Sahana Srinivasa Babu

What Drives Global AI Readiness?

Course: ISE-201 Math Dec. and Data Science
Prof. Khaled Mabrouk

Research Question:
What factors most strongly contribute to a high AI readiness score for a country?

Project Overview & Data Sources

Initially explored **three datasets**:

- Largest Companies in the World
- Worldwide AI Development Rankings
- AI Ghibli Trend (AI-generated image popularity)

After descriptive statistics for all three, I chose the **Worldwide AI Development Rankings** dataset for deeper EDA.

Why I Chose This Dataset

- Provides clear, structured indicators of AI readiness.
- Includes multiple dimensions (Talent, Research, Infrastructure, etc.).
- Directly supports the goal of identifying factors driving AI success across countries.

Dataset Structure & Basis for Scoring

Dataset: **62 countries × 13 columns**

AI dimensions (numeric, 0–100): Talent, Infrastructure, Operating Environment, Research, Development, Government Strategy, Commercial, Total score.

Categorical Columns: Region, Income group, Political regime, Cluster, Country

Basis for scoring :

- Each dimension is a **normalized index from 0 to 100**
  - 0 ≈ extremely weak performance
  - 100 ≈ strong performance among all countries in the dataset
- The **Total Score** is a **composite summary** measure of AI readiness generated by combining all seven dimensions

Handle Missing Data, Handle Outliers And Data Transformation

Handle Missing Data

- Checked for missing values, incorrect types, and duplicates using isnull(), sum(), and duplicated().
- No missing values, no duplicates, and no placeholder missing values.

Handle Outliers

- Inspected numerical features using boxplots.
- Outliers were genuine variations, not errors → **kept them** for further analysis.

. Data Transformation

- No scaling needed (all values 0–100).
- Categorical features **label-encoded**.
- Created new variables (*RnD_ratio*, *tech_strength*, *policy_effectiveness*).
- Grouped data by **Region, Income Group, Political Regime** to compute averages.

Descriptive Statistics Of All Numerical Features

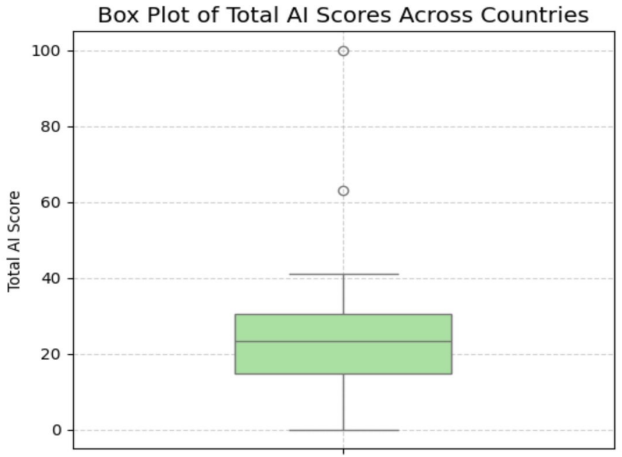
Feature	Mean	Median	Std. Dev.	Skew / Key Insight
Operating Environment	66.93	69.50	20.00	Slight left-skew → Strong for most
Infrastructure	63.50	65.23	20.22	Symmetric → Consistent across countries
Gov. Strategy	57.87	63.93	26.25	Left-skewed → Many strong strategies
Talent	16.80	13.45	15.21	Right-skewed → Low overall; few high outliers
Research	16.61	12.93	17.41	Right-skewed → Most score low
Development	14.82	9.00	19.42	Right-skewed → Slow AI development
Commercial	6.17	2.58	14.03	Right-skewed → Very low for most
Total Score	23.91	23.22	15.12	Symmetric → Moderate differences

- **Infrastructure & Operating Environment** score high and are fairly consistent across countries – **basic** infrastructure and supportive environments
- **Talent, Research, Development, and Commercial** scores are low and highly uneven, with only a few strong performers – human capital, research output, and commercialization **lag** far behind.

Global Distribution of Total AI Scores

Boxplot of Total Score:

- Scores range from **0 to 100** (range = 100).
- Middle 50% of countries cluster around **low to moderate scores**.
- A few outliers (e.g., USA, China) sit at very high scores.



