Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When Does AI Matter? User Experience Across Phases of a VR Flight Simulation

Kaitlyn Tracy

ktracy2@emich.edu

Eastern Michigan University

William Dalian

Eastern Michigan University

Rami Dari

Eastern Michigan University

Ashton Clark

Eastern Michigan University

Ourania Spantidi

Eastern Michigan University

Research Article

Keywords: Virtual Reality, Conversational AI, Fear of Flying, Exposure Therapy, User Experience, Human-Computer Interaction

Posted Date: May 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9543308/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

When Does AI Matter? User Experience Across Phases of a VR Flight Simulation

Kaitlyn Tracy[1], William Dalian[1], Rami Dari[1], Ashton Clark[2], Ourania Spantidi[1]

[1]Department of Computer Science, Eastern Michigan University, E Circle Dr., Ypsilanti, 48197, MI, USA.
[2]Department of Psychology, Eastern Michigan University, 341 Science Complex , Ypsilanti, 48197, MI, USA.

Contributing authors: ktracy2@emich.edu; wdalian@emich.edu; rdari5@emich.edu; aclar114@emich.edu; ourania.spantidi@emich.edu;

Abstract

Prior work has shown that conversational AI companions can improve engagement and reduce anxiety in virtual reality (VR) exposure scenarios involving stressful events such as turbulence. However, real-world experiences such as air travel unfold over multiple phases, including takeoff, cruise, and landing, and it remains unclear whether user engagement and immersion are driven primarily by AI interaction or by the temporal structure of the environment itself. In this study, we extend a VR flight simulation to include a multi-phase flight experience and evaluate user engagement, immersion, attention allocation, and perceived realism across flight phases, both with and without an AI companion. Results indicate that immersion varied significantly across flight phases, with cruise producing higher immersion than landing, while engagement did not differ significantly by phase. The presence of the AI companion did not significantly alter engagement, immersion, attention allocation, or perceived realism across phases. These findings suggest that while AI companions may enhance user experience during stress-inducing events, the temporal structure and environmental realism of the VR experience may play a larger role in shaping immersion during stable conditions. This work contributes to understanding the contextual role of AI companions in immersive environments and has implications for the design of VR training and exposure therapy systems.

Keywords: Virtual Reality, Conversational AI, Fear of Flying, Exposure Therapy, User Experience, Human-Computer Interaction

1 Introduction

Specific phobia is the most common psychiatric diagnosis [1]. Of the phobias, fear of flying is one of the most commonly diagnosed, with prevalence rates ranging from 15% to 40% of the population [2]. As those with flight fear tend to avoid air travel as much as possible [3], social and career consequences can be severe, and fear is unlikely to resolve without treatment [4]. Exposure therapy, the gold standard treatment for phobias including flight [5], involves repeatedly exposing the patient to the feared stimuli until desensitization or anxiety tolerance occurs [6]. For flight fear, however, this is both logistically complex and expensive. A potential solution to the challenges of flight exposure therapy is virtual reality exposure therapy (VRET), which has been proven to be effective in treatment for various phobias, including fear of flying (FoF) [7–9]. As VRET can simulate the flight experience in a more affordable and convenient setting, this allows for greater treatment flexibility and accessibility [8]. VRET has the added benefit of allowing clinicians to control the flight environment, simulating scenarios specific to the patient's fear as well as allowing patients to stop the exposure on demand, in contrast to flight where one is committed to the full exposure experience once the plane leaves the ground. These benefits are reflected in patient preferences, with 80% of those with a specific phobia diagnosis preferring VRET compared to in vivo or "real life" exposures [10].

While virtual flights allow for gradual and more customizable exposures, further developments are needed. Effective exposure therapy relies on emotional involvement to elicit an anxious response [2, 8] and "developing a mental representation of the virtual stimuli as one's own environment" is critical to VRET [8]. This highlights the importance of maximizing engagement and presence in VRET to maximize its effectiveness. When a user is totally immersed, for example in a video game, their feelings are at least partially affected by the virtual environment [11]. The success of exposure therapy relies on its ability to generalize to "real world" scenarios and its ability to elicit the anxiety experienced during those scenarios [12]. For VRET, presence is one of the central mechanisms by which anxiety is elicited [13]. Virtual flights should be immersive and engaging enough to realistically recreate a flight experience, including the accompanying anxiety triggers. The success of VRET for flight fear lies in its ability to both induce an appropriate amount of anxiety but not to such a degree that patients cannot tolerate treatment engagement.

In addition to increasing presence and engagement, it is important to optimize patient buy-in and reduce drop out, which can range from 17% to 27% in phobia exposure therapy [10]. To tolerate the fear that is fundamental to exposure interventions, many patients utilize calming strategies or "safety behaviors" during exposure, including securing the presence of a trusted companion [14]. For example, one may refuse to fly alone but finds flying with their spouse reduces their anxiety to tolerable enough levels that they can endure a flight. While there is debate over whether safety behaviors can hinder long term treatment potential, they can also reduce fear just enough for the patient to complete the exposure [15]. A safe person, for example, can provide reassurance and emotional support that allows the patient to complete a flight when they otherwise would not. The limitations of a "safe person" occur when such a person

is unavailable to accompany the patient, which can be particularly difficult to coordinate with costly flights. Virtual AI companion components in VRET may serve to simulate the emotional support and reassurance of the safe person without the logistical and financial complications. Having such a presence in a VRET environment may both reduce anxiety and allow for greater emotional engagement with the treatment.

