Financial Services Consumer Panel
AN INDEPENDENT VOICE FOR CONSUMERS OF FINANCIAL SERVICES

Email: enquiries@fs-cp.org.uk

Synthetic Data Expert Group
Financial Conduct Authority
12 Endeavour Square
London E20 1JN

17 May 2024

Submitted by email: SyntheticDataEG@fca.org.uk

Dear Sir / Madam,

Financial Services Consumer Panel response to FCA's Synthetic Data Expert Group Report on using synthetic data in financial services.

Introduction

The Financial Services Consumer Panel (the Panel) is an independent statutory body. We represent the interests of individual and small business consumers in the development of financial services policy and regulation in the UK. The consumer voice is critical to this debate as it is financial services consumers who ultimately stand to benefit from, or be harmed by, developments in the use of synthetic data.

On 22 June 2022, the Panel responded to the FCA's Call for Input on synthetic data to support financial services innovation. The Panel welcomes the opportunity to respond to the Synthetic Data Expert Group's (SDEG) report at this time. With this letter, the Panel has taken the opportunity to respond to the key considerations and recommendations from a consumer perspective, as well as respond to the questions addressed in the report.

The Panel observes that the SDEG's report analyses synthetic data issues and use cases largely from a firm perspective and under-focusses on consumer outcomes and engagement. Mindful of the FCA's consumer protection objective-to provide an appropriate degree of protection for consumers-we hope to see consumer outcomes and engagement explored in future reports. To this end we set out in this response the key consumer messages that the Panel has outlined in previous responses and which we feel lack the necessary emphasis in the SDEG's report.

The Panel agrees that Artificial Intelligence holds significant potential in the financial services industry to make products and services more efficient and accessible for consumers. The extent of its potential depends on the availability and accessibility, and fair and ethical use, of data for financial services providers. The Panel recognises that, although synthetic data is artificially generated, it can be a great alternative and/or complementary to real data as it addresses some key consumer protection issues with data collection, usage and storage.

Benefits, challenges, application of use cases

The Panel agrees with the SDEG's assessment of the potential benefits of synthetic data to unlock opportunities in financial services for the benefit of consumers, especially, its potential to:

- overcome challenges associated with expensive and time-pressured data collection.
- overcome challenges associated with data scarcity and availability to test obscure cases or validate new data concepts and scenarios for products and services.
- help mitigate bias by generating diverse data that is representative of a given population creating a more fair and inclusive financial landscape.
- enhance training models with sufficient quantity and quality of data.
- enhance and protect the privacy of consumers without compromising UK/EU GDPR rules and requirements associated with the UK Data Protection Act.
- mitigate many privacy and security concerns linked to internal and external data sharing.

The Panel also recognises the practical challenges and limitations associated with creating and using synthetic data, which if left unmitigated increase the risk of harm to consumers (e.g. by increasing the risk of bias, data theft/loss, breaches of privacy etc.) We consider those challenges to be:

- synthetic data depends on real data to generate and validate.
- synthetic data impacted by the quality of real data or externally sourced synthetic data can lead to unreliable and misleading results.
- synthetic data may not be fit for its intended purpose due to deliberate or unintentional human interference with models to generate synthetic data that includes bias, discrimination, and exclusion.

- synthetic data is not realistic, representative, and reliable enough to compare with real datasets leading to misleading information and results that are unreliable. Depending on how synthetic data is developed it may result in unrealistic and unreliable results where synthetic data contains outliers that could be replicated in synthetic dataset and interfere with results. Removal of outliers and ratio adjustment of real-life data to synthetic data may help improve the accuracy of trained models without compromising data privacy requirements.
- synthetic data is not private enough to protect consumers' and firms' characteristics resulting in the unlawful use of consumer data for processing and sharing. Synthetic data with Privacy Enhancing Techniques (PETs) could still fail to fully protect consumer data.
- synthetic data that is not agile, time-dependent and relevant to economic circumstances and conditions may result in models underestimating or overestimating risks and predictions for consumers of the future.
- there are currently no industry standards and metrics to evaluate synthetic data properties such as privacy, utility, and fidelity levels.
- beyond the FCA's Regulatory Sandbox, Digital Sandbox and TechSprints, there is no open data community currently available to build and share synthetic data and models.[1]
- no specific regulatory standards, governance frameworks or legal requirements are available to validate models with synthetic data and define what good looks like for synthetic data.
- there is a risk that some firms may not be sufficiently equipped to understand the limitations of the creation, use, methodology and maintenance of synthetic data models.
- Firms are not sufficiently resourced with subject matter experts to evaluate, analyse, measure, assess, continuously refine, and validate the privacy, reliability and performance of synthetic datasets and results. Where firms use external subject matter experts, they must manage them appropriately to ensure personal or confidential information is not compromised.
- firms are not sufficiently equipped with synthetic data governance frameworks, centralised systems, and guidelines to ethically consider the appropriateness of use cases, manage privacy nuances and safeguard data integrity and confidentiality. We would like to refer the FCA to comments we made in response to the Digital

[1] One example of another possible model is the one being developed by the Centre for Net Zero and Linux Foundation to democratise access to synthetic data

Regulation Cooperation Forum (DRCF)'s consultations about the fair use and governance of algorithms.

