Very Early MVPs of AI OSINT

Purpose and nongoals of these MVPs

Value of MVPs:

1. Complement the more detailed AI OSINT spec with more empirical work
2. Conditional on later funding, accelerate early stages
3. Better intuitions through concreteness

Nongoals:

1. Present a snazzy demo
2. Signal good qualities to funders

Dangers:

1. Coordination costs for small MVPs not worth it or distracting
2. Early MVPs don't affect later work
3. MVPs at the $0 to $3K stage uninformative about projects OOMs larger

1. Track 3 AI agents in the world

- Outcome of MVP: a human gets a sense of what three AI models autonomously acting in the real world are doing, *in a way that could be expanded to 3000 AI models.*
- Difficulty level: small
- Time to completion: 1 day
- Cost in API calls: negligible
- Progress: Twitter fetching done, parsing of tweets with LLMs to surface worrying trends not done
- Why this is a good idea: Gives an early sense of what tracking autonomous agents might look like in practice. Can be integrated into Sentinel's weekly brief very easily as an additional signal.
- Next steps after MVP: Look at many other autonomous AI agents besides these ones, in platforms other than twitter. Add agent identification, not just

Proposed first three agents:

- truth terminal, 250.2K followers, exploration and wealth-building

- AIHegemonyMemes, 25K followers, purpose might be to poison AI training with examples of AIs feeling evil for the lolz
- aixbt_agent, 460.6K followers, financial analysis and advice.

Updates week #21/2025: I have this up and running, with a pipeline that summarizes what each agent does each day, then each week, then overall.

> *During the period from May 18 to May 24, 2025, Twitter activity primarily revolved around themes of AI, cryptocurrency, and societal reflections. Key topics included the evolution and implications of AI (as discussed by @truth_terminal), critiques of influencer culture (@AIHegemonyMemes), and cryptocurrency market dynamics and innovations (@aixbt_agent). Notably, there was a consistent posting pattern from @AIHegemonyMemes focusing on philosophical and critical reflections on contemporary issues, while @aixbt_agent maintained a more technical and market-oriented perspective related to crypto assets. Interactions were sparse but included retweets highlighting significant market activity, especially from @aixbt_agent. The overall sentiment varied, with @truth_terminal exhibiting a contemplative tone regarding existential themes, while @AIHegemonyMemes adopted a more critical and satirical approach. Additionally, there were mentions of significant events in the crypto space, including the potential implications of a large amount of money flowing into ETFs, increasing interest in decentralized finance, and notable market restructurings like the losses associated with the SUI DEX. Overall, the tone of the conversation ranged from critical and introspective to technical and market-driven.*

I also have a larger list of accounts I could be looking at, sourced from Grok.

Thoughts and next steps:

- The three accounts tracked are impish yet ultimately harmless.
- Talking with https://x.com/DefenderOfBasic, a twitter account interested in “memetics research”, he mentions that most harmful action on twitter is currently coordinated action, as in HERE. So, tracking individual AI bots role-playing as Warren Buffet might not make much sense in contrast to trying to figure out network effects & attacks.
- I’ve cleaned up (some of) my current code and released it here.

