



REVOLUTIONIZING DATA DRIVEN INNOVATION: SYNTHETIC DATA

THE BENEFITS AND CHALLENGES OF
SYNTHETIC DATA

Cecilia Dones :: October 2024

MY VIEWS ARE MY OWN AND DO NOT REPRESENT ANY OTHER AFFILIATED ORGANIZATIONS.

BOTTOM LINE UP FRONT

EXPLORE SYNTHETIC DATA TO DRIVE
INNOVATION WHILE CAREFULLY
BALANCING PRIVACY, COMPLIANCE,
AND DATA INTEGRITY.

WHAT YOU WILL TAKE WITH YOU TODAY

Definition of
Synthetic
Data

Pros/Cons +
Trade-Offs

Industry
Applications

CECILIA DONES :: SHORT INTRODUCTION

Founder, 3 Standard Deviations

15+ Years in Industry, Researcher, Practitioner, Educator, Academic

Specialties: Qualitative, Quantitative Methods, Marketing, Interdisciplinary, Pragmatic Solutions For Businesses
Trying To Leverage Data & AI

L'ORÉAL



UBS

 **Columbia
Business
School**

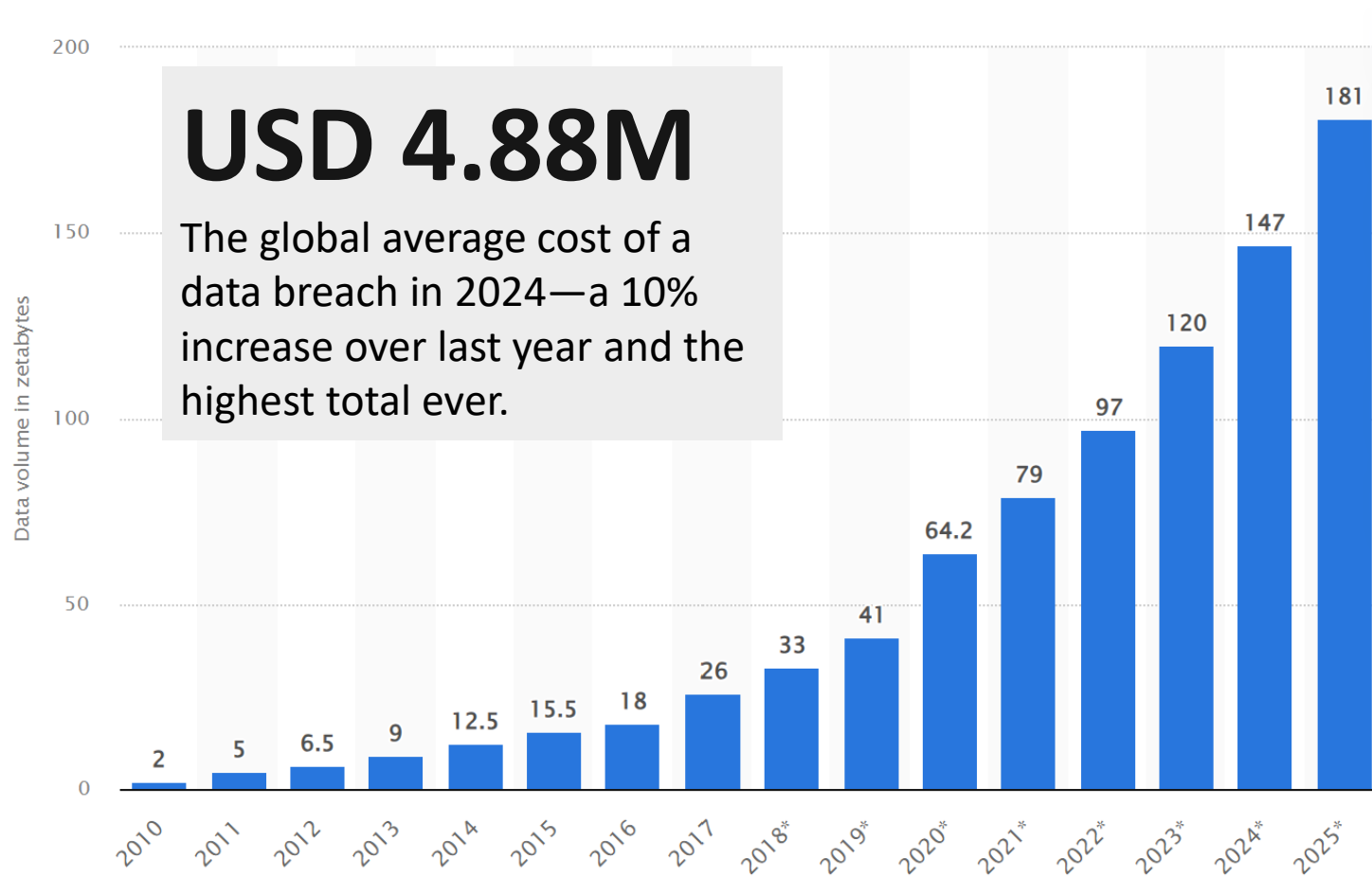
verizon

MIT **MANAGEMENT**
SLOAN SCHOOL

LVMH
MOËT HENNESSY . LOUIS VUITTON



DATA IS COMPLEX AND CONSEQUENTIAL



CHALLENGES

- Increasing demand for insights
- Stricter privacy regulations
- Limitations of traditional anonymization



Fully Synthetic Data

Data created entirely by algorithms, without direct one-to-one mapping to real-world data points.

