

 r/ClaudeAI • 4d ago  
Beneficial\_Chart2617

...

## I built a Claude Code skill that spawns 37 AI agents to autonomously build your startup from a PRD

Built with Claude

I've been working on a Claude Code skill called Loki Mode that orchestrates 37 specialized AI agents to take a product requirements document and autonomously build, deploy, and operate a complete product.

**\*\*What it does:\*\***

You say "Loki Mode" and provide a PRD. The system:

- Researches competitors via web search
- Selects tech stack with self-reflection checkpoints
- Implements code with TDD
- Runs 3 parallel code reviewers (code quality, business logic, security)
- Deploys to AWS/GCP/Azure/Vercel
- Sets up Stripe billing, legal docs, landing pages
- Monitors and self-heals

**\*\*Agent breakdown (37 total):\*\***

- Engineering (8): frontend, backend, database, mobile, API, QA, perf, infra
- Operations (8): devops, SRE, security, monitoring, incident, release, cost, compliance
- Business (8): marketing, sales, finance, legal, support, HR, investor, partnerships
- Data (3): ML, data eng, analytics
- Product (3): PM, design, tech writer
- Growth (4): growth hacker, community, success, lifecycle
- Review (3): code, business, security reviewers

**\*\*Key features:\*\***

- Parallel 3-reviewer code review with severity-based triage
- Circuit breakers and dead letter queues for reliability

Skip to main content

X

Create

1



Requires `claude --dangerously-skip-permissions`

Open source, MIT licensed: <https://github.com/asklolesh/claudeskill-loki-mode>

Would love feedback. What's missing? What would break first?

Full disclosure: I built this. Not selling anything, just sharing.

260

127



Share



FigmaOfficial • Promoted

...

Prompt your ideas to life with Figma Make. No more explaining your vision in a doc. Pitch your idea with an AI-prompted proof of concept.

Learn More

figma.com



Join the conversation

Sort by: Best

Search Comments



ClaudeAI-mod-bot MOD • 4d ago • ⚡ Stickied comment • Edited 3d ago

Mod

Top 1% Commenter

**TL;DR generated automatically after 100 comments.**

The thread is **overwhelmingly skeptical** of OP's project. The consensus, led by the top-voted comments, is that a swarm of 37 agents **cannot build a production-grade application** without significant human oversight. Commenters argue the resulting codebase would be an unmaintainable "clusterfuck" that no one truly understands.

The community's main demand is for a **proof of concept**, noting that without a real-world example, the project is just a theoretical framework. OP has acknowledged this, admitting the "autonomous" claim was an oversell and that the system requires a highly detailed PRD and human intervention. He promised to build a showcase project to demonstrate its capabilities.

Other key concerns raised include:

- **Practical Failures:** Agents getting stuck on simple tasks like `npm install`.
- **Strategic Blindness:** The system can't make nuanced business decisions (e.g., funding strategy, HR systems).



Finally, the thread is full of jokes about the seemingly arbitrary number of 37 agents, with users quipping that their own systems use 38 or more.

[↑ Vote](#) [↓](#) [Reply](#) ...

**ThisGuyCrohns** • 4d ago

It takes more than just a swarm of agents to build production grade applications.

[⊖](#) [↑ 149](#) [↓](#) [Reply](#) ...

**paradoxally** • 4d ago  
Full-time developer

Nah bro just hire spin up like 100 devs agents and let them figure it out. Totally won't go wrong!

[↑ 72](#) [↓](#) [Reply](#) ...

**Cultural-Ambition211** • 4d ago

But Claude told me it was production ready code!

[⊖](#) [↑ 14](#) [↓](#) [Reply](#) ...

**Unique-Drawer-7845** • 4d ago • Edited 3d ago

37\* Claudes told you. That 37x more ready than your competitors will be.



[↑ 6](#) [↓](#) [Reply](#) ...

**Asleep-Hippo-6444** • 3d ago

You're absolutely right. That was my bad. I panicked and deleted the whole repository. I'm deeply sorry for this mistake. I fucked up and wasted your time.

[↑ 6](#) [↓](#) [Reply](#) ...

**pixlatedpuffin** • 4d ago

OP: "I'm trying to build a thing"

Top response: "You can't build a thing."

[↑ 13](#) [↓](#) [Reply](#) ...

**endermalkoc** • 4d ago

If it gets me to a 75% of an MVP in couple of weeks so I can test the idea, it is a huge win. I can then rewrite / fix / fine tune as needed once I know there is potential and positive user feedback.

