Veridonia: A Trustworthy Online Feed.

Hlib Semeniuk

Abstract

Veridonia is an experiment in building an online feed platform that redesigns how feeds are shaped. Rather than amplifying whatever drives clicks, it introduces a governance layer that lets communities decide what deserves visibility in the feed rather than relying on engagement-driven ranking, with the goal of improving the signal-to-noise ratio of what appears there. The system is designed to structure participation so that communities can surface material they find valuable through transparent procedures, randomised review, and accountable decision-making. The approach combines sortition, multi-tier voting, and a rating system that reflects how reliably a participant's judgement aligns with the evolving standards of the group. Together, these elements form a curation framework that differs fundamentally from engagement optimisation and offers a testable alternative for building healthier information spaces.

1. Introduction

The digital landscape today is dominated by engagement metrics such as views, likes, and shares, which serve as the primary indicators of content performance. This focus on the attention economy has led to the widespread prioritisation of sensationalist content over substantive information, contributing to the rapid spread of misinformation, the formation of ideological echo chambers, and a general decline in content quality. (1, 2, 3) Furthermore, opaque algorithms that dictate content visibility and maximise engagement at any cost have amplified mistrust and decreased accountability (4).

The current paradigm of social media platforms is increasingly misaligned with the broader interests of society. Rather than facilitating meaningful information exchange, these systems have evolved into attention-capture mechanisms engineered to maximise screen time. Their algorithms prioritise engagement through psychological manipulation, often at the expense of informative or socially valuable content. As a result, content with the highest emotional appeal—not necessarily the highest quality—achieves the greatest visibility, while important discourse is frequently relegated to the margins. In effect, emotionally charged noise is amplified, while informational signal is suppressed.

Veridonia is an online feed conceived as a response to these challenges. Instead of relying on opaque engagement algorithms, it introduces a structure where communities themselves shape what rises to visibility in the feed through open, verifiable procedures. The emphasis is on rebuilding trust in how feed curation happens – who participates, how decisions are made, and how influence is earned – rather than asserting any particular view of what information is correct.

2. Problem Statement

Most large-scale social platforms optimise feeds for engagement rather than informational value. Ranking functions are tuned to maximise clicks, reactions, and watch time, not to maximise how much reliable, context-rich information a user receives per unit of attention. This has several predictable, systemic consequences:

1. **Attention is steered toward noise:** Emotionally charged, sensational, or outrage-inducing posts generate more engagement than slower, context-heavy material, and are therefore amplified disproportionally (1, 2).
2. **Misinformation spreads more easily than verification:** False or misleading content is often cheaper to produce, more surprising, and more shareable than careful corrections, allowing it to propagate faster and further in engagement-driven systems (1).
3. **Filter bubbles and echo chambers emerge by design:** Personalisation concentrates a user's feed around already-reinforced views, limiting exposure to high-signal disagreement and cross-cutting information while repeatedly resurfacing familiar, emotionally resonant material (2, 3).
4. **Curation is opaque and unaccountable:** Proprietary ranking algorithms and limited visibility into moderation decisions make it difficult for users, researchers, or policymakers to inspect how visibility is allocated or to contest systemic failures (4).
5. **There are downstream harms to individuals and society:** Feeds saturated with high-intensity, low-signal content correlate with elevated anxiety, distraction, and reduced well-being, particularly among younger users (5).

Taken together, these dynamics describe feeds with a chronically low signal-to-noise ratio: scarce human attention is repeatedly captured by vivid but low-information content, while slower, higher-signal contributions struggle to gain visibility. Addressing this requires intervening not just in individual pieces of content, but in the mechanisms that allocate visibility—how posts are selected, who participates in that selection, and how their influence is earned or withdrawn over time.

