



OPEN ACCESS

EDITED BY

Andrea Sanna,
Polytechnic University of Turin, Italy

REVIEWED BY

Dian Novian,
Universitas Negeri Gorontalo, Indonesia

*CORRESPONDENCE

Mar Gonzalez-Franco,
✉ margon@google.com

RECEIVED 19 June 2025

ACCEPTED 23 July 2025

PUBLISHED 09 September 2025

CITATION

Suzuki R, Abtahi P, Zhu-Tian C, Dogan MD,
Colaco A, Gonzalez EJ, Ahuja K and
Gonzalez-Franco M (2025)
Programmable reality.
Front. Virtual Real. 6:1649785.
doi: 10.3389/frvir.2025.1649785

COPYRIGHT

© 2025 Suzuki, Abtahi, Zhu-Tian, Dogan,
Colaco, Gonzalez, Ahuja and Gonzalez-Franco.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Programmable reality

Ryo Suzuki¹, Parastoo Abtahi², Chen Zhu-Tian³,
Mustafa Doga Dogan⁴, Andrea Colaco⁵, Eric J. Gonzalez⁵,
Karan Ahuja^{5,6} and Mar Gonzalez-Franco^{5*}

¹Computer Science Department, ATLAS Institute, University of Colorado, Boulder, CO, United States, ²Computer Science Department, Princeton University, Princeton, NJ, United States, ³Computer Science and Engineering Department, University of Minnesota, Minneapolis, MN, United States, ⁴Google XR, Adobe Research, Basel, Switzerland, ⁵Google, Mountain View, CA, United States, ⁶Department of Computer Science, Northwestern University, Evanston, IL, United States

Innovations in spatial computing and artificial intelligence (AI) are making it possible to overlay dynamic, interactive digital elements on the physical world. Soon, every object might have a real-time digital twin, enabling the “Internet of Things” so as to identify and interact with even unconnected items. This programmable reality would enable computational manipulation of the world around us through alteration of its appearance or functionality, similar to software, but for reality itself. Advances in AI language models have enabled zero-shot segmentation and understanding of the world, making it possible to query and manipulate objects with precision. However, this vision also demands natural and intuitive ways for humans to interact with these models through gestures, gaze, and existing devices. Augmented reality (AR) provides the ideal bridge between AI output and human input in the physical world. Moreover, diffusion models and physics simulations offer exciting possibilities for content generation and editing, allowing us to transform everyday activities into extraordinary experiences. As AR devices become ubiquitous and indistinguishable from reality, these technologies blur the lines between reality and simulations. This raises profound questions about how we perceive and experience the world while having implications for memory, learning, and even behavior. Programmable reality enabled by AR and AI has vast potential to reshape our relationships with the digital realm, ultimately making it an extension of the physical realm.

KEYWORDS

augmented reality, extended reality, artificial intelligence, virtual reality, generative artificial intelligence

1 Unlocking programmable reality

The concept of programmable reality, where the physical world meets the programmable flexibility of the digital realm, might sound like science fiction; however, innovations in augmented reality (AR) and artificial intelligence (AI) are rapidly converging to create environments where the physical and digital elements interact seamlessly and where digital elements can be integrated into or overlaid on the physical world in a dynamic and interactive manner. In this environment, all objects around us are classified, segmented, identified, shareable, and interactable; further, every analog object can become a part of the Internet of Things (IoT) even when it is not truly connected to the internet, has no embedded chip, and is simply sensed and identified by a third entity that helps them connect. As such, every object, person, and space can be considered to have a real-time digital twin (von Willich et al., 2023).

This article aims to provide a vision of this space for the future and help the larger community inform its own roadmap and research objectives. Although significant works are being produced in the extended reality (XR) and AI space (Suzuki et al., 2023) along with some reviews (Hirzle et al., 2023), there is still a need for articles on the envisioned future. As we move from digital content behind screens to users perusing such content in the era of spatial computing, we can still see that all this content is simply overlaid on our physical surroundings. Although spatial computing may be immersive in many cases, it still appears to be detached from reality. The dream of a programmable reality is to interact with physical reality so that it can be computationally manipulated and dynamically altered in a manner similar to programming software.

