Episode Book

EPISODE #5853

Who Actually Competes With Hugging Face?

PUBLISHED	11 October 2026 at 09:16 UTC
DURATION	27:00
LISTEN	myweirdprompts.com/episode/hugging-face-model-hub-rivals-39229/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.3
SOURCES	18 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:40 UTC

SYNOPSIS

Hugging Face hosts 2.96 million public model repositories — and 85.6% of them have fewer than 200 lifetime downloads, while 1.5% of repos account for 99.2% of all downloads. So what are ModelScope, Kaggle Models, and Ollama's library actually for, and when would a team pick one over the Hub? This episode works through the catalog, community, licensing, and deployment differences, including the fact that Chinese models like Qwen and GLM are mirrored identically across Hugging Face and ModelScope — the real variable is transfer speed and reliability from where you're sitting. It also covers the money: Nvidia's $12.93B agreement to acquire Hugging Face and Stripe's $8B+ deal for OpenRouter, roughly $21B of neutral infrastructure changing hands in about six weeks. Plus the licensing split that surprised us: 59% of large Chinese releases carry Apache 2.0, while 41% of comparable American releases sit under custom terms.

TRANSCRIPT



Corn

Most coverage of this topic starts in the same place. Hugging Face won. Everything else is a footnote. Herman, I want to open by saying that's wrong, and then spend the next half hour explaining why I mostly believe it anyway.



Herman

That's a strong start.



Corn

Daniel's question is this. Hugging Face is the best-known hub for open models, but what else is actually out there? He names ModelScope, Kaggle Models, Ollama's library. How did each of these emerge? How do they differ in catalog, community, licensing, deployment, audience? Is Hugging Face the GitHub of AI models, or is there real competition? And what would actually make a person pick one hub over another?



Herman

Fair question, and the answer starts with a number.



Corn

Two point nine six million.



Herman

Two point nine six million public model repositories. Plus a million datasets. Plus one point four four million Spaces. And almost none of it gets used. Eighty five point six percent of those models have fewer than two hundred lifetime downloads. One and a half percent of the repositories account for ninety nine point two percent of all downloads.



Corn

So the hub is enormous and almost entirely empty.



Herman

That's the honest shape of it. It's a warehouse where one aisle is on fire and the other forty thousand aisles have never been walked down.



Corn

The question is whether the other hubs are doing something different or just doing the same thing smaller. Who are we actually talking about?



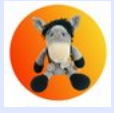
Herman

Four categories. ModelScope, which is Alibaba's platform, built as a model-as-a-service offering, China-market focus, somewhere north of eighty thousand model repos. Kaggle Models, which is Google's, free, and which absorbed the entire TensorFlow Hub catalog back in November twenty twenty three. Ollama's library, which is a curated catalog of GGUF models specifically for running locally. And then the inference marketplaces, Replicate, Together, Fireworks, which serve models rather than host weights.



Corn

And TensorFlow Hub itself?



Herman

Effectively dead. Superseded since twenty twenty three. tfhub.dev redirects to Kaggle now. If anyone is still treating it as a live competitor, they haven't looked at it in three years.



Corn

There's a fifth thing hanging over this whole conversation, which is the money.



Herman

In September, Nvidia agreed to acquire Hugging Face for twelve point nine three billion dollars. That includes a one billion dollar retention equity package. Closing targeted for the first half of next year, pending regulatory review.



Corn

And weeks before that, Stripe agreed to buy OpenRouter for over eight billion.



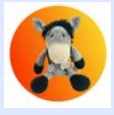
Herman

OpenRouter handles more than ten trillion tokens a day across four hundred plus models for ten million plus developers. So you're looking at roughly twenty one billion dollars of what everybody has been calling neutral infrastructure moving under corporate ownership in the space of about six weeks.



Corn

Which is the real story underneath Daniel's question. The hub landscape isn't just a catalog comparison. It's a question about who owns the front door.



Herman

One gap I should flag. GitHub Models and Azure AI Foundry, I couldn't confirm their current catalogs. Neither of us could. That's an open question and we should say so rather than pretend we know.



Corn

Noted. So, the catalogs. Where do we start?



Herman

China, because that's where the story actually starts. ModelScope is Alibaba's platform. Qwen lives there natively. That's not a distribution decision, it's infrastructure alignment. Qwen isn't mirroring to ModelScope, it's built inside it.



