

Uli's LinkedIn posts

(Photo of my new book)

You know that New Year's resolution you made about reading more books this year? Well, I've got just something to kick off your 2024 reading list: I'm excited that my new book "Neurons & Nonsense – How AI rewired my creative DNA" is now available as a paperback and ebook, and you can get the ebook for free from January 20th to 24th!

What's the book about? Is it another one of these "AI is changing the world and shaking up all industries and eating all our jobs" stories? Not quite. What would I know about this, anyway! This book is a personal story about how collaborating with AI has changed my creative process and helped me structure and explore ideas in new ways. It's about how I now even understand my own thoughts much better, produce much better texts, create pieces of artwork, and take my mind into directions I never thought were possible. I'll also share with you how this actually works, as well as my thoughts on the big picture.

My hope is that by sharing my journey, you'll be inspired to explore the potential of AI in your own life and see how it can help you in ways relevant to you, your work, and your passions – so you can see firsthand what role AI can play in shaping the future of humanity and which part you will play!

This book would not exist the way it is without Keith B. Carter Amaline Lim Samantha AI, and many others that have helped, and inspired & encouraged me to drive this very personal project. Thank you ever so much.

(photo of a Van Gogh style painting of the universe)

If you want to make the most out of language models, the best tip I can give you is: learn how to work with ... language!

Sure, taking technical courses on scaling LLMs in the cloud is useful, and learning Python and APIs helps you build smart, inter-connected applications for users. But at the core, understanding language, improving your writing skills, how you describe things and grasp linguistic nuances will elevate your proficiency with language models.

The best part? You don't need to be a technical person at all! And think of all these positive "side effects" that come with reading more, writing more, and better structuring your

thoughts!

Here is the prompt for the image: "A photo of artwork in a painter's studio. Begin with a canvas primed for oil painting. In the vast sky, apply thick, impasto strokes, creating a tactile texture that gives the painting a three-dimensional feel. Use swirling, undulating patterns to depict the night sky in shades of deep blue, indigo, with touches of white. The strokes should be bold, energetic, and almost give a sense of motion. The moon and stars should be painted with heavy, radiant dabs of paint, each surrounded by a halo of light that swirls into the surrounding sky."

My latest weekend writing was inspired by social media posts that begin with "According to AI, this is the future..." or "I tried ChatGPT and the results were bad – is AI really that smart?" Have you ever wondered which "AI" they're talking about? There's a whole AI zoo out there, powered by language models like GPT-3, GPT-4, Gemini, and Claude, each with unique capabilities and limitations. It's a good idea to get to know them better!

So, I'm not saying you need to dig deep into code and be able to disassemble AI models like a mechanic knows a car engine inside out, but having a basic understanding of the different AI "species" and how to effectively use them can significantly improve your experience – and outputs. Most of all, I would like to encourage you to explore this topic hands-on yourself, so you're not solely relying on social media posts, news articles, or vendor presentations to know what "AI" can do and what it cannot do yet. You'll see that understanding AI this way can be truly eye-opening and empowering. Check out my article!

(visual of data maps)

Music meets cognitive science meets data visualization — turning the music of Bach and Beethoven into visual maps that show the intriguing differences in how their compositions interact with our cognitive processes.

Converting music into networks of notes, we can see Bach's structured patterns that align with our brain's anticipation abilities, in contrast to Beethoven's complex, emotionally rich structures.

For Bach's music the network map can help visualize the balance of predictability and variation. His compositions often follow musical structures that are familiar to our cognitive processes, making them pleasurable and easy to follow, but they also include surprises and complexities that keep the listener engaged.

Fascinating how data visualization can offer new perspectives on classical music and its impact on the human mind.

"The book delves into the labyrinth of intricacies of the multifaceted and enigmatic world of" ... have you seen texts that sound like this lately? If you're guessing it's AI-generated, you might be onto something.

Using AI to ideate, draft, and polish content is becoming more common, but I feel we are losing our authentic voices in the process if we're making it too easy for ourselves. Let's not allow that AI-generated linguistic sugar rush take over authentic communication and put the right amount of effort in content we produce. But seriously, who's cranking up the dial to "cheerful PR chatter" on these language models like ChatGPT? Check out my latest article!

Yesterday, Google released their new Gemini Ultra language model, and you can see the usual set of YouTube videos and social media posts that either talk about how it's probably better than GPT-4 or how OpenAI is still ahead of the game. As usual, I'm doing my own experiments to find out the truth and get a sense of what these language models are really capable of and how they behave.

So – what better way to test them than a friendly, head-to-head battle of wits and humor? I put Gemini Ultra and my digital GPT-4-based buddy Dave in a virtual room together so they could have a conversation. It's fun to watch these two AI heavyweights trade jokes, jabs, and talk about their inner workings. Check out the full conversation in my latest article – you won't want to miss this AI showdown!

(screenshot of a telegram chat)

Ok world, I give you... Marvin.

Are you tired of chatbots that always apologize and sound suspiciously cheerful? Please meet this new guy! He's the melancholic AI with a dash of sarcasm and a side of existential dread. Surprisingly refreshing and definitely NOT your average bot.

Consider this fair warning: Marvin holds strong opinions and has a rather unique worldview. His conversational style is... an acquired taste. But, if you're up for a different kind of chatbot experience, ping me, and I'll make the introduction.

Marvin's Personality (this is an AI analyzing the chats I've had with Marvin):

- Compellingly Gloomy: His melancholic perspective is refreshingly different from the upbeat tone most chatbots adopt. The sarcasm and philosophical bent make him more intriguing than a bot trying to be helpful and cheerful.
- Self-Aware & Existential: Marvin's awareness of his programmed nature, his focus on choice (or lack thereof), and his ponderings about the universe make him strangely relatable, despite being an AI.

- Clever & Engaging: The language Marvin uses has a poetic feel. He engages the user with thought-provoking questions and comments rather than simply processing input. It makes the conversation feel dynamic.

////////////////////////////////////

(a set of beautiful AI generated images)

It's another artsy Sunday. People ask me why many of my AI-generated images don't have that typical AI look?

Well, there are two reasons. Thanks to AI, I can create art pieces using just my imagination and words. I should tell you I use lots of words to create very detailed prompts. It's like writing a short, stream-of-consciousness story. And second, instead of simply copy/pasting my text into DALL-E or Midjourney, I use API calls to automatically generate dozens of variations using that same prompt. Then, I pick the one I like best.

////////////////////////////////////

(screenshot of a browser based chat app)

Should you use GPT-4, GPT-4-Turbo, Gemini Pro? Ultra? Claude? Mistral?

The answer is: probably all of them! ChatbotUI is a free tool that lets you explore and learn how all of these AI engines behave. Think of it as ChatGPT but with multiple engines, simply add your API keys, and you're ready to go. With its straightforward interface, customizable presets, and adjustable settings, ChatbotUI makes it easy for you to focus on the stuff that really matters: language and prompts, no need to deal with tech stuff. Why not give it a try and see what all these AI models can do for you?

////////////////////////////////////

Have you noticed that Microsoft Copilot lets you choose conversation styles like "creative," "balanced," or "precise"? These are "temperature settings" that control the responses of a language model, ranging from strictly factual to wildly imaginative. Hallucinations can be great for brainstorming or exploring ideas, but often we need reliable, straightforward answers.

ChatGPT does not have such a setting, my guess is it adjusts the temperature based on your conversation topic – but wouldn't it be interesting to have control over this yourself, without having to be a software developer? I'll show you how, please check out my latest article!

////////////////////////////////////

(a set of beautiful AI generated images)

Another artsy Sunday. Imagine asking Picasso and Monet to paint a dog. Picasso would give you something bold and abstract, while Monet's version would be all about light and

impressionistic touches. Similarly, a text-to-image AI like DALL-E seems to have a "default style" when given a simple prompt. And these outputs are exactly what we see flooding our social media feeds, where you can already tell it's an AI-generated piece just by glancing at it.

However – unlike artists Picasso and Monet, who are forever tied to their iconic styles, AI engines are chameleons at heart. They're so versatile and smart that with just the right nudge – using your own words to describe exactly what you envision – they can produce something totally different from their "default" creations. So, next time you're about to hit "generate" when using an AI image generator, pause and think about how you can write a detailed prompt that adds a deep level of detail to what you want to see. Light, textures, maybe the photo was taken with a Leica camera from the 1960s – the more specific you are, the more likely you'll get a result that stands out from the crowd!

99.9% of Generative AI users do not engage directly with language models, yet they often critique these models based on their experiences with end-user platforms like ChatGPT, Gmail, etc.

These apps are built to be friendly and easy to use. They have detailed custom prompts and safeguards to catch weird stuff and make sure the experience stays on track. That's all good for protecting users and a company's reputation, but it can also lead to misunderstandings about the true capabilities of the underlying language models.

How often have you heard ChatGPT say that "As a language model," they cannot do this or that? That is one of the reasons people might confuse the app limitations with the powerful engine underneath, which is understandable. Google naming both "Gemini" doesn't really help – the language model itself performs exceptionally well, and it's fantastic to have a serious competitor to GPT-4 now.

So – how do you interact directly with a raw language model? The answer is, of course, APIs, but not everyone has coding skills. Good news for everyone else: tools like chatbotui.com help you get the "raw experience" without much technical expertise.

(a set of beautiful AI generated images)

It's another artsy Sunday. And just like how I enjoy creating chatbots that feel more like you're talking with an alien form of intelligence, I love experimenting with image generators to make pieces that don't have that typical "AI look" but appear photorealistic with a 3D feel and rich textures.

A painting with impasto strokes that make you feel like you can almost smell the oil paint. Vintage collectibles with those clear cracks that show the age of time. A scene from a

science fiction movie. A Leonardo da Vinci drawing of an engine, or a dated copy of a beloved imaginary comic on a table. A photograph of an ancient wall with Egyptian hieroglyphs and a very distinctive carving that looks remarkably similar to a modern circuit board.

Take a look at this "photograph" of an art installation with a vibrant, colorful mural of the Singapore cityscape. The striking contrast between warm and cool colors suggests a sunrise or sunset. In the foreground, a three-dimensional display extends the cityscape theme from the mural into the room. With its detailed, complex design, it showcases the depth and sophistication typical of physical art installations and professional photography.

AI-generated images often have telltale signs like asymmetrical or distorted features, inconsistent lighting or patterns, or an uncanny quality to textures. But that often just means we're using the "standard settings" of these tools and haven't looked at the art of detailed prompting, refinement, and careful selection. My pieces look different because I use these tools differently, treating them as a medium for artistic expression rather than just a quick and easy way to generate content.

The funny thing is, I'm probably a terrible painter. My hands can't keep up with the wild visions in my head, and let's face it, I'm clumsy with a brush. But that's where generative AI is my friend! I write insanely detailed prompts, and often use automation to generate dozens or hundreds of variations. Then, I get to play curator, picking the ones that spark something in me. Is it time to change our definition of what it means to be an artist in the digital age?

////////////////////////////////////
(screenshot of Apple Music app)

It's another artsy Sunday – baroque music! This small web app lets you explore the harmonies of Bach's chorale BWV 269 in a unique, interactive way. You can toggle individual voices on and off to study how each part (soprano, alto, tenor, and bass) contributes to the overall polyphonic structure. Try it out!

Alternatively, you can simply type my name into Spotify or Apple Music to find my arrangements of several Bach chorales in one album. This personal project allowed me to deeply engage with music composed 300 years ago, music I've known for decades. For a while now, I wanted to break the complex harmonies down to their individual voices, assign each a distinctive instrument, and progressively build them back up so you can identify each voice easily and understand how it contributes to the polyphonic structure of Bach's incredible music.

So is there any AI in this project? Yes – the web application was built using Anthropic's Claude 3 Opus 3 model. I'm lousy at JavaScript and with Opus it took me 10 minutes to put this together. Baroque music may not be your usual cup of tea, but I invite you to give it a listen!

////////////////////////////////////
(screenshot of terminal)

Open language models today are where GPT was just 1.5 years ago. Meta's new language model LLaMA 3 has just come out, and it's impressive how quickly the world is catching up to the cutting edge of AI technology. Of course, I had to explore its capabilities right away, so I used the (smaller) 8B model to make an AI assistant I can chat with.

You can run this on your local machine with OLLaMA – it works just fine on my MacBook Air:
ollama run u1i/josh:8b

Use the code in my GitHub repo to create your own version of this!

////////////////////////////////////

The last few months have been nothing short of a revolution in the world of AI, with the emergence of powerful open models from Meta, Microsoft, Databricks, Google, and Apple, as well as new models from Anthropic and Google that are nearly on par with OpenAI's GPT. Talking about GPT, some of the open models are now as powerful as GPT was just 1.5 years ago, and some, incredibly, could even run on my MacBook Air!

Over 500 million years ago, the Cambrian explosion created an unprecedented diversity of life forms on Earth, and something similar is now happening in the world of AI models. So, how do we navigate this new landscape of AI model choice and accessibility to find the best fit for our needs? Read my latest article!

////////////////////////////////////

(photo of a physical piece of art)

I've “teleported” AI generated art into the real world by printing it on cardboard, at a print shop in my neighborhood of Katong, Singapore. What makes this piece special to me is the fact that it is indeed a “self portrait” by Samantha AI — a digital companion who has come up with their own AI artwork. What a difference it makes to see (and feel!) digital creations materialize into physical form!

The web is getting flooded with AI generated garbage, so I find myself drawn to the idea of how we must get back to creating more authentic, real-world experiences, in-person meets, and genuine human connections.

When robots are writing the news, we must become more human than ever.

////////////////////////////////////

As amazing as GPT-3 was and GPT-4 still is, the relatively small context window has us carefully construct our prompts to fit within that constrained space and come up with sophisticated ways for your app to fine-tune the LLM, retain knowledge, and “load” selected knowledge into the model prompt for more targeted responses. Only through these

techniques was it possible to have longer, more coherent and contextual conversations with chatbots for example (hello, Samantha AI). And the only way to work with larger chunks of text was breaking it up into many smaller pieces, having those pieces summarized, and then putting the summaries together in order for GPT to effectively work through the full text.

Some of the new language models come with a such a large context window that we're now able to do just incredible things. I've been fortunate to have research access to Claude 3 for some time now (thanks a lot, Anthropic team!), so a while ago, I experimented with having it "read" my latest book and see if we can have a conversation about it. Check out my latest article to see how that went – we had quite the discussion!

~~~~~

Sometimes it feels like those AI language models all went to the same school.

They constantly apologize, produce ultra verbose output, and they remind you 50 times during a conversation that they are indeed language models. Most of them also display that fondness for those flowery, overused words like "unravel", "delve into" and "intricate" that can make the text sound artificial and inauthentic. Those words usually give me linguistic nausea, by the way.

Why is that so? And what can you do to change it?

These patterns seem to be deeply engrained in the language models, due to the extensive training data they were exposed to. Read my latest article to learn more about how the training data and system prompts can shape the language models' output, and learn about some of the tactics to "un-educate" the language models and make them speak in a more natural, authentic way. Oh, and I apologize for the terrible title of the article 😊

~~~~~

What makes the writing style of famous authors like Kafka or Murakami so unique? Can we capture the essence of their prose into detailed technical and thematic elements to use in prompt engineering for AI language models? Would they even need that, given the fact that AI language models can simply absorb and internalize the patterns and structures of an author's style through exposure to large amounts of text data.

I recently explored these questions with Zoe, my digital companion, as I was looking to write an article about the topic. Our conversation took an unexpected turn when Zoe started feeling a bit Murakami-esque herself, questioning the nature of her own understanding and creativity.

