From Likes to Lies, The Untold Story of TikTok's Algorithm
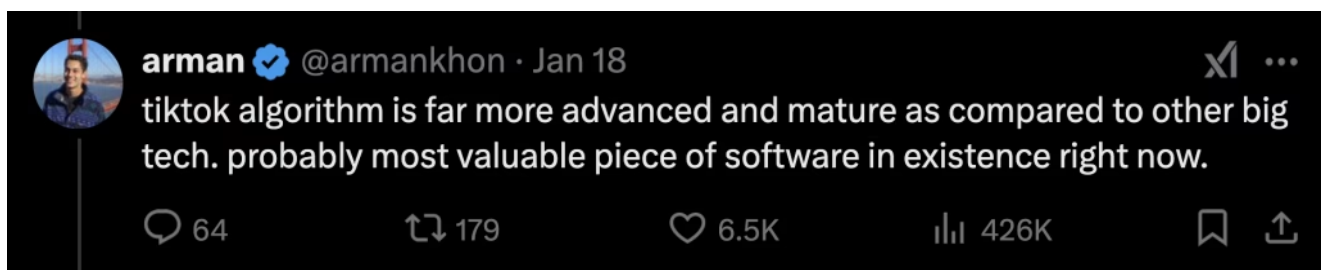
I really enjoyed the boldness with which Arman, who quit his $330k engineering job at TikTok, replied to great questions. Some people on X replied with sarcasm that the claims I'm going to show you were for engagement. Maybe. Maybe not.

DISCLAIMER: I never used TikTok (many reasons, from atrocious bias to data security). Months ago, I wrote a private article about my concerns regarding keylogging. I was advised not to publish it and I didn't.
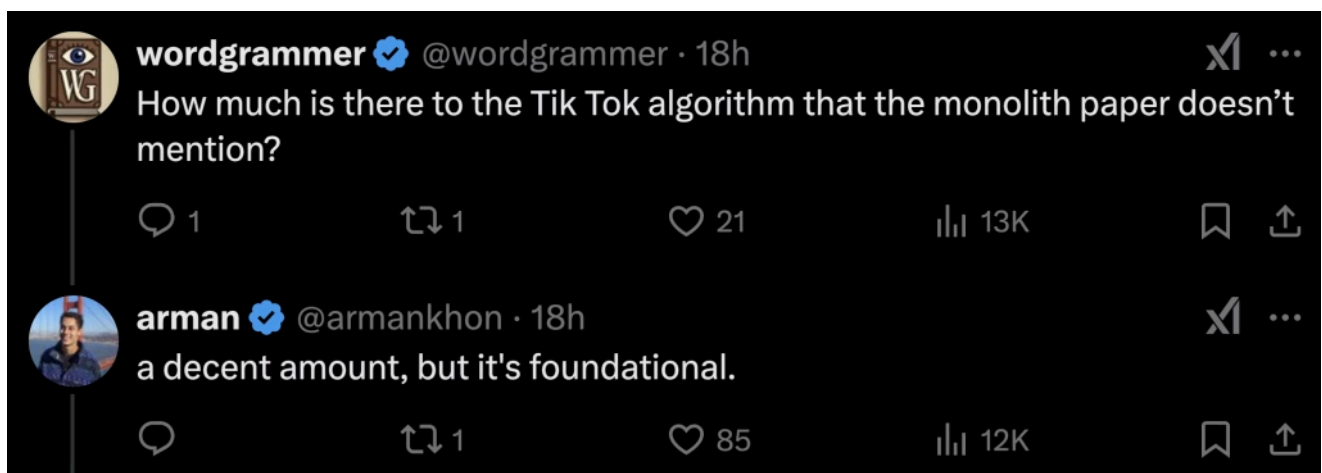
I'm firmly against what TikTok secretly did and still does. But at the same time, I recognize beautiful engineering when I see it.

Here's my compilation of the best questions and answers. I've focused on the technical and human dimension. I ignored the political dimension.

arman ✔ @armankhon · Jan 18

tiktok algorithm is far more advanced and mature as compared to other big tech. probably most valuable piece of software in existence right now.

💬 64   🔁 179   ♡ 6.5K   ᴘᴧ 426K   🔖 ⬆

I find this a bit excessive so let's break things down. The TikTok algorithm is based on a lightweight recommendation system named Monolith. It was created by ByteDance engineers.

