Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on

Advancing Privacy Enhancing Technologies

Syntheticus Ltd

All comments are done by Mr. Aldo Lamberti, Founder & CEO of Syntheticus Ltd. The contact information is:

- Syntheticus Ltd, Mr. Aldo Lamberti, Arnikastrasse 7, 8953 Dietikon, Zürich Switzerland
-

Syntheticus Ltd. is a Switzerland based company that empowers global organizations to democratize access to data, accelerate speed-to-market and mitigate privacy risks by enabling secure collaboration on synthetic data at scale through an enterprise-grade SaaS platform which results in business value and innovation. The Syntheticus Privacy-Enhancing Platform orchestrates the latest technologies such as 'AI-powered Synthetic Data', 'Differential Privacy' and 'Microsoft-based Trusted Execution Environment'.

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.
   a. Better explainability of Deep Learning models ('Generative Adversarial Networks' to generate privacy-preserving synthetic data) to mitigate the "Black Box" assumptions from the users.
   b. Research opportunities are certainly within the realm of Quantum Computing => how easy will it be to re-identify the original real-world data from the artificially generated synthetic data?

2. *Specific technical aspects or limitations of PETs:* Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

a. Pseudonymized data is not free to use according to GDPR. Solely fully anonymized data or synthetic data are compliant with PII requirements and thus free to use (Recital 26 GDPR). According to our personal exchanges with the highest European authorities, the European Data Protection Supervisor (EDPS), even encryption-based technologies are seen as 'not free to use' given the high risk of losing the "key" and thus being able to decrypt the sensitive data.
b. Internal enterprise constraints are IT silos and legal/compliance which hinder Analytics and Software Dev Teams to get full access to the rights data for their daily jobs. Currently either (i) they don't get any access to sensitive information, (ii) they only get anonymized data with lower utility or (iii) they get access to full sensitive data with high privacy/security risks involved.
c. When generating differentially-private synthetic data there is a mathematical tradeoff between utility and privacy. The higher the chosen privacy level, the lower the utility of the synthetic data, and vice versa. The level of privacy can be adjusted by the customer through a parameter (epsilon = privacy budget).
d. The mitigation of bias within the underlying real-world data is an issue. The conditional generation of synthetic data can overcome specific biases and create data sets that are not biased anymore and thus can be used for further advanced analytics without harming minorities and its equity.

3. *Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:* Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

We are active in the following areas with specific applications to bring benefits to stakeholders:

i. Financial Services Industry
    1. Fraud Detection, Anti-Money-Laundering
        a. Challenge: Fraud patterns are different between countries, regions, businesses. It's hard for regulators or private enterprises to detect the anomalies in real-time, given the little amount of training data.
        b. Solution: enterprises and regulators can generate differentially private synthetic data and securely collaborate on that data to better train their individual detection algorithms.
    2. Enterprise Risk Rating
        a. Challenge: the underlying data is limited and thus the enterprise risk rating is lacking forward-looking scenarios.
        b. Solution: by generating an augmented synthetic data set and creating artificial edge cases, those scenarios can be tested right away.

3. Lending Risks
    a. Challenge: When banks run their Machine Learning analytics to classify their customers or to predict default risks, there are inherent biases, such as from gender or ethnicity.
    b. Solution: With the help of synthetic data these biases can be equaled out by generating datasets that have parity between the biases => 50% men and 50% women in a dataset, while mitigating the privacy risks.

ii. Healthcare
    1. One of the most sensitive data is our own health data, which in turn is also of utter interest to unlock value from it for our personal and societal well being
    2. We are active in various Horizon Europe consortia around the topic of opening up access to health data for augmented insights while mitigating privacy risks through synthetic data. One of our partners is Prof. Chou from Stanford University.

iii. Telecommunications
    1. TelCo providers are under constant pressure to digitalize and enhance their products and services. During their app development their engineers obviously need to test the apps before launching to customers. The access to real-world data is limited, and thus the testing is mediocre which results in customer churn => synthetic data is used instead of real-world data for a neatless and compliant app testing for improved apps and happy customers.
    2. Synthetic data allows TelCo providers to interchange between each other insights about customer behavior or business logics to prosper the ecosystem while mitigating privacy risks.

