Statice

Solving healthcare data access challenges with synthetic data

Introduction

From research on new diseases to personalized medicine and studies on drug efficacy, most research and innovation in healthcare relies on the ability to access and analyze data.

However, for most organizations and institutions, data access represents a challenge. Proprietary data formats, siloed IT infrastructures, data governance, regulatory requirements on personal medical data, and growing cybersecurity concerns limit organizations' ability to use data.

To help organizations and health institutions solve data access challenges and drive agility safely, Statice developed a solution that allows teams to generate anonymous synthetic health data that looks and behaves like real data.

Artificially generated from patient records and medical data, the synthetic data produced with Statice's solution contains statistical properties similar to the real data but irreversibly breaks any relationships with actual patients. As a result, synthetic health data offers the highest level of compliance and privacy protection, making it a valuable and safe-to-use data resource.

Synthetic data enables larger-scale data pooling, faster analysis, and cost-effective AI development, allowing organizations and health institutions to safely leverage data while complying with data regulations and healthcare privacy standards.

Use-cases presentation

Synthetic health data is suitable for multiple research and innovation use-cases because it mimics the structure and statistical properties of data gathered from real-life events.

You can use synthetic data as a drop-in replacement for real data in most analysis workflows, like behavior, predictive, or transactional analysis. The privacy protections in place allow you to share data internally or with partners and use it for secondary use-cases.

Roche: sharing synthetic clinical data for Machine Learning applications

The healthcare company Roche collects various clinical and diagnostic data. This data is precious. It notably fuels Machine Learning modeling for predictive medicine and to discover new diseases.

But patient data and medical information is sensitive and can not be processed without consent from the patient for secondary analysis such as research. Additionally, ML applications are data-hungry systems, requiring a new comprehensive dataset for every new clinical trial.

For Roche, traditional anonymization methods are manual and time-consuming. They deteriorate the data quality and granularity.

Together with Roche's data team, we worked to prove the utility and privacy of synthetic health data so they could generate synthetic data from sensitive clinical trial data to serve as training data for machine learning applications.

> *«We worked with Statice on a collaboration investigating the power of synthetic data. This area has been rapidly growing in recent years and we're working to see how to utilize this power in a clinical data setting. We've identified a number of areas where we are interested in, including software and testing, medical data insights, and data sharing.»*
>
> Alex Hughes, Data Sharing Consultant at Roche.

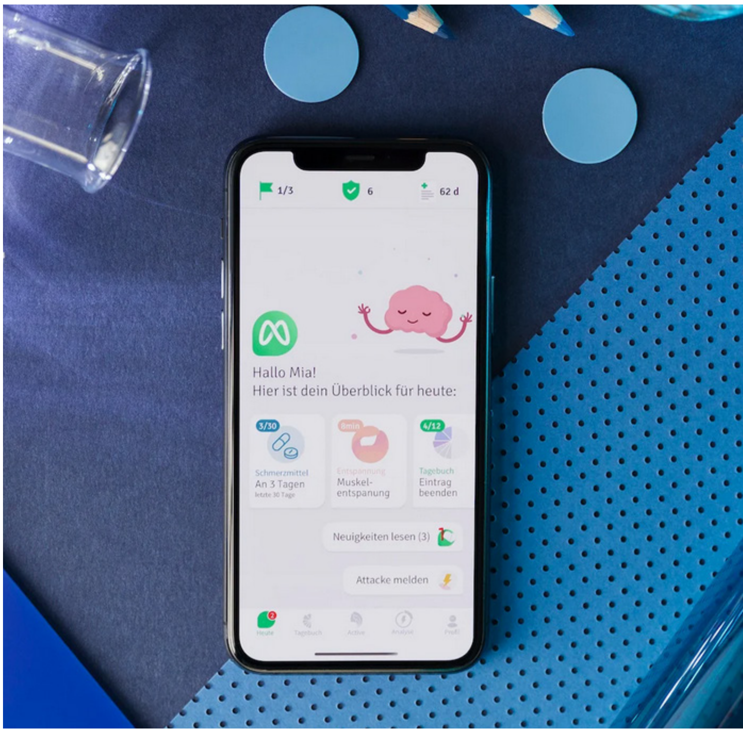
We successfully synthesized laboratory result datasets. It contained new synthetic patients, realistic in terms of statistical distribution but without any possible connections to real patients, confirming that synthetic data was a reliable alternative for healthcare companies to using real patient data.

Compared to traditional anonymization methods, synthetic data is simpler and faster to create for Roche. It offers more robust protection for clinical trial data while still preserving the granularity and general trends of the original data.

M-Sense: anonymizing health user data for research on Migraine

Newsenselab developed M-sense, a migraine monitoring and assistance app for mobile devices that allows people suffering from headaches and migraines to log their symptoms. The team is also committed to helping the scientific community better understand the disease's patterns and causes.
However, users' data represents sensitive medical information regulated by data laws such as the German Digital Healthcare Act or the GDPR. Newsenselab's commitment to their user's data privacy prevented the team from sharing health data with their research partners.

Data masking and pseudonymization methods were not safe approaches for Newsenselab. Even without Personally-Identifying Information (PII), the team knew that datasets could be linked to open data sources to re-identify users.
They implemented the Statice anonymization solution to validate synthetic data as a privacy protection mechanism for user medical data. 170,000 data points went through the privacy-preserving machine learning models to generate an artificial dataset of migraine symptoms over data in a 10-dimensional space.



We ran privacy and utility evaluations on the synthetic so Newsenselab could verify that it provided an equivalent statistical value and the highest level of privacy protection. The results were conclusive, proving their user data, including symptoms frequency and nature over time, could safely be shared with research partners. Additionally, the approach complied with existing personal and health data regulations, guaranteeing the highest level of privacy to their users.

For Newsenselab, synthetic symptoms data could accelerate diagnosis and treatment research through collaboration with other healthcare providers and research groups.

> *«Statice's data privacy solution offers the possibility to advance fundamental research while completely protecting the data of users.»*
>
> Simon Scholler, Head of Research at Newsenselab.

Hackathon and patient data aggregation: anonymized cases from European healthcare organizations

Case A: synthetic patient data for a privacy-preserving hackathon

The research branch of a European hospital decided to organize a hackathon to develop new research ideas on personalized prediction, prevention, and diagnostics.

However, security and privacy concerns and the legal framework around medical data processing in Europe prevented them from sharing patient data.

We worked with the institution to generate synthetic data from clinical observations and patient data. After we demonstrated the anonymity of the data, the institution made the data available for the hackathon and successfully conducted the event.

Case B: aggregating anonymized hospital data for secondary purpose analysis

A European group receives hospital data from its partners. This data is sensitive and cannot be used for secondary purposes, shared off-premise, or combined with other hospital data sets without the patient's consent.

Using Statice, the two partners can anonymize hospital data safely. The synthetic data is safe to be aggregated and can be processed for secondary purposes while remaining compliant with regulatory requirements.

The group data team can create benchmark products from synthetic data, adding a new offering for their health insurance clients.

To discover more about synthetic health data visit www.statice.ai

Statice develops state-of-the-art data privacy technology that helps companies double-down on data-driven innovation while safeguarding the privacy of individuals.

With Statice, companies generate privacy-preserving synthetic data compliant for any type of data integration, processing, and dissemination. Enterprises from the financial, insurance, and healthcare industries drive data agility and unlock the creation of value along their data lifecycle with Statice.