Navigating the AI Frontier: A Comparative Analysis of Fraud Prevention in Financial Payments and Digital Communications

1. Executive Summary

This report provides a comprehensive analysis of global financial payment and 1:1 digital communication transaction volumes, critically examining the feasibility of applying AI-driven fraud prevention techniques, similar to those successfully employed in the payments sector by firms like Stripe, to the vastly different landscape of digital communications. The analysis reveals a colossal disparity in transaction volumes, with communications outnumbering financial payments by orders of magnitude. While AI has proven highly effective in mitigating payment fraud, its direct replication in the communication sphere faces formidable challenges related to scale, the absence of established economic models for funding, and the nuanced risks of financial miscalculation.

The sheer volume of emails, text messages, voice calls, and social media direct messages presents an unparalleled processing challenge. Furthermore, the prevailing "free-at-point-of-use" model for most personal communication channels lacks the inherent payment rails that fund sophisticated AI systems in finance. Applying resource-intensive AI across all communication transactions without a sustainable economic framework could lead to significant operational costs and potential financial losses for service providers. This report argues that while AI is indispensable for combating the growing scourge of spam and fraud in communications, strategies must be carefully tailored to this unique environment. A direct transposition of payment-centric AI models and their associated economic structures is largely untenable, necessitating innovative, context-specific approaches to safeguard digital interactions effectively and economically.

2. The Global Scale of Financial Transactions

The global financial system facilitates a vast number of transactions daily, reflecting the pulse of economic activity worldwide. The shift towards digital and electronic payment methods has been a defining trend, accelerating the volume and velocity of these exchanges.

2.1. Overview of Annual Payment Volumes

The scale of global electronic financial transactions is immense. In 2023, the total global volume of all electronic payment transactions reached approximately 1.39 trillion. This figure is derived from industry data indicating that 266.2 billion real-time payment transactions constituted 19.1% of all electronic transactions during that year.[1] This baseline underscores the sheer number of individual economic interactions occurring digitally.

Looking ahead, the growth trajectory remains strong. Non-cash transaction volumes are projected to approach 3 trillion annually by 2028, signaling continued robust expansion in digital and electronic payments.[2] Within this broader category, real-time payments are a particularly dynamic segment, accounting for 266.2 billion transactions globally in 2023 alone. This represented a significant year-over-year (YoY) growth of 42.2%. This rapid expansion highlights a fundamental shift in how value is exchanged, moving towards instantaneous settlement and greater liquidity. The clarification that the "almost 3 trillion by 2028" figure refers to the

volume of transactions, rather than their total monetary value, is crucial for understanding the scale of interactions.[2]

The substantial YoY growth in real-time payments, far outpacing general economic growth rates, points to a powerful underlying transformation in user and business expectations. There is a clear and accelerating demand for immediacy in financial dealings. This preference for "real-time" interactions is not merely a convenience but is becoming a standard expectation, potentially influencing demands in other digital spheres, including the speed and efficacy of interventions against fraud and spam in communication channels.

2.2. Key Drivers and Trends in Financial Payments

Several interconnected factors fuel the escalating volume of financial payments. A primary driver is the increasing global internet penetration; by the fourth quarter of 2023, 5.30 billion people, or 65.7% of the world's population, were using the internet. This connectivity is increasingly mobile, with 84% of mobile phones in use being smartphones by Q4 2023. This widespread access to digital infrastructure underpins the growth of e-commerce, which inherently relies on electronic payment methods. Concurrently, financial inclusion initiatives worldwide are bringing more individuals and businesses into the formal financial system, often through digital channels.

Within this evolving landscape, digital wallets are rapidly gaining prominence. They are expected to account for over half of all global e-commerce payment methods by 2024, simplifying transactions and enhancing user experience. The growth in "non-cash" transactions, particularly "electronic transactions" and "digital payments", signifies more than just a change in payment habits. It points to the creation of an increasingly data-rich environment. Each digital transaction generates a trail of data—details of the transaction itself, device information, user behavior patterns, and more. This wealth of data is a critical prerequisite for the development and effective operation of AI-driven fraud prevention systems. Payment processors like Stripe, for instance, heavily rely on this expansive dataset to train and refine their AI models. As digital payment volumes grow, so too does the pool of data available, creating a virtuous cycle where AI systems can become progressively more accurate and adaptive. This structured, data-intensive nature of financial transactions provides a fertile ground for AI analysis, a characteristic that may differ significantly from the often unstructured and diverse content found in digital communications.

3. The Immense Volume of 1:1 Digital Communications

While financial transactions are numerous, they are significantly overshadowed by the sheer volume of 1:1 digital communications that occur globally every day. These interactions, spanning various platforms and modalities, form the bedrock of modern personal and professional connectivity. The estimates presented below, synthesized

from available global data, illustrate this staggering scale.

3.1. Aggregated Annual Volumes (Estimates)

The daily and annual figures for 1:1 communication transactions paint a picture of a digital world teeming with interactions, far exceeding the volume of formal financial exchanges.

