Industry Report
27 November 2023

Q* at OpenAI - are we approaching the singularity? Maybe, but not on the front you'd expect!

Research Analysts

Pierre Ferragu
+1 646 681 4616
pierre@newstreetresearch.com

Rolf Bulk
+65 6016 5450
rolf@newstreetresearch.com

Antoine Chkaiban
+44 20 7375 9126
antoine@newstreetresearch.com

Benjamin Harwood
+1 929 379 4139
ben@newstreetresearch.com

Derek Mogull
+1 646 681 4638
derek@newstreetresearch.com

Piyush Kalyani
+44 20 7375 9145
piyush@newstreetresearch.com

Tobias Bähr
+44 20 7375 9111
tobias@newstreetresearch.com

As a follow-up to the short note we published last week-end, we provide here some additional thoughts on what happened at OpenAI in recent weeks, with a particular focus on the much speculated Q* (pronounced Q - Star) model recently tested internally. Although very little is known at this stage, we provide a framework to understand potential implications for the technology value chain.

The facts

Multiple media[1] and innumerable social media conversations[2], often amongst very knowledgeable and respected opinion leaders, have discussed that going into the recent coup, the OpenAI team had internally tested a next-generation AI model, which proved capable of solving math problems at a grade-school level. Although we could argue over how much a grade-school pupil can achieve these days, it was reportedly considered a significant breakthrough.

There is very little known about this new model, and nothing officially. Amongst the voluminous unconfirmed information reported by the press and social media, we curated the following for our readers:

The model is named Q*. Q-learning is a reinforcement learning methodology, in which an AI agent explores possible solutions to a problem by trial and error, step by step, estimating the value of each move it can make, in a graph of possible actions and possible states, following a policy typically learnt during a training phase. In this method, Q is the value of the action taken, as estimated by the policy, and Q* refers to the maximum optimal value that can be achieved from any given state, if the agent follows the optimal path from that state.

Jakub Pachocki, Szymon Sidor, and Ilya Sutskever are at the origin of the model. The former two developed it based on original work from the latter. Pachocki and Sidor were the first two employees to resign after the failed coup, ousting CEO Sam Altman and Chairman Greg Brockman. Sidor worked at OpenAI on the

[1] See Reuters and The Information articles for details

[2] See for example tweets of Yann LeCun & Ate-a-Pi

reinforcement learning part of 'OpenAI Five', the model which achieved supra-human capabilities in the game 'Dota 2', and before that, on causal models at Vicarious AI (models which can generalize causality relationships from limited data.) Pachocki also led the research work on Dota alongside Sidor, and then led the reasoning team at OpenAI. His latest role was as director of research for OpenAI and leader of the pretraining work for GPT-4.

The underlying research involved using computer-generated data rather than real-world data to train the model (as imprecise as that is as a statement.)

Many reference AlphaZero as a source of inspiration for this work, or at least as a relevant comparison. AlphaZero is the model Google's DeepMind developed, after AlphaGo, which was characterized by using reinforcement learning. A simplified way to understand this is that AlphaGo learnt how to play the game of Go from human games, whereas AlphaZero learnt from playing against itself. This is reinforcement learning at its very basic.

Journalists refer to test-time computation, as a mysterious "machine-learning concept". The journalists visibly don't understand what this refers to, but this term must have come up in their investigations... and you will see towards the end of this note that it matters a lot.

The recent hire by OpenAI of Noam Brown is also referenced regularly. His expertise lies in the development of planning, or the ability of AI models to think through possible solutions, vs. producing a solution. He worked for instance on having AI masterfully play the game 'Diplomacy'.

At the APEC conference, the day before the board attempted to fire him, Sam Altman, made a sibylline comment: "*just in the last couple of weeks, I have gotten to be in the room, when we sort of like push the sort of the veil of ignorance back and the frontier of discovery forward*".

