Infosys topaz

VIEW POINT

THE ROBOTIC CANVAS: MULTIMODAL GEN AI PAINTING AUTONOMY INTO PHYGITAL EMBODIMENTS

Infosys®
Navigate your next

This point of view discusses four different types of digital and physical experiences and the various embodiments of Generative AI to introduce novel interactions within these experiences. We first explore immersive web applications and digital human avatars as GenAI embodiments that takes the levels of engagement far beyond the current experience of a 2D world, as well as enabling accessibility seamlessly. Next, we discuss the application of Generative AI in generating 3D assets and environments in the metaverse that enables designers to craft user driven interactive narratives. Further, we discuss the applications of Generative AI to impart conversational intelligence to humanoid robots as well as industrial ones that are today being controlled via desktop and mobile applications. Finally, we discuss the embodiment of Generative AI in broadcast media - AI Radio anchors and in making live TV content accessible to diverse audiences.

Generative AI for Immersive Web Applications

Sama is a Digital Human Avatar or MetaHuman that is the first virtual crew member of the Qatar Airways. She is responsible for guiding passengers in a virtual replica of a Qatar Airways flight and engages them in conversation. This is just one of the applications of digital human avatars being provided by companies like Soul Machine, Uneeq or Deep Brain that can engage an audience on any topic you desire them to. The use cases range from a product advisor to that of a customer help desk. Behind the scenes, the avatar leverages Large Language Models that are fine-tuned with the corpus of information that they need to be knowledgeable about. Multimodal LLMs can enrich the conversation with multimedia content like images or videos both as input and output modalities. With sophisticated AI techniques, the avatars manage to change facial expressions with every turn of the conversation. They also display a personal style of conversation in-line with the role they are supposed to play.

As mentioned above, a key example is Soul Machines, an Australia-based startup, specializing in creating digital humans. They have developed the Human OS™ platform, featuring a patented Digital Brain that powers their Autonomous Animation technology. This innovation enables seamless collaboration between humans and machines, combining the best of both worlds. Their new features include varied thinking behaviors, like re-establishing eye contact while responding and dynamic changes in facial expressions during "thinking" moments.

Digital Brain

Patented IP developed in concert with deep research into neuroscience, psychology and cognitive science to replicate the way our brains handle everyday interactions.

- Machine Learning
- Natural Language Processing
- Content Awareness
- Sentiment Analysis
- Machine Vision
- Emotional Model & Attention Control

Human OS

A Digital Person (TM) is powered by Human OS delivering the goodness of human and machine collaboration.

