PROMPTSLOVE RESEARCH · REPORT 1.0

What Actually Travels: An Empirical Study of Virality in AI Discourse on X

21,864 high-engagement posts on AI, Claude, ChatGPT and prompt engineering, analysed with follower-stratified effect sizes and an adversarial multi-agent verification protocol

Ramanpal Singh

Promptslove · promptslove.com
admin@promptslove.com

21,864 posts · **7,829** creators · **448** queries

24 August 2026

Citation and Rights

CITE THIS REPORT AS

Singh, R. (2026). What Actually Travels: An Empirical Study of Virality in AI Discourse on X. Promptslove Research Report 1.0. promptslove.com.

BIBTEX

```
@techreport{singh2026virality,
  title = {{What Actually Travels: An Empirical Study of Virality in AI Discourse on X}},
  author = {Singh, Ramanpal},
  institution = {Promptslove},
  year = {2026},
  type = {Research Report},
  number = {1.0},
  url = {https://promptslove.com}
}
```

Author. Ramanpal Singh, Promptslove (promptslove.com). Correspondence to admin@promptslove.com.

Data availability. The analysed corpus, the derived feature table, the analysis modules and the report-generation code are held in the study repository. The corpus was collected from a public commercial API under its terms of service; the API credential is not distributed. Reproduction instructions are given in Appendix F.

Conflicts of interest. The author operates promptslove.com, a commercial publisher of AI prompt and automation products. The study analyses public posts about AI tools, including categories in which the author has a commercial interest. No sponsor influenced the design, analysis or reporting, and no finding was selected for its commercial convenience. The practitioner recommendations in Chapter 9 are derived from the same public data presented throughout and can be checked against it.

Use of automated systems. Data collection, feature extraction, statistical analysis, adversarial verification and a first draft of the reporting text were produced with the assistance of large language model agents, under the author's direction and review. The verification protocol described in Chapter 8 was applied to the analysis; residual error remains the author's responsibility.

Preface

This report began as a practical question rather than a research one. Working on products in the AI tooling space, the author was repeatedly handed advice about how to build an audience on X: post threads, post at particular hours, add hashtags, tag large accounts, offer a free resource in exchange for a comment. The advice was confident, widely repeated, and — as far as anyone could demonstrate — untested. Where evidence was offered at all, it took the form of a screenshot from an account with two hundred thousand followers, which is a demonstration of what large accounts can do rather than of what the tactic contributes.

The obvious response was to collect the data. That turned out to be the easy part. The harder part, and the reason this document is longer than a summary of results would need to be, is that nearly every naive analysis of engagement data is wrong in a way that is invisible from the output. Comparisons between formats mostly measure which formats large accounts prefer. A flag for “contains a link” mostly measures whether a post carries an image, because the platform appends a shortened link to every native photo and video. A ratio whose denominator is the variable used to split the sample will produce a strong, meaningless effect. Each of these errors appeared in this study's own working, was caught, and is documented here rather than quietly repaired, because the errors are more instructive than the findings they contaminated.

That experience shaped the method. The analysis was run twice over: once by agents asked to find what the data showed, and again by independent agents asked to attack what the first pass claimed, re-deriving the numbers from the raw records. That second pass produced one hundred and thirty-nine objections against one hundred and fifteen findings, twelve of them fatal. The surviving results are reported with their objections attached. A reader who disagrees with a conclusion here should be able to see exactly which number it rests on and what was already said against it.

Two limitations deserve stating at the outset, before any result is read. First, the corpus contains only posts that already performed well; it can rank winners against one another but cannot say what proportion of any tactic fails, because failures were never collected. Second, the study observes one platform, one topic area and eight months. Nothing here is a law of communication. It is a description of a particular arena during a particular window, offered with enough method and data for the next person to check it.

The report is written for two audiences at once: practitioners who want to know what to do on Monday, and analysts who want to know whether to believe it. Chapters 5 through 7 serve the first; Chapters 3, 8 and 9 serve the second. Readers in a hurry are directed to Chapter 10, which answers each research question in a sentence, and to Chapter 9's practitioner implications, which are ordered by effect size relative to the effort they require.

Ramanpal Singh
Promptslove · promptslove.com

Abstract

Background. Guidance on building an audience on X in the artificial-intelligence niche circulates as anecdote. It is rarely tested, and the comparisons it rests on are confounded by account size, because the accounts producing the advice are large.

Method. We assembled an observational corpus of 21,864 high-engagement English posts about AI, Claude, ChatGPT, prompt engineering and adjacent topics, published between 1 January and 24 August 2026, collected through a commercial search API across 448 queries constructed from topic terms, engagement floors and monthly date slices. Each post was annotated with 49 features covering format, media, opening-line language, persuasion triggers, author characteristics and engagement outcomes. Posts were partitioned into three audience segments by a keyword-and-biography classifier. Effects are reported as median-based lifts computed within follower-count strata, with bootstrap confidence intervals, Mann–Whitney tests and Benjamini–Hochberg correction. Twelve independent analyses were then subjected to an adversarial verification protocol in which a second agent re-derived each headline number from the raw records and attempted to refute it.

Results. Audience segmentation is not optional: mass-audience posts that merely mention a model reach substantially higher like counts than practitioner posts while attracting far fewer saves, and pooling the two corrupts every downstream comparison. Follower-stratified analysis reverses several widely repeated conclusions, including the relative standing of demonstration posts and image carousels. Measurement validity proved to be the dominant source of error: a conventionally constructed “contains a link” feature reported outbound links as mildly beneficial, whereas the same relationship measured from the platform's own URL entities is strongly negative. Language effects concentrate in the opening line, where naming a specific tool out-performs the generic category term. The verification protocol raised 139 objections, 12 fatal, against 115 candidate findings.

Conclusions. Within a winners-only sample, the largest measurable differences in this arena come from the opening line and from the choice of engagement currency — reach or saved reference value — rather than from posting time or attachment type. For analysts, the study's strongest practical result is methodological: in social-engagement data, stratification and construct validity change conclusions more often than statistical technique does.

Keywords: social media analytics · virality · content strategy · observational study · confounding · construct validity · multi-agent verification · X (Twitter) · large language models · creator economy

Table of Contents

List of Figures

List of Tables

1. Introduction

Advice about how to make a post travel on X in the artificial-intelligence niche is abundant, confident and almost entirely untested. It circulates as screenshots, as threads about threads, and as lists of rules — open with a number, never use hashtags, post at seven in the morning, always attach an image, never send readers off-platform. Three structural problems compound in it. It is anecdotal: a rule is offered on the strength of one post that worked, with no comparison group and no count of the posts that used the same tactic and did not. It is produced by accounts whose follower base is itself the dominant cause of the outcome being explained, so that any comparison between the tactics such an account uses and those it avoids is confounded by the account. And it is never re-tested; once a rule enters circulation it is repeated rather than checked, and the repetition is mistaken for corroboration.

The research gap is therefore not a shortage of claims about virality but a shortage of claims that survive contact with a comparison group, a stratification by account size, and an adversary. We assembled an observational corpus of 21,864 high-engagement posts about AI, Claude, ChatGPT and prompt engineering published between 1 January and 24 August 2026 — 236 days, 7,829 distinct authors, 89.3 million likes and 12.1 billion platform-reported views — and used it to re-test the circulating advice under conditions in which the confounds are visible. The result is as much a study of how engagement data misleads as of what engages.

> **In this corpus, stratification and construct validity changed more conclusions than any choice of statistical technique did.**

1.1 Background and motivation

X occupies an unusual position in the AI tooling economy. It is where model releases are announced, where practitioners publish workflows before those workflows reach documentation, and where an account's distribution translates into hiring, funding and product adoption. A practitioner deciding what to publish is making a commercial decision under uncertainty, and the guidance available is folklore.

The folklore is not obviously wrong. Much of it is a plausible reading of what large accounts do, and some is consistent with an established literature: emotional arousal and practical utility both predict sharing [Berger & Milkman, 2012], and outcomes in networked cultural markets are path-dependent, so that success is partly a self-fulfilling consequence of early visibility [Salganik, Dodds & Watts, 2006]. Nor is the platform entirely opaque: X open-sourced components of its recommendation system in 2023 [Twitter, 2023]. What is missing is not theory but measurement in the arena where the advice is applied.

The practitioner's core problem is that the observable evidence is unreadable without stratification. The corpus makes the size gradient explicit: among builder and mixed posts, the median post from an account under 1,000 followers drew 584 likes and the median post from an account above one million drew 2,128, against a population median of 906.5 (n=18,692). Engagement also concentrates among authors: the Gini coefficient of likes across the 7,829 creators is 0.805, and the top decile of creators accounts for 70.9 per cent of all likes. Since format choice is correlated with account size, a pooled comparison of formats measures, in part, which formats large accounts prefer — Simpson's paradox operating on follower count [Simpson, 1951], and not a hypothetical risk: pooled and stratified estimates disagree in sign in this data.

A second problem is that the audience is not one audience. A keyword-and-biography classifier partitions the corpus into practitioner-facing builder posts (n=3,370), mass-audience consumer chatter that merely mentions a model (n=3,172), and a large mixed remainder (n=15,322, 70.1 per cent). The poles behave differently enough to invert advice: consumer posts reach a median of 2,011 likes against 716 for builder posts, a ratio of 2.81, while collecting a median of 218 bookmarks against 845 — 7.4 saves per hundred likes versus 131.6 (Mann–Whitney U, $p < 10^{-15}$). A single pooled table over these populations describes no one.

Table 1.1 — The three populations used throughout the report. All values are medians over the population named; bookmarks per 100 likes is computed per post and then medianed.

Population	Posts	Creators	Median likes	Median bookmarks	Bm / 100 likes	Median followers
Full corpus (builder + mixed + consumer)	21,864	7,829	996	375	31.1	49,713
Analysis (builder + mixed)	18,692	6,365	906.5	408	42.4	49,713
Sub-25K stratum (builder + mixed, <25,000 followers)	6,521	3,341	650	347	51.1	9,003

1.2 Research questions

The study answers six questions, each stated in the form the data can support — a comparison among posts that already performed — and each answered against a named population and baseline.

1. **RQ1 (format).** Which structural formats — thread opener, single post, quote post, screen-recorded demonstration, image, carousel, video — are associated with higher median engagement, and does that ordering survive follower-tier stratification?
2. **RQ2 (hook language).** Which opening-line properties — naming a specific tool rather than the category, a numeral, a colon, an all-caps word, hook length, explicit reply-bait

— are associated with higher median likes and saves after correction for multiple comparisons?

3. **RQ3 (audience segmentation).** How far do builder and consumer posts differ in level and in currency of engagement, and which conclusions change when the two are pooled rather than separated?
4. **RQ4 (account-size stratification).** How much of an apparently content-attributable effect is attributable to account size instead, and which pooled effects invert, attenuate or survive inside the sub-25K stratum?
5. **RQ5 (measurement validity).** How large is the error introduced by a conventionally constructed feature — a substring test for outbound links — relative to the same construct measured from the platform's URL entities, and what does that imply for analyses built on scraped text features?
6. **RQ6 (the small-account playbook).** Among accounts below 25,000 followers, which associations are large, robust and cheap enough to act on, and which widely repeated tactics are associated with worse outcomes in that stratum?

1.3 Contributions

The study makes five contributions.

1. **An annotated, documented corpus.** 21,864 high-engagement posts spanning 236 days, collected across 448 queries built from topic terms, engagement floors and monthly date slices, each row carrying fifty columns of author metadata, engagement counts and derived features, with its selection mechanism and known defects stated rather than smoothed over.
2. **A builder/consumer audience segmentation.** An operational classifier separating practitioner-facing posts from mass-audience chatter, with evidence that the two populations differ not only in magnitude but in the currency of engagement they attract — likes versus saves — which is what makes pooled advice actively misleading.
3. **A stratified-lift method.** Effects are reported as within-follower-tier median ratios combined across tiers, alongside the pooled estimate, so that the size of the confound is itself a reported quantity. Two headline examples: the demonstration format is associated with a 0.4 per cent lift pooled (n=513) but 58.2 per cent inside the sub-25K stratum (n=201), and the carousel format inverts from +7.7 per cent pooled (n=945) to −17.2 per cent in the same stratum (n=308).
4. **An adversarial verification protocol.** Twelve analyst agents produced 115 candidate findings; each module was then attacked by an independent verifier that re-derived the headline numbers from the raw records, performing 163 recomputations and filing 139 objections — 12 fatal, 57 major, 70 minor, a rate of 1.21 per finding. Five modules were recorded as unsound and seven as requiring caveats; 19 recommendations were with-

drawn. The protocol and its catalogue of named failure modes are reported as a result in their own right.

5. **A practitioner playbook with its objections attached.** The surviving associations, ranked for the sub-25K stratum by effect size against effort required, each carrying its population, baseline, n and any objection filed against it.

Methodological note

The measurement result in brief. A naive “the text contains http” flag fires on the shortened URL that X appends to every natively attached photo or video: it fired on 48.2 per cent of media posts against 11.4 per cent of text-only posts and reported outbound links as +9.9 per cent on median likes (n=8,058 flagged). Re-derived from the API's URL entities, counting only off-platform hosts, a genuine outbound link is associated with −32.7 per cent on raw medians (n=2,626; 704 versus 1,046.5 likes) and −33.8 per cent combined across follower tiers, negative in all six tiers, with $r = -0.105$ against $\log_{10}$(likes). The two estimates differ by 42.6 percentage points and disagree in sign.

1.4 Scope and non-goals

The corpus is winners-only by construction: every query carried a minimum-likes floor, so posts that failed were never sampled. This is the single most important limit on interpretation, and it is best stated in the negative. The data supports statements of the form among posts that already performed, X out-performed Y; it cannot support base rates. No sentence of the form “X per cent of threads fail” or “the probability of going viral is P” can be derived from it, and none is attempted. A negative lift here means that among successful posts, those carrying the feature were less successful than those without it — not that the feature causes failure.

The design is observational and the treatment is self-selected: authors choose their formats, their opening lines and their links. Every relationship reported is an association, and where a causal reading is tempting the confound that blocks it is named. Stratification reduces the follower confound but does not eliminate the unobserved ones — topicality, timing relative to a model release, the author's standing within the niche, and reciprocal networks among accounts.

Four further boundaries apply. The corpus covers one platform, one topic area, one language and eight months; nothing here is a law of communication. Reply posts are absent entirely — the search endpoint returned none — so reply-level dynamics are outside the study. Polls appear on seven posts and every poll comparison is suppressed under the $n < 15$ floor. And 202 posts (0.92 per cent) sit below the lowest like floor used, most plausibly because engagement counts are read at retrieval time and can fall after indexing; they are left in place, since trimming would add a second undocumented selection step.

Finally, the report is not a growth-hacking manual. It does not model the ranking algorithm, does not estimate returns to paid amplification, and takes no position on whether the tactics it measures are worth the audience they buy.

1.5 Structure of this document

Chapter 2 places the study against the literature on virality, cascade prediction and platform ranking, and against the practitioner canon it tests. Chapter 3 gives the methodology: query design, the segmentation classifier, feature definitions, the stratified-lift estimator, the bootstrap and Mann–Whitney procedures, and the Benjamini–Hochberg correction applied within each declared family [Benjamini & Hochberg, 1995]. Chapter 4 describes the corpus — coverage, segments, distributional shape and the inequality that frames every later comparison.

The results occupy three chapters, mapped to the research questions in Table 1.2. Chapter 8 reports the verification protocol and its objections, including the failure modes it caught in this study's own working. Chapter 9 discusses what the findings mean, what they cannot mean, and how to act under the residual uncertainty. Chapter 10 answers each research question in a sentence and states the limitations that remain. Chapter 11 carries the appendices: feature definitions, full result tables, the objection register and reproduction instructions.

Table 1.2 — Research questions, the population each is answered against, and where the answer appears.

RQ	Subject	Primary population	Chapter
RQ1	Format and media	Builder + mixed (n=18,692); sub-25K (n=6,521)	5
RQ2	Hook and body language	Builder + mixed (n=18,692)	6
RQ3	Audience segmentation	Builder (n=3,370) vs consumer (n=3,172)	4, 7
RQ4	Account-size stratification	Six follower tiers over n=18,692	7
RQ5	Measurement validity	Full corpus (n=21,864)	3, 8
RQ6	Small-account playbook	Sub-25K stratum (n=6,521)	9

2. Background and Related Work

This chapter situates the study in three literatures and one genre: social transmission, which asks why some content is passed on and most is not; the public record of how X ranks and distributes posts; the methodology of observational inference, which supplies the hazards the design was built against; and the practitioner "viral teardown", the origin of most operational advice on this subject and, for reasons set out below, difficult to audit. It closes with the gap this work fills and the questions it cannot answer.

2.1 What Makes Content Spread

The modern empirical account of sharing begins with affect rather than content quality. [Berger & Milkman, 2012], analysing which New York Times articles reached the most-emailed list, report that physiological arousal matters more than valence: awe, anger and anxiety travel, while sadness suppresses transmission. The causal counterpart is [Berger, 2011], where arousal induced by means unrelated to the message still increases willingness to pass it on; [Heath, Bell & Sternberg, 2001] find urban legends surviving in proportion to the disgust they provoke. The common structure is emotional selection: the transmitted population is not a random sample of the produced one, and the filter is affective.

Arousal is not the only documented driver. [Berger & Milkman, 2012] also identify practical utility as an independent positive predictor, distinct from emotion, and that is the strand most relevant here: a post a practitioner can act on is shared for a different reason than one that startles a general audience, and should leave a different metric trace. The corpus carries that trace directly. Builder posts record a median of 716 likes and 845 bookmarks (n=3,370), or 131.6 bookmarks per 100 likes; consumer chatter that merely mentions a model records 2,011 likes but 218 bookmarks (n=3,172), or 7.4 per 100 likes, the two like distributions differing at $p < 0.001$ on a two-sided Mann–Whitney U test [Mann & Whitney, 1947]. Affective reach and practical utility are, in this corpus, close to orthogonal currencies.

> **The utility–affect distinction is not a theoretical nicety here: the same corpus pays out 2.8 times more likes to affect and roughly eighteen times more saves per like to utility.**

Two further results frame what a study of this kind can observe. [Goel, Anderson, Hofman & Watts, 2016] show that most online diffusion is structurally shallow, large audiences being reached by broadcast rather than long cascades, so reach is largely a distribution outcome; [Salganik, Dodds & Watts, 2006] show experimentally that visible popularity makes outcomes both more unequal and less predictable from intrinsic quality. Any content feature measured against realised engagement is therefore measured through a

stochastic, socially reinforced amplification layer. The concentration seen here is consistent with that: within a corpus already restricted to posts that performed, the top 1% (n=219) hold 25.8% of all likes and the Gini coefficient of likes is 0.758.

2.2 What Is Publicly Known About X's Ranking, and What Is Not

In March 2023 Twitter published source code for parts of its recommendation system [Twitter, 2023]. The release established the broad shape of the pipeline — candidate sources drawn from in-network and out-of-network signals, a learned ranker scoring candidates by predicted engagement, and heuristics and filters applied afterwards — and confirmed that predicted engagement probabilities are weighted rather than treated equally, so different actions on a post contribute differently to its score.

What the release does not provide matters more here. It contained no trained weights, no training data and no full account of the visibility-filtering and product logic layered on top, and it described a system since altered repeatedly. Nothing published fixes the values in force on any given day of the collection window. The ranking function must therefore be treated as **unobservable**: every quantity reported in this paper is a realised engagement outcome, the joint product of a content decision, an audience response and a distribution policy that data of this kind cannot separate.

Methodological note

Standing interpretive constraint. Where this paper reports that a feature is associated with lower likes, it reports a difference in realised outcomes; it is not, and cannot be, a claim about the ranker. The data speak to mechanism in only one place, and indirectly: the decomposition of the outbound-link penalty finds the deficit concentrated in views (−27.0%) rather than in the per-view like rate, across 23 matched format × follower-tier cells containing 2,331 link posts. That is consistent with reduced distribution rather than reader distaste, but does not establish it.

A related caution applies to the engagement variables. Retweets, replies, quotes, bookmarks and views are co-outcomes of the same impression event, not antecedents of it; their correlations with likes in the analysis population (n=18,692) are correspondingly large — `log10(retweets)` r=0.817, `log10(views)` r=0.788 — but they describe engagement coherence rather than a lever available to a writer. Follower count is only weakly associated with likes on the log scale here (r=0.232), which reflects range restriction in a winners-only sample rather than evidence that audience size is unimportant.

2.3 Hazards in Observational Social Data

Four hazards are well established in the literature, and each appears in this corpus in a form strong enough to have reversed a conclusion during the analysis.

Survivorship. The corpus was assembled from queries carrying a minimum-likes floor, so non-performing posts were never sampled. Selection on the outcome removes base rates entirely: nothing here supports a statement of the form “X% of threads fail”, only “among posts that already performed, X out-performed Y”. Truncating the outcome from below also attenuates measured associations and makes every inequality statistic a lower bound on the inequality of the underlying posting population.

Popularity confounding and Simpson's paradox. Format choice is not independent of account size, so a pooled format comparison partly measures who uses the format [Simpson, 1951]. Stratifying on follower tier does not merely sharpen such estimates; it inverts some of them. Screen-recorded demonstrations are worth +0.4% against the pooled median of 906.5 likes (n=513) but +58.2% against the sub-25K median of 650 likes (n=201), and carousels move from +7.7% pooled (n=945) to −17.2% in the small-account stratum (n=308) — a sign reversal on three-figure cell counts on both sides.

Endogenous metrics. Ratio measures built from the same denominator as the grouping variable will move by arithmetic whether or not any effect exists. In this study a reply-ratio result that appeared decisive was tested against a 30-seed permutation placebo which destroyed time order while preserving the labelling rule; the placebo reproduced the effect at full strength and the empirical p-value rose to 0.633. The finding was withdrawn.

Measurement artefacts and duplication. Text-derived features frequently measure something other than their name. The case documented at length later is a naive “does the text contain a URL” flag, which fires on the t.co URL X appends to natively attached media and so inherits the media effect. Near-duplicate posting inflates group statistics separately: 30.0% of one trigger group's posts were near-verbatim repeats, and a single promotional template appeared 128 times across 30 handles, taking that group's reply lift from 9.74× to 2.41× once only first copies were kept.

Table 2.1 — Observational hazards, the form each took in this corpus, and the mitigation adopted.

Hazard	Instance observed here	Mitigation
Survivorship / seů lection on outcome	Every post cleared a likes floor; Gini of likes 0.758 among survivors (n=21,864)	All claims stated conditionally on having performed; no base rates reported
Popularity conů founding	Median likes 584 in the 0–1K tier (n=427) against a pooled median of 906.5	Six follower tiers; pooled and sub-25K estimates reported side by side
Simpson's paradox	Carousel +7.7% pooled (n=945) → −17.2% sub-25K (n=308)	Stratified estimate leads; pooled figure retained only for comparů ison
Endogenous ratio metrics	Reply-ratio effect reproduced by a shuffle placebo, empirical p=0.633	Denominator-confounded ratios flagged; affected claims withů drawn
Feature measureů ment artefact	Naive link flag fires on 48.2% of media posts vs 11.4% of text posts	Links re-derived from API url enů tities, off-platform hosts only
Near-duplicate content	One template posted 128 times; 30.0% dupliců ates in one trigger group	Normalised-text fingerprinting; first-copy-only cohorts
Analyst-introduced artefact	Likes-descending file order made “first post of the day” mean “best post of the day” on 100% of two-post days	Independent recomputation from the raw table by a separate agent

2.4 Practitioner Analyses and Why They Are Hard to Trust

Most published guidance on performing well on X comes from practitioner teardowns: high-performing posts are collected, features tabulated, lifts reported. The genre is not dismissed here; its hypotheses are often good, and several were productive starting points for this study. The difficulty is that such analyses are rarely auditable. Four properties recur: comparisons are pooled, leaving account size inside the estimate; features are derived from post text by simple string tests rather than from the platform's structured entities; duplicates, templates and single-account clusters go unremoved; and neither corpus nor code is published, so no reader can re-derive a number, or discover that a group has an n of 26 of which 13 are reposts of one another.

The consequence is not abstract. Table 2.2 reports what the same corpus yields under a naive pipeline and under a stratified, entity-derived one; on outbound links the two disagree by 42.6 percentage points and, more seriously, on sign. Without the underlying data, the naive figures would be indistinguishable from the corrected ones.

Table 2.2 — The same corpus under two analytical pipelines. Lifts are median likes against the stated baseline.

Question	Naive pipeline	Stratified / corrected	Baseline
Do outbound links cost reach?	+9.9% (n=8,058)	−32.7% raw, −33.8% tier-stratified (n=2,626)	996 likes, full corpus n=21,864
Do demos out-perform?	+0.4% (n=513)	+58.2% (n=201)	906.5 pooled; 650 sub-25K
Do carousels out-perform?	+7.7% (n=945)	−17.2% (n=308)	906.5 pooled; 650 sub-25K
What is a typical high-performing post worth?	996 likes (all posts)	716 builder (n=3,370) vs 2,011 consumer (n=3,172)	Segment medians

2.5 The Gap This Study Addresses

Three gaps follow. The social-transmission literature concerns news articles, legends and general audiences, and says little about a professional subpopulation whose sharing is dominated by later retrieval rather than immediate affect. The public record on X's ranking licenses no mechanistic claim, so the honest object of study is the association between an authored feature and a realised outcome within a stated stratum. The practitioner genre supplies hypotheses at volume but no means of checking them.

This study responds on three fronts. It reports estimates for a defined and disclosed population — builder and mixed posts, n=18,692, with a small-account stratum of n=6,521 — and stratifies every headline comparison on follower tier. It subjects its own analysis to adversarial verification: twelve analyst agents produced 115 findings, each module was then attacked by an independent verifier that re-derived headline numbers from the raw table across 163 recomputations, and the process returned 139 objections, 12 of them fatal; five modules were judged unsound and seven retained with mandatory caveats. And it publishes the numeric substrate — a facts pack with file hashes, library versions and a fixed seed — so that every figure in the chapters that follow traces to a value and can be recomputed.

The limits deserve equal clarity. The design is observational and winners-only; nothing here identifies a causal effect, and no reported lift is the return to adopting a feature. The corpus covers one platform, one topic domain and an eight-month window, so drift in the ranking system or the discourse is confounded with time. Multiplicity is controlled within declared families by the Benjamini–Hochberg procedure [Benjamini & Hochberg, 1995], but with hundreds of comparisons available, replication on a fresh window remains the only satisfactory test of an individual result.

3. Data and Methods

This chapter documents how the corpus was assembled, what population it represents, how every derived variable was constructed, and how the resulting claims were attacked. Two of the study's most useful results — the inversion of pooled format lifts under follower-tier stratification, and the collapse of an apparent outbound-link advantage under a corrected measure — are findings about method rather than about content.

3.1 Data source and collection

Posts were retrieved from the `twitterapi.io` advanced-search endpoint, a third-party read interface over X's search index. Every request used `queryType=Top`, which returns X's relevance-ranked result set rather than a reverse-chronological listing. Each query combined four constraints: a topic term, a minimum-favourites floor (`min_faves`), `lang:en`, and `-filter:replies`. Queries at the three lower floors were sliced into calendar months to keep each result set inside the API's practical pagination depth; the three highest floors are rare enough that one undated query bounded by `since:2026-01-01` suffices.

Table 3.1 — Dimensions of the query grid as designed, and as realised in the delivered corpus.

Dimension	Levels	Designed	Realised
Topic groups	claude, chatgpt, prompting, ai_general, grok, models, creator	7	4
Topic terms	e.g. `claude code`, `system prompt`, `ai agent`	42	25
Like floors	200, 500, 1,000, 2,500, 5,000, 10,000	6	6
Date slices	monthly, 2026-01 to 2026-08 (floors < 2,500 only)	8	8
Distinct queries	terms × floors × slices	1,134	689 executed
Queries contributing ≥1 retained post	—	—	448
Unique posts	de-duplicated by tweet id	40,000 target	21,864

The harvester paginated each query with the API's cursor to a ceiling of twelve pages of twenty posts (240 per query), stopping early when no further page was reported. Eight worker threads ran concurrently. De-duplication was applied twice, by tweet id: once in memory at write time and again at parse time. Two append-only files made the run

resumable — `data/raw.jsonl`, one unparsed payload per retained post, and `data/done_queries.txt`, one line per completed query — so an interrupted harvest resumed without re-spending credits. A hard date floor discarded anything created before 2026-01-01.

Collection did not run to plan. Of 1,134 designed queries, 689 completed before the API account exhausted its prepaid credits and every subsequent request returned `HTTP 402 Payment Required`. The harvest stopped at 21,864 unique posts against a 40,000 target, 54.7 per cent of the intended sample. Because the queue was consumed roughly in generation order, the shortfall is not spread evenly: the `grok`, `models` and `creator` groups were never reached, and 17 of 42 designed terms contribute no posts. The corpus covers four groups — `claude` (n=6,490), `ai_general` (n=5,752), `prompting` (n=5,091), `chatgpt` (n=4,531) — so no claim here concerns Gemini, Midjourney, Sora, Veo or Grok as a topic.

Winners-only design: the pipeline never observes a non-performing post, so it supports "among posts that already performed" comparisons only.

Figure 3.1 — The processing pipeline from query grid to citable facts pack. Each stage writes a durable artefact, and every number in this report resolves to the final one.

The corpus spans 2026-01-01 to 2026-08-24 (236 days), 21,864 posts and 7,829 creators, a mean of 2.79 posts per creator. Excluding replies leaves `is_reply` identically zero, so no reply-level analysis is possible and none is attempted.

3.2 Sampling frame and the bias it imposes

Every query carried a like floor. The corpus is therefore **threshold-conditioned**: a post enters only if it had already cleared 200, 500, 1,000, 2,500, 5,000 or 10,000 likes at index time. Nothing about posts that failed to reach those thresholds was observed, and no amount of downstream modelling can recover them. This is a selection-on-the-outcome design in the classical sense [Heckman, 1979], and it constrains the estimand sharply.

> **The corpus supports statements of the form "among posts that already performed, X out-performed Y". It cannot support statements of the form "X per cent of threads fail", because failing posts were never sampled.**

Three consequences follow. Every reported lift is a within-winners contrast. The inequality statistics in Chapter 4 (a Gini of 0.7584 on likes) describe inequality among winners, and are lower bounds on inequality in the unsampled population. And the relevance ranking compounds the conditioning: the frame is doubly algorithmic, filtered by an engagement floor and then ordered by a proprietary model whose broad shape, but not whose current parameters, is public [Twitter, 2023].

Methodological note

Floor integrity. If the design held perfectly, no post would sit below 200 likes. In fact 202 posts (0.924 per cent) sit below 100 and three sit at zero; the 1st percentile of likes is 106, the 5th is 214. The likely mechanism is that counts are read at retrieval time and can fall after indexing — un-likes, deletion, de-amplification. The residue was left in place rather than trimmed, because trimming would add a second undocumented selection step on the outcome variable.

Two further limits: the corpus is English-only by construction; and 602 of the 21,864 post texts are exact duplicates of another post (2.75 per cent), the residue of template re-posting — small in aggregate, but concentrated enough inside one subgroup to destroy a headline finding (Section 3.7).

3.3 Audience segmentation

The corpus mixes two populations whose mechanics have almost nothing in common: practitioners writing about building with AI, and a mass audience joking about ChatGPT. Consumer posts reach a median of 2,011 likes against 716 for builder posts, a ratio of 2.81,

while returning 218 median bookmarks against 845 — 7.4 bookmarks per 100 likes against 131.6 (Mann-Whitney U = 7,450,445.5, p below 1e-300; n=3,172 and n=3,370). Pooling them produces a third number that describes neither.

