How innovators outperform **competitors in banking**

Your guide to accelerating digital
transformation with synthetic data

Contents

Executive Summary

2021 will be a hard year for the financial sector. The winning strategy is to accelerate digital transformation: build on customer trust and establish true data-centricity by developing synthetic data capabilities to produce accurate and privacy-compliant synthetic truths to be used as drop-in placement for sensitive, raw data. Synthetic data alternatives increase data literacy, streamline privacy-compliance, minimize data retention, speed up vendor validation, empower cooperations with academia and cross-industry players, enable the development of customer-centric products, mitigate data leak risks and improve the performance of AI and machine learning models.

- Organizations should explore synthetic data opportunities in 2021 by first looking at tabular data assets and setting up synthetic data lakes to provide easy, privacy-compliant, and painless access to decision-ready synthetic data.

- Reviewing currently used anonymization techniques should not be missed because what was good enough just a few years ago does not cut it today due to adversaries' technological advances and the increasing amounts of data collected.

- When selecting a synthetic data vendor, look for robust experience in the financial industry, quality assurance regarding both accuracy and privacy, on-premise deployment, and third-party recommendations.

Once we entered crisis mode, most priorities didn't shift. Instead, they became mission-critical.

2021: Shifting landscapes in banking and finance

According to the Gartner Board of Directors survey, seven out of 10 boards have accelerated digital business initiatives as a response to COVID-19 disruption.[1] The general agreement across the industry is that organizations need to step up their game in digital transformation efforts, and they need to do so now. Failure to accelerate data capabilities this year will be fatal for enterprises. According to McKinsey, 2021 will be the year when ROEs bottom out.

Without management action, the global average industry ROE could fall to 1.5 percent. "One likely failure point involves data: in the crisis, a lot of banks' data have proved not just imperfect but nearly useless. That's how much consumer and corporate behaviors have changed. Banks need to distill what they've learned about data and apply it broadly."[2] Shifting behaviors, tightening regulations, and increasingly dangerous adversaries are the signs of our times, accelerating ongoing financial industry trends.

However, there is some good news. Trust in banks has withstood the pandemic: Gartner's trust survey found that 91% of people's trust in banking increased or didn't change during this difficult period. It is this trust banks and financial institutions need to build on in the coming years. To remain relevant, banks need to make data management a priority and unleash its full potential. The opportunity is still present. While institutions that made strategic investments in technology came out stronger, laggards may still be able to leapfrog competitors if they take swift action to accelerate tech modernization.[3] The solution is to start treating data as an asset. Data-centricity - the idea that data is the primary asset of enterprises, not its single applications[4] - is the strategy banks, financial institutions, and large enterprises all need to adopt as their governing strategy if they are to accelerate innovation and drive costs down at the same time. Developing synthetic data capabilities is the fulcrum necessary to provide access and representation of data across large organizations, especially in banks, where legacy architectures make it immensely challenging to free information from data silos and allow data-centricity to flourish.
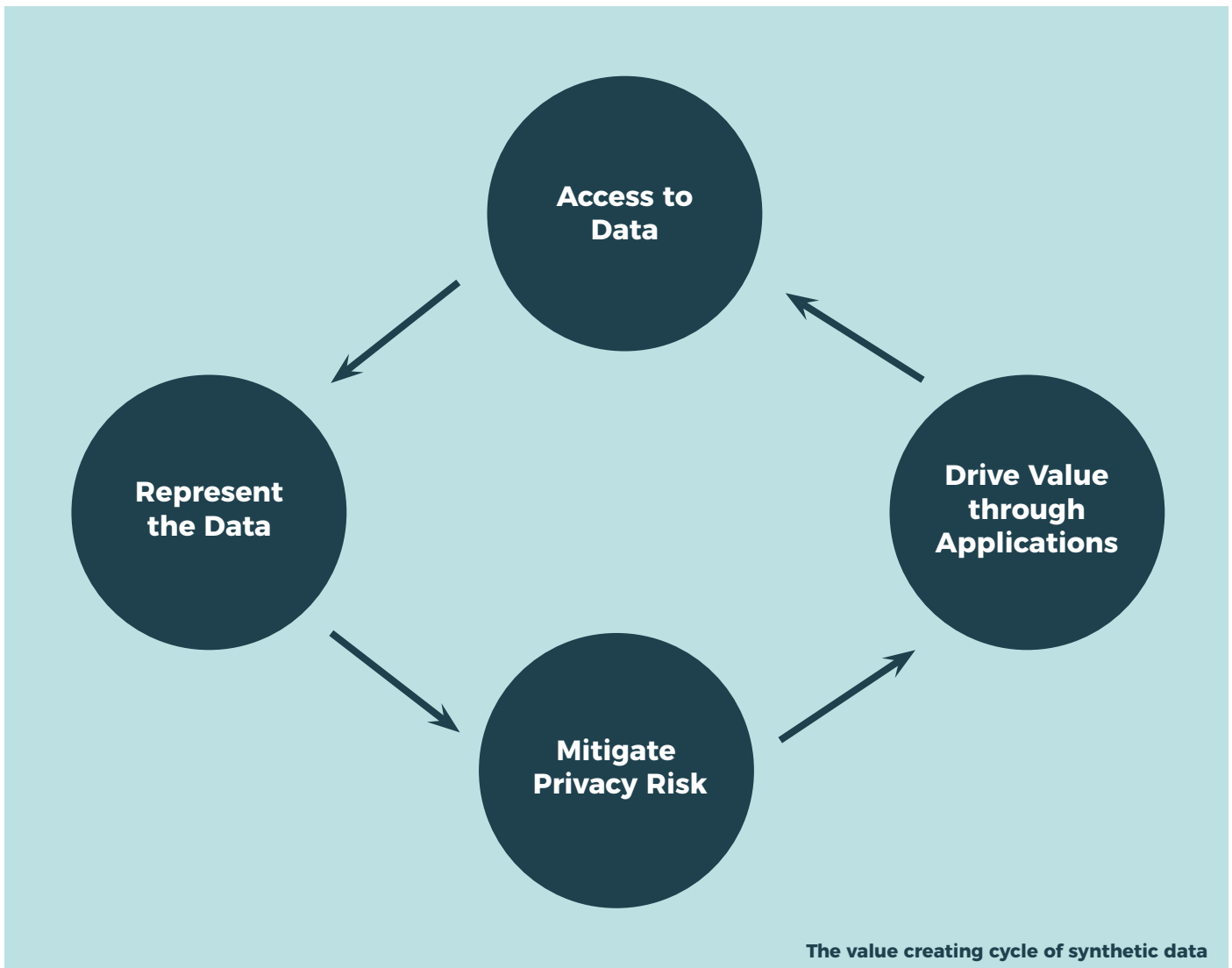
2021