3.2. Voice Calls

Despite the rise of text-based and asynchronous communication, voice calls remain a significant channel. Globally, an estimated **4.9 trillion phone calls** are made annually. This figure is derived from an average of 13.5 billion calls occurring each day across the globe. This persistent use of voice communication also means it remains a vector for fraudulent activities, such as vishing (voice phishing).

3.3. SMS (Short Message Service)

The humble text message continues to be a ubiquitous form of communication. Worldwide, over **23 billion text messages are sent daily**. Extrapolating this daily figure over a year (23 billion * 365 days) results in approximately **8.395 trillion SMS messages annually**. This channel is heavily utilized for both person-to-person (P2P) and application-to-person (A2P) communications, the latter including notifications, alerts, and marketing messages. Its widespread use also makes it a prime target for spam and smishing (SMS-based phishing) attacks.

3.4. Email

Email stands as a foundational tool for both personal and business communication. In

2023, an estimated **347.3 billion emails were sent and received daily** across the globe. Projections indicate a further increase to **361.6 billion daily emails in 2024**. Translating the 2023 daily figure to an annual volume yields approximately **126.76 trillion emails** (347.3 billion * 365 days). For 2024, this would be around 131.98 trillion.

A critical aspect of email volume is the prevalence of spam. Reports from 2023 indicated that around 45.6% of all global email traffic was identified as spam, with some data suggesting this figure rose to over 46.8% by December 2023 (the reference to December 2024 in S31 is likely a projection or typo). Some analyses even suggest that legitimate email constitutes less than 15% of the total global email volume. For the purposes of AI-driven filtering, the *total* volume is the relevant metric, as all messages, whether legitimate or spam, must be processed and analyzed. This massive volume of email, and the high proportion of it being unwanted, means that email filtering systems have long been contending with an enormous "background noise" of undesirable traffic. This historical context has significantly shaped the development of email security solutions. AI tools for email spam have evolved in an environment characterized by extremely high throughput and a high signal-to-noise ratio (where "noise" is spam). This evolutionary path differs from that of financial fraud detection, where fraudulent transactions, while costly, typically represent a smaller fraction of the total transaction volume. Consequently, the tolerance for false positives and false negatives, as well as the cost-benefit analysis for deploying AI in email spam filtering, are intrinsically linked to this existing reality of massive scale and high spam prevalence.

3.5. Social Media Direct Messages (DMs)

Estimating a precise global total for all P2P direct messages across the myriad of social media and messaging platforms is challenging due to platform-specific reporting and the proprietary nature of such data. However, available figures for major platforms provide a sense of the scale:

- WhatsApp users alone send over **100 billion messages daily**. This translates to an astounding **36.5 trillion messages annually**. An older figure mentioning a combined 60 billion messages per day for WhatsApp and Facebook Messenger is likely outdated given WhatsApp's current standalone volume.
- Other major platforms, including Instagram (2 billion monthly active users, MAU), Facebook Messenger (1 billion MAU), Telegram (900 million MAU), and X (formerly

Twitter), also facilitate vast numbers of DMs. While specific global daily DM volumes for these are not consistently reported, their large user bases imply tens, if not hundreds, of billions of additional DMs daily across all such platforms combined. For instance, Instagram and WhatsApp each boast 2 billion MAUs.
- Considering WhatsApp's 36.5 trillion annual messages, a conservative estimate for all other platforms combined (e.g., an additional 10-20 trillion annually) would bring the total for social media DMs to roughly **46.5 trillion to 56.5 trillion annually**.

The proliferation of closed messaging platforms like WhatsApp and other DM services introduces a distinct set of challenges for spam and fraud detection compared to open protocols such as email or SMS. Many of these platforms emphasize end-to-end encryption (E2EE) for P2P messages. While metadata (sender, receiver, timestamp, etc.) might be accessible to the platform, direct content analysis for spam or fraud by third parties—or even by the platform itself in some E2EE contexts—is significantly more complex or restricted. This contrasts sharply with email, where server-side scanning of content (albeit with privacy considerations) is a common practice. As a result, AI solutions designed for DMs may need to rely more heavily on metadata analysis, user reporting patterns, and behavioral signals rather than deep content inspection across the entire network. Alternatively, detection mechanisms might need to be implemented at the device level. This inherent limitation on data access could influence the types of AI models that are most effective and the nature of the data they can be trained on for these closed ecosystems.

3.6. Aggregated Communication Transaction Volume (Annual Estimate)

Summing the estimated annual volumes from these primary communication channels provides a staggering total:

- Voice Calls: ~4.9 trillion
- SMS Messages: ~8.4 trillion
- Email Messages (total, 2023 data): ~126.8 trillion
- Social Media DMs (conservative aggregate estimate): ~46.5 trillion
- **Total Estimated 1:1 Communication Transactions: ~186.6 trillion annually.**

This aggregate figure, while an estimate, starkly illustrates the colossal scale of digital communication when compared to financial payment volumes.

