Hugging Carbon: Visualising the Carbon Emissions of Hugging Face's AI Models

Kevin Yi Chen
Delft University of Technology
Student Number: 4868323

Tiberiu Sabău
Delft University of Technology
Student Number: 5544157

Floris van Veen
Delft University of Technology
Student Number: 5110831

Kah Ming Wong
Delft University of Technology
Student Number: 6348556

ABSTRACT

As artificial intelligence (AI) usage increases, so does its environmental impact, with AI models consuming significant energy and contributing to carbon emissions. However, users often lack accessible tools to visualize the sustainability of different models. This paper introduces *Hugging Carbon*, a web-based platform that visualizes the carbon emissions of AI models hosted on Hugging Face, providing transparency and enabling users to make informed, sustainability-conscious decisions. Through a user study and performance analysis, the platform's usability and effectiveness is validated, showing its potential to raise awareness about AI-related carbon footprints. Future work includes expanding the platform through a large-scale crowdsourced study, enhancing mobile accessibility, and integrating additional features as more data becomes available from Hugging Face. By promoting transparency in AI sustainability, *Hugging Carbon* plays an important role in raising awareness and driving efforts to reduce the environmental impact of AI technologies.

1 INTRODUCTION

Hugging Face stands at the forefront of innovation and collaboration in the rapidly evolving artificial intelligence (AI) and natural language processing (NLP) industry. With a $4.5 billion valuation [2], more than 1,500 companies as customers, and a market share of 30.70% [1], Hugging Face is the industry leader of providing open source NLP technologies and AI models. Millions of users visit their website, and can use any of their over 1.5 million models [6].

While Hugging Face has taken steps toward transparency regarding the environmental impact of AI, including efforts to estimate and disclose the carbon emissions of certain models, this information is not easily accessible nor understandable to non-technical users. Many individuals and businesses rely on these models for daily tasks without fully understanding their environmental footprint. AI technology consumes a significant amount of energy, indirectly contributing to carbon emissions, but the exact scale of these emissions and their long-term impact on climate change remain unclear. While some efforts exist to estimate and reduce AI-related carbon emissions [5], users are typically not presented with precise, comparative data that would allow them to weigh sustainability alongside performance. This lack of visibility means that an individual or business seeking to integrate AI into their workflows has no straightforward way to determine whether an equally effective but more carbon-efficient alternative exists. Moreover, AI users and decision-makers may often make choices without considering the environmental impact of different models, simply because this information is not readily available or easily interpretable. The lack of clear, user-friendly tools to assess and compare energy consumption and carbon emissions means that sustainability is rarely factored into the selection process. Without accessible insights, stakeholders are left unaware of potential alternatives that may offer similar performance with a lower environmental footprint. As AI usage continues to grow, it is crucial for all users to make informed decisions by considering the sustainability of the models they use. While model performance remains important, AI users must also consider the environmental impact, balancing both factors to make more thoughtful decisions.

This paper introduces *Hugging Carbon*[1], a website that visualises the carbon emissions of different AI models on Hugging Face. With this website, the tech industry gains insight into how the state-of-the-art open source AI models play a role in climate change, and it provides transparency for the wider AI community, supporting the ongoing global effort to reduce carbon emissions. Within this report, the proposed solution and implementation are discussed in detail, including the design and development of the platform. To evaluate the effectiveness of the website, two validation methods are employed: a user study to assess usability and impact, and a performance analysis to ensure accessibility and efficiency. The results of these evaluations are promising, indicating that *Hugging Carbon* successfully enhances transparency and awareness of AI-related carbon emissions and can provide users the means to make informed decisions by factoring in the sustainability of the models they use.

2 PROPOSED SOLUTION

In order to provide user-friendly visualisations, our solution consists of a publicly-hosted website with an interactive user interface. The website's main objective is to inform the general public about the relation between AI model performance and carbon emissions. Users should be able to explore, compare, and analyse the carbon emissions of the AI models available on Hugging Face.

The website should meet the following requirements:

- **Intuitive User Interface**: The user interface should be easy to navigate, and the information should be presented in a simple and concise manner that is easy to understand for non-technical people.

[1] https://kahming0.github.io/SSE2/

- **Responsive Design**: The website should be responsive to the actions of the user.
- **Accessible Design**: The website should keep user accessibility in mind.
- **Low Latency**: The website should present the data in a timely manner by efficiently fetching the data.

The website provides the following features:

- **Focus on Education**: Users should be able to look on the website and understand exactly what everything means, even if they do not have a programming background. They should be able to look for additional information as is required - links to additional information, explanations on how the data was gathered, explanations on keywords, etc..
- **Model Search**: Users should be able to look up specific AI models via a search bar in order to view their model details, such as carbon emission and performance on benchmarks.
- **Model Comparison**: Users should be able to compare a selection of AI models side-by-side based on certain metrics, such as carbon emission and benchmark performance.
- **Interactive Visualisations**: The website should provide interactive plots, graphs, and charts for the users.
- **Emission Equivalents**: The website should provide real-world comparisons for the carbon emissions of the AI models (e.g., emissions equivalent to km driven by a petrol car) to help the users understand the scale of the carbon emissions.

3 IMPLEMENTATION

The implementation of the website brings together the design and technical decisions made throughout the development cycle. The decisions are aimed at creating an educative and user-friendly website, which visualises the CO_2 emissions of AI models. This section elaborates first on the website's layout and accessibility considerations, and afterwards dives into the development stack, data processing, and main features offered by the website.

3.1 Initial Website Design

In order to meet all target requirements, a target design was created. This design would be flexible and improvements could be freely implemented, but some targets were laid out. These goals would give an outline to features that would be added, provide a layout for usability, and accessibility features.

Accessibility in this context includes decisions made to allow as many users as possible to visit the website. Time would need to be taken into account however, and as such only relatively simple solutions would be used, such as not solely relying on colour to distinguish differences, or ensuring that the layout of the website is intuitive to all users. Additional accessibility options, such as auditory assistance was not possible due to time constraints.

The initial design can be seen in figure 1. It was decided that the main page would include all the graphs and comparisons, along with some basic information that explains what everything means. The goal was to ensure that any user who visits the page will be able to understand everything that is displayed on the website. This page should be clean, and allow the user to quickly and easily navigate to any comparison they want to see. The information at the top of the website should be unobtrusive, but still clearly provide information to those who wish to see it.

