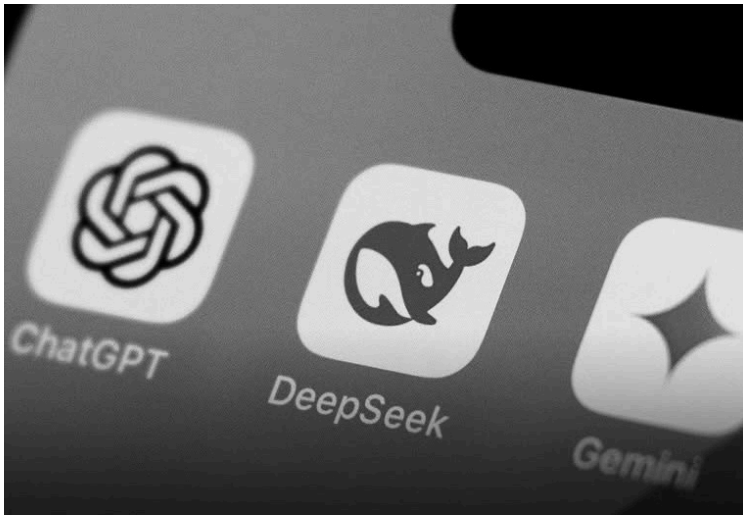
Deep Seeking or Chatting GPT? An AI Product Economics Perspective.

Wilfred Oliver Antwi *February, 2025*

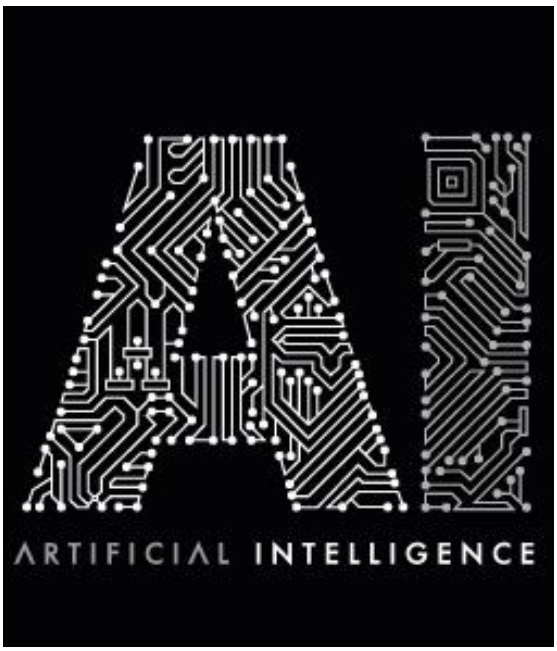


"America expensively innovates, China cheaply multiplies"

A dark horse AI startup has risen from apparent obscurity to ruffle some feathers in the insuperable US AI prominence sending markets reeling in a tailspin.

A seismic knock

The incessant vibrating headlines across major news outlets these past few days of DeepSeek's earthquake R1 release seems to have ruffled some feathers, send some heads rolling, and questioned certain standards and benchmarks. Forbes has indicated that Nvidia's shares were down by 17% shedding a whopping $589 billion in market capitalization as of January, 27 a symbolic gargantuan record single-day value wipeout for any company in memorial. Nasdaq shed 3.1% while S&P 500 fell 1.5%. Other heavyweights in the chip industry such as Broadcom and ARM were not spared as they saw about 10% drops, with Microsoft, Oracle, Palantir, Alphabet and even energy stocks like GE Verona and Siemens being impacted. In total, about 1trillion of BigTech stocks was the value wiped out by DeepSeek's laud and expensive bang. Business as usual "America expensively innovates, China cheaply multiplies". To keep the springboard balanced, US always leads breakthroughs and China engineers faster.



A dark horse

The traditional leadership in the AI craze seems to have been confronted with a new challenge and solution at the same time. Just to scratch the surface of the bruhaha, DeepSeek (深度求索 or Shēndù Qiúsuǒ) the AI darling of the last couple of weeks is a Chinese Large Language Model (LLM) open source company incorporated in Hangzhou, Zhejiang and bankrolled by the Chinese hedge fund High Fly, since its establishment in 2023. DeepSeek has developed and released supposedly a US$6 million R1 AI rival - a mere drop in the ocean compared to the reported cost of its American counterparts considering Microsoft having already invested anything around $13 billion in OpenAI and ChatGPT.

