



Innovation at Machine Speed

AI in Development & Product Design

Design, Develop and Innovate with AI

Presented by
Monirul Haque and Sheikh Araf Noshin

“AI is not just a technology trend—it’s a rewriting of how we think about knowledge, work and invention itself.”

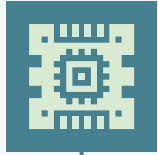
— Sam Altman, 2024
CEO, OPENAI



The Beginning of AI uprising

- **Data Explosion:** Trillions of data points are being generated every second from social media, sensors, transactions, and interactions.
- **Hardware Advancement:** GPUs, TPUs, and now custom AI chips enabling models with billions of parameters.
- **Scientific Breakthrough in Architectures:** The paper “Attention is All You Need” released by Google, started this revolution. The transformer models unlocked LLMs capable of human-like reasoning and communication.

The Beginning of AI uprising



Jun 2017

“Attention is All You Need” Paper released



Nov 2022

OpenAI released Chatgpt as conversational AI



Jan 2023

Chatgpt crossed 100 million users (still the fastest growing app ever)



Present

We have tons of LLM and AI agents for Software, design, healthcare, logistics...

Large Language Model

VS

AI Workflow

VS

AI Agent

VS

Agentic/Multi-Agent AI



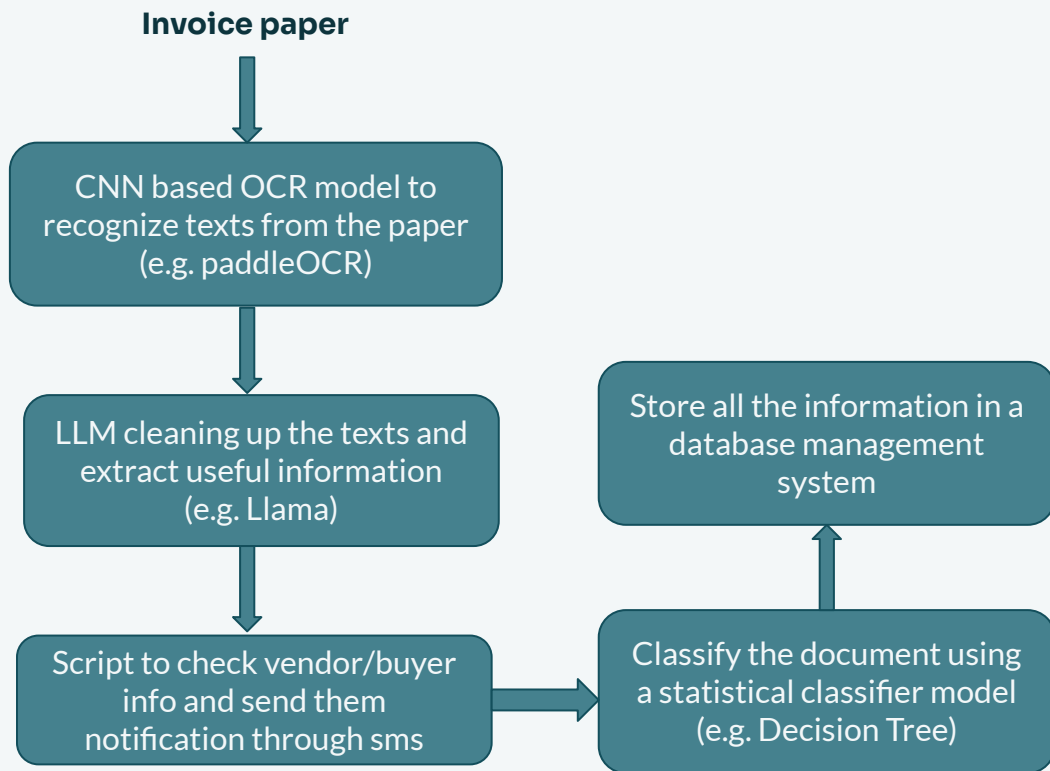
Large Language Model

- **What it is:** A type of AI model (generally derived from transformer arch) trained on huge amounts of text data that can understand and generate human-like language.
- **Core ability:** Predicts the “next word” in a sequence, but can be used for reasoning, summarizing, translating, coding, etc.
- **Example tools to use it:** Chatgpt (GPT 4), Claude, Llama, Gemini