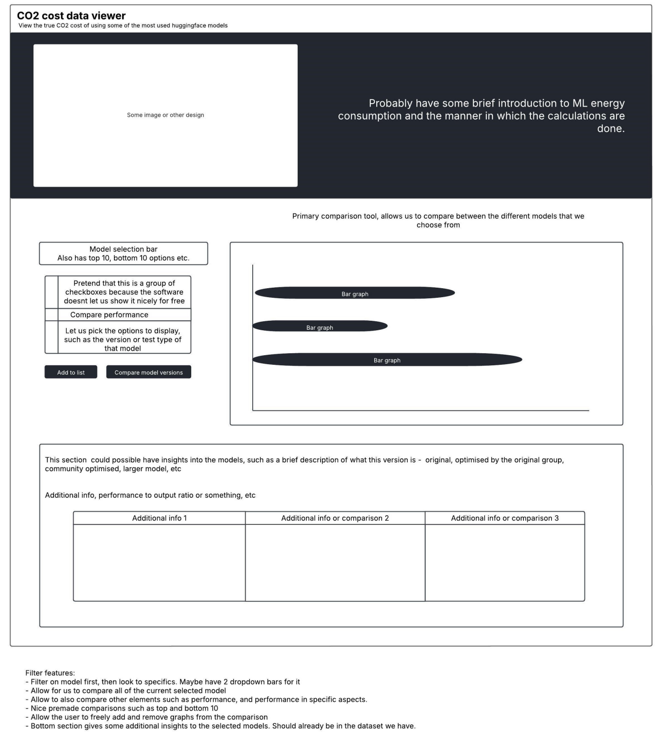


Figure 1: Initial layout of the website

The most important features, the model comparisons, should be in the centre of the website. A number of features should be available, which include comparing the models' CO_2 emissions, comparing the models' benchmark performances, viewing the models' efficiencies, given by the performance per kg of CO_2 released, top and bottom 10 models displays, and a display to easily compare a model's performance vs. CO_2 cost.

Additional information about the selected models is displayed below the comparison graphs. This section provides more details on the carbon emissions, and puts them into perspective by showing real-world equivalence values, such as the number of smartphones charged. These comparisons help the users to better understand the environmental impact these AI models have, and encourage them to consider more environmentally friendly alternatives, which is the aim of the website.

3.2 Website Development

The website is implemented as a static website, and is hosted via GitHub Pages, since it is a simple, accessible, and free option. The front-end is built using React, providing a highly interactive user interface. For the scope of the project, it was set up to be entirely front end based as opposed to making several requests to a back-end API.

The dataset used for the website is Hugging Face's Open LLM Leaderboard[2], which provides the details of all the models on Hugging Face, such as the model performance and carbon emissions. The dataset is taken by parsing the parquet file from the Hugging Face repository. This has the advantage of not requiring constant requests to a back-end and simplifying the website creation process, but makes it more difficult to update when HF makes an update.

Core functionality, such as being able to add models from the available dataset, was implemented first. The visualisation components were developed using the Recharts[3] library, allowing users to hover over the graphs for more detailed information. Four main screens were created: a basic CO_2 comparison, a benchmark comparison display, a ranking of the best and worst models, and a scatter plot displaying every available model's performance vs. CO_2 emissions.

The CO_2 comparison was developed to directly compare the CO_2 cost of each model, and potentially look for better alternatives. This could be seen in a pop-up that appears when a user click on any graph. This pop-up can be seen in Figure 3. This directly follows our goal of educating users of the CO_2 cost of desired models, and potentially offers alternatives with lower CO_2 cost.

Figure 2 shows a comparison between different benchmarks. This displays to users the performance of different models, and they can toggle performance divided by CO_2 cost. This allows users to make more informed decisions between the use cases and efficiencies of their desired models.

The top and bottom 10 emission screen is important to inform users of the range of emissions that these models provide. A desired improvement is to display the efficiencies and other "top 10's".

The final screen is the scatter plot of CO_2 cost vs. performance. This allows users to see the distribution of performance and emissions in one simple display. Despite previous goals, certain elements such as distinguishing different elements by aspects other than just colour could not be done given the time constraints. This display, along with the other images of the website, can be seen in Appendix A.

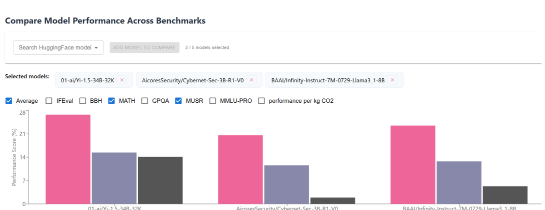


Figure 2: Comparison between different benchmarks.

4 VALIDATION

Three criteria were chosen for the validation of the website, being usability, performance, and accessibility. These criteria will determine whether the website will have a successful launch, and will be assessed using different validation methods, which are elaborated on below.

[2]Hugging Face Open LLM Leaderboard: https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard#/

[3]Recharts Library: https://recharts.org/

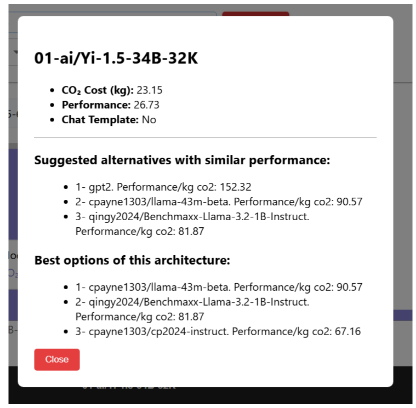


Figure 3: Display of pop-up suggestions.

4.1 User Pilot Study

A small pilot study was conducted in order to test the usability of the website. The main objectives of pilot study was to test whether the website is easy to use and navigate, and presents the information to the user in a clear and concise manner, which are key requirements of the website as mentioned in Section 2.

The pilot study included a sample size of three participants, all of whom are friends of the researchers. They are fluent in English, and two of the three participants study at a non-technical university. Their educational backgrounds are Econometrics and International Business respectively, and the other participant studies Maritime Engineering at Delft University of Technology. The participant with the International Business background has some familiarity with AI models and sustainability due to a specialisation in the innovation of AI technologies. The other participants have limited knowledge of AI and their CO_2 emissions.

At the start of the pilot study, the participants of the pilot study were first informed about the website and why it was made, along with the purpose of the experiment through a consent form. The informed consent can be found in Appendix B. Afterwards, the participants received two tasks that they needed to complete with no time limit in order to prevent stressors, such as time pressure. The two task descriptions given to the participants can be found in Appendix C, and here is summarised version of the two tasks:

- **Compare CO_2 Emissions**: The participant is asked to search for two specific models, compare the details of the models, and choose the model with the lower CO_2 emissions.
- **Compare CO_2 Emissions & Benchmarks**: The participant is asked to search for two specific models, compare the models based on specific benchmarks, and choose the model with the lower CO_2 emissions.

