Dear DeepSeek engineer

critisticuffs

On 13 September 2026, an engineer at DeepSeek published a post on WeChat titled "I had to bury my talent in yesterday".[1] This engineer wrote the main attention kernel[2] of the company's DeepSeek v4.1 model. Their post is about what is happening to this kind of work. They write that within a year, AI went from looking up documentation for them to optimising kernels on its own, and in six months or a year it will probably do their work as well as they do. They keep optimising anyway, partly for the pleasure of it and partly because, if they stopped, other companies would carry on regardless.

They say they do not expect to be "unemployed". In their words, they expect to have to "change careers": to stay employed, but to stop writing kernels themselves and instead become the "mech pilot" of an AI agent that writes them. They compare themselves to a skilled knitter whose quiet afternoons by the window are ended by a knitting machine, and they grieve openly for the work they loved. The post closes with worries about students handing their assignments to AI and about whether power will soon count for more than skill, and with the hope that open, cheap AI, like DeepSeek's, can keep the world from turning into "Cyberpunk 2077". What follows is our reply, quoting the post section by section.

Dear DeepSeek engineer,

1 "There are more gears in my hands, and fewer beats in my heart"

> To fit what industry needs, I'll have to give up the direction I loved for an unknown new one. […] But having what you love taken away really doesn't feel good. The simple joy of sitting quietly at my desk writing kernels for an afternoon may end this summer. I have to bury my talent in yesterday and become a mech pilot. There are more gears in my hands, and fewer beats in my heart.

We read your post, and we want to begin by saying that we do not think your sentimentality is a flaw. An afternoon spent quietly working on a focused task, where you can apply yourself to something you are very good at, can be a real pleasure. For you that meant an afternoon on one kernel,[3] and losing

[1] Original post (13 September 2026); archived copy. An English translation of the post is here. None of us reads Chinese. The quotations are our English versions, put together with the help of several AI translation models and checked against translations of the post already circulating online.

[2] Here, a kernel is a small program that runs on a graphics chip (GPU) and does one basic calculation an AI model needs, over and over, very fast. "Attention" is the calculation by which a model weighs every part of a text against every other part. How well a kernel is written decides how fast, and so how cheaply, a model can be trained and run.

[3] Until recently, kernels were written by hand. You write the kernel in CUDA, a programming language for NVIDIA GPUs, run it and time it. If it is slower than the chip should allow, you might run it under a "profiler", a tool that shows the instructions the chip actually executes and where they sit waiting, for instance, for data to arrive from memory. That waiting is a "stall". Then you work out a way to rearrange the job, say by fetching data earlier or keeping more of the chip busy at once, rewrite the code, and time it again. An afternoon is many rounds of this, most shaving off a few per cent. Now and then you find a trick nobody has tried before. The result is a single number: how fast the kernel runs, often measured against the kernels the GPU manufacturer supplies for the same job.

it is a real loss.[4]

We also think you are right about a lot. You are right that your job may survive, but the work you loved may not. And you are right that it would change nothing if you, on your own, held back: "other companies' models would keep developing and finish me off anyway".

We disagree with you about why you are losing this work. You see a race, but you treat it as a fact of nature, like "the wheels of history", and the machine as what drives it. We want to make two points against this. First, it was incidental that what you loved about the work and what DeepSeek paid for coincided. DeepSeek paid for faster, cheaper models, and your pleasure came along with them. Second, what is happening now is not simply that a machine does your job. Until now your employer owned the kernels, but how they got written was up to you. That control is now passing to your employer, in the form of a machine it owns. Both points are about how your work is organised. Neither is about the technology. The loss is made, not a given.