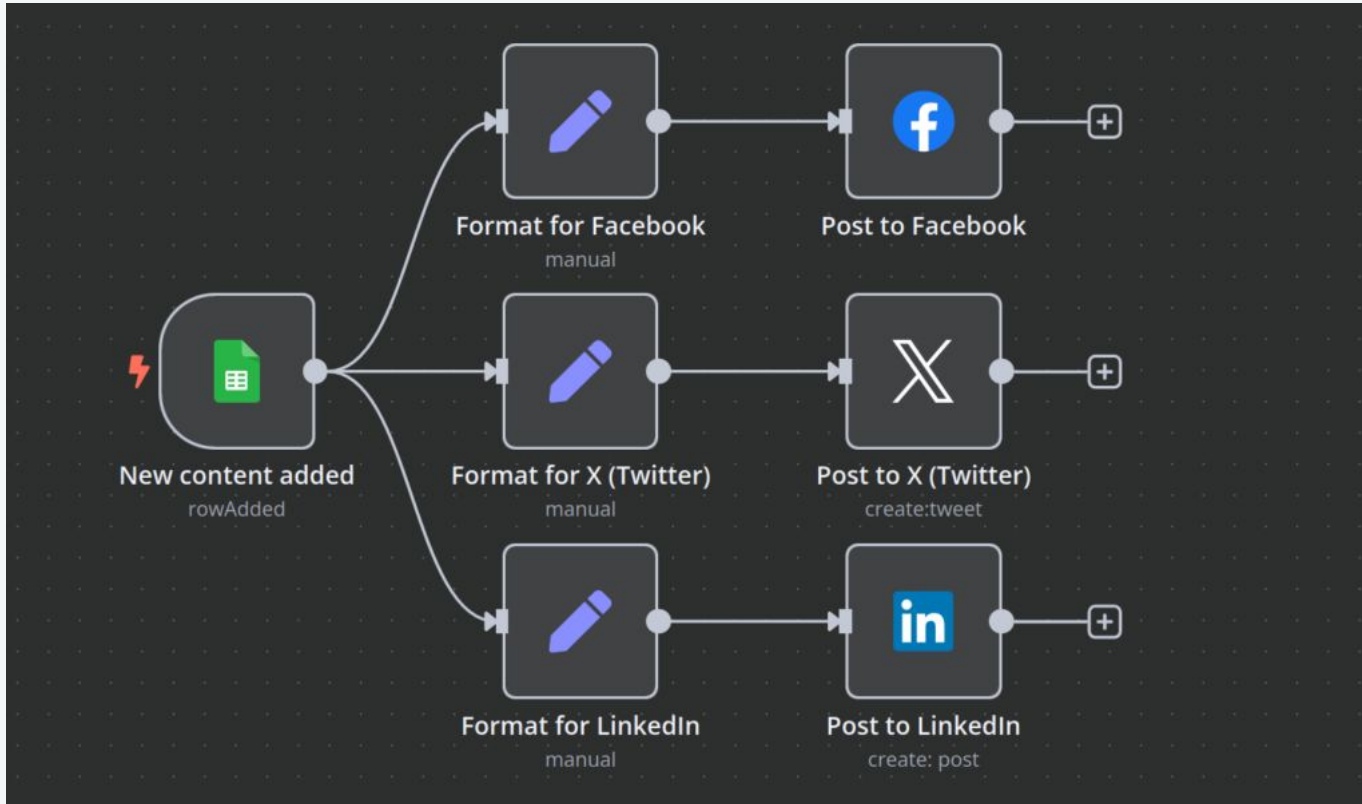
AI Workflow

- **What it is:** A structured sequence of AI models + traditional steps to solve a task. The LLM is one tool inside the workflow, but not the whole thing. We can combine multiple LLM models with different neural network architectures, statistical models and normal logic/python scripts.
- **Core ability:** Combines multiple tools (databases, APIs, LLMs, rules) into a pipeline.
- **Example tools to use it:** n8n, Zapier, ComfyUI, Microsoft power automate