The two tasks were designed to evaluate the two different use cases of the website. The first task asks the participant to compare two different models based solely on their CO_2 emissions, which reflects the use case where users without any knowledge about AI models make an environmentally-conscious decision. The user is curious only if model A or model B is better for the environment from a surface-level perspective without diving into the finer details. The benchmarks are not taken into consideration, because the models might serve entirely different purposes. Comparing benchmarks between such models would not make sense, which is why the first task solely focuses on the environmental impact, disregarding the models' performance.

However, the second task asks the participant to compare two very similar models that do serve the same purpose. The participant has to look into the model benchmark performances of the models as well as the CO_2 emissions. The task involves making a decision that balances both performance and environmental impact, reflects the use case where users with some but limited knowledge try to make an informed decision.

After completing the tasks, the participants fill in a questionnaire, asking them questions regarding the usability, navigability, and understandability of the website and its content. The full questionnaire can be found in Appendix D. The majority of the questions asked are about the usability of the website, and use a numerical scale with 1 being the worst, and 5 being the best score possible. The other questions have no connection with the usability of the website, and are marked with an asterisk. The results can be found in Table 1, and will be discussed in more depth in Section 5.1.

4.2 Performance & Accessibility Audit

The website's performance and accessibility will be validated through the usage of Google Lighthouse[4], which is an open source tool that helps with auditing and improving the quality of web pages. Google Lighthouse performs audits in the categories of performance, accessibility, best practices, and SEO. Each category receives a score out of 100 points. The motivation behind the usage of Google Lighthouse is check whether the website meets the performance and accessibility requirements mentioned in Section 2. The website is accessed via local host using the Google Chrome browser, since Google Lighthouse is built into Chrome DevTools. It analyses the page load, performs various audits for 30 to 60 seconds, and generates an audit report. Google Lighthouse's performance audit [4] is comprised of the following metrics:

- **First Contentful Paint (FCP)**: How long it takes the browser to render the first piece of DOM content after a user navigates to your page.
- **Speed Index**: How quickly content is visually displayed during page load.
- **Total Blocking Time**: The total amount of time that a page is blocked from responding to user input, such as mouse clicks, screen taps, or keyboard presses.
- **Largest Contentful Paint**: When the largest content element in the viewport is rendered to the screen.
- **Cumulative Layout Shift**: The largest burst of layout shift scores for every unexpected layout shift that occurs during the entire lifecycle of a page. A layout shift occurs any time a visible element changes its position from one rendered frame to the next.

Google Lighthouse's accessibility audit [3] is a weighted average of all accessibility audits, where the weights are based on axe user impact assessments [7]. Each accessibility audit is either a pass or fail, as opposed to the performance audit, which uses a point-based system. If a website partially passes an accessibility audit, it is still considered as a fail.

[4]Google Lighthouse: https://developer.chrome.com/docs/lighthouse/

5 RESULTS

5.1 Pilot Study Questionnaire

Given the scores in Table 1, the website performs decently, with all the scores being larger or equal to 3.33, when excluding the non-usability questions marked with an asterisk. Six out of the eight questions regarding the usability of the website in the General Experience and Website Navigation & Features categories have a score larger or equal to 4.0, meaning the website was easy to use in most use cases. However, the lower scores of the remaining two questions regarding usability suggest that the users had some trouble with the benchmarks, real-world comparisons, or both. Overall, the above-average scores indicate that the user interface is intuitive, and the website is responsive to the user, meaning the first two requirements mentioned in Section 2 are met.

The three questions regarding the educative value and content understanding are all larger or equal to 3.33. The score of 4.0 for the question regarding how accessible the website is to non-technical users, suggests that the website is inclusive, and accessible to a broader audience. The lower scores for the other two educative questions should be a key issue that needs to be addressed. However, it should be noted that one of the participants possessed more knowledgable regarding AI models and emissions due to their educational background, as mentioned in Section **??**, which resulted in said person filling in a 2 (a little) for the question regarding how much the participant learned about CO_2 emissions after the experiment. The other two participants filled in a 5 (a great deal) and 3 (a moderate amount), and both participants had limited knowledge about CO_2 emissions of AI models. Therefore, it could be argued that the website does provide great educational value, especially for users with limited domain knowledge.

The overall impression on the website from the participants was very positive, since they all understand the significance of AI models in today's world, while also acknowledging that the rise of generative AI comes at the cost of rising energy consumption, and consequently higher CO_2 emissions. One participant was not knowledgable at all regarding the CO_2 emissions of AI models, meaning the experiment and limited interaction with the experiment provided the participant with a large amount of new information and knowledge. The participant really appreciated that the models they were comparing were aligned next to each other for a better overview. Another participant, who is somewhat knowledgable in the field of AI models and CO_2 emissions, enjoyed the inclusion of multiple benchmarks to compare the AI models, since most models are not a one-size-fits-all solution, but rather specialise in certain domains, such as mathematics or complex reasoning.

5.2 Google Lighthouse Audit

Figure 4 shows the score overview in each of the mentioned categories, and the full audit report can be found in Appendix E.

Question	Average Score
General Experience	
How easy was the website to use overall?	4.00
How easy was it to complete the 1st task (Compare CO_2 Emissions)?	4.67
How easy was it to complete the 2nd task (Compare CO_2 Emissions & Benchmarks)?	3.50
*How clear were the instructions of the 1st task?	5.00
*How clear were the instructions of the 2nd task?	4.00
Website Navigation & Features	
How easy was it to find the models using the search function?	4.00
How easy was it to filter the models based on model type?	4.00
How easy was it to sort the models based on performance?	4.50
How easy were the visualisations (charts/graphs) to understand?	5.00
How easy was it to understand the carbon emissions data shown on the website with the help of real-world comparisons?	3.67
Educational Value & Content Understanding	
*Before using the website, how familiar were you with the concept of carbon emissions from AI models?	3.00
After using the website, how much did you learn about carbon emissions from AI models?	3.33
Did the website help you better understand the environmental impact of AI?	3.67
Did you feel the website content was accessible to non-technical users?	4.00

Table 1: Score overview of the pilot study's questionnaire.

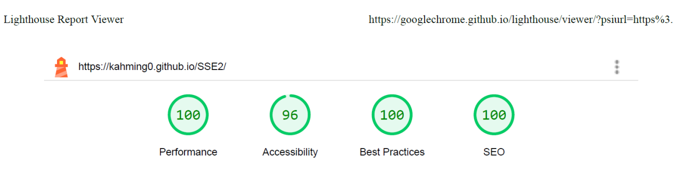


Figure 4: Score overview of the Google Lighthouse audit report.

