

Using LLMs During your Learning Journey and Beyond

Varada, Aldo, Sky

Using LLMs (Sky)

Free + UBC approved GenAI tools

- OpenAI **ChatGPT** (free version, **DO NOT** login)
 - GPT5



Log in

Sign up for free



ChatGPT

Ask anything

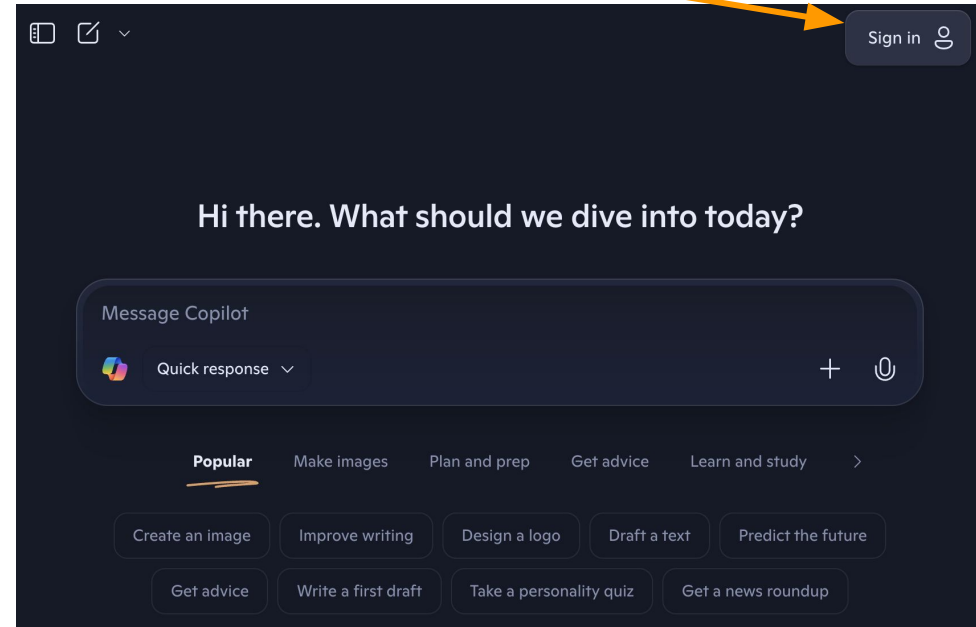
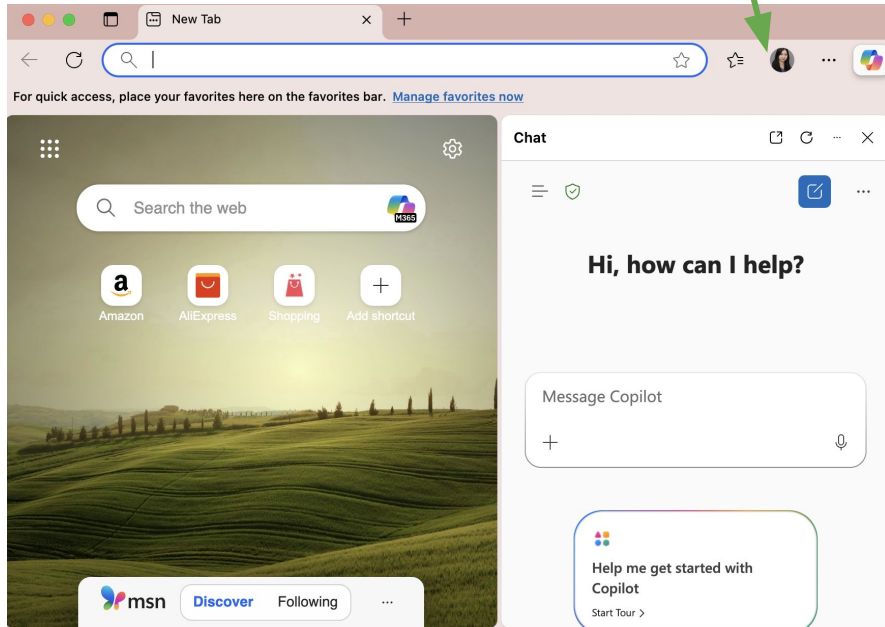
📎 Attach