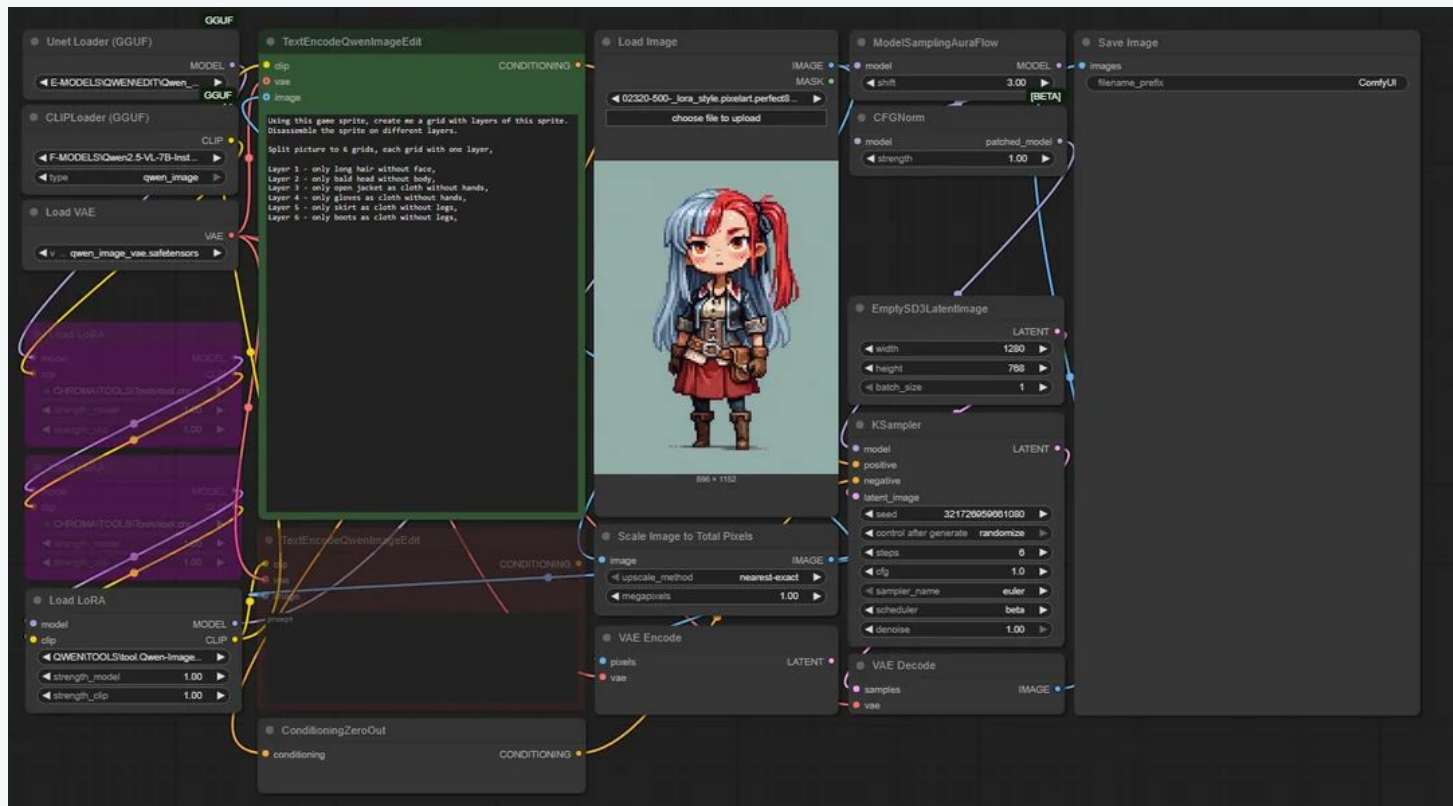
AI workflow example: Commercial Invoice processing



AI workflow example: Content posting (n8n example)



