ChatGPT

List of All LMArena Models

LMArena (Large Model Arena) is a platform that brings together a wide array of AI models for side-by-side evaluation and use. These models span multiple categories – from text-based Large Language Models (LLMs) and chatbots to image generators and even hidden or experimental models used in internal "battles." Below is a comprehensive listing of all LMArena models, organized by category and provider, along with brief context for each group.

Text & Chat Models (LLMs)

This category includes cutting-edge language models from major AI labs. Many of these models are offered in multiple **variants** (e.g., "Pro" for highest performance, "Flash" for faster responses) and **versions** (often indicated by version numbers or dates). Some models have specialized modes like *"thinking"* (deeper reasoning with higher cost) vs. *"no-thinking"* (standard mode) [1]. Below we list the text/chat models by provider:

Google (Gemini & Gemma)

Google's contributions include the **Gemini** series (their flagship next-gen LLMs) and the **Gemma** series (lighter-weight models). Gemini models (e.g. 2.5, 3.0 versions) are state-of-the-art multimodal AI systems with extremely large context windows (up to 1 million tokens) and top-tier performance – in fact, *Gemini 2.5 Pro* ranks #1 on LMArena's leaderboard [2]. They come in variants like *Pro* (maximum capability) and *Flash* (optimized for speed/efficiency). Some have experimental "grounding" or "thinking" modes for enhanced reasoning. The Gemma models (e.g. Gemma-3) are smaller-scale Google models (ranging from ~1B to 27B parameters) intended for instruct-tuned tasks and efficient deployment [3].

- `gemini-2.5-pro`
- `gemini-2.5-pro-grounding-exp`
- `gemini-2.5-flash`
- `gemini-2.5-flash-preview-09-2025`
- `gemini-2.5-flash-lite-preview-09-2025-no-thinking`
- `gemini-2.5-flash-lite-preview-06-17-thinking`
- `gemini-3-pro`
- `gemini-2.0-flash-001`
- `gemma-3-27b-it`
- `gemma-3n-e4b-it`

OpenAI (GPT & O-Series)

OpenAI's models on LMArena include the **GPT-5** series and the more opaque **"O-series."** The GPT-5 family presumably represents OpenAI's next-generation models after GPT-4, offered in various versions: for example, *gpt-5.1* (with a high-precision variant) and *gpt-5-chat* for conversational use. There are also specialized variants like *gpt-5-high-new-system-prompt* (possibly testing new system prompt capabilities) and

scaled-down versions (*mini* or *nano* for efficiency). In addition, LMArena lists certain OpenAI models prefixed with "o" (like **o3** and **o4** series). These **O-series** might be internal code-names or experimental models – for instance, `o3-mini` and its "high" variant appear in OpenAI's model lists [4], suggesting an optimized model with high intelligence at lower cost. The "chatgpt-4o" entry likely refers to an optimized GPT-4 model (possibly via OpenRouter or a specific OpenAI release). Overall, OpenAI's presence includes both public models and these internal prototypes.

- `gpt-5.1 / gpt-5.1-high`
- `gpt-5-chat`
- `gpt-5-high / gpt-5-high-new-system-prompt / gpt-5-high-no-system-prompt`
- `gpt-5-mini-high / gpt-5-nano-high`
- `chatgpt-4o-latest-20250326`
- `gpt-4.1-2025-04-14 / gpt-4.1-mini-2025-04-14`
- `gpt-oss-120b / gpt-oss-20b`
- `o3-2025-04-16 / o3-mini`
- `o4-mini-2025-04-16`

Anthropic (Claude)

Anthropic's **Claude** series is represented with multiple versions. Claude 3 and Claude 4 are major generations, each with sub-variants named after literary forms: **Sonnet**, **Haiku**, **Opus**, etc., which likely correspond to different model sizes or tuning styles. For example, *Claude-3.7 Sonnet (2025-02-19)* is a Claude 3rd-gen model updated in Feb 2025 (with an optional 32k "thinking" mode for extended context), while *Claude-3.5 Haiku/Sonnet (2024-10-22)* are earlier versions from late 2024. The Claude 4 series (Claude 4.0 through 4.5) are more advanced, with *Opus* and *Sonnet* variants often supporting very long contexts (up to 100k or more tokens) and improved reasoning. The "(+ thinking-32k)" or "(+ thinking-16k)" notes indicate versions that allow an even larger context or a special mode for complex reasoning. In summary, Anthropic's models on LMArena range from Claude 3.5 up to the latest Claude 4.5, capturing improvements in capability and context length over time.

