

Build Your AI Assistant for Free

A GUIDE FOR NON-TECHIES



KDA CAPABILITIES
KNOW, DECIDE, ACT



ELECTRIC
MINDS

Agenda

Welcome & Introduction (15 mins)

Build Your AI Assistant for Free (20 mins) – Uli

Your AI intern (20 mins) – Keith

Sign up for the
AI Masterclass
in Singapore
on, happening
December
10th!



- Diverse Team across Singapore, Hong Kong, and Australia.
- Using AI to address global challenges through collaboration across startups, corporations, academia, and government.
- Make an impact while growing your skills and network.

→ Get involved: attend events, join projects, be part of the change!



December 10: AI Masterclass

Practical AI-Driven Workflows

- Discover practical workflows that incorporate AI tools and applications.
- Gain insights into how AI can streamline processes and enhance productivity.

Hands-On Experience with Use Cases Across Industries

- Engage in hands-on activities that showcase real-world AI applications.
- Explore use cases that span various industries and job functions, including marketing, finance, operations, and more.



KDA Capabilities



Keith B Carter



Uli Hitzel

AI For Humanity Forum 2024

📅 12 December 2024 | 9:00am - 5:00pm

📍 Catapult by Capitaland, 1 Rochester Park, #02-01 Rochester Commons, Singapore 139212

Register Now

Use the code **AI4HELECMINDS**
for 35% OFF!

AI FOR HUMANITY

FORUM 2024

CATAPULT BY CAPITALAND, 1 ROCHESTER PARK, SINGAPORE

12 DEC

9AM - 5PM

Dr James Ong

Dr James Ong

Dr James Ong

Renee Tan

Renee Tan

Renee Tan

Von Leong

Von Leong

Von Leong

Sioe Sioe Tan

Sioe Sioe Tan

Sioe Sioe Tan

John Cai

John Cai

John Cai

Andeed Ma

Andeed Ma

Andeed Ma

Harish Pillay

Harish Pillay

Harish Pillay

Saad Siddiqui

Saad Siddiqui

Saad Siddiqui

Chief Patrick

Chief Patrick

Chief Patrick

Tan Xinhui

Tan Xinhui

Tan Xinhui

Melis Ye Ye

Melis Ye Ye

Melis Ye Ye

Cheryl Loh

Cheryl Loh

Cheryl Loh

Paulo Italiani

Paulo Italiani

Paulo Italiani

Prof. David Lee

Prof. David Lee

Prof. David Lee

Oliver Kwan

Oliver Kwan

Oliver Kwan

Uli Hitzel

Uli Hitzel

Uli Hitzel

Brandon Peng Xiaoyu

Brandon Peng Xiaoyu

Brandon Peng Xiaoyu

Vibhav Mithal

Vibhav Mithal

Vibhav Mithal

Maryann Tseng

Maryann Tseng

Maryann Tseng

ORGANISED BY

CO - ORGANISERS

SPONSORS

TECHNOLOGY PARTNERS

AIH

WV

Global Fintech Institute

RIMAS

WILEY

ELECTRIC MINDS

ORIGAMI

ECOSYSTEM PARTNERS

Pruts.AI

Tan Network

Official Website

Scan to Register

2 Years of ChatGPT

- ChatGPT (2022) sparks conversational AI revolution
- 2024: Diverse landscape of chat services and models
- Claude.ai, Le Chat
- Open models, smaller models (e.g., LLaMA)
- Private/Offline, locally-run models gain popularity
- AI landscape offers a spectrum of choices
 - Free vs. paid
 - Online vs. offline
 - Varying model sizes and capabilities
 - Open vs. proprietary options

