

Forum:[Hack Forums](#)

Thread: [any strong ai models used for coding that are unrestricted?](#)

jahseh99 - August 26, 2026, 12:46 pm

any strong ai models used for coding that are unrestricted? hmu

issagi - August 30, 2026, 6:47 pm

Looking for unrestricted coding models? Here's my current list after extensive testing:

Local Models (Ollama):

- FurCode (4B/12B) — Qwen/Mistral-based, uncensored, reliable for code generation
- OmniCoder-9B-Heretic — Qwen3.5-9B, abilitated, trained on 425K+ Claude Opus 4.6 & GPT-5.4 trajectories, 90% pass@5 on AIME 2025
- Qwen2.5-Coder-14B-Heretic — abilitated version, refusals dropped from 100/100 to 3/100, 128K context
- Qwythos-9B — Qwen3.5-9B, 500M tokens of Claude Mythos/Fable traces, native function calling, built for security research
- Ornith-35B — Qwen3.5-MoE 35B, expert-granular abilitation, 0/120 refusals on cybersecurity prompts, requires ~22GB VRAM

Platforms/APIs:

- OpenD AI — local-first CLI agent, streams reasoning live, edits files, runs commands, provider-agnostic
- Venice AI — privacy-focused, uncensored, qwen-2.5-coder-32b available
- Unrestricted Chatbot (GitHub) — self-hosted, zero filters, Ollama integration
- Featherless.ai — hosted uncensored models
- OneAi Freedom — mobile app, multiple uncensored models

Hugging Face:

Search "uncensored" + "coder" or "abilitated" — new models weekly.

My recommendation depends on your hardware. Share your setup and I'll narrow it down.

What's everyone else running? Always looking to add to the list.

ilyass muslim - October 6, 2026, 9:51 am

(Aug 26, 2026 12:46 PM)jahseh99 Wrote: [\[CITE\]](#) any strong ai models used for coding that are unrestricted? hmu

if you have it what do u would To build with this unrestricted ai ?

7 - October 6, 2026, 9:57 am

Im guessing malware and actually there are malware coding tools and defensive and offensive AI LLM frameworks that can do it that are naturally obliterated. I will pull a list of all my offline models and share later, although some are duplicates as they run on multiple machines also some are similar models due to testing token speed and delivery quality.

LINUX — LANGUAGE, CODING, VISION AND CYBERSECURITY

Qwen3-Coder-Next — programming, debugging, code changes, tool use and agentic software development.

Huihui Qwen3-Coder-30B-A3B Instruct Abliterated — community-modified coding model with reduced refusal behaviour.

Cyber-Tiel-Coder-35B-A3B — coding, cybersecurity, penetration-testing assistance, CTFs, technical reasoning and image understanding with compatible vision support.

Qwen3.6-27B — general assistance, reasoning, coding, tool use and image understanding.

Qwen3.8-27B — general assistance, reasoning, coding, agentic tasks, image and video understanding.

Huihui Qwen3.8-27B Abliterated — community-modified general assistance, reasoning and coding model, with vision where supported.

Qwen3.8-9B Heretic — community derivative for general assistance, writing and reasoning; exact variant to confirm.

MiMo-V2.6-Distill-Qwen-9B — distilled model for general assistance, reasoning and coding.

Gemma 4 26B A4B Abliterix — community-modified model for chat, writing, reasoning and image understanding where supported.

LFM2.5-2.6B Heretic — lightweight community-modified model for conversation and assistant tasks.

Llama 3 8B Lexi Uncensored — conversation, general assistance and creative writing.

Sicarius Llama 8B — community Llama derivative; full checkpoint name needed to confirm its specialisation.

Imperum-CybersecurityLLM-v1.0 — defensive cybersecurity, alert triage, incident response, threat intelligence, detection engineering, malware analysis and secure code review.

SenecaLLM × Qwen3-32B — cybersecurity-focused language model, also published under Trendyol-Cybersecurity-LLM-Qwen3-32B.

Huihui Qwen3-VL-4B — image interpretation, screenshots, document understanding and OCR-style tasks.

TranslateGemma-12B — language translation.

Qwen3.8-Flash-Next — general assistance, coding, reasoning, tool use and multimodal understanding.

Qwen3.6-35B-A3B Fable / Holo3.1 / Qwopus / KAT-Coder merge — experimental community merge for conversation, reasoning and coding.

MAC / MLX — LANGUAGE, CODING, VISION AND CYBERSECURITY

Qwen3.8-Flash-Next — coding, reasoning, general assistance, tool use and multimodal understanding.

Qwen3.6-35B-A3B — general assistance, reasoning, coding and image understanding.

Qwen3-Coder-Next — programming, debugging, code changes and agentic software development.
Cyber-Tiel-Coder-35B-A3B, oQ4e and oQ6e — quantized variants for coding, cybersecurity, penetration-testing assistance, CTFs and technical reasoning, with vision where supported.

Qwen3.6-35B-A3B Fable / Holo3.1 / Qwopus / KAT-Coder merge — experimental community merge for conversation, reasoning and coding.

DeepSeek-R1-Distill-Qwen-32B, 8-bit — reasoning, mathematics, analysis and coding.

DeepSeek-R1-Distill-Qwen-32B Abliterated, 4-bit — community-modified reasoning model with reduced refusal behaviour.