The first free chatbot app with around 670billion parameters released on 10 January, 2025 on Apple IOS and Android, by January 27 had chalked success as the most downloaded here in the US outpacing ChatGPT. Further, DeepSeek-R1 is powered by the open source DeepSeek V3 model costing $5.6 million to trained in two months on Nvidia's GPUs despite restrictions on sales of advance Nvidia chips to the PRC. The rhetoric that the shared open-source nature of DeepSeek AI developers ensured that less computing power was needed hence less cost is debatable. Liang Wenfeng claimed that DeepSeek-R1 AI model was trained on around 2800 specialised chips however, Mr. Musk, the tech messiah has pointed out that Alexandr Wang (Scale AI) divulged to him that DeepSeek has around 50,000 NVIDIA H100s.

This if true could have cost DeepSeek anything over a $1.3 billion to acquire (considering a cost of between $27,000 and $40,000 per unit, based on PCIe or SXM). The allegation of DeepSeek having used OpenAI's data in a proprietary technology without authorisation (distillation) seems to be brewing as well. No matter the side of the aisle you are sitting, the debate about computing power needed to develop LLM AI is real. DeepSeek has also launch an image-generation model Janus-Pro-7B touted to outperform the likes of OpenAI's DALL-E-3 and Stable Diffusion by Stability AI. Janus-Pro-7B is claimed to have over 72million high-quality synthetic images combined with natural images with about 7billion parameters within reach.

DeepSeek's release should be "a wake-up call...if you could do it cheaper, if you could do it for less and get to the same end result..."
President D.J. Trump

The conundrum nevertheless necessitates the question of adoption and adaption of AI by medium size and smaller businesses, and individual usages in everyday life. The inevitable question is whether cost reduction can accelerate the adoption of AI across industries. Are there acceptable and agreed ways of producing LLM generative AI, Natural Language Processing (NLP), Deep Learning, Artificial General Intelligence (AGI), and even Artificial Superintelligence (ASI) at a lower cost just to enhance mass adoption? Or is a sort of innovation reserved for BigTech only? In the words of President Trump DeepSeek's release should be "a wake-up call...if you could do it cheaper, if you could do it for less and get to the same end result..."

Just to scratch the surface of what the product economics and adaption could mean for the general LLM AI market, it's imperative to review the elements of costs of AI models. Generally, the cost elements of AI companies taken with a grain of salt and caution of oversimplification could encompass;

1. Research & Development (R&D) covering personnel such as Data Scientists/Engineers with salaries averaging $200k a year per employee here in the US. Machine Learning Engineers with the likes of data scientists, specialising in model development and deployment with averages of $180K a year per employee. Data Analysts for model training 120k a year per person.

2. Perhaps the biggest headache is Data Acquisition & Preparation - a cumbersome process that includes data licensing, obtaining high-quality datasets either niche or proprietary from different sources. Data cleaning and preprocessing including labelling requiring appreciable levels of human and computational power. Additionally, data annotation in the case of supervised learning. Hardware Costs for High-performance computing (HPC) infrastructure, Powerful GPUs (Graphics Processing Units), TPUs (Tensor Processing Units), and other specialized hardware for model training and inference, large-data storage (cloud eg. Azure, AWS, Google) anything around $10k to millions. Software Costs example AI/ML software and tools like TensorFlow, PyTorch, and other AI development frameworks at least $5k+ a year per software.

3. Model Design, Development & Training mainly reserved for the engineer and research teams to experiment, train, tuning and more - the larger the size, the better which also leads to significant cost. Meta Prompting, modality changes, sequencing, automatic prompt, and what have you in the field of Prompt Engineering. Significant time investment is required for model design, experimentation, training, and tuning using the likes of Keras (an API that runs on top of TensorFlow), Scikit-learn (A Python machine learning library), XGBoost, and more. An underlining computational resource with high GPU power especially for LLM model training could skyrocket the cost. It also depends on the model whether GenAI, LLM, etc.

4. Deployment & Maintenance which cover infrastructure costs such as servers, cloud platforms like Amazon SageMaker, & Google Cloud AI platform all comes in. Costs associated with deploying and maintaining AI models in production environments like recurrent costs for hosting, scaling, and maintaining the deployed models. Monitoring, constant updates and upgrades are also crucial to ensure their accuracy and performance over time. Regular updates and upgrades are necessary to improve model performance and address emerging threats such as adversarial attacks which are possible.

