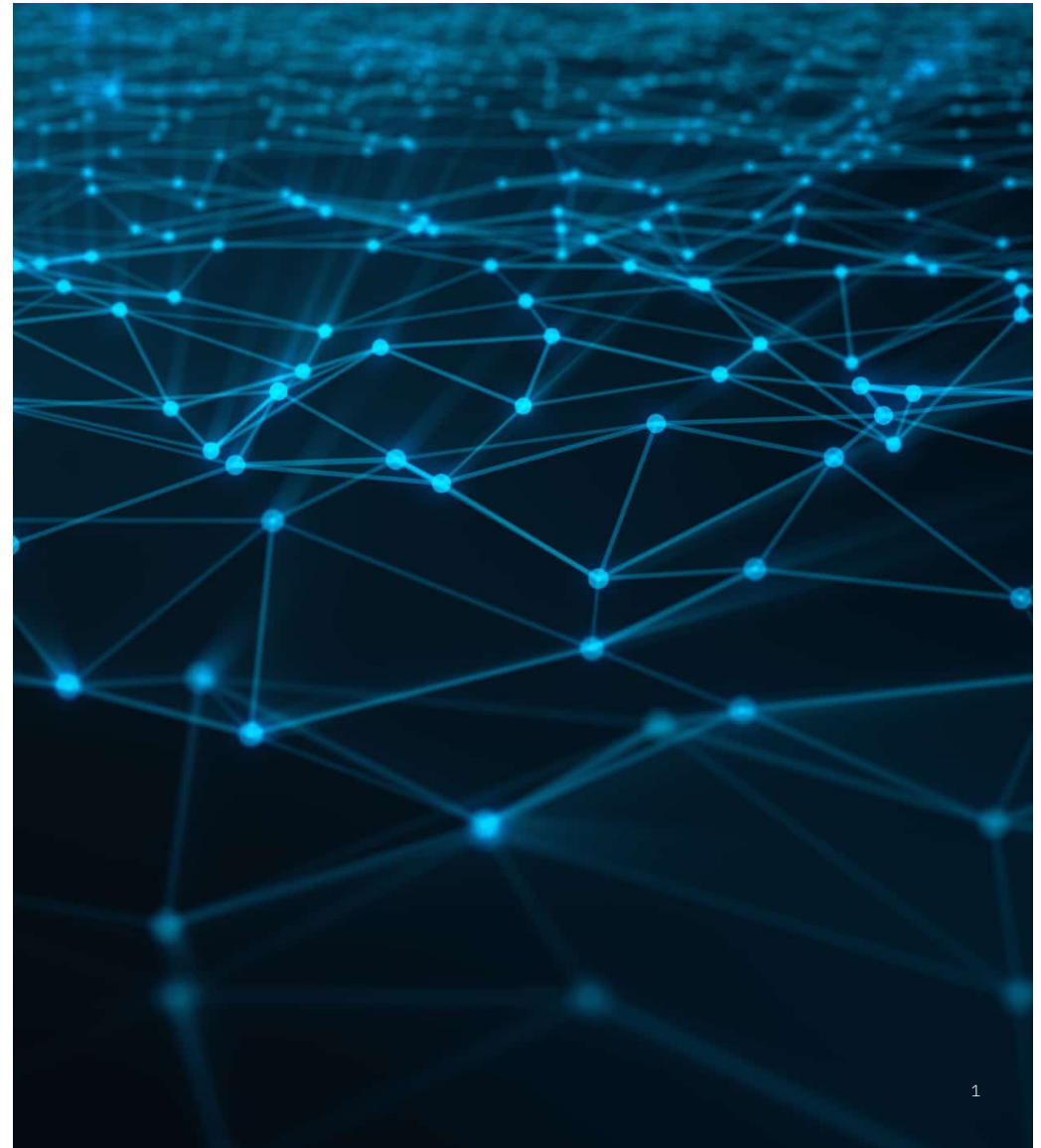# Graph Machine Learning for Financial Crime Analysis

KDD Finance Day
Aug 26, 2024

Kubilay Atasu
Associate Professor, Data Intensive Systems
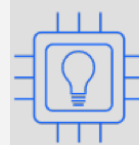*Software Technology Department*
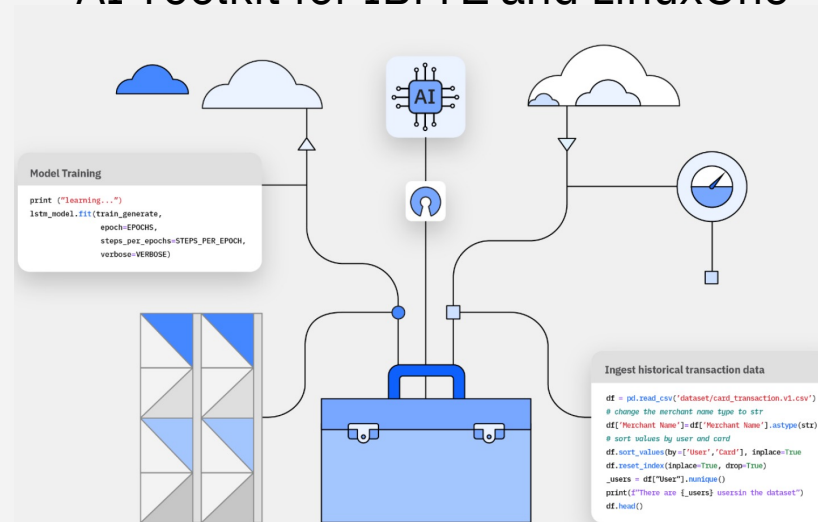
TUDelft — Delft University of Technology

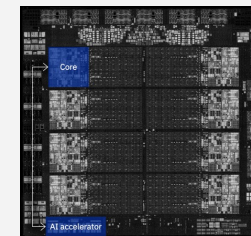# Enabling AI in Financial Transaction Processing

You probably used IBM Z today!

