RESPONSIBLE SYNTHETIC CONTENT FOR THE METAVERSE AGE

Abstract

In recent times, only a handful of technological concepts could grab the attention of our collective imagination, such as the metaverse. The exponential rise in the number of internet-driven devices, cryptocurrencies, and virtual content have made the metaverse ripe for mainstream adoption. The promise of value to multiple industry segments comes with a high data privacy risk, cyber risk, and unfair practices in complex immersive environments.

Organizations would have to look outside traditional methods to build responsible content for the metaverse. This Infosys viewpoint explores the need and design principles for synthetic content for the metaverse. We recommend a model for enterprises to create responsible privacy-first synthetic content to protect end users' personal data and balancing high-quality experiences in the virtual world.

Infosys®
Navigate your next

Why do we need Responsible Synthetic Content for Metaverse?

Metaverse is a symbiotic confluence of mature technologies like Artificial Intelligence (AI), Cryptocurrency and Virtual Reality (VR). Though widely popularized by the gaming and entertainment industry, metaverse now has a growing number of promoters across various domains and use cases. For instance, Boeing1, a world leader in engineering, is heavily betting on physically accurate digital twins to prevent manufacturing problems in the future, such as structural flaws and preventive maintenance.

With every new iteration of innovation on metaverse, we will see an order of magnitude in business benefits as well as the complexity in data privacy, ethics, and sustainability. Creating a hyper-realistic virtual environment that replicates the natural environment, like a shop floor, would need good data utility with minimal risk that avoids any ethics and fairness pitfalls.
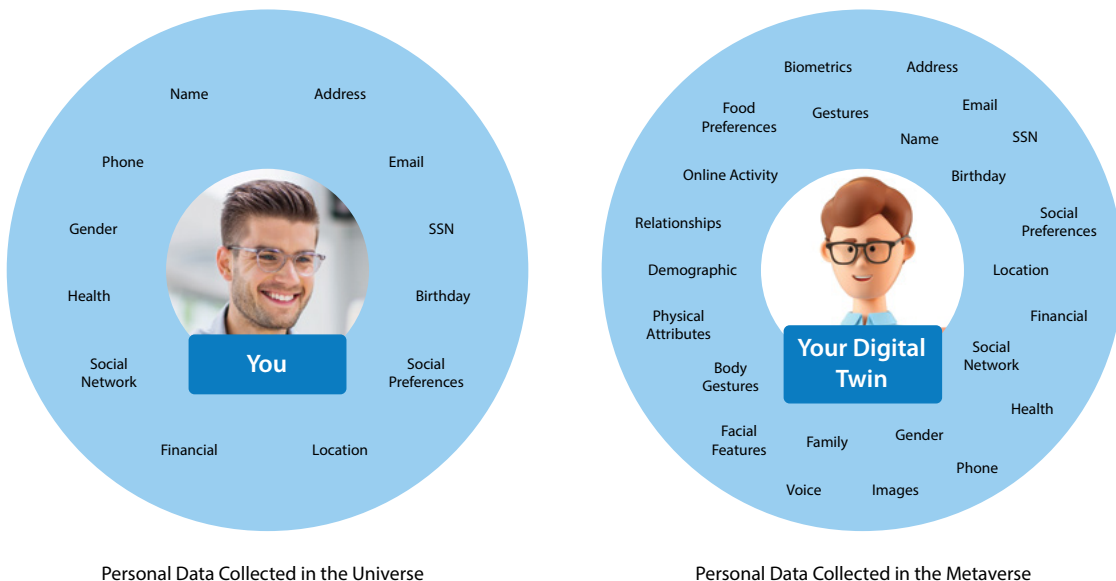

Personal Data Collected in the Universe

Personal Data Collected in the Metaverse

Figure 1 - Personal Data Collected in the Metaverse

Organizations usually run into a "Data Deficit" as it is challenging to collect photos, images, videos, or sound files for building virtual and immersive environments. As the pandemic accelerates the trend towards stricter data privacy regulations and governance, synthetic data offers metaverse the advantage of regulatory compliance. Investments in robust synthetic data technologies, with the right kind of data privacy methods and fairness standards, will help metaverse scale at large and avoid the over-reliance on the personal data of end-users. It will also help organizations create high quality, anonymized and responsible content for their end customers, partners, and the overall community.