3.7. Growth Trajectories and User Engagement Patterns

Similar to the drivers in the financial payments sector, the growth in communication volumes is propelled by increasing internet penetration and mobile adoption. As of 2024, an estimated 5.5 billion people are online, up from 5.3 billion in 2023. The International Telecommunication Union (ITU) reports that while connectivity is increasing, about one-third of the world's population remains offline, primarily in low-income countries and rural areas. Young people are particularly avid internet users, with an estimated 79% of individuals aged 15-24 using the internet globally, compared to the average for the rest of the population. This persistent growth in the user base and the intensity of usage across various communication channels signify that the volume challenge for any comprehensive fraud or spam solution will only continue to escalate.

4. A Stark Comparison: Payments vs. Communication Transactions

The preceding sections have quantified the annual volumes of financial payments and 1:1 digital communications. A direct comparison reveals a dramatic difference in scale, which has profound implications for the application of AI technologies.

4.1. Quantitative Analysis of Volume Discrepancies

The disparity in transaction volumes is not merely incremental but represents a difference of orders of magnitude:

- **Total Electronic Financial Payments:** Approximately **1.39 trillion** annually (based on 2023 data).[1]
- **Total 1:1 Communication Transactions:** Approximately **186.6 trillion** annually (based on estimates for 2023 data from Section 3.6).

This comparison yields a ratio of approximately **1:134** (Communications to Payments). This signifies that for every single electronic financial transaction recorded, there are roughly 134 communication transactions (emails, calls, SMS, DMs) taking place. This direct comparison is central to understanding the core challenge posed by the user's query.

The "value per transaction" is another critical differentiating factor. Financial transactions inherently involve the movement or commitment of monetary value. Consequently, the potential loss from a single fraudulent payment can be immediate, quantifiable, and substantial. This inherently justifies higher security costs on a per-transaction basis. In contrast, the vast majority of individual communication transactions (a single email, a text message, a brief call) have no direct monetary value attached to them. While the aggregate societal and individual cost of scams facilitated through communications is enormous (as noted, over $1 trillion lost by consumers in 2024), the direct monetary cost of a single spam email or an unwanted robocall to the recipient or the service provider is often negligible or zero on an *per-instance* basis. This fundamental economic disparity in per-transaction value profoundly shapes the acceptable cost threshold for AI-driven prevention measures. AI systems in the payments sector can afford to be more resource-intensive, and thus more expensive per transaction, because the value they are protecting is significantly higher. This economic reality is a primary reason why simply replicating sophisticated payment fraud AI models directly into the communication domain is inherently problematic.

4.2. Implications of Scale Differences for AI Processing

The sheer scale difference—processing approximately 134 times more transactions—has significant technical and economic ramifications for AI systems. If AI models of similar computational intensity to those used in payment fraud detection were to be applied to every communication transaction, the demand for infrastructure, processing power, and energy would be vastly greater. This immediately flags a core challenge for the ubiquitous deployment of payment-grade AI in the communications sphere. Such an undertaking would incur significantly higher operational costs, raising serious questions about economic feasibility, potential latency in processing, and the environmental impact associated with the requisite

energy consumption.

To further illustrate this comparison, the following table summarizes the estimated annual transaction volumes:

Table 1: Global Annual Transaction Volume Comparison (Approximate, 2023 Data)

Transaction Type	Estimated Annual Volume	Primary Data Sources
Total Electronic Financial Payments	~1.39 trillion	[1]
Voice Calls	~4.9 trillion	
SMS Messages	~8.4 trillion	
Email Messages (Total)	~126.8 trillion	,
Social Media DMs (Aggregate Estimate)	~46.5 trillion	,
Total 1:1 Communication Transactions	**~186.6 trillion**	
Ratio (Communications : Payments)	**~134 : 1**	

This table provides a clear, concise, and visual summary of the disparate data points, reinforcing the magnitude of difference and setting the stage for subsequent arguments about the challenges of applying similar AI techniques due to this vast disparity in volume.

5. AI-Powered Fraud Prevention in Payments: The Stripe Paradigm

The payments industry, faced with sophisticated and ever-evolving fraud threats, has been a fertile ground for the application of advanced Artificial Intelligence (AI) and Machine Learning (ML). Stripe, a prominent global payments processor, exemplifies this trend with its Radar anti-fraud system.

5.1. Stripe's AI/ML Approach (Radar)

Stripe's Radar system is built upon a foundation of machine learning models trained on an extensive and diverse dataset. This dataset encompasses data from millions of global companies and billions of dollars in payments processed annually. As of 2024, Stripe processed over $1.4 trillion in payments volume, and a crucial aspect of its network effect is that 92% of payment cards used across its platform have been seen before, providing a rich historical context for risk assessment. Radar scans every transaction using thousands of distinct signals. These signals are drawn from multiple layers of the financial stack, including:

- **Data from financial partners:** Stripe collaborates with major card networks like Visa and Mastercard, as well as issuing banks, to incorporate data such as TC40s (fraud reports), SAFE reports (dispute data), and early dispute notifications. This allows Radar to identify potentially fraudulent charges even before a formal dispute is lodged.
- **Rich payment data:** The system leverages detailed information associated with each transaction, such as customer details (e.g., email, name), shipping and billing addresses, and other transaction-specific properties to improve ML performance.
- **Checkout flow data:** Stripe's checkout tools can automatically incorporate buyer patterns from a business's website or mobile app, helping to detect anomalous payment behaviors.

