

Why You've Got Mail: Evaluating Inbox Privacy Implications of Email Marketing Practices in Online Apps and Services

Scott Seidenberger
University of Oklahoma
Norman, OK, USA
seidenberger@ou.edu

Oluwasijibomi Ajisegiri
University of Oklahoma
Norman, OK, USA
oluwasijibomi.ajisegiri@ou.edu

Noah Pursell
University of Oklahoma
Norman, OK, USA
noah.a.pursell-1@ou.edu

Fazil Raja
University of Oklahoma
Norman, OK, USA
fazilraja13@gmail.com

Anindya Maiti
University of Oklahoma
Norman, OK, USA
am@ou.edu

Abstract

This study explores the widespread perception that personal data, such as email addresses, may be shared or sold without informed user consent, investigating whether these concerns are reflected in actual practices of popular online services and apps. Over the course of a year, we collected and analyzed the source, volume, frequency, and content of emails received by users after signing up for the 150 most popular online services and apps across various sectors. By examining patterns in email communications, we aim to identify consistent strategies used across industries, including potential signs of third-party data sharing. This analysis provides a critical evaluation of how email marketing tactics may intersect with data-sharing practices, with important implications for consumer privacy and regulatory oversight. Our study findings, conducted post-CCPA and GDPR, indicate that while no unknown third-party spam email was detected, internal and authorized third-party email marketing practices were pervasive, with companies frequently sending promotional and CRM emails despite opt-out preferences. The framework established in this work is designed to be scalable, allowing for continuous monitoring, and can be extended to include a more diverse set of apps and services for broader analysis, ultimately contributing to transparency in email address privacy practices.

Artifacts: <https://www.doi.org/10.17605/OSF.IO/M36T8>

CCS Concepts

• **Security and privacy;** • **Information systems;**

Keywords

Email Privacy, Email Marketing, Unsolicited Email, Spam.

ACM Reference Format:

Scott Seidenberger, Oluwasijibomi Ajisegiri, Noah Pursell, Fazil Raja, and Anindya Maiti. 2025. Why You've Got Mail: Evaluating Inbox Privacy Implications of Email Marketing Practices in Online Apps and Services. In *Proceedings of the Fifteenth ACM Conference on Data and Application Security*



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

CODASPY '25, Pittsburgh, PA, USA

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1476-4/2025/06

<https://doi.org/10.1145/3714393.3726516>

and Privacy (CODASPY '25), June 4–6, 2025, Pittsburgh, PA, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3714393.3726516>

Taxonomies

We distinguish email sources based the following classifications:

- **Internal:** Emails sent directly by the service's own infrastructure. They typically should pass Sender Policy Framework (SPF) and DomainKeys Identified Mail (DKIM) checks.
- **Authorized Third-Party (ATP):** Emails sent on behalf of the service by authorized providers, authenticated by the service. They typically should also pass SPF and DKIM checks.
- **Unknown Third-Party (UTP):** Emails from unauthorized or unknown entities. If such an email is masqueraded as coming from the service, it typically will fail SPF and DKIM checks.

Unsolicited emails are grouped into these broad categories:

- **Service-Originated Spam (SOS):** Unsolicited internal and ATP emails that pass SPF and DKIM checks.
- **Unknown or Unauthenticated Sender Spam (UUSS):** Emails from UTP sources and emails that fail SPF and DKIM checks (even if the header indicates service-originated in the From: field).

Emails are further categorized by purpose:

- **CRM:** Customer Relationship Management emails aimed at engagement and relationship-building.
- **Promotional:** Emails with overt selling intent, such as discounts or offers.
- **Alerts:** Informational emails, such as account notifications or updates.

1 Introduction

Public perceptions of data privacy, particularly concerning personal data such as email addresses, can be significantly shaped by high-profile data breaches, feelings of powerlessness, and general misunderstandings. Isolated incidents of major data breaches contribute to the growing concern among users about the security of their personal information. For instance, the Yahoo data breach in 2013–2014 affected all 3 billion user accounts, exposing email addresses and other personal data [19]. Similarly, the Equifax breach in 2017 compromised the personal information of approximately 147 million people, including email addresses and Social Security numbers [20]. These incidents have heightened public awareness of vulnerabilities in corporate data handling practices. Such breaches

exacerbate the concept of “digital resignation” where individuals feel powerless to control their personal information, accepting that their email addresses and other data may be shared or exposed without their consent [13]. This resignation is further fueled by past corporate practices that shared data with third-parties [14], before data privacy laws such as GDPR [32] and CCPA [22] were enacted. Hoofnagle et al. [17] found that users are often unaware of the extent to which their email addresses are collected and shared with third-parties, increasing feelings of vulnerability and mistrust. Furthermore, Acquisti et al. [3] found that even users who express privacy concerns can themselves engage in behaviors that compromise their email data, often due to a lack of viable alternatives or understanding.

A significant challenge for users is the difficulty of identifying which app or service may have leaked or shared their email address when they start receiving unsolicited spam or phishing emails. This issue arises because most users generally reuse the same email address across multiple online services and apps. Consequently, when unsolicited communications are received from third-parties, users may not be able to determine the specific source responsible for any unauthorized sharing or leakage of their email information. This difficulty in identifying the source of leaked or shared email addresses is further exacerbated by pervasive email marketing practices, which were often connected to the sharing of email and other personal data with third-parties. Companies often engaged in email marketing strategies that involve partnering with affiliates, utilizing data brokers, or participating in advertising networks to broaden their outreach [26]. These practices frequently entailed sharing or selling users’ email addresses, sometimes without informed consent or clear disclosure. As a result, users could receive unsolicited promotional emails from entities with whom they have no direct relationship, obscuring the boundaries of accountability. The interconnection between email marketing and email sharing practices not only increases the burden of unsolicited communications [4, 31] but also undermines users’ ability to trust that their personal information is handled responsibly.

After the introduction of data privacy laws like GDPR and CCPA, there still remains a significant gap towards systematically auditing online services and apps for their email address sharing and email marketing practices. Email marketing strategies have become increasingly sophisticated, often blurring the lines between legitimate marketing and unauthorized data sharing with third-parties. Without robust auditing mechanisms, it’s challenging to assess compliance with these regulations, hold companies accountable for improper handling of email data, or impose consequences for sending unsolicited emails despite user opt-outs. Developing such a framework is crucial for enhancing transparency and accountability in email marketing practices, ultimately improving user perceptions by rebuilding trust in how their personal information is managed.

Towards addressing this gap, our study offers a structured and scalable framework to assess how online services and apps handle users’ email addresses, from the perspective of an user inbox. Our contributions include the following key steps and insights:

- **Audit of Online Services and Apps:** We conducted a year-long audit of email communications received after registering with the top 150 popular online services and apps, opting out

of any optional communication during the registration process. No unknown or unauthenticated sender spam (UUSS) emails were detected within this time frame, suggesting that opting out during account creation is very effective in limiting unsolicited emails from unknown and unauthenticated entities.

- **Email Marketing Analysis:** Despite no UUSS emails, we collected 4,847 emails received from these online services and apps, directly (internal) or from authorized third-parties (ATP). Our analysis revealed that companies still frequently used customer relationship management (CRM), alert, and promotional emails to engage users, even when opting out of promotional communications. This highlights the persistence of well-established marketing strategies, majority of which may be classified as service-originated spam (SOS).
- **Cross-Sector Insights:** We identified significant differences in email marketing practices across sectors. Consumer industries, such as omnichannel retail, heavily relied on promotional campaigns, while financial services prioritized CRM and alert emails. Some of these sector-specific behaviors offer insights into varying risks to inbox privacy.

