Similarities of Top AI Indexed Countries

Keitaro Ogawa[1], Skye Crocker[2], Colin Hascup[3], Diego Velazquez[4]

I. Introduction

In this project, we will be examining the nuanced landscape of AI adoption across developed countries using the AI Global Index data-set and AI companies data-set from Kaggle. Our curiosity lies on the relationships between the level of AI adoption and non tautological variables (discussed more in hypothesis section). We will further explore its relationships using the presence of AI companies in those countries. In doing so, we aim to capture the patterns in AI usage by showing different visuals and deploying various statistical techniques including logistic regression, k-means clustering, and principal component analysis. Thus, we will uncover insights that contribute to the broader narrative of AI innovation on a global scale.

Some key findings from our analysis include a surprising lack of correlation between Income Group and AI Index, challenging conventional assumptions. High-income countries didn't dominate, prompting further exploration into income distribution. Additionally, a consistent pattern emerged among top AI performers, with Operating Environment, Infrastructure, and Government Strategy playing crucial roles. Additionally we found a strong association between the Total AI Index Score and Political Regime, particularly with Liberal Democracies having the highest score, which is especially interesting since this contradicts some previous literature.

II. Data

Our main data-set is the AI Global Index data-set. This data-set is obtained from Kaggle, which has a comprehensive measure of nations' skills in AI as of April 2023, containing 62 rows (each row contains the data for one country) and 13 columns[5]. Fig1 shows each column in the data-set and its respective data type. While we plan to explore and describe all the features in the data-set, we place a special emphasis and focus on the 'Total Score' variable.

Here is a description of the major numerical variables which we will be looking at.

- **Talent Indicator:** Focuses on the availability of skilled practitioners for the provision of artificial intelligence solutions.
- **Infrastructure Indicator:** Focuses on the reliability and scale of access infrastructure, from electricity and internet, to super computing capabilities.
- **Operating Environment Indicator:** Focuses on the regulatory context, and public opinion surrounding artificial intelligence.

Country	STRING
Talent	FLOAT
Infrastructure	FLOAT
Operating Environment(	FLOAT
Research	FLOAT
Development	FLOAT
Government Strategy	FLOAT
Commercial	FLOAT
Regions	STRING
Cluster	STRING
Political Regime	STRING
Income Group	STRING
Total Score	FLOAT (target)

Fig. 1. "AI Global Index Data"

- **Research Indicator:** Focuses on the extent of specialist research and researchers; investigating the amount of publications and citations in credible academic journals.
- **Development Indicator:** Focuses on the development of fundamental platforms and algorithms upon which innovative artificial intelligence projects rely.
- **Government Strategy Indicator:** Focuses on the depth of commitment from national government to artificial intelligence; investigating spending commitments and national strategies.
- **Commercial Indicator:** Focuses on the level of startup activity, investment, and business initiatives based on artificial intelligence.
- **Total Score:** An aggregated score for a nations AI development and capabilities, calculated based on the previous 7 indicators.

The following tables provide summary statistics for the key numerical variables in the data set. The scores all range from 0 to 100, with varying means and standard deviations.

Metric	Talent	Infrastructure	Operating Environment	Research
Count	62.00	62.00	62.00	62.00
Mean	16.80	63.50	66.93	16.61
Std	15.21	20.22	20.00	17.41
Min	0.00	0.00	0.00	0.00
25%	7.37	55.86	58.11	3.03
50%	13.45	65.23	69.51	12.93
75%	24.57	75.95	80.50	25.41
Max	100.00	100.00	100.00	100.00

TABLE I

Summary statistics for AI-Index data set

In addition, the following categorical variables provide

Metric	Development	Govt Strategy	Commercial	Total score
Count	62.00	62.00	62.00	62.00
Mean	14.82	57.87	6.17	23.91
Std	19.42	26.25	14.03	15.12
Min	0.00	0.00	0.00	0.00
25%	1.20	41.03	0.70	14.81
50%	9.00	63.93	2.59	23.22
75%	19.98	77.95	5.31	30.49
Max	100.00	100.00	100.00	100.00

TABLE II

SUMMARY STATISTICS FOR AI-INDEX DATA SET CONT...

further information about each country that will be useful for analysis.

