| **Document Field** | **Document Value** |
| --- | --- |
| Title | Responsible Data Science Policy - Concepts and Techniques Inventory |
| Document ID |  |
| Version |  |
| Last Updated | [DATE] |
| Last Reviewed | [DATE] |
| Responsible Body | [POLICY OWNER] |

**Table of Contents**

[**Concepts**](#_heading=h.30j0zll) **3**

[Privacy Terms](#_heading=h.1fob9te) 3

[Data Subject](#_heading=h.y83i8mbdg7k6) 3

[Dataset](#_heading=h.3znysh7) 3

[Personally Identifiable Information](#_heading=h.2et92p0) 3

[Pseudo-anonymization](#_heading=h.tyjcwt) 3

[Anonymization](#_heading=h.3dy6vkm) 3

[Re-identification](#_heading=h.1t3h5sf) 4

[K-Anonymity](#_heading=h.4d34og8) 4

[L-Diversity and T-Closeness](#_heading=h.2s8eyo1) 4

[Motivated Intruder](#_heading=h.17dp8vu) 4

[Fairness Terms](#_heading=h.3rdcrjn) 5

[Modeling](#_heading=h.8pe0qrfc1ci6) 5

[Algorithmic Bias](#_heading=h.26in1rg) 5

[Training Data Bias](#_heading=h.lnxbz9) 5

[Target Bias](#_heading=h.35nkun2) 5

[Fairness Metric](#_heading=h.1ksv4uv) 6

[Statistical Parity](#_heading=h.44sinio) 6

[Predictive Parity](#_heading=h.2jxsxqh) 6

[Equal Opportunity](#_heading=h.z337ya) 6

[Theil Index](#_heading=h.3j2qqm3) 6

[Atkinson Index](#_heading=h.1y810tw) 6

[Techniques](#_heading=h.4i7ojhp) **7**

[Types of Techniques](#_heading=h.yiawy6quha5o) 7

[Perturbative Techniques](#_heading=h.2xcytpi) 7

[Non-Perturbative Techniques](#_heading=h.1ci93xb) 7

[Deterministic Techniques](#_heading=h.3whwml4) 7

[Stochastic Techniques](#_heading=h.2bn6wsx) 7

[Technique Inventory](#_heading=h.qsh70q) 8

[Deletion](#_heading=h.3as4poj) 8

[Masking](#_heading=h.1pxezwc) 8

[Shuffling](#_heading=h.49x2ik5) 8

[Rounding](#_heading=h.2p2csry) 8

[Flooring](#_heading=h.147n2zr) 9

[Ceiling](#_heading=h.3o7alnk) 9

[Truncation](#_heading=h.23ckvvd) 9

[Rank](#_heading=h.ihv636) 9

[Percentile or Quantile](#_heading=h.32hioqz) 10

[Parametric Normalization](#_heading=h.1hmsyys) 10

[Hashing](#_heading=h.41mghml) 10

[Cryptographic Hashing](#_heading=h.2grqrue) 10

[Encryption](#_heading=h.vx1227) 11

[Homomorphic Encryption](#_heading=h.3fwokq0) 11

[Collation](#_heading=h.1v1yuxt) 11

[Aggregation](#_heading=h.4f1mdlm) 11

[Frequency Distribution](#_heading=h.2u6wntf) 12

[Embeddings and Transformers](#_heading=h.19c6y18) 12

**Note:** Examples are provided below to illustrate each concept, not to indicate adequacy from a compliance perspective. Just because an example technique is described does not mean it should be performed.

**Note:** Within this document and as used in the Policy and Procedures, concepts may be defined more broadly than their commonly-accepted or colloquial definitions. This is intentional, as these documents are designed to address multiple jurisdictions, sectors, and activities.

# Concepts

## **PRIVACY TERMS**

### Data Subject

A Data Subject may refer to an individual, organization, product, event, or any other person, place, or thing about which data is collected.

**Note:** This definition is scoped more broadly than traditional privacy policies or Data Subject definitions.

### Dataset

A Dataset is information that contains attributes about or relations between one or more Data Subjects.

### Personally Identifiable Information

Personally Identifiable Information is any and all information that could be interpreted as relating to individuals under a statute, regulation, or ruling related to the Organization. Privacy frameworks vary across jurisdictions and industries in their classification of personally identifiable information, and this definition is meant to capture information under all such relevant frameworks.

