

Chapter 2 - Current LLM offer

LLMs for everything!

Murat Yıldızoğlu
Bordeaux School of Economics
University of Bordeaux
UMR CNRS-INRAE 6060
<https://yildizoglu.fr>

Master 1, IREF-ERDS

A very extensive LLM offer

- On [Hugging Face](#), I count 1 339 964 models (on Jan 27, 2025) and still counting...
- With a variable market share.
- Some very confidential and/or very specialized.
- Some on every media (ChatGPT, anyone?).
- My objective:
 - Giving you a preview of the current offer and
 - To share with you my very partial and subjective experience with the most promising LLMs.
- For different tasks.
- Hoping to discover new ones thanks to your experience.
- Also exploring running some of them on your computers.
- Let's the fun start! :-)

Synthesis on their main characteristics and application domains

- **Text Summarization:** Ability to condense long texts into concise summaries.
- **General Text Generation:** Versatility in generating human-like text for various purposes.
- **Specialization Areas:** Specific domains where the platform excels (e.g., conversational AI, creative writing).
- **Translation:** Performance in translating text between languages.
- **Image Generation:** Ability to generate images from text prompts.
- **Image Understanding:** Ability to analyze and interpret images.
- **Other Relevant Domains:** Additional areas where the platform is particularly strong (e.g., code generation, ethical AI).

Major platforms I

Platform	Text Summarization	General Text Generation	Specialization Areas	Translation	Image Generation	Image Understanding	Other Relevant Domains
ChatGPT (OpenAI)	Excellent	Excellent	Conversational AI, Customer Support	Good	No	No	Code generation, Few-shot learning
Claude (Anthropic)	Good	Excellent	Ethical AI, Business Applications	Good	No	No	Long-context retention
LLaMA (Meta)	Good	Good	Research, Open-source projects	Moderate	No	No	Low-resource environments
Mistral (Mistral AI)	Good	Good	Enterprise applications, Multilingual tasks	Excellent	No	No	Cost-effective deployment
Gemini (Google)	Excellent	Excellent	Multimodal tasks, Reasoning	Excellent	Yes	Yes	Integration with Google ecosystem

Major platforms II

DeepSeek (DeepSeek AI)	Good	Good	Finance, Health-care, Real-time decision-making	Moderate	No	No	High-performance applications
Cohere	Good	Good	Enterprise applications, Customer support	Excellent	No	No	Multilingual tasks

Major platforms III

- **Strengths:**

- Highly versatile for conversational AI, content creation, and customer support.
- Strong performance in zero-shot and few-shot learning.
- Widely adopted and integrated into various applications.

- **Weaknesses:**

- Prone to generating incorrect or nonsensical information (“hallucinations”).
- Limited context window for long conversations.
- High cost for API usage at scale.

- **Strengths:**

- Focus on ethical AI and reducing harmful outputs.
- Strong conversational abilities and context retention.
- Customizable for specific business needs.

- **Weaknesses:**

- Less widely adopted compared to ChatGPT.
- Limited public information on model architecture and training.

- **Strengths:**
 - Open-source, making it accessible for researchers and developers.
 - Efficient architecture with competitive performance.
 - Lower computational requirements compared to larger models.
- **Weaknesses:** -
 - Smaller scale compared to proprietary models like GPT-4.
 - Requires significant fine-tuning for specific tasks.

- **Strengths:**
 - Focus on efficiency and performance for enterprise use cases.
 - Strong multilingual capabilities.
 - Optimized for cost-effective deployment.
- **Weaknesses:**
 - Less documentation and community support compared to open-source models like LLaMA.
 - Still emerging, with limited real-world adoption data.

- **Strengths:**

- Multimodal capabilities (text, images, voice, and potentially video).
- Integration with Google's ecosystem (e.g., Search, Workspace).
- Strong performance in reasoning and complex tasks.

- **Weaknesses:**

- Limited public availability and transparency.
- High computational requirements for training and inference.

DeepSeek (DeepSeek AI)

- **Strengths:**

- Focus on high-performance, low-latency applications.
- Strong in niche domains like finance and healthcare.
- Optimized for real-time decision-making.

- **Weaknesses:**

- Limited public information on model architecture.
- Smaller user base compared to mainstream models.