- **Region:** Geographic location on Earth (Americas, Asia-Pacific, Europe, Middle-East, Africa).
- **Cluster:** Predetermined cluster labels for different groups of countries with similar characteristics regarding AI (power players, traditional champions, rising stars, waking up, nascent). **not being used in this report.
- **Political Regime:** Form of government (closed autocracy, electoral autocracy, electoral democracy, liberal democracy).
- **Income Group:** Average income level of population (high, upper middle, lower middle).

Fig2 indicates the second data-set obtained from Kaggle as well. This AI Companies data-set has a trove of information on AI companies worldwide as of 2022, having 3096 rows and 7 columns [8]. We mostly focus on the location variable.

Company Name	STRING
Website	STRING
Location	STRING
Minimum Size Project	INT
Average Hourly Salary	INT
Number of Employee	INT
Percent of AI Service Focus	FLOAT

Fig. 2. "AI Companies Data"

III. LITERATURE REVIEW

AI has recently seen an explosion of development and adoption worldwide. AI can now perform even better than humans in a variety of fields, including image recognition, chatbots, social media, autonomous vehicles, and gaming. On an intricate level, the computational power used to train AI has increased from tens of petaFlops to tens of billions of petaFlops, and investments in AI have increased by over 1000%. [1] Given the potential power of AI, it is imperative for countries to adopt it as soon as possible so they don't fall behind. Yet more than 90% of the hardware for AI is designed and manufactured in only a handful of countries: the US, TWN, CHN, KOR, and JAP. Additionally, a recent paper found that authoritarian regimes perform better in producing AI development outcomes [2]. AI is currently exploding in industries like Financial services, Healthcare, transport and Logistics. These improvements will drive up labor productivity as well as GDP [7]. It is important to look into these patterns in AI development so that we can better understand and promote the development of fair, safe and effective AI around the world.

IV. HYPOTHESIS/GOALS

Our goal is to find the closeness of the Top AI indexed countries to other top-countries. However, this would be intuitive and tautological, because without doing extensive research, we know countries with advanced technology have more AI adoption. It would not be interesting for our analysis. Hence, we have decided to just focus on non-tautological variables, which are government strategies, operating environment, political regimes, and regions. At first sight, we were not sure if there are significant correlations to AI Index, or whether they even have correlations. Also, we looked forward to detecting some outliers countries. Mainly, these were our motivations going into the research.

Moreover, to serve as a baseline, we use the AI Companies data-set to map out the total counts of companies in each country. This leads to our base-hypothesis, which is countries with more AI index correlate (geographically) to their numbers of companies that have AI services.

To sum up, our main object is to gain insights into the top AI indexed nations and find out whether those countries have associations or patterns with the non-tautological variables. These variables provide enough rooms for our detailed visuals.

V. METHODS

To begin with, most of the feature engineering techniques were used in the process of visualizations. For example, we have normalized variables for better usage of data-sets in general, PCA was employed for the dimensionality reduction of data-sets for clustering, and label encodings of string variables were conducted when performing correlation matrix using heatmaps. At the end, we performed the correlation matrix of all variables (including tautological columns) as an indicator of how much non-tautological variables were correlated to the AI Index scores compared to tautological variables.

Regarding the visuals, we explored ranges of plot types, such as geospatial visualization using geopandas, bubble charts, and scatter plots. Each of the plot offer unique perspectives in top AI indexed countries and non-tautological variables.

Consequently, we used logistic regression for the classification. This method was utilized when we predicted each political regime based off of other AI index indicators (results provided in results section). The motivation for using logistic regression lies in its straightforward linear nature. The idea is that if we can predict the political regime of a country based on its other features, then there must be some underlying relationship(s) that make this possible. The linear nature of logistic regression would suggest that any underlying relationships are also straightforward and likely linear. All

variables were used since this was only used as an initial analysis.