On the one hand, the key to unlocking programmable reality lies in a complete high-fidelity understanding of the real world. This means complete understanding and segmentation of the world in a zero-shot manner, with full object–vocabulary access that is being made available with large AI models, as well as new discoveries on the understanding acquired by these models through extensive training. One example of this is the DiffSeg approach to segmentation by clustering of the attention layers (Tian et al., 2024) along with multimodal large language models (LLMs) available with very large sets of tokens for the prompts (Team, 2024); these can be used to provide detail context to the models in a multimodal manner while reducing the need to retrain them repeatedly. Most recently, even time-series models are becoming zero-shot capable with techniques that instrumentalize multiple tasks into a single model, such as the TOTEM architecture (Talukder et al., 2024), thereby providing frameworks for real artificial general intelligence (Morris et al., 2023) and even direct brain interfaces. However, these models stay within a box on a Python notebook in the cloud unless we can interact with them easily. This vision of programmable reality is not one where we can prompt the model using commands on ChatGPT but instead with natural tools used by people that show attention and intent, such as hand and body gestures or gaze, along with existing devices like phones, mice, and keyboards (Gonzalez et al., 2024). These are not efforts shown on a screen but in the real world, through context along with precision, gaze, and pinch actions. For instance, if we are to look at a random object like a toaster and then select and apply an AI model to it to change its appearance or functionality, then we need to have an output that is far more elaborate than a text response to a prompt. Hence, we need to be able to access both the actual digital and real contents with representations that are familiar to the people interacting with the models, such as user interfaces (UIs) or spatial overlays. The core challenge in human–computer interaction (HCI) is in bridging the gap between human thoughts and machine computations. Traditionally, computer science has relied on layers of representational interfaces (such as UIs) to facilitate such interactions. However, with the advent of AI, we have seen a shift toward prompt-based interactions reminiscent of the MS-DOS era, with some users even suggesting total elimination of UIs or replacement of the operating system. Although we agree that AI necessitates a rethinking of our interfaces, the basic interaction model of ChatGPT highlights the critical role of representation in shaping our capacity to engage with computers. Representations that hinder basic cognitive functions like memory,

discovery, and articulation will directly limit our goals with AI. Conversely, well-designed representations will foster intuitive interactions, allowing us to effectively communicate our intentions. An AR can be the right kind of vessel to channel AI outputs and inputs to the appropriate places; this shows the crucial roles of AR and HCIs in successful adoption of AI, which are inseparable sides of the same coin from our perspective as well as necessary for programmable reality.

On the other hand, the final key to converting such a future into full-fledged programmable reality is the capacity to multimodally add, replace, crop, or edit any content so as to clearly differentiate between realities when wearing a set of AR devices versus removing them. Such editing and content generation are now possible with diffusion models such as VEO3 (Bar-Tal et al., 2024) or actual physics-based simulation engines like SORA. This would enable adding dynamics to scenes, changing the style of our total world, living in a comic or inside a Van Gogh painting for an entire day, but not in the metaverse or in the real world, where we could go about our regular human social activities of playing sports, shopping, and hanging out with friends or even working. The quality of passthrough observable in recent times is almost perfect, so much so that it enables us to experience a full life without ever removing these devices. However, if we should choose to do so, there is the option to turn off the passthrough while simply filtering and stylizing the entire reality. If a person were to utilize the full potential of the scene toolings described herein while combining the content generation capabilities with scene segmentation and understanding, the device can completely interface between perception and the real world, akin to a parallel life inside a simulation for those wearing these devices non-stop.

