

Synthetic Data in Healthcare: AI-Generated Medical Imaging

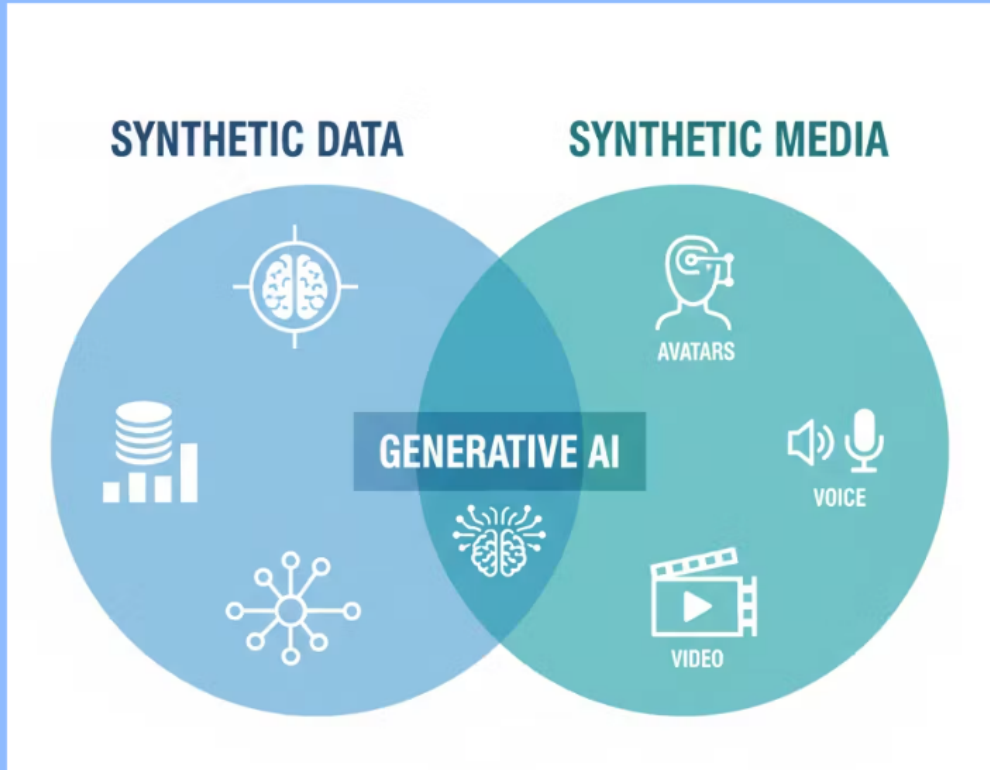


Rayan Sadri

 Carez AI

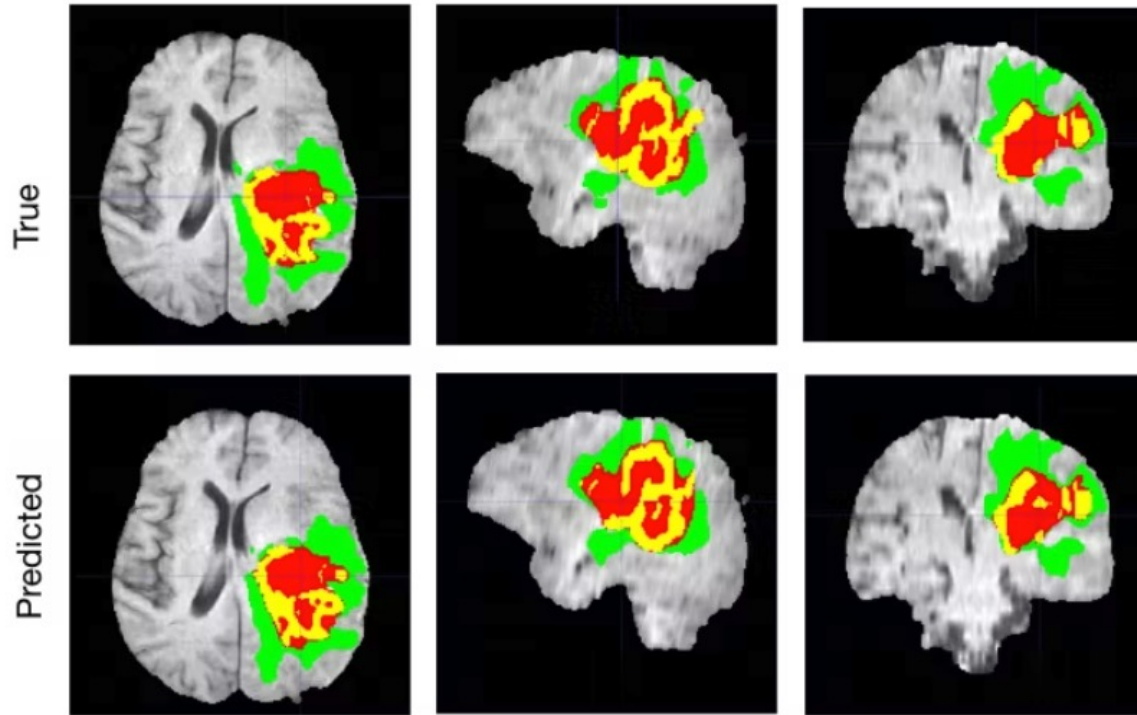
28 October 2025

The Intersection of Synthetic Data and Synthetic Media



- Both fields use generative AI models to create new, artificial content.