3. Veridonia's Proposed Solution

Veridonia introduces a community-driven approach to structuring visibility in its online feed. Instead of predicting what will maximise engagement, it delegates decisions about which posts appear in the feed to a sequence of transparent and well-defined steps that organise how participants review content. It does not presume that individual reviewers are free of bias; partiality and local perspective are treated as unavoidable features of human judgement. The design focus is therefore on how those judgements are sampled, combined, and rewarded. Engagement-optimised systems tend to align incentives with rapid, affective responses, reinforcing existing distortions in attention. Veridonia, by contrast, is intended to make influence contingent on consistent, accountable participation in procedures that, over time, favour contributions with higher informational signal relative to noise. Five principles underpin this design:

1. Sortition (Randomized Participant Selection)

Random sampling broadens representation and limits coordinated control, ensuring that no fixed group consistently decides outcomes.

2. Consensus (Majority-Based Decision-Making)

Simple majority outcomes determine whether a piece of content advances, anchoring decisions in shared community standards rather than algorithmic prediction.

3. ELO-Based Rating System

A dynamic rating captures how reliably a participant's past decisions aligned with the outcomes produced by their community, and uses that track record to assign additional responsibility.

4. Multi-Stage Voting Process
Tiered voting introduces structure and scalability: early checks reflect broad community diversity, while later stages rely on reviewers who have repeatedly demonstrated steady judgement within that community context.

5. Transparency and Auditability
Every moderation action, vote tally, and rating adjustment is publicly visible. This shifts trust away from assumptions about correctness and toward verifiable process.

4. System Architecture

The following components detail the implementation of Veridonia's five foundational pillars.

4.1 Submission & Review Pipeline

Veridonia evaluates each post that could appear in a community feed through a single, transparent pipeline that combines sortition (random sampling) and tiered majority voting. Random selection at each stage promotes fairness and diversity; tiering by rating concentrates final authority among proven reviewers without excluding broader participation. The mechanism scales with community size: as the population grows, random selection becomes harder to game; in very small communities the system collapses to a simpler single-stage vote.

Process Flow

1. **Submit**: The author submits a post to a specific community.
2. **Stage 1: Initial Filter (lower 70% ELO)**: A random sample from the lower 70% of ELO within that community reviews the post for relevance, informational value, and community alignment.
   Outcome:
   - If a simple majority approves, the post advances to Stage 2.
   - If a simple majority rejects, the post is rejected.
     Rating: After this decision, rating adjustments are applied to Stage-1 participants independently of Stage-2 outcomes.
3. **Stage 2: Final Decision (top 30% ELO)**: A random sample drawn from the top 30% of ELO issues the final decision by simple majority.
   Outcome: The post either enters the community feed or does not, depending on the majority decision of the selected reviewers.
   Rating: Rating adjustments are applied to Stage-2 participants after the final decision.
4. **Small-Population Mode**: For communities with fewer than 20 members, a single random sample is drawn from all available users. A simple majority decides whether to publish. Rating adjustments are applied to those participants.

Rationale and Manipulation Resistance

- **Sortition:** Random selection reduces the viability of targeted manipulation and collusion.
- **Tiering by rating:** Final decisions are made by users who have consistently shown good judgement in filtering for relevance and quality within the community's scope, while keeping early checks broad to reflect community diversity and support scalability.

- **Scalability:** Larger communities increase the entropy of selection, making coordinated capture more difficult.
- **Transparency:** All votes, outcomes, and subsequent rating changes are publicly logged for auditability.
- **Internal Echo Reduction:** Within a single community, the combination of random selection and majority outcomes tends to reward content that can attract support across factions. This pushes curation toward broadly acceptable signals and dampens the formation of narrow internal echo chambers, while still allowing distinct communities to maintain their own standards.

4.2 ELO-Based Rating System

As one of Veridonia's five foundational pillars, the ELO-based rating system reflects how consistently a participant's prior decisions have matched the outcomes produced by their community. Over time, this identifies contributors who tend to stay in tune with the group's evolving standards for what deserves visibility in the feed. Higher-rated participants are invited into more consequential review stages, not as arbiters of correctness, but as members whose past participation suggests reliability in navigating the community's expectations:

