Working Paper presented at the

Peer-to-Peer Financial Systems 2026 Workshop

June 2026

Sovereign AI for Financial Intelligence: Multi-Bank Transaction Analysis and AML Detection Using Open-Source Large Language Models

Shokan Orazbekov
Astana IT University

Bekzat Ashirbek
Astana IT University

Akbota Zhamanbay
Astana IT University

Birganym Abilkhaiyr
Astana IT University

Powered by

P2P Financial Systems

exponential Science Foundation

Sovereign AI for Financial Intelligence: Multi-Bank Transaction Analysis and AML Detection Using Open-Source Large Language Models

Shokan Orazbekov, Bekzat Ashirbek, Akbota Zhamanbay, Birganym Abilkhaiyr
Astana IT University, Astana, Republic of Kazakhstan
Emails: Chokan.kz@gmail.com, bekzatashirbek2@gmail.com,
akbotazhamanbai09@gmail.com, abylkhayirbirganym@gmail.com

***Abstract*—Kazakhstan's digital finance ecosystem exemplifies the transformation underway across emerging markets. The Digital Tenge CBDC is operational and integrated with national budget financing and public procurement; the regulated crypto-asset market reported USD 6.8 billion in trading volume on AIFC-licensed platforms in the first nine months of 2025; and Central Asia's first regulated spot Bitcoin ETF launched in August 2025. However, this rapid digitalisation has amplified financial-crime risks—money-mule schemes, structured transit, and cash-out typologies exploit fragmented banking infrastructure where each institution exports data in proprietary formats.**

Financial Intelligence Units (FIUs) in emerging economies face four simultaneous constraints: (1) data-sovereignty requirements prohibit cloud-based AI services, mandating fully on-premise deployment; (2) heterogeneous banking infrastructure lacks standardised reporting schemas; (3) budget limitations render commercial AML platforms (approximately USD 0.5–2 M annually) prohibitive; (4) judicial-traceability requirements demand alerts explicitly referencing Criminal Code articles—pure machine-learning "black boxes" fail regulatory-compliance standards. As CBDCs and tokenised assets proliferate across Central Asia, Africa, and the Middle East, sovereign AI capabilities for financial monitoring become critical for systemic stability. This work demonstrates that open-source large language models can achieve operationally useful AML detection while respecting data sovereignty—a model applicable to forty-plus emerging economies facing similar constraints.

We developed AFM Ingestion, a five-layer architecture prototyped during an internship programme at the Agency for Financial Monitoring of the Republic of Kazakhstan (AFM RK). The system combines: (i) multi-bank ingestion with vector-embedding-based format detection (BGE-M3, 1024-dim, cosine similarity threshold 0.92) for automatic parser selection; (ii) a canonical schema with twenty-one standardised fields and semantic enrichment via PostgreSQL+pgvector; (iii) hybrid transaction classification using approximately fifty prioritised regex rules with embedding fallback (threshold 0.35) across seventeen categories; (iv) eight AML typologies with calibrated thresholds and Criminal Code mappings (Art. 232-1, 190, 218), including dropper-transit accounts, cash-out, structuring, and fan-in drop accounts; (v) a natural-language SQL interface using locally hosted Qwen2.5-Coder-14B with retrieval-augmented generation for Russian-language queries.

Evaluated on an anonymised corpus modelled after real export formats from several Kazakhstani commercial banks (institution names withheld for confidentiality; all identifiers replaced before evaluation): format detection achieved 100% accuracy with automatic format learning; transaction classification reached 97.0% accuracy on approximately 20,000 anonymised transactions (98.2% after a documented bug-fix); the natural-language-to-SQL component achieved 89% exact accuracy on a 100-item Russian-language benchmark with 73% self-repair success and approximately 2.5 s median latency; AML typology detection flagged 3.2% of transactions as suspicious with 88% qualitative precision on an analyst-reviewed sample. Counterparty-network visualisation supports rapid identification of fan-in patterns; automated category breakdowns support audit-trail compliance.