**Synonyms:** Personally Identifying Information, PII, Personal Information, Personal Data

### Pseudo-anonymization

Pseudo-anonymization is the process of transforming data to increase the cost or decrease the probability of inferring selected information about a Data Subject. The term anonymization is used colloquially to describe pseudo-anonymization which results in either a cost exceeding reasonable feasibility or a probability approaching zero.

**Synonyms**: pseudo-anonymisation

**See also**: anonymization

### Anonymization

Anonymization is a term used to describe a pseudo-anonymization process which results in either a cost exceeding reasonable feasibility or a probability approaching zero of inferring selected information about a Data Subject.

**Note:** Anonymization is context-dependent; Personnel should consider the budget and resources of potential threat actors and monitor trends in computation, storage, and data breach that might otherwise reduce the cost of de-anonymizing a Dataset that was previously deemed to be Anonymized.

**Synonyms:** anonymisation

**See also**: pseudo-anonymization

### Re-identification

Re-identification is the process of inferring Data Subject identity from data that has been previously omitted or pseudo-anonymized.

**Synonyms:** de-anonymization, de-anonymisation, identity disclosure

### K-Anonymity

k-anonymity is a measure of the size of the smallest group of Data Subjects indistinguishable from each other by their attributes within a pseudo-anonymized dataset. If a dataset has k=1-anonymity, then there is at least one Data Subject with unique attributes that is at increased risk of re-identification. In general, datasets with larger k-anonymity values have lower probabilities of re-identification than those with lower value. However, approaches based on k-anonymity may be vulnerable to targeted attacks that could be avoided through the use of l-diversity or t-closeness metrics.

**See also:** l-diversity, t-closeness

### L-Diversity and T-Closeness

l-diversity and t-closeness are a family of measures of the distribution of sensitive attributes within groups of Data Subjects in a dataset, such as through the number of distinct values or attribute entropy. In general, datasets with larger l-diversity or t-closeness values have lower probabilities of re-identification than those with lower values. Privacy approaches based on l-diversity and t-closeness are generally less susceptible to attack than those based on k-anonymity. t-closeness is a refinement of l-diversity that provides a higher degree of differential privacy.

**See also:** k-anonymity

### Motivated Intruder

A hypothetical threat actor whose goal is to re-identify one or more Data Subjects from a Pseudo-Anonymized Dataset. This threat actor has access to reasonably-accessible sources of information, such as public records or information leaked via other data breaches. Typically, this hypothetical threat actor is used to construct estimates of the cost or probability of re-identification for a given Data Subject or Dataset.

## **FAIRNESS TERMS**

### Modeling

Modeling is any activity that might be described as inferential, causal, predictive, or generative modeling in nature. The most common activities relate to the creation or management of feature or input data, the preprocessing or engineering of such feature or input data, and clustering, classification, regression, hyperparameter optimization, generation, or simulation using such feature or input data. The output of a Modeling activity is referred to as a Model.

**Example:** A classification algorithm used to screen resumes for hiring is a Model. The process of obtaining the data from an HR system, engineering features from this data, and fitting a decision tree to the data is Modeling.

### Algorithmic Bias

A Model can be described as having Algorithmic Bias when it produces systematically unfair inferences or outcomes. Algorithmic Bias can occur in any type of Modeling or Model.

**See also:** Training Data Bias, Target Bias

**Example:** A classification algorithm is used to screen resumes for hiring. The algorithm systematically rejects older applicants.

### Training Data Bias

A Model can be described as having Training Data Bias when the features or inputs used in Modeling result in Algorithmic Bias. Training Data Bias can arise in a number of ways, such as the manner of Data Subject identification or the inclusion, omission, processing, or annotation of attributes related to the Data Subjects.

**See also:** Algorithmic Bias, Target Bias

**Example:** A classification algorithm is used to screen resumes for hiring. The algorithm is trained on a Dataset that does not include any older applicants.

### Target Bias

A Model can be described as having Target Bias when the targets, labels, or objectives used in Modeling result in Algorithmic Bias. There are various ways in which Target Bias is introduced, such as the framing of the research problem, the manner of Data Subject identification or sampling, or the inclusion, omission, processing, or annotation of attributes related to the Data Subjects.

**See also:** Algorithmic Bias, Training Bias

**Example:** A classification algorithm is used to screen resumes for hiring. The target used to fit the classification algorithm is based on whether the applicant will remain at the Organization for at least 10 years.