***5.* Miscellaneous cost may include legal, licenses, and regulatory** compliance costs associated with complying with data privacy regulations for instance GDPR in Europe, CCPA) and other relevant laws, for instance DeepSeek's Janus-Pro has a commercialised use under the MIT license. Sales and promotion which is last on the ladder of cost all seem to come in.

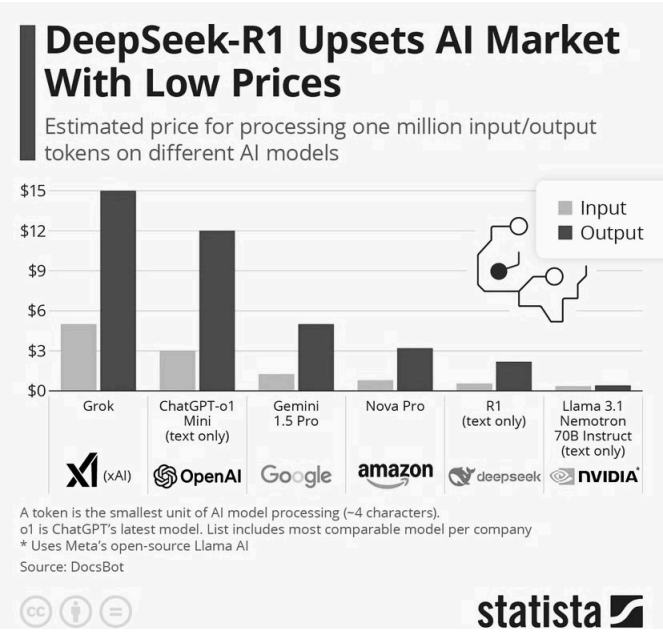
What could this mean for AI Products & Market?

After such a somewhat weighty and ponderous process, cost of production, working capital, current ratio, market tests and breakeven point analysis all comes to play. Notwithstanding, DeepSeek seems to have won a competition lottery out of thin air having crossed these hurdles unscathed and seemingly enjoying the teetering feat. The jury is out that DeepSeek R1 cost $5.6 million to trained in two months (55days) on 2,788 H800 GPUs. Such a cost-effective development approach provides a significant competitive advantage in the LLM AI market.

Could faster time-to-market have played a role in cheaper development costs and accelerated the development process, allowing DeepSeek to bring its AI solution to market more swiftly?

The jury is out that DeepSeek R1 cost $5.6 million to trained in two months (55days) on 2,788 H800 GPUs. At a fraction of the typical development cost of 100million and over, as in the case of ChatGPT-4.

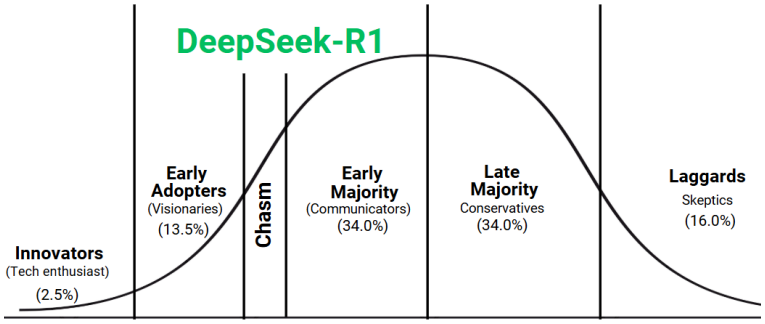
By lessening development costs and achieving a lower break-even point, DeepSeek can allocate and plough back more resources towards marketing, customer acquisition, and product innovation, further solidifying its position in the competitive AI landscape. At a fraction of the typical development cost of 100million and over, as in the case of ChatGPT-4, DeepSeek could reach faster adoption rates hence reaching the masses.



This means offering a compelling price-performance ratio compared to alternative LLM AI solutions. A lower break-even point, achieved through minimized development costs, signifies that DeepSeek requires fewer sales to cover its expenses. Therefore, it comes as a no surprise at all that DeepSeek-R1 offers a significant pricing edge with tokens way below the scale of its LLM peers. Uploading 1million input/output tokens costs between $0.30-0.55 against the $3-15 price tag for models such as ChatGPT, Gemini, Bard, and xAI (Grok). Perhaps a competitor at the bottom of the price ladder is Llama 3.1 of Meta. Statista covers this better and in more details with the table below.

