



# UTD Hack 2025

Nemotron Prize Track

**Ethan Einhorn**

Developer Community  
[eeinhorn@nvidia.com](mailto:eeinhorn@nvidia.com)



## Challenge Statement:

In Addition to Using Nemotron, Your Project Must Include:

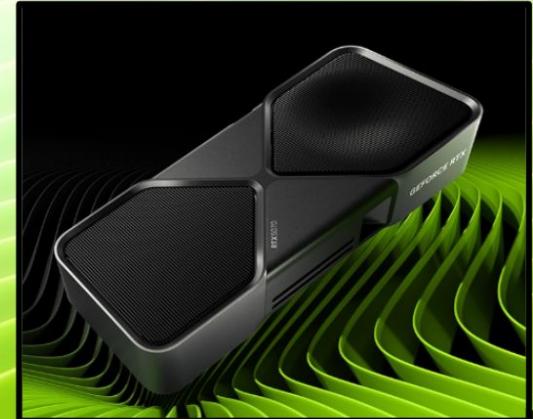
- Beyond a Chatbot - not just responding directly to prompts
- Multi-step workflows - agents that plan and execute complex tasks
- Tool integration - using external APIs/services intelligently
- Real-world applicability - solving actual problems

### Prizing:

First Place:

3 RTX 5080 GPUs

3 NVIDIA Logo Hats



Second Place:

2 RTX 5080 GPUs

3 NVIDIA Backpacks

3 \$100 Brev Credits

Third Place:

3 NVIDIA Backpacks

3 \$200 Brev Credits

# Essential Resources

## Quick Start Workshops (self-paced, deploy in minutes):

[Report Generator Agent Tutorial](#) - Learn agent architecture with LangGraph

[Agentic RAG Tutorial](#) - Build smart retrieval systems

## Refer to the get-started tutorials:

[Build a Report Generator AI Agent with NVIDIA Nemotron on OpenRouter](#)

[Build a Retrieval-Augmented Generation \(RAG\) Agent with NVIDIA Nemotron](#)

## Access Nemotron:

[build.nvidia.com](#) - get API key

Via OpenRouter API Nemotron-nano-9b-v2

On [Hugging Face](#) (open weights)

## Brev Getting Started:

<https://docs.nvidia.com/brev/latest/>

# Getting Started - build.nvidia.com

The screenshot shows the build.nvidia.com interface with a modal window titled "Copy Code to Make an API Request". The modal contains code for generating an API request using the OpenAI library. A green callout box labeled "Step 3: Generate API key" points to the "Generate API Key" button. Another green callout box labeled "API key populates here" points to the "api\_key" variable in the code. A green box labeled "Step 2: Navigate to model and click view code" has an arrow pointing to the "View Code" button in the top right corner of the modal. A green box labeled "Step 1: Log in" has an arrow pointing to the "A" sign-in icon in the top right corner of the main page.

Copy Code to Make an API Request

Python Node Shell API Reference

Using free API for development

Upgrade Generate API Key Copy

```
from openai import OpenAI

client = OpenAI(
    base_url = "https://integrate.api.nvidia.com/v1",
    api_key = "$API_KEY_REQUIRED_IF_EXECUTING_OUTSIDE_NGC"
)

completion = client.chat.completions.create(
    model="nvidia/nvidia-nemotron-nano-9b-v2",
    messages=[{"role":"system","content":"/think"}],
    temperature=0.6,
    top_p=0.95,
    max_tokens=2048,
    frequency_penalty=0,
    presence_penalty=0,
    stream=True,
    extra_body={}
```

Step 1: Log in

Step 2: Navigate to model and click view code

Step 3: Generate API key

API key populates here

# Getting Started - brev.nvidia.com

The screenshot shows the brev.nvidia.com dashboard. At the top, there's a navigation bar with links for NVIDIA, GPUs (highlighted in green), Deployments (new), Launchables, Team, Billing, Docs, Add Credits (highlighted with a green box), and a user ID nca-5hmfb. A green arrow points from the 'Add Credits' button down to the redeem code information.

