Team ID: 1086

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

Abstract

The purpose of this report is to estimate the electrical power required from New Zealanders adopting the new rise of generative AI. Our model assumed proportionality between the number of AI users in New Zealand and the number of AI models that exist for the New Zealand market. Using our model, we used projected data for 2028, to conclude that in five years time, New Zealanders will use 281 MW of power due to their use of AI. This is an increase of 14000% compared to our estimations of New Zealand's power consumption due to AI today. This value (281MW) is roughly equivalent to 3% of New Zealand's total power supply today.

Introduction

With the increase in demand for both manual and creative work, the use of generative AI had risen considerably. With many companies adopting or exploring the implementation of AI tools such as ChatGPT as well as Midjourney, It raises the concern regarding the practicality and efficiency of implementing such a tool.

The cost of running such an AI model is not insignificant, and the electrical power required in the process constitutes a large percentage of that cost. For context, the development of the recent ChatGPT saw a loss of $540 million to its creator, OpenAI. (Yahoo Finance, 2023)

Increased power consumption also raises concerns about the carbon footprint of generative AI. With an increase of power consumption due to AI, there may be concerns about the sources of energy which these companies use, further increasing carbon emissions to the atmosphere.

In this report, we will model the changes in power consumption due to increased usage of generative AI, and provide a projected value for power consumption from AI in five years time.

Interpretation and Definition of Key Terms:

Due to the uncertainty of this question, we intend to keep our variables practical to keep our discussions realistic.

By “adopting”, we will model the electrical power consumption of the New Zealand population over the next 30 years. We can assume that as time goes on, a larger percentage of the workforce will adopt the use of generative AI. These workplaces are primarily high skill labour sectors, e.g: certain creative fields or mass distribution of goods.

“Generative AI” encompasses tools such as OpenAI and Midjourney, and as we see other companies adopting similar models to these, we will primarily reference them for data when compiling this report.

Assumptions

- Adoption of generative AI in NZ will remain largely similar in all developed western countries
- The use of generative AI will have negligible impact on other sources of energy consumption.
- The underlying technology in the near future stays the same (there will be no radical changes to the fundamental way which things work although there may still be optimizations in current technology to make it more efficient).
- The number of models trained and served in New Zealand is proportional to the number of Generative AI users in New Zealand

Below are some more assumptions we made, and our thought process when designing our model.

Designing Our Model

Under the previous assumptions, we can construct a model which can approximate the average power consumption caused by an increased usage of generative AI.

In order to do this, we will consider two variables. Firstly, the number of people who will likely have access to tools such as ChatGPT, and are inclined to use them. We will focus on the changes to the New Zealand population to deduce the number of people in the future who will likely use AI on a daily basis.
We will primarily be looking at population change concerning:

- The age of individuals
- The employment rate of individuals
- The number of these individuals involved in occupations that will likely utilise AI tools now or in the future.

We have selected these variables as we think it is most likely that individuals who would use AI consistently will be:

- Of an age between 15 and 65 years.

    - We have chosen these numbers as individuals below the age of 15 would likely have little incentive to use AI, and individuals above the age of 65 would be in retirement age where they too would not utilise AI. ie. This group of people have no productive need to use these tools.
- Employed or enrolled in an educational institution.
    - It is mostly likely that generative AI would be primarily used to increase productivity in the workplace, or for students, so we do not see a large amount of unemployed individuals being inclined to use AI. Furthermore, unemployment tends to correlate with a lack of wealth, which would inhibit a regular use of AI tools, as they tend to require a device with constant internet connection. As we do not see many unemployed individuals having access to these, we chose to remove them from our category of individuals who would likely use AI regularly.
- Working in non-manual occupations.
    - Individuals who work as carpenters, electricians, masons etc. have not been included. While it is likely that they will have access to AI tools, and potentially do use them recreationally, their effect on power consumption would be negligible as compared to the number of people who use them in the workplace. We would argue that the usage from these individuals would not regularly require these tools, and hence have been excluded.

Our second variable is the increase in complexity of AI tools, and the power required to account for this increase. We have explained how we account for this later in our paper. Once we have a value for the number of New Zealanders in the future who would likely use AI, we can then model the future power consumption by assuming that:

1. Number of models being trained is proportional to the number of people using AI on a regular basis.
2. The energy required to train a model increases over time in a predictable way. (We will describe the increase in power consumption later in calculations.)