These findings contribute to a broader understanding of email marketing and data sharing practices, promoting transparency and accountability in how online services and apps handle users’ email addresses.

2 Related Work

Email and inbox privacy has become a critical issue in the wake of increased data breaches and unauthorized sharing of personal information. Englehardt et al. [14] demonstrated the widespread use of email tracking, allowing third parties to gather user behavior data without consent. Their findings emphasize how email addresses, once shared, become tools for covert tracking, often leading to privacy violations. Similarly, Hoofnagle et al. [17] found that most users are unaware of how their email data is shared with third parties for advertising purposes, heightening concerns about privacy infringement. Plice et al. [23] provided an economic analysis of unwanted commercial messages and their dependence on weak privacy controls. These studies highlight the necessity of open auditing mechanisms to protect email privacy.

Email marketing remains a central tool for customer engagement, but its effectiveness is tied to timing, frequency, and personalization. Thomas et al. [28] found that emails with lower levels of overt persuasion, such as CRM emails, tend to have higher open and click-through rates compared to promotional emails, aligning with the *Persuasion Knowledge Model* [16]. Similarly, Nuutinen [21] emphasized the role of segmentation and personalization in improving email performance, noting that content relevance significantly impacts open and conversion rates. The importance of timing has also been explored extensively. Araujo et al. [7] proposed a machine learning approach using regression models like Random Forest and LSTM to predict the optimal send times for marketing emails. Their study demonstrated that predictive algorithms could improve open rates by sending emails when users are most likely to engage, highlighting the importance of behavioral analytics in email marketing. Abrahams et al. [1], in their multi-industry study of large U.S. franchises, found that industries like hospitality and

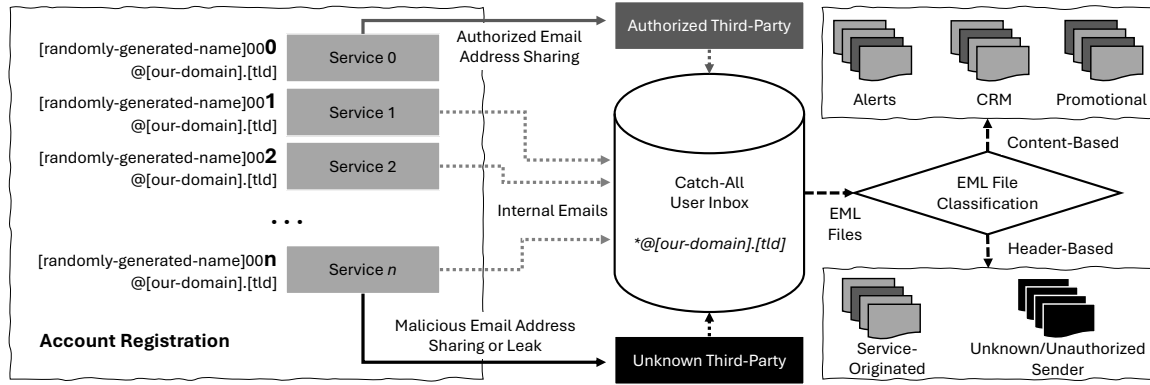


Figure 1: Framework for email collection, processing, and classification to analyze inbox privacy across services.

food service were prolific in email frequency, often sending 2-4 emails per week, compared to sectors like automotive, which sent emails infrequently. These findings provide valuable benchmarks for industry-specific marketing practices. Further research by Yang et al. [33] examined the impact of post-stay email marketing in the hotel industry, identifying how personalized, interactive emails increased customer retention and revisit intention. This aligns with the findings from [9], which highlighted the role of A/B testing, personalization, and frequency in determining email effectiveness across different industries.

Existing literature highlights the need for a balance between effective email marketing and safeguarding inbox privacy. While predictive models and personalized strategies can optimize engagement, they must be deployed ethically to avoid violating user privacy. Prior works emphasize that compliance with data protection laws, along with the adoption of transparent email marketing practices, is critical for maintaining consumer trust. Our study builds on this foundation by analyzing the email marketing and data-sharing practices of popular apps and services. Unlike prior works, which often did in-depth analysis of emails pertaining to a single company or industry type, our study instead focuses on a breadth of services and apps across various sectors.

3 Data Collection Framework

Our framework (Figure 1) is based on the assumption that if email addresses are leaked, shared, or sold by the services or apps we signed up for, those addresses would eventually start receiving spam or unsolicited emails from third-parties. This assumption allows us to passively detect unauthorized sharing of user data through the presence of unsolicited communications, which are typically linked to marketing activities that involve purchasing user data from third-party vendors. It is important to note, however, that this approach has its limitations. Specifically, we cannot detect instances where email addresses are leaked, shared or sold to third-parties if they do not subsequently send spam or other detectable communications.

To comprehensively evaluate the email sharing and marketing practices across different sectors, we selected a diverse set of widely used online services and mobile applications, as follows:

- **Top 100 Online Services:** We compiled a list of the most popular websites worldwide based on sources such as Cisco Umbrella¹ and Statista². Services that were consistently ranked across the two sources were prioritized for inclusion in the final list.
- **Top 50 Mobile Applications:** Similarly, a list of the top 50 mobile applications was compiled by analyzing app rankings from Google Play Store and Apple App Store. Notably, mobile apps that were already included in the top 100 online services (e.g., Facebook) were excluded to ensure non-redundant data. This ensures that our dataset captures a broader spectrum of both online services and apps across platforms.

Account Registration. After finalizing the list of online services and apps, we began the account registration process for each, utilizing standardized contact information. For each identified entity (online service and mobile app), a unique account was created following a standardized registration procedure. Unique email addresses were generated in the following format: **[randomly-generated-name]xxx@[our-domain].[tld]**, with specific ranges of xxx assigned as follows:

- 000-099 for the 100 online services.
- 100-149 for the 50 mobile apps.

This design of our framework facilitates clear differentiation between service types and ensures that each service and app has a unique identifier throughout the experiment. During registration, no additional subscriptions, email notifications, or promotional materials were selected. This was essential for maintaining consistency in data collection, allowing us to focus on analyzing unsolicited marketing emails from the service and app providers.

All received emails were collected in a single catch-all inbox, making it easier to manage, store, and scale the data collection. During the entire experiment, we did not click on any unsubscribe links in the received emails, which also helped maintain a consistent approach to data collection without influencing the outcome. At the end of the 361-day data collection period, all received emails were saved as EML files, containing the email headers, subject, body and any additional metadata, for our analyses.

¹<https://umbrella-static.s3-us-west-1.amazonaws.com/index.html>

²<https://www.statista.com/>

Email Content Classification. To classify the content of the emails, we rely on definitions provided by Thomas et al. [28], where emails were classified as either promotional, CRM, or alerts. These email types convey distinct messages with differing levels of persuasion. It is the appropriate combination of these different types of messages at the right time that drive click-through and eventual conversion, or conversely, abandonment, by the potential customer. To classify the emails that we collected, we used an in-context classification approach with a general purpose LLM (Llama3.1-8b-instruct [5]) with structured output. This approach has been shown to achieve human-level performance on sufficiently large LLMs with clear and concise rules [8]. The full procedure is available in Appendix A. There were 3,346 (69.0%) promotional emails, 1,393 (28.7%) CRM emails, 92 (1.89%) alert emails, and 16 (0.33%) that were unable to be classified as they were unparseable.

