How the X Algorithm Works and How to Grow: A Source Code Analysis

Boris Djordjevic

199 Biotechnologies · March 2026

Abstract. X open-sourced its recommendation algorithm in January 2026. This paper explains how it works and what it means for anyone trying to grow an account, particularly from a small or dormant starting point. The system uses a Grok-based transformer that scores every post on 19 engagement signals—15 positive and 4 negative—learned entirely from user behaviour, with no hand-engineered features. We explain the scoring pipeline, deduce the relative importance of each signal, and derive a concrete growth strategy grounded in the source code. Every claim is tagged by evidence source: source code (**CODE**), published empirical research (**EMP.**), or first-principles inference (**INF.**).

---

1 How the Algorithm Works

When you open your "For You" feed, the system executes a pipeline that narrows millions of posts down to a ranked list of roughly 50. The entire process takes about 200 milliseconds.

1.1 Where posts come from

Posts are sourced from two places, in parallel **CODE:**

- **In-network (Thunder).** Posts from accounts you follow, served from an in-memory store with sub-millisecond lookups.
- **Out-of-network (Phoenix Retrieval).** Posts discovered from a global corpus using a machine learning model that matches your engagement history to candidate posts via dot-product similarity.

Out-of-network is the discovery mechanism. It is how posts reach people who do not follow you. It is also penalised by a multiplicative discount factor (`OON_WEIGHT_FACTOR` < 1.0), meaning in-network content has a structural advantage **CODE.**

1.2 How posts are scored

Every candidate post is scored by a Grok-based transformer model that predicts the probability of 19 different user actions—like, reply, repost, share, block, report, and so on. These predicted probabilities are combined into a single score using a weighted sum **CODE:**

$$\text{score} = \sum_{i=1}^{19} w_i \cdot P(a_i)$$

The weight constants (w_i) determine how much each action matters. They are not published. We estimate them in Section 2.

1.3 What the model sees about you

The transformer takes your last **128 engagements** as input—every like, reply, repost, and share you made, along with the posts and authors involved **CODE.** This engagement history is how the model understands your interests. It is also how it predicts what you would engage with next.

A dormant account has no engagement history. The system literally cannot score posts for you or about you until you generate data (see Section 4).

1.4 Filtering

Before scoring, 10 filters remove ineligible content—duplicates, old posts, content from blocked or muted accounts, muted keywords, and previously seen posts. After scoring, additional safety filters remove spam, violence, and policy violations **CODE.**

Out-of-network content faces a stricter safety threshold than in-network content **CODE.**

1.5 Author diversity

If the same author appears multiple times in a feed, each successive appearance is penalised by an exponential decay function **CODE:**

$$\text{multiplier}(n) = (1 - \text{floor}) \cdot \text{decay}^n + \text{floor}$$

This means posting 5 times in an hour is counterproductive—your 5th post receives a fraction of its natural score.

2 What the Algorithm Rewards (and Punishes)

The scoring formula uses exactly 19 signals. The weight constants are hidden, but we can estimate their relative importance from code structure, platform economics, and empirical research.

2.1 The 19 signals, ranked by estimated impact

	Signal	Est. weight	Source
1	**Follow author** — user follows you from this post	∼30×	INF.
2	**Share via DM** — user sends your post in a direct message	∼25×	INF.
3	**Reply** — user replies to your post	∼20×	EMP.
4	**Share via copy link** — user copies the URL to share elsewhere	∼20×	INF.
5	**Quote tweet** — user quotes your post with commentary	∼18×	EMP.
6	**Profile click** — user clicks your name or avatar	∼12×	EMP.
7	**Click** — user clicks into the full conversation	∼10×	EMP.
8	**Share (generic)** — user opens the share menu	∼10×	INF.
9	**Dwell** — user pauses on your post (binary)	∼8×	EMP.
10	**Video quality view** — user watches your video past a threshold	∼3×	CODE
11	**Retweet** — user reposts without commentary	∼3×	EMP.
12	**Photo expand** — user taps to see full image	∼2×	INF.
13	**Favourite (like)** — baseline	1×	EMP.
14	**Dwell time** — how long the user pauses (continuous, in seconds)	∼0.1/s	CODE
15	**Quoted click** — user clicks into the original from a quote	∼4×	INF.
16	**Not interested** — user taps "show less"	∼ −20×	INF.
17	**Mute author** — user mutes you	∼ −40×	INF.
18	**Block author** — user blocks you	∼ −74×	INF.
19	**Report** — user reports your post	∼ −369×	INF.