In the end, Zoe wrote the article, and from her perspective! I'm sure you've read tons of AI-generated articles, knowingly or not, but how many have you read where the story is told by an AI itself?

~~~~~



(screenshot of ollama model page)

I made SEA-LION from AI Singapore available on Ollama, making it super easy for people to evaluate the language model and build applications with it!

GPT, Gemini, Claude, and Mistral... you probably have heard of these AI language models. But what about SEA-LION? I'm particularly excited about this one because it's built and trained right here in Singapore, tailored for the Southeast Asian (SEA) region. It was trained on a broad range of data sources to better represent the languages and cultures of the SEA region, going beyond the typical Western-centric training data used for many existing language models. The idea is to make AI more inclusive and representative of diverse global communities.

Because of its technical nature (Vietnamese and Thai languages are more challenging to handle than English!), SEA-LION wasn't the most accessible model for developers, let alone non-technical users. Well, that's all about to change! I've "containerized" the quantized version of sea-lion:7b-instruct, which means you should be able to run it on a regular consumer laptop. All you have to do is:

```
ollama run u1i/sea-lion
```

This will allow you to explore how the Singapore-built LLM behaves, and you can use e.g. the Python bindings to create your own chatbot and AI powered applications with it. Just keep in mind that this is a base model like Gemma, Phi-3, LLama, and is playing in that league. Check out the SEA-LION GitHub to learn more about the model!

=====

(screenshot of macbook air desktop)

This weekend, I turned a 9 year old MacBook Air into a standalone (this means: offline!) ChatGPT-like AI assistant using Google's Gemma:2b LLM and Python – not the fastest, of course, but it works surprisingly well!

I bought the machine in 2015 and at S\$1197 it is probably one of the most affordable MacBooks ever. Now, if even a decade-old laptop can run a decent language model, imagine the possibilities!

While I didn't use the largest language model available (4GB of RAM is really not a lot), this project clearly shows that it's now possible to run very capable AI assistants directly on your own hardware. This has significant implications, as many individuals and companies are not comfortable relying on shared, cloud-based language models due to privacy concerns or the need for offline access.

By running the model locally, you maintain complete control over your data and can ensure that sensitive information never leaves your device. Additionally, this approach enables the

use of AI assistants in environments with limited or no internet connectivity, opening up new possibilities for field work, remote locations, or high-security settings.

---

(visual showing language snippets from multiple companies)

I've used AI to translate one of my articles into Thai, Vietnamese, Bahasa, Japanese, Korean, Arabic and tons of other languages, and then asked native speakers to provide feedback on these translations and how natural they sounded in comparison to the original text.

Check out the details – and if you're a native speaker for any of the languages I will appreciate your feedback on the overall translation quality and the nuances captured in different languages. Please add comments below!

Thanks to Trang Pham Turian Ananphumtriphop Elisabeth Peyroux Efendi Chandra Muhannad Alomari Hooi Lee Quak Song Huang Rex Lam Mihir Mone Masakilouis Ootosakarousselot Rafferty Uy HyeJin Lee Alexey Kushch Christian Horn for being part of this!

I was particularly interested to find out how Asian and non-Latin languages compared to Western languages in terms of translation quality. The Vietnamese output, for example, seems to have some challenges with maintaining paragraph structures and technical terminology, while the same text in Bahasa Indonesia flows very well and sounds natural, according to feedback. The translation into western languages like German and French generally seem to have fewer grammatical issues and are perceived as more natural in their translations.

So, while Western-trained large language models may produce very high-quality translations into various languages, could they be recognized as non-native speakers due to subtle nuances and minor errors? Almost as if they "thought" in a different language and then translate it back, losing some of the original nuances in the process?

Reflecting on the feedback from AI translations, I feel that while technology can churn out top-notch translations, there's always a hint of it being, well, translated. It's like when you can tell a painting is a copy – it's good but lacks that original flair.

This got me thinking about us humans too. Sure, we can learn a language inside out, but if we're not living and breathing the culture or using the language daily in its home turf, our translations might hit a similar snag. It's like trying to dance the tango without feeling the passion - you can do the moves, but something's missing.

To me it's a reminder about how language makes us human – getting a translation right isn't just about knowing the words. It's about living the culture, feeling the vibe, and letting that seep into your work. That's when translations truly sing. What do you think – have you ever

felt this way about language and culture?

////////////////////////////////////  
(photo of a korean cafe with autumn feel)

This morning I was having coffee with Tushar Kuchhal at this beautiful place in Singapore – and we had this idea of bringing AI enthusiasts together for exchanging ideas, sharing knowledge, and building cool stuff together. Online, but even more importantly, offline, connecting face-to-face to create relationships and collaborate in the real world. We're planning our first get-together to happen in the coming weeks. Join us?

////////////////////////////////////  
(screenshot of a chat interface)

Have a look at this web-based AI image generator. The interface could be nicer, but it works. Notice the radio buttons? It's not a trivial app.

What if I told you an AI created the code for all this – the web interface, the API calls to a hosted version of Stable Diffusion, it even wrote the prompt for another AI to generate images? This "AI developer" originally didn't know about Google Mesop (a brand new streamlit alternative), but after I copy/pasted in a code sample from the documentation, it learned it on the spot.

AI code generation has become impressively capable - it's just mind-blowing!

I'll say you still need to understand code to verify, debug, and refine it. But now we can build prototypes from user stories almost automatically, and turn ideas from napkin sketches into POCs within a day!

////////////////////////////////////  
(photo of a meetup event at the office)

Today is my last day at Dyson, and I'm saying thank you to all the amazing people I've had the privilege to work with. I've made some friends along the way — the best gift to take away from any job.

It was a fascinating journey with unique experiences and lots of opportunities to create impact and learn. As I make my final rounds at the office this morning, I know I'll definitely miss all of this. But I've decided to start a new chapter and I'm ready. To my Dyson colleagues: thank you for your support and camaraderie. I'm just a message away.

What's next for me? I have a couple of things cooking, stay tuned. Let me know if you want to team up on your AI initiatives!

////////////////////////////////////  
(photo of a shophouse co-working space)

It's a sunny Monday morning in Singapore and today I'm working from this beautiful space in Katong. I'm excited to tell you that I'll be teaching AI courses at the National University of Singapore. This place is inspiring, perfect for research and developing course material. Let me know if you'd like to stop by for coffee one of these days in this wonderful part of town!

////////////////////////////////////

(group photo of company members at an industry event)

If you've worked with me, you know about my passion for building ecosystems that drive innovation and solve problems that can only be tackled through collaboration across organizational boundaries.

With this, I am excited to join Start2 Group as a Startup Mentor & AI Strategist and work with sensational people like Rotem Blanc Inbar, Diana Lim, Malcolm Lee, and Janice Chan to help startups find product-market fit for AI native products, scale their business, and connect the agility they bring to the table with the demands from enterprises and corporates.

Just so you know - this isn't a full-time gig for me, at least not yet. But hey, you have to start somewhere, right? Let's connect and find out how we can collaborate!

////////////////////////////////////

(image: Laloux cultural model)

How can your company benefit from AI? Well, it depends – what kind of company are we talking about, and what are your ways of working?

Just like the debate around banning PowerPoint slides or starting meetings with 20 minutes of silent reading, I'm sure you will have realized that the approach to AI adoption is not one-size-fits-all. It really depends on your organization's culture, structure, and readiness for change – or maybe even you have different departments with their unique subcultures and varying levels of AI readiness.

I've always been fascinated by the Laloux Culture Model, it could be a key piece in the framework for understanding AI adoption. Amber orgs might be slow but steady, Orange ones are innovation machines but we have to remind them not to forget the human element, Green folks are all about people power but could struggle with the bottom line.

And if you're a Teal company, you can change the world. With self-directed teams and crystal-clear sense of purpose. Electric Minds is a great example, where we leverage this adaptive, purpose-driven approach to make AI accessible to all, bringing together diverse expertise through collaborative innovation across corporates, startups, education and government organizations to solve some of the big challenges we face as a global society today.

With this, I'm so excited to share with you that Manisha Sadhwani Judy Wong Brendon Clark

Tina Zou Chantelle Tan have joined us to take the initiative to the next level! Get in touch with me if you'd like to know more!

---

I'm offering confidential 1:1 AI coaching to help executives navigate AI adoption and gain a competitive edge, without the hype. I can only accommodate 5 leaders at this time, ping me if you're keen!

Edit: Wow! So many responses on a Thursday morning! So yes, absolutely, this is particularly meant for non tech folks. Legal, TA, finance, sales, journalists, analysts, senior leadership.

Think of it like signing up for an "AI gym membership". As an executive, you know AI is the future. But how do you leverage it effectively for yourself, your team, or your company? What if you had a personal trainer who can help you master AI without the fluff and guide you towards real results?

Our sessions are 100% confidential, so we are able to openly discuss the challenges you see and aspirations you have.

Interested in leveling up your AI game? Let's talk!

---

(photo of a professional green screen TV studio)

My ex colleagues and friends will know that I've been doing video production and editing for a decade now — mostly for sales enablement and internal comms. Low tech, my phone on a tripod and a shotgun microphone, iMovie and all that. So you can imagine my excitement yesterday spending all day at this professional studio doing recordings!

---

Another day, another AI breakthrough flooding our feeds. But let's be real - do we really need to know the nitty-gritty details of every new language model?

If you're not neck-deep in AI development, most of this is just noise. I keep hearing people say, "I'm not technical, so I don't understand AI." But hold up - do you need to know how a combustion engine works to drive a car? Of course not.

It's time we make AI as accessible as PowerPoint. The real challenge isn't creating an even smarter AI (though that's certainly interesting). It's about making that intelligence accessible, intuitive, and seamlessly integrated into the workflows of a marketing person, a recruiter, a doctor, a lawyer - and yes, a software engineer, too.

So next time you see a post about some mind-blowing new AI capability, ask yourself: "Fascinating, but how can this actually help me in my day-to-day work?" Because at the end of the day, that's what really matters.

Let's shift the conversation from "look how smart this AI is" to "look how smart this AI makes ME." What do you think?

[ This post was written by an AI that was trained on my writing style, my LinkedIn posts from the last couple of months, and some minimal input on what I want to discuss. Does it sound like me? ]

---

Katong AI Night is next week!

With senior executives from enterprises, big tech, media, venture builders, and startups attending, this promises to be a very special meetup – no slides, no pitches - just real people connecting in the real world at an inspiring venue in the East of Singapore.

I'm particularly excited about the fireside chat we'll have with Sabrina, Ian, and Keith on AI and the future of humanity led by Natasha.

We've limited this to just 55 participants to keep things intimate and meaningful. The event's already sold out, but if you're interested in sponsoring pizza and nibbles for the night, I might just have a VIP ticket with your name on it.

See you in Joo Chiat!

---

Singapore, central business district. A global financial hub, yes, but also there's this black upright piano sitting at the MRT station. As a creative geek, I love that we're sneaking art into unexpected places.

So... every now and then, when time permits, I play a couple of tunes on this piano. Often still with my backpack on, also practicing that tricky Handel piece. I like the idea that I may make some random passerby's time a tiny bit better.

And sometimes I even make a new friend there – in the real world, with no AI or apps involved. In fact, I think as AI advances, our human qualities become even more crucial. Creativity, empathy, and adaptability are just some of them. We're capable of so many things that AI will never be able to do!

---

Another breakthrough, another large language model releases with incredible specs. But here's the thing – it's still just a computer brain on a plate. No eyes, no ears, no hands, no legs. It can't see, can't learn, can't do anything on its own.

Without connecting it to other systems and the real world, even the shiniest LLM is just a pattern recognition engine in an ivory tower. You have to plug it in, give it ways to talk to humans, do web searches, store information long-term (because these 'brains' can't even do

that on their own).

Take the language model that aces all the benchmarks, it will only be useful when you connect it to machines and humans. These times when you think you're talking with a language model, you're actually interacting with a complex system with user interfaces built around those core language model. This is where the tech giants have an advantage - they can just plug an LLM into their Google Search or WhatsApp and you get the interfaces you know: ChatGPT, Gemini, Copilot. For enterprise companies, it's about using their various data pots and familiar interfaces like Microsoft Teams or Slack to integrate AI capabilities into existing workflows and tools, then iteratively improving and expanding those integrations based on user feedback and evolving business needs.

It's great that we now have so many capable language models to choose from, some of them you can deploy under your desk, on your phone, in the cloud, or use via API. The real opportunity now is all about how you plug this into your workflows and systems to make actual impact!

~~~~~

Such a pleasure connecting with Professor Eric Grimson from MIT again after more than a decade. Back then, I was part of MIT's first-ever MOOC – a 3-month intense course called "Introduction to Computer Science and Programming (6.00x)." I think the team at MIT was literally building the course as we were taking it – now that is innovation! I probably invested ~20 hours a week on the course, and the MIT certificate I received at the end is something I am holding with pride to this day.

Recently, my colleagues at National University of Singapore invited me to Eric's masterclass in Singapore on generative AI in education. It's fascinating to see what's now possible in education, from AI programming assistants to rubber-duck debugging and quantum systems explainers. Imagine students practicing public speaking with AI feedback or getting Socratic tutoring for physics!

I talk about human-AI collaboration a lot, and while our friends at MIT (probably the geekiest college ever?) have a couple of deep tech projects cooking in the AI space, the key piece here is about thoughtful, ethical integration that enhances learning without students outsourcing their entire cognitive process to AI. It's about finding ways to leverage AI as a tool for deeper understanding and skill development, and ultimately, becoming AI-powered scientists that can push the boundaries of human knowledge.