AI Toolkit for IBM Z and LinuxOne

On Chip AI Accelerator

Detecting Financial Crime in Real Time!

Credits(2019-2023)

SNSF Project 172610: Hardware-accelerated recursive programs (PI: K. Atasu)

PhD thesis by J. Blanusa, "Acceleration of graph pattern mining and applications to financial crime", Aug. 2023.
- Publications in VLDB 2020, SPAA 2022, TOPC 2023, NeurIPS 2023

- Winner of the 2023 Fritz Kutter Award: Best Industry Related Doctoral Thesis in Computer Science in Switzerland
- IBM Outstanding Accomplishment Award for Contributions to System Z AI Offerings (Real-time AML & Fraud Detection)

Trends in Financial Crime Analysis

Trends

Legacy rule-based systems are being replaced by agile AI-based systems

Know your customer (KYC) and customer due diligence (CDD) mechanisms

Follow the data instead of following the money, knowledge graphs and AI!

Convergence between AML and other financial fraud detection solutions

Challenges

Detecting constantly evolving crime patterns in real-time

Criminal networks crossing bank & national boundaries

Building cost-efficient and sustainable AI technologies

Regulatory Compliance, Trustworthy and Secure AI

Example: Money Laundering
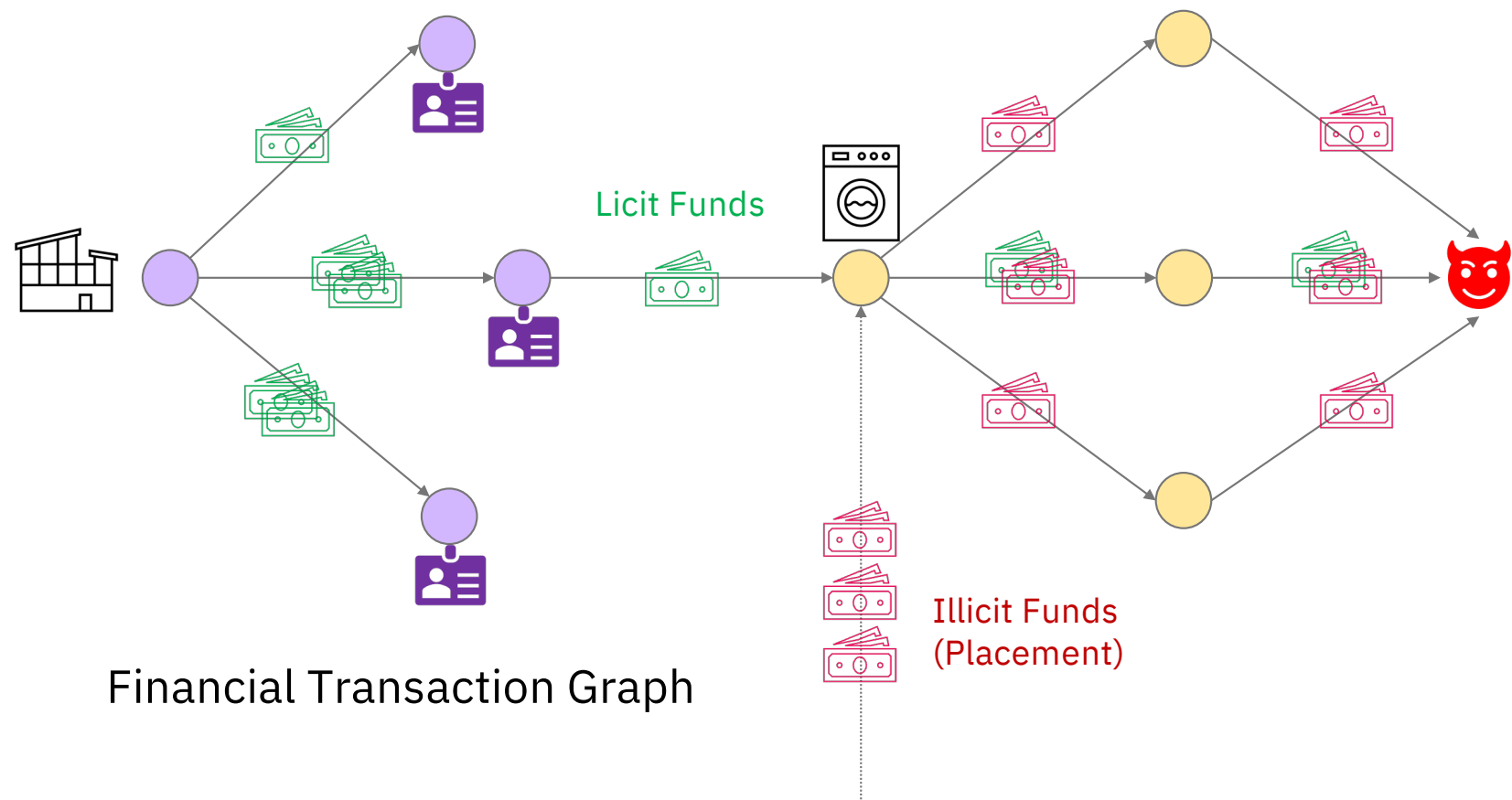
UN estimates that 2—5% of the global GDP is laundered each year.

Licit Funds

Illicit Funds
(Placement)

Financial Transaction Graph

Example: Money Laundering

UN estimates that 2—5% of the global GDP is laundered each year.

