

Technology, Media & Telecommunications

PDPC Launches Proposed Guide to Synthetic Data Generation

Introduction

The Personal Data Protection Commission ("**PDPC**") has launched a Proposed Guide on Synthetic Data Generation ("**Guide**") to help organisations understand Synthetic Data ("**SD**") generation techniques and potential use cases. The Guide will be offered as a resource within the Privacy Enhancing Technology ("**PET**") Sandbox.

PETs are technologies that allow the processing, analysis, and extraction of insights from data without revealing the underlying personal or commercially sensitive data. SD generation is an increasingly utilised PET, in which artificial data is created to drive the growth of Artificial Intelligence ("**AI**") / Machine Learning by enabling AI model training, while protecting the underlying personal data.

However, while SD is generally fictitious data that may not be considered personal data on its own, it still carries the risk of possible re-identification. The Guide thus seeks to address these risks by proposing good practices in the generation of SD. It also includes risk assessments/considerations, governance controls, contractual processes, and technical measures.

This Update highlights the key aspects of the Guide, including a summary of the key considerations and best practices in SD generation.

SD Generation

What is SD? SD is artificial data that has been generated using a purpose-built mathematical model (including AI/ML models) or algorithm. It can be derived by training a model or algorithm to mimic the characteristics and structure of the source data.

What can SD be used for? SD can be used in a variety of ways, including generating training datasets for AI models, data analysis and collaboration, and software testing. It can facilitate the acceleration of research, innovation, collaboration, and decision-making while mitigating concerns about cybersecurity and data privacy.

What are the risks of SD? SD may not always be risk-free as information about an individual in the source dataset, or confidential data, can still be leaked due to the resemblance of the SD to the source data.

Technology, Media & Telecommunications

Key Considerations and Best Practices

The Guide includes a Handbook on Key Considerations and Best Practices in SD Generation. It describes the key considerations and best practices for organisations to reduce re-identification risks of SD through a five-step approach.

Step 1: Know your data

Before any project, the following should be observed:

- Have a clear understanding of the purpose and use cases of the SD and source data to identify possible risks.
- Establish objectives before SD generation to determine an acceptable risk threshold and the expected utility of the data so as to determine the appropriate benchmarks to assess trade-offs between data protection risks and data utility.
- Adjust the benchmarks appropriately after SD generation to meet the business objectives, taking into consideration any trade-offs between data utility and data protection risks, as well as safeguards and controls to mitigate or lower any residual risks.

Step 2: Prepare your data

When preparing the source data for generating SD, it is important to consider the following:

- What are the key insights that need to be preserved in the SD?
 - Identify trends and entity relationships to be preserved for SD generation.
 - Remove outliers if such trends/insights are not necessary.
- Which are the necessary data attributes for the SD to meet the business objectives?
 - Apply data minimisation to select only data attributes that are necessary to meet business needs.
 - Remove or pseudonymise direct identifiers.
 - Generalise granular data or add noise to the data/model if such detailed information is not necessary.
 - Standardise and document format, constraints, and categories of source data in a data dictionary.

Step 3: Generate SD

Organisations should consider which SD generation methods are most appropriate, based on their use cases, data objectives, and types of data. Organisations may then consider splitting the source data into two separate sets for assessing re-identification risks of the SD .

Technology, Media & Telecommunications

After generating SD, it is a good practice to perform the following checks on the quality of the SD:

- **Data integrity** – This ensures the accuracy, completeness, consistency, and validity of the SD as compared with the source data.
- **Data fidelity** – This examines if SD closely follows the characteristics and statistical attributes of the source data.
- **Data utility** – This refers to how well SD can replace or add to source data for the specific data objective of the organisation.

Step 4: Assess re-identification risks

After the SD is generated and utility measurement is assessed to be acceptable, organisations should assess and perform the re-identification risk assessment based on their internal acceptance criteria. Re-identification risks include:

- Singling out attack;
- Linkability attack; and
- Inference attack.

There are various approaches to determine and quantify re-identification risks. The Guide provides examples of such approaches that organisations may choose to adopt.

As there is no universally accepted numerical threshold value for risk level, the Guide provides examples of existing industry guidelines and recommendations for de-identified/anonymised data.

Step 5: Manage residual risks

Organisations should identify all potential residual risks and implement appropriate mitigation controls to minimise these risks, documenting and obtaining management approval as part of the enterprise risk framework. Such risks may include:

- New insights derived from SD;
- Potential impact on groups of individuals due to membership disclosure;
- Parties receiving SD;
- Changing environment; and
- Model leakage.

