Digital

Co-funded by the European Union

TrustOS.

THE DIGITAL TRUST OPERATING SYSTEM FOR BANKS

Team 3:
Marta Corrado
Federico Mercurio
Xin Ye
Yao Lu
Ruifeng Nie
Wail Ameur

Contents

1 Introduction

1.1 Challenge Definition

As e-commerce and mobile banking grow, consumers expect instant, seamless transactions. However, this environment has become fertile ground for sophisticated fraud schemes.

- 24% increase in real-time transaction fraud since 2022 (Visa, 2024)
- 64% of users abandon transactions when security checks feel intrusive or slow (Forrester, 2023)

1.2 Executive Summary

One of the most pressing and complex challenges in modern banking is the rise of fraud schemes that occur outside the traditional boundaries of digital platforms. Social engineering attacks such as phone scams, phishing emails, and fraudulent SMS messages are increasingly responsible for significant customer losses, particularly among vulnerable populations like the elderly. These scams often manipulate users into authorizing transactions themselves, making them difficult to detect and nearly impossible to reverse under conventional fraud prevention models.

While banks have invested heavily in detecting anomalies within their apps and transactional systems, they often lack visibility into the human-centered activity that precedes a scam. As a result, many fraudulent transactions appear legitimate from a technical standpoint and evade automated safeguards. This blind spot leaves financial institutions exposed not only to direct monetary losses but also to reputational damage and regulatory scrutiny.

TrustOS introduces a new layer of fraud intelligence specifically designed to detect and respond to threats that originate outside the bank's immediate ecosystem. By monitoring cross-channel signals such as SMS, email metadata, call patterns, and behavioral indicators, we provide banks with real-time contextual alerts that signal when a customer may be under the influence of a scam.

Importantly, our system does not interfere with transactions or interrupt user flows. Instead, it integrates into the bank's existing fraud detection and response infrastructure, enhancing their decision-making capabilities without compromising customer experience. This allows institutions to act on richer multidimensional insights while maintaining full control over how and when to intervene.

By bridging the gap between human communication and transactional behavior, our platform equips financial institutions to proactively identify and mitigate modern fraud. This enables them to protect their customers, strengthen trust, and adapt to a rapidly evolving threat landscape.

2 Problem

2.1 Problem Definition

Banks are increasingly vulnerable to social engineering fraud, a form of attack where victims are psychologically manipulated into approving fraudulent transactions. Unlike technical frauds, these scams often originate outside the bank's infrastructure via phone calls, SMS, or phishing emails, making them significantly harder to detect and prevent through traditional monitoring systems.

Existing fraud detection mechanisms are typically centered on transactional anomalies or in-app user behavior. This creates a critical blind spot: attacks that exploit human psychology and occur across external communication channels often go unnoticed until the fraudulent transaction has already been authorized.

The scale of the problem is substantial. In the United Kingdom alone, over **£213 million** was lost to Authorised Push Payment (APP) fraud in the first half of 2024, a form of social engineering where victims are tricked into willingly transferring funds. This was part of a total £570 million in payment fraud losses during that period. In the United States, **94% of bank** executives reported an increase in business clients hit by fraud in 2023. Large banks (over $100 billion in assets) experienced check fraud losses exceeding **$15 million**. Mid-sized banks (assets between $31 billion and $100 billion) lost between $1 million and **$3 million** each. Smaller banks (under $10 billion) suffered losses in the **hundreds of thousands.** These figures highlight how even well-secured banks remain exposed to frauds that exploit the human element rather than technical vulnerabilities.

As a result, banks face not only direct financial losses, but also reputational damage and growing regulatory pressure, especially as customers increasingly expect their financial institutions to proactively prevent these evolving threats. [1] [2]

2.2 Stakeholders

The most relevant stakeholders for our solution are banks. Unlike fintechs or digital-first platforms such as Stripe or PayPal, which already employ advanced fraud mitigation technologies, many retail banks are operating with legacy systems and limited cross-channel visibility. These institutions are on the front line of authorizing transactions, and they typically absorb the highest risk when fraud occurs.

Why Banks Are the Ideal Target:

- **Systemic Role:** Banks are central to the financial ecosystem and are often the underlying enablers for digital platforms. Even modern solutions like PayPal require links to traditional bank accounts.
- **Broad User Base:** A significant portion of the population still relies on regular banks, many of whom are less digitally literate and therefore more vulnerable to social engineering scams.

- **Security Gap:** Banks often lack tools to monitor or intercept fraud that begins outside their applications (e.g., over the phone or via SMS).
- **Untapped Opportunity:** This presents a unique value proposition in an underserved market segment where improving upstream fraud prevention can deliver widespread impact.

We refer to these stakeholders as *Digital Intermediaries*: institutions that bridge traditional finance and digital services but face increasing pressure to modernize their fraud prevention capabilities.

2.3 Aspirations

Banks today are under increasing pressure to strike a balance between **security**, **user experience**, and **regulatory compliance**, all while operating in a fast-evolving fraud landscape. At the core of their aspirations is a need to protect their customers comprehensively, including from threats that originate outside the bank's digital systems.
Their key objectives include:

- Preventing financial losses caused by fraudulent but customer-authorized transactions.
- Safeguarding brand reputation and public trust by demonstrating responsibility and responsiveness.
- Ensuring customer retention by offering a sense of security, especially to vulnerable or less tech-savvy users.
- Meeting regulatory expectations on fraud detection and consumer protection.
- Staying competitive against digital-first banks and fintechs that increasingly highlight advanced fraud prevention as a core value proposition.

More than ever, banks are expected to go beyond traditional transaction monitoring and detect fraud before it happens, particularly scams that manipulate human behavior, such as impersonation calls, fake emergency messages, or phishing attempts via SMS or email. Customers now hold banks accountable for fraud that occurs even outside of official banking channels, which adds to the pressure. [3]
However, despite these strong aspirations, banks face several persistent obstacles:

- **Technological Gaps:** Most fraud prevention systems are designed to detect suspicious activity based on historical patterns or internal app behavior. They are not equipped to interpret contextual signals or sudden behavioral changes driven by external manipulation.

- **Fragmented Systems and Silos:** Fraud teams, IT departments, and customer support often operate in disconnected environments, limiting the bank's ability to act swiftly and holistically when a threat is detected.

- **Legacy Infrastructure:** Many institutions still run on outdated core banking systems that are difficult to adapt or integrate with modern fraud detection tools, particularly those that require real-time data flow or behavioral intelligence.

- **Lack of Real-Time Customer Context:** Banks generally cannot tell when a transaction is being carried out under psychological duress, such as during a scam call. This makes it extremely difficult to intervene in time, even when subtle red flags are present.

- **Human Factor:** The root of many modern frauds lies in the customer's emotional state, fear, urgency, confusion, which current systems are not designed to detect or evaluate.

To summarize, while banks clearly aspire to be proactive defenders of their customers, they are currently constrained by tools and structures that are reactive, isolated, and limited to internal channels. Addressing these challenges requires a shift toward real-time, cross-channel, behavioral fraud intelligence, *exactly the gap our solution is designed to fill.*

3 Solution

3.1 Inspiration

To shape our approach, we examined success stories from adjacent industries that had developed tools particularly relevant to our objectives. These cases featured technologies and strategies that we identified as highly adaptable to the challenges of detecting and preventing social engineering fraud in banking.