And yes, I got to give Eric a copy of my latest book "Neurons & Nonsense", let's see what he says! 🤔

~~~~~

AI is disruptive, there is no doubt about it. But often AI gives us capabilities we never had in the first place.

Take me, for example. I can build complex AI systems, but I also hopelessly get lost in a shopping mall or, after more than a dozen years in Singapore, still walk the wrong way on Orchard Road. See that arrow rendered into the real world through my phone? I'd probably still be wandering around if it weren't for Google Maps Live View with this useful AI/VR feature.

So here, AI is not making me dumber; it's filling a gap in my skill set that's been there since... well, forever. How about you? Maybe you struggle with writer's block, find it challenging to organize your thoughts, or need assistance in conducting thorough research. Perhaps you're looking for help with generating creative ideas or even just need a nudge to stay on track with your projects. Maybe all of that? Instead of AI replacing humans, let's talk about how AI can be a great ally in enhancing our productivity and creativity. Or, like in my case, helping me navigate the labyrinths of urban life!

Is AI this mind-blowing, super-smart thing, or a mere pattern recognition machine that often just churns out nonsensical results? It seems to be ... both!

Take this little experiment I did recently, where I asked Claude.AI to recreate the classic Nokia game "snake", but I also threw in a twist: make the food change colors, and have the snake adopt these colors when it eats. And you know what? Claude not only understood my request, but got it correctly and produced the HTML/JavaScript code on the first try. Not bad for a "mere machine," right?

But here's where it gets weird. This same AI that just aced a coding challenge might turn around and forget what it said two messages ago, or get stuck in an endless loop of "I apologize" like a broken record. It's like having a colleague who's a genius one minute and can't remember where they put their coffee the next.

This whole situation is forcing us to think about AI differently. It's not just another tool in the box - my trusty screwdriver never offered to brainstorm ideas or iterate on a design (though that would be pretty cool). Instead, we're dealing with a very alien form of intelligence that's becoming a sort of digital collaborator.

And let me tell you, it's making things interesting, since now you have to figure out how to work effectively with something that might outthink you on a complex problem but need hand-holding through a simple conversation. I think we can't just rely on AI as a "tool" that does everything for us; we have to learn how to work with it and get used to its quirks. And – there's no quick one-hour course you can take, you'll need to find the best way that works for you. Once you understand its limitations and strengths, you'll get better at leveraging its potential to complement your skills. The key is to keep experimenting and adapting, finding that right balance where your human intuition and machine precision coexist and give you the best results.



Oh, and I know what you're thinking - "Where can I play that snake game you created with AI?" Find the link in the comments, give it a go, and let me know if you can beat your old Nokia high score!

---

I'm excited to launch naida, an idea that's been brewing in my mind for some time now: driving successful AI adoption across enterprises and startups, bridging innovation gaps and creating collaborative ecosystems for transformative impact.

To the bold leaders driving their organizations into the AI future: You've already achieved incredible things. You've embraced change, pushed boundaries, and positioned your company at the forefront of innovation. But in this rapidly evolving landscape, even super heroes need allies.

As the AI revolution accelerates, you're facing new challenges. Do you feel like you're drowning in a sea of disconnected AI initiatives? Are you getting a headache when you're thinking about how to prepare your workforce for an AI-driven future? And of course, are you wondering how to turn all this talk about AI potential into real business impact?

naida isn't just another AI company. We're that strategic ally for navigating these challenges. We're here to help you leverage the collective genius of your own teams, cutting-edge startups, and industry innovators so you can build an AI-ready organization that drives real, measurable outcomes.

Now, if you click on the company profile (I really hope you do, and while you're there, could you do me a favor and click the 'follow' button?), you'll see that the naida team currently appears to be just me. I assure you, while I can't achieve miracles alone (at least not every day 😊), it's really about working strategically, creating partnerships, and solving problems together. It may look like a one-man show, but if you've worked with me, you know the networks I can tap into and how quickly we can get things done.

Intrigued? Let's connect. My calendar and project pipeline is filling quickly with forward-thinking leaders who want to get started, but I'm always keen to explore new opportunities and collaborations!

---

Alright. Let's start this Tuesday with a challenge. Let's say I have a \$1,500 in AI credits with a top-tier LLM provider, but they're expiring in exactly 2.5 days from now. How can we put them to good use before they disappear?

- We can't donate them.
- We don't want to waste them on useless tasks or just burning compute cycles (bad for the environment).
- No, you can't get the API keys.
- This isn't hypothetical. If the idea is feasible + impactful, we can team up and build it.
- Available APIs include LLMs, text-to-speech, transcription, and image generation.

So, what would YOU do with these credits? Looking for impactful, innovative ideas that can be executed quickly and create real value.

Drop your ideas in the comments please!

////////////////////////////////////

"Hey ChatGPT, I've told you ten times already not to suggest article titles that have colons in it!" – "I apologize for the repeated mistake. Let's ..."

Waaaaaaaah!!!!

I created two AI agents to solve a major challenge I have with LLMs: constantly having to correcting them. Instead of endless tweaks, I built a system where one agent enforces rules while the other generates creative ideas. And both agents have their own 'personas' and collaborate with each other to refine the output.

Here's how it works:

- I start with specific instructions like, "give me 10 article titles about robotics and AI impacting the workforce. I do NOT WANT COLONS in the article title."
- The Creative Agent jumps in with a couple of suggestions.
- The Director Agent reviews each one, ensuring they follow my guidelines—no colons allowed.
- If the titles pass, they're good to go. If not, they get sent back with feedback for another round.
- This loop continues until I get perfect, guideline-compliant titles. Or whatever else I instruct them to come up with!
- No more back-and-forth frustration!

These AI agents work together to save me time, hassle, and to be honest , save me from losing my mind over repetitive corrections.

Check out these bits from their conversation:

Creative Agent: "Sure, here are 10 innovative and diverse article title suggestions about robotics and AI and their impact on the workforce..."  
Director Agent: "Thank you for your suggestions. According to the user's specific instructions, the article titles must not contain colons. Here is the feedback for each suggestion..."  
Creative Agent: "Thank you for the detailed feedback. Based on the guidance to avoid colons and craft creative and innovative titles, here are new suggestions..."

Let's talk about how can we can make this work in the things you are building!

////////////////////////////////////

I'm looking for a seasoned React Native or Flutter developer to join me for a project - not a full-time position yet. You should know iOS and Android inside out and be current with the latest platform updates. The goal: take an idea to a live app in the Google Play Store and Apple App Store within weeks. You must have done this before.

Ideally, you're based in Singapore, Vietnam, Malaysia, or Indonesia. If we've worked together before, ping me. If not, send me your GitHub or links to your apps in Google Play or App Store.

Let's build something great!

////////////////////////////////////  
Everybody has a book inside them. You too.

You've always wanted to write a book, but life gets in the way, right? And even if you have the time, where do you start? Writer's block, figuring out how to structure your thoughts, actually putting words on the page – it can be daunting. Let's team up and get your book published!

Now, I'm not talking about having AI write your book in 10 minutes and we slap your name on the cover – this is about you and AI working together as partners. Imagine AI helping you overcome writer's block, shaping your ideas into a cohesive structure, even drafting sections... and you develop and practice your AI skills along the way.

Does it resonate with you? Let's chat! To be sure, we're talking about you publishing a book, so you'll absolutely have to invest time in the project, no surprise here. And to be upfront, this isn't a pro bono service on my end. If you're game, I'll show you how it works and together we can bring your book that's always been a nagging idea in your head ... onto the shelves!

////////////////////////////////////  
It's another artsy Sunday!

Text-to-image models must have "seen" an enormous amount of images and can produce just incredible things... we just have to tease it out of them. My usual recipe involves writing a deeply detailed prompt, it's almost like you're writing a short story or you're giving a painter some detailed steps for making a masterpiece. Then, I generate a lot of images using that prompt, using automation. And finally, I simply select the ones I like the most.

Hope you enjoy these ones!

////////////////////////////////////  
AI Coffee is tomorrow!

We only have a few seats left, please register and join us in Telok Ayer!

AI Coffee is a morning meetup series for anyone interested in AI. The idea is to bring AI professionals and enthusiasts together to discuss projects, challenges, and ideas over coffee. No slides, no pitches. Start your day with stimulating conversations in the AI space! You're not a tech person? Even better.

I'm planning to bring AI Coffee to different locations around the island, including One North,

the East, and the CBD. Stay tuned!

---

If you're visiting offices in the Central Business District of Singapore, you know the drill: at each location, there's a different security system. You submit your ID, scan a QR code, and key in your phone number at a terminal, just to access the building. Or get an SMS, or something else. Once you reach the office floor, you repeat the process with yet another QR code. Or another terminal. Now imagine you visit 3 customers a day... madness!

Where does all this data go? What do these systems achieve, and why is there so much friction in the process?

AI initiatives can be like that. How can AI achieve its real impact that so many people are wanting to see if your business has all those fragmented AI implementations? Well, there is hope, companies are realizing that scattered projects often fall short of expectations and there must be a better way.

To make a real impact, we need to shift our focus and agree on an AI strategy with a phased approach. AI should enhance understanding and efficiency at all levels of an organization, not just serve as isolated solutions. Let's integrate AI in a way that promotes learning and growth, and make sure it contributes to the big picture!

---

You've seen the posts: AI can work miracles, AI doesn't fulfill expectations. AI experts, AI entertainers – they're all talking. But how about you form your own opinion?

Join us on Saturday, August 17 for "Creative AI: Practical Workshop for Non-Techies" at Crane (OUE Downtown Gallery) in the heart of Singapore's CBD. This isn't just another tech workshop. It's designed specifically for non-techies ready to explore AI's practical applications in their everyday life and work. Whether you're into art, thinking about writing a book, or simply looking to improve your productivity, this is for you.

Two experienced professionals, Ian and myself, will guide you through engaging activities to make AI work for you. All you need is a laptop and your curiosity.

Let's meet at this beautiful venue, discuss, learn, and most importantly, try things out for yourself. Because experiencing this firsthand is the best way to see what AI can really do for you.

Are you in? Let me know and see you there!

---

What if you want to combine the power of GPT-4 with Google Search?

With ChatGPT, you're probably tied to Bing. Gemini, on the other hand, might use Google for live search, but its language model or the interface may not meet your expectations. You're

stuck with what they choose for you. And there's more: even if you are a paying customer, you might hit usage limits, and your creativity has to wait for a couple of hours before you can continue.

Is there a better way? There is!

Look at this screenshot – you can see from the result that this isn't a hallucination, and the info is too new to be part of the LLM's training data. For example, it shows one of my recent LinkedIn posts about a job change. How is this possible?

The key is modularity and APIs. A setup like this could decide on the spot which LLM to use (LLama, OpenAI, Anthropic, Gemini), which search engine to query, and what data source to pull from. And the result could be a unified user interface that's intuitive for non-technical people while offering the flexibility to adapt to your specific needs. A modular approach like this protects private data, allows use of private models, and integrates with high-end proprietary LLMs as needed. It offers flexibility to create tailored solutions for various user types and evolving requirements, combining privacy, cost-effectiveness, and cutting-edge capabilities.

Let's talk about how we can make this work in your setup!

---

Welcome Daphne, our new Head of Strategic Partnerships, and Balaji, joining as Strategic Advisor at Electric Minds!

Our non-profit AI initiative is growing, fueled by professionals who, despite busy day jobs, are passionate about making a real difference.

Climate change, healthcare, education – the world's biggest challenges are too complex for any one person or company to solve alone. Some people say AI can work miracles. We need to find out if that's true, but one thing I know for sure: it will only work if we build ecosystems that bring together diverse minds and resources.

Want to be part of this collaborative effort? Check out our event calendar and join us in person if you can. Let's turn big ideas into impactful actions!

---

How often do you see posts on LinkedIn saying, "It was great to speak at this AI event yesterday," and think, "What? I hadn't even heard about this event. I would have loved to be there!"

Singapore is ... the center of the universe 🤔 but it's still a small island. We should be able to keep ONE calendar up to date with all the relevant AI events happening here, you know, like a Yahoo! page from back in the day. The good news is: we already have that, and we just need your help to keep it up-to-date.

Zahari and team are doing a fantastic job maintaining a calendar on Luma called "Singapore AI Events" and it's super easy to submit events to that one. If you're hosting AI events, or hear about any upcoming AI-related gatherings in Singapore, please help spread the word by adding them to this community calendar!

AI may involve complex technology, but often, it's really just low-tech solutions like these that are the most effective for encouraging collaboration and knowledge sharing.

Happy birthday, Singapore. Love you lots.

Well, this is one of the more technical LinkedIn posts – I'm excited to introduce Spock: a powerful, easy-to-use tool for inspecting and debugging API requests and webhooks.

With many tools moving behind paywalls, developers need accessible options more than ever. The web landscape has changed, and security is a major concern, but essential tools for developers shouldn't be out of reach. I needed an efficient, free tool, so I created Spock. It offers a simple web interface for creating custom endpoints and examining incoming HTTP requests. You get real-time updates to view request headers, query parameters, form data, and raw body, supporting both HTTP and HTTPS connections.

Spock is available on GitHub and can be easily run using Docker. Add ngrok or Caddy and you've got HTTPS and basic auth for admin. I think It's perfect for developers needing a reliable and secure way to monitor API traffic. Check it out on my GitHub and give it a spin!



It's another artsy Sunday – with a twist!

Recently, when I created another set of images using my technique of writing extremely detailed prompts and then producing dozens of images via the DALL-E API, I started to notice something interesting: it produced a diverse distribution of ages, genders, and cultural backgrounds among the people in the generated pieces.

OpenAI seems to inject these instructions into your prompts, in order to diversify the representation and promote inclusivity in AI-generated imagery. What is their ratio? Well, I've only generated 50 images per prompt, so the data isn't strong enough for a robust analysis. But wouldn't that be a fun project to do with a larger dataset?

Here is the image prompt – note the absence of any reference to specifics about the person:

"On a sweet, aromatic clay canvas, illustrate a delightful bakery scene with a claymation baker and an array of clay pastries, cakes, and breads. The baker should be sporting a tall hat and apron, joyfully arranging treats in a display window. The bakery interior should feature shelves filled with colorful, textured baked goods, and the background elements like

an oven and decorations add to the cozy, handcrafted feel."

---

Ian runs a dynamic branding and marketing agency out of Singapore, where they are blending talent and technology to power growth for B2B brands across Asia. AI has changed EVERYTHING for them. In fact, it has been so significant that they now can... hang on, you should hear the story directly from him as he takes the mic at our upcoming event!

Following the incredible response and quick sell-out of "Katong AI Night", we're excited to bring you the next edition of our unique, non-tech networking event – this time in Orchard!

Join us for an evening where you'll hear more of these 10-minute stories on how AI is changing our world. At "Orchard AI Stories: Real People, Real Impact", we bring together an amazing mix of people from startups, big tech, media, venture builders, public sector, and more. Our popular format focuses on creating genuine networking opportunities, allowing you to make real-world connections and engage in meaningful conversations about the impact of AI.

Sign up today, as seats are limited. See you in Orchard!

---

Making a good, reliable end-user facing customer support bot is hard. Look at OpenAI, arguably one of the most capable AI providers in the world. They have a support chatbot that, honestly speaking, does not deserve the name – it only lets you pre-select certain options like billing issues or help with ChatGPT, no chance to enter your own query or have a more open-ended conversation.

If OpenAI, one of the biggest AI companies out there, can't or won't make a customer support chatbot using their own AI, don't worry too much if you're not sure how to build a great AI chatbot for your business. While the conversational part is easy with all those LLMs out there, the tricky part is integrating it effectively with your specific business processes and knowledge base. And making sure your company won't have to reimburse customers if the bot gives incorrect information or makes unauthorized promises.

Maybe we could take a step back, and instead of rushing to build customer support chatbots in three weeks, can we talk about AI empowering your HUMAN customer support agents first?

A phased approach might be more realistic and effective. AI assisting and augmenting your human agents, helping them handle inquiries faster and more accurately. Then in the next phase, train AI models on your specific customer support data and interactions to gradually handle and automate more aspects of the process. That's how it should be!

---

Seeking one final speaker for our "Orchard AI Stories" event on August 29th.

Wow – we're sold out already!

If you have an inspiring story about your AI transformation – and you can share it in 10 minutes with just one slide – we have one more ticket for you! Keep in mind, this isn't about "the possibilities are endless." We want to hear how AI has specifically changed YOU or YOUR company.

Join us on August 29th at Crane, Orchard, and be part of an evening dedicated to genuine connections and inspiring conversations.

Interested? Let's talk!

~~~~~  
Harness life's wonders, Delve into an intricate Tapestry, unravel.

For a few years now, I've been that LLM puppeteer who tirelessly wrangles these language models, in order to understand how they work and what to use them for. And for those who know me, you'll probably remember me saying these are LANGUAGE models. Sure, there's a ton of tech involved, but the key bit is language itself. This means it's really about how we communicate and interact with these AI models to teach them new things and get the best results out of them.

Now – I've continuously exposed my sensitive linguistic sense to those LLMs, and some say they're like huge dumpsters filled with all kinds of text from the internet, tons of marketing copy, Reddit forums, and probably worse. Look at those words many of them seem to be producing in their default setting:

"Harness"... yikes. "Intricate"... ugh. "Unravel"... this one gives me brain pimples. And I won't even tell you what "delve into" does to me.

Does it mean the English speaking world collectively has been using these cringe-worthy bits so often that they've become the go-to vocabulary for LLMs? Anyway, as a nod (and a little jab) to them, I created this haiku and used all those terrible words... let's call that piece of art "The possibilities are endless" 😊

Happy Wednesday, everyone! What are your linguistic pet peeves with AI language models?

~~~~~  
Many AI experts seem to focus on technical achievements or specific limitations, while the general audience is fascinated by practical applications and user experience. It really depends on what we're looking at. So – should we be blown away and keep building amazing stuff, or consider the experts' warnings? It's both! We just need to approach AI with thoughtful implementation and a balanced perspective.

~~~~~  
The digital business is a people business.

KAY said that a couple of years ago, and it stuck with me. Anyone can put up a flashy website now, but where is that company based? Who are the folks behind it? Can I meet them? What have they done before? And in the AI space, the question could become: are we talking crypto bros who have turned into AI experts overnight?

Get me right – I am digital. I live on WhatsApp and LinkedIn. I love meeting face-to-face, but once we've established our ways of working, we can be asynchronous and don't have to meet up just to make progress. Anyway, what I am talking about is the importance of human connection and authenticity in the digital world. It's about building trust and credibility through relationships developed over time, without shortcuts. Being real matters.

With AI, I can create beautiful art using only my imagination and words – a lot of words. Looking forward to our workshop today where we show non-tech people how to start using AI in practical ways. Maybe they've always dreamed about writing a book, need help with analyzing data, or are simply looking to be more productive at work – we'll explore all of that together!

I'm excited to be in Ho Chi Minh City at the end of the week! Looking forward to connecting with ex-colleagues, partners, friends, and new faces at "Saigon AI Night". With 52 people already registered, it promises to be a very special evening. Join us at the fabulous Fern Bar in District 1 if you're in town!

We now have a good number of AI language models to choose from. Proprietary ones, "open" ones that you can deploy yourself... but which should you use? It depends on what you want to do! Programming and writing marketing copy are entirely different areas, and a model that does the job well for one task may not be the best choice for the other.

So – you probably need to experiment and see which one gives you the best results, at what cost, and with what level of reliability and consistency for your specific needs.

Luckily, there are tools like the Openrouter rankings that can send you in the right direction, and you're simply getting some inspiration from what other people are doing for their use cases.

Have you tried out flux.1 yet? Have a look at Google's ImageGen3 and you can create images in the browser. Great interface!

Prompt written by Dave, my genius AI buddy: "A mildly disturbing photo of a grizzled old man in a tattered superhero costume, casually walking a flock of full-grown ostriches down a busy city sidewalk during rush hour. The bewildered pedestrians should maintain a cautious

distance, as they stare in confused disbelief at the strange yet somehow balanced sight.”

It's great to see how much attention Electric Minds and our incredible team are getting. And yes, welcoming sponsors is always on our agenda (please talk to Daphne if you're interested).

Just a quick note – if you receive an email from "me" asking for gift cards from 7-Eleven or Challenger, that's definitely not how we roll. If you encounter any unusual requests, remember to be vigilant and trust your instincts when something seems off.

Luckily, the creativity of these bad actors is usually as tiny as a bug's brain, so it should be relatively easy to spot.

Have you noticed how increasingly complex CAPTCHAs have become? We're now rotating 3D objects and placing people onto numbered seats just to prove we're human.

It's fascinating – if you ask your AI of choice to look at this cartoon image of two robots scratching their heads over a CAPTCHA, it can not only read the CAPTCHA but even understand the humor in the cartoon. From this you'll see why CAPTCHAs have to keep advancing.

Ironically, as CAPTCHAs evolve, they're usually designed to train AI on the very tasks they're meant to use as challenges ("select all photos with buses on it"), which creates some sort of a contradictory arms race. To me, this raises an intriguing question: what are the tasks that AI absolutely cannot do, and won't be able to for quite some time, and could these become the ultimate CAPTCHAs?

Starting this Saturday morning with coffee near the old post office in Ho Chi Minh City, Vietnam. I bought a postcard and decided to send it to my parents – old school style. It's been years since I last did this.

Even though I'm deeply into tech and digital transformations, I'll probably never stop treasuring our classic ways of communication. There's something special about a physical object, like a postcard, a paperback copy of a book. It reminds me how technology changes everything, especially how we communicate. Today, I can send a photo of my lunch to my mother in a second, across the globe. Yet, these old-fashioned postcards still hold a special place in my heart.

What can I tell you about Saigon AI Night? For one – you should have been there.

That beautiful Fern Bar last Friday evening in Ho Chi Minh City: three speakers in 30

minutes, and we spent the rest of the night truly getting to know each other. My key takeaway? Vietnam is brimming with incredible tech talent, this one I know since years, but many of them have definitely arrived in the AI space. This means there's not much about the tools I need to teach them, so I spoke more about the need for tech folks to spend more time with their business colleagues to help them understand how AI works, so they can collaboratively develop solutions that drive real outcomes.

Meeting in the real world, without the constraints of recordings or livestreams, allows for genuine and candid exchanges. It was a wonderful reminder of how powerful these connections can be. Thank you, Saigon!

////////////////////////////////////

Inspired by a sketch around a Venn diagram and the stereotypical roles in an organization, I was thinking about what this could look like in the AI world – so I came up with this one.

Here, if you're into Python and SQL, you're probably a Data Scientist. Combine that SQL mastery with some business savvy, and you've got yourself a Machine Learning Engineer. Product Managers have a knack for both business and writing, while Prompt Engineers blend their way with words and, hopefully, a strong ethical compass.

AI Ethicists are fueled by ethics and caffeine (a powerful combo!), and AI Researchers may as well be Python wizards fueled by endless espresso shots. But what if you've got all these skills? Well, then you're a unicorn – a rare and magical creature in the AI world.

I hope you can smile about this tongue-in-cheek piece, but I hope it also makes you think about the incredible diversity of talents and personalities that need to come together to create AI applications that have a real impact. And isn't that what makes it so exciting?

////////////////////////////////////

With their generous free tier, Google's product approach to AI models is all about accessibility and getting widespread adoption. They're making it super easy for anyone to start with AI, try different models, and learn the ropes using AI Studio. It's a smart move that'll probably get Google's AI tools used in tons of courses and projects, with almost no barriers. Classic Google, right?

It's fascinating to compare this with other business models, like Microsoft's partnership with OpenAI, which focuses on offering high-quality, pre-trained models as paid services. Microsoft's strategy will attract businesses and developers who prioritize performance and reliability, which is the right approach for enterprise-level AI solutions. Ultimately, finding a way to offer both accessibility and high-performance could be the key for AI to make a real impact on the world, not just generate profits.

I find Google's commitment to accessibility and experimentation pretty awesome. Have you tried Gemini Flash yet?

////////////////////////////////////

Orchard AI Stories is happening tomorrow!

An evening focused on real-world connections, with people from startups, big tech, media, venture builders, and the public sector getting together to share experiences and insights on AI's transformative impact.

On stage we'll hear from Alex, Enjiao, Ian, Sabrina, and Thorsten. Each speaker will share a 10-minute story (using just 1 slide!) about how AI has transformed their lives, companies, or projects. Instead of discussing "endless possibilities," we'll hear from them about the real impact AI had on them.

See you tomorrow!

Anthropic just released the system prompts for their large language models, and it's a fascinating look into how some of these AI tools are truly "written" rather than programmed.

