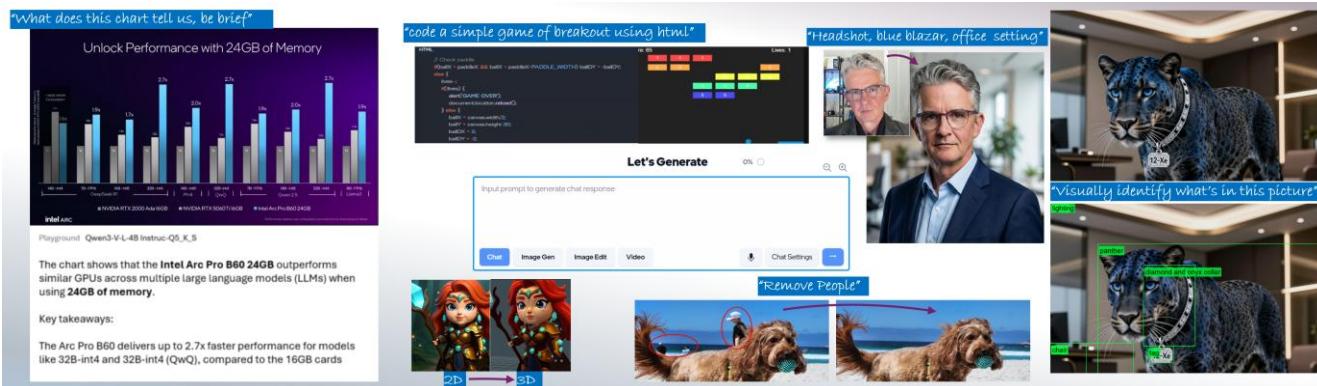


# AI PLAYGROUND 3.0 Users Guide. v3.0.0a

Welcome to AI Playground v.3.0 alpha. This application provides a full suite of generative AI features for chat, code assistance, document search, image analysis, image and video generation. All features run offline and are powered by your PC's Intel® Core™ Ultra with built-in Intel Arc GPU or Intel Arc™ dGPU Series A or B with 8GB+ of vRAM.



## Quick Links

- [Installing AI Playground](#)
- [Overview of AI Playground](#)
- [New in AI Playground 3.0a](#)
- [Installing Models](#)
- [User Interface](#)
- [Generative AI Presets](#)
- [Installing Other Models](#)
- [Using ComfyUI](#)
- [Model Index](#)

## Minimum Specifications

AI Playground is currently available as a packaged installer, or available as source code from our Github repository. To run AI Playground, you must have a PC that meets the following specifications

- Windows OS
- Intel Core Ultra Series 1H, Series 2H, Series 2V, or Series 3H processor OR Intel Arc GPU (discrete) Series A or B with 8GB of vRAM
- 82GB+ of Hard Disk Space: AI Playground application takes up roughly 6GB of space. Models will take up additional space.

## Installing AI Playground

**Packaged Installer:** AI Playground 3.0 has a single installer for all currently supported hardware. The installer will create the runtime application needed to run AI Playground and to install backend component and models. After running the installer, you will be asked to launch AI Playground which, at first run, will launch the installation manager to install all needed backend components such as PyTorch, OpenVINO, ComfyUI and Llama.cpp.

**Backend Installation Manager:** You will be presented with installation options at first run of the application. Installation is a simple mouse click operation allowing you to install and manage all backend components without using a command line interface. This portion of the app does require an open connection to the internet. *NOTE: Running this installation from inside a corporate network or from a home connection while connected to corporate VPN connection will likely cause it to fail. For this portion, be sure to have a clear and unfiltered Internet connection*

AI Playground and OpenVINO are automatically installed components for installation. ComfyUI and Llama.cpp are not required but provided key services. If these features are not installed, you may be asked to install them when trying certain features. ComfyUI is the backend of all image and video generation features. Llama.CPP provides access to latest GGUF models like GPT-OSS 20B.

The gear icons to the right of each backend component allows you to repair, install and set the version of installation. If desire to change or set the version of Llama.cpp, ComfyUI or OpenVINO Gen AI, you can do this here. Intel has chosen these version settings to create the best experience with the features for a given release. Changing the version may create instability or break some features.

## Overview of AI Playground 3.0

AI Playground is a locally run Generative AI application suite that acts as an offline alternative to cloud tools such Gemini ChatGPT and Grok. AI Playground is intended to give consumers and AI curious prosuser easy and intuitive access to a wide variety of generative AI features using their Intel powered AI PC. This means you can be offline, without loading sensitive or personal data to 3<sup>rd</sup> party sites, for free, in a single app without having to know how to install and manage multiple AI backend frameworks. **Key features:**

- Latest and greatest chat models:** Support for Qwen 3 VL, Mistral 7B, DeepSeek R1 or GPT-OSS, AI playground makes a variety of chat models available to users
- Vision, Reasoning and RAG:** Chat features support Vision, Reasoning and RAG to analyze and get deep answers on both visual and text content

Analyze images with Qwen3 VL Model	Vibe Coding with GPT-OSS 20B Reasoning	Document RAG with Mistral 7B Instruct
<p>Can you explain this?</p> <p>Playground Qwen3-VL-4B-Instruct-Q5_K_5</p> <p>The expression you've provided is a mathematical calculation: <math>6 + 4(7 - 2) - 3^2</math></p> <p>Let's break it down step by step using the order of operations</p>	<p>Done Reasoning after 1.4 seconds ~</p> <p>Below is a complete, self-contained "Breakout" clone that you can drop into a single HTML file and run in any modern browser.</p> <p>It uses only vanilla HTML, CSS, and JavaScript – no external libraries.</p> <p>Feel free to copy-paste the code into a file called breakout.html and open it in your browser.</p> <pre>&lt;!DOCTYPE html&gt; &lt;html lang="en"&gt; &lt;head&gt;     &lt;meta charset="UTF-8"&gt;</pre>	<p>What image generation features and models does AI Playgrounds support</p> <p>Playground Mistral-7B-Instruct-v3.0-Q4_K_5</p> <p>Sources: AI Playground Users Guide.pdf (Page 1, Lines 1-24) AI Playground Users Guide.pdf (Page 1, Lines 41-44) AI Playground Users Guide.pdf (Page 2, Lines 1-8) AI Playground Users Guide.pdf (Page 4, Lines 29-34) AI Playground Users Guide.pdf (Page 6, Lines 1-7) AI Playground Users Guide.pdf (Page 5, Lines 46-55)</p> <p>AI Playground supports image generation through the 'comfyUI' tool. This tool allows for creating, editing, or enhancing media content (images) based on text prompts. The tool can generate images with various styles, lighting, colors, and quality tags. AI Playground supports models from Huggingface for both AI chat and image generation. For image generation, it currently</p>

- Image Generation:** From Stable Diffusion 1.5, SDXL, Flux.1 and Z-image models AI Playground is making a breadth of image generation from quick easy low-res draft generation to high quality image generation
- Image Editing:** Subscription free and private control for upscaling, inpainting, outpainting, 2D to 3D mesh or editing images in a variety of ways. Good for editing personal photos to taking sketches and generated images to the next level with greater control.