Highest scoring region: Americas
Lowest scoring region: Africa

Insights:

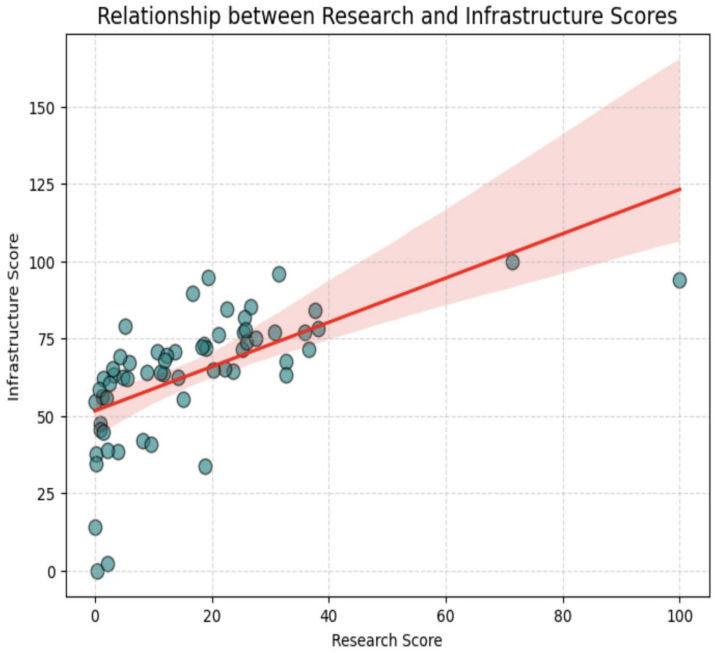
1. The AI landscape is **highly unequal**: a small group of leaders strongly outperforms the majority of countries.

Relationship Between Research and Infrastructure Scores

Countries with **higher Research scores** generally also have **higher Infrastructure scores**.

Insight

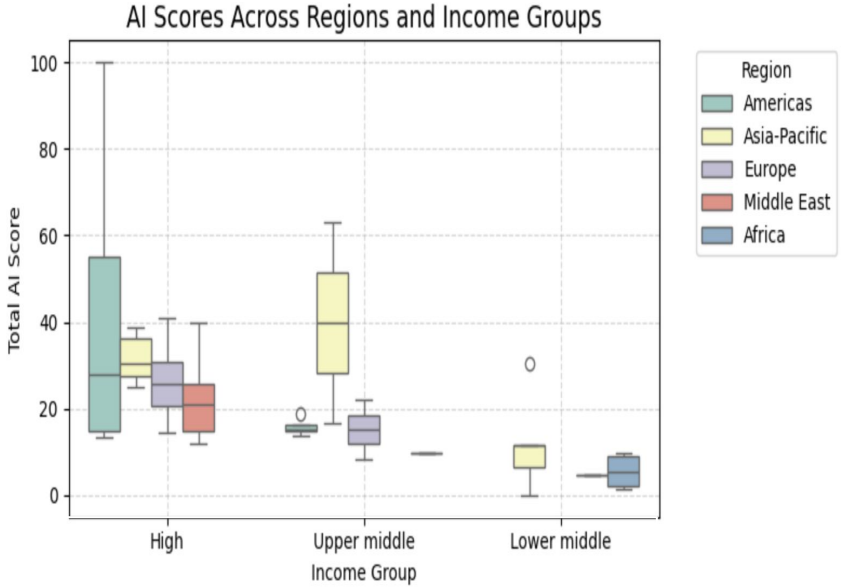
- Most countries cluster at **lower research levels**, but even within that cluster, there is a **positive upward trend** in infrastructure.
- A few countries perform strongly in both areas, indicating well-developed research systems supported by strong infrastructure.
- Strong research systems are usually paired with strong infrastructure.



Regional and Income Group Patterns

- **Average Total Score by Region:**
  - Americas ≈ 29.0 (highest)
  - Asia-Pacific ≈ 25.8, Europe ≈ 25.5
  - Middle East ≈ 19.7
  - Africa ≈ 6.4 (very low)

- **Operating Environment by Income group:**
  - High income: ≈ 71.9
  - Upper middle: ≈ 71.4
  - Lower middle: ≈ 41.7



Insight:

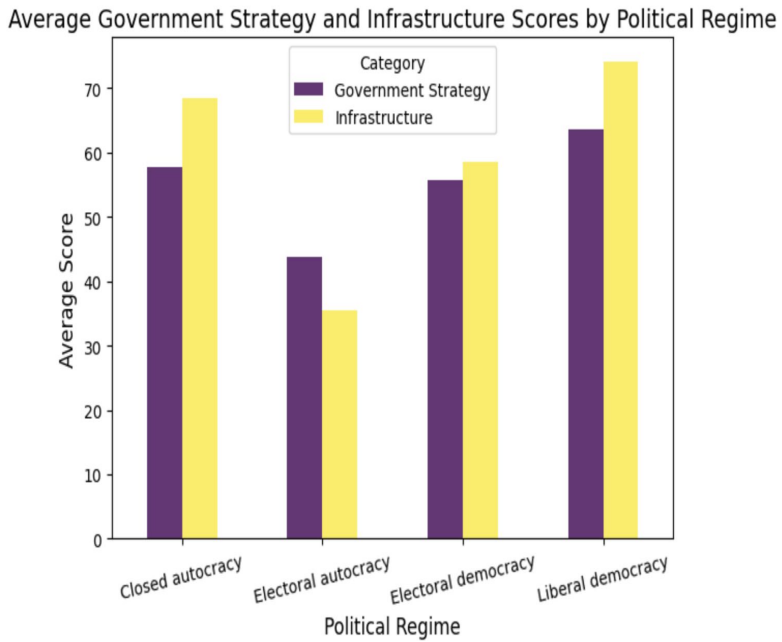
Upper-middle-income countries match high-income countries on Operating Environment, but still lag on Total Score → suggests that **their bottlenecks lie in other areas (talent, research, commercialization)**.

Political Regime, Government Strategy, and Infrastructure

- **Liberal Democracies:** Highest Strategy and Infrastructu scores.
- **Closed Autocracies:** Strong Infrastructure, weaker Strategy.
- **Electoral Autocracies:** Lowest on both.

Insight:

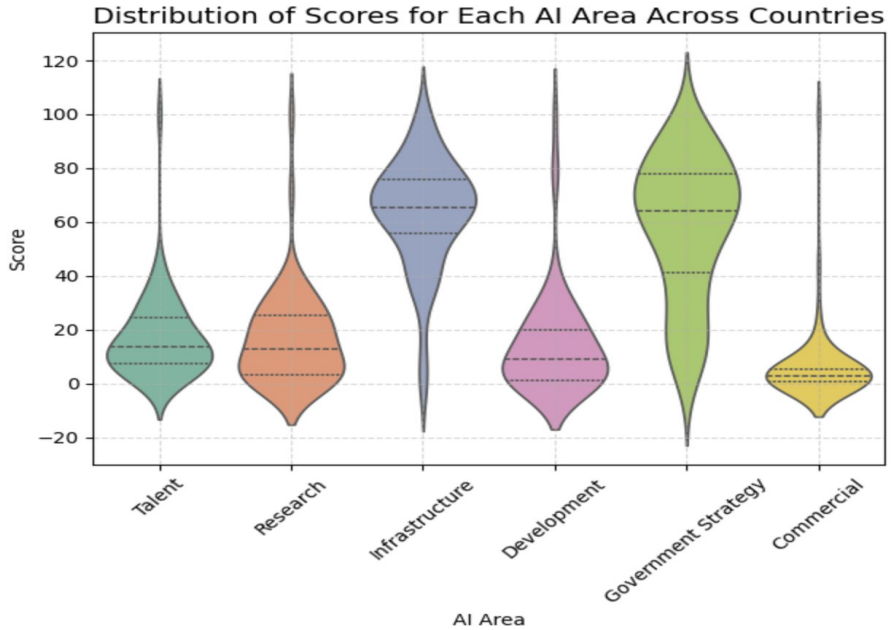
- **Stable, more democratic systems** appear better at combining **long-term AI planning** with **infrastructure investment**.
- However, this does **not guarantee high commercialization or development** (we see gaps later)



Distribution of AI Area Scores

Key Patterns Across Dimensions

- **Infrastructure & Government Strategy**
  - Highest scores overall
  - Also the most uneven across countries
- **Talent & Research**
  - Lower scores but more consistent
  - Most countries perform at modest levels
- **Development & Commercial**
  - Lowest scores among all dimensions
  - Indicate limited progress and weak commercial AI activity