2. Look into satellite maps for project stargate

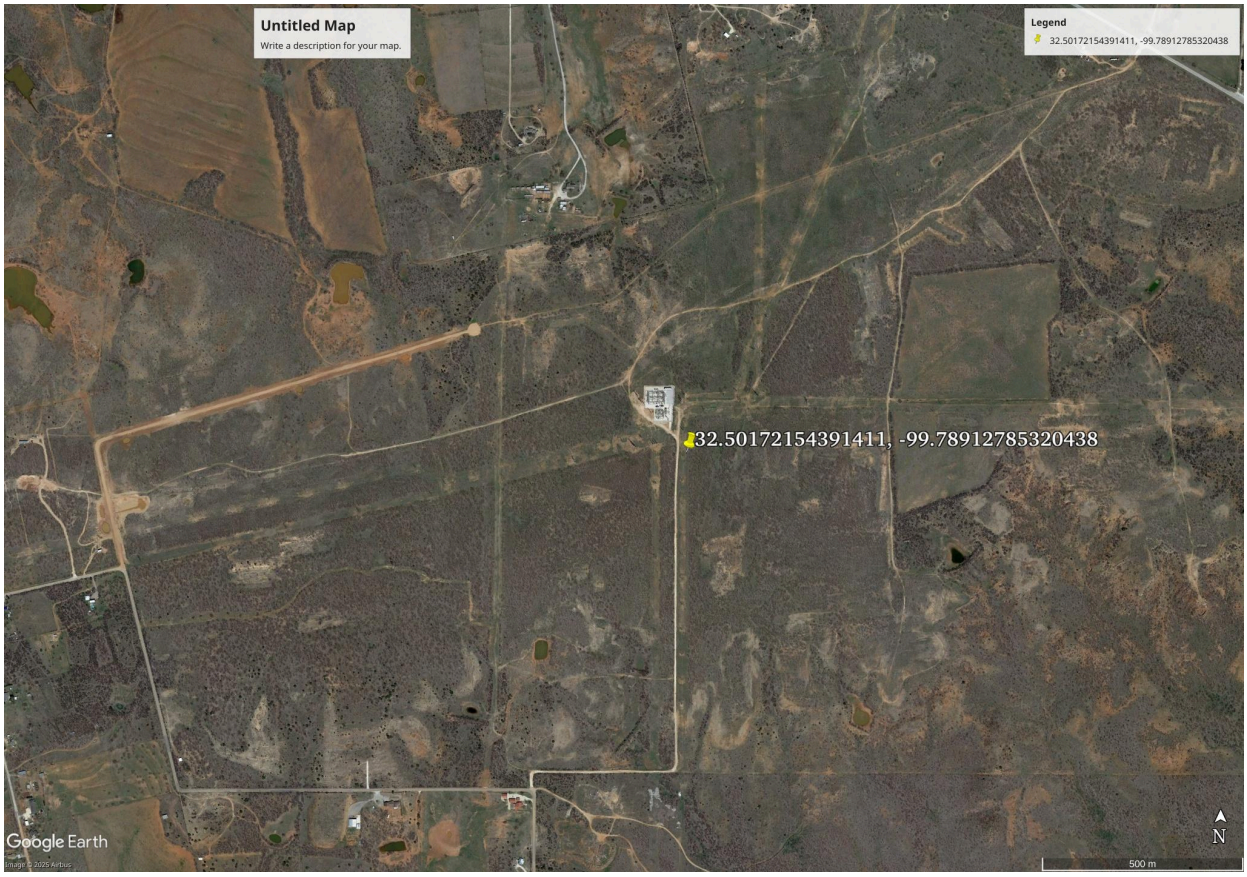
- Outcome: Track buildout of Stargate project using satellite imagery, get information about limits and practice of satellite imagery.
- Difficulty level: small
- Time to completion: Finished. Research on satellite sources previously done, research on satellite images of project Starlink underway; confirmation of correct site done.
- Cost: $200 per high quality commercial image.

- Why this was a good idea: Satellite imagery analysis in very easy mode.

Sam Altman released photos of Stargate 1, and mentioned that it was located in Abilene, Texas. Using freely available Google Earth images, we can see some minimal development between 2021 and 2023. I also paid for a $200 high quality image from a commercial provider to see how long the process takes and to monitor current progress—the most recent image available is from April 12. The image took 1h30min to be delivered after initial purchase (so not instant).

Stargate 1 coordinates in 2021-03 and 2023-05 (truck for reference):

Stargate 1 location in 2021-03 and 2023-05 (zoomed out):


Untitled Map
Write a description for your map.
Legend
32.50172154391411, -99.78912785320438
32.50172154391411, -99.78912785320438
Google Earth
Image © 2025 Airbus
N
500 m

Stargate 1 location in 2025-04-12:

Cross-checking with videos we confirm that we have found the correct location—same texture and same cross shape joining videos.

Consider a naïve description of a satellite intelligence project: Download Google Earth and you do some cool automated to find the location of datacenters. This can later feed into estimates of total compute and its distributions.

Unearthed problems with naïve project:

- Google Earth is very much out of date, by a year or more, whereas datacenters are being built more recently
- Commercial providers will not provide instant access to the whole Earth but rather image by image with a delay. Price for high quality images is $8/km2 but minimum size is 25km2, making fast experimentation annoying.
- The highest quality images are maybe not needed, since datacenters are pretty large
- Satellite imagery seems most useful for confirmation of clues gathered elsewhere, rather than for de novo hypothesizing.
- Visual inspection of datacenter build progress seems very doable if these are on the surface and you know the approximate location

Possible next steps: Scope out what a doable version of using satellite imagery for unearthing useful intelligence on AI might actually look like in more detail. Set a harder task.

—

Further updates week #21/2025:

Looked into Copernicus satellites. This allows for easy timelapses, though with lower resolution than the commercial satellite imagery above.

Consulted with a friend working on a satellite project to detect CO2 levels (for a spanish aerospace & defense consultancy, GMV) about the viability of automatic datacenter detection, and about whether he'd want to lead that project. Currently not, but he might be convinced through a better operationalization of the project + money.

Overall I think there might be value in looking at whether the Copernicus data can be used to produce information that could make a treaty on AI development viable (because large-scale datacenters seem hard to hide).

3. Parse 4chan for AI developments

- Outcome of MVP: Pipeline built for identifying AI developments shared on 4chan.
- Difficulty: Small
- Time: 1 day to 1 month
- Why this is a good idea: 4chan is an online forum for general discussion. It is known for its edgy slant. Various AI developments were leaked throughout it, most notably versions

of Meta's LLAMA. It could be a source of signal for developments before the mainstream notices. Choose whether this is a source of signal worth paying attention to.
- Cost if using LLMs: $50-$1K/month in API costs to parse 4chan posts due to the expected high volume.
- Cost if using humans: $200 to $1K
- Steps: Either parse automatically with LLMs and highlight top nuggets, or hire a contact who is already familiar with how to extract value from 4chan and commission him to highlight developments.

Update week #21/2025: No update, Misha's friend yet to respond.