The use of LLMs as a rater for content classification tasks such as this is an area of active research [12, 27], and recently has been successfully applied to spam email classification tasks [25]. For this task, however, which relies on the LLM to perform comparably to a human rater, we enlisted 5 independent researchers from our institution to classify 50 randomly selected emails using the same instructions given to the LLM. Since there is no ground-truth, we instead evaluate whether the LLM-human inter-rater reliability (IRR) falls within the range of human-human IRR. The average Cohen's Kappa among human-human pairs is $\kappa = 0.406$ (moderate agreement), and the average LLM-human Kappa is $\kappa = 0.479$ (moderate agreement). In other words, the LLM achieves a level of agreement with humans that not only exceeds the human-human average, but also places it well within the variability observed among human raters themselves. This suggests that the LLM is performing comparably to human raters, demonstrating its viability as a reasonable evaluator in this classification task.

Our framework and following analyses address key aspects of data privacy laws, including the right to opt-out of marketing communications and restrictions on unauthorized data sharing practices, and can be extended to evaluate compliance across more diverse regulatory environments.

4 Findings

4.1 Descriptive Statistics

Over the 361-day collection period, we received 4,847 emails from 109 of the 150 services and apps, post registration (Table 2 in Appendix B). Of those emails received, 99.96% passed SPF checks and 81.64% passed DKIM checks. All emails that passed DKIM checks also passed SPF checks. Sender Policy Framework (SPF) verifies the sender's IP address to ensure it is authorized by the domain owner, while DomainKeys Identified Mail (DKIM) adds a cryptographic signature to emails, verifying that the message hasn't been altered in transit and is indeed from the claimed sender. Although ideally, both SPF and DKIM should pass, the successful verification of either check indicates that the email was authorized by the service with high confidence. A detailed manual review of the three emails that failed both SPF and DKIM revealed that these failures were likely due to DNS or server misconfigurations rather than malicious intent.

Post inspection of the emails received by their root domains and the email security checks via SPF and DKIM, we conclude that all emails received were from the original companies' domains with which we had made accounts and no UUSS emails were detected in this dataset.

However, further analysis reveals that the origin of the emails tell a more complex story of the lineage and custody of user data. Initial descriptive statistics indicate that there is not a well-defined parametric distribution that determines the volume of emails received. The mean was 43 emails per domain with a median of 5, a skewness of 3.55 showing the heavy right tail, and kurtosis of 12.92. The data's high skewness and kurtosis suggest a complex, highly skewed, and heavy-tailed nature that is not well captured by parametric distributions. Figure 2 shows that a few large senders make up the dominant amount of email volume received, with the cumulative distribution following the 80-20 Pareto principle, where the total volume of emails received are dominated by a few senders. Just the top 10 root domains make up 63.23% of the total volume.

The emails received do not just vary in the volume sent by companies, but also in the content of the messages being sent. Previous literature shows that strategic email marketing campaigns differ in both *what* the sender says (the content) and *when* they say it (temporal dimension) [29]. Additionally, previous work has shown that not all companies engage equally in strategic email marketing [1], with some being more sophisticated actors than others. Therefore, the following subsections of the analysis are structured as such: (1) We explore a unique contribution of our dataset to the literature on this topic, which is the sender's IP address and associated information about that sender through its Autonomous System Number (ASN). (2) A temporal analysis to show at what times and with what frequency emails are sent. (3) Clustering of companies based on their emails content and temporal patterns to

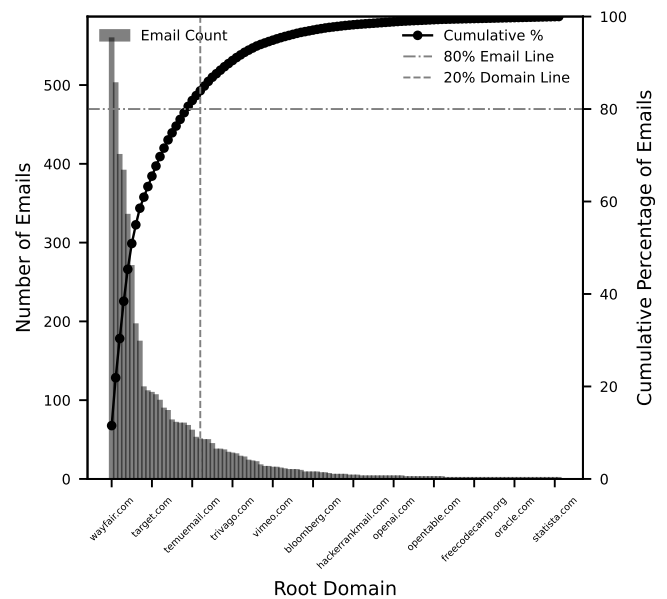


Figure 2: Pareto diagram of total email volume.

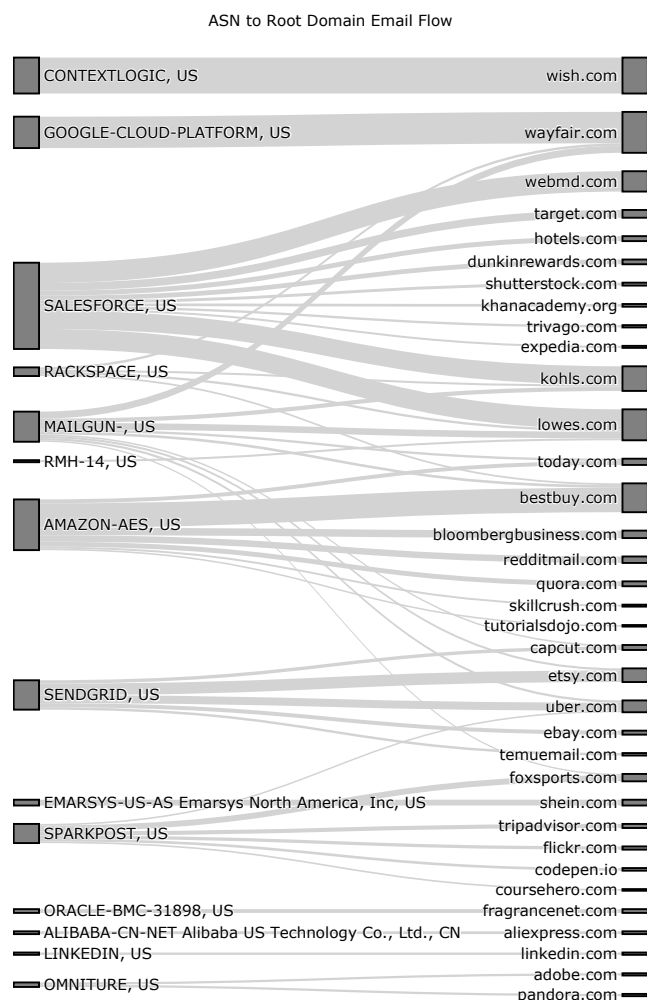


Figure 3: Sankey diagram of the top 50 flows.

determine where companies fall on their level of sophistication or general tactics.

4.2 Senders by Autonomous System

We captured the IP address of the sender of each email, with that we then ran a lookup of the registered autonomous system for each of those IP addresses and were able to tie it back to the company that sent it. Because of this, we are able to identify that **most of the emails received come from authorized third-parties** that specializes in business-to-business marketing and CRM services like *Salesforce* and *Mailgun*. Like the overall email volume, the email volume by ASN also follows the Pareto principle, where a few large providers are responsible for the sending of the large majority of emails. Specifically, the top twenty percent, 8 ASNs, produced 89.35% of the total email volume. Figure 3 shows the evidence that most of the email marketing flow originates from a few large, third-party providers.

