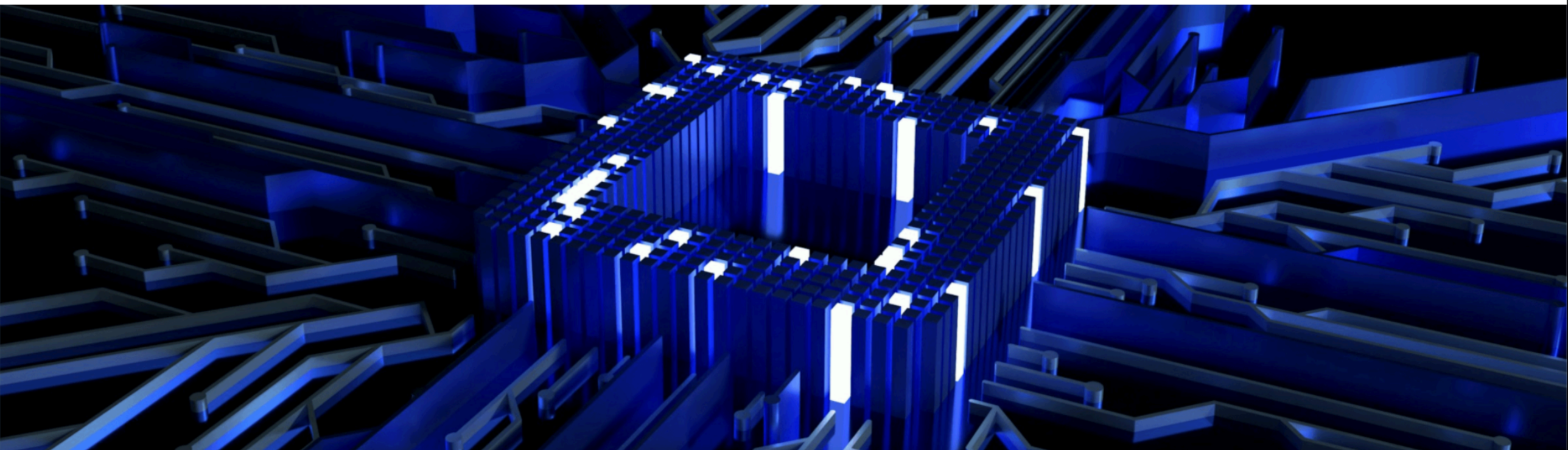
Amazon Trust: A Machine Learning Project for Amazon HackOn 2025

Presenter: Nivesh Jain and Yash Agarwal

01 Problem Statement



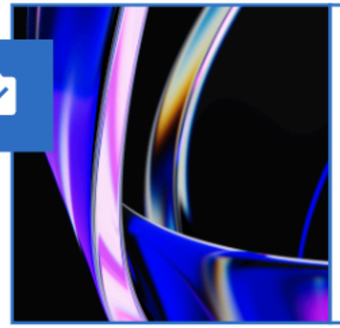
Understanding the Challenge

Rise of Fake Buyers

The rise of automated **buying bots** has introduced a new wave of challenges for e-commerce platforms. These bots simulate real customer behavior but operate at high speed and scale, often buying up stock during flash sales, abusing coupon systems, and manipulating inventory levels. Their behavior is subtle yet disruptive, frequently bypassing basic detection mechanisms. As a result, genuine customers are left with poor experiences, and sellers struggle to maintain fair access to their products.

Rise of Fake Reviews

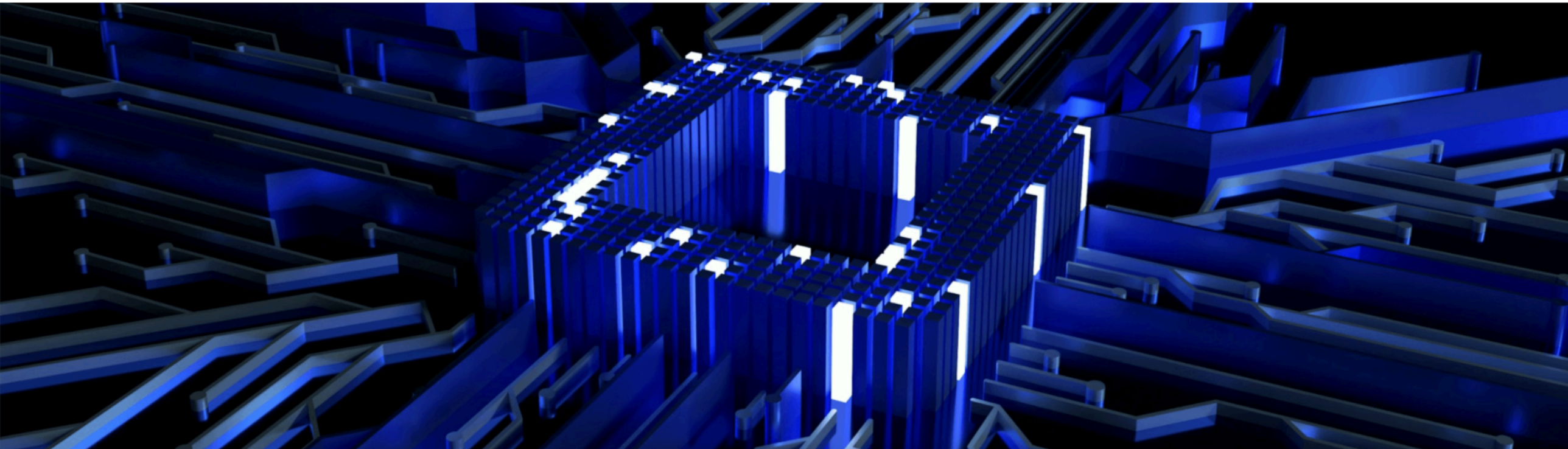
Fake reviews are flooding product listings, often generated using scripts, AI models, or through hijacked accounts. These reviews create a misleading perception of product quality and customer satisfaction. Buyers are misled into making decisions based on false narratives, while honest sellers face reputational damage due to manipulated ratings. Traditional moderation systems fail to scale with the volume and sophistication of such fake content.