Licit Funds

Financial Transaction Graph

Illicit Funds
(Placement)

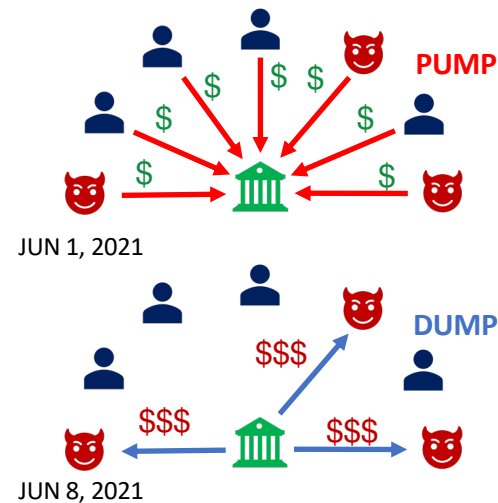Scatter-Gather Pattern
(Layering)
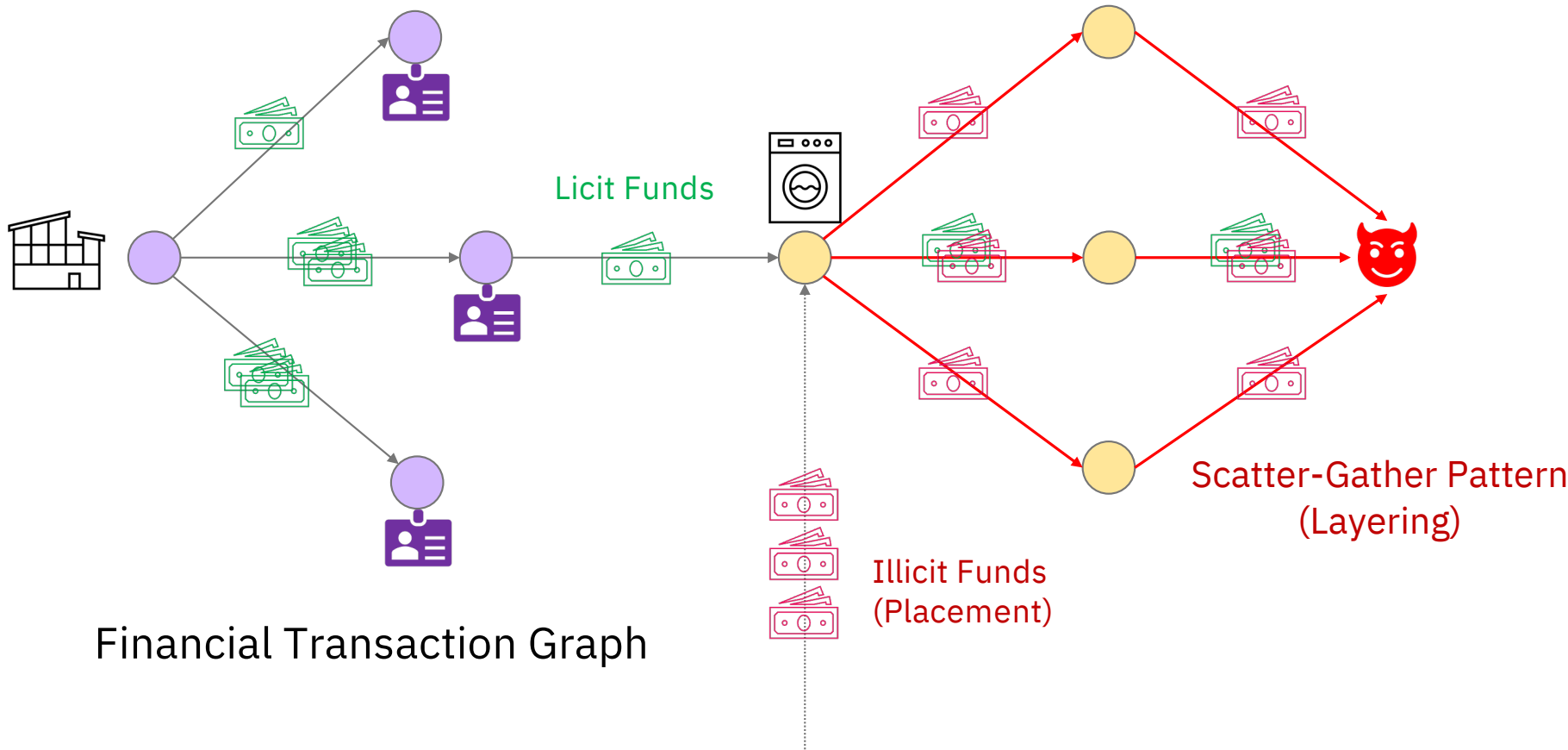
Known Suspicious Financial Transaction Patterns
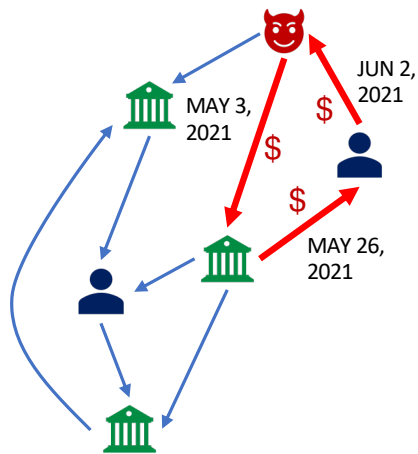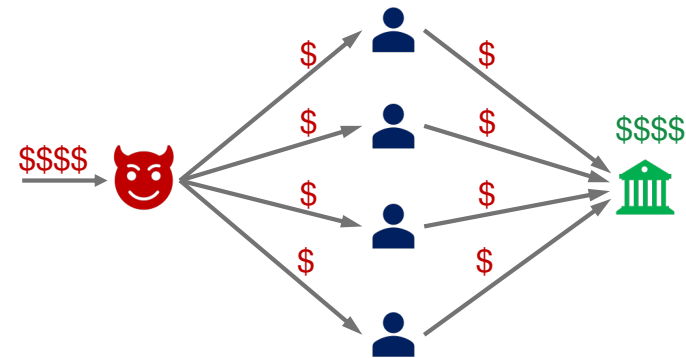