Because of the nature of the first-person experiences provided by these devices along with the bottom-up sensory feeding and perfectly closed loops of motor control owing to the interactivity, it is possible to reduce the body semantic violations to zero (Padrao et al., 2016). At this point, only the top-down higher cognitive function mechanisms will be able to remind us that we are still in a simulation; because we do not relinquish our higher cognition, it is always possible to remember that we can remove these devices. The nature of devices that interface between the real world and our perceptions is such that our brains will still function normally to create experiential memories and learnings, while gaining a secondary set of life experiences that are partially to fully detached from reality. Such devices have the potential to behaviorally train users completely if they were to truly live with the devices on 24/7. Nature was the main behavioral training tool for humans in the past, but this could be replaced by programmable reality in the future.

2 AR is to AI what screens are to computers

Programmable reality represents a remarkable convergence of various technological streams and promises to reshape our interactions with the digital world. Advances in AR and virtual reality (VR) technologies are the cornerstone of programmable reality, offering immersive experiences that seamlessly blend the



FIGURE 1
Examples of some end uses of programmable reality, where people can perform any of their current life tasks in a physical reality that is blended, dynamic, and pervasive with a generated and programmed digital world.

real and digital. Additionally, AI and machine learning (ML) are crucial for creating intelligent and responsive environments that adapt to user interactions. AI provides the backbone for both scene understanding and adaptive content generation, in addition to complex user digitalization for context awareness and dynamic interactions. The concepts can be extended further to areas like real-life versions of digital twins (virtual replicas of physical entities) (von Willich et al., 2023), smart materials with changeable properties on command (Steed et al., 2021), and advanced robotics and AI systems that interact with the physical world in sophisticated ways, e.g., understanding chains of actions, context, and physics (Battaglia et al., 2013). The core idea here is that the elements of our physical environment, whether they be objects, spaces, or even biological entities, can be controlled, transformed, or experienced in new ways through programmable interfaces and engaged with instead of simply being consumed.

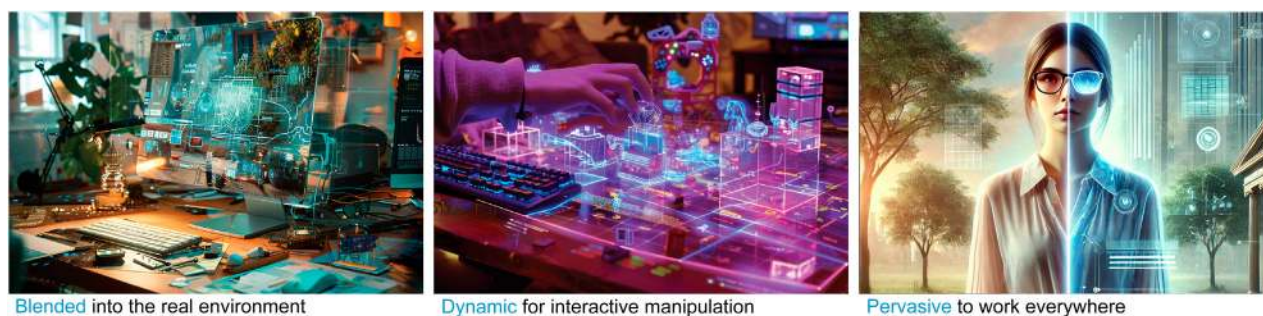
Thus, we consider three principles of a programmable reality enabled by AR and AI: (1) it needs to blend with the real world and not just be overlaid; (2) it needs to be dynamic and interactive, so simple scripted queries from a chatbot such as “What is this?” are not enough; (3) it needs to be pervasive and capable of being always ON to work anywhere (Figure 1). We explore these concepts and create definitions to clarify the need for the capacity to modify our own media as a realistic world wide web that will emerge through immersive technology and augmented devices as they become broadly available in the future. Programmable reality could perhaps become the major form of interaction with digital content and AI for people in their daily lives. To elucidate and provide a possible example of such a world, we present a mock prototype of how a user might program their reality using simple tooling in Figure 2.

