

# A world full of Information Disorder.

In 2019, a book marketed a \$20 "Miracle Mineral Supplement" as a cure for every major disease known to man. It preyed on the desperate, but the reality was far more grim: the FDA debunked it as a dangerous substance that could lead to kidney failure.

So how does this happen? It's not just a technical glitch; it's a fundamental hijacking of how we learn.

Humans are natural **social learners**, constantly scanning our environment to mimic others and understand norms. But when you add **algorithmic amplification**, extreme views suddenly seem legitimate and widely held.

We aren't just looking at a few "fake news" stories; we are living in an **Information Disorder**.

---

To fix the problem, we have to understand the three distinct ways information breaks down.

First, there is **Misinformation**: inaccurate information shared without the intent to harm. Second, there is **Mal information**: genuine information that is deliberately stripped of its context to cause harm. And finally, there is **Disinformation**: false information created with the specific intent to mislead or manipulate, like fabricated crime statistics from non-existent agencies.

These spread through **Echo Chambers**, where our own beliefs are reflected back at us, and **Filter Bubbles**, where algorithms isolate us from any perspective that doesn't drive engagement.

---

Our current mitigation tools are struggling...**Journalistic fact-checking** is nuanced but slow and often too late to stop a viral lie. **Automated AI moderation** is fast but lacks the ability to detect nuance, often leading to unintended censorship.

Unfortunately though, most platforms operate on an Engagement **Algorithm**. These platforms reward virality over accuracy, and because of human nature we engage with things that trigger an emotional response which makes more users post more polarizing content for the views.

With that being said it makes sense for a chronological, network-based feed right? Well.... No. Chronological or network-based feed can actually strengthen filter bubbles by never exposing users to new content.

---

In order to promote a safer environment, I propose a system that focuses on three areas:  
**Algorithm Moderation, A.i. Aided Fact-checking, and Media Literacy.**

**Phase 1: The Downvote.** In the real world, we have "decency standards" that provide social feedback. By implementing a community-driven downvote or flag system, we give the community a voice. When a post reaches a certain threshold of flags, it is prioritized for review.

**Phase 2: A.i. Aided Fact-Checking.** We can't rely on humans alone to handle the speed of the internet.

By merging A.i.'s efficiency in identifying previously debunked claims with a human's ability to interpret context, we can drastically reduce the "lifespan" of false information.

**Phase 3: Pre-bunking.** Think of this as a "driving lesson" for the digital age.

We wouldn't give a kid keys to a car without driving lessons; so why do we give them the internet without **media literacy**?

Pre-bunking (or microdosing) users with examples of deceptive messaging so they can recognize and dispute it in the future, essentially building antibodies against deception.

---

This creates a cycle of accountability. When controversial content is flagged by the community and reviewed by the A.i-Human team, it doesn't just disappear. If deemed a threat, that content is transformed into a **pre-bunking example** to educate the public for the future.

By utilizing the speed of A.i with human judgment and proactive education, we can finally move toward a social media environment that is both safe and trustworthy.