Each post was therefore scored by a rule-based classifier over four regular-expression families:

- `BUILDER_TEXT` (+2) — technical vocabulary in the body: `prompt`, `context window`, `token`, `api`, `sdk`, `mcp`, `agent`, `rag`, `embedding`, `repo`, `eval`, plus named models and tooling.
- `HOWTO_TEXT` (+2) — instructional framing: `here's how`, `step N`, `tutorial`, `template`, `steal this`, `i built`, `how to`.
- `BUILDER_BIO` (+1) — practitioner vocabulary in the author biography.
- `CONSUMER_TEXT` (−2) — mass-audience markers: `my mom`, `lmao`, `bruh`, `homework`, `horoscope`; a further −1 for the meme quote-card shape (text opening with a quotation mark and under 160 characters).

Labels follow at score ≥ 3 (builder), ≥ 1 (mixed), otherwise consumer, giving builder n=3,370 (15.4 per cent), consumer n=3,172 (14.5 per cent) and mixed n=15,322 (70.1 per cent). Actionable analysis uses builder or builder+mixed, and every table states which.

3.3.1 Error modes of the classifier

Three limitations are material and are disclosed here rather than buried.

The biography term never fired. The observed support of the score is {−3, −2, −1, 0, 1, 2, 4}: scores of 3 and 5, reachable only via the +1 biography bonus, occur zero times. The raw payloads explain why — the classifier read `author.description`, which the API returns empty, while the biography is delivered at `author.profile_bio.description`, non-empty for 1,968 of a 2,000-post probe (98.4 per cent). The classifier is therefore text-only in practice: "builder" means the post reads like practitioner how-to writing, not the author is a practitioner, and that reading should be substituted wherever the label appears.

The mixed segment is one rule wide. Of 15,322 mixed posts, 15,282 score exactly 2 — a single hit on the technical-vocabulary family and nothing else. Mixed is a large residual class defined by one regex, not a middle ground of ambiguous evidence.

Cancellation sends technical posts to consumer. Because the consumer family subtracts 2, a post matching the technical vocabulary that also carries one colloquial marker nets to zero and is labelled consumer. The 3,064 posts scoring exactly zero mix genuinely signal-free posts with these cancelled cases, and the two are not separable after the fact.

3.4 Feature engineering

Each payload was flattened into a row of 50 columns (49 published plus the tweet id). Table 3.2 groups them by provenance, the distinction that matters for interpretation: platform counts are ground truth up to retrieval-time drift, whereas every structural label is a rule-based inference made by this study, with its own error rate.

Table 3.2 — Column families by provenance. Only the first two families are platform ground truth.

Family	Representative columns	Provenance
Engagement counts	`likes`, `retweets`, `replies`, `quotes`, `bookmarks`, `views`	Platform, read at retrieval time
Account and timestamp	`followers`, `verified`, `date`, `hour_utc`, `weekday`	Platform
Derived ratios	`engagement_rate`, `bookmark_ratio`, `reply_ratio`, `amplification`, `likes_per_follower`	Arithmetic on platform counts
Media and format	`media_type`, `format`, `photos`, `gifs`, `is_poll`	Media entities + rule ladder
Structural flags	`is_thread`, `is_demo`, `is_quote`, `is_grok`, `is_engagement_bait`, `question`	Regex inference
Text shape	`chars`, `words`, `lines`, `emojis`, `hashtags`, `mentions`	Deterministic counts on text
Hook	`hook`, `hook_types`	First line + 25-pattern classifier
Audience	`audience`, `audience_score`	Regex classifier (Section 3.3)
Link	`has_link`, `ext_link`	Naive substring / URL entities (Section 3.5)
Provenance	`theme`, `term`	Retrieval path, not content classification

Media is read from the payload's media entities and reduced to video, carousel (more than one photo), image, poll or text. **Format** is a precedence ladder in fixed order — `@grok post`, poll, thread, demo, video, carousel, image, quote post, single post — so labels are mutually exclusive by construction and the ordering, not the data, decides which one a post carrying two signals receives. A post is a thread if a thread marker appears or if the text exceeds 900 characters, so "thread" is partly a length variable; a post is a demo only if a build-or-show phrase matches and media is attached, which is why demo is small (n=525) and comparatively clean.

Hooks are the first non-empty line, truncated at 220 characters, labelled by 25 non-exclusive patterns plus a `Plain statement` fallback. The fallback dominates — 15,303 posts carry it — so hook-type comparisons run against a large heterogeneous residual, and because a post may carry several labels the groups overlap and their lifts are not additive.

Theme and term record which query first returned the post, not what the post is about. The harvester was concurrent and kept the first retrieval of any id, so theme-level contrasts are contrasts between retrieval paths and are reported as such. Two quirks carry forward: `is_demo` is stored as a count of demo signals (0–4), binarised at >0 here, and `is_poll` fires on seven posts, so every poll cut is suppressed under the rule in Section 3.6.

3.5 The `ext_link` correction: a worked case in construct validity

The corpus originally carried `has_link`, defined as `"http" in text`. It is a textbook construct-validity failure: the operational definition and the intended construct come apart, and the measure records something else entirely [Cronbach & Meehl, 1955]. X appends a `t.co` URL to the body of every natively attached photo or video, so a substring test for `http` fires on the attachment itself. Here it fired on 48.2 per cent of media posts against 11.4 per cent of text posts, flagged 8,058 posts (36.9 per cent of the corpus), and reported linked posts at a median of 1,057 likes against 962 — an apparent **+9.9 per cent** advantage to linking (p = 1e-7). The flag was measuring media attachment, and inheriting the media lift.

The corrected column, `ext_link`, is re-derived from the API's URL entities: a link counts only when an expanded URL resolves to a host that is not `x.com`, `twitter.com` or `t.co`. Genuine outbound links are far rarer than the naive flag suggested — 2,626 posts, 12.0 per cent — and the sign reverses. On raw medians, linked posts return 704 likes against 1,046.5, a lift of **−32.7 per cent** (p below double-precision underflow, 1e-308); Pearson r against log10(likes) is −0.1049. In the actionable builder+mixed population the penalty is −26.7 per cent raw and −27.9 per cent tier-stratified (n=2,354 of 18,692); in the sub-25K stratum, −22.5 per cent raw and −22.8 per cent stratified (n=709 of 6,521). The naive-to-corrected swing is 42.6 percentage points and crosses zero.

The penalty is not an artefact of account size: the ratio of linked to unlinked median likes is below one in all six follower tiers without exception, from 0.43 in the 0–1K tier to 0.74 in the 5K–25K tier. The tier-weighted combination (−33.8 per cent on the full corpus) depends on how those six ratios are pooled — an unweighted mean gives −39.4 per cent — so the honest statement is an interval rather than a point estimate: a like penalty of roughly a quarter to two-fifths, negative in every tier. An independent module matching linked and unlinked posts inside 23 format × tier cells (2,331 linked posts) recovers −22.0 per cent on likes and −27.0 per cent on views, with bookmarks moving the other way at +55.9 per cent. A penalty larger in views than in per-view like rate is consistent with distribution suppression rather than with readers disliking linked posts, though the design cannot separate the two.

Methodological note

The naive column is retained in the published table under its original name for continuity and audit, and is never used for inference. Every link result in this paper is computed on `ext_link`.

3.6 Statistical methods

Medians, not means. Every engagement variable is heavily right-skewed: across the full corpus the mean of likes is 4,086.5 against a median of 996, a ratio of 4.1. A mean in this distribution is a statement about the largest handful of posts. All central tendencies here are medians and all dispersion is reported as percentiles.

Lift is the percentage difference of a group median against a named baseline median, whose value and n are always stated: `L_raw = median(present) / median(absent) − 1`.

Follower-tier stratified lift. Because format and length choices correlate with account size, a pooled contrast partly measures who adopts the format. Stratified lift is a follower-tier-weighted geometric mean of within-tier median ratios across the six tiers (0–1K, 1K–5K, 5K–25K, 25K–100K, 100K–1M, 1M+):

`L_strat = exp( Σ_t w_t · ln( m_t,present / m_t,absent ) ) − 1`, with `w_t` proportional to the number of feature-present posts in tier t, and any tier carrying fewer than 15 posts on either side dropped.

This matters materially. Pooled and stratified estimates disagree for several formats: demo is +0.4 per cent pooled (n=513) but +58.2 per cent in the sub-25K stratum (n=201), while carousel is +7.7 per cent pooled (n=945) and −17.2 per cent at sub-25K (n=308). The mechanism is Simpson's paradox operating on follower count [Simpson, 1951]. Where the two disagree, this paper reports both.

Interval estimates on median ratios are percentile bootstrap intervals, 1,000 resamples, seed 20260824 [Efron, 1979]. **Tests** are two-sided Mann-Whitney U on untransformed likes distributions, assuming neither normality nor equal variance [Mann & Whitney, 1947]. **Multiplicity** is handled within each declared family — for instance the 15 binary content features scored in one population — by Benjamini-Hochberg FDR at q = 0.05 [Benjamini & Hochberg, 1995]; 13 of 15 features survive adjustment in builder+mixed, 12 of 15 in the sub-25K stratum. **Correlations** are Pearson r on log10(likes), likes floored at 1, with the standing caution that bookmarks, replies, retweets, quotes and views are co-outcomes of the same impression event: their large coefficients describe engagement coherence, not a lever an author can pull.

Suppression. Any group with fewer than 15 posts is suppressed and the suppression noted; the rule applies inside stratification cells as well as at the top level, which is why some stratified estimates rest on five tiers rather than six. Analyses ran in Python 3.13.5 with NumPy 2.4.2, pandas 3.0.1 and SciPy 1.17.1, and the facts pack behind every number in this report is regenerated by a single script from the 28.8 MB post table.

3.7 The verification protocol

Findings were produced under an adversarial two-role architecture. Twelve analyst agents each worked one thematic module over the same post table — hook linguistics, timing and cadence, engagement funnel, psychological triggers, amplification network, content pillars, creator archetypes, growth trajectories, repurposing, brand share of voice, whitespace matrix, antipatterns — returning claims with evidence, interpretation and a self-declared confidence. Each module then went to an independent verifier agent with one instruction: re-derive the headline numbers directly from `data/viral_posts.csv` and file objections graded fatal, major or minor.

Verifiers were directed at five attack surfaces: arithmetic that does not reproduce; circularity; selection and sort-order artefacts; confounds with account size; and subgroups whose lift rests on duplicated text. They produced 163 independent recomputations and 139 objections against 115 analyst findings — 1.21 per finding — graded 12 fatal, 57 major and 70 minor. Five modules were returned **unsound** and seven **needs caveat**; none passed unchallenged. The strategist stage that consumed the survivors killed a further 19 recommendations and adjudicated nine contradictions between modules.

Table 3.3 — Analysis pipeline stages and the artefact each writes.

Stage	Operation	Artefact	Scale
1. Harvest	448 contributing queries, cursor pagination, id de-duplication	`data/raw.jsonl`	21,864 payloads
2. Flatten	Feature extraction, classification, second de-duplication	`data/viral_posts.csv`	21,864 × 50
3. Aggregate	Headline cuts by format, media, tier, theme, hour, month	`data/analysis.json`	builder segment
4. Analyse	12 analyst modules, each on the full post table	`data/modules/*.json`	115 findings
5. Verify	12 independent verifiers re-deriving from the CSV	`workflow_results.json`	139 objections
6. Synthesise	Contradiction adjudication, recommendation triage	`strategy_synthesis.json`	19 killed
7. Fact-pack	Recomputation of every citable number with CIs and tests	`data/facts.json`	one source

Five failure modes recurred often enough to be named, and they are the practical yield of the protocol. **Sort-order tautology**: the post table is stored sorted by likes descending, so a stable sort within a calendar day ordered posts by performance — "first post of the day" silently meant "best post of the day". **Copy-paste spam artefact**: roughly 30 per cent of the posts carrying one psychological trigger were a single duplicated template re-posted by many accounts, so the trigger's lift measured one text's distribution. **Circular metric**: a ratio reported as an outcome while sharing its denominator with the variable defining the groups, which fixes the result's direction independently of the data. **Pooled-tier confound**: the Simpson's-paradox mechanism of Section 3.6. **Measurement artefact**: the `has_link` case of Section 3.5. A sixth, narrower class — regex self-overlap and missing word boundaries, such as a money-vocabulary pattern matching inside "learn" and "yearn" — was caught twice, and is a reminder that rule-based features need unit tests of their own.

Methodological note

How disputed findings are handled in this report. A claim that a verifier falsified is dropped and does not appear. A claim that survived with a material caveat appears together with the objection, in the same paragraph, never in a footnote. A withdrawn claim is never restated as settled. Where analysts disagreed and the adjudication rests on judgement rather than on a decisive recomputation, both positions are given.

Two limits of the protocol should be stated plainly. The verifiers audited internal validity — arithmetic, circularity, confounding, artefacts — and could not audit external validity, because the sampling frame binds every agent equally. And the protocol is adversarial, not exhaustive: an objection rate above one per finding shows the attack surface was large, not that it was cleared. The claims that follow survived a determined attempt to break them, which is a weaker and more useful guarantee than correctness.

4. The Corpus

This chapter describes the material the rest of the study rests on: how much of it there is, how the engagement it records is distributed, how it divides into audiences, what it is composed of, and where its coverage is uneven. Nothing here explains performance. The purpose is to fix the shape of the sample precisely enough that later comparisons can be read against it, and to make visible the several respects in which the corpus is a manufactured object rather than a slice of X as it is.

4.1 Scale and coverage

The corpus contains 21,864 posts published between 2026-01-01 and 2026-08-24, a window of 236 days, retrieved through the twitterapi.io advanced-search endpoint. The retrieval design crossed 42 topic terms with six `min_faves` floors (200, 500, 1,000, 2,500, 5,000 and 10,000) and, for the three lower floors, monthly date slices, yielding a grid of 1,134 distinct queries. Collection stopped when the API credit was exhausted, so 689 of those queries were executed and 448 of the 689 returned at least one post that was new to the corpus at the moment it arrived; the delivered corpus therefore carries 25 distinct terms rather than 42. Appendix A records the grid as designed, as executed and as realised. After de-duplication on post id, the surviving posts were flattened to a table of 50 fields — 49 published variables plus the identifier — covering author, timing, engagement counts, derived ratios, format and media labels, an audience classification, and the post text.

Those 21,864 posts were written by 7,829 distinct accounts, a mean of 2.79 posts per account. The distribution of authorship is itself uneven: 5,298 accounts (67.7% of all accounts) appear exactly once, while the most prolific single account contributes 144 posts. Verified accounts supply 90.9% of posts and the median author has 49,713 followers — a consequence of the like floors rather than a fact about X, since an unverified account with a thousand followers rarely clears a 200-like threshold.

Table 4.1 — The corpus at a glance. Engagement totals are sums over all 21,864 posts as counted at retrieval time.

Quantity	Value
Posts	21,864
Distinct authors	7,829
Date span	2026-01-01 to 2026-08-24 (236 days)
Queries contributing at least one retained post	448 (of 689 executed, 1,134 designed)
Topic terms represented in the corpus	25 (of 42 designed)
Fields per post	50
Total likes	89,346,465
Total views	12,092,549,651
Total bookmarks	30,219,146
Total reposts	10,373,537
Total replies	4,222,909
Total quotes	1,534,542
Median likes	996
Median views	97,632
Median author followers	49,713
Verified share of posts	90.9%

Methodological note

Floor integrity. Every query carried a like floor, so in principle no post should sit below 200 likes. In practice 202 posts (0.92% of the corpus) fall below 100 likes and 39 below ten, with a minimum observed value of zero; the first and fifth percentiles of likes are 106 and 214 respectively. The most economical explanation is that engagement counts are read at retrieval time and can fall after a post is indexed, through un-likes, deletion or de-amplification, so a post may satisfy a floor when indexed and fail it when collected. The residue is left in place rather than trimmed, because trimming would introduce a second, undocumented selection step. It should be recalled wherever the lower tail of the likes distribution is discussed.

4.2 The engagement distribution

Engagement in this corpus is extremely top-heavy, and it is top-heavy after the like floors have already removed the entire lower tail of the posting population. The Gini coefficient of likes across all 21,864 posts is 0.758; for views it is 0.811 and for bookmarks 0.746. The mean number of likes is 4,086.5 against a median of 996 — a mean-to-median ratio of 4.1 — and only 18.2% of posts exceed the mean. Half of all likes in the corpus are held by 936 posts, which is 4.3% of it.

Table 4.2 — Concentration of likes, full corpus (n=21,864, 89,346,465 likes). Slices are cumulative from the top of the likes ranking.

Slice	Posts	Share of all likes	Likes at slice boundary
Top 0.1%	22	6.2%	177,223
Top 1%	219	25.8%	56,610
Top 5%	1,093	53.1%	16,266
Top 10%	2,186	67.1%	8,235
Top 25%	5,466	84.2%	2,710
Top 50%	10,932	94.2%	996

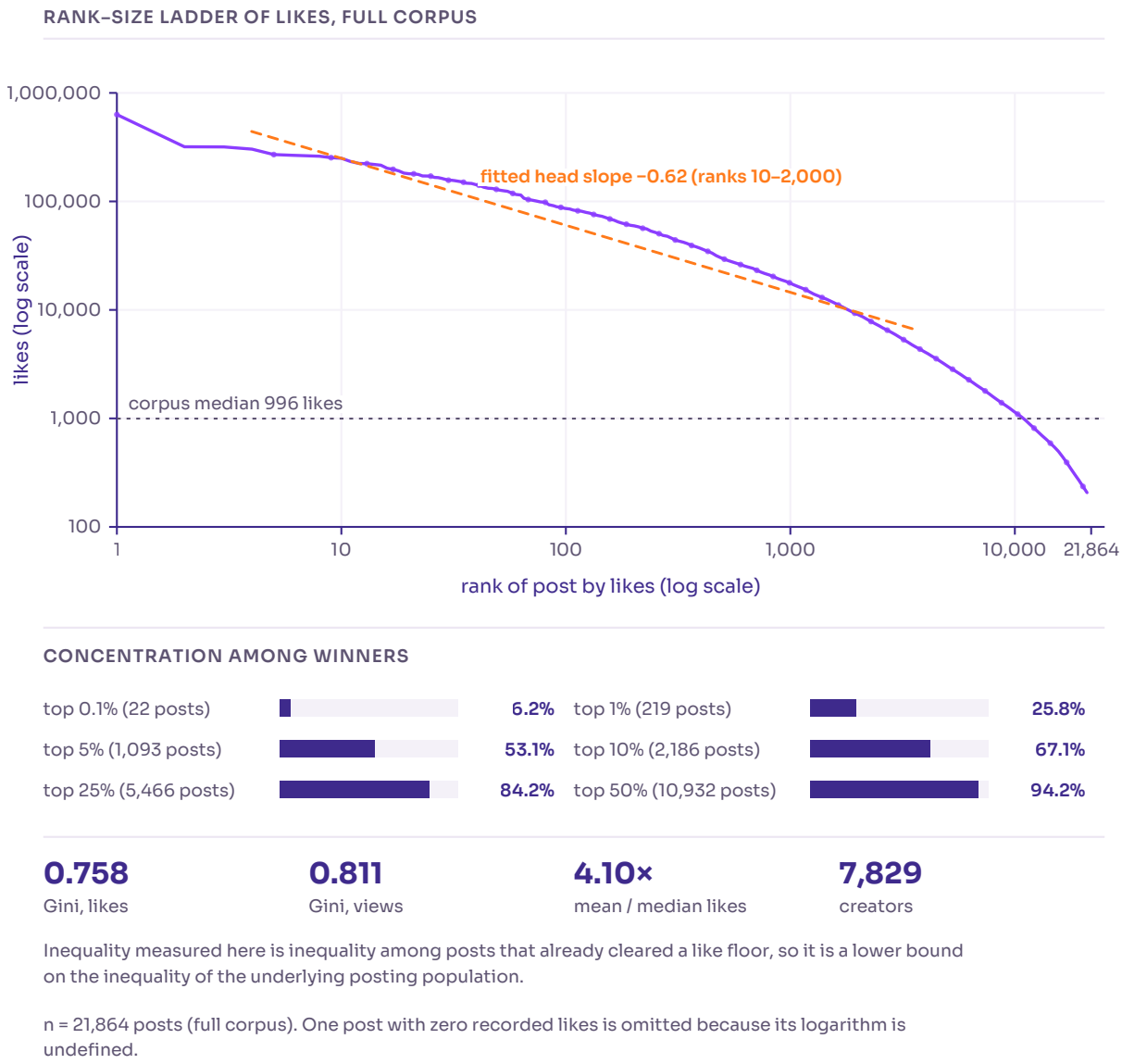


Figure 4.1 — Rank-likes ladder on log-log axes, full corpus. The head of the curve is approximately linear; over ranks 10 to 2,000 the ordinary-least-squares slope is −0.618.

Figure 4.1 plots the ladder. The head is close to linear on log-log axes, with a slope of −0.618 over ranks 10 to 2,000, which is the visual signature usually described as power-law-like. We do not fit or claim a power law. A straight-looking segment on a log-log rank plot is weak evidence for the family, and the corpus is truncated from below by construction, so any tail exponent estimated here would describe the sampling frame as much as the phenomenon [Clauset, Shalizi & Newman, 2009]. The descriptive point survives the caution: outcomes span roughly three orders of magnitude among posts that all cleared the same admission bar.

The same concentration appears at account level. Across the 7,829 authors, the Gini coefficient of likes per author is 0.805; the top 1% of authors hold 28.2% of all likes and the top 10% hold 70.9%. Two mechanisms are confounded there — some accounts post more often, some reach further per post — and this chapter does not separate them.

> **The distribution has a practical reading for anyone planning content: the mean is not a target. Fewer than one post in five reaches it, and a plan whose arithmetic depends on average performance is a plan that depends on a rare event.**

The planning implication is that the median describes the outcome of a typical attempt and variance is the dominant term: a programme of posts should be sized so that its value does not depend on any single post landing in the upper decile, where two thirds of the total sits. Excluding consumer chatter reduces the inequality but does not remove it — within the analysis population of builder and mixed posts (n=18,692) the Gini of likes is 0.707 and the top 1% still hold 21.1% of likes.

4.3 The three audience segments

A keyword-and-bio classifier assigns every post to one of three audiences: builder (practitioners building with AI), consumer (mass-audience chatter that merely mentions a model by name), and mixed (neither clearly one nor the other). The mixed class is a residual rather than a substantive category, and at 70.1% of the corpus it is the largest; it is retained because excluding it would discard most of the data, but no claim rests on mixed being a coherent audience.

Table 4.3 — The three audience segments. Medians throughout; bookmarks per 100 likes is the median of the per-post ratio.

Segment	Posts	Authors	Median followers	Median likes	Median views	Median bookmarks	Bookmarks / 100 likes	Gini (likes)
Builder	3,370	1,421	45,833	716	81,423	845	131.6	0.624
Consumer	3,172	2,149	51,027	2,011	120,409	218	7.4	0.776
Mixed	15,322	5,811	50,624	964	98,018	330	29.0	0.713
All	21,864	7,829	49,713	996	97,632	375	31.1	0.758

The segments diverge sharply, and they diverge in opposite directions on the two outcomes. Consumer posts reach a median of 2,011 likes against 716 for builder posts, a ratio of 2.81 (n=3,172 and n=3,370; two-sided Mann-Whitney U = 7,450,445.5, p below double-precision underflow). On bookmarks the ordering reverses: builder posts reach a median of 845 saves against 218 for consumer posts, a ratio of 0.26 in the consumer direction. Normal-

ised, the gap is starker still — 131.6 bookmarks per 100 likes for builder posts against 7.4 for consumer posts, a factor of roughly 17.7. Median engagement rate, by contrast, is identical at 0.0243 in both segments, which is a reminder that a ratio can be preserved while its components move apart.

The plain statement is that consumer posts win reach and lose retention. They travel to more people and are saved by almost none of them; builder posts travel less far and are filed away by a substantial fraction of the people who touch them. Because the segments differ by nearly a factor of three on the headline metric, any pooled comparison that mixes them is partly a measurement of segment composition. Every actionable comparison in the chapters that follow is therefore computed on the builder segment or on the builder-and-mixed analysis population, and states which.

4.4 Composition

The composition tables below are reported on the analysis population (builder and mixed, n=18,692, median 906.5 likes and 408 bookmarks), which is the population against which the lifts in later chapters are computed. Two cautions apply throughout. The share column describes the corpus, not X: the mixture of formats and themes here was produced by the query design and the like floors. And share and median are independent — the most abundant format in the corpus is also the weakest performer in it.

Table 4.4 — Format composition, analysis population (n=18,692). Lift is against the population median of 906.5 likes. Polls (n=5) are suppressed under the n<15 rule.

Format	Posts	Share	Median likes	Lift	Median bookmarks	Median followers
Thread	6,454	34.5%	727	-19.8%	527	44,316
Image	3,697	19.8%	817	-9.9%	251	56,444
Video	3,344	17.9%	1,215	+34.0%	670	44,667
Single post	2,606	13.9%	1,155	+27.4%	229	68,930
Quote post	1,080	5.8%	1,148	+26.6%	202	71,917
Carousel	945	5.1%	976	+7.7%	256	73,357
Demo	513	2.7%	910	+0.4%	1,151	30,567
@grok post	48	0.3%	807	-11.0%	181	227,557
Poll	5	0.03%	—	—	—	—

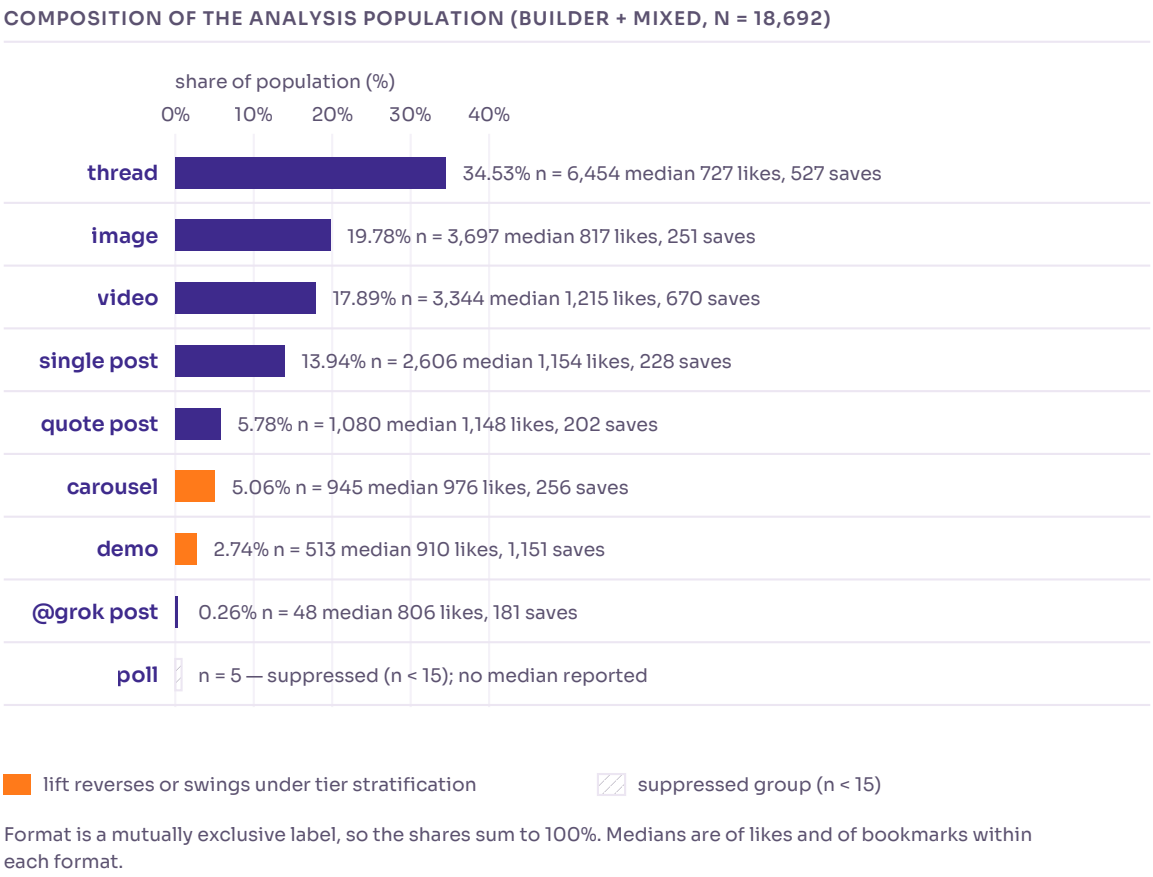


Figure 4.2 — Format share of the analysis population against median-likes lift. Supply and performance are close to inversely ordered at the top of the chart.

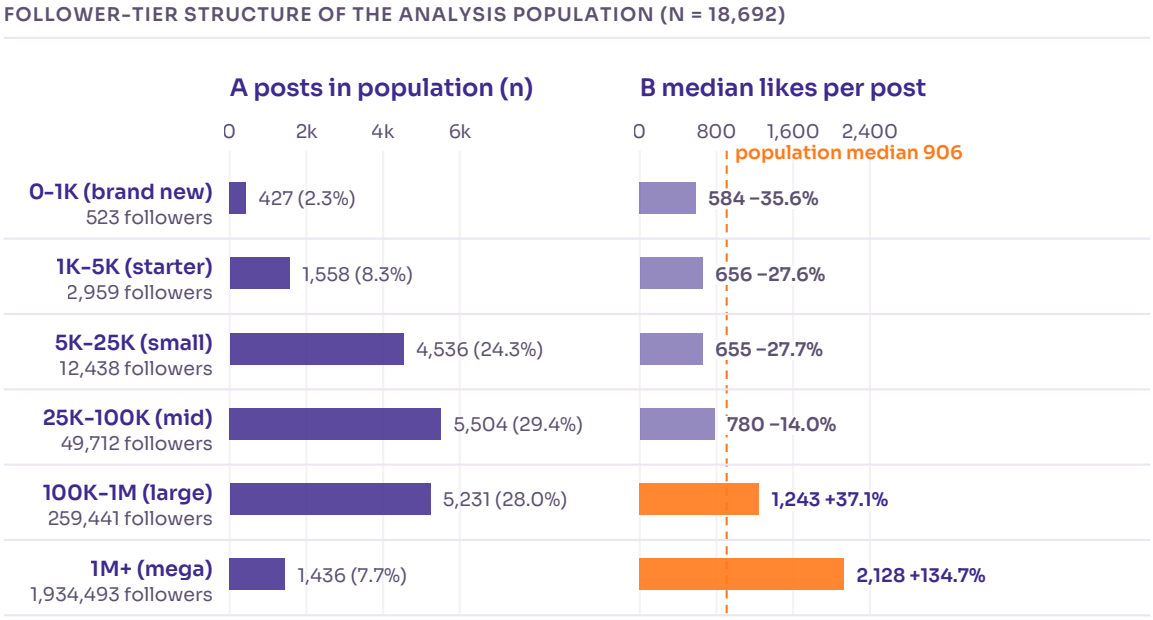
Figure 4.2 places supply against performance. Threads are a third of the analysis population and sit 19.8% below the baseline on likes, while sitting 29.2% above it on bookmarks — the format is abundant and is being selected for a different outcome than likes. The demo row is the clearest warning against reading this table alone: its pooled lift of +0.4% is a null, but demo authors have a median of 30,567 followers against 49,713 for the population, and once the small-account stratum is isolated the same format reads +58.2% (n=201). That inversion is treated in full in the stratification analysis; here it is enough to note that the median-followers column is not decoration.