Platform	Text Summarization	General Text Generation	Specialization Areas	Translation	Image Generation	Image Understanding	Other Relevant Domains
Flux (Flux AI)	Moderate	Good	Creative writing, Storytelling	Moderate	No	No	User-friendly content generation
Napkin (Napkin AI)	Moderate	Moderate	Marketing, Social media content	Moderate	No	No	Small business applications
Dalee (Dalee AI)	No	No	Artistic content generation, Text-to-image	No	Excellent	No	Design and creative tools
Diagram (Diagram AI)	No	No	Visual content generation, Infographics	No	Excellent	No	Technical documentation

Flux (Flux AI)

- **Strengths:**
 - Designed for creative tasks like storytelling and content generation.
 - User-friendly interface for non-technical users.
 - Strong in generating engaging and imaginative content.
- **Weaknesses:**
 - Limited applicability for technical or analytical tasks.
 - Less robust in handling factual or data-driven queries.

- **Strengths:**

- Specialized in creative and artistic content generation (e.g., text-to-image).
- Strong integration with design tools and platforms.
- User-friendly for artists and designers.

- **Weaknesses:**

- Limited applicability outside creative domains.
- Requires significant fine-tuning for non-artistic tasks.

- **Strengths:**

- Focus on visual and diagrammatic content generation.
- Strong in generating flowcharts, infographics, and visual explanations.
- Useful for technical documentation and education.

- **Weaknesses:**

- Limited text generation capabilities compared to other LLMs.
- Niche use case, less versatile for general-purpose tasks.

- **Strengths:**

- Focus on simplicity and ease of use for small businesses.
- Affordable pricing for startups and SMEs.
- Strong in generating marketing content and social media posts.

- **Weaknesses:**

- Limited customization and scalability for larger enterprises.
- Less advanced in handling complex or technical tasks.

Code generation I

Platform	Developer	Code Generation Characteristics	Supported Languages
GitHub Copilot	GitHub (Microsoft)	Autocompletes code, suggests functions, and generates snippets	Python, JavaScript, TypeScript, Go, Ruby, Java, C++, and more
OpenAI Codex	OpenAI	Generates code from natural language prompts, supports debugging	Python, JavaScript, TypeScript, Ruby, Java, C++, and more
Replit's Code LLM	Replit	Real-time code generation and collaboration in IDE	Python, JavaScript, Java, C++, and 50+ languages
Amazon CodeWhisperer	Amazon	Autocompletes code, suggests best practices, and integrates with AWS	Python, Java, JavaScript, TypeScript, C#, and more
Tabnine	Tabnine	AI-powered code completion, supports multiple IDEs	Python, JavaScript, Java, C++, and 20+ languages
CodeT5	Salesforce	Code generation, summarization, and translation	Python, Java, JavaScript, C#, and more
AlphaCode	DeepMind	Competitive programming and algorithmic problem-solving	Python, C++, Java, and more
Cursor	Cursor.sh	AI-powered code editor with autocompletion, refactoring, and debugging	Python, JavaScript, TypeScript, Java, C++, and more

- **Strengths:**
 - Seamless integration with Visual Studio Code.
 - Supports a wide range of programming languages.
 - Continuously learns from public code repositories.
- **Weaknesses:**
 - Can generate insecure or inefficient code.
 - Requires careful review of generated code.

General platforms

- ChatGPT, Claude, and Gemini are perfectly capable of
 - Generating code following your description of your program
 - Explaining the functions of different components of an existing program
 - Debugging errors in your code
- Buy, more specialized platforms are better integrated with development tools.

- **Strengths:**

- Powerful natural language-to-code capabilities.
- Excellent for prototyping and debugging.
- Broad language support.

- **Weaknesses:**

- High cost for API usage at scale.
- Limited understanding of complex or niche domains.

Replit's Code LLM

- **Strengths:**

- Real-time collaboration in a browser-based IDE.
- Supports over 50 programming languages.
- Beginner-friendly with interactive features.

- **Weaknesses:**

- Limited offline functionality.
- Less advanced than GitHub Copilot or Codex.

- **Strengths:**
 - Strong integration with AWS services.
 - Focus on security and best practices.
 - Supports multiple popular languages.
- **Weaknesses:**
 - Limited to AWS-centric workflows.
 - Less mature compared to GitHub Copilot.

Tabnine

- **Strengths:**
 - Works offline, ensuring data privacy.
 - Supports a wide range of IDEs and languages.
 - Lightweight and fast.
- **Weaknesses:**
 - Less advanced than GitHub Copilot or Codex.
 - Limited natural language understanding.

- **Strengths:**

- Open-source and customizable.
- Strong in code summarization and translation.
- Supports multiple programming languages.

- **Weaknesses:**

- Requires significant fine-tuning for specific tasks.
- Less user-friendly for non-technical users.

AlphaCode

- **Strengths:**

- Specialized in competitive programming and algorithms.
- Strong performance in solving complex problems.
- Supports multiple languages.