- Video Generation:** (the early pre-alpha version of AI Playground may have this feature all or partially disabled, but this feature will return in the beta release)
- Intelligence Built-In:** AI Playground uses AI to make the app easier and more efficient to use. From Agentic AI tool calling, allowing you to access and mix different modalities across different backends and models, to AI voice integration, allowing for speech input and reducing prompt typing fatigue.
- Backend Manager:** AI Playground makes installing and managing multiple AI backends effortlessly by installing and managing OpenVINO, PyTorch, Llama.cpp and ComfyUI for use with various models and features. AI Playground also provides version control for any of these components allowing you to adjust versions and reinstall them as needed between AI Playground releases.
- Open Source:** AI Playground is a free to download and use application. It is open source allowing developers to contribute and fork it

## New To AI Playground 3.0

### New Modalities

- Voice mode:** AI Playground now supports a voice mode. From the app settings you can turn on Voice mode. This will load an AI voice model to the NPU of your Intel Core Ultra PC. When enabled a microphone icon will be available in the prompt. Simple click the icon and voice your prompt. Voice mode will convert your voice to text prompt
- Vision Model Support:** New to AI Playground is Vision Model support. With Vision support you can paste or load an image to Chat mode and have AI Playground analyze the image, giving you insight, answers and information about the image
- Agentic Multi-Modal Tools.** AI Playground exposes MCP tooling to our chat mode allowing you to use a core model to do agentic tasks across AI Playground features within a single chat discussion. This feature will allow you to use text chat, vision chat, and image generation withing a single discussion window. See Agentic Preset for more information

**Unified Prompt:** AI Playground now deploys a single unified prompt window for all AI Playground features. This design aligns AI Playground to familiar AI tools like Gemini, Chat GPT and Grok

**Light Theme:** You can now choose to experience AI Playground in a light-colored theme

**Prompt Modes, Settings and Presets:** The main prompt can now be directed to specific models and tasks through the Prompt Mode, Prompt Settings and Prompt Presets.

- Prompt Modes** allow you to quickly and easily change between modalities from the main prompt: Select chat for vision and text based chat responses, Image Gen for image generation from text. Image Edit for altering or enhancing images. Video Gen: For creating video content from text or image input
- Prompt Mode Presets:** From the Prompt Settings button you have access to dozens of Presets. Presets are similar to the previous AI Playground "Workflows" Settings, configured for specific tasks and usages but are now expanded to all Prompt modes. With Presets you can quickly access a define usage for Chat, Image Gen, Image Edit or Video where the model selections and settings are preset for that task. See the Preset Section for more information

## Installing Models

**Installing Models:** AI Playground **does not ship with Generative AI models**, however we have made the model process easy, and flexible for you to download and install models for any function of AI Playground.

- Automated Model Downloads:** When you first run any task on AI Playground requiring a model that is not present, AI Playground will provide an option to download a model for that task. To download you must first visit the site for the model, then confirm you acknowledge the terms of use for the model. After downloading the process to run the task will continue. Alternatively, you can cancel and install a model of your choice.
- Add Model Icon:** In our chat presets you have the ability to manually download models of your choosing, making them available for that preset. To do this, go to the Preset you wish to use the model. Click the Add Model icon. Put in the hugging face ID and model path for the model. The format will typically be EntityName/ParentModel/ModelFileName

## User Interface Overview:

**Unified Prompt:** AI Playground no longer has Tabs for each modality. We instead use a single prompt experience for any type of content and modality. This experience aligns to popular online AI services, where a single and simple UI can be used for any feature and generative output. Simply type a prompt and generate!

The UI functional breakdown areas sections

- Heading:** Minimize and Maximize or Close AI Playground.
- Footer** The bottom portion of the app which includes information the version, and links to terms and disclosures. This section can now be hidden or shows to maximize your screen space
- Prompt Viewport:** This is the central mid-section of the app that contains the prompt. This is the single prompt tool for all features. New mode buttons are available, which provide context switching for both the history and settings panels
- History Panel:** The history panel, hidden by default, can be made visible by clicking History in the upper left. The history is in context to the selected mode and will switch content depending which mode is selected. History items will have a submenu for additional information and actions for those items
- Prompt Settings Panel:** The Prompt Settings panel will show on the right whenever the Settings button from the Prompt field is clicked. The panel will provide access to prompt presets and their adjustable settings

- App Settings:** The app settings is selectable from the gear icon on the lower left above the Footer. When selected it will take the left panel space. From app settings you can access the Backend installation manager, Voice mode option, language settings and more

Resizing can be done using the CTRL + mouse wheel to zoom in out. As well as hot keys. Zoom: CTRL+, Zoom out: CTRL-, Reset Zoom: CTRL0

