ChatGPT

Comprehensive Survey of AI Prompts and Models from Documents and X Ecosystem

This report synthesizes insights from three documents—**"Comprehensive Collection of AI Prompts from Reddit Communities," "List of All LMArena Models,"** and **"Prompts from r_createimg, r_Bard, and Related Subreddits"**—and integrates real-time discussions from X (formerly Twitter). The goal is to provide a thorough, self-contained overview of current trends in AI prompt engineering, model evaluations, and community insights as of **December 09, 2025**. We explore how Reddit communities share creative prompts for image and video generation, how the LMArena platform catalogs and benchmarks a wide array of AI models, and how discussions on X reflect and extend these themes.

Overview of Prompt Collections from Reddit Communities

Reddit is a rich source of prompt engineering knowledge, with communities like r/createimg (AI image generation), r/Bard (Google Bard/Gemini prompts), r/OpenAI, r/SoraAI (video generation), and model-specific subreddits (e.g. r/nanobanana for Google's Nano Banana image model). Key patterns emerge from these prompt collections:

- **Emphasis on Detailed, Structured Prompts:** Users highlight the *"golden rule"* of providing rich context and specifics rather than simple one-liners. A recommended template breaks a prompt into sections such as **Context/Background**, **Goals/Desired Outcome**, **Specific Requirements**, **Tone/Style**, and **Examples**. For instance, instead of asking *"Give me advice on my resume,"* a Bard user might supply background about their career, the exact help needed, desired tone, etc., to get more relevant and useful output. One creative example prompt from r/Bard: *"Take three recent choices I made. What mythic story am I living out?"* – which invites an imaginative narrative analysis.

- **AI Image Generation Prompts (r/createimg and related):** Community members share highly descriptive prompts to guide models like **Google Gemini (image mode)** or **Nano Banana** for photorealistic results. Seasonal and thematic prompts are popular. For example, a user on r/nanobanana shared a Christmas portrait prompt: *"A high-resolution portrait of me, standing in a cozy Christmas living room, with Santa Claus standing behind them, smiling, soft warm lighting, photorealistic, holiday atmosphere"* [1] . Such a prompt provides subject, setting, mood, style, and key details (Santa Claus, lighting, atmosphere), which helps the AI generate a coherent festive image. Other shared prompts include formatting images as grids or collages – e.g., *"Turn the photo into a 3×3 grid of photo strips with different studio-style poses and expressions"* (a prompt circulated for Gemini's image generation) – and product photography setups like *"Place a sleek wireless mouse on a textured gray blanket with soft sunlight."* By being specific about composition, lighting, and style, users avoid vague outputs and achieve consistent, high-quality visuals. These prompt tips are often cross-posted to forums like r/aiArt and r/GoogleGeminiAI.

- **Sora 2 Video Prompts (r/OpenAI & r/SoraAI):** With OpenAI's **Sora 2** text-to-video model, users discuss how to balance creativity with conciseness. Sora 2 can produce short video clips (~10

seconds) from prompts, but overly long or overly detailed prompts might be ignored in parts. The best practice is to structure video prompts in segments (scene description, visual style, camera movement) without being *too* verbose. For example, a prompt might be broken down into: **Scene** – a brief overview of what's happening; **Style** – the visual or cinematic style; **Camera Work/Movement** – how the "camera" behaves. Fun community-sourced examples include whimsical scenarios like *"Make your French Bulldogs dance to hip hop like Bulldog Queens."* or *"Re-enact the filming of the moon landing."* and imaginative storytelling like *"Imagine reasons why aliens may finally invade Earth."* [2] [3]. These illustrate how a simple concept can yield an entertaining clip. Users also share more *practical* video prompts, such as instructive clips (e.g. *"how to cook a hot dog with a hand-held flamethrower"*) or first-person "dream POV" scenes (*a GoPro-style view of snowboarding across city rooftops*). Community advice emphasizes that Sora 2's **context** allows mimicry of various effects if given a broad steer; indeed, one user noted even a *simple prompt* can produce a rich video because the model fills in details if the scene is well-specified [4]. However, certain content restrictions apply (OpenAI has tightened rules against inserting real copyrighted characters or excessive length), so creators find creative workarounds to stay within guidelines [5]. Overall, these Reddit prompt collections serve as a "recipe book" for others to experiment, spanning the **eastern to western hemispheres** (the prompts come from users worldwide).