Impact on Platform

The combined effect of fake buyers and fake reviews severely **undermines trust** on the Amazon platform. It leads to distorted product rankings, unfair competitive advantages, and an unreliable shopping experience. Over time, this erodes consumer confidence, increases return rates, and harms the brand reputation of both Amazon and legitimate third-party sellers. To maintain marketplace integrity, there is an urgent need for adaptive, ML-driven solutions that can detect and mitigate these forms of fraud in real time.

02 Objectives



Significance of the Project

Strengthen User Trust

Trust is the foundation of any successful e-commerce platform. By accurately identifying fake reviews and bot-driven purchases, this project helps Amazon reinforce the authenticity of its marketplace. Shoppers are more likely to rely on product ratings and reviews when they know fraudulent behavior is being actively detected and prevented.

01

Enhancing User Experience

Enhancing user experience is crucial for maintaining customer loyalty and satisfaction. By leveraging machine learning, this project aims to offer personalized recommendations, streamline navigation, and improve search relevance, ultimately creating a more engaging and efficient shopping journey for Amazon users.

02

Protecting Brand Integrity

Protecting brand integrity is crucial for Amazon's reputation and customer trust. This project leverages machine learning to proactively identify and mitigate risks, ensuring consistent brand representation and preventing counterfeit issues, thereby enhancing consumer confidence and loyalty in Amazon's marketplace.

03

Project Goals

Develop Detection Models

To enhance the security and reliability of the Amazon platform, this project focuses on developing comprehensive machine learning models that detect fraudulent activities such as fake reviews and bot-driven purchases. By analyzing user behavior, transaction patterns, and review characteristics, the system enables rapid, automated identification of anomalies. This helps protect Amazon's marketplace integrity, maintain customer trust, and reduce the impact of deceptive practices on both buyers and sellers.

01

Implement User-Friendly UI

The project delivers a user-friendly interface built with Streamlit, allowing seamless interaction with the detection system. The UI simplifies tasks such as data generation, model training, and real-time prediction, ensuring accessibility for analysts, developers, and testers. This focus on usability ensures that fraud detection becomes not only effective but also easy to operate and scale within real-world e-commerce workflows.

02

Expected Outcomes

High Detection Accuracy

Achieving high detection accuracy is crucial for our machine learning model to effectively identify anomalies and enhance user trust. This outcome will not only reduce false positives but also improve customer experience, ultimately driving higher engagement and satisfaction on the Amazon platform.

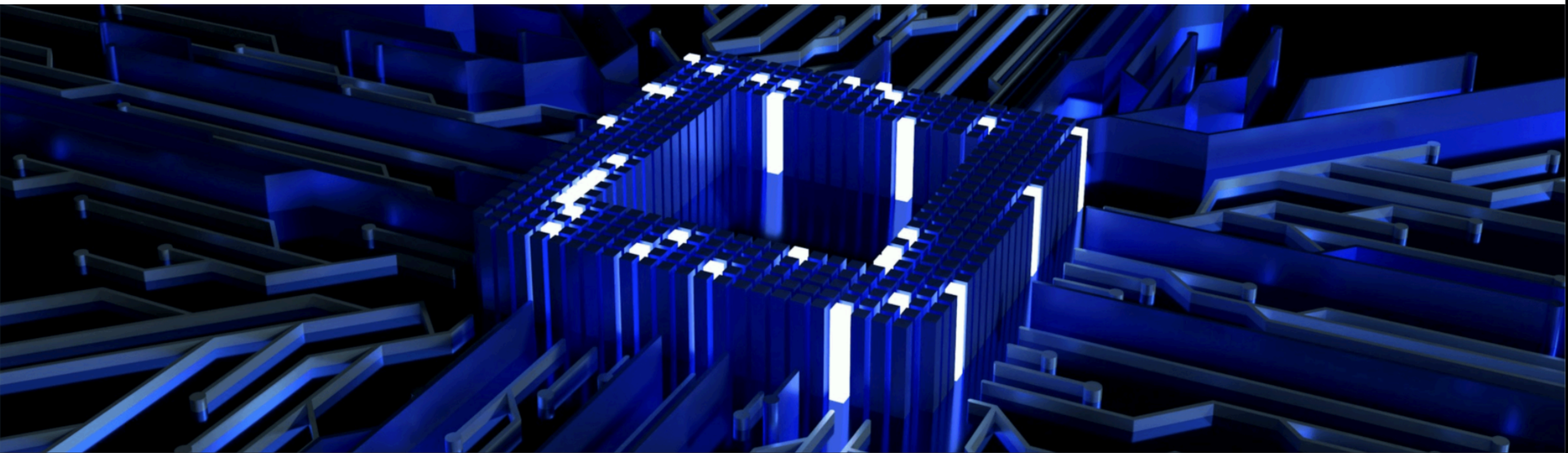
Instant Behavioral Insights

The system will deliver fast, actionable insights through an interactive interface, allowing fraud analysts and developers to visualize suspicious activity as it occurs. This immediate feedback enables quicker testing, evaluation, and model refinement, improving operational efficiency and supporting data-driven improvements.