Technical contributions include the first openly described on-premise AML pipeline for heterogeneous multi-bank statements, eight interpretable typologies with Criminal Code mappings for judicial traceability, Russian-language NL2SQL achieving 89% exact accuracy using a locally hosted LLM, and a self-learning format registry. For regulators, interpretable monitoring at operationally useful quality is achievable without commercial platforms in the USD 0.5–2 M range while maintaining legal traceability. For policy-makers, the architecture bridges Digital Tenge CBDC monitoring with traditional banking surveillance, extends to crypto-asset oversight relevant to Kazakhstan's regulated market, and transfers to forty-plus emerging economies. As Eurasian Economic Union payment integration deepens, shared AML infrastructure built on open-source LLMs could reduce systemic risk while preserving data sovereignty—balancing FATF compliance with geopolitical cloud-dependency constraints. Future work includes labelled-corpus development, graph-based counterparty analysis, cross-border EAEU typologies, and on-chain integration with regulated crypto exchanges.

***Index Terms*—Anti-Money Laundering, Large Language Models, Generative AI, Financial Intelligence, Data Sovereignty, CBDC Monitoring, Text-to-SQL, Fraud Detection, Emerging Markets, Regulatory Technology**

I. Introduction

Money laundering through fragmented retail banking remains one of the hardest enforcement problems for emerging-market FIUs. Heterogeneous bank-statement layouts, absence of standardised reporting schemas, strict data-residency rules that rule out cloud LLMs, and legal requirements for judicial traceability together rule out almost every off-the-shelf commercial monitoring platform. This paper describes a sovereign, on-premise pipeline—prototyped during an internship programme at the Agency for Financial Monitoring of the Republic of Kazakhstan (AFM RK)—that ingests multi-bank statement exports, normalises them, and emits interpretable AML alerts directly mapped to articles of the Criminal Code of the Republic of Kazakhstan.

We claim four contributions: (a) an open architecture for heterogeneous multi-bank-statement ingestion with self-learning format detection; (b) eight interpretable AML typologies with calibrated thresholds and explicit Criminal-Code mappings; (c) a Russian-language NL2SQL analyst interface using a locally hosted open-source LLM; and (d) an end-to-end evaluation on an anonymised corpus modelled after real bank-statement formats. Section II describes the architecture; Section III details the AML typologies; Section IV reports evaluation; Sections V–VI discuss policy implications and limitations.

II. System Architecture

The pipeline comprises five layers, depicted schematically in Fig. 1.

A. Ingestion and Format Detection

Bank-specific adapters combine with a universal Excel extractor that auto-detects data blocks and header rows in unseen workbook layouts. A format registry stores normalised header signatures as 1024-dimensional BGE-M3 embeddings; on each new file the registry performs a cosine-similarity lookup (threshold 0.92) and either selects a known parser or seeds a new format. This self-learning behaviour reduces the marginal cost of each new bank export.

B. Canonical Schema and Semantic Enrichment

Twenty-one canonical fields (`operation_ts`, `amount_kzt`, `payer_iin_bin`, `purpose_text`, direction, currency, counterparty identifiers, and others) are populated via rule-based header mapping with an embedding fallback (cosine threshold 0.85) for novel column names. Each transaction also carries a concatenated `semantic_text` field embedded with BGE-M3 and stored as a `vector(1024)` column in PostgreSQL via pgvector, enabling later hybrid retrieval.

C. Hybrid Transaction Classification

Approximately fifty prioritised regex rules assign each transaction to one of seventeen operational categories (`P2P_TRANSFER`, `CASH_WITHDRAWAL`, `LOAN_ISSUANCE`, `GAMBLING`, `SALARY`, `FX_OPERATION`, `SECURITIES`, `CONTRACT_SETTLEMENT`, and others). Unmatched transactions fall back to embedding-based nearest-category retrieval against a labelled centroid set (cosine threshold 0.35). Each rule carries a stable identifier, enabling end-to-end traceability from the assigned category back to the lexical pattern that triggered it.

D. AML Typology Detection

Eight typologies, each tied to articles of the Criminal Code of the Republic of Kazakhstan, are evaluated on each account window. Section III describes thresholds and statutory mappings.

E. Natural-Language Analyst Interface

A locally hosted LLM (Qwen2.5-Coder-14B served via Ollama) converts Russian-language analyst questions into single-statement `SELECT` queries over a restricted analyst view. Retrieval-augmented prompting draws sample `purpose_text` values from a semantic catalogue and previously validated NL→SQL pairs from a query-history table (both indexed via pgvector). The entire stack runs on-premise; no transaction data leaves the deployment perimeter—a property essential under Kazakhstan's data-residency regime.