### Fairness Metric

A Fairness Metric is a quantitative measure of algorithmic bias in a Model. There are a wide range of such metrics, some of which are focused on specific outcomes or units, such as wealth or salary.

**See also**: Statistical Parity, Predictive Parity, Theil Index

### Statistical Parity

Statistical Parity is a fairness measure based on the difference in a Model’s positive rate based on a protected Attribute. A Model can be said to exhibit Statistical Parity if its positive rate is not statistically-significantly different across selected groups.

**See also**: Predictive Parity, Equal Opportunity, Theil Index

### Predictive Parity

Statistical Parity is a fairness measure based on the difference in a Model’s precision based on a protected Attribute. A Model can be said to exhibit Predictive Parity if its precision is not statistically-significantly different across selected groups.

**See also**: Statistical Parity, Equal Opportunity, Theil Index

### Equal Opportunity

Statistical Parity is a fairness measure based on the difference in a Model’s false negative rate based on a protected Attribute. A Model can be said to exhibit Equal Opportunity if its false negative rate is not statistically-significantly different across selected groups.

**See also**: Statistical Parity, Predictive Parity, Theil Index

### Theil Index

The Theil Index is an inequality measure based on the degree to which an income or wealth distribution varies from a baseline distribution. Larger values correspond to more inequality, not more fairness.

**See also**: Statistical Parity, Predictive Parity, Equal Opportunity, Atkinson Index

### Atkinson Index

The Atkinson Index is an inequality measure based on normalizing the Theil index. Values closer to 1 indicate more inequality, whereas values closer to 0 indicate more equality of income.

**See also**: Statistical Parity, Predictive Parity, Equal Opportunity, Theil Index

# Techniques

## **TYPES OF TECHNIQUES**

### Perturbative Techniques

A Perturbative Technique is a Pseudo-Anonymization process that introduces uncertainty into a Dataset through the alteration of attributes away from their “true” values. Perturbative Techniques should be used with caution, for example, when the resulting values may be used in the control of a physical system.

**Example:** A Dataset contains the age of individuals. The age attribute of each individual is altered by adding or subtracting a random number between 0 and 10. For example, records for three 25-year-old individuals might be altered to have ages 22, 27, and 30.

**See also:** Deterministic Technique, Stochastic Technique

### Non-Perturbative Techniques

A Non-Perturbative Technique is a Pseudo-Anonymization process that introduces uncertainty through the reduction, not the alteration, of attribute information.

**Example:** A Dataset contains the age of individuals. The age attribute of each individual is truncated to the tens place. For example, all records for 25-year-old individuals would indicate that their age is 20.

**See also:** Deterministic Technique, Stochastic Technique

### Deterministic Techniques

A Deterministic Technique is a Pseudo-Anonymization or Modeling process that can be verified to produce the same output for the same inputs.

**See also:** Perturbative Technique, Non-Perturbative Technique

**Example:** A Dataset contains the age of an individual. The individuals are ranked by age, and each individual’s age is replaced by the next youngest age; the youngest individual is removed from the Dataset.

### Stochastic Techniques

A Stochastic Technique is one that may produce different results for the same input, typically due to the use of random number generation.

**Synonym:** Non-Deterministic, Probabilistic

**See also:** Perturbative Technique, Non-Perturbative Technique

**Example:** A Dataset contains the age of individuals. The individuals are ranked by age, and each individual’s age is replaced at random by either the next youngest or next oldest age; the oldest and youngest individuals are removed from the Dataset.

## **TECHNIQUE INVENTORY**

**Note:** The techniques below are not mutually exclusive, and some definitions may overlap with each other. The list below is broadly intended to cover techniques as commonly used or referenced in standards, regulations, or common commercial contracts.

### Deletion

Deletion is a Pseudo-Anonymization technique that removes a Data Subject or attribute from a Dataset. Deletion can affect one or more Data Subjects or one or more attributes. In general, if Deletion is selected for an attribute or Data Subject, Personnel should consider how and why the data was obtained and simply not collect that data from Data Subjects in the future.

**Example:** A classification algorithm is used to screen resumes for hiring. The Dataset obtained from HR contains an age attribute. The age attribute is removed from the Dataset.

