users looked away during mutual gaze.

Paired-sample *t*-tests were used for eye tracking data assuming a normal distribution. Eye-tracking data were analyzed to quantify how gaze awareness influenced visual attention between the user and the avatar. We examined three interaction metrics, *user → avatar gaze*, *avatar → user gaze*, and *mutual gaze*, as well as gaze aversion behavior. Table 1 summarizes descriptive and inferential statistics.

Table 1. Objective gaze metrics in the baseline and gaze-aware conditions ($N = 17$). The table shows the proportion of time, mean episode duration, and episode rate for all gaze directions, with paired-sample *t*-tests.

Metric	Unit	Baseline		Gaze-Aware		$t(16)$	p
		M	SD	M	SD		
User → Avatar							
Proportion of Time	–	0.524	0.152	0.458	0.155	2.215	.042
Mean Episode Duration	s	0.919	0.462	0.752	0.304	2.527	.022
Episode Rate	1/s	0.671	0.335	0.694	0.379	-0.488	.632
Avatar → User							
Proportion of Time	–	0.084	0.024	0.553	0.123	-16.153	<.001
Mean Episode Duration	s	0.975	0.162	1.883	0.430	-10.842	<.001
Episode Rate	1/s	0.089	0.031	0.295	0.049	-16.356	<.001
Mutual Gaze							
Proportion of Time	–	0.033	0.022	0.306	0.103	-12.171	<.001
Mean Episode Duration	s	0.394	0.249	0.511	0.159	-1.889	.077
Episode Rate	1/s	0.088	0.047	0.627	0.277	-8.624	<.001

Participants spent slightly less time looking at the avatar (see Figure 5) when it was gaze-aware ($M = 0.46$) than in the baseline condition ($M = 0.52$; $p = .042$). Mean episode durations (see Figure 6) were also shorter ($p = .022$). This pattern suggests that participants tended to avert their gaze sooner once the avatar established eye contact, a behavior consistent with human conversational norms.

As expected from the model implementation, the avatar looked at the user more often and for longer periods in the gaze-aware condition ($p < .001$). Mutual gaze time increased sharply from 3% to 31% of the total interaction time, demonstrating that the gaze-aware avatar successfully established dynamic eye contact. Mutual gaze episodes mean duration (see Figure 6) increased slightly but not significantly ($p = .077$).

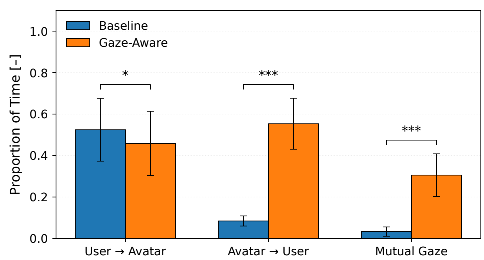

Fig. 5. **Proportion of interaction time** spent in the three gaze metrics under baseline and gaze-aware conditions. Mutual gaze and avatar-to-user gaze increased with gaze awareness. Trend marker: † p < .10. Significance markers: * p < .05, ** p < .01, *** p < .001.

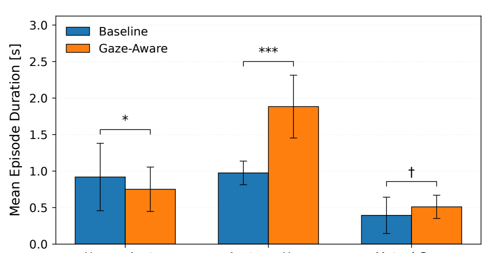

Fig. 6. **Mean gaze episode duration** for user, avatar, and mutual gaze directions. Mutual gaze episodes were slightly longer in the gaze-aware condition. Trend marker: † p < .10. Significance markers: * p < .05, ** p < .01, *** p < .001.

Aversions refer to events in which the avatar looked at the user and the user looked away, thereby breaking eye contact. The aversion rate was significantly higher under gaze-awareness conditions ($p = .003$), while the mean duration of aversion events was shorter ($p = .013$) (Table 2, Figure 8). This indicates that participants looked away more often, but for briefer moments when the avatar maintained gaze behavior consistent with natural conversational regulation of intimacy and attention.