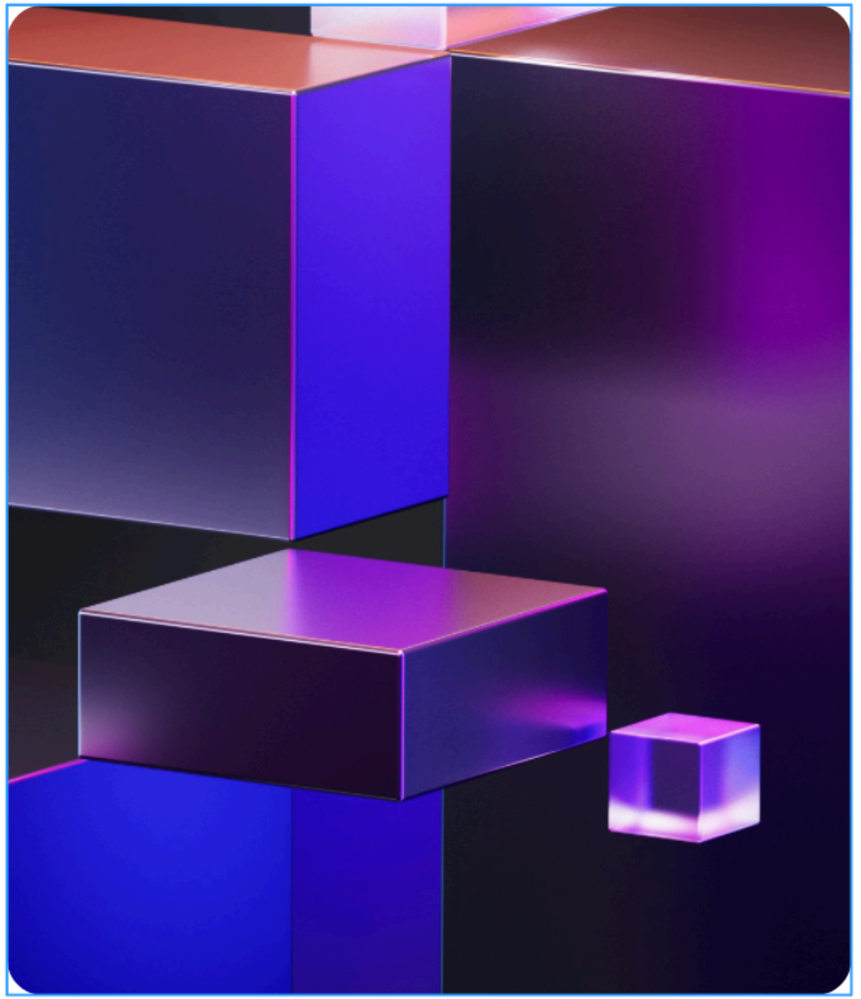
Scalable Fraud Detection Framework

The project aims to create a modular and scalable framework that can be easily extended to new products, categories, or geographies. This adaptability ensures the solution remains effective as Amazon evolves, allowing future integration with seller analysis, regional fraud trends, or multilingual review detection.

03 System Architecture



Architecture Overview



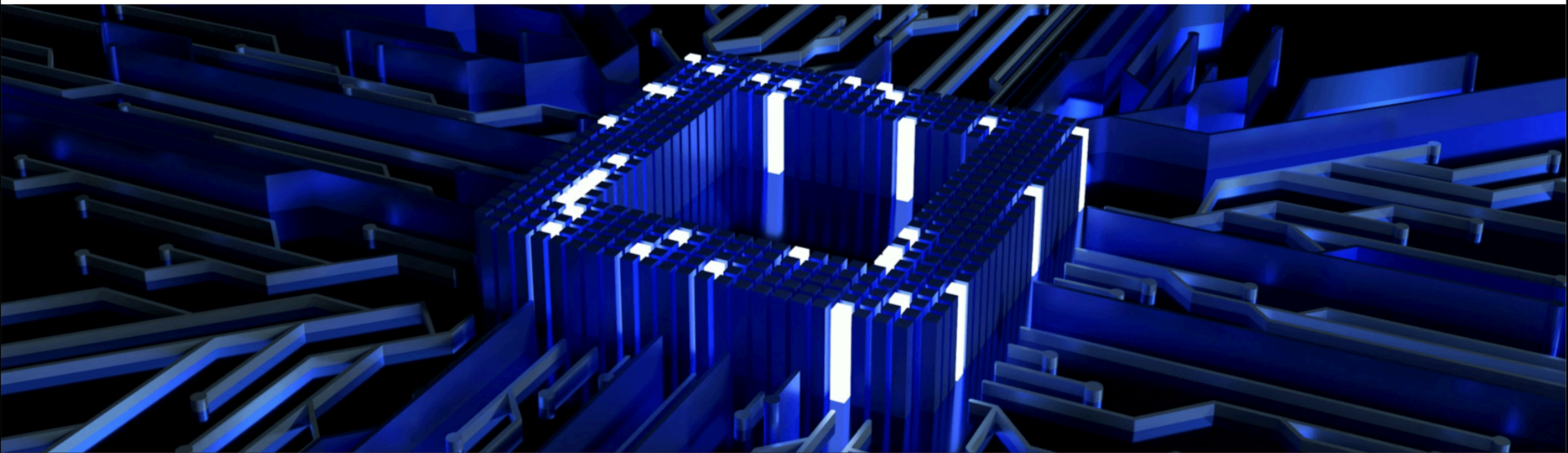
1. Modular Design

The system is built using a modular architecture with two main components: the Fake Review Classifier and the Bot Buying Classifier. Each module operates independently but follows a consistent structure involving data generation, feature extraction, model training, prediction, and evaluation. This design ensures flexibility, ease of maintenance, and straightforward scaling for future fraud detection tasks.