2 The knitter by the window

> An image may help. You are an expert knitter, especially good at patterns and colour combinations. Your sweaters are sturdy and beautifully patterned, rich people from all the villages around come to you for them, and you have made a good deal of money. And you love sitting by the window with a pot of tea, looking out at the green hills and water, the cattle and the chimney smoke, quietly knitting through an afternoon. But one day someone invents a marvellous machine. Give it yarn and a pattern and it knits the sweater by itself, no worse than yours in quality or texture, and far faster. You know your competitors can use it to reach your old level easily, so you have to use it too. You also know that with the skill you have built up over twenty years, even when everyone has a machine, you can still beat your competitors on speed and quality. But the pleasure of sitting by the window listening to the rain, drawing needle and thread, letting time pass slowly, is crushed in the end by the roar of the machine.

A machine that knits in minutes what took an afternoon should leave *more* afternoons free. A village with such a machine has more than before: more sweaters, more free time, or both. Nothing about the machine forbids anyone from sitting by the window with needles and a pot of tea. So why does "the roar of the machine" cost the knitter their afternoons instead of giving them more?

Your own story answers this. The knitter does not knit sweaters to wear them. They knit for "rich people from all the villages around", and for the knitter the sweaters are a means: what they want from them is the money they fetch, which is what the knitter lives on. And they knit in competition with other knitters.

[4] Many people find no pleasure in their work at all, and nothing is wrong with them for that. Most work done for a wage was never arranged to be enjoyed, and finding it a burden is a normal reaction. None of our argument depends on you loving your work.

Once the machine is in use, a sweater sells for what it takes to make one with the machine. Suppose a sweater takes four hours by hand and twenty minutes by machine. While only a few knitters have the machine, they can sell below the old price and still earn more per hour than before. As more knitters buy it, they undercut each other until the price covers little more than the twenty minutes. Nobody decides this alone; it follows from everyone selling against everyone else.

The afternoon by the window is still an afternoon of the knitter's life, but the market now counts it as twenty minutes' work and pays for twenty minutes. The knitter loses the afternoon not because of the machine but because their time is their income, and the time a sweater is worth is now set by the machine. If the knitter knitted only for themselves and their friends, the same machine would simply give them the choice to use their time differently.

There is something else your image leaves out. In the story, the knitter owns the machine. They decide to use it, they decide how, and the extra sweaters are theirs to sell. Your situation is not like that. In your case the machine is the agent, and you do not own it. You never owned the kernels either. Your employer, DeepSeek, does. You do not sell kernels; you sell your ability to work for so many days, and what you do in them is decided by whoever buys it. When your employer gets a machine, it decides what the machine is used for. That is why you expect to "change careers" rather than lose your job: you stay, but you will do whatever work the machine leaves.

3 Making a living

> My judgement is that I won't be "unemployed", but I will have to "change careers". I can probably still make a living, but I may never again get to do the work I once loved. […] Before, my interests, what I was good at and what industry needed were basically aligned. Now AI has made what I'm good at into something it is better at.

What did the industry want from you? Not your enjoyment, and not even your kernels as such, but a result: faster training and inference, more model per yuan of compute. Until now, getting that result usually required people who could work their way through hard technical puzzles, and people who can do that usually find some satisfaction in it. So the pleasure came with the work, but your employer was paying for the speed, not for the pleasure.

This is now ending. Once the firm can get the speed some other way, it no longer needs the work you enjoyed. You put this precisely yourself: you can "probably still make a living", but you have to "change careers". The career is the work: writing kernels. The living comes from the job: DeepSeek employs you and pays you a wage. You can keep the job and lose the work because DeepSeek does not pay you to write kernels. It pays you to make its training and inference cheaper, and until now writing kernels was the way to do that.

4 “Mech pilot”

> So what does changing careers mean? It means giving up kernel design, writing and optimisation, the field I’ve worked in so deeply and loved, to become an agent’s “mech pilot”. […] [I]ndustry’s demand has drifted from “people who can write high-performance kernels” to “people who can use AI to produce high-performance kernels faster”.