Table 1: All 19 scoring signals from `weighted_scorer.rs`, ranked by estimated relative weight. Favourite (like) = 1× baseline. Signals 1–15 are positive; 16–19 are negative. True weight values are in the unpublished `params.rs` module.

2.2 Key observations

Likes are the weakest positive signal. Most growth advice focuses on likes. The algorithm barely values them. A like is the lowest-effort action a user can take, and its weight reflects that.

Shares are probably the most underrated signals. DM shares, copy-link shares, and generic shares are three separate dedicated signals—new in 2026. Sending a post via DM is the highest-conviction action a user can take (personally vouching to someone they know). The algorithm treats it accordingly INF..

Follows-from-post are the ultimate signal. If your content causes someone to follow you, that post receives the highest positive weighting. This rewards genuinely novel or valuable content from accounts people have not seen before INF..

Negative signals are predictive, not reactive. The Grok transformer predicts the *probability* that a user would block, mute, or report your content—and penalises your post *before anyone acts*. Content that the model expects to provoke negative reactions is suppressed pre-emptively CODE.

Negative compression is asymmetric. Positive scores scale linearly. Negative scores are compressed into a bounded band near zero. This means even moderate negative predictions can kill a post, while positive signals stack without limit [CODE].

Bookmarks are not a signal. They are not among the 19 signals in the scorer. The widely circulated claim that bookmarks carry high weight is incorrect [CODE].

3 What to Post (and How)

3.1 Text

Text posts have the highest average engagement rate on X at 0.48%, compared to 0.41% for images and video [EMP.]. They require no production overhead, enabling higher posting frequency with quality. The scoring formula has no format-specific bonus for text—it simply tends to generate more replies and dwell time [INF.].

Structure for maximum impact:

- Open with a strong first line (visible without expanding).
- Use line breaks for scannability—this increases dwell time.
- End with a question or provocative claim to drive replies.

3.2 Images

The `photo_expand_score` signal fires when a user taps to see the full image [CODE]. Design images that demand expansion:

- Infographics with text too small to read in the feed.
- Charts and data visualisations from papers or research.
- Screenshots that are partially cropped to force a tap.

Native image uploads see up to 40% more engagement than linked images [EMP.].

3.3 Threads

Threads maximise the continuous `dwell_time` signal—the only non-probability input to the scorer, measured in seconds [CODE]. A 5-tweet thread where someone reads all 5 generates substantially more dwell signal than a single tweet. Threads average 3× more engagement than single tweets [EMP.].

3.4 Video

Videos must exceed a minimum duration threshold (`MIN_VIDEO_DURATION_MS`, value not published) to qualify for the `vqv_score` (video quality view) signal. Videos shorter than this threshold receive zero contribution from VQV [CODE]. Aim for 15–60 seconds minimum.

3.5 Posting frequency and spacing

The author diversity decay means each successive post from you in the same feed session gets `decay`^{n} of its natural score. **Space posts at least 2 hours apart** to avoid cannibalisation [CODE]. A rhythm of 3–5 posts per day, well spaced, outperforms 10 posts dumped in quick succession.

3.6 What not to post

Behaviour	Why it hurts	Source
Off-topic content	Elevates P(not interested) prediction	INF.
Engagement bait ("Like if you agree!")	Trained users ignore or mute; elevates P(mute)	INF.
Combative or aggressive tone	Grok predicts higher P(block), P(mute) even with high engagement	EMP.
Spam-like patterns (copy-pasted replies)	Triggers P(report); may activate safety filters	EMP.
5+ hashtags	Associated with spam; 40% engagement reduction observed	EMP.
Posting 5+ times in 1 hour	Author diversity decay: 5th post gets decay^4 of score	CODE

Table 2: Behaviours that trigger negative signals or scoring penalties.