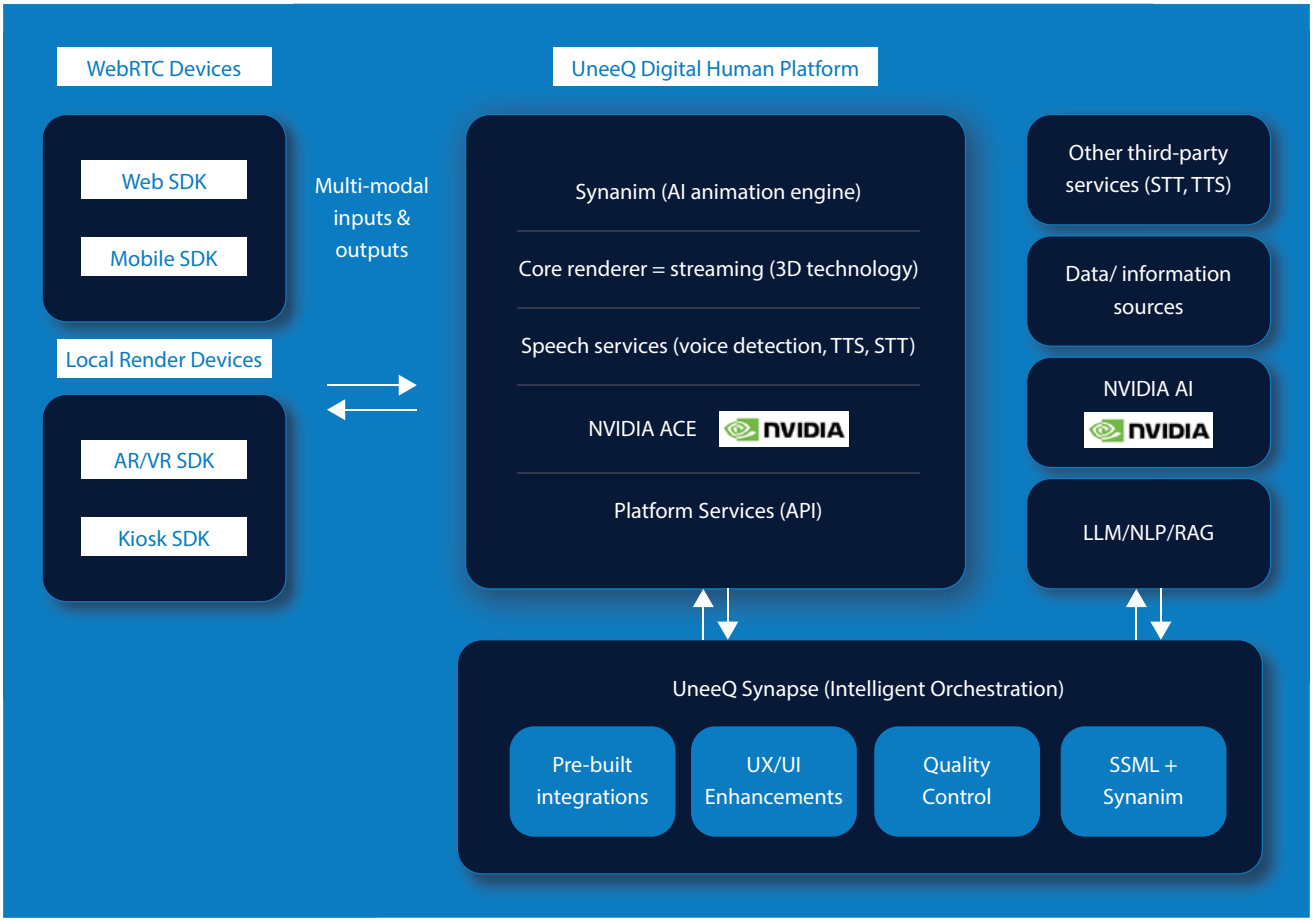
Automation Animation

Ability to autonomously react to external stimuli vs manually programming "natural" "emotions". This capability is unparalleled in the marketplace today.

- Hyper-realistic CGI
- Expression Rendering
- Gaze Direction
- Synthetic Voice
- Real-time Gesturing
- Gestural Personality

Human OS™ Platform

Another example is **UneeQ**, a leading platform recognized for creating lifelike digital characters using AI-powered technology.



UneeQ's Digital Human Platform

Achieving high-quality speech animation for digital characters is challenging, often leading to delays and unnatural expressions. To overcome this, UneeQ integrated NVIDIA ACE, which is a set of technologies designed to bring digital humans to life using generative AI into their AI animation system, Synanim, for smoother real-time interactions. Synanim, short for synthetic animation, powers UneeQ's lifelike digital avatars by orchestrating real-time movements and emotional expressions like friendliness and empathy. It also controls subtle details like breathing and body movements to enhance realism. UneeQ integrates NVIDIA ACE, including the Audio2Face (A2F) service, for realistic facial animation and lip-sync during speech. This creates low-latency, natural conversations with digital characters, delivering a highly immersive user experience.

UneeQ leverages Synanim and NVIDIA ACE to power digital avatars that deliver personalized, emotionally engaging interactions across various scenarios, from e-commerce to customer service and training. This adaptability allows clients to provide exceptional experiences, whether in the cloud or on-premises.

