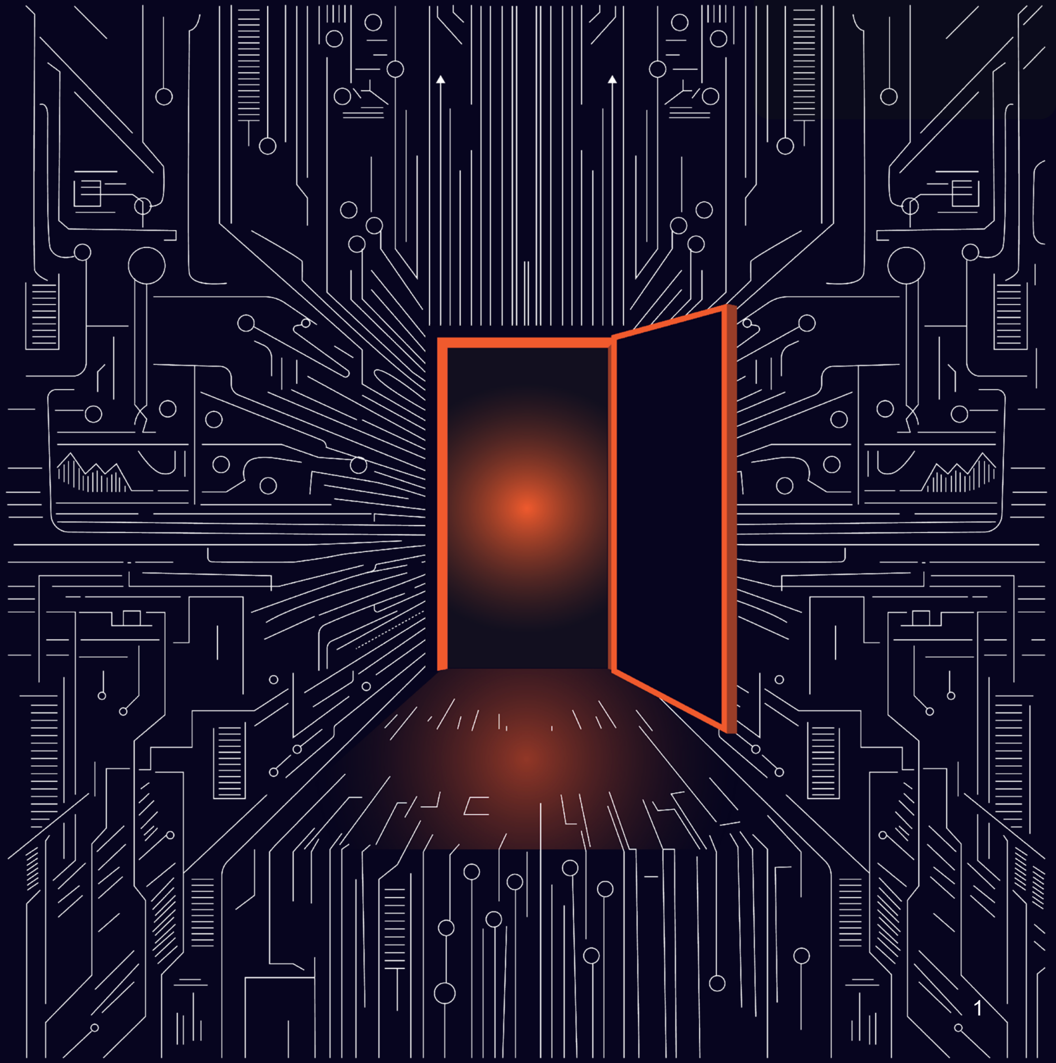
CARDO AI

opportunity from complexity

Data Science

January 2025



GAN-Driven Approaches to Transfer Learning in Tabular Data

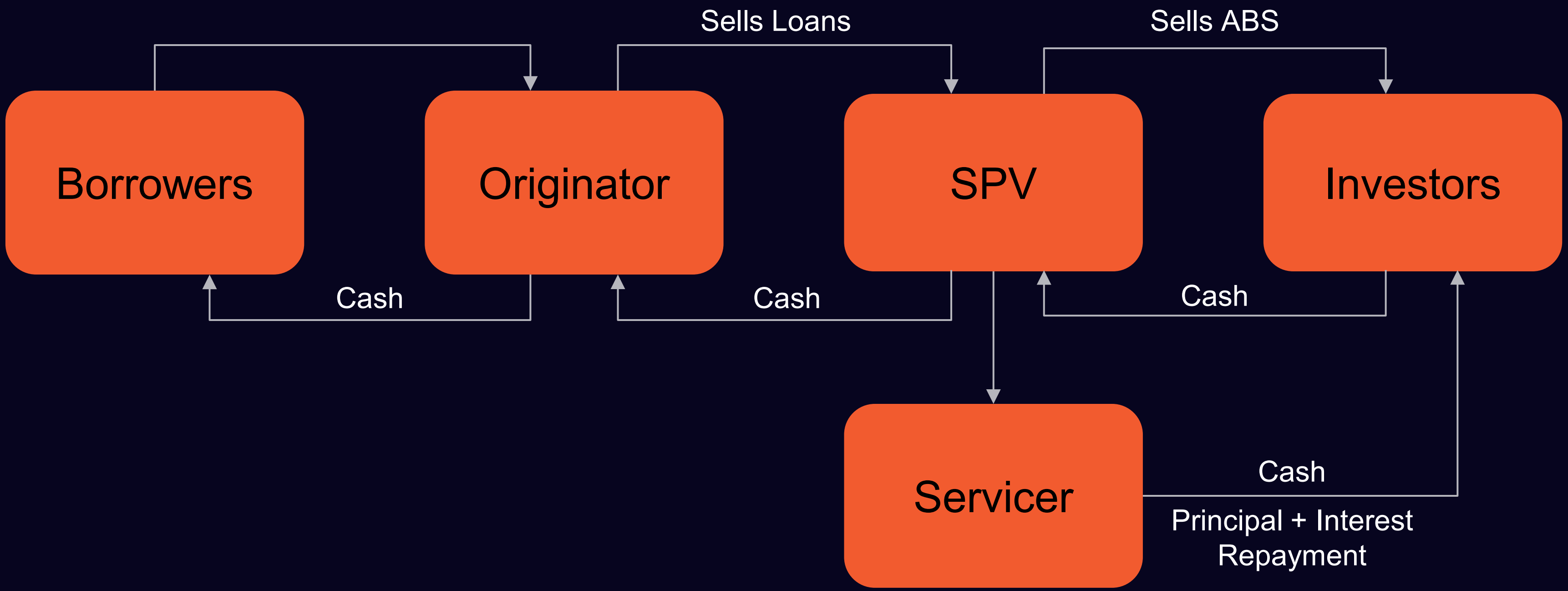
Foundations and Potential Applications in Structured Finance

