Gmail's 2026 Inbox Decision for High-Volume Cold Email: A Three-Layer Authentication, Reputation, and Engagement Model

Francis Davison

Founder, SpamCipher · https://spamcipher.com/insights/how-gmail-decides-inbox-vs-spam

Preprint · SpamCipher Insights · July 2026 · Primary-source desk study

ABSTRACT

This paper examined how Gmail classifies inbound commercial email as inbox or spam, with a focus on legitimate high-volume cold email. Practitioner understanding of this decision relies largely on claims with no documented basis, and the academic literature has studied mass email as abuse; legitimate outbound deliverability has not been its subject. The study was a desk-based synthesis of primary sources: Google's documentation and engineering publications, the RFCs that define email authentication, and peer-reviewed measurement research. The evidence describes a three-layer system. A mandatory compliance floor requires SPF, DKIM, and DMARC, one-click unsubscribe, and a user-reported spam rate below 0.30%. A reputation layer grades each sending domain and IP into four published bands whose definitions Google ties to delivery outcomes. A machine-learning classifier evaluates what Google describes as "thousands of potential signals," with recipient engagement described as central, and issues a per-recipient verdict. The one end-to-end placement measurement in the literature concerns forged senders: moving an impersonated domain from no authentication to a strict DMARC policy reduced the aggregate inbox rate of forgeries from 60.5% to 28.4%, and from 93% to 0% at Gmail. These results bound the anti-impersonation value of authentication and bear on a legitimate sender's own placement only by extension. The paper contributes an integrated, cited, signal-level account of Gmail's inbox decision for legitimate high-volume cold email. The synthesis is associational; no causal claims are made.

Keywords: cold email, email deliverability, Gmail spam filtering, sender reputation, SPF DKIM DMARC, inbox placement, bulk sender requirements, machine-learning classification

1. Introduction

Inbox placement determines the outcome of high-volume cold email. A message routed to the spam folder is unlikely to be read or answered, and the failure is difficult for the sender to observe, because sending tools report such a message as "delivered." The routing decision is made by Gmail's mail-classification system, which Google documents only in part and never as a single mechanism.

Prior work on this decision falls into two groups, and neither describes it adequately for legitimate senders. Practitioner materials circulate content-level rules, among them spam-trigger-word lists, image-to-text ratios, and link-count caps; as Section 6 shows, most of these have no basis in any Google source. The academic literature has treated mass email through the lens of spam and spoofing abuse: the economics of spam campaigns [1], the end-to-end spam value chain [2], measured spoofing attacks [3], provider-based email security [4], and the global adoption of email authentication [5]. That literature, by its own framing, does not study legitimate outbound deliverability. To our knowledge, no integrated, primary-source account of how Gmail decides inbox versus spam for legitimate high-volume cold email has been published.

Research questions. The primary question is descriptive: what signals does Gmail's 2026 mail-classification system use to route a message to the inbox versus spam, and what roles do sender reputation, authentication, recipient engagement, and content play? A secondary question is relational: how is each signal associated with measured inbox placement in the published literature? Because Gmail's classifier cannot be manipulated externally, the study is an exploratory, associational synthesis; findings are labeled as descriptive or associational, and claims that could not be sourced were excluded.

This paper addresses these questions through a structured desk study that (i) catalogues, from primary sources, the signals Gmail evaluates; (ii) reconstructs how these signals combine into the routing decision; (iii) reviews the peer-reviewed measurement evidence on the association between authentication and placement; and (iv) identifies which common practitioner claims have no support in Google's documentation. Every claim is descriptive or associational; where Google does not disclose a mechanism, the gap is stated (Sections 3 and 8). The contribution is an integrated, cited account of the signals in Gmail's inbox decision for legitimate high-volume cold email, together with an assessment of which practitioner claims lack documentary support.

2. Background

Google's stated model. Google describes Gmail's spam defenses as machine learning "powered by user feedback," and reports that these defenses "help block more than 99.9 percent of spam, phishing, and malware from reaching Gmail inboxes" [6, 7]. Google states the filters "look at a variety of signals, including characteristics of the IP address, domains/subdomains, whether bulk senders are authenticated, and user input" [7]. IP-level reputation of this kind has an independent research lineage in the detection of spam-sending infrastructure [8]. Since February 2024, Google has also published hard requirements for senders exceeding 5,000 messages per day to personal Gmail accounts [9], following the announcement it made in late 2023 [10]. This material is distributed across a dozen help pages and blog posts; no single Google document describes the complete mechanism.

