Confidential AI

Robust **E**valuation and **A**lgorithmic **L**earning **M**odels

2025-02-06

The Future That Came Early

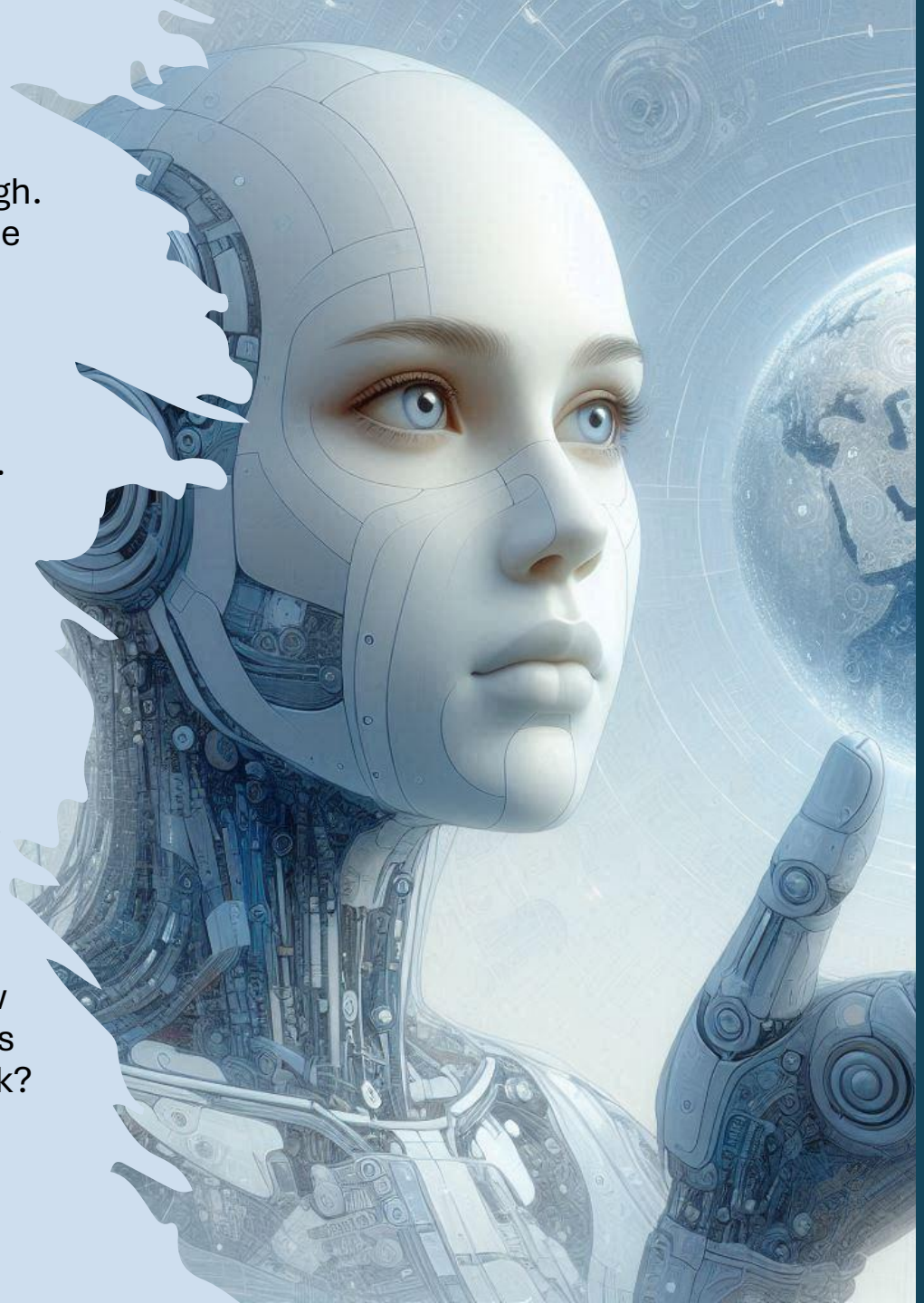
Once, we relied on machines to predict the future. They knew the past well enough. But the world was moving too fast. Predictions fell short, like a sailor staring at the horizon, blind to the storm coming from behind.

Then came a new kind of intelligence—generative AI. It didn't guess. It created. It shaped new worlds, built from nothing but numbers and possibilities. Synthetic data, they called it—flawless copies of reality. You could tear them apart, push them to their limits, without spilling a single secret. We could simulate the world. What would happen if a city fell to a plague? What if a housewife in Connecticut watched her savings vanish as unemployment rose?

But there was danger in creation. Generative systems unlocked a new kind of power. Digital Twins—perfect replicas of people, systems, lives. And with that power came risk. What if we lost control of these copies? What if someone, somewhere, used them to tear down the very thing they were meant to protect? That's why we built Confidential AI. It wouldn't just create. It would safeguard. These twins, these creations, would remain sealed, like secrets locked in a vault. No one would ever reach behind the curtain to see who, or what, they truly were. We had made something powerful, but we had made it safe.

Now we can ask the questions we never dared before. What happens if tomorrow becomes something we can't predict? If the ground beneath us shifts? And in this uncertain future, who will help us navigate the storm, without letting the ship sink?

We stand at the edge of something new. A future shaped by intelligence but guarded by trust.



Approach: ONE PLATFORM, TWO POWERFUL AI SOLUTIONS

- **PAMOLA**: A robust Confidential AI system focused on privacy and security using PET (Privacy-Enhancing Technologies). It ensures data protection with synthetic data generation, Federated Learning (FL), Explainable AI (XAI), and advanced privacy techniques. PAMOLA is your shield for secure AI-driven business operations.

- **AYITA**: A specialized intelligent personal assistant (PA) powered by local LLMs, fine-tuning, and Retrieval-Augmented Generation (RAG). AYITA provides tailored, privacy-safe support for business tasks with expertise in various industries, offering adaptable and precise solutions for enterprise needs.

WHAT IS THE REALM

REALM is a cutting-edge **Confidential AI platform** designed to provide secure, private, and innovative AI solutions for businesses. It offers advanced data privacy features, ensuring that sensitive information remains protected while delivering high-performance AI capabilities tailored to various industries, including finance, healthcare, and e-commerce.

AYITA

1. AI Agent for Enterprise and B2C:

AYITA is a specialized, privacy-focused personal assistant (PA) powered by local LLMs, designed to securely handle business-specific tasks and data with customizable industry expertise.

2. Fine-Tuning & Model Customization:

Seamlessly fine-tune and customize AI models using the latest frameworks and automated hyperparameter optimization tools.

3. Synthetic Data Generation with Differential Privacy:

Create realistic, privacy-preserving data for AI models, supporting multimodal data and generative models like GAN and VAE, with differential privacy to ensure data security.

4. Federated Learning:

Train AI models across decentralized data sources without sharing sensitive information, ensuring data privacy while leveraging distributed datasets.

5. Explainable AI (XAI):

Full transparency of AI decisions with visualization tools and methods like LIME, SHAP, and counterfactual analysis.

PAMOLA

6. Confidential Computing:

Secure, privacy-preserving data processing with hardware solutions (e.g., Intel SGX) and advanced cryptography (MPC, FHE).

WHY REALM?

Generative AI is rapidly evolving and has the potential to transform industries through efficiency and personalization. However, its widespread adoption is slowed by challenges such as poor data quality, centralized control, and growing cybersecurity threats. REALM addresses these issues by providing decentralized, privacy-focused AI solutions, ensuring businesses can innovate securely without compromising on data integrity.

Lack of Quality Data

- Insufficient high-quality data for training AI models.
- Issues with data diversity and representativeness.

AI systems need large amounts of high-quality data to work effectively, but such data is often scarce and difficult to obtain because it is expensive and time-consuming to collect and curate.

According to Gartner, poor data quality costs organizations an average of $15 million per year in losses.

60% of AI projects fail to move past the proof-of-concept stage due to data quality issues (VentureBeat, 2020).

Confidentiality

- Ensuring data privacy and protecting sensitive information.
- Compliance with data protection regulations and standards.

The essence of the privacy issue in AI is that AI cannot function without data, and if the data is real, it often contains personal information. Using such data is problematic due to legal regulations and ethical concerns.

A survey by IBM found that 81% of consumers have lost trust in how companies handle their data.

The GDPR has led to over €275 million in fines for data breaches since its implementation (DLA Piper, 2020).

Excessive Centralization of LLM

- Over-reliance on centralized large language models (LLM).
- Risks of single points of failure and lack of distributed control.

Centralized AI models can become single points of failure, as seen with outages in cloud services impacting numerous applications.

85% of enterprise decision-makers report concerns about AI model centralization and its associated risks (O'Reilly, 2021).

Security and Explainability of Models

- Protecting AI models from adversarial attacks.
- Ensuring AI decisions are transparent and understandable.

The MITRE ATLAS initiative identifies over 500 techniques for adversarial attacks against AI systems.

73% of organizations consider AI model explainability a critical factor in adoption, according to a survey by FICO (2020).

THE MARKET

Generative AI is on the verge of reshaping our world, tackling problems we once thought unsolvable. But its path is fraught with challenges: fragmented data, centralized control, and lurking security threats. Despite this, the market is surging forward, driven by relentless innovation and the promise of profound change.

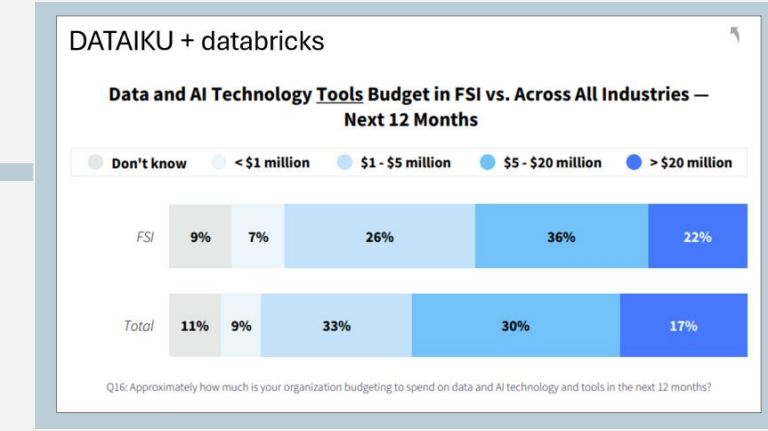
The global AI market was valued at **$196.63 billion in 2023** and is projected to grow with a **CAGR of 37.3%** through 2030.

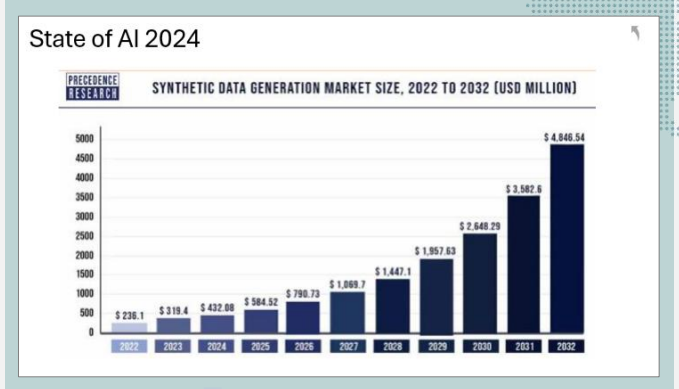
80% of AI startups successfully pass their first year, showing the sector's resilience and potential.

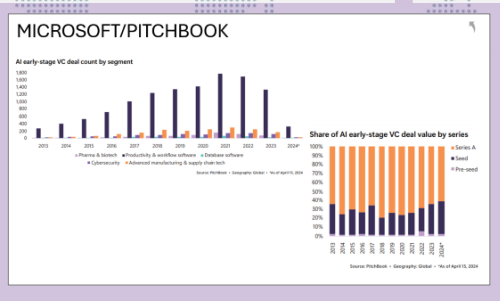
The U.S. leads, securing **over 50% of global AI startup funding** in Q4 2023.

OpenAI received the largest venture capital funding, totaling **$14 billion** in 2023.

AI startups attracted **nearly $50 billion in funding** throughout 2023, significantly outpacing non-AI sectors.







EDGE DELTA

REALM CYBERSECURITY AI

HOW IT WORKS?

REALM is a comprehensive solution aimed at solving data security issues when using AI. It integrates advanced Cyber AI algorithms to ensure data confidentiality and privacy, providing a secure and reliable platform for AI applications. By focusing on transparency and ethical AI practices, REALM empowers organizations in finance and healthcare to harness the power of AI without compromising on security.