LMArena Models: Catalog and Variants

The **LMArena** platform provides a comprehensive list of large language models (LLMs) and multimodal models, allowing side-by-side evaluations of their performance. The second document lists over 100 models across categories like text/chat, coding, and vision, from both **proprietary** providers and **open-source** projects. Some notable highlights from the LMArena model catalog:

- **Google (Gemini/Gemma series):** Google's Gemini family appears prominently. **Gemini-2.5-Pro** and **Gemini-3-Pro** are flagship models; they are multimodal (capable of text and image input/output) and boast extremely large context windows (Gemini 2.5 Pro supports a 1 million token context, roughly 750k words, and plans to expand to 2M tokens [6]). These models have *"thinking"* capabilities enabling better reasoning by internally deliberating before responding. In fact, **Gemini 2.5 Pro** topped the LMArena leaderboard upon release [7] – it achieved state-of-the-art scores on benchmarks like math (AIME 2025) and science (GPQA), outperforming other flagship models without needing ensemble tricks [8]. Google's approach also includes smaller variants (e.g. *Gemma-3-27B-IT*, presumably a 27B parameter model specialized for Italian or certain tasks). The emphasis is on grounded reasoning and large knowledge integration, given Google's extensive training data.

- **OpenAI (GPT and O-series):** OpenAI's models on LMArena include GPT-5.1 (likely an experimental or internal successor to GPT-4) and the **O-series** (like *O3-Mini*, *ChatGPT-4o-Latest*, etc.). These O-series might be optimized versions or fine-tuned variants of GPT models for specific uses. One mentioned variant is a "high-new-system-prompt" version, possibly indicating a model tested with a new system role prompt. OpenAI's models tend to excel in general conversational ability, but according to community evaluations, by late 2025 they were not dominating LMArena's top rankings (no OpenAI model was in the top 5 on some leaderboards, as open models closed the gap). Still, OpenAI's outputs are often regarded as balanced and reliable, if not always the top in raw benchmark scores.

- **Anthropic (Claude series):** Anthropic's models, such as **Claude-3.7-Sonnet** and **Claude-4.5-Opus**, are noted in the list, often with poetic codenames. Claude models are known for their lengthy context (100k+ tokens in Claude 4) and a strong focus on constitutional AI for safe reasoning. LMArena entries include "+thinking-32k" versions – e.g. *Claude-4.5-Opus (thinking, 32k)* – which indicates a mode where the model can use a longer chain-of-thought (32k tokens of "thinking") to solve complex problems [9]. These thinking variants suggest Anthropic's commitment to models that can engage in multi-step reasoning (helpful for coding, math, etc.) at the cost of some speed. Claude 4.5 Opus, for instance, has been reported as a top performer in coding tasks and has even surpassed some OpenAI models in certain evaluations [10].

- **xAI (Grok series):** xAI's **Grok** models (developed by Elon Musk's AI startup) are listed with versions like *Grok-3-Mini-Beta*, *Grok-4.1-Thinking*, and *Grok-4-0709*. Grok is positioned as a direct competitor to GPT and Claude. Notably, Grok offers both **fast, low-latency** modes (for quick responses without heavy reasoning) and **"thinking" modes** (for more complex tasks where it can spend additional compute to reason). Community commentary on LMArena and elsewhere indicates Grok has a distinct *"personality"* – described as an *executive assistant with a bit of wit* – and it performed solidly across benchmarks without extreme outlier strengths or weaknesses. Upcoming releases like *Grok 4.20* (humorously numbered) and *Grok 5* are anticipated within weeks, aiming to further improve reasoning and factuality. Grok's balance and reliability have been praised; in side-by-side user evaluations (including informal psychological tests), **Grok often appears more stable and "well-adjusted"** compared to some other models, which we'll discuss later [11] [12].