However, understanding which companies use which email marketing service, or whether they use a third-party service at all, is not straightforward. Some companies send emails originating from their own registered ASN such as *wish.com* (Wish is the “doing-business-as” name for ContextLogic), Alibaba Group, LinkedIn, and Adobe (Adobe acquired Omniture as part of their web analytics platform). Others employ specialized third-parties, chiefly *Salesforce*, *Mailgun*, *Sendgrid*, and *Sparkpost*. A third group had emails originate from cloud service providers like AWS, GCP, and *Rackspace*. For those that originate from cloud service providers, using only the IP addresses we cannot ascertain what entity is managing the email campaigns at the application layer.

Sender Reputation. While the domain-level analysis confirms authenticity and shows no evidence of unauthorized or unknown senders, the infrastructure-level analysis reveals a potentially more ambiguous trust environment. To deepen our understanding of the quality and trustworthiness of these sending infrastructures, we conducted an IP reputation analysis. We cross-referenced each unique email origination IP against a third-party IP reputation database [2], which aggregates historical spam reports and other abuse metrics. In total, there were 612 spam reports associated with 84 unique IP addresses belonging to 11 ASNs, corresponding to 29 unique companies that were in our study. This allows us to assess the prevalence of reported spam activity associated with the IP ranges, from which we received emails, used by various ASNs.

By coupling the ASN-level volume analysis with the historical IP reputation data, we found that even well-known brands may rely, potentially unknowingly, on email infrastructure that carries varying levels of reported spam activity. The treemap visualization in Figure 4 arranges ASNs hierarchically, with each company’s sending domain nested beneath the ASN that serviced it. The area of each box corresponds to the total number of spam reports for the associated IPs, while the grayscale intensity indicates the magnitude of these reports.

This hierarchical perspective reveals patterns and raises important questions for both research and policy. For instance, some large ASNs have a relatively high volume of reported spam across multiple well-known clients. In such cases, the individual brand may not be involved in illicit or spammy behaviors, yet they inherit the reputation challenges posed by their chosen infrastructure providers. Or, the community-writ-large will deem certain messages as spam and report them as abuse. These findings highlight that brand trust does not necessarily translate into a pristine underlying email infrastructure. Rather, the operational practices, vetting procedures, and network hygiene of intermediary email service providers matter a great deal.

Ultimately, these findings show that the email ecosystem is more complex and federated than it may appear on the surface. While the messages themselves may appear genuine and from the expected sender, the underlying infrastructure often involves multiple intermediaries and networks of varying reputations.

For end-users, this complexity raises concerns about data stewardship and privacy. When personal information and communication preferences flow through third-party service providers, some of which may have been associated with spam-related activity, users face heightened uncertainty about how their data is handled,

who can access it, and to what end. This underscores the need for greater transparency and more privacy-centric policies from both the companies authorizing these messages and the infrastructure providers enabling their delivery.

IP Hopping and Spam Reports. To investigate whether companies associated with higher spam reports are more likely to use a larger number of IPs, a behavior that could suggest “IP hopping” to evade reputation-based spam filters, we examined the correlation between the number of unique IPs per company and the total spam reports associated with those IPs.

Our analysis revealed a strong positive correlation between the number of IPs used by a company and their associated spam reports. Pearson’s correlation coefficient ($r = 0.71, p < .0001$) indicates a linear relationship, where companies using more IPs tend to have higher volumes of spam reports. Spearman’s rank correlation ($r_s = 0.55, p = .002$) further supports this finding, suggesting a monotonic relationship even in the presence of potential non-linearity or outliers.

This result highlights that companies linked to higher spam reports may be employing a strategy of cycling through multiple IP addresses to distribute the reputational impact across their infrastructure. Such IP hopping practices can dilute the effects of blacklistings or abuse reports on any single IP address [24], allowing these companies to maintain email delivery volumes despite the reputational damage incurred by individual IPs.

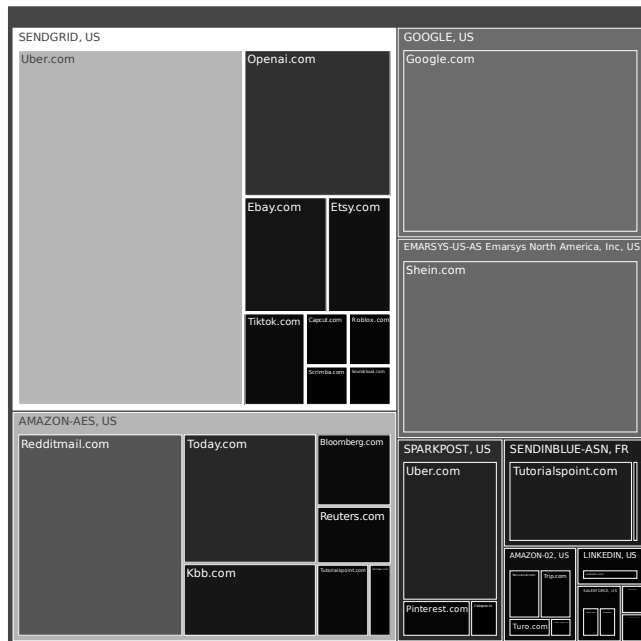


Figure 4: Hierarchical treemap of community reported spam on sender IPs, organized by ASN and associated company domains. Lighter shades indicate higher counts of spam reports, and vice versa.

4.3 Temporal Analysis

The temporal dimension of this analysis examines both the timing (time of day and day of week) and the frequency of emails sent. First, we treat the incoming messages as a time series, decomposing it to understand the general trend and identify hypothesized seasonality as shown in Figure 5. Since account creation dates differ across services and apps, we standardize the series by the number of days since the first email received, which coincides with the account creation date for each app. This standardization provides a uniform time scale for the analysis.

Applying a Fourier transform to the combined time series revealed dominant frequencies, identified at the 2σ level, corresponding to various periodic cycles: annual, semi-annual, quarterly, and weekly. The identified frequencies (f), magnitudes ($|X(f)|$), and periods (T) are as follows:

- Annual cycle: $f = 0.0028$, $|X(f)| = 834.38$, $T \approx 362$ days
- Semi-annual cycle: $f = 0.0055$, $|X(f)| = 479.22$, $T \approx 181$ days
- Quarterly cycle: $f = 0.0138$, $|X(f)| = 308.50$, $T \approx 72.4$ days
- Weekly cycle: $f = 0.1436$, $|X(f)| = 393.39$, $T \approx 6.96$ days

The most relevant frequency for our analysis is the weekly cycle (≈ 7 days), denoted by $f = 0.1436$ with a magnitude of $|X(f)| = 393.39$. This weekly periodicity corresponds closely with the expected operational and behavioral cycles observed in email-sending patterns, reflecting regular weekly activity. We use the 7-day period as the seasonality figure for the time series decomposition.

The decomposition highlights a linear, decreasing trend in the number of emails received over time. Additionally, it shows that the weekly seasonal component explains only 1.17% of the variance in the number of emails received, which is evidence of varied strategies being employed in when emails are sent.

The amount of seasonality in email sending behavior is mediated at the company level and will be explored in greater detail in following sections on the cross-sector and company-level analysis. Given that the observed weekly seasonality in sending patterns we look to specify the general time of day and day of week that email messages are sent. The polar heat maps shown in Figure 6 show that the most popular times messages are sent occur during the mid-week, Tuesday through Thursday, and emails are predominantly sent around midnight or during the afternoon work hours between 1300 and 1500 hours, specifically targeting when users are most active online.