5.3 Questionnaire Results

Questionnaire ratings between gaze-aware and baseline conditions are compared using two-tailed Wilcoxon signed-rank tests for paired samples. Descriptive statistics (mean and standard deviation) are reported for all items.

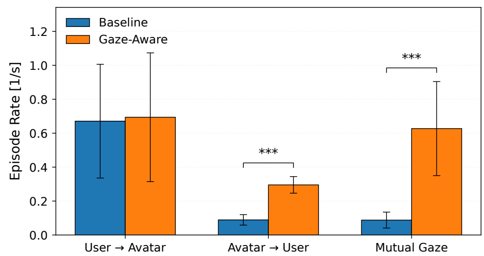

Fig. 7. **Gaze episode rate** for the three gaze metrics under baseline and gaze-aware conditions. Trend marker: † p < .10. Significance markers: * p < .05, ** p < .01, *** p < .001.

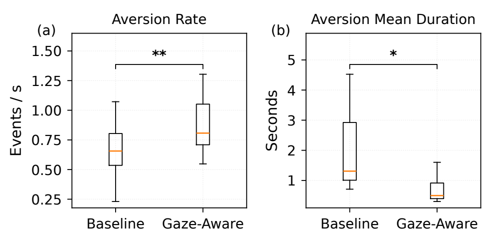

Fig. 8. **Gaze aversion metrics.** Aversion rate (left) was higher and mean duration (right) shorter in the gaze-aware condition, indicating more frequent but shorter gaze breaks. Significance markers: * p < .05, ** p < .01

Table 2. Results of paired t-tests for aversion metrics with mean values and standard deviations.

	$M \pm SD$			
Aversion Metric	Baseline	Gaze-Aware	$t(16)$	p
Rate (1/s)	0.706 ± 0.323	0.968 ± 0.362	-3.54	.003
Mean Duration (s)	2.875 ± 3.203	0.721 ± 0.420	2.80	.013

As shown in Table 3 and Figure 9, the ratings for the gaze-aware avatar were numerically higher for most of the items, although the differences did not reach statistical significance. The greatest difference was observed for perceived naturalness (Q3), with higher scores in gaze-aware condition ($M = 3.41$, $SD = 1.00$) compared to baseline ($M = 2.94$, $SD = 0.75$; $W = 13.5$, $p = .07$). Overall satisfaction (Q4) was also slightly higher for the gaze-aware avatar ($\Delta = +0.29$), while perceived functionality (Q5) was marginally better in the baseline condition ($\Delta = -0.29$). Differences in perceived interest (Q1) and presence (Q2) were minimal.

In addition to per-item ratings, participants directly compared the two versions. Several participants reported that they did not notice a difference. The participants focused most of the time on the quiz and did not pay attention to the avatar's gaze. When asked which avatar appeared more natural, eight participants (47%) preferred the gaze-aware version, six (35%) preferred the baseline, and three (18%) rated both as equally natural ($\chi^2(2) = 2.24$, $p = .33$, $\varphi = .36$). The responses to the question on future use showed the same distribution (Table 4, Figure 10).

In general, subjective ratings and preferences slightly favored the gaze-aware avatar, particularly with regard to the perceived naturalness of gaze behavior.

Table 3. Descriptive statistics ($M \pm SD$) and Wilcoxon signed-rank test results for the five Likert-scale items comparing the baseline and gaze-aware conditions. Δ indicates the mean difference (GA − BL).