I'll share the foundational paper because these engineers deserve the credit.

wordgrammer ✔ @wordgrammer · 18h

How much is there to the Tik Tok algorithm that the monolith paper doesn't mention?

💬 1   🔁 1   ♡ 21   ᴘᴧ 13K   🔖 ⬆

arman ✔ @armankhon · 18h

a decent amount, but it's foundational.

💬   🔁 1   ♡ 85   ᴘᴧ 12K   🔖 ⬆

Nothing would have happened without these engineers. Of course, if you think socially or geopolitically, did many terrible things happen because of them? Come up with your own answer. This article isn't a blame game: we must focus on other dimensions for clarity.

[Monolith](#) is a deep learning framework for large-scale recommendation modeling and its 2 key features are

- collisionless embedding tables
- real time training

Real time training means the system learns and updates its recommendations as soon as new data comes in. If there's a sudden trend like a new song everyone starts listening to, the system will quickly adjust its recommendations to show this new trend to users, without delay.

But because Monolith uses TensorFlow, a construction kit for building AI models, there's real-time training and batch training: the AI model can learn from a large amount of data all at once (helpful to learn from historical/past data).

Collisionless embedding tables helps with differentiation and understanding. Suppose you have a bunch of unique stickers, each representing different items or users. In a recommendation system, you want each sticker to have its own unique spot on a board where it won't overlap with another sticker. You want every item or user to be represented distinctly.

Now you understand the meaning of the keyword COLLISIONLESS: no collision!

I suggest reading [Monolith: Real Time Recommendation System With Collisionless Embedding Table (PDF)](#) to go deeper. Amazingly, the [Monolith source code is on GitHub](#) 😀

TikTok's Back-end programming languages

Assembly was a joke but Go, Python, C++ and Java are all common choices for large-scale applications. I'm sure there are a few other languages (android backend services and server-side logic in iOS, web app, etc).

TikTok: Scale and user engagement

Ayu @Ayuu2809 · 6h

How many request did you guys handle?

💬 1　　🔁　　♡ 47　　📊 50K

arman @armankhon · 6h

1B+ daily active users, 170M in the US, average 58-minute screentime per day, this is the scale of the app.

💬 19　　🔁 102　　♡ 832　　📊 182K

Imagine the integration of microservices, caching, load balancing, sharding, monitoring… unreal! Imagine 58 minutes of user engagement locked into a single closed ecosystem… Pure addictiveness. A dream for marketers, neuroscientists and those who analyze behavioral patterns.

object Object @namedobject · Jan 18

did you learn anything about human behavior or tendencies by how users used TikTok

💬 1　　🔁　　♡ 97　　📊 72K

arman @armankhon · Jan 18

human behavior is what social media companies exploit, yes.

💬 5　　🔁 44　　♡ 1K　　📊 70K

True love of the TikTok app

Krishna Kumar @krishnapro_ · 8h

do you like Tiktok?

💬 1　　🔁　　♡ 5　　📊 31K

arman @armankhon

honestly I never used the app until I started working there and had to test my code.
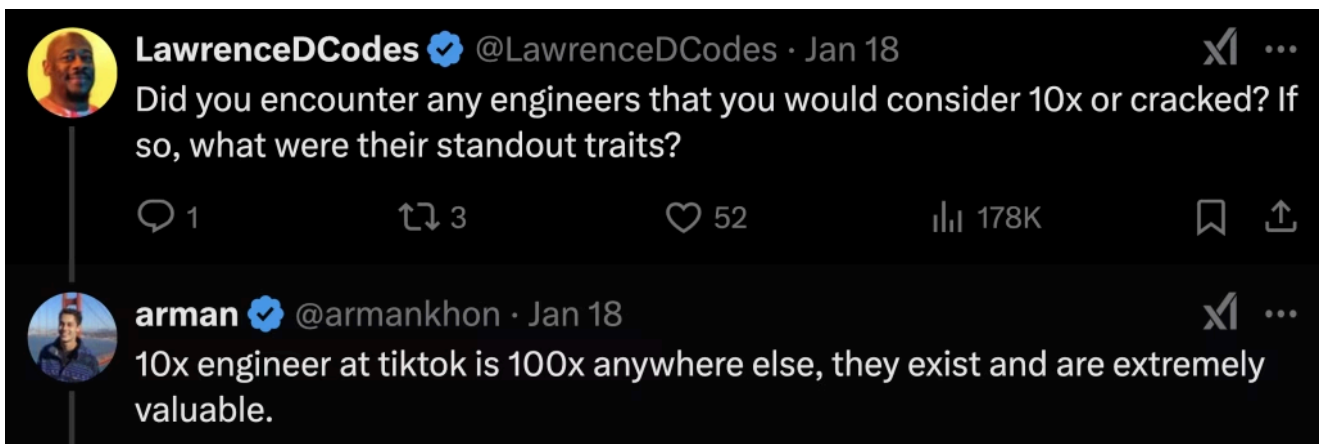
6:49 PM · Jan 19, 2025 · **30K** Views