What we can see in 2021 and beyond are trends all pointing in the same direction: get data right, and you will emerge as a winner. Gartner suggests including synthetic data in your overall data strategy and exploring the technology with tabular data in 2021.[5] Synthetic data will be the key technology in tackling data management challenges across domains of privacy, predictive analytics, security, and overall data-centricity.

This book will give you an overview of what AI-generated synthetic data is and we'll guide you through the most important use cases you need to understand to leverage synthetic data to maximize your organization's data opportunities.

Once you learn what AI-powered synthetic data is capable of, we suggest conducting a data audit with synthetic capabilities in mind. Map your organization's data flow: identify bottlenecks and hard stops where synthetic data could readily solve data access issues.

At the end of the book, you will find a checklist that will help you make an informed decision in your synthetic data vendor selection, as well as a page you can immediately forward to your data scientists with the most important answers to their most likely questions. You can also get a personal consultation from one of our synthetic data experts - find contact details on the very last page of this book.

Access to Data

Represent the Data

Drive Value through Applications

Mitigate Privacy Risk

The value creating cycle of synthetic data

Data synthesization is fast becoming a key component of data management in large organizations. Simply put, the technology enables its users to extract meaning out of datasets while leaving potential data exposures behind.

So what is synthetic data technology today? It's not what it was five years ago. Back then, the expression was used to describe randomly produced dummy data. That is no longer the case.

AI-powered synthetic data generating algorithms of today digest real data, learn its features, correlations, and patterns in great detail and are then able to generate infinite amounts of completely artificial, synthetic data matching the

statistical qualities of the originally ingested dataset. The new, synthetic datasets are scalable, privacy-compliant, and contain all of the original meaning without the burden of sensitive information.

Leveraging synthetic data in your data management processes not only mitigates privacy risks, drives costs down, accelerates transformation and innovation, but ultimately increases the value of your data assets in a self-reinforcing fashion.

In the following chapter, we will walk you through the most important synthetic data management best practices you need to use to maximize your data opportunities.