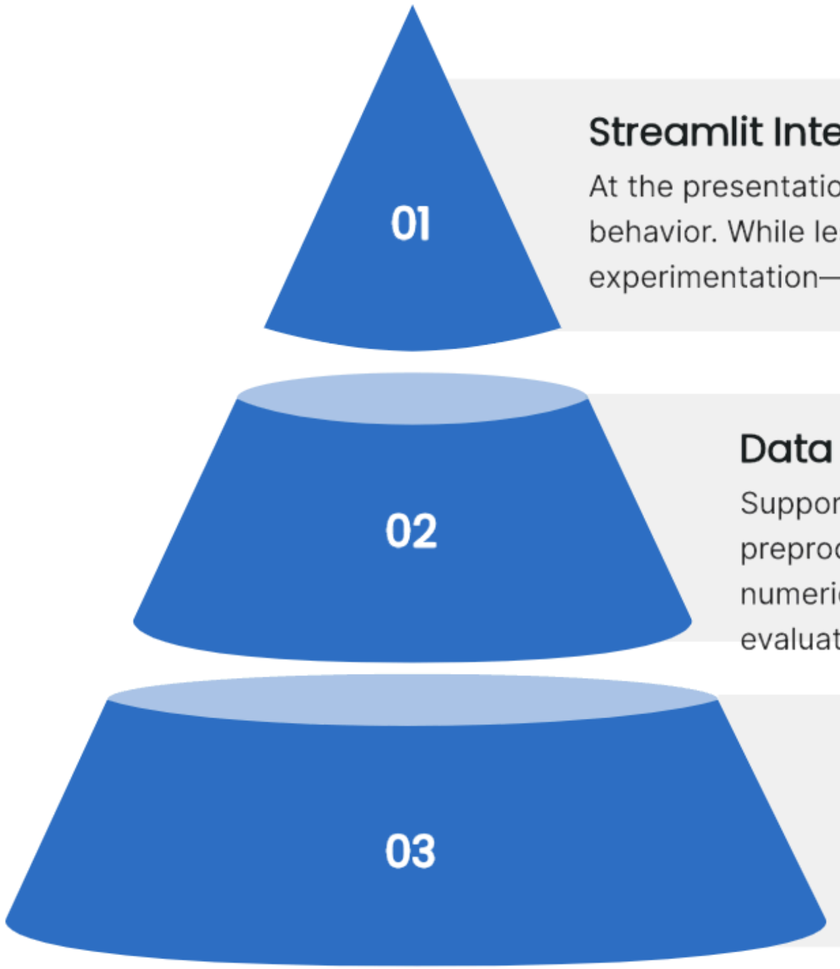
2. Streamlit-Based Interface

A unified Streamlit interface connects the backend logic to the user, enabling seamless interaction with both modules. Users can input simulated data, trigger model training, view real-time predictions, and analyze performance metrics—all through a clean, accessible UI. This tight integration between frontend and backend enhances usability and supports rapid experimentation.

04 Module 1: Bot Detection



Technology Stack



Streamlit Interface and integration

At the presentation layer, the system uses Streamlit to provide an interactive UI for testing, monitoring, and simulating behavior. While less critical to detection accuracy, this layer enhances usability, accessibility, and ease of experimentation—making the tool practical for developers, analysts, and hackathon demonstrations.

Data Processing and Featuire Engineering

Supporting the models is a robust layer of data handling using Pandas, NumPy, and Scikit-learn preprocessing tools. This includes extracting meaningful behavioral features from user actions, scaling numerical inputs, encoding labels, and organizing structured Excel-based datasets for training and evaluation.

Machine Learning Models

At the core of the system are advanced ML algorithms like Random Forest, LightGBM, and ensemble models, which drive the detection of fraudulent reviews and bot activity. These models are responsible for pattern recognition, classification, and risk scoring, making them the most critical layer of the technology stack.

Bot Buying Classifier

Step1

ML-Based Detection Strategy

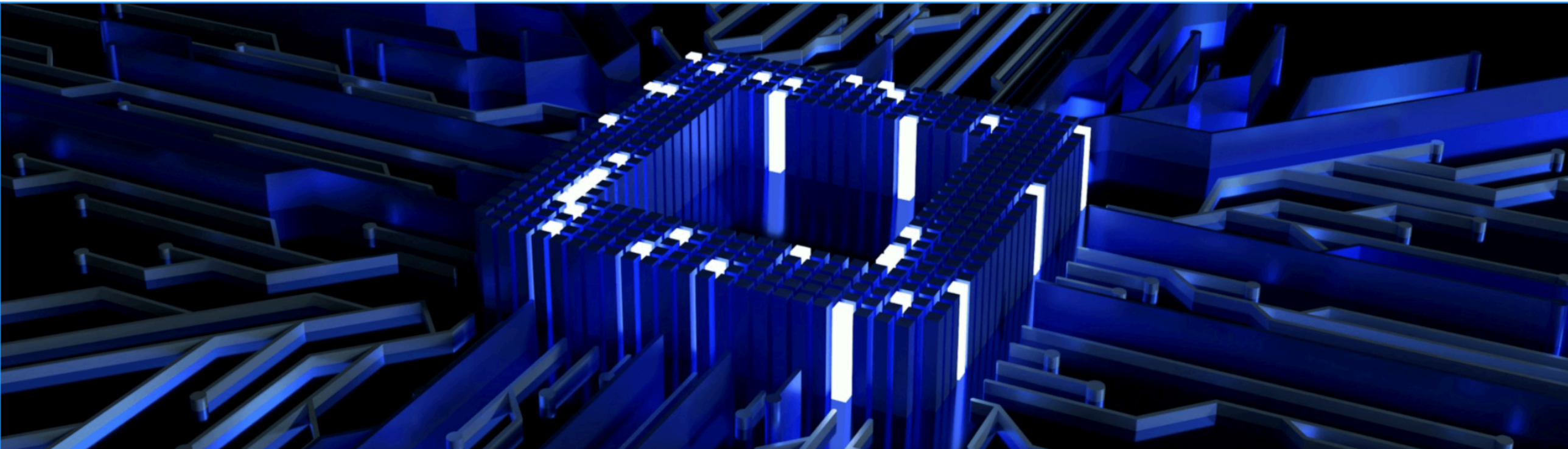
The Bot Buying Classifier uses machine learning to detect behavior anomalies that indicate automated buying. By analyzing metrics like purchase frequency, session timing, and coupon usage patterns, the model distinguishes between genuine user actions and those generated by bots. The approach is adaptive, learning from both simulated and historical data to improve fraud detection over time.