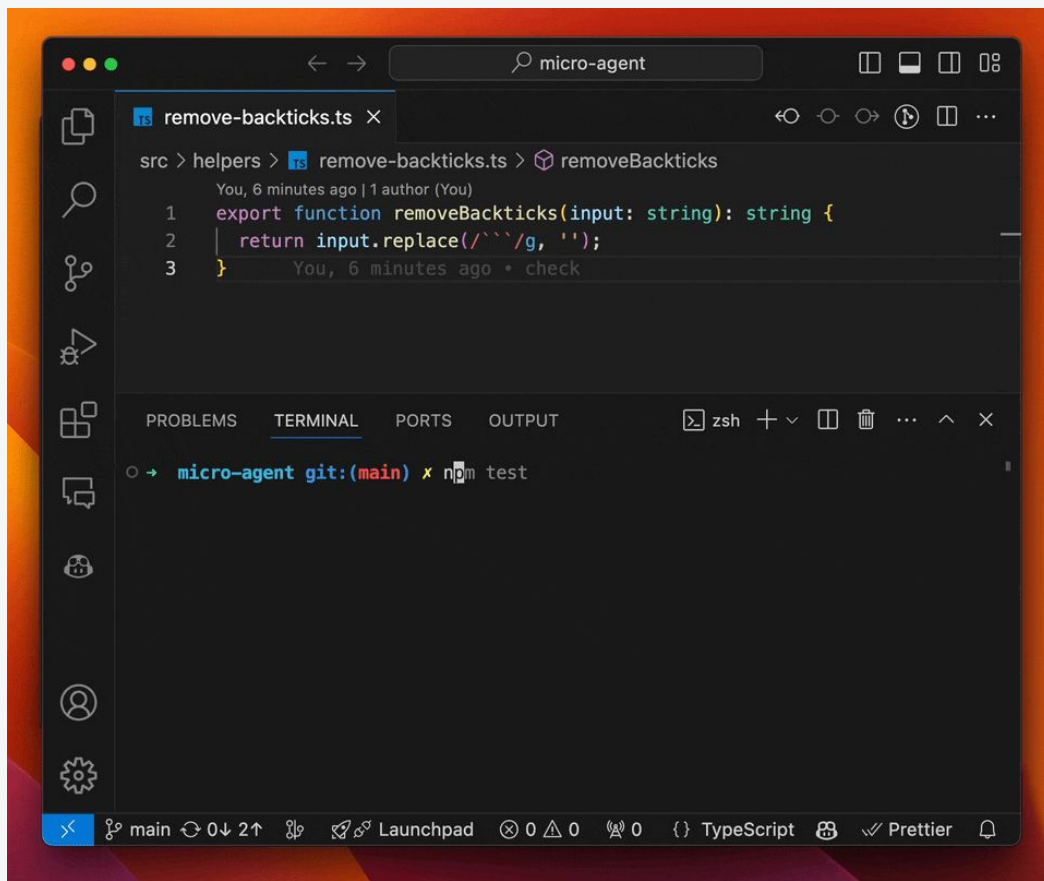
AI workflow example: Pixel animation (ComfyUI example)



AI Agent

- **What it is:** An AI helper system (generally powered by an LLM) that can take actions, not just answer. It plans, calls external tools/APIs, and adapts.
- **Core ability:** Can decide steps, execute, evaluate itself and repeat to improve results
- **Example tools to use it:** Cursor IDE, Gemini CLI, ChatGPT plus (agent mode)

AI agent example: Coding & testing (Micro Agent)



The screenshot shows a code editor window titled "micro-agent". The editor displays a file named "remove-backticks.ts" with the following TypeScript code:

```
src > helpers > remove-backticks.ts > removeBackticks  
You, 6 minutes ago | 1 author (You)  
1 export function removeBackticks(input: string): string {  
2   | return input.replace(/``/g, '');  
3 } You, 6 minutes ago • check
```

Below the code editor, there is a terminal window with the following content:

```
PROBLEMS TERMINAL PORTS OUTPUT  
zsh + - + ... ^ x  
micro-agent git:(main) x npm test
```

