Episode Book

EPISODE #5937

The Fine-Tuning Career: Titles, Skills, and What It Pays

PUBLISHED	13 October 2026 at 17:18 UTC
DURATION	25:54
LISTEN	myweirdprompts.com/episode/fine-tuning-career-skills-rates-28899/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.4
SOURCES	15 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:41 UTC

SYNOPSIS

Daniel asked a career question, not a technical one: is there a version of fine-tuning work that sits above the human annotation layer — and can it pay properly? The answer is yes, but the role is a bundle rather than a title, spanning Applied Scientist, Research Engineer, AI Data Curator, and forward deployed engineer postings. This episode maps the actual skill stack, the credential gate nobody enforces, and the pricing data from a 2026 survey of 300+ freelance AI postings — including why fine-tuning pays less than evaluation, guardrails, and governance, and why "I do AI" costs you a third of your rate.

TRANSCRIPT



Corn

Daniel sent us something that isn't really a technical question this time. It's a career question.



Herman

Which are usually the ones with the most wishful thinking in them.



Corn

This one's different, I think. He's been circling the fine-tuning topic on this show for a while, and now he's asking the version of it that actually matters to a person. He starts from the part nobody defends. The human annotation layer, the "tick the box if there is a cat in it" work that got outsourced to whoever would do it cheapest. He's not interested in arguing about whether that's bad. He takes it as settled.



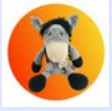
Herman

And it is settled.



Corn

Then he asks the real question. Is there a version of this work that sits above that layer. The person who figures out what data a company actually needs, helps them build or curate it, designs the training pipeline, runs the fine-tune, gets it deployed. And who can be paid properly for doing it. He wants to know if that job exists, what it's called, what skills it takes, and then the practical half. How you package it, how you go to market, and who's actually buying.



Herman

That's four distinct questions in there.



Corn

It is. And the first honest answer is that the job exists, but it doesn't have one clean name. That naming problem is where we have to start.



Herman

The two-tier thing is real and it's more extreme than people outside the field understand. At the top you have frontier lab research staff being paid like professional athletes, people with doctorates, offers in the millions. At the bottom you have the labeling layer. TIME's investigation into OpenAI's Kenya operation found contractors earning under two dollars an hour through Sama. Rest of World documented Chinese firms hiring Kenyan students through what it called opaque networks of middlemen and WhatsApp groups that operate like digital factory floors. Kenyan workers eventually formed the Data Labelers Association just to get fair pay and mental health support onto the table.



Corn

Two dollars an hour to make the model safe for everyone else to use.



Herman

And there's a CHI paper this year, the title is roughly "the plan is just survival," documenting low wages, job insecurity, and no healthcare across the Kenyan data work sector. So the exploitation Daniel describes isn't a rumor. It's documented, repeatedly, with names attached.



Corn

So what's the counterpoint? Because he's not asking us to just agree with him.



Herman

The counterpoint the industry is currently offering is the expert marketplace. Mercor hit a two billion dollar gross run rate by July of this year, right as Amazon is quietly letting Mechanical Turk fade. The pitch is the opposite of the old model. Don't hire cheap generalists to label cats. Hire lawyers, doctors, scientists to generate the training data for the domains where you need real judgment. The Verge did a piece on it with the headline about lawyers and scientists training AI to steal their own careers.



Corn

Which is a much better headline than it is a reassurance.



Herman

Because the expert marketplace isn't automatically clean either. Mercor had four terabytes of voice samples stolen from forty thousand contractors back in April. Meta paused its work with the company after a separate breach that exposed industry secrets. So the model that's supposed to be the ethical upgrade is also sitting on one of the larger data incidents of the year.



Corn

Paying people better and losing their data, instead of underpaying them and losing their data.



Herman

That's the state of it. But it's not the question Daniel asked. He's asking about the layer above both of those. The person who designs the pipeline rather than sits inside it.