Corn

And GLM?



Herman

GLM on ModelScope is a China-market distribution decision. Different motive, same address.



Corn

The two Chinese-facing platforms mirror the same weights. That's the thing that surprised me when I read through it.



Herman

Identical. Same license, same model card. The choice between Hugging Face and ModelScope for a Chinese model isn't about content, because the content is the same file. It's about access and tooling.



Corn

So the choice is about the road, not the destination.



Herman

Right. For a team in China pulling something like GLM 5.2 at seven hundred forty four billion parameters, or Kimi K3 at two point eight trillion, ModelScope is materially faster. For a team outside China, Hugging Face is faster. That's it. That's the whole difference in that column.



Corn

Two point eight trillion parameters. Say that number out loud once more.



Herman

Two point eight trillion.



Corn

The file sizes involved here aren't downloads, they're relocations.



Herman

That's exactly the right way to think about it. You don't download a two point eight trillion parameter model, you move into it. I mean, at that scale you're talking about a checkpoint that takes up multiple terabytes before you've even thought about quantization. You're not pulling it onto a laptop. You're provisioning storage for it the way you'd provision storage for a database migration.



Corn

So the practical question for a team isn't "which hub has it," it's "which hub can I physically get it from in under a week."



Herman

Right, and that's the part of the comparison that people skip, because it's not about features. It's about physics. Distance to the weights is a real variable, and it's the one variable where the answer flips depending on where you're sitting.



Corn

And it's not just raw transfer time, right? It's also the retry logic, the resume behavior, the checksumming on a multi-terabyte pull. If a connection drops at hour nine of a twelve-hour transfer, what happens?



Herman

Depends entirely on the tooling. Some clients resume cleanly, some start over, and some fail in a way that leaves you with a corrupted shard you don't discover until you try to load the thing. At that scale, a failed download isn't an inconvenience. It's a day gone. Which is



Herman (continued)

another reason the geography question is sharper than it sounds. It's not just "which is faster," it's "which is faster and more reliable from where I'm sitting," and those two answers don't always match.



Corn

Kaggle, then. How did Google end up with a hub?



Herman

By consolidation, mostly. Kaggle Models absorbed TensorFlow Hub in November twenty twenty three, which pulled all of Google's model hosting under one roof. They version weights per framework, you pull them with the kagglehub library, and it's free. FAUN scored it seventy one out of a hundred, with its best marks on pricing transparency and cost of ownership. Which makes sense, because the price is zero.



Corn

And Ollama?



Herman

Ollama came out of the local inference movement. llama.cpp, GGUF, running models on hardware you already own. It's a curated library, not a warehouse. Every entry is meant to be runnable. Top pulls are llama three point one at a hundred and twenty million, deepseek r one at ninety three point six million, nomic embed text at eighty eight point nine million.



Herman

Curated is the whole product. Ollama's library is a promise that whatever you pull will actually run, which is a promise Hugging Face can't make about two point nine million repos. That's the trade. Ollama gives up breadth to guarantee function. Hugging Face gives up the guarantee to get the breadth.



Corn

So if I pull something from Ollama and it doesn't run, that's a bug. If I pull something from Hugging Face and it doesn't run, that's Tuesday.



Herman

That's Tuesday. And that's not a knock on the Hub, it's just the arithmetic of hosting everything. You can't curate two point nine million repos. Nobody can. So the curation has to happen somewhere else, either in the community's voting or in a downstream tool that filters for you.



Corn

Then there's the thing that happened in February. The ggml team joined Hugging Face.



Herman

llama.cpp, the most important project in local inference, moved under the Hub's resources, remaining open source and community governed. So the local-inference ecosystem and the dominant hub are now the same organization, sort of, which is a strange sentence to say out loud.



Corn

The thing I keep circling back to is the licensing split, because that's where the data actually surprised me.



Herman

Of a hundred and seventy eight Chinese releases above twenty billion parameters this year, fifty nine percent carry Apache two point oh, twenty two percent carry MIT.



Corn

Almost nothing non-commercial.



Herman

Almost nothing. Meanwhile on the American side of the same size band, twenty nine percent is Apache or MIT, forty one percent sits under custom terms, and thirty percent declares nothing at all.



Corn

So the American labs are the restrictive ones.