The picture on our front page became a very popular meme on Twitter[3] in the last couple of weeks, given the pace of speculation, and is brought to you on the lighter side. Most important sources for the bullet points above are referred to in footnotes. We insist on the fact that none of these reports have been confirmed officially by companies, institutions or individuals involved. This "fact" section is therefore probably the most speculative one we have ever written. It nevertheless genuinely reflects our best understanding of what can reasonably be inferred from the noisy signals of the last week.

Our analysis

Water down sensationalism.

We understand sociological and anthropological dynamics pushing journalists and social medial commentators to sound like recent events are signs of an unprecedented breakthrough towards AGI (Artificial General Intelligence). Facts seem to rather indicate OpenAI remains at the forefront of innovation and is progressing in the direction in which the industry is pushing, but there is no objective evidence of an unusual step-up.

It seems about consensual amongst top researchers in the field that the main avenues of development of AI is to leverage reinforcement learning for the fine-tuning and alignment of large language models. After all, this is what produced the magical ChatGPT bots. The next frontier is to develop the planning capabilities of these models, i.e. their ability to reason towards an answer vs. spitting one straight from their billions of weights. This avenue is being investigated by all major teams, including GPT-5 at OpenAI and Gemini at Google, but no product announcement has been made yet. The recent news flow only seems to indicate that the OpenAI team passed a significant milestone on their journey in that direction.

In our view, it could be equivalent to the breakthrough that ChatGPT was. ChatGPT is a pretrained large language model, fine-tuned with RLHF (reinforcement learning from human feedback.) This tuning is what made ChatGPT feel so casual and good at answering a succession of questions. The chatbot nevertheless still displays significant limitations. Amongst others, it produces answers that are unreliable, or "made-up" (hallucinations) and it doesn't display proper reasoning capabilities (or only anecdotally). With better planning capabilities, such a model could increase significantly in usefulness, approaching more an "agent" role, solving problems or taking actions in multiple steps.

[3] We refer here to X, formerly known as Twitter. It takes us a very long time to adapt to new names.

Understanding planning.

Daniel Kahneman nailed in 2011 the idea of contrasting the two modes along which humans think.

- System 1 is when we operate near-instantaneously, automatically, instinctively, with limited effort or sense of control.
- System 2 starts when we allocate attention and effort to solve a problem.

System 1 is often associated with a reflex, and doesn't require much agency. System 2 is usually associated with agency, making choices, concentrating, and taking multiple steps to solve a problem, or more generally produce a thought.

The definition we prefer for planning is the ability to 'drive' system 2 thinking.

A good example used by Kahneman is solving a multiplication problem, like 17x24. System 1 can assert that 16 is not the right answer. System 2 is required to provide the right answer. It is not an example we picked randomly… our readers will note that the most recent progress OpenAI has made relates to solving math problems… To us this is a strong indication that this progress is about planning capabilities.

How do we get AI to do planning?

AI is already very good at planning, but only on very specific tasks. Deepmind's AlphaZero is the best well-documented example of an AI system doing a good job at System 2. As it plays chess or Go, at every step, it brings the board to a configuration that increases its chances of eventually winning.

Needless to say, it is also a good example because it achieved supra-human capabilities.

OpenAI 5, playing Dota 2 is also a good example of AI planning and delivering supra-human performance at a very specific task. Although the system is more complex to comprehend than chess or a Go board, the principle is the same: At every move, the AI evolves the system into a state, which increases its odds of the AI eventually winning the game.

How do we build such an AI?

In principle, there is nothing complicated, as is always the case in AI (remember we coined the term *'Artificial Stupidity'* in 2015). AlphaZero simply played against itself the equivalent of billions of years of games in order to progressively score all the board configurations that are the most likely to come up in a game, in terms of the probability of winning from this configuration. If you dig one step deeper and increase the level of abstraction, it means the AI selectively crawls through the tree of all possible plays, and remembers how each configuration it visits brings it closer to a win. The important word not to miss here is 'selectively', since the tree of all possible plays cannot be fully crawled. Its size means it could, probably not be fully crawled, even by a computer the size of the universe[4].