Applications:

E-COMMERCE

FINANCE

- Fraud Detection
- Churn Analysis
- Credit Risk Management
- Personal Investment Management

- Organ Transplantation
- Remote patient support
- Telemedicine
- Hospitals Business Processes

HEALTHCARE

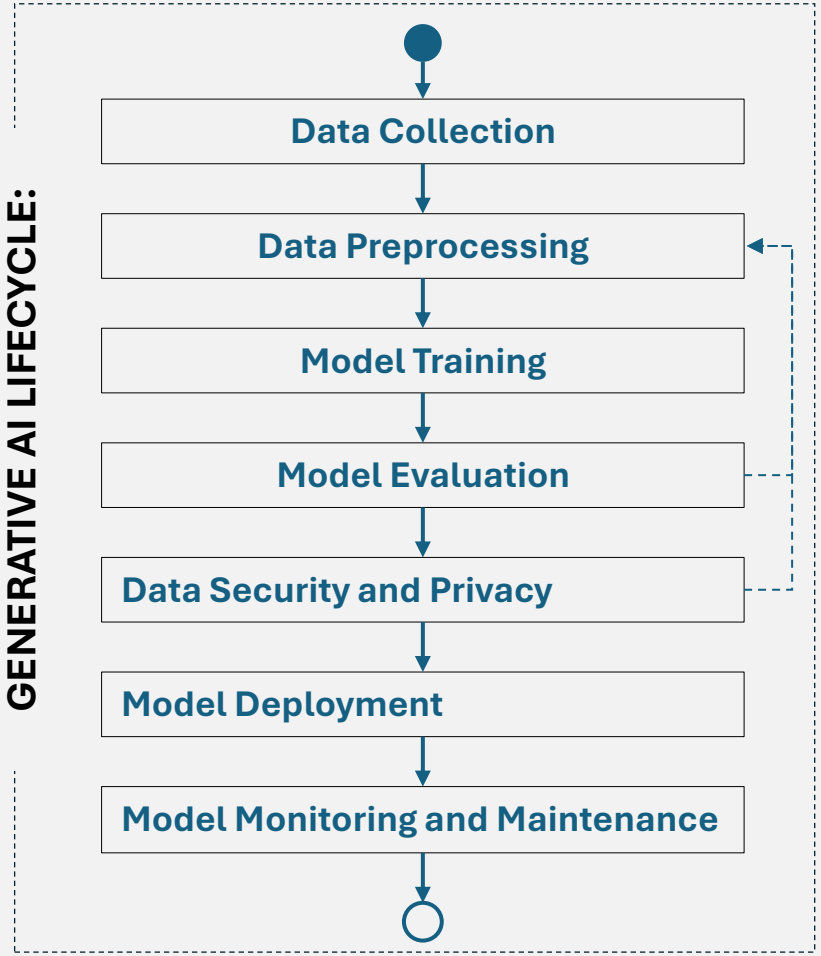
UNDERSTANDING AI

Artificial Intelligence (AI) enables machines to perform tasks that typically require human intelligence, such as clustering, regression, and inference. It is based on mathematical and statistical principles.

Generative AI creates new content, like text, images, and music, based on the data it has been trained on.

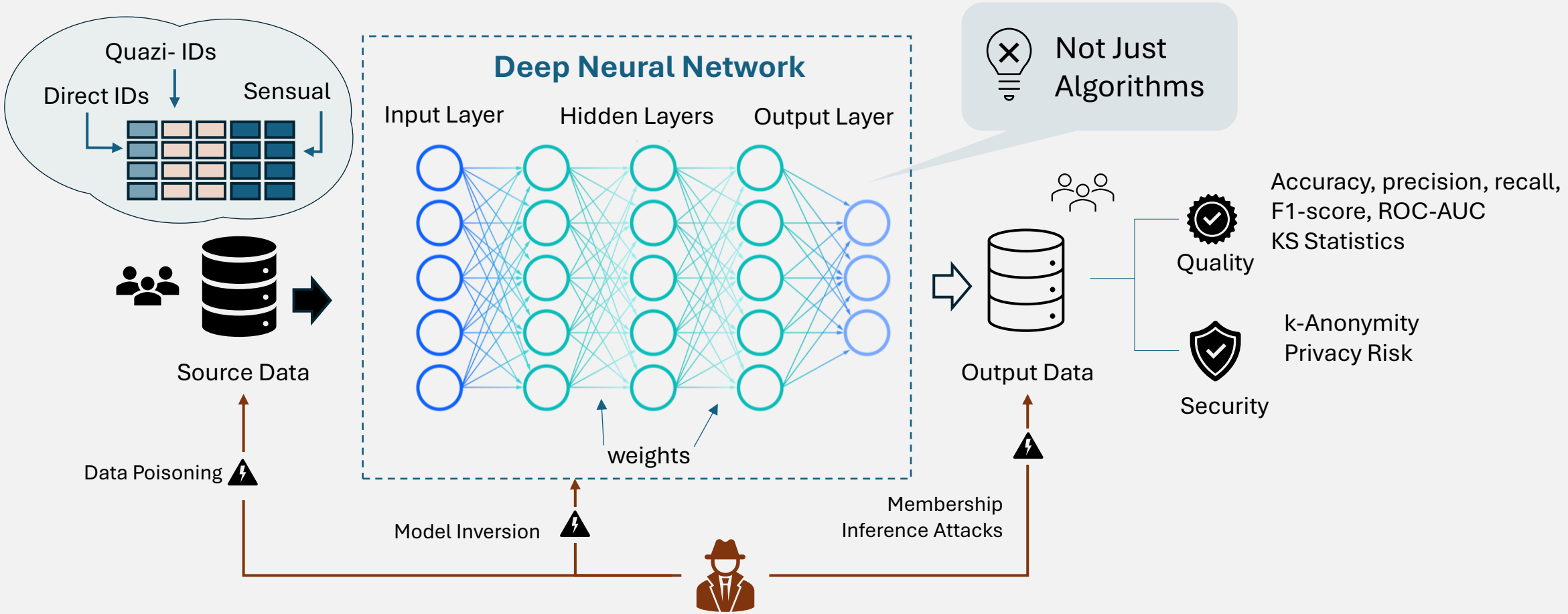
KEY CONCEPTS:

- **Neural Network:** A set of algorithms modeled after the human brain, designed to recognize patterns. Consists of layers of nodes (neurons) that process input data to generate an output.
- **Model:** A mathematical representation of a real-world process. Used by AI systems to make predictions or decisions based on input data.
- **Model Training**: The process of teaching a model to make accurate predictions by exposing it to data. Adjusts the model's parameters to minimize errors in predictions.
- **GANs (Generative Adversarial Networks), Transformers**: GANs create realistic images by pitting two neural networks against each other; Transformers generate coherent text by understanding context.
- **Large Language Models (LLM)** : AI models trained on vast amounts of text data. Understand and generate human-like language, enabling tasks like translation, summarization, and conversation.
- **Machine Learning**: A subset of AI focused on building models that learn from data. Involves training, validating, and testing models to improve their accuracy
- **Deep Learning**: A subset of machine learning that uses neural networks with many layers (deep neural networks) to model complex patterns in large datasets. Powers advanced AI applications, such as image recognition, natural language processing, and autonomous driving.



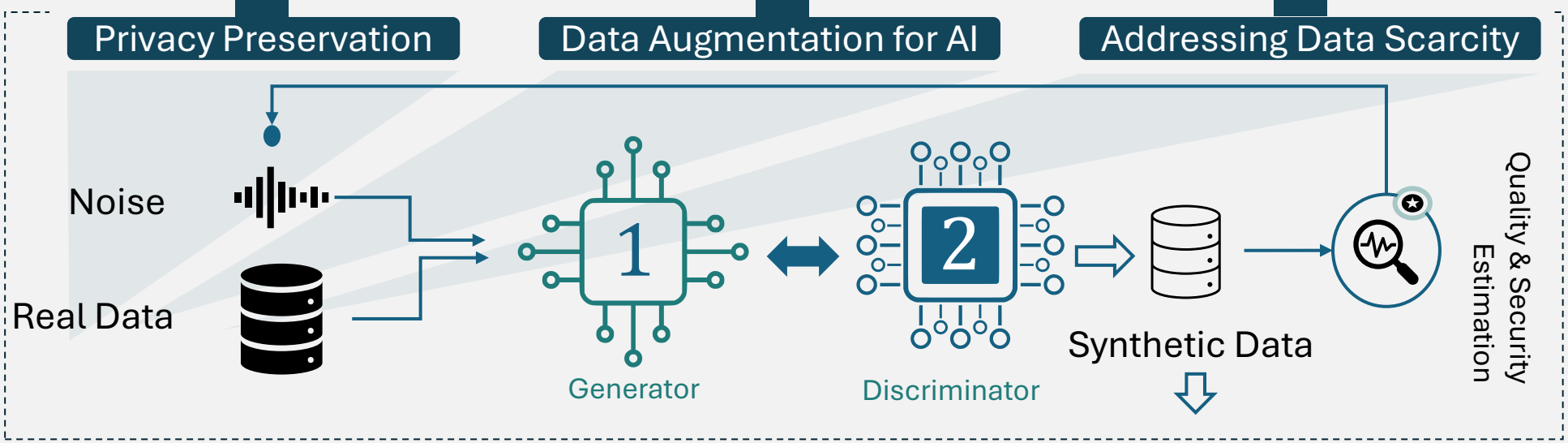
HOW AI MODEL WORKS

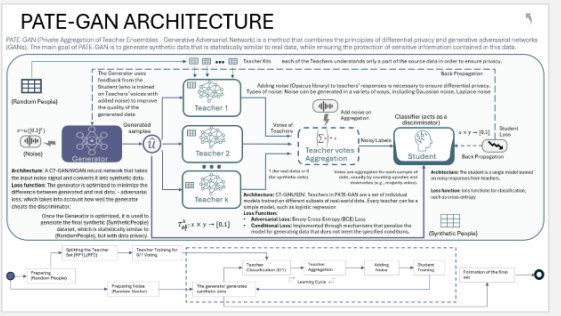
Understanding the intricacies of AI models is essential for leveraging their full potential. These models rely on complex systems and high-quality data to function effectively and securely, making it crucial to grasp their structure, training process, and the importance of robust security measures.

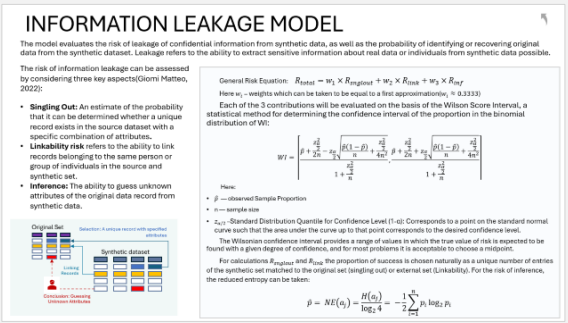


SYNTHETIC DATA

Synthetic data is artificially generated data that mimics real-world data. It is crucial for enhancing privacy, improving machine learning models, and addressing data scarcity issues. Advanced techniques like differential privacy and Generative Adversarial Networks (GANs) play a significant role in creating high-quality synthetic data.







CHALLENGES AND ISSUES:

Quality and Accuracy: Ensuring that synthetic data accurately represents the statistical properties of real data.

Bias and Fairness: Avoiding the introduction of biases during the synthetic data generation process.

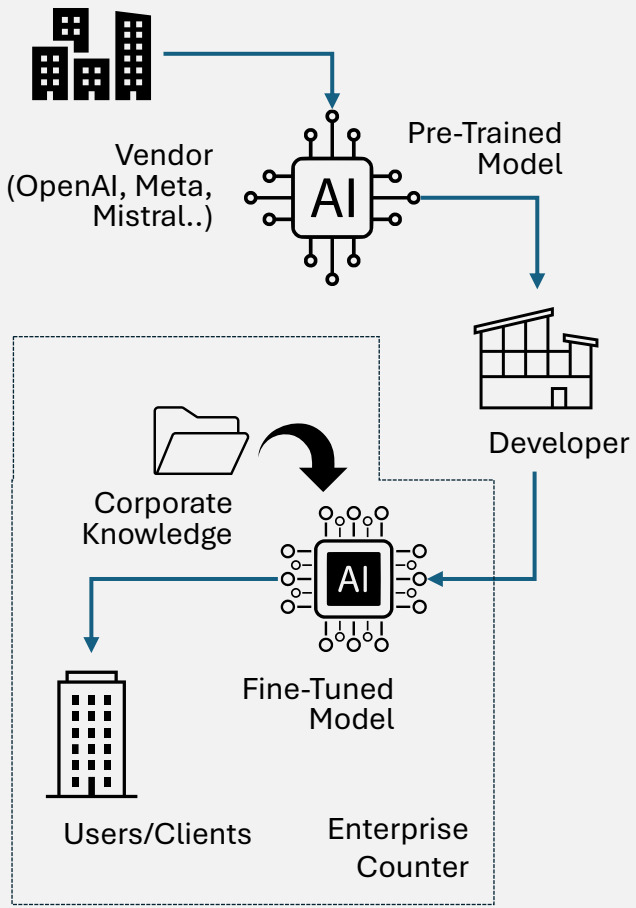
High Cost: The development and maintenance of synthetic data generation systems can be expensive.

Potential Services with Synthetic Data

- **Data Privacy Services**: Offering anonymization and privacy-preserving services alongside synthetic data products.
- **Synthetic Data Generation Tools**: Providing platforms and APIs for generating synthetic data on demand.
- **Consulting and Support**: Helping organizations implement synthetic data solutions and ensuring data compliance.

FINE TUNING

Fine-Tuning is the process of adjusting a pre-trained model on a new, often smaller dataset to specialize it for a specific task. This technique allows leveraging the knowledge learned by the model on a large dataset and adapting it to new, specific requirements.

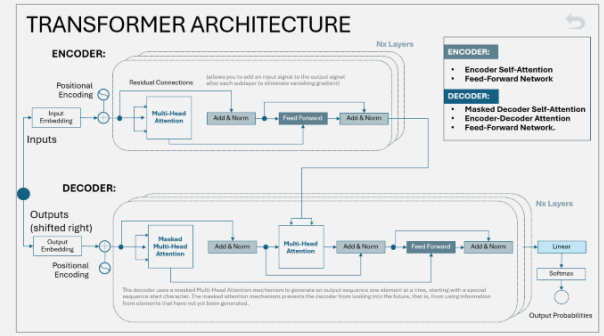


DRIVERS:

Improved Performance: Enhances model accuracy and performance on specific tasks.
Efficiency: Saves time and computational resources compared to training a model from scratch.
Utilization of Limited Data: Makes effective use of smaller, domain-specific datasets.
Customization: Allows models to meet specific business needs and requirements.

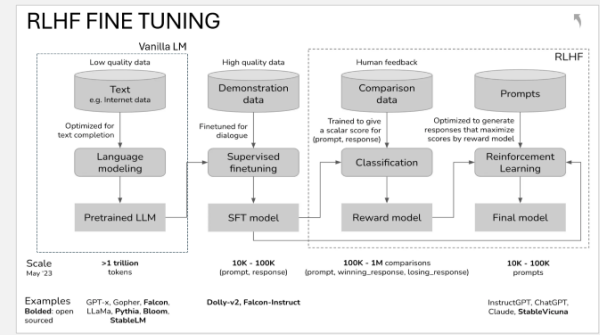
SERVICES:

Custom Model Development: Offering specialized models fine-tuned for specific client needs.
Model Optimization: Services to enhance and optimize existing models for better performance.
Training Data Services: Providing data collection, annotation, and preprocessing for fine-tuning purposes.
Consulting and Support: Assisting organizations in implementing and managing fine-tuning processes.



