Piloting Planetarium Visualizations with LLMs during Live Events in Science Centers

Mathis Brossier[1], Mujtaba Fadhil Jawad[1], Emma Broman[1], Ylva Selling[1], Julia Hallsten[2], Alexander Bock[1], Johanna Björklund[3], Tobias Isenberg[4], Anders Ynnerman[1], Mario Romero[1], Lonni Besançon[1]

[1]Linköping University, Sweden [2]Visualisering Center C, Sweden [3]Umeå University, Sweden [4]Université Paris-Saclay, CNRS, Inria, LISN, France

Abstract

We designed and evaluated an AI pilot in a planetarium visualization software, OpenSpace, for public shows in science centers. The piloting role is usually given to a human working in close collaboration with the guide on stage. We recruited seven professional guides with extensive experience in giving shows to the public to study the impact of the AI-piloting on the overall experience. The AI pilot is a conversational AI-agent listening to the guide and interpreting the verbal statements as commands to execute camera motions, change simulation time, or toggle visual assets. Our results show that, while AI pilots lack several critical skills for live shows, they could become useful as co-pilots to reduce workload of human pilots and allow multitasking. We propose research directions toward implementing visualization pilots and co-pilots in live settings.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in interaction design;*** • ***Computing methodologies*** → *Natural language processing;* • ***Applied computing*** → *Interactive learning environments;*

1. Introduction

The development of LLMs and foundation models capable of reasoning about complex tasks facilitates a natural interaction between machines and humans. Uses of AI assistants are flourishing in various applications of visualization and interaction [BIS*26], such as data science [CWG*25], operative surgery [HWYZ25], forensics [PCB*24], or science communication [JIB*25]. LLM applications are also becoming ubiquitous in academia [SKE*25] and in society [DLA*25; CHGD24]. This development, however, remains controversial within the scientific community [BAR*25] and the public at large [ENBS24]. Beyond societal considerations, the interplay between users and LLM assistants is not well understood, especially when they exhibit anthropomorphic features [SZW*26]. A key limitation of current AI-naturalness in collaboration with humans is the inability of LLM agents to be proactive in a conversation [CWZ*25]. Agent autonomy is an active research topic, such as to improve the *reasoning* autonomy of LLMs [FTD25], but more research is needed to explore *interaction* autonomy. In visualization, proactive agents could support explorative approaches, by showing users contextually relevant visualization without explicit instructions.

Working toward this goal, we implemented system for mixed-initiative visualization steering, that leverages the complementary strengths of human and AI operators. We focus on live visual storytelling events using interactive visualization, driven by a human guide on stage, and a visualization pilot. Our system builds on OpenSpace [BAC*20], an astrophysics visualization software commonly used in planetarium live shows.

To further explore the topic of pragmatics and proactive AI, we compare two modes of operation: a *reactive* mode, in which the AI agent responds to explicit and directed queries; and a *proactive* mode, in which the agent intervenes to perform actions while the human guide is talking, without explicit queries. We conducted a comparative study with five professional planetarium guides and we interviewed two additional experts to better understand human-driven visualization piloting and presentation during live events. Noteworthy, the experts have extensive experience with human pilots and are able to provide comparative observations against the two AI conditions. We explore the strengths and weaknesses of AI-driven visualization steering in both modes, and contribute research directions stemming from the study and participants' experiences as both public guides and visualization pilots. We center our exploration around three research questions:

- RQ1: What are the strengths and weaknesses of AI co-pilots for visualization steering?
- RQ2: What are the strengths and weaknesses of proactive AI pilots compared to reactive AI pilots?
- RQ3: What further developments and research directions should be explored for AI co-piloting in visualization?

2. Related work

Different dimensions of using conversational agents in visualization have been studied [BIS*26]. Most prominently, using natural language to query data, (e.g., text2sql tasks [HYZ*25]) and visualization recommendation (e.g., generation of visualization via code [BM24]) has been studied extensively. Visualization reading has also been studied [DSD*21]. LLMs show promises in tackling domain-specific visualization tasks, e.g. volume rendering [ATW26]. We focus here on an underexplored dimension: using natural language navigation through LLM-augmented visualization software [JIB*25].