4.4 Clustering

The next task is to determine which companies exhibit similar email sending patterns, and then map those patterns to levels of sophistication from prior literature. To accomplish this unsupervised learning task, we first perform a principal component analysis (PCA) on all the features created from the previous analysis. Then, we perform K-Means to assign each company to a cluster.

The analysis began by loading the relevant email and ASN data and selecting the temporal and behavioral features.

- **Hourly Email Frequency:** We computed the count of emails sent by each company per hour of the day (0-23), aggregated across all days.

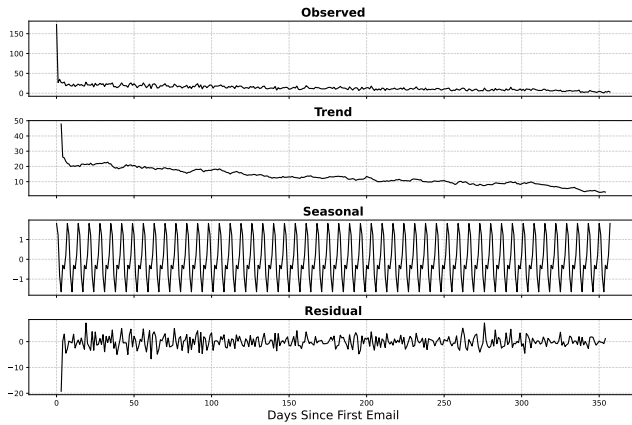


Figure 5: Additive time series decomposition of aggregate emails received.

- **Weekly Seasonality:** We analyzed the weekly email sending pattern by determining the frequency of emails sent on each day of the week (0 = Monday, 6 = Sunday).
- **Professional Marketing Indicator:** Each ASN description was checked against known marketing services (e.g., Salesforce, Mailgun, SendGrid, SparkPost) to create a binary feature indicating the presence of professional marketing practices.
- **Email Content Classification:** We captured the distribution of email content (e.g., promotional, CRM, alert) for each company, allowing us to profile their communication style.
- **Total Email Volume:** The overall number of emails sent by each company was computed to gauge their communication scale.

These features were combined into a comprehensive feature set, standardized using `StandardScaler`, and subsequently reduced to five principal components using PCA, retaining over 80% of the variance in the data.

K-Means clustering was employed to group companies based on their PCA-transformed features. To determine the optimal number of clusters, we evaluated cluster validity using the Silhouette Score, a metric that measures how similar a point is to its assigned cluster compared to other clusters. We tested a range of cluster numbers (2-10) and identified that separating companies into two clusters yielded the highest silhouette score (0.831). Using the PCA loadings, we interpret the characteristics of each cluster to better understand the distinct email marketing behaviors exhibited by companies within each group. The average scores of each cluster on the principal components highlight the dominant features that distinguish one cluster from another.

- **Cluster 0: Low-Activity Generalists**

Cluster 0 exhibits negative scores on Component 1 and slight negative or near-zero scores on other components. This cluster represents companies with generally low email activity, characterized by a lack of pronounced patterns in sending frequency, content, or professional marketing service usage.

The absence of significant influence from any principal component suggests that these companies do not heavily invest in sophisticated or high-volume email marketing strategies. Their approach may be minimalistic or highly targeted with low engagement, reflecting a more traditional or less dynamic email marketing approach.

- **Cluster 1: High-Volume Promotional Marketers**

Cluster 1 stands out with a highly positive score on Component 1 and a negative score on Component 2. This combination indicates companies that engage in extremely high-volume, promotional-focused email marketing but may not prioritize morning sending times. The negative loading on Component 2 suggests these companies might not strategically time their emails for the morning hours. These firms dominate in terms of sheer volume, indicating a strategy that prioritizes broad promotional outreach, likely at the expense of targeted or personalized engagement.

The clustering analysis, visually depicted in Figure 7, reveals distinct segments of companies based on their email marketing strategies, ranging from high-volume promotional tactics to CRM-focused approaches. The PCA loadings provide a clear view of the factors driving these segmentations. From this analysis, there are two clear factor loadings that drive this clustering:

- **Component 1** captures overall email activity and promotional focus. Companies with high scores on this component send a high volume of emails, particularly promotional content, consistently across weekdays. This suggests a strategy that emphasizes frequent promotional outreach.
- **Component 2** represents email sending times, particularly in the morning hours. High scores on this component indicate companies that strategically send emails during morning hours (7 AM to 10 AM), possibly to increase open rates and customer engagement.

We see from this clustering that the dominant discriminator between these companies and their email sending patterns is the volume of specifically promotional emails that they are sending.

4.5 Cross-Sector Differences

To explore potential differences in email sending behaviors across various sectors, we analyzed key metrics such as the total volume of emails sent, the types of emails (promotional, CRM, alert), and the timing of email sending (hourly and daily patterns). Since companies are moving to integrate their sales and marketing channels [30], and there are no uniformly accepted definitions that separate many of these companies [18], we rely on broad categorizations without making definitive claims as to their generalizability. This analysis is to further the argument, supported by our data, that not all companies send these unsolicited marketing emails in the same manner, and that certain sectors may be greater offenders than others.

Email Volume and Type by Sector. A chi-squared test of independence was conducted to examine the relationship between email type and sector. The test revealed a statistically significant association between email type and sector ($\chi^2(14) = 2138.858, p < 0.0001$), indicating that the distribution of email types varies by sector, with results shown in Table 1.

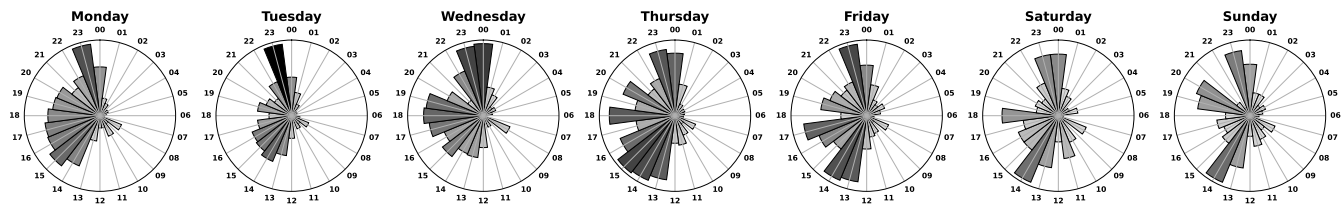


Figure 6: Email frequency by hour of day for each day of the week.

Table 1: Sector email volume and type averages.

Sector	Total	Promo	CRM	Alert	<i>n</i>
Omnichannel Retailers	139.56	133.22	5.89	0.44	9
E-tailer	123.23	104.69	17.31	0.38	13
Financial Services	53.50	0.50	40.00	13.00	2
Digital Services	23.69	7.89	15.17	0.60	35
Online Marketplaces	26.75	20.06	6.63	0.00	16
Communication Platforms	19.63	3.37	14.84	1.26	19
Online Entertainment	15.20	5.10	9.10	1.00	10
Brick & Mortar	19.00	14.60	4.00	0.40	5

The significant chi-squared test suggests that sectors differ in the types of emails they send:

- Retailers, both *Omnichannel Retailers* and online only *E-tailers*, predominantly send **promotional emails**. Examples of omnichannel companies include *Kohls* and *BestBuy*, while e-tailers include *Wish.com* and *Wayfair.com*.
- The *Financials* sector, while only 2 of the 21 companies actually sent emails, those two focus heavily on **CRM** (75.4%) and **alert emails** (24.3%) more than any other sector. This could reflect those companies using emails for non-marketing activities like account notifications.