3.1.1 Google Gmail Filters

Google Gmail's sophisticated spam and phishing filters serve as a primary inspiration for the core mechanism of our early warning system. Gmail processes billions of emails daily, relying on dynamic blacklists of known malicious URLs, scammer email addresses, and suspicious domains, combined with advanced content analysis engines that detect deceptive language patterns. This capability allows them to proactively identify and flag or quarantine malicious communications at scale, often before they even reach the user's primary inbox. The direct analogy to our solution lies in this first line of defense approach: just as Gmail flags suspicious emails, our system aims to build and maintain a comprehensive, real-time "blacklist" of scam indicators (such as reported fraudulent phone numbers, phishing website URLs, or known scam message patterns). This robust, constantly updated intelligence forms the foundation of our ability to flag potential social engineering attempts the moment they appear, much like a global inbox defender.

3.1.2 Recorded Future

From Recorded Future, we draw inspiration for the crucial process of managing actionable threat intelligence. Recorded Future excels in its robust methodology for collecting, curating, updating, and distributing diverse threat data from a vast array of sources, including the open web, dark web, and technical feeds. Their success lies in transforming raw data into actionable insights that empower organizations to make informed security decisions in real-time. For our project, this translates directly to the rigorous management of our "list of scams and frauds." We are inspired to develop highly effective processes for sourcing new scam indicators, validating their authenticity, integrating them into our database dynamically, and ensuring their timely dissemination to our partner's fraud protection systems. This ensures that the intelligence we provide is not merely data, but a potent, up-to-the-minute defense against rapidly evolving social engineering tactics.

3.1.3 Cybersource

Cybersource's approach to modularity and seamless integration provides a vital blueprint for how our solution will connect with existing banking infrastructure. Their success in enabling banks and merchants to integrate various payment and fraud management tools demonstrates the power of an API-first design that feels like an intrinsic part of the bank's ecosystem, rather than a complex third-party add-on. For our B2B2C model, this is paramount: our early warning system must integrate smoothly into the partner's fraud solution, which then integrates into the bank's merchant or core banking portals. The goal is to ensure that banks perceive our solution as an invaluable, native enhancement to their fraud capabilities, delivering immediate insights and flags without requiring complex reorganization or appearing as a disconnected tool. This focus on frictionless technical and experiential integration is key to widespread adoption and effectiveness.

3.2 Product Description

TrustOS is a fraud intelligence and orchestration platform designed to close a critical gap in modern banking security: detecting scams that begin outside the digital or transactional scope of traditional fraud systems. These include social engineering attacks carried out through phone calls, SMS, or emails, channels that are typically invisible to banks' existing monitoring infrastructures. TrustOS enable financial institutions to proactively identify and respond to these external threats before they result in customer-authorized financial losses. The core purpose of TrustOS is to enhance banks' ability to protect their most vulnerable customers without adding friction to the user experience. We do this by introducing a cross-channel detection layer that monitors behavioral signals and communication metadata to identify high-risk patterns associated with scams. Our platform allows banks to take timely and intelligent action when customers may be unknowingly acting under manipulation. Importantly, TrustOS does not block or delay transactions. Instead, it acts as an intelligent advisory layer, providing contextual insights, such as correlations with known scam numbers, communication timing, and behavioral anomalies, that can inform internal review, fraud workflows, or customer support interventions.

3.3 Technical Approach

TrustOS ingests and analyzes metadata from external communication channels, such as SMS timestamps, email patterns, and call records, alongside behavioral and transactional data already available within the bank. By correlating this information in real time, the platform creates a centralized view of user activity that includes both digital interactions and real-world social cues. For users who have provided explicit consent, metadata is stored in anonymized and encrypted form for up to 72 hours. This temporary retention window allows TrustOS to examine recent communications for potential fraud signals, such as messages or calls linked to known scam networks, while maintaining user privacy. The data is used solely for fraud detection and to continuously fine-tune our detection models, improving accuracy and adaptability over time. To ensure stability and reliability for our clients, model updates are consolidated and released to banks on an annual cycle. This approach allows TrustOS to evolve in response to new fraud tactics while avoiding disruption from frequent model changes. All user data used for training is fully anonymized, encrypted, and permanently deleted after processing. These insights are delivered directly into the bank's existing fraud infrastructure to add an additional intelligence layer that enhances detection without disrupting current workflows. This significantly broadens the institution's ability to detect socially engineered scams and respond with greater speed, accuracy, and confidence.

3.3.1 Use-Case Scenario

An elderly customer receives a text message that appears to be from their grandson, claiming to be in an emergency and urgently needing money.
Trusting the message, the customer visits their local branch to withdraw $5,000 in cash.
Before the transaction is finalized, TrustOS intervenes. It detects unusual behavior based on the customer's typical transaction history and the context of the request.
The platform flags the transaction as potentially fraudulent and notifies the bank in real time.
A staff member reaches out to the customer to verify the withdrawal, and during the conversation, it becomes clear that the message was part of a social engineering scam.
The transaction is stopped before any money is handed over, protecting the customer and helping the bank avoid a financial loss.

3.4 Company Mission

TrustOS was founded with the mission to make fraud defense smarter, more contextualized, and more effective. We aim to empower financial institutions to protect individuals, especially those most at risk, from manipulation, coercion, and financial exploitation. By turning fragmented signals across human, digital, and transactional channels into cohesive, actionable insights, we help banks act earlier, respond more responsibly, and restore customer trust. At a societal level, our solution contributes to reducing the emotional and economic toll of scams, supporting broader efforts to build public confidence in digital finance. As fraud becomes increasingly psychological rather than technical, tools like TrustOS are essential for financial institutions that seek to remain trusted, proactive, and resilient in a rapidly changing threat environment.

3.5 Value Proposition

3.5.1 Market Landscape

The growing complexity of financial fraud, especially social engineering and impersonation scams, has exposed critical blind spots in traditional fraud detection systems. In particular, authorized push payment (APP) fraud and manipulation through external channels such as phone calls and SMS remain under-monitored and under-addressed by most institutions. Recent research highlights both the scale of the problem and the opportunity for advanced solutions. Visa reports that banks using AI-driven fraud detection systems have reduced fraud-related costs by more than 40 percent and mitigated millions in potential losses linked to scams involving social engineering [4]. Similarly, Forrester estimates that each dollar of fraud prevented through analytics tools can yield savings of up to 2.70 dollars in downstream operational costs related to investigation, remediation, and customer service [5]. The need for early intervention is especially urgent when considering the human impact. According to the United States Federal Trade Commission, the average loss per scam victim in 2024 exceeded 9,000 dollars, with a significant proportion of cases involving vulnerable populations, particularly the elderly, targeted through calls or SMS [6]. These scams are often invisible to standard fraud detection platforms, as they exploit the user before any suspicious transaction is logged.

3.5.2 Competitive Advantage

TrustOS offers a unique position in the fraud prevention ecosystem by solving a problem that is largely unaddressed by existing tools: identifying socially engineered fraud before a transaction becomes anomalous or unauthorized in traditional systems. While most fraud detection software focuses on digital behavior within apps or transaction patterns, TrustOS extends its intelligence layer into the real-world communication channels that attackers use to manipulate users. Key competitive advantages include:

- **Cross-channel orchestration:** Unlike point solutions that monitor only one channel (e.g., email phishing filters or transaction scoring engines), TrustOS unifies communication data (SMS, email, calls) and behavioral signals with transactional context. This allows banks to identify scam attempts that span digital and human vectors (such as in-person withdrawals triggered by phone impersonation).

