

In conversation with... Omar Ali Fdal, CEO, Staticce



Could you introduce yourself? And what drove you to work on synthetic data?

I'm Omar Ali Fdal, co-founder and CEO of Staticce. We are a Berlin-based company developing Privacy Technology software solutions. We've been working with large organizations from the finance, insurance, and healthcare industries for the last four years. We support their data protection efforts with a particular focus on implementing synthetic data as a privacy mechanism.

I started my career as a research engineer at a major European IT provider for the travel industry. My role was the 2010-equivalent of the modern data scientist, and there I worked daily with large amounts of data. Over the years, I realized the value of secondary data uses and the massive potential of the data economy. At the same time, I experienced the gaps of working with data at the enterprise scale - especially when it comes to privacy and data protection. In 2018, the GDPR was around the corner, and I knew enterprises would need new alternative ways of accessing data safely if they wanted to be compliant. That's when I met a couple of brilliant people who had just started working on this new technology, synthetic data. Their ambition was to enable a privacy-preserving way of working with data. The decision to join them was a no-brainer.

What is synthetic data?

Synthetic data is data artificially created that looks and behaves like real data. Using various algorithms, you generate entirely new data points with statistical properties, structure, and granularity similar to an original dataset. As a result, you get this synthetic data that you can use for most statistical analyses that you had intended for the original data.

In the case of Staticce, we focus our efforts on the generation of tabular and privacy-preserving synthetic data. Tabular means table-like data such as transactions, geolocations, patient records, or time-based customer activity. And privacy-preserving means that we include additional data protection mechanisms to protect individuals' privacy. Because, although the synthetic data generation process irreversibly breaks direct relationships with the original individuals in the data, one still needs to assess the risk of the synthetic data. At Staticce, we do not only generate synthetic data, we also provide tools to help assess that risk.

Where does synthetic data fit in the Privacy Enhancing Technology ecosystem? How do you compare it to other PETs?

The PET ecosystem is relatively large, and new technology developments are constantly blurring its boundaries. It's also hard to position synthetic data precisely as there is no commonly accepted taxonomy to our knowledge. One aspect is sure that all these technologies have one core purpose:

to mitigate privacy risks and protect personal data. And for that reason, synthetic data belongs in that space.

Compared to most PETs, synthetic data provides an interface that is most familiar to all data practitioners, and that is data itself. This is a key aspect, it means that the same pipelines, tools, processes, infrastructure, that is already in place to handle real data, can be used with synthetic data.

Another dimension that is specific to fully synthetic data, is the absence of a direct link between synthetic data records and real ones. There is no key that you can use to retrieve the original record linked to a synthetic record. This means that it offers more guarantees by design. This is important, especially in internal or external data release scenarios.

A final aspect makes synthetic data complementary with other PETs. The training of the synthetic data generation algorithms, is itself a processing. As such, we can make it Differentially Private, or we can run it in a decentralized fashion using Federated Learning.

What value can synthetic data bring to different industries?

Today, all customer-facing industries must process personal data. In parallel, a growing number of countries are adopting stricter modern data regulation laws. As a result, virtually all industries relying on personal or sensitive data to innovate or function can benefit from synthetic data.

We've seen that enterprises from historically regulated sectors, namely finance, insurance, and healthcare face similar data access challenges. Regardless of the vertical, the inability to share and process data impedes innovation.

There are some specificities though. In the finance and insurance industry, challenges start already with internal data access. This is not a surprise when we know that a large proportion of data leaks are caused by internal actors.

In healthcare on the other hand, collaboration and research are the main drivers for medical breakthroughs. Again, due to the high-sensitivity of real data, these exchanges are frozen.

All in all, synthetic data represents safe, easier-to-access data assets for teams who rely on data to operate.

How do you see the future of PETs?

Bright, definitely. Consumers' expectations from service providers and third-parties are increasing. The regulations are adjusting to that expectation all over the world, with GDPR being the most prominent example. Many feared that these trends would put a stop to data-driven innovation. But what happened was the emergence and the productization of a number of technologies that enable this innovation, and make data usage more responsible and transparent than before.

But it's not only the availability of these solutions that is growing, their adoption is too. We have recently seen massive interest and investments. In the synthetic data field alone, companies raised at least 120€millions in the last eleven months. More and more companies are working on developing strong, reliable, and user-friendly technology for enterprises and consumers.

The privacy research community is also growing and I believe that we will continue to see new technologies and techniques arise.

Why are you looking forward to the Privacy-Enhancing Technology Summit?

First of all, we are glad to see that PET Summit 2022 is actively working on bringing together the researchers, technology providers, the users, and the regulators at the same table. We are deeply convinced that our ecosystem can only mature with all these stakeholders participating in the discussion.

And on a personal note, we are excited about returning to Switzerland. My last business trip before the pandemic hit Europe was to Zürich for a customer meeting. Since then, we moved to a full-remote setup, and so did our customers and partners. Nevertheless, we are always happy to meet people in-person and looking forward to a great conference!



Omar Ali Fdal
CEO
STATICE

OMAR WILL BE SPEAKING ON
'THE FUTURE OF SYNTHETIC
DATA AND PRIVACY -
DOWNLOAD THE AGENDA HERE!