Applications of proactive (or autonomous) agents are emerging. While the terminology is not yet fully established, we chose the term "proactive," as it counter-balances well with "reactive." The term is used in context-aware models [YXZ*25], video streaming models [ZSWY25], as well as in applications to visualization [TLT*25]. The first workshop focusing on the topic of proactive LLMs occurred in 2025 [CWZ*25]. We take inspiration from the work by TABALBA et al. [TLT*25], who studied proactive agents in visualization for collaborative work, and JIA et al. [JIB*25], who studied visualization navigation with conversational agents. We combine these approaches, and therefore study proactive visualization navigation during live educational shows about the solar system.

3. System implementation

We build upon our previous early prototype of LLM-steered visualization implementations [BBS*24], which we have significantly expanded it to support the *reactive* and *proactive* modes, described next. We rely on the astrophysics visualization framework OpenSpace [BAC*20] for the rendering. OpenSpace is used notably in planetariums during "tour of the universe" live events with broad audiences but which has generic rendering and explaining capabilities including, e.g., molecular dynamics in contexts [BSB*23]. We control OpenSpace via its Lua API over WebSocket. We use a separate JavaScript process to coordinate between an LLM agent, OpenSpace's API, and peripherals (microphone, remote trigger and graphical interface). We employ OpenAI's low-latency, multimodal model (`gpt-realtime-2025-08-28`) for the natural language reasoning tasks. The LLM model is capable of dispatching *tool calls* to travel to different scene nodes, toggle the visibility of nodes, travel to a geolocation (lat, long, altitude), and change the simulation time and speed. The LLM can often loose track of the instructions during longer sessions. To mitigate this issue, we add few-shot examples as warm-up at the start of conversations, delete undesired model-generated textual messages, and add a short instruction reminder message at the tail of the conversation thread. We provide the LLM prompt structure and textual template in supplemental material.

In *reactive* mode, we feed the audio input from the microphone directly to the LLM. In *proactive* mode, however, we employ a *streaming* speech-to-text technique developed by MACHÁČEK et al. [MP25], which allows us to prompt the LLM in a loop, with a stream of transcribed text chunks. Text chunks comprise typically 1 to 4 words with a latency of around 500ms. We instruct the LLM to dispatch a tool call at each prompt loop turn. If the LLM judges that no action is needed, it produces a special "no-op" call that we simply ignore. In both modes, the LLM does not produce any textual or audio output in addition to dispatching tool calls. This setup imitates the relation between a human pilot and a guide, where only one-way verbal communication from the guide to the pilot is possible.

4. Study methodology

Our study consisted of five one-hour sessions, one for each of five participants in a full-dome planetarium at the Visualization Center C in Norrköping, Sweden. We compared usage of our system of two cases, (*reactive* vs. *proactive*) against the human baseline, which we established through interviews with two additional consulting experts. We collected both quantitative data from the system usage and qualitative data from participant feedback. We performed a thematic analysis from the interviews and an error-case analysis from the system operation logs (subsection 5.4).

4.1. Participants

We recruited five expert participants who are professionals working at a science center. All the participants had experience presenting in the dome to public audiences. Two (P1 and P2) were educators at the facilities and had the most experience engaging with visitors, but a limited technical background. The other three (P3–P5) were engineers and researchers working on the dome software development and had also presented several dome shows. In addition, we recruited two consulting experts with leading roles in science communication (C1) and dome software development (C2). They provided us with qualitative feedback during longer semi-structured interviews on a typical dome show experience with a public audience, including elements that make shows engaging, guide $\leftrightarrow$ pilot teaming, planning of shows, and their opinions on AI. All participants are working at our institution and are exempted from requiring ethical approval or getting paid to participate in the study. Co-authorship was offered to the participants after their individual participation, as is sometimes the case in studies involving domain experts contributing their time (e.g., [RPH*21; SMM12]).