This is where Q-learning and Q* come in. Q-learning is the algorithm of selectively going through a tree of possible states, looking to increase the probability of winning at every step.

Q-learning is a reinforcement learning approach. The algorithm associates a reward to every step through the tree and computes at ever step its Q-value, ie its estimate of the total reward it could get from this state.

Q* is in some ways the maximum of the Q-value, i.e. the total reward the AI would accumulate, were it to make the best move at every step.

Obviously, by naming an AI Q* the team may mean they cracked a very efficient Q-learning algorithm, and this is not easy. To be able to learn how to do this, the AI needs two things that super-intelligent and super-nerdy humans need to figure out: First, an algorithm that allows an efficiently selective tree crawl; second a way to rate each possible move in the tree with a reward. This is where the very simple algorithm becomes challenging to apply to a given problem.

So what did OpenAI do?

Our best hypothesis (after crawling the tree of possible explanations in an efficient and selective way) is that OpenAI developed and implemented a Q-learning algorithm, which crawls through possible answers a language model makes to a prompt, token by token, or group by group of tokens, and somehow selects an optimal answer.

Many in the field are working on this idea, and it is arduous along the two implementation requirements described above.

The first challenge is to define the Q-value of each transition. As the model writes a sentence, each word

[4] We are not sure of what a computer the size of the universe is, or a tree the size of the universe, but here we want to vividly reflect that the number of possible games in Chess is 10^120, i.e. more than the number of atoms in the universe.

comes one after the other, and the model has the choice at every step between several words. How to determine which one, or which ones are the most likely to lead to the best answer? We can easily understand how playing again and again chess will help estimate this for a multitude of board configurations, but it is difficult to intuitively imagine how to define such a function for a language model, especially a universal one, valid for any use case, any conversation.

The second challenge is the size of the tree. The tree for the game of chess is larger than the universe, the tree for a language model, in which each token has tens of thousands of possible values, and the number of successive tokens is the number of words in your text, is infinitely larger.

Here again, the fact that solving math problems seems to be the task at which the new model was tested concurs with our hypothesis. We can intuitively imagine that every step in a math demonstration can be evaluated with regard to whether it can lead to a solution. It would likely use human feedback, with operators grading elements of answers, based on their own knowledge of math and experience at solving problems. This is something for which OpenAI would be very well positioned, with the experience accumulated in the field of RLHF (Reinforcement learning with Human Feedback) for ChatGPT. The difference with ChatGPT is that feedback would not be about picking the best final answer but picking the best path towards a good answer, token by token.

Implications for infrastructure

Let's now address what excites us the most here. What does the introduction of planning in AI systems based on large language models mean for compute, and therefore, compute infrastructure?

Training.

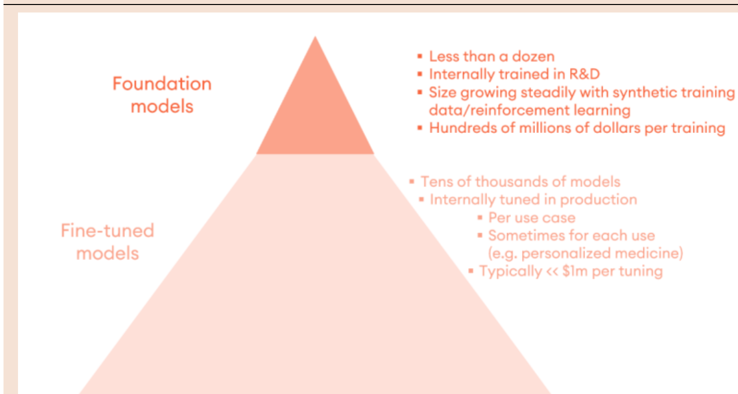
Reinforcement learning can be understood as replacing real-world data by 'compute'. (The attentive reader will note this is one of the comments picked by journalists that we highlighted in the "Facts" section of this note.) Our depth of expertise doesn't allow us to quantify the impact, but we can give a directional hint on a couple of axes:

In a reinforcement model, LLMs, after first training on existing real-world data could then generate their own content, evaluate it, probably in a self-supervised way, and use it for further training and tuning. This means a lot more compute, and, removing any upper limits. On the amount of training that can be made. If the universe of possible states is about infinite (the tree of all what the model can create) and the data available to train the model is self-generated, we would need to see evidence of dramatically diminishing returns before we believe the race for scale could slow.