- `claude-3-7-sonnet-20250219 (+ thinking/thinking-32k)`
- `claude-3-5-sonnet-20241022`
- `claude-3-5-haiku-20241022`
- `claude-opus-4-5-20251101 (+ thinking-32k)`
- `claude-sonnet-4-5-20250929 (+ thinking-32k)`
- `claude-haiku-4-5-20251001`
- `claude-opus-4-1-20250805 (+ thinking-16k)`
- `claude-opus-4-20250514 (+ thinking-16k)`
- `claude-sonnet-4-20250514 (+ thinking-32k)`

xAI (Grok)

xAI's **Grok** models are also included. These are Elon Musk's AI initiative models, which appear in LMArena as Grok v3 and v4 series. The list shows **Grok-3** (with "mini-beta" and "mini-high" variants – likely smaller versions tuned for certain performance levels) and **Grok-4** in several forms. Grok-4.1 is listed with a normal and a "thinking" variant, suggesting an ability to toggle a deep reasoning mode. There are also *fast* vs. *non-reasoning* versions (e.g., `grok-4-1-fast-reasoning` vs `fast-non-reasoning`), indicating some

deployments optimized for speed vs. reasoning quality. A dated entry `grok-4-0709` likely refers to a specific build or checkpoint (perhaps July 09). In summary, Grok models in LMArena cover both the third-generation and an early fourth-generation, with configurations balancing speed and reasoning.

- `grok-3-mini-beta / grok-3-mini-high`
- `grok-4.1 / grok-4.1-thinking`
- `grok-4-1-fast-reasoning / grok-4-1-fast-non-reasoning`
- `grok-4-0709`
- `grok-4-fast-chat / grok-4-fast-reasoning`

Alibaba (Qwen)

Alibaba's **Qwen** family is extensive and appears prominently in LMArena. **Qwen 3** is the latest generation, with models of various sizes and specialist versions: - *Qwen-3 Max* – likely the largest versions; the list shows dated entries (2025-09-23, 09-26, 10-20) which could be successive preview releases of a very large Qwen model ("Max" might imply a multi-hundred-billion parameter MoE model). - *Qwen-3 235B A22B* – a huge 235B-parameter model (possibly mixture-of-experts with 22B experts), listed with instruct, thinking, and no-thinking variants. This corresponds to one of the highest-end open models, with very strong performance [5]. - *Qwen-3 Next 80B A3B* – an 80B model (A3B might denote architecture details) with instruct and thinking versions. - *Qwen-3 30B A3B* – a 30B variant (likely a dense model) with an instruct version [6]. - *Qwen-3 Coder 480B* – a massive 480B model specialized for coding (perhaps MoE), indicating Alibaba built a code-focused model at unprecedented scale [7]. - *Qwen-3 Omni-Flash* – possibly a multi-modal or highly optimized version (the term "Omni" suggests a model capable of various tasks or modalities). - *QwQ-32B* – this appears to be a predecessor or related model (QwQ might be an earlier codename in the Qwen lineage, 32B params). - **Vision models**: Alibaba's list also includes Qwen vision-language models: e.g. `qwen3-vl-235b-a22b` and `qwen3-vl-8b`, plus `qwen-vl-max-2025-08-13`. These are likely multi-modal models (VL = vision-language) of different sizes for image understanding tasks (image captioning, OCR, etc.), complementing the text models.

Overall, Alibaba's Qwen lineup in LMArena spans from powerful text models that excel in reasoning, math, coding, etc., to multi-modal VL models, demonstrating Alibaba's push in both large-scale model performance and broad capabilities.

- `qwen3-max-2025-09-23 / 09-26 / 10-20`
- `qwen3-max-preview / qwen3-max-thinking`
- `qwen3-next-80b-a3b-instruct / thinking`
- `qwen3-235b-a22b (+ instruct/thinking/no-thinking)`
- `qwen3-30b-a3b (+ instruct)`
- `qwen3-coder-480b-a35b-instruct`
- `qwen3-omni-flash`
- `qwq-32b`
- *Vision:* `qwen3-vl-235b-a22b (+ instruct/thinking)`, `qwen3-vl-8b (+ instruct/thinking)`, `qwen-vl-max-2025-08-13`

DeepSeek

DeepSeek is an open-source model provider known for high reasoning and coding performance. LMArena lists DeepSeek v3.2, including a special "thinking" variant: - *deepseek-v3.2* – the base model (likely a general-purpose LLM known for coding ability). - *deepseek-v3.2-thinking* – an enhanced mode of v3.2 that perhaps engages more intensive reasoning or chain-of-thought for complex queries. - *deepseek-v3-0324* – possibly an earlier release (March 24, 2025) or a specific checkpoint version.