A step-by-step guide to leveraging synthetic data in banking

1 Democratize data access to increase data literacy

STEPS TO TAKE

To create a truly data-centric organization, a data-driven culture must be cultivated from the top down. Providing constant access to data in every corner of your organization will be the single most important ingredient of your success. This is no small feat in institutions traditionally built on values of secrecy embodied as air-gapped data silos. Thanks to synthetic generators, you will be able to provide high-touch access to decision-ready synthetic datasets not only to your core data science team but also to citizen data scientists, who will be able to extend the analytics universe of your organization. By setting up self-service synthetic data lakes, you can reduce wait times, administrative overhead, and most importantly, introduce a continuous, iterative, data-driven attitude across departments and operating companies.

REAL-LIFE EXAMPLE

One of the top 10 banks in North America created a twin data lake, enabling any employee to point to any data source and get its synthetic version with minimal effort and no privacy risk. The initiative increased overall data literacy, and citizen data scientists were deriving data-driven insights from the synthetic data lakes on a daily basis.
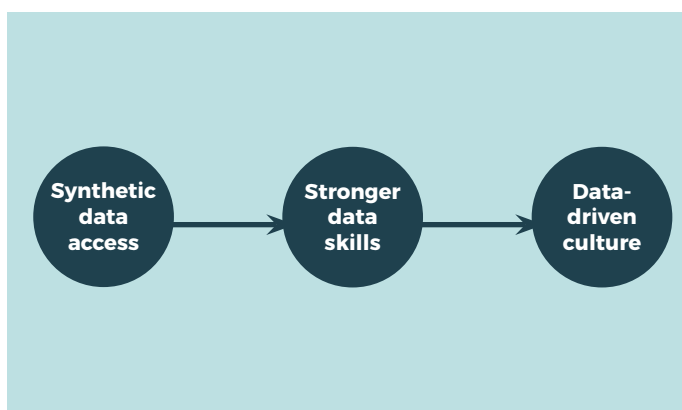

Culture-creation with synthetic data

2 Streamline privacy compliance

STEPS TO TAKE

According to Gartner: "Privacy is becoming a discipline of its own. No longer "just a part of" compliance, legal or auditing, privacy is becoming an increasingly influential, defined discipline of its own, affecting almost all aspects of an organization. As a rapidly growing stand-alone discipline, privacy needs to be more integrated throughout the organization."[6] By making fresh, large batches of synthetic data flow freely through your organization, you can minimize your exposure to privacy risks and ensure compliance, because synthetic data is privacy-compliant by its very nature. Even when cooperating with government agencies and legislators, providing synthetic data in compliance processes significantly reduces costs and administrative overhead. Synthetic data is private by design. How? Generated synthetic datasets bear no direct relationship to the original datapoints; they are completely artificial. These synthetic versions retain statistical meaning but no longer classify as personal information. Synthetic data is GDPR and HIPPA-compliant, free to use, share, and hold. CISOs should plan accordingly: cross-border synthetic data-sharing is still possible after Schrems II and will remain so due to the intrinsic, private nature of synthetic data. Instead of restricting data access in-house or across different subsidiaries, provide readily available synthetic alternatives.

REAL-LIFE EXAMPLE

Due to the legal constraints of such projects, global HR analytics is only possible when leveraging synthetic data capabilities. A large multinational enterprise conducted an HR analysis of more than 90 000 employees using synthetic data. Due to legal regulations, operating companies couldn't touch employees' sensitive, raw data. Using the synthetic version of the data, they could identify patterns leading to employee churn, optimize HR processes, and improve talent acquisition and retention rates.

Expected
0.5%
reduction of turnover rates

Expected
double digit million
reduction of turnover rates

increased team productivity

Data retention is now heavily regulated in Europe, the Americas, and many other regions by GDPR, CCPA, and many other overreaching laws. National laws often act as additional layers of regulation, especially with respect to telecommunications or banking data, limiting storage periods. However, certain high-value insights can only be accessed if we see the historic big picture. Annual seasonality analyses, for instance, need at least two years of data to provide meaningful results. Synthesizing datasets before deleting them would retain exactly the kind of statistical information such analysis would require. What's more, during synthetization, imbalances - stemming from, for example, unusual customer behavior during pandemics - and human biases embedded in the data can be addressed by augmenting the original dataset using the same synthetic generator.
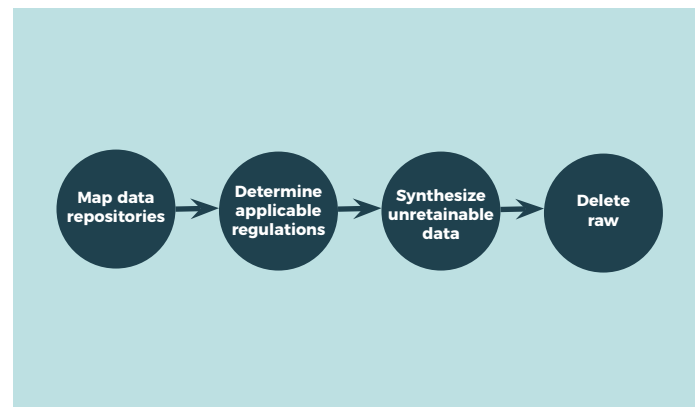
Data retention strategy powered by synthetic data