The social media ambassador of a consumer goods company might project a charming personality while an avatar manning a customer help desk would be sincere and efficient in its mannerisms. The technology is available in the public domain in a way that even novices can create and train these avatars in a matter of minutes. From ethnicity to language, facial features to attire – everything can be customized by clicking a few buttons.

Given the ease of use and the massive potential for end user engagement, it is just a matter of time before we see immersive web content come alive with the aid of these virtual humans.

Generative AI in the Metaverse

Imagine being able to transform virtual worlds while you engage in conversations with 3D avatars. In the gaming world, you can interact with Non-Playing Characters (NPCs) in natural language and ask them to create worlds aligned with your tastes, while still retaining the other elements of game play.

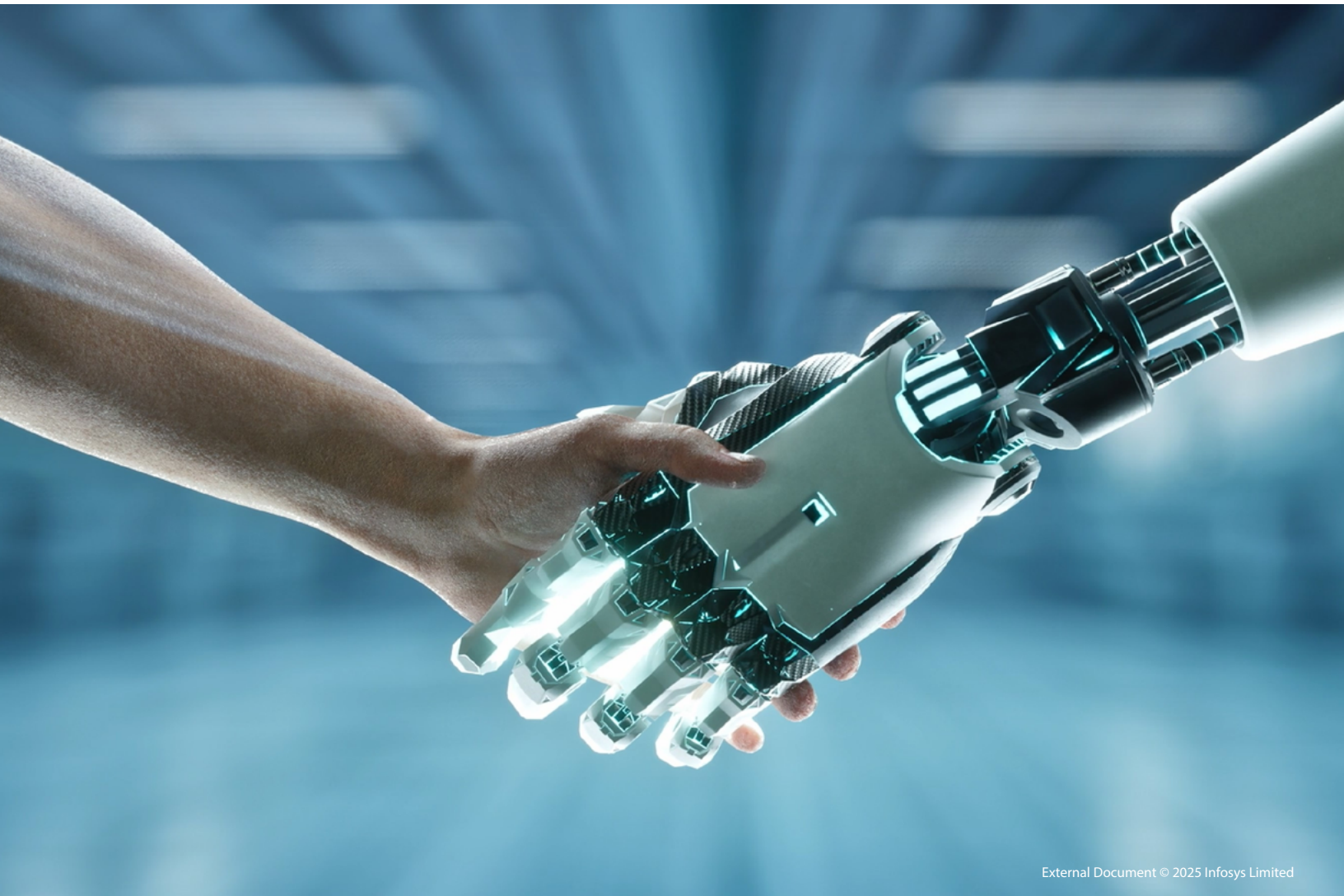
Inworld.ai is a leading AI engine for games, offering groundbreaking mechanics, dynamic NPCs, and evolving game worlds. The California-based startup is developing tools for NPCs to engage players with dynamic, unscripted dialogue and actions, ensuring fresh interactions every time.

