

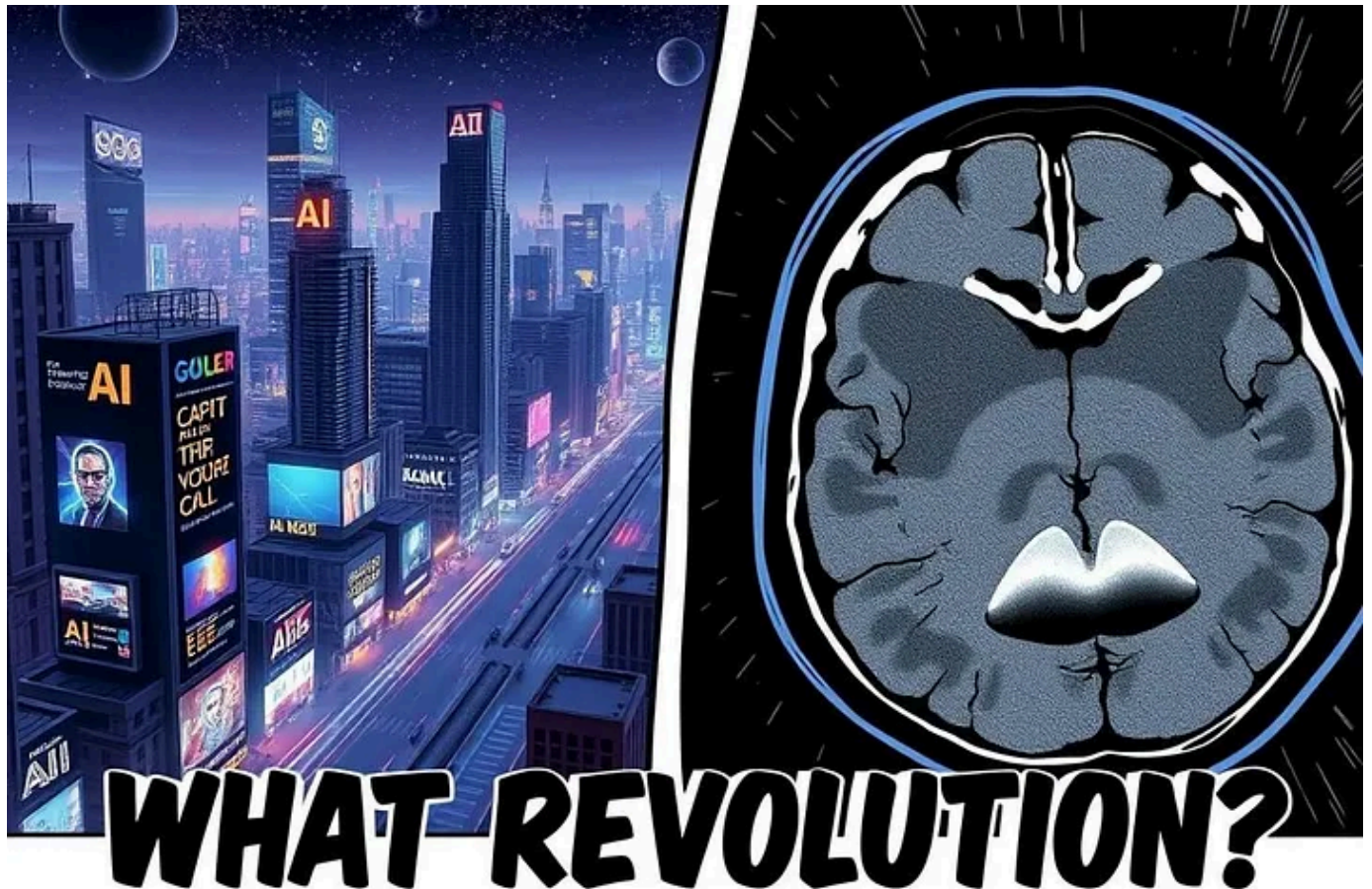
Another week, another AI “Revolution”... Seriously?

Cutting through the noise a simple question in my mind: what matters more, fancy multimodal models or spotting Cancer?