The AI techniques employed by Radar are sophisticated and adaptive. They include device fingerprinting (tying multiple signals to a single device profile), analysis of historical snapshots (to spot recurring patterns over years of data), proxy detection (to identify IP spoofing), and the generation of comprehensive risk scores by combining these multiple signals. A key feature of Radar's architecture is its dynamism; the ML models, including those customized for specific businesses, are retrained daily. This continuous learning process allows the system to quickly adapt to shifting fraud patterns and newly emerging attack vectors. For instance, Stripe uses AI

to estimate the overall prevalence of "card testing"—a common fraud tactic where criminals make small, phony payments to validate stolen card details—and updates its risk systems on a daily basis to counter this threat.

Stripe's AI infrastructure also benefits enormously from what can be described as a nearly "closed-loop" view of transactions within its network, coupled with deep integrations with financial partners. The platform sees a transaction from its initiation by a customer, through its internal processing, and onward to interactions with card networks and issuing banks. Information such as early dispute notifications provides relatively quick and high-fidelity feedback on confirmed fraudulent activity. This tight feedback loop is essential for the rapid retraining of its models and the continuous improvement of their accuracy. This contrasts with the communication domain, where feedback loops are often more fragmented. User reports of spam, while valuable, can be inconsistent, delayed, or subjective. Conclusively confirming that a specific communication was indeed fraudulent or part of a scam is often more challenging than confirming a fraudulent financial charge, making the training of equally powerful AI models in communications a more complex endeavor.

5.2. Effectiveness and Economic Model

The effectiveness of Stripe's AI-driven approach is demonstrable. The company reported that its technology reduced "card testing" attacks by 80% over a two-year period, even as its overall payment volume doubled during that time. This highlights the system's ability to scale and adapt. Similarly, Visa's AI models, operating on similar principles of large-scale data analysis, are credited with preventing nearly $30 billion in fraud annually. Within Stripe's Radar, the introduction of adaptive rules, which combine ML risk scores with real-time issuer responses (e.g., CVC or postal code verification results), has been shown to increase legitimate payment success rates by an average of 1.3 percentage points with minimal impact on fraud rates.

The economic model underpinning Stripe's Radar is intrinsically linked to its core business of payment processing. Businesses that use Stripe pay transaction fees for payment acceptance, and Radar is a fundamental, integrated feature designed to protect both Stripe and its merchants from the financial losses associated with fraud. There isn't a separate, itemized "fee for AI;" rather, robust fraud prevention is a core component of the value proposition offered to merchants. This integrated funding mechanism is crucial: the direct financial cost of fraud in payments—borne by

merchants in the form of lost goods or services, chargeback fees, and by payment processors like Stripe in managing disputes and maintaining network integrity—creates a powerful and direct financial incentive to invest heavily in sophisticated AI prevention systems. This clear financial pain point directly drives the development and continuous improvement of solutions like Radar. This incentive structure differs markedly from the communications sector, where the direct financial cost of a single spam message to the service provider is often negligible, even if the aggregate societal costs of scams are high.

6. The Pervasive Challenge of Fraud and Spam in Communications

Digital communication channels, while indispensable for modern life, are unfortunately rife with fraudulent and unsolicited activities. These threats are diverse, constantly evolving, and have significant financial and societal impacts.

6.1. Types and Impact of Communication-Borne Threats

The spectrum of malicious activities perpetrated through communication channels is broad:

- **Email Spam and Phishing:** Unsolicited commercial emails (spam) remain a massive problem, with categories including marketing/advertising (accounting for nearly 36% of spam), adult content (around 31.7%), and financial solicitations (about 26.5%). More perniciously, email is the primary vector for phishing attacks, with an estimated 96% of such attacks conducted via email. Business Email Compromise (BEC) attacks, a sophisticated form of phishing targeting organizations, resulted in an average cost of $5.96 million per incident in 2021.
- **SMS Spam and Smishing:** SMS messages are used for "smishing" – phishing attacks delivered via text. Proximus Global's international connectivity enabler, BICS, reported tracking 331 million smishing messages worldwide in 2024.
- **Voice Call Fraud (Vishing and Robocalls):** Voice phishing, or "vishing," involves scammers using phone calls to deceive victims. This threat is being amplified by AI, which can generate highly convincing synthetic voices, including cloning the

voices of real individuals, to enhance deception. Robocalls, often delivering scam messages, also persist as a major nuisance and threat.
- **Social Media Scams:** Direct messaging features on social media platforms are exploited for various scams, including AI-powered phishing campaigns and the dissemination of deepfakes (AI-manipulated videos or audio) to defraud or misinform users.

The collective financial impact of these communication-borne threats is staggering. According to the Global Anti-Scam Alliance, scammers successfully stole an estimated $1.03 trillion from consumers globally in 2024 alone, with SMS and phone calls identified as leading delivery methods for these scams. This underscores not only the severity of the problem but also the substantial potential value of effective AI solutions, provided they can be implemented in a sustainable and economically viable manner.

6.2. Current State of AI Application in Communication Security

Artificial intelligence is not a newcomer to the field of communication security; however, its application, sophistication, and economic integration vary significantly across different channels and providers.