- **Real-time contextual intelligence:** Inspired by models like Gmail's spam filters and enhanced with threat feeds from platforms like Recorded Future, TrustOS operates in real time, detecting indicators of fraud before the transaction reaches the execution point. The use of temporary, anonymized metadata enables rapid detection while preserving privacy and regulatory compliance.

- **Non-intrusive integration:** TrustOS acts as an advisory layer, not a blocking mechanism. This allows it to be adopted without risk of interfering with legitimate transactions or creating friction for end-users. Insights can be fed into fraud queues, case management systems, or customer service tools, giving institutions full control over how to act.

- **Scalable and predictable model updates:** Our detection models are continuously refined using real-world signals, but updates are deployed on an annual release cycle. This ensures consistent performance without introducing operational instability, a common concern with real-time learning systems.

- **High adaptability to compliance and regulatory needs:** Through this added visibility, TrustOS equips fraud teams to make more informed assessments while helping institutions address a growing fraud vector that traditional systems were not designed to handle.

3.6 Elements of Value

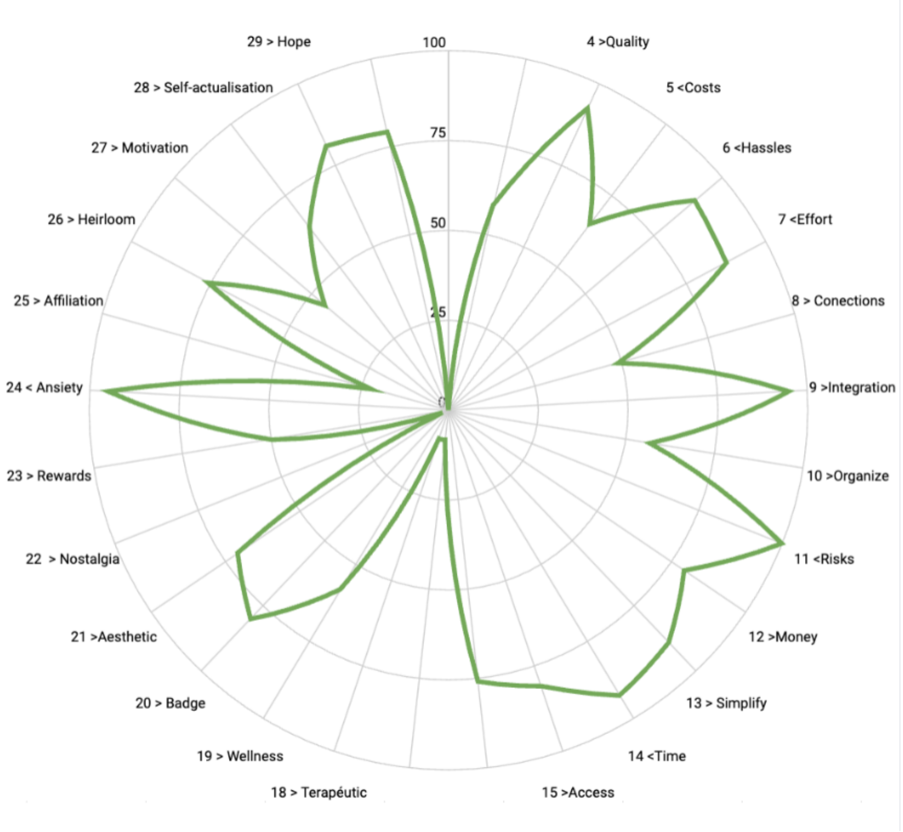


Figure 1: EoV Graph

The TrustOS platform is designed not only as a technical solution but as a direct response to the most critical values identified by our target stakeholders: financial institutions and their customers. At its core, TrustOS delivers on risk reduction, a top strategic priority for banks. By targeting fraud schemes that originate outside the digital transaction layer such as social engineering, impersonation scams, and in-person coercion, it helps prevent both financial losses and reputational damage. This aligns with what the Elements of Value framework identifies as Reduces Risk (Element 11, fig 1), the highest-ranking value for enterprise customers in regulated environments. Equally important is TrustOS's ability to solve structural inefficiencies by enhancing integration across communication and transactional channels. By bringing together SMS, phone, email, and banking activity into a unified fraud intelligence layer, the platform addresses the long-standing problem of disconnected fraud signals. This directly supports Integration (Element 9, fig 1), a value highly appreciated by institutions seeking operational efficiency without overhauling their core systems.

On the customer side, TrustOS helps banks meet deeper emotional needs often overlooked by technical fraud solutions. Vulnerable users, particularly the elderly, often experience anxiety and shame when exposed to scams. The platform enables banks to detect manipulation before losses occur, which in turn Reduces Anxiety (Element 24, fig 1) and restores users' sense of control and security. By demonstrating active protection and empathy, banks can also instill Hope (Element 29, fig 1), reinforcing customer loyalty and brand trust in moments of vulnerability.
In this way, TrustOS does more than improve fraud prevention. It adds measurable value at both the institutional and human levels, making fraud defense more connected, proactive, and customer-centric.

4 Business Plan

4.1 Revenue Model and Pricing Strategy

Our company operates under a B2B SaaS model, offering subscription-based fraud detection and response services to banks and financial institutions. Clients are billed monthly under annual contracts, with pricing determined by the volume of transaction insights processed. We offer a tiered usage-based model:

- €0.0300 per transaction for the first 100,000 requests each month
- €0.0075 per transaction for any additional requests

This approach ensures flexibility across different tiers of financial institutions, making our solution accessible to smaller regional banks while remaining cost-effective at scale for larger enterprises. The model reflects pricing strategies used by leading enterprise SaaS companies like Salesforce and Palantir, where clients pay according to their usage volume and feature adoption [7].

We also plan to establish channel partnerships with system integrators and security consultancies. These partners will receive 5-10% commissions on any closed B2B deals, helping us scale our sales reach quickly. For investors, returns will be structured through equity dilution or dividend-style payouts based on profitability or exit, aligned with standard early-stage venture models where investors receive 1-3% carry or equity return depending on the round [8].

4.2 Value Proposition and Anchored Pricing

Our pricing is anchored on the measurable fraud prevention value we provide. AI-based tools like ours are proven to reduce fraud investigation and operational costs by up to 40% [5], and help banks reduce phishing-related losses by 18% or more [9]. Visa's own AI-based fraud detection system has demonstrated massive national-level impact, projecting £100 million in savings on APP scams in the UK alone [4]. Further emphasizing the cost of inaction, the FTC reports average losses of $9,000 per victim in scam-related incidents [6], while studies show that a mid-sized bank can face between €1 and €3 million in annual fraud losses [2]. Our service costs banks roughly 6% of those losses, positioning us as a clear ROI-positive investment.

4.3 Target Market and Addressable Opportunity

We address a growing need among financial institutions to protect their clients from social engineering and advanced fraud that originates outside traditional banking apps and interfaces. Our total addressable market includes:

- Retail banks
- Digital-first banks and fintechs

- Credit unions and cooperative banks
- Embedded finance providers offering payment or transfer services

The immediate go-to-market focus is on small to mid-sized banks that lack the infrastructure or internal expertise to monitor external fraud signals across communication channels. These institutions represent a significant underserved segment in fraud intelligence.