8 min read · Apr 19, 2025



Fabio Matricardi



Are you fed up with this too?

I cannot stand all of this bull**** of AI marvels anymore.

What exactly is over?

I know, just bear with me. There are going to be a lot of images in this article.

But I consider that, more than ever, **seeing is believing...**

This week all over the socials you may have find posts like this one:



or this one



What is over?

Let's be brutally honest on the so-called advancement toward AGI and amazing Generative AI models:

they are building blocks to feed themselves — no more no less

At least for now.

They can do amazing things, something unbelievable just a few months ago.

Because think about it: apart from extremely limited use-cases like vision-impaired or audio-impaired support, what is the real benefit of having a multi-modal language model?

1. to automatically annotate images for the pre-training phase
2. to automatically annotate the audio for the pre-training phase
3. to be able to scavenge new text data from OCR processed books
4. to transform videos in new data

Everything else is online marketing, *solopreneur* business scamming, theory of *look how you can launch a product in only 2 hours...*

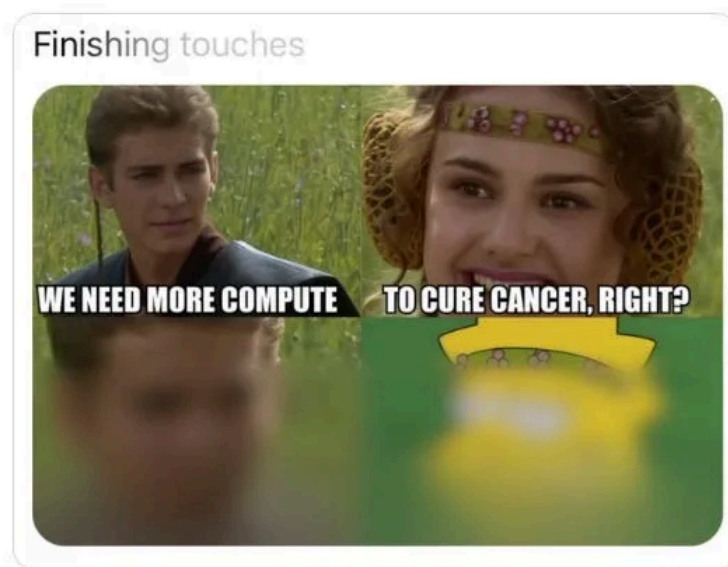
. . .

Why do we need AGI?

The new models are greed of data, for the future to come. We can show everyone a cool generated picture, or the new Qwen able to describe an image.

We can listen to Orpheus-3b generated speech with heart-breaking emotions in the synthetic voice... and then?

Why do we need an AGI? For this?



Or to give these kinds of scary-misinformed clickbait:

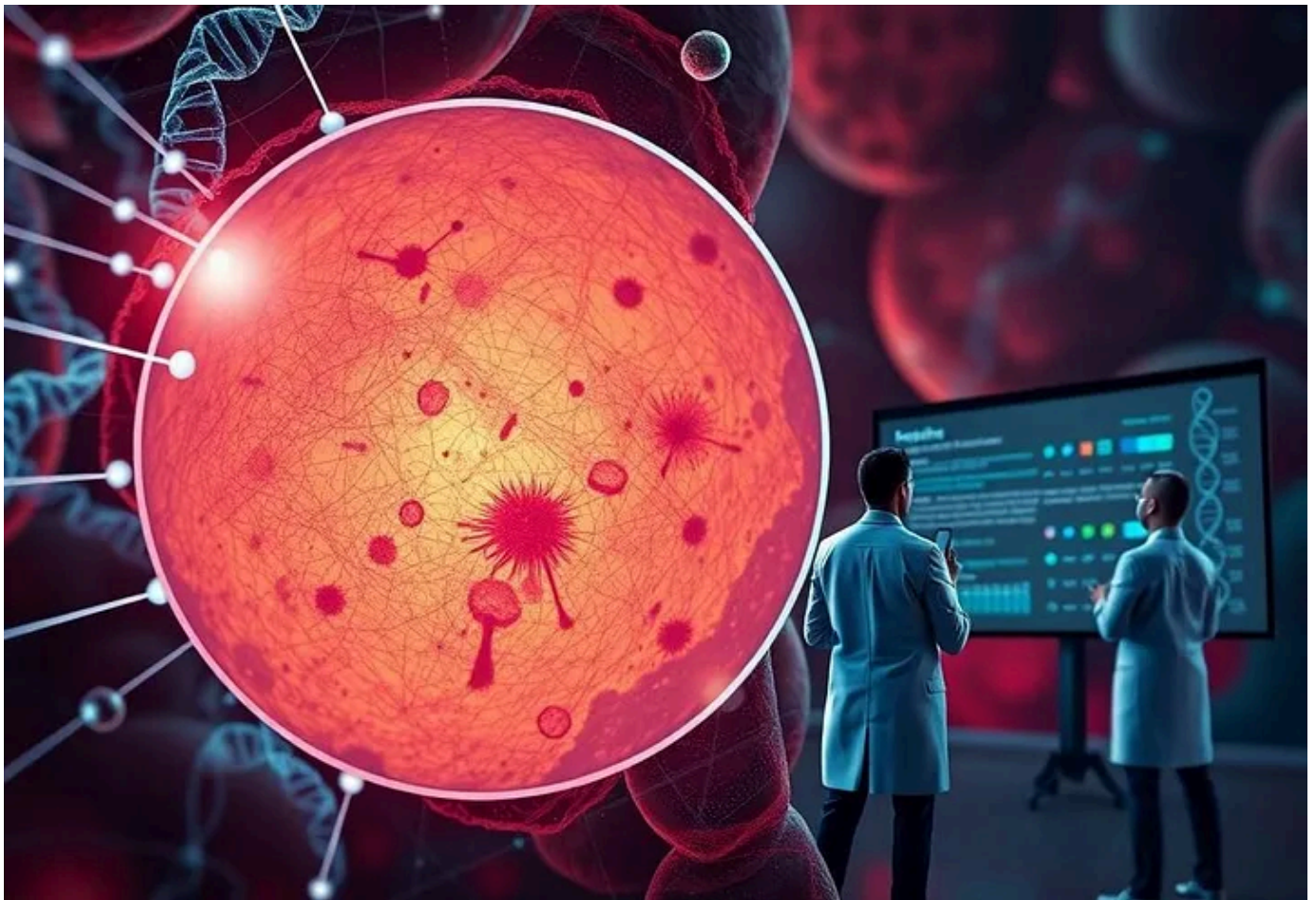


There is a lot of good news, though.

In this article I will show you a few of the latest State of the Art multi-modal powerhouses: this is important, at least to know where the hype is going, and where research is going.

Sometime the two directions do not match.

The good news is exciting (more than the new models).



Cancer diagnosis and drugs discovery with AI

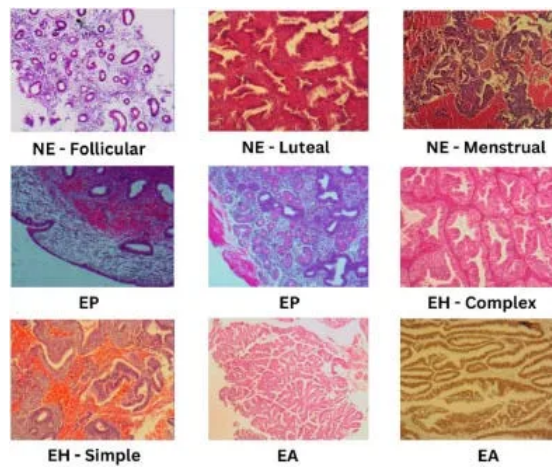
ECgMLP is a new gated MLP model for enhanced endometrial cancer diagnosis. And it works!

Endometrial cancer, affecting the lining of the uterus, is a big concern for women all over the world. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. Traditionally, diagnosing this cancer involves examining tissue samples under a microscope, a process that can be time-consuming and prone to errors. However, a recent breakthrough demonstrates how artificial intelligence (AI) changes the game in this process.