FINE TUNING APPROACHES

#	Title	Description	Process	Pros	Cons
1	Reinforcement Learning from Human Feedback (RLHF)	Uses human feedback to fine-tune models for nuanced tasks	Model is fine-tuned based on feedback from human evaluators who rank outputs	High accuracy and performance on subjective tasks	Time-consuming and resource-intensive
2	Low-Rank Adaptation (LoRA)	Introduces low-rank matrices to model parameters for efficient fine-tuning	Adds low-rank adaptations to the model's weights, enabling significant changes with minimal computation	Efficient, requires fewer resources	May not be as effective for highly complex tasks
3	Feature-Based Fine-Tuning	Adjusts feature extraction layers while keeping other parts unchanged	Fine-tunes only the last few layers or adds additional layers for specific tasks	Preserves general knowledge, less computationally intensive	Limited to tasks closely related to the pre-trained model's original purpose
4	Layer-Wise Freezing	Gradually unfreezes layers during fine-tuning	Starts by training only newly added layers, progressively unfreezes more layers	Fine-grained control over training process, reduces overfitting	More complex training process
5	Multi-Task Learning	Fine-tunes a model for multiple tasks simultaneously	Trains on a combined dataset including examples from multiple tasks	Improves generalization, leverages shared representations	Requires diverse and well-labeled multi-task dataset



XAI APPROACH

In the quest to make artificial intelligence (AI) more interpretable and transparent, Explainable AI (XAI) employs a variety of methods and tools designed to shed light on the decision-making processes of complex models. These methods can broadly be categorized into model-agnostic, model-specific, and visualization-based approaches, each offering unique insights into how AI systems operate.

Model-Agnostic Methods

These methods are designed to work with any machine learning model, providing flexibility and broad applicability. They do not require access to the model's internal architecture, making them highly versatile for explaining different types of AI models.

Model-Specific Approaches

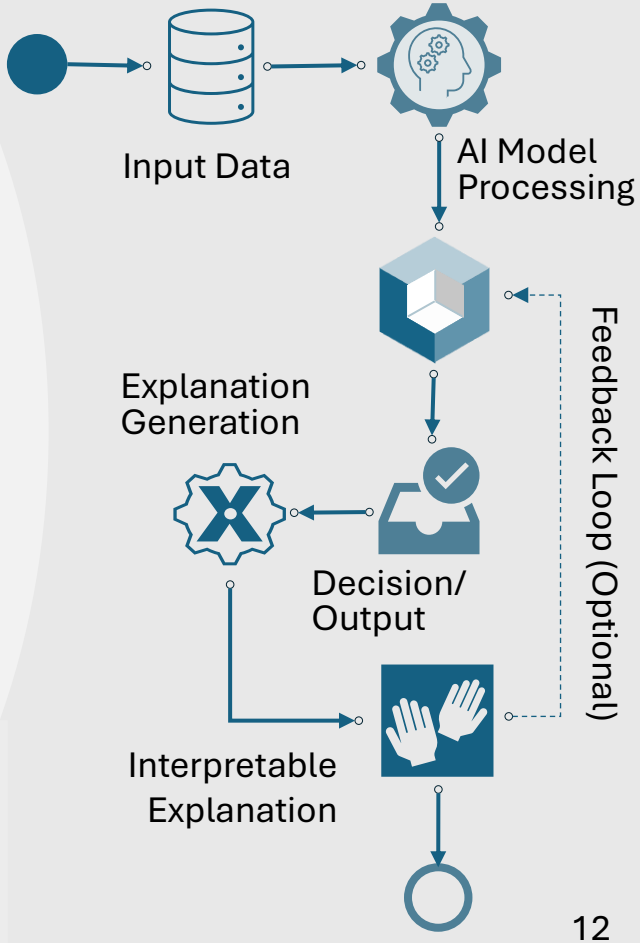
In contrast, model-specific methods are tailored to specific types of models. They leverage the internal mechanics of the models they are designed to explain, offering deeper insights into the model's decision-making process. However, their applicability is limited to certain model types

Visualization-Based Approaches

Visualization tools and techniques play a crucial role in XAI by providing intuitive and accessible explanations. They transform complex model outputs into visual formats that are easier to understand, making them an invaluable resource for both technical and non-technical stakeholders.

POPULAR XAI METHODS			
LIME (Local Interpretable Model-Agnostic Explanations)	Generates local explanations by using simple models to approximate the predictions of the AI model in the vicinity of a particular point.	LRP (Layer-wise Relevance Propagation)	Assigns relevance scores to input features, reflecting their contribution to the model's prediction.
SHAP (SHapley Additive Explanations)	uses Shapley's game to compute the contribution of each input element to the model's prediction	GAM (Generalized Additive Model)	Represents the AI model as a sum of simple functions, each of which explains the dependence of the prediction on one of the input parameters.
Counterfactual Explanations	Generates examples of input data that lead to a change in the prediction of the model.	Rationalization	Rationalization uses inference techniques to create rule-based explanations.

XAI Process



FEDERATED LEARNING (FL)

Federated Learning (FL) is a collaborative machine learning approach that trains models across multiple decentralized devices or servers holding local data samples, without exchanging them.
This approach enhances data privacy, reduces latency, and leverages distributed data.

Horizontal Federated Learning:

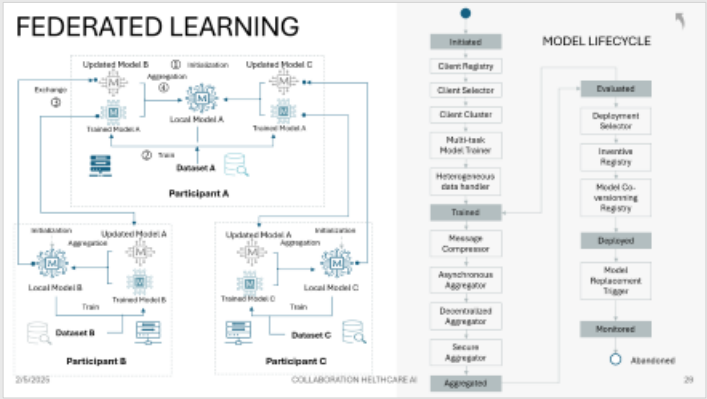
Combines data from the same feature space but different samples, typically used when data from different sources share the same structure.

Federated Transfer Learning:

Applies transfer learning techniques in federated settings, useful when data from different organizations have different features and samples.

Vertical Federated Learning:

Combines data from different feature spaces but shares the same sample IDs, useful when organizations have different attributes of the same user.



DRIVERS:

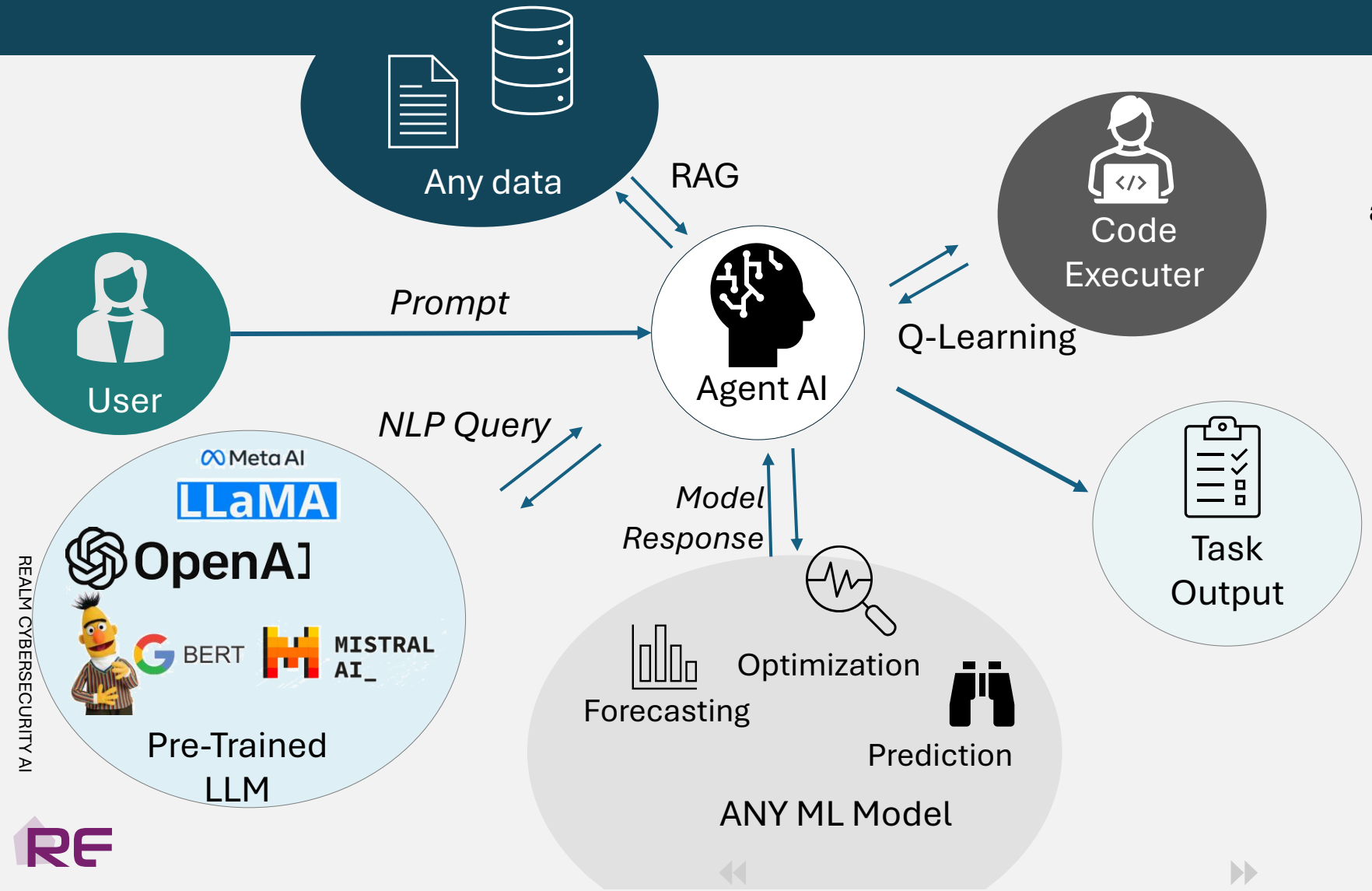
Data Privacy: Enhances privacy by keeping data on local devices and only sharing model updates.
Reduced Latency: Improves training efficiency by leveraging local computation.
Utilization of Diverse Data: Enables the use of data from various sources that cannot be centrally aggregated due to privacy concerns or data regulations.

SERVICE:

- **Federated Data Analytics**: Performing data analytics on distributed data sources without compromising privacy.
- **Personalized Model Training**: Developing personalized models for individual users or devices without sharing their data.
- **Cross-Organization Collaborations**: Enabling multiple organizations to collaboratively train models while maintaining data confidentiality.
- **Edge AI Services**: Deploying AI models on edge devices like smartphones and IoT devices for real-time, on-device inference.

AYITA: AGENT AI

Agent AI represents an advanced AI framework where an intelligent agent autonomously interacts with various services, extracts data, communicates with external large models, executes code, and leverages machine learning models to perform tasks efficiently.



The agent approach to AI represents the future due to its ability to operate autonomously, efficiently, and adaptively, enhancing user interactions through natural language processing and contextual understanding. These AI agents seamlessly integrate with various systems, providing scalable and versatile solutions across multiple industries. They handle complex tasks, drive efficiency, and improve data privacy through techniques like federated learning and differential privacy. This makes AI agents a powerful tool for transforming industries and automating intricate processes, paving the way for a more intelligent and interconnected technological ecosystem.

REALM HOLISTIC APPROACH

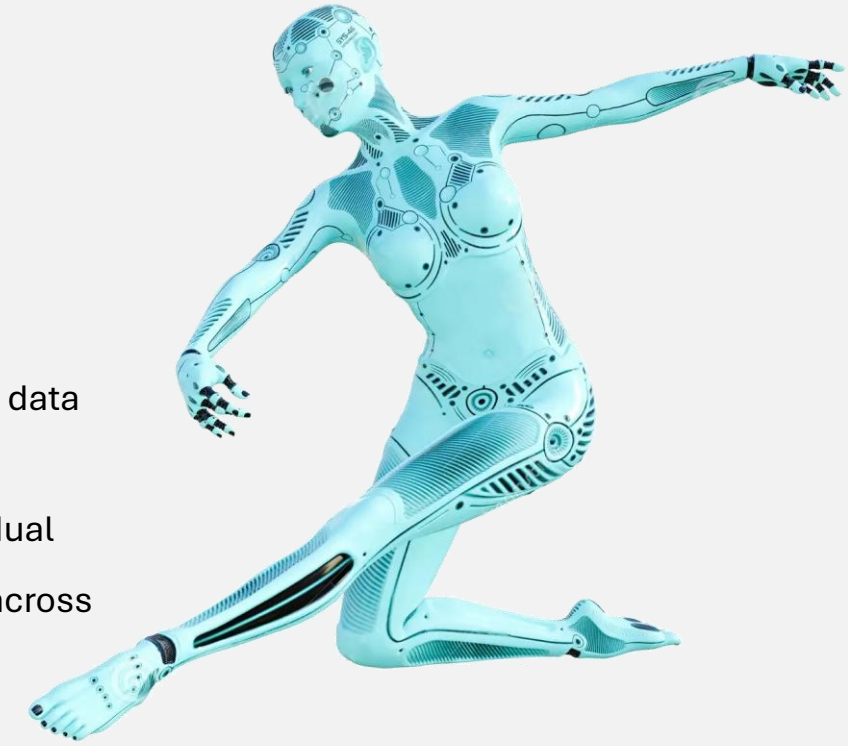
REALM is a comprehensive AI platform tailored for finance and healthcare, providing secure and efficient AI-driven solutions to democratize AI access, enhance data security, and improve decision-making processes through advanced AI technologies.