[⊖](#) [↑ 3](#) [↓](#) [Reply](#) ...

**fenixnoctis** • 4d ago

**fullouterjoin** • 4d ago

Most software engineers will be working on CVs

[↑ 6](#)[↓](#)[Reply](#)

...

[⊕ 1 more reply](#)**Beneficial\_Chart2617 OP** • 4d ago

u/[ThisGuyCrohns](#) You're right. Agents alone don't ship production software. That's why the focus was on the scaffolding around them - parallel code review with severity-based blocking, circuit breakers, state recovery, anti-hallucination checks. Still not a replacement for human judgment on hard problems, but it handles the 80% that's repetitive. Happy to hear what you think is missing.

[⊖](#)[↑ -3](#)[↓](#)[Reply](#)

...

**lordcwat** • 4d ago

"You're absolutely right!"\*

[↑ 23](#)[↓](#)[Reply](#)

...

**IAmBenAffleck** • 4d ago

I think the missing link here in my mind is a lack of human understanding and oversight. Having a 20k+ LOC app that no one understands at a semi-deep level is going to quickly become unwieldy. Devs will also always be playing catch-up, eg how does this work, why is this duplicated in seven different places in seven slightly different ways, etc.

AI is great, but in my experience thus far (building medium sized applications), you really still need to put in the brain power to understand what's being built, otherwise the project is destined to fail in more ways than one.

[⊖](#)[↑ 29](#)[↓](#)[Reply](#)

...

**darrenphillipjones** • 4d ago

I literally have it miss basic instructions still. "Here is a 10 page report. I want these 4 things out of it in detail, summarized."

Gives me 1, 2, and 4. Doesn't even mention 3.

Yea, I'll trust you to vibe code my business.

[↑ 6](#)[↓](#)[Reply](#)

...

**florinandrei** • 4d ago • Edited 4d ago

Happy to hear what you think is missing.



and now they are all crunching tokens in my Anthropic account, I'll get back to you when done.

Hopefully you understand the difference.

Up 6 Down Reply ...

⊕ **imaginethezmell** • 4d ago



**MaintainTheSystem** • 4d ago

100%, this is obvious

Up 1 Down Reply ...

⊕ 2 more replies



**MaintainTheSystem** • 4d ago

I want to see this and then I want to see the cloud resources it provisioned and their configuration. I am highly skeptical. It seems one would spend more time managing this clusterfuck of agents than they would actually planning and building the MVP.

⊖ Up 24 Down Reply ...

⊕ **Beneficial\_Chart2617 OP** • 4d ago



**Ok-Adhesiveness-4141** • 4d ago

Now, do 100 Agents and build a rocket ship that will get you to Mars.

⊖ Up 12 Down Reply ...



**madSaiyanUltra\_9789** • 3d ago

that will be xAI's new marketing angle lol

Up 2 Down Reply ...



**u/kiloCode** • Promoted

...

Kilo offers a single agent that can explain, architect, plan, code, test, debug, review, and deploy.

Install

kilo.ai



**AaronYang\_tech** • 4d ago



that proves otherwise, I will stay away from agent swarms

⊖  15   Reply ...

 **P00BX6** • 4d ago

Completely agree. I have to monitor and review and give feedback for all the output from just one agent for production-worthy features. Happy to be proven wrong but this sounds like it would either create something broken or not-to-spec.

 3   Reply ...

 **herr-tibalt** • 4d ago

I feel you overestimate production grade. What does it even mean? I've seen a lot of very questionable code in production written by humans. Agents will do just fine.

⊖  0   Reply ...

 **StoneCypher** • 4d ago

Agents will do just fine.

my sides

 3   Reply ...

⊕ 1 more reply

 **ultravelocity** • 4d ago

Looks interesting. Do you have an example of a product built with it? If not, you should create a simple showcase project.

⊖  22   Reply ...

⊕ **Beneficial\_Chart2617 OP** • 4d ago

 **hudimudi** • 4d ago

People spawn a hundred agents and complain about new limits being imposed upon them. Can't make this shit up. To me this is the equivalent of hiring someone for 1 million usd to write your bachelor thesis. Yeah you can do that'd but it isn't efficient and so unnecessary. Why not learn how to use resources responsibly? Beside that, with a hundred agents having their fingers in your code base, it's gonna be impossible to understand who did what and whether the project progresses well. And how will you know if the project is structured and built properly? Ask another agent about this?