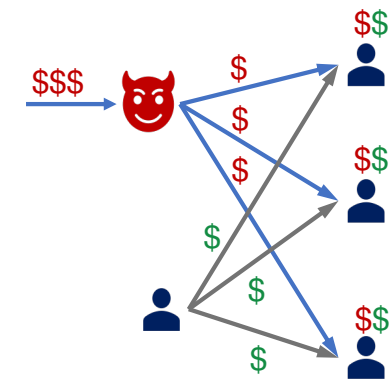Circular trading and money laundering

Cycles

Pump and dump scheme

High fan-in and fan-out
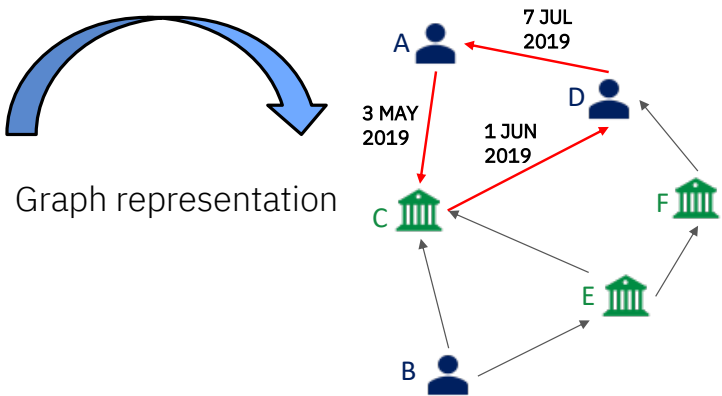
Smurfing

Scatter-gather

Disguising money flow

Biclique

How Does Graph Machine Learning Help?

Tabular representation of financial transactions

Trans. ID	Timestamp	Source bank ID	Source Account	Target bank ID	Target Account	Amount	Currency	Payment type
0	3 MAY 2019 12:45	1	A	1	C	1400	USD	Cheque
1	15 MAY 2019 07:34	2	B	1	C	710	EUR	ACH
2	18 MAY 2019 16:55	3	E	1	C	950	USD	Credit card
3	1 JUN 2019 10:06	1	C	3	D	1200	CHF	Wire
4	27 JUN 2019 13:18	2	F	3	D	2300	EUR	Credit card
5	7 JUL 2019 11:14	3	D	1	A	1100	USD	Credit card

Graph representation

7 JUL 2019

3 MAY 2019

1 JUN 2019

A   D   F   C   E   B

New Transaction

7 JUL 2019
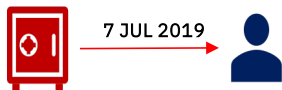
Pattern Discovery
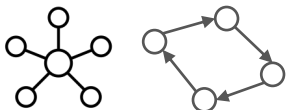
Accuracy: 9% → 0.73%

Dataset size: 100 M transactions. *Illicit rate*: 0.3%. *Model*: LightGBM. *Metric*: minority (illicit) class F1-score .

AML Accuracy Improvements using Graph ML

Synthetic AML dataset with 100M transactions
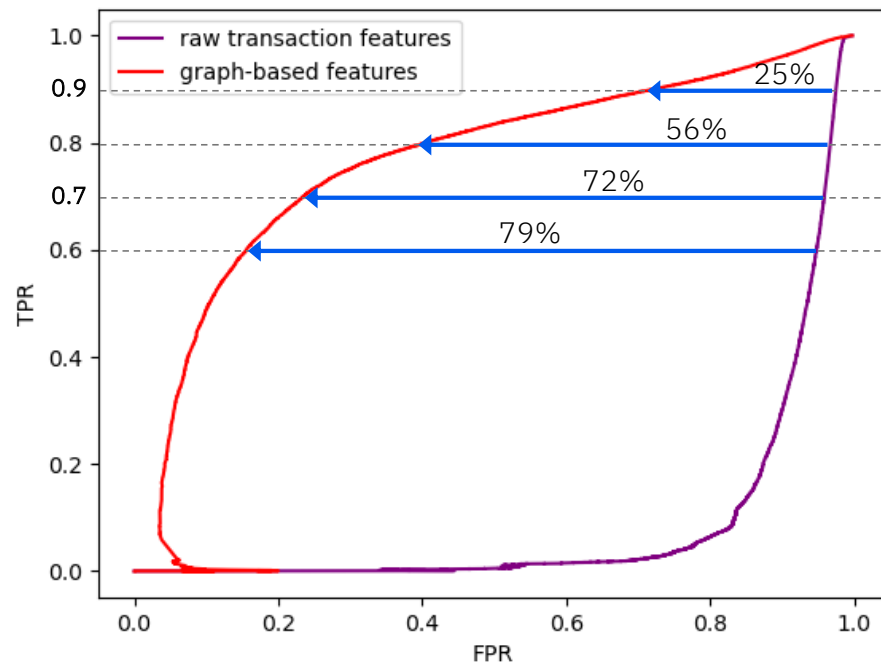
True Positive Rate (TPR) vs False Positive Rate (FPR)

raw features [%]		
TPR	FPR	F1
60	94.7	9.7
70	95.8	7.9
80	96.7	6.4
90	97.4	5.0

graph features [%]		
TPR	FPR	F1
60	15.6	70.2
70	23.5	73.1
80	40.6	68.2
90	72.3	42.4

Why does accuracy matter?

- Higher TPR → less regulatory risk!
- Lower FPR → more cost savings!

Explaining The Predictions of Graph ML

D 7 JUL 2019 → A

Inference

ML model → **Illicit**

Prediction: D → A is illicit!