FIRST AI CYBERSECURITY PLATFORM:

Data Security and Privacy: Ensures stringent data protection and compliance with privacy regulations.
Scalability: Supports scalable AI solutions that can grow with business needs.
Efficiency: Streamlines and automates critical business processes to enhance productivity.
Customization: Offers customizable AI solutions tailored to specific industry needs.

KEY FEATURES:

- **ADVANCED AI CAPABILITIES**
  - **Cyber AI Algorithms**: Utilizes state-of-the-art AI algorithms to tackle complex problems in finance and healthcare.
  - **Generative AI**: Implements Generative Adversarial Networks (GANs) for creating synthetic data and enhancing data privacy.
- **SECURITY AND COMPLIANCE**
  - **Differential Privacy**: Ensures data privacy by adding noise to datasets, maintaining individual data confidentiality.
  - **Federated Learning**: Enables collaborative model training without sharing sensitive data across entities.
- **INTEGRATION AND INTEROPERABILITY**
  - **API Management**: Provides robust APIs for seamless integration with existing systems.
  - **Workflow Orchestration**: Automates workflows and integrates various AI and data processing tasks.



THANK YOU


PERSONAL
DATA
ANONYLZIZATION
TRUST
DIFFERENTIAL PRIVACY

APPENDIX

SYNTHETIC DATA ATTACKS

Attack Type	Description	Consequence
Model Inversion Attack	Adversaries use the model's output to infer sensitive input data.	Exposure of sensitive information, compromising individual privacy.
Membership Inference Attack	Attackers determine if specific data was part of the model's training set.	Potential identification of individual data contributions, leading to privacy breaches.
Data Poisoning	Malicious data is introduced into the training set, affecting learning.	Compromised model integrity, leading to skewed or harmful outputs.
Adversarial Manipulation	Deceptive data input exploits model vulnerabilities, causing wrong outputs.	Eroded trust in model accuracy and potential manipulation for nefarious purposes.
Model Stealing/Extraction	Reverse-engineering a model to replicate its functionality and data.	Unauthorized access and potential misuse of proprietary algorithms and data insights.
Re-identification Attack	Cross-referencing anonymized data with external sources to identify individuals.	Violation of anonymity guarantees, leading to privacy invasions and potential legal ramifications.
Attribute Inference Attack	Using model outputs to infer sensitive attributes of individuals in the dataset.	Exposure of sensitive attributes, leading to privacy breaches and potential misuse of data.

PXP FRAMEWORK

THREE TYPES OF DATA ARE USED:

- Numeric tabular data
- Semi-structured text data
- Time-series

TABULAR DATA

- IDs
- Social Characteristics
- Financial performance

REAL DATASET
PSEUDO-REAL DATA

THE PREPARATION OF HIGH-QUALITY INITIAL DATA ALLOWS YOU TO CONDUCT THE **GENERATION**CORRECTLY

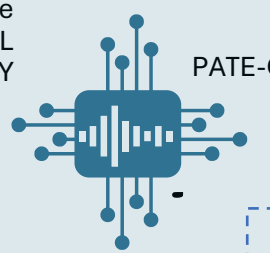
SEMI-STRUCTIRED

Text data

- Financial Attributes
- IDs
- Digital Fingerprint

Используется трансформеры с дифференциальной приватностью

Special architecture with DIFFERENTIAL PRIVACY



PATE-GAN

CODING

DECODING

CALCULATING METRICS

GAN

GAN with support for transient dependencies

WGAN, CT-GAN

Risk Measures already at the generation stage

ATTRIBUTE EXTRACTION

The data contains identifiers and characteristic "stamps"

Learning with the TRANSFORMER Architecture

Training a transformer is computationally intensive and time-consuming

Differential Privacy Core

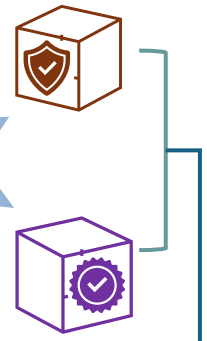
Statistical Tests

Pearson correlation

AUROC, AUPRC

QUALITY MODEL

SYNTHETIC DATASET



RISK MODEL

DIVERGENT MODELS

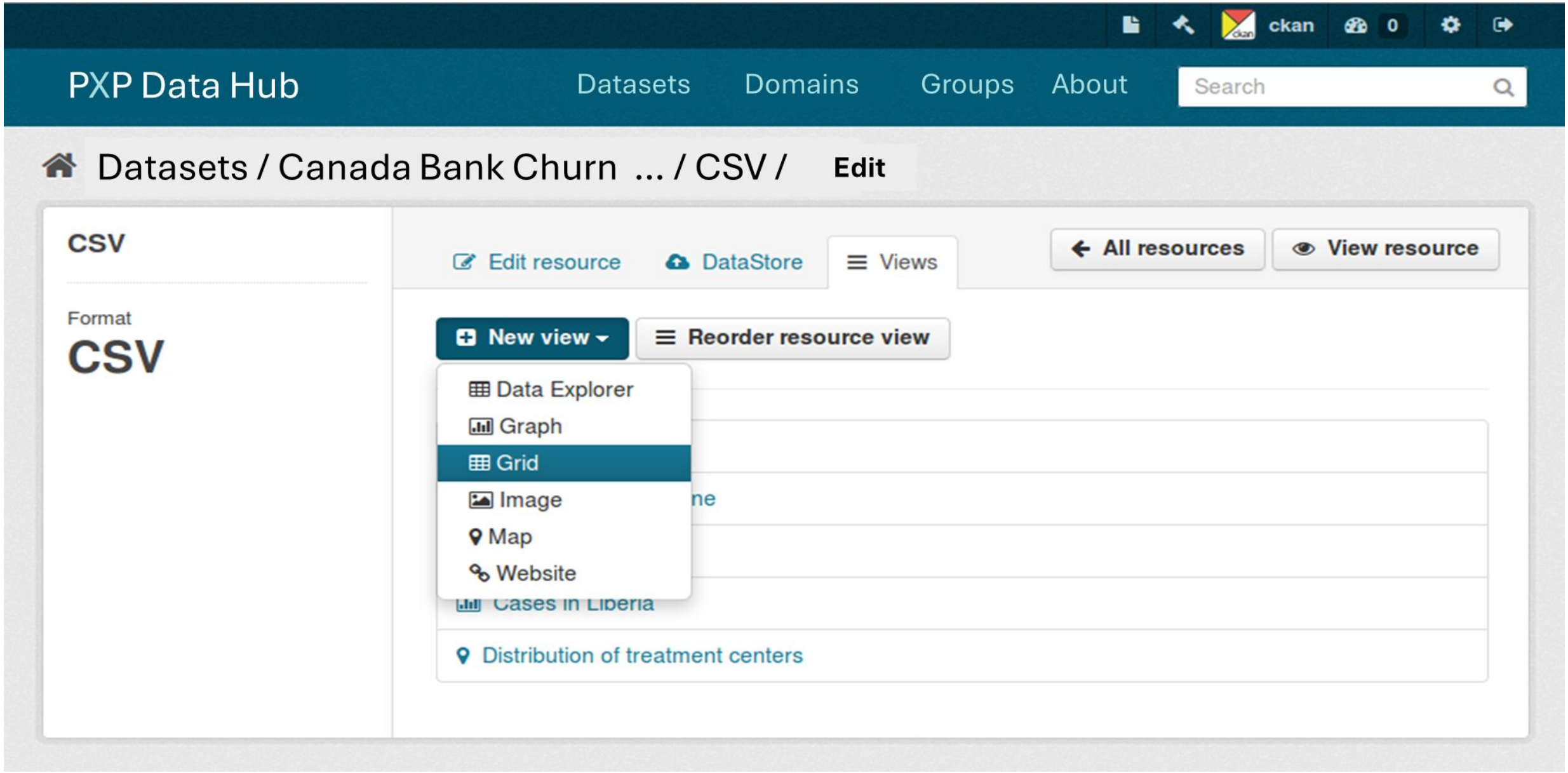
INFORMATION LEAK

WITH DIFFERENTIAL PRIVACY

BASED ON EMBEDDING

PXP PORTAL
Powered By ckan

PXP Portal Data Hub (powered by CKAN)

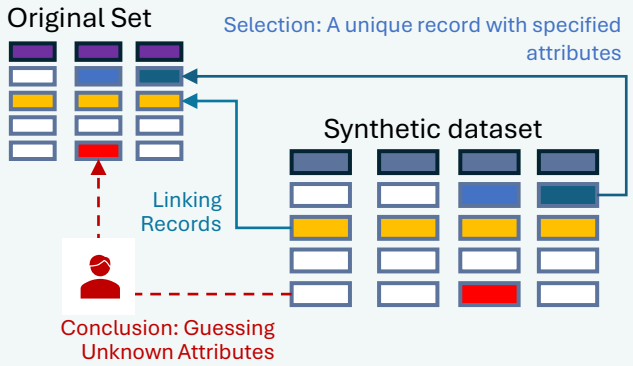


INFORMATION LEAKAGE MODEL

The model evaluates the risk of leakage of confidential information from synthetic data, as well as the probability of identifying or recovering original data from the synthetic dataset. Leakage refers to the ability to extract sensitive information about real data or individuals from synthetic data possible.

The risk of information leakage can be assessed by considering three key aspects(Giomi Matteo, 2022):

- **Singling Out:** An estimate of the probability that it can be determined whether a unique record exists in the source dataset with a specific combination of attributes**.**
- **Linkability risk** refers to the ability to link records belonging to the same person or group of individuals in the source and synthetic set.
- **Inference:** The ability to guess unknown attributes of the original data record from synthetic data.



General Risk Equation: $R_{total} = w_1 \times R_{snglout} + w_2 \times R_{link} + w_3 \times R_{inf}$

Here w_i – weights which can be taken to be equal to a first approximation($w_i \approx 0.3333$)

Each of the 3 contributions will be evaluated on the basis of the Wilson Score Interval, a statistical method for determining the confidence interval of the proportion in the binomial distribution of WI:

$$WI = \left[\frac{\hat{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}}}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}}, \frac{\hat{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}}}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}} \right]$$

Here:

- $\hat{p}$ — observed Sample Proportion
- n — sample size
- $z_{\alpha/2}$ –Standard Distribution Quantile for Confidence Level (1-α): Corresponds to a point on the standard normal curve such that the area under the curve up to that point corresponds to the desired confidence level.

The Wilsonian confidence interval provides a range of values in which the true value of risk is expected to be found with a given degree of confidence, and for most problems it is acceptable to choose a midpoint.

For calculations $R_{snglout}$ and R_{link} the proportion of success is chosen naturally as a unique number of entries of the synthetic set matched to the original set (singling out) or external set (Linkability). For the risk of inference, the reduced entropy can be taken:

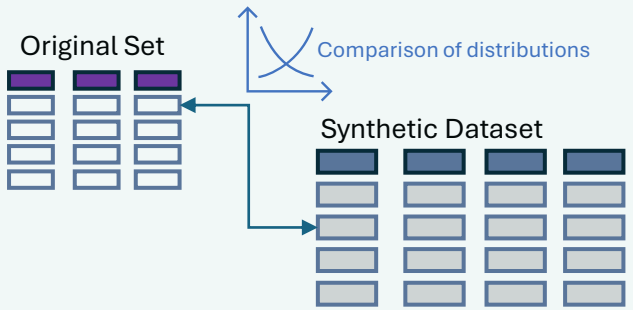
$$\hat{p} = NE(a_j) = \frac{H(a_j)}{\log_2 4} = -\frac{1}{2}\sum_{i=1}^{n} p_i \log_2 p_i$$

DIVERGENT MODELS

Divergent models are based on the idea of estimating the differences between two distributions of data – the original dataset and the generated synthetic set. These models help to quantify how closely synthetic data reproduce the statistical characteristics of the original data, as well as to identify potential information leaks.

Divergent Model Limitation

- **Outlier Sensitivity**: Divergent models may be overly sensitive to outliers, leading to skewed representations or analyses. These models might overemphasize or underrepresent the impact of data points that significantly deviate from the majority of the dataset.
- **Dependency Complexity**: Divergent models often focus on capturing linear relationships between variables, potentially overlooking the more complex, nonlinear interactions. This limitation can result in a partial or superficial understanding of the underlying data dynamics.
- **Interpretation Challenges**: The results produced by divergent models can be intricate and subtle. Interpreting these results correctly requires a nuanced understanding of the model's behavior and the specific context of the data, making it a challenging task that demands expertise and careful consideration.



Apply divergence metrics, such as the Kullback-Leibler divergence (KL divergence) or the Jensen-Shannon divergence (JS divergence), to quantify the differences between distributions.

The normalized **Kullback-Leibler distance** is calculated as the ratio of the KL distance to the maximum possible KL value:

$$\widehat{D}_{KL}(P||Q) = \frac{\sum_{i=1}^{n} P_i \log_2\left(\frac{P_i}{Q_i}\right)}{\max(D_{KL})}$$

Here:

- P,Q – Probability Distribution for Real and Synthetic Data
- $max\ (D_{KL})$ chosen for theoretical or practical reasons, equivalent to the maximum risk threshold.