Challenges of Building Responsible Data for Metaverse

Metaverse data or content can be divided across 5 key layers for any business use case. The rising number of connected devices and experiences means that large quantities of data will be connected across these 5 layers.
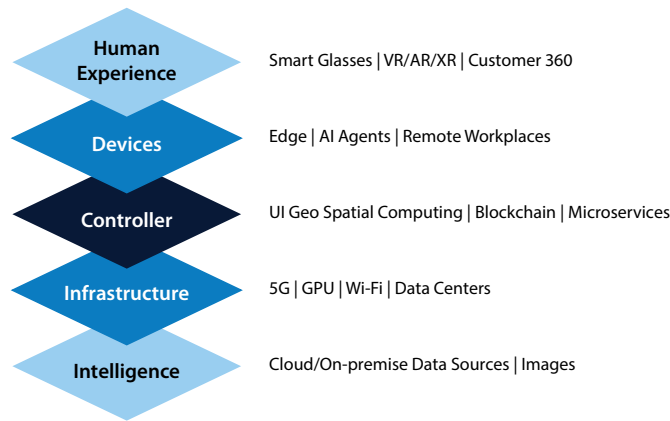
Layer	
Human Experience	Smart Glasses	VR/AR/XR	Customer 360
Devices	Edge	AI Agents	Remote Workplaces
Controller	UI Geo Spatial Computing	Blockchain	Microservices
Infrastructure	5G	GPU	Wi-Fi	Data Centers
Intelligence	Cloud/On-premise Data Sources	Images

Figure 2 - Metaverse Data across 5 Key Data Layers

1. **Human Experience** – The consumer data is essential for user experience. It started with headsets and smart glasses and slowly moved into the world of smartphones and other mobile devices.

2. **Edge Devices** – Inputs from the edge devices could be vital data that would drive the metaverse. It could range from industry field pieces of equipment for measuring temperature or seismic activity to a smart-microwave oven that pre-heats your lunch on time.

3. **Controllers** – Under the finance domain, for instance, financial transactions, digital ledgers, and consumer files can simplify with a microservices-driven metaverse controller framework.

4. **Infrastructure** – Rich content of metaverse would need significant investments in infra to ingest, process, and manage content from multiple sources for a unified user experience.

5. **Intelligence Layer** – This would be a key area of evolution with cleverer decision-making and data-centric services based on the data collected across the other 4 layers.

We will put this into perspective with a 5G telecom use case. Metaverse gives wireless operators an opportunity to monetize their 5G investments rapidly. It is how major players like Verizon are investing in imaginary enterprises and consumer use cases on edge computing services. Moreover, this technology can help reduce the difficulty of laborious tasks, thereby improving productivity. For example, field service agents have to step out for a service outage, but it can be easily powered by metaverse instead, as shown in the figure below.

However, there is a need for proper controls to identify, manage and mitigate privacy and ethical data concerns faced by each layer. Some of the challenges are as follows:

1. **Human Experience** – Personal data protection is necessary as high precision data of users, like personal preferences and GEO location, would be processed by the applications. Robust Personal Data Privacy controls, including consent governance, need to be an integral part of metaverse solutions.

2. **Edge Devices** – Always "ON" devices pose a major threat to field service operators and will have compliance problems.

3. **Controllers** – Cryptocurrency transactions, inventories stockpile data, digital payments, and digital ledgers need clear protection focus on cross-border data flow and should be protected at rest and in transit.

4. **Infrastructure** – Data Privacy first ingestion, processing, and disposal of large volumes of data can be a difficult task. Building breach management mechanisms is an additional challenge.

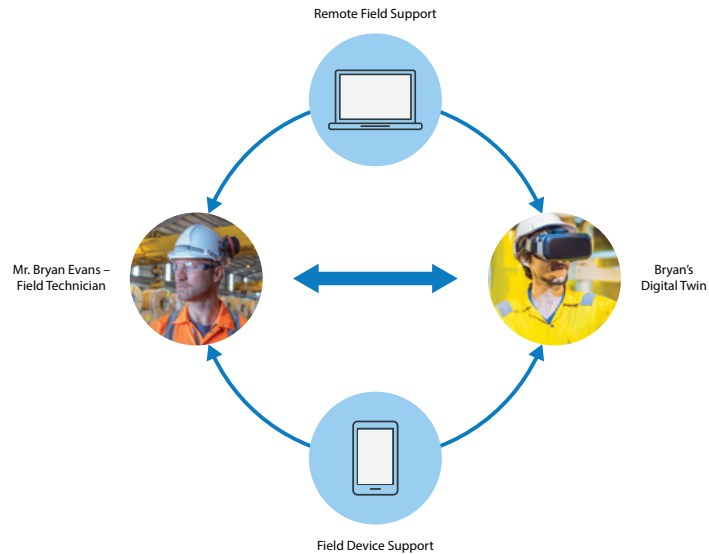5. **Intelligence Layer** – Existing AI models have biases in one form or another.

