Team 1192

[redacted]

5th August, 2023

How much electrical power will be required due to New Zealanders adopting the use of generative AI?

1 Abstract

With the continuous advancement of artificial intelligence and the widespread adoption of generative AI, the electrical power required to keep up with the demand is an important factor to be considered. We aimed to solve the question of how much power would be required in New Zealand with this uptake in generative AI use, using ChatGPT as our model for a generative AI. Through data analysis, we concluded that the electrical power required in 2024 due to this would be 460 kW or 4×10^{10} J per day, across energy use per query, and cooling of data centres.

2 Introduction

Our civilizations have always been driven by rapid technological advancements, and modern times are no exception. The potential posed by artificial intelligence, is boundless, and set to revolutionize information technology as we know it. It allows machines to create and innovate where in the past, only humans were capable of such things. It has caused a seismic shift in not only business and innovation, but also our daily lives. Due to its diverse range of applications, the uptake in use of AI has been massive, with an estimated value of 142.3 billion USD across the global AI market[1], and for most of us generative AI is at the forefront. Generative AI is a branch of artificial intelligence which allows machines to produce human-like content. Its functions range from detailed artwork, lifelike images, emotive literature, and even computer coding, unlocking a revolutionary new set of tools for our use.

However, one must consider the potential threat on the horizon: energy consumption. Data centres and data transmission networks use 2-3% of all energy produced globally[2], which may seem small but in really is massive. AI are likely to cause an increase in this number due to the large number of servers required to run one. With how environmentally conscious New Zealand is, this is a large consideration required when looking to the future of Generative AI use

within our country. Attempting to raise electrical energy production without compromising attempts at switches to renewable energy sources is a difficult task thus it is important to know how much of an impact the current switch to AI technology will have on power resources.

3 Definition/s

We have decided to use specific definitions of variables included within the question to make it possible to come to a plausible answer.

Firstly, we have defined Generative AI as "technology that creates content, text, images, video, code, by identifying patterns in large volumes of data and then creates original material with similar characteristics. Chat GPT is a generative AI."[3] We have also chosen to split Generative AI into two large groups, text generation and image generation, due to the large difference in function and energy usage between them, so that the data can be calculated separately. We have defined text generating AI, as any AI model which solely generates text-based content of any form, and image generating AI, as any AI model which solely generates image or video-based content of any form.

We have chosen to define the phrase "will be required" as an estimate for energy use by the generative AI users on a daily basis in 2024, given current usage data for our first model, and the same for a projected population and AI usage number in 2024.

4 Assumptions

5 Calculations and Models

5.1 Usage Rate of Large Language Models (LLMs) in New Zealand

A large language model (LLM) is a novel type of text-generating model that has recently risen to prominence. We began by focusing on how much New Zealanders were using LLMs. Using ChatGPT, by far the most commonly used of these models, as our standard, we started by constructing a distribution for the number of queries per visit a user will make on ChatGPT. We will call this distribution X with probability distribution function $f(x)$.

We will model this discrete distribution by approximating it with the normal distribution, which we call N, such that $N \sim N(5, 1.5^2)$. We arrived at this value by determining the estimates for how much users queried ChatGPT per visit varied from a lower bound of 1 query per visit to an upper bound of 10 queries per visit, with a mean of 5, based on an average site visit time of approximately 7 minutes and 27 seconds[4]. Assuming a normal distribution where the interval [1,10] represents an interval of three standard deviations on either side of the

mean (reasonable since this covers 0.997 of the normal distribution) , we arrived at a standard deviation of 1.5.

Now we will find an expression for $f(x)$. Since x represents the number of queries per visit, it is an integer. This means for each value of x, the probability is the probability that N is in the interval $[x - 0.5, x + 0.5)$.

$$f(x) = P(x - 0.5 \leq N < x + 0.5)$$

To calculate this we, need to transform N into the standard normal distribution, $Z \sim N(0, 1)$.

$$\begin{aligned} N &\sim N(5, 1.5^2) \\ N - 5 &\sim N(0, 1.5^2) \\ \frac{N-5}{1.5} &\sim Z \end{aligned}$$

which is the standard normal. Therefore the final expression of $f(x)$ is as follows:

$$\begin{aligned} f(x) &= P(x - 0.5 \leq N < x + 0.5) \\ &= P\left(\frac{x - 5.5}{1.5} \leq Z < \frac{x - 4.5}{1.5}\right) \\ &= \Phi\left(\frac{x - 4.5}{1.5}\right) - \Phi\left(\frac{x - 5.5}{1.5}\right) \end{aligned}$$

where $\Phi(z) = P(Z \leq z)$, which can be calculated through code. Note that the probability that $x < 0$ is small enough to be ignored.