Product Adoption Analysis

The relatively popular assessment method and forecast for software product seems to toe the line of that of other consumer product in an available market. Innovators, (Tech enthusiast who spend time trying out newly release versions), the Early Adopters, Early Majority, Conservatives, and Laggards are the main subsequent categories for an available market. These stages are largely influenced by a myriad of other factors such as product targets, and other persona details. This is applied to the case of R1.



In terms of the of Product Adoption Analysis, DeepSeek could be jumping the chasm between Early Adopters (pegged at 13.5%) and Early Majority (34%) underpinned by its cost effectiveness. In the case of an unexpected positive mass adoption attitude that could come in the form of “American consumer revenge” during the bruhaha surrounding the possible ban of TikTok, where other Chinese applications (RedNote & Lemon8) became the alternatives, R1 could fly in the face of other LLM models. Irrationality is quite a predictable behaviour to expect from consumers when pricing becomes the value parity. Innovators (2.5%) hurdles were already overcome by DeepSeek as of 2023 when it caused a little storm in the AI teacup.

DeepSeek commanded early Innovators interests when it released several models including the DeepSeek-V2 series in May 2024 with 4 models, 2 base models and 2 chatbots. A subsequent DeepSeek-V2.5 was added in September the same year and upgraded in December. Clearly DeepSeek is crossing at the speed of light the space that normally exists between Early Adopters who are normally visionaries transitioning, and Early Majority who are slow at adopting new technologies. Inasmuch as R1 will not answer questions on the Tiananmen Square incidence and other sensitive information related to the PCR, the experimentation, adoption, and rise seem rapid. This would not have been the case has the AI craze been a bubble – apparently it is not hence any AI that seems to nurse in users the FoMo syndrome, and cheaper will do the magic.

And this is irrespective of whether R1 collects personal data or not, considering the reaction from users in the recent attempt to ban TikTok, there will always be the crazy adopters. The Time-to-value (TTV) which implies the time taken by a user to realize the expected value from a software product in this case R1, is swift making users (small and medium enterprises, and individual app users) enjoy the freemium rather quickly. This could catalyse mass adoption of AI by small and medium size enterprises because adaption cost would not eat up a chunk of their budgets and abnormal profiteering would not heap cost on the heap of personal users.

And So what?

The US AI market is innovative yet that of a modern Oligopoly – a situation whereby between 3 to 20 large firms dominate and nevertheless about 40% of the market is controlled by 4 to 5 BigTechs. These major players are at the giving and receiving end of all needed skills and resources. Fleetingly, gradually directly and indirectly overshadowing all possible "deep-seeking" startups that could replicate similar shocks caused by DeepSeek. As of this article President Trump had just announced an AI infrastructure investment project worth $500 billion supposedly packed with 100,000 new jobs, together with Oracle, Open AI, and Softbank. Once again, it goes to the BigTech who have enough resources and can foot their own development costs, consequently another rain into the ocean whiles smaller "deep-seekers" still wallow and crawl in development stages. Any educated prophesy that a chunk of this resources will go into acquisition of smaller nascent stage AI companies may come true, leading to mergers and acquisitions instead of low adoption.

The void is left for Chinese "deep-seekers" to fill and obviously they won't disappoint. If AI use in smaller enterprises and ordinary app users is costly, adoption and adaption will be economically grim which will open the doors for alternative AI models no matter where they come from, how shoddy, and clandestine their data collection tactics may be. Its American peers will wobble.

Only 2% of businesses that desire the involvement of AI really succeed in doing so successfully. There are about 30,000 AI tools for about 15,800 tasks with about 5000 job categories already being affected.

To smaller businesses for instance generating less than $20million a year feeling the urgency to adopt and adapt to AI, it is worth taking your time to know whether you really need it in your everyday work or ERP. You don't have to pay a higher cost for that, don't just join the bandwagon with a sick chatbot incapable of addressing real need of your customers. However, if it is a must and not just buying into the fad, consultation, strategic utilization, and cost concerns should be at the crux. Only 2% of businesses that desire the involvement of AI really succeed in doing so successfully. There are about 30,000 tools for about 15,800 tasks with about 5000 job categories already being affected. These will not affect people directly but rather what you do and how you do it. Obviously, majority of the needs are RPAs, Workflows (Human-in-the-loop or not), chatbots, data analyses and insights, deploying tools and prompting, or simply knowing the right AI to use. Finally, know which will work for you whether DeepSeek-R1 or ChatGPT and consider the ergonomics and economics of your firm.