Consider your own description of what the AI can do: it “reads CUDA, PTX and SASS on its own, uses professional tools to analyse the stall time of every instruction, and optimises kernels by itself”. That is your job description. The knowledge that used to exist only in people like you now exists in the agent, which belongs to the firm, and the agent learned it from the work of people like you, your own kernels included.[5]

What is left for you is to set the agent going, read what it produces, catch its mistakes and approve its output. You no longer set the pace; the agent does. Once an agent can produce a kernel in an hour, DeepSeek will not pay for a day, any more than the buyers pay for the knitter’s afternoon.

Marx had a name for this change: the *real subsumption* of labour under capital.[6] It is easiest to see in the trades where it happened first. When capital takes over a kind of work, it at first takes the work process as it finds it. A weaver who used to weave on their own account now works in an employer’s workshop, on the same kind of loom and with the same skill. The employer pays them a wage for their ability to work, so many hours a day, and owns the yarn, the looms and the cloth. Marx called this *formal* subsumption: the employer owns the product and directs the working day, but how cloth gets woven still depends on what the weaver knows.

Real subsumption begins when the employer rebuilds the work process to make it more profitable by cutting its costs. First, the craft is split up. Marx’s example is watchmaking: instead of one watchmaker making a whole watch, each worker makes one part over and over, and only the workshop as a whole, which the employer organises, can make a watch. Then machines take over the parts, wherever a machine costs less than the wages it saves, and the worker fills the gaps the machine leaves. In Marx’s words: “In handicrafts and [workshops], the worker makes use of a tool; in the factory, the machine makes use of him.”[7] What the worker once knew no longer gives them any hold over the process. It

[5] Why kernels? It is often said that machines come for unskilled work and leave skilled work alone, but since the first factories machines have also been used to replace expensive specialists. What makes kernels open to it now is technical. Your post compares you to a speedrunner breaking their own record, and a kernel is judged the same way: by one number, how fast it runs on a benchmark compared with the vendor’s. A machine can test its own attempts against that number and keep whatever is faster, and work with a clear measure like this is currently among the easier kinds to train a machine to do.

[6] The terms are Marx’s. In *Capital*, vol. 1, ch. 16, he writes that a specifically capitalist mode of production “arises and develops spontaneously on the basis of the formal subsumption of labour under capital. This formal subsumption is then replaced by a real subsumption” (Penguin ed., p. 645).

[7] *Capital*, vol. 1: watchmaking, ch. 14, §3; machines used only where they cost less than the wages they save, ch. 15, §2; the quotation, ch. 15, §4, p. 548. Where we write “[workshops]”, Marx writes “manufacture”, by which he means workshop

faces them as the employer's equipment.

Your work had been subsumed only formally: DeepSeek paid you a wage for your ability to work and owned what you made, but how a kernel got fast was still up to you, because that knowledge existed only in people like you. That was the part you loved.

None of this is new to your industry. Your own post mentions the old skill of "writing x86 assembly fluently". Higher-level languages and frameworks each built a piece of specialist knowledge into a tool, and for the firms that used them each was also a way to need fewer specialists, or cheaper ones.[8] Each time, the craft moved to the part that could not yet be automated, and kernels for new chips were one of the last such parts. Now that part has been reached too. You keep the job, but you no longer decide how the work is done. The firm decides, and the machine it owns enforces that decision in every working day.

The pressure does not come from the agent itself. Your employer uses it to need fewer of you, and to get more out of those that remain. This is not something peculiar to DeepSeek. Every firm in this race does the same, and does it on purpose: its interest in cheap kernels is opposed to your interest in your working day, and, as the next section explains, the race rewards the firm that pursues that interest hardest.

5 "Humans have never hesitated when it comes to destroying themselves"

> Humans have never hesitated when it comes to destroying themselves. Why, knowing full well that "the better I write kernels, the faster our new models train and run, the faster their capabilities grow, and the sooner I'll be replaced", do I still optimise kernels as hard as I can? Partly because writing kernels is like playing a game for me, and gives me enormous pleasure. When I invent a new technique or see a kernel's performance climb, I'm as thrilled as a speedrunner breaking their own record. And when I see my kernels far outperform the vendor's official ones, I feel great pride. But a more important reason is that even if I slacked off, or deliberately sabotaged model training, other companies' models would keep developing and finish me off anyway.