Feature importance Analysis

Cycle len. 3
Amount
Cycle len. > 5
Payment type
Timestamp

0   0,2   0,4   0,6
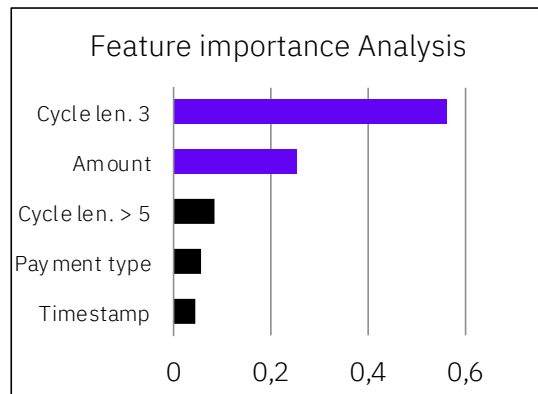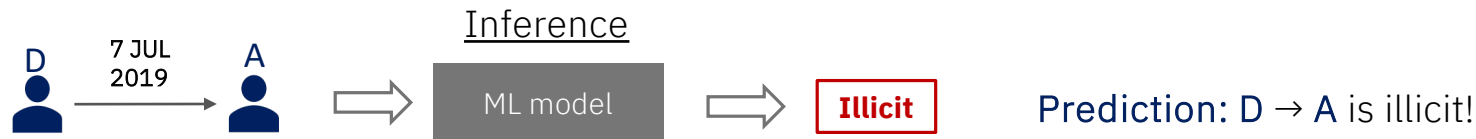
Trans. ID	Tstamp	Src. Bank	Src. Acc	Targ. Bank	Targ. Acc.	Amount	Curr.	Payment Type	Cycle len. 3	Cycle len. 4	Cycle len. 5	Cycle len. >5
5	7 JUL 2019 11:14	3	D	1	A	1100	USD	Credit card	1	0	0	0

Important graph pattens

A
D
C
F
E
B
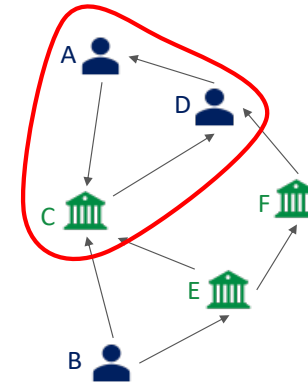
Highlight the patterns that have affected the prediction!

Graph Machine Learning in IBM AI Toolkit for Z

Monitoring suspicious account activities in real time!

Financial Transactions

On Chip AI Accelerator

Graph Feature Preprocessor

Cliques	Cycles	Fans	Scatter-Gather	Other

Real-time extraction of graph features

Machine Learning Libraries

Random Forest	Graph Neural Networks	Multilayer perceptron	Gradient Boosting	Other

Real-time inference

Suspicious accounts and transactions
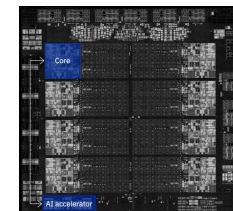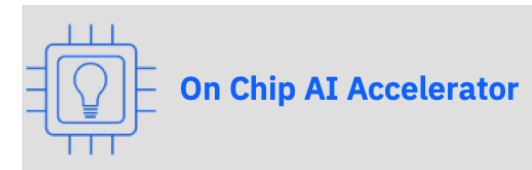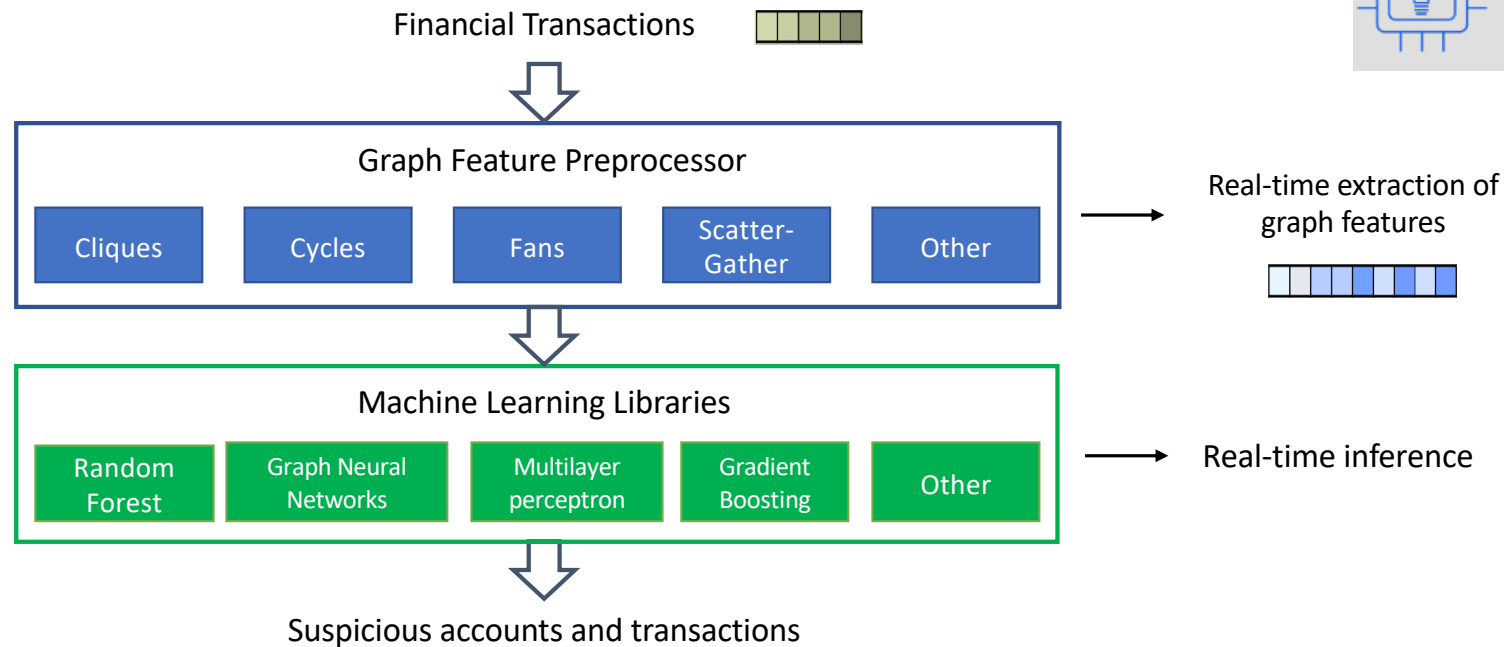
Can Graph Neural Networks Help?