While social entities such as other players or AI agents affect user experiences [11, 16], and despite advancements in AI, no existing work has been found that studies the impact of AI agents on participants' engagement and presence in a VRET environment for FoF. In our prior work [17], we introduced a conversational AI flight companion into a VR flight simulation and examined its impact during turbulence scenarios, which represent high-stress moments in a flight experience. The results showed that the presence of the AI companion increased engagement and immersion and helped reduce anxiety during turbulence. These findings suggested that conversational AI companions can act as a form of social and informational support during stress-inducing virtual experiences, similar to the effects of a "safe person".

However, real-world flight experiences are not defined solely by isolated stress events such as turbulence, but by the temporal progression of multiple phases such as takeoff, cruise, and landing. While turbulence represents a high-anxiety moment, much of the flight experience occurs during relatively stable conditions. Fear of turbulence is common, however, a significant portion of those with flight phobia experience fear outside of turbulent conditions [18]. Therefore, it remains unclear whether user engagement and immersion in VR flight environments are driven primarily by AI interaction or by the temporal and structural realism of the simulated experience itself. This paper extends our prior work [17] by introducing a multi-phase VR flight simulation consisting of takeoff, cruise, and landing, and evaluates user engagement, immersion, attention allocation, and perceived realism across flight phases, both with and without an AI companion.

For the current study, we revisit the original research questions (RQs) from our prior work [17] (RQs 1-4) and introduce additional RQs (RQs 5–7) to investigate the generalizability of AI effects, the role of temporal structure, and the robustness of results to experimental design factors:

RQ1 Does the presence of an AI flight companion impact engagement?
RQ2 Does the presence of an AI flight companion impact immersion?
RQ3 Does the presence of an AI flight companion lower the user's anxiety level?
RQ4 To what extent do participants feel supported by the AI flight companion?
RQ5 Do the effects of AI companions observed in high-stress scenarios generalize to multi-phase VR flight experiences?
RQ6 How does the temporal structure of a VR flight (i.e., takeoff, cruise, landing) influence user experience?
RQ7 Are the effects of AI companions robust to contextual and experimental factors such as flight phase and condition order?

By examining user experience across different phases of a VR flight, this work aims to better understand whether and how AI companions influence user experience within

more realistic, temporally structured environments. In addition, this work contributes to understanding how users perceive and interpret visually simulated environments, particularly in the context of temporally evolving scenes. By examining how immersion and engagement are shaped across different phases of a visual simulation, this study provides insight into the relative roles of environmental realism and interactive AI elements in influencing user perception.

2 Related Work

2.1 Presence & Engagement Measures

Immersion in virtual environments is commonly conceptualized as a multi-level construct consisting of engagement, engrossment, and presence. Engagement refers to the user's active interaction with the virtual environment, while engrossment reflects a deeper level of emotional involvement in which awareness of the real world diminishes. Presence represents the highest level of immersion, characterized by the sensation of being fully situated within the virtual environment [11, 19]. In this work, we adopt this established definition of *presence*.

A variety of instruments have been developed to measure presence and engagement in VR. Early work by [20] employed researcher-administered surveys to assess emotional reactivity and overall comfort. Jennett et al. [19] introduced a widely used self-report instrument, validated through comparisons between immersive and less immersive gameplay conditions, which has since served as a foundation for subsequent measurement approaches. More recent studies have extended these tools by incorporating additional items to capture nuanced aspects of engagement, presence, and related experiential factors, often administering these measures directly to participants in either paper-based or digital formats [11, 16].

Within the context of AI-enhanced VR environments, prior work has explored how artificial agents influence user experience. For example, [21] examined the impact of a GPT-driven teaching assistant on engagement and presence in a VR classroom using self-report measures. Similarly, [22] investigated AI-led virtual reality exposure therapy (VRET) for fear of heights, employing questionnaires that assessed perceived realism, emotional engagement, and the naturalness of the virtual therapist.

Despite these advances, existing studies have not employed measurement instruments specifically validated for AI-driven social presence systems. While prior findings suggest that AI agents can enhance engagement in educational VR settings [21] and show promise in therapeutic contexts such as fear-of-heights interventions [22], there remains a lack of research examining their role in VRET for fear of flying (FoF). This gap is particularly notable given the importance of social reassurance in real-world flight experiences. The present work addresses this limitation by investigating the impact of an AI companion in a domain where passenger–attendant interaction is central to user coping and emotional regulation.

2.2 Fear-of-Flying Measures

Prior research on fear of flying (FoF) has relied heavily on standardized self-report instruments to assess anxiety and related constructs. One commonly used measure is the Fear of Flying Inventory (FFI), often paired with the Questionnaire on Attitudes Toward Flying (QAF), both of which assess FoF intensity and situational fear responses using Likert-scale items [23].

Additional studies have employed complementary instruments, including the Fear of Flying Scale (FFS), the General Fear of Flying Questionnaire, and the Danger Expectancy and Anxiety Expectancy Scales, alongside broader measures such as the Anxiety Sensitivity Index [2]. These tools provide a comprehensive assessment of fear responses across different flight-related scenarios and have been widely used to evaluate the effectiveness of VRET interventions.

While these validated instruments offer robust measures of anxiety, they are typically designed for clinical assessment rather than capturing moment-to-moment experiential changes within dynamic VR environments. As a result, many VR-based studies—including the present work—adopt more lightweight, phase-specific self-report measures to balance measurement sensitivity with participant burden.