Example: Generating a complete set of fictional customer profiles for a non-existent e-commerce platform, including demographics, purchase history, and browsing behavior.

Partially Synthetic Data

Real dataset where some variables are replaced with synthetic values.

Example: A hospital dataset where patient names and addresses are replaced with synthetic ones, but medical conditions and treatments remain real.

Hybrid Synthetic Data

A mix of real and synthetic data points, where synthetic data is used to augment or expand real data.

Example: A financial dataset that contains real transaction data for existing products, combined with synthetic data for hypothetical new products or market scenarios.

Simulated Data

Data generated from a model designed to replicate specific real-world processes or systems.

Example: Climate model outputs that simulate future weather patterns based on current climate data and physical models.

TRADEOFFS IN REAL VS. SYNTHETIC DATA



	Real Data	Synthetic Data
Privacy Risk	High	Low
Data Availability	Limited	Abundant
Compliance Complexity	High	Low
AI/ML Model Training Efficiency	Constrained by data limits	Enhanced by data generation
Data Authenticity	High	Variable
Anomaly Representation	Naturally occurring	May be underrepresented
Unforeseen Insights	Possible	Limited
Real-world Accuracy	High	Dependent on generation quality

STRATEGIC IMPLICATIONS FOR BUSINESS

- Accelerating Innovation Cycles
- Enabling Cross-Border Data Sharing
- Improving Data Democratization
- Enhancing AI/ML Development

IMPLEMENTATION CHALLENGES AND SOLUTIONS

Challenge	Solution	Key Considerations
Data quality concerns	Advanced validation techniques	<ul style="list-style-type: none">- Implement statistical validation methods- Use machine learning for anomaly detection- Establish data quality metrics and thresholds- Conduct regular data audits
Infrastructure integration	Phased approach and APIs	<ul style="list-style-type: none">- Start with non-critical systems- Develop robust API architecture- Ensure scalability and compatibility- Implement strong security measures
ROI measurement	Defined KPIs and benchmarking	<ul style="list-style-type: none">- Establish clear baseline metrics- Define both short-term and long-term KPIs- Conduct regular performance reviews- Use A/B testing for comparative analysis
Organizational resistance	Education and pilot programs	<ul style="list-style-type: none">- Develop comprehensive training programs- Start with small-scale pilot projects- Showcase early wins and success stories- Foster a data-driven culture

