NO PREP

AI in Banking

Fraud Detection and Auto-Approvals

Includes: Student reading, case study, differentiated questions, teacher key

MORE LESSONS LESS PLANNING · MLLP · FREE TODAY · FREE TOMORROW · FREE FOREVER

Name: Period: Date:

AI in Banking: Fraud Detection and Auto-Approvals

The Invisible Security Layer in Your Wallet

Every time you swipe your credit card, something happens in the background that most people never notice: a machine learning model processes your transaction in milliseconds and decides whether it looks legitimate. If the model is confident the transaction is normal, it approves it instantly. If it detects something unusual — a purchase in a city you've never visited, a transaction at 3am, a purchase amount far larger than your typical spending — it can flag it, decline it, or trigger a secondary verification step. This system is called real-time fraud detection, and it is one of the most consequential AI deployments in everyday life. In 2023, credit card fraud losses globally exceeded $33 billion. Without AI, those numbers would be dramatically higher. The models running inside your bank's fraud detection layer are some of the most sophisticated machine

Name: **Period:** **Date:**

AI in Banking: Fraud Detection and Auto-Approvals

The Invisible Security Layer in Your Wallet

Every time you swipe your credit card, something happens in the background that most people never notice: a machine learning model processes your transaction in milliseconds and decides whether it looks legitimate. If the model is confident the transaction is normal, it approves it instantly. If it detects something unusual (a purchase in a city you've never visited, a transaction at 3am, a purchase amount far larger than your typical spending), it can flag it, decline it, or trigger a secondary verification step. This system is called real-time fraud detection, and it is one of the most consequential AI deployments in everyday life. In 2023, credit card fraud losses globally exceeded $33 billion. Without AI, those numbers would be dramatically higher. The models running inside your bank's fraud detection layer are some of the most sophisticated machine learning systems in operation, because the stakes, literally measured in billions of dollars, are high enough to justify enormous investment.

How Fraud Detection AI Works

Fraud detection AI is built on anomaly detection: the model learns what 'normal' looks like for you specifically, and then flags anything that deviates significantly from that pattern. Your normal includes things like: the geographic range where you typically spend money, the time of day you usually make purchases, the types of merchants you frequent, your typical transaction amounts, and the pace at which you normally make purchases (one or two transactions per day vs. a sudden burst of 15 transactions in 30 minutes). The model is trained on enormous datasets of historical transactions, both legitimate and fraudulent. It learns the patterns that distinguish fraud from legitimate spending at a population level, then calibrates those patterns to your individual profile. The more transaction history the model has on you, the

better it gets at distinguishing your unusual-but-legitimate spending (you bought an expensive gift for a birthday) from genuinely suspicious activity (someone stole your card number and is buying gift cards in bulk).

The Cost of False Positives

A fraud detection model can make two types of mistakes. A false negative is when the model approves a fraudulent transaction it should have flagged. This costs the bank and cardholders money. A false positive is when the model declines a legitimate transaction it incorrectly flagged as suspicious. This doesn't cost money directly. But it has real costs in customer frustration, missed purchases, and eroded trust. Banks have to balance these two error types, and the balance matters. A model set extremely conservatively (flag anything unusual) will catch more fraud but will also block legitimate purchases constantly, making the card useless for travel or unusual spending. A model set too permissively will let fraud slip through. Most banks try to optimize for something like: 'catch the vast majority of fraud while keeping the false positive rate low enough that normal customers rarely notice.' The false positive problem disproportionately affects certain customers. Research has found that transactions associated with certain types of merchants, certain geographic areas, and certain spending patterns more common among lower-income consumers tend to trigger higher false positive rates. In other words, the 'normal' that the model was trained on is based on the majority of the dataset, which may not reflect minority spending patterns accurately.

Credit Scoring and Automated Loan Decisions

Beyond fraud detection, AI plays a major role in credit scoring — the system that determines whether you qualify for a loan, credit card, or mortgage and at what interest rate. The traditional FICO score is a formula-based system using a handful of variables: payment history, amounts owed, length of credit history, new credit, and credit mix. AI-based credit scoring goes further, processing hundreds or thousands of variables and using machine learning to find patterns in that data that predict creditworthiness. Companies like Upstart and ZestFinance have built alternative credit models that incorporate non-traditional variables — educational background, employment history, even social network connections in some cases. These models claim to approve more applicants, especially those with thin or nonexistent credit files (common among young people and recent immigrants), while

maintaining comparable default rates. Proponents argue this expands credit access; critics worry about which non-traditional variables are correlated with protected characteristics like race or national origin.

Algorithmic Lending and Discrimination

The possibility that AI credit models discriminate — even unintentionally — is one of the most actively litigated issues in fintech regulation. The Equal Credit Opportunity Act (ECOA) prohibits lenders from discriminating based on race, sex, national origin, and other protected characteristics. But AI models can learn to use variables that are merely correlated with those characteristics and produce discriminatory outcomes without explicitly encoding race or gender at all. This is called proxy discrimination. In 2019, Apple and Goldman Sachs faced allegations that their Apple Card credit algorithm gave significantly lower credit limits to women than to men with comparable financial profiles — including in some cases where spouses shared finances. The companies denied intentional discrimination and cited algorithm design, but the incident prompted a formal investigation by the New York Department of Financial Services. No explicit bias was found in the stated algorithm, but the investigation highlighted how opaque AI systems make it nearly impossible to verify fairness claims from the outside.