DeepSeek models have been noted to rival top closed models on certain benchmarks [8] and are often used for tasks like code generation or analysis. In LMArena, DeepSeek serves as a strong open competitor among the heavyweights.

- `deepseek-v3.2`
- `deepseek-v3.2-thinking`
- `deepseek-v3-0324`

Meta (Llama)

Meta's **Llama** series is present with next-generation versions: - **Llama 3.3 70B** – a 70B-parameter instruct-tuned model (likely an evolution after Llama 2). This is labeled "3.3" and instruct, suggesting Meta continued improving Llama in 2024/2025 with better training and perhaps longer context (70B is the largest size, possibly with 128k context as indicated elsewhere [9]). - **Llama 4 "Maverick" 17B** – an early Llama 4 variant codenamed *Maverick*, with 17B parameters and "128e" (perhaps 128k context, enhanced version). This suggests Meta was testing Llama 4 with a mid-size model before scaling up. "Maverick" could be an internal project name for Llama 4 research.

These models underscore Meta's ongoing LLM development, with Llama 3 and experimental Llama 4 prototypes accessible via LMArena.

- `llama-3.3-70b-instruct`
- `llama-4-maverick-17b-128e-instruct`

Mistral

Mistral AI's models are known for being efficient and high-quality at smaller scales. In LMArena we see: - **Mistral Large 3** – presumably a version of Mistral's large model (might correspond to a 13B or similar model, as Mistral 3 could be an iteration beyond their initial 7B release). - **Mistral Medium (2505 / 2508)** – likely medium-sized models with dates (May and Aug 2025 updates). - **Mistral Small (2506) / 3.1-24B (2503)** – possibly a 3.1 version from March 2025 and a 24B parameter model (which is interesting since Mistral's known release was 7B; perhaps they scaled up or did MoE). - **Magistral Medium 2506** – "Magistral" could be a variant or next-gen codename associated with Mistral (the naming suggests a play on words). It might be an experimental model or an internal test around mid-2025.

In summary, LMArena has multiple Mistral models covering different sizes and versions, reflecting how Mistral AI's technology progressed and was evaluated.

- `mistral-large-3`

- `mistral-medium-2505 / 2508`
- `mistral-small-2506 / 3.1-24b-instruct-2503`
- `magistral-medium-2506`

Other Text Models

Aside from the big players above, LMArena also integrates various models from other organizations and some miscellaneous ones: - **Baidu (Ernie)** – Baidu's foundation model, **ERNIE 5.0**, appears as "ernie-5.0-preview" (with versions dated 2025-11-03 and 11-20, likely early previews around Baidu World conference). Ernie 5.0 was introduced as Baidu's first native *omnimodal* foundation model, excelling in multi-modal understanding, creative writing, and instruction following [10]. Earlier or experimental versions are listed as "ernie-exp" with various dates, indicating iterative releases of Baidu's model. - **Zhipu (GLM)** – Zhipu AI's models like *GLM-4.5* series are included. The entries `glm-4.5`, `glm-4.5-air`, `glm-4.5v`, `glm-4.6` suggest multiple variants or updates of their bilingual generative model (GLM) around 2024-2025. (In context, GLM-4 is an open large model family from China, where "air" or "v" might denote specific fine-tunes or versions.) - **MiniMax** – MiniMax's models (`minimax-m1`, `m2`, and `m2-preview`) are present. MiniMax is another Chinese AI startup, and these models likely correspond to their first and second generation chat models (with M2 Preview being a test release). - **Tencent (Hunyuan)** – Tencent's large model **Hunyuan** appears with `hunyuan-t1-20250711` (perhaps the v1 model as of July 2025) and notably a `hunyuan-vision-1.5-thinking`. The latter suggests a multimodal or vision-enabled model (Hunyuan 1.5) with a "thinking" mode, meaning Tencent had a vision-language model or an image-capable LLM in the Arena. - **Amazon (Nova)** – Amazon's models like `nova-2-lite`, `amazon-nova-experimental-chat`, and `amazon.nova-pro-v1:0` are listed. Nova is Amazon's family of foundation models (following their earlier Titan series). Nova-2 indicates second-generation; "lite" might be a smaller variant, and "pro" a more powerful version. An experimental chat version is also present, reflecting Amazon's ongoing development in generative AI. - **Miscellaneous** – A few models with unique names are included, possibly from smaller providers or open projects: - `command-a-03-2025` – likely Cohere's **Command** model (a large instruct model) as of March 2025, since Cohere's flagship is often "Command" [11]. - `ling-1t` and `ling-flash-2.0` – these could be from an organization or an internal project named "Ling" (perhaps an AI lab or a codename for a multilingual model; the `1t` might hint at 1 trillion tokens trained or similar). - `step-3` – unclear origin, perhaps an academic model (the name is generic). - `ring-flash-2.0` – possibly related to the "Ling" above (ring and ling could be paired codenames) or another project focusing on a "Flash" efficient model. - `intellect-3` – a model likely named for general AI prowess; could be an experimental model from a lesser-known group.