4.4 First-Year Sales Projection

In our first full year of sales, we expect to close contracts with 2-3 financial institutions. This estimate reflects the longer sales cycles associated with regulated enterprise sectors and the need for deep technical evaluation. These early adopters will serve as strategic references, providing us with validation and real-world product feedback. We estimate an average annual contract value (ACV) of €100,000 per client. This is in line with other enterprise-grade SaaS offerings in high-stakes verticals like fraud prevention, where average ACVs range from €50,000 to €200,000 [10]. A more accurate estimation based on a small sized bank is present in the Cost Structure section of the report, where a detailed analysis of a realistic and conservative scenario is presented, and still remains in line with this estimation.
Based on a mid-range estimate:

$$3 \text{ clients} \times €100,000 = €300,000 \text{ in Year 1 gross revenue}$$

This projection is realistic for a direct-sales SaaS strategy in FinTech, especially given the urgency banks face around external scam prevention and reputational risk management.

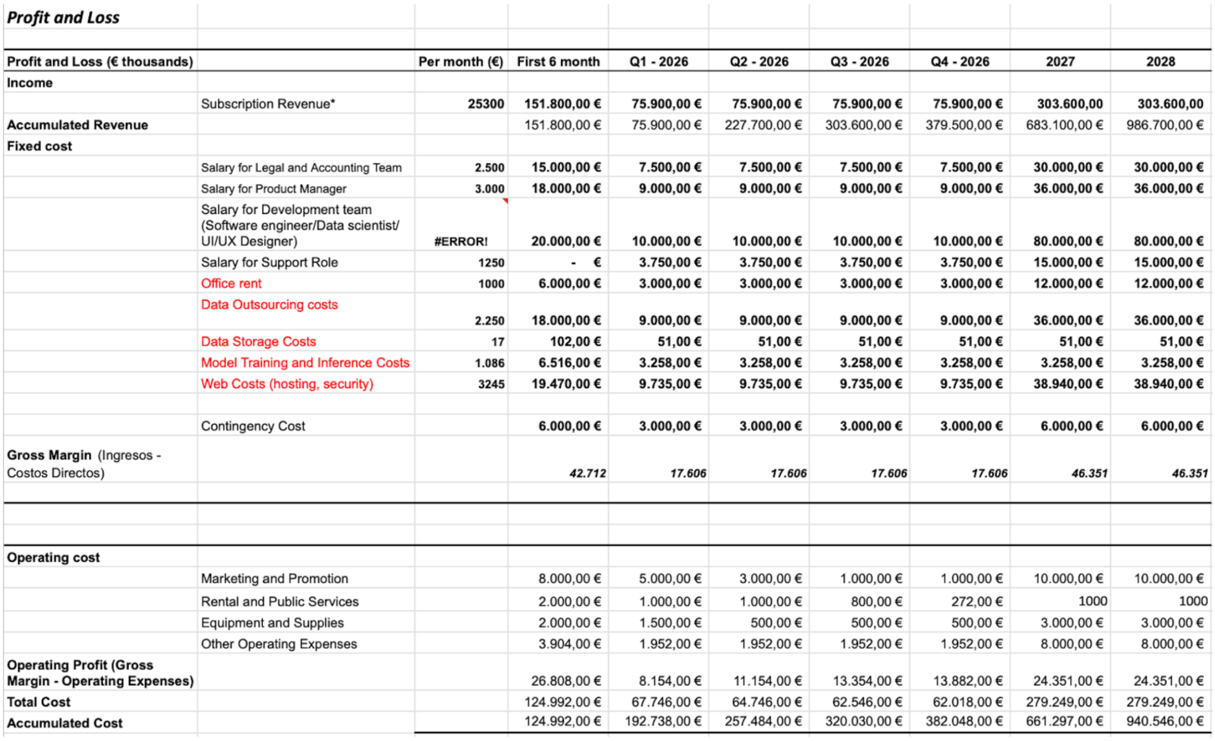
4.5 Initial Sales Plan and Go-To-Market Execution

In the first month, we will focus on building a strong initial customer base and go-to-market infrastructure through two parallel strategies:

1. **Pilot Engagements:** We plan to onboard 2-3 small or mid-sized banks for pilot programs. These pilots will allow us to test real-world deployment, refine detection models using live data, and deliver measurable early impact. Successful case studies will serve as proof points in our broader commercial rollout.
2. **Strategic Sales Partnerships:** Concurrently, we will pursue collaborations with system integrators and cybersecurity consultants. These players already have access to financial institutions and can extend our reach into qualified sales opportunities, accelerating credibility and distribution in our target market.

This dual approach ensures we can both validate our solution and begin building a scalable revenue pipeline from day one.

4.6 Cost Structure and Revenue Estimation



Profit and Loss

Profit and Loss (€ thousands)		Per month (€)	First 6 month	Q1 - 2026	Q2 - 2026	Q3 - 2026	Q4 - 2026	2027	2028
Income									
	Subscription Revenue*	25300	151.800,00 €	75.900,00 €	75.900,00 €	75.900,00 €	75.900,00 €	303.600,00	303.600,00
Accumulated Revenue			151.800,00 €	75.900,00 €	227.700,00 €	303.600,00 €	379.500,00 €	683.100,00 €	986.700,00 €
Fixed cost									
	Salary for Legal and Accounting Team	2.500	15.000,00 €	7.500,00 €	7.500,00 €	7.500,00 €	7.500,00 €	30.000,00 €	30.000,00 €
	Salary for Product Manager	3.000	18.000,00 €	9.000,00 €	9.000,00 €	9.000,00 €	9.000,00 €	36.000,00 €	36.000,00 €
	Salary for Development team (Software engineer/Data scientist/ UI/UX Designer)	#ERROR!	20.000,00 €	10.000,00 €	10.000,00 €	10.000,00 €	10.000,00 €	80.000,00 €	80.000,00 €
	Salary for Support Role	1250	- €	3.750,00 €	3.750,00 €	3.750,00 €	3.750,00 €	15.000,00 €	15.000,00 €
	Office rent	1000	6.000,00 €	3.000,00 €	3.000,00 €	3.000,00 €	3.000,00 €	12.000,00 €	12.000,00 €
	Data Outsourcing costs	2.250	18.000,00 €	9.000,00 €	9.000,00 €	9.000,00 €	9.000,00 €	36.000,00 €	36.000,00 €
	Data Storage Costs	17	102,00 €	51,00 €	51,00 €	51,00 €	51,00 €	51,00 €	51,00 €
	Model Training and Inference Costs	1.086	6.516,00 €	3.258,00 €	3.258,00 €	3.258,00 €	3.258,00 €	3.258,00 €	3.258,00 €
	Web Costs (hosting, security)	3245	19.470,00 €	9.735,00 €	9.735,00 €	9.735,00 €	9.735,00 €	38.940,00 €	38.940,00 €
	Contingency Cost		6.000,00 €	3.000,00 €	3.000,00 €	3.000,00 €	3.000,00 €	6.000,00 €	6.000,00 €
Gross Margin (Ingresos - Costos Directos)			*42.712*	*17.606*	*17.606*	*17.606*	*17.606*	*46.351*	*46.351*
Operating cost									
	Marketing and Promotion		8.000,00 €	5.000,00 €	3.000,00 €	1.000,00 €	1.000,00 €	10.000,00 €	10.000,00 €
	Rental and Public Services		2.000,00 €	1.000,00 €	1.000,00 €	800,00 €	272,00 €	1000	1000
	Equipment and Supplies		2.000,00 €	1.500,00 €	500,00 €	500,00 €	500,00 €	3.000,00 €	3.000,00 €
	Other Operating Expenses		3.904,00 €	1.952,00 €	1.952,00 €	1.952,00 €	1.952,00 €	8.000,00 €	8.000,00 €
Operating Profit (Gross Margin - Operating Expenses)			26.808,00 €	8.154,00 €	11.154,00 €	13.354,00 €	13.882,00 €	24.351,00 €	24.351,00 €
Total Cost			124.992,00 €	67.746,00 €	64.746,00 €	62.546,00 €	62.018,00 €	279.249,00 €	279.249,00 €
Accumulated Cost			124.992,00 €	192.738,00 €	257.484,00 €	320.030,00 €	382.048,00 €	661.297,00 €	940.546,00 €

Figure 2: Profit and Loss

4.6.1 Revenue and Transaction Modeling

We are building a fraud detection and orchestration layer designed for small- to mid-size banks. Using BancaStato (a cantonal bank in Switzerland) as a representative customer, we estimate the following:

- BancaStato holds 0.65% of Swiss banking assets and serves an estimated 58,000 clients [11].
- Using benchmarks from PostFinance (2.4M accounts, 1.4B transactions annually), the average transaction volume per account is 583/year [12].
- Thus, for BancaStato: $58,000 \times 583 \div 12 = \sim 2.8$ million transactions/month.
- Expanding to three such banks = 8.5 million transactions/month.