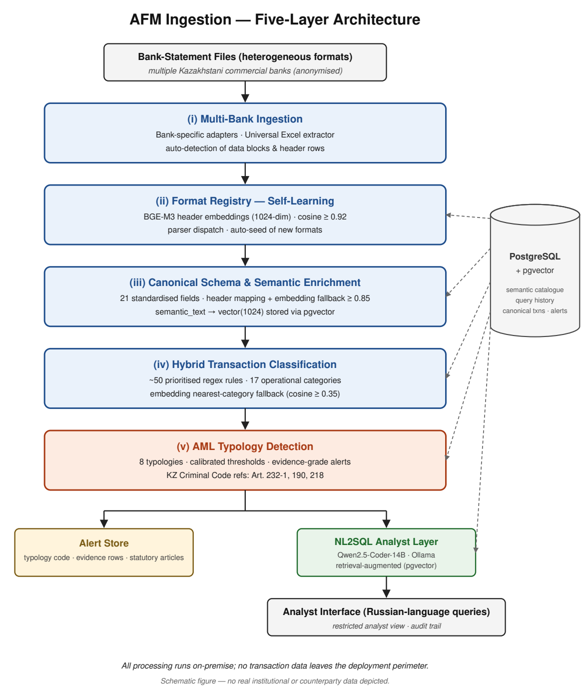

Fig. 1. Five-layer architecture of the AFM Ingestion pipeline. From top to bottom: bank-specific adapters and the universal Excel extractor feed the format registry (pgvector cosine-similarity matching); the canonical mapper produces twenty-one normalised fields and a `semantic_text` embedding; the rule engine assigns one of seventeen operational categories with embedding fallback; eight typology detectors emit alerts with Criminal-Code references; the analyst NL2SQL layer (Qwen2.5-Coder-14B on Ollama, retrieval-augmented from semantic catalogue and query history) serves Russian-language queries against a restricted analyst view. The whole pipeline runs on-premise.

III. AML Typology Detection

Each typology is parameterised by calibrated thresholds and explicitly linked to articles of the Criminal Code of the Republic of Kazakhstan. Each emitted alert carries the triggering typology code, the offending transaction sample, the counterparties implicated, and the applicable statutory articles—a design choice motivated by the need for evidence directly citable in judicial proceedings.

- **Dropper-transit accounts** (Art. 232-1, 190, 218): outgoing/incoming turnover ratio ≥ 0.78; retention ratio ≤ 0.30; same-day-turnover share ≥ 0.30; ≥ 4 distinct credit senders.

- **Cash-out** (Art. 232-1, 218): cash-out share of total outflows ≥ 0.35.
- **Structuring / smurfing** (Art. 190, 218): $\geq 60\%$ of credits below KZT 100,000 with ≥ 4 distinct senders.
- **Fan-in drop account** (Art. 232-1, 218, 190): ≥ 5 distinct senders over ≥ 8 credit transactions into a single account.
- **Rapid balance flush** (Art. 232-1, 218): intraday debit/credit ratio ≥ 0.80 on days with $\geq$ KZT 300,000 inflow.
- **Repeated amount patterns**: multiple credits or debits at near-identical amounts within a short window.
- **High-activity spike**: account-level transaction count or volume exceeding a per-account baseline by a configured factor.
- **Purpose–behaviour mismatch**: stated `purpose_text` inconsistent with observed counterparty and amount profile.

Algorithm 1 illustrates the dropper-transit detector—the most frequently triggered typology in our evaluation—in pseudocode.

