

Unlocking AI/ML Innovation in Healthcare Using Synthetic Data

Vyas Sekar

Carnegie Mellon University



ROCKFISH
DATA



11th INTERNATIONAL CONFERENCE ON
**TRANSFORMING
HEALTHCARE WITH IT**
23-24 February, 2024
Sheraton-Grand, Whitefield, Bengaluru, India

Agencies Explore Policy Barriers, Synthetic Data to Maintain Health Care Privacy

Policy changes and synthetic data could play a key role in eliminating risks while improving data access and security.

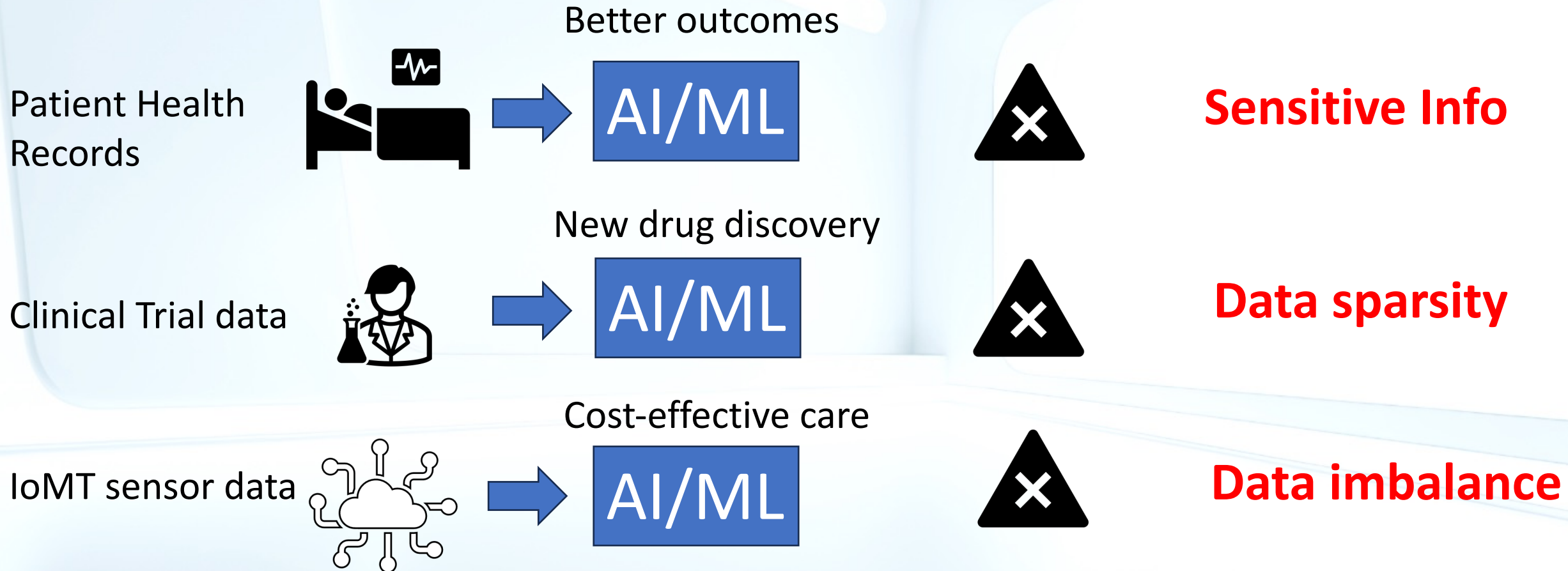
TECHNOLOGY > AI AND AUTOMATION | August 2, 2023

The majority of AI training data will be synthetic by next year, says Gartner

FORBES > INNOVATION > AI

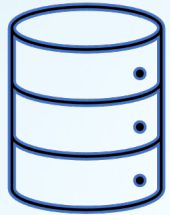
Synthetic Data Is About To Transform Artificial Intelligence