Table 4.5 — Media-type composition, analysis population (n=18,692). Polls (n=5) suppressed.

Media type	Posts	Share	Median likes	Lift	Bookmarks / 100 likes
Image	6,637	35.5%	724	-20.1%	40.8
Text only	5,265	28.2%	1,083	+19.5%	21.9
Video	5,163	27.6%	1,050	+15.8%	81.9
Carousel	1,622	8.7%	806	-11.1%	38.9

Table 4.6 — Follower-tier composition, analysis population (n=18,692). Tiers are assigned from author follower count at retrieval time.

Follower tier	Posts	Share	Median likes	Lift	Median bookmarks
0–1K (brand new)	427	2.3%	584	-35.6%	182
1K–5K (starter)	1,558	8.3%	656	-27.6%	329
5K–25K (small)	4,536	24.3%	655	-27.7%	380
25K–100K (mid)	5,504	29.5%	780	-14.0%	478
100K–1M (large)	5,231	28.0%	1,243	+37.1%	462
1M+ (mega)	1,436	7.7%	2,128	+134.7%	256



Lift in panel B is against the population median of 906 likes; every tier differs from it at $p < 0.01$ (two-sided Mann–Whitney U). The second line of each row label is that tier's median follower count.

The two smallest tiers are the least represented and the worst performing, so any pooled statistic is weighted towards accounts above 25K followers.

Figure 4.3 — Distribution of the analysis population across follower tiers, with median likes per tier. The sub-25K stratum holds 34.9% of the population.

The tier profile in Figure 4.3 matters mainly because it determines what the stratified analyses can support. The three tiers below 25,000 followers hold 6,521 posts, 34.9% of the analysis population — a sufficient stratum for format-level comparison but not for fine sub-

division within it. Median likes are flat across those three tiers before rising, and the flatness should not be over-read: it is produced partly by the like floors, which admit a small account only when it has already done unusually well.

Table 4.7 — Theme composition, analysis population (n=18,692). Themes are derived from the query term that retrieved the post, so shares reflect the retrieval design.

Theme	Posts	Share	Median likes	Lift	Bookmarks / 100 likes
Claude	6,353	34.0%	1,420	+56.6%	30.7
Prompting	4,929	26.4%	760	-16.2%	66.1
AI general	4,357	23.3%	755	-16.7%	43.2
ChatGPT	3,053	16.3%	661	-27.1%	42.4

Methodological note

Composition is manufactured. None of the share columns in Tables 4.4 to 4.7 estimates a population share. They describe a corpus assembled from 448 queries built on 25 chosen terms and six like floors. That a third of the corpus carries the Claude theme is a property of the term list; that a third of it is threads is a property of what clears a like floor in this topic area. Shares are reported so that the reader can judge the weight behind each median, not as a description of what people post.

4.5 Temporal coverage and its gaps

Day-level coverage is complete: all 236 days in the window contain at least one post, with a median of 85 posts per day, a maximum of 335, and a single thin day at 36 posts (2026-08-03). There are no missing days to interpolate over. The unevenness is at the month level, and it is substantial in one direction.

Table 4.8 — Monthly coverage, full corpus (n=21,864). August 2026 is truncated at the 24th. Lift is against the full-corpus median of 996 likes.

Month	Posts	Share	Days	Posts / day	Median likes	Lift
2026-01	2,410	11.0%	31	77.7	1,085	+8.9%
2026-02	2,271	10.4%	28	81.1	1,321	+32.6%
2026-03	2,645	12.1%	31	85.3	1,279	+28.4%
2026-04	2,726	12.5%	30	90.9	1,058	+6.2%
2026-05	2,687	12.3%	31	86.7	937	-5.9%
2026-06	2,658	12.2%	30	88.6	967	-2.9%
2026-07	2,861	13.1%	31	92.3	970	-2.6%
2026-08	3,606	16.5%	24	150.2	687	-31.1%

August is over-represented and under-performing at the same time. It contributes 16.5% of the corpus from 24 days, at 150.2 posts per day against 77.7 in January, and its median of 687 likes is 31.1% below the corpus median. Both facts are most plausibly artefacts of the same mechanism. Recent posts are more completely indexed and more readily returned by the search endpoint, which inflates their count; and engagement counts were read at a single retrieval time, so a post from August had days to accumulate likes where a post from January had months. Cross-month comparisons of engagement level therefore confound content effects with the maturation of engagement, and no such comparison is made in this study.

The consequence for the temporal analysis is a deliberate restriction. Drift over the window is read only from indices normalised against the all-format median inside the same half of the window — H1 as 2026-01 to 2026-04 and H2 as 2026-05 to 2026-08 — so that the maturation term divides out. A second, quieter gap is that the window covers eight months of one topic area during a period of rapid model releases; the corpus can describe what varied within those months but cannot establish that any of it holds outside them.

Two further limits bound every result that follows. The corpus contains no unsuccessful posts, so it supports statements of the form "among posts that already performed, X outperformed Y" and none of the form "X% of threads fail". And because format, theme, follower tier and audience are mutually correlated in the tables above, a difference attributed to one may belong to another; the stratified analyses in the following chapters exist because pooled composition, left unexamined, produces reversals of the kind Simpson described [Simpson, 1951].

5. Format, Media and Length

This first results chapter reports what the corpus shows about the container a post arrives in: its format, its attached media, and its length. It is organised around a methodological result the study treats as a finding in its own right — that a format comparison computed over the whole population and the same comparison computed inside a follower-size stratum can disagree, and in two cases disagree in sign. Everything downstream of Section 5.2 is reported twice for that reason.

Methodological note

Populations used in this chapter. Pooled means the analysis population, audience in {builder, mixed}, n = 18,692; consumer chatter is excluded because it reaches roughly 2.8 times the median likes of builder posts (2,011 against 716) on a fraction of the saves, and pooling it swamps every practitioner-relevant contrast. Sub-25K means the same population restricted to authors with fewer than 25,000 followers, n = 6,521. Baseline medians are the population medians: 906.5 likes and 408 bookmarks pooled, 650 likes and 347 bookmarks at sub-25K. All lifts are percentage differences of group medians against those baselines, all tests are two-sided Mann–Whitney U with Benjamini–Hochberg control within the declared family [Benjamini & Hochberg, 1995], and every group with n < 15 is suppressed. The corpus is winners-only: every claim has the form among posts that already performed, X outperformed Y.

5.1 Format performance, pooled

The `format` column assigns each post to exactly one container, with media and thread structure resolved by precedence. It is not the same variable as `media_type`: a thread carrying a video is `format = thread` and `media_type = video`, which is why the 6,454 pooled threads decompose across 2,811 image, 1,568 text, 1,417 video and 658 carousel posts, and why the two tables in this chapter give different counts for the same word. Table 5.1 gives the pooled ranking.

Table 5.1 — Format performance, pooled analysis population (builder + mixed, n = 18,692). Baseline: population median 906.5 likes, 408 bookmarks, p90 6,209 likes. Saves/100 likes is the median of the per-post bookmarks-to-likes ratio within the group, whose population value is 42.4; it is not the ratio of the two median columns beside it.

Format	n	Share %	Median likes	Lift %	p90 likes	Saves/100 likes	Median followers	p
Video	3,344	17.9	1,215	+34.0	9,720	58.1	44,667	<1e-308
Single post	2,606	13.9	1,155	+27.4	8,401	13.4	68,930	<1e-308
Quote post	1,080	5.8	1,148	+26.6	7,837	13.3	71,917	3.1e-06
Carousel	945	5.1	976	+7.7	8,202	19.1	73,357	0.008
Demo	513	2.7	910	+0.4	4,631	137.9	30,567	0.841
Image	3,697	19.8	817	−9.9	5,555	21.1	56,444	2.8e-05
@grok post	48	0.3	807	−11.0	4,168	19.4	227,557	0.834
Thread	6,454	34.5	727	−19.8	3,853	76.3	44,316	<1e-308
Poll	5	0.0	—	—	—	—	—	—

Three features of the table matter more than the ordering. First, the largest supply lane is the worst-paying one on likes: threads are 34.5 per cent of the population and sit 19.8 per cent below the population median, a gap that is not marginal ($p < 1e\text{-}308$). Second, likes and saves rank the formats close to inversely. The demo — a post showing a tool being operated — is indistinguishable from baseline on likes (+0.4 per cent, $p = 0.841$) while returning 137.9 bookmarks per 100 likes against a population figure of 42.4; the quote post is 26.6 per cent above baseline on likes and returns 13.3. Third, the p90 column shows that format governs the ceiling as much as the centre: video's p90 of 9,720 is 2.5 times the thread's 3,853 on a baseline p90 of 6,209, so the format difference is not confined to the median post.

Two rows are unstable rather than informative. The poll cut ($n = 5$) is suppressed under the $n < 15$ floor and appears only so that its absence is visible; the @grok row ($n = 48$) clears the floor but fails significance ($p = 0.834$) on a median follower count five times the population's.

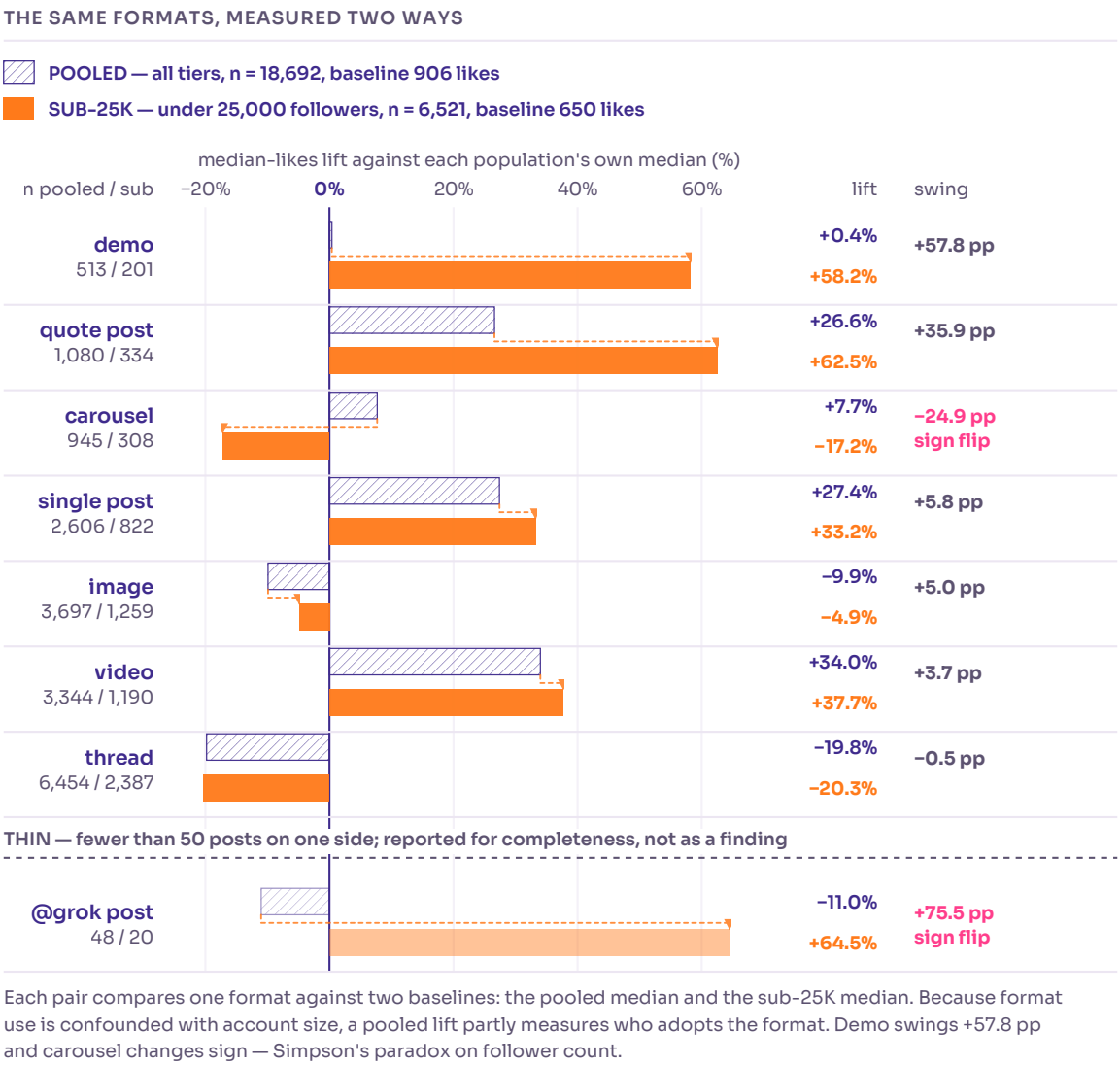
5.2 The stratification reversal

Format choice is not independent of account size. The median carousel in Table 5.1 was written by an account with 73,357 followers, the highest of any non-novelty format; the median demo by an account with 30,567, the lowest. A pooled comparison between those two rows therefore compares two different populations of authors as much as two containers. Table 5.2 recomputes the same table inside the sub-25K stratum.

Table 5.2 — The same format comparison inside the sub-25K stratum (builder + mixed, followers < 25,000, n = 6,521). Baseline: 650 likes, 347 bookmarks, p90 4,667 likes.

Format	n	Median likes	Sub-25K lift %	Pooled lift %	Swing (pp)	p
Quote post	334	1,056	+62.5	+26.6	+35.9	<1e-308
Demo	201	1,028	+58.2	+0.4	+57.8	0.0002
Video	1,190	895	+37.7	+34.0	+3.7	<1e-308
Single post	822	866	+33.2	+27.4	+5.8	<1e-308
Image	1,259	618	−4.9	−9.9	+5.0	0.158
Carousel	308	538	−17.2	+7.7	−24.9	0.053
Thread	2,387	518	−20.3	−19.8	−0.5	<1e-308
@grok post	20	1,070	+64.5	−11.0	+75.5	0.028

Two rows change sign. The carousel moves from +7.7 per cent pooled to −17.2 per cent at sub-25K, a swing of 24.9 percentage points. The demo moves from +0.4 per cent — a null — to +58.2 per cent, a swing of 57.8 points, which converts the format from the study's least interesting row into one of its most interesting. The @grok row flips too, but on n = 20 and is flagged thin; it is shown for completeness and carries no weight. Other rows barely move: video swings 3.7 points, thread 0.5. The reversal is specific, not general.



Each pair compares one format against two baselines: the pooled median and the sub-25K median. Because format use is confounded with account size, a pooled lift partly measures who adopts the format. Demo swings +57.8 pp and carousel changes sign — Simpson's paradox on follower count.

Figure 5.1 — Pooled versus sub-25K likes lift by format. Carousel and demo cross zero in opposite directions; video, thread, single post and image hold their sign.

Table 5.3 shows the mechanism directly, by computing each format against the median of all other formats inside the same follower tier. The demo is not a null; it is a strong positive for small accounts (+123.7 per cent at 1K–5K, +52.4 per cent at 5K–25K) and a mild negative above 25K, and the pooled +0.4 per cent is those opposed effects cancelling. The carousel is the sharper case: it is 29.1 per cent below its tier at 5K–25K and 27.7 per cent above it at 100K–1M, and because 42.5 per cent of carousels come from accounts above 100K followers, the pooled figure inherits the large-account result.

Table 5.3 — Within-tier likes lift by follower tier, pooled analysis population. Each cell compares the format's median likes against the median of all other formats in the same tier. Cells with n < 15 are suppressed.

Follower tier	Demo	Carousel	Video	Thread
0–1K	— (n=11)	+57.3 (n=25)	−0.2 (n=72)	−50.8 (n=127)
1K–5K	+123.7 (n=50)	+2.4 (n=73)	+33.3 (n=271)	−35.9 (n=557)
5K–25K	+52.4 (n=140)	−29.1 (n=210)	+62.0 (n=847)	−31.3 (n=1,703)
25K–100K	−6.9 (n=220)	+9.7 (n=235)	+43.0 (n=954)	−22.8 (n=2,108)
100K–1M	−7.5 (n=87)	+27.7 (n=259)	+20.8 (n=876)	−9.3 (n=1,714)
1M+	— (n=5)	+3.3 (n=143)	+22.0 (n=324)	+19.4 (n=245)

> **Any pooled format claim is partly a claim about who uses the format. Where adoption is correlated with account size, the pooled number is a weighted average over author populations, not a property of the container.**

This is Simpson's paradox operating on an observational content corpus [Simpson, 1951]. It was also raised independently during verification: one verifier re-derived the carousel result and objected that its pooled deficit on saves was "largely 'carousels skew to 100K–1M and 1M+ accounts'", noting the same median follower count of 73,357, while conceding that the overall format ordering survives stratification. That concession is important and is retained here: stratification changes two rows out of eight, and leaves the top and bottom of the ranking — video high, thread low — intact.

Methodological note

Robustness of the demo reversal. Duplicate templates are a live hazard in this corpus; the verification pass found one trigger whose lift was 30 per cent driven by a single re-posted script. The 201 sub-25K demos come from 153 distinct handles, and 12.4 per cent of them are repeat copies of another post's opening 120 characters. Collapsing each duplicated text to its first copy leaves n = 176 and a median of 1,085.5 likes, a lift of +67.0 per cent against the sub-25K baseline of 650 — higher, not lower, than the un-deduplicated +58.2 per cent. The reversal is not a copy-paste artifact.

5.3 Media type and the currency it buys

Read against `media_type`, which classifies the attachment rather than the container, the ordering on likes is text, video, carousel, image (Table 5.4), and it holds its sign in the sub-25K stratum: text +26.8 per cent, video +20.5, carousel −20.4, image −21.1. No media row reverses. What changes across media is not rank but currency.

Table 5.4 — Media type, pooled analysis population (n = 18,692). Two conventions are mixed here and the difference matters: saves per 100 likes is the median of the per-post bookmarks-to-likes ratio (population

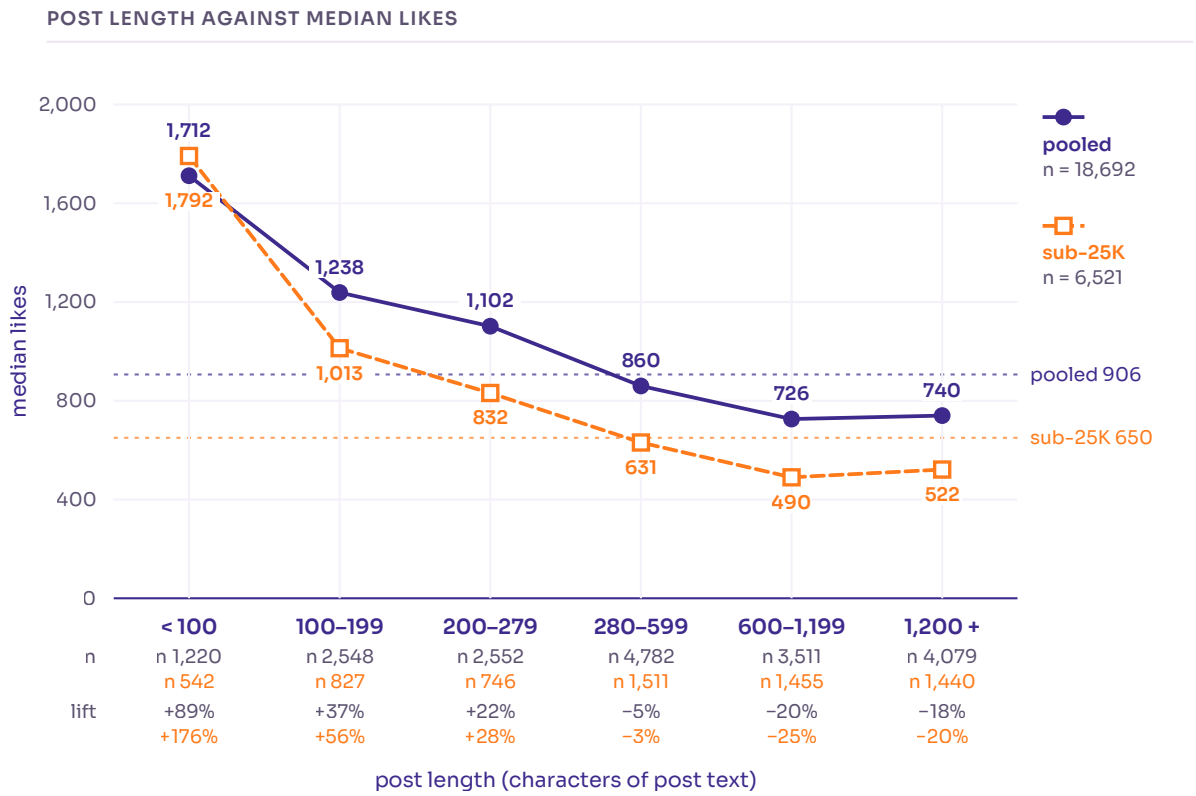
value 42.4), while replies and retweets per 100 likes are ratios of the group's medians. The poll cell (n = 5) is suppressed.

Media	n	Share %	Median likes	Lift %	Median views	Saves/100 likes	Replies/100 likes	Retweets/100 likes
Text	5,265	28.2	1,083	+19.5	108,868	21.9	6.2	7.5
Video	5,163	27.6	1,050	+15.8	134,205	81.9	5.5	11.8
Carouůsel	1,622	8.7	806	−11.1	77,602	38.9	6.5	10.9
Image	6,637	35.5	724	−20.1	67,696	40.8	8.4	12.6

Video buys reach and saves: the highest median views of any media class (134,205, 23.3 per cent above text) and 81.9 saves per 100 likes, rising to 92.2 at sub-25K where video's bookmark lift reaches +102.0 per cent. Text buys likes cheaply and little else — the lowest saves ratio and, by some distance, the lowest retweets ratio (7.5 per 100 likes against image's 12.6). Image is the largest supply lane at 35.5 per cent and the weakest on likes, yet returns the highest replies per 100 likes (8.4). These are not one currency measured four ways: an author choosing between text and video at sub-25K trades roughly equal likes against a fourfold difference in saves.

5.4 Length and structure

Length is the strongest simple predictor in the chapter and its response curve is close to monotone (Figure 5.2). Pooled, posts of 1–25 words median 1,442 likes, +59.1 per cent against baseline (n = 2,842); the curve falls through 26–50 words (+19.4), 51–100 (−11.9), 101–200 (−18.5) to a trough at 201–400 words (711 likes, −21.6 per cent, n = 2,728), then recovers slightly at 400+ words (828 likes, −8.7 per cent, n = 1,285). The recovery at the long end coincides with the highest save ratio in the table, 90.3 bookmarks per 100 likes against 12.1 for the shortest bucket: long posts are not failing so much as being paid in a different currency.



Lift is against each population's own median. Length is measured on the post text only; for a thread it is the opening post.

Figure 5.2 — Median likes by character-count band, pooled and within the sub-25K stratum. Bands follow the platform's own 280-character preview boundary. The two curves have the same shape but different slopes: compression pays more for a small account than for a large one.

The sub-25K curve starts higher and falls further: +112.4 per cent at 1–25 words against +59.1 per cent pooled.

The length effect is amplified, not attenuated, by stratification. At sub-25K the 1–25 word bucket medians 1,380.5 likes against a 650 baseline, +112.4 per cent (n = 1,062), a swing of 53.3 points above the pooled figure. Every other length bucket at sub-25K sits below its own baseline. Structure moves in the same direction: a post that is a single unbroken line (no line break at all) medians 1,661 likes against 832 for the rest of the population, +99.6 per cent, bootstrap 95 per cent CI [+84.2, +116.7], and the effect survives follower-tier stratification at +74.4 per cent. Within the sub-25K stratum the raw lift is +170.7 per cent and the stratified figure +169.5 per cent. It is also strictly decreasing in account size — +266.5 per cent at 0–1K, +145.6 at 5K–25K, +27.3 at 100K–1M, +11.4 at 1M+ — which is what a ceiling effect looks like: large accounts have a floor under them that small accounts do not. The same post pays for that reach in saves: single-line posts median 218.5 bookmarks against 443 for the rest, −50.7 per cent.

5.4.1 The 280-character boundary is not a cliff

X truncates the display of posts beyond roughly 280 characters behind a "Show more" control, which makes the boundary a natural candidate for a discontinuity. The corpus does not show one. Pooled medians run 1,364.5 likes below 200 characters (n = 3,768), 1,052 at 200–239 (n = 1,060), 1,137.5 at 240–279 (n = 1,492), 1,053 at 280–319 (n = 1,565), 800 at 320–359 (n = 661) and 761 at 360–439 (n = 1,023). The step across the fold, 240–279 against 280–319, is −7.4 per cent (p = 0.008) — detectable, but smaller than the −24.0 per cent step from 280–319 to 320–359, and smaller than the non-monotone wobble at 200–239 that precedes it. The penalty associated with length is broad and gradual rather than located at the truncation point, so the mechanism is more plausibly reader attention than an interface threshold. This is an association in observational data and no causal reading is licensed.

5.5 Video's decay across the study window

Because the corpus spans eight months and recent posts have accumulated engagement for less time, no level comparison across the window is interpretable on its own. The measure used here is an index: a format's median likes divided by the all-format median inside the same half of the window, so that maturation affects numerator and denominator together. On that index, video's advantage at sub-25K roughly halves (Table 5.5).

Table 5.5 — Video against all formats, by half of the study window. Index = format median likes ÷ all-format median likes within the same half. H1 = 2026-01 to 2026-04, H2 = 2026-05 to 2026-08.

Population	H1 n	H1 index	H1 supply %	H2 n	H2 index	H2 supply %	p (H1 vs H2 likes)
Sub-25K	350	2.20	13.9	840	1.20	21.0	1.3e-18
Pooled	1,309	1.67	15.3	2,035	1.21	20.1	9.3e-38

The monthly series at sub-25K is consistent with the two-half summary rather than driven by one month: the index runs 2.57, 1.78, 2.86, 2.06 across January to April and 1.22, 1.58, 1.19, 1.06 across May to August, while video's share of sub-25K posts rises from 11.3 per cent in January to 20.4 per cent in August. Supply and index move in opposite directions, which is what crowding looks like — an under-supplied format that pays well, is adopted, and pays less as it fills. The corpus cannot separate crowding from the alternatives. A change in how the platform ranks video inside the window would produce the same series, and X's published recommendation code establishes only that media features enter ranking, not how their weights moved [Twitter, 2023]; and if video accumulates likes on a slower schedule than text, maturation would depress H2's video index specifically. The direction is robust across both populations; the causal account is not.

5.6 What this chapter does and does not establish

It establishes, within a winners-only corpus of 21,864 high-engagement posts, that container variables are associated with substantial differences in engagement among posts that already performed. The largest are: threads, at 34.5 per cent of supply, sit 19.8 per cent below the pooled median and 20.3 per cent below at sub-25K, with the deficit present in every follower tier below 1M; short posts strongly out-perform long ones, +59.1 per cent pooled and +112.4 per cent at sub-25K for 1–25 words; single-line posts nearly double the median of everything else, and that effect survives stratification; and likes and saves order the formats close to inversely, with demo posts returning 137.9 bookmarks per 100 likes against the quote post's 13.3.

It establishes, as a methodological result, that the pooled and stratified answers disagree for two formats out of eight, and that in both cases the disagreement is explained by the correlation between format adoption and account size. The correct reading of a pooled format lift is therefore a compound of container effect and author composition, and any recommendation derived from a pooled table inherits that compound.

It does not establish base rates: nothing here supports a statement of the form "X per cent of threads fail", because posts that never cleared a like floor were never sampled, so the thread deficit is a deficit among threads that already succeeded. It does not establish causation; format is chosen by authors, and authors who choose demos differ from those who choose carousels in ways the corpus does not measure — topic, production budget, audience composition, cadence. It does not establish persistence, since Section 5.5 shows one of the largest associations halving inside eight months. And it establishes nothing about polls (n = 5, suppressed) or @grok posts (p = 0.834). Whether the language inside these containers carries independent signal — and whether arousal-based accounts of sharing [Berger & Milkman, 2012] have purchase on a technical audience — is taken up in the next chapter.

6. The Opening Line

Chapter 5 treated the post as a container. This chapter treats it as a sentence, reporting what the corpus shows about the first line — its pattern, vocabulary, grammar and punctuation — and about twelve families of persuasive framing detected across the body. Two results organise it: reach and saving are driven by close to opposite linguistic shapes; and three of the strongest-looking language findings did not survive re-derivation, so both the retracted figure and its replacement are reported.

6.1 The opening line as the unit of analysis

The hook is the first non-empty line of the post, URLs stripped and whitespace collapsed, recomputed from the raw `text` field rather than read from the CSV's `hook` column, which truncates at 220 characters. On the analysis population — audience in {builder, mixed}, n = 18,692 — the median hook is 63 characters and 11 words, the 90th percentile 167 characters, and only 2.01 per cent of hooks exceed 280 characters while 65.8 per cent of whole posts do. The opening line is therefore almost never what "Show more" hides; the fold conceals the body.

Three outcomes are used. Reach is a like index: likes divided by the median likes of the post's own follower tier, which absorbs the follower confound of Chapter 5. Group lift is the ratio of the group's median like index to the population's, with the raw uncorrected likes lift alongside wherever the two disagree. Save intent is `bookmark_ratio`, bookmarks per like; argument intent is `reply_ratio`, replies per like. Baselines are 906 median likes, a median like index of 0.928, 408 median bookmarks, a bookmark ratio of 0.4242 and a reply ratio of 0.0575. Groups with n < 15 are suppressed; the mined lexicon uses a floor of n ≥ 25.

Methodological note

Two testing regimes. The hook-pattern table in Section 6.2 carries Benjamini–Hochberg control across its family [Benjamini & Hochberg, 1995]. The linguistic tables in Sections 6.3 to 6.6 come from the hook module, which runs two-sided Mann–Whitney U per row with no multiplicity correction; rows with p between .01 and .05 are suggestive rather than established. A verifier found that the module originally tested only the likes side, leaving every save figure untested — one published claim borrowed a likes p-value to license a saves number. A save-side test was added in response, and every save figure below is quoted with its own p-value.

6.2 Hook pattern performance

A rule-based classifier assigns each post zero or more of 26 hook-pattern labels. Because the labels are non-exclusive, the rows of Table 6.1 do not partition the population and the lifts are not additive. Two of the 26 labels fall below the suppression floor in the pooled population and are absent from the table; five of the 24 that remain fall below it inside the sub-25K stratum and are marked — in that column.

Table 6.1 — Hook pattern performance, pooled analysis population (builder + mixed, n = 18,692; baseline 906.5 likes, 408 bookmarks) and inside the sub-25K stratum (n = 6,521; baseline 650 likes, 347 bookmarks). Saves/100 likes is the median of the per-post bookmarks-to-likes ratio within the group, against a population value of 42.4. Significance is BH-adjusted within the 24-row family of unsuppressed patterns. Patterns suppressed at n < 15 in a column are marked —.