- Shared challenges: **trust, bias, transparency, and IP.**
- In healthcare, these technologies converge when we simulate patients, train assistants, or visualize outcomes.

Synthetic Data in Medical Imaging

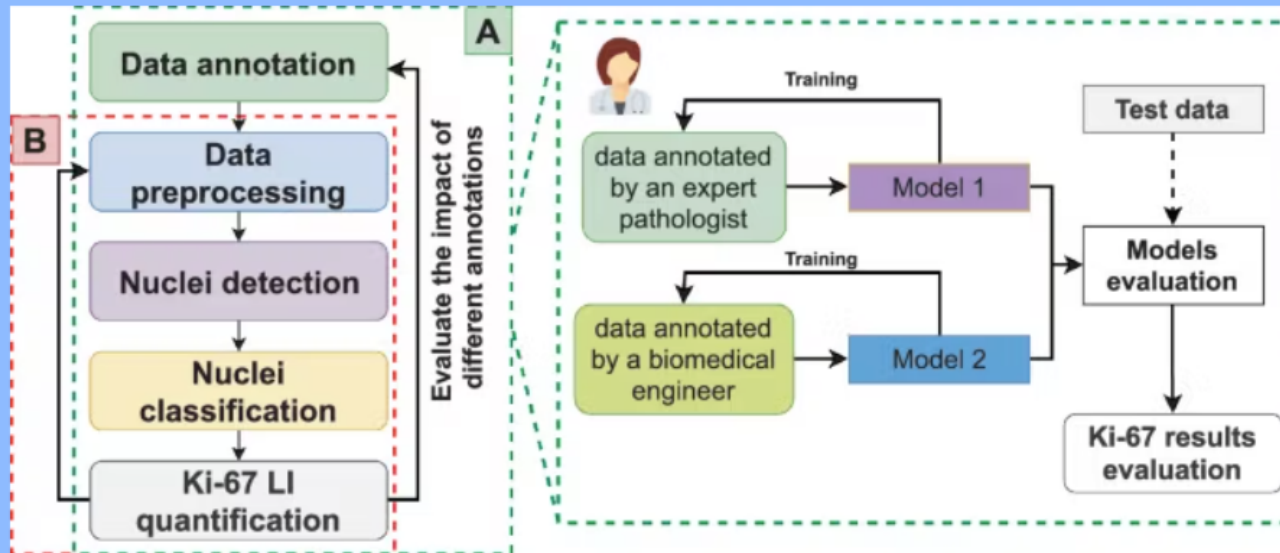


Source: NVIDIA Blog, “Automatically Segmenting Brain Tumors with AI,” Nov 27, 2018. By Nefi Alarcon.