Agenda

1. **Introduction**
    a. What Cardo AI Does
    b. What is Securitisation?
    c. Sketching the problem
2. **Generative Adversarial Networks for Tabular Synthetic Data Generation and Transfer Learning**
    a. Generative Adversarial Networks (GANs)
        i. Vanilla GANs
        ii. f-Divergence Duality
        iii. f-GANs
        iv. Kantorovich-Rubinstein Duality and Wasserstein GANs
        v. Conditional GANs
        vi. Conditional Tabular GANs
        vii. Transfer Learning in GANs

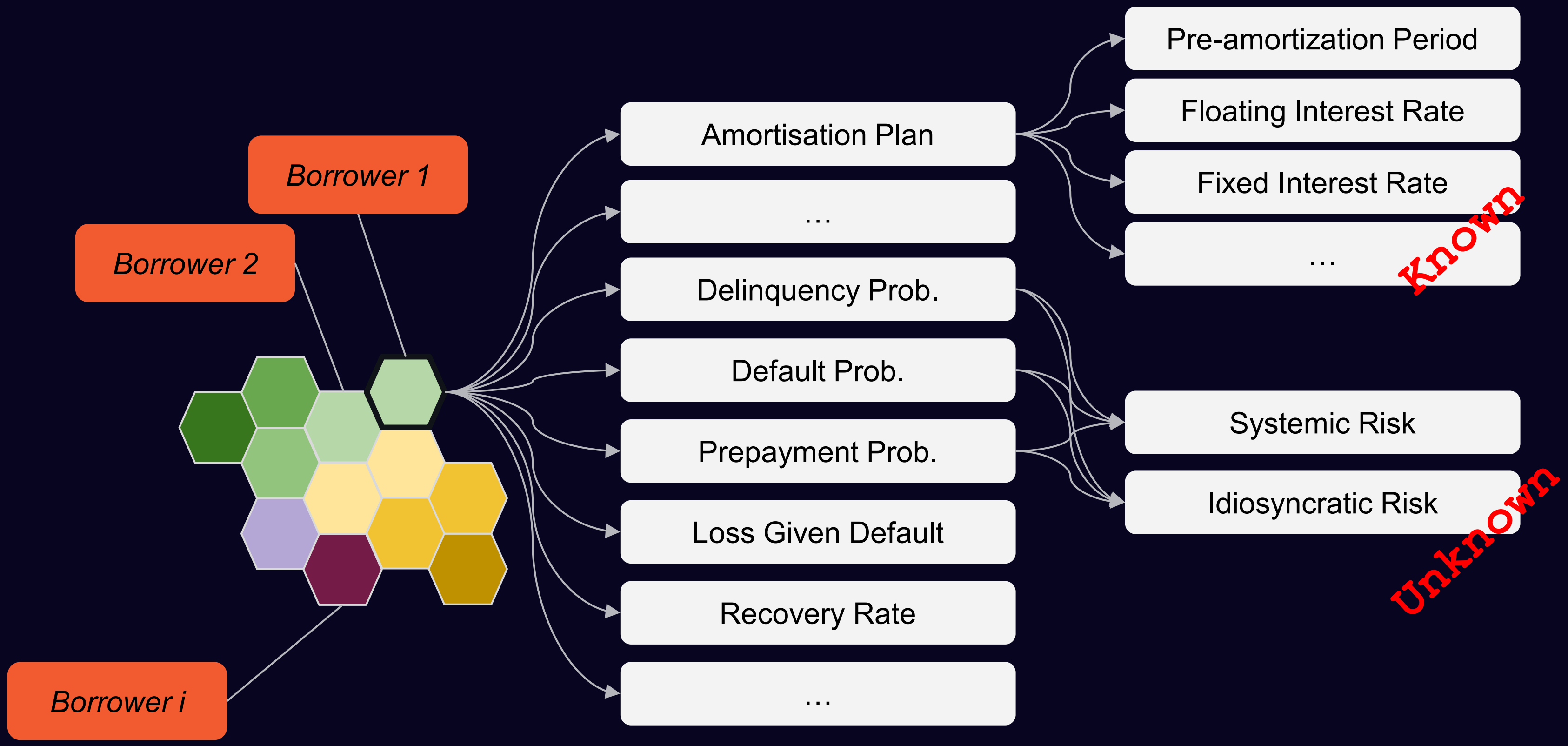
What is Securitisation?

