

INFO204 Assignment 2

Yuki Yoshiyasu 5861229

Part 1. Data Ethics

The purpose of this critique is to determine whether the AI language model “ChatGPT” produced a well-rounded response regarding data ethics within the context of Aotearoa. I will be exploring the response to discuss the sections done well or needs further work.

In the Introductory paragraph of ChatGPT’s response, the following phrase was an accurate response: “The way we handle and use this data raises questions of trust, responsibility, and ethics”. It’s a great sentence as it summarizes the overarching theme of data ethics, which becomes the foundation of the following discussion points of the response. This well-written response is most likely due to the question being straightforward: “What is meant by data ethics”. Since ChatGPT is a large language model, most likely trained with an immense number of definitions of data ethics. Therefore, it can output strings of words that predict the best answer to the question, based on the training data.

Secondly, ChatGPT created two paragraphs to answer the part “How are Data Ethics different from Computer Ethics”, which I believe can be joined into a single paragraph. Since the 1st and 2nd paragraphs overlap on the points discussed. Upon going through the 1st paragraph, the following sentences could be improved on: “Data Ethics revolves around the right and wrong ways of sourcing, handling, processing, and using data.” “It concerns itself with the values, principles, and rules governing the treatment of data, especially in contexts where human rights, privacy, security, and fairness are at stake.” The first sentence can be stated concisely like this: “Data Ethics has a strong focus on curation of data and algorithmic processing”. The second sentence is a wrong response as it includes wrong information about the principles of data ethics. Listing computer ethics principles like human rights, privacy and security which is from an ICT perspective rather than a Data Ethics perspective. ChatGPT’s 2nd paragraph response was exceptional as it gives an example of the differences between computer ethics and data ethics which answers the question given. The lines explaining the main focus on computer ethics are boundless: “Computer ethics focuses on the ethical challenges related to the use of computer technology...” And “Data Ethics specifically targets the ethical use and handling of data”.

Thirdly, the listing of components for a data ethics framework in the 3rd paragraph is the key improvement point. Data ethics frameworks are supposed to promote the human concerns of data usage over technical challenges, known as human-first emphasis (Dick, 2023). The several key components listed in the ChatGPT’s response were: “Transparency”, “consent and control”, “privacy”, “fairness” and “accountability”. Whereas “consent and control” and “privacy” are not the key components of the UK data ethics framework, therefore suggests that we should remove it from ChatGPT’s response.

Fourthly, the data ethics frameworks in NZ and the role of the Data ethics advisory group, which ChatGPT could improve on. The information about the Te Mana Raraunga (Māori data sovereignty network) should be included as it covers several important principles: Rangatiratanga, Whakapapa, Whanaungatanga, Kotahitanga, Manaakitanga, Kaitiakitanga.

Additionally, the generated response should also include that we simply cannot reuse an international data ethics framework since in NZ, we have many stakeholders in data collection and use, and those stakeholders have a claim to ownership/stewardship even if they do not create the data themselves.

Finally, the response delves into the unique landscape of Data Ethics in Aotearoa New Zealand. ChatGPT should add some supporting details about how the following affects the NZ data ethics landscape: “the treaty of Waitangi principles”, which emphasizes on understanding and respecting the rights of Māori. It should also include the Te Mana Raraunga which was Founded in 2015, with Principles of Māori Data Sovereignty released in October 2018. Enabling Māori Data Sovereignty and advancing Māori aspirations for collective and individual well-being.

Stochastic parrots are a term used when a language model can generate realistic human-like responses but don't understand the true meaning of the language. This can be a large concern as these large language models are unreliable due to “garbage in and garbage out” (GIGO), where none of the output can be essentially “trusted”. This unreliability is particularly worrisome for those seeking to safeguard their language and culture, as is the case with the Māori community.

Large language models like Whisper require immense amounts of training data, such as the 680,000 hours of audio it used (OpenAI, N.D.). Unfortunately, Māori language data is scarce due to historical factors like colonization and modernisation of languages to English.

This poses a threat to protecting Māori's scarce cultural significance, which is a key principle of Māori Data Sovereignty. It raises a concern that Māori language data may be acquired by large companies (E.g., Google) with ease, without protection from initiatives like Te Mana Raraunga.

Te Hiku Media is a charitable media organization where they have data protection for Indigenous languages and data protection known as the Indigenous Data Sovereignty and the Kaitiakitanga license. These efforts address data ethics challenges in New Zealand and recognize stakeholders' rights to data ownership and stewardship, even if they didn't create the data themselves. Additionally, Te Hiku Media is most likely the only company with adequate data to create/train a language model and genuinely respects and cares about the data. According to Coffey, “It was enough data to build language tech for te reo Māori, the Māori language – including automatic speech recognition and speech-to-text.”(Coffey,2021).