All these additional models broaden LMArena's coverage, ensuring even niche or region-specific models (like Chinese bilingual models or specialized assistants) are represented for comparison.

- **Baidu:** `ernie-5.0-preview (1103/1120)`, `ernie-exp` *(various dates)* [10]
- **Zhipu:** `glm-4.5`, `glm-4.5-air`, `glm-4.5v`, `glm-4.6`
- **MiniMax:** `minimax-m1`, `minimax-m2`, `minimax-m2-preview`
- **Tencent:** `hunyuan-t1-20250711`, `hunyuan-vision-1.5-thinking`
- **Amazon:** `nova-2-lite`, `amazon-nova-experimental-chat`, `amazon.nova-pro-v1:0`
- **Misc:** `command-a-03-2025`, `ling-1t`, `ling-flash-2.0`, `step-3`, `ring-flash-2.0`, `intellect-3`

Image Generation Models

This category includes text-to-image models, image editing models, and even some text-to-video overlaps. These models allow generation of visual content (images or video frames) from prompts, and some can modify or extend given images. LMArena's image model lineup features contributions from tech giants as well as specialized labs:

Google (Imagen / Gemini)

Google's image models include the **Imagen** series and new **Gemini** multimodal generative models: - **Imagen 4.0** – Google's advanced text-to-image model (successor to Imagen 3). There are multiple variants: *standard generate* (`imagen-4.0-generate-001`), a *fast generate* version for quicker, possibly lower-fidelity outputs, and an *ultra* version for highest fidelity (`imagen-4.0-ultra-generate-001`). These indicate different trade-offs between speed and quality. - **Imagen 3.0** – an earlier generation (`imagen-3.0-generate-002`) likely still included for reference. - **Gemini 3 Pro (Image Preview)** – this appears to be a **Gemini** model applied to image generation. Gemini is primarily an LLM, but Google's Gemini is multimodal by design, so this could be an *image-generation preview* using the Gemini 3 model. It's offered in Standard, 2K, and 4K versions, suggesting it can produce high-resolution images (2K, 4K pixel dimensions). - **Gemini 2.5 Flash (Image Preview)** – another multimodal model variant focusing on fast image generation, possibly using the Gemini 2.5 architecture. - **Gemini 2.0 Flash (Image Generation)** – an earlier preview of image generation capabilities from Gemini 2.0.

In summary, Google's presence covers both dedicated image models (Imagen) and the image-generating capabilities of its latest multimodal LLM (Gemini).

- `gemini-3-pro-image-preview` *(Standard, 2k, and 4k versions)*
- `gemini-2.5-flash-image-preview`
- `gemini-2.0-flash-preview-image-generation`
- `imagen-4.0-generate-001`
- `imagen-4.0-fast-generate-001`
- `imagen-4.0-ultra-generate-001`
- `imagen-3.0-generate-002`

Black Forest Labs (Flux)

Black Forest Labs (BFL) provides the **Flux** series of image generative models. These models, named Flux, seem to have different versions and specializations: - **Flux 2** – includes *Flux-2-Pro*, *Flux-2-Dev*, *Flux-2-Flex*. The "Pro" likely is the full model for high quality, "Dev" might be a developer or experimental version, and "Flex" could indicate a flexible or faster model variant. - **Flux 1 "Kontext"** – includes *Flux-1-Kontext-Pro*, *Dev*, and *Max*. The use of "Kontext" (context) suggests these versions might integrate additional context or conditioning (perhaps using text+image inputs or other metadata). "Max" could be a version aimed at maximum quality or context length.