Authentication protocols. Three protocols define email authentication. SPF (RFC 7208) allows a domain to publish, in DNS, the hosts authorized to send mail on its behalf [15]. DKIM (RFC 6376) attaches a cryptographic signature that receivers verify against a public key in the signing domain's DNS [16]. DMARC (RFC 7489) allows a domain to publish a policy (`none`, `quarantine`, or `reject`) for mail that fails authentication checks aligned with the From-header domain, together with a reporting mechanism [17]. RFC 8058 specifies one-click unsubscribe signaling for list email [18].

The measurement literature. The email-security research community has measured the infrastructure surrounding this decision: the global adoption of transport and authentication security [5], the effectiveness of provider-side enforcement [4], the deployment and misconfiguration of DKIM [11] and of DMARC reporting [12], and, in one study that measures placement directly, the inbox rates of spoofed mail as a function of authentication [3]. This work supplies the empirical evidence reviewed in Section 7. Its subject is spoofing and abuse; it does not measure legitimate cold-email deliverability, and it does not measure sender reputation, which is Gmail-internal and unpublished [3].

3. Methodology

This paper is an analytical desk study: an observational synthesis of public documents and published measurements. No experiment was performed, and the design and its limits follow from that.

Search and inclusion strategy. Primary sources were located through Google's own help and engineering-blog properties (`support.google.com`, `workspace.google.com`, `security.googleblog.com`, `blog.google`) and the IETF RFC series. Peer-reviewed sources were identified by searching dblp, the ACM Digital Library, IEEE Xplore, and the USENIX proceedings for the venues where email-security measurement is published (USENIX Security, IEEE S&P, ACM CCS, ACM IMC, NDSS, RAID, PAM), using the query terms *email authentication*, *SPF*, *DKIM*, *DMARC*, *spam filtering*, *email spoofing*, *inbox placement*, and *email deliverability*, over a 2008 to 2026 window. Inclusion criterion: a source was admitted if it is (a) a Google primary document or defining RFC bearing directly on a signal in scope, or (b) a peer-reviewed measurement study reporting SPF/DKIM/DMARC adoption, enforcement, or message placement. Exclusion: vendor blog statistics and practitioner claims were excluded as evidence and admitted only in Section 6 as labeled examples of unsupported claims. This is a purposive synthesis of the placement-relevant literature, not an exhaustive systematic review; the consequences of that boundary are recorded in Section 9.

Sources and evidence grading. Five source tiers were admitted, in descending weight: (1) Google's own primary documentation and the exact text of its rules; (2) peer-reviewed measurement research; (3) internet standards (RFCs); (4) reputable industry-consensus documents (e.g., M3AAWG); and (5) practitioner claims, admitted only to be examined and labeled. Every non-obvious claim carries a citation. Statements that are our inference are marked as inferences.

Operational definitions. "Inbox placement" denotes delivery to a Gmail inbox tab (Primary, Promotions, Updates, Social, or Forums), as distinct from the Spam folder; the tab distinction is treated in Section 5.3. "Spam rate" follows Google's own definition: "the percent of your messages that are delivered to engaged recipient's Inbox and then marked as spam by the recipient" [13]. The denominator of this definition is deliveries to engaged recipients' inboxes, not all sent messages.

Variables. In the associational frame, the independent variables are the sender's authentication policy, reputation band, recipient-engagement signals, and content signals; the dependent variable is inbox-versus-spam placement. Principal confounds are time of send, message content, sending IP, domain age, and mid-window provider-rule changes; in the one cited placement study these are handled by a crossed design over content types and sending IPs, and no others are controlled because no experiment of our own was run [3].

Reliability and error. As a single-author synthesis, source selection and extraction were performed by one coder, so inter-coder reliability could not be computed; this is a bias source, recorded in Section 9, and is distinct from random-sampling error. The aggregate figures cited carry systematic constraints, most importantly that the one placement study measures spoofing, not legitimate sending, and these constraints are separate from random sampling noise and are not resolved by interval estimates.

Ethics. All sources are public; the synthesis involved no active probing of third-party systems and no personal data.

Threats to validity. *Provider opacity (construct and internal validity):* Google publishes the approach of its classifier but not its features, weights, or thresholds [6, 7]; any claim about how signals combine internally is an inference and is hedged, and no internal-validity (causal) claim is made because there is no manipulation. *Statistical-conclusion validity:* no original inferential test is reported; the one inferential result cited (Section 7) is the source study's own, and its stated statistic and limits are carried rather than re-derived. *Recency:* provider rules change, so every rule is snapshot-dated to 2024 through 2026. *Construct validity of third-party numbers:* vendor "deliverability" statistics use

heterogeneous, undisclosed methods, and are excluded or labeled. *External validity:* findings are specific to Gmail and do not transfer to Outlook or Yahoo without evidence.