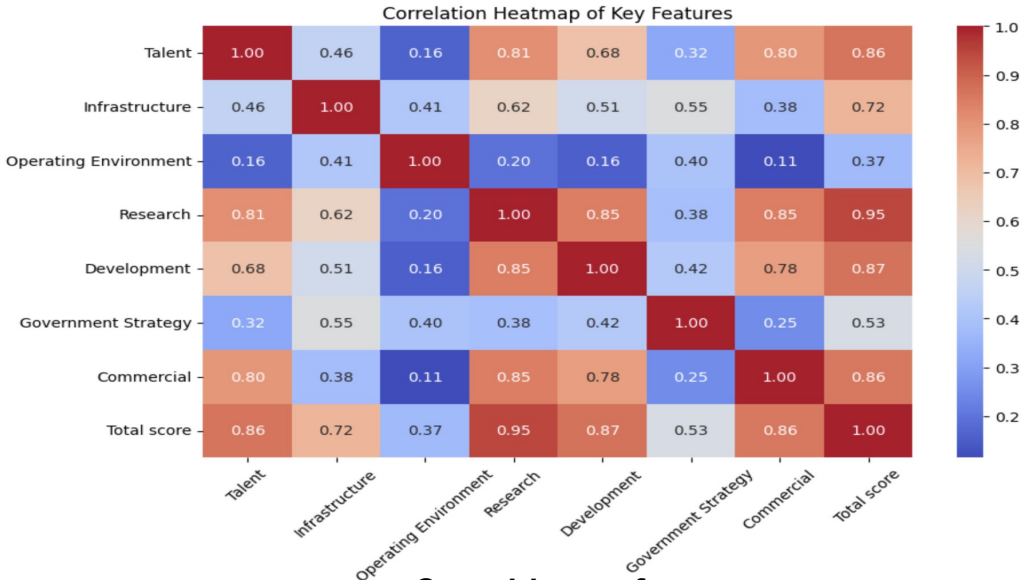
Insight

Countries invest strongly in infrastructure and planning, but **struggle with execution areas** such as talent, research output, development, and commercialization.

Correlation Heatmap: What Actually Drives Success?

*Correlation with **Total Score**:*

- Research: **0.95** (strongest driver)
- Development: **0.87**
- Talent: **0.86**
- Commercial: **0.86**
- Infrastructure: **0.72**
- Government Strategy: **0.53**
- Operating Environment: **0.37** (weakest link)



- Infrastructure and government strategy matter, but their effect is **indirect and weaker** than actual research and deployment.
- Operating Environment alone does **not** guarantee high AI readiness.

Core drivers of success

- *Innovation capacity (Research - 0.95)*
- *Ability to build and deploy AI (Development - 0.87)*
- *Skilled people (Talent - 0.86)*
- *Market uptake (Commercial - 0.86)*

NEW FINDINGS

Countries Fall into Two AI Development Archetypes

Based on score gaps:

1. **Balanced Leaders**
    - Strong in all dimensions
    - Example: China

2. **Unbalanced Aspirants**
    - Strong Infrastructure/Strategy
    - Weak in Research/Talent/Commercial
    - Examples: Saudi Arabia, Colombia

Being “AI-Ready” Does NOT Mean Being “AI-Productive”

Insight

Countries with strong **Infrastructure** and **Government Strategy** can still score low overall—they appear AI-ready but are not AI-productive.

Why

They lack:

- Strong **Research capability**
- Robust **Talent pipelines**
- Sufficient **Development capacity**
- Active **Commercial ecosystems**

Operating Environment Has Very Little Real Impact

Insight

Despite scoring high on average, **Operating Environment** has the weakest correlation with Total Score (0.37), meaning it has little real impact on AI readiness.

Why

- Improving business/regulatory conditions alone **does not translate** into stronger AI capability.
- This challenges the common assumption that a good operating environment directly drives AI success.

Government Strategy Matters Only When Research Systems Is Strong

Insight:
Government Strategy boosts AI readiness **only** when Research scores are high.

Why:

- Gov. Strategy → Total Score correlation is **moderate (0.53)**; Research → Total is **very strong (0.95)**.
- Countries with **high Strategy but weak Research** still have **low Total Scores**.
- Top performers score **high in both Strategy and Research**

Upper-Middle-Income Countries Are Misdiagnosing Their Weaknesses

Insight

Upper-middle-income countries have Operating Environment scores nearly equal to high-income nations but still achieve low Total Scores.

Why

They lack:

- Strong **Talent**
- Deep **Research capacity**
- Effective **Commercialization capability**

This pattern only appears when comparing multiple variables together, making it a **novel analytic finding**.