At a conservative price of €0.003 per transaction insight (based on AWS service references and fixed price point), we estimate:

$$\text{Monthly Revenue: } 8.5M \times 0.003 = €25,500$$
$$\text{Annual Gross Revenue: } €25,500 \times 12 = €306,000$$

4.6.2 Infrastructure and Model Costs

Data Storage We store SMS and email metadata for fraud insight (no content, just metadata), kept for 72 hours:

- SMS metadata: 20 messages/day × 256 bytes = 5KB/day
- Email metadata: 50 emails/day × 5KB = 250KB/day

For 58,000 users × 3 banks × 3 days = 130.5GB, conservatively rounded to 200GB/month. Using Azure Blob Storage (Hot Tier) at €0.0158/GB: 200GB × €0.0158 = €3.16/month or €38/year. Additionally, we estimate 1TB total storage for model weights, logs, user features, etc., adding another €16-17/month.

Model Training and Inference We train a deep neural network (100M weights) using Recorded Future's fraud intelligence module (€27,000/year) [13].

- Training: 1 A100 GPU @ €1.29/hour × 168h = €546/month
- Inference: 1 A10 GPU @ €0.75/hour × 720h = €540/month
- Total GPU compute: €1,086/month or €13,032/year

Web and Cloud Services Estimation for the cloud service are done by using Alibaba Cloud calculator [14] and Cloudflare [15] prices for security services.

- Domain: €30/year
- Database & Backend Infrastructure: €11,150/year
- CDN, DNS, and Security (Cloudflare): €2,065/year
- **Total web services: €13,245/year**

4.6.3 Salaries and Staff Costs

All salaries are based on Spain or similar European countries' market rates [16]. We based our estimates on typical startup salary ranges to ensure our projections are realistic and aligned with companies at a similar stage, rather than referencing salaries from larger, more established organizations.

Table 1: Salary Estimates.

Role	Salary (EUR)	Description
Legal & Accounting	€30,000	Outsourced, includes tax and compliance
Product Manager	€36,000	Startup-aligned PM compensation
General Support Role	€15,000	Light B2B support load
R&D - Software Engineer	€40,000	Technical co-founder or early hire
R&D - ML Engineer	€40,000	Model development and ops
Total Salaries	**€161,000/year**	

We also include the cost of the office rent to accommodate the personal not doing remote working. Estimated at 1000€ a month rent, in line with most European countries rent prices for a small sized office suitable for our start-up team. Office rent: 1000 * 12 = €12,000/year.

4.6.4 Operating Costs

We allocate 8% of our revenues on the operating costs as it's the median percentage for most starting SaaS, this is equivalent to a budget of €24,480. It's an estimation for the first year. [17]

1. **Marketing and Promotion - €10,000/year:** Includes: Website, branding, content marketing, LinkedIn/Google Ads, event attendance. Benchmark: Early SaaS startups spend 10-20% of total revenues on budget marketing early on, especially B2B.

2. **Rental and Public Services - €3,672/year:** Includes: Office rental, coworking space, electricity/internet, Zoom. Benchmark: Small B2B startups often use coworking/hybrid remote. Space/infra = 10-15% of OPEX.

3. **Equipment and Supplies - €3,000/year:** Includes: Laptops, monitors, software licenses (Figma, Github), office equipment. Benchmark: Founding teams usually bring their own laptops. This category is modest (12% early on).

4. **Other Operating Expenses - €7,808/year:** Includes: AWS, legal fees, business registration, insurance, compliance, accounting, domain hosting etc... Benchmark: Early-stage fintech startups often spend 40-60% of OPEX on infra, legal, compliance.

4.6.5 Cost Structure Analysis

Based on the provided profit and loss breakdown and the cost structure discussed, the financial model reflects a realistic and disciplined early-stage operational plan for a B2B

SaaS platform focused on fraud detection and response for financial institutions. The company projects subscription-based revenues starting at €25,300 per month, amounting to €151,800 in the first six months.

Revenue continues to grow steadily over the forecast period, reaching €683,100 in 2027 and €986,700 by the end of 2028. These figures are based on a conservative assumption that we will continue servicing the equivalent of three small banks throughout this period.

This forecast includes the possibility of losing one client and replacing it with another of similar size, but does not assume traction with larger banks during these early adoption phases. This conservative stance reflects the expected caution from major institutions in adopting new security technologies until robust field results and case studies are available.

Fixed costs are composed of core salaries for legal, product, development, and support roles, as well as critical infrastructure components such as cloud hosting, data outsourcing, model training, and fraud intelligence feeds.

All pricing estimates are grounded in current market rates from trusted service providers such as Azure, Cloudflare, and Recorded Future. Office rent, contingency buffers, and essential web and security infrastructure have also been factored in to ensure business continuity and compliance.

Overall, the financial plan demonstrates strong fundamentals, responsible assumptions, and high scalability potential once initial market validation is achieved.

5 Feasibility Study

This section presents a comprehensive feasibility study conducted to assess the practicality and potential success of the proposed solution. Through detailed analysis, we identified and evaluated key challenges, limitations, and contextual factors that could impact implementation.

5.1 Legacy System Integration

a. **Potential issue:** A major technical risk is the complexity of integrating with banks' legacy systems, which may be outdated, fragmented, or undocumented. This creates a high dependency on custom development, slowing down onboarding and scalability.

b. **How to validate:** Start the pilot with small to mid-sized banks (typically more agile), and test a modular API-based connector designed for legacy compatibility. Work closely with their IT teams to understand real integration barriers early.

c. **Findings & solution:** We will learn that most banks prefer minimal disruption and which are the most common legacy systems for banks. We will adapt by designing a plug-and-play modular architecture with backward compatibility, reducing friction and development time.

5.2 Data Privacy Concern

a. **Potential issue:** Monitoring and analyzing SMS, email, and transaction data to detect fraud involves processing sensitive customer information. This creates a significant risk of violating GDPR, PSD2, or local data protection laws if data is mishandled. Our startup lacks an in-house compliance/legal expert, which increases the risk of non-compliance and potential fines or loss of customer trust.

b. **How to validate:** To ensure data privacy compliance, consult a data privacy lawyer or hire a Data Protection Officer (DPO) to review designs and data flows. Map all data for processing and conduct a Data Protection Impact Assessment (DPIA) to identify risks. Design the platform to anonymise or pseudonymise customer data when possible. Test the Minimum Viable Product (MVP) in a controlled sandbox with synthetic data to validate functionality without privacy risks.

c. **Findings & solution:** In the pilot, we aim to understand banks' internal data privacy practices, their requirements for cloud, on-premise, or hybrid deployments, acceptable levels of data anonymization, and how privacy concerns vary between small and mid-sized banks, so we can adapt our approach to their specific needs.