These Flux models are less commonly known publicly, but within LMArena they represent a set of possibly proprietary image generators focusing on contextual and high-quality image synthesis.

- `flux-2-pro`

- `flux-2-dev`
- `flux-2-flex`
- `flux-1-kontext-pro`
- `flux-1-kontext-dev`
- `flux-1-kontext-max`

OpenAI

OpenAI's image generation in LMArena features: - **DALL·E 3** – OpenAI's flagship text-to-image model released in late 2023, known for its superior prompt understanding and high-quality outputs [12]. It's one of the top image models, excelling in creating creative and detailed images from text prompts. - **GPT-Image 1** – An experimental series of models under the GPT branding that generate images. The list shows a base *gpt-image-1*, a *mini* version, and a *high-fidelity* version. This suggests OpenAI had a first iteration of a multimodal GPT that can output images: possibly *GPT-Image* is an internal project combining GPT-style reasoning with image generation. The mini version would be smaller/faster, and the high-fidelity one aimed at very detailed outputs (perhaps at the cost of speed).

These models indicate OpenAI's exploration beyond DALL·E, merging their GPT capabilities into image generation. On LMArena, DALL·E 3 is featured as a top benchmark, while GPT-Image models show the experimental edge of OpenAI's research into multimodal generation.

- `dall-e-3` [12]
- `gpt-image-1`
- `gpt-image-1-mini`
- `gpt-image-1-high-fidelity`

Alibaba (Qwen Image)

Alibaba also provides image generation models under the Qwen umbrella: - **Qwen-Image-Edit** – likely a model specialized for image editing. This could take an input image plus instructions and output a modified image (for example, inpainting or applying a style). - **Qwen-Image-Prompt-Extend** – perhaps a model that extends or enhances prompts for image generation, or generates images with extended prompts. Another interpretation is that it might extend an image beyond its original boundaries (like outpainting) using prompt guidance.

These tools suggest Alibaba's efforts to cover not just text generation (Qwen LLMs) but also image manipulation and enhancement tasks in the Arena.

- `qwen-image-edit`
- `qwen-image-prompt-extend`

Tencent (Hunyuan)

Tencent's **Hunyuan** family, besides text models, includes image generators: - **Hunyuan-Image 3.0** – a text-to-image model (v3.0) presumably with significant quality improvements. - **Hunyuan-Image 3.0-fal** – possibly a variant of 3.0; "fal" might stand for something like *fast iteration* or a specialized fine-tune (the exact meaning is unclear, possibly an acronym for a project or an optimized version). - **Hunyuan-Image 2.1** – an earlier version, indicating that 3.0 is a major update over 2.1.

These models show Tencent's progression in image generative AI, likely used for their internal products or cloud services (Hunyuan is Tencent Cloud's AI brand). LMArena including them means their performance can be compared with other contemporaries.

- `hunyuan-image-3.0`
- `hunyuan-image-3.0-fal`
- `hunyuan-image-2.1`

Wan / Video Models

This subcategory appears to bridge image and video generation, featuring models capable of producing not just static images but possibly animations or edited video frames: - **Wan2.5** – Likely a model from a project named "Wan" (possibly Chinese for "myriad" or just a code name). The entries suggest: - *wan2.5-preview* – a general preview of version 2.5 (perhaps a text-to-video model). - *wan2.5-t2i-preview* – specifically a **Text-to-Image** preview using the Wan2.5 model (so it can generate images from text prompts, indicating multimodal capability). - *wan2.5-i2i-preview* – an **Image-to-Image** preview (the model can take an image and produce a transformed image, e.g. style transfer or variation). These suggest Wan2.5 is a versatile generative model handling both image and potentially video (though the entries listed are image-focused previews). - **vidu-q2-image** – This looks like a model possibly named "Vidu" (which hints at video). "q2" might be a version or quarter indication. It could be a video-to-image or an intermediate step in video generation (like generating video key frames). It's not very clear, but likely an experimental model dealing with video or sequential image generation. - **Reve-v1 / Reve-fast-edit** – "Reve" could be French for "dream", aligning with generative imagery. Reve-v1 might be a base model for image or video generation, and *fast-edit* a variant optimized for quick editing of images or video frames. Possibly a project focusing on real-time video edits or high-speed generation.