**Synonym**: Complete Redaction

### Masking

Masking is a Pseudo-Anonymization technique that replaces or removes a subset of an attribute, typically while preserving the length or structure of an attribute. Masking is most commonly applied to data that is naturally character- or text-based.

**Example, Deterministic:** The last N digits of a postal code care replaced with a character like # or \*.

**Example, Stochastic:** The last N digits of a postal code are replaced with two random letters.

### Shuffling

Shuffling is a Stochastic Pseudo-Anonymization technique that alters the ordering of relationships between Data Subject and attributes.

**Example**: A Dataset contains the age of individuals. Each individual is assigned at random the age of a different individual in the Dataset.

### Rounding

Rounding is a Pseudo-Anonymization technique that alters a numeric attribute by replacing a value with the nearest point on a discretization of the number line, most typically discretized by tens, ones, or tenths place. Tie-breaking in rounding may be Deterministic or Stochastic, for example, by 0.5 up or down at random or by always rounding down. For ordinal values, rounding and binning are typically identical.

**Example:** A Dataset contains the age of individuals. The ages are rounded to the nearest multiple of five; the ages 22, 23, and 27 are transformed to 20, 25, and 25.

### Flooring

Flooring is a Pseudo-Anonymization technique that alters a numeric attribute by replacing a value with the first integer value that it is less than or equal to. Flooring is similar to “rounding down” with integer discretization.

**Example**: A Dataset contains the age of individuals recorded to the first decimal. The ages 22.0, 22.6, and 22.8 are all replaced by the value 22.

### Ceiling

Ceiling is a Pseudo-Anonymization technique that alters a numeric attribute by replacing a value with the first integer value that it is greater than or equal to. Ceiling is similar to “rounding up” with integer discretization.

**Example**: A Dataset contains the age of individuals recorded to the first decimal. The ages 22.0, 22.6, and 22.8 are all replaced by the values 22, 23, and 23.

### Truncation

Truncation is a Pseudo-Anonymization technique that alters an attribute by removing digits or characters, typically from the end of a written number or string. For numeric values, truncation is typically equivalent to flooring.

**Example:** A Dataset records the four-digit year of birth of an individual. The last digit of each year is truncated, resulting in three-digit values. For example, 1971, 1982, and 1993 are replaced by 197, 198, and 199.

**See also**: flooring

### Rank

Ranking is a Pseudo-Anonymization technique that alters an attribute by replacing it with the value’s relative position in an ordering. In the case of attributes that are cardinal or ordinal, the rank is the mathematical rank of the values. Non-numeric attributes that have an ordinal interpretation can be mapped to ordinal values and ranked accordingly.

**Example**: A survey records open-ended feedback from employees about a company’s culture; a sentiment score is calculated for each response. Each individual’s sentiment score is then replaced with its rank relative to others. For example, if three individuals’ feedback scored at 3/10, 6/10, and 5/10, then these values would be replaced with 1, 3, and 2, respectively.

**See also**: Percentile, Quantile, Parametric Normalization

### Percentile or Quantile

A Percentile is a measure of the percentage of numerical attributes that are less than a given numerical attribute in a Dataset. A quantile is simply a Percentile stated in decimal units instead of percentage. Percentiles and Quantiles can be used as a Pseudo-Anonymization technique by altering an attribute to replace it with its corresponding percentile or quantile in a Dataset.

**Example:** A Dataset contains the age of 50 individuals. Each individual’s age is replaced with the corresponding percentile for the distribution of ages in the Dataset. For example, if 35 individuals have an age less than 27, then any individuals with an age of 27 will have their attribute replaced by 35/50 = 70% = 0.7.

**Synonym:** Inverse Cumulative Distribution Function (CDF, ICDF)

**See also:** Rank, Parametric Normalization

### Parametric Normalization

Parametric Normalization is a Pseudo-Anonymization technique that alters an attribute by replacing it with a value that is calculated using a sample or population distribution parameter, such as a standard deviation. Parametric Normalization techniques typically include scaling (division) or shifting (subtraction), such as when normal or Gaussian data is transformed by subtracting the mean and dividing by the standard deviation.

**Example:** A Dataset contains the heights of individuals. Using externally-known population parameters or sample statistics, the height of each individual is replaced by dividing the value by the standard deviation.