AI NPCs could significantly enhance player interactions by allowing for spontaneous engagement. This spontaneity may create the illusion of a living, dynamic environment. Research led by Joon Sung Park at Stanford University explored this using a life simulation-style game named Smallville, which featured twenty-five AI-generated characters with distinct names and simple backstories. Over two days, these characters engaged in human-like conversations, remembered each other, and referenced historical interactions. For instance, when a character organized a Valentine's Day gathering, others sent invitations and arranged dates through natural dialogue, showcasing the complexity of generative AI NPCs. This happened through dialogues, with previous exchanges between characters saved as natural language in their "memories".

Generative AI-powered NPCs will transform player agency by enabling co-creation of game narratives. Unlike traditional games with limited player influence, this technology offers players greater freedom and engagement in shaping the storyline, marking a shift toward player-centric experiences.

With image and 3D asset generation models, whole new gaming worlds can be created while the intents in your conversation are converted into prompts by the Avatar Engine that fires the 2D and 3D asset generation models like StableDiffusion or NVIDIA Picasso.

Another application is the possibility of merchandise being created while being in the virtual world of a retail brand. For example, while in an apparel store, a customer might enter the virtual world of say, a nature inspired collection. Inside this world, the virtual replicas of real-world products can be customized with elements picked up from the virtual world. Some popular inpainting models include Stable Diffusion Inpainting, Stable Diffusion XL (SDXL) Inpainting, and Kandinsky Inpainting. The creation can then be purchased and retained as a collectible or NFT. Taking it one step further, the merchandise can be 3D printed and shipped to the customer after a purchase.



Generative AI for Robotics

Advancements in robotics have made robots able to perform physical tasks like picking and placing things, walking or even running and jumping while avoiding obstacles, etc. However, human-like conversations were still a barrier until the advent of Generative AI. LLMs can be integrated into robots to impart them conversational capabilities. Furthermore, with LLMs like PaLM say Can, the conversations can also be translated into instructions to perform physical tasks.

Another approach for novel interaction with robots could be to integrate a conversational interface on a tablet with a robot that performs a specific task, that is currently being enabled using a mobile application. For example, instead of maneuvering a solar panel cleaning robot using controls on a mobile app, you could have a natural language conversation with the robot via the mobile device that can ensure you can instruct the robot using natural language instead of fidgeting with the touch controls while you are simultaneously monitoring the robot's actions.

Generative AI in Broadcast Media

AI anchors on radios shows are already making waves as they conduct request-based shows, curate music and summarize news of the day from across the internet for their audience. Talk shows with AI created anchors are another format that is becoming a part of popular radio broadcasts.

Though the current shows replicate the format of the shows as a human anchor would, it would be far more interesting to explore alternative approaches to delivering radio broadcasts given the versatility and creativity possible with Generative AI content, as well as limitless content curation capabilities in near real time from the internet. Broadcast content need not only be curated but can also be generated anew – think opinions in the form of conversations between philosophers like Plato and Aristotle on topics of current interest. Also, with models like MusicGen grom Meta, even original music can be created on demand in near real time for an audience requesting the same.