Mathematical Modelling and Calculations

The serving and interference of AI models accounts for a larger amount of power consumption than the training of these models. NVIDIA estimates that 80 - 90% of power consumption of the ML workload is inference processing. Amazon Web services claimed that 90% of the ML demand in the cloud is for inference. (Patterson et al., 2022)

We will say that 10% of the energy comes from training and 90% of the energy usage comes from serving and interference

Power Consumption for a AI Model (Training)

The compute power needed for the latest AI Models are projected to increase exponentially, doubling every 3 - 4 months. It will take 4 months for the computer power to double. (OpenAI, n.d.)

The compute power that a GPU produces per watt of power supplied to it increased by a factor of 9.7 in 20 months. From the A100 that was released in June 2022, which gave a performance of 312 teraflops over 300 watts of FP16 performance (NVIDIA, 2021), to the H100, that was released in March 2023, which gave a performance of 7916 teraflops of FP8 performance over 700W (NVIDIA, 2022), which is has comparable accuracy to FP16 in AI tasks (NVIDIA, 2022), gives a $\frac{7916}{700} \div \frac{312}{300} = 10.87$ compute power per watt increase over 20 months. Thus, how many months it takes for the compute power per watt to double can be calculated by $20 \div (\log_2 10.87) = 5.8$

$$Power\ per\ model = A \times \frac{Compute\ Power\ /\ Model}{Compute\ Power\ /\ Power\ Supplied}$$

Thus the *power per model* can be modelled using the equation.

$$Power\ per\ model = Power\ per\ model\ in\ 2020 \times 2^{\frac{t}{A}} \div 2^{\frac{t}{B}}$$

Where t is the amount of time in years since 2020.

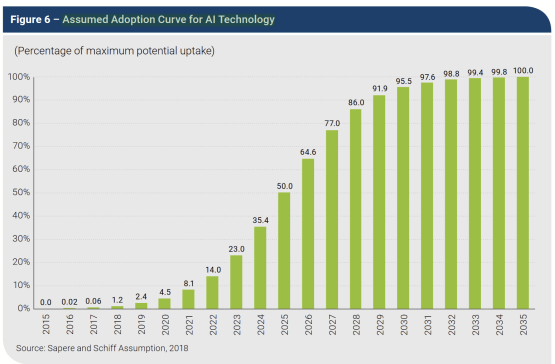
$$A = \frac{4}{12},\ B = \frac{5.81}{12}$$

These two values of A and B will be used later in our calculations.

Calculate the number of people using AI in New Zealand

$$U = N \times p_E \times p_W \times A$$

U is the number of users of AI in New Zealand
N is the New Zealand population who fit the age range of people who will likely use AI (15 - 65 years).



p_E is the number of people in the age range who are employed.
p_W is the number of employed people who work in sectors that use AI or are currently in education.
A is the rate at which individuals may actually adopt the use of AI tools.

The report from Sapere and Schiff assumes the adoption rate of AI follows a standard S curve. (Reid, 2018)

Number of AI Models in New Zealand

We claimed that the number of AI models trained in New Zealand is proportional to the number of users of AI within New Zealand

$$M \propto U$$

Power consumption

It is difficult to obtain information about the power usage of AI in New Zealand, to calculate this information we use the power usage of AI and use the ratio between the New Zealand population and the World population to obtain an estimate for the power usage of AI in New Zealand

$$Power\ Usage\ In\ New\ Zealand\ = Power\ Usage\ in\ the\ World\ \times \frac{New\ Zealand\ Population}{World\ Population}$$

Calculating Power Usage In New Zealand

IEA(2023) states that the power usage of training an AI model only uses 20-40% of the total power usage. It took 1287W to train GPT-3 (Patterson et al, 2021), thus the power usage to operate GPT-3 is given by:

$$Power\ usage\ to\ operate\ GPT\ -\ 3\ = 1287MW\ \div\ 30\%\ = 4290MW\ ()$$

$$\begin{aligned} Power\ Usage\ In\ New\ Zealand\ &= 4290\ \times 10^6\ \times \frac{5.2 \times 10^6}{8 \times 10^9} \\ &= 2.79\ \times\ 10^6\ \text{W} \end{aligned}$$