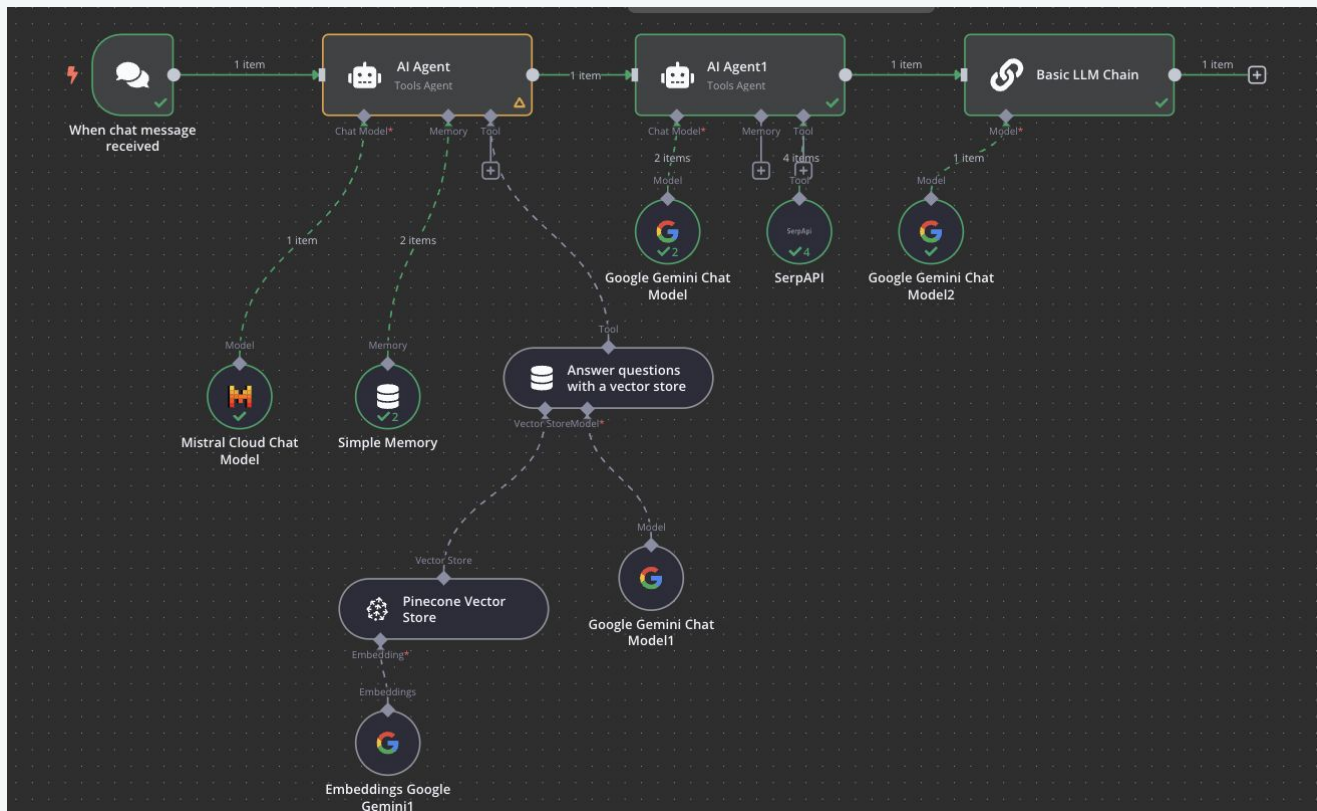
The status bar at the bottom of the editor shows the following information:

- main
- 0 ↓ 2 ↑
- Launchpad
- 0 △ 0
- TypeScript
- Prettier

Agentic AI Or Multi-Agent AI

- **What it is:** Multiple AI agents working in a team. It's like a team of specialists working together to solve a problem.
- **Example tools to use it:** n8n, ComfyUI-CrewAI
- In the previous example of AI Agent we saw an AI agent capable of writing code, executing the code from terminal, testing the outputs, then modifying the code again to fix the bugs. It's just an AI agent specialized in programming. Imagine AI agents for Designing, Programming, QA testing and Documentation working together to create a whole software; that's agentic ai or multi-agent ai.

AI agent example: An n8n example of how perplexity search works underneath



Trending AI tools

- **AI assistants:** ChatGPT, Grok, Claude, Gemini, Microsoft Copilot, Meta AI, Pi (Inflection), Mistral Le Chat, Amazon Q, Perplexity AI, Character.AI
- **Video generation:** Synthesia, Google Veo, OpusClip, Runway Gen-3, Pika, Luma Dream Machine, HeyGen, Descript AI Video, Stable Video Diffusion, Vizard
- **Image generation:** GPT-4o, Midjourney, DALL·E 3, Stable Diffusion XL, Adobe Firefly, Ideogram, Leonardo.ai, Playground v2, Flux (Black Forest Labs), Recraft
- **Meeting assistants:** Fathom, Nyota, OtterPilot (Otter.ai), Fireflies.ai, Supernormal, Sembly, Avoma, Read.ai, tl;dv, Airgram
- **Automation:** n8n, Manus, Zapier, Make (Integromat), IFTTT, Pipedream, Bardeen, UiPath, Automation Anywhere, Microsoft Power Automate, Tines
- **Research:** Deep Research, NotebookLM, Consensus, scite Assistant, ResearchRabbit, Connected Papers, Litmaps, Semantic Scholar