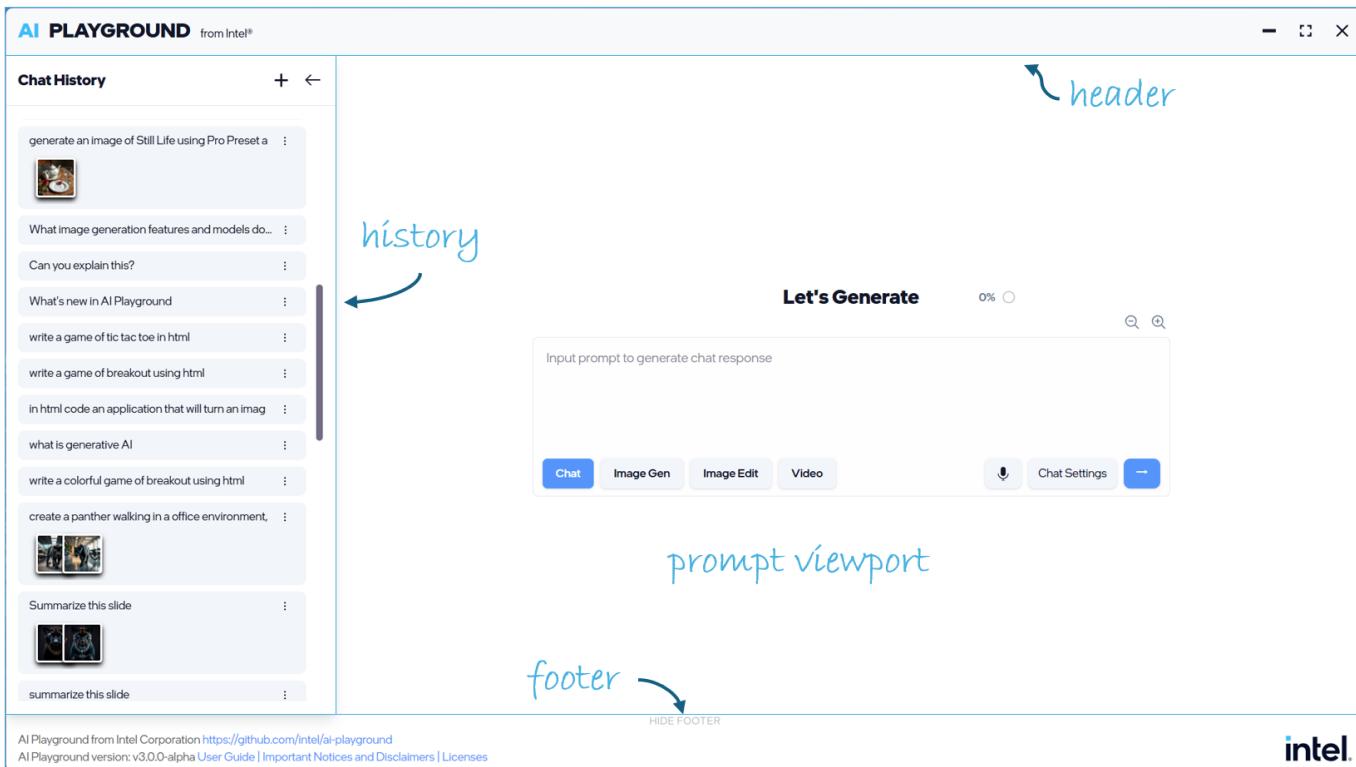


Figure 1: Screenshot of user interface with the light color theme and Chat History Panel and Footer Panel exposed

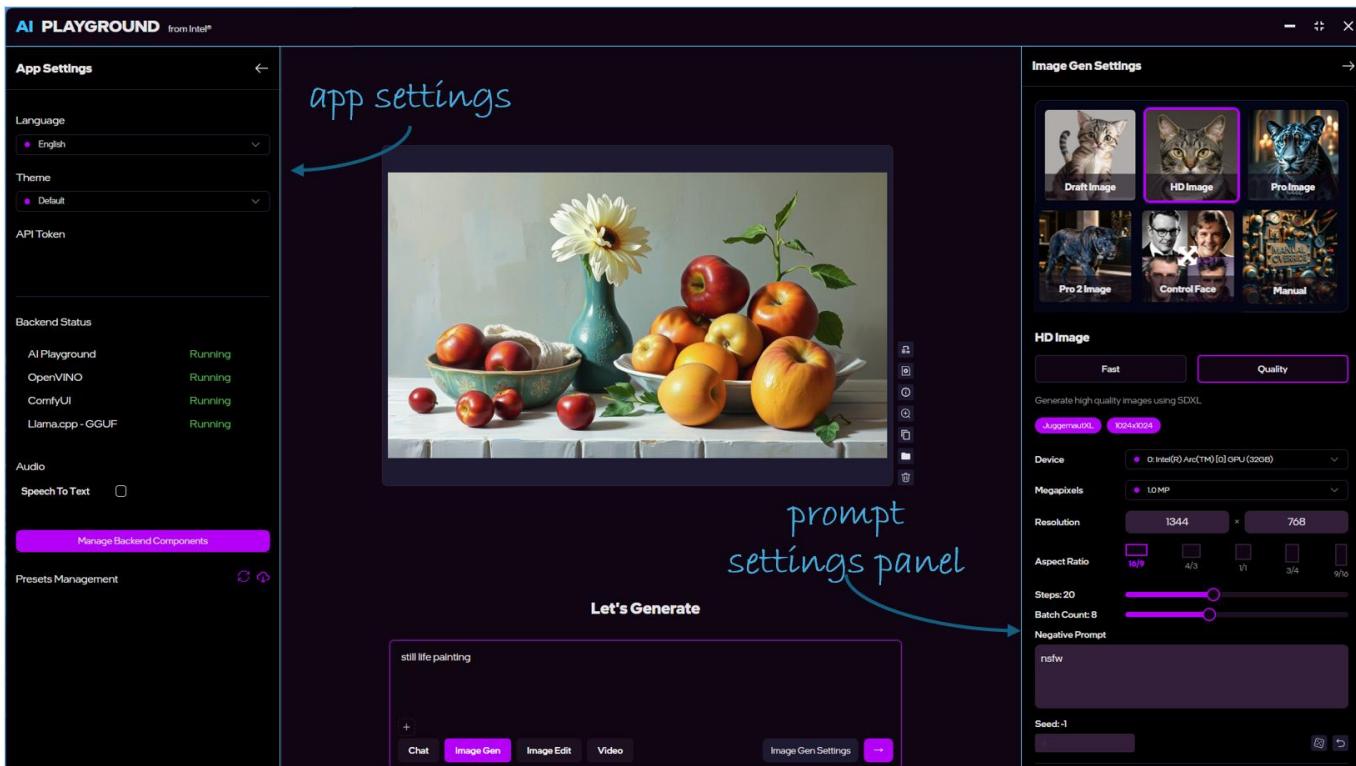
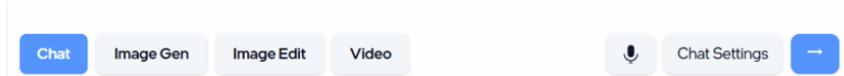


Figure 2: Screenshot of user interface using the default color theme, with the App Settings Panel and Prompt Settings Panel exposed

## Using the new interface:

Using AI Playground is simple. Simply type a request in the main prompt and click the generate arrow button. When you are in a chat discussion, all of the history of this discussion is used along with your prompt. This way the AI remembers what you've been discussing. Below are ways you can use this prompt tool to control the output.

**Mode And Preset Selecting:** Select a Mode, either Chat, Image Gen, Image Edit or Video Gen. Then click the mode settings near the generate button to browse Presets for use. For many targeted usages from Vision Chat to Inpainting, we have Presets where models, settings and input options for dozens of usages are predefined. After selecting, type a prompt and go. See Generative AI Presets below



**Agentic Multi-Modal:** This approach allows you to use both chat and image generation in a single discussion. To do this, select the Chat mode from the Prompt area. Then Select Chat Settings, and choose the Agentic Preset: By default, the Qwen 3 VL model will be used. You can now do the following

- Paste or click the + icon to add images and ask questions about the image
- Click the + icon to add documents to search and analyze content from documents on your PC
- In the prompt ask to generate an image where you describe what you want an image of and you define if the image is Draft, HD or Pro. You can request how many images to generate.
- Click New Discussion from the History panel to start a new discussion without existing context and history

**Voice Mode:** Select the Settings gear to open the App Settings Panel. Select “Speed To Text. At first run a model option will be provided to download. After it is installed and checked on you can now click the microphone in the prompt area to turn on listening mode. Speak your prompt then click the microphone when done.