Researchers have developed a new AI model, called ECgMLP, specifically designed to analyze histopathological images — microscopic pictures of tissue samples — to detect endometrial cancer. And this means improved accuracy and speed of diagnosis, leading to earlier detection and better patient care.

The process begins with enhancing the quality of the tissue images. Sophisticated image processing techniques, including noise reduction and clarity enhancement, are employed to ensure the AI receives the clearest possible data. Then, AI powered segmentation identifies the most important areas of the image, the regions of interest, for further analysis.

The results are remarkable. In tests, the ECgMLP model achieved an accuracy of 99.26% in identifying different categories of endometrial tissue, surpassing previous methods. This level of precision could significantly reduce the need for subjective interpretations by human pathologists, streamlining the diagnostic process.



This AI-driven approach improves image quality using advanced processing techniques and detects cancer with 99.26% accuracy. The model processes images of tissue samples using advanced techniques. It cleans the images, removes noise, and highlights important details. ECgMLP uses special methods like Otsu thresholding and the watershed approach. These help in separating cancerous and healthy tissue clearly.

I believe that finally we have some good AI deeds!

This technology, in my opinion, is a bright example and a promise for future applications in diagnosing other forms of cancer.

Latest AI release

We go back now to the kind of AI Intelligence we have, so far. Top AI labs released their best models (so far) or upgraded their champions.

To summarize in one line, these past AI weeks: Qwen new models ,Google Gemini 2.5 Pro, DeepSeek's V3 AI model and finally OpenAI's Image Generator.

