**Your research on "The AI Moral Code" is an ambitious and important endeavor,** aiming to capture a global perspective on AI governance principles, values, and ethics across diverse sectors (Government, Academia, Business, NGOs, Religious Organizations). The current distribution of your 250 documents shows a clear bias, with the USA (87 documents) and China (19 documents) dominating, while many countries (e.g., Africa, Chile, Iceland, New Zealand, Norway, Portugal, Saudi Arabia, Scotland, Spain, Taiwan, UAE, Ukraine) have only 1 document each. This imbalance suggests that your retrieval process may be skewed toward economically powerful or highly visible nations, which aligns with your observation of potential bias.

**Your idea of weighting documents based on a country’s economic influence and leadership in A**I thinking is a solid starting point, but it can be refined to ensure fairness and inclusivity. Below, I’ll outline a recommendation for weighting your document retrieval process, considering economic power, AI leadership, and representation across sectors and regions.

|  |
| --- |
| **Analysis of Current Data** |
| **Total Documents: 250** |

**Top Contributors: USA** (87), China (19), France (16), UK (24), Germany (8), Canada (8), India (8), Japan (10), Singapore (6), Switzerland (4), Netherlands (3), Australia (3), Finland (3), Belgium (2), Italy (2), Israel (2), Mexico (2), South Korea (2), Spain (2), Sweden (2), Russian Federation (2), Chile (2), Brazil (1), Argentina (1), Denmark (1), EU (19), etc.

Underrepresented Regions: Africa, Middle East (UAE, Saudi Arabia), Latin America (Chile, Mexico, Argentina, Brazil), and smaller European nations (Iceland, Norway, Portugal, Spain) have minimal representation.

Non-Countries: "Africa" and "EU" are aggregates, not single countries, which complicates direct comparison but indicates a need for regional weighting.

The G7, G10, and G20 classifications you’ve used (e.g., G7, G10, and G20 for USA, UK, France, etc.) suggest you’re already considering international groupings, which is a good framework. However, the current count doesn’t reflect the diversity of thought from less economically dominant regions or sectors.

Recommended Weighting Approach

To balance economic influence, AI leadership, and global inclusivity, I suggest a multi-factor weighting model for your document retrieval process. Here’s how you can approach it:

1. Define Weighting Criteria

Assign weights based on the following factors:

Economic Power: Use metrics like GDP (e.g., from the World Bank or IMF) to reflect a country’s economic influence. Higher GDP could increase the weight of documents from that country.

AI Leadership: Assess a country’s contribution to AI research, policy, or innovation. This could be measured by:

Number of AI patents (e.g., WIPO data).

AI research publications (e.g., Scopus or Google Scholar metrics).

Presence of AI policy frameworks (e.g., national AI strategies).

Population: Ensure representation proportional to population size to avoid over-emphasizing small but economically powerful nations.

Geographic Diversity: Add a bonus weight for underrepresented regions (e.g., Africa, Latin America, Middle East) to capture diverse cultural perspectives.

Sector Representation: Weight documents based on the sector (Government, Academia, Business, NGOs, Religious Organizations) to ensure all voices are heard, not just government or business perspectives.

2. Assign Initial Weights

Create a scoring system. For example:

Economic Power Weight (W1): Based on GDP ranking (e.g., USA = 1, China = 2, etc., with inverse ranking for simplicity, so lower rank = higher weight). Normalize to a scale (e.g., 0.1 to 1.0).

AI Leadership Weight (W2): Based on a composite score of patents, publications, and policy presence (e.g., 0.1 to 1.0).

Population Weight (W3): Proportional to population size (e.g., normalized to 0.1 to 1.0).

Diversity Weight (W4): Fixed bonus (e.g., 0.5) for underrepresented regions (Africa, Latin America, Middle East, etc.).

Sector Weight (W5): Adjust based on sector (e.g., Government = 1.0, Academia = 0.9, Business = 0.8, NGOs = 0.7, Religious = 0.6) to prioritize underrepresented sectors.