4. The authentication and compliance floor

The first layer of Gmail's decision is a set of pass/fail requirements. Since 1 February 2024, a sender exceeding 5,000 messages per day to personal Gmail accounts, counted per primary domain, a status that "doesn't have an expiration date" [14], must satisfy all of the following [9, 14]:

- Authenticate with SPF (RFC 7208) and DKIM (RFC 6376), and publish a DMARC record (RFC 7489); the DMARC policy "can be set to none" to comply [9, 15, 16, 17].
- Align the From-header domain with the SPF or DKIM domain [9].
- Use a DKIM key of 1024 bits or longer [9].
- Support one-click unsubscribe per RFC 8058 [18] in marketing and promotional mail, with the `List-Unsubscribe-Post: List-Unsubscribe=One-Click` header, and process unsubscribes within two days [9, 14].
- Keep the user-reported spam rate below 0.30%; Google separately advises keeping it below 0.10% and to "avoid ever reaching a spam rate of 0.30% or higher" [9].
- Maintain valid forward and reverse (PTR) DNS, transmit over TLS, and format messages per RFC 5322 [9, 19].

Yahoo published a parallel set of requirements effective the same month [20]. These requirements codify a long-standing industry consensus on sender authentication [24]. Two details of the requirements are commonly misstated. First, the requirements cite ARC (RFC 8617) nowhere. Industry commentary often describes ARC as required for forwarders, but ARC does not appear in Google's sender guidelines, Google's FAQ, or Google's forwarding best-practices page, which instead warns against breaking DKIM signatures [9, 14, 21, 22]; the interoperability problem ARC addresses is documented separately in RFC 7960 [23]. Gmail's published requirements therefore do not include ARC. Second, enforcement is graduated: Google describes it as "gradual and progressive," and as of November 2025 states it is "ramping up its enforcement on non-compliant traffic," with messages facing "temporary and permanent rejections" [14].

Meeting these requirements establishes eligibility for classification; failing them results in throttled or rejected mail [14]. The routing decision itself is made in the layers described in Section 5.

5. The reputation and classification layers

5.1 Sender reputation, graded in four bands

Google grades each sending domain and IP into one of four reputation bands and reports them in Postmaster Tools. Table 1 gives Google's definitions of the bands and the delivery outcome each definition states [13].

Table 1. Gmail sender-reputation bands (Google Postmaster Tools), verbatim definitions and the outcome each states.

Band	Google's definition (verbatim, with elisions marked)	Stated outcome
High	"History of very low spam rates ... Email from this domain or address is rarely marked as spam by Gmail."	Rarely marked spam
Medium	"History of sending legitimate email, but occasionally sends spam ... a fair deliverability rate, except when there's a notable increase in spam."	Fair, volatile on spikes
Low	"History of sending a significant volume of spam regularly ... likely to be marked as spam."	Likely spam
Bad	"History of sending a high volume of spam regularly ... almost always marked as spam or rejected by the receiving server."	Almost always spam or rejected

The band definitions are Google's own statement that sender reputation maps onto the inbox, spam, or reject outcome. Two caveats apply. The middle band was historically labeled "Fair" and is now "Medium"; the current text is the one cited here [13]. More substantively, Google nowhere states that the four reputation bands are an input feature to the machine-learning classifier. The band definitions describe delivery outcomes, and the classifier is described separately as using "IP address, domains/subdomains, whether bulk senders are authenticated, and user input" [7]. The overlap in the named signals supports an inference that reputation feeds the classifier, and that statement is presented here as an inference, not as a Google quote.

5.2 The machine-learning classifier

Above the floor and alongside reputation sits the classifier that makes the routing call. Google states that "every email has thousands of potential signals," and that "ML allows us to look at all of these signals together to make a determination" [6], a template-and-signal approach with a research lineage in content-based spam classification [26]. Google reports that the system has used "an artificial neural network" since 2015 [25] and TensorFlow-based models since 2019, the latter "blocking around 100 million additional spam messages every day" [6]. The most recent public advance, RETVec (NeurIPS 2023), is a resilient text vectorizer that, deployed in Gmail, "improve[d] the spam detection rate over the baseline by 38% and reduce[d] the false positive rate by 19.4%," which Google called "one of the largest defense upgrades in recent years" [27, 28].

Two properties of this classifier are relevant to a cold-email sender. First, it is designed against adversarial input: RETVec is explicitly built to resist "homoglyphs, invisible characters, and keyword stuffing" [28], which bears directly on the content claims examined in Section 6. Second, it is personalized: Google states the system "personalize[s] our spam protections to each user," adding that "what one person considers spam another person might consider an important message" [6]. The same message can therefore receive different verdicts for different recipients.