4.2. Study protocol

We ran one-hour sessions with each participant (P1–P5) in the dome. We asked the participants to conduct two simulated dome shows to be able to compare the two interaction modalities (*reactive* vs. *proactive*) with the LLM assistant. We followed with a 20-minute semi-structured interview. Due to some instability with the software, the first author had to intervene several times during the presentation. For participant P4, we could not reliably run the software, so we instead decided to have them be spectator of P5's experience and we interviewed both participants together.

For the simulated show we used a shortened version of a typical show, which consists of a "tour of the Solar System," starting from a location on Earth and visiting the moons and planets of the Solar System. We gave participants freedom to explore and discuss to let them understand and adapt to the system's functionality and limitations. We compared two system uses: in *reactive* mode the guides held a hand-triggered microphone, which they could use to perform direct voice commands such as "Go to the Moon." In *proactive* mode the system was constantly reacting to the guide's speech, without an explicit trigger. Mentioning the Moon, e.g., would initiate a flight to it. Each test case lasted up to 10 minutes. We alternated the starting case between participants to limit bias. P1, P3, and P5 started with the *reactive* case, P2 and P4 with the *proactive* case. After the first test case, we conducted a short ($\approx$5min) unstructured interview to collect the participant's immediate feedback while preparing the system for the next test case. After the second case, we conducted a 20-minute semi-structured interview.

4.3. Data collection and analysis

We collected both data from the system logs, as well as the qualitative feedback from the participant interviews. From the system we logged the session audio and transcripts, the end-to-end latency, and the nature, number and success rate of system interventions. At this stage, we focus on reporting on the insights from the interviews.

5. Results

5.1. Comparison with a human pilot

Regarding their expectations and enjoyment, some participants expressed clear enjoyment. It is a "cool concept" (P1), "more fun than anticipated. [The system is] listening to me and [doing] stuff as I am talking" (P2). We then asked participants about the current and projected capabilities of AI pilots in dome shows. All participants agree that AI pilots are missing many nuances compared to humans, which explains that while they sometimes behave correctly, they are far from having the robustness to sustain a full show autonomously. P1: "A human pilot can guess what I want to say and anticipate". P2 adds: "sometimes we open questions from the audience [...]. With specific things it may fail".

We examined what makes the piloting role fundamentally human. Expert participants (P1, P2, C1), emphasized the importance of explicitly acknowledging the pilot during the show. Being a live event from both the guide and the pilot sides, there is "something to it" as a show for an audience (P2). It "underlines that it's a live presentation" and it "compensates for the limited interaction with the audience" (C1). Two participants (P1, P3) also explained that the AI lacked proper *pacing*: "it's all about the pace. If I talk about the ISS for 5 minutes it's going to be boring. But the first 5 minutes traveling slowly there would make it more interesting. Buildup" (P3). The pilot is constantly moving the camera, like driving a car. As in movies, the scene composition and motion also supports storytelling and immersion, and camera control must be precise, as jittering or too fast movements can cause discomfort. Due to their poor temporal awareness and their inherent latency, LLM systems struggle with reactive, fluid, and precise camera control.

We further asked about the relationship between the guide, the pilot and the public. There are only limited interactions possible, but they are nonetheless important. The guide expert (C1) explained that perceiving the audience's engagement is possible, yet hard to formalize. A strong signal is the breathing (gasps, silences), as well as feeling "something in the air." Reaction to jokes is the easiest way to probe for engagement. Such cues are difficult to measure for any machine. From the booth behind the stage where the pilot sits, it is very hard to read anything from the audience.

5.2. Comparison between reactive and proactive cases

The preference between the two modes was mixed. Due to intermittent system failures, it was sometimes difficult for participants to imagine the best-case scenario. P1 preferred the proactive mode, as it felt more like interacting with a human because of reading context cues. P2, P3, and P5 preferred the reactive mode because it was more predictable and reliable. The reactive mode was more reliable, but added to the mental load of participants. There was better flow without direct questions and addressing to the AI. Further, remembering to press the trigger increased cognitive load which made it harder to sustain conversation flow (P1, P2). The proactive mode thus felt more fluid or natural, but less reliable in live conditions.