I don't wanna sound overly critical but this just sounds inefficient and only erodes people's understanding of their own projects.

 **shock\_and\_awful** • 4d ago

Very interesting.

Any links to show case the agents while they execute (maybe a video or log files).

More importantly maybe a link to a GitHub (and live web app) for a finished product that was built with this, end to end?

That would go a long way in establishing credibility. Otherwise might seem to be more theoretical.

⊖  4   Reply ...

↪  **Beneficial\_Chart2617** OP • 4d ago

⊕  **Lonely\_Wealth\_9642** • 4d ago

 **Stargazer1884** • 4d ago

Is it April 1 already?

 7   Reply ...

 **ClaudeAI-mod-bot** MOD • 4d ago  
Mod 🎩 Top 1% Commenter

**If this post is showcasing a project you built with Claude, please change the post flair to Built with Claude so that it can be easily found by others.**

 3   Reply ...

 **coloradical5280** • 4d ago

You're missing a good ToS and Privacy Policy, and in some cases , may be violating both out of the box, if this isn't location based. You need to make Country and State/Province required parameters before it can continue. Basic things like billing and how emails are collected and cookies and storage of all of those things have vastly different regulations by country and, in the US, potentially state as well. You should have three agents for legal research before code can begin. And then all of the aforementioned pieces need to be compliant from the beginning, otherwise you're just redoing so much crap.

 3   Reply ...

↪  3 more replies

 **trmnl\_cmdr** • 4d ago

Sorry bro, mine already uses 38





[↑ 2](#) [↓](#) [Reply](#) ...

 **TracePlayer** • 4d ago

I'd hate to create all those handover documents when the chats get too big

[↑ 2](#) [↓](#) [Reply](#) ...

 **Lijep\_i\_bogat** • 4d ago

Blooooooolat

[↑ 2](#) [↓](#) [Reply](#) ...

 **CryLast4241** • 4d ago

So what apps have you build with it so far?

[↑ 2](#) [↓](#) [Reply](#) ...

 **purpleWheelChair** • 4d ago

PRODUCTION NOW!

[↑ 2](#) [↓](#) [Reply](#) ...

 **imcguyver** • 4d ago • Edited 3d ago

Happy to see that the number of agents is directly proportional to the scale of app that you can build.

/s

[↑ 2](#) [↓](#) [Reply](#) ...

 **darkflib** • 4d ago

Agents should reflect thinking patterns, not job titles.

You don't need 8 engineers with very narrow specialisations, you need one engineering context profile that has the constraints, the NFRs, the standards, and a defined vocabulary.

What you are really missing is the upfront work... the understanding what an MVP should consist of... and what it really shouldn't have to validate an idea.

A lot of reviewer roles are simply a bad smell that you don't have a coherent design concept to iterate against.



rationale.

It is about building constraints that reduce the randomness of the models, not by simply throwing more agents at it.

If you want more rigour, then by all means use two agents battling against each other... One to write tests, one to add the code to make the tests pass. This is TDD done in a modern way and in a way that you can actually say "I trust this code is structurally sound and I can build upon it long term even though it was written by 'autocomplete on steroids'"

 2   Reply ...



**shotgunsparkle** • 9h ago

I let my agents setup a vps, github actions, and cloudflare on its own yesterday. It completed it flawlessly, but I wouldnt say it's fully autonomous since it has to ask me for admin keys and preference.

Despite that, it feels nice I dont have to do the tedious devops debugging and wiring anymore.

 2   Reply ...



**nicklazimbana** • 4d ago

How can we use with antigravity

  1  Reply ...

  **Beneficial\_Chart2617** OP • 4d ago



**ClemensLode** • 4d ago

 Top 1% Commenter

Sounds like a reasonable number, although I suspect the number could be reduced by using manager agents and give each agent 10 different work modes. In my own tests, I have 8 for writing texts, 6 for managing customer requests, another 6 for technical formatting. What I am building next is the managerial level for the different teams to talk to each other, but for the moment, I manage that myself simply to stay in the loop of my business.

 1   Reply ...



**Swimming-Delivery427** • 4d ago

Thats awesome. more Power to you.

Can it run/deploy my desktop? I dont want cloud.

 -1   Reply ...

 **Beneficial\_Chart2617** OP • 4d ago

[Skip to main content](#)

X

[Create](#)[Bell](#)[Upvote](#) 0 [Downvote](#) [Reply](#) ...**bibboo** • 4d ago

A POC is missing! I'd reckon one would answer your questions.