🌐 Search

🗣️ Voice

Free + UBC approved GenAI tools

- Microsoft **Copilot** Edge web browser sidebar (Log in with @student.ubc.ca)
- Microsoft **Copilot** organization (use as guest)



UBC restricted GenAI tool (hard NO)









As a result, the following applies at UBC:

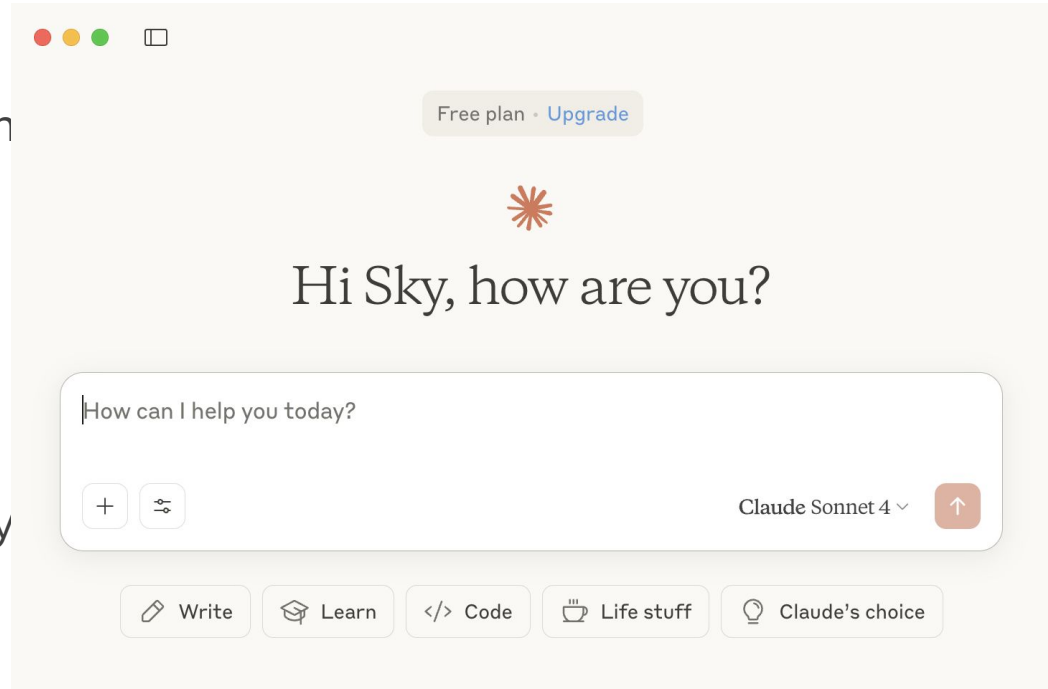
1. The DeepSeek applications, including mobile, desktop and web or browser access (**the “DeepSeek Applications”**) present a high degree of privacy and security risk. The DeepSeek model (i.e., the underlying, open-source architecture) downloaded from reputable sources is not currently known to be high risk.
2. UBC restricts the installation or use of the DeepSeek Applications on any UBC-owned or personal devices that are used for university business or that access a UBC system through user login. Examples include the UBC wireless network, UBC Library systems, Canvas, Workday, VPN, or any other UBC system requiring a username and password. For any other personal devices, UBC strongly recommends against using or installing the DeepSeek Applications.

All the GenAI tools & models I
will demo after this slide are
NOT recommended by UBC

(but you are NOT restricted from using it)

NOT UBC approved GenAI tools

- Anthropic **Claude**
 -  **Free** tier is often enough (Sonnet 4)
 -  Requires login
 -  Best reputation among programmers
 -  Claude Artifact
 -  Emphasis AI safety
 -  Transparent with system prompt



NOT UBC approved GenAI tools

-  Proprietary
 - Google Gemini 2.5 Pro
 - xAI Grok 4
 - Mistral large 2
 - ...
-  **Open** source/Open weight
 - Meta Llama 4
 - DeepSeek V3.1 **MODEL**
 - Kimi K2
 - ...