Algorithm 1 Dropper-transit detection

Require: account A, time window W, transactions $T(A, W)$
Ensure: alert(A) with typology code and statutory articles

in_total $\leftarrow \sum\{\text{amount} \mid t \in T,\ \text{credit}\}$
out_total $\leftarrow \sum\{\text{amount} \mid t \in T,\ \text{debit}\}$
balance_end $\leftarrow$ balance(A, end of W)
intraday_turn $\leftarrow \sum_{d \in W} \min(\text{in}(A, d), \text{out}(A, d))$
$n_{\text{senders}} \leftarrow |\{\text{distinct } \texttt{payer\_iin\_bin} \mid t \in T,\ \text{credit}\}|$
ratio_out_in $\leftarrow$ out_total$/\max(\text{in_total}, 1)$
retention $\leftarrow$ balance_end$/\max(\text{in_total}, 1)$
same_day $\leftarrow$ intraday_turn$/\max(\text{in_total}, 1)$
if ratio_out_in ≥ 0.78 **and** retention ≤ 0.30 **and** same_day ≥ 0.30 **and** $n_{\text{senders}} \geq 4$ **then**
  emit alert(A, typology=`dropper_transit`, articles=[232-1, 190, 218], evidence=top-$k(T, \text{by amount})$)
end if

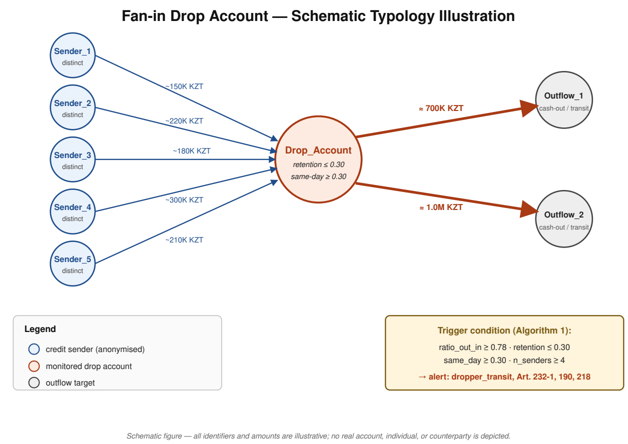

Fig. 2. Schematic illustration of the fan-in drop-account typology. Five anonymised sender accounts (Sender_1 …Sender_5) push credits of comparable magnitude (illustrative values $\sim$ 150–300 thousand KZT) into a single drop account (Drop_Account); the drop account shortly afterwards emits two large debits to Outflow_1 and Outflow_2. Edge thickness encodes amount; node colour encodes role. All identifiers and amounts are illustrative; no real account, individual, or counterparty is depicted.

TABLE I
REPRESENTATIVE NL→SQL EXAMPLES (PARAPHRASED; BENCHMARK ITSELF IS IN RUSSIAN).

#	Intent (paraphrase)	Latency	Correct?
1	Top-10 largest debits in Q2	1.8 s	yes
2	Count of utility transfers in January	2.1 s	yes
3	Find transactions semantically similar to “leasing payout”	3.0 s	yes
4	Cash withdrawals above KZT 1 M	1.6 s	yes
5	Average inflow amount per category	1.9 s	yes
6	Deposit-related transactions for 2024 (self-repair invoked: bare year initially treated as amount; corrected on retry)	3.4 s	yes (after repair)

IV. EVALUATION

Evaluation was conducted on an anonymised corpus modelled after real export formats from several Kazakhstani commercial banks. Bank identities are withheld for confidentiality; all personal identifiers (IIN, BIN, account numbers, names of natural and legal persons, authorisation reference numbers) were replaced before evaluation. Aggregate corpus size is on the order of 10^4 transactions across a multi-year window; per-bank breakdowns are omitted for the same reason.

A. Natural-Language-to-SQL Quality (Russian)

We curated a 100-item Russian-language benchmark spanning five intent classes: aggregate counts, top-N rankings, time-windowed slices, topic search, and counterparty lookups. Table I shows representative examples (intent paraphrased; no real entity names).

B. Aggregate Component Metrics

Table II summarises the four core components. The post-fix transaction-classification figure reflects a corrected regex-ordering bug discovered during evaluation; AML qualitative precision was assessed by manual review on a stratified random sample of flags.

C. Category Distribution After Classification

Table III reports approximate shares of each operational category across the anonymised evaluation corpus, rounded to whole percentage points. The non-zero UNCLASSIFIED residual reflects transactions whose `purpose_text` matched no regex rule and whose embedding-fallback distance exceeded the configured threshold; these are routed to a separate analyst review queue.

TABLE II
AGGREGATE EVALUATION METRICS. COUNTS ARE ROUNDED; PER-BANK, PER-ACCOUNT, AND PER-PERIOD BREAKDOWNS ARE OMITTED FOR CONFIDENTIALITY.