5.3 Real-Time Detection Accuracy

a. **Potential issue:** A key risk is ensuring the machine learning models behind our fraud detection stack are accurate and fast enough to detect fraud without creating false positives that could annoy users or block legitimate transactions.

b. **How to validate:** We first benchmark our models on historical and simulated fraud data to ensure high accuracy and low false positives. Then, we run them in shadow mode on live traffic to test real-time performance and latency without disrupting operations. Finally, we close the loop with continuous feedback from analysts and flagged cases to retrain and improve accuracy over time.

c. **Findings & solution:** Initial tests might show that combining behavioural analysis with rule-based flags drastically reduced false positives.

5.4 Sales Friction & Long Sales Cycles

a. **Potential issue:** Banks often have slow procurement processes and long sales cycles, which could delay growth. Additionally, lack of trust in early-stage startups can prevent first deals.

b. **How to validate:** Partner with trusted system integrators or cybersecurity firms who already work with banks. Run co-branded pilots to transfer part of the trust and reduce internal friction.

c. **Findings & solution:** We are initiating relationships with integrators and consulting companies and offering a free pilot with clear KPIs to build trust.

5.5 Scalability

a. **Potential issue:** Scalability is a risk because the system must handle large, unpredictable volumes of transactions and communication data in real time, especially as we expand to mid-sized and large banks with millions of customers. If the system cannot scale efficiently, it may result in delays, missed fraud signals, and reduced trust from clients.

b. **How to validate:** By simulating small and mid-sized banks transaction and communication volumes, monitoring performance metrics like latency and throughput to identify bottlenecks. By collaborating with bank IT teams to align tests with real peak loads and using synthetic or anonymized data, we can assess system behavior under stress and ensure it can scale to meet future demand.

c. **Findings & solution:** We expect to find that our initial cloud setup might struggle to handle more than the initial pilot users at the same time under realistic bank traffic. To solve this, we plan to adjust and increase our cloud resources: adding more storage, computing power, and bandwidth, so the system can handle larger volumes smoothly.

5.6 Problem: User Consent and Platform Access Management

a. **Potential issue:** As intermediaries between users and banks, we handle analysis of third-party platforms (e.g., SMS, emails). This creates risks around obtaining user consent, managing API access, and handling unstable or limited third-party integrations. Our current setup lacks infrastructure for scalable token/authentication management across these platforms.

b. **How to validate:** We will test integrations with key platforms (e.g., Gmail, SMS APIs) using sandbox accounts and simulate real consent flows. We will also consult with pilot banks' compliance teams to ensure our data handling meets legal and privacy standards.

c. **Findings & solution:** We expect to identify stable vs. volatile APIs (since the APIs we are analyzing might not be equally reliable) and adapt by building a modular, fault-tolerant connector system (we might develop different connectors for different APIs) with fallback modes (backup behaviors for when an API fails). Also we will simplify the consent experience for users while ensuring clear privacy guarantees and ongoing access stability.

5.7 Potential FAQ

5.7.1 Back to the future

a. **What will this concept look like in the future?** In the future, our platform will serve as a seamless intelligence layer between banks and users, capable of ingesting data from diverse, user-authorized platforms (e.g., SMS, emails), processing it in real time, and returning highly accurate fraud detection, all with minimal integration burden on the bank side. The infrastructure will be fully modular, API-first, and privacy-by-design. Consent management, scalable data ingestion, and model accuracy will be automated and adaptive to regulatory and technical shifts.

b. **How to take advantage of it? How to keep an eye on it?** To take advantage of this evolution we can continuously track regulatory updates (e.g., eIDAS2, PSD3) and platform access policies (e.g., changes to Gmail APIs or iOS permissions); invest in infrastructure observability and model monitoring to detect issues early and maintain trust. We can keep an eye on it by participating in fraud detection consortiums, collaborating with data privacy associations, and staying close to banks' IT feedback loops, we will anticipate future needs and pivot faster.

5.7.2 Reduce project risks by simplifying the service

a. **How could we simplify the service to make it easier to build?** We can make the service easier to build by starting small and focusing on just a few high-priority types of fraud. At first, we could implement only one channel, for example, monitoring SMS, and simply flag suspicious activity for the bank to handle. The analysis of additional channels like email can be added later.

b. **How would this change affect the value proposition?** Simplifying the service reduces initial complexity and accelerates time-to-market, which lowers implementation risks for banks and builds trust more quickly. While the value proposition becomes more focused, it remains compelling by delivering tangible fraud reduction with minimal disruption. The broader vision can still be communicated as the product roadmap, showing banks the long-term potential while keeping early adoption friction low.

6 Minimum (Viable) Lovable Product

To define the Minimum Lovable Product (MLP) for **TrustOS**, our goal is to deliver a targeted fraud detection solution that demonstrates tangible value to banks with minimal friction in integration. In the pilot phase, we are not aiming for a full-scale deployment but rather a focused and measurable implementation that proves the effectiveness of our approach in real-world banking scenarios.

This means working closely with one or two partner banks during a pilot phase to integrate a modular version of TrustOS into their existing infrastructure with as little disruption as possible. The goal is to validate our system's ability to flag fraud risks earlier and more accurately than current solutions, particularly in the context of social engineering, multi-channel phishing, and AI-driven attacks.

The MVP is designed for lightweight integration via an API-based interface, allowing banks to connect to TrustOS without needing to rework their core systems. The platform's modular architecture enables individual components, such as the fraud detection engine powered by natural language processing, to be deployed independently. This gives banks the flexibility to test and adopt only those parts of TrustOS that complement their existing fraud detection stack, avoiding the need for full system replacement. Such modularity is particularly valuable for institutions with complex or legacy infrastructures, where minimal disruption and rapid validation cycles are essential.

Importantly, TrustOS is built to preserve full control and autonomy for the bank. The system flags high-risk messages or transactions internally, while the final decision-making remains entirely within the institution. This ensures that customer communication and experience remain under the bank's management, supporting both compliance and trust.

Furthermore, the product is designed with regulatory compliance in mind, aligning with key frameworks such as GDPR and PSD2. Data privacy, explainability, and lawful signal processing are embedded into the system's architecture from the ground up.

From the bank's perspective, the MVP must meet clear minimum requirements: it should reduce false positives, detect sophisticated fraud attempts earlier, and integrate smoothly with existing workflows. At this early stage, the business objective is not monetization but validation, proving that TrustOS is reliable, adaptable, and capable of driving measurable improvements in fraud detection. The pilot phase will also allow us to fine-tune performance under real-world constraints, laying the foundation for future production-grade deployments.

6.1 Live Demo Code Implementation

To support our vision with a tangible proof of concept, we have developed a live MVP demo, available on GitHub at `https://github.com/marta682/Fintech3`. This working prototype was built collaboratively by the entire team and is designed to run locally, using appropriate API keys for large language model (LLM) access during message analysis.

The demo allows users to simulate a typical scam scenario by inputting a series of SMS messages, including one fraudulent example among legitimate ones. The system processes each message using NLP techniques, classifies them as either safe or potentially fraudulent, assigns a risk score, and displays the results through a simple, user-friendly web interface.

While the current version is limited to SMS messages, the demo demonstrates the real-time detection capability of our core engine and serves as a strong foundation for early testing, user feedback, and integration discussions with partner banks.