Normalized **Euclidean distance** between sets:

$$d_{NE}(u,v) = \sqrt{\sum_{i=1}^{n}\left(\frac{u_i}{\|u\|} - \frac{v_i}{\|v\|}\right)^2}$$

Here:

- u, v are vectors representing the original and synthetic sets
- $\|\cdot\|$ is Euclidean norm

The normalized Euclidean distance lies in the range $(0; \sqrt{2})$, Where 0 means the identity of vectors, and $\sqrt{2}$ is its orthogonality. In this regard, the risk can be defined as $R = 1 - \frac{d_{NE}}{\sqrt{2}}$

The **Jaccard Index** also measures the proximity between two sets based on the ratio of their intersection to the union:

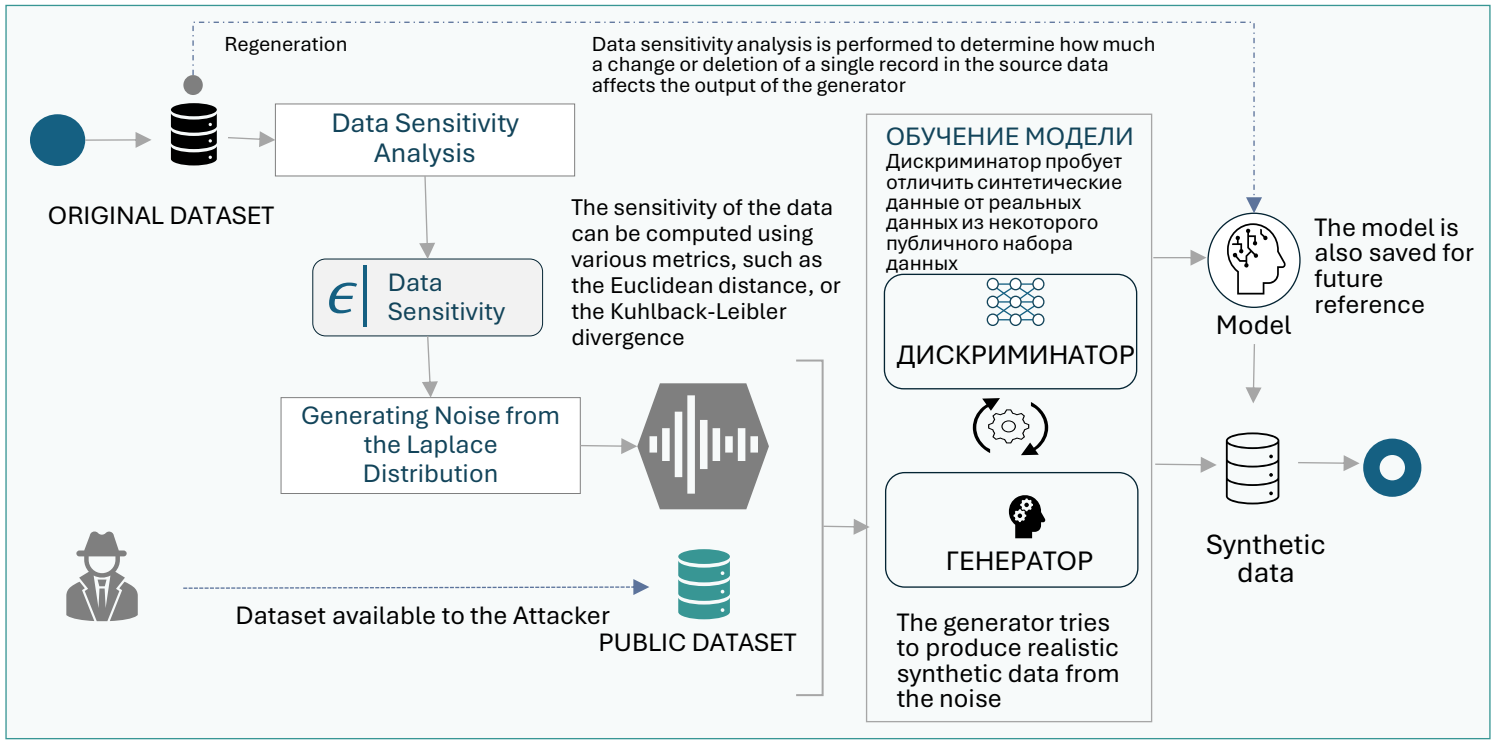
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Here $|\cdot|$ is cardinal dataset Number.

DIFFERENTIAL PRIVACY

The differential privacy model is an approach to protecting the privacy of individual data in a data set by allowing data analysis to be conducted without revealing specific information about individual individuals.
A brief description of the main aspects of the differential privacy model:

- **Differential privacy** is a formal definition of privacy that ensures that the addition or removal of a single item from a data set will not have a significant impact on the results of the data analysis.
- **Confidentiality**: Differential privacy techniques provide a mechanism whereby conclusions drawn from data do not reveal sensitive information about individuals, making the results of the analysis virtually indistinguishable, regardless of the presence or absence of a specific record in the data.
- **Noise mechanisms**. Noise-adding mechanisms are often used to achieve differential privacy. These can be a variety of methods, including adding Laplace or Gaussian noise to the results of data queries.
- **Privacy budget**. The differential privacy model uses the concept of a "privacy budget," commonly referred to as a ε (epsilon). A low ε value corresponds to a higher level of privacy, but it can reduce the accuracy of the analysis results.



The risk value can be calculated in this model:

$$R_{DP} = e^{\epsilon}$$

Here, ε is a differential privacy parameter. The formula assumes that the synthetic data is generated by a differentially closed algorithm that ensures that the output does not change materially if any particular record in the source data is changed or deleted. The formula also assumes that the attacker has unlimited basic knowledge and supporting information, and that the attacker can perform any type of attack on the privacy of synthetic data.

When it comes to generating synthetic data using AI, differential privacy is one of the most effective approaches to ensuring data privacy.

RISK MODEL FOR SEMI-STRUCTURED DATA

1. **Vectorize text with FastText, Word2Vec, or GloVe:** These models create vector representations of words by learning on large corpora of texts and capturing the semantic relationships between words. Word2Vec uses contextual words to predict the current word (CBOW) or the current word to predict its context (Skip-gram), while GloVe builds a word co-occurrence matrix and factorizes it.

   Word2Vec uses contextual words to predict the current word (CBOW) or the current word to predict its context (Skip-gram), while GloVe builds a matrix of word occurrence and factorizes it.

$$w_i^T w_j + b_i + b_j = \log(X_{ij})$$

Here X_{ij} - element of the word co-occurrence matrix, w_i ,w_j is Word Vectors, b_i, b_j - Offsets for words.

```
import fasttext
import fasttext.util
ft = fasttext.load_model('cc.en.300.bin')
word = 'computer'
word_vector = ft.get_word_vector(word)
print(f"Vector representation of a word'{word}':\n{word_vector}")
```

2. Retrieving the Ad Vector: Averaging or summing word vectors: Convert an ad to a vector by averaging or adding the vector representations of all the words in the ad.

$$V_{ad} = \frac{1}{n}\sum_{i=1}^{n} v_i$$

3. Integration of additional attributes: Explicit attributes (price, area, etc.) are converted into numerical vectors using techniques such as one-hot encoding, scaling, or embedding. Vectors are concatenated

$$V_{final} = [V_{ad}; V_{attr}]$$

4. Comparing Vectors Using Cosine Similarity: calculate the cosine similarity

$$cosine_{similarity}(V^A, V^B) = \frac{V^A \cdot V^B}{\|V^A\| \, \|V^B\|}$$

can be conveniently calculated using the scikit-learn library

Cosine similarity immediately provides a score for a risk for which a threshold can be set. For example, if you want to consider vectors similar when the cosine similarity is greater than or equal to 0.7 (which corresponds to an angle of about 45 degrees or less), you can set such a threshold. This means that the closer the cosine similarity value is to 1, the smaller the angle between the vectors and the greater their similarity.

Reliable cleaner / cleaning services / airbnb cleaner

150 $ per service

Location
GTA
Richmond Hill
Barrie
Orillia.

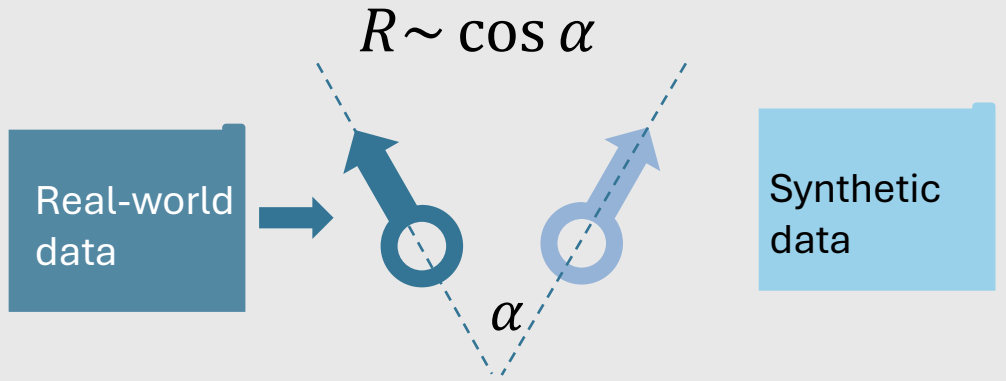
Contact Karla Garcia

☎ : 647--490—5XXX

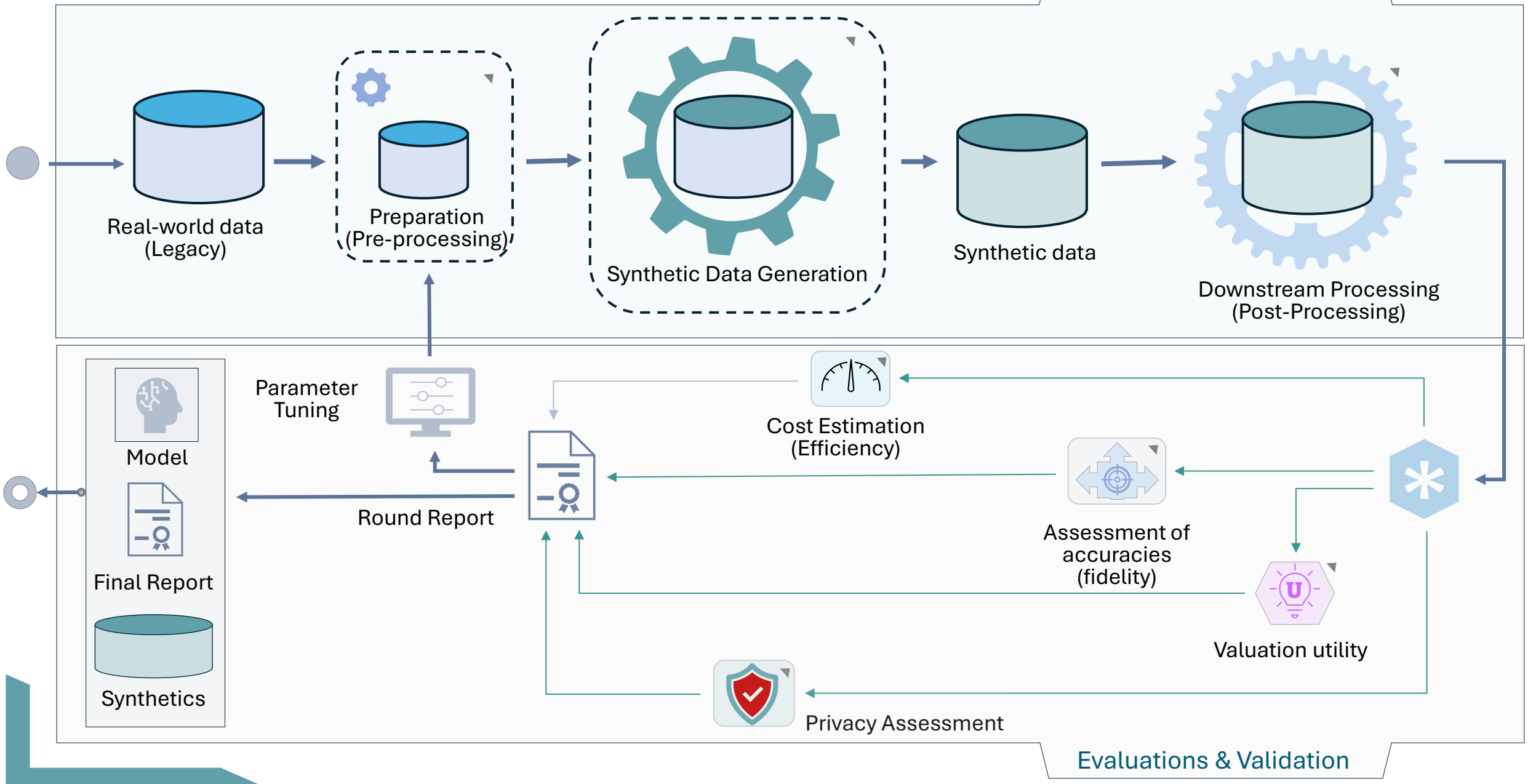
Details

We specialize in:
Deep Cleaning Post Construction
Renovation Airbnb cleaning and Hosting Residential Cleaning