Step2

Behavioral Features & Input Signals

The system relies on various data points including transaction time, IP address reuse, burst activity, time taken for each action, and device fingerprinting. These features are engineered to capture subtle behavioral patterns unique to bots. By analyzing these inputs, the model accurately flags suspicious activity while minimizing false positives, maintaining platform integrity and fairness.

05 Module 2: Review Detection



Fake Review Classifier

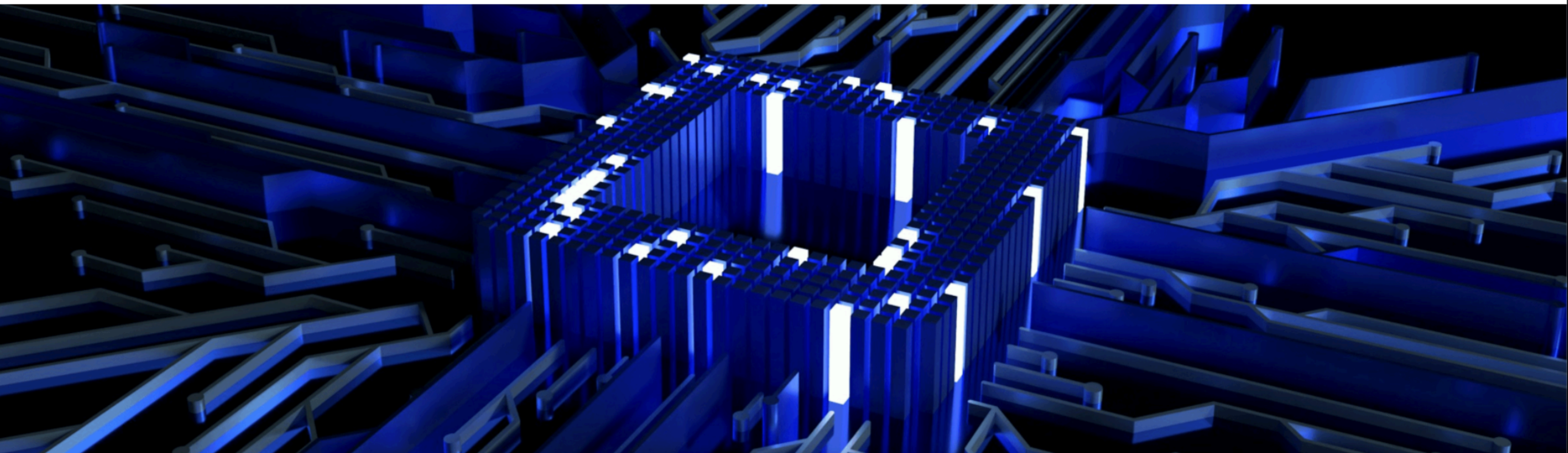
ML-Powered Review Analysis

The Fake Review Classifier uses supervised machine learning to detect manipulated or non-genuine product reviews. The system is trained to recognize patterns found in AI-generated text, scripted templates, and reused (hijacked) reviews. By leveraging models like Random Forest and ensemble methods, the system can identify subtle linguistic and behavioral cues that distinguish fake content from genuine customer feedback.

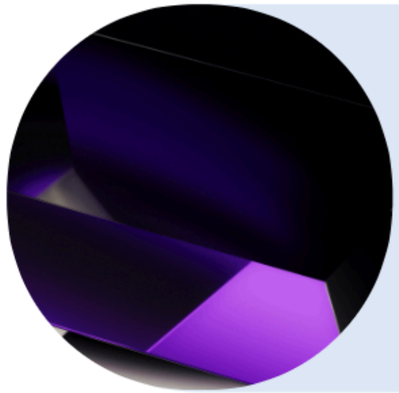
Textual & Behavioral Features

To support accurate detection, the system extracts both textual and contextual features from reviews. This includes TF-IDF vectors, review length, time of posting, account quality, IP address, and the time taken to write the review. These diverse data points enable the model to evaluate not just the content of the review, but also the behavior behind its submission—ensuring precise classification across AI, script, hijacked, and human-authored reviews.

06 Tools & Tech Stack

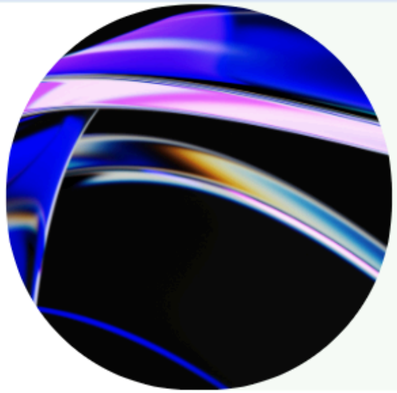


Technologies Used



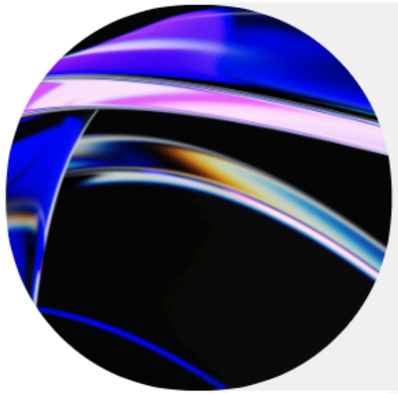
Data Processing & Management

The project relies heavily on **Pandas** and **NumPy** for data manipulation, feature extraction, and transformation. Preprocessing tools like **StandardScaler** and **LabelEncoder** ensure the input data is clean and model-ready. Additionally, **OpenPyXL** is used to manage Excel files that store simulated purchases and reviews, making the system both traceable and easy to update during testing.