Component	Metric	Value
Format detection	Accuracy on held-out layouts	100%
Format registry	New-format self-seeding rate	1 / unseen layout
Transaction class.	Accuracy ($n \approx 2 \times 10^4$)	97.0% / 98.2% post-fix
NL→SQL	Exact accuracy ($n =$ 100, RU)	89%
NL→SQL	Self-repair success on first failure	73%
NL→SQL	Median end-to-end latency	$\approx 2.5\,$s
AML detection	Suspicious-flag rate	3.2%
AML detection	Qualitative precision (analyst sample)	88%

TABLE III
APPROXIMATE CATEGORY DISTRIBUTION ACROSS THE EVALUATION CORPUS.

Operational category	Share of corpus
P2P_TRANSFER	≈ 22%
CARD_PAYMENT	≈ 18%
CASH_WITHDRAWAL	≈ 11%
INTERNAL_TRANSFER	≈ 10%
SUPPLIER_SETTLEMENT	≈ 8%
SALARY	≈ 7%
TAX_PAYMENT	≈ 6%
LOAN_ISSUANCE/REPAYMENT	≈ 5%
FX_OPERATION	≈ 3%
SECURITIES/DEPOSIT	≈ 2%
GAMBLING/OTHER	≈ 2%
UNCLASSIFIED (fallback)	≈ 6%

V. DISCUSSION AND POLICY IMPLICATIONS

The evaluation supports three claims relevant to emerging-market FIUs. First, heterogeneity of bank-statement layouts is the dominant ingestion challenge: a small embedding-based format registry substantially reduces the marginal cost of each new bank export and removes the need for hand-coded adapters for every institution. Second, eight calibrated, statute-anchored typologies produce alerts that are directly citable in judicial proceedings, addressing the interpretability gap that disqualifies most pure-ML AML systems from regulatory use. Third, a locally hosted open-source LLM is sufficient for Russian-language NL2SQL at quality useful to analysts, removing a major blocker for jurisdictions with strict data-residency rules.

For policy-makers, the architecture suggests that interpretable monitoring at operationally useful quality is achievable without commercial platforms in the USD 0.5–2 M annual range. As Kazakhstan integrates the Digital Tenge with budget execution and public procurement, and as the regulated crypto-asset market expands, a sovereign on-premise pipeline can bridge traditional-banking monitoring with CBDC and on-chain forensics. As Eurasian Economic Union payment integration deepens, shared open-source AML infrastructure could reduce systemic risk while preserving data sovereignty—balancing FATF compliance against geopolitical cloud-dependency constraints.

VI. LIMITATIONS AND FUTURE WORK

Limitations: (i) ground-truth suspicious-versus-not labels are not yet available, so reported precision is qualitative rather than statistical; (ii) typology thresholds are calibrated against the evaluation corpus and may transfer poorly across jurisdictions or customer segments; (iii) the NL2SQL interface has not yet been evaluated with non-technical analyst users in a controlled study; (iv) per-bank evaluation breakdowns and longitudinal stability results are not reported in this paper for confidentiality reasons.

Planned extensions include a labelled evaluation corpus developed in cooperation with domain supervisors; graph-based counterparty analysis for cross-account schemes; extension to cross-border remittance flows under EAEU integration; and integration with on-chain data from regulated Kazakh exchanges, in order to extend typology coverage to the crypto on-ramp layer.

ACKNOWLEDGMENT

The system was prototyped during an internship programme at the Agency for Financial Monitoring of the Republic of Kazakhstan (AFM RK). All evaluation data is anonymised: bank identities are withheld, and all personal and corporate identifiers (IIN, BIN, account numbers, names of natural and legal persons, authorisation reference numbers) were replaced before evaluation. No real counterparty, individual, or transaction figure is reproduced in this paper; all illustrative examples and schematic figures use synthetic identifiers and indicative magnitudes only. Views expressed are those of the authors and do not necessarily represent the position of AFM RK or the authors' institution.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.
[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technologies (NAACL-HLT)*, vol. 1, Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186.
[3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, Dec. 2020, pp. 1877–1901.
[4] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, Jun. 2020.
[5] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale *et al.*, "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint*, arXiv:2307.09288, Jul. 2023.