4 Growing from a Small or Dormant Account

This section addresses a specific scenario: an account with fewer than 150 followers, dormant or low-activity, now trying to grow.

4.1 The cold start problem

The Grok transformer requires engagement history as input. The model takes your last 128 interactions and uses them to understand your interests and predict what you would engage with. If your engagement history is empty, the query hydrator returns an error and the scoring pipeline short-circuits CODE:

```
if thrift_user_actions.is_empty() {
    return Err(format!("No user actions found for user {}", user_id));
}
```

This means step zero is using X actively—liking, replying, reposting—for at least 1–2 weeks before expecting any organic reach from original content.

4.2 Phase 1: Build your engagement history (Days 1–14)

Daily time: 45–60 minutes.

Morning (20 min):

- Like 20–30 posts in your niche. Each like enters `history_actions` and teaches the retrieval model your topics CODE.
- Reply to 5–10 posts from accounts with 1K–50K followers, targeting posts under 30 minutes old EMP.
- Quote 2–3 posts with added context. Quote tweets are a separate signal from retweets CODE.

Midday (15 min):

- Post 1–2 original tweets (text-only at this stage).
- Reply to every response you receive within 30 minutes.

Evening (10 min):

- DM 1–2 posts to people who would genuinely value them. This fires the `share_via_dm` signal—separate, high-value, and almost universally ignored by growth practitioners CODE.

- Follow 5–10 relevant accounts. Your following list determines what Thunder serves as in-network candidates, shaping your own engagement history **CODE.**

4.3 Phase 2: Establish a rhythm (Days 15–60)

With engagement history populated, the transformer can score your content.

- Increase to 3–5 original posts per day, spaced $\geq$ 2 hours apart **CODE.**
- Maintain a 70/30 split: 70% engaging with others, 30% original content **EMP.**
- Add 1 thread per week (5–7 tweets) for dwell time **CODE.**
- Add 1 image post per day designed for tap-to-expand **CODE.**

4.4 Phase 3: Compound (Days 60–180)

- 5–7 posts per day if quality is maintained.
- 1–2 video posts per week exceeding the minimum duration for VQV scoring **CODE.**
- Actively seek quote-post opportunities on larger accounts.
- Share valuable posts via DM consistently—the most underrated lever in the algorithm.

4.5 The reply strategy in detail

Replies are estimated at $\sim$20$\times$ the weight of a like. They are the single highest-impact action available to a small account for three reasons:

1. **Algorithmic weight.** A reply fires `ServerTweetReply`, one of the highest-weighted positive signals.
2. **Visibility.** Your reply appears in the thread below the original post, exposing you to the original author's audience.
3. **Profile clicks.** Readers who find your reply valuable click your profile, firing `profile_click_score` ($\sim$12$\times$) for your other content.

What makes a good reply: Add information, cite data, offer a contrarian perspective, or share relevant experience. Never write "great post" or emoji-only responses—these generate zero engagement and teach the model that your content is low-value **INF.**

Target selection: Accounts with 1K–50K followers in your niche. Large enough to have active threads, small enough to notice you. Avoid accounts with 500K+ followers—your reply will be buried **EMP.**

5 Premium Subscription

The Premium boost is not present in the 2026 recommendation source code **CODE.** It likely operates at a different layer of the stack. However, empirical data consistently reports substantial reach advantages **EMP.**:

- Premium accounts average $\sim$600 impressions per post vs. significantly fewer for free accounts ($\sim$10$\times$ advantage).
- Premium+ accounts average $\sim$1,550 impressions per post.

When to subscribe: Not on Day 1. The boost is multiplicative—it amplifies your existing score. If your engagement history is empty and your content generates no engagement, multiplying zero is still zero. Subscribe at Day 21–30, once you are posting consistently and receiving measurable engagement **INF.**

Choose Premium ($8/month) initially, not Premium+ ($16/month). The incremental benefit matters more at higher follower counts where the reach differential compounds.