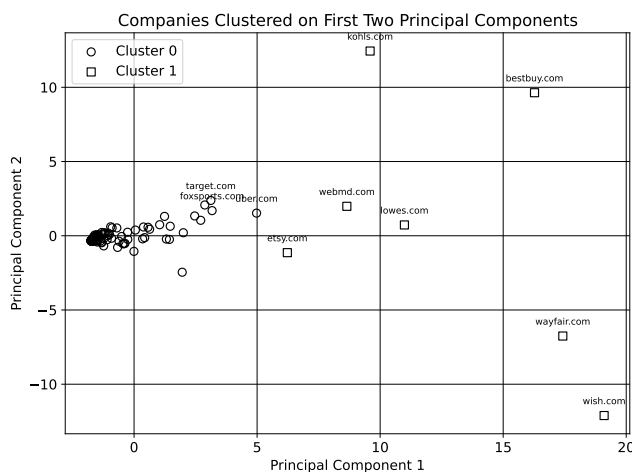


Figure 7: Cluster membership by top two principal components.

Variability Among Individual Companies. A one-way ANOVA was conducted to compare the average emails sent per company by sector. The results indicated statistically significant difference ($F(7, 101) = 3.5095, p = 0.002$), suggesting that the average email volume per company differs across sectors. Kruskal-Wallis H tests for both hourly and daily email sending patterns revealed no statistically significant differences between sectors. This indicates that sectors do not differ significantly in terms of when they send emails.

The significant differences in email volume and types between sectors implies that we may be able to direct future work towards those companies and sectors that are specifically sending promotional advertisements to consumers—retail and product-oriented businesses. While individual companies within the same sector may adopt diverse strategies in terms of email frequency and timing, there are some clear patterns that different industries utilize email communications to achieve their specific business objectives.

5 Discussion

5.1 New Insights

This study provides a comprehensive analysis of email marketing strategies employed by the most popular online websites and mobile applications. By collecting emails where each unique email address corresponds to a single company, we have ensured a precise lineage and isolation of source company to end user. This methodological approach, combined with a year-long email data collection period encompassing 150 apps and services, offers both breadth and depth to our dataset, distinguishing our work from prior studies. Our research contributes to the ongoing conversation about how companies utilize digital channels to market to and maintain relationships with their customers. The extensive dataset allows us to observe real-world practices over an extended period, providing insights into not only the frequency and content of communications but also the strategies underlying them.

The findings reveal no evidence of USS emails from unknown or unauthorized third-parties and a significant portion of consumer concerns regarding excessive promotional emails can be mitigated by opting out of marketing communications where possible. The data also shows that the bulk of SOS emails originate from a small subset of companies, supporting the Pareto principle where approximately 80% of the emails come from 20% of the senders. This concentration suggests that *aggressive email marketing is not a universal practice but is driven by specific companies, often within certain sectors like retail.*

ASN Mapping Shows Reliance on Authorized Third-Parties. A novel aspect of our study is the analysis of email provenance via the originating ASNs. By tracing the sender IP addresses to their corresponding ASNs, we uncover the reliance of companies on authorized third-party marketing and CRM services. Major intermediaries such as *Salesforce*, *Mailgun*, *SendGrid*, and *SparkPost* are identified as key players facilitating the bulk of email marketing traffic.

This approach sheds light on the ecosystem of email marketing infrastructure, highlighting how companies leverage specialized external platforms to execute their communication strategies. A complete picture, however, is elusive as many of the emails originate from public cloud providers such as *AWS* and *GCP*, so what third-party is operating its service hosted on these hyperscalers isn't able to easily be determined just from the IP address. It also raises questions about data privacy and the extent to which customer information is shared with third-party providers, a topic that warrants further investigation.

The “What” and “When” Matter. Our temporal analysis indicates that while the overall volume and timing of emails are relatively consistent across sectors, the content and purpose are tailored to industry-specific strategies. For instance, retailers predominantly send promotional emails, aligning with their goal of stimulating immediate purchases. In contrast, financial service companies and communication platforms focus on CRM and alert emails, reflecting the importance of customer engagement and timely information dissemination.

The identification of distinct clusters through principal component analysis (PCA) and K-means clustering reveals that companies adopt varied email marketing tactics ranging from high-volume promotional outreach to personalized CRM communications. This diversity underscores the need for nuanced strategies that consider both the timing and content of emails to effectively engage customers.

5.2 Legal and Ethical Considerations

The findings raise important legal and ethical considerations regarding compliance with privacy policies and marketing preferences. Despite regulations like the CCPA, CAN-SPAM and GDPR, consumers continue to report receiving unsolicited emails. Our study suggests that opting out during registration can reduce the volume of such emails, but also highlights that some companies may not fully adhere to opt-out preferences.

Aggressive marketing practices, especially those that aim to “wear down” consumers through persistent communications, can erode trust and infringe upon consumer rights. Prior research has demonstrated that excessive emailing can lead to customer fatigue and negative brand perception [11]. Companies must balance their marketing objectives with respect for consumer autonomy and privacy.

Policymakers and regulatory bodies may find these insights valuable in assessing the effectiveness of current regulations and identifying areas where enforcement or additional guidelines are necessary. Our study identifies that a vast majority of marketing emails are facilitated by third-party services such as *Salesforce*, *Mailgun*, and *SendGrid*, with 89.35% of emails being sent through

the top 8 ASNs. *This centralization indicates a potential focal point for improving regulatory oversight.* Policymakers should consider these patterns when evaluating regulatory frameworks, ensuring that opt-out mechanisms are straightforward, transparent, and honored in practice. Enhancing regulations to specifically reference and include the use of third-party marketing services and enforcing compliance can further improve consumer experiences and uphold consumer rights.

5.3 Limitations and Future Work

The richness of our dataset extends beyond metadata to include the full subject lines and body content of emails. This presents opportunities for future research into the specific language, persuasive techniques, and thematic elements employed in email marketing. To do this, future work can explore the content of emails in greater detail through Natural Language Understanding (NLU) techniques [6, 10]. Additionally, most of the emails in the dataset contain images that could be rendered and then analyzed [15, 34], yielding insights into specific persuasion techniques that are employed. Moreover, our dataset can support studies on the effectiveness of different email strategies by correlating communication patterns with customer engagement metrics, where available. Finally, investigating the role of cultural, regional, or demographic factors in email marketing practices also presents an avenue for further study. Future work could also investigate whether smaller companies, which may face less regulatory scrutiny compared to popular online services and apps, are more likely to share or leak users' email addresses without consent.

While our study analyzes the correlation between spam reports and IP usage, identifying potential IP hopping behaviors, it does not yet evaluate whether such practices align with the companies' stated privacy policies. Future work can focus on bridging this gap by cross-referencing email-sending behaviors with documented privacy commitments. This analysis would help determine if companies engaging in practices like IP hopping are acting inconsistently with their stated policies, providing deeper insights into accountability and compliance within the email ecosystem.

A key component of our framework is its low cost of ongoing maintenance and scalability. Once deployed, it can monitor any number of online services and apps passively. A key advantage of this is that we can detect any future changes in the inbound email messages. So for example, *if a company in the monitored set has a breach that leaks customer email addresses and then those addresses are used for spam campaigns or sold to unauthorized third-parties, our system would be one of the first authoritative detection sources of such a breach.*

6 Conclusion

This study offers a detailed examination of email inbox privacy and marketing practices among leading online services and apps, highlighting key patterns in communication strategies and the use of third-party services. Our findings contribute to a better understanding of how companies balance marketing objectives with consumer preferences and the implications for trust and engagement in the digital marketplace. By shedding light on these practices, we aim to inform both industry and policy discussions, promoting approaches

that respect consumer rights while enabling effective email marketing communications.