Lastly, for our cluster technique, we applied both K-means and Spectral clustering to the AI index data-set with all numerical variables (excluding total score) and with label encoded categorical variables. These are both popular and established clustering techniques that will hopefully provide hidden patterns and insights into the data by comparing the 'Total Score' of the clusters created against other variables, which is why that variable was left out. We chose the number of clusters by using an internal validation metric, the Calisnki-Harabasz index, to compare the ratio of between cluster and within cluster scatter for each clustering result with different values of n. The results of the analysis are in the next section, but we end up using n=2 clusters since it had the best score for both clustering methods. Interpretation, including visualization using PCA, of the actual clustering can be found in the results section.

VI. Results

A. AI Indexed Countries and AI Companies Comparison

As we discussed in the introduction, we see obvious patterns with the countries with more technology and GDPs having high AI index. According to Fig3, top 10 countries were US(1), CHN(2), UK(3), CAN(4), ISR(5), SGP(6), KOR(7), NLD(8), DEU(9), and FRA(10). With the exception of CHN, Europe and North America contain the highest indexes, which again makes sense due to the technological advancements in those regions. Also, we notice that the US and CHN are the two AI powerhouse of AIs. These countries play an important role as we are analyzing the consistent outliers.

In Fig4, we also have a world map showing the number of AI companies in each country by colors. Comparing to Fig3, we have US, UK CAN, CHN, and DEU as commonalities and most are concentrated in Europe. On the other hand, some differences with the AI index map are that IND, POL, AUS UKR, PAK are in the top 10 for counts of AI companies. While RUS and countries in Oceania regions had relatively high AI index, they have less AI companies as seen in Fig4. Also, nearly 30 percent of the total AI companies are based in USA, 14 percent in IND, and 11 percent in UK, all of which are top AI indexed countries.

An intriguing article we found claimed that companies worldwide exhibit hesitancy in adopting AI technologies, with adoption rates ranging from 16 percent to 23 percent. This caught our attention as we expected to see exponential rise in the use of AI. Some notable barriers include technological issues like poor data quality and high costs, as well as generic indicators such as company size influencing adoption[4]. Regarding this, we can claim that only certain countries don't exhibit these trends, such as the countries listed as the top for high numbers of AI companies.

B. Clustering results

First, in order to determine the number of clusters for our unsupervised learning algorithm, we ran multiple clusters

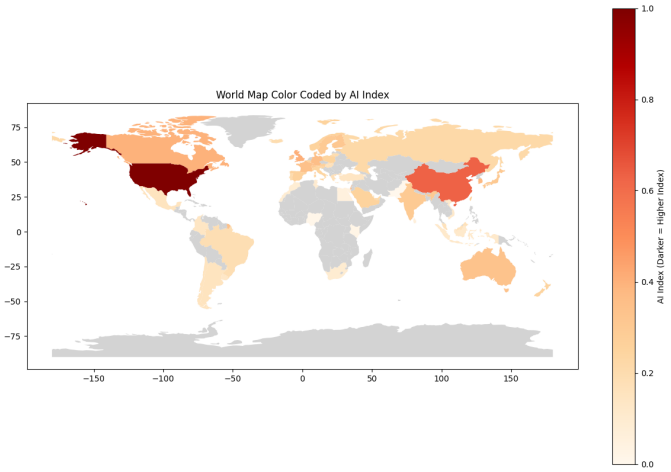

Fig. 3. "AI Index Scores Map"

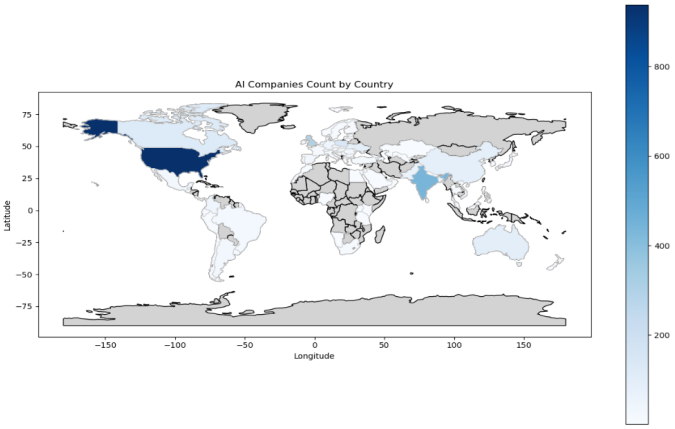

Fig. 4. "AI Companies Counts Map"

and compared their Calinski-Harabasz scores. As seen in Table III, n=2 clusters had the best score for both clustering algorithms.