The Guide provides examples of best practices that organisations can consider implementing to manage residual risks posed by using SD.

Technology, Media & Telecommunications

Concluding Words

The Guide provides helpful insight on the best practices that organisations may adopt when generating SD. The initiative indicates PDPC's acknowledgement of the value of SD in PET efforts, while also addressing the associated risks of re-identification.

As SD is still being actively researched and developed, the Guide is not intended to stand as a comprehensive or in-depth review of the technology or its assessment methods. The Guide is intended to be a living document, and will be updated to ensure its recommendations remain relevant.

The full Guide is available [here](#).

For further queries, please feel free to contact our team below.

Contacts



Rajesh Sreenivasan
Head, Technology, Media &
Telecommunications

T +65 6232 0751

rajesh@rajahtann.com



Steve Tan
Deputy Head, Technology,
Media & Telecommunications

T +65 6232 0786

steve.tan@rajahtann.com



Benjamin Cheong
Deputy Head, Technology, Media
& Telecommunications

T +65 6232 0738

benjamin.cheong@rajahtann.com

Click [here](#) for our Partners in our Technology, Media and Telecommunications Practice.

Please feel free to also contact Knowledge Management at eOASIS@rajahtann.com

Our Regional Contacts

RAJAH & TANN SOK & HENG | *Cambodia*

Rajah & Tann Sok & Heng Law Office

T +855 23 963 112 / 113

F +855 23 963 116

kh.rajahtannasia.com

RAJAH & TANN 立杰上海

SHANGHAI REPRESENTATIVE OFFICE | *China*

Rajah & Tann Singapore LLP

Shanghai Representative Office

T +86 21 6120 8818

F +86 21 6120 8820

cn.rajahtannasia.com

ASSEGAF HAMZAH & PARTNERS | *Indonesia*

Assegaf Hamzah & Partners

Jakarta Office

T +62 21 2555 7800

F +62 21 2555 7899

Surabaya Office

T +62 31 5116 4550

F +62 31 5116 4560

www.ahp.co.id

RAJAH & TANN | *Lao PDR*

Rajah & Tann (Laos) Co., Ltd.

T +856 21 454 239

F +856 21 285 261

la.rajahtannasia.com

CHRISTOPHER & LEE ONG | *Malaysia*

Christopher & Lee Ong

T +60 3 2273 1919

F +60 3 2273 8310

www.christopherleeong.com

RAJAH & TANN | *Myanmar*

Rajah & Tann Myanmar Company Limited

T +95 1 9345 343 / +95 1 9345 346

F +95 1 9345 348

mm.rajahtannasia.com

GATMAYTAN YAP PATACSIL

GUTIERREZ & PROTACIO (C&G LAW) | *Philippines*

Gatmaytan Yap Patacsil Gutierrez & Protacio (C&G Law)

T +632 8894 0377 to 79 / +632 8894 4931 to 32

F +632 8552 1977 to 78

www.cagatlaw.com

RAJAH & TANN | *Singapore*

Rajah & Tann Singapore LLP

T +65 6535 3600

sg.rajahtannasia.com

RAJAH & TANN | *Thailand*

R&T Asia (Thailand) Limited

T +66 2 656 1991

F +66 2 656 0833

th.rajahtannasia.com

RAJAH & TANN LCT LAWYERS | *Vietnam*

Rajah & Tann LCT Lawyers

Ho Chi Minh City Office

T +84 28 3821 2382 / +84 28 3821 2673

F +84 28 3520 8206

Hanoi Office

T +84 24 3267 6127

F +84 24 3267 6128

www.rajahtannlct.com

Rajah & Tann Asia is a network of legal practices based in Asia.