3 Review of Prior Work & Extension Overview

Our prior study [17] and the extension introduced here are similar in their system design, specifically the use of *Planogen* [24] to generate the virtual plane cabin, the plane cabin environment itself, and the concept of integrating an AI in-flight companion. This section outlines the similarities and provides an overview of the novel choices made for the extension. The design, procedure, and measures differ between the two studies. We provide a synopsis of the prior work's methodology in Subsection 3.1. The specifics of our extension's methodology are illustrated in Section 4.

3.1 Design & Procedure of Prior Work

Our original pilot study evaluated the impact of a conversational AI agent on user experience during VR flight simulations under low and high turbulence levels. We employed a within-subjects repeated measures design with four sequential conditions: low and high turbulence *without* an AI companion, followed by low and high turbulence *with* an AI companion. During the phase in which participants had access to the AI companion, there would be a virtual trigger cube mounted on the seat in front of the participant, as a means to interact with the AI companion. For the original study, participants remained in the same virtual environment while they experienced 15 seconds each, low turbulence followed by high turbulence, first without the AI companion, and then with the AI companion. The transition between *without* and *with* the AI companion involved the investigator clicking a checkbox to add the virtual trigger cube to the scene so that the participant could interact with the AI companion. Before adding the AI companion to the scene, a questionnaire was administered regarding the participant's experience with the prior conditions (*without* AI), then the same questionnaire was administered again after the second round of turbulence

conditions (*with* AI). Participants concluded the study with a final questionnaire after they had removed the headset.



(a) User's viewpoint showing a generated passenger in the cabin.



(b) Central-aisle perspective of the procedurally generated cabin interior.

Fig. 1: The generated aircraft cabin interior: (a) seated-passenger view, and (b) aisle-center view. Reproduced from [17].

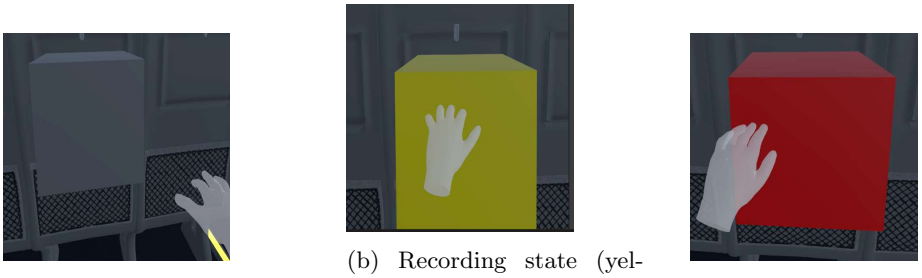
3.2 VR Plane Application

The VR application was developed in Unity 2022.3.45f1 using *Planogen* [24], a procedural generation plug-in. The virtual cabin featured 200 seats and 20 procedurally generated passengers. Results from the generated plane are shown in Figure 1. The Meta Quest 3 headset was used with hand tracking to maximize immersion. Turbulence was simulated by applying randomized x- and z-axes camera rotations to the XR rig: low turbulence ranged from $[-1°, 1°]$, and high turbulence from $[-5°, 5°]$. As turbulence is often a contributing factor to anxiety in those with FoF, it was implemented as rotational camera perturbations. These were feasible to implement within the Unity environment while maintaining stability and minimizing cybersickness risk.

In our original study, participants could interact with a GPT-based AI companion (model used was gpt-3.5-turbo) via a virtual trigger cube mounted on the seat in front of them. When touched, the cube changed color to indicate its state (gray: idle, yellow: recording, red: processing), and recorded the participant's voice via speech-to-text (Figure 2). The participant's question was appended to a prompt to prime the companion to be "a warm, calming in-cabin support presence on a virtual airplane" which is there to "reassure and assist passengers" in a "friendly, empathetic, and confident tone".

3.3 Extension Novelties

For our current study, we sought to investigate the effects of temporal progression by simulating a "full flight" experience with a multiphase exposure composed of takeoff, cruise, and landing phases. The measurements were applied within-phase, so after the participant experienced the duration of a phase, the investigator would read them a series of questions on a 5-point Likert-scale. The same set of questions was asked immediately after each phase, and another set was asked in regards to the complete



(a) Idle state (gray). (b) Recording state (yellow). (c) Processing state (red).

Fig. 2: Trigger cube states for initiating AI interaction. Reproduced from [17].

flight experience. This practice was followed for each flight, once for the flight with the AI companion, and once for the flight without the AI companion. The order in which we administered the flights was randomized for each participant.

Due to the nature of the current study's multi-factor and multi-phase design, our statistical analysis differs from our prior work. We analyzed the responses at the phase level, utilizing ANOVA tests to evaluate whether means differ across different factors and conditions (*AI presence* and *flight phase*).

In addition to our prior work's RQs, we posed three more RQs specific to the current work. These three additional RQs, stated in Section 1, ask about AI's effects across *flight phases* and whether or not the order in which the flights (*with* AI, *without* AI) are administered.

The following section, 4, expands on these novelties and details the design and procedure of the study conducted as an extension on our prior work.

Component	**Prior work [17]**	**This work**
Analysis Level	Block-Level	Phase-Level
Design	2 blocks within-subject	Multi-phase within-subject
RQs	4	7

Table 1: A comparison between our prior work [17] and this work.