Figure 6 displays more details on performance part of the audit report. The perfect score in this category indicate that the website performs up to standard in terms of handling payloads and responding in a timely manner, providing a great experience to the user. The 0.2 s under the First Contentful Paint metric indicates that it only takes 0.2 seconds to show the first useful component to the user when they first visit the website. The Largest Contentful Paint shows that the largest component of the website is fully loaded after 0.6 seconds. The Total Blocking Time of 70 ms means that the website is only blocked and not responsive to the user for 70 milliseconds. The 0 under the Cumulative Layout Shift indicates that there are no layout shifts present on the website that could disturb the user experience. The 0.7 s in the Speed Index shows that it only takes 0.7 seconds for the full website to be visually displayed. All in all, the audit report strongly indicates that the website has incredibly low loading times to render, presents the data in a timely manner to the users, and responds to the user in near real-time. All this properties support that the website meets the third and fourth requirement mentioned in Section 2.

The near-perfect score in the accessibility category underscores that the website has been carefully designed with accessibility in mind, as stated in the requirements of Section 2. However, there is still room for improvement, since the website did not score perfectly in this category. The audit report indicates an issue with the sizing and spacing of the component containing more information at the top of the website. Figure 5 highlights the specific area that needs to be addressed in order to provide the best possible experience to all users. By creating more space between the drop-down menus, and increasing the size of the drop-down menu buttons, the website could score a perfect 100 in this category as well.

The two perfect scores in the best practices and SEO categories are promising results as well. The perfect score for best practices emphasises that the website's design and implementation are up to the modern Web development standards. The perfect SEO score suggests that the website would show up near the top of search engine result pages, gaining more visibility, and creating more awareness for the CO_2 emission of AI models.

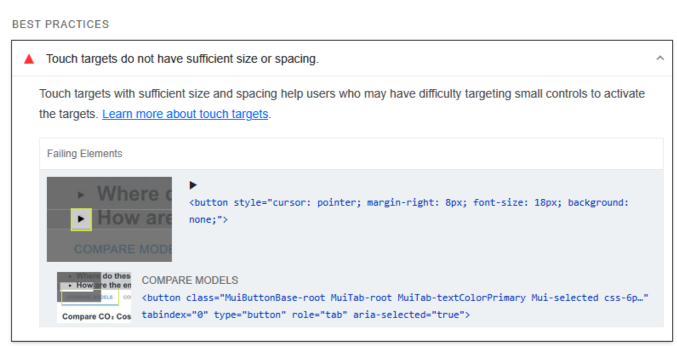


Figure 5: Accessibility improvement of the Google Lighthouse audit report.

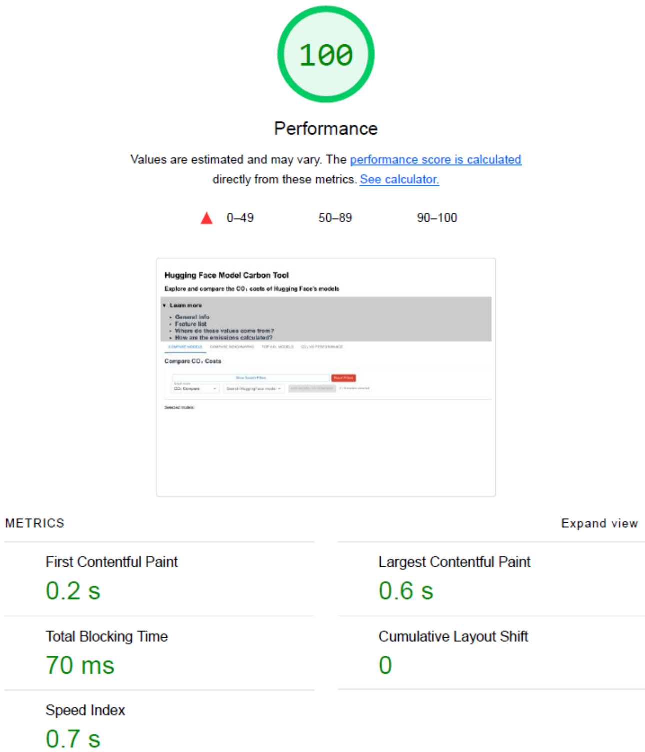


Figure 6: Performance of the Google Lighthouse audit report.

6 DISCUSSION

From our experience building this tool, we have noticed that a limited amount of Hugging Face repositories report any information regarding energy-efficiency and carbon emission. Through manually searching, the amount of repositories that reported their carbon emission and made it easily accessible, was exactly one. This is quite concerning in the perspective of energy-awareness, given the fact that Hugging Face is the *melting pot* of AI, where models of all shapes, sizes, and philosophies converge. Enabling researchers and developers to experiment and share. ChatGPT, one of the most popular generative AI tool, too, put no effort in making their users aware of the impact their usage has on the environment. Many people, including some of our study participants, accept the notion that AI models is bad for the environment, and, furthermore, that both training and prompting consumes a lot of energy. However, no clear answer were given when asked for an estimate. Another lingering issue includes the sentiment that any talk regarding CO_2 was a "waste of time" and that the value of AI models far outweighs their energy usage. Though this may hold some truth, it neglects to acknowledge the impact the energy usage has on the environment, disregarding any sustainable practice.

One key challenge encountered during the development of *Hugging Carbon* is the heterogeneity of AI models. Models vary widely in their architecture, the data used for the training, the training phase itself, and their evaluations. This makes it difficult for AI models to be directly compared, since AI models could serve different purposes. Even with the comparisons in CO_2 emissions and the multiple benchmarks, it does not show the full picture. Users with limited domain knowledge would not immediately recognise the nuances that exist when comparing different AI models. The performance scores and CO_2 emissions need to be contextualised in order to provide relevant comparisons. In order to create more fair comparisons between AI models, and increase the awareness of the CO_2 emissions of AI models, future tools should include the categorisation of models based on their intended purpose. This would allow the wider AI community to gain a more detailed understanding of the CO_2 emissions in specific use cases, increasing the already rising innovation in the domain of sustainable AI.

The pilot study results indicate that there is a need for tools, such as *Hugging Carbon*, that allow non-technical people to gain more insight into the sustainability side of AI technologies. AI is getting implemented everywhere nowadays, but little attention is paid to what the impact of all these AI implementations has on the environment right now, and in the future. It is mentioned in the news every day, but for the average non-technical person, who uses AI for their homework, job application, or gym routine, they only interact with one side of the AI coin. With the increasing usage of AI technologies, and the nearing deadline to stay below the tipping point with regards to the temperature increase and climate change, even more emphasis should be put onto the sustainable aspect in order to create more efficient AI models in the future, curb the increasing energy consumption of AI technologies, and mitigate climate change.