5.3. Combining human and AI

A future approach could lie in piloting cooperation between a human and an AI agent. Three directions emerged from our study. First, we established that AI was rather good at toggling assets based on context, while it performed poorly with camera motions and pace control. Conversely, human pilots have to simultaneously steer the camera and toggle assets. An AI co-pilot could assist by preparing the next actions ahead of time. To illustrate, P3 shared an anecdote: "[the public was] excited about satellites which we didn't prepare for. I was frantically adding assets about satellites. It would be great if AI could add all of them at once." Second, utilizing speech control in combination with keyboard and mouse could facilitate pilots' multitasking abilities, especially for tasks easy to describe with words, but harder to actuate with traditional inputs, e.g., "show all available satellites at once." Finally, LLM agents can be helpful during on-boarding and training of new pilots, as well as preparation of controls ahead of the show, a step which is currently very time consuming when experimenting with new live shows.

5.4. Error case analysis

Action correctness is a spectrum. One may argue that a direct query, such as "Show me the Earth" is objective and there is no interpretation possible. In 3D visualization software, however, the level of zoom and the location in focus have a definitive impact on the quality of the action performed. Indeed, we observed that our system frequently zoomed in on a location that lies on the dark side of the planet, or is too zoomed-in/out to show an overview or the details of interest (see Figure 1). A qualitative analysis of error cases (both their characteristics, and prevalence) is needed to inform on the cases where LLM-based navigation or interaction can fail. From our observations, we characterize a spectrum comprising the following error dimensions for AI pilot actions, with examples from our study:

- **Detection.** How sensitive and specific action intent detection is. *Failure example:* Mishearing the guide. *Means of evaluation:* counting true positives, false positives, and guide intents.
- **Reasoning.** How well the system understands users' implicit or explicit intent. *Failure example:* Asking to show the Moon "as we see it from Earth". *Means of evaluation:* can use reasoning benchmarks [LHZ*25].
- **Context.** Contextual elements include current visualization state, audience, and conversation history. *Failure example:* Showing the dark side of a planet. *Means of evaluation:* reasonably can only be coded by humans, but can be trained on examples.
- **Naturalness.** Whether the action feels human-like. *Failure example:* Poor steering of the view. *Means of evaluation:* subjective human assessment.

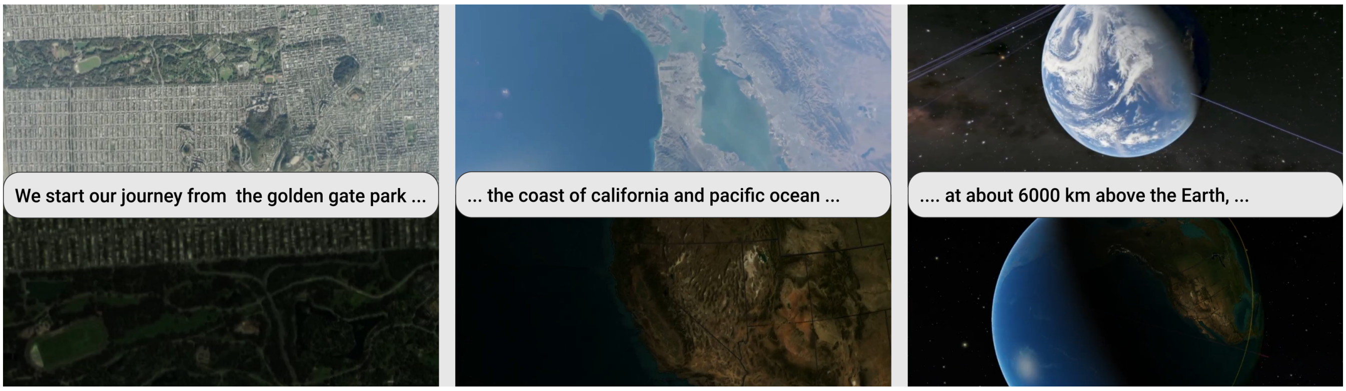

Figure 1: *Comparison of visualizations produced by a human guide (top) and an LLM proactive mode (bottom) from a planetarium show recording. While the LLM selects correct locations, the visualization is often too dark (nighttime) and sometimes imperfectly zoomed.*