4 Methodology

Our study employed a 2 (*Condition:* AI vs. No-AI) × 3 (*Phase:* Takeoff, Cruise, Landing) within-subject design. Each participant experienced both conditions in a randomized order to minimize potential order effects. Across conditions, the VR environment and phase structure remained consistent; the only difference was the availability of the AI companion. In the AI condition, the companion was continuously available throughout the entire flight. Participants were informed that they could interact with the companion at any point, including during transitions between phases. The full procedure is described in Subsection 4.4.

4.1 Participants

Thirty-two participants (9 female, 21 male, 1 non-binary, and 1 preferred not to answer; $M_{age} = 25.03$, $SD = 4.96$) were recruited from a university's College of Arts & Sciences, including undergraduate and graduate students as well as faculty. Informed consent was obtained from all participants. Pre-screening was conducted to identify individuals with severe anxiety disorders, PTSD, or extreme motion sickness; however, no participants were excluded. Background information, including prior VR and flight experience, was collected via a pre-questionnaire. This sample size is consistent with recommendations for pilot studies examining usability and experiential outcomes [25, 26], and matches that of our prior work.

4.2 Virtual Environment

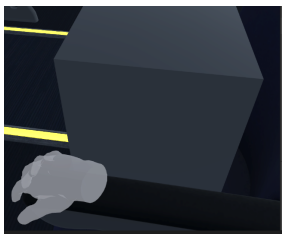
The virtual environment in this extension builds upon the original system, with two primary modifications: (1) the addition of an external airport environment to support a full-flight experience, and (2) adjustments to the placement and design of the AI interaction mechanism.

To simulate a complete flight sequence, we introduced a low-poly airport environment visible through the aircraft windows. During takeoff, the aircraft moves forward and gradually ascends, with the ground scene transitioning into a cloud layer. At cruising altitude, the external view becomes obscured by clouds to reinforce the perception of altitude. During landing, the clouds dissipate as the aircraft descends, revealing the ground environment before coming to a stop on the runway. These visual transitions were designed to enhance realism and support users' interpretation of flight progression. Such environmental cues are known to play a critical role in shaping immersion by providing continuity and reinforcing perceived motion.

In response to usability issues observed in our prior study, we repositioned the AI interaction trigger cube. Previously located on the back of the seat in front of the participant, the cube occasionally resulted in unintended activations due to turbulence and natural hand movement. In this study, the cube was relocated to the seat to the participant's right, reducing accidental interactions while maintaining accessibility. The previous and updated placements are shown in Figures 2 and 3, respectively.

We also refined the cube's visual feedback. The recording state color was changed from yellow to green to improve clarity and consistency, particularly in contrast with the red processing state. The updated interaction states are shown in Figure 3. Aside from these changes, the aircraft interior—including cabin layout and procedurally generated passengers—remains identical to that used in our prior work.

During the AI-assisted condition, participants could interact with a GPT-based AI companion (gpt-3.5-turbo) via the trigger cube. When activated, the cube changed color to indicate its state (gray: idle, green: recording, red: processing) and captured the participant's speech using automatic speech recognition. The transcribed input was combined with a system prompt that framed the agent as "a warm, calming in-cabin support presence" designed to reassure and assist passengers in a friendly and empathetic manner. Additional contextual information—including the current



(a) Idle state (gray).



(b) Recording state (green).



(c) Processing state (red).

Fig. 3: Trigger cube states for initiating AI interaction (extension).

flight phase and turbulence state—was also incorporated into the prompt. Responses generated by the AI were displayed as text near the participant's field of view.

The Takeoff and Landing phases were programmatically set to last approximately 45–60 seconds each. The Cruise phase was manually timed within the same duration range to allow for the administration of post-phase questionnaires before transitioning to the next phase. Phase transitions were controlled by the investigator via keyboard input.

4.3 Measures

Participants first completed a pre-questionnaire capturing demographic information, prior experience with VR and air travel, fear-of-flying (FoF) concerns, and screening criteria (see Subsection 4.1).

During the VR experience, participants responded to a set of four 5-point Likert-scale items following each phase (Takeoff, Cruise, Landing), assessing perceived engagement, immersion, anxiety, and attention. These items are provided in Appendix A.

Following each full flight (AI and No-AI conditions), participants completed a post-flight questionnaire consisting of six 5-point Likert-scale items. These items assessed perceived realism (e.g., completeness and naturalness of the experience), overall engagement and immersion, and affective responses such as anxiety and perceived safety. This enabled within-subject comparisons across conditions (Appendix B).

After completing the AI-assisted condition, participants completed a final post-VR questionnaire consisting of five Likert-scale items and three open-ended questions. These items assessed perceptions of the AI companion, including trust, emotional safety, perceived support, and clarity of understanding. The open-ended responses provided additional qualitative insight into participant experiences.

4.4 Procedure

The study was conducted in the Embedded AI Systems Lab at Eastern Michigan University. Upon arrival, participants received a questionnaire packet containing a condition order code used to determine whether they began with the AI or No-AI condition. After completing the pre-questionnaire and confirming eligibility, participants were seated and fitted with a Meta Quest 3 headset.

The investigator, positioned nearby, controlled the VR simulation and administered verbal prompts. Participants were informed that they would complete two full flight experiences, each consisting of Takeoff, Cruise, and Landing phases. During each phase, they remained seated and experienced the simulated environment.