Question	W	n_{eff}	p	Δ	BL $M \pm SD$	GA $M \pm SD$
Q1 Interest	12.0	7	.706	+0.06	3.65 ± 0.79	3.71 ± 0.92
Q2 Presence	10.5	7	.527	+0.12	3.47 ± 0.87	3.59 ± 0.94
Q3 Natural	13.5	11	.070	+0.47	2.94 ± 0.75	3.41 ± 1.00
Q4 Satisfac.	3.0	6	.096	+0.29	3.53 ± 1.07	3.82 ± 1.13
Q5 Funct.	3.0	6	.096	−0.29	4.41 ± 0.87	4.12 ± 0.93

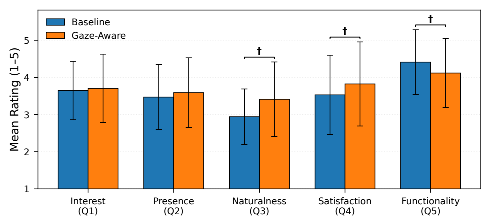

Fig. 9. **Mean Likert ratings (± SD) for items Q1–Q5.** Trend marker: † p < .10. Significance markers: * p < .05, ** p < .01, *** p < .001. Scores for the gaze-aware condition were numerically higher for most aspects, particularly naturalness (Q3).

Table 4. Participant choices in the comparative questions and corresponding chi-square test results.

Outcome	**Baseline**	**Gaze-Aware**	**Equal**	**Test**	**p**	**Effect size**
Perceived naturalness	6 (35%)	8 (47%)	3 (18%)	$\chi^2(2) = 2.24$	.33	$\varphi = .36$
Future-use choice	6 (35%)	8 (47%)	3 (18%)	$\chi^2(2) = 2.24$	.33	$\varphi = .36$

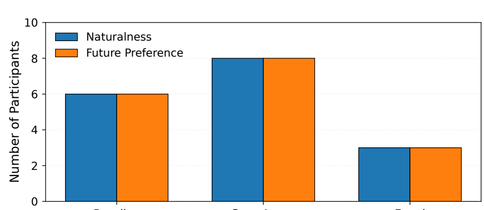

Fig. 10. **Participant preferences** for perceived naturalness and preferred future use. Slightly more participants favored the gaze-aware avatar, although differences were not statistically significant.

6 Discussion and Limitations

The study examined how gaze awareness in an AI-driven avatar influences user perception and gaze behavior during real-time interaction on a 2D screen. The results show that gaze awareness substantially increased mutual eye contact between the user and the avatar. The participants adapted their gaze patterns accordingly, showing more frequent but shorter aversions when the avatar returned their gaze. These results suggest that reactive gaze behavior affects users on a largely subconscious level, reinforcing the importance of eye contact as a social cue even in human–AI interaction.

6.1 Extending Gaze Research to Real-Time AI Interaction

Previous research has shown that avatars with responsive gaze behavior improve social presence and conversational realism in immersive or scripted environments [Bee et al. 2010b; Garau et al. 2001; Wang and Gratch 2010]. Our findings confirm several of these effects — mutual gaze increased markedly in gaze-aware condition, and participants unconsciously adapted their gaze in human-like ways. This supports long-standing models of gaze as a non-verbal regulator of intimacy and turn-taking [Argyle and Dean 1965; Kendon 1967].

At the same time, our study extends this body of work in several ways. We demonstrate that realistic social gaze effects occur even on conventional 2D screens, suggesting that dynamic eye movements can overcome limitations such as the Mona Lisa effect. Unlike previous work using pre-scripted or Wizard-of-Oz dialog [Kipp and Gebhard 2008], our system enabled unscripted, real-time interaction driven by a large language model. That comparable gaze dynamics emerged under these conditions indicates that established social mechanisms generalize to modern AI-driven avatars. Furthermore, the discrepancy between participants' self-reports and objective gaze data highlights the implicit nature of social gaze, offering a methodological cue for future research. Finally, the use of a simple finite-state model demonstrates that even lightweight gaze implementations can meaningfully affect user behavior, providing practical design insight.

6.2 Implications for AI-Driven Avatars

The behavioral data show that users adapted their gaze patterns in systematic, human-like ways when interacting with a gaze-aware avatar. Participants looked away sooner once the avatar met their gaze, mirroring intimacy regulation strategies observed in human conversation. These reactions occurred largely without conscious awareness, indicating that reactive gaze behavior triggers implicit social responses even in non-immersive, AI-mediated settings. This finding forms the basis for several implications for the design and study of gaze-aware avatars.