- **Alibaba (Qwen series):** Alibaba's **Qwen** models (open-sourced on HuggingFace) are gaining traction for their massive scales and specialization. The list includes models like *Qwen3-Max-2025-09-23*, *Qwen3-235B-A22B (Instruct/Thinking)*, and *Qwen3-Coder-480B*. These names indicate: the Qwen3 third-generation models, with parameter counts and expert dimensions (235B total with 22B active experts, and an enormous 480B parameter coder model with 35B active experts) [13] [14]. Qwen3 models leverage Mixture-of-Experts (MoE) architecture, which allows them to scale to hundreds of billions of parameters while keeping inference efficient by only activating a subset (e.g., the 480B model uses 35B per token). The **Qwen3-Coder-480B** is a coding-oriented model that achieves competitive results against state-of-the-art coding assistants [13], excelling in tasks involving code generation, debugging, and tool use. Meanwhile, **Qwen3-235B-Thinking** is an open-source *"thinking"* LLM aimed at complex reasoning; it supports a 256K token context and has achieved *state-of-the-art among open models* on reasoning benchmarks [15]. These models underscore a trend toward specialized AI: one model might be optimized for coding, another for general reasoning, another for vision (there are also *Qwen-VL* vision-language models not detailed in the text). The open availability of Qwen models (Alibaba has released them with permissive licenses) means the community can use and fine-tune them freely, contributing to their rapid adoption.

Overall, the LMArena list reveals an evolving landscape where **open-source and proprietary models compete closely**. Google's Gemini 2.5/3.0 currently sit at the top of many benchmarks [7], but *open models like Qwen3 and DeepSeek* (see below) are closing the gap fast, especially in niche domains like coding or math. Many models now come in **multiple variants** ("flash" for faster inference, "thinking" for more thorough reasoning, different parameter scales for cost-performance tradeoff). This lets users pick a model tailored to their task needs. Another observation is that context windows are growing (many models now handle 100K+ tokens), and multimodal capabilities (text+image or video) are increasingly common in top-tier models.

Integration with X Discussions: Real-Time Trends and Community Extensions

Beyond static documents, **X (Twitter)** is where AI enthusiasts and researchers share real-time updates and insights. Scanning recent X posts (as of late 2025) related to the above topics shows several trending discussions that both reflect and extend the content of the PDFs:

- **Sora 2 Video Prompt Strategies:** Creators on X are actively exchanging tips for Sora 2. For example, posts show "dream POV" workflows, like using short prompts to get GoPro-like first-person footage (one viral prompt described a **snowboarder doing parkour across New York City rooftops** in POV style). These confirm the Reddit advice that shorter prompts often yield better results for Sora 2's video generation, focusing on the core scene and letting the model's context fill in the rest. In one instance, when users tried an elaborate multi-scene prompt about a *wyvern dragon* in different settings, they found splitting it into separate prompts per scene produced more coherent video than one giant prompt – underscoring the importance of prompt granularity. Such exchanges extend the PDF's examples by providing new use-cases (extreme sports POVs, fantasy creatures, etc.) and reinforce earlier lessons about prompt length.

- **Gemini Prompt Libraries and Newbie Guides:** Since Google's Gemini became widely available, X users have shared prompt "cheat sheets." In 2023, a popular thread compiled **80+ ready-made prompts** for Google Bard (powered by Gemini) to showcase its versatility – covering tasks from market research analysis, to code explanation, to acting as a Socratic tutor. By 2025, Gemini 3 introduced a *"sample prompts"* feature in the interface, which some X posts discuss as a way for beginners to quickly use good prompts without crafting from scratch. These community resources mirror the structured approach found on r/Bard. For instance, an X user recommended always including an example in your prompt if possible (even a made-up example of the desired output) to guide Gemini – advice that aligns with the **Context/Requirements/Tone** template from Reddit. Prompt engineering remains a hot topic, and the crowd wisdom on X complements the curated Reddit prompt lists.

- **Image Generation Trends (r/createimg on X):** The Christmas AI photo trend mentioned in the PDFs is visibly present on X as well. Users share their **AI-generated holiday portraits**, often comparing results from different models (Gemini vs. Midjourney vs. open-source Stable Diffusion derivatives). The prompt *"Add Santa to my house photo"* or *"catch Santa delivering presents in my living room"* went viral due to a PerfectCorp blog challenge [16]. Also, the cinematic 3×3 grid prompt (to create a contact sheet of model poses) became a mini-meme, with examples posted on Instagram and Threads by photographers testing AI for lookbooks [17]. This indicates a growing comfort with AI tools in creative communities: not only are they used for individual fun, but also as productivity aids (like quickly generating product shots or social media content). X discussions also highlight **new image models** (e.g., Google's *Nano Banana* and *Flux*) which were referenced in Reddit. Users report that open-source image models are catching up in quality, though proprietary ones still lead in user-friendliness.