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs:* Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking

authorities under the Office of Management and Budget, the Federal Trade Commission, and **financial regulatory bodies**, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (*e.g.,* Other Transaction Authority) to procure PETs for development.

a. Office of Management and Budget
    i. United States Digital Service
        1. Challenge: in order to improve federal websites and its applications, a thorough testing workflow needs to be in place. For software testers, internal or external (in foreign jurisdictions), access to test data is mission-critical. However, sharing sensitive test data should be avoided in order to avoid privacy breaches
        2. Solution: synthetic data is used as test data instead of real-world citizen data. Software testers can build better products and maintain them, while mitigating privacy risks.
b. Federal Trade Commission
    i. Bureau of Consumer Protection
        1. Challenge: private organizations are applying more and more advanced analytics (Machine/Deep Learning) to unlock insights and value from its collected consumer data. The used algorithms are private and won't be disclosed.
        2. Assumption: The Bureau of Consumer Protection has little overview over the used algorithms and whether they're benign or malign versus the consumers.
        3. Solution: The Bureau of Consumer Protection can install a synthetic data sandbox and then let the private algorithms perform on it to create a market benchmark and possibly detect outliers that potentially are working against the wills of the consumers.
    ii. Office of Technology Research and Investigation
c. SEC U.S. Securities and Exchange Commission
    i. The Financial Conduct Authority (FCA) from the United Kingdom started with pilot projects on the exploration of synthetic data and its benefits for the financial sector one year ago.
    ii. According to the FCA, the benefits are manifold and promise to unleash the synthetic data economy:

1. enable greater data sharing for the purposes of competition, without undermining data protection laws that are in place to protect consumers.
2. alleviate the data access challenge by making synthetic financial data available to participating firms
3. enable more effective testing and product development

iii. Therefore, the FCA opened up a 'Call for Input' on specific questions around synthetic data to explore market attitudes towards synthetic data, and its potential for opening data sharing between firms, regulators and other public bodies. Syntheticus answered to this 'Call for Input' with their deep expertise in the field.

iv. The SEC can build upon the previous experiences of the FCA and explore the setup of synthetic data within the financial sector.

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

6. *Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:* This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and

territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

a. The introduction of a standard on the quality of synthetic data is of utter importance. The questions from the public and private sphere are most of the time around the guarantee (both technical and legal) on how good synthetic data is. For that, we're in close contact with the highest European organization, the European Data Protection Supervisor (EDPS), to work on some initial thoughts about a potential standard.
b. The 'Privacy Shield" with the EU is an on-going issue that needs to be reflected on.

7. *Risks related to PETs adoption:* Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

a. The risk analysis of synthetic data is an ongoing research field, given the above mentioned trade-off between utility and privacy.
b. We at Syntheticus implement quality assurance and validation steps after every synthetic data generation. However, there could still be the risk of re-identification of the original data through 3rd party data and dedicated inference attacks. Such a re-identification would cause tremendous damage to the individuals or business information present in the original real-world dataset.
c. There is no 'silver bullet' PET that can deliver the highest utility without any privacy risk. It's a balance act and needs to be carefully taken into consideration for every individual use case. Anyhow, policy and governance measures need to be in place and also enforced, in order to mitigate the risks.

8. *Existing best practices that are helpful for PETs adoption:* Information about U.S. policies that are currently helping facilitate adoption as well as best practices that facilitate responsible adoption. This includes existing policies that support adoption, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This also includes information about where and when PETs can be situated within tiered access frameworks for accessing restricted data, ranging from publicly accessible to fully restricted data.

9. *Existing barriers, not covered above, to PETs adoption:* Information about technical, sociotechnical, usability, and socioeconomic barriers that have inhibited wider adoption of PETs, such as a lack of public trust. This includes recommendations on how such barriers could be overcome. Responses that focus on increasing equity for underserved or marginalized groups are especially welcome.

a. Gartner => Target market adoption for synthetic data is <5% for enterprises (Innovation Insight for Synthetic Data, Choudhary Farhan, Published 7 February 2022)
b. The privacy breaches throughout the last 10 years have been tremendous. But still, the enforcement of fines is lacking on a large scale. If there's no enforcement, privacy breaches remain a small risk in the overall enterprise calculations. The EDPS, with its leader Wojciech Wiewiórowski, is trying to counteract that topic.
c. Given the global privacy breaches, public trust has suffered. The trust needs to be re-established first - PET's will play a crucial role in this endeavor.
d. A global, or at least a more regional legal framework concerning privacy is needed. The EU, with the GDPR, takes a leading role in that. Other regions are taking similar approaches (China, India, Brazil). The USA instead is rather fragmented, having privacy regulations on state level

(California, New York, Massachusetts, etc), and also on specific sectors (HIPAA for health data or GLBA for financial data). This could be a recommendation to overcome barriers of technical and legal usability.

10. *Other information that is relevant to the adoption of PETs:* Information that is relevant to the adoption of PETs that does not fit into any of the topics enumerated above.