References

- [1] Alan S Abrahams, Tarun Chaudhary, and Jason K Deane. 2010. A multi-industry, longitudinal analysis of the email marketing habits of the largest United States franchise chains. *Journal of Direct, Data and Digital Marketing Practice* 11 (2010), 187–197.
- [2] AbuseIPDB. [n. d.]. AbuseIPDB - IP address abuse reports and blacklist checking. <https://www.abuseipdb.com/>. Accessed: 2024-12-09.
- [3] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.
- [4] Leonie Eline Agema. 2015. *Death by e-mail overload*. B.S. thesis. University of Twente.
- [5] AI@Meta. 2024. Llama 3.1 Model Card. https://github.com/meta-llama/llama-models/blob/main/models/llama3_1/MODEL_CARD.md.
- [6] James Allen. 1988. *Natural language understanding*. Benjamin-Cummings Publishing Co., Inc.
- [7] Carolina Araújo, Christophe Soares, Ivo Pereira, Duarte Coelho, Miguel Ângelo Rebelo, and Ana Madureira. 2022. A novel approach for send time prediction on email marketing. *Applied Sciences* 12, 16 (2022), 8310.
- [8] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 1877–1901.
- [9] Camelia Budac. 2016. Theoretical approaches on successful email marketing campaigns. *Ovidius University Annals: Economic Sciences Series* 16, 2 (2016), 306–311.
- [10] Massimo Canonico, Luigi De Russis, et al. 2018. A comparison and critique of natural language understanding tools. In *Cloud Computing 2018*. 110–115.
- [11] Junyu Cao and Wei Sun. 2019. Dynamic learning of sequential choice bandit problem under marketing fatigue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3264–3271.
- [12] Youngjin Chae and Thomas Davidson. 2023. Large language models for text classification: From zero-shot learning to fine-tuning. *Open Science Foundation* (2023).
- [13] Nora A Draper and Joseph Turow. 2019. The corporate cultivation of digital resignation. *New media & society* 21, 8 (2019), 1824–1839.
- [14] Steven Englehardt, Jeffrey Han, and Arvind Narayanan. 2018. I never signed up for this! Privacy implications of email tracking. *Proceedings on Privacy Enhancing Technologies* (2018).
- [15] Hao Fei, Shengqiong Wu, Hanwang Zhang, Tat-Seng Chua, and YAN Shuicheng. 2024. Vitron: A unified pixel-level vision llm for understanding, generating, segmenting, editing. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- [16] Marian Friestad and Peter Wright. 1994. The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research* 21, 1 (1994), 1–31.
- [17] Chris Jay Hoofnagle, Jennifer M Urban, and Su Li. 2012. Privacy and modern advertising: Most us Internet users want 'do not track' to stop collection of data about their online activities. In *Amsterdam Privacy Conference*.
- [18] Christiane Lehrer and Manuel Trenz. 2022. Omnichannel business. *Electronic Markets* 32, 2 (2022), 687–699.
- [19] New York Times. 2017. All 3 Billion Yahoo Accounts Were Affected by 2013 Attack. <https://www.nytimes.com/2017/10/03/technology/yahoo-hack-3-billion-users.html>.
- [20] New York Times. 2017. Equifax Says Cyberattack May Have Affected 143 Million in the U.S. <https://www.nytimes.com/2017/09/07/business/equifax-cyberattack.html>.
- [21] Sami Nuutinen. 2015. How to Increase the Efficiency and Reach of Email Marketing: Case Onninen Finland. (2015).
- [22] Stuart L Pardau. 2018. The california consumer privacy act: Towards a european-style privacy regime in the united states. *Journal of Technology Law & Policy* 23, 1 (2018), 2.
- [23] Robert K Pllice, Oleg V Pavlov, and Nigel P Melville. 2008. Spam and beyond: An information-economic analysis of unwanted commercial messages. *Journal of Organizational Computing and Electronic Commerce* 18, 4 (2008), 278–306.
- [24] Anirudh Ramachandran, Nick Feamster, and Santosh Vempala. 2007. Filtering spam with behavioral blacklisting. In *Proceedings of the 14th ACM conference on computer and communications security*. 342–351.
- [25] Sergio Rojas-Galeano. 2024. Zero-Shot Spam Email Classification Using Pre-trained Large Language Models. In *Workshop on Engineering Applications*. Springer, 3–18.
- [26] Justin Sherman. 2021. Data brokers and sensitive data on US individuals. *Duke University Sanford Cyber Policy Program* 9 (2021).
- [27] Hanneke Theelen, Joyce Vreuls, and Jim Rutten. 2024. Doing Research with Help from ChatGPT: Promising Examples for Coding and Inter-Rater Reliability. *International Journal of Technology in Education* 7, 1 (2024), 1–18.
- [28] Jacquelyn S Thomas, Chaoqun Chen, and Dawn Iacobucci. 2022. Email marketing as a tool for strategic persuasion. *Journal of Interactive Marketing* 57, 3 (2022), 377–392.
- [29] Jacquelyn S Thomas, Chaoqun Chen, and Dawn Iacobucci. 2022. Email marketing as a tool for strategic persuasion. *Journal of Interactive Marketing* 57, 3 (2022), 377–392.
- [30] Manuel Trenz, Daniel Veit, and Chee-Wee Tan. 2020. Disentangling the impact of omnichannel integration on consumer behavior in integrated sales channels. (2020).
- [31] Milos Vacek. 2014. Email overload: Causes, consequences and the future. *International Journal of Computer Theory and Engineering* 6, 2 (2014), 170.
- [32] Paul Voigt and Axel von dem Bussche. 2017. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer.
- [33] Kiseol Yang, Ji Hye Min, and Krishna Garza-Baker. 2019. Post-stay email marketing implications for the hotel industry: Role of email features, attitude, revisit intention and leisure involvement level. *Journal of Vacation Marketing* 25, 4 (2019), 405–417.
- [34] Yanzhe Zhang, Ruiyi Zhang, Jiuxiang Gu, Yufan Zhou, Nedom Lipka, Diyi Yang, and Tong Sun. 2023. Llavav: Enhanced visual instruction tuning for text-rich image understanding. *arXiv preprint arXiv:2306.17107* (2023).

A LLM Classification Workflow

The following code demonstrates the LLM classification workflow used to categorize emails as promotional, CRM, or alert.

```

1 tagging_prompt = ChatPromptTemplate.from_template(
2     """
3     You are an email classifier system. Your task is to classify the
4     following email message as either promotional, CRM, or alert.
5     Respond only with the JSON formatted classification based on
6     the provided schema below, containing only the
7     classification results and not the schema itself. Do NOT
8     output the schema or any definitions-only provide the
9     classification response.
10
11     Here is the JSON schema of the response you must strictly follow:
12
13     {schema}
14
15     Here are the classification definitions (these are for your
16     reference only, do not include them in the output):
17
18     promotional: Emails offering price discounts intended to induce
19     short-term actions such as purchasing. These are overt
20     selling attempts with a call to action.
21
22     CRM: Emails that engage, inform, or enhance the company's
23     relationship with the customer. These emails are not overtly
24     sales-focused but aim for longer-term goals, like brand
25     building or engagement.
26
27     alert: Emails containing various notifications, updates, or alerts
28     based on consumer preferences. These are informational and
29     do not directly attempt to sell.
30
31     Classify the following email strictly based on these definitions:
32
33     Email:
34     {input}
35     """
36 )
37
38 # Create the structured output format class
39 class Classification(BaseModel):
40     sentiment: str = Field(..., enum=["promotional", "CRM", "alert",
41                                     ""])
42     confidence: int = Field(
43         ...,
44         description="describes how confident the classification is
45         , the higher the number the more confident in the
46         classification",
47         enum=[1, 2, 3, 4, 5],
48     )
49     rationale: str = Field(description="The primary driver of the
50                             classification based on the definitions")

```

Listing 1: LLM classifier prompt and code.