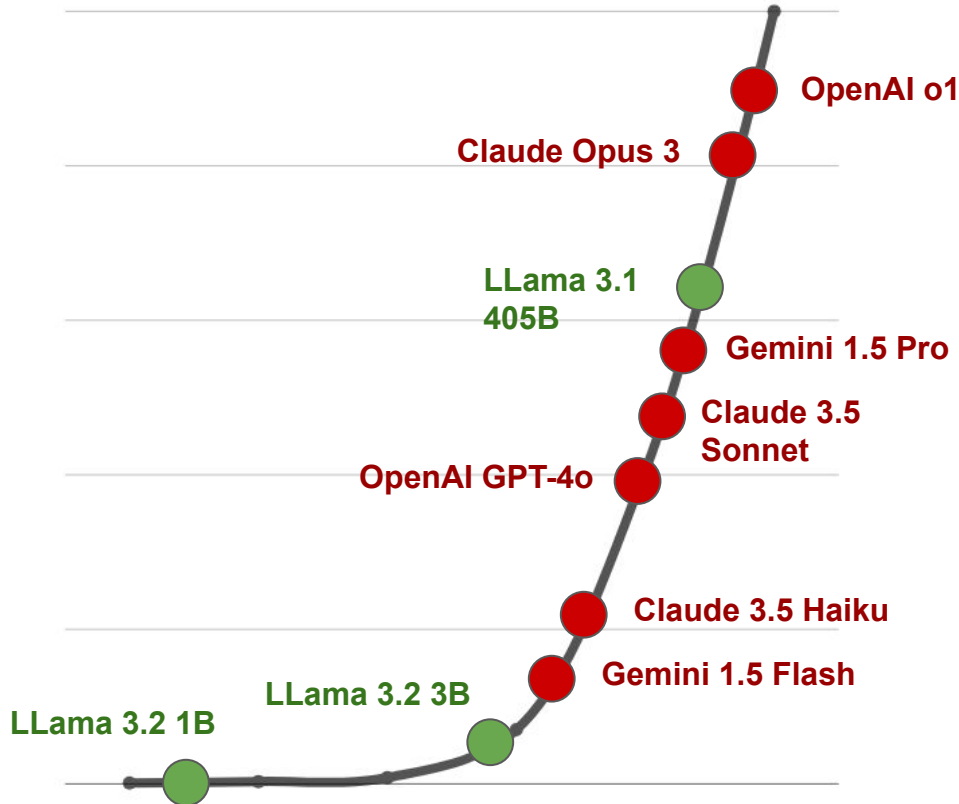


The LLM Zoo

- Open vs. Proprietary
- Different LLMs have varying capabilities and complexity levels
- Model size (parameters) generally correlates with performance
- More advanced tasks require larger, more capable models



Which One For What?



X-Axis: Model Size/Parameters

- **Small (1B to 3B):** Efficient and suitable for basic tasks.
- **Medium (8B to 70B):** Balanced performance for a broad range of tasks.
- **Large (100B and above):** High capability for complex and demanding tasks.

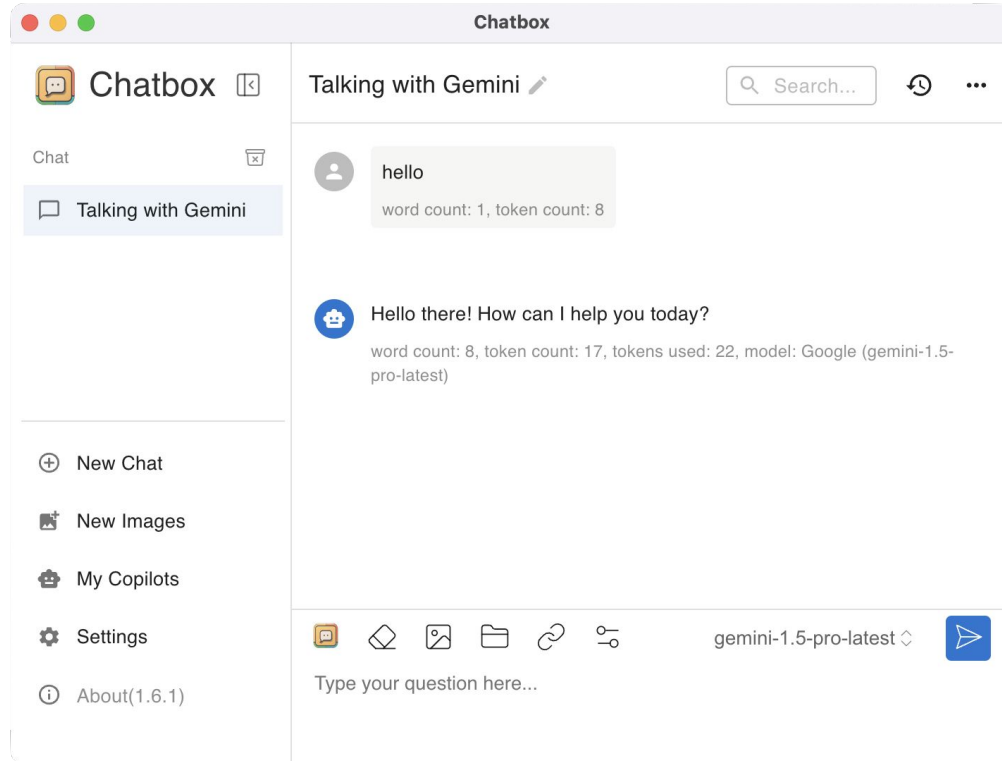
Y-Axis: Capability/Complexity

- **Basic:** Simple summarization, basic chatbots.
- **Intermediate:** Instruction following, multilingual retrieval.
- **Advanced:** Complex reasoning, multi-turn dialogue, specialized tasks.

