

From Chaos To Clarity: How Data Quality & AI Are Reshaping Email Targeting



Sherie Driscoll



Jess Adams

Cambridge University Press & Assessment

Effects of poor data

New data shows **email databases now decay at least 28% per year**, up from the ~22% cited from 3 years ago - a roughly 25% faster rate of list decay.

[Zero Bounce](#)

Average **cost of \$15k USD per million emails sent** for a business due to deliverability problems

[Litmus](#)

One in six legitimate emails don't reach the inbox

[Litmus](#)

The problem of poor data quality

The deliverability problem: bad data kills sender reputation

- Senders maintaining <1.5% bounce rate see 10-12% higher inbox placement
- Exceeding industry benchmarks of a hard bounce rate of 0.21 or a soft bounce rate of 0.7 will damage sender reputation

Metrics lie when data is poor

- UK GDPR and PECR (and increasingly many international ePR) requires documented consent records. List decay (high bounce rates) means expired consent = compliance breach
- Penalties can be up to €20M or 4% of global revenue + ICO action

Compliance and privacy risks

- UK GDPR and PECR (and increasingly many international ePR) requires documented consent records. List decay (high bounce rates) means expired consent = compliance breach
- Penalties can be up to €20M or 4% of global revenue + ICO action

The “creepy problem” – the personalisation paradox

- Consumers want relevance but worry about data privacy and mistrust misuse.
- 75% want and engage with personalised content but ~70% worry about data privacy and think companies collect too much personal data.

Why good data is critical today

ePrivacy Regulations define data quality and
AI-powered inboxes demand accuracy, relevance, and trust alongside that.

ePrivacy Regulations define data quality

GDPR/ePrivacy
force regular
data audits
(consent
records, opt-in
proof)

**Stale data =
expired consent
= compliance
risk**

**List decay
demands
systematic
refresh** and re-
engagement
cycles

**Data
minimisation:**
Only keep data
you can justify
using

AI is redefining email quality standards

Gmail 2025: Real-
time sender
reputation scoring
is based on user
engagement
frequency and reply
patterns

**Engagement
signals** (opens,
clicks, replies)
weighted most
heavily by AI-driven
Machine Learning
filters

**Mailbox provider
validation:** Gmail,
Outlook, Yahoo
require
SPF/DKIM/DMARC;
and MAGY inbox
placement at 86%

**User feedback
loops** continuously
train filters;
complaints from
stale/irrelevant
sends will tank
scores

**Poor data (invalid
addresses, stale
contacts) = poor
engagement
signals = poor
inbox placement**

What is 'quality data'?

Let's look at the three pillars of email data excellence...

Accuracy: getting the basics right

Correct and current data about who your subscribers are:

- Valid email syntax (no typos, formatting errors)
- Current role, title, organisation information
- Correct engagement preferences and segmentation attributes

Why it matters: Reduces bounces, improves targeting signals to AI filters



Completeness

Rich data that enables smarter segmentation & personalisation:

- First/last name, job title, industry, company size
- Engagement history, purchase stage, behaviour signals
- Preference centre data (topics, frequency, channels)

Why it matters: More context enables AI to rank relevance more accurately. Personalisation only works when you have the context.



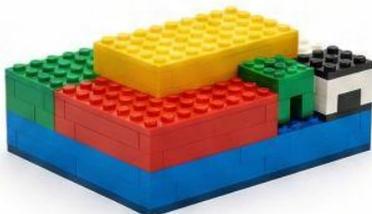
RAW GUEST DATA (Messy Input)



SORTED DATA (By Source → Color Coded)



ARRANGED → UNIFIED GUEST PROFILE (360° View)



STORY LEVEL → SEGMENTATION (Guest Personas)