Machine Learning Framework

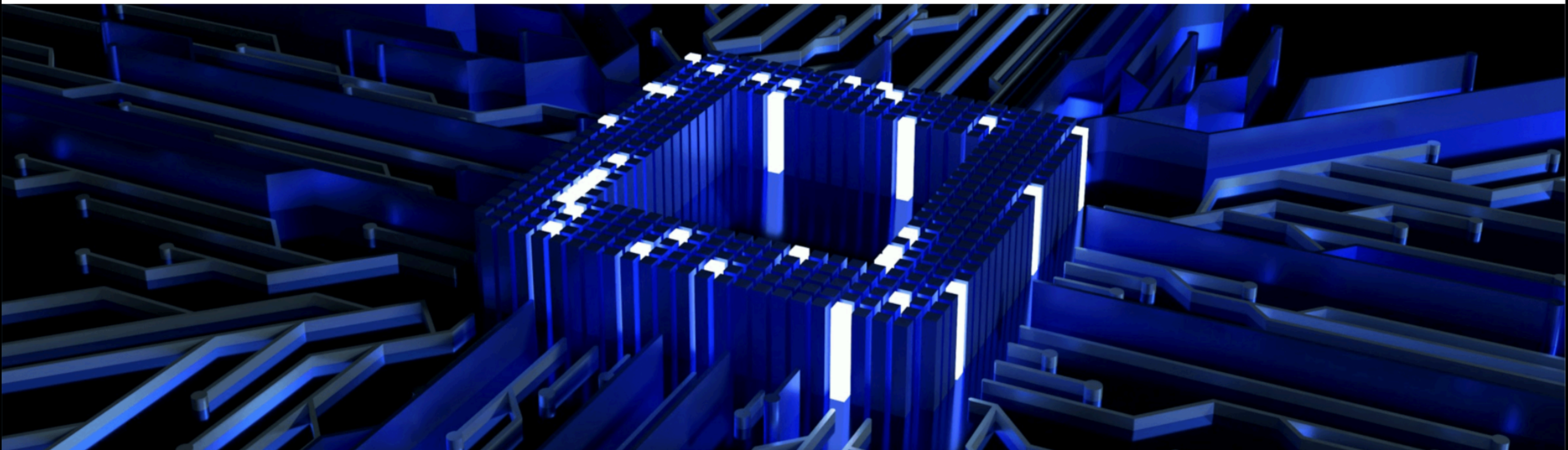
At the heart of the system are machine learning libraries such as **Scikit-learn**, **XGBoost**, and **LightGBM**, which power both the bot detection and fake review classification models. These tools enable robust pattern recognition, classification, and risk scoring using advanced techniques like ensemble learning and TF-IDF vectorization. They are essential for training and deploying models that can adapt to evolving fraud behaviors.



User Interface & Integration

A clean and functional interface is built using **Streamlit**, allowing users to generate reviews, simulate purchases, trigger model training, and view predictions in real time. Custom **CSS** enhances the UI's visual appeal, while the **Gemini API** adds AI-generated review capabilities. Together, these technologies ensure that the system is not just intelligent but also easy to interact with and demonstrate.

07 Impact



Impact

cost

Implementation Cost

The **Bot Buying Classifier** involves moderate implementation cost. It requires building behavioral simulations, extracting structured features (like IP reuse and transaction bursts), and training models on synthetic and real interaction data. On the other hand, the **Fake Review Classifier** has a slightly higher cost due to the integration of natural language processing techniques, AI-generated content (via Gemini API), and more complex text-based feature engineering.

Cost

VS

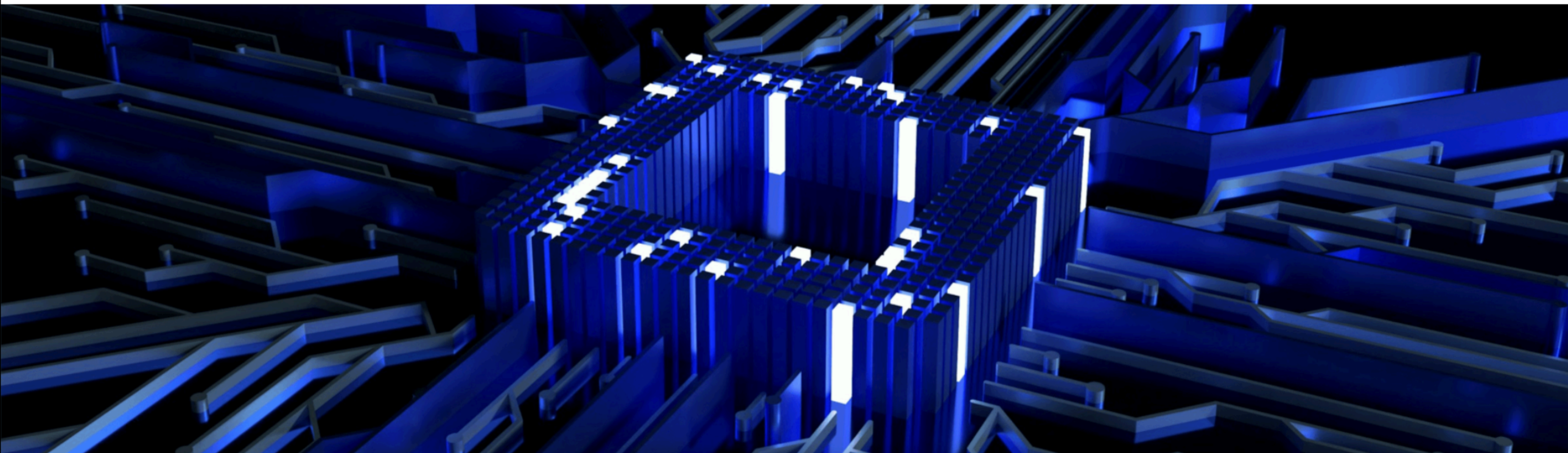
Benifits

Benifits

Impact and Benifits

The **Bot Buying Classifier** offers high operational value by preventing stock manipulation, ensuring fair access to products, and protecting the platform during high-traffic events like flash sales. Meanwhile, the **Fake Review Classifier** provides an even greater long-term benefit by safeguarding customer trust, improving product credibility, and supporting Amazon's reputation for authenticity—making it a critical tool for marketplace integrity.