You call this humans destroying themselves. But it is not humans in general who are racing here. It is firms, each trying to get ahead of the others with its models. Whatever a firm competes on, it uses the agent if a kernel made by the agent and a pilot costs it less than one you write by hand.

production based on the division of labour, before machinery. Marx's name for what the worker sells is labour-power, the ability to work, sold for a fixed period (ch. 6).

[8] On how firms arrange production so that workers' special skills matter less and less, branch by branch, see "Does capital always aim to suppress wages?".

DeepSeek is under pressure too, but a different one: a firm that kept paying for hours its competitors have stopped paying for would get less out of the same money than they do, and fall behind. DeepSeek puts the same pressure on its competitors, and meets it at your expense, by paying for fewer of your hours.[9]

That is also why the better you write kernels, the sooner you are replaced. Your kernels belong to the firm, and it uses them to train and run its models faster, the ones that write kernels included.

6 Communism or 2077?

> In the society of the future, will power matter more than technical skill or IQ? […] As AI develops, future society may tend toward one of two extremes: communism or Cyberpunk 2077. In the first, the productive forces are greatly liberated and living standards rise markedly […]. In the second, a few tech companies control most resources; only a tiny minority can use the most advanced AI and technology […], while most people can only use very weak AI. Climbing out of your class becomes ever harder: you need the strongest AI first, which makes it a vicious circle. […] So I still believe frontier intelligence should be supplied to everyone, openly and cheaply. I don't trust Anthropic or OpenAI to do that. Above all, I don't want Anthropic to hold the most advanced AI or AGI. […] It is also why I chose DeepSeek and have stayed: we research powerful, fast AI that benefits everyone, and we open-source it. Perhaps that can pull the world back a little from the 2077 end.

You ask whether power will matter more than technical skill. But your case shows that it already does. In your 2077, class is a ladder, a ranking by income and access, and only those with the strongest AI can climb it. But as you acknowledge ("even harder"), class already exists. Class divides those who own the machines from those who have to sell them their ability to work. Those who own the machines decide what talent is used for, and your kernels, however good, are written for them.

Your answer is to make the best models open. But open-sourcing is itself a move in the race. For a firm competing with richer rivals, giving its models away is a way to win users and developers, to make its model the one others build on, and to undercut competitors who charge for theirs. Governments back such firms for the same reason when they want their country's industry to keep up with another's.[10] Cheap access for everyone can be a real result of this. But it is not the aim. It is a means of competing, and it lasts as long as it serves that purpose.

[9] On why a wage is a cost to be cut, not a share in output, and why a society that needed less work from its members would otherwise count as rich, see "Jobs, Growth, Justice – an alternative that isn't".

[10] Firms say as much themselves. Explaining in 2024 why Meta gives its Llama models away, Mark Zuckerberg wrote that "selling access to AI models isn't our business model", so "openly releasing Llama doesn't undercut our revenue", and that "the path for Llama to become the industry standard is by being consistently competitive, efficient, and open generation after generation". He also argued that America's leading companies should "work closely with our government and allies" to "achieve a sustainable first-mover advantage over the long term". ("Open Source AI is the Path Forward", 23 July 2024.)

Now take your solution at its word. Suppose DeepSeek succeeds: the best model is open and cheap for everyone. "Everyone" includes every firm that employs kernel engineers. Each of them now has a cheap kernel agent, and its engineers go the way you are going. Your own firm would have handed it to them. Whether DeepSeek or Anthropic holds the best model decides which firm comes out ahead. It does not change why a firm uses the agent instead of you: it costs less.

7 "May all the beauty be blessed"

> May the world of the future be well. May all the beauty be blessed.

A machine that can do in minutes what took you an afternoon could be good news wherever work is done to satisfy needs: fewer hours of necessary work, and more afternoons for the problems people find beautiful, kernels included. It arrives as a loss for you because your work is done for a firm's profit and under a firm's direction.