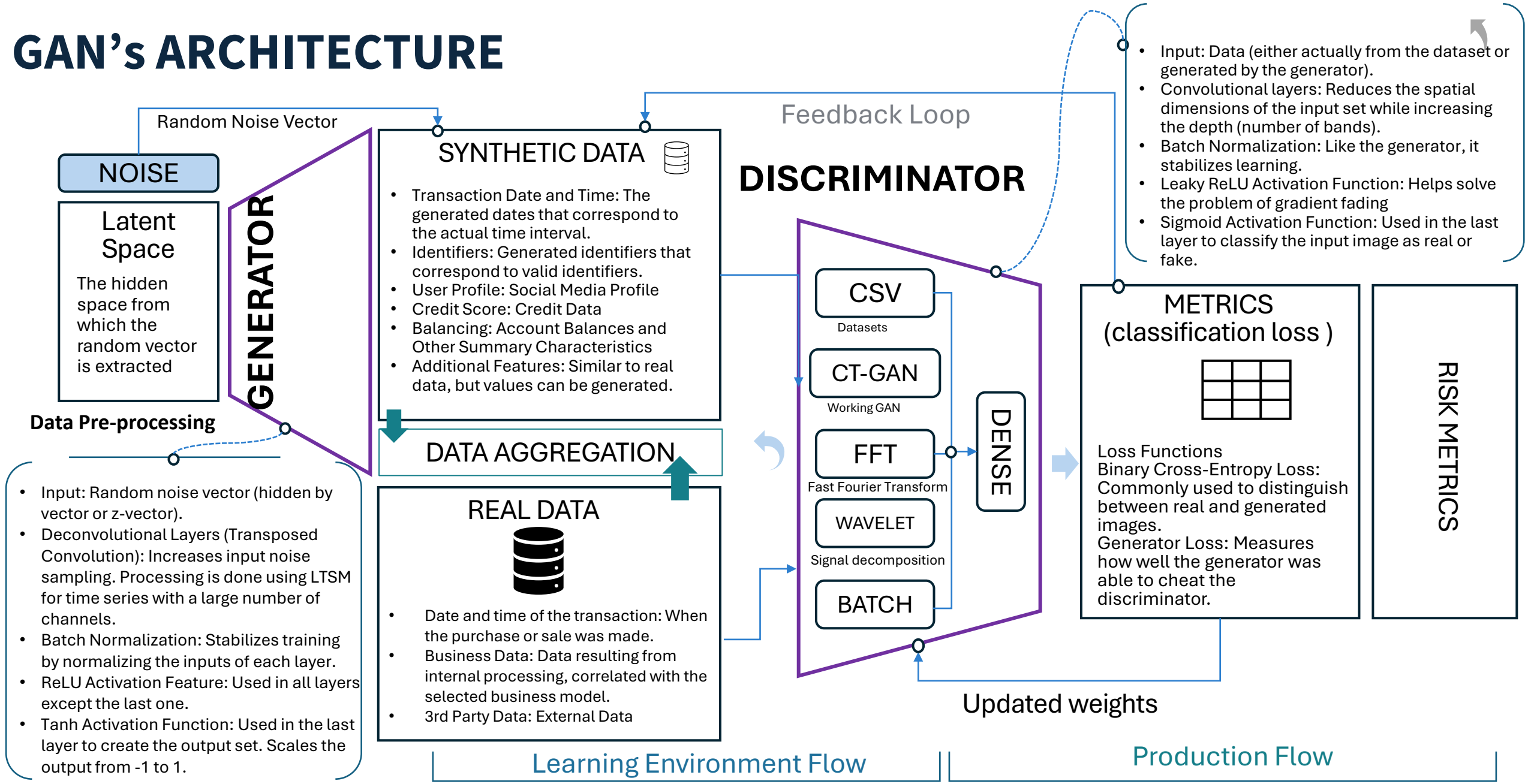
Shine cleaning services we have Pleasure to service in Aurora ,Richmond Hill,Newmarket,Bradford,Innisfil and barrie more the 10 Years expirence with are family businness with love to taken care the most precious thing people have your cozy and beautiful homes With services businnes,comercial,industrial,move out move in ,school,daycare,dentist,arquitect etc Please fill free to call as for free estimated anything with are Placer to services everyone.



GENERATION LIFE CYCLE

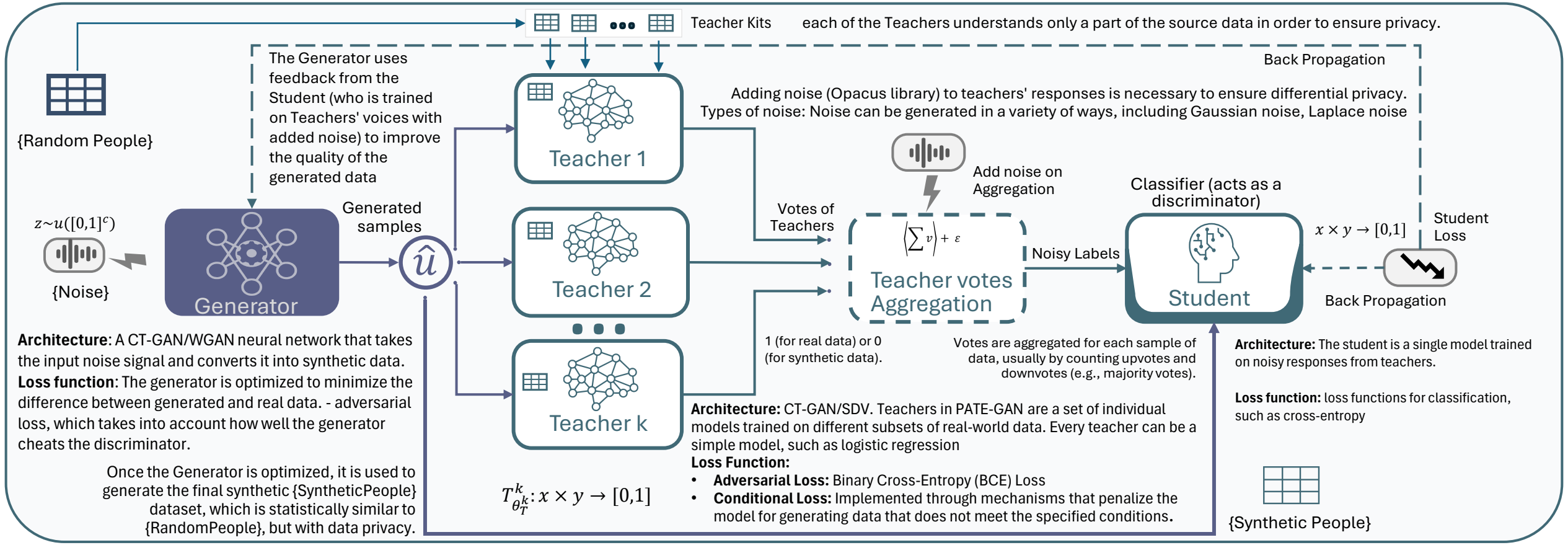


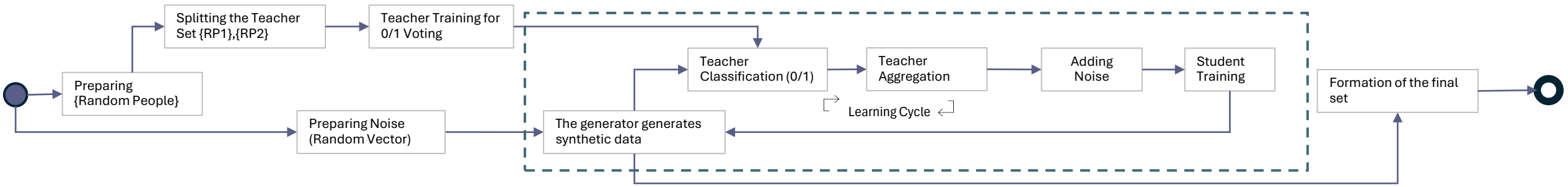
GAN's ARCHITECTURE



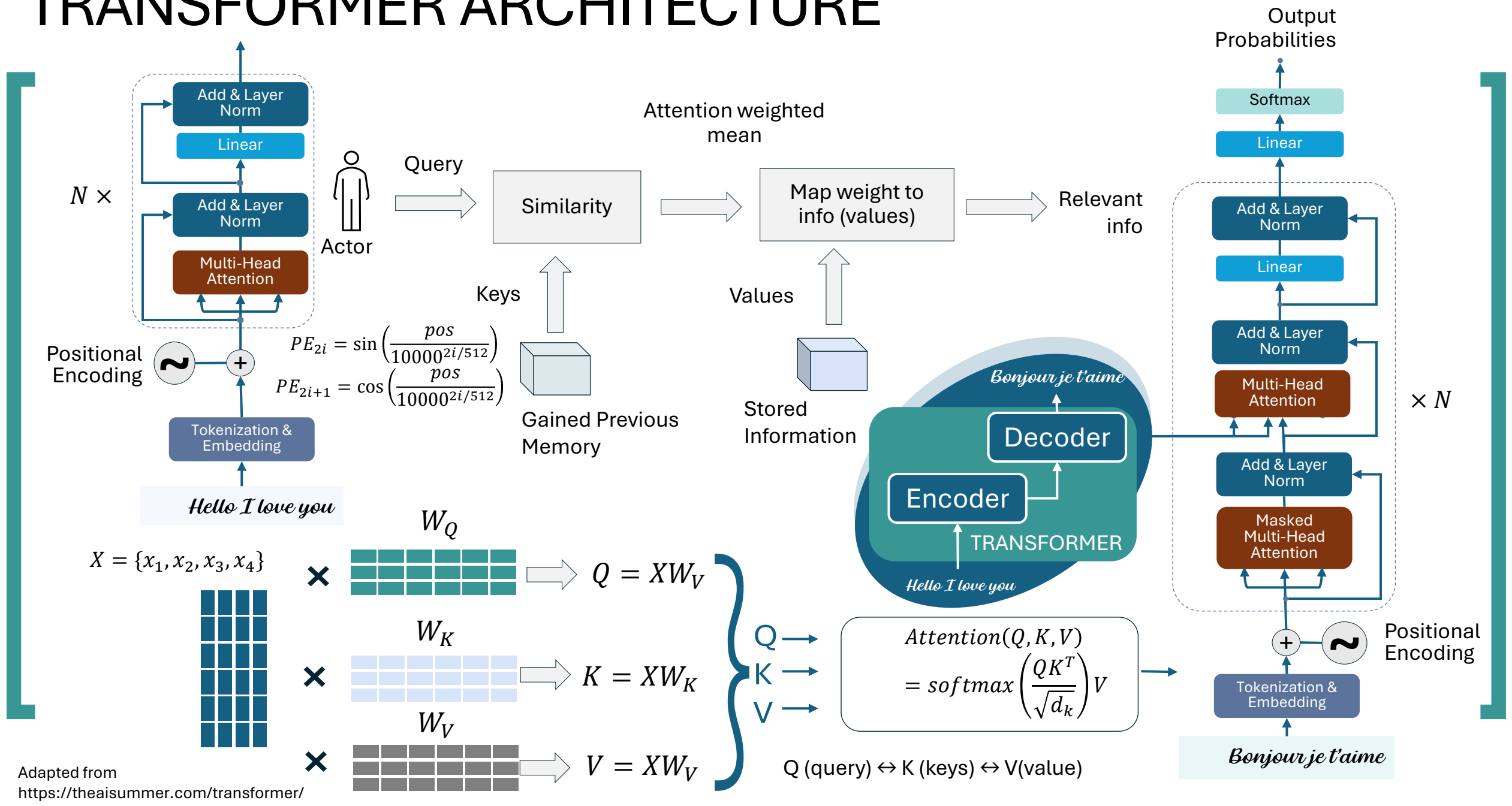
PATE-GAN ARCHITECTURE

PATE-GAN (Private Aggregation of Teacher Ensembles - Generative Adversarial Network) is a method that combines the principles of differential privacy and generative adversarial networks (GANs). The main goal of PATE-GAN is to generate synthetic data that is statistically similar to real data, while ensuring the protection of sensitive information contained in this data.