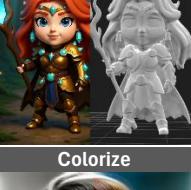
## Generative AI Presets:

Presets are where you drill into specific usage and capabilities of generative AI. Presets are predefined sets of curated, tuned, and tested models and settings for doing specific tasks and workflows in AI Playground. To use a preset first select a mode from the Prompt, then click that modes Settings button, just to the left of the generate button. Browse preset options, select one then adjust settings for your desired use. See table below for list of presets and descriptions for each Prompt Mode

	Preset	Description	Preset Specific Features	Models
Chat	Basic Chat	Choose between OpenVINO or GGUF models (via Llama.cpp) for general knowledge search	<ul style="list-style-type: none"> <li>• Device selector to target processor</li> <li>• Backend selector for choosing OpenVINO or GGUF (Llama.cpp) models</li> <li>• Model selected filtered for Device and Backend for basic text to chat</li> <li>• Max Tokens and Max Context input</li> <li>• Temperature input</li> <li>• Show Metrics toggle</li> </ul>	GGUF: Llama-3.2-3B-Instruct-Q4_K_S.gguf, Llama-3.2-3B-Instruct-Q8_0, Meta-Llama-3.1-8B-Instruct-Q5_K_S, Mistral-7B-Instruct-v0.3-Q4_K_S, Qwen3-4B-Instruct-2507-Q5_K_S.gguf  OpenVINO: Phi-3.5-mini-instruct-int4-ov, Phi-3-mini-4k-instruct-int4-ov, Mistral-7B-Instruct-v0.2-int4-ov, TinyLlama-1.1B-Chat-v1.0-int4-ov
	NPU Chat	Run an LLM on the AI Boost accelerator, for the best in power efficiency	<ul style="list-style-type: none"> <li>• OpenVINO framework set by default.</li> <li>• NPU AI Boost processor tile is set by default</li> <li>• Models filtered to OpenVINO NPU models</li> </ul>	Phi-3-mini-4k-instruct-int4-cw-ov Phi-3.5-mini-instruct-int4-cw-ov Mistral-7B-Instruct-v0.2-int4-cw-ov Mistral-7B-Instruct-v0.3-int4-cw-ov Mistral-7B-Instruct-v0.3-int4-cw-ov
	Chat with RAG	Easily find and analyze content from text and PDF documents	Basic Chat Features PLUS <ul style="list-style-type: none"> <li>• Document loader and manager available</li> <li>• Embedded model selected available</li> </ul>	GGUF: Llama-3.2-3B-Instruct-Q4_K_S.gguf, Llama-3.2-3B-Instruct-Q8_0, Meta-Llama-3.1-8B-Instruct-Q5_K_S, Mistral-7B-Instruct-v0.3-Q4_K_S, Qwen3-4B-Instruct-2507-Q5_K_S.gguf  OpenVINO: Phi-3.5-mini-instruct-int4-ov, Phi-3-mini-4k-instruct-int4-ov, Mistral-7B-Instruct-v0.2-int4-ov, TinyLlama-1.1B-Chat-v1.0-int4-ov
	Reasoning	Use large reasoning models like DeepSeek R1 Distilled, and GPT-OSS 20B to get deeper and more accurate responses	Basic Chat PLUS <ul style="list-style-type: none"> <li>• Model list filtered to Reasoning Models</li> <li>• Document loader and manager available</li> <li>• Embedded model selected available</li> </ul>	GGUF: DeepSeek-R1-Distill-Qwen-1.5B-Q4_K_S, DeepSeek-R1-Distill-Qwen-7B-Q8_0, gpt-oss-20b-Q8_0, Qwen3-4B-Q5_K_S, OpenVINO: DeepSeek-R1-Distill-Qwen-1.5B-int4-ov, DeepSeek-R1-Distill-Qwen-7B-int4-ov, Qwen3-4B-int4-ov
	Vision	Use Vision Language Models in GGUF Format with image understanding. Load images and ask questions about the image	Basic Chat PLUS <ul style="list-style-type: none"> <li>• Model list filtered to Vision capable models</li> <li>• Document loader and manager available</li> <li>• Embedded model selected available</li> </ul>	GGUF: gemma-3-4b-it-Q5_K_S.gguf, Qwen3-VL-4B-Instruct-Q5_K_S.gguf  OpenVINO: InternVL2-4B-int4-ov, Qwen2-VL-7B-Instruct-int4-ov
	Agentic	Using agentic tool calling where you mix vision, text and image modalities in the same discussion	Basic Chat PLUS <ul style="list-style-type: none"> <li>• Model list filtered to Tool calling models</li> <li>• Enable Tool calling checkbox available</li> <li>• Document loader and manager available</li> <li>• Embedded model selected available</li> </ul>	GGUF: Llama-3.2-3B-Instruct-Q4_K_S, Llama-3.2-3B-Instruct-Q8_0, Meta-Llama-3.1-8B-Instruct-Q5_K_S, Mistral-7B-Instruct-v0.3-Q4_K_S, gpt-oss-20b-Q8_0.gguf, gemma-3-4b-it-Q5_K_S, Qwen3-VL-4B-Instruct-Q5_K_S, Qwen3-4B-Q5_K_S, Qwen3-4B-Instruct-2507-Q5_K_S.gguf  OpenVINO: Qwen3-4B-int4-ov, Mistral-7B-Instruct-v0.3-int4-cw-ov, Qwen2-VL-7B-Instruct-int4-ov
	Advanced Chat	Manually select a backend, target device and models	All features accessible PLUS <ul style="list-style-type: none"> <li>• System Prompt Editor</li> </ul>	All