7 LIMITATIONS

The main limitation is the lack of integration with Hugging Face's API, meaning the website is not able to display the most up-to-date data in real-time to the users. Using Hugging Face's API, only the first 100 models were fetched, which is a small fraction of their over 1.5 million models. The website would not be useful at all in this situation, which is why we opted to parse all the models via the parquet file manually. This way the website will be slow during the parsing process, but fast for every request that follows. To solve this issue, a proper integration with Hugging Face's API needs to be created, which would allow the website to fetch the newest data whenever a new model is added to Hugging Face. This way users are able to observe the performance, CO_2, and other details of the newest and most trending AI models. Another limitation of the current implementation is that if any data is updated on Hugging Face, the parquet file needs to be downloaded and parsed again, followed by a new deployment of the website. This provides more problems to maintainability than a simple API call would.

Another limitation is that the focus was put on desktop users during the development of the website. The overall design of the website was created using the dimensions of desktops and laptops, which impacts the usability and navigability for mobile users. A key improvement to enhance the mobile responsiveness, and redesign the website layout would serve as another future work. Areas to focus on in particular would be font sizes, chart layouts, and navigation menus, which would ensure a pleasant and enjoyable experience for users on smaller devices. This would make the website more inclusive and accessible to a broader audience.

Furthermore, the pilot study involved a relatively small sample size of three participants. Due to the time constraints, it was not possible to increase the scope of the pilot study. The results of the questionnaire and feedback from the participants are valuable regardless of the sample size. However, the experiment needs to have a larger group of participants in order to gather more representative results.

8 CONCLUSION & FUTURE WORK

The development of *Hugging Carbon* marks a significant step toward improving transparency around AI-related carbon emissions. By providing an accessible platform that visualises the environmental impact of AI models, the users are enabled to make more informed decisions regarding model selection by factoring in sustainability. The validation studies conducted indicate that our tool is both effective and user-friendly, reinforcing the need for sustainability considerations in AI workflows. As AI usage continues to increase, tools such as *Hugging Carbon* will play an important role in promoting responsible and sustainable AI usage.

Moving forward, there are several key areas for improvement. If we continue as a team, our next steps would focus on refining and expanding the platform to improve its performance and usability. Individual examples would be to introduce more areas where users can get suggestions of better models to make use of, adding more instances to the 'top 10' page, and expanding on the additional information on the bottom of the page. Another important step is transitioning from the initial pilot study to a large-scale crowdsourced study. By gathering user feedback from a diverse and unbiased set of AI practitioners and stakeholders, we can identify areas where the platform needs improvement—whether in data representation, model comparison methods, or the overall user experience. Another important enhancement is extending *Hugging Carbon* beyond desktop support to provide a fully responsive mobile experience, since that will increase accessibility for the users. Furthermore, the website can be expanded to offer additional features for users, depending on both community needs, and the availability of new data from Hugging Face in the future.

All in all, *Hugging Carbon* serves as a foundation for greater transparency in AI sustainability. With continued development, user feedback, and collaboration with the AI community, this platform can evolve into a comprehensive resource that not only informs but also drives meaningful action toward reducing the environmental impact of AI technologies and increasing sustainability.

REFERENCES

[1] 6sense. 2025. *Hugging Face Market Share.* https://6sense.com/tech/nlp-and-text-analytics/hugging-face-market-share Accessed: 2025-03-20.

[2] CNBC. 2023. *Google, Amazon, Nvidia and other tech giants invest in AI startup Hugging Face, sending its valuation to $4.5 billion.* https://www.cnbc.com/2023/08/24/google-amazon-nvidia-amd-other-tech-giants-invest-in-hugging-face.html Accessed: 2025-03-20.

[3] Google Developers. 2019. *Lighthouse accessibility score.* https://developer.chrome.com/docs/lighthouse/accessibility/scoring Accessed: 2025-03-31.

[4] Google Developers. 2019. *Lighthouse performance scoring.* https://developer.chrome.com/docs/lighthouse/performance/performance-scoring Accessed: 2025-03-31.

[5] Hugging Face. 2025. *AI Energy Score Leaderboard.* https://huggingface.co/spaces/AIEnergyScore/Leaderboard Accessed: 2025-03-20.

[6] Hugging Face. 2025. *Hugging Face Models.* https://huggingface.co/models Accessed: 2025-03-20.

[7] Deque Labs. 2025. *axe-core: Rule Descriptions.* https://github.com/dequelabs/axe-core/blob/develop/doc/rule-descriptions.md Accessed: 2025-03-31.

A UNDISPLAYED WEBSITE ELEMENTS

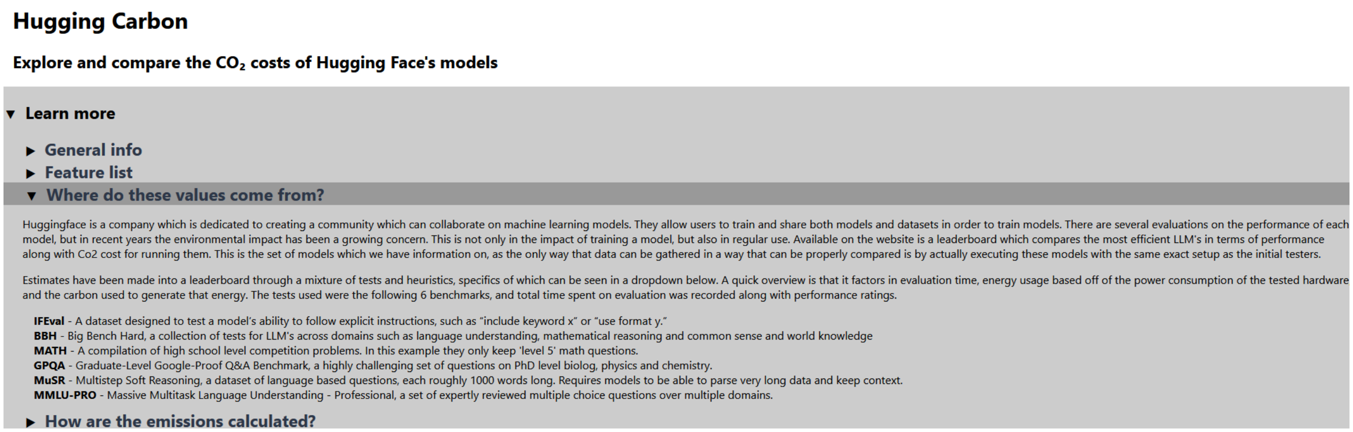


Figure 7: Display of the information section.