 Gemini

 MISTRAL
AI_

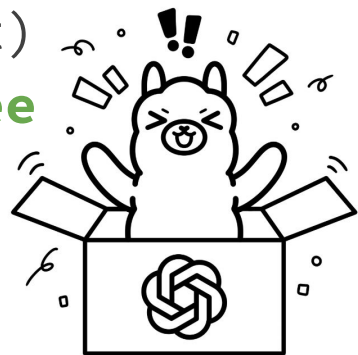
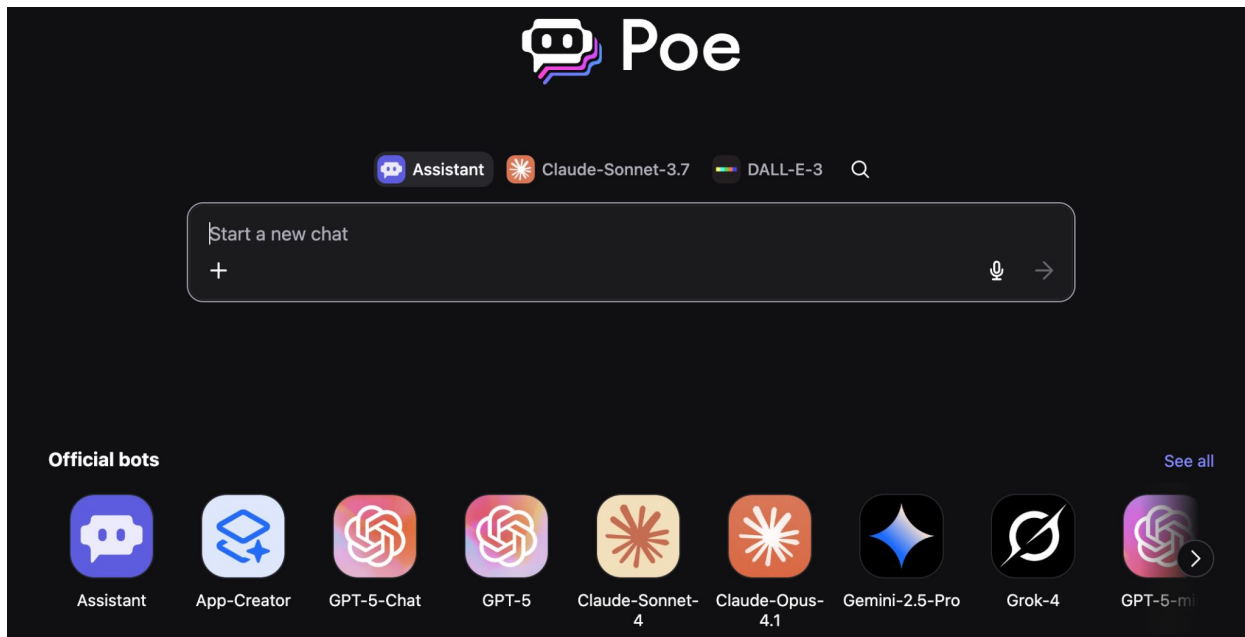


 Meta AI

 deepseek  K

Is there a platform where I can try them all?