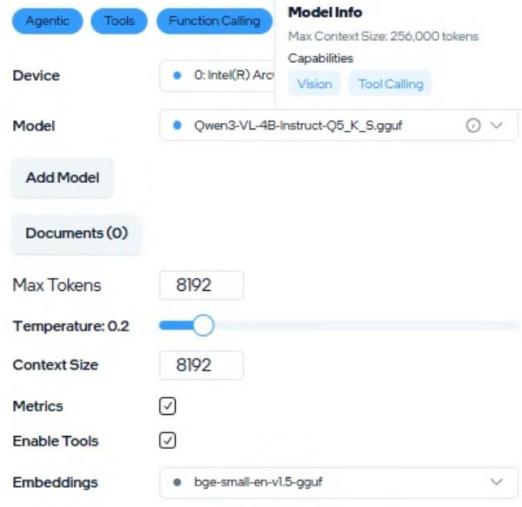
## Presets Continued

Preset	Description	Preset Specific Features	Models
Image Gen	<p><b>Draft Image</b></p>  <p>Rapidly generate images using Stable Diffusion 1.5 up to 720x720 pixels [Fast and Standard]</p>	<ul style="list-style-type: none"> <li>Variant Selector – Fast or Standard mode</li> <li>Device selector to target processor</li> <li>Megapixel – for 0.25 and 0.5 megapixel</li> <li>Aspect ratio to set the shape/ratio of the image</li> <li>Resolution – display &amp; manually set resolution</li> <li>Steps: Manually enter generation steps</li> <li>Batch Count: # of images to generate</li> <li>Negative Prompt: Enter things not to see</li> <li>Seed: Set seed, randomly generate</li> <li>Show Preview: ON as default</li> <li>Safety Checker: Moderation set to 0.5</li> </ul>	Dreamshaper 8 (SD1.5)
	<p><b>HD Image</b></p>  <p>Generate high quality images using SDXL up to 1024x1024 pixels [Fast and Standard]</p>	<p>Same As Draft Except:</p> <ul style="list-style-type: none"> <li>Variant Selector – Fast or Quality mode</li> <li>Megapixel – 0.25, 0.5, 0.8, 1.0 megapixel</li> </ul>	JuggernautXL (SDXL)
	<p><b>Pro Image</b></p>  <p>Precisely generate images, with text and adherence to prompt using Flux.1-Schnell and Z-image Turbo at 1 to 1.5 Megapixel [Standard and Quality]</p>	<p>Same As Draft Except:</p> <ul style="list-style-type: none"> <li>Variant Selector – Standard or Quality mode</li> <li>Megapixel – 0.25, 0.5, 0.8, 1.0, 1.2, 1.5 megapixel</li> <li>No Negative Prompt</li> </ul>	Flux.1-Schnell Q4, Q8
	<p><b>Pro 2 Image</b></p>  <p>Generate extremely high quality images using the Z-image-turbo model</p>	<p>Same As Draft Except:</p> <ul style="list-style-type: none"> <li>No Variants Selector</li> <li>Megapixel – 0.25, 0.5, 0.8, 1.0 megapixel</li> <li>No Negative Prompt</li> </ul>	Z-Image-Turbo
	<p><b>Control Face</b></p>  <p>Generate HD images where you control its likeness. Load an image with a face to guide the likeness, then describe the character to generate</p>	<p>Same As Draft Except:</p> <ul style="list-style-type: none"> <li>No Variants Selector</li> <li>Megapixel – 0.25, 0.5, 0.8, 1.0 megapixel</li> <li>No Negative Prompt</li> <li>Active FaceSwap checkbox</li> <li>Load Image input</li> </ul>	JuggernautXL (SDXL), Inswapper
	<p><b>Manual</b></p>  <p>Advanced mode to manually control the model, CFG, sampler and more. But be sure resolution and settings fit the model selected</p>	<p>Same As Draft Except:</p> <ul style="list-style-type: none"> <li>Variant Selector – Fast or Quality mode</li> <li>Megapixel – 0.25, 0.5, 0.8, 1.0 megapixel</li> <li>Model Selector</li> <li>Sampler Selector</li> <li>Scheduler Selector</li> <li>CFG Scale Slider</li> </ul>	User defined

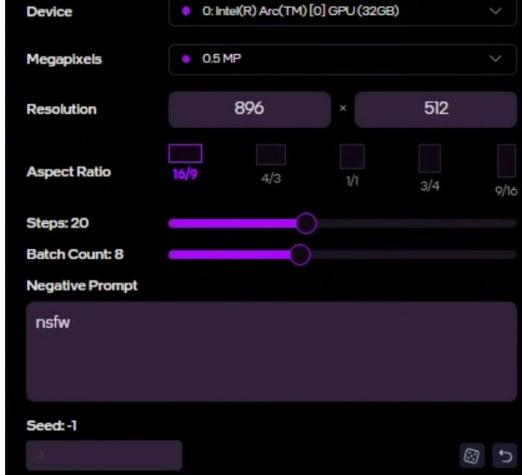
Preset	Description	Preset Specific Features	Models
<b>Upscale</b> 	Load an image and set a scale up to 4X to increase the resolution of that image	<ul style="list-style-type: none"> <li>• Device selector to target processor</li> <li>• Load Image</li> <li>• Upscale Factor Slider</li> </ul>	ESRGAN
<b>Edit By Prompt</b> 	Add, remove, edit, or combine elements in photos by describing what to change [Fast and Standard]	<ul style="list-style-type: none"> <li>• Variants: Fast, Quality</li> <li>• Device selector to target processor</li> <li>• Megapixel – for 0.25, 0.5, 0.8, 1.0 megapixel</li> <li>• Aspect ratio to set the shape/ratio of the image</li> <li>• Resolution – display &amp; manually set resolution</li> <li>• Steps: Manually enter generation steps</li> <li>• Seed: Set seed, randomly generate</li> <li>• Use Resolution of Reference Checkbox</li> <li>• Load Image (Left)</li> <li>• Use Second Reference Checkbox</li> <li>• Load Image (Right)</li> </ul>	Flux.1-Kontext
<b>Sketch To Photo</b> 	Turn rough sketches into stunning photos. Load a sketch and describe the location, weather and environment [Fast and Standard]	<p>Same as EditByPrompt Except</p> <ul style="list-style-type: none"> <li>• Batch Count Slider</li> <li>• Canny Low Setting Slider</li> <li>• Canny High Setting Slider</li> <li>• Control Net Strength</li> <li>• Load Reference Image (one image only)</li> <li>• </li> </ul>	JuggernautXL (SDXL), Canny-ControlNet
<b>Change Face</b> 	Edit an image to have a likeness you control. Load image with a face you want to change, then load a reference face to guide the likeness	<p>Same as EditByPrompt Except</p> <ul style="list-style-type: none"> <li>• No resolution, aspect ratio settings, steps, or Seed</li> <li>• Load Image (Original)</li> <li>• Load Image (New Face)</li> </ul>	Inswapper
<b>Inpaint</b> 	Regenerate portions of an image by masking an area that will change based on your prompt [Draft and HD]	<p>Same as EditByPrompt Except</p> <ul style="list-style-type: none"> <li>• Variants: Draft, HD modes</li> <li>• Batch Count Slider</li> <li>• Negative Prompt</li> <li>• Show Preview: ON as default</li> <li>• Safety Checker: Moderation set to 0.5</li> <li>• Load Image (Original)</li> <li>• Mask editor, Inpainting</li> </ul>	SD1.5_inpaint, JuggernautXL-inpaint
<b>Outpaint</b> 	Extend an image beyond its borders by sizing and positioning an image in a new space, then generate new content around it [Draft and HD]	<p>Same as EditByPrompt Except</p> <ul style="list-style-type: none"> <li>• Variants: Draft, HD modes</li> <li>• Batch Count Slider</li> <li>• Negative Prompt</li> <li>• Show Preview: ON as default</li> <li>• Safety Checker: Moderation set to 0.5</li> <li>• Load Image (Original)</li> <li>• Mask editor, Outpainting</li> </ul>	SD1.5_inpaint, JuggernautXL-inpaint
<b>Image To 3D Model</b> 	Create a 3D mesh from a 2D image ready for 3D tools like Blender [Preview and Standard]	<p>Same as EditByPrompt Except</p> <ul style="list-style-type: none"> <li>• Variants: Preview, Standard</li> <li>• No resolution, aspect ratio settings, steps, or Seed</li> <li>• Load Image (Reference)</li> <li>• Number of Chunks</li> <li>• Octree Slider</li> <li>• CFG Scale</li> <li>• Voxel Algorithm</li> </ul>	Hunyuan3d 2.0
<b>Colorize</b> 	Add color to a black and white photo	<p>Same as EditByPrompt Except</p> <ul style="list-style-type: none"> <li>• No resolution, aspect ratio settings, steps, or Seed</li> <li>• Model Selector (Colorize)</li> <li>• Load Image (Original)</li> </ul>	