- **Email Security:** AI is increasingly used to enhance traditional email filtering mechanisms (which often rely on rule-based systems or Bayesian probability). Modern AI approaches leverage Machine Learning (ML), Natural Language Processing (NLP), and deep learning techniques to analyze email content, context, tone, sender reputation, and user interaction patterns. This allows for more accurate and adaptive spam and phishing detection. Companies such as Abnormal Security and EmailTree.ai specialize in AI-driven email security solutions for businesses.
- **SMS Security:** In the SMS domain, AI-powered solutions like Proximus Global's 365guard employ advanced analytics and dynamic threat detection engines. These systems are designed to counter spam, fraud, and smishing by continuously learning and adapting to regional threat landscapes and novel attack vectors employed by fraudsters.
- **Voice Call Security:** AI is applied to analyze call patterns (e.g., unusual call frequency or duration), keywords within conversations, and other behavioral clues to identify robocalls and vishing attempts. Consumer applications like Hiya and

Truecaller utilize AI for spam call blocking, often incorporating user-reported data. Speech analytics platforms, such as CallMiner Eureka, can detect scam-related phrases or emotional cues indicative of manipulation during calls. Paradoxically, AI is also being exploited by attackers to create more convincing vishing attacks through sophisticated voice cloning technologies.

- **Social Media DM Security:** Major social media platforms inherently use AI algorithms to detect and remove spam, fake profiles, and scam attempts within their ecosystems, including direct messaging services. Third-party AI tools, like Google's Perspective API for scoring comment impact and Microsoft's Azure Cognitive Services for content moderation, also offer capabilities applicable to social media content. However, attackers are also leveraging AI to craft highly personalized and convincing phishing messages disseminated via social media channels.

This overview indicates that AI is an active component in the defense against communication fraud. A significant "arms race" is clearly underway, where AI is wielded by both attackers and defenders. Malefactors use AI to generate more believable fraudulent content, automate their attacks at scale, create deepfakes, and craft personalized phishing lures designed to bypass traditional filters. In response, defenders deploy increasingly sophisticated AI to detect these evolving threats. This dynamic creates a continuous cycle of innovation on both sides, implying that any AI solution in the communications sphere must be exceptionally adaptive and constantly updated, which adds to its inherent complexity and operational cost. Unlike some types of payment fraud that might exploit static vulnerabilities in systems, communication fraud is increasingly dynamic, often relying on sophisticated social engineering tactics that require more advanced behavioral AI for effective detection.

Furthermore, while the effectiveness of current AI in communication spam and fraud filtering is improving, it is by no means foolproof. These systems still face challenges, including the occurrence of false positives (legitimate messages mistakenly flagged) and the ability of spammers to devise new techniques to evade detection. Traditional filtering methods have well-documented limitations that AI aims to overcome, but the journey to near-perfect accuracy is ongoing. The existence of a competitive market for third-party AI-based security solutions for email and messaging suggests that the default protections offered by platform providers may not be sufficient for all users or all threat levels. This indicates a potential gap between the current state of AI in communications security and an ideal, highly effective, universally deployed system. This gap may be attributable to a combination of technical challenges, the cost of deployment at scale, or the sheer diversity of communication content and legitimate

user intent, which makes accurate classification inherently difficult.

7. Applying Payment-Grade AI to Communications: A Critical Assessment

The success of AI in mitigating payment fraud, as exemplified by Stripe's Radar, naturally raises the question of whether similar approaches can be effectively and economically applied to the communication domain. However, a critical assessment reveals fundamental differences that make a direct transposition of these strategies problematic.

7.1. The Volume Conundrum: Technical and Processing Challenges at Scale

As established in Section 4, the volume of 1:1 communication transactions (estimated at ~186.6 trillion annually) is approximately 134 times higher than that of electronic financial payments (~1.39 trillion annually). Attempting to apply AI models with computational demands comparable to Stripe's Radar to every email, SMS, voice call, and direct message would necessitate an unprecedented scale of processing power. This raises immediate and serious concerns regarding:

- **Latency:** Real-time analysis of such vast data streams could introduce unacceptable delays in message delivery or call connection.
- **Infrastructure Cost:** The capital expenditure for servers, storage, and networking equipment would be astronomical.
- **Operational Cost:** Ongoing expenses for maintenance, software updates, and expert personnel would be immense.
- **Energy Consumption:** The environmental impact and financial cost of powering such a massive AI infrastructure would be substantial. S49 notes, for example, that the energy required for a ChatGPT query is significantly higher than for a standard Google search, illustrating the energy demands of complex AI.

This sheer scale difference is a primary hurdle, suggesting that strategies optimized for the relatively lower volume (though high value) of payment transactions may not

be directly scalable to the hyper-volume environment of communications.

7.2. The "No Payment Rail" Dilemma: Funding Sophisticated AI Without Direct Transaction Fees

A crucial distinction lies in the economic models. Stripe's Radar AI system is funded as an integral part of its core payment processing services; businesses pay transaction fees, and robust fraud prevention is a key component of the value they receive. This creates a direct revenue stream linked to transaction volume that can support the development and operation of sophisticated AI.