### Hashing

Hashing is a Deterministic Pseudo-Anonymization technique that transforms attributes of an arbitrary size to a fixed-length value. This technique is not intended to be reversible, though it is often intended to be approximately unique. Hashing is most commonly applied to binary data or data that is naturally character- or text-based.

**Example Algorithms:** MD5, RIPEMD, SHA1, SHA-2, SHA-3, BLAKE2, BLAKE3

**Example:** The SHA2-256 digest of the word “policy” is 823412d1eacb67956220e532959f0104603057c88704863ca38e7cd188fda812.

**Synonym:** One-Way Encryption (colloquial), Hash

**See also:** Cryptographic Hashing, Encryption

### Cryptographic Hashing

A Hashing technique can be described in a given context as a Cryptographic Hashing technique when it is not practically feasible to determine the input that produced a given Hashing output. Cryptographic Hashing techniques should have transparent models for estimating the costs of producing collisions and probabilities of collision.

**Synonym:** Cryptographic Hash

**See also:** Hashing

### Encryption

Encryption is a Deterministic Pseudo-Anonymization technique that transforms attributes of an arbitrary size into a different representation of arbitrary size that is not reversible by an actor without access to specific information, typically referred to as key(s). Encryption can be described as “strong” when it is not practically feasible to determine the original attribute value as a result of cost or probability. Some Encryption techniques may be designed to allow for verification of some information about a message, such as public-key signatures.

**Example Algorithms:** Substitution Ciphers, Triple DES (3DES), Blowfish, Twofish, AES, RSA,

**Synonyms:** Cipher

**See also:** Hashing, Homomorphic Encryption

### Homomorphic Encryption

An Encryption technique can be described as Homomorphic in a given context when it supports computation without decryption. Such encryption can be described as partially- or fully-homomorphic depending on the types of computations that are supported.

**Example Algorithms:** BFV, CKKS

**Example:** Under a homomorphic system supporting addition, two encrypted numbers can be added together by a party without the ability to decrypt them. When the result of this computation is decrypted by a party with such ability, the answer will be the correct sum.

**See also:** Encryption

### Collation

Collation is the process of combining information from two or more Datasets to create a new Dataset. Collation can often improve the ability to achieve Pseudo-Anonymity as seen through k-anonymity or l-diversity measures. Collation can also reduce the impact of Training Data Bias by increasing the sample size or diversity of certain Attributes or outcomes.

**Example:** Five companies pool their hiring data together to create a larger Dataset of candidates and employee history. Models trained from this collated Dataset are less likely to reproduce any bias in the decisions of a single company’s hiring process, and higher degrees of differential privacy are more likely to be obtained.

### Aggregation

Aggregation is the process of combining information from two or more Data Subjects. In the case of Aggregation with Data Subjects, Aggregation is typically combined with functions such as a count, sum, average (mean), median, minimum, maximum, range, standard deviation, or other statistical function.

**Example Implementations:** SQL GROUP BY statements, split-apply-combine patterns, map-reduce patterns

**Example:** A Dataset contains the age and country of residence of individuals. Individuals are grouped by their country of residence and then the average age within each group is calculated. Each individual’s age could then be replaced with their respective average for their country of residence.

**Synonyms:** binning, bucketing

### Frequency Distribution

Frequency Distribution is a Pseudo-Anonymization technique that is typically used to transform character- or binary-based information into a list of unique elements and the number of times each element occurs. Common elements include single characters, single words (tokens), single bytes, or tuples of consecutive words (n-grams). In cases where n-grams are used, Personnel should consider whether increasing n can decrease the pseudo-anonymity of the resulting feature data.

**Synonyms:** histogram, distribution, frequency table

**Example:** A Dataset contains the text of an email from one individual to another that reads “We can get lunch tomorrow so we can talk.” The text is replaced with the frequency distribution count(we)=2, count(can)=2, count(get)=1, count(lunch)=1, count(tomorrow)=1, count(so)=1, count(talk).

### Embeddings and Transformers

Embeddings and Transformers are Pseudo-Anonymization techniques that are typically used to transform higher-dimensional data such as images or text into lower-dimensional representations for subsequent clustering or classification. Many common embedding or transformer models are either fully or partially pre-trained. Personnel should take care to ensure that any pre-trained models used do not violate this Policy or related Procedures.

**Example**: A Dataset contains the text of resume biographies received from applicants. The text of each biography is replaced by a vector corresponding to the sum of GLoVe vectors for all words in the biography.