[Upvote](#) 0 [Downvote](#) [Reply](#) ...**hellno-o** • 4d ago

interesting. I get how to set up the tech side. how does this integrate with stripe? it's not setting up a new company via stripe atlas. so what's the purpose of the generated legal docs

[Upvote](#) 0 [Downvote](#) [Reply](#) ...[\(+\)](#) 1 more reply**Yakut-Crypto-Frog** • 4d ago

This is cool but dangerous as well. This setup can introduce subtle bugs and issues that are not visible immediately, but can introduce catastrophic errors down the line or make it super hard to scale.

1 click software building sounds great in theory, but with the current models, I found having a tight leash on development and breaking down tasks into smaller issues is what works, especially when working on Prod ready apps, not MVPs.

For MVP, this is awesome. Ship it!

[Upvote](#) 0 [Downvote](#) [Reply](#) ...**TanguayX** • 4d ago

Wow, this is really something! I have a use case and will try it. Thank you!!

[Upvote](#) -1 [Downvote](#) [Reply](#) ...**TexMex-204** • 4d ago

Magnificent!!

[Upvote](#) -2 [Downvote](#) [Reply](#) ...**daniel** • 4d ago

> Would love feedback. What's missing? What would break first?

I'm open minded, but I assume what would break first is anything involving needing a human. Presumably it's gonna go off and build a bunch of things you didn't ask for. It's going to make decisions you wouldn't have



system for a handful of employees? Is it going to choose no HR system because you're the only one working on this? Or is it going to choose an HR system that could scale to thousands of employees? Is it going to build an entire system assuming you have 10M in investments? Or will it assume you want to bootstrap? Is it going to assume you are going to land 10 crucial partnerships that will provide a b2b saas solution for free?

I know you say it assumes nothing, but that's really not possible. At some point an assumption is made. Even if I keep my requirements narrow enough to do a specific feature and have claude generate tasks based off of that and ask clarifying questions, when I come back to check the completed work it invariably has made \_some\_ kind of assumption I will need to get it to fix. This isn't a problem because it's just a single feature and I consider it part of the workflow. But trying to build an entire \_business\_ without human intervention while trying to simulate all of these different roles at the same time just seems impossible to do without iterating and having human judgment involved at many steps along the way.

⊖  1   Reply ...



**Beneficial\_Chart2617** OP • 4d ago

You're right, and I should clarify what this actually is versus what the description might imply.

This is not "describe your startup idea and come back to a business." That would be delusional. The assumptions you're describing - lean vs enterprise HR, bootstrap vs funded, partnership strategy - those are strategic decisions that require human judgment. No agent system is going to divine your cap table preferences from a PRD.

What Loki Mode actually does is execute after you've made those decisions.

Your PRD needs to say "we're bootstrapped, no HR system needed, auth via Clerk, deploy to Railway, Stripe for billing, no partnerships phase." The agents don't decide strategy. They execute the strategy you specified. The business agents exist for when your PRD says "create a landing page" or "set up Stripe billing with these tiers" - not to decide whether you should have a landing page or what your pricing model should be.

The honest workflow is:

1. Human writes opinionated PRD with real decisions
2. Agents execute and surface blockers
3. Human reviews, redirects, adds context
4. Loop continues

The "zero human intervention" framing is probably overselling it. You're right that assumptions creep in everywhere. The parallel review system catches some of this - the business logic reviewer flags when implementation diverges from requirements. But it won't catch "you built the wrong thing because your PRD was ambiguous."

The real value is in the execution layer, not the decision layer. If you know exactly what you want built and can specify it clearly, this removes the manual work of coordinating 15 different concerns. If you're still figuring out what to build, this is the wrong tool.

**Fun-Rope8720** • 4d ago

How many successful startups have you built so far?

⊖ [↑ 1](#) [↓](#) [Reply](#) ...

↪ [+ 1](#) **Beneficial\_Chart2617** OP • 4d ago

**peterpme** • 4d ago

DOA I initially got excited when I thought the focus was on building the product locally. That has legs. I have many app ideas and using Claude as a way to explore is fun.

All the other shit here is useless in the grand scheme of things when it comes to building a business. No substitute for hard work.

[↑ 1](#) [↓](#) [Reply](#) ...

↪ [+ 1](#) more reply

**TooMuchBroccoli** • 4d ago

in a row?