Data Bottlenecks in Healthcare AI/ML



New Opportunity: Synthetic Data

Source
Data



Training
Algorithm



Deep
Generative
Models



Synthetic
Data



Sensitive Info
Data sparsity
Data imbalance

Privacy preserving

Hide sensitive information

Amplify

Create more data on demand

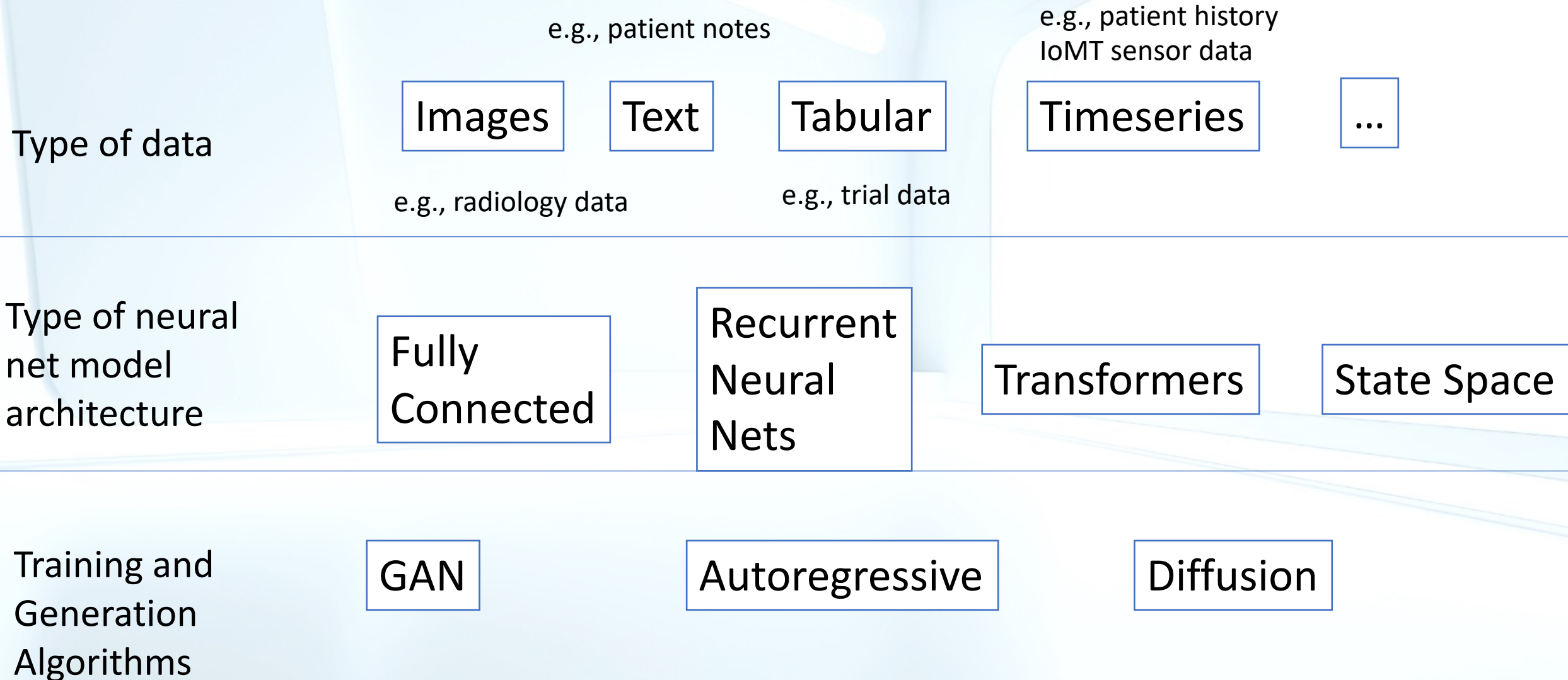
Flexible

Create new scenarios, missing data

Art/Science of Generative Modeling



11th INTERNATIONAL CONFERENCE ON
**TRANSFORMING
HEALTHCARE WITH IT**
23-24 February, 2024
Sheraton-Grand, Whitefield, Bengaluru, India



Example Uses In Healthcare



11th INTERNATIONAL CONFERENCE ON
**TRANSFORMING
HEALTHCARE WITH IT**
23-24 February, 2024
Sheraton-Grand, Whitefield, Bengaluru, India

Simulation and Prediction Research:

e.g., evaluate policy impact on visit rates, prescription, and referral rates

Hypothesis, Methods, and Algorithm Testing:

e.g., predict longitudinal change in A1c among adults with type-2 diabetes

Epidemiological Study/Public Health Research:

e.g., evaluate effectiveness of interventions

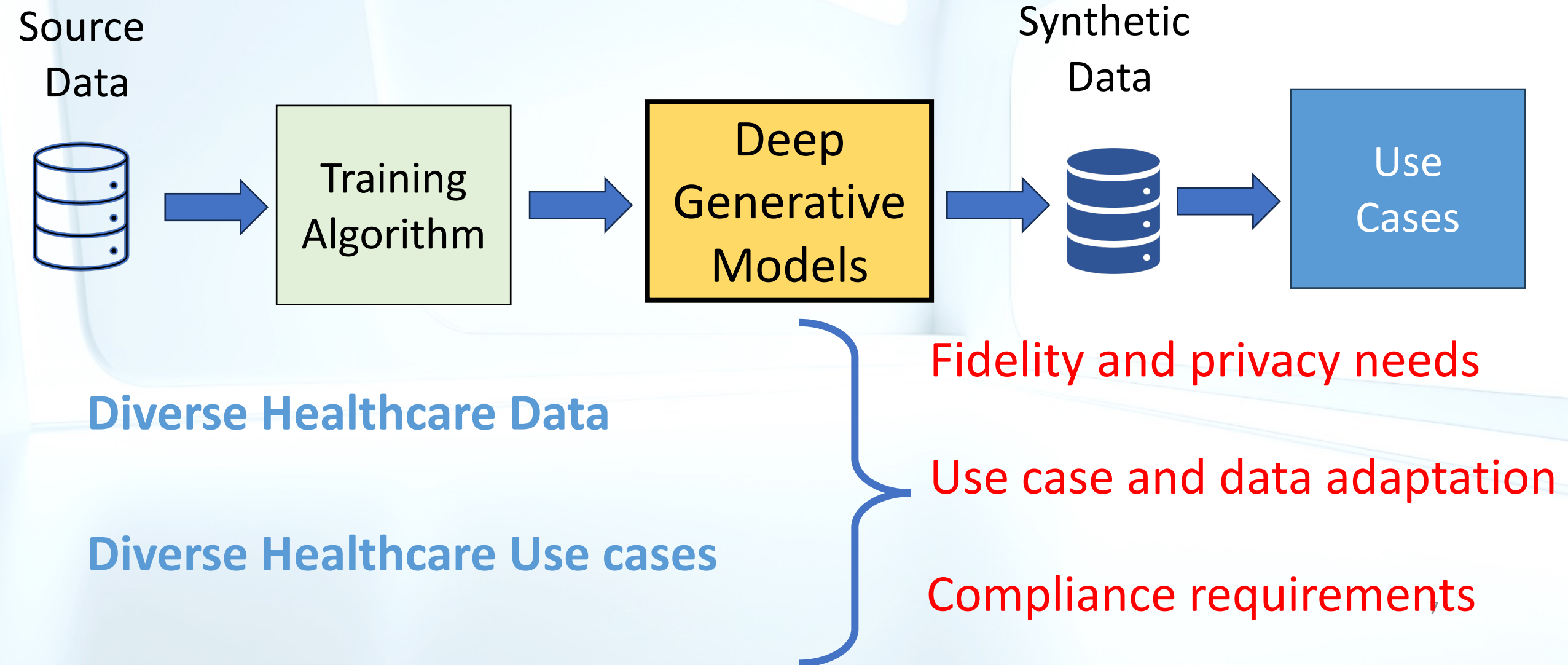
Health IT Development and Testing:

e.g., test software applications in accessing clinical data

Public Release of Datasets:

e.g., prevent personal identification in released datasets

Mind the Gap: Practical Challenges



Fidelity considerations in Healthcare



11th INTERNATIONAL CONFERENCE ON
**TRANSFORMING
HEALTHCARE WITH IT**
23-24 February, 2024
Sheraton-Grand, Whitefield, Bengaluru, India

*Statistical
Properties*

*Downstream
Tasks*

*Domain
Constraints*

Privacy considerations in Healthcare



11th INTERNATIONAL CONFERENCE ON
**TRANSFORMING
HEALTHCARE WITH IT**
23-24 February, 2024
Sheraton-Grand, Whitefield, Bengaluru, India

*Attribute
Privacy
(Protecting Values)*

*Sample-Level
Privacy
(User Privacy)*

*Distributional
Privacy
(Business Secrets)*

Unleash Benefits of Synthetic Data!

Healthcare
AI/ML
Practitioners



Diverse
Healthcare
Datasets



Diverse
Healthcare
Use Cases



***Rockfish
Platform***

Data and use case adaptive

Outcome centric

Secure and privacy preserving

Takeaways



11th INTERNATIONAL CONFERENCE ON
**TRANSFORMING
HEALTHCARE WITH IT**
23-24 February, 2024
Sheraton-Grand, Whitefield, Bengaluru, India

- Many exciting healthcare innovations are blocked by access to data
- Synthetic data can unlock the potential of AI/ML in healthcare
- Early academic work has shown promise, guides intuition, and design
- Exciting time for democratizing use of synthetic data in healthcare
- Would love to hear about your use cases and collaborate!