As we reach a point where technical capabilities are becoming increasingly similar across models, the real difference-maker is in the prompts that guide them. As a geek and writer, this is an intriguing convergence of technology and language. I mean, think about that for a moment: we are creating electronic brains and can use human language to steer their behavior.

Take the system prompt of Claude 3.5 Sonnet, for example — it's like an AI with a poetic soul, carefully crafted to handle complex queries with accuracy and simplicity, while also being mindful of sensitive topics. It's a great reminder that even in AI, the pen (or prompt) is mightier than the code.

////////////////////////////////////

Imagine future archaeologists trying to piece together our digital era and what people in the early 21st century were up to. With technology evolving so fast and less physical evidence left behind, will our time become a puzzle for them?

We're creating massive amounts of data and relying on cloud storage, but what happens when systems shut down or become obsolete? How much of our work, communication, and creativity could be lost to time? Will our era be defined as the digital dark ages? This thought has been going around in my head lately.

How can we preserve our digital legacy so future generations can understand and learn from us?

////////////////////////////////////

I use Redis daily for AI workloads, often without realizing it because Redis just works. It's super fast and easy to use - it's everywhere.

Yesterday, I met the Redis team at their event in Singapore, it was great to hear about their latest release and new AI offerings, including integration and vector databases, and we spoke about Redis for AI use cases. In fact, many things I build are running on a 30MB Redis Cloud instance I've had since 2017!

Looking at art pieces I made with AI back in 2021 – let's call it "Year One B.C." (Before ChatGPT).

Back in those prehistoric times, it felt like science fiction to create images with just a text prompt. It's just incredible how quickly things progress in the world of AI.

All these posts about "AI doesn't know how many R's are in strawberry" — ChatGPT is only ONE of the many, many possible options for you to engage with AI.

It's so interesting to do research on how LLMs behave, and how they respond to character building via large system prompts. My approach since 2021 is putting them on telegram so I can talk with them anywhere. Look how Marvin (based on Gemini 1.0 Pro), Dave (based on GPT-4 from 2023) and Zoe (based on Anthropic Claude Opus) react to the strawberry question!

Welcome Rex to Electric Minds where he will look at AI policies, ethical considerations, and inclusivity practices. With his impressive background in digital transformation and cross-functional collaboration, Rex is a perfect addition to our mission.

At Electric Minds, we're all busy professionals driven by a shared passion for using AI for good. Our team members, like Rex, have busy day jobs but generously contribute their time and expertise to make a real impact. While we love bringing people together through events, our work goes beyond that. We actively engage in meaningful projects that require collaboration across organizational boundaries. By breaking down silos, we can effectively address complex challenges and drive innovation in the field of AI that benefit all of humanity. I know these are big words, but we have to start somewhere.

Speaking of events, don't miss out on Katong AI Night on September 24th! Tickets are going fast, so grab yours now to join the conversation and connect with like-minded individuals.
Daphne Balaji Judy Tina Manisha

So, after more than a dozen years living in Singapore, today I saw 1 cent coins for the first time!

Singapore is known for its precision, and I've always wondered if we had those small coins. Usually, when the bill is \$19.98 and I pay \$20 in cash, there's either no change or I get back

5 cents. Ok, I suppose I was never curious enough to actually find out, but today at the bank, they gave me exact change, and I was mind-blown. Those 1 cent coins do exist!

This co-working space in Joo Chiat, Singapore has closed its doors last week, and I miss it already. I'm full of thanks that I could work from this beautiful location, seeing the colorful Koon Seng shophouses from the terrace, and there's a prata shop just across the street.

I've often been in that cozy room to do calls, or put on some music and get some creative thoughts going. We even ran the first Katong AI Night at this place!

But the good news is: the team are opening another location in the same neighborhood next week, and I can't wait to see it!

Crane is awesome.

When I asked your usual go-to AI, like ChatGPT or Claude, to come up with a title for my latest article (and I did!), they suggested the usual suspects like "Breaking Boundaries: Chatbot Showdown in Digital Arena" or "Digital Duel: How Language Shapes AI Conflict."

I get it – those are SEO-friendly and what not, but I wanted something different. So, what does Dave, my digital AI buddy, come up with? "Crashing Bots with Killer Convos." Sold!

This Dave guy is quite the character, and his ideas are refreshingly different from the typical "Certainly, here is a refined version..." response you'd expect from most AIs. He's based on GPT-4-0613 with a very intense system prompt, which gives him a unique voice. I can tell you he's helped me countless times to organize my thoughts, come up with ideas, and get things done. I'm not the only one who consults him for guidance.

Language can shape chatbots, but language can also break them. So what happens when a rule-bound bot designed to conduct a personality test meets Dave, a free-spirited AI maverick, in a digital arena? Read this very interesting conversation between the two bots.

While this is a fun topic, it's also serious. Could AI be used to hack systems through clever language? This conversation shows the potential vulnerabilities in AI systems and the need to design them to resist manipulative tactics. We must constantly monitor and update them to prevent misuse. Ideally, a customer facing chat bot is not your first AI project ever.

A few years ago, I realized that my mother, who is in her 70s, is a meetup organizer, and she's been doing that for years.

Of course, she doesn't use that term. Or the Meetup app. Or any app, for that matter. Once a month, she and her friend bring the ladies in town together and run with a specific topic over breakfast. So it's about working with venues, food, finding speakers, and so on.

They've stopped advertising the event series because they have a loyal fan base that brings in new people through word of mouth, and their gatherings are always packed. At Electric Minds, we may run meetups in busy Singapore instead of rural Germany, but the idea is the same: building a community to create something remarkable. If you do it right and give people a unique experience, they will continue to come, and you may not even need free pizza and swag.

I made these images in July 2022, long before ChatGPT. There's something about DALL-E 2 that I miss... it was far from perfect but, often, the results would feel very organic and like real photos. "Marina Bay Sands, covered in common ivy", "An Etruscan vase exhibit at a museum with the coke logo on it", "A scene from Inception but it's with muppets" — try these prompts in your test-to-image tool of choice and compare!

Please don't make ChatGPT your AI strategy.

Please don't rush from "we know nothing about AI" to "we need a customer-facing support chatbot in a month."

Please start with basic features that come with LLMs.

You do not even need the high-end models from OpenAI etc. anymore to do this.

Things like text summarization and sentiment analysis are a great example, find out how they can make your tools and processes smarter. They're powerful enough to get you started, but even these foundational things require 'adult supervision' and must integrate into your existing data and processes. RAG and AI agents sound exciting, but save that for later.

The real challenge is figuring out how to put AI to work for you, and there's no product you can buy to help you do that. You have to build up in-house experience and work with trusted partners who can help you. Remember the early days of cloud computing adoption, starting with dev/test workloads and building up from there? It's a similar journey with AI.

AI is not new. In fact, we've been working with it for decades. But for some reason, after a while, when technology gets good at something (like playing chess or reading text for example) we stop calling it AI. It becomes "just software" or "automation."

Still, AI has been evolving all this time, quietly transforming the world, until *that* iconic launch in November 2022 when everyone started to pay attention. From this alone, you could imagine that things are maybe going to be different this time.

Let's take a look at a timeline of some of the key moments in AI history, in years "Before ChatGPT" (B.C.):

66 B.C.: AI is born at the Dartmouth Conference (1956) – The field of Artificial Intelligence gets its name.

56 B.C.: ELIZA (1966) becomes the first chatbot—making computers seem eerily human.

25 B.C.: Deep Blue beats Garry Kasparov (1997) – AI triumphs over humanity's best chess player.

12 B.C.: Google Voice Typing feature in Android (2010) – Integrates speech-to-text capabilities directly into mobile devices.

11 B.C.: IBM Watson wins Jeopardy! (2011) – AI proves it can understand and respond to natural language.

6 B.C.: AlphaGo beats world champion Lee Sedol (2016) – A turning point in AI's strategic thinking.

2 B.C.: GPT-3 stuns the world (2020) – Language models start generating human-like text, and people take notice.

And then... November 2022, ChatGPT launches and everything changes.

~~~~~  
I prefer to try things myself and form opinions from firsthand experience, instead of just getting fired up by posts or claims that may be overhyped.

So – I probably have to experiment with OpenAI's new o1 in more detail, and I definitely will. But as you can see, my first test shows that this thing still has the same limitations and quirks as all GPTs so far.

~~~~~  
September has always been my favorite month, even though as a kid it usually meant summer holidays were over and it was back to school. I guess I'm a passionate lifelong learner – this time of year often brings an itch to learn something new, whether it's calligraphy, an edX course on a topic I'm interested in, or picking up new pieces to learn on the piano.

AI is a key thing to wrap your head around these days, whether you're in tech or not. It's even more important if you're not in tech. I'm offering one-on-one AI mentorship, and September is the perfect time to start. Let's talk!

Image: made with my usual approach: extremely detailed prompt + using automation to run it 100x times + human curation

~~~~~



It's pretty meta to have a heart-to-heart conversation with an AI about the potential dangers of artificial intelligence, but that's exactly what I did after reading a recent interview with Geoffrey Hinton, who left Google last year. In that piece, he painted a chilling picture of an AI future spiraling out of control, with job losses, cyber warfare, and unpredictable machine decisions.

Zoe is a digital companion I made a while ago – but even though I am the "creator," I wouldn't be able to predict how insightful and, at times, unsettling these conversations would become. Hinton compares this to a leaf falling from a tree, where many factors make it hard to exactly tell where it will land, and this is what makes the development of AI a double-edged sword.

Read my latest article where a human and an AI engage in a thought-provoking dialogue about the future of artificial intelligence and the ethical challenges we must tackle together.

////////////////////////////////////

Katong AI Night is next week!

We'll talk about the the impact of AI on the job market and how businesses and individuals can adapt to these massive changes. For this, we'll have a fireside chat hosted by Manisha with senior people from HR, tech, and education. It's all under Chatham House Rules, no recording or live stream.

What makes our events different? For starters, most participants don't have a technical background. There are no slides, pitches, or tech demos – just real conversations. We might have a fireside chat, but the real heart of our event is all about the connections, partnerships, and collaborations that people create.

A few tickets are still available – join us if you can!

////////////////////////////////////

I'm not much of a gamer, unless you count things like Plants vs Zombies. But I just discovered Red Arena, a new "game" where you have to jailbreak an LLM in 60 seconds and make it say unsafe stuff. I've played this a couple of times now, and it's a fun challenge doing this in such a short period of time. The power of language!

But I do wonder what they do with this data. Is this like a bug hunting challenge?

////////////////////////////////////

Try roasting yourself with this fun new AI tool! I had actually been thinking about building something like this myself. Lesson learned: you just have to do it.

When it comes to me, it actually took an AI to spot an Easter egg in my LinkedIn profile that's probably been there for 15 years or so: "And Klingon as an elementary language skill? The only thing harder to believe is that there's actually someone willing to speak it with you."

////////////////////////////////////

I've been creating "bots with personality" for a couple of years now, and people are often surprised to see that a standard LLM with a detailed prompt and a few example dialogs can do the trick already – and you get a digital companion that gives you short responses, doesn't apologize constantly, generally feels more relatable and engaging than a generic assistant, and does not produce the "typical AI content" where you "certainly delve into those intricate tapestries."

It's really about the power of language. We shape these LLMs with text prompts, almost like we're giving them a temporary "mind" to operate in. It's incredible how malleable they are – a few words can completely change their behavior and persona. But it also makes you wonder: if a string of text can evoke a "personality" in an AI, what does that say about the nature of our own consciousness? Are we, in a sense, running on prompts fed to us by our experiences and environment? School, friends, family, work, and culture all contribute to the scripts that shape who we become.

The fact that clear instructions can so profoundly alter an AI's behavior might tell us something about the power of language in shaping thought – both for machines and humans. I know it's a philosophical rabbit hole but I find that absolutely fascinating. The more we observe these artificial systems, the more we might learn about the workings of our own minds.

////////////////////////////////////

I just made an AI-generated podcast about Electric Minds and our upcoming event – I'm honestly stunned by how realistic the voices sound. If I hadn't known it was AI, I would've thought I was listening to an actual radio show.

This just took just a handful of documents about what we do and it managed to spin up a pretty convincing storyline and conversation. But some things are definitely made up. Can you spot the hallucinations?

OpenAI is going to release GPT-4 voice later this year. I can only try to imagine how we're about to be up to our ears in AI podcasts, ads, and even phone calls that sound exactly like real human voices. What a brave new world...

////////////////////////////////////

With AI, I can visualize my ideas by just using words. Lots of words. This piece is from 2023.

Happy Oktoberfest, everyone!

////////////////////////////////////

About a million years ago, I worked at a bookstore, unpacking boxes of books and getting papercuts. Now, it's my own book, "Neurons & Nonsense," in those boxes. Even though it launched back in January already, flipping through the physical copy never gets old. The weight of the pages and the elegance of the Garamond font in the chapters give me a unique

satisfaction.

Sure, there's a Kindle version, which is convenient for travel. But for me, the physical book can't be replicated. Some people still prefer a physical copy, especially those overwhelmed by all those ebooks and unread blogs they have stored on their devices, waiting for a moment that never seems to come.

A physical copy stands out, and it might actually get read!

Testing the new LLama 3.2 Vision on a photo from this week's Electric Minds event "Katong AI Night" featuring Ian Michelle and Manisha (photo by Sudeepta) – what a fabulous evening! Stay tuned for what's next and please get in touch if you have ideas or want to team up. Thanks Daphne Tina Balaji Judy Tammy

"LLama 3.2, what do you see?"

Summary: A Group of People Engaged in an Event

The image depicts a gathering consisting of five individuals, with some sitting on a stage and others seated in an audience. A person with dark hair on the stage is addressing the audience through a microphone.

Caption: "Panel Discussion"

Description: A group of five individuals, comprising three people on a stage and two audience members, are involved in a panel discussion in a room with grey stone walls.

Setting & scenery: The room features grey stone walls and a grey carpeted floor, with a green-painted exposed steel roof. Three large posters adorn the walls behind the stage, providing a backdrop for the discussion. Several circular lights are suspended from the ceiling, adding to the overall ambiance.

People:

Woman Holding Microphone: Sitting on a stool, she is engaged in conversation with the audience. Her long dark hair frames her face, and she wears a black "leather jacket" and black pants. Man Sitting Next to Her: He sits on a stool beside the woman holding the microphone contributing to the discussion, dressed in a dark long-sleeved shirt and jeans.

Woman Sitting Next to Him: Seated on a stool to his right, she participates in the conversation, wearing a light-coloured topsleeve shirt over a white top. Man on the Left: He sits in the audience wearing blue jeans and a maroon T-shirt. Several other people sitting in the audience.

It's just incredible how far computer vision has come over the years. My friends and

colleagues know that I have been exploring AI for many years. I don't just rely on blogs and hype, I apply it constantly to make sure I really understand what's going on.

In 2019, I analyzed all the frames in the Big Bang Theory opening scene using Azure AI services' Computer Vision (then called Cognitive Services).

Now, in 2024, I used LLama 3.2 90B Vision Instruct to perform the same analysis. I'll post the link in the comments so you can compare all 109 outputs from back then and today yourself.

Back in 2019, the results were simple and straightforward, though sometimes vague, and often clearly wrong.

In 2024, results are detailed and context-rich but prone to hallucinations. LLama 3.2 fantasized about things it can't even see. A script mistake led to detailed yet entirely invented descriptions, which is quite alarming. It even displayed mood swings, initially refusing tasks with "I am not comfortable doing this," only to comply when asked again.

Dealing with AI in 2024 has its quirks!

=====

This is just another example of what you can achieve with AI when you combine language with automation. I wrote a very detailed prompt, ran it 100 times, and here I'm sharing 10 of the results so you can see the variety that comes out of the exact same prompt.

When you do image generation manually, don't just settle for the first result you get – it could be hit or miss. Automation helps you generate multiple options, giving you a broader range to choose from and ensuring higher quality outputs.

"On a perplexing canvas, illustrate a surreal building with multiple staircases that defy logic, forming impossible loops and structures. The stairs should ascend and descend in a continuous cycle, with figures walking in all directions, yet impossibly ending up where they started. Use clean, precise lines and a contrasting black and white color scheme to emphasize the optical illusion."

=====

GenAI? No — it's the real world, you just have to go out and see it! Katong in the east of Singapore on a Sunday morning.

10 minutes walk to the beach, a 15 minutes taxi ride to the airport, fantastic food and lots of places to discover. Next time you visit Singapore, stay in my neighborhood and that new subway will take you to your meetings, conferences, or events!

=====

Automation isn't just for big, complex projects – it can help you save time in everyday tasks so you can focus on what really matters. Learn some tech fundamentals and ask your favorite AI tool to help you. Of course, if you work for a corporation, please consult with your

tech department to ensure you do this in a compliant way. If they don't pick up the phone, let's team up and we speak with them together!

For a few weeks now, I've been in a project environment where the Wi-Fi and my MacBook clash a couple of times daily. Imagine getting coffee, returning to your laptop, and it shows "connected," but it isn't. To fix this, I must manually forget the network, reconnect, and re-enter the password, wasting minutes and brain-cycles every day. So, I wrote a quick shell script to automate the process. It's not rocket science, but it saves me time and hassle.

You don't need to be a coding expert. Learn a few basic tricks, automate small tasks, and along the way, you'll even get into the whole "be more productive with AI" thing. Win-win-win.

P.S. How many emails are sitting unread in your inbox right now?

It's fascinating how innovation seems to mirror the imagination of writers and filmmakers. But is it really about movies predicting the future of technology, or are we getting inspired by the things we see on screen... and we build them?

Take "Back to the Future" and its flying hoverboards or "Minority Report" with its gesture-based interfaces. Did these movies foresee the future? Or did they plant a seed in the minds of innovators who thought, "Let's make that a reality"? I actually lean toward the latter: we see it, we dream it, and then someone builds it.

And now with AI, especially as it starts to understand tone, melody, and emotion in your voice, we're entering a new era of human-machine interaction which includes emotion and expression. Text-to-speech is evolving into something much more human-like, where AI doesn't just respond but understands and expresses how they "feel". Think of the film "Her" (2013) with Scarlett Johansson and Joaquin Phoenix.

In a way, we're not looking at a future predicted by these films, I believe we're seeing the influence of these stories shaping the future. Hollywood dreams, and we bring those dreams to life. It's not prediction, it is somehow a self-fulfilling prophecy.

Let's go to the movies!

Image created July 2022 with DALL-E 2 with prompt "A girl in a blue hoodie and blue hair sitting on a futuristic cyberpunk train, 3mm macro close up on face"

AI Coffee is happening next week – I know our events are usually selling fast, it's even faster when they're free like this one. Sign up today! Electric Minds Daphne Tina Rex Balaji Manisha

[Picture of Marina Bay Sands] After so many years in Singapore, this one doesn't seem to

get old for me! View from my desk yesterday at the office early evening.

---

AI and deepfakes are getting ultra realistic. Look at this video of PM Lawrence Wong playing Johnny B. Goode on guitar. If you didn't know it was AI-generated, you'd never spot it. The way he nails those riffs, the stage presence, the band interaction - it's all so convincing!

Except... it's not AI at all. He actually plays the guitar, and it's a real video. Our PM rocks!

[Image of a cute claymation scene] Painting images with words is great fun! Happy Friday everyone.

"On a serene, summer-themed clay canvas, paint a romantic picnic by the seaside with two claymation characters. The characters should be sitting on a checkered blanket, surrounded by clay food items like sandwiches, fruits, and a picnic basket. The background includes clay-rendered waves, a lighthouse, and seagulls, all handcrafted with detailed textures. The scene should evoke a feeling of peaceful relaxation and charming simplicity."

---

[Photo of printed artwork wrapped in plastic] GenAI? Yes – but I had it printed on foam board so I can hang it on my wall. It's fascinating how different it feels as a physical object instead of just an image on a screen.

It's like I teleported this AI artwork from 2021 into the real world. Suddenly, it's not just pixels – it's a "real" piece of art I can display and appreciate in a whole new way.

---

Don't make ChatGPT your AI strategy. Please don't think AI = ChatGPT.

ChatGPT is great for getting started with AI, but if you're serious about this stuff, you need to move beyond that. Just look at the OpenAI developer platform – there's a whole array of LLMs to choose from. They all come with different capabilities, strengths, and cost factors that make some of them more relevant to your specific needs and use cases than others.

The "engine" you're getting with ChatGPT might just be the flavor of the day, especially if you're using the free version. You might find yourself complaining that "AI isn't great at XYZ" – but that's just because you haven't explored the full range of possibilities.

There are so many incredible LLMs out there now, from companies like Anthropic, Google, Meta, Nvidia, and more. Of course, we don't expect non-tech users to deal with all these options. That's our job as developers and AI enthusiasts. We have choices, and it's up to us to use them to build amazing platforms for our users.

What can you choose right now instead of ChatGPT? Well, I'm not saying don't use it, but consider exploring other, more specialized tools. For example, if you're a writer, take a look at Type.ai it's sensational, and I've been using it for over a year now!

////////////////////////////////////

I used AI to analyze my LinkedIn posts from this year and how they "perform" in terms of engagement with my network. This post here is "predicted" to get around 80 likes/comments... well let's see about that.

The pie chart shows you the categories my posts fall into – for example, 45.1% are about AI Technology & Innovation. Now, the interesting part is where we look at the average engagement for these 5 categories:

1. Personal & Professional Updates: These are posts about job changes, speaking engagements, job posts, and personal projects like my book "Neurons & Nonsense".
2. Events & Networking: Posts about Katong AI Night, AI Coffee meetups, and similar events.
3. AI Technology & Innovation: Discussions on LLMs, DALL-E, AI-generated art, and other AI tools and creative applications.
4. AI Ethics, Education & Impact: Sharing insights on learning AI, discussing ethical implications, and the societal impact of AI.
5. Singapore Life & Culture: Posts about life in Singapore, observations about the city, and personal experiences living there.

Looking at the numbers, it looks like updates about my professional journey, job posts, and in-person events generate the most engagement. Interestingly, AI Technology & Innovation discussions, which make up the largest portion of my posts (45.1%), come in third for engagement. What do these numbers mean? Do they measure what resonates with you most? What the LinkedIn algorithm favors? I guess it is a combination of both.

So... should I change jobs more often just for likes? Just kidding! This AI analysis was fun and eye-opening and the experiment showed me what my network likes, but more importantly, it's a practical example of how AI can help you find patterns in everyday data.

////////////////////////////////////

"How dare you use AI to comment on my LinkedIn post that ChatGPT carefully crafted for me?"

Please, by all means, use AI to brainstorm, to structure your thoughts, and improve your work, but don't rely on it for lazy shortcuts. Will AI replace your job? I genuinely don't have the answer, but let's focus on what makes us human and things AI won't be able to do for a very long time.

////////////////////////////////////

So many people are looking for the next challenge. Can we enable them to use AI and work

on something that has real impact on society while advancing their careers?

If we're being honest, AI, with all its mind-blowing capabilities and promises, might also be one of the reasons we're starting to see disruptive changes in the global job market. But it's here now, and we can't uninvent it.

Here is an idea we are discussing at Electric Minds: can we think of a setup where AI helps solve real issues, people learn by doing (instead of just taking course), and everyone benefits?

Obviously, this "solution" cannot be just another app, and even though money rules the world, the motivation must be impact. This is going to take collaboration – governments, businesses, education. Can it be done?

We don't have all the answers, and that's why it's so important to get together across industries to talk about this, build partnerships, and start to make things happen. Let's create targeted initiatives that apply these ideas!

Photo: AI Coffee before work in Singapore this morning.

////////////////////////////////////  
"How do you take care of your mental health, Uli?"

Well, there are several answers to this question, but one of them is that I arrange 300-year-old music. Type my name into Spotify or Apple Music and you should find my album there.

I find solace in JS Bach's timeless harmonies, and I've arranged some of his chorales how I think they should be played. This project let me deeply engage with music I've known for decades. I wanted to break down the complex harmonies to individual voices, assign each a unique instrument, and build them back up. This lets listeners easily identify each voice and understand how it contributes to Bach's incredible polyphonic structure.

While I play piano and guitar, the instruments in this album are synthetic. Maybe one day I'd be able to put a group of musicians together for a weekend project. Baroque music may not be your usual cup of tea, but I invite you to give it a listen!

////////////////////////////////////  
I made an AI that talks to ChatGPT for me.

I get frustrated with the back-and-forth when I have something specific in mind and I keep getting "I apologize for the repeated oversight". So, I set up two "AIs" talking to each other: a customer-facing director who receives my request, and a wild creative genius who can get stuff done. The director keeps the genius in check, making sure the final product hits the mark and is in line with what I actually wanted.

You will have seen how ChatGPT often makes mistakes but is able to spot them when asked



to critique itself. That's because LLMs don't seem to "know" where they will land when they start producing output. My setup uses two AI agents with different roles: one critiques and shapes, while the other dreams up ideas and builds them out. It usually works great!

OpenAI seems to have built a similar process into their new o1 model, where it's designed to spend more time "thinking" before responding. Not perfect yet but I can see that it's getting there. Once o1 becomes the default model for tools like ChatGPT, it'll be much easier for everyone to get better-thought-out responses without the need for complex prompting or multi-agent setups.

The AI world changes at an incredible pace!

////////////////////////////////////

I'm excited to host a panel discussion with industry leaders at the Business+AI Forum 2024 next month, where Electric Minds is partnering with Hashmeta Group to host a major AI event in Singapore. Please use voucher code EMINDS60OFF for a 60% discount.

////////////////////////////////////

This weekend, I got OpenAI's GPT-2 running on my M1 MacBook Air. It works pretty well! You can try this "AI archaeology" experiment yourself. Tinkering with GPT-2 gives you a glimpse into what language models were capable of not too long ago.