Raw Data

Sorted Data

Validated and compliant data

Personalised experiences



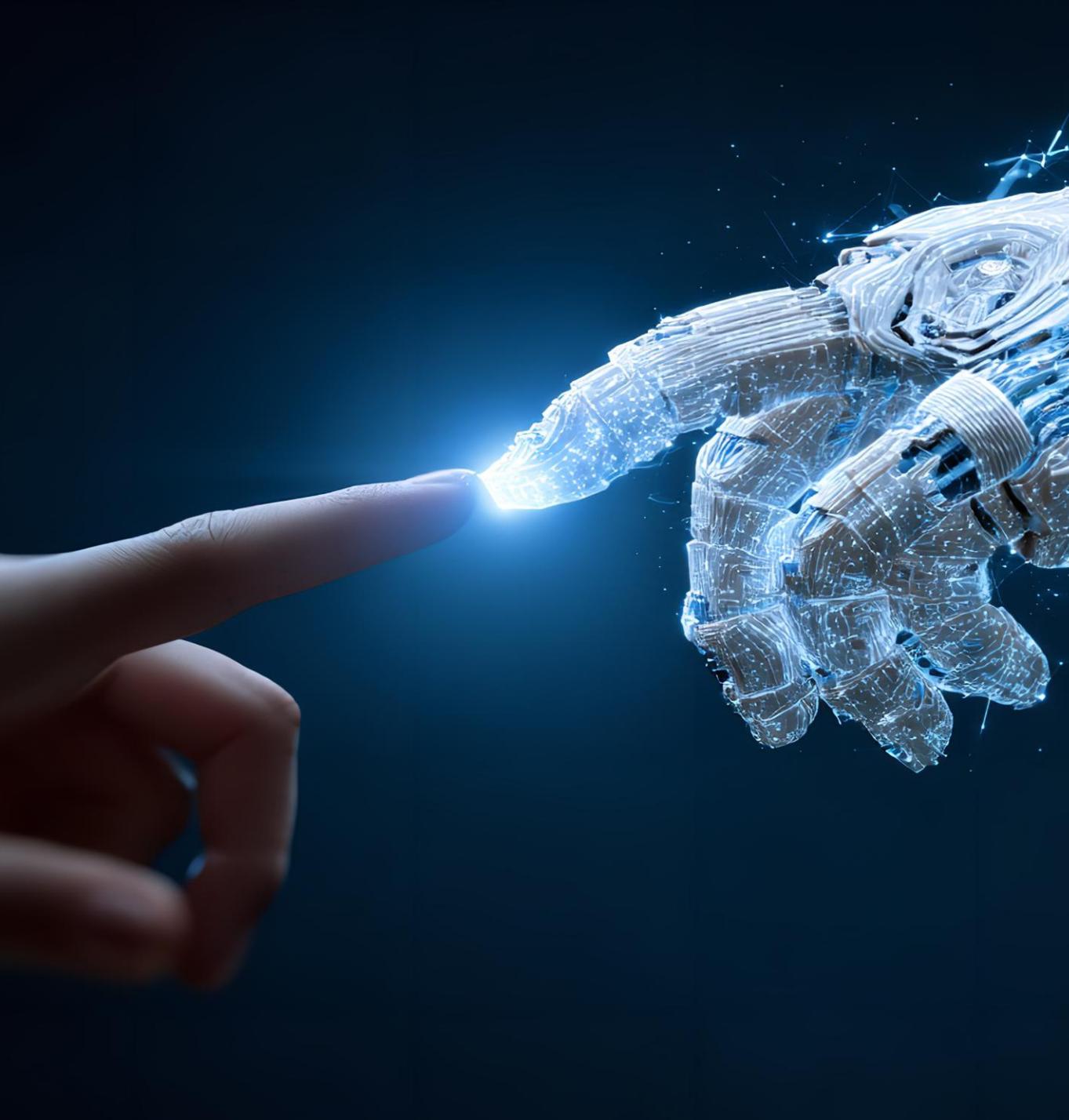
Freshness: data has an expiration date

Personal data has a peak freshness window. Your data is most valuable immediately after collection.

- Email lists decay at 28% annually (2025 benchmark); tech/B2B even faster due to job churn
- Re-engagement windows: Monitor 60-90 days inactivity; win-back campaigns achieve 14-29% success
- Contact attribute refresh: Update quarterly

Why it matters: PECR requires proof of consent; AI filters heavily favour recent engagement signals.





AI works best with humans in the loop

- AI can **identify patterns**, but humans make **judgment calls**.
- AI flags suspicious patterns – you decide action
- AI suggests send times – you validate for your audience
- AI personalises content – you review for tone and brand fit
- AI identifies churned contacts – you decide re-engagement strategy

AI & Automation: Your Quality Multipliers

Strategies to operationalise data quality



VALIDATION

Real-time email verification at signup blocks bad data before it enters your system



ENRICHMENT

AI fills gaps – finding missing job titles, company details, engagement signals



HYGIENE

Automated list cleaning removes bounces, inactive addresses, duplicates

Your data readiness framework

ASSESS

- Audit your lists:
How many bounces?
Missing fields?
Last validation?
- What's your compliance status?
Identify the biggest risks
- Calculate cost impact – quantify cost of wasted sends, deliverability loss and compliance exposure
- Set Objective and Key Result Targets – what does success look like.

CLEAN

- Remove bad data
- Validate your email send lists
- Suppress high risk contacts
- Identify the peak period for reengagement and win back or remove those contacts
- Enrich missing data with AI from trusted sources to unlock better targeting and segmentation

ENRICH & MAINTAIN

- Once you've cleaned your data – stay on top of it. Continuously look to enrich your database with current interests. AI pipelines can help keep data fresh
- Automate validation of new contacts
- Set governance rules around frequency caps, reengagement windows
- Monitor health and track against your industry benchmarks

Takeaways

- Data quality rests on **three pillars**: Accuracy, Completeness, Freshness
- AI-powered inboxes and privacy regulations both demand better data - it's no longer optional
- The roadmap is simple: **Assess your baseline, Clean your list, Maintain ongoing hygiene**

+10-12% higher inbox placement with <1.5% bounce rate

Validity 2025

>+29% open rates with subject line personalisation

Omnisend 2025

> +41% CTR with accurate personalisation

Omnisend 2025

Thank you

Connect with us on LinkedIn....