Phase	No-AI Condition	AI Condition
Takeoff	VR exposure (no AI interaction)	VR exposure + AI available
Cruise	VR exposure (no AI interaction)	VR exposure + AI available
Landing	VR exposure (no AI interaction)	VR exposure + AI available

Table 2: Condition × phase design structure.

In the No-AI condition, participants experienced the full flight sequence without access to the AI companion. After each phase, the investigator verbally administered the post-phase questionnaire. At the end of the flight, participants responded to the post-flight questionnaire.

In the AI condition, the same procedure was followed, with the addition of the AI companion. Participants were first given instructions and a brief opportunity to practice interacting with the system. They were informed that they could engage with the AI at any point during the flight, including between phases. The interaction process was explained as follows:

> "To interact with the companion, move your hand into the cube until it turns green. While it is green, it is recording your voice. When you are done speaking, remove your hand. The cube will turn red while your request is processed, and a response will appear shortly."

Participants were reminded that the AI could provide informational and emotional support, but could not perform physical actions.

After completing both conditions, participants removed the headset and completed the final questionnaire. They were then debriefed and dismissed.

5 Results

In this section, we first revisit the original research questions (RQs 1–4) in the context of the multi-phase flight scenario, and then examine the extension-specific research questions (RQs 5–7) to further investigate the role of temporal structure and experimental factors.

5.1 Revisiting Original RQs

In our original study [17], we examined how the presence of an AI companion affected participants' perceived engagement, immersion, anxiety, and support during a VR flight experience. For this extension, we revisit these research questions within a more realistic, multi-phase flight scenario.

Contrary to our prior findings, AI presence did not significantly impact engagement, immersion, or anxiety in the multi-phase setting. A repeated-measures ANOVA

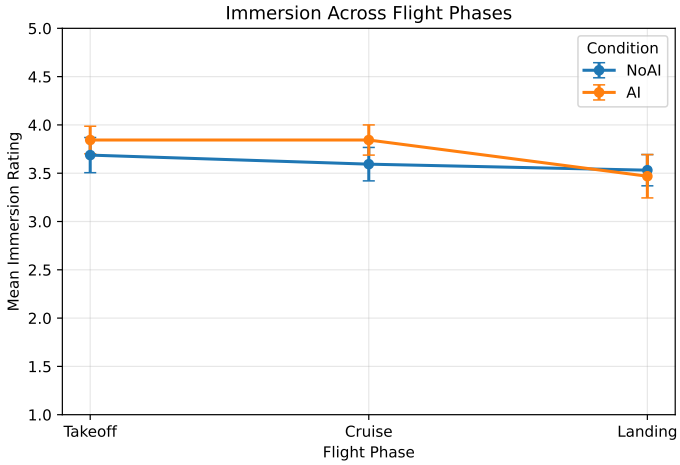


Fig. 4: Mean immersion ratings across flight phases (takeoff, cruise, and landing) for the AI and No-AI conditions. Immersion varied significantly across phases ($F = 3.78$, $p = 0.028$), while differences between conditions remained small and non-significant. Error bars indicate standard error.

revealed a statistically significant main effect of *flight phase* on immersion ($F = 3.78$, $p = 0.028$), indicating that immersion varies across takeoff, cruise, and landing. This was the only statistically significant effect observed across the measured factors. These findings suggest that the structure and progression of the experience may play a more prominent role in shaping user experience than the presence of an AI companion. As shown in Figure 4, immersion varied across phases, while differences between the AI and No-AI conditions remained small.

RQ1: Does the presence of an AI flight companion impact engagement?

No statistically significant effect of AI on engagement was observed across phases ($p > 0.05$). While small positive trends favored the AI condition, particularly during landing ($d = 0.29$), these effects did not reach significance. This contrasts with prior findings and suggests that AI-driven engagement effects may be context-dependent.

RQ2: Does the presence of an AI flight companion impact immersion?

No statistically significant differences in immersion were observed between the AI and No-AI conditions across any phase ($p > 0.05$). Small positive trends were observed during takeoff and cruise; however, these effects were inconsistent and did not reach statistical significance, indicating that AI-related improvements in immersion do not generalize to the multi-phase setting.

RQ3: Does the presence of an AI flight companion lower the user's anxiety level?

No statistically significant reduction in anxiety was observed in the AI condition across any phase (all $p > 0.05$). Effect sizes were small ($|d| \leq 0.22$), suggesting that previously observed reductions in anxiety may not generalize under more structured conditions.