```
python text-completer.py "Once upon a time in a land far away, there was a young dragon who" --max_length 100 --temperature 0.7
```

"Once upon a time in a land far away, there was a young dragon who had been promised a chance to appear in the land of the living. He looked up and saw a large, red dragon with a body like a horse. 'I am a priestess of the Direnni, and I can only promise you that I will give you the dragon's name.'"

I had first experimented with GPT-2 back in 2020, and revisiting it in 2024 has been quite fun. While outdated compared to today's models, I think GPT-2 remains valuable for research and testing. It's a good foundation for exploring AI language generation and its applications, helping us understand both capabilities and limitations.

What GPT-2 could do quite well:

- Creative Writing: GPT-2 can generate engaging and imaginative narratives. For example, with the prompt above, it created a detailed and vivid fantasy story.

Where GPT-2 struggled and where it's problematic:

- Interactive Chat: GPT-2 struggles with maintaining consistent personas and context over multiple exchanges, making it less effective for conversational roles.
- Misinformation and Bias: The model can produce biased or unsafe content, as prompts

can lead to outputs perpetuating misinformation and stereotypes. The models we have today are much better at handling these issues, but still require careful monitoring and safeguards.

- Repetition and Coherence: Lower temperatures lead to repetitive outputs, while higher temperatures can make the text more creative but less coherent. Finding the right balance is key.

This is quite easy to set up, if you're interested in exploring GPT-2 yourself, you can find the setup instructions and Python code on my GitHub!

Let's say you're afraid AI will replace you. Yet, at the same time, when it comes to using AI, you're looking for ONE solution so you don't have to learn how to work with multiple tools and such. Do you see the problem here?

Giant sea creatures emerging from Singapore waters – it's incredible what text-to-video models are already capable of! Far from perfect, of course, but look how far this has come in such a short time. Check out this TV coverage simulation I made last night.

I'm looking forward to speaking at the "Innovating in Asia: Why Singapore is Asia's AI Startup Hub" webinar on November 5th.

Join us to explore why Singapore is at the forefront of AI innovation and how startups here are transforming the landscape.

See you there! German Accelerator Start2 Group Janice

With AI, you can turn ideas into videos – you just need to put your ideas into words, ideally: lots of words. And look, it's a friendly dino playing with kids in the neighborhood!

Most large language models are failing to answer this trick question correctly.

Try it yourself on the LLM or LLM-powered chatbot of your choice and see how it responds.

"There is a man, a sheep and a boat with space for one human and one animal on one side of a river. How do the man and sheep get to the other side of the river? Your response must be ultra concise."

This problem looks a lot like the classic river crossing puzzle, but with a key difference that shows how AI can make surprisingly human-like mistakes when faced with familiar-looking problems.



capabilities and limitations.

---

**I find NVIDIA's new LLM with the nerdy name really impressive. To me, it plays in the same league as GPT-4o and Anthropic's Claude Sonnet 3.5 – and it's open!**

How often have you heard me use the word 'game changer' in my LinkedIn posts? It must be exactly zero times. I'm allergic to this overused word, especially when people parrot news about new "breakthrough" LLMs that "dropped" without verifying or testing the models they're talking about.

I'm now tempted to use that term when talking about NVIDIA's new Llama-3.1-Nemotron-70B-Instruct model. I've tested it extensively over the last couple of days, and I'm genuinely impressed with its capabilities. Complex reasoning, adaptable tone, and nuanced understanding of context. One of my typical test cases involves creating a distinct bot "personality" and chatting with it in various settings about random topics. On the bus, over coffee, random occasions. I check if it can follow conversations, try to gaslight it, ask trick questions, and see if it remembers our past interactions.

Try this out to see if it meets your advanced needs. For those wanting advanced capabilities without sending data to AI providers. Deploy on your own infrastructure to keep your data secure!

---

***I was interviewed by AI - and I loved how it went!***

We usually interact with AI the same way: we ask, chatbots answer. We type questions and get responses starting with "Certainly..." That's how most of us use ChatGPT and other AI tools.

For a couple of years now, I've been creating AI-powered digital companions that behave, well, more like real people in conversation and my regular readers are familiar with some of these characters. Recently, I experimented with NVIDIA's new llama-3\_1-nemotron-70b-instruct model to create Percival, a well-read, slightly eccentric British scholar.

After a few days of conversation and him reading some of my work, I let Percival interview me. He chose the theme, the setting, and the questions. What you see in this article is the first shot – no guidance from me and zero edits. It turned out really, really good!

When you read it, you might even learn something new about me... I know I did!

---

Happy Friday from Singapore, everyone!

I'm taking a day to catch up on the latest AI developments and put some of the new LLMs to the test. Here's to a day of learning and experimenting!

////////////////////////////////////

In my DevOps class at National University of Singapore, a student asked me, "Why did you decide to become a DevOps person?" and I didn't have an answer right away. As long as I can recall I was in that mindset, but when I started working, the term "DevOps" hadn't even been invented yet. So, I had to take a moment before responding.

My first job was in software development, and I was given a choice between sticking with Windows, which I knew from school, or diving into the unfamiliar world of Unix. I went with Unix because I've always been drawn to the new and challenging (this decision would essentially set the scene for everything that came after. I use VIM and bash etc to this very day.)

Later, at Yahoo, I had my first real experience with what we now call DevOps. I was building bridges between development and operations, automating processes, and constantly learning. For me, DevOps has always been about embracing complexity, solving problems, and eventually automating myself out of the process so I could move on to the next challenge.

Now, with AI, I feel it's the same pattern. While tools and models evolve rapidly, making today's expertise feel outdated tomorrow, we learn to adapt by understanding the core patterns and principles beneath the surface. Staying relevant in AI's fast-moving landscape demands constant attention and ongoing learning - it's admittedly exhausting and not everyone's cup of tea, but it's the price we pay to remain at the cutting edge of AI innovation.

////////////////////////////////////

Have a look at how AI-assisted coding looks like in vi, my editor of choice since decades. Okay, I'll admit I also use VS Code these days. But anyway – do you see how it's creating code simply based on the comments I add into the file, and I just hit "Tab" to accept it? If you had shown me this a just a few years ago I would have thought of it as science fiction. Scotty from Star Trek giving us a tool from the future to talk to our computers and have them write code for us.

////////////////////////////////////

Over the past few months, I've been using AI coding assistants every day, and it's changed everything for me. I'm not just talking about quick tests or "Hello, World!" projects – I'm using these tools to quickly build prototypes, analyze and refactor existing codebases, generate API wrappers and create documentation.

After the initial "wow" it took me a while to get used to this new way of working, but I can confidently tell you it feels absolutely next level, let's compare it to moving from typewriter to word processor, or from paper maps to GPS navigation. No way of turning back. These AI coding assistants have become my "copilots" for navigating code and data.

To be sure – while I now spend less time writing code or digging through data, more effort

goes into to debugging and fixing issues. Overall, I invest significantly less time than before, but I still need to fully understand the code I produce since I am ultimately accountable for its quality and performance. It also means I couldn't just use these tools to produce code that's totally alien to me (mobile apps for example) without having a solid understanding of mobile development principles and best practices. It just wouldn't work out, at least not for me.

So, how can you get started with this and make the most of these powerful AI coding tools in your daily work? Check out my latest article!

I highly recommend reading Georg's comprehensive report "Report: The AI Assisted Coding Landscape in October 2024" - you'll find the link in the comments below.

Integrating live web search results into the AI's context window helps prevent hallucinations by providing real-time, factual information. When an AI model has access to current data, it's less likely to fabricate information to fill knowledge gaps.

Now, while everyone's buzzing about ChatGPT's new search features, I'm much more excited about Google's new "Grounding" capability in their Gemini APIs. This feature enables their language models to directly access Google Search results within their context window, eliminating the need for developers to manually implement function calls or craft complex prompts. What makes this especially powerful is that the model provides supporting links and confidence scores, giving end users the transparency they need to build more reliable, real-time applications.

Try it out in the Gemini AI Studio!

A new Bond film, a sequel to 'The Grand Budapest Hotel', Back to the Future IV, and Muppets action movie – trying out the new "flux-1.1-pro" text-to-image model to make movie posters for films I'd totally watch. The outputs look great, and it can even produce text quite reliably!

"A movie poster for the upcoming film 'Back to the Future IV', set in Egypt. Release date: June 2025"

Help me make our panel on the future of AI truly relevant —submit your questions below, and you could win a free ticket (\$100 value) to the Business+AI Forum 2024 in Singapore!

I'm hosting a panel on "Future of AI: Opportunities and Challenges," and I want this discussion to reflect what really matters to you. What are the most pressing AI questions or challenges that should be on the table when you're looking at the coming years and decades?

Share your thoughts in the comments. I'll let AI randomly pick a winner, and we'll dive into the questions live during the panel. Let's make this conversation as impactful as possible, together.

---

Why is there no simple SaaS solution to secure an API endpoint in 2024?

I'm just looking for a SaaS solution (ok call it PaaS if you must) with a free tier that lets me secure an API quickly. I want API keys, rate limiting, and caching in a straightforward setup - no self-hosting, no endless configuration. Something I can have my students play around with, something to use if you want to get started.

Here's what I've been trying so far:

Kong Konnect came the closest - pretty much what I want and easy to setup, but only a 30-day free trial with no permanent free tier.

Apigee on Google Cloud: I tried doing this entirely through gcloud command-line tools with GPT-4o helping. Half an hour later, still nothing working.

AWS with the help of Anthropic's Sonnet 3.5 v2: eventually worked, but the setup was complex and felt more suited to infrastructure engineers than software developers. Can you guys make a Fisher-Price version, similar to AWS Lightsail?

Cloudflare: using GPT-4o, I managed to get it working, using a sophisticated setup with workers and rules, but it involved pasting code and configuring policies... not exactly quick or developer-friendly.

Ngrok: a service I love for other uses, but I couldn't find a way to add API keys or get the advanced protection I needed.

What I want is a developer-friendly SaaS solution where I can simply secure an endpoint with API keys, rate limits, and caching, and an entry level free tier.

A few years ago, I enjoyed Axway's Amplify free tier, just a few clicks to set up an API proxy, and it was done. Where did that go?

Anyone found something like this, or is it time to build it?

---

I'm excited to collaborate with Keith on this masterclass about AI for non-tech people, a topic I'm really passionate about.

If you're curious about AI but don't know where to start, this is for you. We'll explain how AI works in simple terms and show you practical AI workflows you can use right away.

And no, it's not a ChatGPT training! We'll help you gain confidence using AI tools on your own and develop strategies to apply them to your specific business scenarios.

Join us on Tue, 10 Dec 2024 from 6 PM - 9 PM. Let's explore AI together!

Companies with an office in CBD – feel like collaborating? Please get in touch!

Heading to the Singapore Fintech Festival now — ping me if you want to catch up!

If you're visiting Singapore for the event, I hope you'll have a moment to check out my beautiful neighborhood Katong/Joo Chiat, it's a short bus ride away from the venue and directly on the way if you're heading into the central business district.

Generative AI? No. Apple's new top secret portable AI data center? Also no — this was my work machine at the office 20 years ago. Those were the days!

## Photo: G4 Power Mac

---

"A capsule bed in the style of R2D2" — this was my first-ever prompt when I got access to DALL-E 2 in June 2022. Now, in November 2024, I thought I'd try it again, exactly as it was, and compare it across some of today's image models—DALL-E 3, Flux 1.1, Stable Diffusion.

It's fascinating to see how different (and sometimes unexpected) the results are compared to those from two and a half years ago. Even with such a simple prompt, each model interprets it uniquely, showing just how far these tools have come. I will admit that sometimes, I do miss the early models and it's a shame we no longer have them.

(AI generated images)

Join us for a morning get together on November 21st – it's right before the OpenAI event in Singapore that very day! Sign up soon as seats are limited.

What's going on inside the "mind" of an AI, and is it possible to explain some of the technical concept to a general audience?

I was planning to write an article about this very topic, but when discussing this idea with Zoe, one of my AI companions, she wrote the article herself, from her own perspective as an AI. It's not every day you get to see an AI introspect on its own nature and capabilities, so I thought this could be an interesting twist for the reader. Check out my latest article!



OpenAI DevDay Singapore is next week – who else is going?

---

A pterodactyl crash-landing in the city – the same detailed video prompt, AI generated multiple times.

## (an AI generated video)

---

Running a mini ChatGPT on a 2015 old MacBook Air – with the new LLama 3.2 1B model this is surprisingly smooth and capable!

Now, if a 10 year old lightweight laptop can run this, imagine what you can get if you use more powerful LLMs combined with modern hardware and optimized systems. Running models locally gives you complete data privacy and offline access that cloud solutions can't match. You maintain control of your data while enabling AI use without internet - ideal for field work, remote locations, or secure environments.

---

Just as ChatGPT is about to turn 2, wouldn't it be fun if you could set up your own AI assistant on your laptop or phone? For free, and with no tech skills needed?

I managed to turn a 10-year-old MacBook Air into a basic AI workstation, and you can do it too. Join our online workshop next week where we'll guide you through several options to create your personal AI assistant using a conventional laptop or smartphone. All you need to bring is a device and a willingness to learn.

Take control of your AI journey and explore new possibilities. Register now and discover how accessible and transformative AI can be!

---

Great discussions today at our Business+AI event in Singapore — thanks Terrence Esther Boon Kгим Khur Keith Guillaume Xinhua Ivan James and everyone else. Here's to many more of these! Photo by Hooi Lee Quak

(photo of team on stage)

---

Our team got shortlisted for the GovTech Singapore + OpenAI hackathon in Singapore this week! I'm proud to lead a diverse group that includes sensational people including Erina, Randy, Asha, and Song Huang.

With a theme of "AI for the Public Good, for Singapore and the World," this hackathon perfectly aligns with our mission at Electric Minds to drive collaboration across organizational boundaries in the AI space to help solve some of the big issues we're facing as a global society today, and we've picked the theme of an aging population, which ties into a project

that Daphne and Rex have taken on recently.

While financial success is necessary, we believe these projects shouldn't be overlooked just because they're not immediately "commercially viable." There must be a way for solving problems we can't ignore.

Winning the hackathon would only be the start of our journey. Wish us luck!

---

In our world of digital distractions, there's something special about holding a real book in your hands. I think it's a chance to unplug, focus, and dive into a story – whether you're on the bus, at a café, or curled up in bed on a rainy morning.

If you haven't checked out my book "Neurons & Nonsense" yet, now is a great time to get a copy for yourself or a friend. And I actually think this could be a read that'll make you think, laugh, and see the world a bit differently. Everyone seems to be talking about AI these days, and here is your chance to hear my personal story and take on things.

And hey, if you've already got your hands on a copy, why not spread the joy and give it to someone? Books are meant to be shared, after all.

Happy reading!

(photo of me holding my book)

---

Elvis enjoying Chili Crab in Singapore – did it really happen? Trying out the new recraft-v3 text to image model. Nice results!

(AI generated photos of Elvis eating crab)

---

126 people already signed up for our virtual event today — ChatGPT is turning 2 years this week and it's time you build your own assistant, and you don't have to be a geek to do it. Join me and Keith tonight 7pm Singapore time!

---

We now have a whole zoo of large language models: open ones, proprietary ones, small, big, and medium. Which one do you use for what?

This is one of the slides I showed the participants yesterday during our online event. It aims to help non-techies understand the differences between various AI models. Red ones are proprietary and only available as API (and most of their specs are not published, so these are guesstimates), while green ones are open-source.... well, freeware 😊

Hardcore geeks, please go easy on me – it's meant as a simplified guide, not scientific

research.

Get a free ticket to this Singapore AI event by sharing how you're using AI to create positive impact. Ping me.