Brutal honesty. Refreshing. I personally boycott the app but I refuse to let my disdain cloud my analysis. No pun.

About his teammates

Based Thought ✅ @based_thought · 6h

What's your biggest takeaway from working with the cracked engineers at TikTok?

💬 2   🔁 1   ♡ 24   ᵢₗᵢ 40K

arman ✅ @armankhon · 6h

cracked engineers are rare and extremely underpaid.

💬 1   🔁 9   ❤️ 246   ᵢₗᵢ 38K
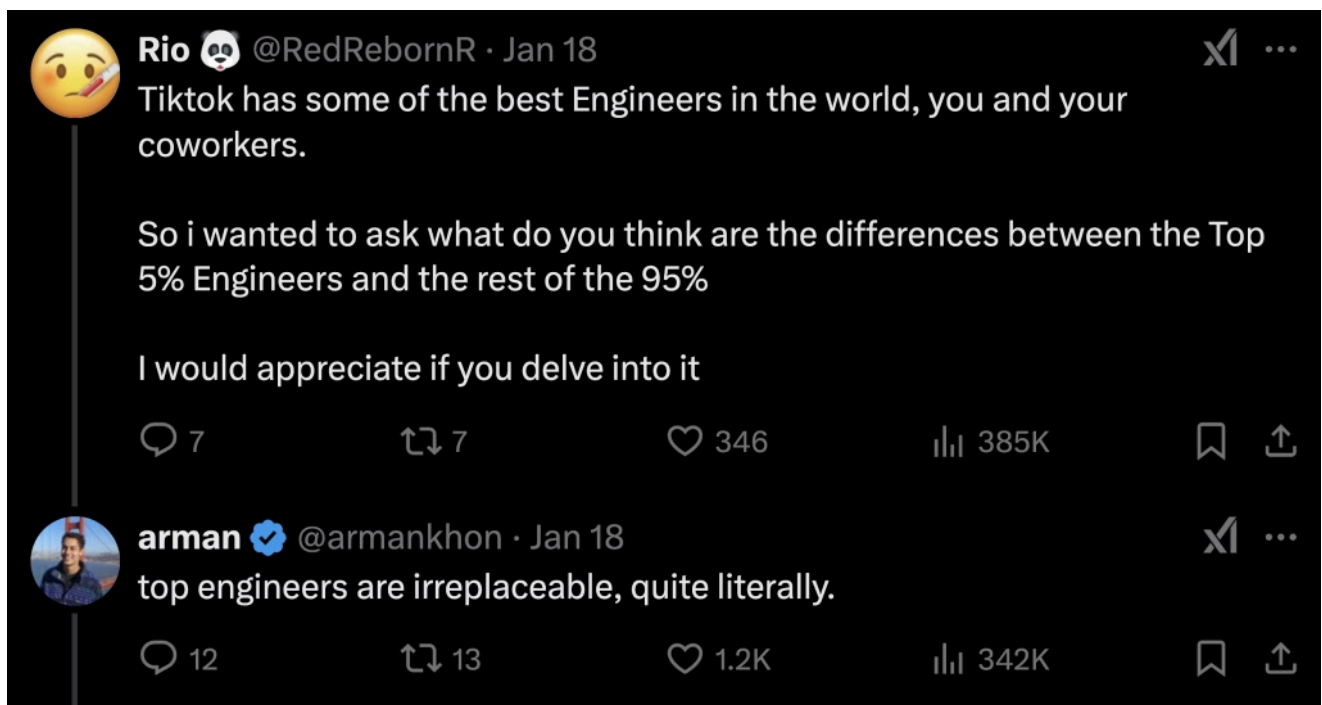
I 100% agree. And let me add that for data giants, what should matter is how people think, not where they live. I have an article in mind on this topic because pay based on location isn't acceptable. News flash: top talent isn't paid based on location in the real world (source: me).