Even minimal gaze cues can convey attentiveness and responsiveness without complex animation. The observed subconscious responses suggest that designers can use lightweight gaze models to foster social presence in 2D applications such as digital tutors, companions, or customer service agents.

Participants subconsciously adjusted their gaze to the avatar's behavior, indicating that gaze functions as a low-level regulatory signal. Designers can use such cues to communicate turn-taking or listening states in multi-modal systems. For example, brief gaze aversions could indicate thinking pauses or attention shifts without explicit verbalization.

The confident gaze style of the avatar worked well in our neutral quiz setting, but other contexts may require more subtle or deferential behavior. System designers could parameterize gaze frequency and duration to express different personalities or to match user preferences and comfort levels.

The effects observed on 2D screens demonstrate that meaningful social gaze phenomena can be studied without immersive technology. This opens opportunities for accessible, reproducible research setups and supports the development of real-time, gaze-aware AI systems that operate across everyday platforms.

7 Future Work

Our work provides many starting points for future research.

Avatar Appearance. As mentioned above, there are many possible style appearances for the avatar. The question of how this influences users' reactions remains open and could be explored in future research.

Multi-modal Feedback. The study deliberately focused on gaze as the sole reactive modality. Future work could examine how gaze interacts with other nonverbal cues such as facial expressions, gestures, or prosodic variation. Combining these channels may yield a more nuanced understanding of how multi-modal feedback shapes social presence and user comfort.

Adaptive Gaze Personalities. Our avatar followed a confident, engagement-seeking gaze strategy. Exploring alternative strategies, such as shy, neutral, or submissive gaze patterns, could reveal how avatar personality and user traits interact. Such work could inform adaptive systems that tailor gaze behavior to user preferences or situational context.

Technical Developments. Although our prototype relied on a dedicated eye tracker, advances in computer vision-based gaze estimation may soon enable similar functionality using standard webcams. Evaluating the accuracy, latency, and user acceptance of such lightweight solutions will be important to scaling gaze-aware interfaces to everyday devices.

Alternatively, future everyday devices could provide integrated eye trackers, as is already the case for many VR glasses. The question of whether additional hardware costs are justified depends on the customers who buy the devices and the number of existing gaze-aware avatars.

Emotion and Personality Modeling. A longer-term challenge is to understand how personality traits and emotional states manifest in body movements, facial expressions, and gaze patterns. Insights from affective computing and personality psychology [Digman 1990; Völkel et al. 2020] could inform the development of avatars that not only react to user gaze but also express consistent social personas. Such models may also enable a more precise analysis of user behavior in human-AI interactions.

Ethical and Experiential Considerations. As gaze-aware avatars become more realistic, questions of transparency, user autonomy, and emotional attachment become increasingly important. Because gaze operates below conscious awareness, it can subtly influence users' sense of connection and emotional comfort. Developers should therefore apply gaze feedback transparently and ethically, ensuring that it enhances understanding and trust rather than manipulation.

Future studies should examine how reactive gaze influences trust and empathy, and how to design gaze behavior that supports meaningful, respectful interaction without fostering dependence or manipulation.

8 Conclusion

This work demonstrated that a gaze-aware avatar can influence user eye movements and interaction behavior, even when users are unaware of the underlying gaze responsiveness. Participants subconsciously regulated their visual attention in response to avatar's gaze mirroring patterns, found in human conversation. These findings confirm that gaze operates as an implicit social signal in human–AI interaction and show that such effects occur even on conventional 2D screens during unscripted, real-time dialog.

Implementing gaze awareness thus represents an important step toward more natural and socially tuned AI avatars. Our results indicate that even minimal reactive gaze can elicit human-like behavioral responses, providing a lightweight yet powerful design principle for conversational systems. Because gaze operates below conscious awareness, designers must consider its ethical use and its potential influence on user perception and emotional comfort. By combining empirical evidence with technical feasibility, this work lays the foundation for developing AI-supported avatars that communicate more naturally, transparently, and responsibly.