Google does not publish the classifier's feature set, weights, or thresholds, and RETVec is only the vectorizer front-end, not the deployed spam model [27]. The 38% and 19.4% figures are Google's self-reported internal results and have not been independently reproduced [28]. They are reported here as such.

5.3 User feedback and inbox categories

Google's material describes the recipient's own actions as central to filtering. Google states that user feedback "is key to this filtering process, and our filters learn from user actions" [7]. In the separate

context of tab sorting, Google states that "the most important" signal "is your direct input" [29]; that statement concerns the sorting subsystem and is not extended here to the spam verdict. For the spam verdict itself, the named positive signals are specific: a message from an address in the recipient's contacts is "less likely to be marked as spam" [9]; a reply is "another way to indicate you and the sender are familiar" [29]; and marking "not spam" means "future emails from that sender won't go to Spam" [30]. The named negative signal is the spam report, which for the individual recipient means "emails from the same sender might be sent to the Spam folder in the future" [30], and which bulk senders receive in aggregate through Gmail's feedback loop, where a rising complaint rate degrades domain reputation [31].

The Promotions tab is distinct from the Spam folder. Google defines Promotions as a delivered inbox category, "Deals, offers, and other promotional emails," and describes tabs as an ML "classification system ... based on a variety of signals" applied to delivered mail [32, 29]. Google publishes no sentence stating that Promotions is not spam; the distinction follows from the definitions themselves, since Promotions is a delivered inbox tab and Spam is removal from the inbox, and it is presented here as an inference from those definitions, not as a quotation.

6. Content signals

Google's stated content rules are few, and they concern deception [33]. Google does say: do not use "HTML and CSS to hide content" [9]; keep web links "visible and easy to understand" [9]; do not send links whose URL "doesn't match the description" [34]; and it scans URLs and attachments through Safe Browsing and dedicated models [35, 36]. URL-reputation filtering of this kind, and its measured limits, is a research field in its own right [37, 38]; the threat it targets is credential phishing and malware delivery [39]. Impersonation, which Google calls "spoofing," "might" cause a message to be marked as spam [9]. These signals are sourced to Google.

No Google source was found for the following widely circulated content claims, and each is labeled accordingly.

- Spam-trigger-word lists ("free," "guarantee"). Google names no prohibited words; its stated signals are reputation, authentication, and user feedback [7]. A legacy contribution from older statistical word-probability filters may exist at the margin [40], but a curated word list is not a documented lever.
- Image-to-text ratio thresholds (the "60/40 rule"). No Google source specifies a ratio. An all-image message can create rendering and accessibility problems that depress engagement, an indirect effect, and no Google source states a ratio-based spam score.
- Hard link-count caps ("one link only"). Google requires that links be visible and legitimate and scans their reputation [9, 36], but names no maximum. Widely cited figures ("6+ links, 73% more likely spam") trace only to marketing blogs and are unverified.
- `Precedence: bulk` headers, capital letters, or exclamation marks as direct triggers. Google's only near-relevant statements concern "misleading" headers and using emoji "to imitate graphic elements" [9]; neither is a stated penalty on formatting as such.

This pattern is consistent with Section 5.2: RETVec was built to resist adversarial text manipulation [28], and the content signals Google does state concern deception and user safety, with word choice and formatting absent from them.

7. Authentication policy and impersonation: the placement evidence

One study in the literature measures inbox-versus-spam placement end to end, and its subject is spoofing, not legitimate sending. Hu and Wang (USENIX Security 2018) sent 52,500 controlled messages across 35 popular providers, crossing the impersonated domain's authentication profiles, content types, and sending IPs, and recorded inbox versus spam for each forged message [3]. Table 2 reports the placement rates by the impersonated domain's published policy.

Table 2. Hu and Wang (2018) [3]: inbox rate of a forged message as a function of the impersonated domain's published authentication policy. The aggregate column is measured across 35 providers; the Gmail column reports Gmail alone. Rates are the share of forged messages reaching the inbox.

Impersonated domain's policy	Aggregate inbox rate	Gmail inbox rate
No SPF/DKIM/DMARC	60.5%	93%
Relaxed / `p=none`	47.3%	66%
Strict / `p=reject`	28.4%	0%

The impersonated domain's published policy was strongly associated with the forgery's placement, although it did not eliminate delivery in aggregate. Moving the impersonated domain from no authentication to a strict `p=reject` policy roughly halved the aggregate inbox rate (60.5% to 28.4%, a 32.1 percentage-point reduction). Expressed as odds, an unauthenticated domain had about 3.9 times the odds of a forgery reaching the inbox relative to a strict-policy domain (odds ratio approximately 3.86; from `p=none` the odds ratio is approximately 2.26). Hu and Wang report the underlying association as significant by chi-square ($p < 0.00001$) but do not publish per-cell counts, so the point odds ratios are reported here and confidence intervals are not computed, because they are not derivable from the aggregated proportions alone [3].