From here, we run the algorithms and perform PCA to visualize the results in two dimensions. At first glance, there appears to be two obvious clusters as well as two outliers, the US (top left) and CHN (top right) in Fig5. We can also see that Total AI index score improves diagonally moving from the bottom right of the graph to the top left. This information will be useful as it serves as the basis for analyzing our clustering results.

Figure 6 shows several plots using PCA where the countries are labeled based on other categorical variables. From these graphs we have the following findings:

- **Statistical Clusters:** Both clustering algorithms divide the countries into two obvious groups. (Left = group 1, Right = group 0) Excluding outliers, group 1 has generally higher AI index scores.
- **Regions:** There seems to be a high concentration of European countries in group 1, while most of the African countries are in group 0. Otherwise there doesn't seem to be too much correlation between the geographic

TABLE III
CLUSTERING RESULTS

n	K-means (CH Score)	Spectral (CH Score)
2	32.05	29.03
3	25.29	24.83
4	25.80	19.72
5	24.01	19.43
6	23.60	18.58
7	22.36	16.90
8	23.42	16.65
9	23.43	15.48
10	22.25	14.37

region and the cluster of the country.

- **Income Group:** Group 1 is dominated by high income. This is not too surprising, since we saw that group 1 generally has higher scores and income is one of our tautological, less interesting variables. Nevertheless, it is interesting that a few high income countries are in the other group.
- **Political Regime:** A strong cluster of liberal democracy is observed in group 1. This is an interesting finding that we will analyze further, especially since it contradicts some of the literature reviewed earlier [6]. Most of the electoral autocracies are in group 0, but otherwise there is good diversity between the political regimes and their cluster.
- **"Clusters":** Power players are outliers. Rising stars and traditional champions are in group 1, while nascent countries are in group 0. The waking up group is spread out between the two clusters. One thing that is interesting is that the predetermined clusters are divided into five groups while our analysis found it better to divide into two groups.

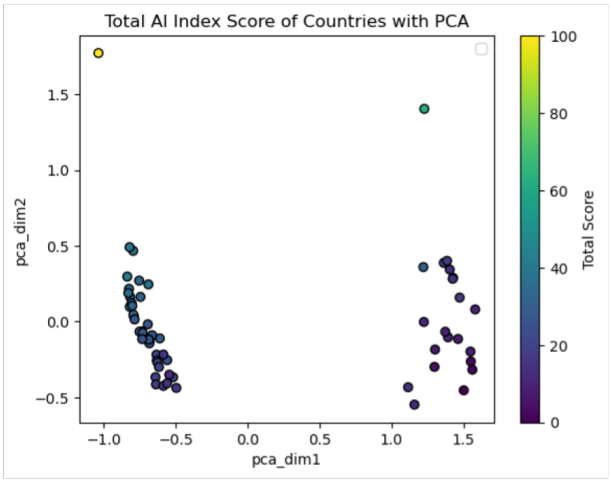

Fig. 5. "PCA - Total AI Index score"

For further analysis of the distribution and nature of political regimes and their relationship to AI indicators, we now show the results of our logistic regression classifier in Fig7. The model achieved an accuracy of 77.42%, which is

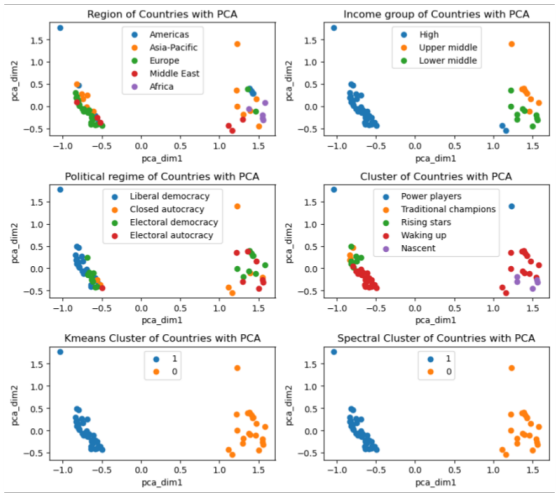

Fig. 6. "Cluster Results with PCA"

relatively good considering that the largest category (liberal democracy) only comprises 43.55% of the data set (27 out of 62) which would serve as a baseline for guessing randomly. This would indicate that there is likely some sort of linear relationship / boundary between AI indicators and political regime.