In stark contrast, most P2P communication channels lack such an inherent per-transaction fee structure that could fund intensive AI scrutiny:

- **Email:** Services like Gmail and Outlook are typically free for individual users, with costs absorbed by the provider (often monetized through advertising or other services) or borne by businesses for corporate email solutions.
- **SMS:** Text messaging is usually bundled into mobile phone plans. Mobile Network Operators (MNOs) may absorb the costs of basic SMS filtering or offer enhanced security as a value-added service to retain customers or meet regulatory requirements.
- **Social Media DMs:** These services are "free" to end-users, with the platform's operational costs, including security measures, funded by its overall business model (e.g., advertising, data analytics, premium features for businesses).

This absence of direct "payment rails" for individual P2P communications is a central challenge. Without a clear and accepted revenue stream tied to each message or call, financing universally deployed, highly sophisticated AI systems becomes a major economic hurdle. While S51 notes that spam incurs costs for organizations through lost employee working time, and spam filters can reduce these costs (implying an indirect ROI for businesses that pay for email services), this does not translate to a direct funding model for P2P communications for the general public. The idea of making spam sending economically unprofitable by charging a nominal fee for every email sent has been discussed but has not seen widespread adoption, particularly for P2P interactions, due to potential impacts on legitimate communication and user acceptance.

The "free" nature of many widely used communication services has conditioned users

to expect zero-cost interactions for basic P2P messaging and calls. Attempting to introduce direct fees specifically for enhanced AI security on these channels—such as a per-message fee or a general security surcharge—would likely face significant user resistance. This is unlike the payments domain, where transaction fees are an established and accepted norm for merchants, and sometimes indirectly for consumers. This makes the "payment rails" issue not merely a technical one of lacking a mechanism to charge, but also a profound market acceptance and business model challenge. Financial losses could indeed occur if platforms invest heavily in costly AI solutions and then either fail to monetize them effectively or lose users due to the introduction of new, unwelcome fees.

7.3. Risk of Financial Miscalculation: Potential for High Operational Costs, False Positives, and User Impact

Deploying payment-grade AI across the communication spectrum carries substantial risks of financial miscalculation:

- **High Operational Costs:** As discussed, the sheer scale implies massive, ongoing operational expenditures. If these costs are not adequately offset by new direct revenue, clear and substantial indirect cost savings (e.g., significantly reduced customer support loads due to fewer fraud incidents), or tangible improvements in user retention and engagement that boost other revenue streams (like advertising), service providers could face significant financial losses.
- **Impact of False Positives:** At the volumes seen in communications, even an extremely low false positive rate (where legitimate messages are incorrectly flagged as spam or fraudulent) would affect a massive absolute number of communications. This could disrupt critical business correspondence, interfere with important personal interactions, lead to widespread user frustration, and ultimately drive users away from a platform. The "cost of error" for a misclassified email or DM, while different from a misclassified payment, can still be very high in aggregate or for specific crucial communications.
- **Consequences of False Negatives:** Conversely, if an expensive AI system fails to catch sophisticated fraud or widespread spam campaigns, users still suffer harm, and the platform or service provider faces reputational damage and potential loss of trust, which can also translate into financial losses.

These factors highlight that financial losses are not solely about the direct cost of

implementing AI but also encompass the broader economic consequences of its imperfections when operating at such an unprecedented scale.

7.4. Accuracy and the "Cost of Error" in High-Volume, Low-Individual-Value Systems

While AI systems like Stripe's Radar aim for high precision (minimizing false positives) and recall (minimizing false negatives), the tolerance for error and the definition of "cost of error" differ between payments and communications. In payments, a false positive might mean a legitimate transaction is declined, causing customer inconvenience and potentially lost sales. A false negative means a fraudulent transaction is approved, leading to direct financial loss.

In communications, the "value" of an individual message is often subjective, contextual, or non-monetary. A false positive could mean a critical job offer email lands in spam, or an urgent family message is never seen. A false negative means a user is exposed to a scam or unwanted content. The challenge is to optimize AI for a context where individual data points (messages) are incredibly numerous and often of low intrinsic "value" in isolation, yet the system as a whole must maintain a high level of trustworthiness and protect users from significant aggregate harm.

The diversity of content and intent in communications is also far greater than in payment transactions. Payment transactions, while varied, ultimately relate to the exchange of value and follow relatively structured protocols and data formats. Communications, on the other hand, encompass an almost infinite variety of topics, tones, languages, slang, cultural nuances, and intents (informational, social, transactional, persuasive, malicious, etc.). Training an AI to accurately and consistently distinguish legitimate intent from malicious intent across billions of unique, often unstructured, interactions is an immense technical challenge. This inherent complexity increases the risk of both false positives and false negatives and likely requires more sophisticated (and potentially more computationally expensive) NLP and contextual understanding capabilities than are typically required for identifying payment fraud patterns, which are often more reliant on structured data signals (e.g., transaction velocity, geographic anomalies, device characteristics).