Together, these models in LMArena highlight the frontier of text-to-video and advanced image-to-image generation. They may be in early preview stages ("preview" implies not fully production) and are included in the Arena for evaluation against more established image models.

- `wan2.5-preview`
- `wan2.5-t2i-preview (Text to Image)`
- `wan2.5-i2i-preview (Image to Image)`
- `vidu-q2-image`
- `reve-v1 / reve-fast-edit`

Other Visual Models

In addition to the above, LMArena features a collection of other visual generation/editing models, many with creative code names. These are likely specialized models from various research groups or smaller companies, each with unique capabilities:

- **ReCraft-v3** – Possibly an image **re**construction or **re**creation model. It might excel at tasks like re-drawing or editing existing images (the name suggests "crafting" images anew).
- **Ideogram-v3-quality** – *Ideogram* is known as an AI that can generate text in images (like stylized text or logos). This entry likely refers to a version focused on high-quality outputs. It suggests the

model excels at producing images containing legible and well-rendered text or symbols (a notoriously hard task for image generators).
- **Seedream & Seededit** – These sound like models from the same family:
- *Seedream-3, Seedream-4.5, Seedream-4-high-res-fal* – possibly generative models (dream = image generation) where v4.5 is an improved version, and a "high-res fal" variant for high-resolution output (perhaps "FAL" stands for a specific high-res technique or acronym).
- *Seededit-3.0* – likely an image editing model (complement to Seedream), version 3.0, used for altering images based on instructions.
- **Mai-image-1** – Could be a model by an organization or person named "Mai", or an acronym (e.g. Multi-modal AI?). Version 1 suggests a first iteration. Not much context, but included as a distinct model in the Arena.
- **Photon** – Perhaps a model focusing on photorealistic image generation or manipulation (photon evoking light/photos). It might be a renderer or an enhancer for realism.
- **Lucid-origin** – Name implies generating vivid, "lucid" imagery. Could be an art-focused model or a framework output.
- **Hazel** – There are *hazel-gen* (generation) and *hazel-edit* models:
- *Hazel-Gen-2 / Hazel-Gen-4* – likely two versions of a generative model named Hazel (maybe v2 and v4).
- *Hazel-Edit-2 / Hazel-Edit-6* – corresponding editing models for images, where v6 is a later iteration with improved editing capabilities.
- **HiDream-e1.1** – Possibly "HiDream" model (play on 'hydra' or 'dream'). E1.1 might denote an experimental or engine version for image creation.
- **Tangerine** – Possibly a code name for a vibrant image generator (the name is whimsical; it might produce vivid "fruity" colors or be a small-scale model).
- **Ghost-Pepper** – Another creatively named model, which could imply generating spicy or intense visuals (or could be a novelty model for ghostly images or augmented reality).

Many of these are likely internal or lesser-publicized models included in LMArena for comprehensive benchmarking. They cover niche capabilities (text-in-image, high-res outputs, editing functions) that complement the major image models.

- `recraft-v3`
- `ideogram-v3-quality`
- `seedream-3 / seedream-4.5 / seedream-4-high-res-fal`
- `seededit-3.0`
- `mai-image-1`
- `photon`
- `lucid-origin`
- `hazel-gen-2 / 4`
- `hazel-edit-2 / 6`
- `hidream-e1.1`
- `tangerine`
- `ghost-pepper`

Hidden / Anonymous / Battle Models

In LMArena's competitive evaluations (such as blind comparison battles or leaderboard tests), model identities are sometimes obscured. The platform uses **codenames** or **anonymous labels** for models to

ensure fairness and prevent bias. Below are groups of such hidden or internally named models that have appeared in the Arena:

The "Beluga/Phantom" Series

These are code names that likely correspond to specific models entered anonymously: - **beluga-1128-1** – “Beluga” could be a codename (perhaps for a model known for being large or “whale” themed). The number might indicate a date (Nov 28) or an internal version. - **phantom-1203-1** – “Phantom” with a date (Dec 03) suggests an anonymized model tested around that time. - **phantom-mm-1125-1** – Another Phantom variant, possibly “mm” indicating a multi-modal version, tested on Nov 25.

These names might hide well-known models (for example, a GPT-4 or Claude entry under a fake name) during blind battles. The Beluga/Phantom series was used to keep model identity secret while gathering human preference data.