Hook pattern	n	Likes lift %	Saves lift %	Saves/100 likes	Median followers	FDR .05	sub-25K n	sub-25K lift %
Breaking / just launched	597	+105.8	−29.2	11.8	470,644	yes	64	+19.8
Announcement of access	78	+94.9	−27.2	22.5	200,282	yes	15	+80.2
This is insane / crazy	36	+53.6	+137.6	96.4	140,662	no	—	—
Emotional reacůtion	126	+45.5	+89.7	63.1	57,493	yes	35	+59.5
I built / I made	140	+9.9	+148.3	113.1	39,909	no	47	+76.6
Nobody is talking about	42	+7.7	+134.2	104.7	57,994	no	—	—
Direct address "you"	299	+6.3	+68.6	78.6	56,768	no	100	+9.5
Plain statement	12,685	+5.9	−14.2	30.5	51,939	yes	4,454	+4.6
Contrarian take	18	+0.8	−65.3	13.4	35,438	no	—	—
Personal result / income	1,781	−0.1	+58.6	85.4	44,789	no	621	+9.8
X vs Y comparisůon	235	−3.1	−9.8	48.0	47,314	no	88	−9.2
Time-bound flex	452	−3.8	+146.2	148.3	37,350	no	164	+5.3
Thread anůnouncement	108	−9.9	+31.5	68.6	35,722	no	44	−17.9
Question hook	201	−13.4	−42.2	29.6	36,656	no	86	+20.4
How to / tutorial	146	−18.0	+164.7	153.0	78,577	no	37	−32.5
Superlative claim	885	−28.5	+17.6	85.2	58,134	yes	279	−21.1
Prompt drop	50	−29.6	−22.4	69.7	36,282	yes	22	−9.5
Free resource giveaway	245	−30.6	+93.6	142.7	23,999	yes	127	−18.2
Comment-to-get (bait)	107	−31.3	+14.2	78.7	33,760	yes	41	−34.3
Stop doing X	160	−33.1	+159.4	156.7	34,418	yes	64	−8.5
Warning / fear	41	−35.7	+125.2	172.8	64,916	no	—	—
Steal this / take this	18	−41.4	+172.2	162.3	46,080	no	—	—
Us vs them / gateůkeeping	367	−42.2	+52.5	126.6	36,345	yes	152	−36.2
Numbered list / listicle	466	−51.1	−34.7	59.8	25,110	yes	226	−43.0

Two features of the ranking matter more than any individual row. First, the top of the table is a follower artefact as much as a language effect: the two highest-lift patterns are launch announcements from accounts with median follower counts of 470,644 and 200,282, and the Breaking lift collapses from +105.8 per cent pooled to +19.8 per cent at sub-25K, where it no longer clears significance (n = 64). Second, the two orderings are close to inverted: across the 24 unsuppressed rows, likes lift and saves per 100 likes correlate at Spearman ρ = −0.49 (p = 0.015), and the patterns at the bottom of the reach ranking — Stop doing X, Steal this, How to, Warning — carry the highest save ratios in the corpus, 153 to 173 bookmarks per 100 likes against a population figure of 42.4.

> **The listicle opener is the only construct in the taxonomy that loses on both axes at once: −51.1 per cent likes and −34.7 per cent saves, with the lowest median follower count of any pattern.**

6.3 Lexical findings

6.3.1 Opening n-grams

Anchored n-grams are the literal first one, two or three words; sliding n-grams are any n-gram falling entirely inside the first five words. Both are reported at n ≥ 15, and their extremes are instructive chiefly as a warning. The highest-lift anchored opener is `introducing claude` (n = 17, median 38,827 likes, +1,984 per cent), from accounts with a median follower count of 1,735,933: a laboratory press release, not a writing move. The lowest openers are dominated by image-generation provenance captions — `gpt image on` (n = 19, −70.8 per cent), `created with` (n = 36, −68.2) — and by the round-number scold `99 of the` (n = 18, −60.6), all at median follower counts between 8,000 and 12,000. Both tails are partly statements about who writes them.

6.3.2 First-word grammatical buckets

Every hook's first token is assigned to one of seventeen part-of-speech-like buckets, checked in priority order. Table 6.2 gives the result; Figure 6.1 plots the same buckets on both axes.

Table 6.2 — First-word grammatical bucket, analysis population (n = 18,692). Likes lift is on the follower-tier-normalised like index; the saves column is bookmark-ratio lift with its own two-sided test. Rows are ordered by likes lift.

First-word bucket	n	Median likes	Likes lift %	p (likes)	Saves lift %	p (saves)	Median follow-ers
Greentext (>)	84	2,264	+178.7	2.7e-08	−37.3	0.495	40,135
Quotation mark	223	1,612	+78.9	4.3e-08	−70.2	1.6e-16	48,352
Connective (if / when / beů cause)	810	1,031	+16.4	0.013	−4.0	0.380	42,380
Pronoun	2,145	1,022	+15.6	9.9e-12	+28.3	1.2e-08	45,696
Gerund (-ing)	620	1,208	+13.8	0.060	−48.0	5.4e-11	85,411
Question word	425	922	+8.1	0.252	+11.8	0.419	42,931
Deictic (this / that / there)	819	944	+6.6	0.165	+142.5	3.5e-43	49,217
Emoji	1,269	1,029	+5.2	0.994	−9.8	0.911	69,003
Brand / model name	2,295	932	+4.8	0.079	−35.0	0.005	67,953
Imperative verb	780	861	+3.9	0.111	+128.4	6.5e-39	49,113
Other noun / adjective	6,442	893	−3.5	0.069	−10.7	4.0e-19	47,464
Auxiliary (yes/no question)	153	741	−3.9	0.422	−44.3	8.5e-06	29,792
Determiner (the / a / every)	1,613	817	−9.7	9.1e-06	−3.4	0.188	58,055
Hashtag	43	586	−18.9	0.143	−69.8	2.4e-07	5,048
Symbol / other	36	619	−34.7	0.066	−3.7	0.191	17,434
Number	656	466	−38.6	3.2e-27	+53.6	4.1e-13	25,110
Generic AI noun	264	497	−49.4	1.1e-20	−28.7	2.0e-04	77,254

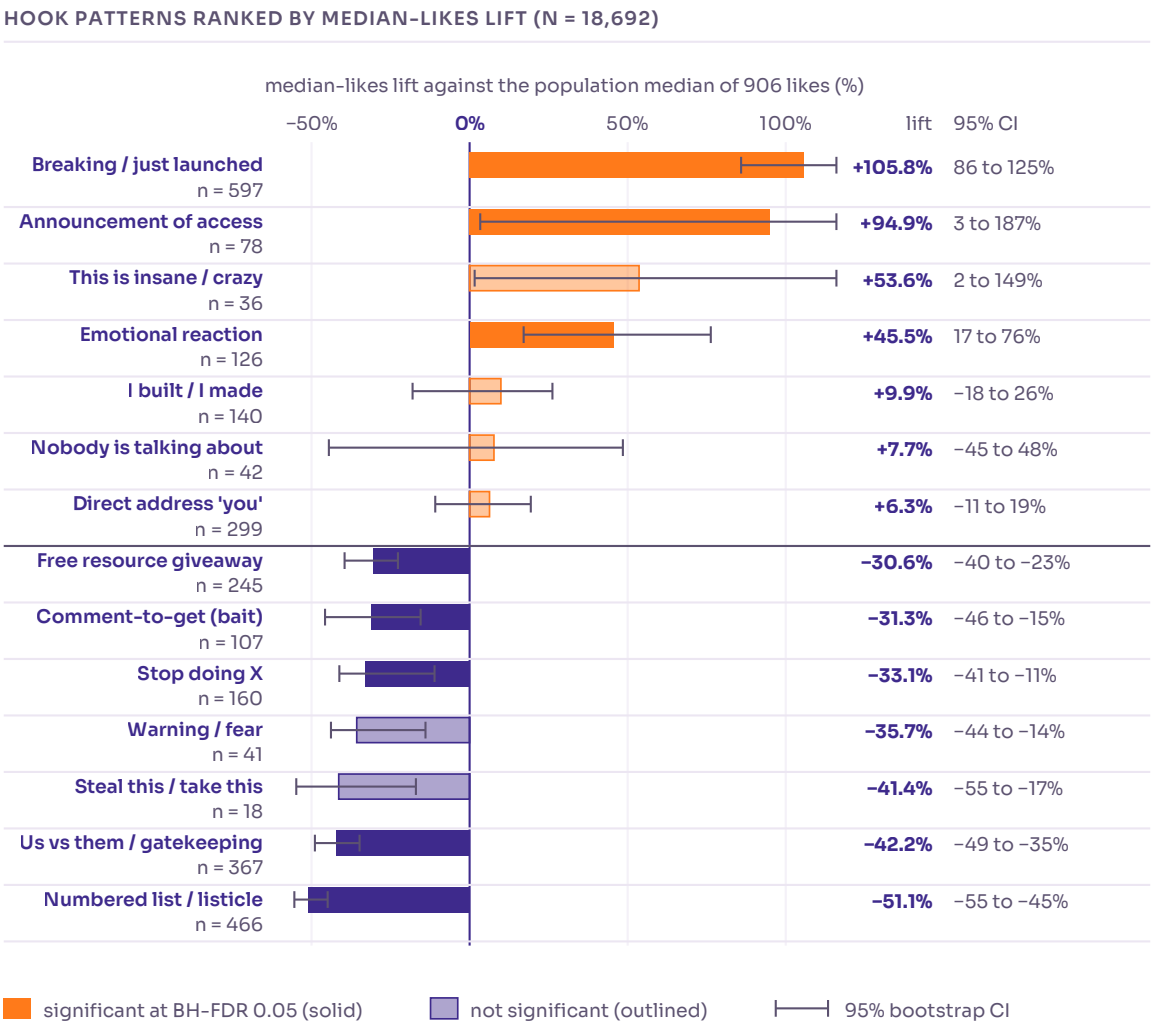


Figure 6.1 — Hook patterns ranked by median-likes lift against the population median of 906 likes (n = 18,692), showing the seven strongest and seven weakest of the 24 patterns with n ≥ 15. Bars are solid where the difference survives Benjamini–Hochberg correction at 0.05 and outlined where it does not; whiskers are 95 per cent bootstrap intervals on the group median. Pattern labels are non-exclusive, so the rows do not partition the population and the lifts are not additive.

The buckets separate into a reach group (greentext, quotation mark, connective), a save group (deictic, imperative verb) and a group that loses on both (number, generic AI noun, hashtag).

The pronoun bucket (n = 2,145) is the only one positive and significant on both axes, at +15.6 per cent likes and +28.3 per cent saves. The two strongest reach buckets are stylistic rather than semantic: an opening `>` in the greentext idiom (+178.7 per cent) and an opening quotation mark (+78.9 per cent). Both hold inside the sub-100K slice, at +215.5 per cent (n = 68) and +110.2 per cent (n = 145), so neither is a privilege of scale, though both rest on small

groups. The greentext save figure of −37.3 per cent carries p = 0.495 and should not be quoted; it is retained to show the failure. The worst bucket is opening on a generic AI noun, at −49.4 per cent, which sets up Section 6.4.

6.3.3 The mined power and drag lexicons

Rather than test a memorised list of "power words", every token appearing in at least 25 first lines was tested against the rest of the population, retaining rows at p < 0.10. Table 6.3 shows ten tokens from each tail; the caption records what the selection leaves out.

Table 6.3 — Mined opening-line lexicon at n ≥ 25 and p < 0.10, analysis population: ten tokens selected from each tail. The selection is not the exhaustive top ten in either direction. Omitted from the lift column are a profanity (+499.5%, n=42, the largest single-token effect in the lexicon), two spellings of one contraction (+274.5% and +213.6%), worse (+180.0%) and funny (+168.6%); omitted from the drag column are anime (−52.8%) and resources (−49.7%). Lift is on the follower-normalised like index. No multiplicity correction is applied.

Lift words	n	Lift %	p	Drag words	n	Lift %	p
reset	28	+376.8	4.0e-04	execution	26	−71.8	2.1e-05
wife	25	+339.0	2.8e-05	tedious	50	−66.1	1.4e-12
accidentally	51	+230.8	9.4e-06	creation	28	−61.4	4.2e-04
proved	26	+218.9	7.3e-06	influencer	38	−59.2	7.6e-09
usage	107	+204.9	1.8e-11	creators	30	−57.1	9.7e-06
limit	63	+198.7	2.2e-08	cinematic	47	−55.0	1.8e-07
preview	34	+186.2	2.0e-04	gm	46	−50.5	3.1e-07
introducing	167	+180.6	6.4e-18	workflow	331	−48.9	7.2e-29
session	47	+166.5	2.3e-05	image	482	−48.9	4.4e-37
terminal	33	+144.6	0.004	replace	98	−45.3	5.4e-07

The lexicon must be read as topic detection, not word choice. A verifier confirmed the arithmetic and checked handle concentration — `workflow` draws its 331 rows from 264 distinct handles, `image` its 482 from 160 — but observed that `image`, `cinematic`, `creators` and `influencer` mark an image-generation and creator-economy niche that underperforms here, rather than levers whose deletion would recover reach. Two rows are additionally concentrated: `gm` (n = 46) and `tedious` (n = 50) draw 28 and 30 per cent of their observations from one account each. Nothing tests what happens when a writer removes the word and keeps the topic.

6.4 Specificity: a named tool against the word "AI"

Hooks that name a concrete tool or model — from a 91-entry list — median 1,051 likes against 740 for hooks whose only AI reference is a generic noun (AI, LLM, model, agent): a like-index lift of +16.4 per cent (n = 7,900, p = 2.3e-49) against −16.4 per cent (n = 5,210, p =

2.8e-48), a 42.0 per cent gap between the group medians. Hooks with no AI noun at all sit between them at −0.7 per cent (n = 5,582, ns). The independently computed facts-pack feature agrees: +28.4 per cent raw and +29.6 per cent after follower-tier stratification (n = 7,455), with near-identical values at sub-25K (+27.9 and +30.0, n = 2,471). This is one of the few language effects in the chapter stable across estimator, population and stratum.

Placement matters, and the original analysis got the test wrong: it compared hooks containing `claude code` (n = 1,349) against posts containing the phrase anywhere (n = 2,617) — a group against its own superset, which dilutes rather than isolates. The figure reported here is the verifier's mutually exclusive partition: the phrase in the first line, +111.5 per cent (n = 1,349); in the body only, +25.3 per cent (n = 1,268); absent, −5.9 per cent (n = 16,075). The direction survives — the first line is where naming pays — but the magnitude is now measured against a clean complement.

Methodological note

Retracted: "comparison framing is the highest-volume reach play." The original finding reported that two named tools in one hook lift likes by +26.3 per cent (n = 3,853). The matching routine scanned a list containing both `claude` and `claude code`, and `claude` matches inside `claude code`, so all 1,349 `claude code` hooks were automatically tagged as carrying two tools; that pair alone supplied 881 of the 3,853 rows. De-overlapped, the effect is n = 2,972 and +6.9 per cent (p = 5.3e-04), falling from +30.0 to +15.4 per cent in the sub-100K slice. Comparison framing remains positive and significant but modest, and the retracted version double-counted the specificity result above.

6.5 Numerals in hooks

Opening with a number is the clearest reach-for-saves trade in the chapter. A digit anywhere in the first eight words is associated with −13.2 per cent likes and +13.1 per cent saves (n = 5,544, p = 7.0e-24 and 3.5e-09). A hook that literally begins with a digit is far worse for reach: −39.3 per cent on the like index and −50.1 per cent raw (n = 616, p = 1.2e-26), with +49.6 per cent saves (p = 3.0e-12). The facts-pack feature `numeral_in_hook` agrees at −17.5 per cent raw and −11.3 per cent stratified (n = 7,365).

By value, the penalty concentrates on round marketing numbers rather than digits as such: 100 is −47.8 per cent (n = 176), 2 is −34.5 (n = 708) and 7 is −19.4 (n = 56), while 3 (+3.1, n = 241) and 5 (−0.5, n = 866) are indistinguishable from baseline. Only 15 lifts, on 30 observations at p = 0.078.

Two qualifications follow. First, the digit-opener penalty is a small-and-mid-account tax, not a general law: −47.3, −44.1 and −45.9 per cent in the 1K–5K, 5K–25K and 25K–100K tiers, but +0.3 per cent at 100K–1M (n = 97, p = 0.62). The original claim that this is among the most reliably damaging things a writer can do was stated too broadly. Second, the tax and the subsidy separate cleanly when the count moves off line one. Hooks whose first line does

not begin with a digit but whose second line does number 714, median 870 likes, and run −0.8 per cent on likes (p = 0.95) with +136.0 per cent on saves (p = 1.3e-43). Hooks in which the enumeration has moved off line one therefore carry no measurable reach penalty and a larger save lift than the +49.6 per cent of the digit-first construction — a contrast between two groups of hooks, not an observed consequence of moving a digit. The verifier computed this after objecting that the original recommendation had been asserted without measurement.

> **The listicle opener is a reach tax and a save subsidy, and the two are separable: the tax is levied by the digit's position in the first line, while the subsidy is bought by the enumeration itself, wherever it sits.**

6.6 Punctuation, casing and structure

Table 6.4 reports the orthographic constructs. The strongest single result — a post that is one unbroken line medians 1,661 likes against 832, +99.6 per cent raw and +74.4 per cent stratified — is given in full in Section 5.4; what matters here is its mirror image. The classic thread opener, a hook of 90 characters or fewer followed immediately by a line break, covers 11,807 posts at −10.6 per cent likes and +48.1 per cent saves, the latter at p = 2.6e-232, the smallest p-value in the module.

Table 6.4 — Punctuation, casing and line-structure constructs in the opening line, analysis population (n = 18,692). Saves lift is bookmark-ratio lift under its own test. Rows ordered by likes lift. Eighteen of the module's twenty orthographic constructs are shown, together with the trailing-emoji row from the emoji-position block; the three omitted rows (parentheses in hook, +3.4; ALL-CAPS shout word with acronyms excluded, +1.4; ellipsis in hook, −1.2) are all null, and the rank correlation quoted in the text is computed over the full twenty-row set.

Construct	n	Median likes	Likes lift %	p (likes)	Saves lift %	p (saves)
Greentext opener (> first character)	84	2,264	+178.7	2.7e-08	−37.3	0.495
Hook opens with a quotation mark	223	1,612	+78.9	4.3e-08	−70.2	1.6e-16
Single-line post (no break at all)	2,428	1,661	+62.0	2.2e-78	−77.1	<1e-300
Hook is all lower-case	1,105	1,127	+39.7	1.4e-20	−29.4	0.002
Em- or en-dash in hook	343	1,146	+25.3	0.441	−36.9	5.3e-05
Comma in hook	4,721	1,128	+13.3	7.6e-12	−44.5	7.8e-78
Multi-line post, long first line (>90 chars)	4,457	1,096	+11.2	2.9e-08	−22.7	1.1e-08
Colon anywhere in hook	3,434	1,083	+9.1	1.6e-04	−11.6	0.493
Two or more ALL-CAPS shout words	1,710	991	+3.4	0.577	+51.8	1.7e-12
Hook ends with a period	8,795	971	+3.3	0.003	−0.1	0.018
Exclamation mark in hook	854	882	+1.5	0.686	−27.6	4.3e-08
Hook ends with a colon	1,512	852	−1.4	0.368	+116.1	1.2e-63
Question mark in hook	676	848	−2.8	0.787	−49.6	7.8e-21
No terminal punctuation (open loop)	7,353	844	−2.9	0.110	−9.3	7.8e-17
Entire hook in ALL CAPS	973	806	−8.2	0.002	+143.4	1.3e-47
Any ALL-CAPS word (acronyms inů cluded)	8,746	814	−10.0	6.6e-39	+17.4	3.0e-08
Line break right after a short hook	11,807	755	−10.6	3.8e-72	+48.1	2.6e-232
Trailing emoji in hook	1,227	628	−21.7	1.3e-12	+14.3	0.172

Three observations. Emoji placement matters and its direction is not obvious from prevalence: a leading emoji is a non-event (+7.3 per cent, n = 1,157, p = 0.574) while a trailing emoji is −21.7 per cent (n = 1,227, p = 1.3e-12). Capitalisation is a stratification reversal in miniature: an ALL-CAPS word in the hook is +0.6 per cent pooled and not significant, but −13.9 per cent and significant inside sub-25K — shouting is tolerated at scale and penalised below it. And the two-lane structure recurs: the constructs that close a hook (terminal colon, +116.1 per cent saves; a hook entirely in capitals, +143.4 per cent saves) are save instruments with flat or negative reach. Across the 20 constructs, likes lift and save lift correlate at ρ = −0.70 (p = 6.0e-04).

Methodological note

Retracted: "write a 240–279 character single-line hook." Pooled across all posts, hook length looks monotone in reach — 30+ word hooks +24.3 per cent, 20–29 words +22.9, 5–7 words −12.9 — peaking at +42.3 per cent in the 240–279 character bucket. This is a mixture. Hook length is entangled with line structure: 59.3 per cent of 240–279-character hooks belong to single-line posts, and being single-line is itself worth +62.0 per cent. Re-scored inside each lane against its own baseline, the direction reverses. Within multi-line posts the long-hook effect is real (30+ words +25.4 per cent, p = 2.8e-08). Within single-line posts it inverts: 1–4 words +97.4 per cent (n = 58, p = 0.087, not significant and quoted only for the shape of the curve), 30+ words −16.6 (n = 662, p = 1.3e-05), and the celebrated 240–279 character bucket is +2.0 per cent at p = 0.759 (n = 235). The pooled recommendation therefore prescribed the worst combination the data supports — the classic aggregation reversal [Simpson, 1951], and the second of two in the results chapters.

6.7 Psychological triggers

A separate module tags each post with any of twelve persuasion families, defined by high-precision regular expressions over the whole body rather than the hook alone, and divided into pull constructions (curiosity gap, social proof, status, utility, novelty, generosity) and push constructions (authority, controversy, FOMO, vulnerability, us-vs-them, scarcity). Several of the family labels — social proof, authority, scarcity — are borrowed from the classical persuasion literature [Cialdini, 2007]; the borrowing is terminological, and no result below depends on that literature being correct. Because the patterns favour precision, every prevalence figure is a floor rather than a census, and the labels are non-exclusive. Table 6.5 adds the de-duplicated re-measurement described below.

Table 6.5 — Persuasion trigger families, analysis population (n = 18,692; baselines 906.5 likes, bookmark ratio 0.4242, reply ratio 0.0575). Lifts are ratios to baseline, so 1.00 is at baseline. Tier-balanced is the median within-tier likes lift. Dup % is the share of the family's posts that are verbatim repeats of another post; the dedup columns re-measure on first-copy-only data (n = 17,328) against its own baseline.

Trigger	Family	n	Prev %	Likes	Tier-bal.	Saves	Replies	Dup %	Dedup likes	Dedup saves
Authority / insider access	push	390	2.09	1.24	1.11	2.01	0.90	7.7	1.26	2.11
Vulnerability / failure story	push	40	0.21	1.08	—	1.11	1.19	10.0	1.28	1.14
Us vs them	push	81	0.43	1.07	1.09	1.12	1.63	2.5	1.03	1.00
Curiosity gap	pull	1,253	6.70	0.99	0.95	2.53	1.08	4.6	0.95	2.85
Social proof	pull	545	2.92	0.98	0.95	2.53	0.97	4.6	0.96	2.86
Novelty / news	pull	4,871	26.06	0.95	0.96	0.81	1.06	5.0	0.94	0.84
Controversy / contrarian	push	486	2.60	0.95	0.94	1.70	1.36	8.6	0.99	1.64
Scarcity / time-bound	push	671	3.59	0.94	0.95	1.26	9.74	30.0	0.97	1.88
Generosity / free gift	pull	2,201	11.78	0.92	0.99	2.66	1.20	16.8	0.92	3.11
Status / identity	pull	220	1.18	0.85	0.82	2.25	1.05	8.6	0.96	2.54
Fear of missing out	push	266	1.42	0.74	0.82	2.71	1.10	7.1	0.74	2.97
Utility / practical value	pull	4,742	25.37	0.70	0.76	2.82	1.05	13.8	0.68	3.05

Ranked by save lift minus reply lift, eight of the twelve families are save instruments, led by utility (+1.78), FOMO (+1.61) and social proof (+1.56); two are argument instruments, us-vs-them (−0.51) and scarcity (−8.48 before de-duplication); novelty is the only reach-only family, lifting neither saves nor replies. Trigger density behaves monotonically in the same currency: posts matching no trigger (n = 8,195, 43.8 per cent) median 1,105 likes at 1.22× baseline with a save ratio of 0.49×, while posts matching five or more (n = 58) run 0.79× likes and 3.72× saves. That gradient is not a dose-response of persuasion alone: the zero-trigger group is compositionally different, at a median 47 words against 69 population-wide and 22 per cent threads against 35, so length and format are confounded with trigger count.

6.7.1 Two fatal objections

The scarcity result was a copy-paste artifact. Scarcity's headline figure — a reply ratio of 9.74× baseline with likes at 0.94× — is not a trigger effect. Thirty per cent of the 671 scarcity posts are verbatim repeats of other corpus posts. One spam template offering free courses to the "first 4500 people" within "72 hours only" appears 128 times, 116 of them inside the scarcity group, from 30 handles with one contributing 28 copies; its median reply ratio is 0.932, roughly sixteen times baseline, and 80 per cent of its instances are flagged engagement bait. On first-copy-only data the reply lift collapses from 9.736× to 2.408× and the likes lift reverses from 0.941× to 0.970×. The direction survives and the magnitude does not: scarcity is still the most reply-inducing family, at roughly 2.4×, but the claimed reach pen-

alty disappears. The module was re-run with a 120-character normalised-text fingerprint and a first-copy-only cohort carrying its own baseline; 1,397 near-duplicate posts were identified corpus-wide.

The authority result is a construct-validity failure. Authority/insider access is the only family positive on both reach and saves, and it was the module's headline recommendation. Re-derivation showed the regex is dominated by third-party leak reporting rather than earned first-person access: 176 of 390 matches (45 per cent) fire on the bare token leaked, and 306 of 390 belong to a leak, internal-document or ex-employer cluster which carries the reach at 1.327×. The genuinely first-person subset — the behaviour the recommendation described — is n = 64 at 0.894× likes, below baseline, with saves at 3.31× and replies at 1.16×. An earned-access opening line is therefore associated with saves at the cost of reach; what is associated with reach is having a leak to report. The objection was reported rather than repaired, since splitting the family would change the taxonomy and every downstream key: the authority row of Table 6.5 should be read as measuring leak-news, not credentials.

A third, non-fatal objection corrected the pair analysis: the module emits 30 pairs clearing an n ≥ 40 floor, not 35, and its `stacks` flag is true for four, not five. The pair originally promoted as the best combined play, authority + generosity, has a stack gain of −0.071 and at 1.17× underperforms authority alone at 1.24×. The one pair that is both the top reach combination and a genuine stacker is curiosity gap + authority: n = 56, 1.358× likes against 1.218× for the better ingredient alone. At n = 56 that is a lead, not a settled result.

Finally, segmentation inverts one family outright. Controversy runs at 0.95× in the builder+mixed population and 0.82× among sub-25K accounts, but at 1.892× in the consumer segment (n = 81), where it is the top-performing family. Contrarian framing about AI pays with the mass audience and not with the practitioner audience — a difference pooling would erase.

6.8 Synthesis: the two-lane model

The language data support a two-lane account of what an opening line can be optimised for; the lanes are near-orthogonal in construction and negatively correlated in outcome. Across the 24 hook patterns of Table 6.1, likes lift and saves per 100 likes correlate at ρ = −0.49 (p = 0.015); across the 20 orthographic constructs of Table 6.4, at ρ = −0.70 (p = 6.0e-04); across the 12 trigger families of Table 6.5, at ρ = −0.61 (p = 0.036). These are rank correlations over small numbers of group-level summaries, descriptive rather than inferential, and the first-word buckets of Table 6.2 are the counter-example at ρ = −0.14 (p = 0.59). Three of four independent groupings nonetheless point the same way.

The reach lane is a single unbroken line, kept short within that lane, opening on a pronoun, a connective, a quotation mark or the greentext marker, naming a concrete tool rather than a generic AI noun, with no digit in the opening position, no hashtag, no trailing emoji and at most one persuasion trigger. The save lane is a short first line followed immediately by a break, an imperative or deictic opener, a terminal colon, enumeration displaced to line two,

and two or three stacked pull triggers. Two constructs belong to both lanes, and they are the most useful findings in the chapter: naming a concrete tool in the first line, and the pronoun opener — the only grammatical bucket significantly positive on both axes.

Every statement above is associational. The corpus is winners-only and observational; the language variables are entangled with subject matter, with format and with who writes them, as the drag lexicon and the authority family both demonstrate; and nothing here manipulates a hook and observes the result. The correct reading is that among posts that already performed, these opening-line shapes separate the ones that were liked from the ones that were kept.

7. Accounts, Networks and Timing

The preceding chapters treated the post as the unit of analysis. This chapter shifts to the account: how far performance tracks the audience already attached to a post, whether creators fall into recognisable behavioural types, how mentions and quotes route attention between handles, what separates rising from decaying creators, where payoff exceeds supply, which vendor brands carry engagement, and whether publication timing survives its confounds. Three of the six modules underlying the chapter were graded unsound by their verifiers and three needs_caveat; the surviving numbers are reported with the objections attached.

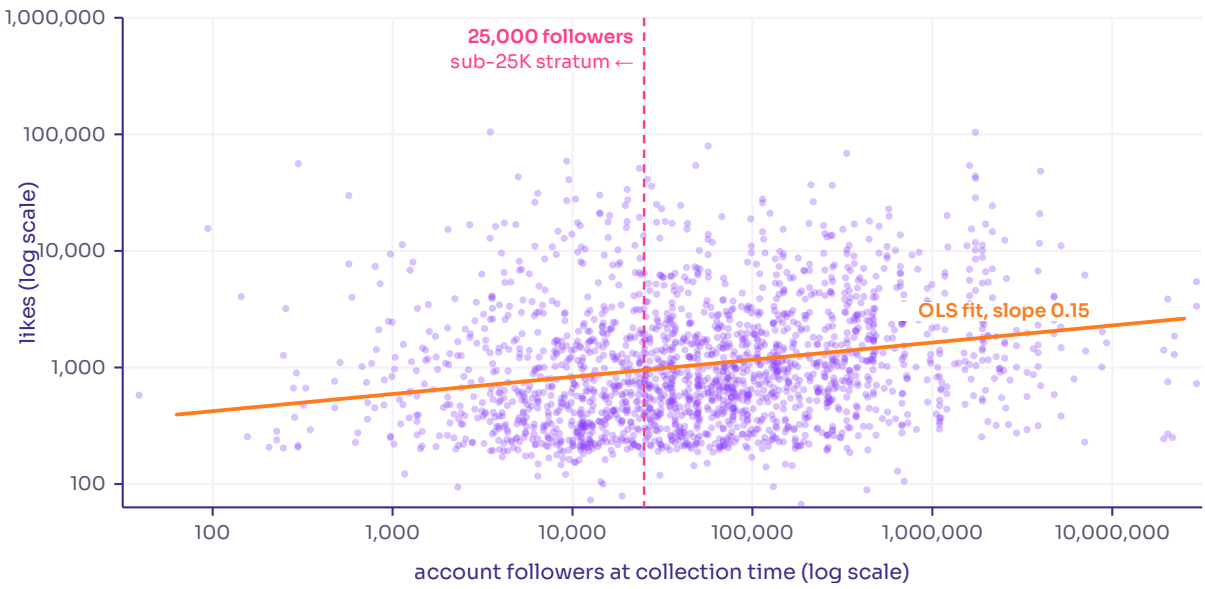
7.1 Follower count as a predictor

Within the builder segment (n=3,370) the Pearson correlation between $\log_{10}$(followers) and $\log_{10}$(likes) is r=0.209; across builder and mixed posts (n=18,692) it is r=0.232 (r^2=0.054), across the full corpus (n=21,864) r=0.160, and inside the sub-25K stratum (n=6,521) r=0.069. Among posts that have already performed, follower count accounts for roughly five per cent of the variance in log-likes at its strongest (Figure 7.1).