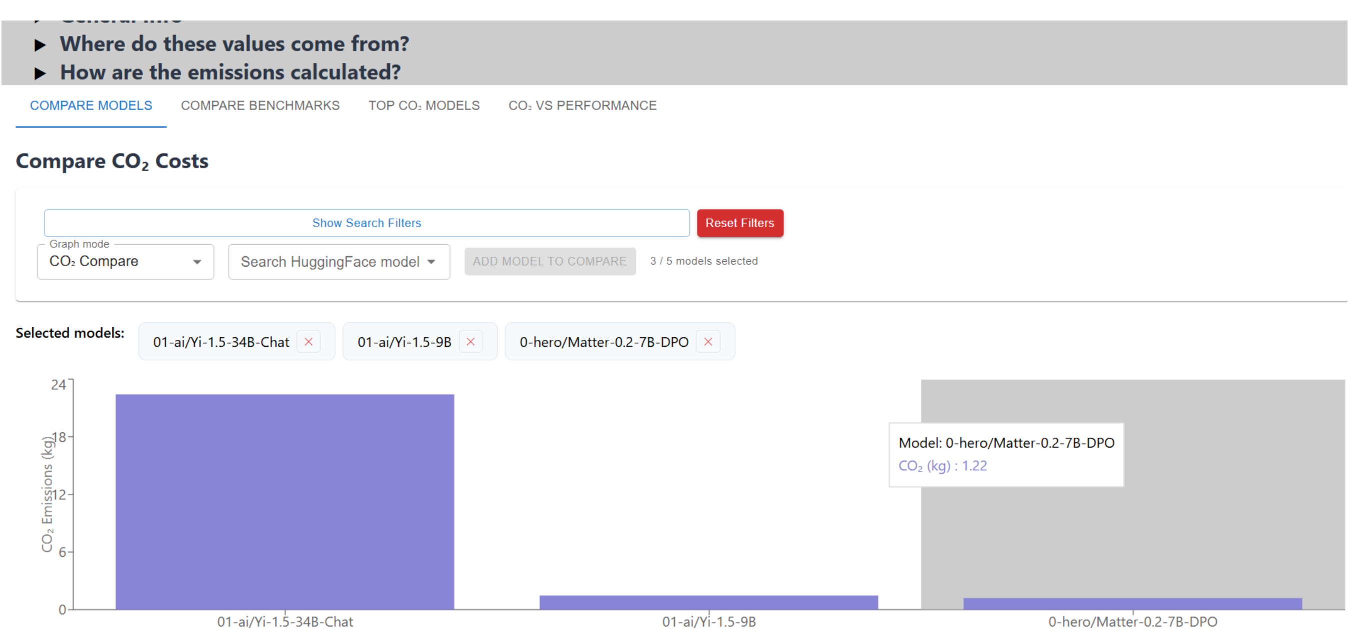


Figure 8: Display of the basic graph comparison.

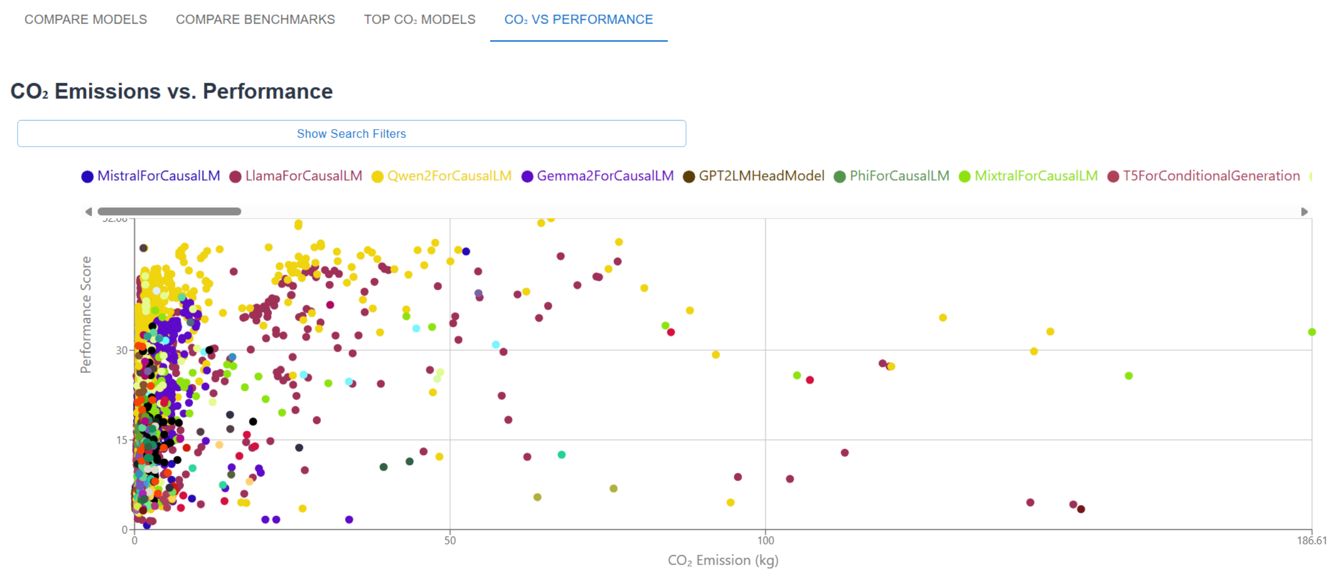


Figure 9: Display of the scatter plot.