How Fraud Detection Models Are Trained

The machine learning models that banks use to detect fraud are trained on massive historical datasets of transactions labeled as either legitimate or fraudulent. The training process involves feeding the algorithm millions of examples so it can identify the statistical patterns that distinguish normal spending from suspicious activity. One persistent challenge is class imbalance: fraudulent transactions represent a tiny fraction of all transactions — often less than one tenth of one percent — which means the model sees vastly more examples of normal behavior than fraud. Without careful handling, a model trained on imbalanced data might classify everything as legitimate just to achieve high accuracy, which is catastrophically wrong for fraud prevention purposes. Data scientists use techniques like oversampling fraudulent cases, undersampling legitimate ones, or generating synthetic fraud examples to ensure the model develops genuine sensitivity to the patterns it most needs to catch.

The Regulatory Landscape for AI in Financial Services

The use of AI in banking is heavily regulated, and the regulatory environment is evolving rapidly as the technology advances. In the United States, the Equal Credit Opportunity Act and the Fair Housing Act prohibit lending decisions that discriminate based on protected characteristics, and these laws apply to algorithmic systems just as they do to human loan officers. Banks must be able to explain adverse credit decisions to applicants — a requirement that creates direct tension with the use of complex machine learning models whose decision logic is not easily interpretable. The European Union's AI Act, which entered into force in 2024, classifies AI systems used in credit scoring as high-risk, requiring transparency, human oversight, and regular auditing. These regulatory frameworks are shaping how banks design and deploy their AI systems, often pushing them toward more interpretable models even when opaque ones might perform better by narrow accuracy metrics. Regulatory compliance has become a central design constraint, not an afterthought, in how financial institutions build and deploy AI-powered credit and fraud systems.

Key Information: AI in Banking

Detail	Information
Global Fraud Losses (2023)	$33 billion in credit card fraud globally
Fraud Detection Method	Anomaly detection — compares each transaction against the user's personal baseline pattern
False Negative	A fraudulent transaction the model incorrectly approves
False Positive	A legitimate transaction the model incorrectly declines
Traditional Credit Score	FICO score — based on 5 variables (payment history, amounts owed, history length, new credit, mix)
AI Credit Scoring	Processes hundreds of variables; companies like Upstart use non-traditional factors

Detail	Information
Proxy Discrimination	When an AI uses variables correlated with protected characteristics, producing discriminatory outcomes indirectly
Apple Card Investigation	2019 allegations of gender-based credit limit disparities; investigated by NY Dept. of Financial Services

Case Study: Kezia's Declined Card

Kezia is 19 and just started college in Boston after growing up in rural Mississippi. She has a debit card linked to her checking account and a new student credit card with a $500 limit. It's October, and she's traveling for the first time to visit a friend in Chicago. At a restaurant in Chicago, her credit card is declined. She tries again. Declined. She pays with her debit card, embarrassed, and calls her bank when she gets back to the hotel. The customer service rep tells her the transaction was flagged by their fraud detection system. 'It was a purchase in a state you've never spent in before, at a restaurant, in the evening,' the rep explains. 'The system saw that as unusual.' Kezia isn't sure whether to be impressed or angry. 'So my card got blocked because I traveled?' she asks. 'Sort of,' the rep says. 'The model didn't know you were traveling.'

1. Using what you know about how fraud detection AI works, explain precisely why Kezia's transaction was flagged. What signals did the anomaly detection model respond to?

2. Kezia says she's not sure whether to be 'impressed or angry.' What is the strongest argument for each reaction? Which do you find more compelling, and why?

Multiple Choice

1. What is a 'false positive' in the context of fraud detection AI?

   a) When the AI incorrectly identifies a merchant as high-risk

   b) When the AI declines a transaction that is actually legitimate

   c) When the AI approves a transaction that is actually fraudulent

   d) When the AI fails to update a user's spending profile after a major life change

2. What is 'proxy discrimination' in the context of AI credit scoring?

   a) When an algorithm applies different interest rates to different geographic regions to manage default risk based on regional historical lending data

   b) When a credit model is trained on a dataset that is too small to support statistically accurate predictions

   c) When an AI uses variables correlated with protected characteristics and produces discriminatory outcomes without encoding race or gender directly

   d) When a lender explicitly instructs an algorithm to treat applicants differently based on their race

3. What did the 2019 Apple Card investigation reveal about AI fairness in financial services?

a) That Apple had intentionally programmed explicit bias into its credit algorithm to discriminate

b) That the Equal Credit Opportunity Act does not apply to AI-based lending decisions

c) That AI credit models consistently outperform traditional FICO scoring methods on all key equity metrics

d) That opaque AI systems make it extremely difficult to verify fairness claims from outside the company

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ Anomaly Detection	A) A traditional formula-based credit score using five variables including payment history and amounts owed
2. ___ False Positive	B) A fraud detection method that learns what 'normal' looks like for a specific user and flags significant deviations
3. ___ Proxy Discrimination	C) A fraud detection error where a legitimate transaction is incorrectly declined as suspicious
4. ___ FICO Score	D) A fintech company using machine learning to score creditworthiness using non-traditional variables beyond the FICO model
5. ___ Upstart	E) Discriminatory outcomes produced when an AI uses variables merely correlated with protected characteristics

Fill-in-the-Blank

1. Credit card fraud losses globally exceeded _____ billion dollars in 2023.

2. A fraud detection model that declines a legitimate transaction has committed a false _____ error.

3. The Equal Credit Opportunity Act (ECOA) prohibits lenders from discriminating based on race, sex, national origin, and other _____ characteristics.

Short Answer

Answer each question in 2-4 complete sentences.

1. Explain how anomaly detection works in fraud AI. Why does the model get better the longer you use your card?

_____

_____

_____

2. The reading says false positives 'disproportionately affect certain customers.' Explain why this happens and what it reveals about a potential fairness problem with fraud detection AI.

_____

_____

_____

3. Should banks be required to explain to customers exactly why an AI model declined their transaction or credit application? What are the arguments for and against this kind of algorithmic transparency?