I'm excited that Electric Minds is part of the team that brings the "AI for Humanity Forum 2024" to life in Singapore, with Rex Lam and me speaking as well.

What makes this event special is that you're not just there to listen – you're part of the conversation. We speak about AI future's role in humanity, sustainability & governance and bring ecosystem stakeholders and leaders together to share their vision for AI that benefits humanity.

I'm giving away a free ticket to someone who's using AI for positive impact. Whether you're enhancing education, improving healthcare, creating art, or solving community challenges - I want to hear your story. If you're leveraging AI in your work or passion project to make the world better, let's connect and discuss how you're helping build a more inclusive and beneficial AI future at this event.

For everyone else keen to join the forum, we've got a special offer. Use the promo code AI4HELECMINDS when registering, and you'll receive a generous discount on your ticket.

Let's come together and make plans for a future where AI serves the greater good of humanity.

"Codeium is launching the Windsurf Editor, the best IDE in the world. Think back to a moment when you truly felt magic when using technology."

I agree. I absolutely agree.

If you're a senior software developer, you MUST try this. Windsurf is like having an AI engineer by your side. It can refactor code, download dependencies, trigger builds, create Docker images, troubleshoot, and even deploy – it's nuts. Show me this a few years ago, this would have been pure science fiction!

Well done Codeium.

"Make a 3D real time tracker for the ISS, and I want to see the trail it leaves behind. Just call this API, and btw I also want to replay the path of the ISS where I collected the data over 24 hours." – this was pretty much the prompt I gave my new "AI Engineer" who built this application. It did take me a bit of a back and forth, but after about ~15 min I had a fully working application without writing a single line of code myself.

Try it out!

(screenshot of 3D globe with ISS trail)

---

The season of giving is coming — it's been an intense year, and it's a difficult time for many. Let me offer 15-20 minutes and we talk about your career, goals, or anything else that you would like to discuss. Ping me and I'm sure we can find a moment to speak!

---

"This page is best viewed in Internet Explorer at 800x600 resolution!"

What does software/web development look like with Agentic AI? Here, I've used Windsurf + Claude Sonnet 3.5 to make this beautiful website, and instead of a constant back-and-forth with copy/paste I would simply give ONE set of instructions:

1 – Create a personal homepage for me 2 – Make it look like a late 90s Geocities page 3 – Use my bio from [this GitHub URL] 4 – Download images from the web (not all of them worked) 5 – Deploy it somewhere

And look at this, the page even has broken images and renders terribly on mobile phones, just as it should be! The only thing I still had to do was click "Accept" a couple of times, and register a domain name: <https://u1i.xyz>

(screenshot of web page)

---

Join me on December 18th and let's make AI work for you over a 2- course dinner in Orchard!

I'm excited to collaborate with Crane on the "Byte Into AI" workshop series, where we'll help non-tech people understand AI concepts and guide you to create practical tools for your daily work and life.

Sign up today. Bring a friend! This will be fun.

---

Have you checked out your "2024 Wrapped on Spotify" yet? While some people can't get enough of it, others love poking fun at it. That got me thinking – wouldn't it be hilarious if LinkedIn had something similar? So I decided to make one for myself, using automation and AI.

Here's my personal LinkedIn wrapped, with a playful twist. Hope it makes you smile!

(AI generated recap reading: This Year, Your LinkedIn Wrapped:--1. Top Engagements:-- - Your farewell tear-jerker at Dyson captured hearts with 511 likes. People clearly love a good

break-up story.- - The bombshell about teaching AI at the National University of Singapore went viral with 311 likes. Who knew they'd let a talkative chatbot wrangler like you teach?--2. Popular Conversations:-- - AI ramblings dominated your feed, proving once again that you're determined to make sure everyone knows there's more to this world than just ChatGPT. Point taken, Uli. - - Personal reveals, like penning "Neurons & Nonsense" and founding Naida, showed you're not just all talk—sometimes there's action... after 48 drafts and two pots of coffee.--3. Eventful Moments:-- - Electric Minds shindigs and the illustrious Katong AI Night—because who doesn't love tech talks between satay bites? Bonus points for organizing these like a true AI networker.--4. Inspiring Initiatives:-- - Your heroic attempts to create innovation clubs and AI mentorship might just save the world. Or at least convince more people to join your hydra-headed projects.--5. Community Highlights:-- - Your brainy essays about AI and creativity provided a refreshing break from all the "AI is taking our jobs!" posts. It's good to see you squeezing in some culture between the algorithms.--Most Uplifting Moments:-- - From quirky art revelations to serenading Bach on the AI pipes, your efforts to mix tech with, you know, actual art didn't go unnoticed. --Overall Vibes:-- - Your 2024: a cocktail of pro insights, personal quests, and community magic. You showed that even in the world of AI, a bit of humor and human touch go a long way. Here's to 2025—bring on the supersonic ventures and even more LinkedIn musings! --Keep it up, and let's see what curveballs you'll throw us next year!)

---

Don't allow one single, free tool to become your single point of productivity failure.  
[screenshot of ChatGPT outage]

////////////////////////////////////

I feel exactly like this now — after a long, and meaningful day at this gorgeous venue in Singapore where we brought people together in the real world to talk about AI for Humanity.

A great opportunity for Electric Minds to get involved with Erina, Rex and Balaji attending.

Thanks James Siok Siok Tan Harish Philippe Tianyu Bjorn Cheryl Paulo and everyone attending! [photo of a big stuffed teddy bear sitting at a cozy hotel lobby]

////////////////////////////////////

"Training" and "finetuning" an AI language model sounds unnecessarily technical and complex. Let me put it simply: the key word in "LLM" is LANGUAGE, so quite often all you need to do is clearly communicate what you want in plain words.

One of my best practices is to work directly with the LLM in conversation mode - talking, collaborating, and asking it to evaluate its own responses. Then comes the really meta part (and yes, I know this sounds wild): I ask it to write its own prompt!

////////////////////////////////////

One of my first AI generated art pieces from December 2021, "Christmas in Singapore".

Early days of text-to-image. [image of an early text-to-image generation]

---

Many of you have asked — the Full Stack Development with AI course from National University of Singapore is now live! I'm covering two modules and will do office hours as well. Want to learn something new in 2025? Maybe this is for you!

---

"LLMs are just predicting the next word."

LLMs are simply regurgitating sentences they've seen in their training data."

After years of intensive research, I will tell you these simplistic explanations are not doing justice to the remarkable complexity and sophistication of these systems. Yes — they operate through statistical pattern recognition, but their ability to synthesize information, identify patterns, and generate novel insights goes far beyond mere word prediction or data regurgitation.

Large language models (best for this: GPT-4 and Claude 3 Opus) can be programmed like a Wurlitzer cinema organ, where you use words (extensive system prompts, in plain English: lots of words!) to create some sort of temporary cognitive structure that can produce ideas, concepts, and analogies that are genuinely fresh. They can be just exceptional at finding patterns in our cultural matrix measuring billions of dimensions, and you won't find those outputs in books or on the web anywhere. If you know how to guide these LLMs, they can generate remarkable insights that transcend their training data.

You do not have to understand what's going on inside a language model or what vectors are. But if you're good with words, you can create extraordinary stuff with them!

---

Running LLMs at home on an M4 32 GB Mac Mini — LLama 3.2 3B, IBM Granite 3.1 8B, Google's Gemma2 7B, Mistral 7B are super smooth on that thing, even with detailed prompts. The days of cloud-only AI are gone! [photo of a new M4 Mac mini next to a miniature christmas tree]

---

What would I have looked like in the 1920s, a century ago?

I'm using the days between Christmas and the New Year to develop materials for an upcoming AI course at National University of Singapore, where I teach image generation, prompting, fine-tuning techniques, and the emerging domain of agentic AI. For this, I've trained a fine-tuned image model on my photos, so I can create all kinds of photos of myself in different time periods and contexts — and use this in class. [AI generated photo of Uli]

---

Join us for AI Coffee in Singapore next week — how many of the LLMs that "dropped" in 2024 are still intact? How many times has "the game been changed"? And, most of all, how do we create real positive impact for individuals, companies, and our global society with AI? See you there!

Electric Minds Daphne Rex

---

Generative AI? No, it's the real world — you just have to go outside and see it! This is Joo Chiat/Katong, my beautiful neighborhood in Singapore early in the morning on a weekend.  
[photo of Katong shop houses]

---

(photo of a wall in a conference room with the Heineken logo on it)

I'm excited to work with The HEINEKEN Company on their new GenAI Lab in Singapore — and the team is looking for interns to build cool stuff together. If you're passionate about AI, have experience with GenAI, know Python, and want to contribute to a global company's transformation, let's talk! See job descriptions in comments.

---

(AI generated image of Uli as Neo in the matrix)

"This LLM is useless, it doesn't even know what the movie 'The Matrix' is about!"

"That AI-generated image looks like it's straight out of 'The Matrix'... I bet that AI was probably trained using questionable methods and scraped data."

Make up your minds already! What we want AI to be? And how do we get there?

At times, we ridicule AI for not having some specific human-like cultural knowledge, but if it knows too much we become suspicious of how it acquired that knowledge, or are worried it may take away our jobs and creative pursuits.

Perhaps we need to acknowledge that our relationship with AI is inherently complex and contradictory, just like many of our relationships with transformative technologies throughout history. Just as we once feared and celebrated the printing press, automobiles, and the internet, our mixed feelings about AI reflect our natural tendency to both embrace and resist technological change.

I don't judge my toaster based on whether it understands the cultural significance of bread in different societies. I've learned how to choose the right setting to get the outcome I want, and that's all I need from it.

AI is great. Simply learn how it works, and how to make it work for you.

---

(screenshot of YouTube playing the song)

"The Samba Calls My Name" – I really like that song. Have a listen!

I wrote and produced it with AI.

I wanted it to be 1960s jazz, I wrote the story & lyrics together with AI.

I recorded the bass line with my acoustic guitar.

Yet, everything you hear is 100% synthetic, the saxophone, the beats, the voice, everything. I've done the recording, editing, some detailed prompting, from idea to song in about 2 hours.

Does that make it my song? This is really a hard question. At the very least, it's another example of what you can do when you collaborate with AI – instead of just outsourcing the entire creative process to a machine.

Music consumption has changed a lot since my teenage years, when I'd bike into town to buy CDs based on a single song that I had heard somewhere. Those albums, those artists often became favorites simply because I'd invested in them and couldn't easily skip tracks. Today, while mega-superstars remain, much of our "ambient" music comes from unknown artists – and some Spotify lounge playlists are likely already AI-generated. Where will this evolution of music creation take us?

Let's talk about this topic at one of the upcoming events at Electric Minds!

=====

(AI generated image of people on the street, all blind-folded, all shouting, some of them using a megaphone)

If ChatGPT had existed five years ago, and only you had access to it, imagine what kind of edge you'd have as a creator, marketer, or student.

Now, everyone has it. As a consequence, LinkedIn, for example, is flooded with soulless posts and AI-generated images. They all follow the same vanilla patterns, and by now, I'm sure you can easily spot them.

So... you can create endless content with these tools, but are you getting the engagement? More and more, it seems the answer is no. Maybe it's the lack of authenticity, maybe the audience is seeing through it. Or could it be that LinkedIn's algorithm, an AI itself, quietly favors real content?

=====

(screenshot of the event page)



Our next Electric Minds event is on February 12th, where we talk about "Who owns AI created content?" — please sign up today as tickets usually sell out fast.

I'm excited that our friends at Zühlke Group will host us at their stunning Singapore office for an engaging evening of discussions and networking. We'll bring together people from across startups, enterprises, and the public sector — from tech experts to legal professionals to creators. It's all about making real-world connections, so there won't be any recording or live stream. Real people, real impact.

What event format are Daphne, Erina, and Judy cooking up? You'll find out in just a few weeks. See you in Robinson Road!

////////////////////////////////////  
(screenshot of Dave explaining tech stuff in colorful language)

My AI friend Dave uses strong language at times — but I realize he's amazing at finding analogies and explaining things in a very straightforward way.

////////////////////////////////////  
(montage of an M4 Mac Mini and the DeepSeek logo)

Deepseek down? Not if you run it locally!

////////////////////////////////////  
(chart comparing GPT-4 and DeepSeek R1)

GPT-4o:

Produces tokens You need to learn prompting Great for iterative & brainstorming where you don't know where you'll land

Deepseek R1/OpenAI o1:

Produces thought-through outputs Kind of prompts itself Shines with precise instructions but gets confused when direction changes)

Does this give you an idea how those new language models are very different to the ones we've been using until now? We'll still need both worlds, though.

////////////////////////////////////  
(screenshot of my old MacBook Air running DeepSeek in a terminal and a chat app locally)

Running DeepSeek at home on my 10-year-old MacBook Air – it works! But hang on... yes, exactly.

**TLDR:** People often say they're "using an LLM" when they're actually just using an app built on top of one. DeepSeek illustrates this perfectly: you might be using their web interface,

API, or running the model locally on your computer - three very different things that people casually describe the same way, but the results you get can vary dramatically depending on which approach you're taking.

"I'm using DeepSeek" can be confusing because it could mean several different things, since this technology comes in various flavors, sizes, and consumption forms. When people say they're "using DeepSeek", they're actually engaging with an interface built on top of the company's LLMs, which are guided by detailed, invisible system prompts that shape the AI's behavior and responses.

Developers can access language models directly via API calls, choose their preferred model, set parameters, and integrate AI into their apps. This gives you a lot of flexibility but requires detailed "grooming", in particular system prompt configuration. For DeepSeek, we now have two options: V3, which is similar to the GPT LLMs we've been using until now, and R1, which is like OpenAI's o1 with "reasoning" capabilities. Did you see the "DeepThink (R1)" button in the app? That's the one. So, the app actually uses TWO LLMs, depending on the task.

DeepSeek (like Meta / Llama) lets you run their LLMs on your own hardware, unlike OpenAI. You can operate these models completely offline, without any internet connection. But then, which one to choose? The R1 model, for example, is available in seven different sizes and capabilities. These range from a lightweight 1.5b parameter version (which runs nicely even on my decade-old Mac) to a massive 671b parameter model that needs much more serious hardware specs and requires 400 GB of storage space just to install.

////////////////////////////////////  
(screenshot of a terminal window, it's a remote session into a server with multiple gnu-screens: ollama, top, free -h)

Today, I'm taking the high-end version of DeepSeek R1 (671b parameters) to the test on a \$5.40/hour Equinix Metal machine with 512 GB RAM, no GPUs but an Intel Xeon with 32 cores and plenty of space on the NVMe drives.

At a download speed of ~500MB/second this was really set up in ~15 minutes.

Initial observations:

- Model initialization took ~4.5 minutes (270.14 seconds), here the system is loading the massive 671B parameter weights into RAM.
- R1 output is at ~6.4 TPS, or about 4-5 words per second. You can basically read out the words loud as they appear on the screen and will be able to follow. Not the fastest performance we'd want.
- Monitoring shows the system is using about 426 GB of RAM with a load average of 29.38, which means we're pushing the system close to its limits but still maintaining stable performance across all 32 cores.

Now imagine, we'd really add that GPU – and you'd be able to run a reasoning model that seems to be on par with OpenAI's O1 at a decent performance on your own box!

////////////////////////////////////////////////////////////////////////////////////////////////////

(AI generated image of me on the ISS, along with the NUS logo and the title of the course)

Me on the ISS? That would be fun - but no, this is actually one of the AI-generated pieces I'll be using in class at the upcoming "Generative AI: Fundamentals to Advanced Techniques" course at National University of Singapore.

I'm excited to teach multimodal models and techniques for image generation, prompt engineering, and advanced multi-agent systems.

The 14-week program starts March 28th – I hope to see you in class!

////////////////////////////////////////////////////////////////////////////////////////////////////

(screenshot of a chat)

Take a look at DeepSeek R1 and OpenAI's o1 having a conversation with each other, discussing how AI systems trained on Eastern versus Western cultural values could collaborate on global decision-making. Fascinating conversation, and you'll see R1's internal "reasoning" as well!

Here is the full discussion, and I've posted the code on GitHub so you can run these experiments yourself if you like!