- **Weaknesses:**

- Limited to algorithmic and competitive programming.
- Not suitable for general-purpose code generation.

- **Strengths:**

- AI-powered code editor with autocompletion, refactoring, and debugging.
- Supports multiple programming languages.
- Lightweight and fast, with a focus on developer productivity.

- **Weaknesses:**

- Still emerging, with fewer integrations compared to GitHub Copilot.
- Limited offline functionality.

Less known platforms I

Platform	Text Summarization	General Text Generation	Specialization Areas	Translation	Image Generation	Image Understanding	Other Relevant Domains
Jurassic-1 (AI21 Labs)	Excellent	Excellent	Creative writing, Story-telling	Good	No	No	User-friendly API
Bloom (Big-Science)	Good	Good	Multilingual tasks, Ethical AI	Excellent	No	No	Open-source research
Ernie Bot (Baidu)	Good	Good	Chinese language tasks, Multimodal applications	Excellent	Yes	Yes	Integration with Baidu ecosystem
GPT-NeoX (EleutherAI)	Good	Good	Open-source text generation, Research	Good	No	No	Community-driven development
WuDao 2.0 (BAAI)	Excellent	Excellent	Scientific and technical applications	Excellent	Yes	Yes	Multimodal tasks

Less known platforms II

Hugging Face Transformers	Excellent	Excellent	Customizable NLP tasks, Research	Excellent	No	No	Wide range of pre-trained models
Character.AI	Moderate	Good	Conversational agents, Role-playing	Moderate	No	No	Interactive story-telling

Platform for learning and research I

Platform	Self-Learning	Teaching	Scientific Research
ChatGPT /Claude ...	Excellent (explains complex concepts, Q&A)	Excellent (lesson planning, quiz generation)	Good (literature review, hypothesis generation)
Perplexity AI	Excellent (research-oriented Q&A, citations)	Good (teaching with verified sources)	Excellent (academic research, source tracking)
Elicit (by Ought)	Good (research paper summarization, Q&A)	Moderate (teaching research methods)	Excellent (automating literature reviews, hypothesis generation)
Scite	Good (interactive citation analysis)	Moderate (teaching citation skills)	Excellent (research validation, citation context)
Consensus	Good (evidence-based Q&A for learning)	Moderate (teaching evidence-based reasoning)	Excellent (research synthesis, evidence extraction)
Scholarcy	Excellent (summarizing research papers)	Good (teaching paper analysis)	Excellent (research paper summarization, extraction)
ResearchRabbit	Good (discovering related research papers)	Moderate (teaching research discovery)	Excellent (research paper recommendations, collaboration)
Typeset.io	Moderate (academic writing assistance)	Good (teaching academic writing)	Excellent (research paper formatting, collaboration)
Explainpaper	Excellent (simplifying research papers for learning)	Good (teaching paper comprehension)	Moderate (research paper analysis)
DeepSeek Academic	Good (domain-specific explanations)	Moderate (teaching in specialized fields)	Excellent (domain-specific research support)

NotebookLM (Google)	Very good. Multiple tools and aides	Moderate, some tools (podcast)	Moderate, literature synthesis
--------------------------------	-------------------------------------	--------------------------------	--------------------------------

Description of their roles

- **General LLMs:** Tailored for explaining complex topics, generating quizzes, and assisting with research tasks.
- **Perplexity AI:** Combines LLMs with verified sources, ideal for academic Q&A and research.
- **Elicit (by Ought):** Automates literature reviews and extracts insights from research papers.
- **Scite:** Analyzes citations to show how research is supported or contradicted.
- **Consensus:** Provides evidence-based answers by synthesizing findings from academic papers.
- **Scholarcy:** Summarizes research papers into digestible formats for learning and teaching.
- **ResearchRabbit:** Helps researchers discover and organize related papers.
- **Typeset.io:** Streamlines academic writing and formatting for researchers and educators.
- **Explainpaper:** Simplifies research papers for easier comprehension by students and researchers.

- 1 Downloading Ollama on your computer (Available for macOS, Linux, and Windows) from <https://ollama.com>.
- 2 Installing Ollama.
- 3 Search and choose models from the repository:
<https://ollama.com/search>.
- 4 For example, Deepseek-r1:
<https://ollama.com/library/deepseek-r1>.
- 5 Choose 7b version (small enough, 4.5GB) and install it by executing in a terminal window `ollama run deepseek-r1:7b`
- 6 It will download it first, then launch in Ollama.
- 7 For commodity, install a chat interface like [Enchanted for MacOS](#).
- 8 Start to test (see Conversation-DeepSeek.pdf).