Visualising global AI state actors

Data Visualization Project Milestone 1
Supervised by: Pr. Laurent Vuillon, EPFL
March 29, 2024

GROUP MEMBERS

Quentin	Malo	Anne
Esteban	Ranzetti	Silvestre de Sacy

I. DESCRIPTION OF THE PROJECT (PROBLEMATIC)

Given the recent development of the AI industry and its apparent impact on our society, it is relevant to see which countries are the current leading actors and why. The dataset we are using is from the article "Artificial Intelligence on the World Stage: Dominant Players and Aspiring Challengers" [1].

The main objective of the project is to make an interactive map to efficiently show the different factors that impact a country's capabilities in the AI sector and hence ease the decision-making process in this industry. In the storytelling we want to show which countries form leading cliques, with a great importance on the "why?". Our target audience is not only the general public, but individuals from inside the sector, and institutions or governmental agencies likely to have an interest in AI. This is why we aim for a sleek, modern, and professional look.

II. DATASET

Link to AI Index Dataset (Kaggle)

The dataset we plan to use is the AI Global Index by Tortoise Media. In addition to the overall index, it provides more detailed information about three main features of the 62 countries it analyzes: AI implementation, innovation, and investment. To do so it combines 143 indicators into seven indexes in a very comprehensive way. In our analysis, we may want to weigh these indicators differently to put more emphasis on certain aspects than others(it should require some additional work though). The dataset only including 62 countries may seem not enough but all the main protagonists are there and it could help us not to overwhelm our public.

A big advantage of this dataset is that it is very clean. It almost can be used as is It has a usability score of 10 out of 10 on Kaggle. While the completeness and credibility ensure us that the data are relevant, usability allows us to load and interpret them easily, without much preprocessing. Another very convenient feature is that the variables are already normalized on a 0 to 100 scale.

We would like to point out that we don't necessarily limit ourselves to this dataset. These are the fundamentals of our work but according to the development of the project we could add other datasets, especially approaching geopolitics.

III. EXPLORATORY DATA ANALYSIS

In this section, we perform some basic data exploration. We can directly observe that the data will allow us to build interesting statistics. Figure 1 shows the top 5 leading state actors globally. What is interesting is that using the dataset we can construct a ranking of leading regions globally, as shown in Figure 2. Finally, we give some more general statistics based on each individual class on which the total score depends, as shown in Figure 3. Overall, we can see that the individual classes do not have centered means and that there is fluctuating intra-class variance. We will therefore need to communicate this to the target audience. We can also see that coupling this data with another dataset could be very interesting. It is important to note that not all countries of the world are included in the dataset, as we only have 62 entries.

---

[1] intersog.com/blog/ai-dominant-players-and-aspiring-challengers/

Rank	Country	Total score
1	United States of America	100.00
2	China	62.92
3	United Kingdom	40.93
4	Canada	40.19
5	Israel	39.89

Fig. 1: Top 5 countries by index score

Rank	Region	Average Region Score
1	Americas	29.031250
2	Asia-Pacific	25.792143
3	Europe	25.493103
4	Middle East	19.656667
5	Africa	6.426000

Fig. 2: Global regions ranked by average score

IV. RELATED WORK

Given the leading theme of our project, we found the following works from which we could draw some inspiration.

First, the AI Global - 'Where in the World is AI?' [2] map, also showing data related to AI. This map shows the countries that are using AI in various domains between 2005 and 2021. This map is interesting because it has different colours to show the different domains which we also want to do to have a clear visual distinction between the different indicators we will represent. Figure 4

Then, the 'Global Conflict Tracker'[3] map. This map doesn't show data related to AI but to the different conflicts in the world. This map is interesting for its design. The use of soft colours for the map enables to highlight key information with darker tones. Compared to the first one it looks more professional which is more relevant to our project. Figure 5

This dataset has mainly been used with machine learning prediction algorithms. Our project will be more informative and build on data to highlight the information it gives.

---

[2] map.ai-global.org

[3] cfr.org/global-conflict-tracker

	Talent	Infrastructure	Operating Environment	Research
Mean Value of the index	16.80	63.50	66.93	16.61
Median Value of the index	13.45	65.23	69.50	12.93
Standard Deviation of the index	15.21	20.22	20.00	17.41
Country with the highest value (100)	United States of America	China	Saudi Arabia	United States of America
Country with the lowest value (0)	Qatar	Nigeria	Egypt	Colombia

	Development	Government Strategy	Commercial	Total score
Mean Value of the index	14.82	57.87	6.17	23.91
Median Value of the index	9.00	63.93	2.58	23.22
Standard Deviation of the index	19.42	26.25	14.03	15.12
Country with the highest value (100)	United States of America	Canada	United States of America	United States of America
Country with the lowest value (0)	Qatar	South Africa	Qatar	Pakistan