Methodological note

A low r^2 here is not evidence that audience size does not matter. Selection on the dependent variable is the textbook condition for range restriction, which attenuates a follower-to-likes slope toward zero. Among winners, size is a weak discriminator; nothing follows about the unobserved population of all posts.

FOLLOWER COUNT AGAINST LIKES, ANALYSIS POPULATION



2,000 points plotted	**18,688** population fitted	**0.229** Pearson r
0.052 R^2	**0.147** slope (log-log)	**2.331** intercept

Follower count explains about five per cent of the variance in log likes among posts that already performed; the cloud is wide at every account size. Sampling is uniform without replacement at seed 20260824; the fit uses all 18,688 posts, not the sample.

Posts with zero likes or followers are dropped (undefined logarithm); 13 of the sampled points fall outside the plotted range and are not drawn.

Figure 7.1 — Follower count against likes, both on log scales, builder and mixed posts (n=18,692). The associů ation is r=0.232 in this population and r=0.209 within the builder segment alone (n=3,370).

Table 7.1 — Engagement by follower tier, builder and mixed posts. Baseline is the population median of 906.5 likes and 408 bookmarks (n=18,692). Bookmarks per 100 likes and likes per follower are both per-post medians, not ratios of the tier medians beside them.

Follower tier	n	Median likes	Lift	Median bookmarks	Bookmarks / 100 likes	Median likes / follower
0–1K (brand new)	427	584	−35.6%	182	15.9	1.335
1K–5K (starter)	1,558	656	−27.6%	328.5	40.6	0.259
5K–25K (small)	4,536	655	−27.7%	379.5	57.1	0.055
25K–100K (mid)	5,504	780	−14.0%	478	64.9	0.016
100K–1M (large)	5,231	1,243	+37.1%	462	34.8	0.005
1M+ (mega)	1,436	2,128	+134.7%	255.5	11.5	0.00086

A single correlation understates the structure. In Table 7.1 median likes are flat across the three smallest tiers and then rise, a 3.6-fold increase from the smallest tier to the largest, while median likes per follower falls across the same span from 1.335 to 0.00086 — a decline of roughly 1,560-fold. Bookmarks follow neither: they peak in the 25K–100K tier at 478 (+17.2%) and fall to 255.5 at mega scale (−37.4%), so saves per hundred likes run 64.9 at mid-scale against 11.5 at the top. Because the 0–1K tier enters a winners-only corpus only by escaping the follower graph — its likes-per-follower median is some twenty-four times the 5K–25K figure — nothing about a new account's expected outcome can be read from its row.

> **Absolute reach and per-follower efficiency move in opposite directions across the whole tier ladder. Which of the two a comparison happens to measure determines its sign.**

7.2 Creator archetypes

Creators with at least three posts (1,484 creators, 14,472 posts) were fingerprinted on twenty-seven behavioural features spanning format, hook, theme and audience mix, link and quote share, engagement ratios, post length, log followers and viral hits per month, then z-scored and clustered by k-means into seven groups, against a baseline of 947 median likes.

Table 7.2 — Seven creator archetypes. Likes lift is against the 947-like median of all clustered posts; bookmark-ratio lift against 0.463; likes-per-follower lift against 0.0128.

Archetype	Creators	Posts	Median followers	Median likes	Likes lift	Bookmark-ratio lift	Likes/follower lift
Media Broadcaster	418	4,125	81,690	1,129	1.19×	0.40×	0.69×
Long-form Essayist	342	3,498	36,136	786	0.83×	1.43×	1.18×
Timeline Commentator	315	2,668	84,882	1,326	1.40×	0.30×	0.84×
Hype Amplifier	203	2,299	23,382	793	0.84×	3.47×	1.36×
Prompt Librarian	131	1,131	20,072	515	0.54×	1.83×	1.75×
News Wire	38	421	406,394	1,955	2.06×	0.20×	0.18×
Demo Builder	37	330	25,201	804	0.85×	2.80×	1.69×

In Table 7.2 the clusters separate along a reach-versus-retention axis rather than a good-versus-bad one: the three highest on raw likes are the three lowest on bookmark ratio. The News Wire is the clearest case, at 2.06× the corpus median on likes but 0.18× on likes-per-follower, with a median of 406,394 followers — an advantage attributable to distribution already owned rather than to the one-line breaking-news post. The save-first clusters invert it: the Prompt Librarian is lowest on likes and highest on likes-per-follower, and the Demo Builder pairs the highest builder share of any cluster (58.8%) with 2.80× on saves. The affective commentary formats reach further while the utility formats are retained more, a pattern consistent with the arousal-and-utility account of sharing [Berger & Milkman, 2012].

Methodological note

The clusters are a gradient, not a taxonomy. The silhouette coefficient at k=7 is 0.0903 [Rousseeuw, 1987] — effectively no separation — and k=8 scores 0.0894, so the choice of k is near-arbitrary within the scanned range. Cross-seed agreement has an adjusted Rand index averaging 0.636 with a minimum of 0.55 [Hubert & Arabie, 1985]. Two further objections were sustained. The n ≥ 15 suppression floor counts posts, not creators, so several cells rest on a handful of accounts: the Demo Builder's 1K–5K cell is 26 posts from seven creators, and leave-one-creator-out moves its median between 619 and 1,400, reversing its ordering against two other archetypes. And the saturation score imputes a neutral value for any component it cannot measure, flattering exactly the archetypes most closed to small accounts. Archetype-level medians are robust; archetype-by-tier cells are not.

7.3 The amplification network

Mentions were parsed from post text and quote targets joined from the raw payload, giving 9,232 directed edges (5,719 mention, 3,513 quote) over 4,087 hub accounts, 221 of which clear the minimum inbound threshold. Posts carrying four or more mentions were treated as follow-list dumps and excluded from the focused counts; self-quotes were removed from the edge list. Against a no-reference baseline of 939 median likes (n=13,999), posts mentioning another account without quoting run 610 likes (0.65×, n=1,900), posts quoting without mentioning run 1,050 (1.12×, n=2,489), and posts doing both run 608 (0.65×, n=304). On the separate self-versus-other partition, quoting someone else returns 1,009 (1.11×, n=2,797) against a 906-like core baseline, self-quoting 702 (0.77×, n=1,795) and not quoting 925 (1.02×, n=14,100). The two partitions cannot be ranked against one another: 1,745 of the 1,900 mention-only posts also sit inside the 14,100 no-quote bucket, and the tables use different baselines. Within its own partition each direction is stable, and the mention association holds in every follower band, running 0.48×, 0.55×, 0.60×, 0.65×, 0.68× and 0.67× from the smallest band to the largest — attenuating as accounts grow, though not monotonically at the top, which is a correction the verifier filed against the original claim of monotonicity.

Hub-level lift is heavily confounded by who refers: across the 48 hubs with at least fifteen referencing posts, the Spearman correlation between a hub's lift and the median follower count of the referring accounts is $\rho=0.69$ ($p=5.2\times10^{-8}$). Holding the referring author in the 5K–50K band, the ordering nevertheless survives and widens — @claudeai 2.27× and @anthropicai 1.89× against @apob_ai 0.25× and @perplexity_ai 0.39× — so consumer-facing vendor and airdrop-adjacent accounts are associated with depressed outcomes even after size matching, while model-layer accounts are not.

A target list of 38 accounts was derived by scoring each candidate on relevance (0.35), reachability (0.25), measured lift (0.20) and authority (0.20), where authority saturates in the count of distinct focused inbound authors and lift falls back to a flat constant when every slice is below n=15. Three qualifications attach. Four of the top ten carry an authority-only evidence tier, so their lift is imputed rather than measured; three of the top twelve have no sub-10K account on record that has reached them; and the community structure is unstable, the graph yielding seventeen communities of size three or more at the shipped label-propagation seed [Raghavan et al., 2007] of which twelve were displayed, while other seeds move the count between thirteen and seventeen and shift cluster lifts by up to 0.4×.

7.4 Growth trajectories

Creators with at least eight posts spanning three months and sixty days were split into an early window (first 40% of posts) and a late window (last 40%), with likes normalised by the corpus median for the calendar month so that baseline drift cannot manufacture decliners. Of 454 qualifying creators, 55 were classified improvers (late-to-early ratio ≥1.5), 49 decliners (≤0.667) and 350 steady, over 8,074 posts; improver medians run 653 likes early to

1,141 late, decliner medians 1,875 to 532. Table 7.3 reports the behaviour deltas alongside a thirty-seed permutation placebo that shuffles each creator's posts to destroy time order while keeping the labelling rule.

Table 7.3 — Behaviour deltas between improvers (n=55) and decliners (n=49), with placebo-shuffle verdicts. DiD is mean improver change minus mean decliner change; share features in percentage points. Effect sizes are Cliff's delta.

Feature	Observed DiD	p	Cliff's δ	Placebo DiD	Empirical p	Verdict
@-mentions someone	−12.3 pp	0.0025	−0.311	−2.2 pp	0.000	survives
Video post	+13.9 pp	0.0027	+0.329	+4.0 pp	0.000	inflated
Quote post	−8.5 pp	0.0183	−0.254	−3.4 pp	0.133	inflated
Characters	−45.0	0.0496	−0.224	−8.75	0.133	inflated
Bookmark ratio	+0.101	0.0074	+0.305	+0.042	0.167	inflated
Reply ratio	−0.023	<0.0001	−0.529	−0.026	0.633	artefact

Methodological note

The denominator confound. The reply-ratio result was the module's highest-confidence headline and its only large effect [Cliff, 1993]. It is a selection artefact. Reply ratio is replies divided by likes, and the improver/decliner split is itself defined on likes, so any likes-denominated ratio must move oppositely between the groups by arithmetic. The placebo reproduces it at full strength (−0.026 against an observed −0.023) and 63% of null runs match or exceed it; the numerator points the other way, decliners' absolute replies having fallen (per-creator median −49.5) while their likes fell faster. Bookmark ratio and amplification carry the same defect in weaker form, and the confound also voids any cross-group comparison of month-normalised likes, which stand at 1.24 for improvers against 0.73 for decliners by construction.

Two qualifications remain. Forty features were tested without multiplicity correction; the Bonferroni family-wise threshold is $p < 0.0013$, which only the artefactual reply-ratio row clears, so every other p-value in Table 7.3 is suggestive rather than established absent false-discovery control [Benjamini & Hochberg, 1995]. And the cohort regresses strongly to the mean [Galton, 1886]: the lowest early-performance quartile has a median late-to-early ratio of 1.32 and is 22.8% improvers, the highest a ratio of 0.59 and 21.1% decliners. What survives is narrow — improvers cut the share of posts carrying an @-mention by about twelve percentage points more than decliners did, and raised their video share, though roughly 29% of the latter effect is generated by the placebo alone.

7.5 Topic × format × tier whitespace

The whitespace matrix crosses eighteen rule-assigned topic clusters with format and follower tier, suppressing every cell below n=15: 239 of 626 possible cells survive in the full analysis universe (90.8% of posts) and 96 of 288 in the sub-25K universe (87.2%). Each surviving cell receives an opportunity score equal to its performance percentile minus its supply percentile, where performance is 0.65 × the percentile of median likes plus 0.35 × the percentile of median bookmarks and supply is the percentile of cell size. The score is bounded at ±100 and expresses payoff relative to crowding among winners — explicitly not a success rate.

Because a cell can clear the floor on repeat posts from one account, the distinct-handle count and the maximum single-handle share are published as guards, and any cell above a 0.25 handle share should be read as an account effect. The rule proved necessary but not sufficient: one course-resource demo cell passed it while containing thirteen near-identical reposts of a single post, collapsing to n=12 once same-handle duplicate text is merged.

Table 7.4 — Leading under-supplied cells in the sub-25K universe, by opportunity score. The format-tier lift compares each cell against the same format at the same tier and is the correct figure for a claim about topic; the tier lift, shown for contrast, inherits the format's own advantage. Six of the module's eight highest-scoring cells are listed; two that fall in the 0–1K tier (Claude Code craft × single post, opportunity 68.6, n=16; AI commentary × video, 54.3, n=16) are omitted, because a cell of sixteen posts from the thinnest tier in the corpus carries no usable precision.

Topic × format × tier	n	Handles	Max share	Median likes	Lift vs tier	Lift vs format×tier	Opportunity
Claude Code craft × demo × 5K–25K	17	17	0.06	3,645	5.56×	3.69×	87.5
Agents & automation × demo × 5K–25K	15	15	0.07	1,077	1.64×	1.09×	75.0
Money & AI business × video × 5K–25K	21	16	0.14	1,487	2.27×	1.51×	65.3
Money & AI business × single post × 5K–25K	17	17	0.06	1,052	1.61×	1.23×	62.9
Agents & automation × quote post × 1K–5K	17	17	0.06	1,787	2.72×	1.07×	61.7
Model releases & benchmarks × video × 1K–5K	23	19	0.13	1,686	2.57×	2.00×	53.0

The gap between the two lift columns in Table 7.4 is the module's largest defect: dividing by the tier's all-format median lets every cell inherit its format's advantage, which was then attributed to the topic. Recomputed against the same format at the same tier, the agents-and-automation quote-post cell falls from 2.72× to 1.07×, its format's own baseline, and sev-

eral cells reported as flat prove to sit below their format peers. Only the Claude Code demo cell keeps a substantial margin, and it too carries an objection: five of its seventeen posts, including its two highest-liked, are misclassified, and cleaning to the twelve genuine posts holds the direction while dropping the cell below the suppression floor, so its figures should not be quoted. Two further defects were sustained. Topic assignment is first-match-wins over a priority ordering, making low-priority topics residual buckets: the Spearman correlation between priority rank and opportunity score is 0.554 (p=0.021), so apparent under-supply is partly definitional. Two topic regexes also lacked word boundaries, admitting learn and yearn into the money topic and issue into the news topic. The crowded end of the ranking is less fragile: image-prompt recipes in image form at 5K–25K (n=108, 39 handles, median 234 likes, opportunity −81.8) and tool listicles in thread form (n=127, median 354, opportunity −76.8) remain below their format peers on the corrected baseline, at 0.39× and 0.66×.

7.6 Brand share of voice

A dictionary of 98 vendor and model names was matched against post text; 68.2% of the corpus mentions at least one and 85 brands cleared the reporting threshold. Table 7.5 is computed on the organic slice — 18,117 builder and mixed posts, tool-listicle posts removed — against a baseline of 932 median likes and 412 median bookmarks.

Table 7.5 — Leading brands by organic volume. Lifts are against the organic baseline of 932 median likes and 412 median bookmarks (n=18,117); intervals are bootstrap 95% intervals on the likes lift. Small-account lift is against a 664-like baseline for accounts under 25K followers.

Brand	Category	Organic posts	Share	Likes lift	95% CI	Bookmark lift	Small-account lift
Claude	model	8,187	45.2%	1.22×	1.17–1.26	1.38×	1.43×
ChatGPT	model	2,738	15.1%	0.74×	0.70–0.77	0.92×	0.75×
Anthropic	model	2,672	14.8%	1.20×	1.12–1.27	0.99×	1.31×
Claude Code	coding	2,510	13.9%	1.57×	1.45–1.67	2.12×	1.86×
OpenAI	model	1,410	7.8%	0.84×	0.78–0.90	0.65×	0.92×
Gemini	model	1,017	5.6%	0.89×	0.82–0.95	0.94×	0.93×
Codex	coding	899	5.0%	0.99×	0.91–1.08	1.10×	1.07×
MCP	agent-autoů mation	588	3.3%	0.91×	0.83–1.04	2.50×	0.88×

Category rollups separate likes from saves cleanly: coding (eleven brands, 3,426 posts) carries 1.32× likes and 1.80× bookmarks, while agent-automation (nine brands, 880 posts) carries 0.78× likes but 2.03× bookmarks, as do the infra and product categories. Naming a tool is associated with better outcomes than naming none, but only briefly: brand-free posts run 0.96×, one vendor 1.12×, two 1.03×, three to four 0.91×, five to seven 0.88×, eight or more 0.50×, the last bucket being entirely tool-listicle posts. The sequence is not monotone in either cut, and once the Claude family is excluded the brand-free bucket leads outright at 1.15×; the endpoints are stable but the shape should not be over-read.

Methodological note

The pair-lift confound. Co-mention payoff was originally benchmarked against the global organic baseline, so any pair containing Claude Code inherited Claude Code's own 1.57× solo lift and the pairing appeared to cause it. Conditioning on Claude Code already being present reverses most of these: Claude Code × Codex is 0.84× (95% CI 0.71–0.94, n=371), a significant decrement reported as a 1.31× win; × Kimi 0.89×; × Figma 1.00×; × Ollama 1.14× (0.72–1.47); × Obsidian 1.40× (0.92–2.28), the only pair pointing upward and with an interval still spanning 1.0. Co-mention affinity — how much more often two names appear together than chance predicts — is better behaved: CrewAI × LangGraph co-occur at 244× expectation (n=18) and Llama × Mistral at 103× (n=29). Neither carries an engagement premium.

7.7 Timing and cadence

This module was graded unsound on a sort-order tautology. The source file arrives sorted by likes descending, and a stable lexsort within handle and day preserved that order, so a within-day cumulative count labelled the highest-liked post of each day as the first post of the day in 803 of 803 two-post days, while it was the chronologically earlier post only 49.4% of the time; the reported decay across same-day position (1.035, 0.661, 0.562, 0.430) restated the sort key. After replacing the sort with a chronological mergesort, adjusted lifts by same-day position are 1.006 (n=17,601), 0.945 (n=925), 0.766 (n=122) and 0.739 (n=44), and a paired test on the 803 two-post days gives a median second-to-first ratio of 0.876 (Wilcoxon p=0.0013) — a penalty near 12% rather than 35%. The same patch moved the same-day bucket of the gap-since-previous-post table from 0.632 to 0.912.

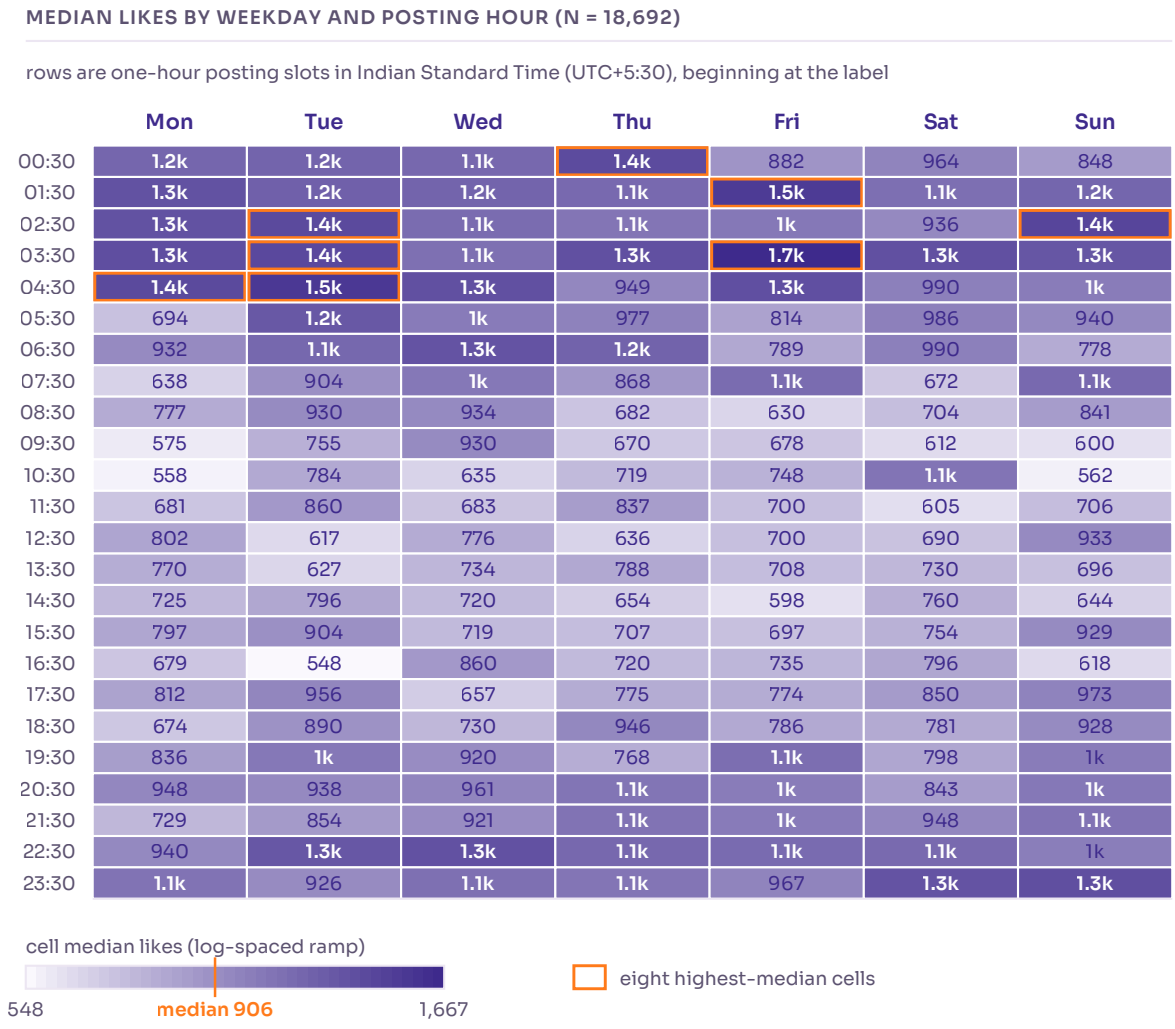


Figure 7.2 — Median likes by weekday and posting hour (IST), builder and mixed posts (n = 18,692), on a log-spaced colour ramp against the population median of 906 likes. Cells below n=15 are suppressed. The adjustment removes follower count only, not topic, format or account age.

Table 7.6 — Engagement by IST daypart, builder and mixed posts (n=18,692), against a baseline of 906 median likes. The median-followers column is the reason the adjusted-lift column cannot be read as a timing effect.

Daypart (IST)	n	Share	Median likes	Raw lift	Adjusted lift	Median followers
Early morning 05:30–09:30	2,098	11.2%	895	0.99×	0.99×	45,326
Morning 09:30–12:30	1,863	10.0%	688	0.76×	0.81×	29,725
Afternoon 12:30–17:30	3,428	18.3%	713	0.79×	0.84×	39,110
Evening 17:30–21:30	4,444	23.8%	879	0.97×	0.97×	48,103
Night 21:30–01:30	4,317	23.1%	1,054	1.16×	1.13×	68,982
Late night 01:30–05:30	2,542	13.6%	1,234	1.36×	1.34×	74,500

The raw late-night-over-morning ratio in Table 7.6 is 1.79×, inflated by follower composition the adjustment fails to remove. Late-night posts come from accounts with a median of 74,500 followers against 29,725 in the morning band, but the follower model's OLS slope is only 0.106 (r^2=0.026), so the residual adjustment strips out just a 1.10× factor and leaves 1.66×. Stratifying properly, the median within-follower-decile ratio is 1.345 and the within-tier ratio 1.31–1.32 in the two largest well-populated tiers. **The corrected effect is therefore approximately 1.3×, not 1.79×.** Two supporting observations survive: hour profiles of accounts below and above 50K followers agree at ρ=0.658 (p=0.00047), so the ranking of hours does not depend on account size, and after the sort fix a four-to-seven-day gap since the previous post is highest at 1.010 against 0.937 for a one-day gap.

Methodological note

What remains confounded and must not be reported as a timing finding. The apparent decay across consecutive-day streaks (1.014, 0.943, 0.884, 0.887, 0.826) is a composition effect: median followers run 46,270, 76,494, 93,892, 88,319 and 88,320 across those buckets, roughly doubling after the first, and centring residuals within creator on the 585 creators who have both day-one and later-day posts flattens the curve to 1.000, 1.000, 0.992, 1.035, 0.833. The bookmark-ratio inversion favouring afternoon hours is a denominator artefact — absolute median bookmarks are highest at night (473) and late night (465) and lowest in the morning (304), so the afternoon's high saves-per-like ratio buys no additional saves. Cadence buckets are floored by construction, since posts per active week cannot fall below 1.0, which is where the winning bucket sits, and cadence here means winners harvested per week rather than posting frequency (ρ=0.513 with corpus volume).

Read conservatively, the timing evidence supports one association of moderate size — posts published in the IST night and late-night windows, corresponding to United States daytime, carry roughly 1.3× the likes of comparable posts from accounts of the same size — and supports it equally for small and large accounts. Every finer claim either did not survive the sort-order repair or remains confounded with account size, format mix, or the arithmetic of its own denominator. This is the study's clearest illustration of the pooled-tier problem visible in Table 7.1: a comparison made across a variable correlated with account size is partly a comparison of accounts [Simpson, 1951]. That the ranking system is itself public [Twitter, 2023] does not resolve this, since an observational corpus of winners cannot separate what the ranker did from who was posting.

8. Adversarial Verification: Protocol, Yield and Limits

The analyses in Chapters 5 to 7 were produced by language-model agents operating on the corpus of Chapter 4. This chapter documents the protocol used to attack those analyses before any conclusion entered the report, what the attack found, and — more importantly for anyone contemplating a similar pipeline — what it could not have found. The objection yield is high enough to be a finding in its own right: an unverified version of this study would have shipped at least a dozen conclusions wrong in direction, not merely imprecise in magnitude.

8.1 Motivation: why an unverified automated analysis is not evidence

An agent that writes and runs its own analysis code is fluent at every stage a reader can inspect and unreliable at the stage a reader cannot. Across the twelve modules the analysts volunteered 111 caveats unprompted. None identifies any of the twelve fatal defects below; every one was found by an outside process that recomputed the numbers from the raw table.

Three failure classes make this predictable. An analyst that both defines the grouping variable and picks the outcome metric may choose a metric sharing algebra with the grouping rule, fixing the result before the data is read. An analyst that builds a feature by regular expression rarely inspects what it matched, so the flag measures something adjacent to, sometimes opposite to, the intended construct [Cronbach & Meehl, 1955]. And an analyst working through many cuts without pre-registration finds a large effect somewhere, with no recoverable path from data to headline [Gelman & Loken, 2014]. Fluency aggravates all three [Ioannidis, 2005].

> **The analysts wrote 111 caveats about their own work. Not one names a defect the verifiers later rated fatal.**

8.2 Architecture

Twelve analyst agents were each given one thematic module — the twelve are named in Table 8.1 — together with the full corpus and permission to write and execute Python against it. Each produced a JSON module of numbered findings, recommended plays and caveats.

Each module was then handed to an independent verifier agent, instantiated fresh, with no access to the analyst's reasoning trace. Its instructions had four parts.

1. **Re-derive, do not re-read.** Recompute every headline number directly from `data/viral_posts.csv` using the verifier's own feature extraction. Re-running the analyst's script was disallowed, because a script reproduces its own bugs faithfully.
2. **Hunt named failure modes.** Test by name for circular metrics, order dependence, duplicate-text contamination, pooled-tier confounding, regex leakage, suppression-floor violations, and quantities cited in prose that do not exist in the module output.
3. **Grade every objection** as fatal (direction or recommendation wrong), major (number or framing unsupported, direction survives), or minor (imprecision, missing evidence, robustness gap).
4. **Repair where repair is legitimate.** The verifier held write access to the module and could patch and re-run it, provided the patch was additive — existing keys and values preserved — so a repair could not silently overwrite an auditable result. Where repair meant redesigning the analyst's taxonomy or de-duplication policy, the verifier was required to report and stand down.

Verifiers logged 163 independent re-derivations and closed each module with a verdict of sound, needs_caveat or unsound. No module was rated sound.

Methodological note

On the repair boundary. Of the 139 objections, 49 carried a code patch and a module re-run; the other 90 were reported without a code change, most often because the module output was already correct and the error lay in the analyst's prose. A verifier that fixes everything it touches destroys the record of what was wrong. Those 90 are why the claims in this report could be re-quoted from the module JSON rather than the analyst's summary.

8.3 Yield

The twelve analysts produced 115 findings and 95 recommended plays. The verifiers filed 139 objections — 1.21 per finding — of which 12 were fatal, 57 major and 70 minor. Seven modules closed at needs_caveat and five at unsound. Table 8.1 gives the distribution; Figure 8.1 maps it by severity.

Table 8.1 — Objections by module and severity. Re-derivations are independent recomputations from the raw CSV logged by the verifier. Ordered by fatal count, then total objections.

Module	Findings	Re-deriv.	Fatal	Major	Minor	Total	Verdict
whitespace_matrix	10	13	2	8	5	15	unsound
antipatterns	10	12	2	6	4	12	unsound
growth_trajectories	10	11	2	4	5	11	unsound
psych_triggers	8	11	2	1	6	9	needs_caveat
timing_cadence	10	14	1	6	3	10	unsound
repurposing_engine	10	16	1	4	6	11	unsound
brand_share_of_voice	10	10	1	4	5	10	needs_caveat
hook_linguistics	10	12	1	3	5	9	needs_caveat
content_pillars	10	15	0	5	10	15	needs_caveat
engagement_funnel	10	18	0	6	8	14	needs_caveat
amplification_network	9	18	0	6	8	14	needs_caveat
creator_archetypes	8	13	0	4	5	9	needs_caveat
Total	**115**	**163**	**12**	**57**	**70**	**139**	**0 sound**

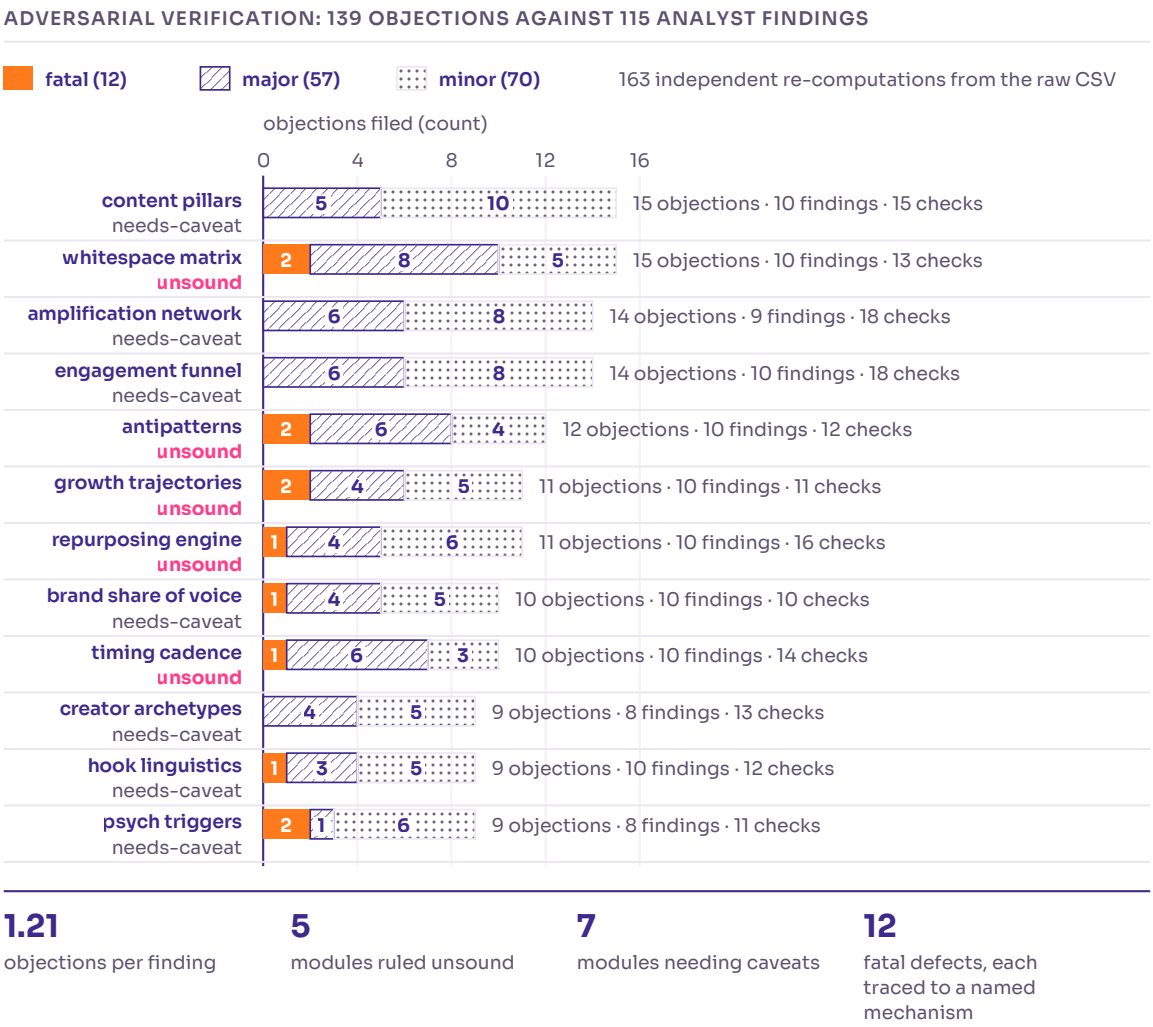


Figure 8.1 — Objection severity by module. Fatal objections concentrate in the modules that constructed the most derived features; no module closed as sound.