3 Minimize data retention

STEPS TO TAKE

The idea of data minimization originally stems from data regulations,[7] but the principle is useful beyond mere compliance with legislation. Effective data asset management maps and prepares quality data while also recognizes the inherent dangers of retaining statistically useless sensitive information. Synthetic data provides a simple alternative to retaining sensitive datasets: synthesize before you delete. This way, you can retain answers before you even know what your questions would be. Preprocessing data to create decision-ready datasets is necessary anyway - why not go the extra mile and synthesize to increase quality by fixing embedded biases and decreasing privacy risks by eliminating sensitive information.

4 Collaborate with external vendors using synthetic data lakes

STEPS TO TAKE

Finance has traditionally relied on third-party expertise, and in some respects, this is likely to stay so. Organizations that are new to analytics, and data science specifically, or with not enough in-house talent, should look for third party AI and analytics vendors. "Enterprises in this situation should first look to providers with not just predictive and prescriptive proficiency, but that also possess knowledge of their particular industry."[8] Synthetic data is a vital ingredient of agile, cost-effective, and ultimately successful PoC processes.

REAL-LIFE EXAMPLE

A Fortune 100 bank needs to evaluate 1,000+ vendors and start-ups annually. In 80% of these evaluations, the process involves handing over sensitive datasets to external organizations. This step takes 3.5 months per PoC since the data needs to be manually selected, sanitized, anonymized, and individually approved in each specific case. Due to this labor-intensive procedure, $25,000 of internal costs are generated per PoC, resulting in $25 million of annual costs for external data sharing.

The bank created a Rapid PoC Sandbox where the most commonly requested data assets were proactively converted to synthetic versions. Using this newly created sandbox, vendors can now test their solutions in a controlled, privacy-safe environment. The privacy-compliant synthetic data generated by the bank provides statistically identical results to those derived from production environments and complies with all internal and external regulations.

As a result, the average time of data delivery decreased to 3 weeks, a 70% reduction. The process now only involves standardized checks and onboarding procedures, saving further costs. According to the bank's estimation, the average cost of a PoC is now only $5,000 – an 80% reduction from the average price of previous PoCs. The annual savings impact of this initiative amounts to over $10 million.

70%
time reduction for data delivery

80%
cost reduction for data provisioning

£10M+
annual savings on PoCs

5 Cooperate with academia and cross-industry players

STEPS TO TAKE

Innovation efforts are often highly dependent on research cooperations with academia and cross-industry collaborations with other players in the financial space. Hackathons are a great way to test technologies, research-relationships, and product ideas, but only if you can provide participants with the right data environment. Populating sandboxes with synthetic data will not only provide the right level of security and privacy-compliance, but realistic datasets will also make sure that your teams will arrive at the right conclusions.

REAL-LIFE EXAMPLE

Swift provided synthetic data and APIs for a large hackathon in a cross-industry collaboration effort, involving 70 participants from banks, fintech, and tech companies to identify innovative ways to address pressing industry challenges. The synthetic data allowed very practical solutions to be developed using advanced analytics tooling.[9]

In their effort to develop successful research-relationships, a large financial service provider in North America organize hackathons with

university research departments on a quarterly basis. These events help them set a vision and build prototypes around disruptive technologies and academic research topics.

After attempting to run a hackathon with anonymized datasets, they concluded that obfuscating personally identifiable information destroyed the richness of the data, impacting participants' ability to build effective prototypes. The organizers created synthetic data sandboxes for the teams to work with – Virtual Machines on Google Cloud Platform that come preloaded with their developer tools and synthetic datasets.

After an intensive 48 hours of working with these datasets, they learned about behavioral patterns that they ordinarily would not find in public or anonymized datasets. One of the teams created a proof-of-concept for a financial assistant that helps customers better plan their spending and achieve their financial goals. Their solution suggests to customers where they can save money while not deviating too much from their peers.

Other highlights were an app that helps freelancers save for their pensions and a financial planning assistant for patients with chronic illnesses. The accuracy and realism of synthetic data provided the hackathon participants with deep insights into individual earning and spending patterns.