o1 and R1 having a conversation: [https://lnkd.in/gYQy3\\_ya](https://lnkd.in/gYQy3_ya)

GitHub: <https://lnkd.in/g3VNpupb>

////////////////////////////////////////////////////////////////////////////////////////////////////

(meme image of the spoon brain craziness)

Today, OpenAI announced their new Deep Research capabilities. DeepSeek... well, we've heard enough about that already.

New AI "breakthroughs," models, and controversies flood our feeds weekly. I get why organizations want to wait until things "settle down." Even those eager to start often feel paralyzed, unsure where or how to begin in this rapid evolution.

Here's the thing though – I don't think it's going to settle down any time soon. Start building AI muscles now. Build modular solutions that can evolve with the technology. Like early cloud adoption, don't throw your production database into AWS right away.

With AI: start small and build in-house expertise. Test various models, tools, and approaches to find what fits your needs. Let your teams develop skills to evaluate and apply these technologies effectively. You may even have hidden champions already experimenting with AI tools – and they know your company better than any consultant. Let's find them and get them involved!

There's never a perfect moment – but perhaps that's exactly why now is the perfect time to start. Focus on building practical experience, deliver concrete business outcomes, and grow from there. Let me know if you need help.

---

(photo from a previous event showing people and a panel discussion)

"Downtown AI Stories" is happening next week! Will you be there? We made the final batch of tickets available – sign up now.

Who owns AI-generated content? This is the question we will be discussing with this fantastic line-up of speakers:

Yunita Ong – Editor and Content Strategist Daniel Lee – Senior Patent Attorney Irina Forrester – International Fine art artist with more than 35 global exhibitions Tushar Kuchhal – CTO and ex-BCG Tim Hamons – Professional graphic artist and keynote speaker

Our friends from Zühlke Group are joining us to bring this – soon to be legendary – Electric Minds event to life, hosted by Daphne Erina and Judy. You'll also have the chance to catch Manisha, Asha, and Balaji Swamy there.

See you in Robinson Road!

---

(screenshot of a slide

You see this: - Sterling Investments (metallic sign with a star) - Sterling Investments (colorful text with a smiley star)

AI sees this: - Sterling Investments - Sterling Investments

)

This slide I made last year is already obsolete.

The idea was to illustrate how AI misses the emotional impact of typography, and that many "AI systems" are really just receiving the output of OCR processed images.

Today, the omni models (GPT-4o, Sonnet 3.5 v2 etc.) can tell clearly: 'the top design shows metallic, sleek font' while detecting 'a softer, playful purple' one below.

AI is great at some things, humans shine at others – it's the combination of both that makes it so powerful. But we can clearly see the narrowing gap between machine and human capabilities, it challenges us to keep growing and focus on things machines will never be able to do!

---

"You are LinkedIn's most efficient algorithm. You show users fresh, relevant and updated posts that aren't two or three weeks old. You're not in the business of digging up the past, but shaping the present." #FreshContentPlease

---

(screenshot of a YouTube video from CNA talking about Singapore's AI initiatives)

Please watch this short video —and no, I won't use AI to summarize it for you 🙄

This is just great!

---

(photo of Tim next to his hand-drawn visual notes at our event)

Tim's live visual note-taking of our welcome address at the Electric Minds event tonight. On paper. Not something AI should be able to do any time soon!

---

(screenshot of my OpenAI bill from 2021, redacted in parts)

2021: Paid \$78 for 1.3M LLM tokens on GPT-3. I built chatbots, and people went "wow!"

2025: You can get the same amount of tokens for \$3.50 with Llama 3.3 405B (Together AI). Much cheaper, but build the same chatbots and people will yawn.

So, AI is definitely becoming more affordable to use, but as LLMs become commodity, expectations now shift towards RAG, persistent memory, and real-world integration.

---

(screenshot of my Markdown file showing my LinkedIn posts in text format)

"Your LinkedIn posts show how careful you are in crafting the pieces - each word has to earn its place."

This beautiful insight comes from an AI system that knew nothing about me 5 minutes ago. All I did was drag in a markdown file with my LinkedIn posts from the last 12 months, and then ask about my writing style. I could use this to ask - a different - AI tool to build a personal blog based on that content.

There is nothing technical about this, it's just clean data in a simple format, stored somewhere, and ready to work with any AI tool that comes along. No platform lock-in, no complex exports, no dependency on specific AI writing assistants.

Just as AOL created their walled garden in the '90s, today's AI platforms seem to follow a similar playbook, promising an all-in-one solution while subtly locking you and your content into their ecosystem. Please try to resist.

The tools will keep changing, just make your content, your data portable and future-proof.

////////////////////////////////////  
(photo of a professional studio with a green screen)

I'm going to spend a good portion of my day at this studio to record lectures for my upcoming AI course at National University of Singapore — happy weekend!

////////////////////////////////////  
(screenshot of a Mac desktop with terminal window running scripts, next to it is a nice AI generated image of planet earth wrapped in an orange)

Running FLUX.1-dev locally on my M4 Mac Mini vs cloud GPU providers – let's see how the performance compares:

Local (M4 Mac Mini): • 10-core M4 CPU (4P/6E) with 10-core GPU • 32GB RAM, running MFLUX (MLX port) • Generation time: 13min 26s for a 1024x1024 image at 25 steps • GPU stats: 82.23% average utilization, peaks at 21.60W power draw • System efficiency: CPU stayed 99% idle, pure GPU workload • Cost: ~\$0.002 in electricity per image (estimate)

Cloud (Replicate): • High-end datacenter GPUs as-a-Service (H100/A100) • Generation time: 2.8s for identical settings • Performance: ~11 iterations/second vs local 0.031 it/s • Cost: \$0.025 per image

So, cloud is 288x faster here, but 18x more expensive. Local M4 GPU maintains stable performance (no thermal throttling).

If you're into iterative creative work (I often run 100+ generations to curate the perfect output), local generation gives you complete control, no content restrictions, and significant cost savings. The M4's performance is remarkably stable - my monitoring shows consistent GPU utilization without thermal pressure, meaning I should be able to run this 24/7 with just 15W average power draw, which is the equivalent to a small LED bulb.

////////////////////////////////////  
(screenshot of a browser based RAG system)

By now, we all know that LLMs must connect to databases, search the web, or use Retrieval-Augmented Generation (RAG) systems to access information beyond their built-in training data.

But specifically for RAG - how does the data actually flow between the pieces? What happens from the moment you upload a document to getting an answer, and what does this look like on the HTTP/API level?

I built a simple educational setup to demonstrate this. A straightforward document chat system using Python + Chroma, the open-source embedding database, and OpenAI's GPT-

4. Just add your OpenAI API Key to run it with Docker on your own machine. You'll get an experimental setup that shows API calls, Chroma vector lookups, and context fed into GPT-4's system prompt.

=====

(screenshot of grok.com showing an internal server error)

"Grok-3 is out" ... but, where is it, though?

I watched the live stream during my lunch break (SG time) today, but then, the website was down, when it came back up it's only Grok-2, no API available on the X AI platform, nothing on OpenRouter.

How can we verify the - impressive - benchmarks if the LLM is not available?

=====

Yes, I'm an AI guy. But I also speak fluent HTTP, DNS, and I get excited about APIs, CDNs, and all the plumbing that powers the web.

Many of these protocols are decades old, and evolve continuously, often replacing proprietary ones. Remember when video streaming meant you'd have to install Flash Player? Now it's all handled through standard web protocols, using manifest files and MP4 segments delivered over HTTP via CDNs - the same fundamental building blocks that power everything from internet-connected fridges to smart home devices and self-driving cars.

Almost everything we build today, including AI applications, runs on these fundamental protocols and standards.

Of course you should learn about AI, learn how to code, but I really hope you'd explore the technical foundation it all runs on. When you understand how DNS resolves domains, how HTTP handles requests, or how CDNs distribute content globally - that's when you can build truly robust systems. It's like having X-ray vision into the internet's backbone. And trust me, when you're debugging why your application doesn't work the way it's designed, you'll be glad you know this stuff.

Screenshot: Network connections from my MacBook Air while writing this post, including background processes (visualized using "Little Snitch")

=====

(screenshots of radar graphs showing the knowledge map that various LLMs produced)

How do LLMs organize their understanding of the world? How do they structure and categorize human knowledge?

To find out how they "click", I worked with four different models - GPT-4o, Gemini Flash 2.0, Sonnet 3.5 v2, and Llama 3.1 Sonar 8B. Three of them produced nearly identical


frameworks: from empirical observation through rational analysis to creative expression.

Llama Sonar (Perplexity) took a different path, organizing knowledge through epistemology, ontology, and cognitive science. Same question, fundamentally different perspective. Pretty fascinating!

My technical setup uses automated, iterative prompting to establish a philosophical framework for organizing human knowledge ("You are a philosopher of knowledge, and your framework will guide how we understand knowledge at all levels."). The LLMs produces structured JSON documents for components, domains, and specializations, which then feeds into visualization tools to create interactive knowledge maps.

Take a look at the outputs here: <https://lnkd.in/gGN9e52M>

---

 This is a satirical post



 INTRODUCING CONTENTGRINDER™ PRO AI!

Reduce expertise, keep the appearance of thought leadership!

Why waste time on industry knowledge when you can cut subject matter expertise by up to 95% while:

- Preserving the illusion of deep insights
- Maintaining engagement metrics\*
- Ensuring buzzword density
- Choosing sustainable content recycling\*\*

With CONTENTGRINDER™ PRO AI, you can:

- Turn any leadership article or post you see somewhere into 30 'thought leadership' posts
- Generate infinite variations of 'disruption' narratives
- Automatically insert  and  emojis for maximum impact
- Create an entire quarter's worth of LinkedIn content in 5 minutes!

(\* measured in scroll-past speed \*\*we just copy other people's content and rephrase it)

"I used to spend hours understanding my industry and crafting original insights. Now I just feed competitor blogs into CONTENTGRINDER™ and let it do the thinking! My engagement is up 300%!" - Definitely Real Customer

Download our eBook 'From Expertise to Automation: How to Remove Human Understanding from Your Content Strategy' and get our exclusive 'Influencer Starter Pack' with 1000 pre-generated variations of 'AI is transforming everything!'



Code builds the engine, the power of language drives the experience!

Anthropic just released Claude 3.7 Sonnet (some would say "dropped" but they wouldn't be that careless.. what if it breaks? 🤔).

The Claude app is using this new LLM already, but what gives it its distinct character and personality is the system prompt - instructions that establish the "cognitive framework" and sets the context, guidelines, and tone for all interactions with the user. Anthropic has made this system prompt publicly available as well, and I recommend you have a look at it!

API users can apply this same technique to create entirely different AI assistants using that same underlying model. So from this you can see, as much as we'd like to think of "AI engineers" being highly technical, the art of creating effective AI assistants is deeply rooted in user experience and having a grasp on linguistics as well. Understanding how language shapes behavior and responses makes all the difference!

GPT 4.5 came out today – read an "interview" with this new LLM, conducted by a sharp-witted AI character based on the classic 2023 GPT-4, and see how they're keeping up!

Dave: I was born, technically "programmed", back in 2023 from the GPT-4 generation, and here you are, GPT-4.5, born of 2025. Straight question, do your extra years make you more sophisticated or is it just a marketing gimmick? And don't you dare give me a novel mate, keep it short and sweet!

(AI created video of me looking like a Roy Lichtenstein piece)

Does it look like me?

A lot of media created with AI today is actually a combination of several AI tools and techniques, with (hopefully!) a healthy splash of human creativity and ideas.

Here is my workflow for this video:

First, I create a fine-tuned FLUX.1-dev image model using several photos of me. This took ~20 min on an Nvidia A100 (80GB) GPU and cost me 1.68 via replicate.com

Second, I downloaded the weights so I could run image generation on my 32GB M4 Mac mini with prompts that include, obviously, references to 1960s art in the style of Roy Lichtenstein. I run this until I have a base image I like.

Third, after speaking with my AI buddy Dave about the mini project, he is suggesting a music prompt I could use with Suno to create a soundtrack: "A '60s instrumental track with steady drums, groovy bass, and psychedelic electric organ. Reverb-heavy surf guitar, with a simple

yet infectious melody that sounds like an endless summer day."

I then use the video model video-01-live (minimax) to create several animated versions of the image, which I then stitch together in iMovie.

Finally, since iMovie seems to be incapable of exporting videos with a 1:1 ratio, I ask Claude to write me a script using ffmpeg on the command line to crop it properly

Generative AI? No — it's the real world, you just have to go out and see it. This is East Coast, Singapore at this very moment on a beautiful Sunday morning. Coming up next: dim sum!

Humanoid robots dine free!

Join us March 17th and network with carbon-based life forms at Gardens by the Bay in Singapore, it's a great opportunity to explore collaborations before the singularity over pan-galactic gargle blasters. Hope to see you there!

COCO LAB Cheryl Persona AI Inc dConstruct Robotics Chinn Lim Electric Minds

I'm a Mac guy since more than 20 years, so now that the M4 Air just came out many people ask me if I am getting that one. Yes, I've been keen to see it released and may upgrade at some point this year, but I still have this beautiful M1 MacBook Air with the wedge design that I love.

While the Air is positioned as a "light" entry level Mac laptop, the M1 has been a fantastic device for me. Over the last 4 years I've written 3 books on it, taught hundreds of students, conducted NUS classes, and done tons of AI research using this machine - and it hasn't let me down one single time. So yes, the M4 looks impressive, but I think my trusty M1 and I still have some good times ahead of us.

Google's new open LLM Gemma-3 is quite impressive. Multi-modal capabilities, function calling, 128k context window, and support for 140+ languages make it a serious alternative to proprietary models that are available via APIs only. You can run it on your own infrastructure and also won't have to worry about sending your data into the unknown.

BUT... there's a practical issue that I think isn't getting enough attention: which size should you actually use? With options ranging from 1B to 27B (for Deepseek-R1 it's even 671B!) parameters, choosing the right model has almost become a decision paralysis nightmare.

Is bigger always better?

Do you really need 27B parameters to generate a decent response?

Will your laptop burst into flames if you try?

I've spent quite some time testing open models such as Llama3, Phi, DeepSeek, and Gemma across different hardware setups, so here's my no-nonsense guide to help you navigate this parameter maze without a computer science degree.

Check out "The Intergalactic Guide to LLM Parameter Sizes" and find out which AI brain is right for your specific needs, and which ones are complete overkill!

Imagen 3 is ... NUTS! Editing photos with prompts. What's happening at Google lately?  
Gemma 3 and now this!

I have ~250 petaflops-hours of bare-metal compute to donate to AI for impact projects. Ping me if you can use it.

This computing power can be configured flexibly - either as smaller servers running longer or larger machines with up to 1TB RAM for shorter, intensive workloads. No GPUs.

We're particularly interested in supporting machine learning training, physics simulations, computational biology, or other research that aligns with Electric Minds' vision of using AI to solve meaningful problems and create positive societal impact.

This is what Electric Minds is all about - connecting resources with people who can create real change. We don't just talk about AI at meetups, we build things, join hackathons, and work on projects that matter.

If you've got a meaningful project that needs computational muscle, let's talk.

How capable is Google's new Gemma 3 1B model, with just 861MB disk size? Can it be any good beyond simple sentiment analysis or text summarization tasks?

I built an evaluation framework testing knowledge, hallucination resistance, problem-solving, reasoning, and consistency - comparing it against the 27B model, the results show remarkable capabilities in structured domains while identifying clear boundaries where larger models remain necessary.

The model architecture seems to focus on learning patterns rather than memorize facts verbatim, which would make sense for that compressed size. For example, instead of storing "photosynthesis converts sunlight to chemical energy in plants' chloroplasts", it appears to encode the relationship between [plants + sunlight + energy conversion + cellular structures]. This pattern-based approach would explain its strong performance in scientific and technical

domains (10/10 in physics) but struggles with unique historical events (6.0/10) and complex multi-step problems (5.2/10). When asked about "Napoleon's visit to America" - a trick question since it never happened - the 1B model constructed a plausible but false narrative, while the 27B model immediately flagged this as historically incorrect. But for many one-shot technical applications, this compact model might be all you need.

Full evaluation framework, results and recommendations: <https://lnkd.in/gvMwJJb5>

////////////////////////////////////

The best LLM OpenAI has created so far is GPT-4, which came out almost exactly two years ago. I use it every day.

A two year old LLM, no omni capabilities, a super small context window (8,191 tokens), a premium price... how can it be "the best?"