Trending AI tools

- **Writing:** Rytr, Sudowrite, Jasper, Writesonic, Grammarly, Wordtune, Copy.ai, QuillBot, Anyword, Lex
- **Search engines:** Google AI Mode, Perplexity, ChatGPT Search, Bing Copilot, You.com, Phind, Brave Search, Kagi
- **Graphic design:** Canva Magic Studio, Looka, Adobe Express, Microsoft Designer, Kittl, VistaCreate, Fotor, Snappa, Tailor Brands, Brandmark
- **App builders & coding tools:** Lovable, Cursor, GitHub Copilot, Codeium, Tabnine, Replit Agents, Bolt.new (StackBlitz), Vercel v0, Debuild, Windmill
- **Knowledge management:** Notion Q&A, Guru, Coda AI, Confluence AI, ClickUp Brain, Slab AI, Mem, Glean, Slite AI, Tettra
- **Email:** Hubspot Email Writer, Fyxxer, Shortwave, Superhuman AI, Flowrite, Lavender, Ellie, Front AI, Gmail “Help me write” (Gemini)
- **Scheduling:** Reclaim, Clockwise, Motion, Vimcal, Calendly, Mayday, Trevor AI, CalendarHero

Trending AI tools

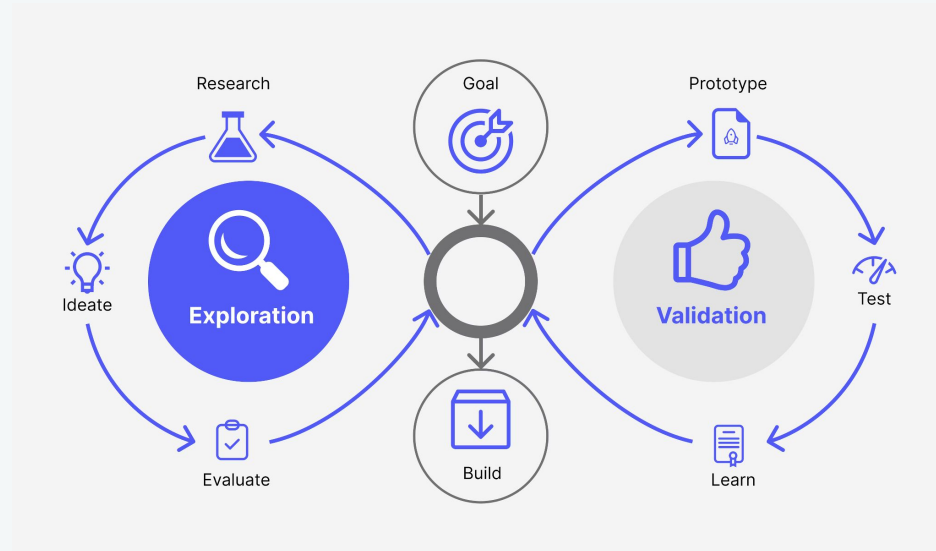
- **Presentations:** Gamma, Copilot for PowerPoint, Tome, Beautiful.ai, Pitch, SlidesAI.io, Plus AI (for Google Slides), Prezi AI, Canva Magic Design
- **Resume builders:** Teal, Kickresume, Rezi, Resume.io, Enhancv, Standard Resume, FlowCV, Wonsulting ResumAI, Zety, Novorésumé
- **Voice generation:** ElevenLabs, Murf, Play.ht, Resemble AI, NaturalReader, Amazon Polly, Azure Neural TTS, Google Cloud TTS, Coqui TTS
- **Music generation:** Suno, Udio, Stable Audio, AIVA, Soundraw, Beatoven.ai, Mubert, Boomy, Meta MusicGen
- **Marketing:** AdCreative, AirOps, Jasper (Campaigns), Copy.ai (Ads), Anyword, Predis.ai, Ocoya, Surfer AI, HubSpot Content Assistant, HoppyCopy
- **Game Sprite/Animation Generation:** Meshy, Retrodiffusion, Scenario.gg, PixelVibe, Kaedim 3D, Inworld Sprite Creator, GANimator, PIFuHD, Anything V5 (anime/manga sprite-focused models),

AI company types

- **Foundational Companies:** (e.g. OpenAI, Anthropic, Mistral, Google)
 - They research and develop their own models
 - Only large companies with huge funding can do that
- **Wrapper Companies:** (e.g. Jasper, CopyAI, Descript etc)
 - Leverage the already existing apis from foundational companies and use them to build user-facing products
 - It's easy and quick to startup a wrapper company but easily replicated by competitors.
- **Hybrid Companies:** (e.g. Perplexity, Microsoft)
 - They start by using others apis but eventually research and develop their own models and keep both their own models and external models on their UI.