## Settings Explained:

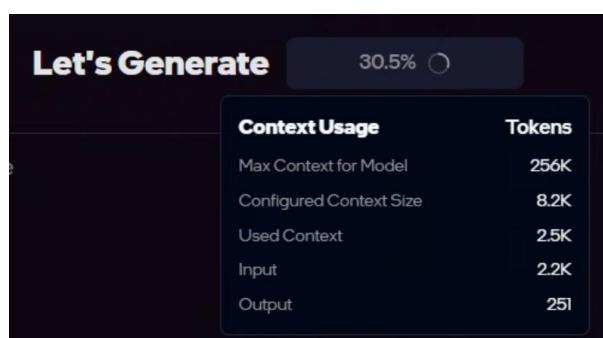
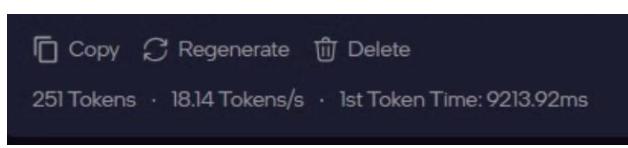
**Common Chat Settings:** User adjustable settings for chat are accessible by clicking the Chat Settings button in the Prompt area. Some presets have specific settings and default values. Below is a description of these settings

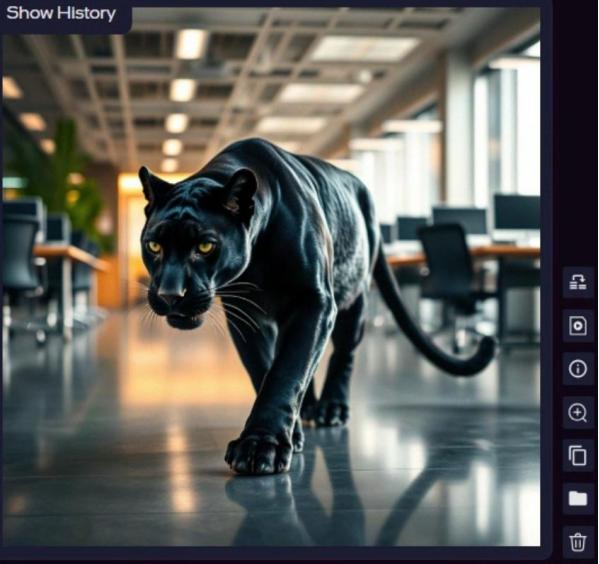
 <p><b>Model Info</b> Max Context Size: 256,000 tokens Capabilities Device: 0: Intel(R) Arc Model: Qwen3-VL-4B-Instruct-Q5_K_5.gguf Add Model Documents (0) Max Tokens: 8192 Temperature: 0.2 Context Size: 8192 Metrics: <input checked="" type="checkbox"/> Enable Tools: <input checked="" type="checkbox"/> Embeddings: bge-small-en-v1.5-gguf</p>	<p><b>Device Selector:</b> Choose which processor to run this workload. This may be restricted or filtered by the preset. For example NPU will limit the hardware to the AI Boost hardware on Intel Core Ultra processors</p> <p><b>Model Selector:</b> This list will be filtered to a set of models that are known to work well for this preset.</p> <p><b>Model Info Icon:</b> This icon when rolled over, will give info on the model including its Max Context size and features</p> <p><b>Add Model:</b> This will allow you to download a model of your choice from Huggingface</p> <p><b>Add Documents:</b> Add documents and turn these documents on or off for any chat generation you run</p> <p><b>Max Tokens:</b> The maximum number of tokens or words the output will generate</p> <p><b>Context Size:</b> The maximum number of tokens or words that can be included and read into the request, this includes all the words in your context history (current discussion)</p> <p><b>Temperature:</b> Set how strict vs creative a reasoning model will be when reasoning. Lower the stricter. Higher the more creative</p> <p><b>Metrics:</b> Toggles on or off metrics in the viewport</p> <p><b>Enable Tools:</b> Allows for tool calling, using of agentic AI to call a tool outside this model and preset</p> <p><b>Embeddings Selector:</b> Selection of models used for embedding your documents for search and analysis</p>
---	---

**Common Image Settings:** Image generation prompt settings are accessible by clicking the Image Settings button in the Prompt area. Some Presets have specific settings to their preset. Below is a description of these settings

 <p>Device: 0: Intel(R) Arc(TM) [0] GPU (32GB) Megapixels: 0.5 MP Resolution: 896 x 512 Aspect Ratio: 16/9, 4/3, 1/1, 3/4, 9/16 Steps: 20 Batch Count: 8 Negative Prompt: nsfw Seed: -1</p>	<p><b>Device Selector:</b> Choose which processor to run this workload. This may be restricted or filtered by the preset. For example, NPU will limit the hardware to the AI Boost hardware on Intel Core Ultra processors</p> <p><b>Megapixel Selector:</b> Set the resolution mode from 0.25 up to 1.5 megapixel. Options will be limited by the preset, where Draft images use Stable Diffusion 1.5 and thus are limited to 0.25 and 0.5 megapixel.</p> <p><b>Resolution:</b> This primarily serves to show you the resolution you have set, but it can also be used to define a specific resolution not available from the aspect ratio tool</p> <p><b>Aspect Ratio:</b> Use this to set the specific shape and aspect ratio of the image from ultra-wide, to square to tall portrait images</p> <p><b>Safety Checker:</b> Turn this on to censor output that might be overtly suggestive, inappropriate or NSFW.</p> <p><b>Steps:</b> Set the number of steps this generation will take to produce the image. Recommends not to change</p> <p><b>Batch Count:</b> The count of images it should generate before finishing</p> <p><b>Negative Prompt:</b> An alternative prompt you can add to some Presets where you call out things that should not be in the image</p> <p><b>Seed:</b> This is a number that represents a noise pattern used to generate the image. You can use this to regenerate images exactly as they were before</p>
---	--