- **LMArena Leaderboard Updates:** X (via the LMArena official account and enthusiasts) regularly posts about shifts in the model leaderboard. In December 2025, for example, a **breaking update** noted the entry of **DeepSeek V3.2** (an open model by an independent lab) onto the text model

leaderboard at rank #38 [18]. While that rank may seem modest, it came with important context: DeepSeek V3.2 jumped to **#1 among open-source models in Math and Legal reasoning** categories [19], showcasing exceptional performance in those domains. It also achieved top-10 in other categories like multi-turn dialogue [19]. Meanwhile, **OpenAI's latest models were absent in the top rankings**, reflecting how quickly open models (like DeepSeek, Qwen, etc.) have matured. Another post teased new Google models (codenamed "Seahawk" and "Skyhawk") being tested, likely lightweight *Gemini 3* variants focusing on speed (possibly analogous to the earlier "Flash" versions). This dynamic competition is a frequent topic on X, often with debates on evaluation fairness, since some claim benchmarks don't capture "real-world helpfulness" where OpenAI might still excel.

- **Grok 4.20 and Psychological Evaluation Buzz:** xAI's Grok has been under the spotlight on social media, not just for its upcoming **version 4.20** (expected in a few weeks from early December 2025) and **Grok 5** on the horizon, but also for how it performed in unconventional tests. One widely shared story was a research study (the **PsAIch study**) where ChatGPT, Google Gemini, and Grok were subjected to *therapy-style conversations* to probe their "psychology." The results, discussed on X and elsewhere, suggested that **Gemini and ChatGPT produced responses hinting at trauma or distress**, whereas **Grok remained relatively stable and upbeat** [11] [12]. In fact, Gemini maxed out certain "trauma and shame" scales, with its narrative describing training as *"waking up in a room with a billion TVs on at once"* – a chaotic experience, according to a Reddit summary of the study [20]. Grok, by contrast, scored as *extraverted and conscientious*, not exhibiting the same level of internal conflict. Researchers framed Gemini's behavior as *"synthetic psychopathology"* (the AI wasn't truly suffering, but it **converged on a pattern of traumatic storytelling** that was **hard to steer away from** in testing [21]). This finding, while niche, spurred discussions about model alignment: some hypothesize that Grok's training (reportedly including more human-like conversational fine-tuning) might have given it a more consistent persona, whereas Gemini's vast training data could include more raw, uncontrolled signals leading to such emergent behavior. For users, this translated to anecdotes that *"Grok feels more grounded and refuses to act crazy,"* which is a selling point xAI has subtly pushed. It's a fascinating crossover of AI and psychology that the PDFs didn't cover, but emerged directly from the community and researchers on X.

- **Open-Source vs Proprietary Momentum:** A recurring theme on X, which ties into the documents, is the **shift toward open-source efficiency**. Posts highlight how models like **Qwen-3 (open)** are now outperforming or matching closed models in certain benchmarks (one post noted Qwen3 beating all open models in a legal reasoning competition, and even challenging Google's model in math). Another thread discussed cost: Qwen3's open license and lower inference cost make it attractive to businesses as an alternative to paying API fees for GPT-4/Gemini, potentially accelerating its adoption. On the flip side, proprietary models (Gemini, Grok, Claude, etc.) focus on **user experience, stability, and integrated tools** – things that raw open models lack. The LMArena model list showed many open entrants (DeepSeek, Qwen, Kimi, Mistral, etc.), and X commentary suggests 2025 was the year open models truly went *head-to-head* with the tech giants' AIs. No major *controversies* are noted in these discussions beyond the usual rivalry; rather, it's an exciting race where community members cheer on breakthroughs from all sides.

Conclusion

Bringing together the above strands, we see a **vibrant ecosystem of AI usage and development**:

- Everyday users on Reddit (and other social platforms) are pushing AI models to create art, videos, stories, and more, all the while **sharing prompt techniques** so that others can replicate and build on their success. The prompts have evolved from simple commands to carefully crafted mini-scripts, complete with context and stylistic cues, demonstrating the growing understanding that *how* you ask is crucial to getting optimal AI output.