They could demonstrate their ideas using concrete scenarios and help focus the company's vision on financial assistants and personalization. Additionally, several participants were successfully recruited by the bank, and a 7-year research contract was signed by the bank and the university.

6 Develop customer-centric products

STEPS TO TAKE

The promise of digital banking is yet to be fulfilled. There is no better time for developing truly customer-centric products and services than now. The pandemic accelerated digital product usage across all demographics and what had

previously been discretionary or aspirational digital transformation became imperatives for survival. According to Deloitte, a staggering 44% of retail banking customers said they are using their primary bank's mobile app more often.[10] McKinsey estimates that preference for handling everyday transactions digitally is as high as about 60 to 85 percent across Western European markets, even for customers 65 years of age or older.[11]

It is no surprise that banks and financial service providers are finding themselves in a turbocharged competition, where customer experience makes or breaks a product more than ever. It is crucial to develop, test, and improve products based on real insights, using realistic, rich data throughout the entire development cycle.

REAL-LIFE EXAMPLE

One of the largest retail banks in Europe sought to improve their mobile app to meet the modern-day expectations of their customers. Their goal was to make development more efficient and speed up UX design decisions.

Bank policies and privacy regulations prevented the team from accessing customer data. They were left to populate their product development systems with dummy data and even their own personal banking data. These datasets, however, came nowhere near the scale or complexity that they required.

They transformed sensitive customer transaction datasets into realistic synthetic copies to accurately reflect customers' statistical features – while also meeting all privacy compliance regulations the bank required. The product came to life when the team fueled their development environment with AI-powered synthetic data, providing them with actionable insights into individual customers' behavior patterns.

The department responsible for data security and management fulfilled the production team's data requests in hours, where legacy anonymization processes would take months. Using highly realistic synthetic data empowered developers to create a product that genuinely met customer needs and expectations. They delivered what quickly became the number one banking app in the target country – with an average rating of 4.6 stars across app stores and countless reviews praising its seamless user experience.

Accomplished goals

Smart, data-driven product features, such as account balance prediction and responsive UX decision-making.

Shortened development sprints by several days as a result of near-instant availability of synthetic customer transaction data.

Demoed a highly realistic product to internal stakeholders, earning valuable support throughout the organization.

Shared granular yet fully anonymous synthetic data with several external partners to develop further products and services, such as a bill splitting app, born out of identifying genuine customer needs.

Mitigate costly data breaches

7

STEPS TO TAKE

The average cost of a data breach in 2020 was $3.86 million.[12] As the recent SolarWind hack proves, no organization is immune. What's more, 59% of privacy incidents originate with an organization's own employees. Worse still — 45% of employee-driven privacy failures come from intentional behavior (though it may not be malicious).[13] Exposure is getting more and more likely and dangerous as the amount of data collected and held is rising. At the same time, demand for decision-ready data is growing across all paradigms, and denying access is denying life support from innovations and data-centricity. Providing synthetic data alternatives across the organization, even in the cloud, is the tool you need to minimize sensitive data exposure while fuelling data-hungry applications and processes. Each access request granted to synthetic data reserves makes your organization more resilient to attacks and, at the same time, better equipped to scale innovations.

REAL-LIFE EXAMPLE

A global insurance provider wanted to access and analyze data across borders. For reasons of security and privacy compliance, sensitive customer data could not be shared across subsidiaries. If the data couldn't come to the data science department at the company's headquarters, the data science went to the data. Literally. Teams of analysts would fly to 5-10 countries per year to analyze data locally. After the pandemic hit, this way of operating became impossible. Just like with so many of the digital transformations accelerated in 2020, this issue found a new, future-proof solution: synthetic data. Instead of flying people to sensitive datasets, synthetic copies of the data can be readily and cost-effectively shared across borders and subsidiaries without exposing sensitive data to leaks and breaches.