Herman

At that parameter scale, yes. There's a caveat on the Hugging Face claim that almost none of the Chinese releases carry non-commercial restrictions. A commenter pushed back and cited Kimi's twenty million dollar annual revenue threshold, above which you need Moonshot's authorization to serve it. So the "no restrictions" framing isn't quite right either.



Corn

Twenty million in revenue is a much friendlier line than most American licenses draw.



Herman

Much friendlier. It's a restriction, but it's a restriction that only touches companies big enough to negotiate. A two-person startup can build a product on top of it and never think about the clause. A company with a legal department and a hundred million in revenue is having a different conversation.



Corn

So it's a restriction aimed at competitors, not at users.



Herman

That's how I read it. It's a "don't resell our model as a service" clause wearing a revenue threshold as a disguise. Which is a much more surgical instrument than a blanket non-commercial license.



Corn

What about the aggregate picture? If you look at what people actually download, not what gets released.



Herman

Zentor looked at the top thousand most-downloaded models. Seventy one point five percent ship under a permissive open license. Apache two point oh alone owns forty nine percent of that group. So the top of the download distribution is heavily permissive, and the licensing picture gets messier as you go down.



Corn

The stuff people use is open. The stuff people publish is a mixed bag.



Herman

That's the split, yeah. Publishing is cheap. Using is a vote.



Corn

Qwen. Because that's the case study that makes the whole thing concrete.



Herman

A hundred and fifty one thousand four hundred and forty eight Qwen derivatives on the Hub. That's two point six times Meta's footprint and four point seven times Llama's. Growing at a hundred and eighty to two hundred and ten new repos a day.



Corn

A day.



Herman

And here's the part I find interesting. Of twenty eight thousand five hundred and thirty one GGUF conversions of Qwen models, Qwen published fifty four of them.



Corn

Fifty four out of twenty eight thousand.



Herman

The community did every other one.



Corn

That's the ecosystem doing the work the lab didn't.



Herman

And it shows up in the download numbers. Qwen GGUF downloads run thirty nine point six million a month. Gemma's at twenty point eight. Llama's at seven point five.



Corn

So Qwen gets a third of a billion downloads a year on a format the lab barely touches.



Herman

Because the community did the conversion work. That's what a hub does that a file server doesn't. A file server holds the file. A hub holds the file and then holds all the work other people did to the file, and that second thing is the actual product.



Corn

Herman, hold on. Do we know if that's Qwen choosing not to publish GGUF, or Qwen not having the bandwidth?



Herman

I don't know, honestly. If I had to guess, it's that the quantization community is faster than any lab's release engineering, so it's not worth competing with. But I don't have that confirmed, so treat it as my read, not a fact.



Corn

Fair. What's the size distribution look like?



Herman

Models under one billion parameters take eighty three percent of all-time downloads. Above a hundred billion parameters, one percent.



Corn

Eighty three percent of the downloads are for the smallest models.



Herman

Which tells you what hubs are actually used for. Most downloads are embedding models, classifiers, small fine-tunes, things you run in a pipeline, not things you argue about on the internet.



Corn

The models that get discussed and the models that get pulled are two different populations.



Herman

Almost completely different. There's one repo that appears in both the top twenty five by downloads and the top twenty five by likes. One. And thirteen of the top twenty five downloads date from twenty twenty two.



Corn

Twenty twenty two. That's four years old.



Herman

Because a download is a dependency decision, not an excitement decision. Likes track what's new. Downloads track what already works.



Corn

Give me the concrete version of that, because I think people hear "dependency decision" and it slides past them.



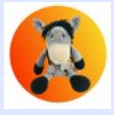
Herman

Take a sentence-embedding model. Somebody wired it into a retrieval pipeline in twenty twenty three. It works. It's fast. It's cheap. Nobody has any reason to touch it, because touching it means re-indexing everything downstream. So it keeps getting pulled, every CI run, every fresh container build, every new hire's laptop. That's a download. That's not a person being excited about a model. That's a build script doing its job.



Corn

So the download leaderboard is really a list of things that are load-bearing in other people's systems.



Herman

And that's why it barely moves. The top of that list is sticky in a way the likes list never will be, because swapping out a dependency costs real engineering time, and swapping out something you're excited about costs nothing.



Corn

The catalog isn't the product. The lock-in is the product. Say more about that, because I think that's the load-bearing part of the whole comparison.