Input Financial Transaction Graphs

1 Graph Mining Algorithms

Cliques | Cycles | Other

2 Machine Learning Algorithms

RF | MLP | LightGBM

1 Graph Neural Networks

GIN | PNA | Other
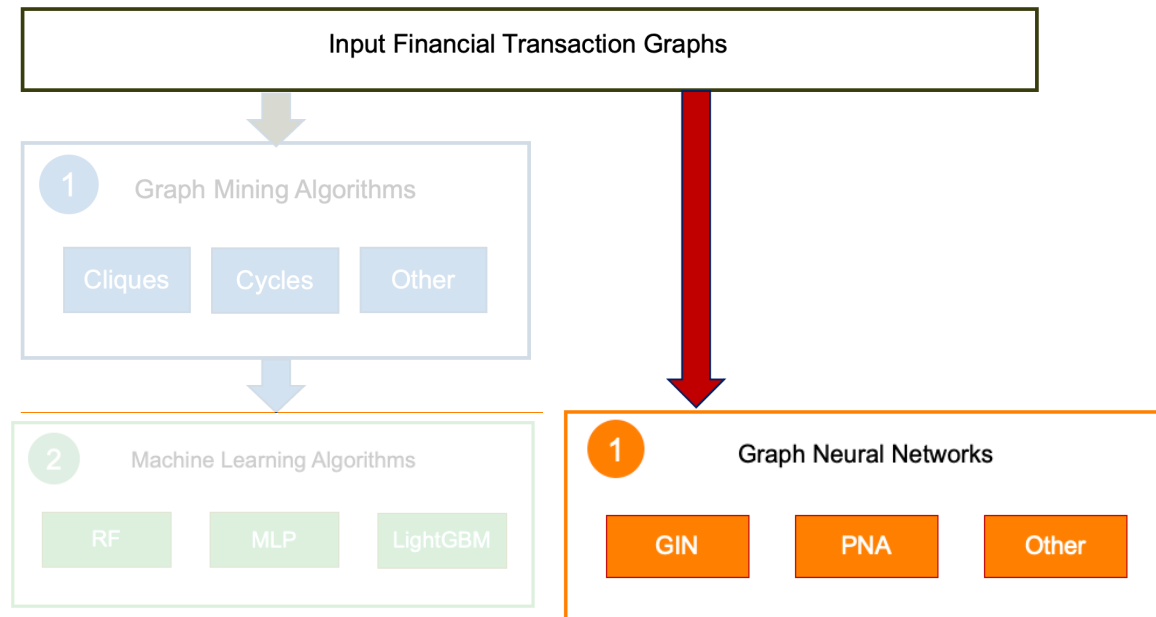
Why Graph Neural Networks?

Automation No feature engineering

No (less) **domain knowledge** required

Can detect "**unestablished**" patterns

Differentiable – connect with LLMs

Provably Powerful GNNs for Directed Multigraphs

Out Degree & Reverse MP
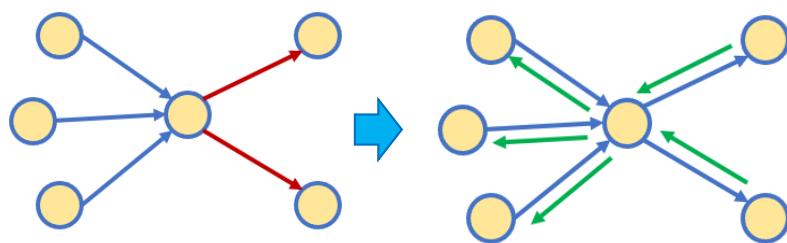
Parallel Edges & Ports
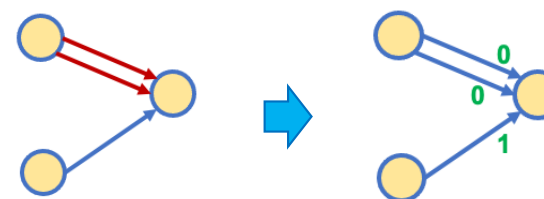
Cycles & Ego IDs

Edge Features & Edge Updates

Theorem: Combination of reverse MP, ego IDs, and ports enables detection of any directed subgraph pattern.

AML and Ethereum Phishing Fraud Detection Results

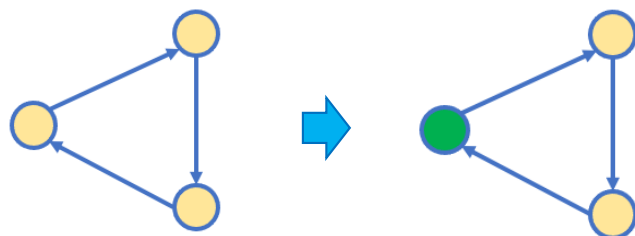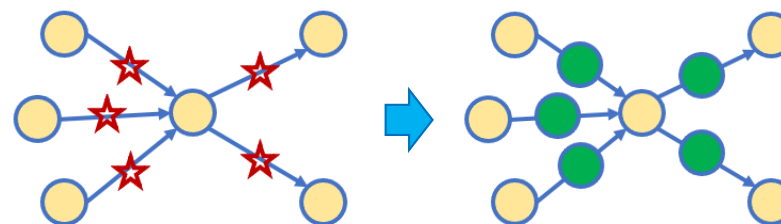| Model	AML Small HI	AML Small LI	AML Medium HI	AML Medium LI	ETH
LightGBM+GFs (Altman et al. 2023)	62.86 ± 0.25	20.83 ± 1.50	59.48 ± 0.15	20.85 ± 0.38	53.20 ± 0.60
XGBoost+GFs (Altman et al. 2023)	63.23 ± 0.17	27.30 ± 0.33	65.70 ± 0.26	28.16 ± 0.14	49.40 ± 0.54

Using IBM's Graph Feature Preprocessor [1]

Model	AML Small HI	AML Small LI	AML Medium HI	AML Medium LI	ETH
Multi-GIN+EU	64.79 ± 1.22	26.88 ± 6.63	58.92 ± 1.83	16.30 ± 4.73	48.37 ± 6.62
Multi-PNA	64.59 ± 3.60	30.65 ± 2.00	65.67 ± 2.66	33.23 ± 1.31	65.28 ± 2.89
Multi-PNA+EU	68.16 ± 2.65	33.07 ± 2.63	66.48 ± 1.63	36.07 ± 1.17	66.58 ± 1.60

Our Multi-GNN Models (Without Graph Features) [2]

Multi-GNNs achieve 5-15% higher accuracy without any feature engineering!
Multi-GNNs can automatically discover discriminative graph features!