- `beluga-1128-1`
- `phantom-1203-1`
- `phantom-mm-1125-1`

The "Raptor" Series

“Raptor” series comprises several versions: - **raptor** – The base name appears with various date codes: 1110, 1119, 1123, 1124, 1202 (likely indicating builds on Nov 10, 19, 23, 24 and Dec 02). These could be iterative improvements of a model nicknamed Raptor. - **raptor-llm** – Specific to language model versions, with dates 1017, 1024, 1125, 1205 (Oct 17, Oct 24, Nov 25, Dec 05). This suggests a timeline of an “LLM” variant of Raptor being tested across the fall of 2025. - **raptor-vision** – Vision-capable versions of Raptor, with 1015 and 1107 (Oct 15 and Nov 07) as version markers. Possibly these integrated image understanding or multimodal input.

The Raptor series likely corresponds to a research model (or set of models) that underwent rapid iteration. By disguising it as “Raptor” in the Arena, developers could compare it against others without revealing it was their own model until ready.

- `raptor (base, 1110, 1119, 1123, 1124, 1202)`
- `raptor-llm (1017, 1024, 1125, 1205)`
- `raptor-vision (1015, 1107)`

The "EB / X1" Series

These codenames likely refer to internal experimental models or specific challenge entries: - **EB45** – Possibly stands for a project name (EB could be a prefix, maybe “EagleBeta” or something fictional). We see *EB45-turbo*, *EB45-turbo-vl-0906*, *EB45-vision*. “turbo” suggests a faster version, “vl” might stand for vision-language (multimodal) dated Sept 06, and a pure vision variant. These could be entries from a team named EB45, or an internal reference to a class of models. - **X1** – This suggests a series of models labeled X1. There’s *x1-1-preview-0915* (perhaps the first preview on Sep 15) and *x1-turbo-0906*. Similar to EB45, X1 might be another internal model or a competition submission series, with a turbo version for speed.

Both EB45 and X1 series are likely not publicly known model names; they exist within LMArena to anonymize cutting-edge experiments (which might later be released under real names or might be combinations of known base models with special tuning).

- `EB45-turbo`
- `EB45-turbo-vl-0906`
- `EB45-vision`
- `x1-1-preview-0915`
- `x1-turbo-0906`

Anonymous IDs

In some cases, LMArena simply uses generic labels for anonymous models. These are placeholder names that reveal nothing about the model's origin: - **anonymous-1111, 1010, 915, 922, 925** – A series of anonymous models identified only by dates (e.g., 11/11, 10/10, 9/15, 9/22, 9/25 of presumably 2025). These could be multiple entries of various models (possibly user-contributed or under evaluation) where anonymity was preserved. - **lmarena-internal-test-only** – A label clearly indicating a model used purely for internal testing on the platform, not meant for public use. - **not-a-new-model** – A tongue-in-cheek label (perhaps to quell speculation when an entry was just a baseline or duplicate model tested blind). - **stephen-v2 / stephen-vision-csfix** – These appear to use a person's name "Stephen" as a codename, perhaps an internal project or a developer's custom model. *vision-csfix* might indicate a vision model with some fix applied (CS could mean "common sense" or a bug fix code). Using a first name could have been a quick placeholder to track a test model.

These anonymous IDs ensure that during evaluation, participants or analysts focus on performance and behavior rather than brand. Only after testing would organizers reveal (or not) which actual model corresponded to each anonymous tag.

- `anonymous-1111, 1010, 915, 922, 925`
- `lmarena-internal-test-only`
- `not-a-new-model`
- `stephen-v2 / stephen-vision-csfix`

Abstract Codenames

Finally, LMArena's list shows a set of **abstract codenames**. These likely represent either: 1. *Placeholder names for blind evaluation* – models from various sources given random names. 2. *Internal prototype models* – developed by the Arena team or partners, named creatively.