6. Limitations

Several limitations of our study can impact the transferability and reproducibility of the work. First, the software instability. There were several instances where the system would become unresponsive, in which case an engineer intervened to restart it. We communicated this limitation to the participants during briefing. Second, the small sample size (n=7) of participants from the same institution. We thus frame this work as an explorative study, that points at several research directions built from the experience of participants and our observation, rather than making statistical and generalizable claims. Still, although our sample size was small, this is common in studies involving expert participants in our field; moreover, the expertise of all participants makes their feedback particularly valuable and allows us to draw important insights into the use of LLMs as facilitators for live science shows.

7. Directions for future work

In light of our experiment and interviews, we establish the following directions for future work.

7.1. Visualization interaction recommender systems

A clear feedback from the participants was that the AI co-pilot cannot replace human pilots (and it was not our intent). However, they may prove useful as recommender systems to, for example, speed up human pilots and reduce their cognitive load by digging through assets, and simplifying the on-boarding of novice pilots. Visualization recommender systems (i.e., capable of producing visualizations based on user requests), including natural language and LLM-based ones, have been studied extensively [VHS*17]. Recommender systems for *interaction* specifically, have, to our knowledge, not been studied before. Such *interaction recommender systems* would consist of an interface which suggests triggerable actions based on the current context, possibly including the guide's speech, and other interactions and context cues such as gestures.

7.2. Multitasking with multimodality

Participants highlighted that multitasking was frequent during shows. Typically, the pilot would control the camera path and prepare the upcoming assets simultaneously. This problem is amplified by the need for constant motion. Continuous rotation around objects improves depth perception (P4) and audience engagement. To facilitate multitasking, OpenSpace's camera motion has inertia, which allows the pilot to set it in motion and let go of the controls shortly while focusing on another task. When asked about their opinion on multimodal keyboard/mouse and speech interaction, P4 and C2 indicated that the current workflow is sufficient and that speaking out loud in the booth could disturb the audience.

7.3. Use of AI during show preparations

Live events are always planned by both the pilot and the guide. According to P3 and C2, live shows are composed of small bricks of storytelling assembled together. Adapting to slightly different shows is easy for the pilot. Designing new assets and storylines, however, requires more work and software knowledge. An AI pilot could help during the testing and brainstorming in a low-risk environment.

8. Conclusion

We explored the concept of LLM visualization pilots for guided interactive presentations. Our study indicates that, while such pilots are far from reaching the capabilities of their human counterparts, they may find uses as co-pilots, with aims of reducing the cognitive load, enabling multitasking, and facilitating on-boarding of new pilots. Such systems, functioning in the background, need greater levels of autonomy than typical conversational chatbots, for which interaction patterns are too obtrusive in live events. Our results highlight that autonomous LLMs appear as an interesting research direction to facilitate discrete interaction with visualization by guides during live events in education contexts.

9. Acknowledgments

This work was partially supported by the Marcus and Amalia Wallenberg Foundation (2023.0130), and the Knut and Alice Wallenberg Foundation (2019.0024). We thank NVAB and all the dome-supporting crew at Visualisering Center C.

References

[ATW26] AI, KUANGSHI, TANG, KAIYUAN, and WANG, CHAOLI. "NLI4VolVis: Natural Language Interaction for Volume Visualization via LLM Multi-Agents and Editable 3D Gaussian Splatting". *IEEE Trans Vis Comput Graph* 32.1 (Jan. 2026), 46–56. ISSN: 1941-0506. DOI: 10.1109/TVCG.2025.3633888. (Visited on 03/27/2026) 2.