3 Interactivity levels

Because humans interact with the environment physically, creating a programmable reality involves categorizing the different ways in which users can engage with and manipulate their environment in a digitally augmented or programmable context. The aspects of such an interactive space include new input and output methods that are subtle yet easily accessible (Pfeuffer et al., 2024; Mao et al., 2025); these aspects also allow interactions with existing objects to blend the range encompassing the world, objects, and devices with the human body. Consider creating a chain of physics understanding capable of telling us the temperature of our coffee, not because of some “super” vision capability but as a human using heuristics, along with the capacity to abstract layers of information and memory by considering the volume, brewing time, and when it was brewed (Zhu-Tian et al., 2023). Above all, the interactivity will be defined by the type of programmability that we provide to the physical world. Accordingly, we note four main levels of interactivity that are becoming popular research spaces, namely static augmentation, dynamic augmentation, dynamic interaction, and collaborative environments.

3.1 Static augmentation

Fixed digital enhancements are possible for a major portion of our physical reality, even parts that comprise analog objects. The main aspect of this type of augmentation is the identification of objects and their contexts, which can be achieved through multimodal LLMs (Abreu et al., 2025), computer vision segmentation combined with extensive open vocabulary (Liu et al., 2025), or embedding of ID markers that are perhaps even hidden to the naked eye (Dogan et al., 2022). Although embedding technology may appear to be cumbersome, the reality is that we live



Our World Becomes Programmable



FIGURE 2

Basic principles of programmable reality: it cannot be simply overlaid but needs to be blended; it cannot be static but needs to be dynamic while allowing interactive manipulation, where query chatbots with prompts are not enough. Artificial intelligence must be pervasive and work anywhere, with always ON and awareness modes. This would allow everyone to program their reality quickly.

in highly fabricated worlds with extensively manufactured portions, for which cheap marker solutions can be integrated into the manufacturing pipeline. Nonetheless, we should be mindful of overdoing static augmentation as not all augmented words need to be crowded and as programmable reality can be used to declutter spaces (Gonzalez-Franco and Colaco, 2024).

3.2 Dynamic augmentation

Current analog devices have no voice in the digital world; to make such analog devices smarter and more interactive in the virtual world, we need new methods of communication that can transmit their current digital states rather than just static payloads. For example, a fire alarm could continuously display its battery status, or a book on a bookshelf could constantly update a digital bookmark (Ahuja et al., 2019). To achieve this level of interactivity, the process normally begins with IoT sensors collecting raw data from the environment. These sensors serve as the eyes and ears of the digital world, capturing the nuanced states of analog devices in real time. However, the IoT concept does not mean that each analog object can become a part of the IoT network as many of these environmental sensing devices can be centralized into a single entity. Following this environmental gathering via XR and other connected existing devices, AI models come into play and are tasked with the job of deciphering the raw data gathered from all objects (both IoT and not) to make them shareable (Allen et al., 2025). In essence, this makes every object an IoT as its data could be sensed using wearable glasses on the go rather than via internal sensors. Finally, AI model pipelines will understand and analyze the gathered data to transform them into meaningful information for each user, ensuring that the physical environment can change accordingly.

3.3 Dynamic interaction

Real-time modifications and interactions with digital elements will become more available as objects and people become more digital or become connected to the internet, e.g., through IoT or because users wear XR devices that can understand reality and help with direct inputs. Our interactions with these objects, screens, or panels will increasingly resemble our current digital interactions. Additionally, if we have robust identification of a user body that is also digitized well, the use of analog objects with overlaid digital content could create very strong dynamic interactions (Suzuki et al., 2020). This is feasible to the point that one might be able to change people's representations through immersive interfaces, e.g., by changing the appearance or face of a person on demand, perhaps one can appear to be a famous actor or actress on demand with deep fakes (Pataranutaporn et al., 2021). However, not all uses will be dystopian as XR re-rendering capabilities can also empower users to block ads and other undesired elements in virtual and physical environments (Katins et al., 2025); this could have positive effects on cognitive noise via diminished reality (Cheng et al., 2022).