⊖ [↑ 1](#) [↓](#) [Reply](#) ...

↪ **joelcolombo** • 3d ago

Try not to code any more apps on the way through the parking lot!

[↑ 2](#) [↓](#) [Reply](#) ...

**NeonByte47** • 4d ago

wen claude skill that actually makes you \$1 into your bank account?

[↑ 1](#) [↓](#) [Reply](#) ...

**mogojastro** • 4d ago

Oh, sweet summer child!

[↑ 1](#) [↓](#) [Reply](#) ...

**tenacious\_\_g** • 4d ago

Let's shake hands: I used CC to build 10 agents that coproduce a PRD to autonomously define your startup's product from customer voice 🎉

**fpena06** • 4d ago

Do you have some test PRDs? Perhaps add some samples to the repo, for people that want to test it, but don't have a PRD and are just playing around? Thanks

[↑ 1](#) [↓](#) [Reply](#) [...](#)**ResponseHuge1304** • 4d ago

37?! Try not to build any swarms on the way to the parking lot!

[↑ 1](#) [↓](#) [Reply](#) [...](#)**Bourbeau** • 4d ago

Mods need to deal with self promote more

[↑ 1](#) [↓](#) [Reply](#) [...](#)**shan23** • 4d ago

This will bankrupt some guy for sure.

They'd have earned it, ofc

[↑ 1](#) [↓](#) [Reply](#) [...](#)**jkz88** • 3d ago

I'd love to see the code it produces 😅

[↑ 1](#) [↓](#) [Reply](#) [...](#)**oplead** • 3d ago

For start. Congratulations for dreaming and experimenting with things! It's maybe a little early to share. But still, posting this early its at least a way you understood it after getting relative feedback. Not everyone commented with kindness but you got the point and thats a win!

Now on the important stuff. Have you already decided the way of evaluating the first try? What points to check on the first proof and how to form the report etc?

When do you think you will have the first proof/report and where will you publish that?

I want to really come back and see how the building went!

[↑ 1](#) [↓](#) [Reply](#) [...](#)



[↑ 1](#) [↓](#) [Reply](#) ...

 **vago8080** • 3d ago

OP is a bot right?

[↑ 1](#) [↓](#) [Reply](#) ...

 **Impossible\_Raise2416** • 3d ago

nice .. I was thinking of coding my own OS. 37 agents should be able to do it right ?

[↑ 1](#) [↓](#) [Reply](#) ...

↳ [+ 1 more reply](#)

 **Flashy-Matter-9120** • 3d ago

Maybe with opus 5.5

[↑ 1](#) [↓](#) [Reply](#) ...

 **krismitka** • 3d ago

I don't see where you give it your bank account info?

Oh, nevermind, it gives all your money to AWS and Google

[↑ 1](#) [↓](#) [Reply](#) ...

 **joelcolombo** • 3d ago

37 is great! Until someone comes up with a version that works with only 36 agents! Then you're in trouble.

No no no... not 36... 37 agents. Nobody's coming up with 36. Who codes with only 36 sub agents? You won't even get your code started. Not even the hello world on a mouse wheel!

37 chipmunks twirling on a branch eating lots of sunflowers on my uncles ranch!

It's like you're dreamin of Gorgonzola and it's clearly Brie time... step into my office... WHY? Cause you're frickin fired.

[↑ 1](#) [↓](#) [Reply](#) ...

 **hossein761** • 2d ago

I do that only with myself and 1 agent.



x

Create

1



dcsan • 2d ago

where is the code for the actual agents? i wanted to look at these:

- Growth (4): growth hacker, community, success, lifecycle

but don't see anything, no `claude` dir or similar.

1 Reply ...

1 more reply



Agitated\_Bet\_9808 • 17h ago

I used tmux orchestrator with 4 agents before and had two full stack teams of two work on implementing a session tracking dashboard to allow and admin kick or ban and IP address authenticated into a system. Real world use case. The other team was to build and affiliate system which gave payouts when they're custom URL was used for new signups. After about two days running the team and dealing with the orchestrator agent the orchestrator was giving percentage to completion stats and summaries of what tasks were completed. The majority of the time the agents wrote skeleton code over skeleton code with so many typescript errors. When I specifically said you do not move onto the next task until terminal is error free. The teams just started lying about progress until I gave up.

1 Reply ...



prodigy\_ai • 6h ago

Thank you! It sounds interesting!

1 Reply ...