## GPU Instances

Provision GPU instances, monitor logs, SSH with the Brev CLI, access JupyterLab, and more!

[Create New Instance ↗](#)

**My Instances**

All Mine Team Search Instances

BrevGPT Finetune Workspace Create your first instance ↗  
ID: 1234 · User: Jensen Huang (1234) · Created: 4/16/2024, 10:54:16 PM

Compute Running  
**A100** NVIDIA A100 (80GB)  
2 GPUs x 24 CPUs | 340GB  
120GB GCP  
\$2.48/hr

Container Built  
nvidia/cuda:12.2-devel-ubuntu22.04  
View Logs

Notebook  
Your FineTune NoteBook  
Preview Notebook  
[Access Notebook](#)

\$60 Redeem code  
**utdhackathon-2025**

# Ideal Projects for This Track

**Winning projects will showcase true agentic behavior:**

- Multi-Agent Systems:** Build teams of specialized AI agents  
(like Report Generator: Research Agent → Outline Agent → Writer Agent → Editor)
- Agentic RAG:** Systems that intelligently decide WHEN to retrieve information, not just HOW (perfect for domain-specific assistants)

**ReAct Pattern Workflows:** Agents that Reason → Act → Observe in loops to solve problems iteratively (like automated debugging or technical support)

**Tool-Calling Applications:** Leverage Nemotron's exceptional ability to use external APIs and tools (finance analysis, DevOps automation, content creation)

**Multi-Modal Agents:** Combine Nemotron reasoning with VLMs  
(visual analysis + logical decision-making)

**Agent Simulation & Evaluation:** Use Nemotron to generate realistic test scenarios and evaluation pipelines

## NVIDIA - NEMOTRON

“

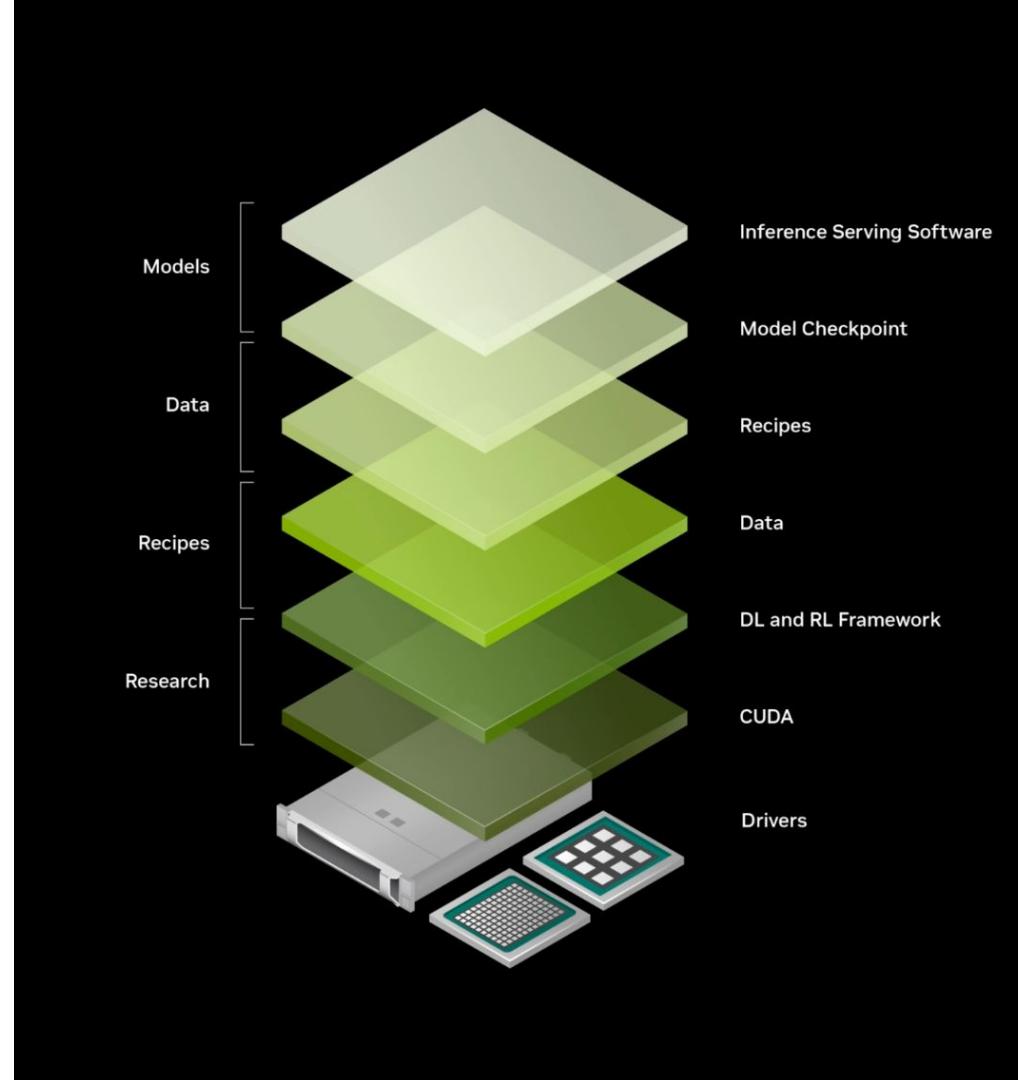
The most open approach to AI development the world has ever seen.