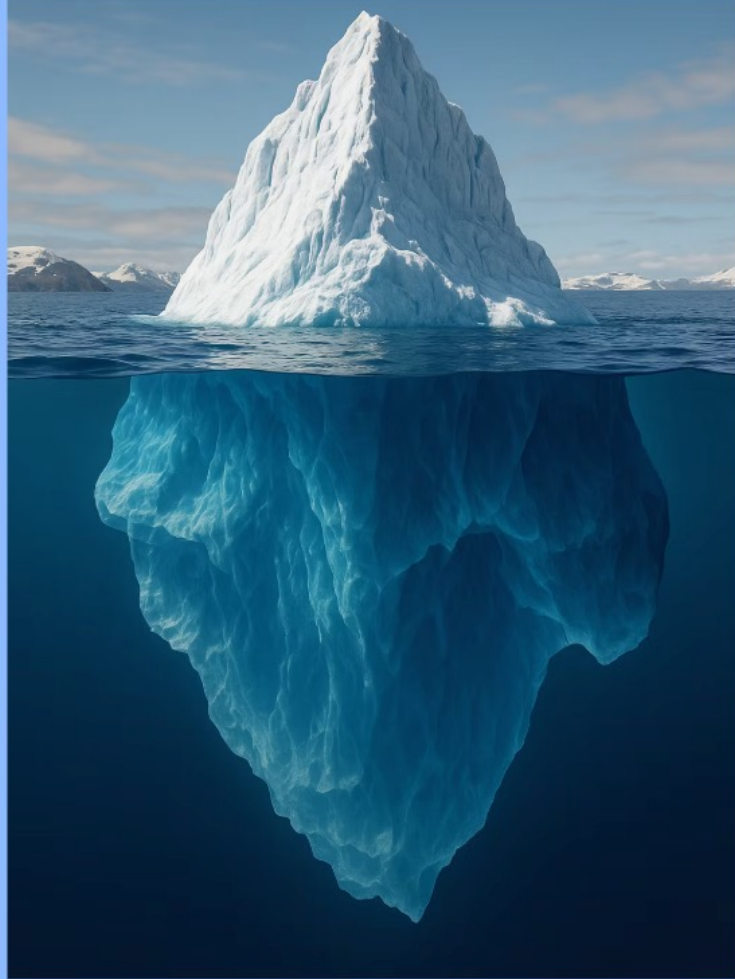
- AI-generated **imaging media** (CT, MRI, ultrasound) help support clinical decisions.
- Synthetic visuals can help **detect patterns earlier** and enhance model training.
- Improves diagnosis speed and accuracy while protecting patient privacy.
- Expands expert-level insights to hospitals and clinics with limited specialists.

How is Medical AI built?

- Collect hundreds of thousands of medical images from hospitals and patients
- Anonymize the data to protect privacy
- Specialists spend years labeling what is in each scan (slow and expensive)
- AI models learn from these labeled images to recognize disease
- Progress is limited by how much high-quality labeled data we can obtain



Source: Benaggoune et al., 2022 (Springer)



The Challenge with training data.

- These datasets come from a limited number of hospitals, scanners, and patient groups
- Rare diseases and edge cases are often undersampled
- Models struggle when they encounter patients who look or scan differently than their training data
- As a result, AI that works in the lab can **fail in the real world**

From Real to Synthetic: Teaching AI to Imagine

Generated from real-world data, but not directly collected.