Your sadness is real, and we share your wish that the world of the future be well. But as long as work is done for profit, firms have to use machines to cut labour costs, and a machine that saves work will go on reaching the people who do the work as a loss.

Wishing you well,

Critisticuffs

8 Appendix: Translation of the full blog post[11]

A few days ago DeepSeek v4.1 was released, pushing the ceiling of small-model capability up another notch.

AI is developing far faster than anyone expected. From the first ChatGPT, which could only babble through a chat with a context of a few thousand tokens, to reasoning models like OpenAI o1, DeepSeek R1 and Kimi K1.5 Thinking took just two short years. From reasoning models to today's agents, which can run commands fluently in all kinds of harness tools and complete complex tasks, took only a year and a half. It is hard to imagine what AI will become if we wait another one, two or three years: how powerful it will be, whether it will be able to evolve itself, whether it will have spread deep into fields like embodied intelligence.

[11] Original post (13 September 2026); archived copy. The English version was made with the help of several AI models and checked against existing translations; readers of Chinese should treat the original as authoritative.

8.1 AI is getting better and better at writing kernels

AI has progressed just as fast in my own field, kernel design and writing. In a single year it has gone from a little assistant that could look things up in the docs, read code and find bugs for me, to a kernel master that reads CUDA, PTX and SASS on its own, uses professional tools to analyse the stall time of every instruction, and optimises kernels by itself. I believe that soon it will also be able to design kernel schedules on its own, evaluate how different schedules perform, and implement and optimise them.

Of course I'm proud of DeepSeek v4.1's success. After all, I wrote its main attention kernel[12], and its excellence is a vote of confidence in my kernels. But the wheels of history roll on, and nobody can stop technology. I know very well that in six months or a year, kernels written by AI will very likely be as good as mine, or better. AI can think 300 tokens a second, type a command in half a second and write a piece of code in twenty seconds; I can't. AI can keep improving in model depth, thinking intensity, number of tool calls (how often it interacts with its environment), even parallelism; I can't.

Humans have never hesitated when it comes to destroying themselves. Why, knowing full well that "the better I write kernels, the faster our new models train and run, the faster their capabilities grow, and the sooner I'll be replaced", do I still optimise kernels as hard as I can? Partly because writing kernels is like playing a game for me, and gives me enormous pleasure. When I invent a new technique or see a kernel's performance climb, I'm as thrilled as a speedrunner breaking their own record. And when I see my kernels far outperform the vendor's official ones, I feel great pride. But a more important reason is that even if I slacked off, or deliberately sabotaged model training, other companies' models would keep developing and finish me off anyway. "Of course I hope I won't be revolutionised. But if I must be revolutionised, I want the one doing it to be me." When everyone is this set on destroying themselves, I have no choice but to join this cruel arms race.

8.2 What about me?

When the day comes that AI really writes kernels better than I do, what will become of me?

My judgement is that I won't be "unemployed", but I will have to "change careers". I can probably still make a living,[13] but I may never again get to do the work I once loved.

I once made a judgement about how the times would change and where that would leave me. Because things change so fast (AI, above, is a good example), I can't possibly predict what will happen in five or ten years; but whatever happens, I believe that my vision, judgement, initiative and intelligence

[12] "Main attention" covers only the MQA attention with head dim = 512. It does not include the indexer that selects the top-k important tokens; that part was written by other (also very strong) colleagues (and their AI agents).

[13] Translator's note: literally "I can probably keep my rice bowl", a common idiom for having a secure livelihood. The same idiom is used again a few paragraphs later.

will keep me at the table, and put me back at the crest of the wave. But that judgement only guarantees I won't be "unemployed". It doesn't guarantee I won't have to "change careers". If anything, it encourages me to change careers so as not to become unemployed.