References

Rumi Allbert, Nima Yazdani, Ali Ansari, Aruj Mahajan, Amirhossein Afsharrad, and Seyed Shahabeddin Mousavi. 2025. Evaluating Speech-to-Text x LLM x Text-to-Speech Combinations for AI Interview Systems. arXiv:2507.16835 [eess.AS] https://arxiv.org/abs/2507.16835

Michael Argyle and Janet Dean. 1965. Eye-Contact, Distance and Affiliation. *Sociometry* 28, 3 (1965), 289–304. http://www.jstor.org/stable/2786027

Nikolaus Bee, Johannes Wagner, Elisabeth André, Fred Charles, David Pizzi, and Marc Cavazza. 2010a. Interacting with a gaze-aware virtual character. In *Proceedings of the 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction* (Hong Kong, China) *(EGIHMI '10)*. Association for Computing Machinery, New York, NY, USA, 71–77. doi:10.1145/2002333.2002345

Nikolaus Bee, Johannes Wagner, Elisabeth André, Thurid Vogt, Fred Charles, David Pizzi, and Marc Cavazza. 2010b. Discovering eye gaze behavior during human-agent conversation in an interactive storytelling application. In *International Conference on Multimodal Interfaces* (Beijing, China) *(ICMI-MLMI '10)*. Association for Computing Machinery, New York, NY, USA, Article 9, 8 pages. doi:10.1145/1891903.1891915

Evgenia Boyarskaya, Alexandra Sebastian, Thomas Bauermann, Heiko Hecht, and Oliver Tuescher. 2015. The Mona Lisa Effect: Neural Correlates of Centered and Off-Centered Gaze. *Human brain mapping* 36 (02 2015). doi:10.1002/hbm.22651

Ryan Canales, Eakta Jain, and Sophie Jörg. 2023. Real-Time Conversational Gaze Synthesis for Avatars. In *Proceedings of the 16th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Rennes, France) *(MIG '23)*. Association for Computing Machinery, New York, NY, USA, Article 17, 7 pages. doi:10.1145/3623264.3624446

Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill (Eds.). 2000. *Embodied Conversational Agents*. MIT Press, Cambridge, MA.

David Dembinsky, Ko Watanabe, Andreas Dengel, and Shoya Ishimaru. 2024. Gaze Generation for Avatars Using GANs. *IEEE Access* 12 (2024), 101536–101548. doi:10.1109/ACCESS.2024.3430835

J M Digman. 1990. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology* 41, Volume 41, 1990 (1990), 417–440. doi:10.1146/annurev.ps.41.020190.002221

Irenäus Eibl-Eibesfeldt. 1997. *Die Biologie des menschlichen Verhaltens* (3., überarb. und erw. aufl., [nachdr.] ed.). Piper, München [u.a.]. 1118 S. pages. Auf der Rückseite des Titelblattes: 4. Aufl., 13. - 14. Tsd. 1997.

Maximilian C. Fink, Seth A. Robinson, and Bernhard Ertl. 2024. AI-based avatars are changing the way we learn and teach: benefits and challenges. *Frontiers in Education* 9 (2024). doi:10.3389/feduc.2024.1416307

Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. 2001. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 309–316. doi:10.1145/365024.365121

Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 529–536. doi:10.1145/642611.642703

E. Goffman. 1963. *Behavior in Public Places*. Free Press, New York, NY 10020, USA.

Alexandra Hoffmann, Sabrina Schiestl, Philipp Sinske, Matthias Gondan, Pierre Sachse, and Thomas Maran. 2024. Sharing and Receiving Eye-Contact Predicts Mate Choice After a 5-Minute Conversation: Evidence from a Speed-Dating Study. *Archives of Sexual Behavior* 53 (02 2024). doi:10.1007/s10508-023-02806-0

Kazuya Izumi, Shuhey Koyama, and Yoichi Ochiai. 2025. AnimeGaze: Real-Time Mutual Gaze Synthesis for Anime-Style Avatars in Physical Environments via Behind-Display Camera. arXiv:2503.06324 [cs.HC] https://arxiv.org/abs/2503.06324

Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26 (1967), 22–63. doi:10.1016/0001-6918(67)90005-4

Michael Kipp and Patrick Gebhard. 2008. IGaze: Studying Reactive Gaze Behavior in Semi-immersive Human-Avatar Interactions. In *Intelligent Virtual Agents*, Helmut Prendinger, James Lester, and Mitsuru Ishizuka (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 191–199.