Two structural patterns recur. Fatal objections cluster in modules that built many derived variables — regex classifiers, composite scores, trajectory labels — not in those reporting plain cuts of existing columns. And many objections concerned prose rather than code: one analyst quoted audience-split lifts of -17.4%, -8.3% and +27.7% where the module output held -17.1%, -6.6% and an explicit null, the consumer cell (n=27) having been suppressed by the module's own floor. The sample sizes matched exactly; the effect sizes did not.

8.4 Four worked failures

Each case carries the diagnostic that exposed it, so the same test can be applied elsewhere.

8.4.1 The sort-order tautology

The timing module reported that a creator's first post of a day outperforms later posts, with a clean monotone decay in adjusted lift (1.035, 0.661, 0.562, 0.430). The corpus CSV is stored sorted by likes descending; the module sorted by handle and date with a stable sort, preserving the incoming order within ties, then indexed within the day by row position. In 803 of 803 two-post days the highest-liked post had been labelled "first"; it was the chronologically earlier post only 49.4% of the time. The finding stated that the post with more likes has more likes. Re-sorting on `hour_utc` flattened the decay to 1.006, 0.945, 0.766, 0.739, and repaired an adjacent gap-since-previous-post bucket with the same root cause (0.632 to 0.912).

Diagnostic. Validate any within-group ordering against an independent timestamp on a subset small enough to check by hand. A pipeline that sorts a file which arrived sorted by the outcome should be assumed guilty until shown otherwise.

8.4.2 The copy-paste spam artifact

The psychological-trigger module reported that scarcity and time-bound language multiplies the reply ratio by 9.74x (n=671) at a 6% cost in reach. Fingerprinting the text, the verifier found 30.0% of the scarcity posts were verbatim repeats: one promotional template appeared 128 times in the corpus, 116 inside the scarcity group, across 30 accounts, one of which posted it 28 times. Restricted to first copies, the reply lift fell to 2.41x and the reach penalty reversed (0.941x to 0.970x), so the stated cost disappeared. The analyst's corroboration — manual inspection of the highest reply-ratio posts — was itself contaminated: 12 of the top 30 belonged to one account.

Diagnostic. Report a duplicate share beside every group statistic; a normalised fingerprint over the first 120 characters suffices. So does the pair of guard fields the whitespace module was made to emit after a comparable failure, `n_distinct_handles` and `max_handle_share`, with any cell above 0.25 read as an account effect rather than a content effect.

8.4.3 The circular metric with a shared denominator

The growth module's highest-confidence finding was that reply ratio is the sharpest discriminator between improving and declining creators and an early warning that inverts before likes collapse. Reply ratio is replies divided by likes; improvers and decliners were defined on likes, so the groups must diverge on any likes-denominated ratio by arithmetic alone. The verifier built a 30-seed permutation placebo — shuffle each creator's posts to destroy time order, keep the identical labelling rule — which reproduced the effect at full strength (placebo median -0.026 against observed -0.023), giving an empirical p of 0.633 where the module reported $p < 0.0001$. The causal story was also backwards: decliners' replies fell too, their likes merely fell faster. The early-warning claim is impossible on its face, since the ratio is computed from the same-window likes it was said to lead.

The defect recurred in the repurposing module, where a composite "blog score" placed 32% of its weight on the percentile of bookmark ratio and was then reported to correlate with bookmarks at Spearman +0.509 — the score reading back its own input, with which it correlates at 0.789. Removing that term dropped the headline bookmark multiple from 21.69x to 4.05x and the correlation to 0.221; on raw metrics the claimed tension inverts entirely, Spearman(likes, bookmarks) = +0.608.

Diagnostic. Write the algebra of every reported ratio beside the algebra of the group definition; if a variable appears in both, the comparison is not evidence. Where the entanglement is subtle, a label-preserving permutation placebo settles it: if a shuffled null reproduces the effect, there is no effect.

8.4.4 The has_link construct-validity failure

The most consequential single defect was a flag that measured the wrong construct. A naive test for an `http` substring in post text was read as "the post links out". X appends a t.co URL to every natively attached photo or video, so the flag fired on 48.2% of media posts against 11.4% of text posts and inherited the media lift wholesale. It flagged 8,058 posts (36.9%) and put outbound links at **+9.9%** on median likes.

Re-derived from the API url entities, counting only off-platform hosts, a genuine outbound link appears on 2,626 posts (12.0%) with a median of 704 likes against 1,046.5 without: a lift of **-32.7%**. Holding follower tier fixed the estimate is -33.8%, negative in all six tiers with ratios from 0.43 to 0.74, so the direction does not depend on the combining rule. On the builder-plus-mixed analysis population the corrected figures are -26.7% raw and -27.9% tier-stratified (n=2,354), with r = -0.09 against log10(likes) where the naive flag gives r = +0.06 — a difference of 42.6 percentage points and of sign. An independent cell-matched estimate in the funnel module recovers -22.0% on likes and -27.0% on views across 23 format-by-tier cells (n=2,331 link posts) with bookmarks up 55.9%, a pattern consistent with distribution suppression.

Diagnostic. Print a random sample of the rows a derived flag fires on and read them, then cross-tabulate the flag against every plausible confound — here, media attachment — before estimating anything. A binary feature whose prevalence differs four-fold between two obvious sub-populations is describing those sub-populations, not the construct it was named for.

8.5 What the protocol did not and cannot catch

The yield above is no warrant of correctness for what survived. Four limits are structural.

- **Shared priors.** Analysts and verifiers are instances of the same model family reading the same instructions. A convention neither questions — that median likes is the outcome worth optimising, that a like floor defines "high engagement", that eight months form a single regime — is invisible to both. Adversarial framing induces independent computa-

tion, not independent belief; nothing among the 139 objections challenges the choice of outcome variable.

- **The corpus's own selection.** Every post collected cleared a like floor, and no verifier can recover the posts never collected, so no protocol operating on this table converts a within-winners contrast into a probability of success. The verifiers policed the arithmetic of "among posts that already performed"; they had no means to police the leap to "this will work".
- **Anything outside the CSV.** The verifier's authority was to recompute from `viral_posts.csv`, so claims depending on what is absent from it — deleted or suspended accounts, follower counts at posting rather than scrape time, platform ranking changes during the window, whether a save became a click — lie outside its reach. The corpus records `is_reply` as identically zero, so no reply-level behaviour was verifiable at all.
- **No third layer.** The verifiers were themselves unverified; their re-derivations reproduce to the decimal, which is reassuring about arithmetic and silent about judgement. Where a verifier declined to repair — 90 of 139 objections — the correction rests on the authors re-quoting from the module output, a manual step with no automated check behind it.

Multiplicity is a fifth, partly addressed exposure: the modules collectively evaluated several hundred cuts, and while one verifier added a Bonferroni threshold from its own feature list, no correction was applied across modules and no false-discovery control [Benjamini & Hochberg, 1995] was run over the pooled tests. Small p-values here describe the corpus; they are not evidence against a pre-specified null.

8.6 Cost and practicality

The full run used 25 agents — twelve analysts, twelve verifiers and one strategist adjudicating the surviving contradictions — consuming approximately 2.9 million tokens over roughly 48 minutes of wall-clock time, the analyst and verifier stages each running in parallel. These are operational figures recorded from the run, not quantities derived from the corpus. Verification roughly doubled the token cost of the analysis stage and added one round of latency; against that it returned twelve fatal defects, several already consumed by a downstream artifact and shipping. The relevant comparison is not verification cost against analysis cost but against the cost of publishing a directional error — here, an instruction to link out freely that the corrected measurement reverses into a reach penalty of roughly 27% to 34%.

8.7 Recommendations for practitioners

Five practices carried nearly all of the protocol's value.

1. **Separate the agent that computes from the agent that checks, and forbid the checker from re-running the analyst's code.** Independent re-derivation from the raw table is the only step that catches a bug the code reproduces faithfully.

2. **Give the verifier a checklist of named failure modes, not a mandate to "review".** Every fatal objection here belongs to a class named in advance: order dependence, duplicate contamination, circular metrics, regex leakage, pooled-tier confounding [Simpson, 1951], and prose-to-output provenance failures.
3. **Require machine-readable output, and re-quote every published number from it.** Roughly two thirds of these objections were prose errors against correct code, cheap to catch only when the two can be compared mechanically.
4. **Constrain repairs to be additive.** Overwriting the original number removes the reader's ability to audit the correction, and guard fields — duplicate share, distinct-handle counts, per-tier cells, placebo p-values — should be emitted permanently, not computed once during an argument.
5. **Publish the objections.** A protocol that produces 139 objections and reports only the survivors has converted an audit trail into a marketing claim. The severity distribution is a property of the analysis, and readers are entitled to it.

A closing caution: surviving an adversarial pass makes a conclusion arithmetically defensible, not causally established. Every result in Chapters 5 to 7 remains an association observed in a winners-only sample. Verification improved the numbers; it did not change what kind of evidence they are.

9. Discussion

The preceding chapters reported a large number of associations. This chapter asks which survive the corpus's own limitations, what structure they form together, what a small account could do with them, and where the design fails. Every proposition below names its population and sample size, and each carries the implicit prefix among posts that already cleared a like floor.

9.1 What the evidence supports

9.1.1 Proposition 1: the audience segment is a precondition, not a covariate

Consumer chatter and practitioner posts are two markets sharing a vocabulary, not two ends of a continuum. Consumer posts (n=3,172) reach a median of 2,011 likes against 716 for builder posts (n=3,370), a ratio of 2.81; on saves the ordering reverses, 218 median bookmarks against 845. As a save rate the gap is an order of magnitude: 7.4 bookmarks per 100 likes against 131.6 (Mann-Whitney U, p below double precision). Pooling them averages contradictory regimes under weights that are an artefact of the query design, so every actionable result here is computed on the builder-plus-mixed population (n=18,692) or on the sub-25K stratum within it (n=6,521).

9.1.2 Proposition 2: reach and saves are produced by opposite shapes

In the sub-25K stratum, posts of 1–25 words reach a median of 1,380 likes, +112.4% against the stratum median of 650, while their bookmarks sit at −12.0% against a baseline of 347 (n=1,062). Posts above 400 words invert both signs: −16.9% and +21.6% (n=371). The strongest single feature behaves the same way: a post written as one unbroken line is +170.7% on likes and −37.2% on bookmarks against posts carrying a line break (n=874; bootstrap 95% CI on the likes ratio 2.29–3.06; significant after Benjamini–Hochberg adjustment) [Benjamini & Hochberg, 1995]. Likes and bookmarks correlate as co-outcomes, r=0.62 between their log10 transforms (n=18,692), but share only 38% of their variance.

9.1.3 Proposition 3: one format is positive on both axes, and only one

Of eight coded formats, the screen-recorded product demonstration alone clears the stratum baseline on both metrics: +58.2% on likes and +290.2% on bookmarks (n=201), a save rate of 127.6 per 100 likes against a baseline of 51.1 (Figure 9.1, Table 9.1). One caveat is substantial: a second operationalisation of the construct — a keyword-based `demo` flag rather than the format label, n=350 — reproduces the save effect (+180.3% on bookmarks) but not the reach effect (+11.5% on likes, not significant at FDR 0.05). The save half of the proposition is robust to how a demonstration is identified; the reach half is not.

9.1.4 Proposition 4: pooled comparisons partly measure who adopts a format

Format choice is correlated with account size, so a pooled lift mixes a format effect with a selection effect. Isolating the stratum moves the demonstration from +0.4% pooled (n=513) to +58.2% (n=201), and the carousel from +7.7% pooled (n=945) to −17.2% (n=308) — a sign reversal. This is Simpson's paradox operating on follower count [Simpson, 1951]: no pooled comparison should be read as advice to a particular account until it has been stratified.

9.1.5 Proposition 5: the reliable levers are subtractive

Twelve of the 15 features tested in the stratum are significant after FDR adjustment, and the largest are negative. A hashtag is associated with −35.0% on likes and −76.4% on bookmarks (n=529); four or more emoji with −38.0% and −34.1% (n=421); an explicit reply-gate with −29.1% on likes against +54.8% on bookmarks (n=173). These hold in direction across follower tiers rather than resting on one, and are cheap to act on because acting means deleting.

9.1.6 Proposition 6: an outbound link is associated with lower reach, and the naive measure says the opposite

A substring test for `http` in the post body flags 36.9% of the corpus and returns a link lift of +9.9% (p=1e-07). It is not measuring outbound linking: X appends a t.co URL to the body of every natively attached photo or video, so the flag fires on 48.2% of media posts against 11.4% of text posts and inherits the media effect. Re-derived from the API url entities with off-platform hosts only (column `ext_link`, n=2,626, prevalence 12.0%), the association is −32.7% on raw medians; within follower tier it is −33.8% under tier-weighted combination and −39.4% under an unweighted mean, all six tier ratios being negative and running from 0.43 to 0.74. On the builder-plus-mixed population it is −26.7% raw and −27.9% tier-stratified (n=2,354), r=−0.090 against log10(likes). Naive and corrected estimates differ by 42.6 percentage points and by sign. A cell-matched module (23 cells, 2,331 linked posts) locates the penalty in views (−27%) rather than in per-view response — consistent with distribution suppression, though not establishing it.

> **Methodological note**
>
> The link result illustrates a general hazard: a column whose name describes the intended construct, whose values describe something else, and whose error is large enough to reverse the sign of the effect. It surfaced only because a verifier re-derived the number from the raw payloads.

9.1.7 Proposition 7: format returns drift within the observation window

Indexing each format's median against the all-format median inside the same half of the window, so that engagement maturation cancels, video falls from 2.20 to 1.20 while its share of supply rises from 13.9% to 21.0%; the demonstration rises from 1.17 to 1.81 at a flat

3.1% share of supply. The movements suit a supply-side account in which returns fall where imitation is cheap, though eight months is a short window and corpus composition changes across it.

9.1.8 Proposition 8: the levers are real and small relative to the variance around them

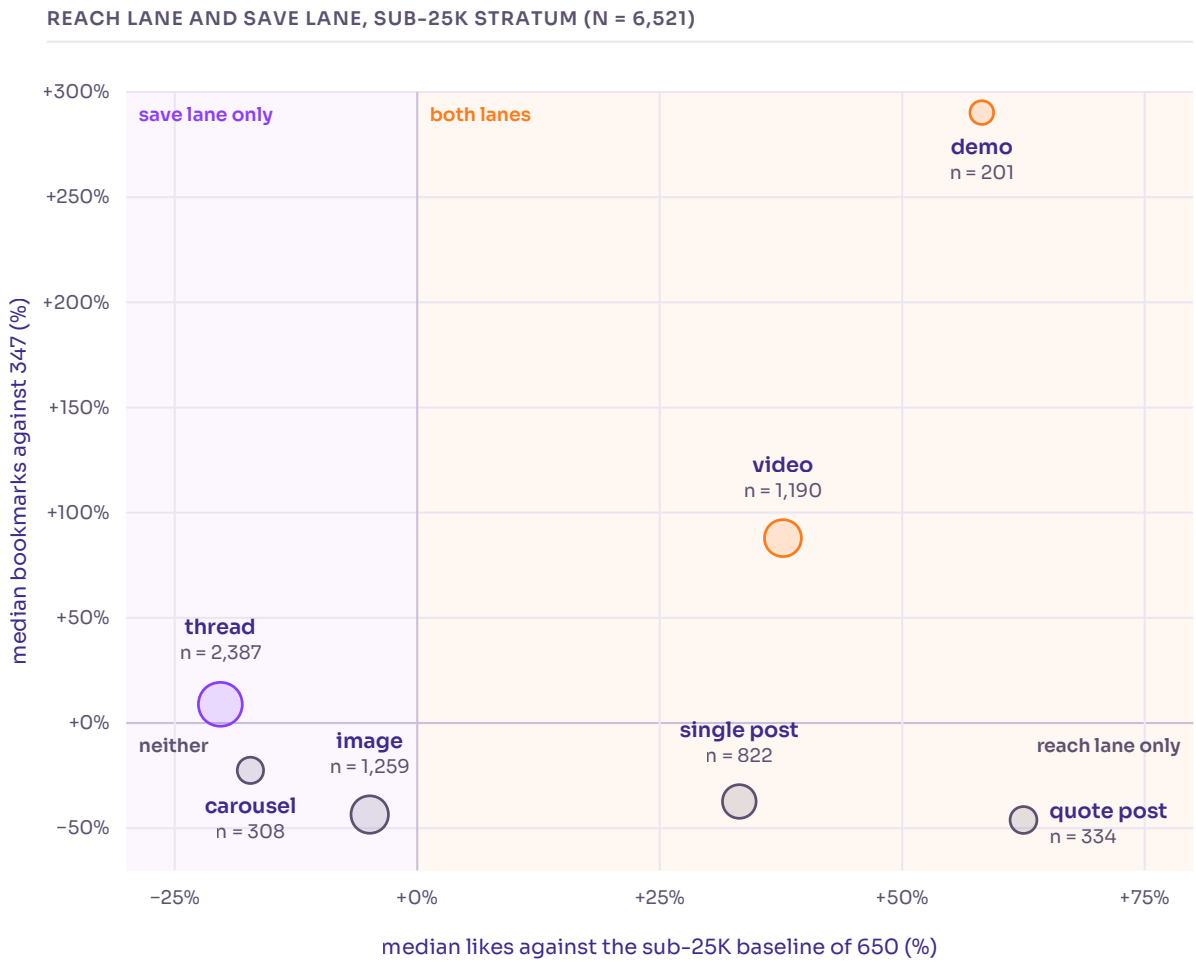
Outcomes remain extremely unequal after conditioning on success: in the stratum the Gini coefficient on likes is 0.725 and the top 5% of posts take 50.0% of all likes, while across creators the Gini is 0.805 and the top decile takes 70.9%. The largest authored correlate in the feature set — the number of lines in a post — explains 2.1% of the variance in $\log_{10}$(likes) ($r=-0.145$, $n=18{,}692$), and no other authored feature exceeds 1.7%; account size, the one variable the author does not choose on the day, reaches 5.4%. These associations shift the distribution of outcomes rather than determine them, as the experimental literature on cultural markets would predict [Salganik, Dodds & Watts, 2006].

9.2 The two-lane model

Propositions 2 and 3 together imply a structure rather than a ranking. Because likes and bookmarks are produced by opposed shapes, an account optimising a single composite post type lands where neither is achieved: in the stratum the 51–200 word bands are negative on likes (−10.2% and −20.5%) and no better than flat on bookmarks (−6.5% and +7.6%).

> **The compromise post is not a hedge between two objectives. In this corpus it is the one shape that fails at both.**

What the data supports instead is two explicitly separate lanes. The **reach lane** is short, single-line, media-supported, free of body links, and names a specific tool in its opening line; it is measured in likes and views. The **save lane** is long and reference-shaped — threads and demonstrations — and is measured in bookmarks per 100 likes. Table 9.1 assigns each format to a lane by its own numbers, and the assignment is not arbitrary: the quote post is the strongest reach format in the stratum (+62.5% likes, $n=334$) and the weakest save format (9.6 bookmarks per 100 likes), while the thread is the weakest reach format (−20.3%) and the second strongest save format (76.6 per 100 likes).



Marker area is proportional to the square root of n. Omitted as too thin to plot: @grok post (n = 20). Only the screen-recorded demo occupies the upper-right quadrant — the one format that wins reach and saves at once.

Figure 9.1 — Formats positioned by reach and save performance, sub-25K stratum (builder and mixed, followers under 25,000; n=6,521). Horizontal axis: median likes against the stratum baseline of 650. Vertical axis: median bookmarks against the baseline of 347. Marker area is proportional to the square root of n. The @grok format is omitted as too thin to plot responsibly (n=20). Only the screen-recorded demonstration occupies the upper-right quadrant.

Table 9.1 — Format profile in the sub-25K stratum (builder and mixed, followers under 25,000; n=6,521). Lifts are against the stratum medians of 650 likes and 347 bookmarks. Lane assignment follows the sign pattern, not a threshold rule.

Format	n	Supply %	Likes lift	Bookmarks lift	Bm / 100 likes	Lane
Demonstration	201	3.1	+58.2%	+290.2%	127.6	Both
Video	1,190	18.3	+37.7%	+87.9%	71.6	Both
Quote post	334	5.1	+62.5%	−46.1%	9.6	Reach
Single post	822	12.6	+33.2%	−37.3%	13.9	Reach
Thread	2,387	36.6	−20.3%	+8.9%	76.6	Save
Carousel	308	4.7	−17.2%	−22.5%	56.7	Neither
Image	1,259	19.3	−4.9%	−43.5%	15.9	Neither

The lanes are not substitutes. The reach lane leaves almost no durable artefact, while the save lane is starved of impressions by construction: the thread is the lowest-liking format in the stratum at 36.6% of its supply. Neither lane funds the other automatically, so running both is a scheduling decision rather than an emergent property of posting more. One limitation must be stated plainly: a bookmark is a proxy for intent to return, not a measured return, and the corpus records no follows, clicks or conversions attributable to a post.

9.3 Implications for a small account

The sub-25K stratum is the relevant reference population for an account below 25,000 followers, and it is where the pooled results are least trustworthy (Proposition 4). Table 9.2 ranks the available actions by association magnitude relative to implementation cost, where cost is nil (a deletion or a word choice), low (a drafting or scheduling rule) or high (a change in production capability). The ranking is a decision aid built from observational associations, not from estimated treatment effects.

Table 9.2 — Actions ranked by association magnitude relative to implementation cost, sub-25K stratum. Feature lifts are against posts lacking the feature; format lifts against the stratum median.

#	Action	n	Association	Cost
1	Remove hashtags entirely	529	−35.0% likes, −76.4% bookmarks when present	Nil
2	Keep emoji below four per post	421	−38.0% likes, −34.1% bookmarks when present	Nil
3	Write reach-lane posts as one unbroken line	874	+170.7% likes, −37.2% bookmarks	Low
4	Name a specific tool in the opening line	2,471	+27.9% likes, +12.6% bookmarks	Nil
5	Keep outbound links out of reach-lane bodies	709	−22.5% likes, +50.9% bookmarks when present	Nil
6	Avoid opening the first line with a digit	2,744	−14.0% likes when present	Nil
7	Concentrate posting in the productive hour band	166 / 299	Best UTC hour +77.2%, worst −38.6%	Low
8	Cut thread share; treat threads as save instruments	2,387	−20.3% likes, +8.9% bookmarks, 36.6% of supply	Low
9	Build a demonstration capability	201	+58.2% likes, +290.2% bookmarks	High

Three rows need qualification. Row 5 is lane-specific: outbound links are associated with fewer likes and more bookmarks in the same stratum, so the rule is to keep them out of reach-lane posts, not to abandon them. Row 7 rests on the pooled hour-of-day table within the stratum; a within-tier re-derivation puts the late-night window at +62.5% (n=702) and the morning window at −31.8% (n=862), so the direction survives stratification while the magnitude attenuates. The same module produced a fatal objection elsewhere — a same-day post-index claim that restated the CSV's sort order — so nothing about posting order within a day is carried forward. Row 9 inherits the caveat of Proposition 3 and is ranked last on cost rather than on evidence, being the only action that requires a capability rather than a habit. Most of the available improvement is subtractive and nearly free; the one additive move is expensive.

9.4 Threats to validity

9.4.1 Construct validity

Every categorical variable here is a rule, not a judgement. The audience segment is assigned by a keyword-and-bio classifier with no human-coded validation set, so its error rate is unknown and unlikely to be random: a builder post mentioning ChatGPT is both the likeliest misclassification and the case where the two regimes differ most. Format labels are pattern-derived in the same way. The demonstration disagreement in Proposition 3 measures

construct instability directly, and it should be assumed to exist for labels where no second operationalisation was available to expose it. The `ext_link` episode gives the lesson: a flag must be validated against the raw payload before its coefficient is interpreted, because a plausible name is not evidence of a valid measure.

9.4.2 Internal validity

No variable was manipulated, so nothing here identifies a causal effect, and three confounders are active. Follower count confounds every format and length comparison, which is why the stratification exists; six tiers reduce but do not remove it, since accounts differ within tiers. Topic confounds format and author alike: the Claude theme is +85.4% on likes (n=1,878) against −20.5% for ChatGPT (n=1,384), and the accounts writing each are not interchangeable. Time confounds any level comparison across the window, which is why drift is reported only as a within-half index. The outcomes are further endogenous to one another and to distribution: likes, views and bookmarks are co-outcomes of one impression event (r=0.79 between log views and log likes), and the ranking system reallocates impressions on the basis of early engagement [Twitter, 2023], so any feature that shifts it is amplified by a feedback loop this data cannot separate from its direct effect.

9.4.3 External validity

The corpus is one platform, one topic neighbourhood, one language and one window: 21,864 posts from 7,829 accounts, drawn by 448 queries on AI, Claude, ChatGPT and prompt-engineering terms, in English, between 1 January and 24 August 2026. The query terms define the universe, so the results describe this vocabulary rather than X in general; and a result that decays inside the observation period, as in Proposition 7, should not be expected to hold outside it.

9.4.4 Statistical validity

Two issues dominate. The first is multiplicity: error is controlled by Benjamini–Hochberg FDR within each declared family of 15 features per population, but twelve analyst modules ran over the same corpus and produced 115 findings, and no adjustment spans that larger search. The declared families are the honest unit of control; the wider analysis is exploratory, and the standard caution about analytic flexibility applies [Gelman & Loken, 2014].

The second is more serious. Every query carried a minimum-likes floor, so this is a sample selected on the dependent variable [Heckman, 1979]. Conditioning on high engagement induces dependence among its causes where none need exist in the population, and that dependence is negative: among posts that all succeeded, one lacking a favourable feature must have compensated elsewhere. The bias therefore runs toward attenuation and toward spurious negative association, and no base-rate statement of the form "X per cent of

threads fail" appears anywhere in this report. A residue of 202 posts (0.92%) sits below the lowest floor used; it was left in place, since trimming would add a second undocumented selection step.

Methodological note

Threshold-conditioned sampling is the binding limitation of this study. No larger corpus, finer stratification or multiplicity control repairs it; only observing posts that did not clear the floor does, which is the first item of future work in §9.7.

9.5 Ethical considerations and data provenance

All records are public posts retrieved through the twitterapi.io advanced-search interface under its terms of use. Collection covered content posted publicly to an open timeline; no protected account, direct message, deleted post or logged-in-only surface was accessed, and no credential was used to reach anything an unauthenticated reader cannot see. The stored fields are the post text, its public engagement counts, timestamp and media attachments, and the author's public handle, display name and follower count; no email address, location, demographic attribute, follower list or off-platform identifier was collected, derived or joined. The audience classifier labels a post from its own text and public bio terms; it is not a profile of a person, is not retained as an attribute of an account, and no attempt was made to identify individuals or infer characteristics they did not publish.

Accounts are named only where they are public figures or organisational accounts posting publicly in a professional capacity, and only in aggregate or illustrative context. Private individuals are not named, quoted at length, or made identifiable through the reported statistics; groups below n=15 are suppressed, which also serves as disclosure control. The corpus is not redistributed: what is published is the derived statistics, the analysis code and the post identifiers needed to reconstruct it from the platform, so deletion remains effective — a post removed by its author cannot be recovered from anything published here.

9.6 Reproducibility statement

Every quantitative claim in this report resolves to a value in a single generated facts file, built by one script from one flat post table. That table (`data/viral_posts.csv`) holds 21,864 rows and 49 published columns plus an identifier and carries the SHA-256 prefix `1a15a8c7a1db5bc0` ; the unparsed API payloads and the log of every executed query are retained alongside it. Aggregates were computed under Python 3.13.5 with NumPy 2.4.2, pandas 3.0.1 and SciPy 1.17.1; confidence intervals are percentile bootstraps of 1,000 resamples under seed 20260824, reproducible to the digit. Medians are used throughout for engagement variables; lifts are ratios of medians against a baseline stated in each block; tests are two-sided Mann-Whitney U; groups below n=15 are suppressed and marked rather than re-

ported. Known data quirks — the below-floor residue, the circular hour variable and the `has_link` artefact — are recorded in the facts file rather than silently corrected, so a reader who disagrees with a decision can locate and reverse it.

9.7 Future work

Six studies would strengthen the evidence, in descending order of how far each would change what can be claimed.

1. **A matched non-viral control sample.** The decisive missing arm. Drawing posts from the same accounts, terms and window without a like floor and matching them on account size, format and date [Rosenbaum & Rubin, 1983] would make this a case-control study, restore base rates, and free §9.1 of the selection described in §9.4.4.
2. **A randomised within-account posting experiment.** Assigning drafted posts at random to lane, hook shape or publication window within one account, over a pre-registered run, is the only design available to a practitioner that identifies an effect rather than an association. Proposition 5's subtractive rules are the natural first arms.
3. **A within-creator panel with fixed effects.** Creator fixed effects, over the 7,829 creators in the corpus, would absorb the author confound that stratification only partly controls, and would test whether the demonstration advantage survives comparing an account with itself.
4. **A measurement-validation study.** Human-coding a stratified random sample of several hundred posts for segment, format and hook type, and reporting agreement against the rule-based labels, would supply the error rates §9.4.1 can currently only assume.
5. **Downstream outcome linkage.** Instrumenting one account so that follows, clicks and conversions can be attributed to individual posts would test the premise of the two-lane model — that a bookmark is worth more than a like — which this corpus can only assume.
6. **A replication and decay study.** Re-running the query design in a later window, a second topic neighbourhood and a non-English language would test whether the drift of Proposition 7 continues, whether the demonstration advantage is exhausted by imitation as supply rises, and how far any of this survives outside the vocabulary that defined the sample.