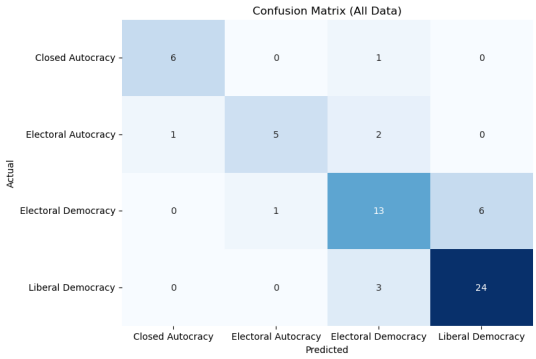

Fig. 7. "Confusion Matrix of Political Regime Linear Classifier"

C. Government Strategy

After we have sorted out the top AI indexed countries, we turned our first focus to the Government Strategy. Through the scatter plot in Fig8, we can see a global trend of higher Government Strategy and low AI index score. However, we would like to mention the outliers: CHN and the US, both of which has high AI index score compared to other top 10 countries. The U.S. in particular, has very high AI index correlation to lower government strategy score. This suggests a commonality in the importance placed on government initiatives and policies to foster AI development.

Accordint to Ndou in 2002, government strategy in the form of ICT (Information and Communication Technology) involves the deliberate policies that a government undertakes to leverage ICT for economic development, societal progress, and global competitiveness [6]. This strategy is correlated to

nations with high GDP growth (at the time was the US, DEU, FRA, and Uk). Hence, historically these nations have high GDPs along with high government strategy and given that these nations still have high AI index in the world-maps, government strategies value have a decent association to AI Index.

Moreover, this independent behavior of these countries can be explained through other variables like, Political Regime. This variable can affect the implementation of the strategy; for instance, we can see that liberal democracies might demonstrate better implementation of policies compared, such as to autocratic regimes. We didn't find any important patterns corresponding to the regions of these countries. But in general, the trends of low AI index score may associate to countries' over-regulation, lack of flexibility, ethical concerns, limited collaboration with the AI industry, inadequate investment in education, and other factors part of the government strategy.

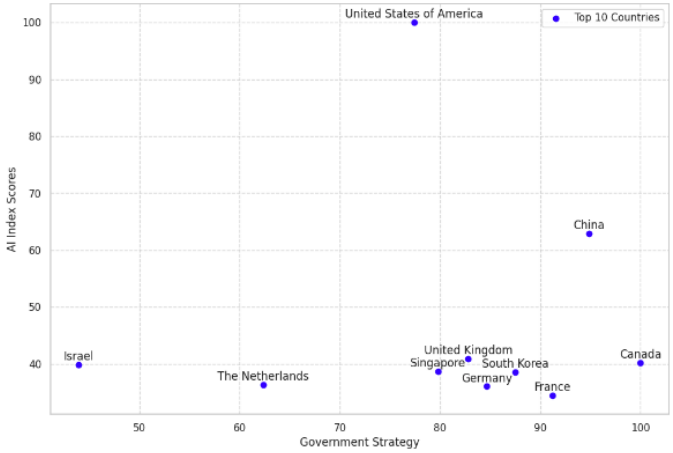

Fig. 8. "Government Strategy and AI Index Correlation Plots"

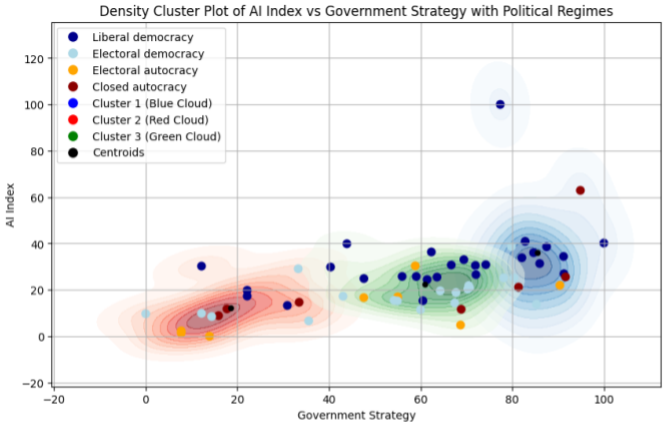

Fig. 9. "Density Cluster Plot of AI Index vs Government Strategy with Political Regime"