We then constructed the distribution for the number of visits per day a Chat GPT user will make. For this, we observed that ChatGPT is estimated to have around 100 million users, and that the website has around 60 million unique visitors per day [10][11], giving a probability of 0.6 that a user visits the site on any given day. Since the number of visits is a positive integer, we can model it with the Poisson distribution. Utilising a Poisson distribution with λ =0.6, we obtain the following distribution, which we will call Y.

$$g(y) = \frac{0.6^y}{y!} e^{-0.6}$$

We then combined these two sets of data to form a final distribution for the number of queries per day for ChatGPT users.

To find the number of queries sent per day, we need to multiply the number of visits per day with the number of queries per visit. We will call the probability distribution of the number of queries per day Z.

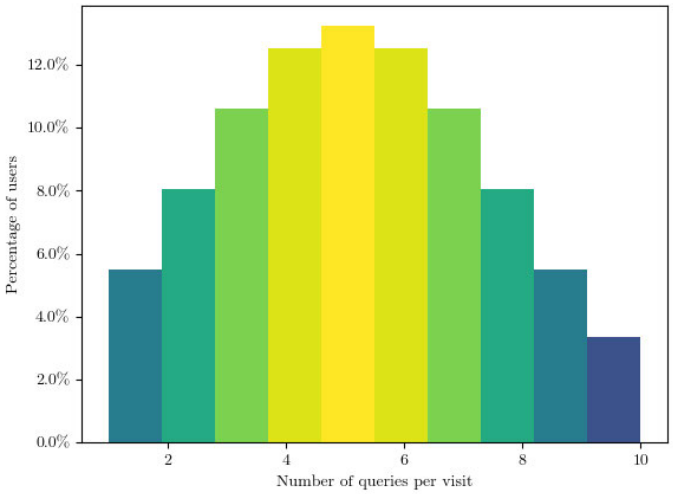

Figure 1: Distribution of number of queries per visit of a user to ChatGPT

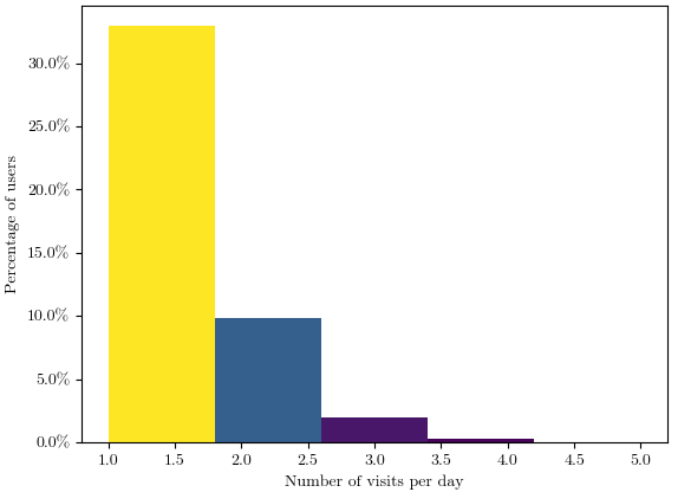

Figure 2: Distribution of number visits per day per person

To find the probability of a specific value z of the distribution Z, we need to find the probability of *all* pairs of values of x and y such that $xy = z$, which we can calculate through the following formula.

$$h(z) = \sum_{\forall x,y \text{ s.t. } xy=z} f(x)g(y)$$

Since the probability distributions X and Y are assumed to be independent, we can multiply the probabilities f(x) and g(y). Then we simply add all the probabilities, f(x)g(y), for which $xy = z$ is satisfied.

For example, to find $P(Z = 6)$ we will need to add up $f(1)g(6)$, $f(2)g(3)$, $f(3)g(2)$, and $f(6)g(1)$, since these are all the pairs of values of (x,y) which multiply to 6.