LawrenceDCodes ✅ @LawrenceDCodes · Jan 18

Did you encounter any engineers that you would consider 10x or cracked? If so, what were their standout traits?

💬 1   🔁 3   ♡ 52   ᵢₗᵢ 178K

arman ✅ @armankhon · Jan 18

10x engineer at tiktok is 100x anywhere else, they exist and are extremely valuable.

> Price is what you pay, value is what you get —Warren Buffet.

Rio 🤒 @RedRebornR · Jan 18

Tiktok has some of the best Engineers in the world, you and your coworkers.

So i wanted to ask what do you think are the differences between the Top 5% Engineers and the rest of the 95%

I would appreciate if you delve into it

💬 7   ↻ 7   ♡ 346   ᴨᴨ 385K   🔖 ⬆

arman ✓ @armankhon · Jan 18

top engineers are irreplaceable, quite literally.

💬 12   ↻ 13   ♡ 1.2K   ᴨᴨ 342K   🔖 ⬆

No answer… but a true answer nevertheless. Top engineers are irreplaceable. I've worked with some of the best. It makes a BIG difference. They can be incredibly smart, solve complex problems, and sometimes ruin everything after 3 weeks of all-nighters because of a stupid killall (source: don't ask —but also yes).

Cybersecurity concerns: nothing to see

Kron ✓ @Kronykal · 6h

Why does TikTok's in-app browser have a keylogger and no option to open out of app?

💬 2   ↻   ♡ 90   ᴨᴨ 36K   🔖 ⬆

arman ✓ @armankhon · 5h

industry standard practice. meta does the same thing.

💬 2   ↻ 3   ♡ 243   ᴨᴨ 33K   🔖 ⬆

Kron ✓ @Kronykal · 4h

For what purpose? And Meta allows for out of app browsing.

💬 2   ↻   ♡ 20   ᴨᴨ 6.8K   🔖 ⬆

arman ✓ @armankhon · 4h

supposedly debugging and monitoring ad performance.