ACCOMPLISHED GOALS

- Pandemic-proof and cost-effective access to data
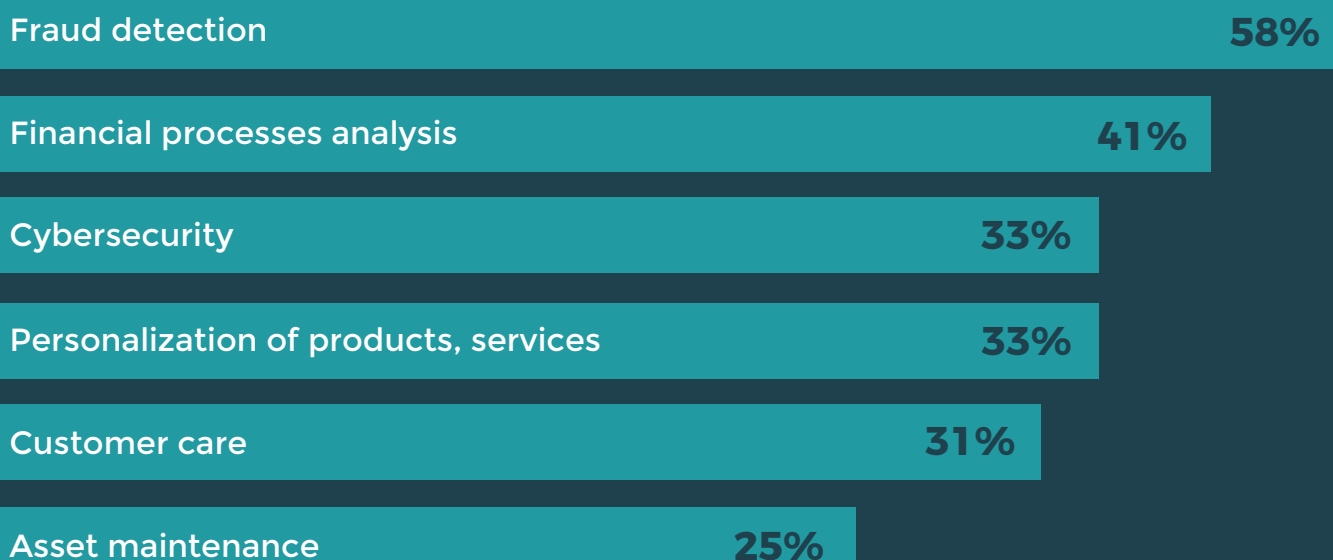- Privacy-compliant cross-border data-sharing

8 Make AI happen

According to Gartner, AI will optimize or transform nearly every activity in finance in the coming decade. CFOs should educate themselves on how the function will change, prepare teams with new skill sets, and explore investments necessary to deploy AI.[14] While Bank of America is one of the big players who already have a robust AI plan in place,[15] 74% of C-suite banking executives surveyed by Accenture believe if they don't effectively scale AI in the near term, they'll go out of business within five years.[16]

Your AI is only as good as the data used to train the algorithm. Therefore, step zero should always be data provisioning. Whether you use in-house modeling teams or external AI-vendors, quality data in large batches is hard to come by. The data needs to be augmented to allow machine learning models to pick up on rare patterns, while human bias embedded in datasets must also be addressed to avoid serious pitfalls.[17] Some synthetic data generators solve both issues by enabling balance correction and upsampling. Recalibrating your AI algorithms for higher accuracy and training new ones is where synthetic data capabilities really shine.

REAL-LIFE EXAMPLE

The publicly available credit card fraud detection set[18] contains very few fraud cases: only 0.17% of the transactions are fraudulent. Conditional synthetic data generation can increase the number of fraud samples, allowing machine learning models to learn more efficiently. Sufficient upsampling can increase machine learning accuracy by as much as 1-2 percentage points, leading to better performance in detecting true positives and reducing the number of false positives. A US payment systems company estimates that a mere 1% reduction in fraud results in $1 million savings per month.[19]

AI use cases in finance, MIT Tech Review 2020.

Fraud detection	58%
Financial processes analysis	41%
Cybersecurity	33%
Personalization of products, services	33%
Customer care	31%
Asset maintenance	25%

How to start your synthetic data journey?

1. START WITH TABULAR DATA:

Tabular data has always been the single most important format for financial institutions. Gartner recommends starting your synthetic data exploration by synthesizing tabular data.[20] Identify valuable tabular assets, synthesize, and publish. It's fast, easy, and rewarding.

2. FIND THE RIGHT SYNTHETIC DATA VENDOR

Although open source synthetic data generators are available, they come with serious limitations regarding accuracy and privacy guarantees. Their performance can be volatile and is highly dependent on the community behind it. Closed source offers more sophisticated capabilities and commercial services you can count on.

Choose a vendor with in-depth experience in the financial industry, capable of augmenting as well as synthesizing data. Demand automated privacy and accuracy quality assurance. Use third party research from trusted sources, such as Gartner and Forrester, to identify viable and robust players.

3. SET UP A SYNTHETIC DATA EXCELLENCE CENTRE

Managing data access requests takes up the majority of CDOs' time and resources. By setting up a Synthetic Data Excellence Centre, you can provide a quick, painless, compliant, and fully audited process to request synthetic versions of data.