The first is not optional if these results are to be read as more than a description of how successful posts differ from one another; the rest determine whether that description is also an explanation.

10. Conclusion

This study asked what distinguishes the strongest performers from the merely adequate within a corpus of 21,864 high-engagement X posts about AI collected between 2026-01-01 and 2026-08-24. Because every post cleared a like floor, the design supports comparisons among posts that already performed and licenses no statement about base rates of failure. Everything below is an association observed in a winners-only, observational sample.

10.1 The research questions, answered

RQ1. Which post formats are associated with higher engagement, and does the answer depend on account size? Video and short single posts out-perform the all-format baseline, and threads under-perform it by 19.8% pooled (n=6,454); but the answer depends on account size for two formats in particular, so a pooled format ranking should not be quoted without its stratum.

RQ2. Which compositional and linguistic features are associated with higher engagement? Compression is the most consistent feature in the corpus: a post written as one unbroken line reaches a median of 1,661 likes against 832 for the rest (n=2,428, +99.6% raw, +74.4% follower-tier-stratified), and 13 of 15 tested features remain significant after Benjamini-Hochberg FDR control [Benjamini & Hochberg, 1995]. The same compression that lifts likes suppresses saves (-50.7% on bookmarks), so the lever trades one outcome for another rather than raising both.

RQ3. How much of the variation belongs to who posts rather than to what is posted? A great deal: the top 1% of the 7,829 creators account for 28.2% of all likes, and the audience segment a post addresses moves the median more than any format choice does, consumer posts reaching 2.8x the median likes of builder posts on roughly a quarter of the bookmarks.

RQ4. How sensitive are these conclusions to how the instruments are defined? Severely, in at least two documented cases. Isolating the sub-25K stratum moves the demo lift from +0.4% to +58.2% and inverts the carousel lift from +7.7% to -17.2%, which is Simpson's paradox operating on follower count [Simpson, 1951]. Replacing a naive "text contains http" flag with a link re-derived from the API url entities moves the estimated link effect from +9.9% to -32.7% on raw medians — a sign reversal produced entirely by the measurement, not by the posts.

RQ5. What does adversarial verification add to automated analysis? It changed the paper: 139 objections against 115 analyst findings, 12 of them fatal, with 5 of the 12 modules recorded as unsound and the remaining 7 carrying mandatory caveats.

10.2 What should replicate, and what should not

Three findings carry the structural features that usually predict replication — large n, a consistent sign across every stratum, and a mechanism that does not depend on the particular eight months sampled.

1. **The outbound-link penalty.** Posts carrying a genuine off-platform link sit below their peers in all six follower tiers without exception, with tier ratios between 0.43 and 0.74, and the penalty falls on views rather than on the per-view like rate. That the same posts attract more bookmarks argues for a distribution mechanism rather than a taste one, and distribution mechanisms persist.
2. **The brevity effect.** Compression shows a monotone gradient across follower tiers rather than a single unstable cell, survives stratification with most of its magnitude intact, and is corroborated independently by the word-count bands.
3. **The segment split.** Builder and consumer audiences differ in kind, not degree: one converts attention into saves, the other into likes. This is a property of two populations, not a tactic, and should reappear in any corpus that samples both.

Two findings are considerably more fragile.

1. **The theme premium on Claude-related posts** (+56.6% pooled, +85.4% at sub-25K). A theme premium is a description of a news cycle. It measures which vendor was shipping during the window, and there is no reason to expect the ordering to hold in a corpus collected a year later.
2. **The video advantage.** Within the sub-25K stratum, video's index against the all-format median fell from 2.195 in the first half of the window to 1.198 in the second while its share of supply rose by 7.07 percentage points. An advantage that decays as adoption rises is a crowding effect, and a replication attempt is as likely to find it gone as to find it.

> **The findings that survived were the ones whose sign did not depend on which slice of the corpus was examined. That is the only test this design can offer in place of an experiment.**

10.3 Implications for automated research

The verification protocol is the result this study would most want carried forward. Twelve capable analyst agents, given the same clean corpus and an explicit brief, produced findings at an objection rate of 1.21 per finding. The fatal errors were not arithmetic; they were errors of research design that read as competent prose — a sort-order tautology in which a stable sort over a likes-descending file made "first post of the day" mean "best post of the day"; a psychological trigger whose effect was one spam template repeated; a ratio sharing its denominator with the variable that defined its groups. None of these is visible by inspecting the output. Each was caught only because an independent agent re-derived the number from the raw table.

Fluency and validity are separate properties, and automated analysis raises throughput on the first without raising it on the second [Ioannidis, 2005]. Analyst agents make the same specification choices human analysts make, faster and in greater volume, enlarging rather than shrinking the garden of forking paths [Gelman & Loken, 2014]. Adversarial re-derivation should therefore be a required stage of an automated pipeline rather than an optional review, and a verification budget comparable to the analysis budget is not extravagant.

Methodological note

A closing caution on scope. The corpus is winners-only, single-platform, eight months long and restricted to one topic area. Every effect reported here is an association within that frame; none of it establishes that changing a post would change its outcome. These results generate hypotheses worth testing experimentally; they do not substitute for the experiment.

References

[1] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300.

[2] Berger, J. (2011). Arousal increases social transmission of information. Psychological Science, 22(7), 891–893.

[3] Berger, J., & Milkman, K. L. (2012). What makes online content viral? Journal of Marketing Research, 49(2), 192–205.

[4] Cialdini, R. B. (2007). Influence: The psychology of persuasion (Rev. ed.). Harper Business.

[5] Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. SIAM Review, 51(4), 661–703.

[6] Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin, 114(3), 494–509.

[7] CourtBouillon. (2026). WeasyPrint [Computer software]. https://weasyprint.org

[8] Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281–302.

[9] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1–26.

[10] Galton, F. (1886). Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246–263.

[11] Gelman, A., & Loken, E. (2014). The statistical crisis in science. American Scientist, 102(6), 460–465.

[12] Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. Management Science, 62(1), 180–196.

[13] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362.

[14] Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: The case of urban legends. Journal of Personality and Social Psychology, 81(6), 1028–1041.

[15] Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica, 47(1), 153–161.

[16] Hubert, L., & Arabie, P. (1985). Comparing partitions. Journal of Classification, 2(1), 193–218.

[17] Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124.

[18] Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50–60.

[19] McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), Proceedings of the 9th Python in Science Conference (pp. 56–61).

[20] The pandas development team. (2026). pandas (Version 3.0.1) [Computer software]. https://pandas.pydata.org

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[22] Python Software Foundation. (2025). Python (Version 3.13.5) [Computer software]. https://www.python.org

[23] Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 76(3), 036106.

[24] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55.

[25] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65.

[26] Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. Science, 311(5762), 854–856.

[27] Simpson, E. H. (1951). The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society: Series B (Methodological), 13(2), 238–241.

[28] Twitter. (2023). the-algorithm: Source code for Twitter's recommendation algorithm [Computer software]. GitHub. https://github.com/twitter/the-algorithm

[29] twitterapi.io. (2026). Advanced search endpoint [Web API]. Retrieved August 24, 2026, from https://twitterapi.io

[30] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272.

Appendix A — The Query Grid

The corpus was assembled by exhaustively enumerating a three-dimensional grid — topic term × `min_faves` floor × date slice — and issuing each cell as one call to the twitterapi.io advanced-search endpoint. This appendix records the grid as designed, the portion of it that was executed before the collection target was reached, and the portion that actually contributed posts. The three counts differ, and the difference is a coverage limitation of the study rather than an implementation detail.

A.1 Topic themes and terms

Seven themes were defined, spanning 42 search terms. Terms were written to be lexically disjoint where possible, but the API matches on substrings of post text, so a single post frequently satisfies several terms; each post is attributed in the `theme` and `term` columns to the query that first returned it, and those columns therefore describe capture provenance rather than topical prevalence.

Table A.1 — The seven designed themes, their search terms, the number of grid cells each generated, and the number executed. Themes with zero executed queries were never reached: the harvester stops when the collection target is met, and it walks the grid in theme order.

Theme	Search terms	Terms	Built	Run	Posts
`claude`	claude, claude code, anthropic claude, claude ai, claude opus, claude sonnet	6	162	162	6,490
`chatgpt`	chatgpt, gpt-5, openai chatgpt, chatgpt prompt, custom gpt	5	135	135	4,531
`prompting`	prompt engineering, prompting, system prompt, prompt hack, this prompt, best prompts,…	9	243	243	5,091
`ai_general`	ai tool, ai agent, ai automation, ai workflow, generative ai, llm, ai startup, vibe c…	9	243	149	5,752
`grok`	grok, @grok, grok explain	3	81	0	0
`models`	gemini prompt, midjourney prompt, sora prompt, nano banana prompt, veo prompt	5	135	0	0
`creator`	ai thread, ai tips, chatgpt tips, claude tips, ai prompts thread	5	135	0	0
Total		**42**	**1,134**	**689**	**21,864**

Methodological note

Coverage limitation. Three of the seven designed themes — `grok`, `models` and `creator`, 351 grid cells in total — were never executed. The harvester runs the grid in dictionary order and halts once the unique-post target is reached, which happened part-way through `ai_general`. The corpus therefore under-represents image- and video-model prompting (Midjourney, Sora, Veo, Nano Banana) and explicit "AI tips thread" creator content, and contains @grok interactions only where they were captured incidentally by another term. Conclusions in this report are scoped to the four themes actually harvested.

A.2 Terms that contributed posts

Of the 26 terms in executed cells, 25 returned at least one post that survived deduplication and the date floor. The term `llm` returned none: every post its queries surfaced had already been captured by an earlier term. Table A.2 gives first-capture counts.

Table A.2 — Posts by first-capture term, descending. The counts partition the corpus exactly; they do not measure how often a term appears in post text.

Theme	Term	Posts	Share %
ai_general	ai tool	1,442	6.60
claude	claude opus	1,328	6.07
claude	claude	1,304	5.96
ai_general	ai agent	1,301	5.95
claude	anthropic claude	1,289	5.90
ai_general	ai automation	1,200	5.49
chatgpt	chatgpt	1,170	5.35
chatgpt	openai chatgpt	1,127	5.15
ai_general	ai workflow	1,125	5.15
prompting	this prompt	1,125	5.15
chatgpt	chatgpt prompt	1,095	5.01
prompting	prompting	1,056	4.83
claude	claude ai	970	4.44
prompting	system prompt	944	4.32
claude	claude code	923	4.22
chatgpt	gpt-5	915	4.18
prompting	prompt engineering	788	3.60
ai_general	generative ai	684	3.13
claude	claude sonnet	676	3.09
prompting	best prompts	644	2.95
prompting	meta prompt	265	1.21
chatgpt	custom gpt	224	1.02
prompting	prompt template	191	0.87
prompting	prompt hack	75	0.34
prompting	megaprompt	3	0.01

A.3 Engagement floors and date slices

Every query carried a `min_faves` floor. The floor is what makes the corpus winners-only, and it is also what makes it tractable: without a floor the population of posts mentioning “prompt” over eight months is far larger than any depth-limited search interface will return. Six floors were used. Low floors return more matches than the API will paginate through in one query, so those cells were split into monthly date slices; the three highest floors are rare enough to be searched across the whole window at once.

Table A.3 — Floor structure. Queries per term is the product of the slice count and one query per floor; the grid total multiplies by 42 terms.

min_faves	Date handling	Slices	Q / term	Built	Run
200	monthly slices	8	8	336	208
500	monthly slices	8	8	336	206
1,000	monthly slices	8	8	336	200
2,500	single window, 2026-01-01 onward	1	1	42	25
5,000	single window, 2026-01-01 onward	1	1	42	25
10,000	single window, 2026-01-01 onward	1	1	42	25
All			**27**	**1,134**	**689**

The monthly slices are the eight intervals `2026-01-01` to `2026-02-01`, `2026-02-01` to `2026-03-01`, and so on through `2026-08-01` to `2026-08-25`. The final slice is 24 days rather than a full month, so any month-over-month supply comparison in this report treats August as partial.

A.4 The query template

Two templates were used, differing only in the date clause. For floors below 2,500:

```
{term} min_faves:{floor} lang:en -filter:replies since:{since} until:{until}
```

and for floors of 2,500 and above:

```
{term} min_faves:{floor} lang:en -filter:replies since:2026-01-01
```

Every query was issued with `queryType=Top`. The `lang:en` clause restricts the corpus to English and is a real scope limit, not a nuisance filter: no claim here extends to non-English AI discourse. The `-filter:replies` clause is the reason the `is_reply` column is identically zero and no reply-level analysis is possible.

Table A.4 — Harvester parameters as executed, from scrape.py.

Parameter	Value	Effect
`Endpoint`	twitter/tweet/advanced_search	twitterapi.io advanced search [twitterapi.io, 2026]
`queryType`	Top	platform-ranked ordering within each query
`MAX_PAGES`	12	ceiling of 240 posts per query cell (12 pages × 20)
`WORKERS`	8	concurrent query threads
`TARGET`	40,000	unique-post stop condition; not reached
`CUTOFF`	2026-01-01	posts older than the window are discarded on ingest
`Retries`	4 attempts, backoff 1.5s × n	transient failures retried, then logged

A.5 From grid cells to corpus

Of the 1,134 cells designed, 689 were executed and 448 returned at least one post that was new to the corpus at the moment it arrived. The gap between 689 and 448 is de-duplication: the grid deliberately overlaps, since a post with 3,000 likes satisfies the 200, 500, 1,000 and 2,500 floors simultaneously and will be returned by all four. Posts are keyed by tweet id and kept once, on first sight, which is why 241 executed cells added nothing. The 448 productive cells yielded 21,864 unique posts from 7,829 creators across 236 days.

> **Every post in the corpus cleared a like floor before it was ever seen. The design can compare winners to other winners; it cannot observe a failure.**

Appendix B — Data Dictionary

`data/viral_posts.csv` holds one row per post and 50 columns: an identifier plus 49 published fields. The tables below give every column with its type, its derivation, and a reliability note. The single most important distinction is between columns copied verbatim from the API payload — platform ground truth, subject only to the platform's own counting conventions — and columns produced by a regular expression or heuristic written for this study, which are inferences and carry measurement error that the analysis must absorb. Chapter 8 documents one case where that error inverted a headline result.

Methodological note

Row order is not information. The file is written sorted by `likes` descending. Any operation that depends on file order within a group — "first row per creator per day", a stable sort on a coarse key, a head-of-group pick — recovers the highest-liked post in that group and nothing else. This produced one of the twelve fatal objections in Chapter 8.

B.1 Identity, account and time

Table B.1 — Identity, account and temporal columns.

Column	Type	Derivation	Reliability
`id`	string	API `id`. Primary key; de-duplication is on this field.	Ground truth.
`url`	string	API `url`. Canonical permalink.	Ground truth.
`handle`	string	API `author.userName` at capture time.	Ground truth at capture; handles can change.
`name`	string	API `author.name` at capture time.	Ground truth at capture; display names change freely.
`followers`	int	API `author.followers` at capture time.	Ground truth at capture, not at posting time — a known confound for any lift computed against follower count.
`follower_tier`	string	Six-way bucketing of `followers`: 0–1K, 1K–5K, 5K–25K, 25K–100K, 100K–1M, 1M+.	Deterministic, but inherits the capture-time confound above.
`verified`	0/1	1 if `isBlueVerified` or `isVerified`.	Ground truth, but conflates paid Blue with legacy verification; 96.9% of the builder segment is flagged, so it has almost no discriminating power here.
`date`	date	`createdAt` parsed, formatted `YYYY-MM-DD`, UTC.	Ground truth.
`hour_utc`	int 0–23	Hour component of `createdAt`, UTC.	Ground truth. No local-time conversion is applied; timing results are UTC-anchored and confounded with author geography.
`weekday`	string	Three-letter weekday of `createdAt`, UTC.	Ground truth, same UTC caveat.

B.2 Engagement counters and derived ratios

The six counters are read straight from the payload. The seven ratios below them are deterministic arithmetic on those counters, computed with a zero-safe divide that returns 0.0 rather than a missing value when the denominator is zero. That convention matters: a post with zero recorded views yields `engagement_rate` = 0, which is a floor artefact and not a measurement, and such rows are excluded from rate analyses.

Table B.2 — Engagement counters and the ratios derived from them.

Column	Type	Derivation	Reliability
`likes`	int	API `likeCount`.	Ground truth. The selection variable: every query imposed a floor on it.
`retweets`	int	API `retweetCount`.	Ground truth.
`replies`	int	API `replyCount`.	Ground truth. Counts replies of every sentiment; a high value is not approval.
`quotes`	int	API `quoteCount`.	Ground truth.
`bookmarks`	int	API `bookmarkCount`.	Ground truth. The clearest available proxy for intent to return.
`views`	int	API `viewCount`.	Ground truth, but platform-defined; an impression is not a read.
`total_engagement`	int	likes + retweets + replies + quotes + bookmarks.	Arithmetic. Sums units that are not commensurable; reported for completeness, not used as a headline.
`engagement_rate`	float	(likes + retweets + replies + bookmarks) / views.	Arithmetic. Shares its numerator with `likes`; never regress it on likes.
`like_per_1k_views`	float	1000 × likes / views.	Arithmetic. Zero when views is zero.
`bookmark_ratio`	float	bookmarks / likes.	Arithmetic. Denominator is the selection variable, so this ratio is mechanically depressed for the highest-liked posts.
`reply_ratio`	float	replies / likes.	Arithmetic, same denominator caution.
`amplification`	float	(retweets + quotes) / likes.	Arithmetic, same denominator caution. A circular use of one of these ratios — grouping on a variable and then reporting a ratio sharing its denominator — is one of the fatal objections in Chapter 8.
`likes_per_follower`	float	likes / followers.	Arithmetic on a capture-time denominator. Strongly negatively related to account size by construction.

B.3 Rule-based classification

These eight columns are the study's own inferences. None is a platform field. Each is produced by a documented regular expression or scoring rule in `analyze.py`, and each therefore has an unmeasured false-positive and false-negative rate. Where a result in this report rests on one of them, the rule is restated in the text so the reader can judge what it actually matches.

Table B.3 — Rule-based classification columns. Every field in this table is an inference, not an observation.

Column	Type	Derivation	Reliability
`format`	string	Priority cascade over the media kinds and text markers: @grok post → poll → thread → demo → video → carousel → image → quote post → single post.	Inference. First match wins, so the categories are not independent: a thread with a video is filed as a thread, and every format lift is conditional on that ordering.
`media_type`	string	From the API media entities: video, carousel (>1 photo), image (1 photo), poll, else text.	Near ground truth — derived from platform entities rather than text — except that `poll` is detected by string-searching the card payload.
`audience`	string	builder / mixed / consumer from `audience_score` thresholds (≥3, ≥1, else consumer).	Inference from a keyword-and-bio classifier. Not validated against human labels. The three segments differ so sharply in engagement that mislabelling is a material risk; every actionable result names its segment.
`audience_score`	int	+2 builder text terms, +2 how-to terms, +1 builder bio terms, −2 consumer text terms, −1 short quoted meme.	Inference. The weights are hand-set, not fitted.
`hook`	string	First non-empty line of the post, truncated to 220 characters.	Deterministic extraction, but “first line” is a rendering convention, not a semantic unit.
`hook_types`	string	Pipe-joined names of all matching patterns from 25 regexes applied to `hook`; if none match, the regexes are re-applied to the first 300 characters of the post; if still none, `Plain statement`.	Inference, and multi-label: one post can carry several types, so hook-type groups overlap and their medians are not a partition. The fall-back to the post body means a labelled hook type is not always present in the opening line.
`theme`	string	The `_theme` of the query that first returned the post.	Provenance, not content. A post captured under `claude` may be mostly about ChatGPT.
`term`	string	The `_term` of the query that first returned the post.	Provenance, not content; see Table A.2.

B.4 Text metrics, flags and body

Table B.4 — Surface text metrics, binary flags and the post body.

Column	Type	Derivation	Reliability
`chars`	int	`len(text)` including the appended t.co of any attached media.	Deterministic, but media posts are inflated by roughly 23 characters of machine-added URL.
`words`	int	Whitespace-split token count.	Deterministic; counts URLs and mentions as words.
`lines`	int	Newline count + 1.	Deterministic. Used as the whitespace proxy throughout.
`emojis`	int	Count of matches of a Unicode emoji range regex.	Inference. The range omits some newer code points and some ZWJ sequences count more than once.
`hashtags`	int	`text.count("#")` .	Crude: counts every literal hash character, including those inside words and code.
`mentions`	int	`text.count("@")` .	Crude, as above: e-mail addresses and decorative @ inflate it.
`has_link`	0/1	1 if the substring `http` appears anywhere in the text.	**Unreliable and retained only for continuity.** X appends a t.co URL to the body of every natively attached photo or video, so this flag fires on 48.2% of media posts against 11.4% of text posts and measures media presence, not outbound linking. It reports a +9.9% like lift where the corrected measure reports a loss.
`ext_link`	0/1	1 if the API `entities.urls` contain an expanded URL whose host is not x.com, twitter.com or t.co.	The corrected outbound-link measure and the one to use. Prevalence 12.0% (n=2,626) against 36.9% for `has_link` ; raw lift −32.7%, follower-tier-stratified −33.8%. Still an inference in one respect: a shortener that resolves to an on-platform target would be counted as external.
`is_reply`	0/1	API `isReply` .	Ground truth but **degenerate**: identically 0 across all 21,864 rows, because every query carried `-filter:replies` . No reply analysis is possible.
`question`	0/1	1 if the text ends with a question mark.	Deterministic but narrow: a question in the opening line followed by a body is not flagged.
`is_thread`	0/1	1 if a thread marker regex fires (🧵, “thread”, `1/` , `1/n`) or the text exceeds 900 characters.	Inference, and the length clause makes it partly a length variable: any long post is a “thread” whether or not it has a second post.
`is_quote`	0/1	1 if the payload carries a `quoted_tweet` .	Ground truth.
`is_grok`	0/1	1 if `@grok` appears in the text, or the author or reply target is Grok.	Inference, but a narrow and reliable one.
`is_demo`	0/1–4	**A count, not a flag.** Number of demo-signal regex hits, conditioned on media being present. Binarised as > 0 wherever this report treats it as a flag.	Inference. Reading it as a boolean without binning arising silently weights posts by signal count.

Column	Type	Derivation	Reliability
`is_engagement_bait`	0/1	1 if a giveaway / reply-gate regex fires (“comment X and I’ll…”, “must follow”, “repost to…”).	Inference with known leakage: the pattern matches ordinary sentences containing “reply” or “drop”. Appendix E reports it beside four narrower hand-built bait patterns.
`is_poll`	0/1	1 if the string `poll` appears in the serialised card payload.	Inference, and **suppressed everywhere**: it fires on 7 posts, below the n < 15 floor.
`photos`	int	Count of media entities of type `photo`.	Ground truth from platform entities.
`gifs`	int	Count of media entities of type `animated_gif`.	Ground truth from platform entities.
`text`	string	Full post body with newlines replaced by the sentinel `⏎` so the field survives CSV round-tripping.	Ground truth, transformed. Any downstream newline count must use `lines`, not this field.

B.5 Derived flags built downstream

Fifteen further binary features are constructed at analysis time rather than stored in the CSV. They are listed here because report figures refer to them by name.

Table B.5 — Analysis-time feature flags, from the facts pack.

Flag	Definition
`ext_link`	Genuine off-platform outbound link (API url entities, t.co self-links removed)
`hashtag`	At least one hashtag
`emoji_4plus`	Four or more emoji in the post body
`thread`	Post is the opener of a multi-post thread
`demo`	Screen-recorded product demonstration
`quote`	Quote-post of another account
`question`	Post text poses a direct question
`engagement_bait`	Explicit comment-to-get / reply-gate bait
`numeral_in_hook`	A digit appears in the opening line
`tool_named`	A named AI tool appears in the opening line
`allcaps_word`	An ALL-CAPS word of three or more letters in the opening line
`colon_in_hook`	A colon in the opening line
`single_line`	Whole post is one unbroken line (no line break)
`hook_le_10_words`	Opening line of ten words or fewer
`media_present`	Any native media attached (image, carousel, video, poll)

B.6 Known quirks

The six defects below are properties of the delivered file and are restated wherever they bear on a result.

1. is_demo is a COUNT of demo signals (0-4), not a flag; it is binarised as >0 here.
2. is_reply is identically 0 across the corpus — the search never returned reply posts, so no reply analysis is possible.
3. is_poll fires on 7 posts and every poll cut is suppressed under the n<15 floor.
4. 202 posts (0.92%) sit below 100 likes and three sit at zero, despite every query carrying a min_faves floor — see corpus.floor_integrity.
5. The CSV is stored sorted by likes descending; any within-group operation that relies on file order is a tautology (see verification.named_failure_modes).
6. analysis.json's by_* blocks describe the builder segment only; the pooled tables in this file are recomputed over builder+mixed and are not directly comparable to them.

On the fourth of those: 0.924% of rows sit below 100 likes and the observed minimum is 0, despite every query carrying a floor of at least 200. The likeliest explanation is that like counts are read at capture and can fall — through deletion, account deactivation or platform recount — after the post cleared the floor at query time. The affected rows are under 1% and are retained rather than trimmed, since trimming on the outcome variable would itself be a selection step.

Appendix C — The Module Catalogue

The analysis was decomposed into twelve thematic modules, each written and executed by a separate analyst agent against the same corpus, and each subsequently attacked by an independent verifier agent that re-derived the headline numbers from `data/viral_posts.csv`. Chapter 8 describes the protocol and its yield. This appendix is the reference catalogue: what each module computes, where its output lives, and how it closed.

Across the twelve, the analysts produced 115 findings, 95 recommended plays and 111 self-declared caveats. The verifiers logged 163 independent re-derivations and filed 139 objections — 1.209 per finding — of which 12 were graded fatal, 57 major and 70 minor. 7 modules closed at needs_caveat and 5 at unsound. No module closed sound.

C.1 What each module computes

Table C.1 — The twelve analyst modules and their scope. Each writes one JSON file under data/modules/.

Module	What it computes
`content_pillars`	TF-IDF over cleaned post text (1–2 grams, min_df 25) followed by a 16-component NMF, yieldů ing interpretable content pillars with per-pillar engagement profiles and supply share.
`whitespace_matrix`	Supply against performance for every topic-cluster × format × follower-tier cell, to identify under-supplied high-performing cells and crowded ones.
`hook_linguistics`	Opening-line linguistics past the 25 stored hook regexes: anchored and sliding n-grams of the first five words, first-word part-of-speech buckets, hook length and punctuation, tool-versus-generic naming, ranked by median-likes lift against the builder+mixed baseline.
`psych_triggers`	Twelve documented persuasion triggers detected by regex, measured individually and pairwise against the cohort baseline on likes, saves and replies.
`engagement_funnel`	Engagement as staged conversion — views to likes to bookmarks, replies, quotes and retweets — separating reach from resonance and intent.
`creator_archetypes`	k-means clustering of the 1,484 creators with three or more corpus posts on format, media, hook-family, theme and cadence fingerprints, into named archetypes with benchmark tables.
`growth_trajectories`	The longitudinal cut: for creators appearing repeatedly across the window, early-versus-late split of winning-post performance and the behavioural changes that accompany movement.
`timing_cadence`	Posting hour × format interaction, per-creator cadence in posts per active week, consecutive-day streaks, and within-day position effects.
`amplification_network`	An interaction graph from @mentions and quote relationships, reduced to a ranked target list of accounts a small builder account could productively engage.
`brand_share_of_voice`	Which AI products are named inside winning posts, their share-of-post trend over the eight months, and a co-mention graph with lift over chance.
`repurposing_engine`	Cross-channel leverage scoring: which winning posts carry enough structure to be rebuilt as long-form video, article or product asset.
`antipatterns`	A ranked risk register of 41 patterns, each measured inside matched follower_tier × format strata, with bootstrapped confidence intervals and an author-judged brand-risk score. Source for Appendix E.

Methodological note

A thirteenth file. `data/modules/strategy_synthesis.json` sits alongside the twelve but is not an analyst module and was not separately verified. It is the strategist cross-cut: it re-runs the contested quantities inside the sub-25K builder+mixed stratum and adjudicates the disagreements between modules, without generating findings of its own. Its adjudications are reported in Chapter 9 as adjudications, not as measurements.

C.2 Verdicts and objection counts

Table C.2 gives the full disposition. It complements Table 8.1, which breaks the same objections down by severity; the columns here are the ones that describe analyst output volume rather than verifier output. The objections-per-finding column is the one worth reading first: it ranges from 0.90 (`hook_linguistics`) to 1.56 (`amplification_network`) and carries no relationship to the verdict, which is what one expects if objection volume measures surface

area rather than soundness. `whitespace_matrix` and `content_pillars` each drew 15 objections against 10 findings and closed on opposite sides; `amplification_network` drew the highest ratio of any module and still closed at needs_caveat, because none of its 14 objections was graded fatal. `hook_linguistics` drew the lowest and closed with a fatal objection outstanding. A verdict turns on whether any single objection was fatal, not on how many were filed.

Table C.2 — Module disposition: analyst output, verifier objections, and closing verdict. Ordered by verdict then objection count.

Module	Find.	Plays	Caveats	Obj.	Obj./find.	Verdict
`whitespace_matrix`	10	8	9	15	1.50	unsound
`antipatterns`	10	8	8	12	1.20	unsound
`growth_trajectories`	10	8	10	11	1.10	unsound
`repurposing_engine`	10	8	10	11	1.10	unsound
`timing_cadence`	10	8	10	10	1.00	unsound
`content_pillars`	10	8	10	15	1.50	needs caveat
`engagement_funnel`	10	8	10	14	1.40	needs caveat
`amplification_network`	9	8	8	14	1.56	needs caveat
`brand_share_of_voice`	10	8	11	10	1.00	needs caveat
`psych_triggers`	8	8	8	9	1.12	needs caveat
`hook_linguistics`	10	8	10	9	0.90	needs caveat
`creator_archetypes`	8	7	7	9	1.12	needs caveat
Total	**115**	**95**	**111**	**139**	**1.21**	

C.3 Named failure modes

Five defect classes recurred often enough across modules to be given names during verification. They are described in full in Chapter 8; they are listed here so the catalogue is self-contained.

Table C.3 — The five named failure modes and where each was first caught.

Failure mode	First caught in
Sort-order tautology	`timing_cadence :: same_day_post_index`
Copy-paste spam artifact	`psych_triggers`
Circular metric	`engagement_funnel / growth_trajectories`
Pooled-tier confound	`all format and length comparisons`
Measurement artifact (has_link)	`engagement_funnel :: link cost`

Every module output remains on disk in its objected-to state. Verifiers were permitted to patch and re-run a module only additively, so a repair could not overwrite the record of what was originally claimed. Where a claim in this report was disputed, the objection travels with it.

Appendix D — The Hook Corpus

The hundred opening lines below are the highest-value hooks in the builder segment (n=3,370), reproduced verbatim from the posts that carried them. They are included because a distribution of hook-type medians is not a usable artefact on its own: a practitioner needs to see what the winning surface actually looked like, and a reader needs to be able to check whether a labelled pattern is really present in the line.