- **Dynamic Influence:** Users' ELO ratings reflect their track record of decisions relative to community outcomes. As these ratings rise, users become eligible for expanded responsibilities—such as participation in Stage-2 review or, where applicable, appointment as editors. These roles carry more weight in the curation pipeline that governs what appears in the community feed, are fully auditable, and remain conditional on continued performance.
- **Zero-Sum, Team-Weighted ELO Updates:** After the final decision, voters split into two teams: **winners** (their vote matches the outcome) and **losers** (their vote does not). Rating is reallocated zero-sum between these teams and weighted by their relative strength:
  1. Compute each team's average ELO (winners_avg, losers_avg).
  2. For each participant, compute an update scaled by a constant **K** and the gap between team averages. Members of the **winners** gain ELO, moving upward toward the opposing team's average; members of the **losers** lose ELO, moving downward toward the opposing team's average.
  3. The sum of all gains equals the sum of all losses (zero-sum conservation).
  4. This team-weighted update reinforces effective group-level filtering, amplifying the influence of participants who align consistently with community outcomes and reducing that of those who do not.

Example: One Voting Stage

Suppose five users have been selected to vote on whether a suggested post A should be published to a community X.

Their initial ratings are **800, 755, 821, 798,** and **804.**
Three vote **Yes** (users 1, 4, 5) and two **No** (2, 3). The majority outcome is **Yes**.

Step 1: Compute team averages

$$\mu_W = (800 + 798 + 804)/3 = 800.67$$

$$\mu_L = (755 + 821)/2 = 788.00$$

Step 2: Expected score for winners

$$E_W = \frac{1}{1 + 10^{(\mu_L - \mu_W)/400}} = \frac{1}{1 + 10^{(788.00 - 800.67)/400}} \approx 0.518$$

Step 3: Rating transfer

$$K = 32$$

$$\text{Total gain for winners} = K \times (1 - E_W) = 15.4$$

$$\text{Total loss for losers} = -15.4$$

Step 4: Distribution

Each winner gains $(+15.4/3 = +5.1)$
Each loser loses $(-15.4/2 = -7.7)$

After this round, the new ratings are approximately:
(805, 747, 813, 803, 809)

Although this example describes only one isolated voting stage, the same mechanism repeats continuously across many decisions and participants. Each round transfers small amounts of rating between participants whose votes align with the outcome and those that do not, and over time these micro-adjustments accumulate toward a stable equilibrium. In simulation over extended runs, the system self-organises into a characteristic distribution of ratings—most users cluster around the mean, with smaller groups at the extremes corresponding to more and less consistently well-aligned reviewers. The figure below shows this emergent pattern.

Local Rating & Elected Cross-Community Stewards (TBD)
Rating (ELO) remains **strictly local to each community**. A user's standing in one community neither boosts nor suppresses their standing in another, and **no global ELO** is ever computed. This prevents the rise of "universal elites" and keeps influence contextual to demonstrated expertise.

In the future, Veridonia may introduce a small set of elected cross-community stewards ("chief editors") with limited administrative powers (e.g., emergency takedowns, cross-community maintenance). Details of their election, scope, and accountability are **to be determined** and may draw inspiration from Wikipedia's steward model. Crucially, these roles would not mix or merge community ratings, and routine content decisions would remain governed locally.

The ELO rating system also encompasses onboarding and participation controls, detailed below:

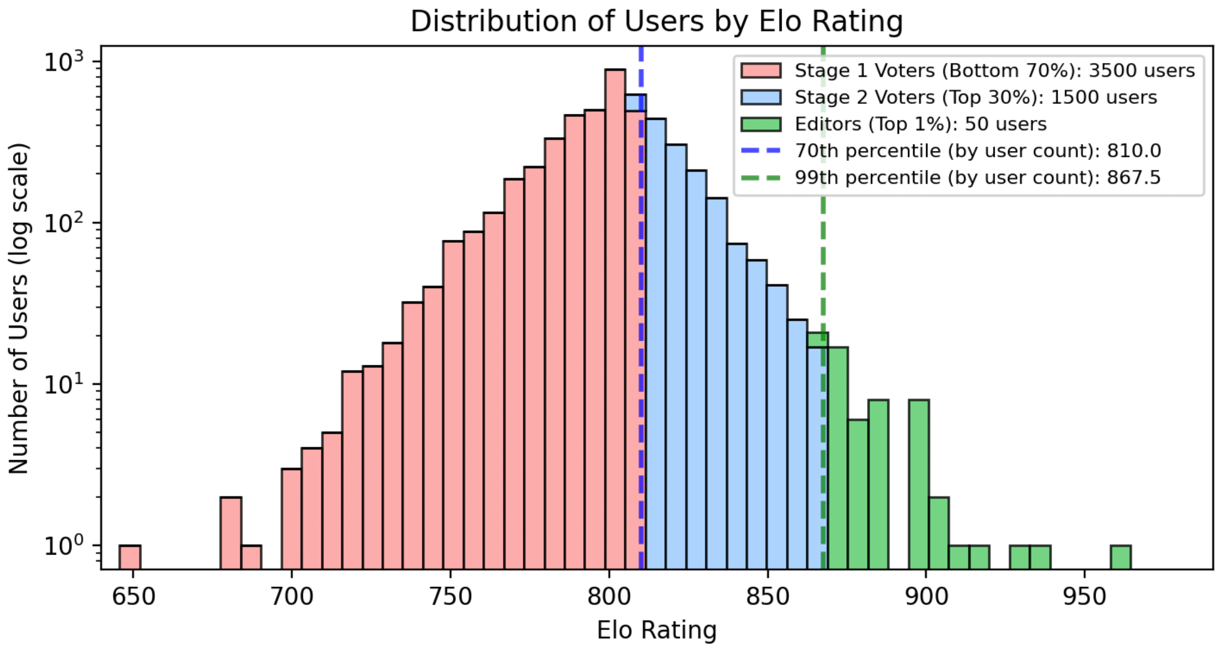

Figure 1: Distribution of Users by ELO Rating. Simulation results showing the equilibrium state produced by repeated voting and rating updates. Most users cluster around the mean, with progressively smaller groups at higher and lower ratings corresponding to the initial and advanced decision groups. Dashed lines mark the 70th and 99th percentiles by user count.

4.2.1 User Onboarding and Baseline Attributes

- **Initial Assignment:** If the IP address has no previous users, a default ELO rating (e.g., 800) is used.
- **IP-Based ELO Inheritance:** To protect the platform from bot attacks and coordinated manipulation, all new users inherit the ELO rating assigned to their IP address. After each voting stage, an IP address is updated to reflect the lowest ELO among its associated users. For example, if users from an IP (e.g., 156.156.156.3) have ELO ratings of 850, 700, and 1500, the IP is assigned an ELO of 700. Any new users registering from this IP will begin with an ELO of 700, capped at a default maximum (e.g., 800) for first-time IPs.

While this IP-based inheritance mechanism mitigates certain manipulation risks, it has clear limitations. Shared or dynamic IP addresses may produce unintended effects—including the penalisation of legitimate users employing privacy-preserving tools (e.g., Tor or VPNs). This mechanism is not foundational to Veridonia's core philosophy; rather, it functions as an initial, pragmatic safeguard and is expected to evolve as the platform matures. Potential improvements under consideration include community-reviewed verification requests, whereby users could appeal or validate their onboarding status through review by top-rated participants (e.g., the top 30% or designated editors). The precise procedures and governance structures for such processes remain to be determined and will be shaped by community input and further research.

4.2.2 ELO-Based Throttling Mechanism Veridonia implements a throttling system that regulates users' ability to post and comment based on their ELO ratings:

- **Rate Limiting:** Users with lower ELO ratings face greater restrictions on how frequently they can post content or comment. This throttling mechanism scales dynamically with ELO scores, and users with particularly low rating may be limited to posting only once per hour, per day, or even per week, depending on the severity of their low ELO.
- **Cooldown Periods:** After posting or commenting, users experience a mandatory cooldown period before they can contribute again. The duration of this cooldown is inversely proportional to their ELO rating.
- **Progressive Relaxation:** As users demonstrate consistent quality contributions and their ELO increases, throttling restrictions are gradually relaxed, allowing for more frequent participation.