08 Results & Metrics



Performance Analysis

Model Performance Results

High Accuracy Achieved

Both classifiers demonstrated strong performance, with accuracy exceeding **90%** in detecting bots and classifying fake reviews during testing with labeled data.

Low False Positives

Fine-tuning of feature thresholds and model parameters helped reduce false positives, ensuring genuine users and reviews were not mistakenly flagged.

Evaluation Metrics Used

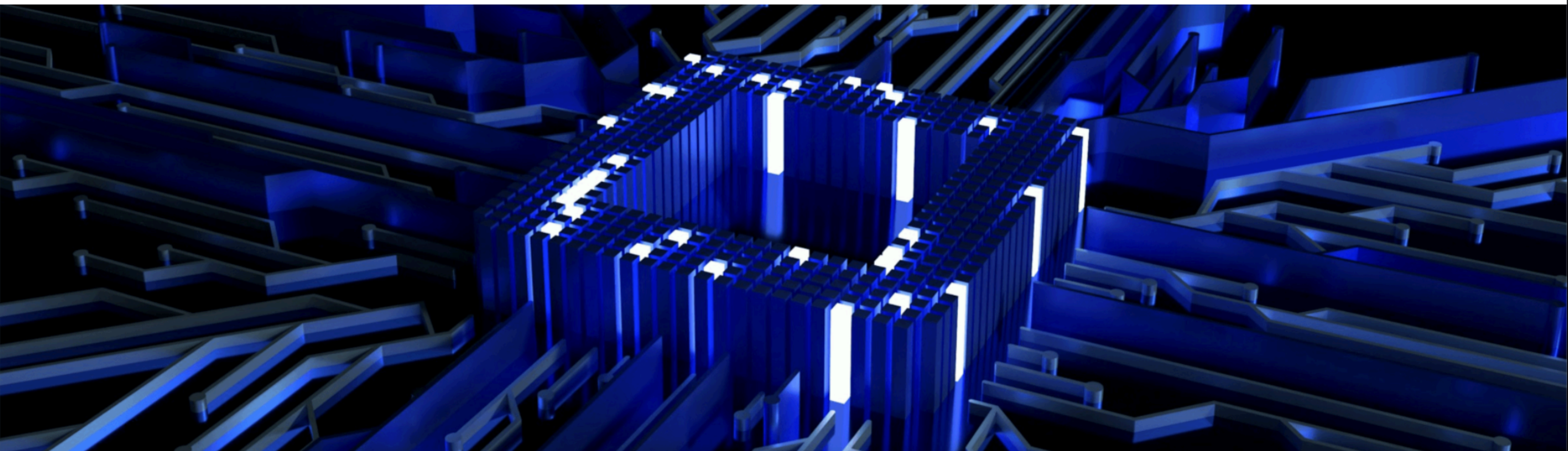
Precision, Recall, and F1-Score

These metrics were used to assess classification effectiveness—especially important for balancing fraud detection and minimizing impact on real users.

Confusion Matrix and ROC-AUC

Visual and numeric evaluation tools like the **confusion matrix** and **ROC-AUC score** helped validate model robustness and decision boundaries.

09 Future Work



Enhancements to Models

Incorporating New Data Sources

To improve detection accuracy and adaptability, the models will be enhanced by **incorporating new data sources** such as user session logs, device fingerprints, behavioral heatmaps, and cross-product activity. These enriched inputs will help capture deeper fraud signals. Additionally, we aim to **upgrade the algorithms** using advanced techniques like anomaly detection, ensemble stacking, and semi-supervised learning to better handle evolving bot and review fraud tactics. Together, these enhancements will make the system more robust, scalable, and harder to bypass.

Improving Algorithms

To boost the system's fraud detection capability, we plan to enhance the models by **incorporating new data sources** such as user session behavior, device metadata, purchase timing patterns, and cross-product activity logs. These additional signals will provide a more comprehensive view of user intent. Simultaneously, we will focus on **improving the algorithms** by integrating advanced techniques like ensemble stacking, anomaly detection, and deep learning models for text and behavior analysis. These upgrades will ensure higher accuracy, faster adaptation to new fraud patterns, and improved performance across both modules.