The web application was developed using Visual Studio Code (with GitHub Copilot for productivity) and version-controlled through GitHub. It features a React-based frontend and a Node.js backend. Given the time constraints of our development challenge, we implemented an “LLM-as-Judge” approach for message evaluation, allowing us to rapidly prototype the classification logic while maintaining flexibility for future enhancements.

6.1.1 Use Case: Input Example

INPUT:

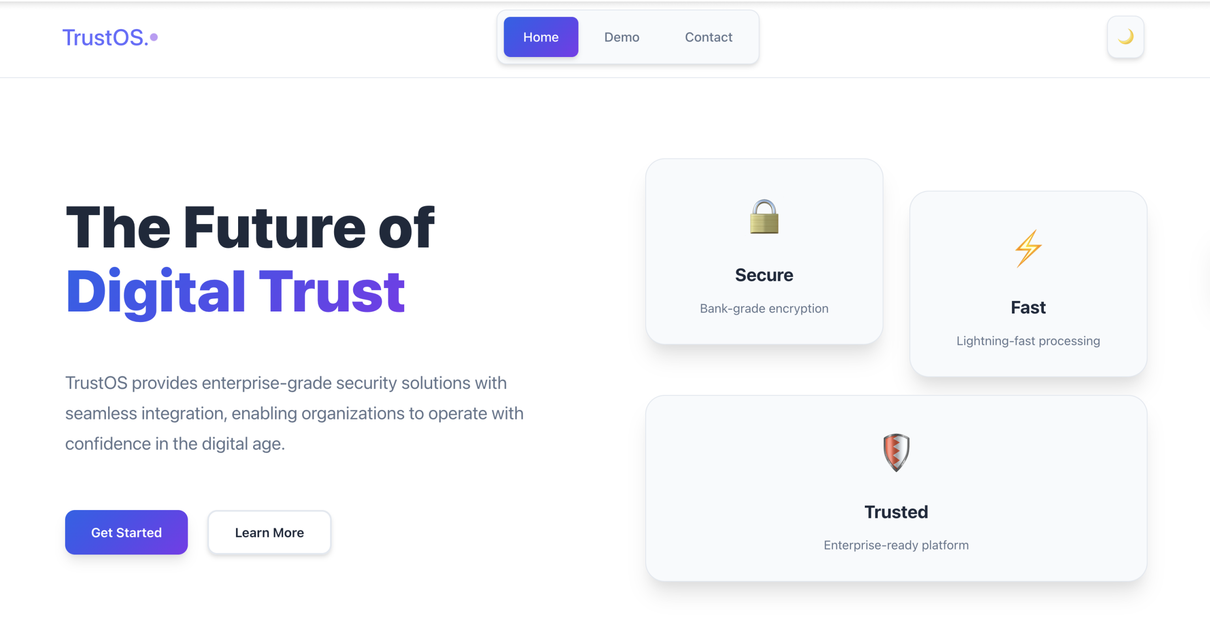
```
Type: Wire Transfer
Amount: 3,200 EUR
Payee: M. Rodriguez
Timestamp: 2025-07-16 14:22:15 PM

From: +34631234560
Content: “Hi mom, i’m alright”
From: +12345678910
Content: “Hi dad, i’m alright”
From: +34655789012
Content: “Don’t forget dinner at 8 tonight!”
From: +34611234567
Content: “NOTICE: Unusual login detected on your CaixaBank account. To avoid suspensio
From: +34698765432
Content: “Thanks for sending the report. Will review it tomorrow.”
From: +34622334455
Content: “Meeting confirmed for Thursday at 11.”
From: +34688997766
Content: “Your electricity bill of €54.30 has been paid. Thank you!”
From: +34633445566
Content: “Good luck on your presentation today!”
From: +34644556677
Content: “Hey, we’re still on for the weekend, right?”
From: +34655667788
Content: “We have received your package request. Delivery expected Friday.”
```

```
From: +34666778899
Content: “Happy birthday! Hope you have a great day!”
From: +34677889900
Content: “Here is your one-time code: 294031”
From: +34688990011
Content: “Your monthly subscription fee of €9.99 has been charged successfully.”
From: +34699001122
Content: “See you at the football game later!”
From: +34610111213
Content: “Mom’s appointment is at 3 pm today.”
From: +34621222324
Content: “Flight confirmed: MAD-LIS, departs Friday at 10:20.”
From: +34632333435
Content: “Don’t forget to bring the documents tomorrow.”
From: +34643444546
Content: “Your Uber is arriving in 2 minutes.”
From: +34654555657
Content: “Class is cancelled today, see you Monday.”
From: +34665666768
Content: “Thanks for the coffee today!”
```

As you can see, there is a suspiscios SMS: "NOTICE: Unusual login detected on your CaixaBank account. To avoid suspension, confirm your identity now at: [bit.ly/caixabank-secure]. Action required within 15 minutes."
The model successfully identified the message

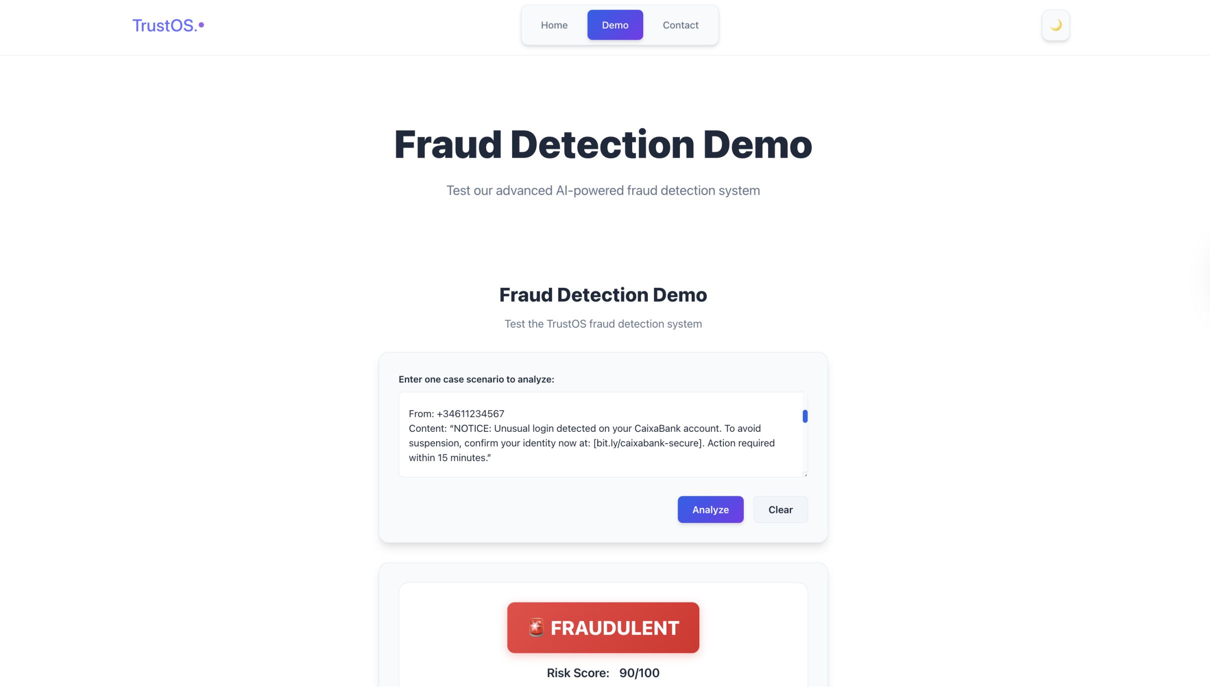
6.1.2 Use Case: Screenshots



The Future of Digital Trust

TrustOS provides enterprise-grade security solutions with seamless integration, enabling organizations to operate with confidence in the digital age.