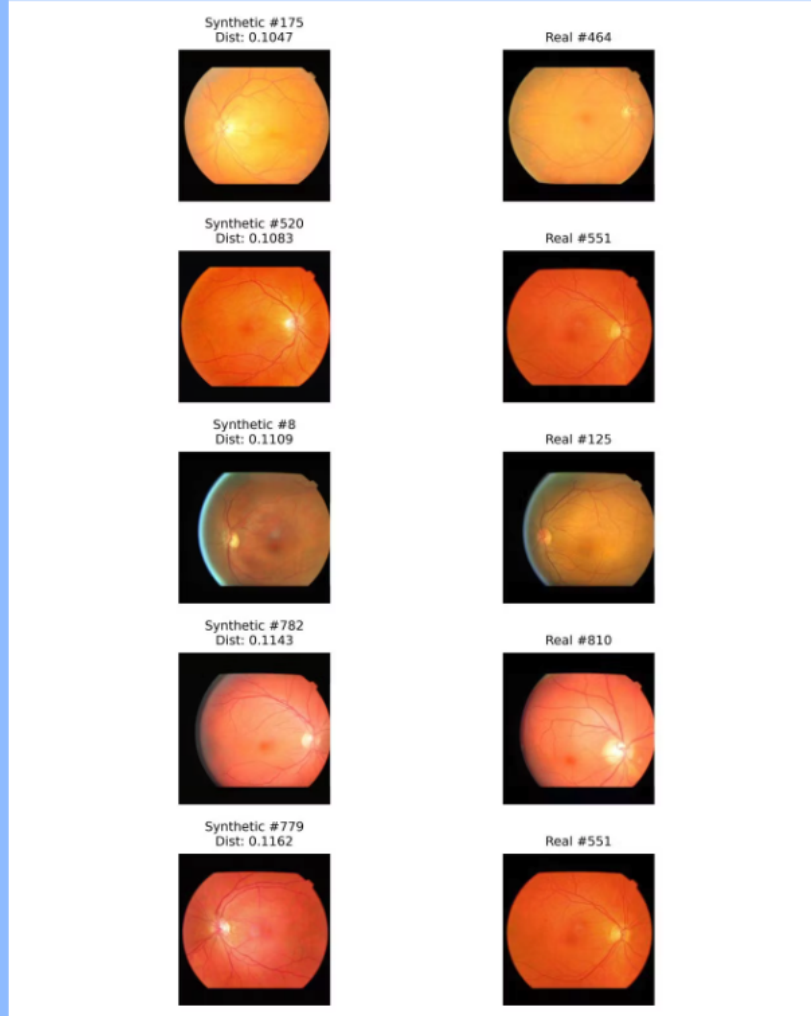
Real Data

AI Model

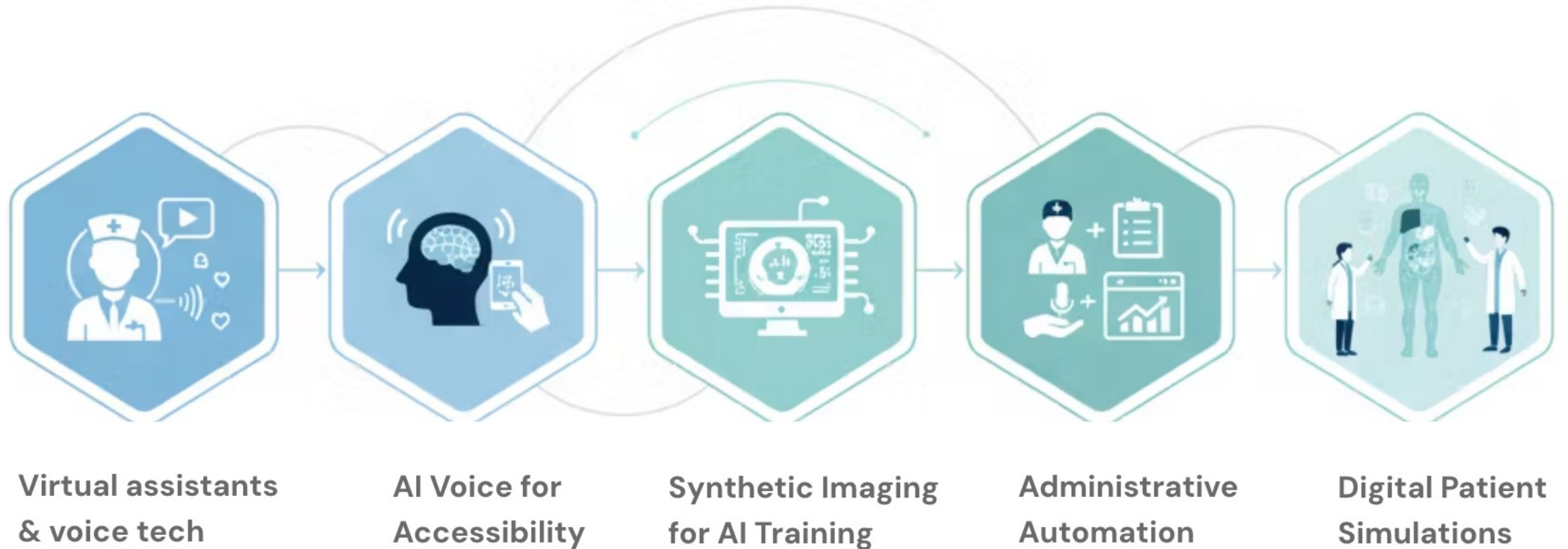
Synthetic Data

Synthetic Medical Imaging: Safe Data for Smarter AI

Synthetic data in imaging is information that is artificially created instead of collected from real patients. It looks and behaves like real medical images, but it does not come from an actual person.



Main Applications of Synthetic Data in Healthcare



Why Synthetic Data Matters for Privacy and Progress

Real patient data is tightly regulated (GDPR , HIPPA), making access and sharing very difficult.

Synthetic data is artificially generated and does not correspond to real individuals, this reduces the risk of exposing personal health information.

Because it is non-identifiable, synthetic data protects patient privacy while still allowing researchers and AI developers to work safely and responsibly.



Challenges



- Ensuring data accuracy and validation for real-world reliability.
- Managing bias and fairness in AI-generated content.
- Addressing data ownership and IP rights of synthetic outputs.
- Building trust through transparency and ethical frameworks.

Conclusion: The Path Forward

- Synthetic data can bridge data gaps while protecting privacy.
- Collaboration between industry, academia, and regulators is essential
- Ethical guidelines and IP frameworks will shape responsible innovation.
- The goal: better, safer, and more accessible healthcare through AI.

Thank you