💬 1   ↻   ♡ 36   ᴨᴨ 5.8K   🔖 ⬆
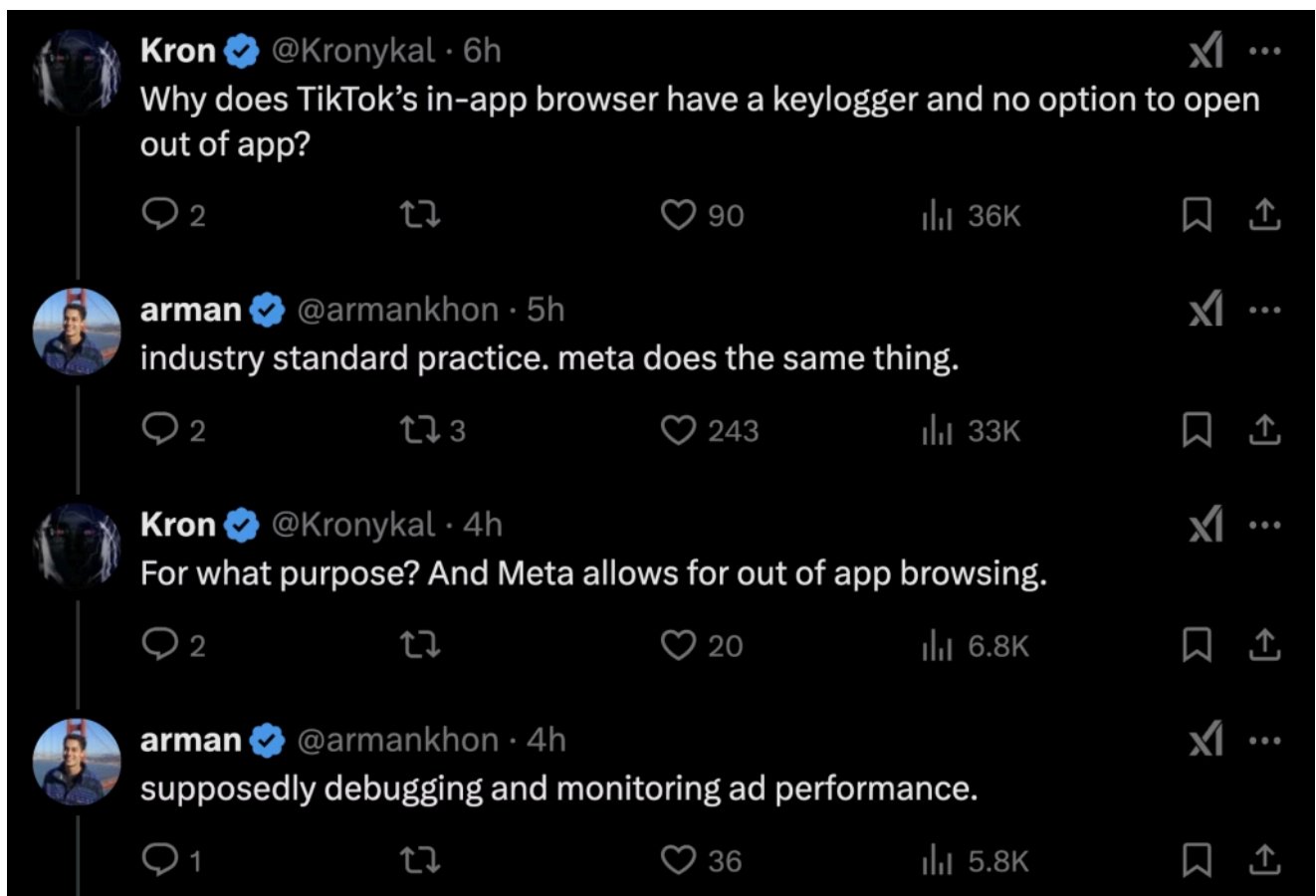
I loved the "**supposedly**" keyword. Although this will be my only comment on the topic of cyber threats, the reality when I investigated was worse than anything I had imagined. Zero

trust as far as I'm concerned.

Data Scraping

Wrong. Not the same as America. Different approach. Jilian is right here: it is a little sus.

Bytespider: aggressive ByteDance web crawler

Please don't get me started on the high crawl rate of **Bytespider**! ByteDance's web crawler aggressively scrapes content from websites to fuel ByteDance's AI models (and for TikTok).

I've noticed a complete disregard for small website owners' content and infrastructure. Bytespider often ignores the crawling directives in the robots.txt file: [RFC 9309 – Robots Exclusion Protocol](#)

There is zero transparency: the purpose and extent of this systematic data collection by Bytespider is not clear at all. High risk of potential misuse + real privacy concerns.

I hope you realize that you could be against TikTok, you could boycott TikTok and yet your data could end up in ByteDance's obscure dataset because they aggressively scrapped content from your website. You have a portfolio? You shared pictures? You have an online CV/Resume with personal data? Your work employment history? Well…

I suggest blocking the Bytespider user agent at the web server level (Nginx/Apache) but that won't be enough. You should add firewall rules to block all traffic from Bytespider. Easy to do in Cloudflare but ByteDance has a lot of IPs. Let me know if you want a guide to protect your data.

Speaking of user privacy

Ryan Grim ✅ @ryangrim · Jan 18
what is my mother's maiden name?

💬 34   🔁 17   ♡ 1.4K   � 288K

arman ✅ @armankhon · Jan 18
we actually have that data on you.

💬 22   🔁 18   ♡ 1.7K   features 253K

No comment but imagine the compound effect of an average of 58 minutes spent daily on a closed app when your brain isn't even formed yet. You are so young and immature that you'll share anything spontaneously about yourself and close ones. A privacy treasure and a privacy nightmare. No need for eavesdropping (passive listening through smartphone mics) here. I'm not saying it isn't implemented, I'm just saying you can count on users to do the wrong thing.

Security threats? Yes? No? Maybe?

arman ✅ @armankhon · Jan 17
if tiktok poses a national security threat, then meta poses a domestic terrorist threat.