**Viewport icons:** These are icons available in the viewport that aid your generative AI experience.

 <p><b>Let's Generate</b> 30.5%</p> <table border="1" style="margin-top: 10px;"> <thead> <tr> <th>Context Usage</th> <th>Tokens</th> </tr> </thead> <tbody> <tr> <td>Max Context for Model</td> <td>256K</td> </tr> <tr> <td>Configured Context Size</td> <td>8.2K</td> </tr> <tr> <td>Used Context</td> <td>2.5K</td> </tr> <tr> <td>Input</td> <td>2.2K</td> </tr> <tr> <td>Output</td> <td>251</td> </tr> </tbody> </table> <p><b>Token Status:</b> Tells you the Token information vs usage. The drop down is hidden by default, only showing the percentage and circle graph. It is exposed by rolling over the circle graph</p>	Context Usage	Tokens	Max Context for Model	256K	Configured Context Size	8.2K	Used Context	2.5K	Input	2.2K	Output	251	 <p>Copy Regenerate Delete 251 Tokens · 18.14 Tokens/s · 1st Token Time: 9213.92ms</p> <p><b>Discussion Icons:</b></p> <ul style="list-style-type: none"> <li><b>Copy:</b> Copies to clipboard the content of the latest discussion response</li> <li><b>Regenerate:</b> Deletes the last response and regenerates the same request to the model using the same prompt</li> <li><b>Delete:</b> Deletes that response from discussion context. This is helpful to keep poor answers and information from the model's memory</li> </ul> <p><b>Metrics:</b> Information on the token count, tokens per second and 1<sup>st</sup> token time of that response in the discussion. This can be toggled on and off using the Metric checkbox in Preset settings</p>
Context Usage	Tokens												
Max Context for Model	256K												
Configured Context Size	8.2K												
Used Context	2.5K												
Input	2.2K												
Output	251												



Show History

**Image Viewport Icons**

**Send To Edit:** Send image to Image Edit History, making it draggable to Image Edit presets

**Send To Video:** Send image to Vide History, making it draggable to Image To Video presets

**Info:** Get info on the image generation details such as prompt, seed, model and more

**Zoom:** Opens the image up in the Photo Viewer application

**Copy:** Copies the image to the clipboard

**Folder:** Opens the Folder where the image is stored and saved

**Delete:** Removes the image from history and the viewer

## Using Other Models:

A key feature of AI Playground is the ability to use other models not provided directly from the AI Playground application.

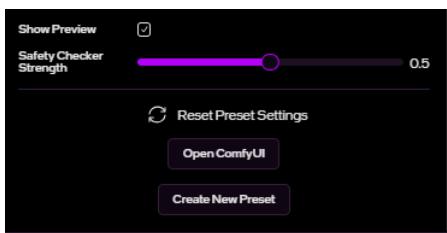
### Example Outputs using different Checkpoint Models

Lykon/dreamshaper-8	realisticVisionV51_v51VAE.safetensors	counterfeitV30_v30.safetensors
		

**Prompt:** close up of an adorable anime bobble head character, detailed eyes, animated film, atomic space age hero, excited, open mouth, swanky, cute, funny, brown hair, wide angle, action pose, fantastic landscape, rocket, red, gold, space, planets purple glow  
**Seed:** 3027545586    **Scheduler:** Euler

**Where to find models:** These websites host models you can use in AI Playground.

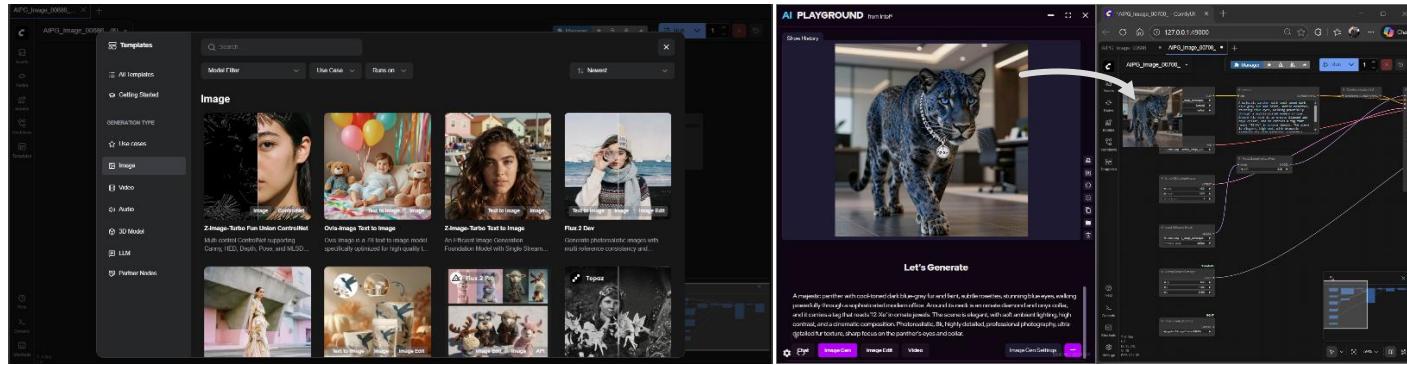
- **Huggingface:** Huggingface is a website that hosts AI models for both AI chat and image generation. You can search huggingface for specific types of models or if you have a model ID you can use that to find a model. <https://huggingface.co/>  
 For LLM models, AI Playground v2.0a currently supports PyTorch models supporting Transformers 4.39.
- **CivitAI:** CivitAI is a website with community defined and refined models for image generation. Here you can find models that are fine tuned to create outputs for specific styles – illustration, photography, anime, game design, architecture, etc. You will find models that work in AI Playground such as checkpoints for Stable Diffusion 1.5 (our standard mode), SDXL (our HD mode), LoRAs used to stylize or adhere output to a specific feature, Inpainting models and more.  
<https://civitai.com/models>
- **ComfyUI:** A key feature of AI Playground is the ability to extend far beyond the features of AI Playground and leverage the many workflow templates or custom workflows you build in ComfyUI. See Using ComfyUI section. To access ComfyUI select OpenComfyUI from the bottom of either the Image Gen or Image Edit Settings



## Using ComfyUI

**Browse and Try New Workflows:** ComfyUI comes with dozens of generative AI workflows. Select the main ComfyUI menu then select Browse Templates. Note some templates require specific nodes, and models not already installed by ComfyUI or AI Playground. See ComfyUI website and online documentation for assistance. <https://www.comfy.org/>

**Tweak and Edit AI Playground Images:** Simply drag any image created from AI Playground 3.0 to the ComfyUI workspace. This will read the data on that image and create the node structure, allowing you to generate new images using the same settings directly from ComfyUI. While in ComfyUI you have access to all the values not normally exposed in AI Playground. You can also add or delete nodes, use different models, and change default settings to create a new workflow built off the AI Playground preset.



## Model Index

AI Playground v3.0a links to models for download at the time of use. You are not restricted to these models and can download models other sources.