New Zealand AI Power Usage In The Future

$$\begin{aligned} Power\ Usage\ &= Power\ per\ Model\ \times\ Number\ of\ Models \\ Power\ Usage\ &= Power\ per\ Model\ \times C\ \times AI\ users \end{aligned}$$

where C is the constant of proportionality as $M \propto U$

$$Power\ Usage_{now} = C \times Power\ per\ Model_{now} \times AI\ users_{now}$$

$$Power\ Usage_{future} = C \times Power\ per\ Model_{future} \times AI\ users_{future}$$

$$Power\ Usage_{future} = C \times Power\ per\ Model_{now} \times 2^{\frac{t}{A}} \div 2^{\frac{t}{B}} \times AI\ users_{future}$$

(where t represents the time since now in years)

$$Power\ Usage_{future} = C \times Power\ per\ Model_{now} \times 2^{\frac{t}{A}} \div 2^{\frac{t}{B}} \times \frac{AI\ users_{future}}{AI\ users_{now}} \times AI\ users_{now}$$

$$Power\ Usage_{future} = Power\ Usage_{now} \times 2^{\frac{t}{A}} \div 2^{\frac{t}{B}} \times \frac{AI\ users_{future}}{AI\ users_{now}}$$

Calculating our $\frac{AI\ users_{future}}{AI\ users_{now}}$ ratio:

To find $AI\ users_{now}$, we use the following values:

N_{now} is the population of New Zealanders who are between the age of 15 to 65. (Infometrics, 2022)

$$N_{now} = 3.565 \times 10^6$$

$p_{E\ now}$ is the percentage of the population between ages of 15 to 65 that are employed. (Stats NZ, 2023)

$$p_{E\ now} = \frac{2.927 \times 10^6}{3.565 \times 10^6} \times 100 = 82.1\%$$

$p_{w\ now}$ is the percentage of that population who are working in non-vocational jobs (who are more inclined to use AI). (Stats NZ, 2013)

$$p_{w\ now} = 61.0\%$$

Hence

$$\begin{aligned} \text{AI } users_{now} &= N_{now} \times p_{E\ now} \times p_{w\ now} \times A \\ &= 3.565 \times 10^6 \times 82.1\% \times 61.0\% \times 23\% \\ &\approx 410 \times 10^3 \end{aligned}$$

Looking at the projection of the number of New Zealanders aged 15 - 65 in 5 years, there will be a 0.9% increase in the percentage of the NZ population that are between this age range. (PopulationPyramid, 2021) The total population of NZ in 5 years is projected to be at 5.36 million. (Statista, 2023)

$$\begin{aligned} N &= 5.36 \times 10^6 \times (64.80\% + 0.90\%) \\ &= 3.52 \times 10^6 \end{aligned}$$

By the projection that the employment rate of New Zealand will increase by 1.3% every year (Ministry of Business, Innovation and Employment, 2019), we have

$$p_E = 82.1\% \times 1.013^5 = 87.6\%$$

And under the assumption that the percentage of the working population working in non-vocational jobs staying constant,

$$p_w = 61.0\%$$

A is projected to be

$$A = 86.0\%$$

hence

$$\begin{aligned} \text{AI } users_{future} &= N \times p_E \times p_w \times A \\ &= 3.52 \times 10^6 \times 87.6\% \times 61.0\% \times 86.0\% \\ &\approx 1.62 \times 10^6 \end{aligned}$$

So our $\frac{AI\ users_{future}}{AI\ users_{now}}$ ratio is

$$\begin{aligned} \frac{AI\ users_{future}}{AI\ users_{now}} &= \frac{1.62 \times 10^6}{410 \times 10^3} \\ &\approx 3.95 \end{aligned}$$

Substituting into the Power $Usage_{future}$ equation, with $t = 5$:

$$Power\ Usage_{future} = Power\ Usage_{now} \times 2^{\frac{t}{A}} \div 2^{\frac{t}{B}} \times \frac{AI\ users_{future}}{AI\ users_{now}}$$

$$Power\ Usage_{future} = 2.79 \times 10^6 \times 2^{\frac{5}{(\frac{4}{12})}} \div 2^{\frac{5}{(\frac{5.81}{12})}} \times 3.95$$

$$Power\ Usage_{future} = 281 \times 10^6\ W$$

Conclusion

From our research and calculations based on mathematical models we expect New Zealand will use 281 MW of power in 5 years which is roughly equivalent to 3% of New Zealand current power generation.