3.4 Collaborative environments

Multiuser interaction in a shared digital/physical space is a complete area in itself (Wang et al., 2024; Grønbaek et al., 2023). We share the world with billions of other humans, and technology has undoubtedly helped us connect more with people located remotely. However, we sometimes achieve such connections at the expense of being present in our physical spaces. In programmable reality, content anchored on colocated physical spaces offers possibilities to connect better with people and to

interact *in situ* while achieving growth (Kitson et al., 2024) across temporal boundaries and in asynchronous ways (Deutch, 1997). When we dive a bit deeper into the spatial dynamics of these new realities, we find that there are local (personal space) interactions that occur within the immediate physical spaces of the users as well as remote (global) interactions occurring over long distances, which affect or involve distant environments or even users.

Some spaces can also undergo different depths of integration, while some surface-level integration may be available initially through simple overlays or enhancements without deep interactions with the physical world, mostly as a result of the application of current AI. A secondary layer of more complex interactions in which the digital and physical elements are intertwined (e.g., smart materials and robotics) will then appear in a more progressive manner; this hints at real potential to change how every object is manufactured and consumed, even the analog ones (Dogan et al., 2024). The same changes are expected for user autonomy in these programmable realities, while scripted experiences with predefined interactions and limited user control will be predominant initially, perhaps owing to the lack of access to raw passthrough data or limited application programming interfaces (APIs). It is likely that user-created reality will be within reach with the current generative AI capabilities, where users will have extensive control over the creation and manipulation of digital elements in their environments.

4 Shared reality: guidelines for implementation

Ensuring that programmable realities are operable across different platforms will be essential for creating shared realities. This is true for not just parallel realities that seem to be byproducts of the internet, like echo chambers in social media feeds. In a way, this will be even more critical for social cohesion; we now know that the digital realm can change our real-life behaviors, so we should account for such changes. Achieving compatibility involves addressing a range of technical, standardization, and user experience considerations. We should collaborate to develop and adopt universal standards for file formats, communication protocols, and data representations, including engaging with international standards organizations (such as the IEEE and W3C) to create and maintain these standards. Furthermore, achieving strong build capabilities on top of open-source frameworks and tools can foster community-driven development and compatibility, as open-source tools are more easily adapted and integrated across various platforms.

The integration of physical and digital realities raises significant concerns regarding data privacy and cybersecurity. There should be strong emphasis on APIs and software development kits (SDKs) as the end users and app developers cannot be expected to compile their own code directly on the devices so as to prevent data privacy issues. It is also worth considering that such developers may not have raw access to the sensors, including the cameras that will allow AR to preserve the privacy of the end users; this means that the roles of comprehensive APIs and SDKs that are platform-agnostic will be highly relevant. The current camera passthrough caps can also be eased in this manner without allowing raw access to everyone.

Additionally, universal design principles may be needed to ensure that the user interfaces and experiences are consistent and intuitive across different platforms; this most likely requires close collaborations among the industry players like hardware manufacturers, software developers, content creators, and other stakeholders in the AR/VR/mixed reality ecosystem, with a focus on interoperability. By addressing these aspects, developers and organizations can create programmable realities that offer seamless, integrated experiences across a wide range of devices and platforms. In the end, we will need to address the question of “Who programs the reality?” and the answer to this must be democratic: everyone.

5 Conclusion

This article aims to formalize and converge several emerging areas of technology toward a future of programmable reality. We are aware that our view is not exhaustive but only offers a framework to understand the myriad ways in which programmable reality could manifest and how users might interact within it. As we stand on the brink of a new era of immersive computing in collision with AI and edge infrastructure, interdisciplinary collaboration as well as thoughtful research and development will be key to unlocking the transformative power of programmable reality.