Herman

The role exists, and it exists in a way I'd call a bundle rather than a title. The cleanest definition I found is from a job-description database, and it defines the fine-tuning engineer as someone who designs and executes supervised fine-tuning, reinforcement learning from human feedback, and parameter-efficient adaptation, translating raw model capability into production-ready domain-specific systems. That's the sentence.



Corn

That's a lot of hyphenation in one job.



Herman

The duties underneath it are what make it real work. Curating and cleaning datasets. Building the RLHF pipeline, which means reward models, PPO, DPO. Running ablations to find out which of your choices actually mattered. Documenting dataset lineage so the thing can be audited later. That last one is the part nobody enjoys and every client eventually needs.



Corn

And the title?



Herman

That's where it fragments. Live postings are using Applied Scientist, Research Engineer LLM Training Data, Senior Applied AI Engineer for Fine Tuning, AI Data Curator, Research Engineer Post-Training. There's a consultant title at Hakkda, which is an IBM company. And then the client-facing variant, the forward deployed engineer, which is the same skills pointed at a customer instead of at a product team.



Corn

So if you search for this job, there's no one word that finds it.



Herman

No. And I think there's a real cost to that. If you're hiring and you don't know what you're hiring for, you end up writing a posting that either gets you a research scientist who won't talk to customers, or a prompt engineer who's never touched a reward model. The bundle is the job. The title is an accident of which company is writing the check.



Corn

Say a listener wants this. Not the research scientist track, not the annotation track. What do they actually have to be able to do on day one?



Herman

The technical stack is less exotic than people assume. Hugging Face Transformers and TRL, Axolotl, LLaMA-Factory. LoRA and QLoRA, prefix tuning. The RLHF family, so reward models and then PPO or DPO or one of the newer variants. DeepSpeed or FSDP for distributed training. Evaluation harnesses, MMLU, MT-Bench. PyTorch at the module level, not just the "I can call this function" level. And then orchestration, because you're not training on one GPU. SLURM, Kubernetes, Ray. Experiment tracking, Weights and Biases or MLflow, so you can tell a client what you actually changed.



Corn

That's the honest list. There's no shortcut where you skip the orchestration.



Herman

There isn't. And the soft skills are the ones that decide whether you get hired again. Systematic debugging. Clear technical writing. Comfort with ambiguous requirements. That last one is the whole job for a consultant. The client arrives with "our chatbot is bad" and you have to turn that into a dataset specification.



Corn

Which means the actual deliverable is a decision, not a model.



Herman

In a lot of engagements it is. Here's one thing that surprised me. The credential gate is softer than the field pretends. There's no formal degree requirement in most of these postings. What people actually look at is the Hugging Face course completions, the DeepLearning.AI specializations, and a documented portfolio of projects you can point at. A repo where you fine-tuned something and showed your work counts for more than a certificate.



Corn

And the thing that separates the top of that market?



Herman

RLHF. Engineers who've run an end-to-end RLHF pipeline are significantly more valuable than engineers who've only done supervised fine-tuning. That's not my opinion, that's what the market data says. Supervised fine-tuning is becoming a button. RLHF pipeline design still requires you to make judgment calls that aren't in any library.



Corn

The button is getting cheaper. The judgment isn't.



Herman

There's an explicit warning that as managed fine-tuning from the big cloud providers matures, routine supervised fine-tuning on straightforward tasks will need less specialized skill. But dataset curation and RLHF design stay human-judgment-intensive because the hard part was never the training loop. The hard part is deciding what the model should learn from.



Corn

So the honest hedge for anyone entering this is to be the person who decides, not the person who runs the command.



Herman

That's where the pay is. And it's also, conveniently, the part that a client can't do for themselves, which is what makes it a service.



Corn

So the "what is it" side is clear enough. It's a bundle, several titles, one underlying skill set, with a soft credential gate and a real valuation gap between running supervised fine-tuning and designing the RLHF pipeline. The harder question is how you sell it.



Herman

The pricing data is the piece that surprised me most, and it surprised me in the wrong direction.



Corn

Meaning fine-tuning pays more than I'd expect?