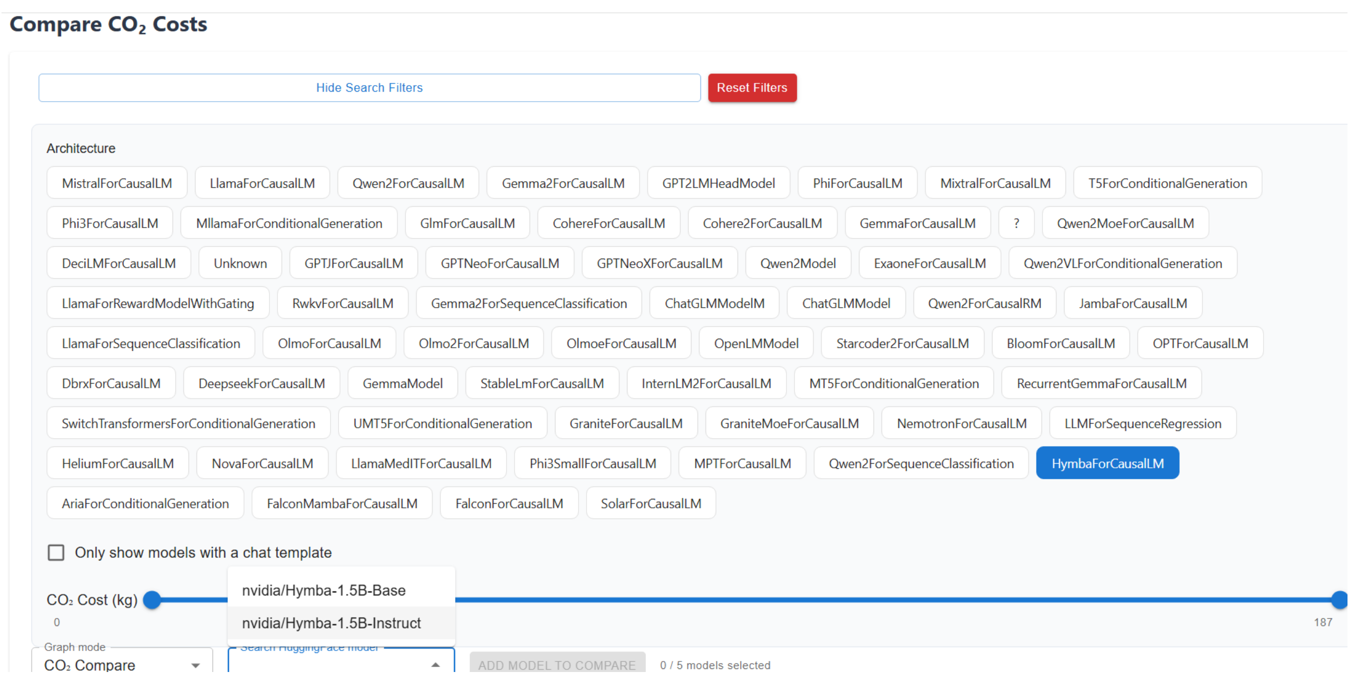


Figure 10: Display of the available filter options.

B PILOT STUDY: INFORMED CONSENT

Informed Consent

You are being invited to participate in a research study for a university course. The study aims to investigate the usability of an artificial intelligence (AI) visualisation website. Before deciding whether to participate, it is important for you to understand the purpose of the study, what it will involve, and how your data will be handled. Please read the following information carefully.

Purpose & Tasks:
The purpose of this study is to determine how user-friendly and effective the visualisation website is in presenting the AI models and its details to the user. During the experiment, you will complete two tasks, which involves interacting with the website.

Confidentiality:
All data collected during the experiment will be treated as confidential and stored securely. Your responses will be recorded anonymously, and no personally identifying information will be collected.

Voluntary Participation:
Your participation in this study is completely voluntary. You are free to withdraw at any time without penalty and without providing a reason.

Figure 11: Informed consent given to the participant of the pilot study.

C PILOT STUDY: TASKS

Task 1: Compare CO2 Emissions

1. Search for the models **AI4free/t2** and **BoltMonkey/DreadMix**.
2. Compare the two models.
3. Choose the model with the lower CO2 emissions (kg).

Figure 12: Task 1 given to the participant of the pilot study.

Task 2: Compare CO2 Emissions & Benchmarks

1. Search for the models **Qwen/Qwen1.5-110B-Chat** and **Qwen/Qwen2-57B-A14B-Instruct**.
2. Compare the two models based on the following benchmarks: **Average**, **BBH**, and **MATH**.
3. Choose the model with the lower CO2 emissions (kg).

Figure 13: Task 2 given to the participant of the pilot study.

D PILOT STUDY: QUESTIONNAIRE

Questionnaire

1) General Experience

How easy was the website to use overall?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

How easy was it to complete the 1st task (Compare CO2 Emissions)?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

How easy was it to complete the 2nd task (Compare CO2 Emissions & Benchmarks)?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

How would you rate the clarity of the 1st task's instructions?

- ☐ Very unclear
- ☐ Unclear
- ☐ Neutral
- ☐ Clear
- ☐ Very clear

How would you rate the clarity of the 2nd task's instructions?

- ☐ Very unclear
- ☐ Unclear
- ☐ Neutral
- ☐ Clear
- ☐ Very clear

2) Website Navigation & Features

How easy was it to find the models using the search function?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

Figure 14: Page 1 of the questionnaire filled in by the participant of the pilot study.

How easy was it to filter the models based on model type?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

How easy was it to sort the models based on performance?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

How easy were the visualisations (charts/graphs) to understand?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

How easy was it to understand the carbon emissions data shown on the website with the help of real-world comparisons?

- ☐ Very difficult
- ☐ Difficult
- ☐ Neutral
- ☐ Easy
- ☐ Very easy

3) Educational Value & Content Understanding

Before using the website, how familiar were you with the concept of carbon emissions from AI models?

- ☐ Not at all familiar
- ☐ Slightly familiar
- ☐ Moderately familiar
- ☐ Very familiar
- ☐ Extremely familiar

After using the website, how much did you learn about carbon emissions from AI models?

- ☐ Nothing
- ☐ A little
- ☐ A moderate amount
- ☐ A lot

Figure 15: Page 2 of the questionnaire filled in by the participant of the pilot study.

☐ A great deal

Did the website help you better understand the environmental impact of AI?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Very
☐ Extremely

Did you feel the website content was accessible to non-technical users?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Very
☐ Extremely

4) Feedback

What did you find most useful or effective about the website?

___

What did you find most confusing or difficult to use?

___

Do you have any suggestions for improving the website?

___

Figure 16: Page 3 of the questionnaire filled in by the participant of the pilot study.

E GOOGLE LIGHTHOUSE AUDIT REPORT

https://kahming0.github.io/SSE2/

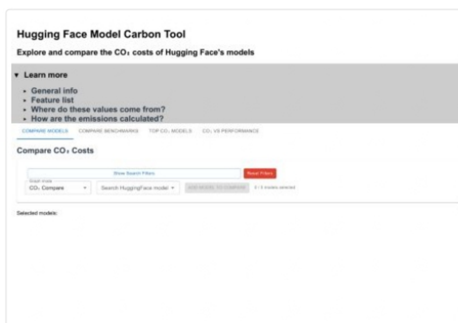
100	96	100	100
Performance	Accessibility	Best Practices	SEO

100

Performance

Values are estimated and may vary. The performance score is calculated directly from these metrics. See calculator.

0–49 50–89 90–100



METRICS Expand view

First Contentful Paint
0.2 s

Largest Contentful Paint
0.6 s

Total Blocking Time
70 ms

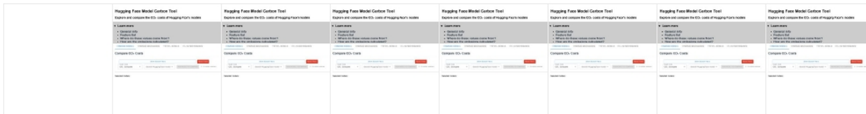
Cumulative Layout Shift
0

Speed Index
0.7 s

View Treemap

Figure 17: Page 1 of the Google Lighthouse audit.

Lighthouse Report Viewer https://googlechrome.github.io/lighthouse/viewer/?psiurl=https%3...