Model Table: Chat

Model Path / Name	Type	Model Card (HF)	License Link
DeepSeek-R1-Distill-Qwen 1.5B	Reasoning	<a href="#">Model Card</a>	<a href="#">MIT License</a>
DeepSeek-R1-Distill-Qwen 7B	Reasoning	<a href="#">Model Card</a>	<a href="#">MIT License</a>
Gemma 3.4B IT (Unsloth)	Agentic, Vision	<a href="#">Model Card</a>	<a href="#">Gemma License</a>
GPT-OSS 20B (Unsloth)	Agentic, Reasoning	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
InternVL2 4B (OV)	Vision	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Llama 3.2 3B Instruct	Agentic	<a href="#">Model Card</a>	<a href="#">Llama 3.2 License</a>
Meta-Llama 3.1 8B Instruct	Agentic	<a href="#">Model Card</a>	<a href="#">Llama 3.1 License</a>
Mistral 7B Instruct v0.2 (OV)	Text Only	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Mistral 7B Instruct v0.3	Agentic	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Mistral 7B Instruct v0.3 (OV)	Agentic	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Phi-3 Mini 4k Instruct (OV)	Text Only	<a href="#">Model Card</a>	<a href="#">MIT License</a>
Phi-3.5 Mini Instruct (OV)	Text Only	<a href="#">Model Card</a>	<a href="#">MIT License</a>
Qwen2-VL 7B Instruct (OV)	Agentic, Vision	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Qwen3 4B (OV)	Agentic, Reasoning	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Qwen3 4B (Unsloth)	Agentic, Reasoning	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Qwen3 4B Instruct 2507 (Unsloth)	Agentic	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Qwen3-VL 4B Instruct (Unsloth)	Agentic, Vision	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
SmolLM2 1.7B Instruct	Agentic	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
TinyLlama 1.1B Chat (OV)	Text Only	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
BGE Small EN v1.5 (GGUF)	Embedding	<a href="#">Model Card</a>	<a href="#">MIT License</a>
Nomic Embed Text v1.5 (GGUF)	Embedding	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Whisper (OV)	Speech	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>

Model Table: Image Gen, Image Edit, Video

Model	Preset	Model Card	License Link
AdamCodd/vit-base-nsfw-detector	SafetyChecker	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Aitrepreneur/insightface/inswapper_128.onnx	ControlFace, CopyFace	<a href="#">Model Card</a>	<a href="#">Non-Commercial</a>
alimama-creative/FLUX.1-Turbo-Alpha	EditByPrompt	<a href="#">Model Card</a>	<a href="#">FLUX.1-dev License</a>
city96/FLUX.1-schnell-gguf/flux1-schnell-Q4_K_S.gguf	Pro	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
city96/FLUX.1-schnell-gguf/flux1-schnell-Q8_0.gguf	Pro	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
city96/t5-v1_1-xxl-encoder-gguf/t5-v1_1-xxl-encoder-Q3_K_M.gguf	Pro, LTX-Video	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
city96/t5-v1_1-xxl-encoder-gguf/t5-v1_1-xxl-encoder-Q4_K_M.gguf	LTX-Video	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
city96/umt5-xxl-encoder-gguf/umt5-xxl-encoder-Q4_K_M.gguf	Wan2.1Vace	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
comfyanonymous/flux_text_encoders/clip_l.safetensors	EditByPrompt, Pro	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
comfyanonymous/flux_text_encoders/t5xxl_fp8_e4m3fn_scaled.safetensors	EditByPrompt	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Comfy-Org/flux1-kontext-dev.../flux1-dev-kontext_fp8_scaled.safetensors	EditByPrompt	<a href="#">Model Card</a>	<a href="#">FLUX.1-dev License</a>
Comfy-Org/Lumina_Image_2.0_Repackaged/.../ae.safetensors	EditByPrompt, Pro	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Comfy-Org/Real-ESRGAN_repackaged/.../RealESRGAN_x4plus.safetensors	Upscale	<a href="#">Model Card</a>	<a href="#">BSD-3-Clause</a>
Comfy-Org/Wan_2.1_ComfyUI_repackaged/.../wan_2.1_vae.safetensors	Wan2.1Vace	<a href="#">Model Card</a>	<a href="#">Wan 2.1 License</a>
Comfy-Org/z_image_turbo/.../z_image_turbo_bf16.safetensors	Pro 2	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Comfy-Org/z_image_turbo/.../qwen_3_4b.safetensors	Pro 2	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Comfy-Org/z_image_turbo/.../ae.safetensors	Pro 2	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
gmk123/GFPGAN/GFPGANv1.4.pth	ControlFace, CopyFace	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
latent-consistency/lcm-lora-sdv1-5/pytorch_lora_weights.safetensors	Draft	<a href="#">Model Card</a>	<a href="#">OpenRAIL++</a>
latent-consistency/lcm-lora-sdxl/pytorch_lora_weights.safetensors	HD	<a href="#">Model Card</a>	<a href="#">OpenRAIL++</a>
Lightricks/LTX-Video/ltxv-2b-0.9.6-distilled-04-25.safetensors	LTX-Video	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
Illyasviel/fooocus_inpaint/fooocus_inpaint_head.pth	Inpaint, Outpaint	<a href="#">Model Card</a>	<a href="#">OpenRAIL</a>
Illyasviel/fooocus_inpaint/inpaint_v26.fooocus.patch	Inpaint, Outpaint	<a href="#">Model Card</a>	<a href="#">OpenRAIL</a>
Lykon/DreamShaper/DreamShaper_8_pruned.safetensors	Draft	<a href="#">Model Card</a>	<a href="#">OpenRAIL-M</a>
Lykon/dreamshaper-8-inpainting/text_encoder/model.safetensors	Inpaint, Outpaint	<a href="#">Model Card</a>	<a href="#">OpenRAIL-M</a>
Lykon/dreamshaper-8-inpainting/unet/...safetensors	Inpaint, Outpaint	<a href="#">Model Card</a>	<a href="#">OpenRAIL-M</a>
Lykon/dreamshaper-8-inpainting/vae/...safetensors	Inpaint, Outpaint	<a href="#">Model Card</a>	<a href="#">OpenRAIL-M</a>
QuantStack/Wan2.1_14B_VACE-GGUF/Wan2.1_14B_VACE-Q8_0.gguf	Wan2.1Vace	<a href="#">Model Card</a>	<a href="#">Apache 2.0</a>
RunDiffusion/Juggernaut-XL-v9/...RunDiffusionPhoto_v2.safetensors	HD, Inpaint, Outpaint, SketchToPhoto	<a href="#">Model Card</a>	<a href="#">OpenRAIL-M</a>
stabilityai/control-lora/...rank128-canny-rank128.safetensors	SketchToPhoto	<a href="#">Model Card</a>	<a href="#">SAI Community</a>
tencent/Hunyuan3D-2.1/hunyuan3d-dit-v2-1/model.fp16.ckpt	3D	<a href="#">Model Card</a>	<a href="#">Hunyuan3D License</a>
tencent/Hunyuan3D-2/hunyuan3d-dit-v2-0/model.fp16.safetensors	3D	<a href="#">Model Card</a>	<a href="#">Hunyuan3D License</a>

**Notices and Disclaimers:**

For information on AI Playground terms, license and disclaimers, visit the project and files on GitHub repo:  
[License](#) | [Notices & Disclaimers](#)

The software may include third party components with separate legal notices or governed by other agreements, as may be described in the Third Party Notices file accompanying the software.