Loans → Tradable Securities



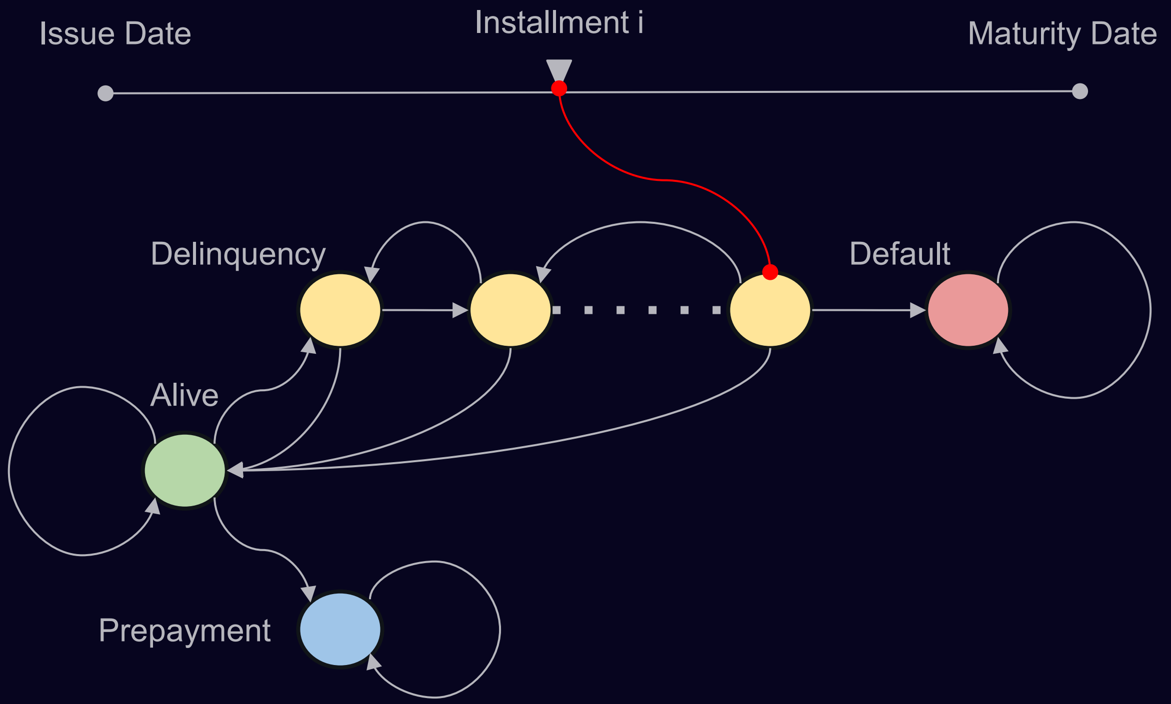
CARDO AI
opportunity from complexity

Asset Pool Cash Flow

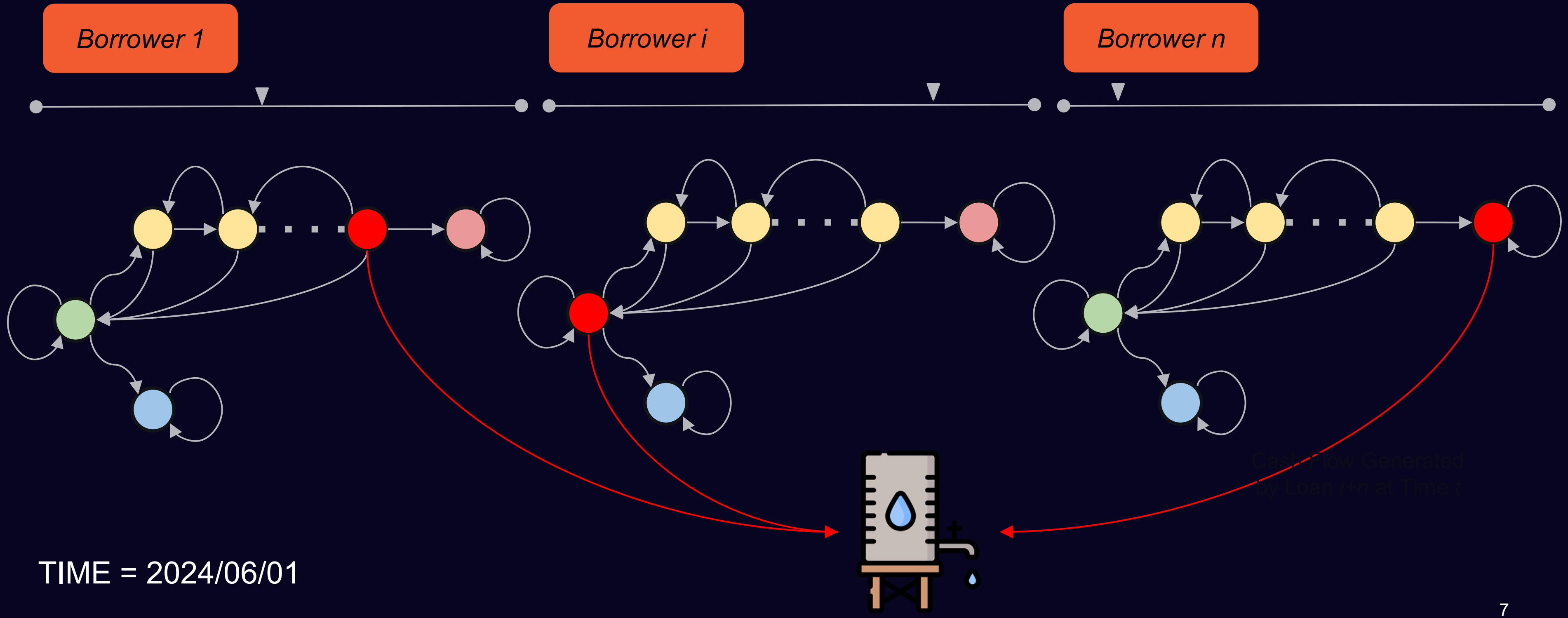


Asset Pool Cash Flow as a Markov Process

Borrower 1



Asset Pool Cash Flow as a Markov Process

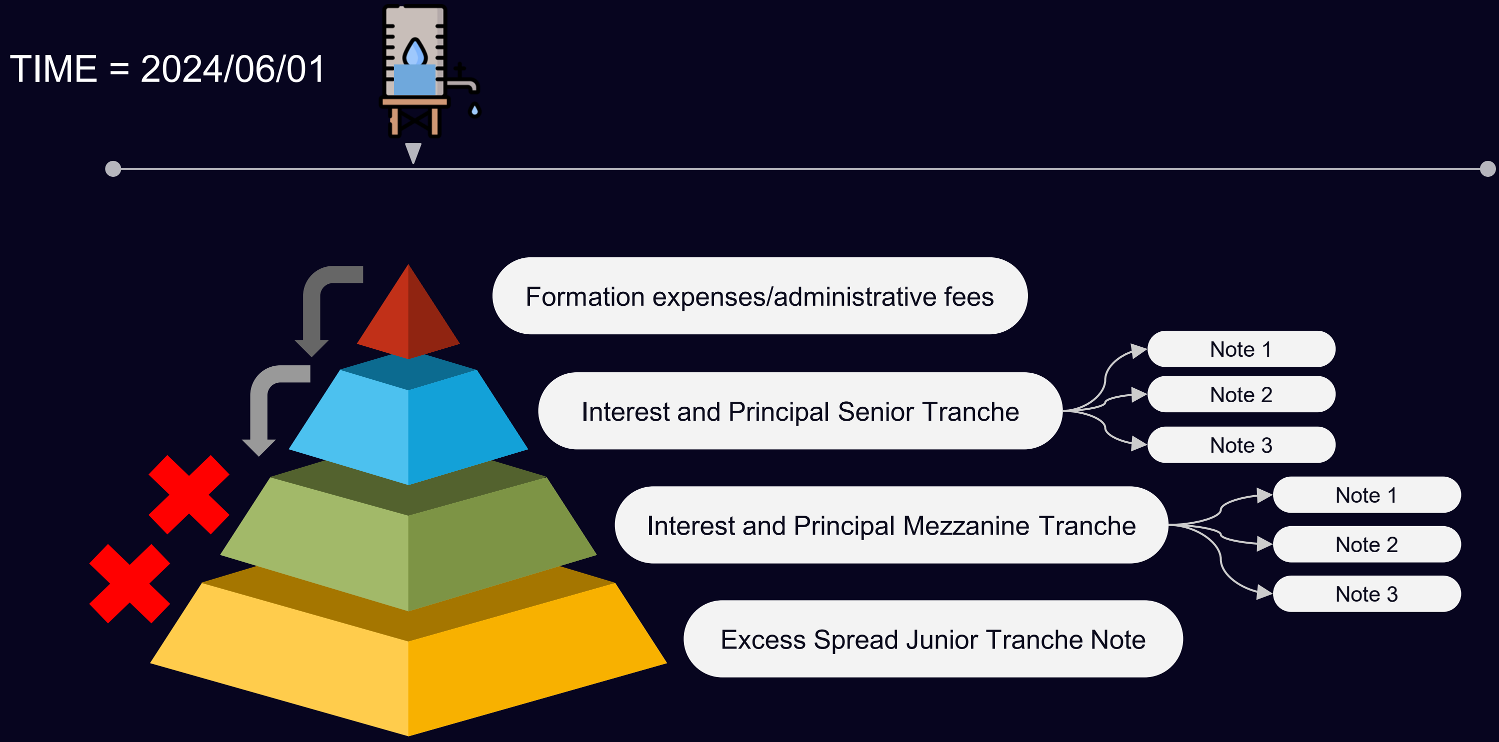


TIME = 2024/06/01

CARDO AI
opportunity from complexity

Cash Waterfall Structure

Cash Waterfall Structure



CARDO AI
opportunity from complexity

What are the Different Asset Classes in Securitisation?

- Mortgage-Backed Securities (MBS)
  - Residential Mortgage-Backed Securities (RMBS)
  - Commercial Mortgage-Backed Securities (CMBS)
- Asset-Backed Securities (ABS)
  - Auto Loans
  - Credit Card Receivables
  - Student Loans
  - Equipment Leases
  - Consumer Loans
  - Small Medium Enterprises Loans
- Collateralized Debt Obligations (CDOs)
  - Collateralized Loan Obligations (CLOs)
  - Collateralized Bond Obligations (CBOs)
- Whole Business Securitization
  - Revenue from entire businesses or franchises
- Future Flow Securitization
  - Securities backed by expected future receivables