Show audits relevant to: All FCP LCP TBT CLS

DIAGNOSTICS

- ▲ Reduce unused JavaScript — Potential savings of 145 KiB
- Enable text compression — Potential savings of 122 KiB
- Avoid serving legacy JavaScript to modern browsers — Potential savings of 0 KiB
- Serve static assets with an efficient cache policy — 2 resources found
- ○ Avoid long main-thread tasks — 4 long tasks found
- ○ Avoid chaining critical requests — 1 chain found
- ○ Minimize third-party usage — Third-party code blocked the main thread for 0 ms
- ○ Largest Contentful Paint element — 560 ms
- ○ Avoid non-composited animations — 4 animated elements found

More information about the performance of your application. These numbers don't directly affect the Performance score.

PASSED AUDITS (27) Show

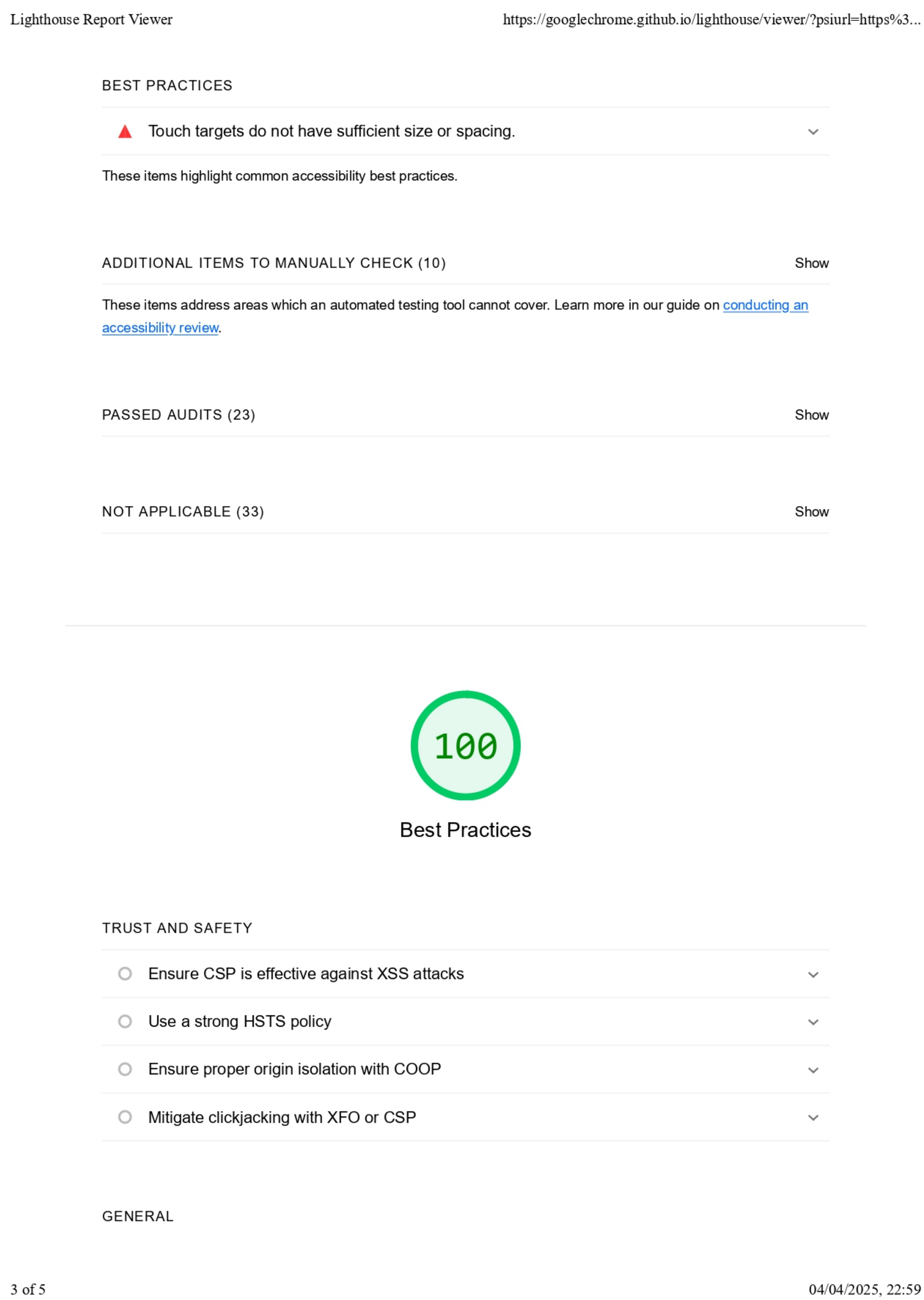
96

Accessibility

These checks highlight opportunities to improve the accessibility of your web app. Automatic detection can only detect a subset of issues and does not guarantee the accessibility of your web app, so manual testing is also encouraged.

2 of 5 04/04/2025, 22:59

Figure 18: Page 2 of the Google Lighthouse audit.



Lighthouse Report Viewer https://googlechrome.github.io/lighthouse/viewer/?psiurl=https%3...

BEST PRACTICES

▲ Touch targets do not have sufficient size or spacing.

These items highlight common accessibility best practices.

ADDITIONAL ITEMS TO MANUALLY CHECK (10) Show

These items address areas which an automated testing tool cannot cover. Learn more in our guide on [conducting an accessibility review](#).

PASSED AUDITS (23) Show

NOT APPLICABLE (33) Show

100

Best Practices

TRUST AND SAFETY

○ Ensure CSP is effective against XSS attacks

○ Use a strong HSTS policy

○ Ensure proper origin isolation with COOP

○ Mitigate clickjacking with XFO or CSP

GENERAL

3 of 5 04/04/2025, 22:59

Figure 19: Page 3 of the Google Lighthouse audit.

Lighthouse Report Viewer https://googlechrome.github.io/lighthouse/viewer/?psiurl=https%3...

Detected JavaScript libraries

PASSED AUDITS (14) Show

NOT APPLICABLE (2) Show

100

SEO

These checks ensure that your page is following basic search engine optimization advice. There are many additional factors Lighthouse does not score here that may affect your search ranking, including performance on Core Web Vitals. Learn more about Google Search Essentials.

ADDITIONAL ITEMS TO MANUALLY CHECK (1) Show

Run these additional validators on your site to check additional SEO best practices.

PASSED AUDITS (7) Show

NOT APPLICABLE (3) Show

Captured at Apr 4, 2025, 10:53 PM GMT+2

Emulated Desktop with Lighthouse 12.4.0

Single page session

4 of 5 04/04/2025, 22:59

Figure 20: Page 4 of the Google Lighthouse audit.

Initial page load | Unknown | Using HeadlessChromium 134.0.6998.165 with lr

Generated by **Lighthouse** 12.4.0 | File an issue

Figure 21: Page 5 of the Google Lighthouse audit.