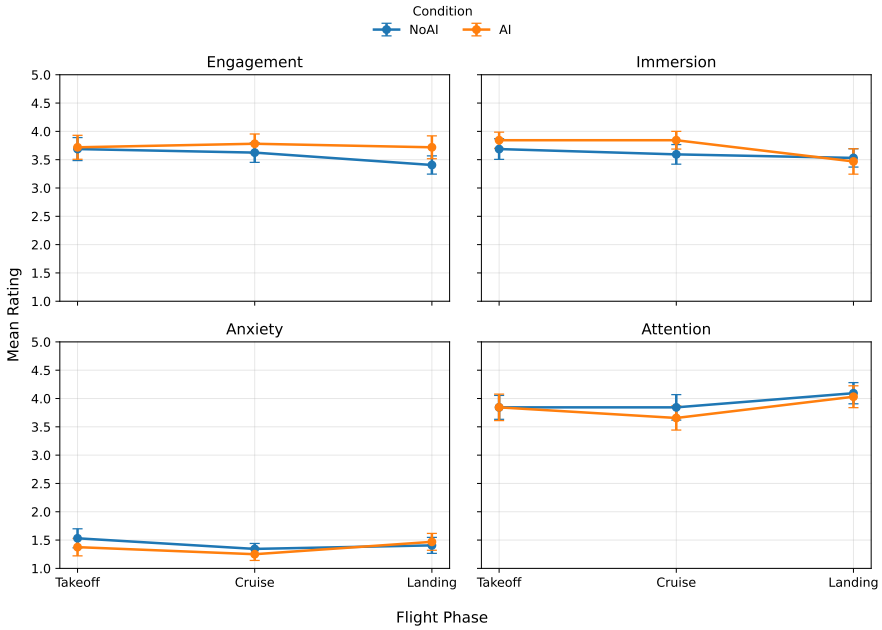


Fig. 5: Mean ratings for engagement, immersion, anxiety, and attention across flight phases for the AI and No-AI conditions. Across all constructs, ratings were broadly similar between conditions, with immersion showing the clearest variation across phases. Error bars indicate standard error.

RQ4: To what extent do participants feel supported by the AI flight companion?

Participants reported moderate levels of perceived support from the AI companion (composite mean = 3.71), with consistently positive ratings across measures of trust, safety, and understanding. This indicates that participants perceived the AI as supportive, even in the absence of measurable effects on engagement, immersion, or anxiety.

5.2 Extension-Specific RQs

Given that the effects of the AI companion did not generalize to the multi-phase scenario (Section 5.1), we next explore whether aspects such as the temporal structure and experimental design of the experience play a more prominent role in shaping user outcomes. The following RQs are designed to investigate these factors. Figure 5 summarizes the phase-level patterns across all measured constructs and highlights the similarity between AI and No-AI conditions.

RQ5: Do the effects of AI companions observed in high-stress scenarios generalize to multi-phase VR flight experiences?

Across all phases, no statistically significant differences were observed between the AI and No-AI conditions for engagement, immersion, or anxiety (all $p > 0.05$).

While small positive trends favoring the AI condition were observed in certain cases, such as engagement during landing ($d = 0.29$), these effects were inconsistent and did not reach statistical significance. This pattern is visually reflected in Figure 5, where ratings remain closely aligned between conditions across phases and constructs, reinforcing the absence of generalizable AI effects.

These findings suggest that the benefits of AI companions observed in prior work under isolated high-stress conditions do not generalize to more complex, multi-phase flight experiences.

RQ6: How does the temporal structure of a VR flight influence user experience?

We conducted a repeated-measures ANOVA that revealed a statistically significant main effect of flight phase on immersion ($F = 3.78$, $p = 0.028$), indicating that immersion differs across takeoff, cruise, and landing. In contrast, no significant phase effects were found for engagement, anxiety, or attention (all $p > 0.05$). Figure 5 further illustrates that this phase-related variation is most pronounced for immersion, while the remaining measures exhibit relatively stable patterns across phases.

These findings suggest that the temporal structure of the experience plays a meaningful role in shaping immersion, independent of AI presence.

RQ7: Are the effects of AI companions robust to contextual and experimental factors such as flight phase and condition order?

No significant interaction effects between AI presence and flight phase were observed for any measured construct, including engagement, immersion, anxiety, and attention (all $p > 0.05$), indicating that the effect of AI does not depend on the phase of the experience. Consistent with this, no distinct divergence between AI and No-AI trends across phases is visible in Figure 5.

Additionally, no significant order effects were observed (all $p > 0.10$), suggesting that the results are robust to condition order and are not influenced by novelty, fatigue, or learning effects.

Taken together, these results indicate that while immersion is influenced by the temporal structure of the experience, AI presence does not produce measurable changes across engagement, immersion, anxiety, or attention in this multi-phase setting.

6 Discussion

6.1 Summary of Key Findings

This study revisited the original research questions (RQs 1–4, Section 5.1) and introduced additional research questions (RQs 5–7, Section 5.2) to examine the generalizability of AI effects and the role of temporal structure in a multi-phase VR flight experience.

For RQ1–RQ3 (Section 5.1), no statistically significant effects of AI presence were observed on engagement, immersion, or anxiety (all $p > 0.05$). While small positive trends favoring the AI condition were present in some cases (e.g., engagement during landing, $d = 0.29$), these effects were inconsistent and did not reach statistical significance. For RQ4 (Section 5.1), participants reported moderate perceived support from

the AI companion (composite mean = 3.71), indicating that the AI was consistently perceived as supportive despite the lack of measurable behavioral effects.

For the extension-specific RQs (Section 5.2), RQ5 revealed that the effects of AI observed in prior work did not generalize to the multi-phase setting, as no significant differences between AI and no-AI conditions were observed across phases. RQ6 identified a statistically significant main effect of flight phase on immersion ($F = 3.78$, $p = 0.028$), indicating that immersion varies across takeoff, cruise, and landing. This was the only statistically significant effect observed in the study. Finally, RQ7 showed that the results were robust to contextual and experimental factors, with no significant interaction effects between AI and phase (all $p > 0.05$) and no significant order effects (all $p > 0.10$).