Providers favored delivery: 34 of 35 providers, Gmail among them, delivered at least one forged message to the inbox; only Hotmail blocked all of them [3]. For Gmail specifically, the measured inbox rate for the forgery fell from 93% against an unauthenticated domain to 0% under `p=reject` [3]; strict sender-side DMARC on the impersonated domain was, for Gmail, decisive against forgery. Receiver-side DMARC checking added little on its own: providers performing full authentication showed a 39.0% inbox rate versus 39.3% for SPF/DKIM-only providers, a 0.3 percentage-point difference that is not significant ($p = 0.495$) [3]. In this study, placement varied with the impersonated domain's published policy, while receiver-side checking showed no significant difference.

The scope of this evidence is limited. Hu and Wang measure whether a forged message survives; their numbers speak directly to authentication's role in defeating impersonation in a domain's name, and only by extension to the baseline value of a legitimate sender authenticating its own domain. They also, like the rest of the literature, do not measure reputation. The remaining studies measure the deployment conditions that constrain authentication as a signal. Adoption remained low and rising (SPF near 45 to 51%, DMARC roughly 1% in 2015 climbing to about 11% by 2020, mostly `p=none`; all three of SPF, DKIM, and DMARC on only about 3% of domains) [5, 41], with subsequent studies confirming both the slow uptake of anti-spoofing protocols [43] and the persistence of SPF misconfiguration at scale [42]. Enforcement was historically weak: in 2015, invalid-DKIM mail "was allowed to successfully complete SMTP delivery" (RFC 5321) at most providers [4, 44]. The DKIM signal is noisy (about 84% of DKIM-enabled domains use weak keys) [11], and the DMARC reporting loop that

would let senders tighten policy safely is itself frequently misconfigured (about 26% of reporting configurations) [12].

On this evidence, authenticating one's own domain is a requirement under the sender guidelines and is positively associated with placement in the reviewed literature. The literature does not support a claim that authentication or reputation causes inbox placement for a legitimate sender in the experimental sense, and no such claim is made here.

8. Discussion

The evidence reviewed in Sections 4 through 7 combines into a three-layer decision (Figure 1). A compliance floor (Section 4) determines eligibility: a sender that fails authentication, formatting, or the 0.30% spam-rate line faces throttling or rejection. A reputation layer (Section 5.1) grades the sender's history into four bands whose definitions Google ties to the inbox, spam, or reject outcome. A personalized machine-learning classifier (Sections 5.2 and 5.3) makes the per-recipient call, with recipient engagement described by Google as a central signal.