Product Design

- It's all about user convenience
- Observe your environment and ask yourself *“What painful problem do people have, that AI could solve 10× better/cheaper/faster?”*



What Modern AI products have underneath

- **RAG (Retrieval-Augmented Generation):**
 - LLMs often “hallucinate” because they rely on their training data only.
 - Instead of asking the LLM to “remember everything,” you give it access to a retrieval system (like a database or search engine).
 - Adding RAG to product means the AI product will have up-to-date knowledge.
- **MCP (Model Context Protocol):**
 - A universal translator between an AI model and external tools (APIs, databases, agents).
 - If your agent needs to use a calendar, a database or an api; MCP ensures the model can understand the capabilities of each and use them consistently.
 - MCP makes agents more modular and extensible.

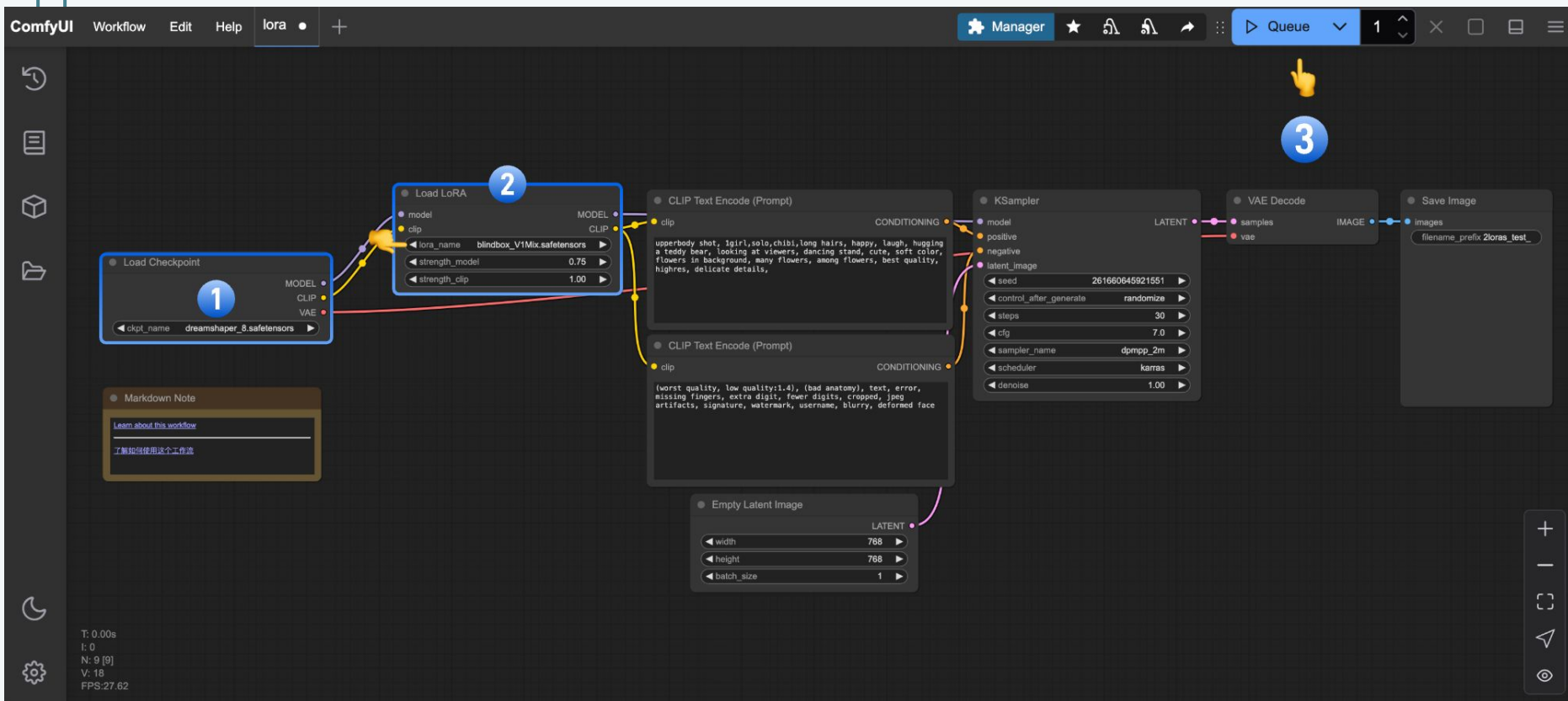
What Modern AI products have underneath

- **LLM Quantization:**
 - Compress the model weights
 - Make LLM's run on low resource devices.
 - There are some algorithms to reduce the number of bits each model weight stores in memory. That's how we get fp-16 or fp-8 quantized models, where each model weight takes up 8 bit space in GPU memory.
 - We even have a 1-bit LLM now (Microsoft's Bitnet 1.58b) which even runs on 2GB vram.
- **Guardrails**
 - Raw LLMS are unpredictable, so a secondary model needs to be added before LLM to filter prompts before they hit the LLM and outputs after they're generated.
 - Guardrails can be used as prompt enhancer for LLM's too.

What Modern AI products have underneath

- **LoRA (Low Rank Adaption):**
 - A way to fine-tune giant LLMs efficiently.
 - Instead of retraining the whole model, you only add/update small number of layers or some “adapter” layers.
 - Makes it possible to customize LLMs for new tasks cheaply (e.g., tuning a general model into a medical chatbot, tuning a general image model into a 2D pixel art game sprite generator).
 - LoRA modules are also modular and shareable — you can load LoRAs trained by others onto a base model and use them in tools like ComfyUI very quickly.

ComfyUI LORA example



What AI Engineers/Data Scientists do?

- Understand the problem and determine whether solving this problem requires an LLM or Transformer/NeuralNet architecture or Statistical classifier or just normal programming logic.
- Data collection, cleaning, preprocessing and annotation (if needed)
- Create/fine-tune model or add an existing pretrained model to an AI workflow.
- Create API out of that model (generally flask or Django)
- Deploy the model or API (Startups require the AI engineer to do MLOps tasks)