Figure 3 - Smart Field Service - Illustration

Creating responsible data for training and deploying models in production would not be easy.

Each layer brings in a unique set of challenges with respect to data management. Though these challenges exist in the current AI world, they get amplified in the metaverse since it entails designing virtual, immersive, and inter-operable environments that would create a surreal experience for the end-users.

The only reliable alternative is to synthesize content using good quality training data, images, and environments. Gartner predicts that by 2025, synthetic data will reduce personal customer data collection, avoiding 70% of privacy violation sanctions. Datagen is talking about a new popular job called "Synthetic Data Engineer" this year (2022) which underlines the significance of generating and processing synthetic data to build robust AI models. Manual synthetic content generation from the real world is expensive and has privacy concerns that may be averted using synthetic data.

What does Infosys Recommend?

Today, industry experts speculate that the metaverse will revolutionize nearly every industry and function, ranging from health care imaging to payments, gaming, entertainment, and even remote working. One potential disruption the metaverse would bring about is the new digital workplace known as the "Horizon Worlds", which has the potential to impact the labor market and eliminate the geographical boundaries. The user can jump into a virtual destination and do whiteboarding instead of a boring video call. But this needs enough data collected from the user for their digital avatar to understand their personal preferences, GEO location details, and behavior. Without new rules of censorship, privacy controls and regulatory enforcement, large-scale adoption of metaverse will not be possible.

Furthermore, we would need multiple pictures and videos to simulate an office or an external location. This training data for simulation should be hand-labelled for algorithms to understand and effectively work.

To add value to the metaverse, we strongly recommend Infosys Enterprise Synthetic Data Generation Model, which leverages a responsible data framework combined with well-evolved Infosys intellectual property, Infosys Enterprise Data Privacy Suite (iEDPS), for:

1. **Utility of synthetic data** - Product development teams within a large organization often wait for 6 -12 months to get the data for product development. It is an arduous process to provision data from different stakeholders. Product teams often get

outdated or worthless data. Today for rapid innovation in the metaverse, organizations need high data utility without the risk of personal data breach and algorithmic bias.

2. **Quality of synthetic data** - With metaverse, AI will undergo a significant paradigm shift in which traditional and model-centric approaches to AI development will give way to data-centrism. It implies that the data scientists would focus on the quality of their training data as a parameter of performance than the quality of their model.

3. **Context of synthetic data** – Synthetic data will be successful only when built with the human context at the center. This data should be responsible, sustainable and private-first on an enterprise scale.
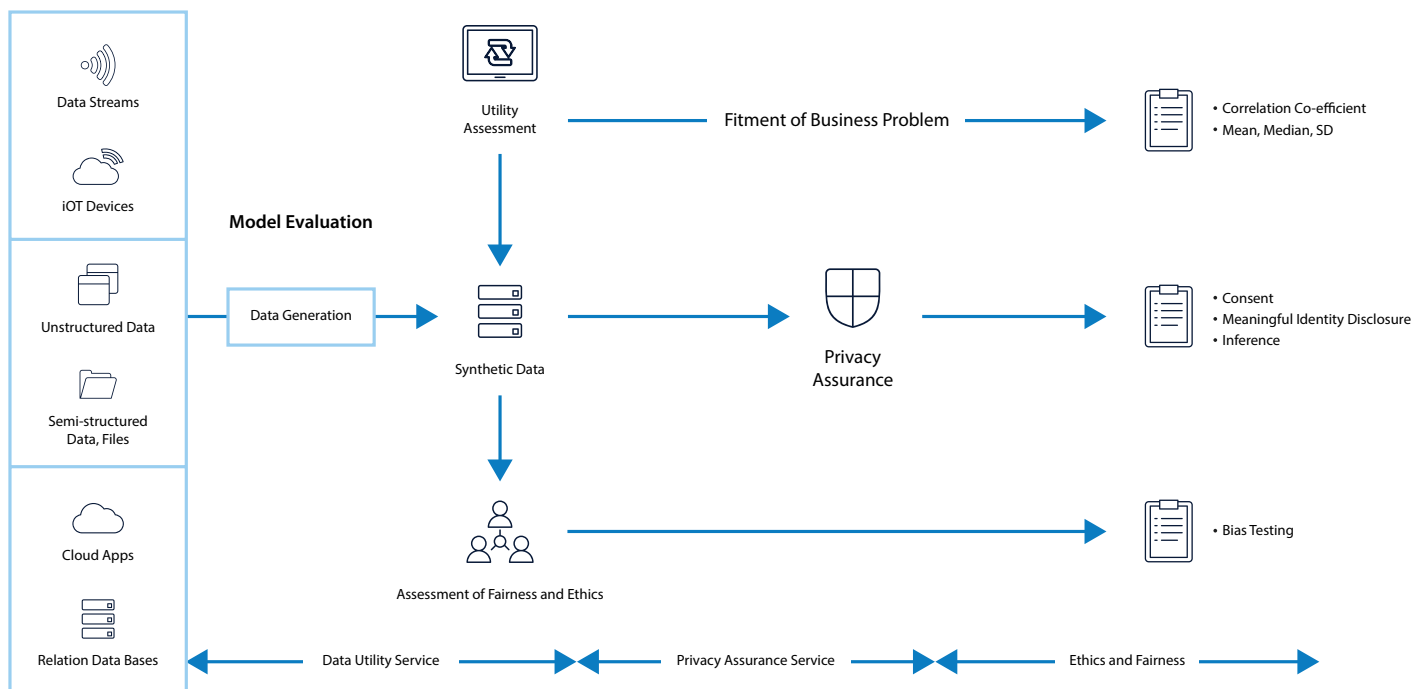