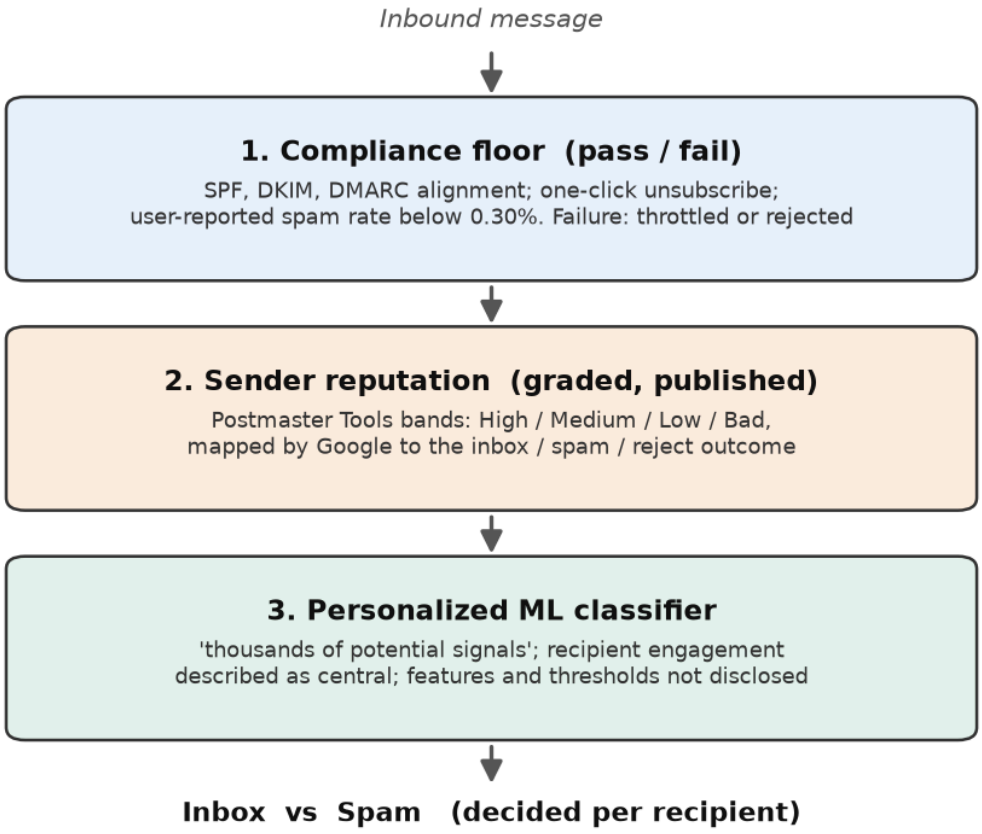

Figure 1. Gmail's inbox-versus-spam decision as synthesized in Sections 4 and 5 from Google's documentation. The diagram shows the three layers in the order they apply to an inbound message: a pass/fail compliance floor, a sender-reputation layer graded in four bands published in Postmaster Tools, and a personalized machine-learning classifier that produces a per-recipient inbox-or-spam outcome. Quoted phrases are Google's.

Three implications for high-volume cold-email senders follow from this evidence. First, the actions with documented consequences precede message content: meeting the Section 4 authentication and formatting requirements [9], and holding the user-reported spam rate below the advised 0.10%, the one threshold Google quantifies [9, 13]. Established sender best-common-practice documents codify the same

discipline [45]. Second, because Google describes engagement as central and the verdict as personalized [7, 29], list selection and message relevance bear more directly on placement than content-level adjustments, which the classifier is trained to resist [28]. Third, placement in the Promotions tab is delivery to the inbox [32]; it should be measured and analyzed separately from placement in the Spam folder.

These findings are in line with the measurement literature. Hu and Wang found that a domain's own authentication policy sharply limited impersonation in its name without eliminating it [3], and the adoption and enforcement studies show that authentication, as a sparse, weakly enforced, and noisy signal, cannot account for the full decision [4, 5, 11]. Sender reputation and the classifier's internal features are unmeasured in the literature and undisclosed by Google, and the account given here stops at that boundary.

9. Limitations

This is an observational synthesis of public evidence, with the following limits. (1) Opacity: Gmail's classifier features, weights, and thresholds are undisclosed, so the account of the classification layer is limited to Google's stated approach, inferences beyond Google's words are marked, and the internal combination of signals cannot be verified. (2) No original placement data: no controlled measurement of our own is presented; the one placement study relied on measures spoofing, not legitimate cold email, and generalizes to the legitimate-sender case only by extension. (3) Reputation is unmeasured in the entire literature; the treatment of it rests on Google's qualitative band definitions alone, so the relative weight of reputation in the decision cannot be assessed. (4) Purposive coverage: the synthesis is a targeted review of the placement-relevant literature, not a PRISMA-complete systematic review, so a study not surfaced by the search could refine the associational picture. (5) Recency: provider rules change; every rule here is dated, and enforcement details may have evolved since the 2024 through 2026 window. (6) Gmail specificity: none of this transfers to Outlook, Yahoo, or other providers without separate evidence. (7) Statistical precision and coding reliability: the effect sizes derived from Hu and Wang (Section 7) are point estimates; the source publishes no per-cell counts, so confidence intervals are not derivable and the odds ratios should not be read as precise. As a single-author synthesis, source selection and extraction were not cross-coded, so selection bias cannot be excluded (Section 3).

10. Conclusion

This paper set out to describe, from primary sources, how Gmail decides whether legitimate high-volume cold email reaches the inbox or the spam folder. On the descriptive question, the evidence supports a three-layer account: a pass/fail authentication and formatting floor, including the 0.30% user-reported spam-rate limit; a sender-reputation layer graded in four published bands; and a personalized machine-learning classifier in which Google describes recipient engagement as central. The widely circulated content-level rules examined in Section 6 have no support in Google's documentation. On the relational question, the one direct placement measurement in the literature concerns forged mail: a domain's published DMARC policy was strongly associated with whether forgeries in its name reached the inbox, and this bears on legitimate senders only by extension. A controlled measurement of authentication and reputation against live Gmail placement for legitimate senders has not been published. Such a study, a seed-panel measurement with a stated hypothesis, full methodology, effect sizes with confidence intervals, and a named reference standard, is a direction for future work.

11. Declarations

Competing interests. The author is the founder of SpamCipher, a commercial cold email and deliverability platform. This is a material competing interest. The analysis draws only on primary and peer-reviewed sources; no product is evaluated, recommended, or required by any result.

Author contributions. Francis Davison is the sole author and performed the conception, source identification and grading, analysis, and drafting.

Data availability. This paper presents no original dataset; all cited evidence is publicly available at the referenced locations.