[6] J. Bai, S. Bai, Y. Chu, Z. Cui, K. Dang, X. Deng, Y. Fan, W. Ge, Y. Han, F. Huang *et al.*, "Qwen technical report," *arXiv preprint*, arXiv:2309.16609, Sep. 2023.

[7] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier *et al.*, "Mistral 7B," *arXiv preprint*, arXiv:2310.06825, Oct. 2023.

[8] V. Zhong, C. Xiong, and R. Socher, "Seq2SQL: Generating structured queries from natural language using reinforcement learning," *arXiv preprint*, arXiv:1709.00103, Aug. 2017.

[9] T. Yu, R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Roman, Z. Zhang, and D. Radev, "Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task," in *Proc. 2018 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, Oct.–Nov. 2018, pp. 3911–3921.

[10] J. Li, B. Hui, G. Qu, J. Yang, B. Li, B. Wang, B. Qin, R. Cao, R. Geng, N. Huo, X. Zhou, C. Ma, K. C. C. Chang, F. Huang, R. Cheng, and Y. Li, "Can LLM already serve as a database interface? A big bench for large-scale database grounded text-to-SQLs," in *Proc. Advances in Neural Information Processing Systems (NeurIPS) — Datasets and Benchmarks Track*, vol. 36, New Orleans, LA, USA, Dec. 2023, pp. 42330–42357.

[11] M. Pourreza and D. Rafiei, "DIN-SQL: Decomposed in-context learning of text-to-SQL with self-correction," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, New Orleans, LA, USA, Dec. 2023, pp. 36339–36348.

[12] D. Gao, H. Wang, Y. Li, X. Sun, Y. Qian, B. Ding, and J. Zhou, "Text-to-SQL empowered by large language models: A benchmark evaluation (DAIL-SQL)," *Proc. VLDB Endowment*, vol. 17, no. 5, pp. 1132–1145, Jan. 2024.

[13] B. Wang, C. Ren, J. Yang, X. Liang, J. Bai, L. Chai, Z. Yan, Q.-W. Zhang, D. Yin, X. Sun, and Z. Li, "MAC-SQL: A multi-agent collaborative framework for text-to-SQL," in *Proc. 31st Int. Conf. Computational Linguistics (COLING)*, Abu Dhabi, UAE, Jan. 2025, pp. 540–557.

[14] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proc. 2019 Conf. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, Nov. 2019, pp. 3982–3992, doi: 10.18653/v1/D19-1410.

[15] J. Chen, S. Xiao, P. Zhang, K. Luo, D. Lian, and Z. Liu, "BGE M3-Embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation," in *Findings of the Association for Computational Linguistics: ACL 2024*, Bangkok, Thailand, Aug. 2024, pp. 2318–2335.

[16] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, Dec. 2020, pp. 9459–9474.

[17] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," *arXiv preprint*, arXiv:1908.10063, Aug. 2019.

[18] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann, "BloombergGPT: A large language model for finance," *arXiv preprint*, arXiv:2303.17564, Mar. 2023.

[19] H. Yang, X.-Y. Liu, and C. D. Wang, "FinGPT: Open-source financial large language models," in *Proc. FinLLM Symposium at IJCAI 2023*, Macao, China, Aug. 2023.

[20] M. Weber, G. Domeniconi, J. Chen, D. K. I. Weidele, C. Bellei, T. Robinson, and C. E. Leiserson, "Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics," in *Proc. KDD '19 Workshop on Anomaly Detection in Finance*, Anchorage, AK, USA, Aug. 2019, pp. 1–7.

[21] E. Altman, J. Blanuša, L. von Niederhäusern, B. Egressy, A. Anghel, and K. Atasu, "Realistic synthetic financial transactions for anti-money laundering models," in *Proc. 37th Conf. Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, New Orleans, LA, USA, Dec. 2023, pp. 29851–29874.

[22] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. B. Schardl, and C. E. Leiserson, "EvolveGCN: Evolving graph convolutional networks for dynamic graphs," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 34, no. 4, New York, NY, USA, Feb. 2020, pp. 5363–5370, doi: 10.1609/aaai.v34i04.5984.