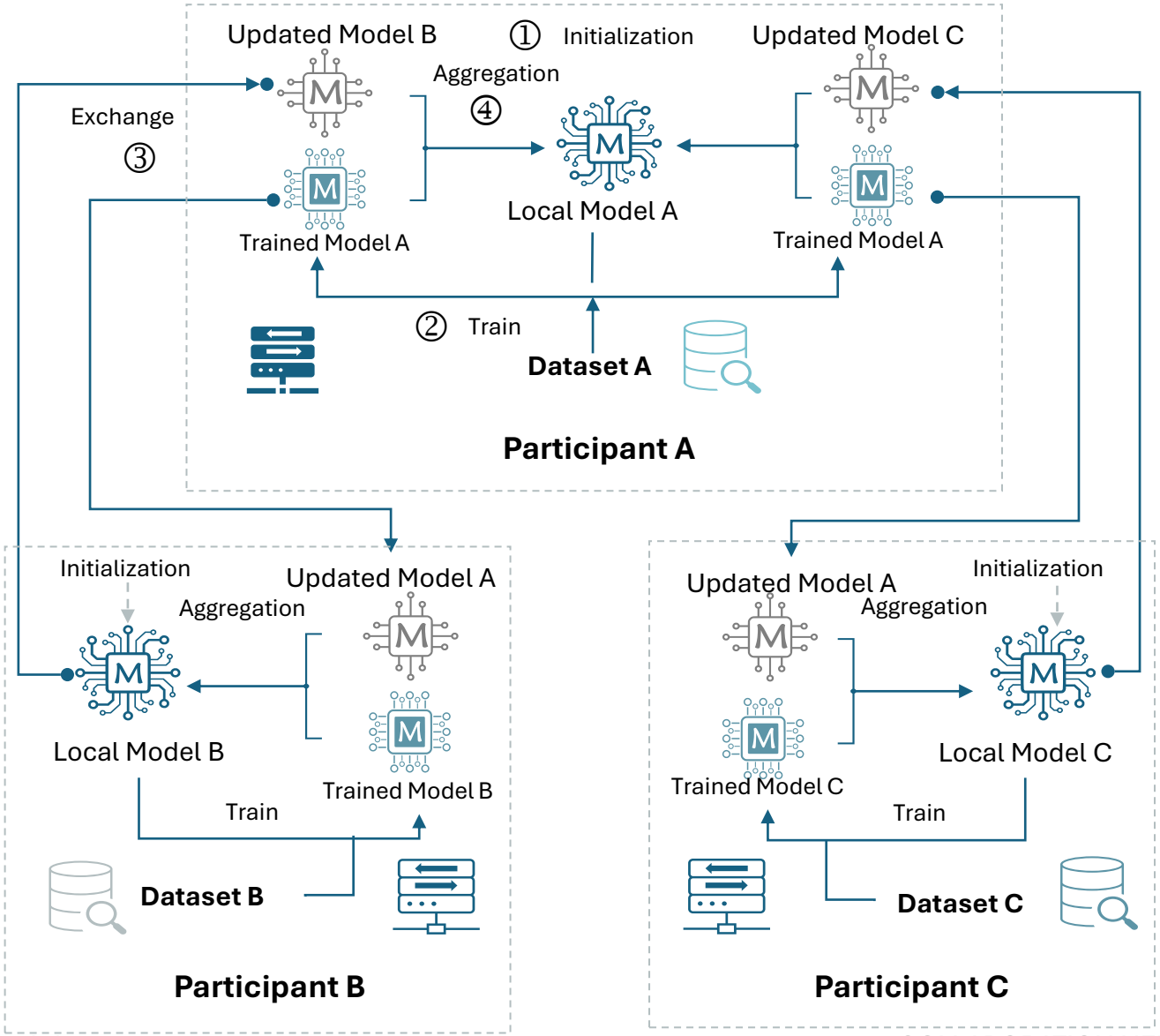


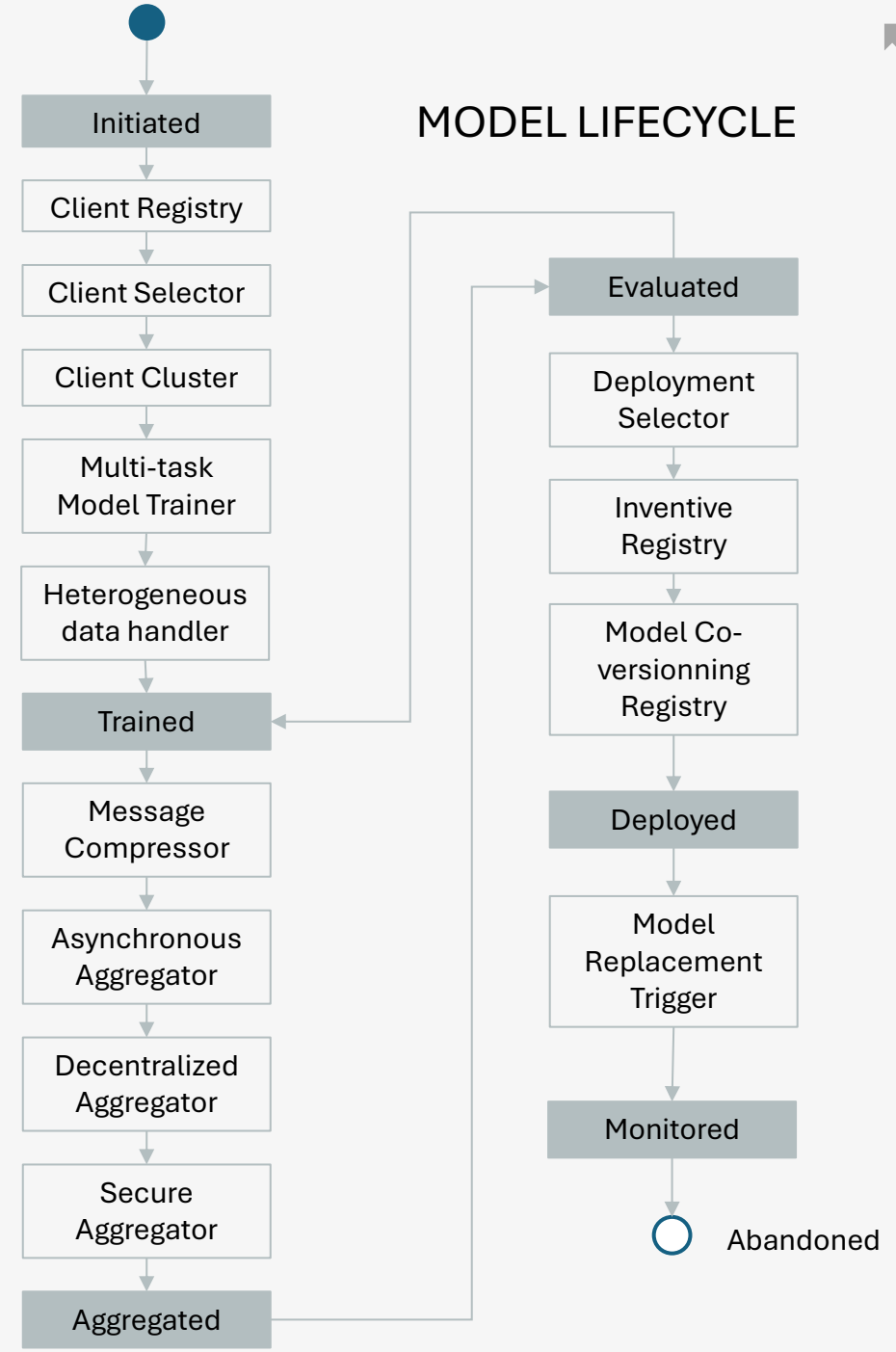
TRANSFORMER ARCHITECTURE



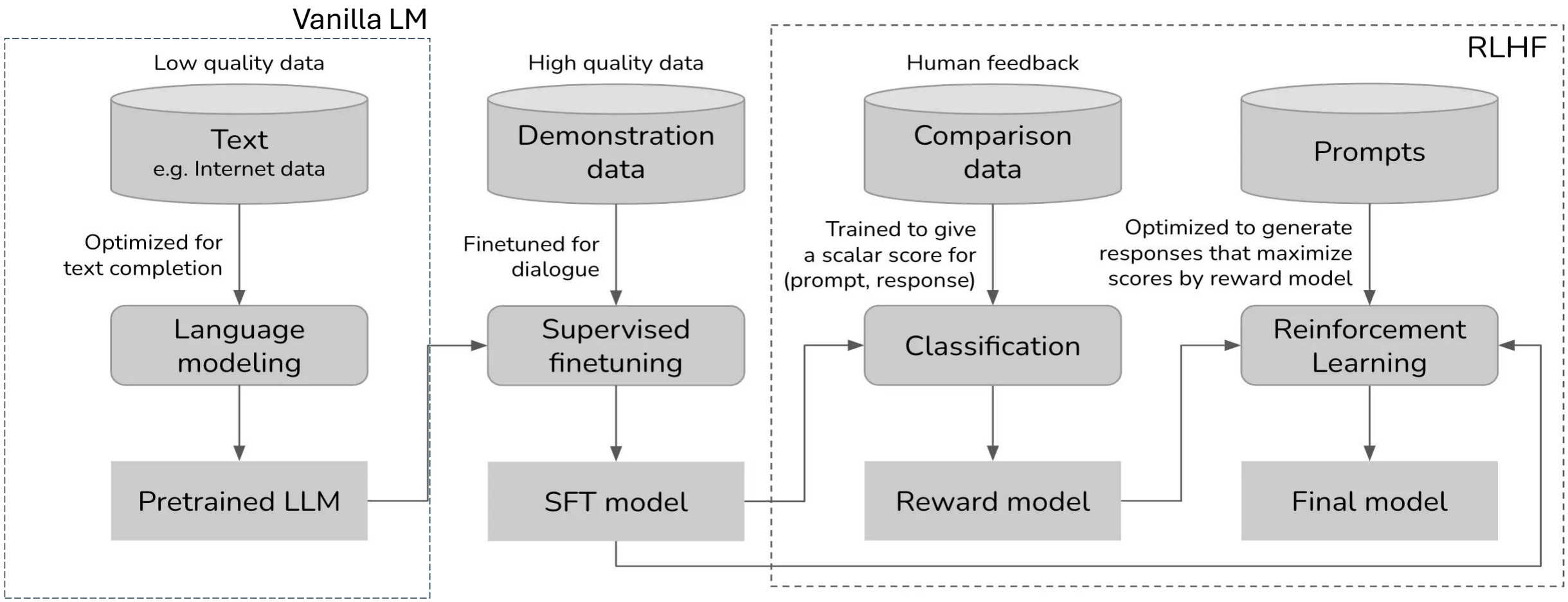
Adapted from
https://theaisummer.com/transformer/

FEDERATED LEARNING





RLHF FINE TUNING

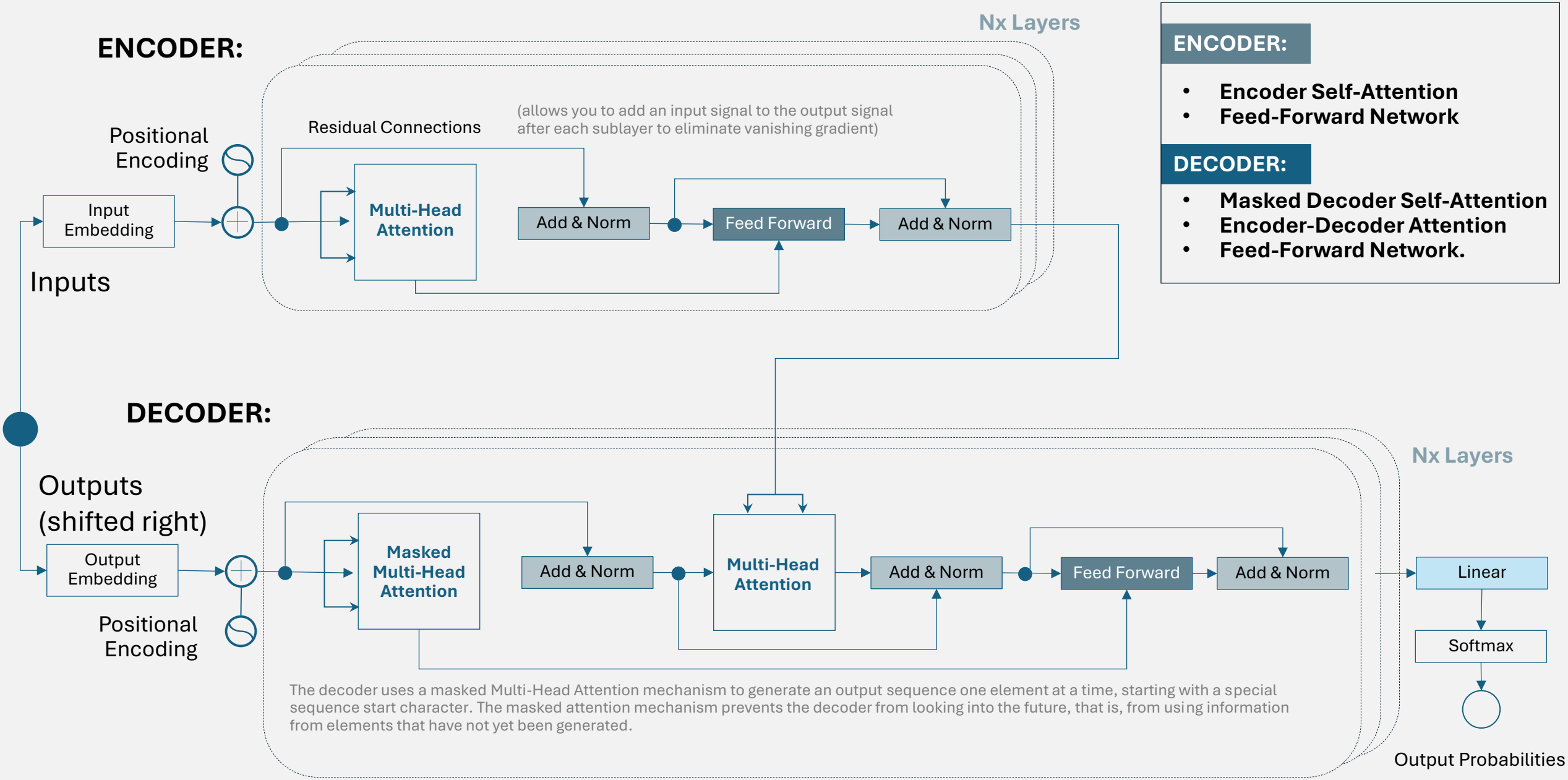


Scale May '23	**>1 trillion** tokens	**10K - 100K** (prompt, response)	**100K - 1M** comparisons (prompt, winning_response, losing_response)	**10K - 100K** prompts
Examples **Bolded**: open sourced	GPT-x, Gopher, **Falcon**, LLaMa, **Pythia**, **Bloom**, **StableLM**	**Dolly-v2, Falcon-Instruct**		InstructGPT, ChatGPT, Claude, **StableVicuna**

FINE TUNING APPROACHES

#	Title	Description	Process	Pros	Cons
1	Reinforcement Learning from Human Feedback (RLHF)	Uses human feedback to fine-tune models for nuanced tasks	Model is fine-tuned based on feedback from human evaluators who rank outputs	High accuracy and performance on subjective tasks	Time-consuming and resource-intensive
2	Low-Rank Adaptation (LoRA)	Introduces low-rank matrices to model parameters for efficient fine-tuning	Adds low-rank adaptations to the model's weights, enabling significant changes with minimal computation	Efficient, requires fewer resources	May not be as effective for highly complex tasks
3	Feature-Based Fine-Tuning	Adjusts feature extraction layers while keeping other parts unchanged	Fine-tunes only the last few layers or adds additional layers for specific tasks	Preserves general knowledge, less computationally intensive	Limited to tasks closely related to the pre-trained model's original purpose
4	Layer-Wise Freezing	Gradually unfreezes layers during fine-tuning	Starts by training only newly added layers, progressively unfreezes more layers	Fine-grained control over training process, reduces overfitting	More complex training process
5	Multi-Task Learning	Fine-tunes a model for multiple tasks simultaneously	Trains on a combined dataset including examples from multiple tasks	Improves generalization, leverages shared representations	Requires diverse and well-labeled multi-task dataset

TRANSFORMER ARCHITECTURE



ENCODER:

- **Encoder Self-Attention**
- **Feed-Forward Network**

DECODER:

- **Masked Decoder Self-Attention**
- **Encoder-Decoder Attention**
- **Feed-Forward Network.**

Privacy Hub

Scan for sensitive data types, apply generators to protect them, and track your progress in the current workspace, all within the Privacy Hub.