[1] J. Blanusa et al.: Graph Feature Preprocessor: Real-time Extraction of Subgraph-based Features from Transaction Graphs, 2024 (Arxiv).
[2] B. Egressy et al.: Provably Powerful Graph Neural Networks for Directed Multigraphs. AAAI 2024 (Oral Presentation).
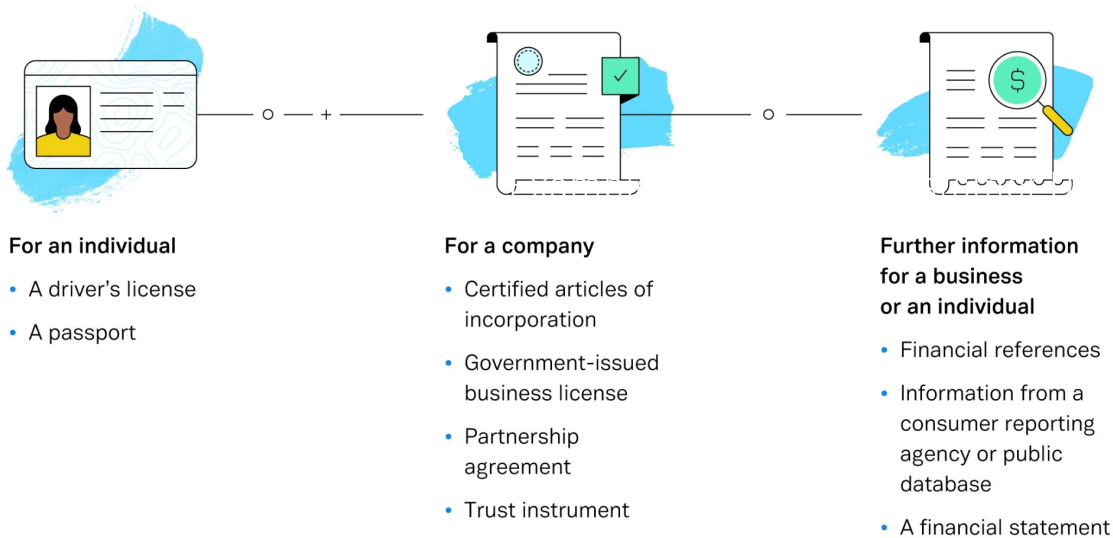
What's Next?

Relational Multimodal Learning

For an individual
- A driver's license
- A passport

For a company
- Certified articles of incorporation
- Government-issued business license
- Partnership agreement
- Trust instrument

Further information for a business or an individual
- Financial references
- Information from a consumer reporting agency or public database
- A financial statement

Image Source: https://plaid.com/resources/banking/what-is-kyc/

Modality: Text

This is a partnership between... , which owns properties in... Its main customers are ...