6.2 Interpretation of Results

Taken together, these findings suggest that the effects of AI companions in VR flight environments are context-dependent and do not readily generalize to more realistic, temporally structured experiences. While our prior work demonstrated that AI could enhance engagement and immersion and reduce anxiety during turbulence, these effects were not replicated in the present study. This discrepancy can be explained by differences in context. Turbulence represents a high-stress, attention-demanding scenario, whereas the multi-phase flight introduces extended periods of lower stimulation (such as the cruising phase), during which the relative influence of the AI companion may be diminished.

The only statistically significant finding in this study was the effect of flight phase on immersion (RQ6, $F = 3.78$, $p = 0.028$). This highlights the importance of temporal structure in shaping user experience, particularly in how users perceive and interpret changes in the visual environment. The progression of visual cues across phases appears to play a more significant role in driving immersion than the presence of an AI companion. It also suggests that immersion is more strongly driven by the progression and dynamics of the environment than by the presence of an AI companion. Notably, no interaction effects were observed (RQ7), indicating that the influence of the AI companion does not vary across different phases of the experience. This lack of phase sensitivity suggests that the AI companion operates as a static element within the environment rather than adapting to changes in user state or contextual demands.

Additionally, the absence of order effects (RQ7) strengthens the validity of these findings, indicating that the observed results are not attributable to novelty, fatigue, or learning effects. Together, these results suggest that simply adding an AI companion is insufficient to meaningfully alter user experience in more complex and ecologically valid VR scenarios.

Despite the lack of significant effects on engagement, immersion, and anxiety, participants consistently perceived the AI as supportive (RQ4, mean = 3.71). This points to a disconnect between subjective perception and measurable outcomes. Users may interpret the AI companion as socially present and supportive without this perception translating into quantifiable changes in experience metrics. This finding suggests that AI companions may primarily influence how users interpret and frame the experience, rather than directly altering their behavioral or affective responses.

6.3 Limitations & Future Work

Several limitations should be considered when interpreting these results. First, the sample size ($n = 32$) may limit statistical power, particularly for detecting small effects. While effect sizes were generally small, some observed trends (e.g., engagement during landing, $d = 0.29$) suggest that subtle effects may exist but were not detectable in the present study.

Second, even though the AI companion used in this study adapted to different phases of the flight, it did not receive any feedback regarding the user's state. As indicated by the lack of interaction effects (RQ7), the influence of the AI companion remained consistent across phases, which may have limited its potential impact. Future work could explore adaptive AI systems that respond to user anxiety or their questions to provide more targeted support.

Finally, while this study increased ecological validity by introducing a multi-phase flight structure, further work is needed to explore how different types of stressors (e.g., announcements, other passengers in the plane) interact with AI presence. Future studies could also investigate individual differences, such as the severity of fear of flying for each user, as potential moderators of AI effectiveness.

Overall, this work highlights the importance of evaluating AI systems under realistic conditions and suggests that future research should focus not only on whether AI improves user experience, but also on when and how such improvements occur.

7 Conclusion

This work revisited the impact of conversational AI companions in VR flight environments by extending a prior turbulence-focused study to a more realistic, multi-phase flight experience. Contrary to earlier findings, the presence of an AI companion did not significantly affect engagement, immersion, or anxiety across takeoff, cruise, and landing. Instead, the only statistically significant effect observed was the influence of flight phase on immersion ($F = 3.78$, $p = 0.028$), highlighting the role of temporal structure in shaping user experience. These results suggest that AI-driven benefits identified in isolated, high-stress scenarios do not readily generalize to more ecologically valid and temporally structured environments. While participants consistently perceived the AI as supportive, this perception did not translate into measurable improvements in experience metrics, pointing to a distinction between subjective interpretation and observable outcomes. Overall, this study contributes to a more nuanced understanding of conversational AI in immersive environments by identifying the limits of its effectiveness and emphasizing the importance of experience design and contextual factors. Future work should explore adaptive AI systems that respond to user state and advanced environmental context, as well as investigate individual differences that may mediate AI effectiveness. More broadly, these findings highlight the importance of visual simulation design and user perception in immersive environments, suggesting that environmental realism and temporal visual structure may play a more critical role than interactive AI elements in shaping user experience.

Appendix A Post-Phase Questions

1. I felt engaged during this phase.
   (1 - Not at all engaged, 5 - Extremely engaged)
2. I felt immersed during this phase.
   (1 - Not at all immersed, 5 - Extremely immersed)
3. This phase increased my anxiety.
   (1 - Not increased at all, 5 - Extremely increased)
4. I was highly focused on monitoring the environment during this phase.
   (1 - Not focused at all, 5 - Extremely focused)

Appendix B Post-Flight Questions

1. The flight felt like a complete and realistic experience.
   (1 - Not at all, 5 - Very much so)
2. Transitions between flight phases felt natural.
   (1 - Not at all natural, 5 - Extremely natural)
3. Overall, I felt engaged during this flight.
   (1 - Not at all engaged, 5 - Extremely engaged)
4. Overall, I felt immersed during this flight.
   (1 - Not at all immersed, 5 - Extremely immersed)
5. Overall, this flight increased my anxiety.
   (1 - Not increased at all, 5 - Extremely increased)
6. I felt confident that nothing bad or dangerous would happen during the flight.
   (1 - Not confident at all, 5 - Extremely confident)

References

[1] Kessler, R.C., Petukhova, M., Sampson, N.A., Zaslavsky, A.M., Wittchen, H.-U.: Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the united states. International journal of methods in psychiatric research **21**(3), 169–184 (2012)