*Bryan Catanzaro, VP of Applied Deep Learning Research. NVIDIA*

”

# Nemotron

Is more than models. It's a focus on optimizing for our AI ecosystem to provide the best quality of results across our technology stack.



# What is Nemotron Best For?

**NVIDIA Nemotron models** are purpose-built for agentic AI - the next generation of intelligent systems that don't just answer questions, but reason, plan, and take action. Unlike general-purpose LLMs, Nemotron excels at:

- Advanced reasoning and multi-step problem-solving
- Function calling to interact with external tools and APIs
- Autonomous decision-making within agent workflows
- Retrieval-augmented generation (RAG) for knowledge-based tasks
- Multi-agent orchestration where specialized agents work together together

## Models

Deploy and scale models on your GPU infrastructure of choice with NVIDIA NIM inference microservices

Filter by text

Sort By: Most Recent

Publisher

Use Case

NIM Type

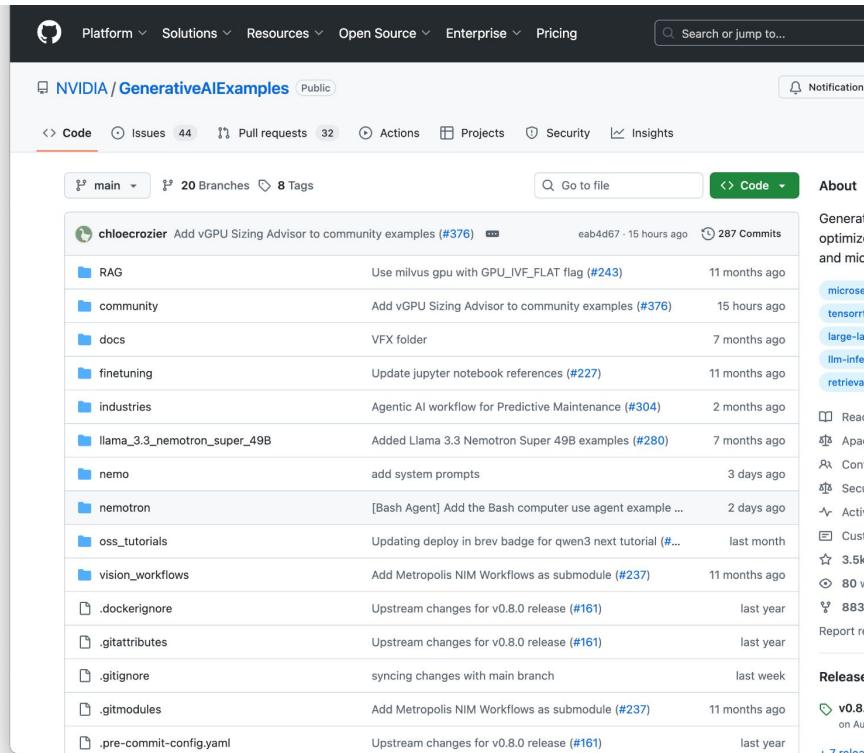
[Clear Filters](#)