The Panel agrees with the applications of synthetic data use cases highlighted in the report for the benefit of consumers, such as the ones listed below:

- combat money laundering and fraud by developing synthetic datasets to train legitimate and illicit money laundering or fraud detection models.
- create diverse data to mitigate selection bias in credit scoring, lending applications and product offerings.
- create data labels to enhance training datasets.
- meet privacy requirements and prevent real consumer data from being used and shared during tech sprints and data sharing.
- overcome intra and inter-organisational boundaries of data sharing allowing cross-organisational and cross-disciplinary collaboration to detect and prevent illicit, fraudulent transactions and scams. The latter can be extremely useful in organisations with legacy data systems and silos.
- create standardised datasets to facilitate public collaboration for innovation and community research into societal challenges.

The Panel believes that the most important use cases of synthetic data within financial services are helping generate data to protect consumers from fraud and scams, as well as helping prevent personally identifiable consumer data from being used and shared.

The Panel observes that the SDEG's report analyses synthetic data use cases largely from a firm perspective and under-focusses on consumer outcomes and engagement. We hope to see this explored in future reports.

Consumer protection: key consumer messages

In line with FCA's operational objective to secure an appropriate degree of protection for consumers, the Panel would like to reiterate the key consumer messages that we have set out in previous responses in the context of synthetic data that we feel lack the necessary emphasis in the report:

- Synthetic data could facilitate greater innovation and allow consumers a broader choice between products and services as it provides a levelling playing field between new entrants and incumbent financial institutions in terms of data wealth. For this benefit to be realised for consumers, there mustn't be any unreasonable barriers to accessing, switching or cancelling products and services.

- Synthetic data systems and procedures should be designed and delivered with consumers in mind to minimise consumer harm (through collection, use and sharing of data) and increase positive outcomes for consumers.
- Consumer representation throughout the development of systems, processes and procedures related to synthetic data is essential. Understanding consumer needs, experiences and potential risks will keep regulators and firms focused on consumer outcomes and ensure that synthetic data is being used fairly and ethically to provide benefits for consumers.
- Firms should only use synthetic data as a substitute for real data where it is more appropriate to do so e.g. where real data is unavailable or of insufficient quality, or where synthetic data better protects privacy and/or confidentiality.
- Firms may sometimes utilise synthetic data to inform decisions regarding the design and delivery of products and services that can benefit consumers. Firms may also use synthetic data to model the value and impact of their decision-making on consumers with different characteristics of vulnerability.
- If firms are to use synthetic data to inform decisions or design and deliver products, they should make it transparent to consumers about where synthetic data was deployed within their customer journey and where human judgment was added as an overlay. Firms should also put consumers in a position to understand what personal information is collected, and recorded, the purpose of collecting that data, and how it will be used (if any) to develop synthetic data (including what data fields are used to drive decisions and shared. The FCA should make clear to firms that their use of synthetic data in product and service design and that such use is communicated to consumers as firms' responsibilities under the Consumer Duty.
- Firms must ensure that synthetic data is fair, representative, and unbiased when making real-world decisions based on analysis from synthetic data. Algorithms should promote inclusion and reduce exclusion. The underlying source data informing algorithms and synthetic data should be accurate, current, inclusive, and representative so that they do not create unreasonable barriers or discriminate against individuals or groups of consumers, including those in vulnerable circumstances or with protected characteristics.
- Firms should regularly test their AI models to ensure the design of products and services that use synthetic data are inclusive and

considerable of all their customers including those with vulnerable characteristics.

- Firms should also maintain sufficient control and understanding of the synthetic data and algorithms they use so that they are capable of assessing and monitoring them to understand how they are impacting consumers. The firm's governance arrangements should expressly include procedures and arrangements for governing the use of algorithms and synthetic data. This will ensure that firms can be held accountable, by consumers and regulators for how their use of data and algorithms impacts consumers and affects outcomes.

Conclusion

The Panel agrees that the insights in the FCA's SDEG Report have helped gain a better understanding of practical use cases of synthetic data in financial services along with practical challenges and opportunities of using synthetic data.

The Panel continues to believe that synthetic data presents a valuable resource in financial services to better understand and protect consumers. The Panel believes that responsible innovation can lead to improved market integrity and downstream benefits for financial services consumers of products and services.

However, the Panel observes that the SDEG's report analyses synthetic data and use cases largely from a firm perspective and under-focusses on consumer outcomes and engagement. Mindful of the FCA's consumer protection objective, we hope to see consumer outcomes and engagement explored in future reports.

The Panel also believes that the insights on this report highlighted the importance of fostering international collaboration between organisations, regulators, law enforcement, academics, and other subject matter experts to generate and test emerging technologies for societal benefit. In line with the efforts of the National Data Strategy (2022) and the UK Government's recent publication of its Pro-Innovation strategy, the Panel continues to encourage the FCA to work with fellow domestic (through the DRCF) and international regulators (through the Global Financial Innovation Network) and other relevant private and public institutions to ensure a consistent and joined up approach on the ethical and permissible use of algorithmic systems and data.

The Panel appreciates the FCA's commitment to innovation and in particular the use of data and technology in the financial services industry and would like to thank the FCA for the opportunity to respond to this report. We look forward to future engagements with the FCA on the use of synthetic data as the thinking develops.

Yours sincerely,

Helen Charlton

Chair of the Financial Services Consumer Panel