The codenames are evocative (often two-word combinations or compound words). They don't reveal any specific info about the model, which is intentional. Some examples: - **aegis-core** – "Aegis" suggests protection or shield, perhaps a model focusing on alignment/safety. - **blackhawk, skyhawk, seahawk** – various "hawk" names, possibly a series of fast or vision-focused models (hawk eyesight?), or just cool codenames. - **blitzphase, dashspark, swiftflare, rushstream** – all imply speed or sudden action, maybe fast reasoning models or rapid generators. - **bridge-mind, integrated-info** – suggests combining knowledge or modalities ("bridge"), maybe multimodal integrators. - **dark-dragon, silentnova, whisperfall, winter-wind** – poetic names, could be larger models with these as project names. - **evo-logic, gauss,**

newton – scientific/mathematical references (Gaussian, Isaac Newton), possibly logical reasoning models. - **flying-octopus** – whimsical, perhaps a multitasking or multi-armed (multi-tool) agent. - **frame-flow** – sounds like a video or sequential data model. - **leepwal** – this one is unclear (could be an anagram or inside joke). - **micro-mango** – whimsical, maybe a small model (micro) with a fun code word. - **monster, neon** – straightforward names, maybe indicating a very large model ("monster") or a flashy experimental one ("neon"). - **monterey** – possibly referencing the location (Monterey), used as a codename for a model (companies often use place names for projects). - **nightride-on / v2** – "Nightride" sounds like a project name, with a version 2 as well. Could hint at a model built for nighttime or dark-themed content (or just a random cool name). - **rain-drop, redwood, route66** – nature or journey themed names; Redwood might hint at a large model (big tree), Route66 at a pathfinding or long-sequence model. - **silvandra** – not obviously meaningful, could be a fantasy name. - **sunshine-ai** – a cheerful name, perhaps an aligned AI model (sunshine implying harmlessness). - **voltwhirl, viper** – high energy names, possibly rapid or "biting" models. - **whisperfall** – evokes something with language (whisper) and falling (maybe a model that "distills" or compresses info quietly). - **winter-wind** – another poetic name, potentially a model tuned for a specific domain or simply a codename.

All these abstract names indicate instances where models were entered into LMArena without revealing their true identity, often for blind evaluation. They make the Arena feel like a code-named battleground, where each model is a contender with a secret identity. Only the organizers or the model submitters know which actual AI system corresponds to which codename.

- `aegis-core`
- `blackhawk`
- `blitzphase`
- `bridge-mind`
- `dark-dragon`
- `dashspark`
- `evo-logic`
- `flashstride`
- `flying-octopus`
- `frame-flow`
- `gauss`
- `holo-scope`
- `integrated-info`
- `leepwal`
- `micro-mango`
- `monster`
- `monterey`
- `neon`
- `newton`
- `nightride-on / v2`
- `rain-drop`
- `redwood`
- `route66`
- `rushstream`
- `seahawk`

- `silentnova`
- `silvandra`
- `skyhawk`
- `sunshine-ai`
- `swiftflare`
- `voltwhirl`
- `viper`
- `whisperfall`
- `winter-wind`

Conclusion: The above lists capture all models present in LMArena, covering a broad landscape of AI systems up to late 2025. From top-tier corporate models (Google's Gemini, OpenAI's GPT-5, etc.) to niche and experimental ones, LMArena serves as a comprehensive arena to compare AI models. Each codename or version tells a part of the story of rapid AI evolution – and the diversity of names reflects the many organizations and ideas competing in the AI space [2] [5]. This exhaustive catalog underscores how 2024–2025 saw an explosion of model releases, with LMArena acting as a centralized ring where they all can be evaluated on equal footing.

Sources: LMArena model list (user-provided compilation); OpenAI, Google, Anthropic model info; performance index for select models [5] [12]; Baidu AI Cloud report on ERNIE 5.0 [10]; OpenRouter model descriptions [2] [1]; others.

---

[1] [2] [4] [6] [8] openrouter.ts
https://github.com/AmoyLab/Unla/blob/8464abbace6b55b8f36a5dacec631c7255c288b4/web/src/config/ai-models/openrouter.ts

[3] [5] [7] [11] list_of_AI_Models.md
https://github.com/raveendratal/PysparkRaveendra/blob/3367f3723e76273eef63dccccc9e7d5cc184165f/list_of_AI_Models.md

[9] MODELS.md
https://github.com/nexicore-digitals/nexus-llm/blob/03e33d0cef03088233b4ecbd28ccc2ff70a16702/MODELS.md

[10] BIDU_3Q25_presentation.txt
https://github.com/vanillefatale/earnings-transcripts/blob/bee0cfec05cf1b998c83d48eb2ab54c27bd338a8/generator/1-done/BIDU_3Q25_presentation.txt

[12] CURATED-MODELS.md
https://github.com/Thirteen88/ish-automation/blob/970ec3e76e93fb90a402d5dd824e8f5e64c58cc5/CURATED-MODELS.md