So what does changing careers mean? It means giving up kernel design, writing and optimisation, the field I've worked in so deeply and loved, to become an agent's "mech pilot". Before, my interests, what I was good at and what industry needed were basically aligned. Now AI has made what I'm good at into something it is better at, and industry's demand has drifted from "people who can write high-performance kernels" to "people who can use AI to produce high-performance kernels faster". To fit what industry needs, I'll have to give up the direction I loved for an unknown new one. I believe that with my understanding of engineering, of what the models above need and of the hardware below, I can keep producing kernels with high quality and efficiency, and I know I may come to love this new direction (or I may not). But having what you love taken away really doesn't feel good. The simple joy of sitting quietly at my desk writing kernels for an afternoon may end this summer. I have to bury my talent in yesterday and become a mech pilot. There are more gears in my hands, and fewer beats in my heart.

An image may help. You are an expert knitter, especially good at patterns and colour combinations. Your sweaters are sturdy and beautifully patterned, rich people from all the villages around come to you for them, and you have made a good deal of money. And you love sitting by the window with a pot of tea, looking out at the green hills and water, the cattle and the chimney smoke, quietly knitting through an afternoon. But one day someone invents a marvellous machine. Give it yarn and a pattern and it knits the sweater by itself, no worse than yours in quality or texture, and far faster. You know your competitors can use it to reach your old level easily, so you have to use it too. You also know that with the skill you have built up over twenty years, even when everyone has a machine, you can still beat your competitors on speed and quality. But the pleasure of sitting by the window listening to the rain, drawing needle and thread, letting time pass slowly, is crushed in the end by the roar of the machine.

I know it's hard, but there's nothing to be done. I can still make a living, but the old love will most likely have to go. My reason and my feelings are fairly separate: when a problem calls for reason I can be very rational, but sometimes my emotional side shows. I remember that when I moved out of the rented flat I'd lived in for a year, I cried my heart out, unable to part with the memories. Saying goodbye today to the era of hand-written kernels and human-brain optimisation is undoubtedly crueller than that.

I don't know whether any readers feel the same, but I suppose that's just how it has to be.

8.3 What about people?

As AI keeps advancing, a few questions worry me.

Won't today's students mostly prefer to have AI do their assignments, especially the practical labs?

Imagine two choices. You can slog through a lab for eight hours and maybe still not get full marks. Or you can start an AI model and have it write full-marks code in a few minutes, for a few cents. Which will most students choose?

That will leave many students with seriously inadequate engineering ability: organising code, building systems, anticipating future requirements and designing for them in advance, abstracting, and so on. As AI keeps getting stronger, is this "engineering ability" still necessary? Will it be left behind by the times, like the old skill of "writing x86 assembly fluently", or will it always be valuable, like "understanding the whole computer system, from software to systems to hardware"? If the latter, that's dangerous: someone with poor engineering ability, paired with AI, can turn out spaghetti code several times faster than before, planting all kinds of hazards in our systems and making the world even more of a makeshift operation.

In the society of the future, will power matter more than technical skill or IQ?

Perhaps only the times themselves can answer these questions.

8.4 Conclusion

As AI develops, future society may tend toward one of two extremes: communism or Cyberpunk 2077. In the first, the productive forces are greatly liberated and living standards rise markedly (I'll leave it at that, or I'm afraid this won't get past the censors). In the second, a few tech companies control most resources; only a tiny minority can use the most advanced AI and technology and come close to "mechanical ascension", while most people can only use very weak AI. Climbing out of your class becomes ever harder: you need the strongest AI first, which makes it a vicious circle.

Guess: if Anthropic held the world's most advanced AI forever, would the future become communism or 2077? Go on, guess.

So I still believe frontier intelligence should be supplied to everyone, openly and cheaply. I don't trust Anthropic or OpenAI to do that. Above all, I don't want Anthropic to hold the most advanced AI or AGI. To exaggerate a little, that would be no less serious than letting Hitler get the atomic bomb before the Allies. It is also why I chose DeepSeek and have stayed: we research powerful, fast AI that benefits everyone, and we open-source it. Perhaps that can pull the world back a little from the 2077 end.

May the world of the future be well. May all the beauty be blessed.