The following table provides a comparative analysis of AI fraud/spam prevention in

payments versus communications:

Table 2: AI Fraud/Spam Prevention: Payments vs. Communications – A Comparative Analysis

Feature	Payments (e.g., Stripe)	Communications (General P2P)
Typical Annual Transaction Volume	~1.39 trillion	~186.6 trillion
Primary Funding Model for AI Security	Integrated into transaction fees	Provider-absorbed; advertising; potential premium tiers
Direct Monetary Value per Transaction	Typically high; direct financial value	Often zero or very low for individual messages
Data Richness for AI Training	High; structured transaction data, network data	Variable; often unstructured; privacy constraints
Nature of "Loss" from Fraud/Error	Direct financial loss; lost sales	User harm (scams); annoyance; reputational damage
Tolerance for False Positives	Low to moderate (impacts sales/UX)	Varies; can be high impact for critical messages
Cost of AI Implementation (Per Tx)	Justifiable by value protected	Potentially prohibitive due to low per-tx value
Primary Incentive for AI Investment	Direct reduction of financial fraud losses	User protection; platform trust; regulatory compliance

This comparative analysis underscores the fundamental differences that make a simple "lift and shift" of payment AI strategies to the communication domain highly problematic.

8. Strategic Considerations for Mitigating Communication

Fraud/Spam with AI

Given the challenges of volume, funding, and complexity, applying AI to combat fraud and spam in communications requires nuanced and tailored strategies rather than a monolithic, payment-inspired approach. The goal is to harness AI's power sustainably and effectively.

8.1. Exploring Sustainable Funding Models

Addressing the "no payment rail" dilemma for P2P communications necessitates exploring alternative or supplementary funding mechanisms for sophisticated AI security:

- **Tiered Service Models:** Communication providers could offer basic levels of spam and fraud filtering for free, with more advanced, AI-powered protection available as a premium feature. This might be particularly attractive to businesses relying on these channels or to individuals seeking enhanced security.
- **Provider-Absorbed Costs as a Differentiator:** Many communication providers (ISPs, MNOs, social media platforms) already treat a certain level of security as a cost of doing business. They could further invest in AI security, funded by their primary revenue streams (e.g., subscriptions, advertising, data monetization for enterprise services), and use superior safety and trust as a key market differentiator.
- **Industry Coalitions and Data Sharing:** Collaborative efforts among industry players to share anonymized threat intelligence and potentially even aggregated data for training AI models could lead to more effective and cost-efficient solutions for all. However, such initiatives face significant competitive, legal, and privacy hurdles.
- **Regulatory Mandates and Incentives:** Governments could play a role by mandating minimum security standards for communication services or by providing incentives for the adoption of advanced anti-fraud technologies. Compliance with such regulations could then justify the necessary investment.
- **Indirect Monetization through Enhanced User Experience:** For platforms heavily reliant on user engagement (e.g., social media), improved security that reduces spam and fraud can lead to higher user satisfaction, increased time

spent on the platform, and greater retention. These positive outcomes can indirectly boost primary revenue streams like advertising or in-platform e-commerce.

The most viable path forward likely involves a hybrid approach. Different levels of AI intensity and diverse funding models may be applied to different segments of the communication ecosystem. For instance, A2P messaging (such as business SMS notifications or alerts) often has clearer monetization paths and a stronger business case for robust security to protect brand reputation and customer trust. Enterprise email security is already an established market with paid solutions. P2P consumer communications remain the most challenging segment to fund directly for highly advanced, resource-intensive AI protection. Therefore, future solutions will likely not be monolithic but rather a patchwork of specialized AI applications, with varying degrees of sophistication, funded through a combination of these diverse mechanisms. The "mistake" to avoid is not the use of AI itself, but rather *how* it is applied and funded without considering these segmental differences.

8.2. Balancing Security, User Experience, and Cost

A "boil the ocean" strategy—applying the most computationally intensive AI to every single message, call, or DM—is likely economically unviable and potentially detrimental to user experience due to latency or false positives. A more pragmatic and sustainable approach involves carefully balancing security effectiveness, user impact, and operational cost:

- **Risk-Based Application of AI:** Implement AI systems that apply varying levels of scrutiny based on assessed risk. For example, messages containing suspicious links, specific financial keywords, originating from unknown or untrusted contacts, or exhibiting anomalous patterns could trigger more intensive AI analysis. Lower-risk communications (e.g., between known and trusted contacts) might undergo lighter-touch AI filtering. Stripe itself employs adaptive rules based on risk scores to manage its payment fraud detection, a principle that can be adapted to communications.
- **Focus on User Empowerment and Feedback:** Provide users with intuitive tools to easily report spam and fraud, manage personal blocklists and allowlists, and customize the sensitivity settings of their filters. This user-generated feedback is invaluable for training AI models and can augment their effectiveness.

- **Transparency and Control:** Communicate clearly with users about how AI is being used to protect them, what types of data are analyzed (while respecting privacy), and what options they have to control these features.
- **Continuous Optimization and Iteration:** Regularly evaluate the performance of AI systems, carefully monitoring the trade-offs between security effectiveness (reduction in successful fraud/spam), the user friction caused by false positives, and the ongoing operational cost of the AI solution. This iterative approach allows for refinement and adaptation over time.