Funding. None external; produced by SpamCipher.

References

[1] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, S. Savage, "Spamalytics: An Empirical Analysis of Spam Marketing Conversion," ACM CCS 2008. https://doi.org/10.1145/1455770.1455774

[2] K. Levchenko et al., "Click Trajectories: End-to-End Analysis of the Spam Value Chain," IEEE Symposium on Security and Privacy 2011. https://doi.org/10.1109/SP.2011.24

[3] H. Hu and G. Wang, "End-to-End Measurements of Email Spoofing Attacks," USENIX Security 2018, pp. 1095-1112. https://www.usenix.org/conference/usenixsecurity18/presentation/hu

[4] I. Foster, J. Larson, M. Masich, A. Snoeren, S. Savage, K. Levchenko, "Security by Any Other Name: On the Effectiveness of Provider Based Email Security," ACM CCS 2015. https://doi.org/10.1145/2810103.2813607

[5] Z. Durumeric et al., "Neither Snow Nor Rain Nor MITM...: An Empirical Analysis of Email Delivery Security," ACM IMC 2015. https://doi.org/10.1145/2815675.2815695

[6] Google Workspace, "Spam does not bring us joy: ridding Gmail of 100 million more spam messages with TensorFlow," 2019. https://workspace.google.com/blog/product-announcements/ridding-gmail-of-100-million-more-spam-messages-with-tensorflow

[7] Google Workspace, "An overview of Gmail's spam filters." Accessed 2026-07-06. https://workspace.google.com/blog/identity-and-security/an-overview-of-gmails-spam-filters

[8] G. Stringhini, T. Holz, B. Stone-Gross, C. Kruegel, G. Vigna, "BotMagnifier: Locating Spambots on the Internet," USENIX Security 2011. https://www.usenix.org/conference/usenix-security-11/botmagnifier-locating-spambots-internet

[9] Google, "Email sender guidelines," Gmail Help. Accessed 2026-07-06. https://support.google.com/a/answer/81126

[10] Google, "New Gmail protections for a safer, less spammy inbox," The Keyword, 2023. https://blog.google/products-and-platforms/products/gmail/gmail-security-authentication-spam-protection/

[11] C. Wang et al., "A Large-scale and Longitudinal Measurement Study of DKIM Deployment," USENIX Security 2022. https://www.usenix.org/conference/usenixsecurity22/presentation/wang-chuhan

[12] M. I. Ashiq, W. Li, T. Fiebig, T. Chung, "You've Got Report: Measurement and Security Implications of DMARC Reporting," USENIX Security 2023. https://www.usenix.org/conference/usenixsecurity23/presentation/ashiq

[13] Google, "Postmaster Tools dashboards," Gmail Help. Accessed 2026-07-06. https://support.google.com/mail/answer/14668346

[14] Google, "Email sender guidelines FAQ," Gmail Help. Accessed 2026-07-06. https://support.google.com/a/answer/14229414

[15] S. Kitterman, "Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1," RFC 7208, IETF, 2014. https://datatracker.ietf.org/doc/html/rfc7208

[16] D. Crocker, T. Hansen, M. Kucherawy (Eds.), "DomainKeys Identified Mail (DKIM) Signatures," RFC 6376, IETF, 2011. https://datatracker.ietf.org/doc/html/rfc6376

[17] M. Kucherawy, E. Zwicky (Eds.), "Domain-based Message Authentication, Reporting, and Conformance (DMARC)," RFC 7489, IETF, 2015. https://datatracker.ietf.org/doc/html/rfc7489

[18] J. Levine, "Signaling One-Click Functionality for List Email Headers," RFC 8058, IETF, 2017. https://datatracker.ietf.org/doc/html/rfc8058

[19] P. Resnick (Ed.), "Internet Message Format," RFC 5322, IETF, 2008. https://datatracker.ietf.org/doc/html/rfc5322

[20] Yahoo, "Sender Best Practices," Yahoo Sender Hub. Accessed 2026-07-06. https://senders.yahooinc.com/best-practices/

[21] K. Andersen, S. Blank, J. Levine (Eds.), "The Authenticated Received Chain (ARC) Protocol," RFC 8617, IETF, 2019. https://datatracker.ietf.org/doc/html/rfc8617

[22] Google, "Best practices for forwarding email to Gmail," Gmail Help. Accessed 2026-07-06. https://support.google.com/mail/answer/175365

[23] F. Martin, E. Lear, T. Draegen, E. Zwicky, K. Andersen (Eds.), "Interoperability Issues between DMARC and Indirect Email Flows," RFC 7960, IETF, 2016. https://datatracker.ietf.org/doc/html/rfc7960