As mentioned above, planning is unlikely to be universal anytime soon, but is likely to be leveraged for a lot of task-specific or use-case-specific fine tuning. Here again it gives an open-ended outlook to compute growth. As techniques to combine LLMs and planning capabilities increase, the amount of fine tuning on the giant ever-growing models described above - which generate their own training data, will increase a lot, this time driven by the breadth of adoption, in addition to the size of models.

Exhibit 2 shows how we see this training market growing as a result: a handful of ever-growing giant foundation models and the proliferation of thousands of variants of these models, fine-tuned to specialized tasks.

Ex 2 – Foundation models will lead to thousands of fine-tuned models



Source: NSR analysis

Inference.

We see for inference something completely new at play. At the time of AlphaZero, inference compute did not go up with innovation, on the contrary. AlphaGo was running a tree search at every stroke and was therefore running inference on a huge infrastructure: 1920 CPUs and 280 GPUs. That is a midsize datacenter of its own. AlphaZero did not need a tree search anymore as it was able to select the next best move just from what it had learnt with reinforcement learning. (In a way, the innovation of AlphaZero meant moving the tree search from inference to training.) As a result, AlphaZero is running inference on half a server, with 4 TPU chips. This is 3 orders of magnitude (1,000x) less than AlphaGo.

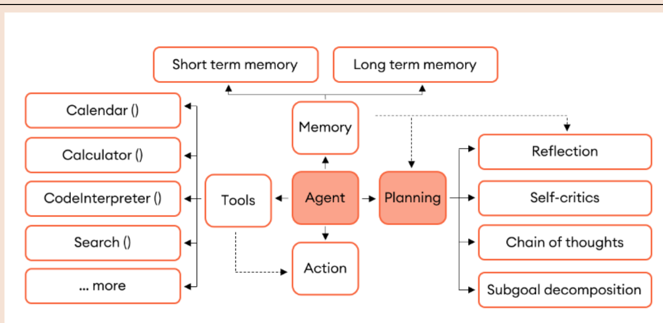
Here, as planning capabilities are developed for large language models, we see an inverse evolution: much more inference compute. Here again, our depth of expertise and the open-ended nature of the field doesn't allow us to quantitatively measure the impact, but we can make the following observations:

In press reports, test-time computation is mentioned. In layman terms, it means inference compute. This probably means Q* has materially increased inference compute in order to achieve its breakthrough.

Exhibit 3 (borrowed from the excellent write up on LLM autonomous powered agents published this summer by Lilian Wang[5]) displays well how inference compute must go through the roof: Instead of producing a text token by token, the agent enters into a planning mode in which it will reflect, assess its own production, improve it, chain thoughts, and decompose its tasks in multiple subtasks.

Intuitively, we believe an agent based on an LLM looking for a better, more thought-through answer than the one that comes to mind instantly (System 2 vs. System 1) is unlikely to be trained, and even less so pre-trained in a generic way, but more likely to run trial and error and optimization routines at inference time, prompt by prompt. Our hypothesis is that Q* operates this way, developing step by step math demonstrations.

Ex 3 – Planning capabilities increases inference compute demand



Source: Lillian Wang paper on LLM powered autonomous agents, and NSR analysis.

Compute infrastructure.

We see two main implications. One is quantitative, the other is qualitative.

On the quantitative front, in both training and inference, we see the development of planning capabilities resulting in an enormous increase in compute requirement, driven by both training and inference, the former playing out now, the latter potentially very soon.