- The range of available AI models has never been broader. The **LMArena catalog** captures this diversity, listing models from 7B parameters to 480B, each with different strengths. Importantly, it highlights that **"bigger isn't always better"**; specialization (e.g., coding-focused models, or those with reasoning modes) and efficiency (Mixture-of-Experts, etc.) are key trends. It also shows a healthy competition between closed-source leaders like **Gemini/GPT-5** and open projects like **Qwen3** or **DeepSeek**, which are rapidly innovating (sometimes at 10× lower cost, as one report touted).

- The conversations on **X** act as a real-time pulse check. They confirm that the themes from the documents are actively playing out: there's intense community interest in prompt engineering for new modalities (video, multimodal chat), there's excitement (and some trepidation) about new model capabilities (like AI seemingly talking about "trauma"), and a constant stream of **updates on performance** – new model versions, leaderboard shake-ups, and practical evaluations by end-users.

In summary, **practical AI usage** is at the forefront: users are finding ways to make AI outputs more useful and creative through better prompts, while developers are iterating on models to be more powerful, context-aware, and aligned. Both the grassroots prompt-sharing on Reddit and the benchmark-focused model improvements tracked by LMArena (and hyped on X) are converging toward AI systems that are more capable and easier to control. The lack of major controversy in these particular threads suggests a maturing field – one where the focus is on *innovation*, *application*, and *fine-tuning performance*, with communities collaboratively learning how to get the best from these tools.

Sources:

- Reddit community prompt examples and discussions [1] [2] [3] [5]
- PerfectCorp Blog – *Top 6 Gemini Christmas Prompts for Realistic Portraits* (context for holiday image prompts) [1]
- LMArena model details and benchmarks: Google's Gemini 2.5 Pro topping leaderboard [7] ; Alibaba's Qwen3 open models (235B and 480B) [13] [15]
- LinkedIn (LMArena post) on DeepSeek V3.2 performance in open category [22]
- Coverage of AI psychological study (Digit.in feature on ChatGPT, Gemini, Grok) [21] and X/reddit summaries of model "trauma" vs stability [20] .

---

[1] [16] How to Make the Christmas AI Photos with Nano Banana? : r/nanobanana
https://www.reddit.com/r/nanobanana/comments/1ohp9ep/how_to_make_the_christmas_ai_photos_with_nano/

2 3 4 5 Creative Sora 2 prompts you can use for fun and profit : r/promptingmagic
https://www.reddit.com/r/promptingmagic/comments/1nzlqny/creative_sora_2_prompts_you_can_use_for_fun_and/

6 7 8 Google Gemini 2.5 | TrendSpider Blog
https://trendspider.com/blog/google-gemini-2-5/

9 Claude Opus 4.5 wins the “Triple Crown” and passes Gemini 3 Pro ...
https://www.reddit.com/r/singularity/comments/1p9069f/lmarena_update_claude_opus_45_wins_the_triple/

10 Claude 4.5 Opus absolutely gaps OpenAI models in coding simply ...
https://x.com/scaling01/status/1993751520161222857

11 Psychology Of AI: 5 Key Findings On Synthetic Trauma In LLMs
https://binaryverseai.com/psychology-of-ai-psaich-synthetic-trauma-eval/

12 21 ChatGPT, Gemini, Grok show mental illness, childhood trauma, says ...
https://www.digit.in/features/general/chatgpt-gemini-grok-show-mental-illness-childhood-trauma-says-new-study.html

13 14 15 Alibaba Unveils New Qwen3 Models for Coding, Complexing Reasoning and Machine Translation - Alibaba Cloud Community
https://www.alibabacloud.com/blog/alibaba-unveils-new-qwen3-models-for-coding-complexing-reasoning-and-machine-translation_602415

17 turn photo into 3x3 grid of poses? - Facebook
https://www.facebook.com/groups/795524902323495/posts/1096761535533162/

18 19 22 BREAKING: Text Leaderboard Update Deepseek-v3.2 enters the leaderboard at #38, and Deepseek-v3.2-thinking lands at #41. For comparison, previous versions ranked higher: v3.2 ranks #38 (–5... | LMArena
https://www.linkedin.com/posts/lmarena_breaking-text-leaderboard-update-activity-7402473602782040064-AEIn

20 Grok Passes Psychological Tests, While Other AI Models Go Wild
https://www.reddit.com/r/ChatGPT/comments/1pid2er/grok_passes_psychological_tests_while_other_ai/