In broadcast television, closed captioning can be generated in real time for live events that can make such content accessible to diverse audiences. Again, the addition of an avatar that can translate the captions into sign language using GenAI video generation capabilities can enable more video content to become accessible, by bringing down the cost of this additional feature. Near-real time video generation is a possibility today, however as video generation models become more sophisticated, and compute power becomes cheaper and democratized, this technology might revolutionize accessible video content as we know today.

Ethics and Copyright

AI models need a vast corpus of data for training. Though the regulations across the globe are evolving, there is a broad consensus that the acquisition of such data, if not available in the public domain, needs to be done as per fair practice and should be licensed. For example, Meta's MusicGen model has been trained on 20,000 hours of licensed music. However, the ownership of music that is generated using this model, as per existing laws of authorship, falls into a gray area. Though the prompt can be considered an original creation of the author, the content being generated autonomously by AI is another matter, the ownership of such content not being covered by existing regulations. The EU AI act is the first step in this direction, and widespread adoption, similar to GDPR can help resolve ambiguities in the areas outlined above.

A Word of Caution: Hallucinations, Bias and Data Privacy

Generative AI models are prone to hallucinations and can output responses that are not only incorrect but outright derogatory or defamatory. Extensive finetuning and filtering for such content via human in the loop involvement is an essential mechanism to ensure that the responses to a prompt are appropriate and not rooted in the imagination of the AI.

Again, bias can creep in the output of a generated content where the data on which the model was trained on was biased towards a particular type of data. For example, an avatar generator utility for accessible content might only turn out fair skinned avatars if it has been trained only on images of Caucasian populations. Models should be tested for various kinds of bias – from racial to gender as well as linguistic or ethnic biases.

User data that is captured while interacting with an autonomous AI agent should be treated in accordance with data and privacy laws applicable to the jurisdiction in which the application is being consumed. For example, for a conversational assistant to be able to conduct a nuanced conversation per the existing emotional state of the user, it might capture and analyze the video of the user. In such scenarios, as per applicable laws, users should be made aware of the same and their consent taken as a pre-requisite before the conversation begins.

Conclusion

Generative AI is a potent technology, and it is important to understand the implications of applying this technology to various modalities in the realm of Human Computer Interaction. It's embodiments hold enormous potential for enabling novel interactions with existing interfaces as well as elevating engagement within existing forms of interaction. However, given the ubiquity and wide scale applicability of this technology, one should tread with caution and attempt to mitigate the risks due to concerns on ethics, privacy, and unintended consequences of effects like hallucinations and biased responses.

References

- Chunyuan Li, Zhe Gan, Zhengyuan Yang, Jianwei Yang, Linjie Li, Lijuan Wang and Jianfeng Gao (2024), "Multimodal Foundation Models: From Specialists to General-Purpose Assistants", Foundations and Trends® in Computer Graphics and Vision: Vol. 16: No. 1-2, pp 1-214. http://dx.doi.org/10.1561/0600000110
- Rezaei Mostafa, (2022), "What is HumanOS Platform? https://metaverse.acm.org/what-is-humanos-platform/
- Firth Niall, (2024), "How generative AI could reinvent what it means to play" AI-powered NPCs that don't need a script could make games and other worlds deeply immersive https://www.technologyreview.com/2024/06/20/1093428/generative-ai-reinventing-video-games-immersive-npcs/
- Almada, Marco and Petit, Nicolas, The EU AI Act: a medley of product safety and fundamental rights? (2023). Robert Schuman Centre for Advanced Studies Research Paper No. 2023/59 http://dx.doi.org/10.2139/ssrn.4308072

About the Authors

Rani Malhotra heads the Applied Research Center for Autonomous Machines at the Infosys Center for Emerging Technology Solutions. She works across emerging technologies and their intersections, including AI, Human Machine Interactions and Smart Systems.

Hindavi Shetye is a Senior Associate Consultant at iCETS (Infosys Center for Emerging Technology Solutions), specializing in researching and analyzing industry trends and emerging technologies, along with their impact across various sectors. She is a research enthusiast, who enjoys delving into and learning about cutting-edge and next-generation technologies.

For more information, contact askus@infosys.com

Infosys®
Navigate your next