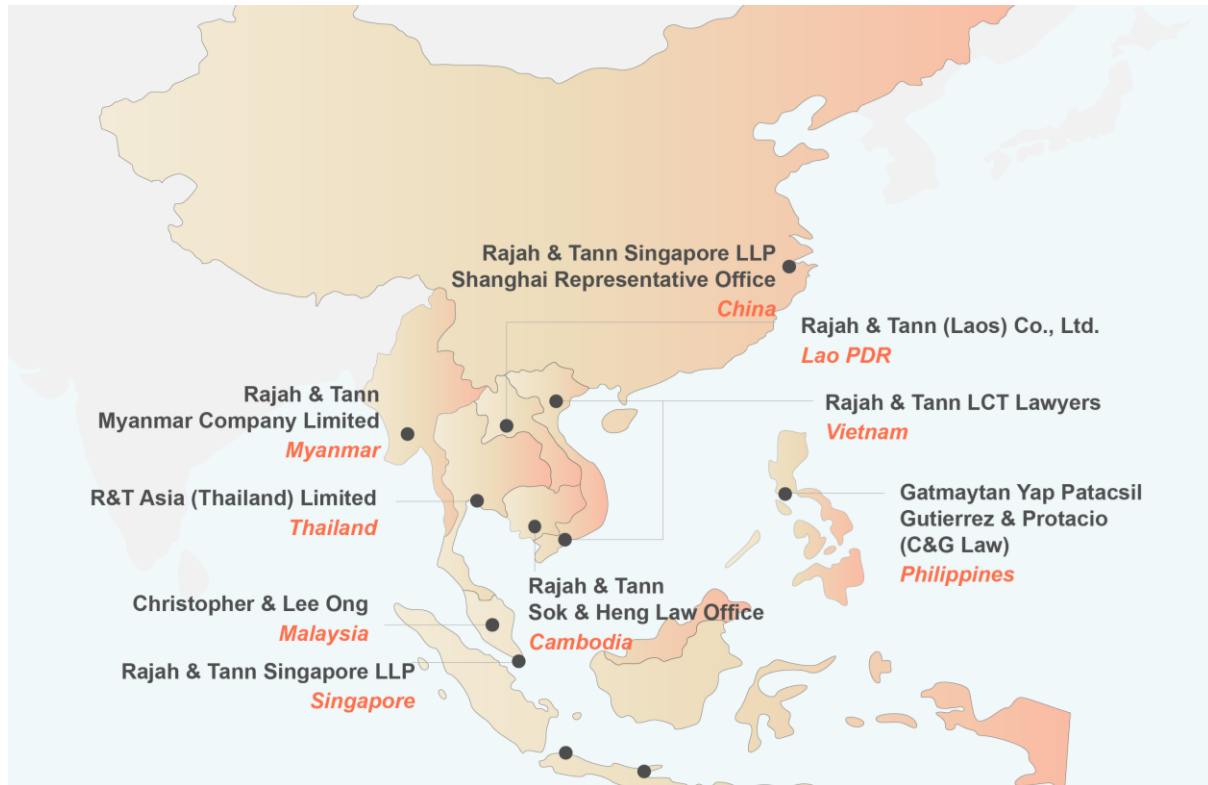
Member firms are independently constituted and regulated in accordance with relevant local legal requirements. Services provided by a member firm are governed by the terms of engagement between the member firm and the client.

This update is solely intended to provide general information and does not provide any advice or create any relationship, whether legally binding or otherwise. Rajah & Tann Asia and its member firms do not accept, and fully disclaim, responsibility for any loss or damage which may result from accessing or relying on this update.

Client Update: Singapore

2024 JULY

Our Regional Presence



Rajah & Tann Singapore LLP is one of the largest full-service law firms in Singapore, providing high quality advice to an impressive list of clients. We place strong emphasis on promptness, accessibility and reliability in dealing with clients. At the same time, the firm strives towards a practical yet creative approach in dealing with business and commercial problems. As the Singapore member firm of the Lex Mundi Network, we are able to offer access to excellent legal expertise in more than 100 countries.

Rajah & Tann Singapore LLP is part of Rajah & Tann Asia, a network of local law firms in Cambodia, China, Indonesia, Lao PDR, Malaysia, Myanmar, the Philippines, Singapore, Thailand and Vietnam. Our Asian network also includes regional desks focused on Brunei, Japan and South Asia.

The contents of this Update are owned by Rajah & Tann Singapore LLP and subject to copyright protection under the laws of Singapore and, through international treaties, other countries. No part of this Update may be reproduced, licensed, sold, published, transmitted, modified, adapted, publicly displayed, broadcast (including storage in any medium by electronic means whether or not transiently for any purpose save as permitted herein) without the prior written permission of Rajah & Tann Singapore LLP.

Please note also that whilst the information in this Update is correct to the best of our knowledge and belief at the time of writing, it is only intended to provide a general guide to the subject matter and should not be treated as a substitute for specific professional advice for any particular course of action as such information may not suit your specific business and operational requirements. It is to your advantage to seek legal advice for your specific situation. In this regard, you may call the lawyer you normally deal with in Rajah & Tann Singapore LLP or email Knowledge Management at eOASIS@rajahtann.com.