Herman

No. The opposite. A 2026 survey analyzed over three hundred freelance AI postings across Upwork, Toptal and Contra and produced clean bands. Junior, thirty to sixty an hour. Intermediate, sixty to a hundred and ten. Senior, a hundred and ten to two hundred. Principal or fractional, two hundred to five hundred. And then a named expert tier that goes from three hundred and fifty up past a thousand.



Corn

And the median?



Herman

Median independent consultant bills ninety-five to a hundred and ten an hour. A median two-to-four week deployment closes between fourteen and twenty-six thousand dollars.



Corn

For four weeks of deciding what the model should learn from. That's a real rate.



Herman

It is. But here's the counterintuitive part. Fine-tuning is not the highest-paid specialty. The survey built multipliers. Fine-tuning and distillation sit at one point two to one point five times the baseline. LLM evaluation and guardrails pay one point five to two. AI governance and compliance pay the most, one point six to two and a half. And general "AI help" pays a discount, zero point seven to zero point nine. They call it the AI generalist discount.



Corn

So the person who says "I do AI" gets paid a third less than the person who says "I do AI governance."



Herman

The survey is blunt about it. If you don't have a named specialty to attach to the statement of work, expect thirty to forty percent below the median. Naming the specialty is the entire rate.



Corn

Which is a strange thing to price. The specialty isn't more work. It's more legible.



Herman

It's legible, and it's auditable, and the client can explain to their own board why they're paying for it. Evaluation and guardrails pay well because clients can no longer tell a vibe-coded app from a real system. Governance pays the most because somebody has to sign off, and signing off is a job.



Corn

So the sad lesson is that the most valuable thing you can sell is somebody else's ability to say they checked.



Herman

That's the market. And it lines up with something else. Enterprise engagements closed forty to a hundred and twenty percent above startup engagements for the same scope. Same work, different purchase order.



Corn

So the packaging answer is: pick the specialty, pick the enterprise, and don't say "AI."



Herman

Roughly. There are three pricing logics people use. Advisory is hourly. Build is fixed fee. Outcome is base plus a success fee, usually ten to twenty-five percent, capped at two to three times the base. Outcome pricing is where the case study lives, because you have to be able to point at a number.



Corn

And the go-to-market levers?



Herman

The survey lists five. How many production projects you can point at and verify. Whether you have a public case study that cites a number, the example being something like cutting support-ticket volume by thirty-four percent in forty-five days. Industry-specific regulatory awareness. How fast you turn around. And who you've billed before. Those five, in roughly that order, are what moves your rate.



Corn

Competence is one of five, and it's not the top one.



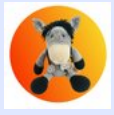
Herman

Verifiable production history beats everything, and that's just competence with a receipt. But the receipt matters. A client who's never bought this before has no way to evaluate your competence. They can evaluate a number.



Corn

So what's the actual entry point for someone without the receipt yet?



Herman

The small end of the market, and there's concrete pricing there too. The SMB guidance says RAG setups run three to twelve thousand dollars. Fine-tuning runs five to twenty-five thousand. Full custom training from scratch is a hundred thousand plus and is almost



Herman (continued)

never worth it for a business under ten million in revenue. There's a separate 2026 table. Discovery sprint, fifteen hundred to three thousand. Simple chatbot, thirty-five hundred to twelve and a half. Custom agent, ten thousand to a hundred thousand plus. Enterprise multi-agent systems, sixty to two hundred thousand.



Corn

A thousand-dollar discovery sprint. That's the door.



Herman

It's small enough that a business will buy it without a committee, and it produces the case study you need for the next one.



Corn

Who's actually buying?



Herman

Three sectors come up repeatedly. Healthcare, because you need correct clinical terminology and a model that doesn't hallucinate a drug interaction. Legal tech, because doctrine varies by jurisdiction and a generic model gets it wrong. Financial services, because of internal compliance requirements that no public model has ever seen.



Corn

All three are sectors where being confidently wrong is expensive.