[23] M. Cardoso, P. Saleiro, and P. Bizarro, "LaundroGraph: Self-supervised graph representation learning for anti-money laundering," in *Proc. 3rd ACM Int. Conf. AI in Finance (ICAIF '22)*, New York, NY, USA, Nov. 2022, pp. 130–138, doi: 10.1145/3533271.3561727.

[24] B. Egressy, L. von Niederhäusern, J. Blanuša, E. Altman, R. Wattenhofer, and K. Atasu, "Provably powerful graph neural networks for directed multigraphs," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 38, no. 10, Vancouver, BC, Canada, Feb. 2024, pp. 11838–11846, doi: 10.1609/aaai.v38i10.29069.

[25] A. N. Eddin, J. Bono, D. Aparício, D. Polido, J. T. Ascensão, P. Bizarro, and P. Ribeiro, "Anti-money laundering alert optimization using machine learning with graphs," in *Proc. AAAI Workshop on AI in Financial Services: Adaptiveness, Resilience & Governance*, Feb. 2022, pp. 1–9.

[26] J. Lin, X. Guo, Y. Zhu, S. Mitchell, E. Altman, and J. Shun, "FraudGT: A simple, effective, and efficient graph transformer for financial fraud detection," in *Proc. 5th ACM Int. Conf. AI in Finance (ICAIF '24)*, Brooklyn, NY, USA, Nov. 2024, pp. 292–300, doi: 10.1145/3677052.3698648.

[27] Z. Chen, L. D. Van Khoa, E. N. Teoh, A. Nazir, E. K. Karuppiah, and K. S. Lam, "Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review," *Knowledge and Information Systems*, vol. 57, no. 2, pp. 245–285, Nov. 2018, doi: 10.1007/s10115-017-1144-z.

[28] M. Jullum, A. Løland, R. B. Huseby, G. Ånonsen, and J. Lorentzen, "Detecting money laundering transactions with machine learning," *Journal of Money Laundering Control*, vol. 23, no. 1, pp. 173–186, Jan. 2020, doi: 10.1108/JMLC-07-2019-0055.

[29] F. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, and S. Ferretti, "Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics," *Electronic Markets*, vol. 33, no. 1, art. 37, Jul. 2023, doi: 10.1007/s12525-023-00654-3.

[30] B. Oztaş, D. Cetinkaya, F. Adedoyin, M. Budka, G. Aksu, and H. Dogan, "Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry," *Future Generation Computer Systems*, vol. 159, pp. 161–171, Oct. 2024, doi: 10.1016/j.future.2024.05.027.

[31] F. Labanca, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, "Amaretto: An active learning framework for money laundering detection," *IEEE Access*, vol. 10, pp. 41720–41735, Apr. 2022, doi: 10.1109/ACCESS.2022.3167699.

[32] J. Lorenz, M. I. Silva, D. Aparício, J. T. Ascensão, and P. Bizarro, "Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity," in *Proc. 1st ACM Int. Conf. AI in Finance (ICAIF '20)*, New York, NY, USA, Oct. 2020, pp. 1–8, doi: 10.1145/3383455.3422549.

[33] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 1024–1034.

[34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learning Representations (ICLR)*, Toulon, France, Apr. 2017, pp. 1–14.

[35] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. 6th Int. Conf. Learning Representations (ICLR)*, Vancouver, BC, Canada, Apr./May 2018, pp. 1–12.

[36] L. Bellomarini, E. Sallinger, and G. Gottlob, "The Vadalog System: Datalog-based reasoning for knowledge graphs," *Proc. VLDB Endowment*, vol. 11, no. 9, pp. 975–987, May 2018, doi: 10.14778/3213880.3213888.

[37] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint*, arXiv:1901.03407, Jan. 2019.

[38] K. Singh and P. Best, "Anti-money laundering: Using data visualization to identify suspicious activity," *International Journal of Accounting Information Systems*, vol. 34, art. 100418, Sep. 2019, doi: 10.1016/j.accinf.2019.06.002.

[39] Financial Action Task Force and Egmont Group, "Concealment of beneficial ownership," FATF, Paris, France, Tech. Rep., Jul. 2018.

[40] Financial Action Task Force, "Money laundering and terrorist financing through the physical transportation of cash," FATF, Paris, France, Tech. Rep., Oct. 2015.