B List of Received Email Domains

Table 2 shows the details and distribution of emails received from 109 different root domains.

Table 2: Distribution of emails received from 109 different root domains.

Root Domain	Sector	Cluster	Total Emails	Promotional	CRM	Alert
adobe.com	Digital Services	0	37	24	13	0
airtable.com	Digital Services	0	2	0	2	0
aliexpress.com	E-tailer	0	37	33	3	1
apartmentfinder.com	Online Marketplace	0	1	0	1	0
apartments.com	Online Marketplace	0	8	1	7	0
autozone.com	Brick and Mortar	0	13	7	5	1
bandcamp.com	Digital Services	0	4	1	3	0
bestbuy.com	Omnichannel	1	391	376	12	3
bloomberg.com	Financials	0	8	0	5	3
bloombergbusiness.com	Financials	0	99	1	75	23
booking.com	Online Marketplace	0	2	0	2	0
canva.com	Digital Services	0	3	1	2	0
capcut.com	Communication Platforms	0	71	19	49	1
cars.com	E-tailer	0	15	6	9	0
cnet.com	Digital Services	0	1	0	1	0
codepen.io	Digital Services	0	49	1	47	1
codewars.com	Digital Services	0	3	0	3	0
coursehero.com	Digital Services	0	22	7	15	0
discogs.com	Communication Platforms	0	8	0	7	1
discord.com	Communication Platforms	0	3	0	3	0
doordash.com	Online Marketplace	0	3	3	0	0
dunkinrewards.com	Brick and Mortar	0	70	61	8	1
ebay.com	E-tailer	0	70	57	13	0
etsy.com	E-tailer	1	196	115	81	0
expedia.com	Digital Services	0	28	24	3	1
flickr.com	Communication Platforms	0	44	13	31	0
foodnetwork.com	Online Entertainment	0	1	0	1	0
foxsports.com	Online Entertainment	0	111	39	63	9
fragrancenet.com	E-tailer	0	50	41	9	0
freecodecamp.org	Digital Services	0	1	0	1	0
geeksforgeeks.org	Digital Services	0	12	7	5	0
github.com	Digital Services	0	2	0	1	1
google.com	Communication Platforms	0	3	0	2	1
hackerrankmail.com	Digital Services	0	4	0	4	0
homedepot.com	Omnichannel	0	1	0	1	0
hotels.com	Online Marketplace	0	89	72	17	0
imdb.com	Online Entertainment	0	2	0	2	0
instagram.com	Communication Platforms	0	14	0	13	1
kayak.com	Digital Services	0	3	0	3	0
khanacademy.org	Digital Services	0	49	21	28	0
kohls.com	Omnichannel	1	335	330	5	0
leetcode.com	Digital Services	0	8	5	3	0
linkedin.com	Communication Platforms	0	31	1	30	0
lowes.com	Omnichannel	1	411	404	6	1
lyftmail.com	Online Marketplace	0	7	1	6	0
nationalgeographic.com	Online Entertainment	0	1	0	1	0
newegg.com	E-tailer	0	1	1	0	0
nike.com	Omnichannel	0	1	0	1	0
openai.com	Digital Services	0	3	0	3	0
opentable.com	Digital Services	0	2	0	2	0
oracle-mail.com	Digital Services	0	5	1	4	0
oracle.com	Digital Services	0	1	0	1	0
orbitz.com	Online Marketplace	0	15	14	1	0

Continued on next page

Root Domain	Sector	Cluster	Total Emails	Promotional	CRM	Alert
pandora.com	Communication Platforms	0	36	17	19	0
pinterest.com	Communication Platforms	0	33	1	32	0
planetfitness.com	Brick and Mortar	0	1	0	1	0
plex.tv	Online Entertainment	0	4	0	4	0
pons.com	Digital Services	0	1	0	1	0
pornhub.com	Online Entertainment	0	3	0	3	0
quizlet.com	Digital Services	0	5	0	5	0
quora.com	Digital Services	0	67	0	63	3
realtor.com	Online Marketplace	0	7	2	5	0
reddit.com	Communication Platforms	0	1	0	1	0
redditmail.com	Communication Platforms	0	86	0	65	20
reuters.com	Communication Platforms	0	2	0	2	0
ring.com	Omnichannel	0	5	1	4	0
roblox.com	Online Entertainment	0	1	0	1	0
roku.com	Communication Platforms	0	3	0	3	0
scrimba.com	Digital Services	0	17	12	4	1
shazam.com	Communication Platforms	0	1	0	1	0
shein.com	E-tailer	0	106	93	12	1
sheinnotice.com	E-tailer	0	1	0	1	0
shopify.com	Online Marketplace	0	1	0	1	0
shutterstock.com	Digital Services	0	52	20	31	1
skillcrush.com	Online Entertainment	0	27	12	15	0
snapchat.com	Communication Platforms	0	5	0	5	0
soundcloud.com	Communication Platforms	0	11	3	8	0
southwest.com	Brick and Mortar	0	1	0	1	0
springernature.com	Omnichannel	0	2	0	2	0
statista.com	Digital Services	0	1	0	1	0
steampowered.com	E-tailer	0	1	0	1	0
target.com	Omnichannel	0	109	87	22	0
temu.com	E-tailer	0	3	0	3	0
temuemail.com	E-tailer	0	61	40	20	1
thumbtack.com	Online Marketplace	0	11	1	10	0
tiktok.com	Communication Platforms	0	1	0	1	0
today.com	Digital Services	0	116	112	3	1
topcoder.com	Digital Services	0	1	0	1	0
trafone.com	Brick and Mortar	0	10	5	5	0
trip.com	Online Marketplace	0	2	0	2	0
tripadvisor.com	Online Marketplace	0	74	32	42	0
trivago.com	Online Marketplace	0	32	26	6	0
turo.com	Online Marketplace	0	1	0	1	0
tutorialsdojo.com	Digital Services	0	21	15	5	1
tutorialspoint.com	Digital Services	0	23	16	7	0
twitch.tv	Communication Platforms	0	6	0	6	0
uber.com	Online Marketplace	0	174	169	4	0
udacity.com	Digital Services	0	1	0	1	0
usnews.com	Digital Services	0	2	0	2	0
vimeo.com	Communication Platforms	0	14	10	4	0
walmart.com	Omnichannel	0	1	1	0	0
wayfair.com	E-tailer	1	559	520	27	2
waze.com	Digital Services	0	2	0	2	0
webmd.com	Digital Services	1	270	9	250	11
wish.com	E-tailer	1	502	455	46	0
xhamster.com	Online Entertainment	0	1	0	1	0
xvideos.com	Online Entertainment	0	1	0	0	1
yelp.com	Digital Services	0	11	0	11	0
zillow.com	Online Marketplace	0	1	0	1	0