Sketching the problem

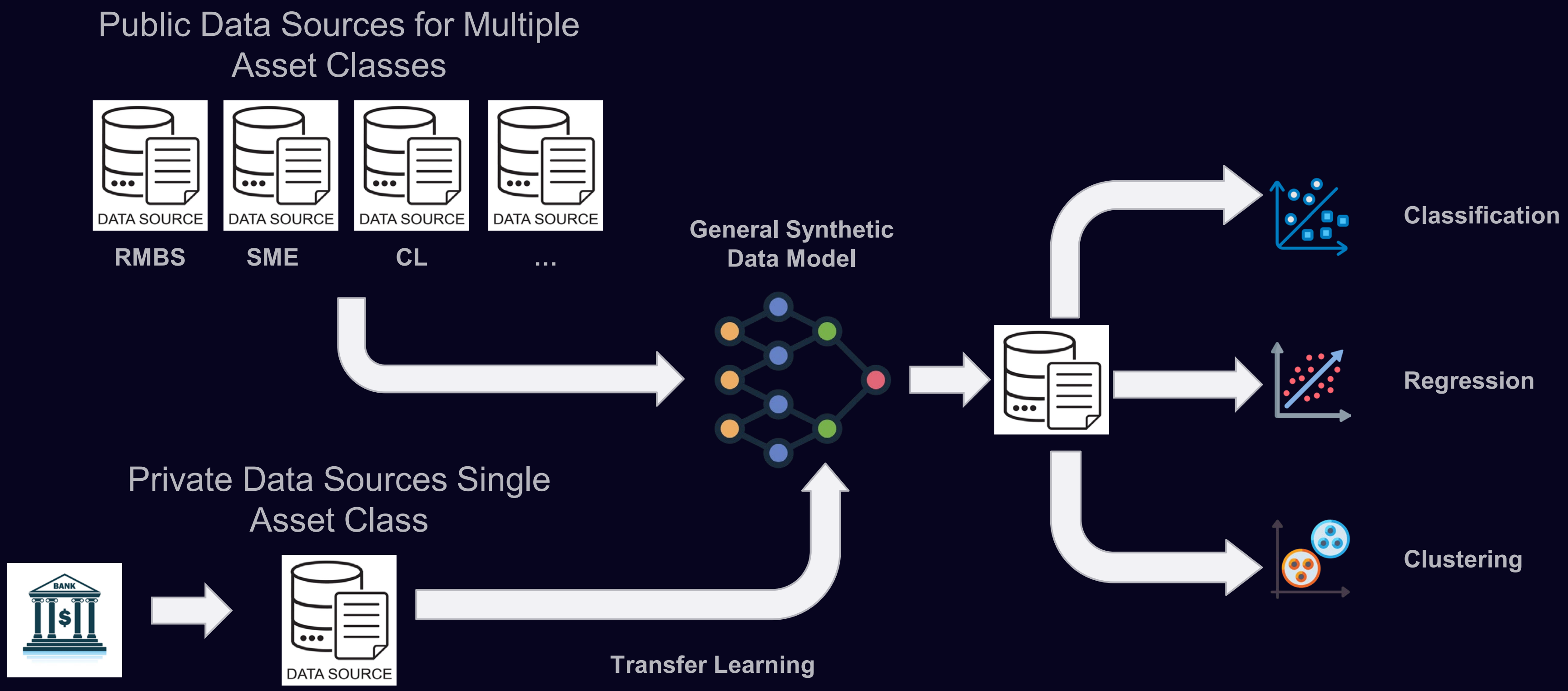
CARDO AI
opportunity from complexity

In the context of loan default classification, we aim to leverage **synthetic data generation for tabular transfer learning**.

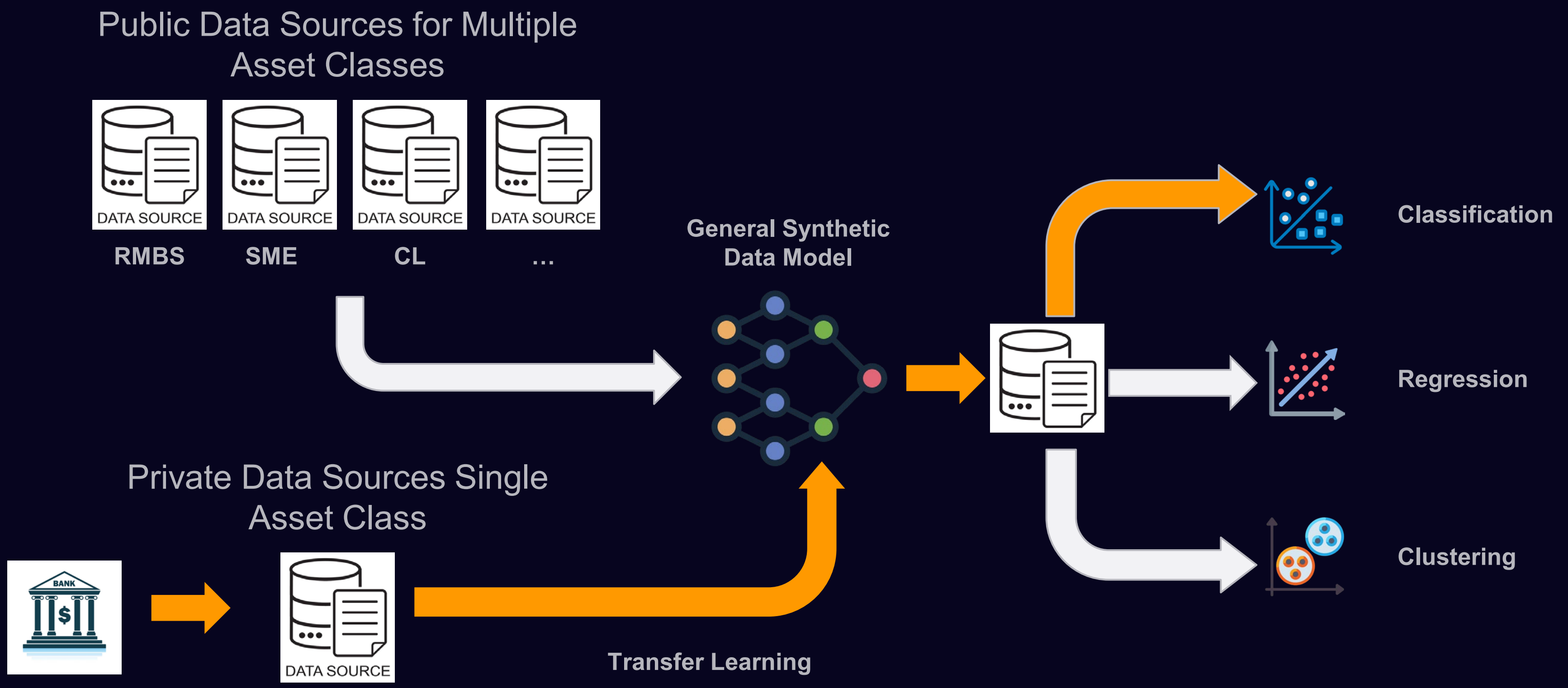
This approach aims to overcome several challenges. **Unbalanced datasets** are a significant issue, as default cases often constitute a minority, making model training difficult. Moreover, banks do not share **proprietary datasets**, seeking predictions without exposing sensitive loan information. **Privacy** is another key concern—borrower data must remain confidential and not directly used, adhering to privacy regulations. The **lack of historical data for new asset classes** further complicates model training. Additionally, **datasets can be noisy**, with past data containing inconsistencies and errors that affect model accuracy.

By implementing tabular synthetic data generation and transfer learning techniques, we seek to generate new balance datasets, protect privacy, address proprietary concerns, supplement data for new asset classes, and mitigate noise.

