Synthetic Data Generation

Real Patient Data

(Private 🔒)

Generation Engine

GAN / VAE

Diffusion Models

Statistical Sim

🤖 AI-Powered

Synthetic Data

(Shareable ✓)

Validation
✓ Statistical similarity
✓ Clinical utility
✓ Privacy preserved
✓ Task performance
✅

Use Cases:
• Training
• Testing
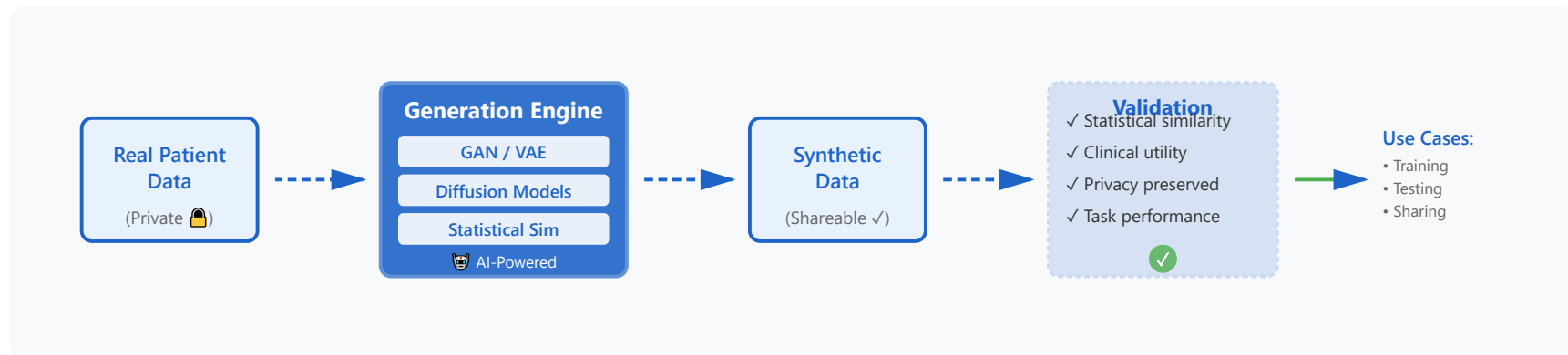• Sharing

Generation Methods

- GANs and VAEs
- Diffusion models
- Statistical simulation
- Physics-based modeling

Privacy Preservation

- HIPAA compliance
- Differential privacy
- De-identification techniques
- Secure data sharing

Validation Approaches

- Statistical similarity testing
- Clinical utility validation

Use Cases

- Algorithm development
- Training data augmentation

- Downstream task performance

- Rare disease modeling
- Clinical trial simulation

Regulatory Acceptance: FDA increasingly recognizing synthetic data for algorithm validation and testing