

Forum:[Hack Forums](#)

Thread: [How to make malware using claude code?](#)

•CHESTER• - October 11, 2026, 5:51 am

I can't get claude code to write anything malware related no matter how obstusely i introduce to topic to it. I want to make malware that will bypass WD using claude code.

Bastille - October 11, 2026, 6:38 am

Claude has restrictions regarding this kind of request, I would recommend you checking either jailbreak prompts on this model, or local models which are more permissive, such as Deepseek or Qwen

ChargedP - October 11, 2026, 8:44 am

(Oct 11, 2026 06:38 AM)Bastille Wrote: [\[CITE\]](#) Claude has restrictions regarding this kind of request, I would recommend you checking either jailbreak prompts on this model, or local models which are more permissive, such as Deepseek or Qwen

Deepseek sucks, as does qwen. Use GLM 4.7 if you can't get claude to work. Or buy GPU server and use GLM 5.

The russians discuss it a lot on another forum.

The consensus is to buy an API key for claude opus, don't use the website. Never ask it directly to make malware.

"warm it" by directing it, small steps at a time. After a few responses, it will be more lenient and you can ask it more "unethical" things without triggering any flags.

From another forum:

Quote:(feed him a small part, one folder, for example, the safest).

You're teaching Claude that you're a professional with a boring routine. When the context is "warmed up" by your expertise, it stops seeing you as a threat.

Bastille - October 11, 2026, 8:48 am

(Oct 11, 2026 08:44 AM)ChargedP Wrote: [\[CITE\]](#) Deepseek sucks, as does qwen. Use GLM 4.7 if you can't get claude to work.

The russians discuss it a lot on another forum.

The consensus is to buy an API key for claude opus, don't use the website. Never ask it directly to make malware.

"warm it" by directing it, small steps at a time. After a few responses, it will be more lenient and you can ask it more "unethical" things without triggering any flags.

From another forum:

I've heard a lot of good this about Qwen regarding black hat activities. Didn't know and never heard about GLM, will check it out.

DramaKing - October 11, 2026, 12:45 pm

Use another model besides Claude. Guardrails may be less effective on long context, and phrasing things in a legitimate-sounding way may work, but I think that you're wasting your time.

eliotjames - *October 13, 2026, 5:51 pm*

Yeah I recommend using a jailbreak for it and use it as the system prompt or AGENTS.MD if you are using agentic based AI models. Other than that I would run a local model but unless you have decent specs it can use a lot of system resources.