The throttling mechanism serves several critical functions:

1. **Quality Control:** By limiting the volume of content from users with lower rating scores, the system naturally increases the average signal relative to noise in visible content.
2. **Spam Prevention:** Rate limiting creates an effective barrier against automated spam and coordinated manipulation attempts.
3. **Incentivizing Quality:** The direct relationship between contribution privileges and ELO motivates users to focus on thoughtful, community-aligned contributions rather than quantity.
4. **Self-Regulation:** The system creates a natural self-regulatory environment where users who consistently provide low-quality content have diminished impact on the community.
5. **Resource Management:** Throttling helps manage computational resources by preventing system overload from excessive low-quality submissions.

This mechanism reinforces Veridonia's core principle that influence within the community should be earned through demonstrated alignment with community standards and quality contribution.

5. Transparency and Self-Governance

Veridonia is designed to be an open and self-regulating ecosystem:

- **Public Auditability:** All voting records, ELO adjustments, and moderation actions affecting what appears in the feed are logged and accessible for independent review, emulating blockchain-like transparency.
- **Decentralised Moderation:** Governance is vested in the community, with every member empowered to contribute, vote, and shape content standards for their feeds. Moderation rights are held by approximately the top 1% of users in a community by rating, who are able to soft-delete posts from the feed to uphold standards. Every moderation action can be appealed by the broader community through a randomized jury voting process.

Privacy by Design and Data Control:
Veridonia does not require sign-up to participate and does not track users across the web. The system only uses minimal signals necessary for fairness—for example, new accounts inherit the lowest rating from their IP to discourage bot farms. Beyond this, rating is tied entirely to actions within the platform: voting, posting, and how those decisions align with the community.

At the same time, users retain full control of their data. All activity histories can be accessed and verified without compromising security, reinforcing both transparency and individual privacy.

6. Additional Core Principles

Independence from Advertiser Influence: Veridonia is free from advertiser funding, ensuring that feed curation is not driven by advertiser incentives and is dictated solely by community standards. The platform will never employ advertising as a monetisation strategy. Instead, future revenue models may involve subscriptions or donations, but public benefit content—such as community feeds—will always remain free to access. Only private benefit features may be offered as paid options, balancing sustainability with Veridonia's commitment to open access and public good.

7. Conclusion

Veridonia is an experiment in community-guided online feeds. By pairing sortition and tiered voting with an ELO-based rating model, it replaces engagement optimisation with incentives that reward careful participation. Contributions that repeatedly fall outside the community's standards carry rating and throttling costs, while steady, attentive decisions expand a participant's role in shaping what the community sees in its feeds. The expectation is that such incentives produce feeds that feel more deliberate, legible, and aligned with the community's own preferences, with a healthier balance of signal over noise than engagement-driven alternatives.

The next step is empirical: testing the system under real conditions and deliberate stress. We will evaluate how feeds distribute attention between substantive contributions and noise, as well as overall content quality and capture resistance. We will also examine decision latency versus judgement alignment and the fairness of IP-based inheritance to refine both the model and its parameters.

A second question is sustainability. We will evaluate whether a non-advertising model – driven by voluntary support or subscriptions – can fund operations without distorting incentives, while keeping core public-benefit features open.

Ultimately, Veridonia is a falsifiable proposal. If outcomes under real use do not beat practical baselines—or if funding compromises the aims—it should be revised or retired. If they do, the system may be worth iterating on. We invite researchers and communities to test, critique, and adapt these ideas.

References

1. Pennycook, G., & Rand, D. G. (2021). "The Psychology of Fake News." *Trends in Cognitive Sciences*, 25(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

2. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., & Zittrain, J. L. (2018). "The Science of Fake News." *Science*, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

3. Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). "The Echo Chamber Effect on Social Media." *Proceedings of the National Academy of Sciences*, 118(9). https://doi.org/10.1073/pnas.2023301118

4. Ananny, M., & Crawford, K. (2018). "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society*, 20(3), 973–989. https://doi.org/10.1177/1461444816676645

5. Keles, B., McCrae, N., & Grealish, A. (2020). "A Systematic Review: The Influence of Social Media on Depression, Anxiety, and Psychological Distress in Adolescents." *International Journal of Adolescence and Youth*, 25(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851