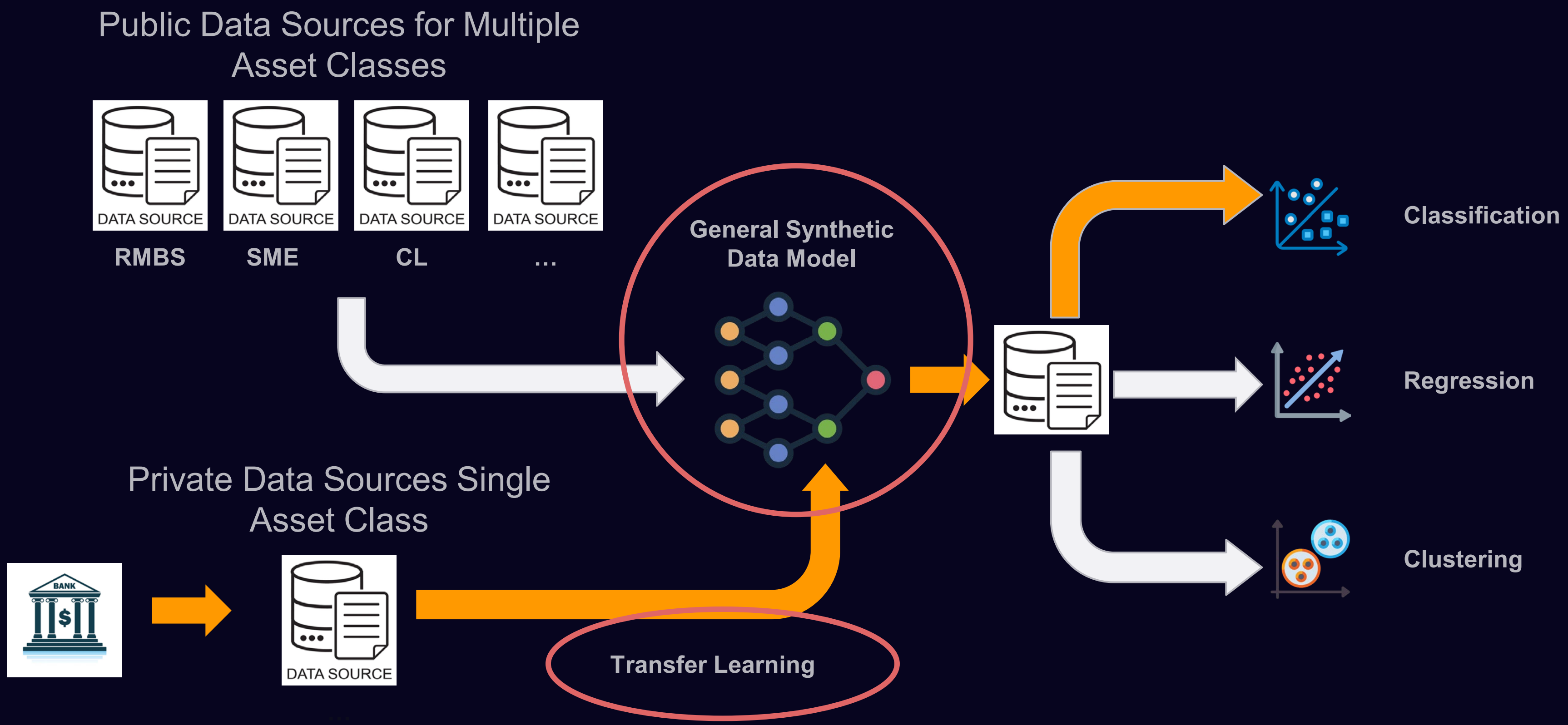
Sketching the problem



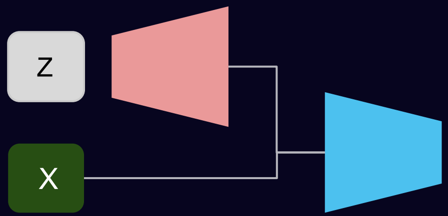
Sketching the problem



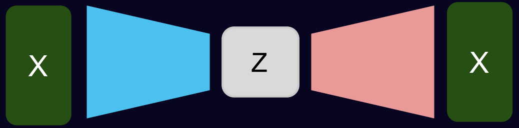
Sketching the problem



Deep Generative Models for Synthetic Data Generation

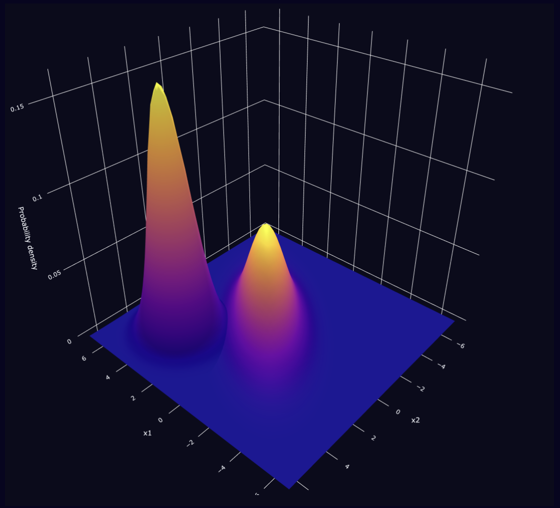


Generative Adversarial Network



Variational Autoencoder

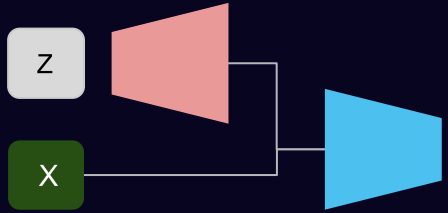
Diffusion Model



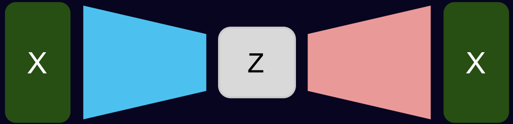
Generative models operate by analyzing patterns and distributions within their training data to generate new data from user inputs. Through training, **the model learns to identify the joint probability distributions of features in the dataset.** It then uses this knowledge to produce new data samples that closely resemble the original training data.