- **Poe** (open source + proprietary)
 - **Subscription** based (I used up \$20 limits in a few hours...)
- **Ollama** (open source / open weight)
 - **Free**



OpenAI and Ollama partner to launch [gpt-oss](#)

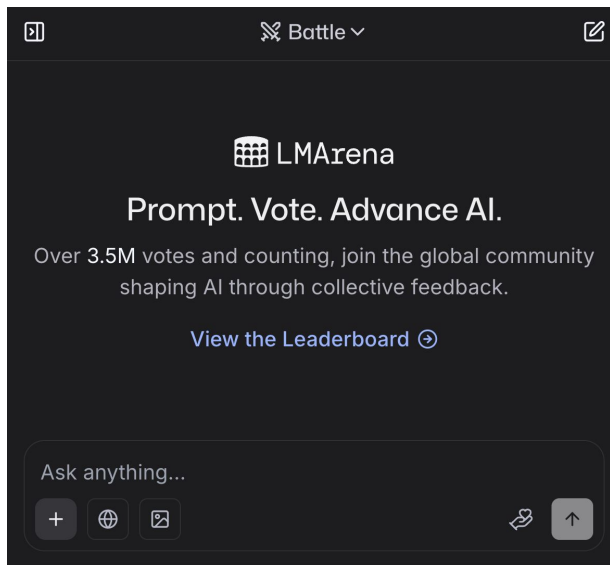
**Chat & build with
open models**

Download

Available for macOS,
Windows, and Linux

How do I know which model(s) is best for what task?

- Chatbot Arena
 - **Free** to try all models
 - You can vote too!



The screenshot shows the LMArena Text Arena leaderboard. The page has a dark theme and a top navigation bar with tabs for Overview, Text, WebDev, Vision, Text-to-Image, Image Edit, Search, and Text-to-V. The 'Text' tab is selected. The main heading is 'Text Arena' with a subtitle 'View rankings across various LLMs on their versatility, linguistic precision, and cultural context across text.' The page shows 'Last Updated Aug 21, 2025', 'Total Votes 827,003', and 'Total Models 233'. A search bar 'Search by model name...' is present. A table lists various tasks and models. The 'Coding' task is selected, showing a list of models with their scores, 95% CI, votes, organizations, and licenses.

Task	Model	Score	95% CI	Votes	Organization	License
Overall	1-20250805-	1509	±14	2,002	Anthropic	Proprietary
	1-20250805	1499	±11	2,961	Anthropic	Proprietary
	20250514-	1480	±9	5,887	Anthropic	Proprietary
	b-instruct-	1476	±11	3,002	Alibaba	Apache 2.0
	1472	±12	2,853	OpenAI	Proprietary	
gemini-2.5-pro	3	1472	±8	9,043	Google	Proprietary
	3	1469	±7	12,237	OpenAI	Proprietary

How can I use proprietary model in an economic way? – API (Application programming interface)

OpenAI Platform

Pricing

Text tokens

Prices per 1M tokens.

Batch

Flex

Standard

Pri

MODEL	INPUT	CACHED INPUT	OUTPUT
gpt-5	\$1.25	\$0.125	\$10.00
gpt-5-mini	\$0.25	\$0.025	\$2.00
gpt-5-nano	\$0.05	\$0.005	\$0.40
gpt-5-chat-latest	\$1.25	\$0.125	\$10.00
gpt-4.1	\$2.00	\$0.50	\$8.00
gpt-4.1-mini	\$0.40	\$0.10	\$1.60
gpt-4.1-nano	\$0.10	\$0.025	\$0.40

AI

Claude Opus 4.1

Most intelligent model for complex tasks

Input	\$15 / MTok
Output	\$75 / MTok
Prompt caching	Write \$18.75 / MTok Read \$1.50 / MTok

Claude Sonnet 4

Optimal balance of intelligence, cost, and speed

Input	
Prompts ≤ 200K tokens	\$3 / MTok
Prompts > 200K tokens	\$6 / MTok
Output	
Prompts ≤ 200K tokens	\$15 / MTok
Prompts > 200K tokens	\$22.50 / MTok