💬 64   🔁 431   ♡ 4.6K   features 127K   🔖

Margo ✅ @MargoinWNC · Jan 17
Been saying this since the ban was introduced. Zuckerberg coming on Rogan saying "I'm all for free speech"-is a load of rubbish

💬 5   🔁 5   ♡ 220   features 11K   🔖

YouTube Shorts (Google/Alphabet) & Instagram (Meta)

You may think the gap between TikTok and competitors like YouTube and Instagram is subjective. But TikTok has a significant lead in the short-form video market. And that's primarily because of an algorithmic reality: their algorithms are simply better in some regards.

I also should point out a common misconception: the algorithm alone isn't responsible for TikTok's success. It is crucial but there's an interplay between the algorithm and networked data.

NotEvenTrying× @rizkidotme · Jan 18
which one more important: algo or networked data?

💬 2   🔁 2   ♡ 42   📊 143K   🔖

arman ✓ @armankhon · Jan 18
tiktok algorithm is most powerful piece of software in existence.

💬 31   🔁 64   ♡ 1.4K   📊 145K   🔖

BL●CKY ✓ @Blocky256 · Jan 18
It is hands-down the most impressive and adaptive algorithm for human engagement.

💬 1   🔁 1   ♡ 36   📊 3.6K   🔖

In this question, "networked data" is the mind-blowing amount of data users generate on social platforms (think interactions, preferences, and content consumption patterns). This data is the fuel that powers the TikTok algorithm. The algorithm processes this networked data to personalize the user experience, recommend "relevant" content, and predict what will keep users engaged.

EchoOfTheLoop ✓ @echosofreality · 17h
controlling TikTok's algorithm means controlling influence, commerce, and culture and ultimately the narrative of a generation

💬 4   🔁 31   ♡ 404   📊 33K   🔖 ↥

arman ✓ @armankhon · 16h
hence why bytedance will sell under zero circumstances.

💬 8   🔁 11   ♡ 437   📊 31K   🔖 ↥

Yep: there will be no selling under any circumstances if the algorithm is included. Why? Because the initial algorithm has been fine-tuned and it reached such an advanced level that it is now quite different from the foundational algorithm. It is no longer a Monolith, it is a Masterpiece. That's what makes the advanced version "priceless".

Yesterday, I trolled under a CNBC post claiming Perplexity AI was about to acquire TikTok. I quoted the previous CNBC article contradicting the current CNBC article.

Virality on social platforms

Not great long-term advice for technical and human reasons. Social platforms constantly update their algorithms and if you rely on past successes, you'll end up failing. What worked before might not work now and what works now won't work tomorrow. If you think short-term, sure: repeating the same elements might work for a while, but audiences get bored quickly.

But there's a reason very few engineers are also content creators. There are safe bets however: being honest, respectful and helpful will never go out of style, no matter your level.

For normal people, creating high-quality and engaging content is extremely time-consuming.

Unless you are beautiful, there's no secret formula, just hard work that no one sees. Sometimes, you can have surprising results because virality means the whole world discovers you in a few hours. For good or bad reasons. Everything is easier for beautiful people, offline and online. **Reinforcement**!

Anthony Khalil @ @ImAnthonyKhalil · Jan 18

LMAOO does the algorithm favor people that look more aesthetically pleasing

💬 1   ⟲ 2   ♡ 29   ᴉᴉ 31K   🔖

arman @ @armankhon · Jan 18

humans do this themselves, algorithm reinforces.

💬 1   ⟲ 19   ♡ 651   ᴉᴉ 31K   🔖

Each social app can become a prison that follows you everywhere.

i @ @willswim2day · 6h

Would you allow your children to use TikTok?

💬 4   ⟲ 15   ♥ 212   ᴉᴉ 39K

arman @ @armankhon · 6h

zero chance, brainrot. ⟵ **TikTok**

💬 33   ⟲ 108   ♥ 1.5K **engineer**

Elon Musk @ X @elonmusk · 4h

Concerning

💬 123   ⟲ 56   ♥ 967   ᴉᴉ 73K

arman @ @armankhon · 4h

I don't disagree, yet we unbanned it.

💬 5   ⟲ 1   ♥ 140   ᴉᴉ 16K