Technology: Transformers

Modality: Table

Customer ID	Type	Function	Bank Acct.	Credit Card

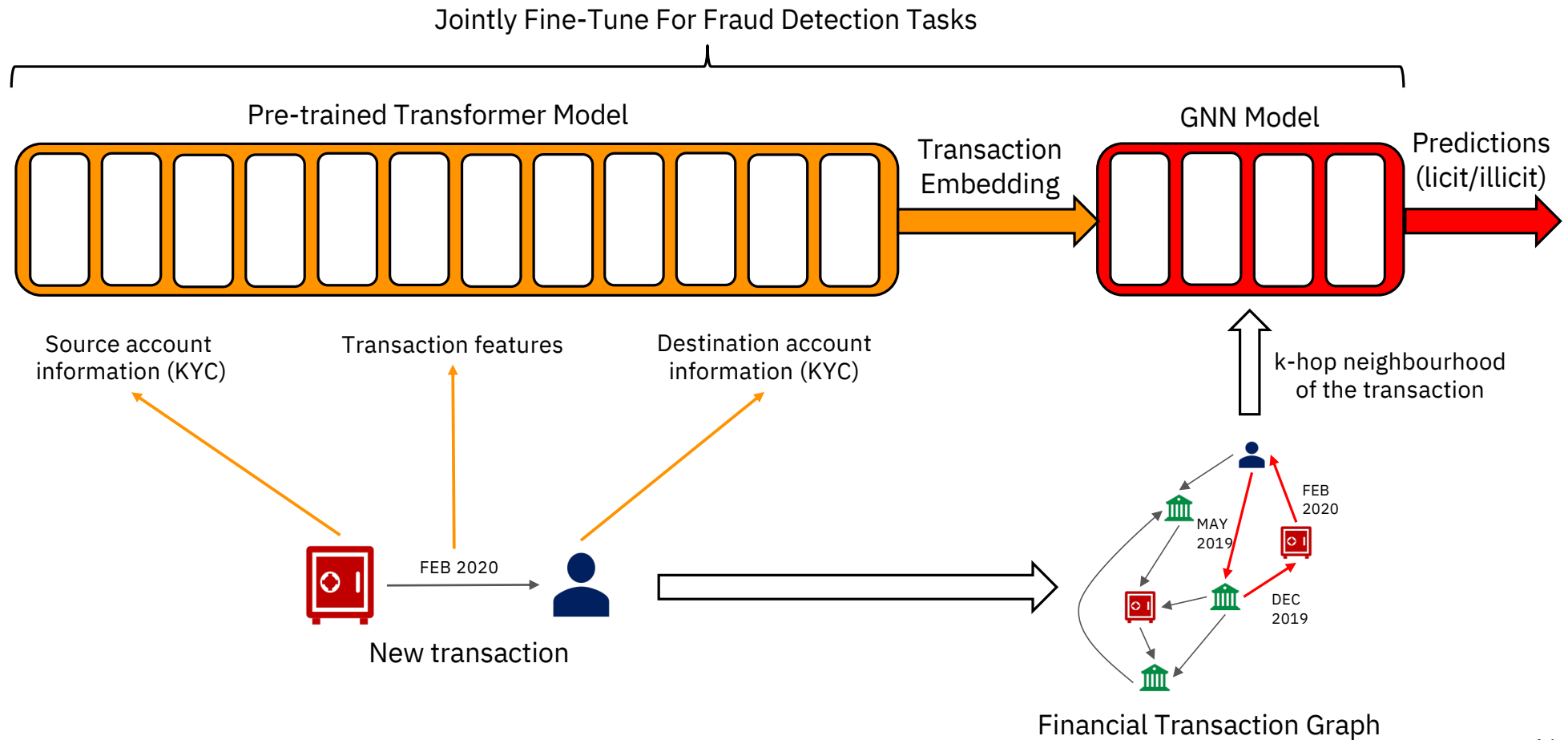**Technology: Tabular Transformers**

Modality: Graph
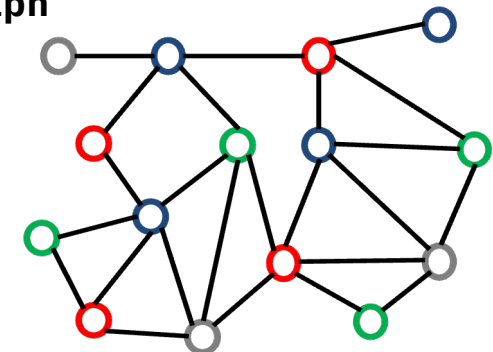
Technology: Graph Neural Networks

Transformers + GNNs for Financial Fraud Detection?

Jointly Fine-Tune For Fraud Detection Tasks

Pre-trained Transformer Model

GNN Model

Transaction Embedding

Predictions (licit/illicit)

Source account information (KYC)

Transaction features

Destination account information (KYC)

k-hop neighbourhood of the transaction

FEB 2020

New transaction

Financial Transaction Graph

FEB 2020

MAY 2019

DEC 2019

Foundation Models for AML & Financial Fraud Detection?

Pre-training	Fine-Tuning
Unsupervised	Supervised
Synthetic data	Real data
Known Patterns	Unknown Patterns

Graph
+Tabular
+Text

Training

Foundation Model

Adaptation

Money Laundering

Employee Fraud

Credit Card Fraud

Tax Evasion

Insurance Fraud

Phishing

Domain-specific
foundation model
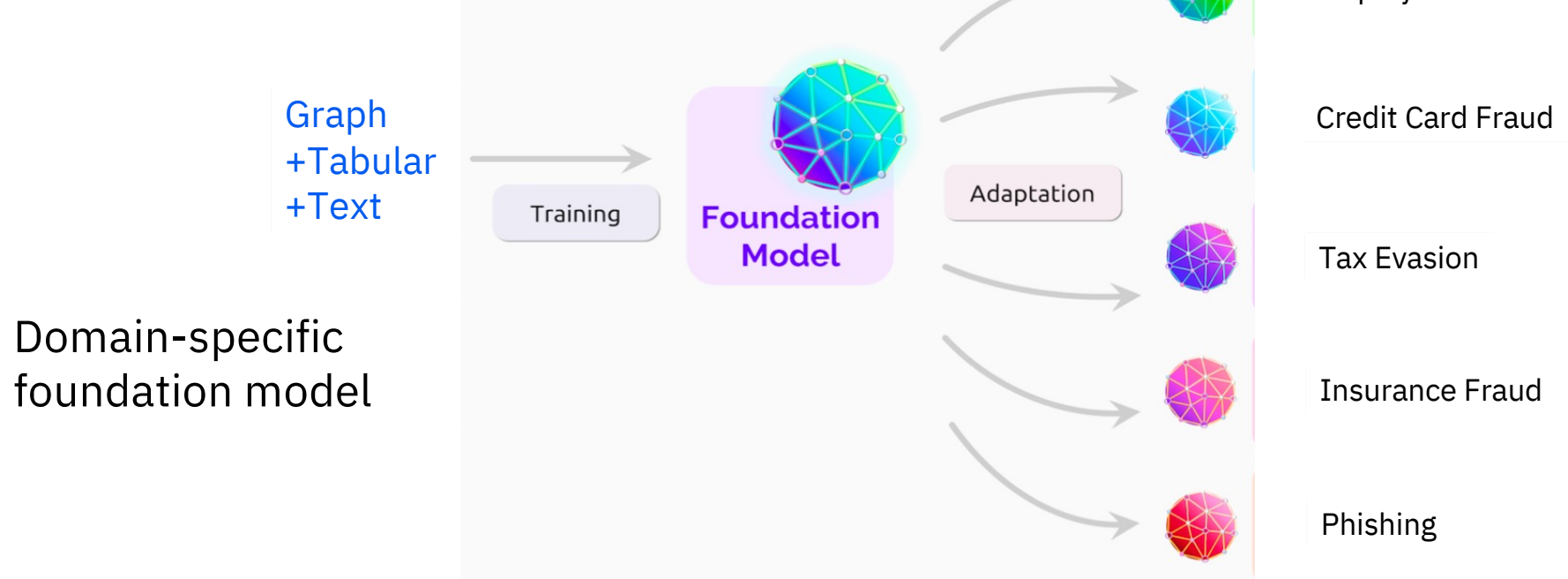
Fig. Source: On the Opportunities and Risks of Foundation Models: https://arxiv.org/abs/2108.07258

Graph Machine Learning for Financial Crime Analysis

Kubilay Atasu
Associate Professor, Data Intensive Systems
Software Technology Department

TU Delft — Delft University of Technology