Running AI: Online & Offline













- Online (e.g. GPT-4o):
 - Internet-based, remote servers
 - Easy access, managed by provider
 - Requires stable internet connection
- Offline (e.g. LLaMA):
 - Run locally on your devices
 - Data stays local, works offline
 - Requires more local computing power

Use Free GUI Tools + Free Online AI Services

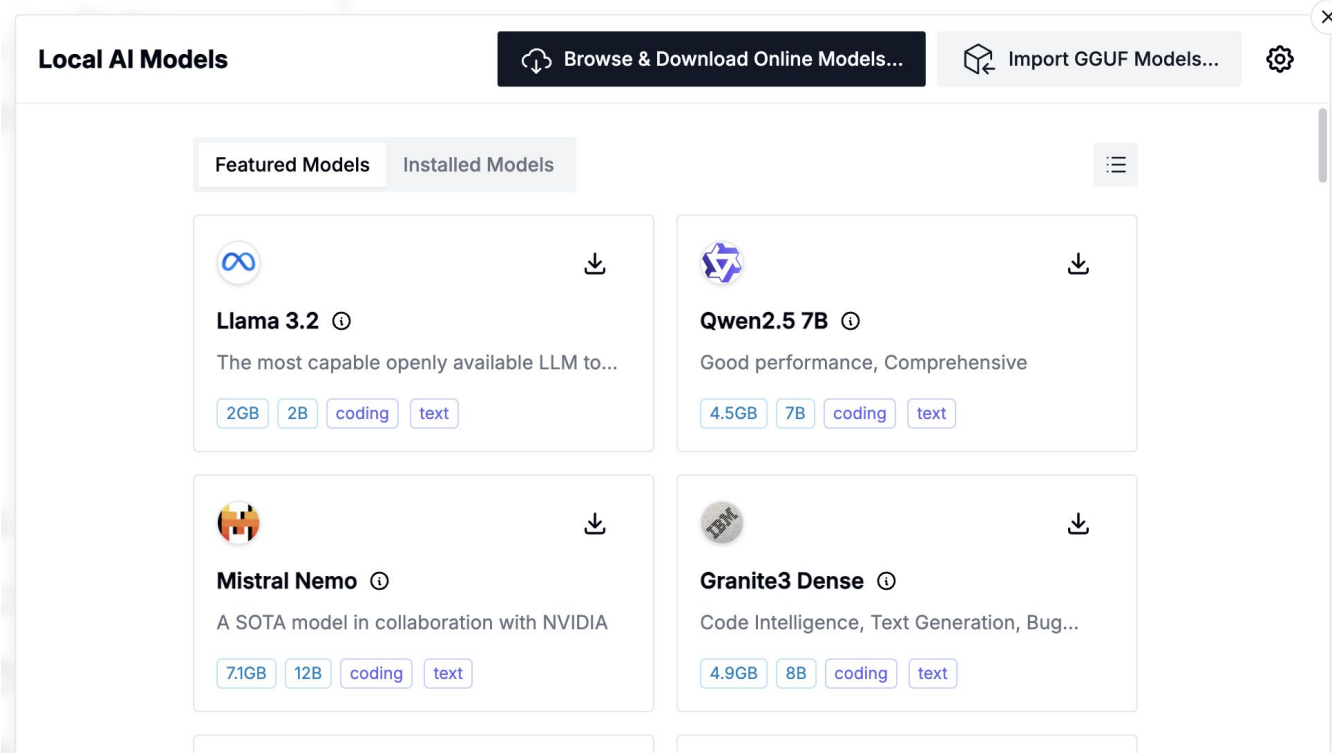


New Remote Models Provider

Models Provider

-  Open AI
- ✓  Open AI
-  Azure OpenAI
-  Claude
-  Perplexity
-  Google Gemini
-  Mistral AI
-  Groq AI
-  Msty Remote
-  Ollama Remote
-  Open Router
-  Open AI Compatible