Expansion Plans

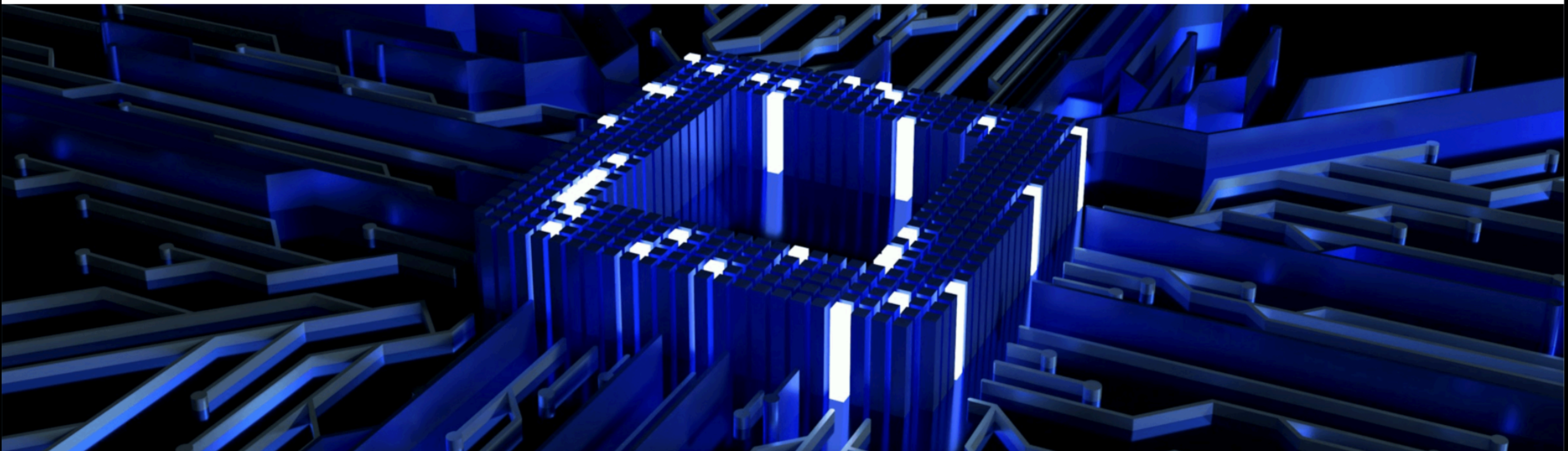
New Market Segments

To support Amazon's growth in new market segments, our system can be extended to detect region-specific fraud patterns in reviews and purchases. By training models on localized data, we can uncover emerging threats unique to new demographics or product categories. This adaptive approach ensures platform trust and customer safety as Amazon expands into untapped geographies or niche verticals, reinforcing its reputation in both established and emerging markets.

Additional Features

To strengthen fraud detection, an advanced **real-time risk scoring system** will be integrated into both modules. This feature will assign each transaction or review a dynamic score based on behavioral anomalies, such as burst activity, unusual timing, or AI-generated patterns. These scores help prioritize moderation efforts, automate flagging of suspicious actions, and improve overall system transparency — ensuring quicker response to fraud and greater user trust.

10 Thank You Slide



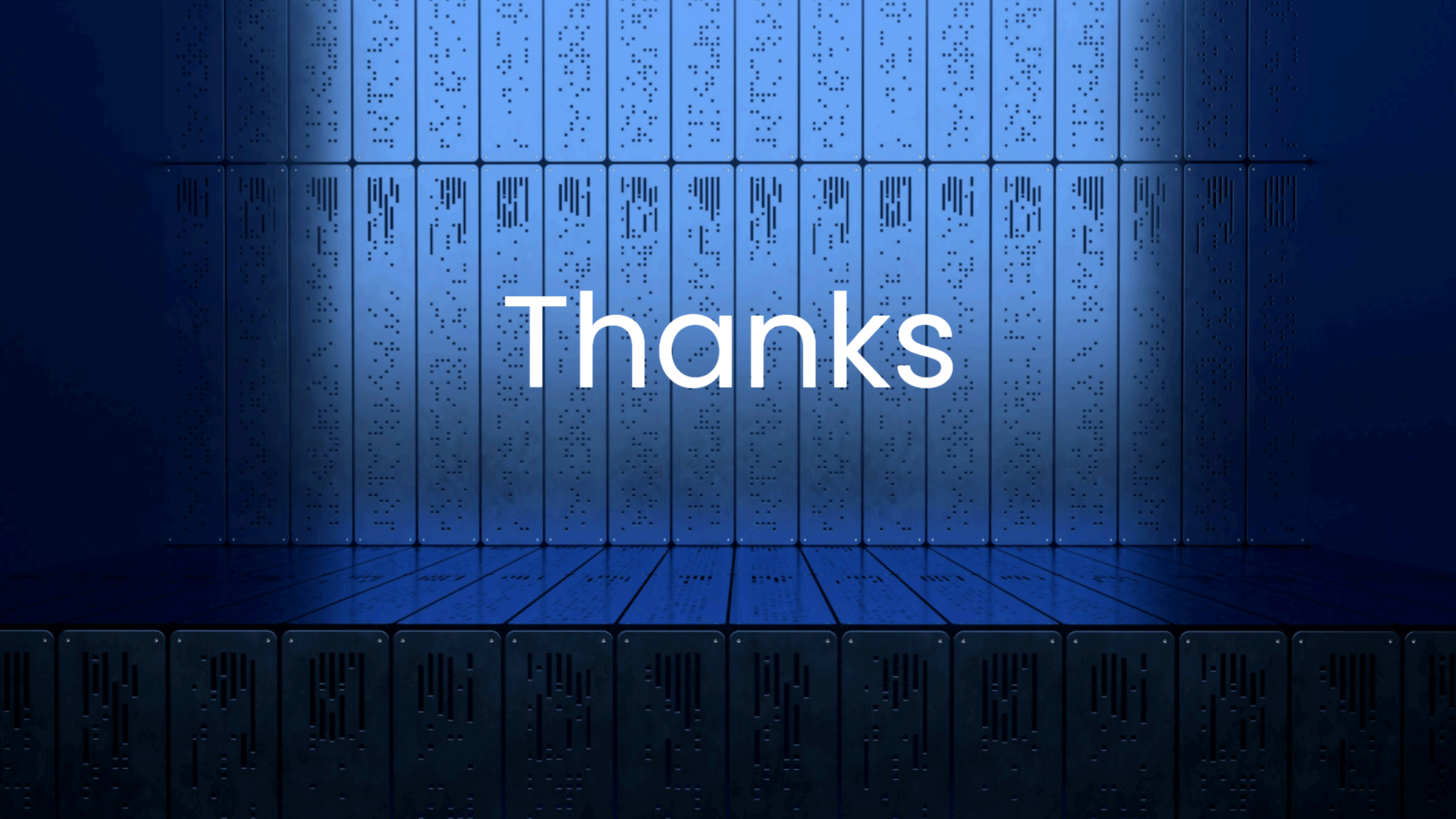
Acknowledgments

Team Contributions

We extend our gratitude to the dedicated team members who contributed to the success of the Amazon Trust project. Their collaborative efforts, innovative ideas, and unwavering commitment were instrumental in achieving our goals and enhancing machine learning applications for Amazon HackOn 2025. Thank you for your invaluable support.

Support from Amazon HackOn

We extend our heartfelt gratitude to Amazon HackOn 2025 for their invaluable support. Your resources, mentorship, and collaborative spirit significantly enhanced our project, fostering innovation and excellence in our machine learning initiatives. Thank you for believing in our vision.


Thanks