[BAC*20] BOCK, ALEXANDER, AXELSSON, EMIL, COSTA, JONATHAS, et al. "OpenSpace: A system for astrographics". *IEEE Trans Vis Comput Graph* 26.1 (2020), 633–642. DOI: 10/ggsrbm 1, 2.

[BAR*25] BINZ, MARCEL, ALANIZ, STEPHAN, ROSKIES, ADINA, et al. "How should the advancement of large language models affect the practice of science?": *PNAS* 122.5 (2025), art. no. e2401227121. DOI: 10/g9mc86 1.

[BBS*24] BROSSIER, MATHIS, BOCK, ALEXANDER, SCHÖNBORN, KONRAD, et al. "In space, no one (but AI) can hear you scream". *Posters at IEEE VIS*. 2024. DOI: 10/hbk46z 2.

[BIS*26] BROSSIER, MATHIS, ISENBERG, TOBIAS, SCHÖNBORN, KONRAD, et al. *State of the art of LLM-enabled interaction with visualization*. arXiv preprint 2601.14943. 2026. DOI: 10/hbk462 1.

[BM24] BASOLE, RAHUL C. and MAJOR, TIMOTHY. "Generative AI for visualization: Opportunities and challenges". *IEEE Comput Graph Appl* 44.2 (2024), 55–64. DOI: 10/nm2j 2.

[BSB*23] BROSSIER, MATHIS, SKÅNBERG, ROBIN, BESANÇON, LONNI, et al. "Moliverse: Contextually embedding the microcosm into the universe". *Comput Graph* 112 (2023), 22–30. DOI: 10/gr7bbz 2.

[CHGD24] CHKIRBENE, ZINA, HAMILA, RIDHA, GOUISSEM, ALA, and DEVRIM, UNAL. "Large language models (LLM) in industry: A survey of applications , challenges, and trends". *Proc. HONET*. Los Alamitos: IEEE Computer Society, 2024, 229–234. DOI: 10/qn2d 1.

[CWG*25] CHEN, JUNTONG, WU, JIANG, GUO, JIAJING, et al. "InterChat: Enhancing generative visual analytics using multimodal interactions". *Comput Graph Forum* 44.3 (2025), art. no. e70112. DOI: 10/g9qr4q 1.

[CWZ*25] CHATTERJEE, SHUBHAM, WANG, XI, ZHANG, SHUO, et al. "ProActLLM: Proactive conversational information seeking with large language models". *Proc. CIKM*. New York: ACM, 2025, 6894–6897. DOI: 10/qn2b 1, 2.

[DLA*25] DREKSLER, NOEMI, LAW, HARRY, AHN, CHLOE, et al. *What does the public think about AI? An overview of the public's attitudes towards AI and a resource for future research*. Social Science Research Network preprint 5108572. 2025. DOI: 10/qn2c 1.

[DSD*21] DAVILA, KENNY, SETLUR, SRIRANGARAJ, DOERMANN, DAVID, et al. "Chart mining: A survey of methods for automated chart analysis". *IEEE Trans Pattern Anal Mach Intell* 43.11 (2021), 3799–3819. DOI: 10/gh6rcz 2.

[ENBS24] EOM, DAYEON, NEWMAN, TODD, BROSSARD, DOMINIQUE, and SCHEUFELE, DIETRAM A. "Societal guardrails for AI? Perspectives on what we know about public opinion on artificial intelligence". *Sci Public Policy* 51.5 (2024), 1004–1013. DOI: 10/qnz8 1.

[FTD25] FERRAG, MOHAMED AMINE, TIHANYI, NORBERT, and DEBBAH, MEROUANE. *From LLM reasoning to autonomous AI agents: A comprehensive review*. arXiv preprint 2504.19678. 2025. DOI: 10/hbk46w 1.

[HWYZ25] HAO, PENGFEI, WANG, HONGQIU, YANG, GUANG, and ZHU, LEI. "Enhancing visual reasoning with LLM-powered knowledge graphs for visual question localized-answering in robotic surgery". *IEEE J Biomed Health Inf* 29.12 (2025), 9027–9040. DOI: 10/qnz7 1.