8.3. The Role of Regulation and Industry Collaboration

External factors, including regulation and industry-wide collaboration, can significantly influence the feasibility and adoption of AI-driven security measures in communications.

- **Regulation:** Thoughtful regulation can establish baseline security standards for communication service providers, create clearer frameworks for action against malicious actors, and potentially foster a market for advanced security solutions by setting expectations for due diligence.
- **Industry Collaboration:** Sharing threat intelligence, developing common standards for identifying and reporting new fraud tactics, and even collaborating on the development of foundational AI models (where feasible and legally permissible) could improve the overall efficacy of security measures and reduce the duplicated effort and cost for individual providers. Initiatives like Proximus Global's 365guard, presented as an industry-level solution for MNOs, point towards the potential of such collaborative frameworks.

Crucially, the human element remains indispensable in this complex equation. AI alone, no matter how advanced, is not a silver bullet. User education on recognizing evolving scam tactics, the provision of clear and accessible reporting mechanisms for suspicious activity, and the availability of human oversight for complex cases or appeals against AI-driven decisions are all essential components of a comprehensive and trustworthy security strategy. Social engineering attacks, which AI itself can unfortunately enhance, often prey on human psychology, making awareness and critical thinking vital. User reporting, as seen with tools like Truecaller, serves as a valuable data source for training AI and rapidly identifying novel attack campaigns. An over-reliance on AI without robust, human-centric safeguards and feedback loops

could lead to user distrust, system failures, and ultimately, continued financial losses from fraud that exploits these gaps.

9. Conclusion and Expert Recommendations

The remarkable success of AI-driven systems like Stripe's Radar in combating payment fraud offers valuable lessons in leveraging large-scale data, adaptive machine learning, and integrated economic models. However, this report concludes that a direct transposition of such AI strategies to the far more voluminous and economically distinct domain of 1:1 digital communications would be ill-advised and fraught with financial risk. The communication landscape's vastly greater transaction volume (approximately 134 times that of electronic payments), its predominantly "free-at-point-of-use" P2P economic models lacking direct payment rails, and the sheer diversity of its content present unique and formidable challenges.

Nevertheless, AI is not merely an option but an essential tool in the ongoing battle against the pervasive and costly problems of communication-borne fraud and spam. The key to avoiding significant financial missteps lies not in shying away from AI, but in making strategic, economically sound, and contextually appropriate decisions regarding its application.

Recommendations for deploying AI in communication security include:

1. **Adopt Risk-Based AI Deployment:** Implement tiered AI scrutiny, applying more computationally intensive and sophisticated models to higher-risk communications (e.g., those involving financial keywords, unknown senders, or suspicious links) while using more efficient, lighter-touch AI for lower-risk interactions. This optimizes resource allocation and minimizes user friction.
2. **Explore Hybrid and Sustainable Funding Models:** Move beyond the expectation of a single, transaction-fee-based model. Investigate and implement a mix of funding strategies, including premium tiered services for enhanced protection, treating robust security as a core operational cost funded by primary revenue streams (especially for platform providers), fostering industry collaborations for shared infrastructure or intelligence, and aligning with regulatory frameworks that may incentivize or mandate security investments.
3. **Prioritize Computationally Efficient AI for Mass-Market Application:** Given the extreme volumes, research and deploy AI algorithms that are optimized for

efficiency and scalability, minimizing the per-message processing cost without unduly sacrificing accuracy for the majority of communications.

4. **Invest in AI Excelling in Contextual Understanding and Anomaly Detection:** The unstructured and diverse nature of communication content requires AI that can go beyond simple keyword spotting or pattern matching. Advanced NLP, behavioral analytics, and anomaly detection capabilities are crucial for accurately identifying nuanced threats within a wide array of legitimate communication styles.
5. **Foster Industry Collaboration and Robust User Feedback Mechanisms:** Encourage the sharing of anonymized threat data and best practices among service providers. Critically, empower users with intuitive tools for reporting suspicious activity and provide transparent feedback loops, as user input is invaluable for training AI and identifying novel threats.
6. **Commit to Continuous Adaptation in the AI "Arms Race":** Recognize that fraudsters will continually adapt and also leverage AI. Defensive AI strategies must therefore be dynamic, with ongoing investment in model retraining, research into new detection techniques, and a proactive stance against emerging threats.

In conclusion, the mistake would be to either underestimate the necessity of AI in securing digital communications or to misjudge the unique operational scale, ignore the distinct funding realities, or underestimate the multifaceted "cost of error" when applying sophisticated AI without a clear, sustainable, and context-aware plan. Strategic, tailored, and economically viable AI deployment is paramount to protecting users and preserving trust in the digital communication ecosystem.

Works cited

1. Prime time for real-time global payments report | ACI Worldwide, accessed May 7, 2025, https://www.aciworldwide.com/real-time-payments-report
2. Global Payments Market Report 2025: Digital Payments Growth ..., accessed May 7, 2025, https://www.businesswire.com/news/home/20250305938222/en/Global-Payments-Market-Report-2025-Digital-Payments-Growth-Continues-to-Accelerate-with-Market-Set-to-Exceed-USD-3-Trillion-by-2028---ResearchAndMarkets.com