Figure 4 - Synthetic Data Framework - Data Utility, Privacy Risk Assessment, Ethics and Fairness Quotient

We can agree that the metaverse will be a new phase of commerce and socialization on the internet. As time and technology move on, the results of the success of data privacy policies, software, and hardware will become apparent. Without robust synthetic data investments, we would see multiple personal data breaches and eventual distrust in the potential benefits of this technology.

About the Authors

Kavitha Sundararajan is a banking domain expert with more than a decade of experience, especially in the lending area. As a Lead Consultant in Infosys Center for Emerging Technology Solutions (ICETS), she is responsible for managing large transformation deals through pre-sales interactions with various stakeholders, developing marketing strategies, and building product roadmaps.

Karthik Nagarajan is an Industry Principal Consultant at Infosys Center for Emerging Technology Solutions (ICETS) and has more than 15 years of experience in data privacy, customer experience and business development. He is currently part of the product team of Infosys Enterprise Data Privacy Suite.

About us

The incubation center of Infosys called 'Infosys Center for Emerging Technology Solutions' (iCETS) focuses on the incubation of NextGen services and offerings by identifying and building technology capabilities to accelerate innovation. The current areas of incubation include AI & ML, Blockchain, Computer Vision, Conversational interfaces, AR-VR, Deep Learning, Advanced Analytics using video, speech, text, and much more.

To know more, please reach out to iCETS@infosys.com.

References

1. https://airwaysmag.com/innovation/boeing-next-aircraft-metaverse/

2. https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/

3. https://research.aimultiple.com/tech-trends-2022/amp/

4. https://analyticsindiamag.com/how-synthetic-data-might-impact-ai-in-2022/

5. https://www.pega.com/insights/articles/podcast-recap-vr-spaces-metaverse-and-building-synthetic-data

6. https://towardsdatascience.com/how-ai-will-shape-the-metaverse-4ea7ae20c99

7. https://vmblog.com/archive/2021/12/21/synthesis-ai-2022-predictions-synthetic-data-will-enable-data-centric-ai-and-the-metaverse-what-to-expect-in-2022.aspx#.YepRWOpBzIU

8. https://siliconangle.com/2021/12/27/report-meta-boosts-metaverse-push-help-synthetic-data-startup-ai-reverie/

9. https://www.globenewswire.com/news-release/2021/11/30/2343152/0/en/The-Metaverse-and-AI-Edge-Cases-Will-Drive-Synthetic-Data-Boom-Top-Predictions-for-2022-by-Synthetic-Data-Innovator-Datagen.html

10. https://www.gartner.com/en/articles/you-ll-be-breaking-up-with-bad-customers-and-9-other-predictions-for-2022-and-beyond

11. https://telecoms.com/513411/qualcomm-invests-in-the-european-metaverse/

For more information, contact askus@infosys.com

Infosys®

Navigate your next

Infosys.com | NYSE: INFY

Stay Connected