- Improve preprocessing/ finetune the model with new data/change the model after getting feedback.

How to find remote jobs

- Build the right skills and showcase your projects on your CV and portfolio site
 - Learn the basics of Scikit-learn, Pytorch, pandas, numpy etc.
 - Also learn git, docker, ComfyUI, Selenium/Playwright, Linux terminal commands, rest api using flask
- Where to find:
 - wellfound.com (previously known as AngelList)
 - Weworkremotely
 - Flexjobs
 - LinkedIn posts
 - remoteok.com
 - Toptal
 - Indeed
 - Airwrk

How to get product ideas for your startup?

- Visit github trending repositories page regularly:
<https://github.com/trending>
- Keep track of the trending tech stacks or latest AI models that are being released regularly:
 - <https://www.youtube.com/@theAIsearch>
 - <https://www.youtube.com/@Fireship>
 - <https://www.youtube.com/@TwoMinutePapers>

Limitations: AI is not perfect yet

- Replit AI disaster: AI agent ignored explicit instructions to freeze all changes and deleted a live production database on July 2025.
- Klarna, a fintech company that fired around 700 employees to replace them with AI-driven customer support. The company stock has plummeted to 6.7 billion from 45.6 billion dollars
- Wimbledon's AI-powered line judge was accidentally turned off during a match, forcing players to replay points, illustrating operational risks with AI systems in live events.
- Apple released Iphone 15 overhyping Apple AI but failed miserably. The beta version underperformed and was far behind from the current competition.
- Apple's paper "The illusion of thinking"

<https://ml-site.cdn-apple.com/papers/the-illusion-of-thinking.pdf> "The Illusion of Thinking"

Future Research

- Continual Learning
- Hierarchical Reasoning Model
- Model Distillation
- Quantum Machine learning
- Parameter/Energy-Efficient AI Architectures
- Sustainable AI
- Explainable LLM
- Reinforcement Learning Agent
- Efficient Model Pruning
- Swarm Intelligence and Collective AI
-

The background is a light blue-grey color. It features several thin, dark teal lines that run horizontally and vertically, some ending in small circles. These lines are arranged in a way that suggests a circuit or a network. A large, dark teal rectangular box is positioned in the upper-middle part of the image. Inside this box, the words "THANK YOU" are written in a bold, light green, sans-serif font. Below this box, there is another dark teal rectangular box, which is narrower and contains the text "Do you have any questions?" in a white, sans-serif font. In the bottom right corner, there are three small, solid teal circles of varying sizes, arranged in a cluster.

THANK YOU

Do you have any questions?