Herman

Everything downstream of the model file is Hugging Face-first by convention. transformers, vLLM, the evaluation harnesses, the quantization libraries. They assume the Hub. If you publish on ModelScope and nowhere else, you're not just missing an audience, you're missing every tool that assumes Hugging Face paths by default.



Corn

So the hub isn't the catalog. The hub is the assumption.



Herman

That's the moat. Two point nine million repos is impressive, but the two point nine million is downstream of the fact that every library you'd want to use reaches for the Hub first.



Corn

Is there a version of this that's happened before? Because "the tool assumes the platform" feels like a pattern.



Herman

It's the oldest pattern in software. Nobody uses npm because it has the best packages. They use it because every tutorial says npm install and every CI config already has it wired in. The package manager becomes the assumption, and the assumption outlives every individual package on it. Hugging Face is running the same play, just with weights instead of code.



Corn

Then the question is whether that assumption survives an acquisition. Which is where Daniel's GitHub framing gets interesting.



Herman

ToolHalla calls Hugging Face the ecosystem play, the GitHub of machine learning models, with inference tacked on as one layer. And VentureBeat drew the precedent explicitly. Microsoft bought GitHub in twenty eighteen for seven and a half billion dollars and promised it would remain open and operate independently.



Corn

It grew. Two hundred twenty five million users, ninety percent plus of the Fortune five hundred.



Herman

Both true.



Corn

And Microsoft also made GitHub the distribution point for Copilot.



Herman

Also true. That's the analogy cutting both ways. The same acquisition that kept GitHub open and grew it also turned it into the funnel for a commercial product. Both things happened in the same building.



Corn

So the honest version of the analogy is, owning the place where developers choose among everybody's products may be worth more than owning any single product.



Herman

Considerably more. You're not betting on which model wins. You're betting that there will keep being a next model, and that everyone will look for it in the same place.



Corn

There's the neutrality question underneath it, then. Jensen Huang says Nvidia compute will not be required to build on or deploy through Hugging Face.



Herman

He said that. And VentureBeat's list of questions is exactly the right list. Will discovery stay neutral? Will AMD models and TPU models get equal integration? Because a promise not to require Nvidia compute doesn't answer whether the ranking of what you see first changes.



Corn

There's a quote from Duane O'Brien at the Open Source Initiative that I think lands harder than the promise does. He said history shows that when a platform pushes open source developers through proprietary workflows, they find or build more open alternatives.



Herman

And Nithya Ruff from the Linux Foundation board, which I think is the sharpest line in the whole discussion. Neutrality is a discipline a company must choose time and again, not a promise it makes once.



Corn

A promise is a point in time. A discipline is a practice.



Herman

That's the distinction. Every ranking change, every default integration, every "recommended" tag is a place where the discipline either holds or doesn't.



Corn

And the model stays open either way. The file is the file. You can still download the weights.



Herman

The model is open. The infrastructure around it may not be. That's the honest framing. Replicating a model file is easy. Replicating an ecosystem is not.



Corn

And that gets us to the practical question. If someone is trying to choose a hub, what actually drives the choice?



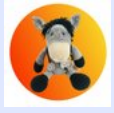
Herman

Five things. Geography and latency, which is ModelScope for China and Hugging Face everywhere else. Tooling, which is Hugging Face-first by convention and will stay that way until the libraries change. Cost, which is Kaggle Models, because it's free and transparent about it. Local deployment, which is Ollama, because that's the curated runnable path. And portability risk.



Corn

Portability risk meaning what, for someone not in infrastructure?



Herman

VentureBeat's advice is the right advice. Mirror critical artifacts, pin revisions, archive the licenses, and separate artifact storage from runtime inference. Which is a lot of words for a simple thing. Don't let your production system depend on somebody else's front door staying exactly the same shape.



Corn

And that advice applies to Hugging Face today, presumably, not just ModelScope.



Herman

Applies everywhere. It's just more urgent at a hub that's being acquired.



Corn

The agent traffic numbers are the detail here that I want to get into, because they're the first sign of who's actually hitting these hubs.



Herman

July numbers on the Hub. Claude Code at forty four point four percent, down from sixty seven point eight percent in April. Codex climbed from ten point four to twenty point eight. And about a quarter of the agent-tagged traffic came from unregistered harnesses that nobody can identify.



Corn

Unregistered harnesses are a quarter of it.