This actually fits well with what we have observed in the last 12 months, in terms of training: We expect revenues from Nvidia and Broadcom related to AI compute to grow 66% and double respectively, between 2023 and 2024, i.e. between the world of Bard and GPT-4 and the world of GPT-5 and Gemini, who are both expected to feature planning capabilities. In our conversations with practitioners, at various steps of the value chain, from chip manufacturers to frontier developers, we understand that training and the increase in model complexity is still the primary driver of compute growth today, while inference, even in the context of a very fast adoption of AI services, has remained a distant secondary driver until very recently.

Inference is what now gives us cautious but increased confidence in further growth in compute investments, beyond 2024. Nvidia management already mentioned on the last earnings call, last week, that inferencing "*is now a major workload for NVIDIA AI computing*". Imagine what happens next, as these planning agents are deployed: it will at the same time accelerate the growth of compute per inference *and* the adoption of services, i.e. the volume of inference!

The second implication we see is more qualitative: the dream of doing better than GPUs with in-house chips may take a serious blow.

As models embed planning capabilities to enable agents, there is a good chance that algorithms for both training and inference will go through significant changes and diversify significantly. This will narrow the window of opportunity to develop alternative chips for a simple reason: There is no free lunch in chip architecture. A new ASIC can do materially better than a state-of-the-art GPU (i.e. an order of magnitude better) only if it pushes specialization further. We showed this several years ago in our research. Investing in an in-house chip can make sense when a use-case generates workloads in very high volumes, which will

[5] LLM Powered Autonomous Agents

remain very similar over a long period of time. If, on the contrary, it is unclear what the workloads of tomorrow will be, and if workloads of today may be replaced by significantly different ones, using the off-the-shelf best-in-class GPU is likely to remain the best option for most, for quite some time.

Conclusion

We don't see an AI singularity (a point where AI becomes sentient and supra-human "in general", at any task), but we potentially see a singularity, just ahead of us, in terms of 1) the compute resources required to continue to "push back the veil of ignorance", and 2) how much these compute-hungry innovations can further accelerate the pace of adoption of compute-hungry AI agents.

The combination of these two would create a singularity... in the pace of deployment of additional silicon. Never forget, Ilya Sutskever said in 2019: *"It is pretty likely that the entire surface of the earth will be covered with solar panels and datacenters".*

We feel maybe we understand better today what he meant at the time

Relevant Research

- OpenAI drama - Our read, and implications. – (19 November 2023) - Link
- Microsoft unveiled two in-house chips. Our take. – (16 November 2023) - Link
- OpenAI DevDay: implications for tech infrastructure – (7 November 2023) - Link
- The latest in advanced packaging - take-aways, replay & preferred slides – (17 October 2023) – Link
- Who can afford that many AI clusters? - Conference Call Slides. – (2 October 2023) – Link
- Generative AI: Updated thoughts on the back of our Big Idea Conference. – (25 September 2023) – Link
- Nvidia: which clients drove most of the surge in datacenter revenues? – (6 September 2023) – Link
- Nvidia, Arista, Broadcom: How much more room for A.I. revisions? – (15 August 2023) – Link
- How much does a 1% increase in A.I. server penetration add to WFE spending? – (3 August 2023) – Link
- What does TSMC doubling CoWoS capacity tells us about Nvidia and AI chip growth? – (14 June 2023) – Link
- Meta's MTIA? Best illustration that succeeding with an in-house chip is challenging... – (23 May 2022) – Link
- Cluster economics of Nvidia GPU vs. Google TPU: How different? Implications for market structure? – (22 May 2023) – Link
- Power semiconductors: Our updated thoughts following the PCIM conference – (16 May 2023) – Link
- AMD & Microsoft designing a custom A.I. chip? Our quick take. – (5 May 2023) – Link
- Amazon shareholder letter: Our read on the momentum of in-house accelerators – (17 April 2023) – Link
- Memory share of Wallet in AI: lower than in traditional servers – (14 April 2023) – Link
- Is Memory a good way to play the ramp of A.I. clusters and Hyperscale Capex? – (6 April 2023) – Link
- What drives Hyperscale capex going forward? A deep dive in the economics of Cloud – (27 March 2023) – Link
- Intel DCAI webinar: our 3 takeaways. – (30 March 2023) – Link
- Nvidia GTC 2023: a few very interesting things we learnt... – (22 March 2023) – Link
- ChatGPT early deployments this year: potentially very material for Nvidia – (7 February 2023) – Link
- ChatGPT - Implications of large language models for tech infrastructure (NVDA, TSMC, ...). – (24 January 2023) – Link
- Tesla A.I. Day: 7 take-aways you won't read elsewhere. – (3 October 2022) – Link
- Semiconductor Big Idea Conference 2022: Our 9 key take-aways & slides. – (23 September 2022) – Link
- Sapphire Rapids, Genoa, ARM: Updated thoughts on server CPU roadmaps – (17 August 2022) – Link
- Dojo, TPU, Cerebras, IPU, GPU et al. Future of Exascale computing architectures –(25 March 2022) – Link
- Giant AI chips & computers, quantum... First Semiconductor Big Ideas Conference. - (13 September 2021) – Link
- Packaging innovation (III) – Networking: the 5-year journey towards the single-chip switch - (24 June 2020) – Link
- Packaging innovation (II) – Can chiplet architecture challenge the CPU Status Quo? - (28 January 2020) – Link
- Packaging innovation (I) – the upcoming datacenter GPU war - (16 January 2020) – Link
- Chip Architecture and manufacturing: Why does Intel have 97% share in Server? - (24 September 2019) – Link