Stefan Klein. 2025. The effects of human-like social cues on social responses towards text-based conversational agents—a meta-analysis. *Humanities and Social Sciences Communications* 12 (2025). https://api.semanticscholar.org/CorpusID:282523151

Michael Koller, Astrid Weiss, Matthias Hirschmanner, and Markus Vincze. 2023. Robotic gaze and human views: A systematic exploration of robotic gaze aversion and its effects on human behaviors and attitudes. *Frontiers in Robotics and AI* Volume 10 - 2023 (2023). doi:10.3389/frobt.2023.1062714

Peter Kullmann, Theresa Schell, Timo Menzel, Mario Botsch, and Marc Erich Latoschik. 2025. Coverage of Facial Expressions and Its Effects on Avatar Embodiment, Self-Identification, and Uncanniness. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 3613–3622. doi:10.1109/TVCG.2025.3549887

Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler. 2002. Eyes alive. *ACM Trans. Graph.* 21, 3 (July 2002), 637–644. doi:10.1145/566654.566629

Yong Ma, Heiko Drewes, and Andreas Butz. 2022. How Should Voice Assistants Deal With Users' Emotions? arXiv:2204.02212 [cs.HC] https://arxiv.org/abs/2204.02212

Amama Mahmood, Junxiang Wang, Bingsheng Yao, Dakuo Wang, and Chien-Ming Huang. 2025. User Interaction Patterns and Breakdowns in Conversing with LLM-Powered Voice Assistants. *International Journal of Human-Computer Studies* 195 (2025), 103406. doi:10.1016/j.ijhcs.2024.103406

Thomas Maran, Marco Furtner, Simon Liegl, Theo Ravet-Brown, Lucas Haraped, and Pierre Sachse. 2020. Visual Attention in Real-World Conversation: Gaze Patterns Are Modulated by Communication and Group Size. *Applied Psychology* 70 (11 2020), 1602–1627. doi:10.1111/apps.12291

Masahiro Mori, Karl MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19 (06 2012), 98–100. doi:10.1109/MRA.2012.2192811

Samer Al Moubayed, Jens Edlund, and Jonas Beskow. 2012. Taming Mona Lisa: Communicating gaze faithfully in 2D and 3D facial projections. *ACM Trans. Interact. Intell. Syst.* 1, 2, Article 11 (Jan. 2012), 25 pages. doi:10.1145/2070719.2070724

Thomas O. Nelson and Louis Narens. 1980. Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior* 19, 3 (1980), 338–368. doi:10.1016/S0022-5371(80)90266-2

Kristine L. Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 12, 5 (10 2003), 481–494. doi:10.1162/105474603322761289

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. arXiv:2304.03442 [cs.HC] https://arxiv.org/abs/2304.03442

Trong Thang Pham, Tuong Do, Nhat Le, Ngan Le, Hung Nguyen, Erman Tjiputra, Quang Tran, and Anh Nguyen. 2024. Style Transfer for 2D Talking Head Generation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 7500–7509. doi:10.1109/CVPRW63382.2024.00745

Mashrur Rashik, Mahmood Jasim, Kostiantyn Kucher, Ali Sarvghad, and Narges Mahyar. 2024. Beyond Text and Speech in Conversational Agents: Mapping the Design Space of Avatars. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1875–1894. doi:10.1145/3643834.3661563

Shane Rogers, Craig Speelman, Oliver Guidetti, and Melissa Longmuir. 2018. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific Reports* 8 (03 2018). doi:10.1038/s41598-018-22726-7

Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-Based Conversational Agents Using the Psycholexical Approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376210

Ning Wang and Jonathan Gratch. 2010. Don't just stare at me!. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1241–1250. doi:10.1145/1753326.1753513

Mark West, Rebecca Kraut, and Chew Han Ei. 2019. *I'd blush if I could: closing gender divides in digital skills through education.* UNESCO. doi:10.54675/RAPC9356