Since we know the expressions for $f(x)$ and $g(y)$, we get that

$$\begin{aligned} h(z) &= \sum_{\forall x,y \text{ s.t. } xy=z} f(x)g(y) \\ &= \sum_{\forall x,y \text{ s.t. } xy=z} \left[\Phi\left(\frac{x-4.5}{1.5}\right) - \Phi\left(\frac{x-5.5}{1.5}\right)\right] \frac{0.6^y}{y!} e^{-0.6} \end{aligned}$$

We used this data to make an average for the number of ChatGPT queries made per user, per day in New Zealand. This is the expected value of the combined distribution.

$$\mathbb{E}(Z) = \sum_{\forall z} z \times h(z)$$

This gives the approximately 2.798 queries per user per day.

5.2 Energy Consumption per Unit Time

There were two parts we considered to the energy consumption rate for Generative AI: fundamental energy requirements and energy requirements per query.

The energy required to run a single LLM query is approximately 0.00396 kWh[6], but on a national scale, tools such as ChatGPT consume a truly titanic amount of power. However, one complication encountered is that New Zealand's population has grown dramatically over the last decade and will continue to grow as such in future.

In 2021 (a period in which internet usage demographics were not significantly different to now), New Zealand comprised around 5,129,727 users with access to the internet[7], or approximately 0.1% of the 5.3 billion internet users worldwide[8].

Additionally, LLMs such as ChatGPT are trained on input data, consuming a comparable amount of power to queries. This process does not occur continuously, but at intervals, as new versions of the language models are released.

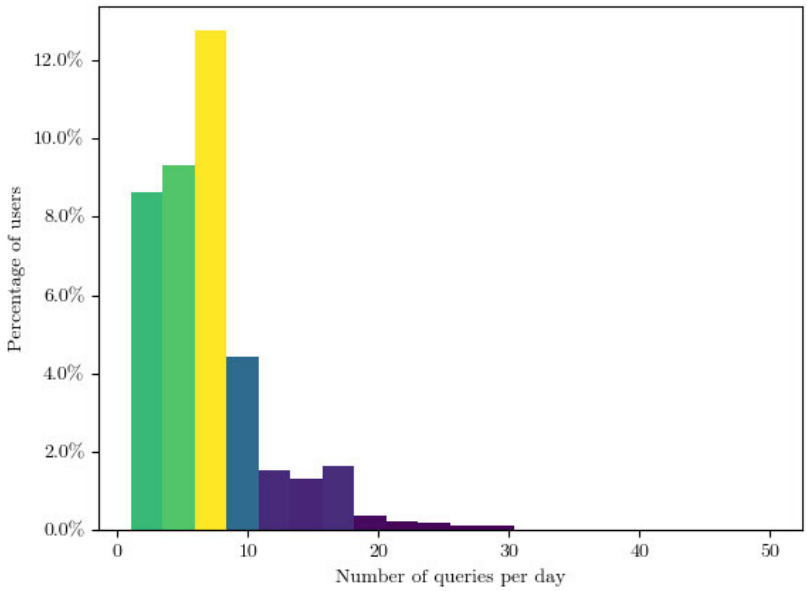

Figure 3: Distribution of Number of Queries per Day per ChatGPT user.

This process requires an amount of energy on the order of 1GWh[5]. With the increasing technology sector in New Zealand, we can assume that in the future this training can and will occur in NZ, particularly with the construction of data centre currently taking place in Auckland[9], thus we have decided to include this into our 2024 estimate.

Based on this research we can state that 0.00396kWh is used per query and an additional 1GWh is used at intervals for training.

An additional 40% energy expenditure is incurred in the form of the energy required to cool down data centres (and consumer devices)[12].

5.3 Chat GPT Users in New Zealand and Estimated Users in 2024

However, the population of New Zealand has not remained stagnant over the course of the last decade. Rather, it experienced a period of accelerated growth pre-pandemic, and now this rate has slowed once more. To compensate for this, we approximate the population growth by means of a 6th-degree polynomial (correlation coefficient 0.9937). A lesser degree did not adequately fit the curve, and a higher degree led to overfitting.

$$\begin{aligned} &-0.000000001894899218627570 9t^6 \\ &-0.000001052441706967139 4t^5 \\ &+0.030592792187053804t^4 \\ &-14.476829528429 96t^3 \\ &-103353.220353692 42t^2 \\ &+56201810.86236322t \\ &+82294447795.12036 \end{aligned}$$

Where t is the year. Substituting in t=2024, we arrive at a population of 5313146, compared to the 5228100 of 2023, representing a 1.62% increase in population. Assuming the proportion of internet users remains relatively constant, we would expect to see around the same increase in the number thereof.