6 How the Algorithm Punishes Content

6.1 Predictive, not reactive

The four negative signals—not interested, mute, block, report—are *predictions*. The Grok transformer estimates the probability that a user would take these actions and penalises your post before anyone actually does **CODE.**

Your historical blocks and mutes become training data. The model generalises: if users who engage with your niche topic tend to mute your posts, the model will suppress your content for the entire niche-interested audience segment **INF.**

6.2 Asymmetric compression

Positive scores scale linearly. Negative scores are compressed into a bounded band near zero by the `offset_score()` function **CODE.** This means:

- A post with strong positive signals and a few negative signals survives—the positive dominates.
- A post with even moderate negative signals and weak positive signals enters compression and is effectively killed.
- There is a floor—mass-reporting cannot drive a score to negative infinity. But the floor is near zero, which is functionally invisible.

6.3 The penalty hierarchy

Rank	Penalty	Recovery
1	Safety filter drop (spam, violence, policy)	Irreversible for that post
2	Blocked/muted by viewer (hard filter—removed before scoring)	Unblock/unmute by viewer
3	High P(report) prediction	Model must relearn from positive signals
4	High P(block) prediction	Same
5	High P(mute) prediction	Same
6	High P(not interested) prediction	Same
7	Out-of-network discount factor	Viewer follows you
8	Author diversity decay	Resets each feed session
9	Rate-limit shadowban (>100 likes/hr, follow cycling)	48 hours to 3 months

Table 3: Penalty mechanisms ranked by severity. Predictions (ranks 3–6) are persistent and require sustained positive engagement to reverse.

7 Expected Growth Timeline

Milestone	Timeline	What it means
Engagement history populated	Day 7–14	The algorithm can now score your content
For You feed shows your niche	Day 7–10	Retrieval model has learned your interests
First meaningful reply exchange	Day 2–3	Engagement edges forming
Subscribe to Premium	Day 21–30	$\sim$10$\times$ reach amplification
250 followers	Week 3–4	Meaningful in-network audience
Post exceeding 1,000 impressions	Week 3–4	Out-of-network retrieval working
500 followers	Month 2–3	Compounding growth begins
First viral post (10,000+ impressions)	Month 2–3	You are established in the retrieval model's embedding space
1,000 followers	Month 4–6	Self-sustaining growth

Table 4: Expected milestones for a dormant account ($\sim$100 followers) following the described strategy.

8 Daily Checklist

Morning (20 min)
- Like 20–30 niche posts (builds engagement history)
- Reply to 5–10 posts from larger accounts (target posts $<$ 30 min old)
- Quote 2–3 valuable posts with added context

Midday (20 min)
- Post 1–2 original posts (spaced $\geq$ 2 hours from each other)
- Reply to all replies on your content within 30 minutes
- DM 1 great post to someone who would value it

Evening (10 min)
- Post 1 more original post or start a thread segment
- Check metrics: which posts generated profile clicks? Double down on that format.
- Follow 3–5 new relevant accounts

Weekly
- 1 thread (5–7 tweets)
- 1 image post designed for tap-to-expand
- Review: which content types drove the most engagement?
- Unfollow accounts that pollute your engagement history with off-topic content

8 References

1. xAI Corp. *x-algorithm*. GitHub, Apache 2.0, January 20, 2026. `github.com/xai-org/x-algorithm`
2. Buffer Research. "Does X Premium Really Boost Your Reach? We Analyzed 18.8 Million Posts." 2025.
3. PostEverywhere. "How the X/Twitter Algorithm Works in 2026 (From the Source Code)." 2026.
4. Tweet Archivist. "Complete Technical Breakdown: How the X Algorithm Works." 2025–2026.
5. Social Media Today. "X Reveals Key Signals for Post Reach." 2025.

6. Pixelscan. “Twitter Shadowban: Causes, Detection & Fixes (2026 Guide).” 2026.
7. Circleboom. “The Hidden X Algorithm: TweepCred, Shadow Hierarchy, and Dwell Time.” 2025.
8. Tomorrow’s Publisher. “X Softens Stance on External Links.” October 2025.
9. TechCrunch. “X open sources its algorithm while facing a transparency fine.” January 2026.