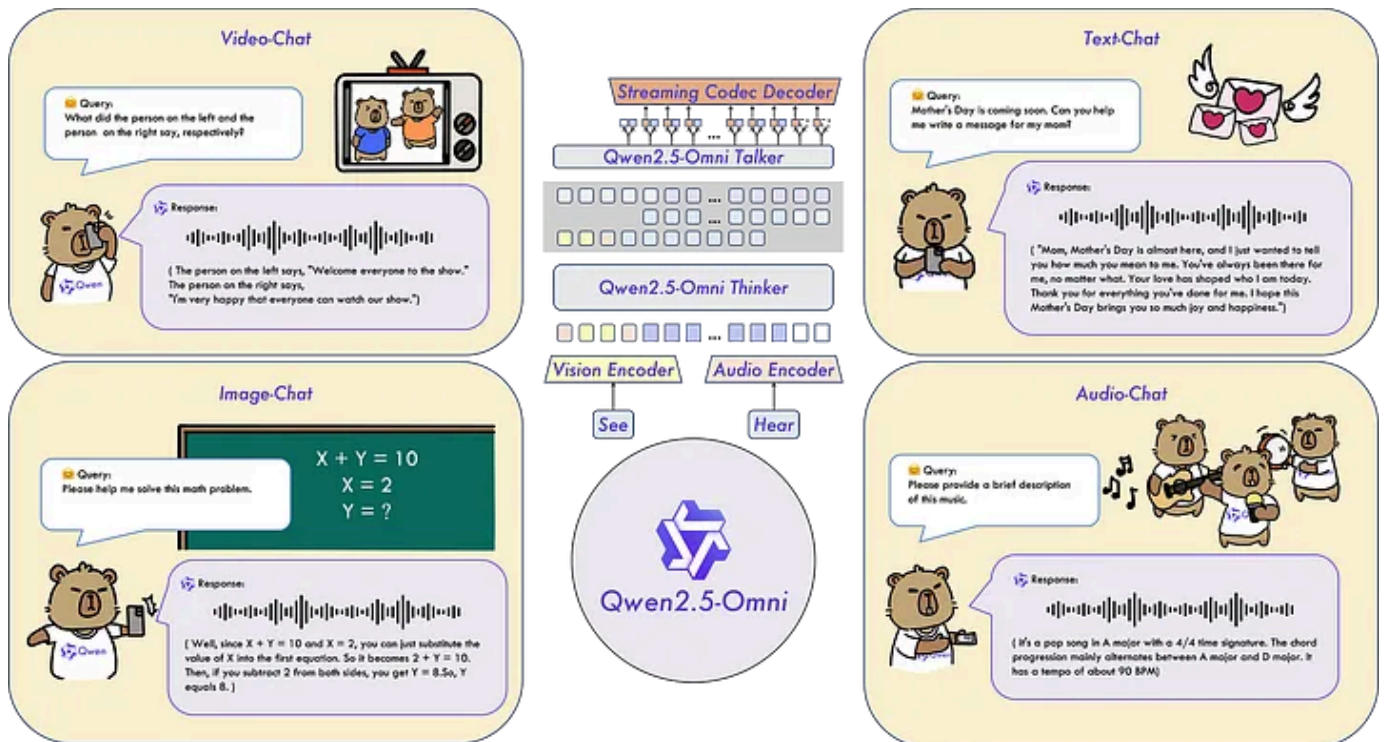


Gemini 2.5 is Google's most intelligent reasoning model to date

On March the 25th, Google announced Gemini 2.5, its latest and most intelligent AI model to date. The first release of the 2.5 generation will be Gemini 2.5 Pro, which is a thinking model (capable of reasoning through thoughts before

responding) just like Gemini 2.0 Flash Thinking. Pro will come with a 1 million token context window, and Google says this will expand to 2 million soon.

According to the tech giant, Gemini 2.5 outperforms some of the most popular AI models across a range of benchmarks, including reasoning, science, and mathematics. The company achieved this feat by “combining a significantly enhanced base model with improved post-training.”



Qwen2.5 Omni is the Chinese All In One

Qwen has dropped three new models this week:

- QVQ-Max
- Qwen2.5-Omni-7B
- Qwen2.5-VL-32B-Instruct

On March the 27th Qwen released **Qwen2.5-Omni**, the new flagship end-to-end multimodal model in the series. Designed for comprehensive multimodal perception, it seamlessly processes diverse inputs including text, images, audio, and video, while delivering real-time streaming responses through both text generation and natural speech synthesis.

You can try the latest model directly in [Qwen Chat](#) and choose Qwen2.5-Omni-7B. The model is now also openly available on [Hugging Face](#), [ModelScope](#), [DashScope](#), and [GitHub](#).

If you are a deep investigator, check out the technical documentation available in the [Paper](#).

There is also a captivating interactive showcase in the Hugging Face [Demo](#) app, for free.

First of all, this model (and QVQ) is designed with an end-to-end multimodal architecture to perceive diverse modalities, including text, images, audio, and video, while simultaneously generating text and natural speech responses in a streaming manner. **Real-Time Voice and Video Chat** are supported with immediate output.

One peculiarity is the **End-to-End Speech Instruction Following** feature: Qwen2.5-Omni shows performance in end-to-end speech instruction following that rivals its effectiveness with text inputs.

Quite the focus on mobile apps and use-cases.



Where is the marvel?

Then there was DeepSeek's V3 model thrown into the mix, plus whatever tweaks OpenAI did to its image generator around the same time.

Honestly, it feels like trying to drink from a firehose these days.

Blink, and you miss three “groundbreaking” releases. It’s exhausting just trying to keep score, let alone figure out what actually matters.

Does anyone else feel like we’re just getting bombarded? Before you can even properly kick the tires on one model, understand its quirks and real uses, BAM! The next “bigger, better, faster” version lands on your doorstep, demanding attention.

It's less like steady progress and more like whiplash, greedy feeding of the hype...

Which brings me back to my main point.

While these massive models fight for the spotlight, doing their multi-modal song and dance, remember that ECgMLP thing? The cancer diagnosis tool? That quiet advancement feels a whole lot more substantial than another chatbot that can describe your cat video, doesn't it?

. . .

Conclusions

So, where does that leave us?

Drowning in demos, dazzled by specs, but maybe missing the point.

My two cents? Keep half an eye on the big model race if you must, but focus on where AI is actually solving real problems, not just generating hype.

That's probably where the truly interesting story is unfolding, away from all the noise.