Author contributions

RS: Writing – review and editing, Writing – original draft. PA: Writing – review and editing, Writing – original draft. CZ-T: Writing – review and editing, Writing – original draft. MDD: Writing – review and editing, Conceptualization. AC: Writing – review and editing, Conceptualization. EG: Writing – original draft, Writing – review and editing, Conceptualization. KA: Writing – review and editing, Writing – original draft. MG-F: Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

Authors AC, EG, KA, and MG-F were employed by Google.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. The figures in this paper were partially AI generated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abreu, S., Do, T. D., Ahuja, K., Gonzalez, E. J., Payne, L., McDuff, D., et al. (2025). "Parse-ego4d: toward bidirectionally aligned action recommendations for egocentric videos," in *ICLR 2025 workshop on bidirectional Human-AI alignment*.
- Ahuja, K., Paredy, S., Xiao, R., Goel, M., and Harrison, C. (2019). "Lightanchors: appropriating point lights for spatially-anchored augmented reality interfaces," in *ACM symposium on user interface software and technology*, 189–196.
- Allen, R. M., Barski, A., Berman, M., Bosch, R., Cho, Y., Jiang, X. S., et al. (2025). Global earthquake detection and warning using android phones. *Science* 389, 254–259. doi:10.1126/science.ads4779
- Bar-Tal, O., Chefer, H., Tov, O., Herrmann, C., Paiss, R., Zada, S., et al. (2024). Lumiere: a space-time diffusion model for video generation. *arXiv:2401.12945*. doi:10.1145/3680528.3687614
- Battaglia, P. W., Hamrick, J. B., and Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci.* 110, 18327–18332. doi:10.1073/pnas.1306572110
- Cheng, Y. F., Yin, H., Yan, Y., Gugenheimer, J., and Lindlbauer, D. (2022). "Towards understanding diminished reality," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–16. doi:10.1145/3491102.3517452
- Deutch, D. (1997). *The fabric of reality*. New York: Viking.
- Dogan, M., Taka, A., Lu, M., Zhu, Y., Kumar, A., Gupta, A., et al. (2022). "Infraredtags: embedding invisible AR markers and barcodes using low-cost, infrared-based 3D printing and imaging tools," in *CHI Conference on Human Factors in Computing Systems*, 1–12. doi:10.1145/3491102.3501951
- Dogan, M., Gonzalez, E., Colaco, A., Ahuja, K., Du, R., Lee, J., et al. (2024). Augmented object intelligence with XR-objects ACM UIST
- Gonzalez, E., Patel, K., Ahuja, K., and Gonzalez-Franco, M. (2024). "Xdtk: a cross-device toolkit for input interaction in XR," in *IEEE VR*.
- Gonzalez-Franco, M., and Colaco, A. (2024). Guidelines for productivity in virtual reality. *ACM Interact. Mag.* 31, 46–53. doi:10.1145/3658407
- Grønbaek, J. E. S., Pfeuffer, K., Velloso, E., Astrup, M., Pedersen, M. I. S., Kjær, M., et al. (2023). "Partially blended realities: aligning dissimilar spaces for distributed mixed reality meetings," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. doi:10.1145/3544548.3581515
- Hirzle, T., Müller, F., Draxler, F., Schmitz, M., Knierim, P., and Hornbæk, K. (2023). "When XR and AI meet-a scoping review on extended reality and artificial intelligence," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–45. doi:10.1145/3544548.3581072
- Katins, C., Strecker, J., Hinrichs, J., Knierim, P., Pfleging, B., and Kosch, T. (2025). "Ad-blocked reality: evaluating user perceptions of content blocking concepts using extended reality," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18. doi:10.1145/3706598.3712330