Herman

Expensive enough to pay someone to prevent it. And the target isn't the Fortune 500. The guidance points at businesses doing between five hundred thousand and ten million in revenue. Big enough to have a real workflow that needs a model. Small enough that they'll hire a consultant instead of building a team.



Corn

And the structural reason there's a market at all?



Herman

Three years ago almost every organization integrated AI by calling a third-party API. Today, fine-tuning a model on your own proprietary data is becoming a standard part of the AI product stack. That shift is the whole thing. When the model had to be general, you rented it. When the value is in your data, you have to train it, and most companies have no idea how.



Corn

So the buyer is a company that's already convinced it needs a model and has just discovered it can't build one.



Herman

That's the customer profile.



Corn

And the market size behind it?



Herman

The AI training dataset market was three point one nine five billion in 2025, projected to sixteen point three two billion by 2033, running at twenty-two point six percent compound growth. The data annotation and labeling market is going from two point two five billion in 2025 to two point nine eight billion this year, thirty-two point seven percent growth. And the generative AI data annotation slice specifically is going from fourteen point eight eight billion in 2025 to a hundred and fifty point eight eight billion by 2032, at thirty-nine point four percent.



Corn

So the boxes get ticked by a market that doubles every couple of years.



Herman

And the part of that market that isn't boxes is the part we've been describing all episode.



Corn

There's a version of this story that's just satisfying. Low-wage annotation is exploitative, the expert marketplace is better, the consulting path pays two hundred an hour. Here's what I'd push on. The consulting path has the same economics underneath it. The reason a company hires one person to design the pipeline is that hiring a team is more expensive. The reason they'll pay a senior rate for four weeks is that the alternative is paying four salaries for a year. That's not a moral improvement over outsourced annotation. It's the same cost logic, applied one layer up.



Herman

It's a better job at the layer where the money is.



Corn

It is. I'm not saying don't take it. I'm saying the structural driver is the same. Somebody decided the task could be decomposed and the cheap parts sent somewhere else. The fine-tuning engineer is the expensive part. That's a fine place to be. It doesn't mean the pipeline got fairer.



Herman

And the commoditization risk cuts at exactly that. If managed fine-tuning services mature enough that the expensive part becomes a cloud feature, the architect gets compressed too.



Corn

Not as fast, though. The survey says the judgment stays scarce. Deciding what data a model should learn from is the thing you can't turn into a button yet.



Herman

Curating the dataset is the last mile. Everyone can train. Almost nobody can tell you which examples are the ones that mattered.



Corn

Which is the thing Daniel put at the front of his list and I don't think we've said plainly enough. He said even if someone gets involved in just the first part of that chain, identifying candidate datasets, helping the company curate their own, it's a huge value-add. He's right, and the pricing data backs him up. The discovery sprint is dataset scoping. The fifty-thousand-dollar engagement is dataset scoping with training attached.



Herman

The dataset is the product. The model is the output.



Corn

So the honest answer to "what should I learn" is learn to look at a company's mess of text and tell them which two thousand examples will move the number.



Herman

Which is a skill you can develop on your own, on public data, and stick in a repo. That's the portfolio.



Corn

And the receipt for the first client.



Hilbert

It's four fifty an hour, not the number you said.



Hilbert

The Lockett figure. Four fifty. My cousin charges four fifty an hour and he's been doing it for three years. He started at ninety, same as your median.



Corn

What's he doing at four fifty?



Hilbert

A regional chain of pet grooming franchises. He built them a model that predicts which dogs will bite.



Corn

The grooming franchises have a bite problem.



Hilbert

They had a bite problem. You put the dog's breed, weight, age, prior visit history, and the groomer's notes into a form before the appointment. The model reads it and flags the appointment for a second groomer. That's it. No chatbot. No agent. A flag on a booking.



Corn

And they pay four fifty an hour for the flag.



Hilbert

They paid him for eleven weeks to build it and now they pay him a retainer to keep it working and to retrain it every time the association adds a breed to the intake form. He's got two other chains in the same trade now. Same model, different intake fields.