CARDO AI
opportunity from complexity

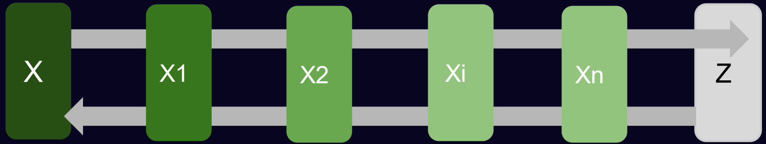
Deep Generative Models for Synthetic Data Generation



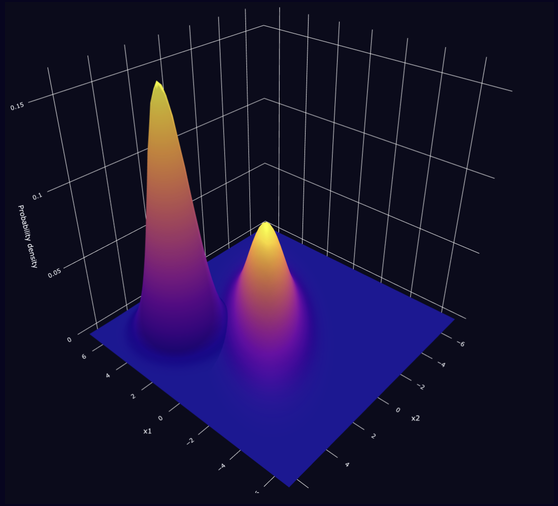
Generative Adversarial Network



Variational Autoencoder

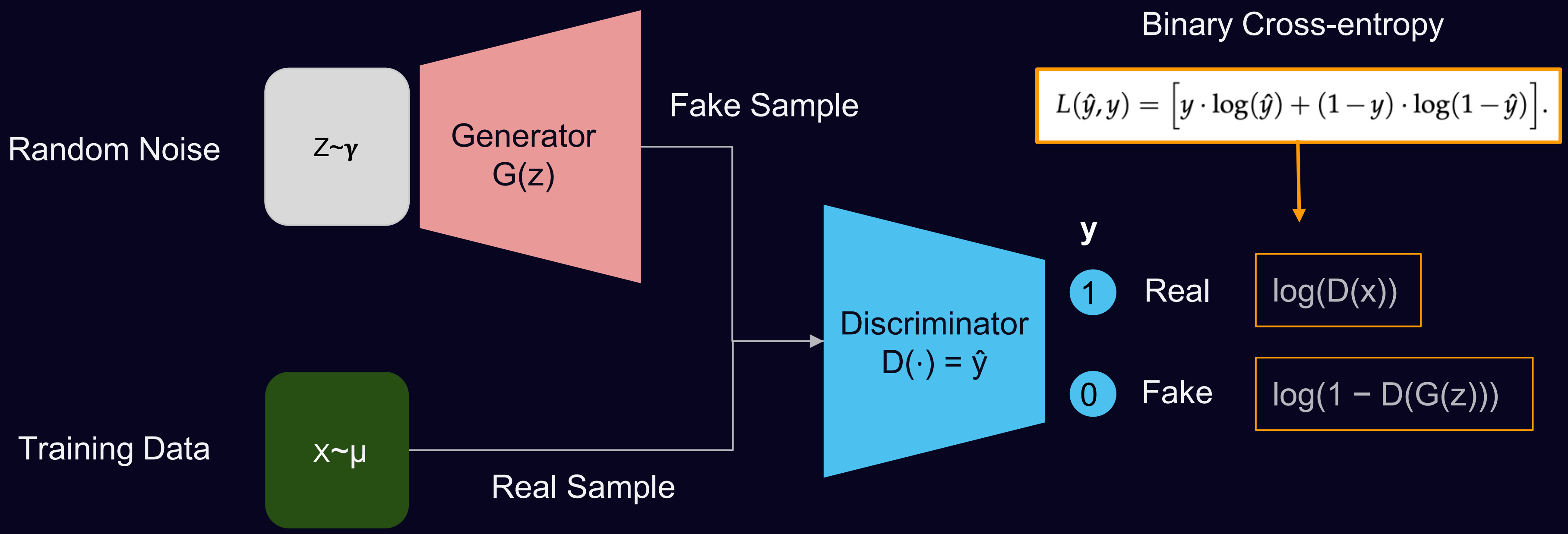


Diffusion Model

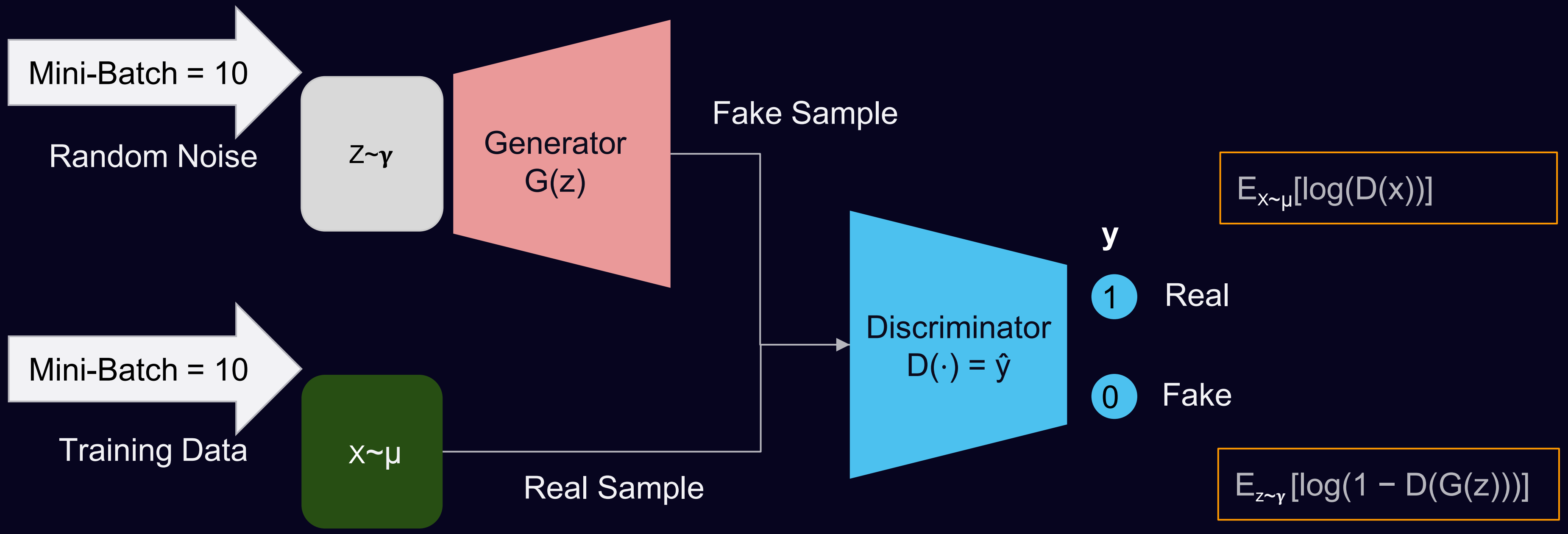


Generative models operate by analyzing patterns and distributions within their training data to generate new data from user inputs. Through training, **the model learns to identify the joint probability distributions of features in the dataset.** It then uses this knowledge to produce new data samples that closely resemble the original training data.

Generative Adversarial Network

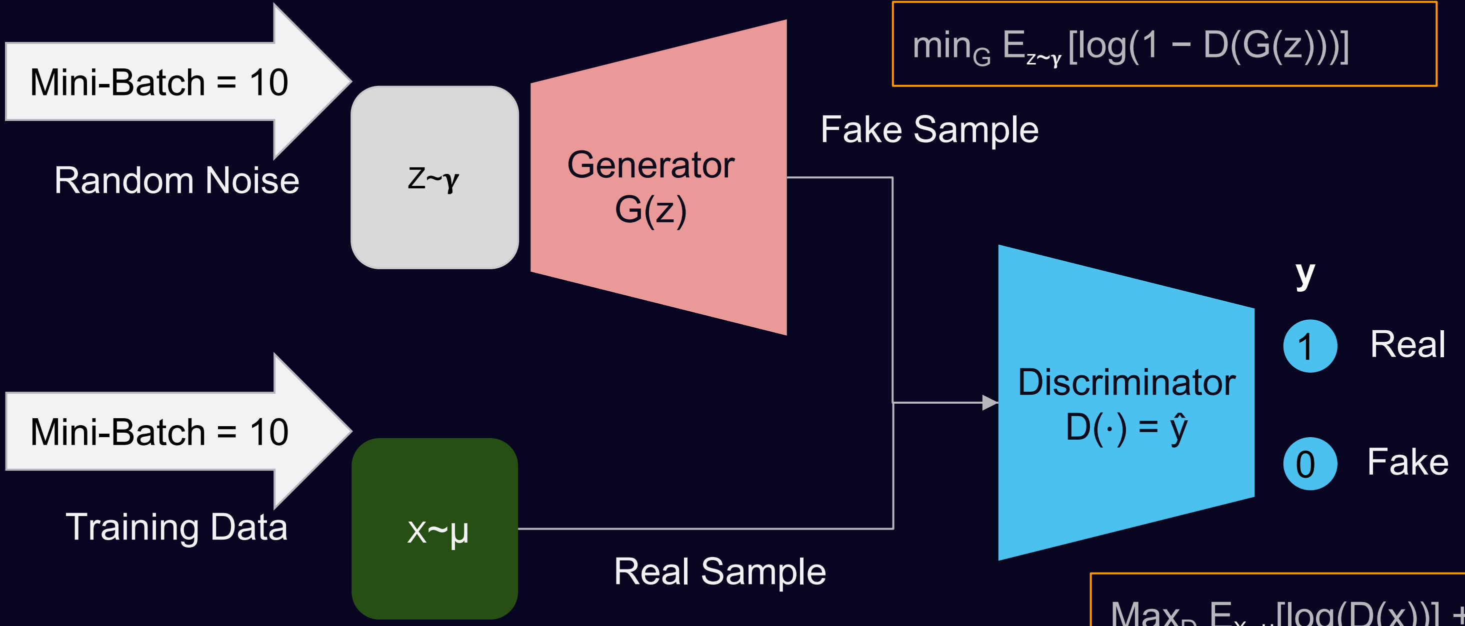


Generative Adversarial Network



CARDO AI
opportunity from complexity

Generative Adversarial Network



$$\min_G E_{z\sim\gamma}[\log(1 - D(G(z)))]$$

$$\mathrm{Max}_D\, E_{x\sim\mu}[\log(D(x))] + E_{z\sim\gamma}[\log(1 - D(G(z)))]$$

CARDO AI
opportunity from complexity

Generative Adversarial Networks for Tabular Synthetic Data Generation and Transfer Learning

- https://stefanopenazzi2.github.io/2025/01/05/F-Divergence-Duality.html
- https://stefanopenazzi2.github.io/2025/01/05/F-GANs.html
- https://stefanopenazzi2.github.io/2025/01/05/Kantorovich-Rubinstein-Duality.html
- https://stefanopenazzi2.github.io/2025/01/21/C-GANs.html
- https://stefanopenazzi2.github.io/2025/01/21/CT-GANs.html
- https://stefanopenazzi2.github.io/2025/01/22/TransferLearning-GANs.html

CARDO AI
opportunity from complexity

www.cardoai.com

/cardoai

/cardoai_

opportunity from complexity

All opinions expressed in this presentation are solely those of the author and do not necessarily reflect the views or positions of the Cardo AI Corporation.