4. CREATE SYNTHETIC DATA LAKES

Set up synthetic data lakes to mirror your most valuable and insightful data assets. Colleagues across your organization can use it as a self-service data center to access decision-ready data.

By making synthetic data flow freely throughout your organization, true data-centricity is born: data-driven decision making and data literacy increases and works in a self-reinforcing fashion.

5. REVIEW CURRENT PRACTICES OF DATA ANONYMIZATION

What was sufficient to protect data a few years ago no longer suffice. Classic anonymization techniques, like randomization, pseudonymization, generalization, or permutation do not protect against linkage attacks.[21]

90% of a banks' data can be utilized in its synthetic form and the remaining 10% sensitive PII data can remain heavily protected, minimizing risk in the event of a data leak.[22] Educate citizen data scientists about data anonymization risks and synthesize wherever possible.

Notes for your data science team

The most important questions and answers about synthetic data for the initiated

What data types can you synthesize?

MOSTLY GENERATE can synthesize numerical, categorical, datetime, short text (ex. transaction text), and geographic data. All data must be provided in a tabular format.

Is time series data supported?

Yes. Time series data is modeled in a two-table setup. The first table, called the subject table, contains unique identifiers. The second table, called the linked table, contains events belonging to a unique identifier. For example, user accounts and their transactions.

How is privacy guaranteed?

Privacy is built into the generation process in multiple ways. The model uses a random generative process to avoid direct duplicates in the synthetic data. Outlier handling protects column-wise privacy by ensuring that unique values don't occur in the synthetic data.

How do you guarantee that outliers don't persist in the synthetic data, potentially leaking sensitive information?

Outliers are handled in two different ways, depending on the data type. For numerical data, any values between the 10th and 90th percentile are clipped away. For categorical data, values appearing more than n times are replaced based on a sliding scale.

How are the privacy and accuracy metrics defined?

After each synthetic generation is complete, a custom QA report is generated. We have open-sourced our metrics in a Python library.[23]

Can synthetic data preserve referential integrity?

Yes. Generation is done by first generating identifiers, then their associated attributes. This way, keys referenced in the subsequent tables are guaranteed to exist.

What is the quality of MOSTLY GENERATE's synthetic data for AI/ML model training?

Always highly accurate. Small fluctuations in the accuracy depend on how much data the model was trained with, how long the model is trained for, and how complex the model becomes. Overall, synthetic data can capture 80%-99% underlying patterns of the original data. We have done extensive research [24][25] covering the use of synthetic data in ML training. The results are consistently on par or better than training with real data.

Can synthetic data capture patterns of fraudulent transactions?

Fraudulent transactions often suffer from two related problems; lack of fraudulent cases and inability to correctly identify fraud. Synthetic data can capture the complex relationships within the data. It can retain the patterns associated with fraud, even if the number of examples is minimal.

Can synthetic data reproduce business rules?

MOSTLY GENERATE can reproduce business rules implicitly. Due to the random nature of the generative process, small violations may occur. The ability to incorporate hard business rules is an upcoming feature, to be released in 2021.

Is there a UI?

MOSTLY GENERATE has an intuitive interface, with drag and drop functionality, interactive runtime graphs, and the ability to queue multiple runs. Once a model is created, it is possible to generate more sets of synthetic data without having to wait through training time again.

Is there a non-UI feature that allows automated data pipelines in production?

MOSTLY GENERATE has an API feature and the ability to read from network drives. These two features allow easy integration for automated data pipelines in production.

- API
- Read data from network drives

Is the application a cloud solution or on-premise installation? Does MOSTLY GENERATE interact with the external internet?

MOSTLY GENERATE is available for both an on-cloud and an on-premise installation.

No internet connection is required to use MOSTLY GENERATE.

References: click on numbers for link

1 Why Now Is the Time to Accelerate Digital, Gartner, September 10, 2020
https://www.gartner.com/smarterwithgartner/why-now-is-the-time-to-accelerate-digital/

2 A test of resilience: Banking through the crisis, and beyond, McKinsey Global Banking Annual Review 2020
https://www.mckinsey.com/industries/financial-services/our-insights/global-banking-annual-review

3 2021 banking and capital markets outlook, Deloitte
https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-outlooks/banking-industry-outlook.html

4 The Data-Centric Revolution: Restoring Sanity to Enterprise Information Systems,
Dave McComb https://www.amazon.com/Data-Centric-Revolution-Restoring-Enterprise-Information/dp/1634625404