Fig. 3: Statistics for each class that the total score depends on given in the dataset
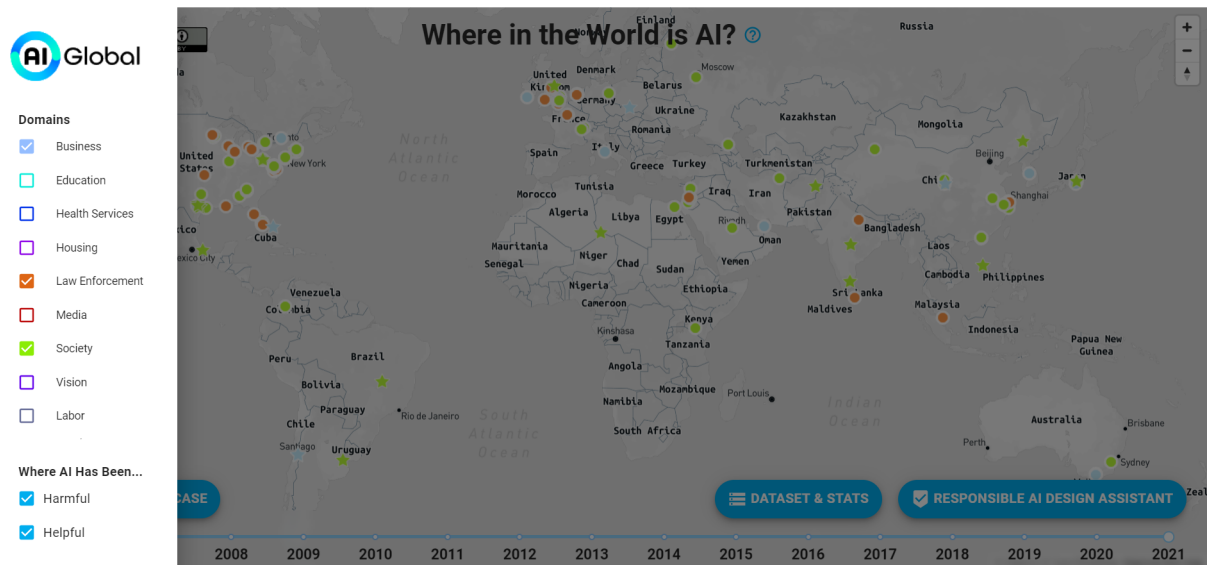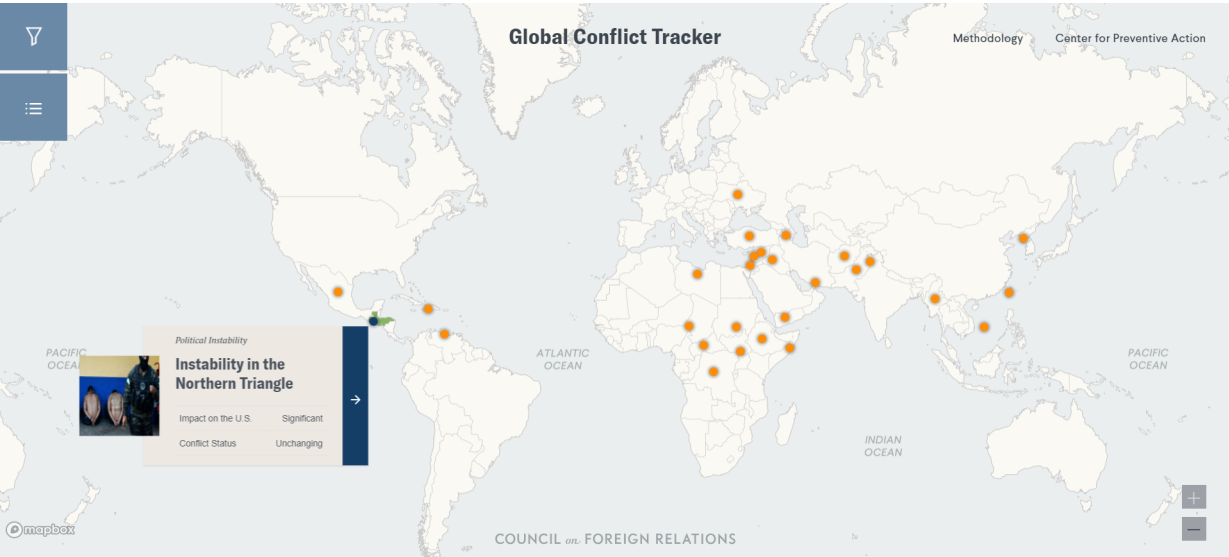

Fig. 4: View of AI Global - Where in the world is AI?

Fig. 5: View of Global Conflict Tracker