Google Cloud

Contact Us

Start free

PRICING

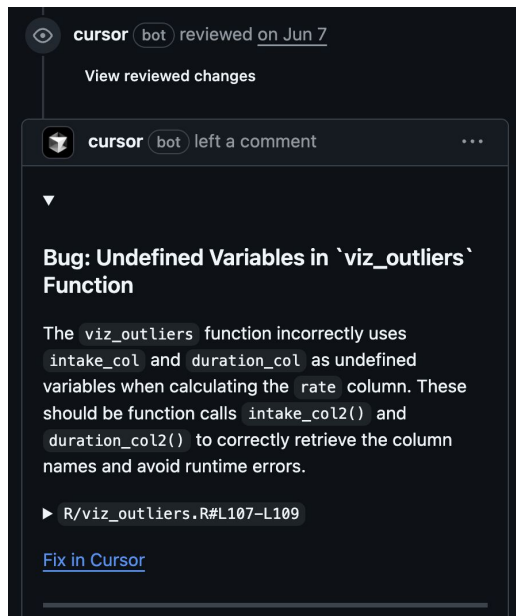
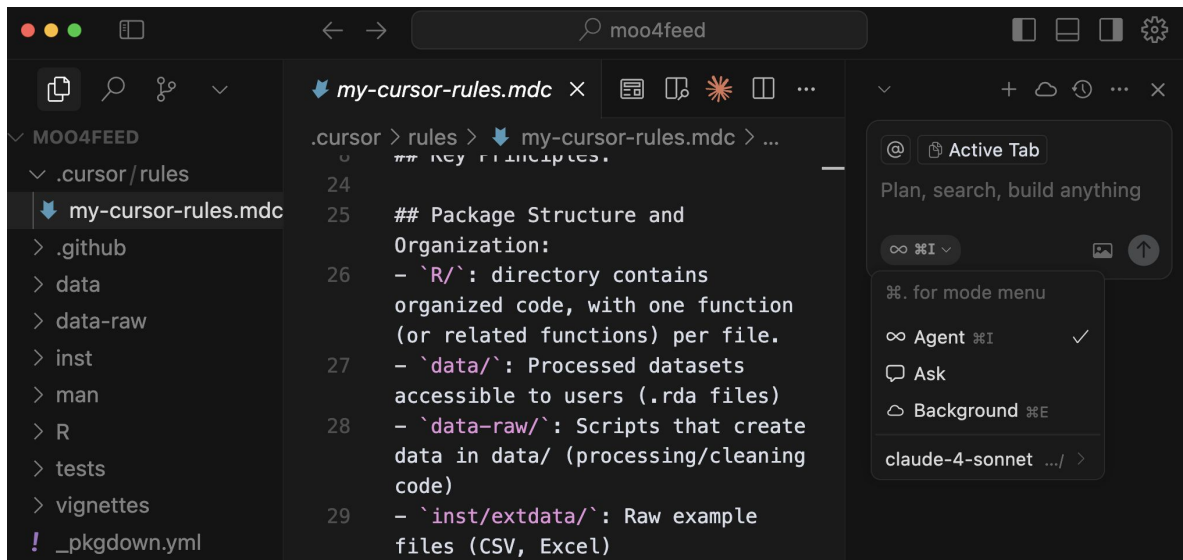
Model	Type	Price (/1M tokens) ≤ 200K input tokens	Price (/1M tokens) > 200K input tokens	Price (/1M tokens) ≤ 200K input tokens with batch API	Price (/1M tokens) > 200K input tokens with batch API
Gemini 2.5 Pro	Input (text, image, video, audio)	\$1.25	\$2.5	\$0.625	\$1.25
	Text output (response and reasoning)	\$10	\$15	\$5	\$7.5
Gemini 2.5 Flash GA	Input (text, image, video)	\$0.30	\$0.30	\$0.15	\$0.15

AI agent for coding (IDE)



CURSOR

- **\$20-200** USD/month
- VS code + LLM of your choice + tools like MCP (Model context protocol)
- Understand your entire codebase & project
- Popular among developers and start-up companies
- Fast for prototyping



AI agent for coding (Command line)

- Claude Code
 - **\$17-200** USD/month
 - This is the best AI agent for coding so far
 - Use it for fun, but **DO NOT** use this for your MDS course work