[HYZ*25] HONG, ZIJIN, YUAN, ZHENG, ZHANG, QINGGANG, et al. "Next-Generation Database Interfaces: A Survey of LLM-Based Text-to-SQL". *IEEE Tran Knowl Data Eng* 37.12 (2025), 7328–7345. DOI: 10/hbjqnp 1.

[JIB*25] JIA, DONGGANG, IRGER, ALEXANDRA, BESANÇON, LONNI, et al. "VOICE: Visual oracle for interaction, conversation, and explanation". *IEEE Trans Vis Comput Graph* 31.10 (2025), 8828–8845. DOI: 10/g9r73s 1, 2.

[LHZ*25] LU, JIARUI, HOLLEIS, THOMAS, ZHANG, YIZHE, et al. "ToolSandbox: A stateful, conversational, interactive evaluation benchmark for LLM tool use capabilities". *Proc. NAACL*. Kerrville: ACL, 2025, 1160–1183. DOI: 10/qn24 3.

[MP25] MACHÁČEK, DOMINIK and POLÁK, PETER. *Simultaneous translation with offline speech and LLM models in CUNI submission to IWSLT 2025*. arXiv preprint 2506.17077. 2025. DOI: 10/hbk46x 2.

[PCB*24] POORYOUSEF, VAHID, CORDEIL, MAXIME, BESANÇON, LONNI, et al. "Collaborative forensic autopsy documentation and supervised report generation using a hybrid mixed-reality environment and generative AI". *IEEE Trans Vis Comput Graph* 30.11 (2024), 7452–7462. DOI: 10/g8p63q 1.

[RPH*21] ROGERS, JEN, PATTON, AUSTIN H., HARMON, LUKE, et al. "Insights from experiments with rigor in an EvoBio design study". *IEEE Trans Vis Comput Graph* 27.2 (2021), 1106–1116. DOI: 10/ghgt5p 2.

[SKE*25] SAJJADI MOHAMMADABADI, SEYED MAHMOUD, KARA, BURAK CEM, EYUPOGLU, CAN, et al. "A survey of large language models: Evolution, architectures, adaptation, benchmarking, applications, challenges, and societal implications". *Electronics* 14.18 (2025), art. no. 3580. DOI: 10/hbd2sw 1.

[SMM12] SEDLMAIR, MICHAEL, MEYER, MIRIAH, and MUNZNER, TAMARA. "Design study methodology: Reflections from the trenches and the stacks". *IEEE Trans Vis Comput Graph* 18.12 (2012), 2431–2440. DOI: 10/f4fv7x 2.

[SZW*26] SCHLESENER, ELIZABETH A., ZIOLKOWSKI, MARCIN, WONG, SAI KEUNG, et al. "'Am I understood?': How the interplay between embodiment and theory of mind behavior affects LLM-based conversational agents on perceived trust, anthropomorphism, presence, usability, and user experience". *ACM Trans Interact Intell Syst* (2026). To appear. DOI: 10/qnz9 1.

[TLT*25] TABALBA, RODERICK S., LEE, CHRISTOPHER J., TRAN, GIORGIO, et al. "A pragmatics-based approach to proactive digital assistants for data exploration". New York: ACM, 2025. DOI: 10/p9c2 2.

[VHS*17] VARTAK, MANASI, HUANG, SILU, SIDDIQUI, TARIQUE, et al. "Towards visualization recommendation systems". *ACM SIGMOD Rec* 45.4 (2017), 34–39. DOI: 10/gfxrs9 4.

[YXZ*25] YANG, BUFANG, XU, LILIN, ZENG, LIEKANG, et al. "ContextAgent: Context-aware proactive LLM agents with open-world sensory perceptions". *NeurIPS Posters*. 2025. DOI: 10/hbk463 2.

[ZSWY25] ZHANG, YULIN, SHI, CHENG, WANG, YANG, and YANG, SIBEI. *Eyes wide open: Ego proactive video-LLM for streaming video*. arXiv preprint 2510.14560. 2025. DOI: 10/hbjqp6 2.