[24] M3AAWG, "Email Authentication Recommended Best Practices," Messaging, Malware and Mobile Anti-Abuse Working Group, 2020. https://www.m3aawg.org/sites/default/files/m3aawg-email-authentication-recommended-best-practices-09-2020.pdf

[25] The official Gmail Blog, "The mail you want, not the spam you don't," 2015. https://gmail.googleblog.com/2015/07/the-mail-you-want-not-spam-you-dont.html

[26] A. Pitsillidis, K. Levchenko, C. Kreibich, C. Kanich, G. Voelker, V. Paxson, N. Weaver, S. Savage, "Botnet Judo: Fighting Spam with Itself," NDSS 2010. https://www.ndss-symposium.org/ndss2010/botnet-judo-fighting-spam-itself/

[27] E. Bursztein, M. Zhang, O. Vallis, X. Jia, A. Kurakin, "RETVec: Resilient and Efficient Text Vectorizer," NeurIPS 2023, arXiv:2302.09207. https://arxiv.org/abs/2302.09207

[28] Google Security Blog, "Improving Text Classification Resilience and Efficiency with RETVec," 2023. https://security.googleblog.com/2023/11/improving-text-classification.html

[29] Google Workspace, "How Gmail sorts your email based on your preferences." Accessed 2026-07-06. https://workspace.google.com/blog/productivity-collaboration/how-gmail-sorts-your-email-based-on-your-preferences

[30] Google, "Report spam in Gmail," Gmail Help. Accessed 2026-07-06. https://support.google.com/mail/answer/1366858

[31] Google, "Feedback Loop," Gmail Help (Google Workspace Admin). Accessed 2026-07-06. https://support.google.com/a/answer/6254652

[32] Google, "Organize your emails into categories," Gmail Help. Accessed 2026-07-06. https://support.google.com/mail/answer/3094499

[33] Google, "Top 10 Gmail sender issues," Gmail Help. Accessed 2026-07-06. https://support.google.com/mail/answer/15256272

[34] Google, "Avoid and report phishing emails," Gmail Help. Accessed 2026-07-06. https://support.google.com/mail/answer/8253

[35] Google Workspace, "Advanced phishing and malware protection." Accessed 2026-07-06. https://knowledge.workspace.google.com/admin/gmail/advanced/advanced-phishing-and-malware-protection

[36] Google, "Safe Browsing," Google for Developers. Accessed 2026-07-06. https://developers.google.com/safe-browsing

[37] K. Thomas, C. Grier, J. Ma, V. Paxson, D. Song, "Design and Evaluation of a Real-Time URL Spam Filtering Service," IEEE Symposium on Security and Privacy 2011. https://doi.org/10.1109/SP.2011.25

[38] A. Oest, P. Zhang, B. Wardman, E. Nunes, J. Burgis, A. Zand, K. Thomas, A. Doupé, G.-J. Ahn, "Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale," USENIX Security 2020. https://www.usenix.org/conference/usenixsecurity20/presentation/oest-sunrise

[39] K. Thomas et al., "Data Breaches, Phishing, or Malware? Understanding the Risks of Stolen Credentials," ACM CCS 2017. https://doi.org/10.1145/3133956.3134067

[40] M. Sahami, S. Dumais, D. Heckerman, E. Horvitz, "A Bayesian Approach to Filtering Junk E-Mail," AAAI Workshop on Learning for Text Categorization, 1998. https://cdn.aaai.org/Workshops/1998/WS-98-05/WS98-05-009.pdf

[41] D. Tatang, F. Zettl, T. Holz, "The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws," RAID 2021. https://doi.org/10.1145/3471621.3471842

[42] S. Czybik, M. Horlboge, K. Rieck, "Lazy Gatekeepers: A Large-Scale Study on SPF Configuration in the Wild," ACM IMC 2023. https://doi.org/10.1145/3618257.3624827

[43] H. Hu, P. Peng, G. Wang, "Towards Understanding the Adoption of Anti-Spoofing Protocols in Email Systems," IEEE Secure Development Conference (SecDev) 2018. https://doi.org/10.1109/SecDev.2018.00020

[44] J. Klensin, "Simple Mail Transfer Protocol," RFC 5321, IETF, 2008. https://datatracker.ietf.org/doc/html/rfc5321

[45] M3AAWG, "Sender Best Common Practices, Version 3.0," Messaging, Malware and Mobile Anti-Abuse Working Group, 2015. https://www.m3aawg.org/documents/en/m3aawg-sender-best-common-practices-version-30

Competing interest: the author is the founder of SpamCipher, a commercial cold email and deliverability platform. This analysis is grounded only in primary and peer-reviewed sources. Canonical version: https://spamcipher.com/insights/how-gmail-decides-inbox-vs-spam