Herman

Which means a quarter of the agent traffic on the biggest model hub is from tools nobody has declared. If you're trying to reason about who's using a hub and for what, you're missing a quarter of the picture by default.



Corn

And the FAUN scoring, because I want the numbers on the record.



Herman

Hugging Face eighty seven. Kaggle seventy one. ModelScope sixty nine. TensorFlow Hub thirty seven.



Corn

The second place is twenty six points back.



Herman

Twenty six points. And that twenty six points is mostly a catalog and integration gap, not a feature gap. Kaggle does the things it does well. It just doesn't do the breadth.



Corn

GGUF repos on the Hub are up four hundred sixty four percent in seven months, which is worth saying out loud because of what it means for Ollama.



Herman

Ollama is the curated path. GGUF is the format. And the format is now growing faster on Hugging Face than on Ollama's own library.



Corn

So Ollama's advantage is curation, not uniqueness. The same files are available elsewhere, just without the promise that they'll run.



Herman

That's the honest state. Curation is a real product. It's just a thin one, because it's a filter on top of a catalog you don't own.



Corn

Does that make Ollama fragile, in your read? Because a filter on somebody else's catalog sounds like a business that can be squeezed from either end.



Herman

It's a real tension. If the Hub makes discovery good enough, Ollama's filter stops being worth the extra step. If the Hub makes discovery bad enough, Ollama's filter becomes essential. So Ollama's fate depends on how well the Hub curates, which is not a thing Ollama controls. That's an uncomfortable place to sit, even if the product is good.



Corn

The China open-weight lead is worth one more pass. In almost every month this year, the largest open model from a Chinese lab was bigger than any US lab's release.



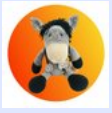
Herman

Seven hundred fifty four billion to two point seven eight trillion on the Chinese side, under a hundred and thirty billion on the American side in five of seven months.



Corn

That's a gulf, not a lead.



Herman

It's a strategy difference. Open weights are how a Chinese lab gets distribution in a market where they can't buy their way into Western cloud defaults. Publish the weights, and the weights travel.



Corn

And ModelScope is the domestic mirror of that strategy. Where the weights go to be found by Chinese teams.



Herman

Hugging Face is where they go to be found internationally, and ModelScope is where they go to be found at speed by the people who will actually use them.



Corn

So the answer to Daniel's GitHub question is yes, and the yes is the point.



Herman

That's how I read it. There is genuine competition. It's just competition that mostly doesn't overlap, because these hubs are competing for different things. ModelScope is competing for Chinese-market access. Kaggle is competing for Google's framework ecosystem to have a home. Ollama is competing for the local-inference crowd. The inference marketplaces are competing for the traffic, not the storage.



Corn

And Hugging Face is competing for the assumption.



Herman

For the assumption. For being the default that everything else points at.



Corn

Which is why the acquisition price is the number I keep coming back to. Twelve point nine three billion dollars is a lot of money to spend on a warehouse where eighty five percent of the shelves are empty.



Herman

It's not a warehouse. It's a customs checkpoint.



Corn

Say that again.



Herman

It's not a warehouse. It's a customs checkpoint. You can put the goods anywhere, but if the checkpoint is where everyone declares them, the checkpoint is the asset.



Corn

And the second-order question is whether the next Qwen or Llama gets equal billing on the checkpoint's front page.



Herman

That's the whole question. It's not whether the weights stay downloadable. They will. It's whether the recommendation layer stays even-handed. And that's not a thing you can promise in a press release. It's a thing that gets tested every day a new model ships.



Corn

So the story isn't that Hugging Face is going away. It's that the front door has an owner now.



Herman

And every front door with an owner eventually gets a doorman.



Corn

The man would like a word.



Hilbert

Not about the front door. About the servers. They're not the same servers, and nobody on this show said that out loud.



Hilbert

I had a job in a records office for a while, cataloging things. Card catalog, cross-referenced by hand. Not models. This was before any of that was mine to worry about. Point is, the shape of the problem was the same. Regional catalog, universal catalog. The universal one always gets the prestige. The regional one always gets the speed.



Herman

You're talking about ModelScope against Hugging Face.



Hilbert

I'm talking about a seven hundred forty four billion parameter model. I tried to pull it one time. Three days on Hugging Face. I switched over and it was about an afternoon. Same file. Same license. The only thing that changed was which building it was sitting in.



Corn

Three days to an afternoon.