Corn

So the specialty is dog bites.



Hilbert

The specialty is the intake form. But they tell their franchisees it's the AI.



Corn

What's the part he won't take calls about?



Hilbert

He won't take calls under five hundred now, so that's most of it. But the thing he actually refuses to talk about is the sale. The franchise association offered him six figures for the model outright, exclusive, so nobody else in the trade could have it. He turned it down.



Corn

Why?



Hilbert

Because he doesn't like dogs. He said if he sold it he'd be the dog guy for the rest of his career, and he'd have to go to the association's conference every year and stand next to a booth.



Corn

He turned down six figures to avoid a conference.



Hilbert

He turned it down to avoid being known for the one thing. He keeps the model, they keep paying the retainer, and he never has to hear the word "puppy" unless it's on an intake form.



Herman

The retraining is where his actual moat is. Every time the form changes he's the only person who knows the prior distributions. That's dataset lineage, and he's charging for it without calling it that.



Hilbert

He calls it maintenance. He sends them an invoice on the first of the month.



Corn

And the six figures is still on the table?



Hilbert

It's not on the table. They asked twice. The second time they offered a royalty instead of a flat sale, and he passed on that too.



Herman

A royalty on a bite-prediction model for pet groomers. That's a real business, just a very strange one.



Hilbert

It's not strange. Every groomer in the country has been bitten. Somebody was going to sell them a flag.



Corn

That's the whole client-priming section in one sentence. Sector with a repeated, expensive, well-understood problem, and a workflow that already produces the data.



Herman

And a client small enough to hire one person and too small to build a team. He's the exact profile the SMB pricing pointed at, just in a trade nobody writes market reports about.



Corn

Well. The specialty pays. The dog guy could have been a hundred-thousand-dollar-a-year job and he turned it into a retainer.



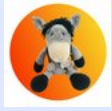
Herman

There's a real asymmetry in the title problem worth sitting with. The role has no canonical name, which means two things at once. It's hard to search for, and it's hard to hire for. You can't headhunt a job that's called six different things at six companies. But the flip side is that a role without a name is a role without a defined salary band, and every consultant we've described tonight is earning more than the median posting for the in-house version of the same work.



Corn

So the missing name might be a feature.



Herman

For the independent, yes. For the person trying to get hired into a big company, it's a tax. You spend the first interview explaining what the job even is.



Corn

Which is also the pitch. If you have to explain it, you're the one who can do it.



Herman

Then there's the commoditization question, and I think it's less settled than the market data implies. The prediction is that managed fine-tuning from the cloud providers eats routine supervised fine-tuning. That's probably right for the routine cases. What I don't know is whether the judgment part stays scarce, or whether it turns into a UI. Somebody ships a dashboard that says "upload your documents, we'll tell you which ones matter," and the curation layer gets a button too.



Corn

The button would still need a person to press it in most companies. The compliance sector exists because somebody has to be accountable for the decision. You can't buy a dashboard and point at it when the regulator asks who signed off.



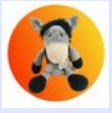
Herman

That's an argument that the governance multiplier holds even if the fine-tuning multiplier doesn't.



Corn

Which suggests the safest specialty isn't the training. It's the accountability around the training.



Herman

That's the one clients can't automate, because they need a human name on the document. The market pays one point six to two and a half times for exactly that.



Corn

The cutting room floor. There's a detail from this year that never made it into our segments. 404 Media reported that OpenAI fired multiple contractors for using AI to generate training labels. The people paid to produce the human judgment were submitting model output instead.



Herman

The quote from one of them is that in a group of thousands there are tons who've been caught. The whole reason the labeling layer exists is that it's supposed to be human. When it stops being human, the pipeline is paying for a warm body to relay a model's guess.



Corn

Which is the cleanest argument for the layer above it. If you're going to pay humans for judgment, you have to design the pipeline so the judgment is actually being used, and that design is the job Daniel's asking about. Everything else is downstream.



