The Alan Turing Institute

HSBC
Synthetic Data Engineer

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

The Data and Architecture Office (DAO) was established to transform HSBC into a truly data-driven organisation by ensuring that every design decision has a built-in data consideration. As DAO we will provide leadership that enables data and analytics to be a key driver of our business strategy. Working closely in partnership with senior stakeholders and delivery teams all over the world to connect the business processes, data, products, services, and technology that underpin customer value and financial success.

Group Business Intelligence & Analytics is a new Group-level function. Business Intelligence & Analytics unlocks additional value from our data providing insights to improve customer experience, increase revenue, optimise capital, enhance risk management, deliver regulatory compliance and reduce costs.

Group Business Intelligence & Analytics will offer the following services:

•   End-to-end production of management information, reporting and business intelligence.

•   End-to-end production of business analytics and data integrity analytics use cases.

•   A self-service visualisation platform.

•   A self-service analytics & artificial intelligence platform.

Role Description and Responsibilities

The Synthetic Data Engineer Intern is responsible for developing GCP PaaS (Platform as a Service) for synthetic data generation, solving cutting edge problems and working with the rest of the team to productionise the solution. This would be done using the already existing blocks and API, however generalisation and scaling to enterprise level will require creativity. We are looking for people who are interested in cloud engineering (namely GCP), and have curiosity for innovation and AI/ML.

The Synthetic Data Engineer Intern will partner with architecture and cloud teams to develop practical solutions and custom modules for Synthetic Data as a Service and will need to ensure that all cloud solutions follow security and compliance controls.

The Synthetic Data Engineer Intern will work in a fast-paced environment where everything

evolves very quickly and will maintain momentum by promoting agile working and ensuring we constantly deliver for our clients. We collaborate by listening to new ideas, as well as voicing our own, in order to deliver and develop the best service for our clients.

Expected Outcomes

Synthetic data refers to artificial information created to serve as substitute data for developers and businesses.

Gartner predicts that by 2024, 60% of the data used for the creation of AI and analytics projects will be synthetically generated.

To response to the growing needs, HSBC initiated **Synthetic Data as a Service** project. Its goal is to create capability for synthetic datasets generation, that preserve both the utility and privacy of our data. The generated synthetic data would be used for rapid prototyping and experimentation internally and support external collaboration with start-ups, academics and technology partners.

The approach to achieve this to combine internal and external capability, develop in-house synthetic engine and expertise, and combined with leading academia and vendors.

Currently, we have several synthetic data generation tools for structured and unstructured data, conceptual architecture with data flows and approvals, and API for Synthetic Data as a Service user (to upload real data, get information on the capabilities, and report on privacy and utility evaluation). We also collaborate with the Alan Turing Institute on several synthetic tools creation.

The outcome being to pilot Synthetic Data as a Service PaaS in HSBC this year.

Supervision and Mentorship

The synthetic data engineer intern will work in an agile scrum team alongside Cloud Engineers, Architects and Data Scientists to deliver GCP PaaS (Platform as a Service) for synthetic data generation.

The Intern will be supervised by an experienced Innovation Research Lead, they will be part of the Data Science Chapter. The Chapter will provide them with a network of people to consult with as well as provide exposure to different technologies and use cases.

Key skills to be gained from the role:

- Real-world, hands-on experience using Google Cloud technologies (cloud infrastructure, API, web services) essential for future career as Cloud Engineer
- Experience in platform development in a large financial enterprise
- Experience in synthetic data generation using advanced Machine Learning methods (GANs, autoencoders)

Person Specification

The ideal Intern will have a passion for data to improve customer experiences and change businesses.

- They will have a foundational level knowledge of cloud platforms and data management and security;
- basic/intermediate coding ability in a number of computer languages (e.g. Python);

- strong problem-solving, written and oral communication skills.

Internship Logistics

Salary: £30,000 p.a pro rata.
Internship length: 6 months, full time.
Location: Remote with occasional travel to London to work alongside colleagues or for training.