D. Government Strategy and Political Regime Co-analysis

Subsequently, we analyze further into Government Strategy along with Political Regime. In Fig9, it shows the detail of the relationship between government strategy index and AI index, with the data points color-coded by country government regime type. Cluster 1 is particularly intriguing as it comprises a significant number of both liberal democracies and closed autocracies, a surprising mix given its association with the highest AI index scores. Interestingly, this cluster includes only a minimal percentage (5.6%) of electoral democracies.

Cluster 2 presents a diverse composition, featuring a balanced representation of all four regime types, with a notable concentration of electoral democracies, each regime contributing about 20-30% to this cluster. In contrast, Cluster 3 encompasses a mixture of all regime types but has the smallest proportion of liberal democracies. It also contains only 3.4% of closed autocracies, which is unexpected considering the general assumption that such governments might typically exhibit lower AI indexes. This data provides valuable insights into the complex interplay between political regimes and AI development.

Moreover in Fig10, we notice the majority of top AI indexed countries have Liberal Democracy as their political regime. Then, the question arose: "Why?"

This question led the team to read other articles. We observed that liberal democracy countries have combinations of internal management strategies and the alignment of technological development[3]. In other words, liberal democracies typically provide an environment conducive to research and development, encouraging both public and private sectors to invest in cutting-edge technologies, including AI. In addition, nations with a strong commitment to individual freedom have a tendency in adopting and regulating AI technologies. For example, US and UK actively address the challenges posed by AI. This commitment not only enhances societal trust but also promotes a balance between technological progress and safeguarding individual rights.

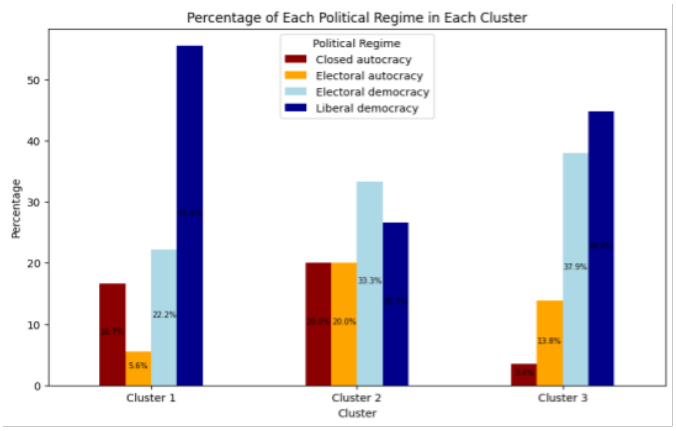

Fig. 10. "Political Regime Distribution by Cluster"

E. Operating Environment Correlation Plots Analysis

Now, we delve into Operating Environment. While a high operating environment generally suggests favorable conditions for AI development, it's essential to acknowledge that

there could be instances where it might have unintended challenges. Some of these challenges are regulatory stringency which prevents agile innovation, ethical and privacy concerns that limit the application of AI in areas where the balance between innovation and ethical consideration is delicate. Another factor is the global competition, which is challenging to compete globally with strict regulations, especially if other nations strike a balance.

Fig 10 illustrates two outliers that clearly demonstrate the effect of the operating environment score. We can see the US as the perfect example of a balanced operating environment score. In terms of this variable, we can see that the US is not even among the top 5, but is the country with the highest AI index. This can be due to the open space for innovation while maintaining regulations. On the other hand, we have CHN, the country with the highest operating environment score. This high score can be translated to regulations and legal moves that slow down the growth of the industry in the country.