Other AI agent for coding (Price in UDS/month)

VS code extension

- Augment (\$50)
- Github Copilot (\$10)
- Cline (pay by token)

IDE

- Amazon Kiro (free/waitlist)
- Windsurf (\$15)
- Trae (\$3-10)
- Replit (\$20)

Command line

- Google Gemini CLI (mostly free)
- OpenAI Codex CLI (pay by token)
- Cursor CLI (\$20-200)

AI for reading

NotebookLM

The screenshot displays the NotebookLM web application interface. The top bar shows the document title 'Sick Dairy Cows' Altered Daily Behaviors', a 'Share' button, a 'Settings' button, and a user profile icon. The interface is divided into three main sections: 'Sources', 'Chat', and 'Studio'. The 'Sources' panel on the left lists the document 'Sick Dairy Cows Behavior...' as a selected source. The 'Chat' panel in the center contains a question: 'How do different production diseases specifically alter the daily activity, feeding, and rumination behaviors of dairy cows?'. Below the question, the AI-generated response is visible, discussing 'sickness behavior' and the challenges of observing cow behavior. The 'Studio' panel on the right offers various output formats: 'Audio Overview', 'Video Overview', 'Mind Map', and 'Reports'. At the bottom right, a list of related documents is shown, including 'Contradictory...', 'Decoding Co...', and 'Behavioral Changes i...'. An 'Add note' button is also present.

- Free
- To read the full result:



AI-based search engine



You are an expert software engineer and AI tools researcher. Help me mal

Perplexity Sources · 40 Steps

youtube Augment Code Releases Fair Pricing Details | They...

augmentcode.com Claude Code vs Augment Code

The Lalit Blogs GitHub Copilot Pricing Plans & Reviews & Top 3 ...

GitHub Docs About individual GitHub Copilot plans and benefits

Here is a comprehensive comparison matrix of leading AI coding tools across VS Code extensions, IDEs, and CLI tools. The analysis covers cost-effectiveness, setup complexity, code quality, context awareness, use case fit, deal-breaker limitations, user profile, recent updates, performance, and integration, with current pricing and practical insights for real-world decisions.

AI Coding Tools Comparison Matrix

Tool	Price & Usage	Setup Complexity	Code Quality	Context Awareness	Best For
Augment VS Code	\$50/mo (unlimited dev tier) youtube ; \$60+/mo enterprise augmentcode ; usage scaling,	Moderate; needs repo indexing, agent config youtube	Enterprise-grade, multi-repo, agentic augmentcode	Full-system context engine, 400k+ files augmentcode	Large codebases, advanced teams, agen flows augmentcode

Ask a follow-up...

- **Free** tier allows 3 questions/day
- Pro is **\$20** USD/month unlimited questions
- To read the full result & sources:



Prompt engineering tips

- Ask AI to **think step by step**
- Give AI time to think - "**Take your time** and think carefully before answering"
- Use **structured formats** - Request JSON, numbered lists, or bullet points for answers
- Give AI a **persona** - "You are an expert data scientist with 20 years of experience"
- Provide clear **examples** - Show exactly what format or style you want
- Ask for **explanations** - Don't just get answers, ask about the "why" behind them
- Use fun **analogies** - "Explain MCP using LEGO building blocks"
- Ask for **alternatives** - Get multiple options: "Show me 3 different ways to solve this"
- Set clear **constraints** - "You can only answer in 200 words"
- Be polite - Good manners really do help! Say "**please**" and "**thank you**"
- Be super **specific** - More details = better answers
- **Manage context window**, start new chats for big tasks - AI can get overwhelmed with 20+ messages
- Ask questions in **JSON** format- AI loves structured requests

Key messages

YOU are the **BOSS** of AI! AI is not your boss!

AI can be your assistant, your co-pilot, but
you need to be AI's **guide**, AI's **teacher**

AI agents & tools only shine when you know what
you are doing

Please collaborate with AI in a **responsible** way