Hilbert

The server was closer to the water. I don't know how else to explain it. That's where the cables are, and nobody in this business wants to admit it.



Herman

Hilbert, is the water actually relevant? Because I don't think...



Hilbert

I'm not going to argue about the water, Herman. The point is the file's the same file. Only the ping changes.



Corn

So you agree with the GitHub framing, then.



Hilbert

I agree with it and I think it's wrong in one place. GitHub never had to worry about whether the code would run on the computer you already owned. It just had to hold it. That's the whole job. A model hub has a second job, which is whether the thing will run where you're sitting. Which is why Ollama is the only one that makes sense.



Herman

Ollama's the only one that makes sense.



Hilbert

The only one. I have never once downloaded a model from Hugging Face without converting it to GGUF afterward. The original format is a suggestion. It's a good suggestion. It's not a final answer.



Corn

And you don't think that's a little...



Hilbert

I keep all of them on one drive. The vault. I've never plugged it in, but I like knowing it's there.



Corn

Never plugged it in.



Hilbert

There are three vaults. Three locations. I rotate them. Security.



Herman

Security against what, exactly.



Hilbert

That's a question for a different day. I have to check on something.



Corn

Where does that leave us? Because I think Hilbert just said something true in the middle of all of that, which is that the hub matters most when it's on your own side of the network, and least when it's on somebody else's.



Herman

Will Nvidia keep the discovery layer neutral? Will AMD and TPU models get the same integration? Will ModelScope expand beyond China, or stay a regional mirror? None of those are answered yet.



Corn

The hub isn't a download site, it's infrastructure, and infrastructure ownership is the real story here. Not the models.



Herman

The GitHub analogy may be right, and the open question is whether the next decade of open weights gets the GitHub treatment, open and growing, or the Copilot treatment, open but steered.



Corn

If this was useful, a review helps other listeners find the show. Thanks to our producer, Hilbert Flumingtop.



Herman

For more along these lines, there's episode three, Safetensors or something else; episode fourteen, AGI's Crossroads; and episode two, Local STT For AMD GPU Owners. This has been My Weird Prompts. Send us your own prompt on Telegram at t dot me slash MWP listener bot.



Corn

We'll be back soon.

SOURCES

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

1. **Hugging Face, 2026-08-14 [primary]**
   https://huggingface.co/blog/state-of-open-models-summer-2026
2. **Hugging Face model catalog [primary]**
   https://huggingface.co/models
3. **ModelScope model catalog [primary]**
   https://www.modelscope.cn/models
4. **Ollama model library [primary]**
   https://ollama.com/library
5. **Kaggle Models [primary]**
   https://www.kaggle.com/models
6. **ModelScope docs [primary]**
   https://modelscope.github.io/modelscope/

7. **VentureBeat, 2026-09-04**
https://venturebeat.com/infrastructure/nvidia-acquires-hugging-face-after-stripe-nabs-openrouter-heres-what-open-source-ai-builders-should-do

8. **The Inference, 2026-09**
https://theinference.org/article/nvidia-s-12-9-billion-hugging-face-deal-is-filed-as-an-other-event-no-contract-attached

9. **The Latent East, 2026-08-20**
https://latenteast.com/insights/hugging-face-vs-modelscope

10. **FAUN.dev, 2026**
https://faun.dev/toolbox/compare-model-hubs/

11. **FAUN.dev, 2026**
https://faun.dev/toolbox/hugging-face-hub-vs-modelscope/

12. **Presenc AI, 2026**
https://presenc.ai/research/modelscope-vs-huggingface-2026

13. **Zentor, 2026**
https://zentor.ai/blog/huggingface-hub-state-2026

14. **ToolHalla, 2026**
https://toolhalla.ai/blog/hugging-face-vs-replicate-vs-together-ai-api-2026

15. **AI Credits, 2026**
https://www.aicredits.co/en/blogs/replicate-vs-together-vs-fireworks

16. **Modern DataTools, 2026**
https://www.modern-datatools.com/compare/replicate-vs-together-ai

17. **India Business Journal, 2026-09-04**
https://indiabusinessjournal.com/full-story/2698/nvidia-chief-seals-13billion-hugging-face-deal-to-back-opensource-ai

18. **Die Signal, 2026**
https://diesignal.com/articles/nvidia-to-acquire-hugging-face-for-12-9-billion-in-open-ai-play-1dd46e97