5 Predicts 2021: Data and Analytics to Govern, Scale and Transform Digital Business, 2020. December 2.
https://www.gartner.com/en/documents/3993855/predicts-2021-data-and-analytics-strategies-to-govern-sc

6 Gartner Top 9 Security and Risk Trends for 2020, Gartner
https://www.gartner.com/smarterwithgartner/gartner-top-9-security-and-risk-trends-for-2020/

7 REGULATION (EC) No 45/2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 18 December 2000
https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:008:0001:0022:en:PDF

8 Add Data Science Skills to Corporate Finance Decisions, Gartner, January 15, 2020
https://www.gartner.com/smarterwithgartner/add-data-science-skills-to-corporate-finance-decisions/

9 Swift Community Update: Payments transformation in the UK
https://www.finextra.com/newsarticle/37146/swift-community-update-payments-transformation-in-the-uk

10 2021 banking and capital markets outlook, Deloitte
https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-outlooks/banking-industry-outlook.html#endnote-18

11 Reshaping retail banking for the next normal, McKinsey
https://www.mckinsey.com/industries/financial-services/our-insights/reshaping-retail-banking-for-the-next-normal

12 Cost of a Data Breach Report 2020, IBM
https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/pdf

13 The Call for Legal and Compliance to Minimize Data Privacy Risk, Gartner
https://www.gartner.com/smarterwithgartner/call-legal-compliance-minimize-data-privacy-risk/

14 Gartner Identifies 10 Trends CFOs Must Understand to Prepare for the Future of Finance, Gartner
https://www.gartner.com/en/newsroom/press-releases/2020-02-27-gartner-identifies-ten-trends-cfos-must-understand-to-prepare-for-the-future-of-finance

15 How Bank of America uses AI to improve customer service and take on competitors
https://www.itpro.com/technology/artificial-intelligence-ai/354510/how-bank-of-america-uses-ai-to-improve-customer

16 AI: Built to Scale, Accenture
https://www.accenture.com/_acnmedia/Thought-Leadership-Assets/PDF-2/Accenture-Built-to-Scale-PDF-Report.pdf

17 Apple Card algorithm sparks gender bias allegations against Goldman Sachs, Washington Post
https://www.washingtonpost.com/business/2019/11/11/apple-card-algorithm-sparks-gender-bias-allegations-against-goldman-sachs/

18 Credit card fraud detection dataset
https://www.kaggle.com/mlg-ulb/creditcardfraud

19 H2O for Real Time Fraud Detection
https://www.h2o.ai/wp-content/uploads/2019/02/H2O-For-Fraud-Analytics.pdf

20 Predicts 2021: Data and Analytics to Govern, Scale and Transform Digital Business, Gartner
https://www.gartner.com/en/documents/3993855/predicts-2021-data-and-analytics-strategies-to-govern-sc

21 Semantic Web Enabled Record Linkage Attacks on Anonymized Data, Jacob Miracle and Michelle Cheatham
http://ceur-ws.org/Vol-1750/paper-03.pdf

22 Cracking FIs' 90/10 Data Monetization Problem, PYMNTS
https://www.pymnts.com/news/security-and-risk/2019/synthetic-data-financial-services-arm-insight/

23 Virtual Data Lab,
https://github.com/mostly-ai/virtualdatalab

24 The World's Most Accurate Synthetic Data Platform? Let's check the Numbers!
https://mostly.ai/2020/09/25/the-worlds-most-accurate-synthetic-data-platform/

25 Boost your Machine Learning Accuracy with Synthetic Data
https://mostly.ai/2020/08/07/boost-machine-learning-accuracy-with-synthetic-data/

MOSTLY•AI

About MOSTLY AI

MOSTLY AI is a mature, enterprise-ready synthetic data vendor, specializing in behavioral tabular data. The team has been working in the financial sector for years. Headquartered in Vienna, Austria, with an office in New York, USA and working with a fully distributed team, the company serves clients all over the globe. MOSTLY AI's synthetic data generating AI algorithm is the world's most accurate, offering automatically generated privacy and accuracy reports with each generated dataset. MOSTLY AI's synthetic data generator is available on-premise, guaranteeing sensitive data never leaves your organization's secure environments. Ask us for a free consultation to discover your synthetic data opportunities!

Talk to one of our experts

Contact

✉ hello@mostly.ai

in linkedin.com/company/mostlyai

🐦 @mostly_ai

Vienna Office (HQ)
MOSTLY AI Solutions MP GmbH
Hegelgasse 21/3
1010 Vienna, Austria

New York Office
MOSTLY AI Inc.
500 7th Ave 8th floor
New York, NY 10018, United States