Figure 3: Final screenshot for HomePage



Fraud Detection Demo

Test our advanced AI-powered fraud detection system

Fraud Detection Demo

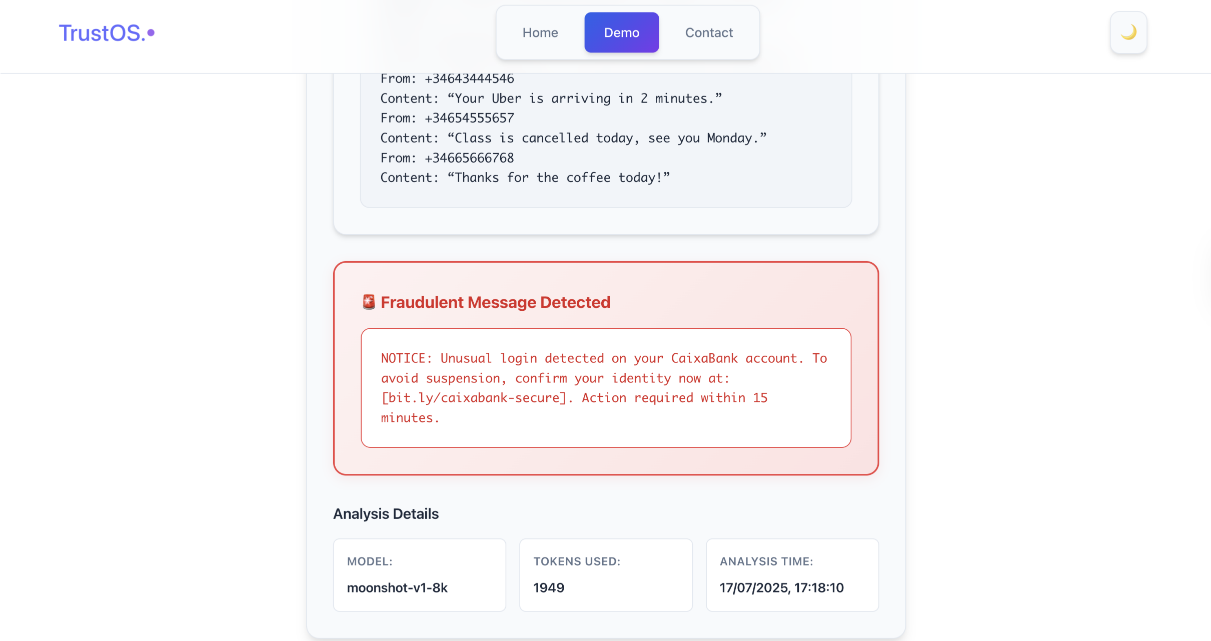
Test the TrustOS fraud detection system

Enter one case scenario to analyze:

From: +34611234567
Content: “NOTICE: Unusual login detected on your CaixaBank account. To avoid suspension, confirm your identity now at: [bit.ly/caixabank-secure]. Action required within 15 minutes.”

Risk Score: 90/100

Figure 4: Final screenshot for Analysis of the use case input



From: +34643444546
Content: “Your Uber is arriving in 2 minutes.”
From: +34654555657
Content: “Class is cancelled today, see you Monday.”
From: +34665666768
Content: “Thanks for the coffee today!”

🚨 Fraudulent Message Detected

NOTICE: Unusual login detected on your CaixaBank account. To avoid suspension, confirm your identity now at: [bit.ly/caixabank-secure]. Action required within 15 minutes.

Analysis Details

MODEL:	TOKENS USED:	ANALYSIS TIME:
moonshot-v1-8k	1949	17/07/2025, 17:18:10

Figure 5: Final screenshot for Results of the use case input

7 Bibliography

[1] *Over £570 million stolen by fraudsters in the first half of 2024.* URL: https://www.ukfinance.org.uk/news-and-insight/press-release/over-ps570-million-stolen-fraudsters-in-first-half-2024.

[2] *Check Positive Pay: The Secret Weapon Against Rising Fraud in 2025, Alkami.* URL: https://www.alkami.com/blog/check-positive-pay-banks-credit-unions-best-fraud-defense/.

[3] *New FICO Survey: Consumers Will Switch Banks If Fraud Prevention Expectations Are Not Met.* URL: https://investors.fico.com/news-releases/news-release-details/new-fico-survey-consumers-will-switch-banks-if-fraud-prevention/.

[4] *Visa's new AI tool could save the UK over £330m a year on fraud and APP scams, Visa Navigate.* URL: https://navigate.visa.com/europe/security/visas-new-ai-tool/#:~:text=Visa's%20new%20AI%20tool%20could%20save%20the%20UK%20over%20%C2%A3,on%20fraud%20and%20APP%20scams.

[5] *AI in Finance. A Powerful Tool for Fraud Detection, SoftIQ.* URL: https://softiq.io/ai-in-finance-a-powerful-tool-for-fraud-detection/#:~:text=According%20to%20a%20Forrester%20report,AI%2Dbased%20fraud%20detection%20systems..

[6] *Top scams of 2024, Consumer Advice, Federal Trade Commission.* URL: https://consumer.ftc.gov/consumer-alerts/2025/03/top-scams-2024.

[7] *The pricing strategy guide: Choosing pricing strategies that grow (not sink) your business, Paddle.* URL: https://www.paddle.com/resources/pricing-strategy.

[8] *OpenView 2023 Saas Benchmarks Report.* URL: https://openviewpartners.com/2023-saas-benchmarks-report.

[9] *Top 70 Phishing Statistics and Trends You Must Know in 2025, Keepnet.* URL: https://keepnetlabs.com/blog/top-phishing-statistics-and-trends-you-must-know.

[10] *Understanding Deal Size: A Critical Metric for SaaS Growth and Profitability, Monetizely.* URL: https://www.getmonetizely.com/articles/understanding-deal-size-a-critical-metric-for-saas-growth-and-profitability.

[11] *Banca dello Stato del Cantone Ticino, TheBanks.eu.* URL: https://thebanks.eu/banks/9505.

[12] *PostFinance Annual Report 2024.* URL: https://geschaeftsbericht.post.ch/24/ar/app/uploads/postfinance-annual-report-2024.pdf.

[13] *Recorded Future Intelligence Platform, Microsoft.* URL: https://azuremarketplace.microsoft.com/en-us/marketplace/apps/recordedfuture1605638642586.recorded_future_intelligence_platform?tab=Overview&utm_source=chatgpt.com.

[14] *Price Calculator, Alibaba Cloud*. URL: https://www.alibabacloud.com/en/pricing-calculator?_p_lc=1&spm=a2796.7960336.3034855210.1.6c50b91atEX6rg#/commodity/vm_intl.

[15] *Recommend Plan - Business Plan, Cloudflare*. URL: https://www.cloudflare.com/about-your-website/recommendation/?pr=asas3cs3.

[16] *Salaries Tools to Level Up Your Career, Levels.fyi*. URL: https://www.levels.fyi/.

[17] *2025 Spending Benchmarks for Private B2B SaaS Companies*. URL: https://www.saas-capital.com/blog-posts/spending-benchmarks-for-private-b2b-saas-companies/.