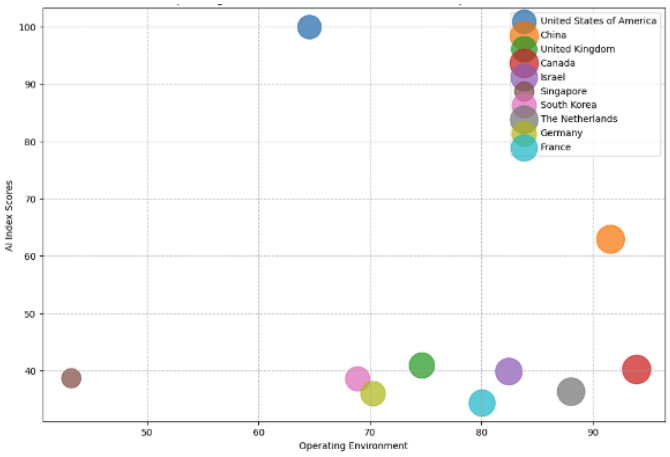

Fig. 11. "Operating Environment and AI Index Correlation Plots"

F. Final Result: Correlation Matrix and Spider Plots

In the hypothesis section, we have defined non-tautological variables to be Political Regime, Regions, Operating Environment, and Government Strategy. After we have performed descriptive analyses, we are concluding the results by displaying the correlation matrix that includes all the variables (tautological columns, too).

From Fig12, we break it down into three tiers: positive, neutral, and negative associations. For positively affected columns to AI index, we have most of tautological variables as expected. For non-tautological variables, we see Government Strategy having medium correlation, and Operating Environment and Political Regime (majority was liberal democracy) have weak correlations.

However, our most unexpected result was the limited association of Income Group (a tautological variable) to AI Index score. Within Income Group, high income group has the majority voting, which would make sense tautologically. Therefore, this may suggest that countries in our data-set had more low income countries than high income by large amount. In other words, this finding prompts further exploration and suggests that the distribution of income among countries in our data-set may not align with conventional assumptions.

Next, we found a strong association between AI Index Total Score and Political Regime, where liberal democracies have the highest score. This contrasts with previous literature which suggested that authoritarian regimes produce better outcomes regarding AI development. Future analysis may want to look into this discrepancy more in-depth for a more conclusive determination.

Additionally, the spider graph in fig13 visually represent the AI capabilities of the top countries in two categories: a pattern emerges regarding the shape and size of these graphs. For most countries, excluding the US, CHN, and South Korea, the graphs exhibit a similar shape. This similarity is primarily evident in three key areas: Operating Environment (OE), Infrastructure, and Government Strategy. These factors appear as prominent spikes in the graphs, indicating their high levels in these countries.

Next, we found a strong association between AI Index Total Score and Political Regime, where liberal democracies have the highest score. This contrasts with previous literature which suggested that authoritarian regimes produce better outcomes regarding AI development. Future analysis may want to look into this discrepancy more in-depth for a more conclusive determination.

The consistency in these areas among the top countries, with the exception of the mentioned outliers, suggests that these elements - OE, Infrastructure, and Government Strategy - are crucial in achieving a high AI Index. This pattern implies a strong correlation between these factors and the overall success in AI development within these nations.

G. Key Takeaways

Our findings can be used in guiding policymakers, businesses, and researchers. For example, policymakers can craft strategies that prioritize key variables like Government Strategy and Operating Environment. But, it is important to note that these two variables only had significant correlations on the consistent outliers, which were CHN and the US. As PwC claims that the greatest economic gains from AI will be predicted to be in China (26 percent boost to GDP in 2030) and North America (14.5 percent boost), equivalent to a total of 10.7 trillion dollars and accounting for almost 70 percent of the global economic impact[7].

For businesses, understanding the factors correlated with high AI Index scores provides a roadmap for growth. Companies can tailor their investments and initiatives based on the identified key nations, thereby enhancing their competitiveness in the AI landscape. In particular, the consistency in Infrastructure, Operating Environment, and Government Strategy among top-performing countries may positively impact their businesses on a global scale.