|  |   |  |
|--|---|--|
| <br>nvidia<br><b>nvidia-nemotron-nano-9b-v2</b><br>High-efficiency LLM with hybrid Transformer-Mamba design, excelling in reasoning...<br><a href="#">thinking budget</a> <a href="#">chat</a> <a href="#">reasoning</a> +1 | <br>nvidia<br><b>llama-3.3-nemotron-super-49b-v1.5</b><br>High efficiency model with leading accuracy for reasoning, tool calling, chat, and instruction...<br><a href="#">chat</a> <a href="#">math</a> <a href="#">advanced reasoning</a> +3 | <br>mistralai<br><b>mistral-nemotron</b><br>Built for agentic workflows, this model excels in coding, instruction following, and function calling<br><a href="#">language generation</a> <a href="#">chat</a> +3                                      |
| <br>nvidia<br><b>llama-3.1-nemotron-nano-vl-8b-v1</b><br>Multi-modal vision-language model that understands text/img and creates informative...<br><a href="#">doc intelligence</a> +3                                     | <br>nvidia<br><b>llama-3.1-nemotron-nano-4b-v1.1</b><br>State-of-the-art open model for reasoning, code, math, and tool calling - suitable for edge agents<br><a href="#">edge</a> <a href="#">tool calling</a> <a href="#">chat</a> +3       | <br>nvidia<br><b>llama-3.1-nemotron-ultra-253b-v1</b><br>Superior inference efficiency with highest accuracy for scientific and complex math reasoning, coding...<br><a href="#">chat</a> <a href="#">math</a> <a href="#">advanced reasoning</a> +3 |

# Learn from Past Winners

Agent examples [here](#)  
[github.com/NVIDIA/GenerativeAIExamples](https://github.com/NVIDIA/GenerativeAIExamples)

- Route Optimization Agent (NVIDIA NeMo Hackathon): Multi-agent system coordinating logistics with specialized agents for constraint extraction and route planning
- Automated Bug Fixer (Dust Hackathon): Non-technical users describe bugs, agent autonomously diagnoses and fixes code
- Multi-Agent Research System (lablab.ai): Orchestrated deep research with autonomous agent collaboration



The screenshot shows the GitHub repository page for 'NVIDIA / GenerativeAIExamples'. The repository has 20 branches and 8 tags. The main branch is selected. There are 287 commits across various pull requests. Some notable commits include:

- chloecrozier Add vGPU Sizing Advisor to community examples (#376)
- RAG Use milvus gpu with GPU\_IVF\_FLAT flag (#243)
- community Add vGPU Sizing Advisor to community examples (#376)
- docs VFX folder
- finetuning Update jupyter notebook references (#227)
- industries Agentic AI workflow for Predictive Maintenance (#304)
- llama\_3.3\_nemotron\_super\_49B Added Llama 3.3 Nemotron Super 49B examples (#280)
- nemo add system prompts
- nemotron [Bash Agent] Add the Bash computer use agent example ...
- oss\_tutorials Updating deploy in brev badge for qwen3 next tutorial (#...)
- vision\_workflows Add Metropolis NIM Workflows as submodule (#237)
- .dockerignore Upstream changes for v0.8.0 release (#161)
- .gitattributes Upstream changes for v0.8.0 release (#161)
- .gitignore syncing changes with main branch
- .gitmodules Add Metropolis NIM Workflows as submodule (#237)
- .pre-commit-config.yaml Upstream changes for v0.8.0 release (#161)

---

“

Happy Hacking!

---

”