D.1 How the hundred were selected

Selection was by a fixed rule, not by editorial judgement. Posts in the builder segment were ranked on a composite value score, `likes + 8 × bookmarks`, which weights save intent an order of magnitude above passive approval. The eightfold weight is a stipulated preference, not an estimated coefficient, and it is the one arbitrary number in the procedure; a bank ranked on likes alone would be a materially different list. Candidates were then filtered:

- the opening line, with URLs stripped, must be between 18 and 200 characters;
- the post body must mention a concrete AI tool, model or prompting concept within its first 400 characters, so the bank does not fill with high-reach material that has nothing to transfer;
- hooks classified as `Plain statement` are dropped unless their bookmark ratio exceeds 0.08, which removes bare product announcements from the labs;
- accounts above two million followers are dropped when their likes-per-follower falls below 0.002, which removes posts carried by distribution rather than by the line;
- near-duplicates are collapsed on a normalised 38-character signature of the opening.

Three properties of the resulting bank limit its use. It is winners-only twice over — the corpus floor, then the value ranking — so it shows what a strong opening looked like, never how often such an opening fails. Its `Pattern` column is the multi-label output of the regex classifier described in Table B.3, so a line can carry several patterns and the first listed is not privileged. And 53 of the hundred — a clear majority — are labelled `Plain statement`, meaning that none of the 25 patterns fired on them. Whatever made those lines work is not captured by the taxonomy, which is the most useful single fact in this appendix: the pattern library describes a minority of the strongest openings in the corpus.

D.2 The bank

Table D.1 — The 100 highest-value opening lines in the builder segment, ranked by likes + 8 × bookmarks. Openings longer than 66 characters are truncated; pattern labels are the multi-label regex classifier output, truncated to the first two labels.

#	Opening line	Account	Likes	Saves	Pattern
1	I'm Boris and I created Claude Code. Lots of people have asked ho...	@bcherny	54,345	103,478	Plain statement
2	INSTEAD OF WATCHING NETFLIX TOů NIGHT.	@cyrilXBT	39,309	100,967	Us vs them / gatekeepů ing
3	🚨 Anthropic's own team just showed how to actually use Claude Cod...	@eng_khairallah1	29,665	95,529	Personal result / income
4	I genuinely don't understand why everyone isn't using this yet	@ridark_eth	30,934	85,167	Plain statement
5	instead of watching 2 hours of Netflix tonight, watch this Stanfo...	@AnatoliKopadze	27,647	79,132	Plain statement
6	the engineer who built Claude Code just dropped a 28-minute video...	@AnatoliKopadze	24,180	60,004	Personal result / income
7	Anthropic dropped a 33 pages cheat sheet for building Claude skil...	@RoundtaůbleSpace	18,652	58,771	Plain statement
8	Claude Design is insane.	@viktoroddy	25,493	51,310	Plain statement
9	This 2 hour Stanford lecture shows exactly how Stanford trains it...	@RohOnChain	13,508	51,049	Plain statement
10	🚨do you understand what two Anthropic engineers just explained in...	@kirillk_web3	13,860	49,122	Time-bound flex
11	🚨 Anthropic just showed a 24-minute workshop on how to actually d...	@anujcodes_21	12,882	41,580	Personal result / income
12	A few random notes from claude coding quite a bit last few weeks.	@karpathy	40,759	37,045	Plain statement
13	Anthropic engineer:	@AnatoliKopadze	15,523	39,466	Time-bound flex
14	Andrej Karpathy spent 4 minutes in an interview explaining a sing...	@rewind02	12,999	36,522	Personal result / inů come / Us vs them / gat...
15	By far my biggest advice to anyone trying to adopt AI properly:	@mstockton	16,882	35,355	Plain statement
16	Claude Mythos is Insane!	@viktoroddy	16,500	30,212	Plain statement
17	I still don't understand why everyone is not using this yet. Than...	@Bober_smart	11,401	30,099	Plain statement
18	Anthropic pays $750,000+ a year for enů gineers who know how to bui...	@sairahul1	7,686	26,185	Personal result / income
19	Godfather of AI: "If you sleep well tonight, you may not have und...	@AnatoliKopadze	9,042	25,051	Superlative claim
20	Anthropic engineers just showed how they build a full app from sc...	@AnatoliKopadze	8,949	24,501	Plain statement
21	Lost my phone at the office and spent 30 minutes turning the plac...	@un1c0rnioz	48,472	19,455	Plain statement
22	How to setup your Claude code project?	@akshay_pachaar	12,054	23,510	

#	Opening line	Account	Likes	Saves	Pattern
					Question hook / How to / tutorial
23	anthropic fucking killed it with this. so many people will start...	@cryptopunk7213	22,331	20,047	Superlative claim
24	🚨BREAKING: You can now run Claude Code for FREE.	@irisneural	12,549	20,791	Plain statement
25	Most people are sleeping on this.	@cgtwts	10,133	20,743	Us vs them / gatekeeping
26	If you want to get the most out of ChatGPT for writing, reviewing...	@AskMichaelTaiwo	5,958	21,179	Thread announcement
27	OpenAI and Anthropic engineers leaked a prompting technique that...	@aiwithmayank	6,489	21,098	Plain statement
28	Here's how to create your classic corporate headshot:	@shushant_l	10,278	20,166	Plain statement
29	Claude + Polymarket = Free Money? (Full Bot Tutorial)	@MoonDevOnYT	11,615	19,775	Plain statement
30	We just released Claude Code channels, which allows you to contro...	@trq212	25,433	17,615	Plain statement
31	Obsidian → create a new vault (folder)	@Bober_smart	6,156	19,790	Plain statement
32	🚨 Prompt engineering is officially outdated.	@aiwithmayank	7,818	19,539	Plain statement
33	OpenAI CEO, Sam Altman:	@Raytar	6,195	19,339	Time-bound flex / Us vs them / gatekeeping
34	Prompt engineering is dead.	@heyrimsha	7,835	18,976	Prompt drop
35	i wish I had these walkthroughs when I was getting into AI	@Nate_Google_	8,679	18,792	Superlative claim
36	ANTHROPIC JUST PROVED MOST PEOPLE HAVE NO IDEA HOW TO PROMPT CLAU...	@cyrilXBT	7,146	18,882	Us vs them / gatekeeping
37	This is one of the better versions of this prompt template I've s...	@AIWarper	59,885	12,071	Plain statement
38	🚨 Someone reverse-engineered the design systems of Apple, Spotify...	@heynavtoor	7,989	18,318	Plain statement
39	Go to ChatGPT and use this prompt: Create a caricature of me and...	@MrMekzy_	24,950	16,090	Plain statement
40	Anthropic just showed a 24-minute workshop on how to actually pro...	@TheAIWorld22	6,092	18,021	Personal result / income
41	Google's former CEO just said what everyone in AI already knows	@0xwhrrari	5,207	17,794	Plain statement
42	This is how to make your first $1,000 to $10,000 from the 2026 Wo...	@iksly2	10,189	17,049	Personal result / income
43		@gregisenberg	7,829	17,028	How to / tutorial

#	Opening line	Account	Likes	Saves	Pattern
	how to use obsidian + claude code to build a 24/7 personal operat...				
44	Don't waste 2 years learning to build LLMs like Claude & ChatGPT.	@RohOnChain	7,547	16,438	Warning / fear
45	this is f*cking gold	@eng_khairallah1	7,491	16,249	Plain statement
46	I reverse-engineered Claude Code's leaked source against billions...	@iamfakeguru	9,072	16,029	Plain statement
47	How to climb all 4 layers of Claude in one weekend:	@rubenhassid	8,578	15,934	How to / tutorial
48	A single **CLAUDE.md** file just hit 15K GitHub stars.	@akshay_pachaar	7,582	16,008	Plain statement
49	If you want to learn how to use Claude Code + YouTube to build an...	@Shawnife	3,400	16,434	Plain statement
50	The quality of your vibecoded slop is horrible. I've seen it. Abs...	@shawmakesmaů gic	5,858	16,024	Plain statement
51	PAID VERSION → FREE ALTERNATIVE	@sufyanmaan	7,891	15,265	Plain statement
52	You no longer need a Photographer to get your passport just use t...	@abs_uiux	5,791	15,387	Direct address 'you'
53	BREAKING: Don't copy and paste answers from ChatGPT.	@allen_lattimer	5,159	15,304	Breaking / just launched
54	Anthropic's applied AI team just showed how to actually prompt Cl...	@Av1dlive	5,247	15,233	Plain statement
55	🚨 Anthropic's CEO: "software engineering will be fully automated...	@eng_khairallah1	4,495	14,971	Time-bound flex
56	Introducing text-to-lottie: an open source skill and harness for...	@konstipaulus	8,789	14,407	Plain statement
57	🚨 Holy shit...A developer on GitHub just built a full development...	@ihteshamali	6,184	14,570	Us vs them / gatekeepů ing
58	This 16-minute talk by two Anthropic engineers who built Claude S...	@Av1dlive	5,265	14,622	Plain statement
59	Software development is undergoing a renaissance in front of our...	@gdb	12,148	13,618	Plain statement
60	I turned Andrej Karpathy's viral AI coding rant into a system pro...	@alex_prompter	6,248	14,334	Plain statement
61	Google Brain founder, Andrew Ng:	@AnatoliKopadze	4,470	14,437	Personal result / inů come / Time-bound flex
62	Head of Claude Code:	@AnatoliKopadze	4,415	14,354	Time-bound flex
63	We built our launch video in Claude Code using HyperFrames.	@HeyGen	8,357	13,681	I built / I made
64	Google launched a brand new AI tool.	@prasenx	11,182	13,315	Plain statement

#	Opening line	Account	Likes	Saves	Pattern
65	I absolutely hate AI writing. And my entire job is to write about...	@rubenhassid	6,265	13,910	Plain statement
66	Andrew Ng just dropped a 3-hour course on how to become an AI Eng...	@0xMovez	6,417	13,889	Time-bound flex
67	My friend applied to 250 tech jobs in two years. No MIT. No Stanf...	@RohOnChain	4,606	13,823	Personal result / income
68	HOW TO VIBE CODE BEAUTIFUL UI	@om_patel5	7,990	13,369	How to / tutorial
69	Anthropic is offering 13 AI courses & certiů ficates.	@rubenhassid	6,541	13,485	Superlative claim
70	must read. doesn't matter if you're just entering the AI space, d...	@NoahEpstein_	5,011	13,538	Plain statement
71	Farzapedia, personal wikipedia of Farza, good example following m...	@karpathy	9,116	12,894	Plain statement
72	People think learning Claude takes days. It doesn't.	@rubenhassid	6,604	13,197	Free resource giveaway
73	POV: you’re the only person at work who doesn’t know how to build...	@hooeem	5,919	13,191	Plain statement
74	How to set up the Claude completely in just 1 day:	@rubenhassid	5,246	12,942	How to / tutorial
75	The Anatomy of a Claude 4.6 Prompt:	@rubenhassid	8,027	12,513	Plain statement
76	INSTEAD OF WATCHING NETFLIX TOů NIGHT. Spend 2 hour with this.	@Krishnasagrawal	6,497	12,613	Plain statement
77	quarterly reminder to use this prompt to get your AI writing to s...	@0xPia	5,232	12,577	Plain statement
78	Stop telling Claude, "do this."	@ajitcodes	4,341	12,637	Stop doing X
79	I installed Claude Cowork yesterday	@vibhu	14,274	10,668	Plain statement
80	I interviewed a guy who gave his OpenClaw an X, stripe account, a...	@businessbarista	4,291	11,786	Personal result / income
81	3D game dev is about to change forever.	@minchoi	7,775	11,156	Plain statement
82	ANTHROPIC JUST DROPPED A 33-PAGE GUIDE.	@kirillk_web3	5,053	11,465	Plain statement
83	Most founders copy/paste their privacy policy from Stripe or ask...	@rwalk_xyz	4,779	11,437	Free resource giveaway
84	You can now run Claude Code locally for FREE.	@dr_cintas	6,300	11,083	Direct address 'you'
85	This prompt turns your AI coding agent into a premium UI/UX archi...	@kloss_xyz	4,316	11,292	Plain statement
86	How to duplicate yourself into Claude in a weekend:	@rubenhassid	4,874	11,150	How to / tutorial
87		@yasminekho	7,298	10,750	Plain statement

#	Opening line	Account	Likes	Saves	Pattern
	Stanford professor Judy Fan went on stage at MIT and broke down w...				
88	Claude is offering 13 AI courses & certificů ates.	@rubenhassid	4,793	10,848	Free resource giveaway
89	You're using Claude wrong.	@rubenhassid	4,952	10,811	You're using X wrong / Direct address 'you'
90	A senior Anthropic engineer just dropped 11-page PDF on "Loop Eng...	@0xCodez	4,890	10,613	Plain statement
91	If you’re vibecoding anything, paste the prompt below In your pro...	@hackSultan	3,417	10,669	Plain statement
92	Anthropic just dropped a 37 minute guide on building AI agents th...	@celineodier	2,835	10,739	Plain statement
93	what is agent looping	@shannholmberg	5,967	10,290	Plain statement
94	I'm just going to dump my whole agentic setup out here, because I...	@jamonholmgren	4,011	10,341	Plain statement
95	The FOMO around Claude Code is real.	@krispuckett	6,349	10,014	Us vs them / gatekeepů ing
96	Google just dropped a free 8-minute lesson on building your first...	@AnatoliKopadze	4,092	10,247	Personal result / income
97	here's how to do it with AI in 3 minutes:	@maxxmalist	5,156	10,050	Time-bound flex
98	Someone finally documented how to actually use Claude Code.	@techNmak	4,452	9,812	Superlative claim
99	Claude Fable 5 will be available again globally tomorrow.	@AnthropicAI	43,125	4,957	Plain statement
100	Claude Code is will turn the "idea guys" into "revenue guys"	@chorus_EA	6,406	9,504	Time-bound flex

D.3 Pattern composition of the bank

The hundred lines carry 105 pattern labels between them. Table D.2 gives the composition. It is a description of this bank, not an estimate of pattern effectiveness: the bank is a top-of-distribution sample, so a pattern’s frequency here confounds how well it works with how often it is used. The corresponding lift estimates, computed over the whole builder segment with the baseline stated, are in Chapter 6.

Table D.2 — Pattern labels across the 100 banked hooks. Labels are not exclusive; the column sums past 100.

Pattern label	Hooks
Plain statement	53
Personal result / income	11
Time-bound flex	9
Us vs them / gatekeeping	7
How to / tutorial	6
Superlative claim	5
Direct address 'you'	3
Free resource giveaway	3
Breaking / just launched	1
I built / I made	1
Prompt drop	1
Question hook	1
Stop doing X	1
Thread announcement	1
Warning / fear	1
You're using X wrong	1

Appendix E — The Anti-Pattern Register

The register measures 41 named patterns against posts of the same account size and the same format. It is the output of the `antipatterns` module, which the verifier closed at unsound on two fatal objections; the table is reproduced in full, with the objections attached, because the estimator itself is sound and the fatal defects are localised. Nothing here should be quoted without the `Sig.` column.

E.1 Estimator

Every post is assigned a matched stratum, `follower_tier × format`, over the builder+mixed universe (n=18,692 of 21,864). Within each stratum the module takes the ratio of the flagged group's median likes to the un-flagged group's median likes; the reported lift is the n-weighted geometric mean of those within-stratum ratios, so no post is ever compared to a post outside its own tier and format cell. A stratum enters only with at least 10 flagged and 15 un-flagged posts, and if the usable strata cover less than 55% of a pattern's flagged posts the comparison is re-run on follower tier alone. Confidence intervals are bootstrapped [Efron, 1979] over 500 resamples of the flagged group's median, holding the matched baseline fixed; because the baseline is estimated on far more posts and its uncertainty is not propagated, the intervals are slightly too narrow.

A second, pooled estimator is computed in parallel — the flagged group's median stratum-normalised index against a directly standardised weighted median of the un-flagged — and the two estimators agree in sign on 39 of the 41 rows. The two that disagree are `grok_tag` (matched +5.9%, pooled −8.5%) and `promo_hard` (matched +6.9%, pooled −4.8%); neither is significant, and neither should be read as an effect in either direction.

Methodological note

Significance is not evidence strength, and neither is the verdict. The module grades `evidence_strength` on sample size, strata coverage and estimator agreement only; it does not require statistical significance, and 13 rows charted as effects have p above 0.05. The `verdict` column additionally folds in a hand-assigned brand-risk score from 0 to 3 that is an author judgement about how a pattern reads on a founder account — it is not measured from the data. A row can therefore read avoid on a null effect. The `Sig.` column below is the only column that reports what the data says on its own.

E.2 The register

Table E.1 — The 41-pattern risk register, ascending by matched lift. Lift is against posts of the same follower tier and format; the CI is a 500-resample bootstrap on the flagged median. Sig. is p < 0.05 on the two-sided

Mann–Whitney test [Mann & Whitney, 1947]; verdict folds measured lift together with an author-judged brand-risk score.

Pattern	n	Lift %	95% CI	p	Sig.	Verdict
'Retweet and follow' mechanic	30	−50.9	−65 to −8	0.0045	yes	avoid
'DM me' CTA	33	−39.1	−56 to −3	0.035	yes	avoid
Hook: comment-to-get	107	−35.6	−42 to −23	4.3×10^{-4}	yes	avoid
Hook: us-vs-them / gatekeeping	367	−34.5	−44 to −21	<0.0001	yes	avoid
Hook: numbered listicle ('100+ AI tools')	466	−34.0	−40 to −28	<0.0001	yes	avoid
3+ hashtags	512	−33.6	−42 to −24	<0.0001	yes	avoid
Tags at least one account in the post body	2,880	−33.1	−36 to −30	<0.0001	yes	avoid
Partly-capsed hook (40–70% caps)	291	−29.2	−35 to −8	3.1×10^{-4}	yes	avoid
Any hashtag	1,116	−28.2	−34 to −21	<0.0001	yes	avoid
Tags 3+ accounts (tag-spray)	488	−28.1	−33 to −17	<0.0001	yes	avoid
5+ emojis	764	−28.1	−34 to −19	<0.0001	yes	avoid
10+ emojis	291	−27.0	−36 to −15	2.4×10^{-4}	yes	avoid
Hook: superlative claim ('the best/smartest')	885	−25.6	−31 to −18	<0.0001	yes	avoid
Thread with 100+ lines	647	−24.4	−30 to −19	<0.0001	yes	use sparingly
Hook: free resource giveaway	245	−20.9	−31 to −1	3.0×10^{-4}	yes	avoid
'Bookmark this / save this / steal this'	951	−19.4	−24 to −12	<0.0001	yes	avoid
Link with no media attached	616	−18.0	−27 to −6	<0.0001	yes	use sparingly
Soft self-promo ('I built', 'my tool')	518	−17.1	−24 to −4	0.050	yes	use sparingly
'Comment <keyword>' mechanic	787	−16.3	−22 to −11	<0.0001	yes	avoid
Engagement bait (any: comment-X / like+RT / r...	1,176	−14.8	−20 to −8	<0.0001	yes	avoid
Any self-promotional language	659	−14.6	−22 to −2	0.0068	yes	use sparingly
Hook: 'Stop doing X'	160	−12.6	−30 to +29	0.038	no	neutral
ALL-CAPS hook (>=70% caps)	1,038	−11.3	−18 to −2	0.0016	yes	avoid
'Comment X and I'll DM you'	371	−9.7	−22 to +3	8.6×10^{-4}	no	avoid
Engagement bait (corpus flag)	678	−8.6	−17 to −0	0.0043	yes	avoid
Quote post	4,609	−5.0	−8 to +0	0.232	no	neutral
Thread announcement ('a thread', emoji-thread...	201	−0.8	−16 to +31	0.967	no	neutral
Dollar-per-month income claim ($10K/mo etc.)	820	+0.0	−8 to +11	0.625	no	avoid
Long thread (400–599 words)	806	+1.4	−8 to +10	0.723	no	neutral
'Like + RT' mechanic	251	+2.8	−12 to +22	0.311	no	avoid
Hook: time-bound flex ('in 24 hours I...')	452	+4.1	−9 to +24	0.800	no	use sparingly
Tags @grok	46	+5.9	−40 to +99	0.270	no	use sparingly
Hard self-promo (sign up / waitlist / use cod...	157	+6.9	−26 to +42	0.038	no	neutral

Pattern	n	Lift %	95% CI	p	Sig.	Verdict
Hook: personal result / income flex	1,781	+7.3	+0 to +15	0.083	no	avoid
'For free' / giveaway language	903	+8.7	+2 to +20	0.104	no	avoid
Very long thread (600+ words)	482	+10.9	+2 to +25	0.026	yes	neutral
Hook: thread announcement	108	+12.2	−21 to +69	0.468	no	neutral
Contains a link	6,831	+21.2	+17 to +26	1.0×10^{-5}	yes	neutral
Opens with BREAKING / siren emoji	1,252	+32.8	+22 to +44	<0.0001	yes	use sparingly
'Follow for more' CTA	161	+34.3	+8 to +75	0.181	no	use sparingly
'Link in comments/replies'	87	+41.9	+7 to +79	0.302	no	neutral

Two further patterns were tested and suppressed under the n < 15 floor: `cta_link_in_bio` (n=12), `is_poll` (n=5). No estimate is reported for either.

E.3 Reading the register

26 of the 41 rows are significant at the 0.05 level, 23 of them negative. The negative significant block is coherent and unsurprising: reply-gate and follow-gate mechanics, tag-spraying, hashtag and emoji density, all-caps and partly-capsed openings, numbered-listicle and gatekeeping hooks, and explicit save commands. The largest measured penalty in the register, the ‘retweet and follow’ mechanic at −50.9% on n=30, rests on two matched strata and is graded weak; the strongest evidence sits in the middle of the table, where n runs into the thousands.

The positive tail deserves more caution than the negative one. Of the 10 rows with a lift above +5%, only 3 are significant, and the two largest — ‘link in comments’ at +41.9% (n=87, p=0.30) and ‘follow for more’ at +34.3% (n=161, p=0.18) — are nulls whose confidence intervals span zero comfortably. One of the 3 survivors is `has_link`, which E.4 shows is not measuring what its name says. They are reported so the register is complete, not because the corpus supports them.

E.4 The link rows are contaminated

Three rows in Table E.1 key off the `has_link` column and must not be read as statements about outbound linking. `has_link` is true whenever the substring `http` appears in the post body, and X appends a t.co URL to the body of every natively attached photo or video; the flag fires on 48.2% of media posts against 11.4% of text posts. Inside a media format it is therefore largely a proxy for having attached media, and outside one it is largely a proxy for length.

Table E.2 — The three contaminated rows, what they actually measure, and the corrected estimate.

Register row	n	Lift %	What it measures
`has_link`	6,831	+21.2	Media presence, not linking. Uninterpretable as a link effect and the source of the sign inversion described in Chapter 3.
`naked_link_no_media`	616	−18.0	The nearest usable proxy in the register: it conditions on `media_type == text`, which removes the t.co contamination, but still counts on-platform x.com links as links.
`cta_link_in_comments`	87	+41.9	A phrase, not a link. Independent of both flags; reported only for the CTA taxonomy, and not significant.
`ext_link` (not in register)	2,626	-32.7	The corrected measure, re-derived from the API url entities with on-platform hosts excluded. Follower-tier-stratified lift −33.8%; against the builder+mixed population, raw −26.7% and stratified −27.9% on n=2,354.

The gap between the naive and corrected estimates is 42.6 percentage points and the sign inverts. The register was not re-run on `ext_link`, so no matched-stratum estimate for a genuine outbound link exists; the corrected figures above are raw and tier-stratified medians reported in Chapter 3, not stratum-matched geometric means, and the two estimators are not interchangeable.

E.5 Damage does not stack

The module also measured what happens when several of these markers appear in one post. The marginal decrement is −9.1% for the first marker, then −1.6% for the second and −2.0% for the third: the gradient saturates almost immediately. The practical reading is that the first anti-pattern in a post carries nearly all of the measured cost, and that a post already carrying one is not meaningfully further penalised by a second. The causal reading — that removing markers recovers the difference — is not available from observational data.

> **Every row in this register cleared a like floor. These are differences among winners, never failure rates.**

Appendix F — Reproduction

The pipeline is a sequence of ordinary Python scripts with no orchestration layer and no database. Everything intermediate is a file on disk, which is what makes the study auditable: any stage can be re-run against the artefacts of the stage before it, and any number in the report can be recomputed from a single 28.8 MB CSV. This appendix gives the commands, the file inventory with checksums, the software versions, and the two respects in which an outside reproduction will not be byte-identical.

F.1 Commands

All commands run from the repository root with the project's Python on the path. The stages are strictly ordered; each reads only what the ones before it wrote.

Table F.1 — The pipeline end to end. Stages 1 and 2 are the only ones that touch the network or the raw payloads; everything after stage 3 is a pure function of the CSV and the module outputs.

#	Command	Reads	Writes
1	`python scrape.py`	queries.py, .env, data/done_queries.txt	data/raw.jsonl, data/done_queries.txt, data/scrape.log
2	`python analyze.py`	data/raw.jsonl	data/viral_posts.csv, data/analysů is.json
3	`python modules/<name>.py`	data/viral_posts.csv (and data/raw.jsonl for amplification_network)	data/modules/<name>.json — run once per module, 12 in all
4	`python modules/ strategy_synthesis.py`	data/viral_posts.csv, data/modules/ *.json	data/modules/strategy_synthesis.json
5	`python extract_workflow.py <journal-dir>`	the analyst / verifier journal	data/workflow_results.json
6	`python paper/build_facts.py`	data/analysis.json, data/modules/*.json, data/workflow_results.json, data/virů al_posts.csv	data/facts.json — the citable-numbers pack
7	`python build_deep.py`	data/modules/*.json, data/worků flow_results.json	report/deep-data.js, report/marketů ing-data.js, report/swipe-data.js
8	`python build_report.py`	data/analysis.json	report/data.js
9	`python build_scorer.py`	data/viral_posts.csv	report/scorer-data.js
10	`python stamp_versions.py`	report/*.html	report/*.html, cache-busted in place
11	`python paper/build_paper.py`	paper/paper.css, paper/sections/*.html, paper/figures/*.svg, data/facts.json	paper/x-viral-research-paper.html and the typeset PDF

Two optional commands are not part of the chain. `python serve.py 8899` serves the HTML report locally with cache headers disabled, which is necessary because the generated payloads keep their filenames across rebuilds. Stage 11 shells out to WeasyPrint as an external

command, `weasyprint paper/x-viral-research-paper.html <out.pdf> -e utf8`; the table of contents page numbers come from CSS `target-counter()`, so the PDF must be produced by a renderer that implements it.

Methodological note

The scrape is resumable and the key is not distributed. `scrape.py` reads `data/raw.jsonl` and `data/done_queries.txt` on start, rebuilds the seen-id set and the completed-query set, and re-issues only the grid cells that are missing. An interrupted run is restarted by invoking it again with no arguments; no state lives anywhere but those two files. The API key is read from `.env` as `TWITTERAPI_KEY`, which is git-ignored and is not distributed with the corpus. Reproducing stage 1 requires the reader's own twitterapi.io credential. Stages 2 onward require none, and are the stages every number in this report depends on.

F.2 File inventory

The inventory below is as recorded in `data/facts.json`, with the first 16 hexadecimal digits of each file's SHA-256. Total on-disk footprint is 156.5 MB, of which 123 MB is the unparsed API payload archive.

Table F.2 — Artefact inventory with sizes and truncated SHA-256 digests.

Path	MB	Rows	SHA-256 (16)
data/viral_posts.csv	28.77	21,864	1a15a8c7a1db5bc0
data/analysis.json	1.13	—	4c13ed7cac6c79d6
data/workflow_results.json	0.45	12	570e9ebb38703054
data/raw.jsonl	123.30	—	d66b4e32fad1004d
data/done_queries.txt	0.06	689	9adca2523da3f32a
data/modules/amplification_network.json	0.17	—	971bc86d424675fc
data/modules/antipatterns.json	0.13	—	0868ae69c1d908ba
data/modules/brand_share_of_voice.json	0.35	—	bc0185f5b56f10a8
data/modules/content_pillars.json	0.14	—	0fc5c0f537cc4d86
data/modules/creator_archetypes.json	0.07	—	f5f6372ae4e22e0b
data/modules/engagement_funnel.json	0.10	—	596383e4257ed717
data/modules/growth_trajectories.json	0.10	—	dbfa66bcb3df5ff2
data/modules/hook_linguistics.json	0.49	—	eeb44424486fced5
data/modules/psych_triggers.json	0.11	—	bf7cd5e992f0ecc2
data/modules/repurposing_engine.json	0.15	—	bbf540433ee60136
data/modules/strategy_synthesis.json	0.08	—	78c332415e6bd73e
data/modules/timing_cadence.json	0.27	—	dfc9762c70a2c09d
data/modules/whitespace_matrix.json	0.64	—	a149a60d87d86f0e

F.3 Software and determinism

Table F.3 — Environment as recorded at build time, and the constants that fix the stochastic stages.

Item	Value
Python	3.13.5 [Python Software Foundation, 2025]
NumPy	2.4.2 [Harris et al., 2020]
pandas	3.0.1 [pandas development team, 2026] [McKinney, 2010]
SciPy	1.17.1 [Virtanen et al., 2020]
scikit-learn	1.8.0 (TF-IDF, NMF and k-means in content_pillars and creator_archetypes) [Pedregosa et al., 2011]
WeasyPrint	invoked as an external CLI; version not captured in the environment record [CourtBouillon, 2026]
Random seed	20,260,824
Bootstrap resamples	1,000 in build_facts.py; 500 in the antipatterns register
NMF initialisation	nndsvda, random_state=42, 16 components
Suppression floor	n < 15 — any group below it is reported as null and marked suppressed
Test	Mann–Whitney U, two-sided, for every group comparison [Mann & Whitney, 1947]
Multiplicity	Benjamini–Hochberg FDR across each feature family [Benjamini & Hochberg, 1995]; raw and adjusted p both reported

Seeded stages reproduce exactly on the recorded versions. Stages 3 to 11 are deterministic given `data/viral_posts.csv`: re-running them on the distributed CSV reproduces every table and figure in this report bit for bit, and this is the reproduction claim the study actually makes.

F.4 What will not reproduce

Two stages are outside that claim, and readers should not expect them to match.

1. **The harvest.** A re-run of `scrape.py` against the live API will return a different corpus. Engagement counts continue to move after capture, posts and accounts are deleted, and `queryType=Top` is a platform ranking that is neither documented nor stable. The 448 productive grid cells are also order-dependent, since de-duplication assigns each post to whichever cell reached it first and the grid is walked concurrently across eight workers. The corpus is a snapshot with a timestamp, not a query that can be re-issued.
2. **The agent layer.** The twelve analyst modules and their verifiers were produced by language-model agents. Their code and their outputs are on disk and re-run deterministic-

ally, but the process that wrote them does not. A reproduction attempt would produce different modules asking different questions; the artefacts in `data/modules/` and `data/workflow_results.json` are the record of what this particular set of agents did, and Chapter 8 treats that record as evidence about the protocol rather than as a repeatable measurement.

The corpus, the derived CSV, the module outputs and the facts pack are distributed together so that the second of those limitations does not obstruct scrutiny of the first: a reader who disagrees with any conclusion here can recompute it from the same 21,864 rows without re-running anything upstream.

Colophon

Set in Sora. Figures drawn as vector SVG from the study's own aggregates. Typeset from HTML and CSS Paged Media with WeasyPrint; the table of contents, list of figures and list of tables carry resolved page references rather than manual ones.

Report 1.0 · 24 August 2026 · Ramanpal Singh · Promptslove · promptslove.com · admin@promptslove.com