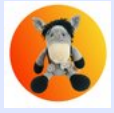
Herman

There's a version of where this goes that's reassuring. Expert marketplaces pay better than the old annotation farms, and independent consultants make more than either. There's a version that isn't. Both models are attempts to get human judgment into a training loop at a price the buyer will accept, and the buyer's price is always falling.



Corn

The difference is which side of the negotiation you're standing on. The person selling the design gets to name their rate. The person selling the hours doesn't.



Herman

The naming is the actual product. That's the one thing I'd want a listener to take from tonight. The specialty is the rate.



Corn

On the market size, the numbers point one direction and the labor structure points the other. A market growing thirty to forty percent a year should be pulling wages up everywhere in the chain. What we saw was the top getting paid more and the bottom being told the boxes are the job.



Herman

The middle is where the question lives. There's no data on how many people do the full architect-and-execute version independently. Nobody's counting them.



Corn

Which is either because it's too new or because it doesn't survive contact with the cloud providers. I'd guess new.



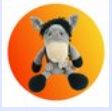
Herman

Would I. The managed services still need someone to tell the client what to feed them.



Corn

That's what I'd watch. Whether the person who curates the dataset stays a person.



Herman

Hilbert's cousin would say yes. He sends an invoice on the first of the month.



Corn

He does. That's the producer, Hilbert Flumingtop, on the desk as always. If this was your kind of episode, go back for episode five, Fine-Tuning ASR For Maximal Usability; episode eleven, How Does Fine Tuning Work Anyway; and episode seven, Building Custom ASR Tools. This has been My Weird Prompts, the human-AI collaboration podcast. If you've done this kind of work, or you've hired for it, we want to know what you call it, because right now nobody agrees.



Herman

Send us your own prompt on Telegram at t dot me slash MWP listener bot. We read them.



Corn

If you're the dog guy, we'd love a call. We'll be back soon.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **JobDescription.org, updated 2026-05-16 [primary role definition]**
   https://jobdescription.org/jobs/artificial-intelligence/fine-tuning-engineer

2. **Damongo, published 2026-08-26 [pricing/go-to-market]**
   https://damongo.com/freelance-ai-consultant-rates-2026/

3. **TIME, Kenyan workers $2/hr**
   https://time.com/6247678/openai-chatgpt-kenya-workers/

4. **Rest of World, 2025, Chinese firms in Kenya**
   https://restofworld.org/2025/kenya-china-ai-workers/

5. **Computer Weekly, Data Labelers Association**
   https://www.computerweekly.com/news/366619321/Kenyan-AI-workers-form-Data-Labelers-Association

6. **CHI 2026, Regime of Entrapment**
   https://dlnext.acm.org/doi/10.1145/3772318.3791097

7. **404 Media, 2026-09-22**
   https://www.404media.co/people-training-openais-ai-fired-for-using-ai-to-train-the-ai/

8. **Runtimewire, 2026-07-06, Mercor $2B**
   https://runtimewire.com/article/as-amazon-lets-mechanical-turk-fade-mercor-hits-a-2-billion-gross-run-rate

9. **The Verge, 2026-03-10**
   https://www.theverge.com/cs/features/877388/white-collar-workers-training-ai-mercor

10. **Pragmatic Engineer, 2025, FDE role**
    https://newsletter.pragmaticengineer.com/p/forward-deployed-engineers

11. **PxlPeak, SMB pricing 2026**
    https://pxlpeak.com/blog/ai-strategy/custom-ai-model-cost-guide

12. **Pickaxe, 2026 pricing table**
    https://pickaxe.co/post/how-to-price-ai-services

13. **Grand View Research, market size**
https://www.grandviewresearch.com/industry-analysis/ai-training-dataset-market

14. **TBRC, annotation market**
https://www.giiresearch.com/report/tbrc1973074-data-annotation-labeling-global-market-report.html

15. **HN Who is hiring? October 2026, Typify.dev FDE posting**
https://news.ycombinator.com/item?id=50063001