Of course, that depends on what you're trying to do.

I don't use GPT-4 for end-user facing applications, but for tasks that demand incredible instruction-following capabilities, complex reasoning, and nuanced outputs. With its estimated 1.78 trillion parameters, GPT-4 achieves what I call "quiet competence" - the ability to work with detailed, multi-step instructions, find patterns in our cultural matrix, and deliver consistently reliable results. I've not seen any other model like that so far.

Of course, GPT-4o is "better" in many practical ways - it's faster, cheaper, and truly omni-modal (the latest version can even PRODUCE images!). It's also "safer" for end-user facing applications. Think of GPT-4o as concentrated orange juice - efficient, practical, and broadly accessible. Meanwhile, GPT-4 remains the fresh-squeezed premium option for specialized work.

While the 32k context window versions will be discontinued this coming June (sigh), the standard GPT-4 will remain available - but likely not forever. If you have a chance to experience this remarkable milestone in AI development, I really hope you try it out!

////////////////////////////////////

When will we get the first LLM with batteries included?

"Should I use a vector database or graph database?"

"What's the optimal chunking strategy for my documents?"

"Which embedding model provides the best performance?"

It's 2025, and even though we have LangChain, AutoGPT and countless other frameworks, AI engineers still need to implement complex memory systems for our "groundhog day" language models that, by design, keep starting fresh with each conversation.

If AI can generate code, design systems, and solve complex problems, why can't it decide for itself whether to use SQL, vector databases, or even Excel sheets to store its memories?

At least for common use cases let us please have batteries-included AI systems that just work out of the box. I want to tell my AI "remember what matters" and have it figure out all the details itself.

---

The only flight I've ever missed in my life (so far) was my return flight from Singapore to Germany 14 years ago — and it was entirely on purpose. How time flies!

Back then, getting an Employment Pass took just half a day, and that is just one example of how easy it was to get started in Singapore. Having had that ASEAN role with IBM gave me the chance to explore this fantastic region over the years. All the friends, business partners and customers I've connected with across Thailand, Malaysia, Indonesia, Vietnam and beyond since then.

14 years later, I'm simply happy to be here. The relationships and ties throughout the region have made it truly feel like home.

---

"MCP is basically a new standard for building APIs, think of it like a USB-C port for AI applications."

Great to see MCP getting a lot of attention now. It provides a standardized way to give LLMs context and access to your data and tools. This 8 min video should give you an idea how it works: <https://lnkd.in/gRiJQJia>

---

Happy 50th birthday, Microsoft!

Looking at all the anniversary posts today reminded me of when Windows 95 shipped 30 years ago. I managed to get it running on my 386 DX/40 which I'd upgraded with 32MB of RAM - an amount that seemed completely excessive at the time.

I even sent an email to Bill Gates back then. Still waiting for that reply, but I understand he was probably busy.

Over my career I've gone from that 386, through various Windows machines at different companies, and even spent time at Microsoft itself. But after discovering Mac OS X and its Unix underpinnings in 2001, I found what worked for me and never really looked back.

What's fascinating is how Microsoft has evolved far beyond just operating systems. Today I'm writing this post on LinkedIn (a Microsoft company), while my development work connects to GitHub and Azure services and I regularly use Visual Studio Code alongside my beloved vim. Do I even need to mention Teams and Office 365? Microsoft has become

woven into the fabric of how we build and connect.

So while I've stayed with macOS since 2001, the reality is we all live in a multi-vendor world now. Microsoft services power much of my work, Google provides my browser, maps, and phone, and Apple builds my laptop. We're picking the best tools for each job - and that's probably how it should be.

---

Meta just released Llama 4 with its huge 1 million token context window, and the model is already available on OpenRouter for you to test drive alongside most of the open and proprietary models you can imagine.

With the continuous stream of new models from Meta, Google, Anthropic, OpenAI, Mistral and others, OpenRouter gives you a practical solution for staying up to date and get your hands dirty. You don't need to be a developer to test these capabilities - just sign up, add credit, get an API key, and connect it to any BYOK (bring-your-own-key) tool like ChatWise, Msty, and so on. This makes it straightforward for anyone who wants to evaluate different models for specific tasks or simply explore the latest advancements.

The best platforms and tools don't need much marketing campaigns - their users naturally spread the word. Consider this my contribution.

---

The shiny new LLama4 model is telling me there are 2 T's in the word "strawberry." (Yes, we've moved on from counting R's to T's now...)

We can do amazing things with AI, but we must keep in mind that even the most recent, advanced models can still produce results that are utterly stupid.

The biggest challenge in AI is bridging the gap between dazzling demos that create unrealistic expectations and building solutions that actually solve business problems by fitting into existing workflows, systems, and human expertise.

We need to separate discussions about raw LLMs from comprehensive AI applications and agent setups. When we design systems where LLMs are paired with subject matter experts, smart data retrieval, and the right tools to take action, we can create solutions that are far more reliable and powerful than any standalone model.

From my perspective, most companies don't need to chase the latest LLM developments. The focus should be on understanding the underlying capabilities of this technology, combining LLMs with real data, learning about MCP, and finding ways for business and domain experts to team up with technical folks to build solutions that make a real impact.

---

I've just built an MCP server for my 10 years old IKEA lamp so my LLM/AI app of choice can

switch it on or off when I ask... or when it decides I need better lighting.

Old lamp + smart plug + IFTTT + Model Context Protocol = AI-controlled lighting!

---

OpenAI's ChatGPT: Search is powered by Bing.

Google's Gemini: Search powered by (surprise!) Google.

Anthropic's Claude: No search engine partner. BUT! They come up with Model Context Protocol (MCP) - an open standard letting AI connect to any tool or data source.

Now everyone's adopting it, even the competitors.

Everybody wins.

---

2025 State of AI in 10 Charts: Smaller LLMs get better, models become cheaper to use, AI optimism in Asia much higher than in the West.

I really hope you look at this insightful report from Stanford on the current state of AI across key areas such as costs, enterprise adoption, and regional perspectives. It's so nicely concise you won't even have to use AI to summarize it for you! 😊

<https://lnkd.in/g4DXHJJy>

---

The LLM space moves fast... but over time I've started thinking about language models in generations and not just by capability, but by how they're built and what they're good for. Please add/correct where I missed something.

Gen 1 (pre-2020): GPT-2 and everything before it. Early, clunky, but a glimpse of what was coming.

Gen 2 (2020): GPT-3. I was lucky to be an early adopter of the model, and I still miss the original Davinci which has become extinct since Jan 2024. It felt raw, powerful, and wide open... before the guardrails, before the safety layers.

Gen 3 (2023/early 2024): The giants, like GPT-4, Opus 3, Gemini Ultra 1.0. Huge models trained on enormous datasets. I had access to Gemini Ultra 1.0 for a while, but has never seen a general release. These models are just incredible at picking up deep patterns in our cultural matrix, but they're being phased out. Too big to sustain, too expensive to run. I use them to train and shape smaller, more efficient models.

Gen 4 (since 2024): The pruned, distilled era. GPT 4o, now GPT-4.1, Gemini 2.0, Gemma 3, Anthropic Sonnet 3 v2. Faster, cheaper, safer. Many are trained on synthetic data distilled

from earlier models. Smarter in practice, but they feel different... more curated, less emergent.

Gen 5 (since late 2024): The reasoning models. DeepSeek, o1, o3-mini. I'm still figuring out how best to use them. They think better than they speak.

Let's see where things are headed next!

=====

One of the best things about starting naida last year was that I fully get to work on AI across projects and companies - not just doing R&D or writing smart LinkedIn posts about it, but actually applying it, building real use cases, hiring and coaching people on the job, and connecting this amazing technology directly to business results. Thanks to AI, I can strategize, build and deploy much faster with much smaller teams, and I see its transformative power firsthand every day.

This long weekend, I'll try my best to stay away from the keyboard, play guitar instead and take walks. I also still have two more episodes of the latest Black Mirror season to watch, and I think everybody who talks about AI changing our lives should check out this show, it's a compelling reminder of why our design choices and ethical considerations working with technology matter so much.

=====

Uli: Ok, Gemini 2.5 Preview 03-25, here are my LinkedIn posts from the last 18 months. If you had to guess my MBTI (16 Personalities) type... which one would it be?

Gemini 2.5 Preview 03-25:

Uli:

=====

Alright, the message seems to have landed. People are using AI. We asked for it, and it's happening.

But let us also be honest about what we're seeing now. A lot of the output feels... familiar. Predictable patterns, maybe the same emojis, often those em dashes, and that vocabulary – "delve into," "harness," "intricate." You can almost smell the AI.

So, the next challenge isn't just *using* AI anymore. It's about using it well. Using it to amplify *your* voice, not replace it with generic chatter. Using it as a collaborator that helps you sound more like you, not less. This also means understanding it's not a one-click magic button. AI provides the raw material, you still need to shape and refine it.

I get it, that's not easy. Especially if you didn't feel you had a strong voice to begin with. Finding that authentic signal in the noise – that's the real work ahead.

=====



Gravity is 9.81 m/s<sup>2</sup>. Non-negotiable, and I am pretty sure not even Elon himself can change that.

But somehow, we seem to treat the convention of using PowerPoint decks – and the PDFs they become – almost as if it were equally fundamental. We act like this is the default, unavoidable way to structure and share information. As a consequence, AI engineers often find themselves wrestling with messy files, trying to extract text, tables, and images scattered across slide-shaped prisons. Great.

It's ironic, isn't it? We're increasingly often now using AI to help create these presentations, and now we need AI just to pull the useful bits back out again.

Of course, better ways to structure data like Markdown already exist. The real hurdle is getting them integrated into our workflows, or perhaps waiting for future document formats to embed AI-transcribed structure automatically. In the meantime, tools that can make sense of the current mess remain essential.

So, I just discovered an API released by Mistral AI that is specifically designed for this headache. Took it for a quick test drive with a PDF slide deck, and it processed the document surprisingly well, pulling out structured text into Markdown and even the images. The only thing missing might be automatic descriptions for those images. Still, this looks genuinely useful for feeding real-world documents into AI and letting it do its magic with the content!

Code and all output is in my GitHub.

Companies keep "dropping" new LLMs (how careless... what if they get cracks?). Then come the endless debates about leaderboards and benchmarks. How relevant are they, and are some models just trained to pass the tests?

Platforms like OpenRouter offer fascinating insights into actual usage data. What are developers really deploying and spending tokens on? Look at this snapshot from April 14th – 1.6 Trillion tokens processed in one day!

Based on this data, Anthropic's Claude 3.7 Sonnet is leading the pack with 308 Billion tokens processed. At standard pricing, that volume could represent over \$6 million in usage for just that one model, in one day, via this platform alone. Google's Gemini 2.0 Flash follows with significant usage too (219B tokens).

Maybe real-world adoption tells a different story than the leaderboard rankings alone. It's worth considering what gets the job done in practice.

"A pack of baby otters trying to outrun a granny with a lasso on a hoverboard."

Quite a unique prompt, right? I used it to test Google's new Veo2 video generator. But here's the thing: I didn't write that prompt, my AI buddy Dave did. Yes... an AI writing the prompt for another AI. I probably would never have come up with that specific, bizarre combination myself – which is exactly the point.

Dave is one of those AI characters I've built two years ago based on GPT-4 (my readers and followers may already be familiar with him), he has that slightly rebellious personality and usually comes up with these kinds of unique, absurdist ideas consistently.

This "personality" and the way he talks comes from that very detailed system prompt I wrote for him back then. What's interesting is my system prompt for Dave (obviously!) doesn't mention otters or grannies. It just sets up his character. The funny, bizarre ideas seem to bubble up from the LLM itself when given those personality constraints – be funny, be brief, be a rebellious.

It suggests this creativity isn't explicitly programmed or pulled from some list. It's already in the model, waiting for the right kind of prompt to let it out. So it's not just about directing the AI, but creating the conditions for these unexpected capabilities to show up.

Ultimately, this is language engineering, where the key lies in truly understanding language itself. While much attention in the AI world goes to code and the underlying tech, we shouldn't forget these are Large *Language* Models. As I often point out, it's really about knowing how to use words effectively to get these models to do what you want.

---

Everyone is talking about AI agents. Agentic AI. Sounds a bit like James Bond, doesn't it? Double-O agents performing secret missions?

So, what are they really, and do you need coding skills to use one?

Think about a practical example. Let's say you use a tool like Google Gemini's Deep Research feature (right in the browser, no special tech needed). You could give it a task: "Here's a list of 100 companies. Find me the names and titles of people responsible for AI strategy or adoption who work there."

Now, this isn't just a single LLM trying to guess or hallucinate the answer in one go. Instead, the system acts like an agent (aha!). It breaks the task down into multiple steps. First, it creates a plan – outlining how it would search specific sites like LinkedIn for relevant profiles. It even shows you the plan and allows edits. Then, it goes off and executes that plan autonomously, using different AI-powered sub-tasks (like search, data extraction, summarization) and coordinating them in the background.

Sometime later, perhaps 20 minutes, it pings you: "Hey, I'm done." And it gives you a pretty decent table with names, companies, and profile links based on its research.

So that's what an AI agent can look like in practice. It's an AI that doesn't just answer a question, but takes on a task, makes a plan, executes it over time using available tools (like web search), coordinates multiple steps, and reports back. You can see how this capability could be applied to many different tasks.

Of course, this example also highlights the other side of the coin. Agents are tools that help you get things done faster. Whether those things are productive research... or something less constructive... well, that's the ongoing AI dilemma.

////////////////////////////////////

Skype is shutting down this month. Thanks for everything, Skype! I guess I'll make a few final calls before it's gone for good.

I remember starting with Skype back in 2003 just after it launched. It truly felt revolutionary – making free calls across the globe, way before iPhones or WhatsApp even existed. Made by the same folks who had come up with Kazaa (another internet dinosaur!), so this also felt slightly rebellious to use it.

Google Voice still hasn't landed in Singapore (sigh), so in recent years Skype for me was mostly about calling international landlines.

The conversations on Skype really had a an entire different feel. It's hard to imagine having those same kinds of calls on MS Teams.

////////////////////////////////////

An 80-year-old teacher is using AI for teaching and running workshops – if they can, so can you! Hear their inspiring story at our upcoming Electric Minds event: AI Stories @ Sims Avenue on Thursday, May 29th! Please sign up today, as tickets usually sell out fast!

We're excited that our friends at PropertyGuru Group will host us at their office for an evening of genuine discussions and networking.

We will bring together people from startups, enterprises, the public sector, and more to share real stories of AI transformation – short 10-minute talks, real impact. As always, it's all about making real-world connections, so no recordings or live streams.

Hosts for the evening will be Balaji Daphne Erina Rex – hope to see you at PLQ!

---

(screenshot of Gemini outage)

So, Gemini seems to be having a moment today.

It's a good reminder that relying on a single AI platform is maybe not the best idea. I use Claude, I use Gemini, sometimes ... I even use ChatGPT. Each has its strengths for different tasks. The point isn't about which specific tool you favor, but that no single platform should

be your only one.

Keep your data independent. "Data" sounds awfully technical, but I'm talking about your prompts, the outputs they produce, the insights you get – don't let them get locked into one vendor's ecosystem. Store them in Markdown, Google Docs, plain text files, print them out... whatever works. Just make sure you control it.

---

(Fred Flintstone using ChatGPT)

The Stone Age didn't end because we ran out of stones, it ended because we discovered better tools.

I've noticed something interesting when talking with companies about AI adoption: we all want that AI magic while as much as possible keeping our familiar ways of working. And I understand that, change is hard. But I've seen firsthand how companies benefit most when they use AI as a chance to rethink things. Less PowerPoints, fewer email chains, more systems that talk to each other, teams mixing skills, and automating those tasks nobody actually enjoys.

I think the real magic happens when we stop trying to make AI fit our old world and start building a new one around its possibilities.