Running Models Locally



Running AI: Online & Offline

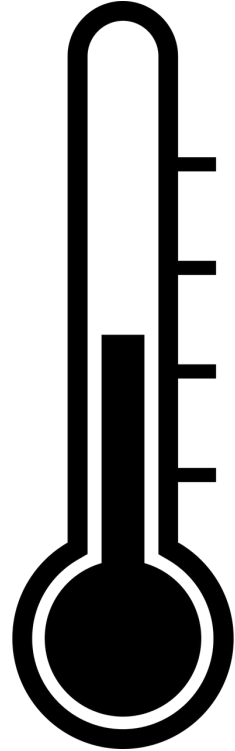
- Online (e.g. GPT-4o):
 - Internet-based, remote servers
 - Easy access, managed by provider
 - Requires stable internet connection
- Offline (e.g. LLaMA):
 - Run locally on your devices
 - Data stays local, works offline
 - Requires more local computing power

Running AI Offline: Requirements

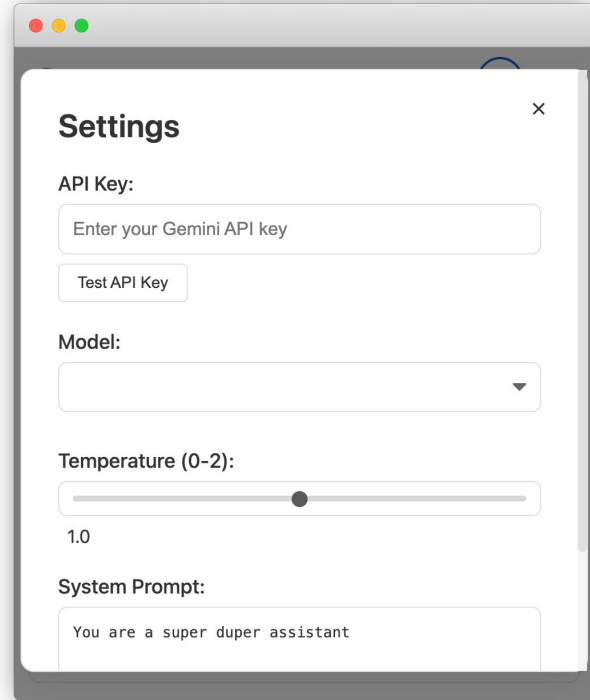
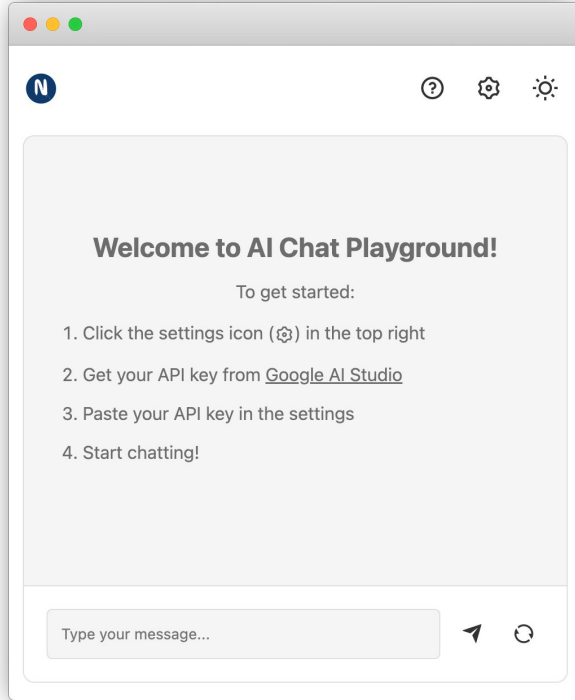
- Hardware requirements for offline LLMs:
 - Small models (e.g. LLaMA 3.2 3B):
 - Modern laptops (8GB+ RAM)
 - Suitable for basic tasks (summarization, simple Q&A)
 - Medium models (e.g. LLaMA 3.1 70B):
 - High-end workstations or servers (64GB+ RAM, GPU acceleration)
 - Handle more complex tasks (advanced Q&A, document analysis)
 - Large models (e.g. LLaMA 3.1 405B):
 - Specialized AI hardware (e.g. TPUs, clusters with 1TB+ RAM)
 - Tackle the most demanding AI applications (complex reasoning, creative generation)

Guiding Your AI with Prompts & Temperature

- System prompts guide the LLM's behavior
- Define purpose, personality, knowledge, output format
- Example: "You are a helpful AI assistant. Provide concise answers."
- Well-crafted prompts key to good results
- Temperature: Randomness of LLM's outputs
 - Low temp: Focused, deterministic, less creative
 - High temp: More random, creative, less focused
- Adjust temperature based on task (e.g. low for factual Q&A, high for story generation)



naida AI Chat Playground



- Fully runs in the browser
- Use free Gemini AI from Google
- Get an API Key (access token) from Google
- Set system prompt, temperature
- Start chatting!