- Technology Infrastructure Coverage initiation - (03 May 2018) – Link

Disclosures

12 month historical recommendation changes are available on request

This report was produced by New Street Research LLP. 100 Bishopsgate, 18th Floor, London EC2M 1GT Tel: +44 20 7375 9111 Regulatory Disclosures: This research is directed only at persons classified as Professional Clients under the rules of the Financial Conduct Authority ('FCA'), and must not be re-distributed to Retail Clients as defined in the rules of the FCA.

This research is for our clients only. It is based on current public information which we consider reliable, but we do not represent that it is accurate or complete, and it should not be relied on as such. We seek to update our research as appropriate, but various regulations may prevent us from doing so. Most of our reports are published at irregular intervals as appropriate in the analyst's judgment.
This research is not an offer to sell or the solicitation of an offer to buy any security in any jurisdiction where such an offer or solicitation would be illegal. It does not constitute a personal recommendation or take into account the particular investment objectives, financial situations, or needs of individual clients.
All our research reports are disseminated and available to all clients simultaneously through electronic publication to our website.

New Street Research LLC is neither a registered investment advisor nor a broker/dealer. Subscribers and/or readers are advised that the information contained in this report is not to be construed or relied upon as investment, tax planning, accounting and/or legal advice, nor is it to be construed in any way as a recommendation to buy or sell any security or any other form of investment. All opinions, analyses and information contained herein is based upon sources believed to be reliable and is written in good faith, but no representation or warranty of any kind, express or implied, is made herein concerning any investment, tax, accounting and/or legal matter or the accuracy, completeness, correctness, timeliness and/or appropriateness of any of the information contained herein. Subscribers and/or readers are further advised that the Company does not necessarily update the information and/or opinions set forth in this and/or any subsequent version of this report. Readers are urged to consult with their own independent professional advisors with respect to any matter herein. All information contained herein and/or this website should be independently verified.

All research is issued under the regulatory oversight of New Street Research LLP.

Sales:

Charlie Gaynor
charlie@newstreetresearch.com

Zach Monsma
zach.monsma@newstreetresearch.com

Ethan Lacy
ethan.lacy@newstreetresearch.com

Mark Franks
mark@newstreetresearch.com

Michael Chambers
michael@newstreetresearch.com

Daniel Gilroy
daniel.gilroy@newstreetresearch.com

Dimitri Livchits
dimitri@newstreetresearch.com

Heather Broffman
heather@newstreetresearch.com