In conclusion, the data ethics topics, and frameworks I have covered in this discussion are curation of data, human-first emphasis, cultural preservation, data ownership and sovereignty.

Part 3. Evaluating the optimized 2024 Formula One Calendar

This is a review of my proposed calendar 2024 F1 Calendar (constructed in part 2), which delves into four potential flaws I have identified in my proposal focusing on the real-world impacts.

The first flaw I have encountered is about the following: “Number 1, Abu Dhabi has a signed contract that guarantees them the final race position until ...” My optimization will face a legal problem, as it does not account for this where the final race position is at Circuit Gilles Villeneuve. Henceforth, I should add a stopping point at Abu Dhabi while optimizing the F1 schedule, like how I added the motorsport valley as the starting point for the races. This emphasizes that the binding contract must be respected and cannot be altered due to my lower carbon emission optimization.

The second flaw that will potentially affect my optimization is “Number 3, Teams are free to use whatever logistics they require to travel to a race...” This could arise as a potential problem as air & water-based transportation is unreliable and can be easily affected by the surrounding environment. Therefore, depending on the budgets the team has left and extreme weather, the routes may have to be reconsidered. Additionally, whether the transportation path exists between the circuits is another point I will have to research when providing the optimization of the F1 calendar. Considering the real-world impacts, especially regarding environmental factors, carbon emissions must be reduced to a minimum. Another point to consider is that we can choose a cleaner route which can be selected via a longer haul freight rather than a short flight, as the longer haul flights tend to be more efficient with per km emissions.

Thirdly, another point hindering my real-world use of my optimization is: “Number 4, In 2021, a cost cap of US \$140 million per season was imposed” This could become a potential flaw as my optimization path includes several long travels, which may become a financial burden for the teams. Upon viewing the 2-OPT (optimized calendar) graphical representation in part 2 of this assignment, the single trip distance between Brazil (Interlagos) and Australia (Albert Park) was approximately 16,300 km (calculated from Google Maps). The flights between the two circuit locations will most likely become a challenge in terms of cost. As a solution to this, in the optimizing process we should cluster the races in regions to minimize trip distance, reducing financial strain on the teams and creating cleaner routes. Additionally, sharing transportation methods between teams can also be cost effective.

Fourthly, the most severe flaw I have encountered is: “Number 5, In August of every year, the entire championship shuts down for at least 3 weeks” Realistically thinking, the teams’ employees will be all over the globe visiting their families instead of staying in a set location for multiple weeks, therefore my optimization of the F1 calendar will be unreasonable. As most of the teams are based in the UK, a suggestion I can make is for all the teams to board the same plane to the UK (saves cost and better for carbon emissions). This will be a simple change in the optimization code as it’s just adding a round trip. However, this suggestion will not be suitable as racers will most likely fly back in private jets.

In the big picture, the fans travelling to these F1 races will cause majority of the pollution which is out of our hands to optimize.

Finally, the overarching challenge with my F1 calendar optimization is that TSP is the basis for my tour optimizing process. The setback of utilizing this is that points can only be visited once and are not suitable for the back-and-forth travelling the F1 calendar requires. Whereas an alternative mechanism of the “capacitated vehicle routing problem” should be implemented. CVRP is a multiple-route-node-service-combination problem that finds the optimal routes for multiple vehicles visiting a set of locations, which fits best for optimizing the F1 calendar. The biggest advantage of this algorithm is that it assists teams in lowering transportation expenses by streamlining routes and minimizing the number of vehicles needed to meet demand. Resulting in cost savings (lower fuel expenditure) and an increase in efficiency (shortening travelling distance) (Upper,2023).

Bibliography:

Te Hiku Media (06/05/2022). Data Sovereignty and the Kaitiakitanga License <https://tehiku.nz/te-hiku-tech/te-hiku-dev-korero/25141/data-sovereignty-and-the-kaitiakitanga-license>

Open AI (N.D.) Robust Speech Recognition via Large-Scale Weak Supervision <https://cdn.openai.com/papers/whisper.pdf>

Google maps <https://maps.google.com/>

Coffey.D (28/04/2021) Māori are trying to save their language from Big Tech Wired <https://www.wired.co.uk/article/maori-language-tech>

Dick. G (2023). Introduction to Data Science. INFO204 Lecture Notes 16 [info204-lect16.pdf](#)

Dick. G (2023). Introduction to Data Science. INFO204 Lecture Notes 17 [info204-lect17.pdf](#)

Upper (12/05/2023). What is Capacitated Vehicle Routing Problem (CVRP) <https://www.upperinc.com/glossary/route-optimization/capacitated-vehicle-routing-problem-cvrp/#:~:text=Benefits%20of%20Using%20CVRP&text=CVRP%20systems%20can%20help%20businesses,finding%20optimal%20routes%20for%20deliveries>