[41] Eurasian Group on Combating Money Laundering and Financing of Terrorism (EAG), "Legalisation (laundering) of the proceeds of cybercrime, including through the use of electronic money or virtual assets and the infrastructure of their providers," EAG Working Group on Typologies, Moscow, Russia, Typology Rep., 2021.

[42] Bank for International Settlements, “CBDCs: an opportunity for the monetary system,” in *BIS Annual Economic Report 2021*, ch. III, Basel, Switzerland, Jun. 2021, pp. 65–95.
[43] R. Auer and R. Böhme, “The technology of retail central bank digital currency,” *BIS Quarterly Review*, Bank for International Settlements, Basel, Switzerland, pp. 85–100, Mar. 2020.
[44] R. Auer, G. Cornelli, and J. Frost, “Rise of the central bank digital currencies: Drivers, approaches and technologies,” *BIS Working Papers*, no. 880, Bank for International Settlements, Basel, Switzerland, Aug. 2020.
[45] R. Auer and R. Böhme, “Central bank digital currency: The quest for minimally invasive technology,” *BIS Working Papers*, no. 948, Bank for International Settlements, Basel, Switzerland, Jun. 2021.
[46] Bank for International Settlements Innovation Hub, “Project Tourbillon: Exploring privacy, security and scalability for CBDCs,” BIS, Basel, Switzerland, Final Report, Nov. 2023.
[47] BIS Innovation Hub Hong Kong Centre, Hong Kong Monetary Authority, Bank of Thailand, Digital Currency Institute of the People’s Bank of China, and Central Bank of the United Arab Emirates, “Project mBridge: Connecting economies through CBDC,” Bank for International Settlements, Basel, Switzerland, Joint Report, Oct. 2022.
[48] T. Adrian and T. Mancini-Griffoli, “The rise of digital money,” *IMF FinTech Notes*, no. 19/001, International Monetary Fund, Washington, DC, USA, Jul. 2019.
[49] T. Mancini-Griffoli, M. S. Martinez Peria, I. Agur, A. Ari, J. Kiff, A. Popescu, and C. Rochon, “Casting light on central bank digital currency,” *IMF Staff Discussion Note*, no. SDN/18/08, International Monetary Fund, Washington, DC, USA, Nov. 2018.
[50] European Central Bank, “Report on a digital euro,” Eurosystem High-Level Task Force on Central Bank Digital Currency, Frankfurt am Main, Germany, Oct. 2020.
[51] National Bank of Kazakhstan and National Payment Corporation of Kazakhstan, “Digital Tenge Project Report (White Paper): Results of feasibility study of introduction of Digital Tenge,” NBK, Astana, Kazakhstan, Final Report, Dec. 2022.
[52] G. Impavido, “The Kazakhstan Digital Tenge Project,” *IMF Staff Country Reports*, vol. 2024, no. 047, Selected Issues Paper, International Monetary Fund, Washington, DC, USA, Feb. 2024.
[53] A. Kumar, A. Chhangani, and J. Lipsky, Eds., “Central Bank Digital Currency Tracker,” Atlantic Council GeoEconomics Center, Washington, DC, USA, Jul. 2025.
[54] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, vol. 54, *Proceedings of Machine Learning Research*, Apr. 2017, pp. 1273–1282.
[55] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, “Deep learning with differential privacy,” in *Proc. 2016 ACM SIGSAC Conf. Computer and Communications Security (CCS ’16)*, Vienna, Austria, Oct. 2016, pp. 308–318, doi: 10.1145/2976749.2978318.
[56] The Astana Times, “Digital Tenge enhances budget transparency in Kazakhstan,” 2025. [Online]. Available: https://astanatimes.com/2025/07/digital-tenge-enhances-budget-transparency-in-kazakhstan/
[57] The Astana Times, “Kazakhstan’s regulated crypto market reaches USD 6.8 billion in 2025,” 2025. [Online]. Available: https://astanatimes.com/
[58] CoinDesk, “Fonte Capital launches Central Asia’s first spot Bitcoin ETF on the Astana International Exchange,” 12 Aug. 2025. [Online]. Available: https://www.coindesk.com/
[59] Cointelegraph, “BitGo backs Central Asia’s first spot Bitcoin ETF in Kazakhstan,” 2025. [Online]. Available: https://cointelegraph.com/