[2] Mühlberger, A., Herrmann, M.J., Wiedemann, G., Ellgring, H., Pauli, P.: Repeated exposure of flight phobics to flights in virtual reality. Behaviour Research and Therapy **39**(9), 1033–1050 (2001) https://doi.org/10.1016/S0005-7967(00)00076-0

[3] Wilhelm, F.H., Roth, W.T.: Clinical characteristics of flight phobia. Journal of anxiety disorders **11**(3), 241–261 (1997)

[4] Van Gerwen, L.J., Diekstra, R.F.: Fear of flying treatment programs for passengers: An international review. Aviation, space, and environmental medicine **71**(4), 430–437 (2000)

[5] Barlow, D.H., Conklin, L.R., Bentley, K.H.: Psychological treatments for panic

disorders, phobias, and social and generalized anxiety disorders. A guide to treatments that work, 409–461 (2015)

[6] Öst, L.-G., Brandberg, M., Alm, T.: One versus five sessions of exposure in the treatment of flying phobia. Behaviour Research and Therapy **35**(11), 987–996 (1997)

[7] Klein, R.A.: Virtual reality exposure therapy (fear of flying): From a private practice perspective. CyberPsychology & Behavior **1**(3), 311–316 (1998)

[8] Rothbaum, B.O., Hodges, L., Watson, B.A., Kessler, G.D., Opdyke, D.: Virtual reality exposure therapy in the treatment of fear of flying: A case report. Behaviour Research and Therapy **34**(5-6), 477–481 (1996)

[9] Wiederhold, B.K., Jang, D.P., Gevirtz, R.G., Kim, S.I., Kim, I.-Y., Wiederhold, M.D.: The treatment of fear of flying: a controlled study of imaginal and virtual reality graded exposure therapy. IEEE transactions on information technology in biomedicine **6**(3), 218–223 (2002)

[10] Garcia-Palacios, A., Botella, C., Hoffman, H., Fabregat, S.: Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. Cyberpsychology & behavior **10**(5), 722–724 (2007)

[11] Sierra Rativa, A., Postma, M., Van Zaanen, M.: The influence of game character appearance on empathy and immersion: Virtual non-robotic versus robotic animals. Simulation & gaming **51**(5), 685–711 (2020)

[12] Foa, E.B., Kozak, M.J.: Emotional processing of fear: exposure to corrective information. Psychological bulletin **99**(1), 20 (1986)

[13] Baños, R.M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., Rey, B.: Immersion and emotion: their impact on the sense of presence. Cyberpsychology & behavior **7**(6), 734–741 (2004)

[14] Clark, D.M., Ehlers, A.: An overview of the cognitive theory and treatment of panic disorder. Applied and preventive psychology **2**(3), 131–139 (1993)

[15] Craske, M.G., Treanor, M., Conway, C.C., Zbozinek, T., Vervliet, B.: Maximizing exposure therapy: An inhibitory learning approach. Behaviour research and therapy **58**, 10–23 (2014)

[16] Liszio, S., Emmerich, K., Masuch, M.: The influence of social entities in virtual reality games on player experience and immersion. In: Proceedings of the 12th International Conference on the Foundations of Digital Games, pp. 1–10 (2017)

[17] Tracy, K., Dalian, W., Dari, R., Spantidi, O.: "are we going to crash?": The effect of an ai presence in user experience during a vr flight. In: International

Symposium on Visual Computing, pp. 468–479 (2025). Springer

[18] McNally, R.J., Louro, C.E.: Fear of flying in agoraphobia and simple phobia: distinguishing features. Journal of Anxiety Disorders **6**(4), 319–324 (1992)

[19] Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A.: Measuring and defining the experience of immersion in games. International Journal of Human-Computer Studies **66**(9), 641–661 (2008) https://doi.org/10.1016/j.ijhcs.2008.04.004

[20] Mousas, C., Anastasiou, D., Spantidi, O.: The effects of appearance and motion of virtual characters on emotional reactivity. Computers in Human Behavior **86**, 99–108 (2018) https://doi.org/10.1016/j.chb.2018.04.036

[21] Tracy, K., Spantidi, O.: Impact of gpt-driven teaching assistants in vr learning environments. IEEE Transactions on Learning Technologies **18**, 192–205 (2025) https://doi.org/10.1109/TLT.2025.3539179

[22] Obremski, D., Wienrich, C.: Autonomous vr exposure therapy for anxiety disorders using conversational ai—ethical and technical implications and preliminary results. Journal of Computer and Communications **12**(4), 111–138 (2024)

[23] Rothbaum, B.O., Anderson, P., Zimand, E., Hodges, L., Lang, D., Wilson, J.: Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. Behavior Therapy **37**(1), 80–90 (2006) https://doi.org/10.1016/j.beth.2005.04.004

[24] Tracy, K., Kouzelis, L.R., Dari, R., Spantidi, O.: Planogen: A procedural generation framework for dynamic vr research environments. Virtual Worlds **4**(33) (2025) https://doi.org/10.3390/virtualworlds4030033

[25] Bowman, D.A., McMahan, R.P.: Virtual reality: How much immersion is enough? Computer **40**(7), 36–43 (2007) https://doi.org/10.1109/MC.2007.257

[26] Leon, A.C., Davis, L.L., Kraemer, H.C.: The role and interpretation of pilot studies in clinical research. Journal of Psychiatric Research **45**(5), 626–629 (2011) https://doi.org/10.1016/j.jpsychires.2010.10.008