Total Weight for a country = W1 + W2 + W3 + W4 + W5 (normalized to sum to 1.0 across all countries).

3. Calculate Target Document Allocation

Total Weighted Share: Sum the weights for all countries to get a total (e.g., 10.0 if normalized).

Target Documents per Country: Multiply the total documents (250) by each country’s weight. For example, if USA has a weight of 0.2, it should get 250 × 0.2 = 50 documents.

Adjust for Current Count: Compare the target number to your current count. If USA has 87 but the target is 50, reduce emphasis on USA; if Africa has 1 but the target is 5, increase retrieval from Africa.

4. Practical Implementation

Data Sources:

GDP: World Bank (2024 data).

AI Leadership: AI Index (Stanford), WIPO patent data, national AI strategy reports.

Population: UN estimates (2025).

Retrieval Strategy:

Use targeted searches (e.g., Google Scholar, policy databases, NGO archives) with keywords like "AI ethics [country]" or "AI governance [sector]."

Collaborate with local experts or institutions in underrepresented regions to identify influential documents.

Set a minimum threshold (e.g., 5 documents) for each region to ensure minimal representation.

Iterative Review: After retrieving additional documents, re-evaluate the distribution and adjust weights if needed.

5. Example Weighting (Hypothetical)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country | GDP Rank (W1) | AI Leadership (W2) | Population (W3) | Diversity (W4) | Sector Avg (W5) | Total Weight | Target Docs (250 × Weight) |
| USA | 0.9 | 1.0 | 0.9 | 0.0 | 0.85 | 3.65 | 91.25 |
| China | 1.0 | 0.9 | 1.0 | 0.5 | 0.80 | 4.20 | 105.00 |
| France | 0.7 | 0.6 | 0.2 | 0.0 | 0.90 | 2.40 | 60.00 |
| UK | 0.8 | 0.7 | 0.2 | 0.0 | 0.85 | 2.55 | 63.75 |
| India | 0.95 | 0.8 | 1.0 | 0.5 | 0.75 | 4.00 | 100.00 |
| Africa | 0.3 | 0.2 | 1.0 | 0.5 | 0.70 | 2.70 | 67.50 |
| ... | ... | ... | ... | ... | ... | ... |  |

...

(Weights are illustrative; normalize to sum to 1.0 across all countries.)

Adjust current counts (e.g., reduce USA from 87 to 91, increase Africa from 1 to 67) by seeking more documents from underrepresented areas.

Thoughts on Your Idea

Your suggestion to weight by economic power and leadership in AI thinking is valuable because:

Economic Power reflects a country’s capacity to influence global AI policy and innovation.

AI Leadership ensures you capture cutting-edge perspectives.

However, relying solely on these could perpetuate the current bias toward G7/G20 nations. Adding population and diversity weights addresses this by:

Giving voice to populous nations (e.g., India, China) that may not yet lead in AI but have significant cultural influence.

Ensuring underrepresented regions (e.g., Africa, Latin America) contribute to a global moral code.

Additional Considerations

Sector Balance: Your current data may over-represent Government documents (e.g., from USA, China, France). Actively seek Academia, NGO, and Religious perspectives, especially from diverse regions.

Qualitative Influence: Beyond quantity, assess document influence (e.g., citations, adoption by international bodies like UN or OECD) to prioritize quality over sheer numbers.

Dynamic Adjustment: Revisit weights annually as AI governance evolves (e.g., Africa’s AI ecosystem is growing rapidly).

Final Recommendation

Adopt a weighted sampling approach combining economic power, AI leadership, population, and diversity, with a minimum regional threshold. Start by collecting data for the weighting factors, calculate target document counts, and adjust your retrieval process to close gaps (e.g., target 5-10 documents from Africa, Latin America, and the Middle East). Collaborate with global networks (e.g., UNESCO, World Bank) to identify influential but underreported documents. This will help you build a more balanced and comprehensive "AI Moral Code" that reflects worldwide perspectives.

Let me know if you’d like help calculating specific weights or refining the process further!