Discussion

It is important to understand that generative AI (particularly as a productive tool) is an extremely recent tool, with little data on it as compared to other industry

tools. Our model was made with the limited information available on AI tools as they are a relatively new technology, particularly the rate of improvement of these tools, and how that will affect the power consumption.

It is possible that the improvements in AI will level off some time in the near future and that for practical applications there is no point attempting to advance AI and new large scale AI's won't be trained anymore reducing the power used by AI.

Many experts speculate that at the current rate of development, AI models will become way too complicated for the hardware and require an extremely large amount of power. It is very likely that new technology will be invented which is able to keep up with the developments in AI technology while using less power. This means that there is little possibility that the power consumption will continue to increase at such a drastic rate in the future. It is also likely that the development of AI will show down and AI systems will be forced to become more optimised / efficient.

However it is impossible for us to determine how technology will advance in the future. Especially for more than 20 years from now.

It is not necessarily true that the number of people using AI is proportional to the number of AI models being developed in New Zealand. It is possible that as AI advances that a few AI models become dominant while other models stop being trained. There are other possibilities and it is impossible to predict exactly how the number of models being trains relate to the number of AI users.

While the power consumption may be high, AI technology could be used to make optimizations, inventions and aid in the development of technology which reduces the power used for other things. The report does not take into account the potential savings in energy which utilising Generative AI could cause.

References

Infometrics. (2022). Population Age Composition. https://ecoprofile.infometrics.co.nz/new%20zealand/Population/AgeComposition

Ministry of Business, Innovation and Employment. (2019). *Medium to long-term employment projections: Looking ahead to 2028*. Ministry of Business, Innovation and Employment.

NVIDIA. (2021). *Nvidia a100 gpus power the modern data center*. NVIDIA. https://www.nvidia.com/en-us/data-center/a100/

NVIDIA. (2022a). *Nvidia h100 tensor core gpu*. NVIDIA. https://www.nvidia.com/en-us/data-center/h100/

NVIDIA. (2022b, September 14). *Nvidia, arm, and intel publish fp8 specification for standardization as an interchange format for ai*. NVIDIA Technical Blog. https://developer.nvidia.com/blog/nvidia-arm-and-intel-publish-fp8-specification-for-standardization-as-an-interchange-format-for-ai/

OpenAI. (n.d.). *AI and compute*. Retrieved August 5, 2023, from https://openai.com/research/ai-and-compute

Patterson , D., Gonzalez , J., Hölzle , U., Le , Q., Liang , C., Munguia , L.-M., Rothchild , D., So , D., Texier , M., & Dean, J. (2022). *The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink*. arxiv.org.

Patterson , D., Gonzalez, J., Liang , C., Munguia , L.-M., Rothchild , D., So , D., Texier , M., & Dean, J. (2021). *Carbon emissions and large neural network training*. arxiv.org. https://arxiv.org/ftp/arxiv/papers/2104/2104.10350.pdf

PopulationPyramid. (2021). *Population Pyramids of the World from 1950 to 2100*. PopulationPyramid.Net. https://www.populationpyramid.net/new-zealand/2021/

Reid, B. (2018). *Artificial Intelligence: Shaping a future for New Zealand*. Forum New Zealand. https://www.mbie.govt.nz/dmsdocument/5754-artificial-intelligence-shaping-a-future-new-zealand-pdf

Statista. (2023, April). *New Zealand—Total population 2018-2028*. Statista. https://www.statista.com/statistics/436377/total-population-of-new-zealand/

Stats NZ. (2013, May 2). *Skill levels of new zealand jobs | stats nz*. https://www.stats.govt.nz/reports/skill-levels-of-new-zealand-jobs

Stats NZ. (2023). *Employment rate | stats nz*. https://www.stats.govt.nz/indicators/employment-rate/

Yahoo Finance. (2023, May 5). *ChatGPT cost a fortune to make with OpenAI's losses growing to $540 million last year, report says*. Yahoo Finance. https://finance.yahoo.com/news/chatgpt-cost-bomb-openais-losses-125101043.html