Additionally, some large countries do not have access to ChatGPT. Compensating for their populations, we find that the proportion of New Zealand internet users is in the ballpark of 1.36%.

With usage of ChatGPT having stabilised at around 100 million users, we might determine that the number of ChatGPT users in New Zealand is on the order of 702000 regular users, assuming the proportions remain the same. Accounting for a 1.62% increase in population raising this to around 714000 users by the same time next year, we may now calculate the final daily energy consumption of ChatGPT.

6 Final Results

$$\begin{aligned}\text{energy consumption} &= \text{energy consumed per query} \\ &\times \text{number of chatGPT users} \\ &\times \text{number of queries per day} \\ &\times (1 + \text{additional energy expended in cooling}) \\ &= 11078\text{kWh per day}\end{aligned}$$

This is approximately 4×10^{10} joules per day, or about 460kW.

7 Addendum: Image Generation Models

By far the two most widely adopted forms of generative AI have been the use of it for text generation, and the use of it for image generation. Nonetheless, adoption of the former has utterly eclipsed that of the latter. Even the largest AI image generation service, Midjourney, has less than 13 million accounts using it, most of which are thought to be duplicates.

What's more, image generation is not applicable to such a broad range of situations as text generation, and consequently users of image generation also use it dramatically less.

Nonetheless, it is worth calculating, as an addendum. The time taken to generate an image on a high-end consumer GPU (the only GPUs that can currently run it) is on the order of seconds. Equally, the wattage of those GPUs is also relatively low, in the 230–300-watt range, as one may observe in the table below.

GPU	Generation time (S)	Power usage (W)	Energy usage (J)
A100 80GB PCIe	6.49	300	1947
RTX 3090	7.96	360	2865.6
RTX A6000	8.09	300	2427
RTX A5500	8.55	230	1966.5
RTX 8000	12.3	260	3198

These values are very small compared to the many kilowatt-hour values needed for generative models.

8 Conclusion

Given our calculations and assumptions, we can say that $4x10^{10}$ joules of electrical energy will be required per day in 2024 due to the adoption of generative artificial intelligence by New Zealanders, due to an increase in AI users in New Zealand, and a switch to using local data centres. We arrived at this answer by research and calculating estimates for daily generative AI queries, energy use per query, and population growth by 2024.

References

[1] *97+ ChatGPT Statistics & User Numbers in July 2023 (New Data).* URL: https://nerdynav.com/chatgpt-statistics/ (visited on 08/05/2022).

[2] *Artificial Intelligence (AI) worldwide - statistics & facts.* URL: https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/ (visited on 08/05/2022).

[3] *chat.openai.com Traffic and Engagement Analysis.* URL: https://www.similarweb.com/website/chat.openai.com/#overview (visited on 08/05/2022).

[4] *Data Centres and Data Transmission Networks.* URL: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (visited on 08/05/2022).

[5] *DCI Data Centers commences construction on Auckland's largest data centre, set to inject another* $400 million into local economy.. URL: https://dcidatacenters.com/news/dci-starts-akl02-construction/ (visited on 08/05/2022).

[6] *Digital 2021: New Zealand.* URL: https://datareportal.com/reports/digital-2021-new-zealand (visited on 08/05/2022).

[7] *Digital 2021: New Zealand.* URL: https://datareportal.com/reports/digital-2021-new-zealand (visited on 08/05/2022).

[8] *Estimating the carbon footprint of BLOOM, a 176B parameter language model.* URL: https://arxiv.org/pdf/2211.02001.pdf (visited on 08/05/2022).

[9] *How much energy do data centres use?* URL: https://dataspan.com/blog/how-much-energy-do-data-centers-use/ (visited on 08/05/2022).

[10] *How much energy does ChatGPT consume?* URL: https://medium.com/@zodhyatech/how-much-energy-does-chatgpt-consume-4cba1a7aef85 (visited on 08/05/2022).

[11] *Limits of Artificial Intelligence and Why it Matters.* URL: https://www.auckland.ac.nz/en/news/2023/05/23/limits-of-ai-and-why-it-matters.html (visited on 08/05/2022).

[12] *Number of ChatGPT Users 2023*. URL: https://explodingtopics.com/blog/chatgpt-users# (visited on 08/05/2022).

9 Appendix

9.1 Code Used