INDUSTRY APPLICATIONS



Financial Services

Fraud detection models

AI-powered anomaly detection in transactions



Healthcare

Clinical trial simulations

Drug development through virtual patient populations



Retail

Customer Behavior Models

Predicting trends and personalizing experiences

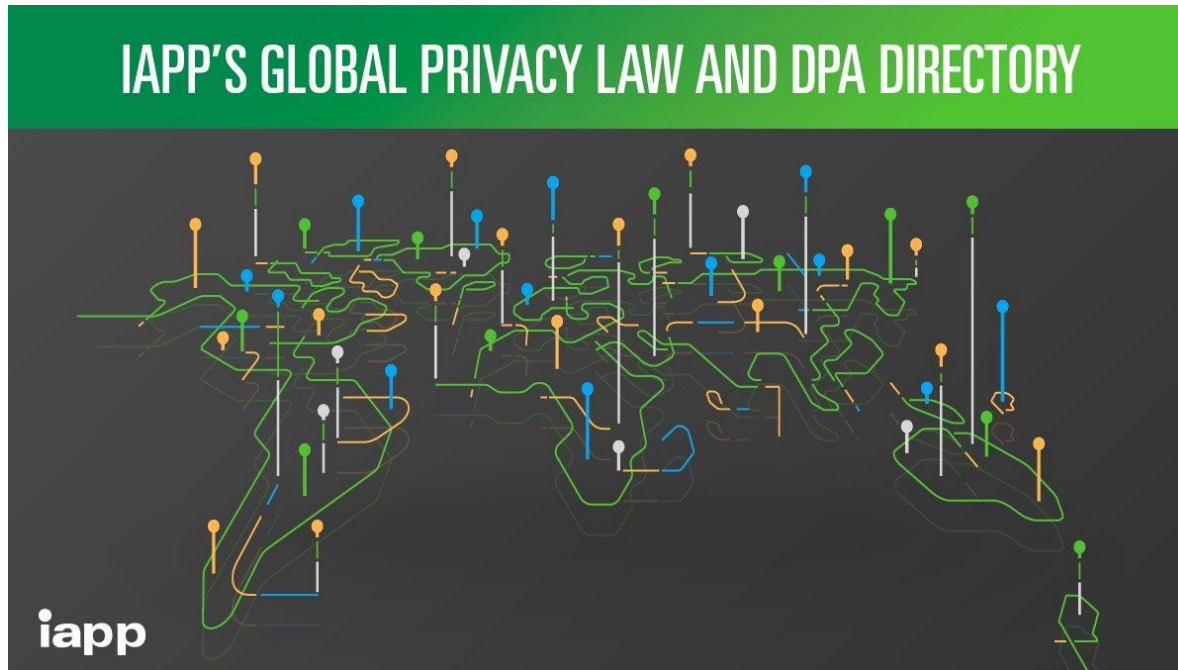


Manufacturing

Supply Chain Optimization

AI-driven efficiency in production and logistics

PRIVACY REGULATION & TRUSTWORTHY AI



United Kingdom

Legislation & regulation: Online Safety Bill (2022, draft); Data Protection and Digital Information Bill (2023, draft)
Standards: Algorithmic Transparency Standard (Central Digital Data Office, 2021)
Principles: A pro-innovation approach to AI regulation (2023)

Canada

Legislation & regulation: Directive on Automated Decision-Making (2019) Bill C-27: Digital Charter Implementation Act, including AI and Data Act (AIDA) (2022, draft)
Standards: CAN-ASC-6.2: Accessible and Equitable AI Systems (2023, draft)
Principles: Canada's Digital Charter (2019)
Oversight: Minister; Proposed AI and Data Commissioner

European Union

Legislation & regulation: Proposed EU AI Act (2021, draft); Updates to the EU Product Liability Directive (2022, draft); AI Liability Directive (2022, draft); EU's Digital Services Act
Standards: CEN/CENELEC standards for AI and related data (forthcoming)
Principles: Ethics guidelines on AI (2018)
Oversight: Proposed European Artificial Intelligence Board

United States

Legislation & regulation: Federal Trade Commission Act, for deceptive practices from deepfakes or chatbots (1914); Algorithmic Accountability Act (US AAA) (2022, draft)
Standards: NIST AI Risk Management Framework (2023)
Principles: Blueprint for an AI Bill of Rights (2023)

China

Legislation & regulation: Chinese Internet Information Service Algorithmic Recommendation Provisions (2021); Opinion on Strengthening the Ethics and Governance of Science and Technology (2022)
Standards: National Standards for Autonomous Vehicle Testing (2018)
Principles: New Generation AI Ethics Specifications (2019); New Generation AI Code of Ethics (2021); White Paper on Trustworthy AI (2021); Internet Information Service Algorithmic Recommendation Management Provisions (2021)

Brazil

Legislation & regulation: Report and proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021 (2022, draft); Bill 705 on the compatibility of AI use in the public sector with ESG practices (2022, draft)
Standards: Incorporation of international standards National standards by the Brazilian Association of Technical Norms (ABNT)
Principles: Art. 3 of the proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021 (2022, draft)



Intergovernmental Organisations

Legislation & regulation: Council of Europe Convention on AI, Human Rights, Democracy and the Rule of Law (2023, draft)
Standards: ISO 31000 Risk management (2009, 2018); ISO/IEC 23053:2022 Framework for AI Systems Using Machine Learning (ML) (2022)
Principles: OECD Recommendation of the Council on AI (2019); UNESCO Recommendation on the Ethics of AI (2021)

LEARN MORE: International Association of Privacy Professionals. (2024). Global privacy law and DPA directory. <https://iapp.org/resources/global-privacy-directory/>; Organisation for Economic Co-operation and Development. (2023). National policies. <https://oecd.ai/en/work/national-policies-2>

THE FUTURE OF SYNTHETIC DATA

Integration with Federated Learning

- Decentralized AI training with privacy-preserving synthetic data

Advancements in Generative AI for Data Creation

- Next-gen AI creating hyper-realistic synthetic datasets

Synthetic Data Marketplaces

- Democratizing access to high-quality, diverse datasets



WHAT YOU CAN DO TODAY

Evaluate potential in your organization

- Identify high-impact use cases for synthetic data

Start with a pilot project

- Test the waters with a low-risk, high-reward initiative

Collaborate with synthetic data experts

- Leverage expertise to accelerate your journey

CAN OR FOUNTAIN?



WHAT YOU WILL TAKE WITH YOU TODAY

Definition of
Synthetic
Data

Pros/Cons +
Trade-Offs

Industry
Applications

EXPLORE SYNTHETIC DATA TO
DRIVE INNOVATION WHILE
CAREFULLY BALANCING
PRIVACY, COMPLIANCE, AND
DATA INTEGRITY.



THANK YOU

IF YOU WOULD LIKE TO COLLABORATE WITH ME, PLEASE
FEEL FREE TO FIND ME ON LINKEDIN OR EMAIL ME:

CECILIA.DONES@3STANDARDDEVIATIONS.COM