- Kitson, A., Ahn, S. J., Gonzalez, E. J., Panda, P., Isbister, K., and Gonzalez-Franco, M. (2024). "Virtual games, real interactions: a look at cross-reality asymmetrical co-located social games," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–9. doi:10.1145/3613905.3650824
- Liu, X., Jia, D., Liu, X. C., Gonzalez-Franco, M., and Zhu-Tian, C. (2025). "Reality proxy: fluid interactions with real-world objects in MR via abstract representations," in *ACM UIST 2025*.
- Mao, H., Gonzalez-Franco, M., Phadnis, V., Gonzalez, E. J., and Chatterjee, I. (2025). "Restfulraycast: exploring ergonomic rigging and joint amplification for precise hand ray selection in XR," in *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, New York, NY, USA (New York, NY: Association for Computing Machinery), 28–39. doi:10.1145/3715336.3735677
- Morris, M., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., et al. (2023). Levels of agi: operationalizing progress on the path to agi. *arXiv:2311.02462*. doi:10.48550/arXiv.2311.02462
- Padrao, G., Gonzalez-Franco, M., Sanchez-Vives, M., Slater, M., and Rodriguez-Fornells, A. (2016). Violating body movement semantics: neural signatures of self-generated and external-generated errors. *Neuroimage* 124, 147–156. doi:10.1016/j.neuroimage.2015.08.022
- Pataranutaporn, P., Danry, V., Leong, J., Punpongsonan, P., Novy, D., Maes, P., et al. (2021). AI-generated characters for supporting personalized learning and well-being. *Nat. Mach. Intell.* 3, 1013–1022. doi:10.1038/s42256-021-00417-9
- Pfeuffer, K., Gellersen, H., and Gonzalez-Franco, M. (2024). Design principles and challenges for gaze+ pinch interaction in XR. *IEEE Comput. Graph. Appl.* 44, 74–81. doi:10.1109/mcg.2024.3382961
- Steed, A., Ofek, E., Sinclair, M., and Gonzalez-Franco, M. (2021). A mechatronic shape display based on auxetic materials. *Nat. Commun.* 12, 4758. doi:10.1038/s41467-021-24974-0
- Suzuki, R., Kazi, R., Wei, L., DiVerdi, S., Li, W., and Leithinger, D. (2020). "Realitysketch: embedding responsive graphics and visualizations in AR through dynamic sketching," in *ACM symposium on user interface software and technology*, 166–181.
- Suzuki, R., Gonzalez-Franco, M., Sra, M., and Lindlbauer, D. (2023). "XR and AI: AI-enabled virtual, augmented, and mixed reality," in *Adjunct proceedings of the 36th annual ACM symposium on user interface software and technology*, 1–3.
- Talukder, S., Yisong, Y., and Gkioxari, G. (2024). Totem: tokenized time series embeddings for general time series analysis. *arXiv preprint arXiv:2402.16412*
- Team, G. (2024). Gemini 1.5: unlocking multimodal understanding across millions of tokens of context. *Google Deep*.
- Tian, J., Aggarwal, L., Colaco, A., Kira, Z., and Gonzalez-Franco, M. (2024). *Diffuse, attend, and segment: unsupervised zero-shot segmentation using stable diffusion*. New York, NY: CVPR.
- von Willich, J., Günther, S., Matvienko, A., Schmitz, M., Müller, F., and Mühlhäuser, M. (2023). "Densingqueen: exploration methods for spatial dense dynamic data," in *ACM symposium on spatial user interaction*.
- Wang, C. Y., Ofek, E., Kim, H., Panda, P., Won, A. S., and Franco, M. G. (2024). "Avatarpilot: decoupling one-to-one motions from their semantics with weighted interpolations," in *2024 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct)* (IEEE Computer Society), 588–591.
- Zhu-Tian, C., Chiappalupi, D., Lin, T., Yang, Y., Beyer, J., and Pfister, H. (2023). RL-label: a deep reinforcement learning approach intended for AR label placement in dynamic scenarios. *IEEE Trans. Vis. Comput. Graph.* 30, 1347–1357. doi:10.1109/tvcg.2023.3326568