Run Sensitivity Scan | Download Scan Log | Download Privacy Report | Learn more about Privacy Hub →

Completed
4 months ago

Subsetting
In-use

Post Job Actions
Not in-use

At-Risk Columns
8 (~7%)
Review the sensitive columns and protect them with generators.
Open in Database View →

Protected Columns
26 (~23%)
Protected columns are using generators to mask or anonymize your data.
Open in Database View →

Not Sensitive Columns
79 (~70%)
These columns are not flagged as sensitive and are not protected.
+53
Open in Database View →

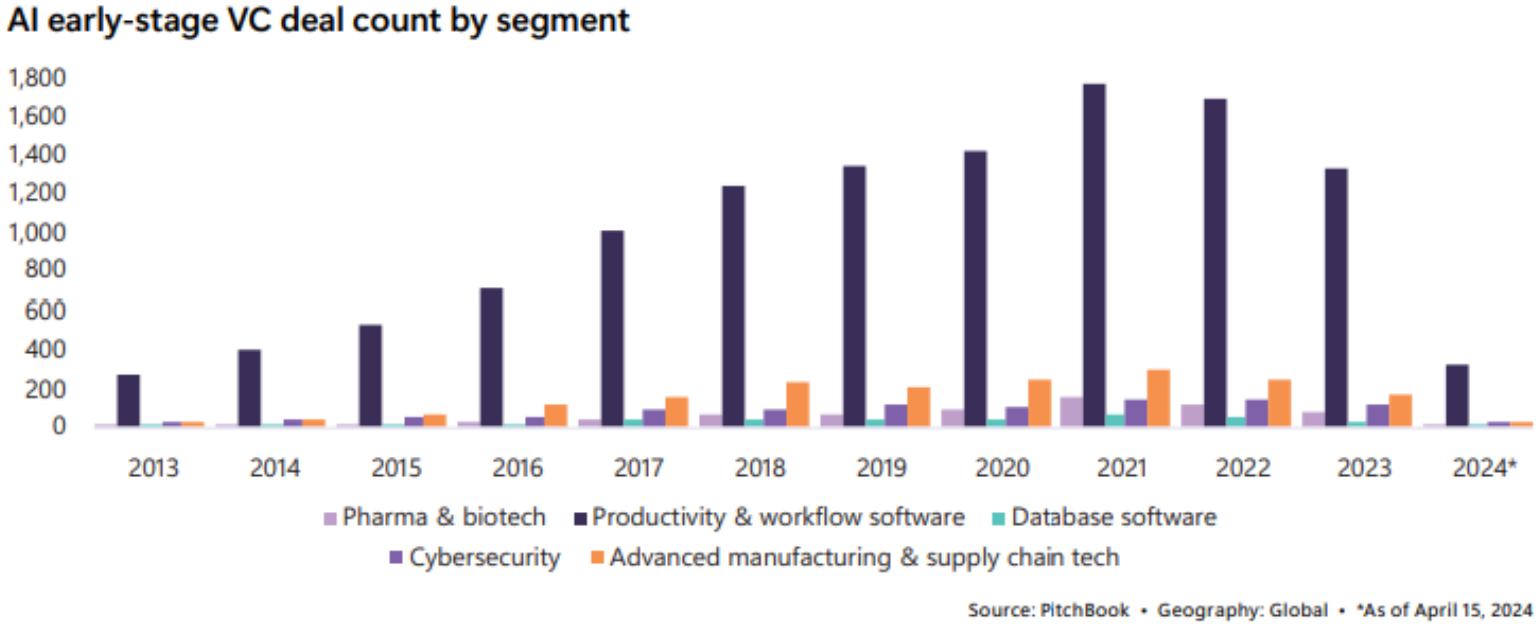
Generation Report

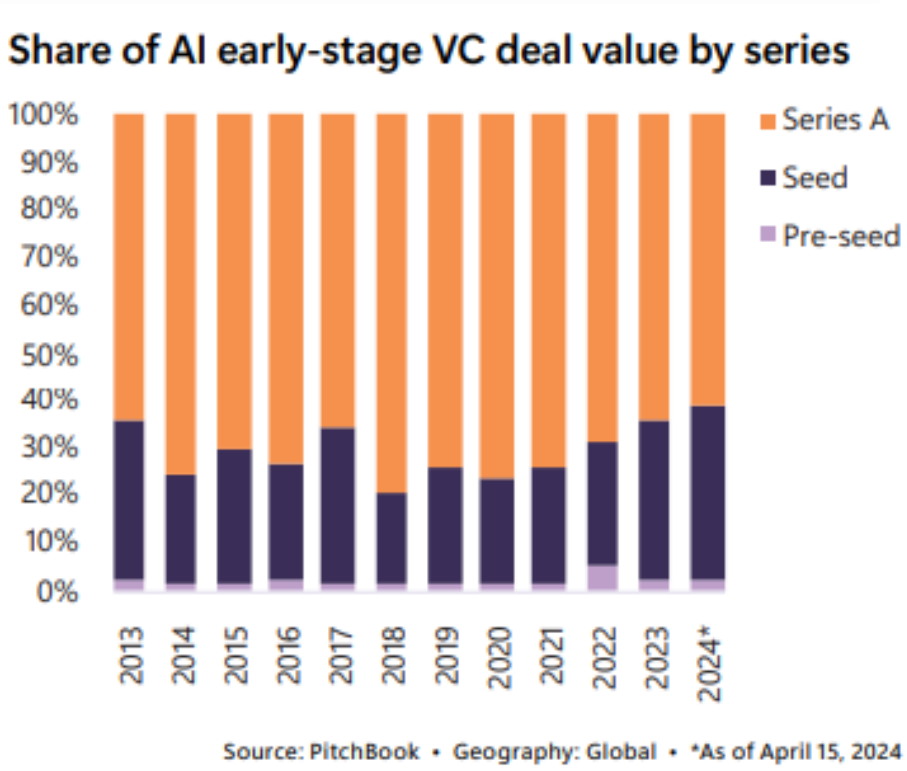
Database Tables

Filter by name... | Filter by Schema

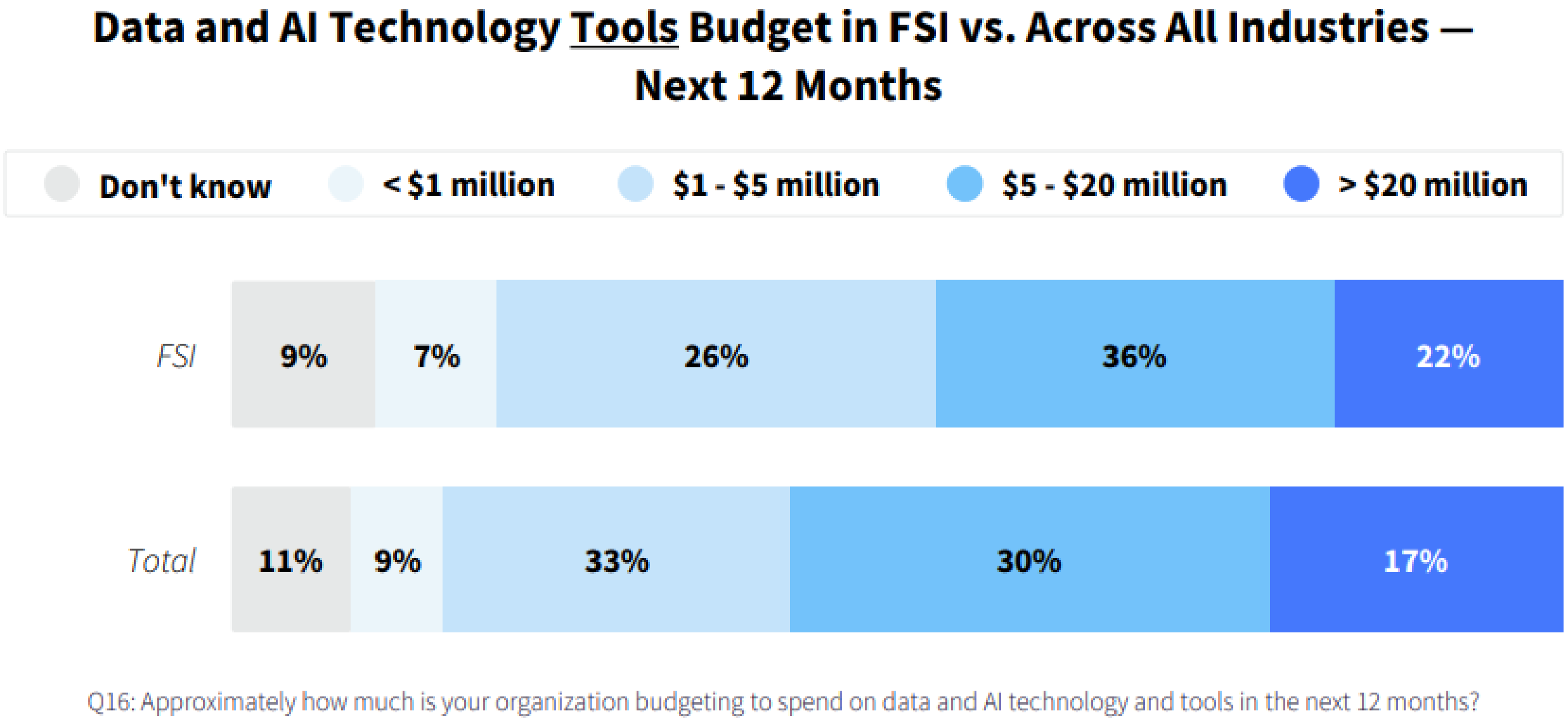
Name	Not Sensitive	Protected	At-Risk	Privacy Status
customers public	7	5	3	
customers_legacy public	10	5	1	
employees public	2	4	1	

MICROSOFT/PITCHBOOK



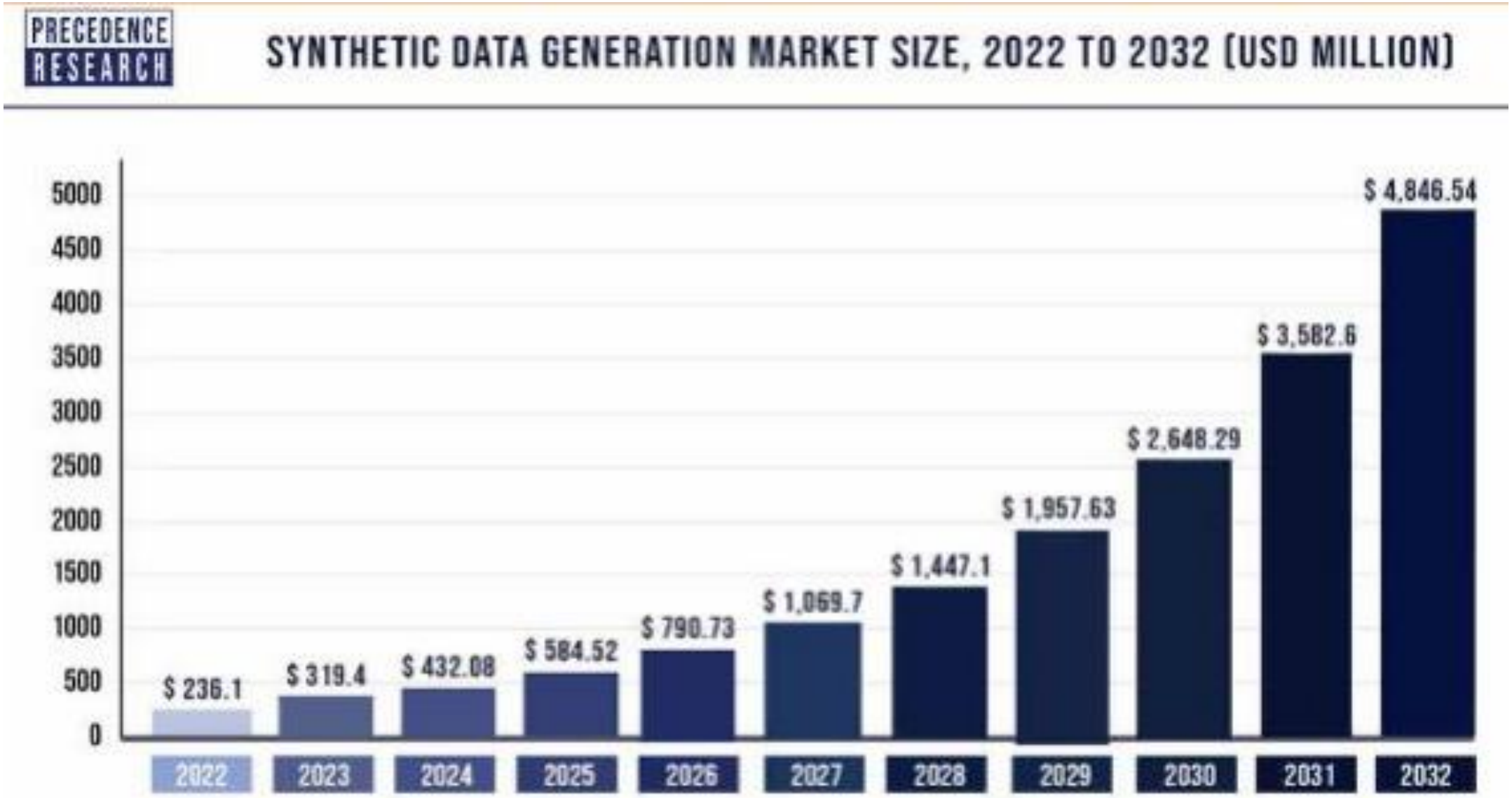


DATAIKU + databricks



Q16: Approximately how much is your organization budgeting to spend on data and AI technology and tools in the next 12 months?

State of AI 2024




AYI


VectorStock
VectorStock.com/28470497


CREATIVEDESIGN
dreamstime.com
ID 277982778 © Asenbayu