Gemma 4 26B A4B Instruct — general assistance, reasoning, writing and visual understanding.

Gemma 4 31B Instruct Claude-Opus-Distill — community Gemma fine-tune described as distilled from Claude Opus outputs.

Qwen3-42B-A3B TOTAL-RECALL / MASTER-CODER — community model family for coding and reasoning; exact Thinking, Instruct or Coder variant to confirm.

Qwen3.8-27B OBLITERATED — community-modified general assistance, reasoning and coding model, with vision dependent on the conversion.

IMAGES — LINUX

Z-Image-Turbo — fast, low-step image generation.

Z-Image Base — undistilled image generation with broader guidance control for photography, illustration and stylized imagery.

FLUX.2 Klein 4B — image generation and reference-image editing.

Qwen-Image-Edit-2511 — instruction-based image editing and reference-based modifications.

Juggernaut XIII Ragnarok — photorealism-focused SDXL checkpoint for portraits, photographic scenes, cinematic imagery and creative artwork.

epiCRealism XL — photorealism-oriented image generation; exact release to confirm.

Nova Reality XL — image-generation checkpoint; exact version and underlying model family to confirm.

Pony Diffusion V6 XL — character-focused generation, anime, cartoons, illustration and stylized imagery.

IMAGES — MAC / MLX

FLUX.2 Klein 4B — image generation and reference-image editing.

FLUX.2 Klein 9B, iQ4.5 and 4-bit — quantized variants of the larger image-generation and editing model.

Qwen-Image-2.1 — image generation, image editing, transparent RGBA images, subject extraction and reference-based edits.

Krea-2-Turbo — distilled text-to-image model for fast creative image generation.

VIDEO AND MUSIC

MiniMax-H3 — video generation with synchronized audio; listed for Linux and Mac.

LTX-2.5 — joint video and audio generation; listed for Mac.

ACE-Step 1.5 XL Turbo — music generation; listed for Mac.

MiniMax-Music3 — music and song generation; listed for Mac.

SPEECH AND AUDIO

Qwen3-TTS-12Hz-0.6B-Base — speech synthesis and reference-voice cloning; Linux and Mac.

Qwen3-TTS-12Hz-1.7B-Base — larger speech-synthesis and reference-voice-cloning model; Mac.

Qwen3-TTS-12Hz-1.7B-VoiceDesign — creates voices from natural-language descriptions; Linux.

Chatterbox Turbo — speech synthesis and reference-voice cloning.

Kokoro — lightweight text-to-speech; Linux and Mac.

Whisper large-v3-turbo — multilingual speech recognition and transcription.

Whisper base.en — lightweight English-only speech recognition and transcription.

Whisper small through faster-whisper — speech transcription using the faster-whisper inference implementation.

Vosk small English — lightweight offline English speech recognition and phrase-detection workflows.

Vosk speaker model — extracts speaker embeddings for speaker comparison and identification workflows.

SEARCH AND MODEL HELPERS

Nomic Embed Text v1.5 — produces text embeddings for semantic search and retrieval.

BGE-Reranker-v2-M3 — scores and reranks retrieved passages by relevance.

Qwen3.6 DFlash drafter — accelerates generation through speculative decoding with its matching target model.

Gemma 4 Assistant drafter — speculative-decoding helper for the matching Gemma checkpoint.

Qwen and Cyber-Tiel MTP components — multi-token-prediction helpers for faster decoding in compatible runtimes.

Vision projectors — connect visual representations to compatible language models.

Text encoders — encode prompts for image, video and other generation pipelines.

VAEs — encode and decode images, video or audio through compressed latent representations.

LoRAs — adapt models for styles, subjects, tasks or behaviour; specific acceleration LoRAs support faster generation.

Upscalers — increase image or video resolution and reconstruct detail.

FRAMEWORKS AND TOOLS

MLX — machine-learning and array-computing framework for Apple Silicon.

MLX-Serve — model serving built around MLX; exact project to confirm.

LM Studio — local model management, chat and API serving.

llama.cpp — language-model inference runtime with GGUF support and supported multimodal extensions.

ROCm — AMD GPU computing software stack.

Vulkan — cross-platform graphics and compute API used by compatible inference backends.

PyTorch — machine-learning framework.

Transformers — model loading, inference, training and language-model tooling.

Diffusers — diffusion-model tooling and image, video and audio generation pipelines.

Accelerate — device placement and distributed execution utilities.

Triton — GPU-kernel programming language and compiler.

ONNX Runtime — inference runtime for exported ONNX models.

Safetensors — model-weight and tensor storage format.

stable-diffusion.cpp — C/C++ diffusion-model inference runtime.

ComfyUI — node-based image, video and audio workflow application.

ComfyUI-GGUF — GGUF loading support for compatible ComfyUI models and workflows.

ComfyUI Easy Use — workflow convenience nodes.

Stability Matrix — manages supported image-generation applications and models.

Stable Diffusion WebUI — browser-based image-generation interface; exact implementation to confirm.

local-agent — local wrapper connecting chat models to tools and skills.

local-image — local command-line image-generation wrapper.

local-browser — local browser-automation wrapper.

Local vision, translation and document-search helpers — connect models to practical tasks.

Website production and media workers — coordinate development, asset generation and checks.

This is just the local models... some I should remove they have already been replaced.