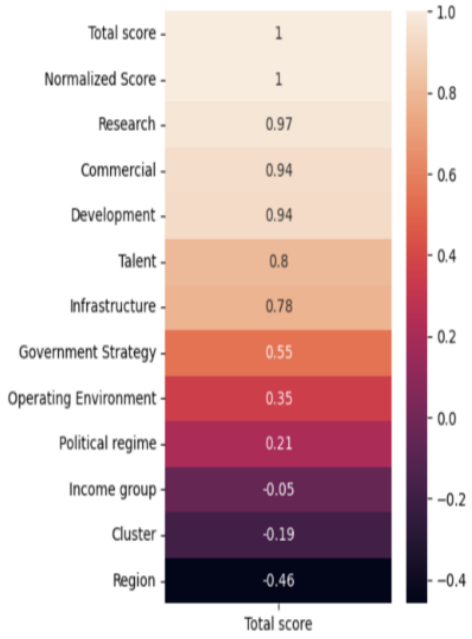

Fig. 12. "Correlation Matrix"

VII. Appendix

References

[1] Charlie Giattino, Edouard Mathieu, Veronika Samborska, Max Roser, "Artificial Intelligence," Our World in Data, 2023. https://ourworldindata.org/artificial-intelligence

[2] Filgueiras, Fernando. "The Politics of AI: Democracy and Authoritarianism in Developing Countries."1 Journal of Information Technology Politics, vol. 19, no. 4, 2022, pp. 449-464. DOI: 10.1080/19331681.2021.2016543

[3] Wright, Nicholas D. Artificial Intelligence and Domestic Political Regimes: Digital Authoritarian, Digital Hybrid, and Digital Democracy, 2019, www.jstor.org/stable/resrep19585.9.

[4] Kinkel, Steffen, et al. "Prerequisites for the Adoption of AI Technologies in Manufacturing – Evidence from a Worldwide Sample of Manufacturing Companies." Prerequisites for the Adoption of AI Technologies in Manufacturing – Evidence from a Worldwide Sample of Manufacturing Companies, 17 Aug. 2021, www.sciencedirect.com/science/article/pii/S0166497221001565.

[5] Meleshenko, Kateryna. "Ai Global Index." Kaggle, 26 Apr. 2023, www.kaggle.com/datasets/katerynameleshenko/ai-index.

[6] Ndou , Valentinanbsp; Dardha. E – GOVERNMENT FOR DEVELOPING COUNTRIES: OPPORTUNITIES AND CHALLENGES, 2002.

[7] PricewaterhouseCoopers. "PWC's Global Artificial Intelligence Study: Sizing the Prize." PwC, www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html. Accessed 15 Dec. 2023.

[8] Vineeth. "AI Companies Dataset." Kaggle, 16 June 2022, www.kaggle.com/datasets/vineethakkinapalli/ai-companies.

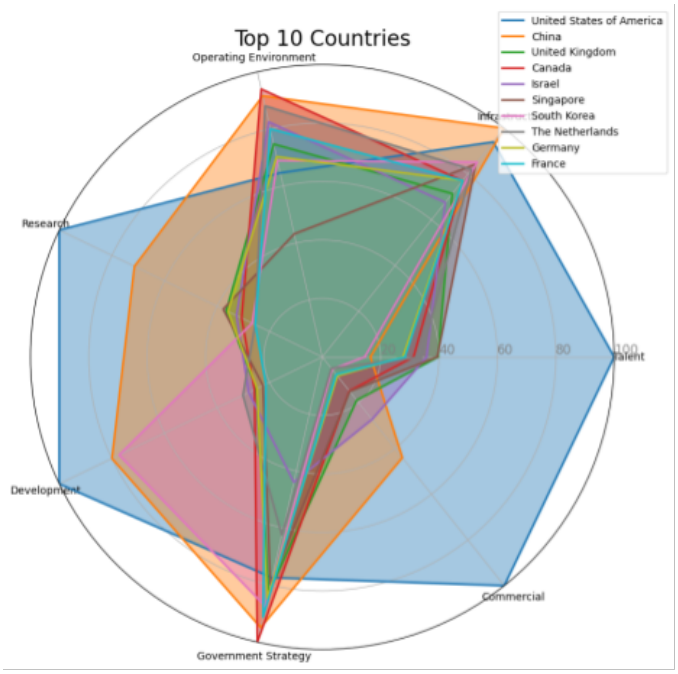

Fig. 13. "Score Distribution of Top 10 AI index Countries"