| Problem Chosen | 2026HSB MCM/ICM Summary Sheet | Team Number |
|---|---|---|
| B | | MI2601258 |

# w This Is Title


## Summary

Here is the summary of paper.

Please replace this text with your own summary. This summary should briefly describe the problem you are addressing, the methods you used, and the key results or conclusions you reached. Make sure to keep it concise and informative.



**Keywords:** KFC V50 Chicken

# Contents

# 1 Intorduction

## 1.1 Background

In the contemporary era, artificial intelligence (AI) has emerged as one of the core domains of global technological competition, exerting profound and systemic influences on economic development, social progress, and national security. With the acceleration of a new wave of technological revolution and industrial transformation, AI technologies are fundamentally reshaping traditional industrial structures, modes of production, and governance systems, and have gradually become a key indicator of a nation's scientific strength and overall competitiveness.

Against this backdrop, countries around the world have elevated artificial intelligence to a strategic priority at the national level, continuously increasing investments in algorithmic research, computing infrastructure, data resource development, and the expansion of application scenarios, with the aim of securing a leading position in the global AI competitive landscape.

## 1.2 Problem Restatement

This study aims to quantitatively evaluate national artificial intelligence (AI) development capabilities, compare global competitiveness, and analyze future development trends through a systematic mathematical modeling framework. The problem is decomposed into four sequential and interrelated tasks:

**Task 1: Factor Identification and Correlation Analysis**

Relevant data are collected and integrated to identify the key factors influencing national AI development. These factors are quantified, and their intrinsic correlations and interaction mechanisms are analyzed using statistical and visualization methods.

**Task 2: Comprehensive Evaluation and Ranking**

Based on the quantified factors and their correlations obtained in Task 1, a multi-criteria evaluation model is constructed to assess and rank the AI competitiveness of ten selected countries.

**Task 3: Competitiveness Trend Prediction**

Using historical data from 2016 to 2025, the future evolution of AI development factors during the period 2026–2035 is predicted. The evaluation model established in Task 2 is then applied to analyze the dynamic changes in national competitiveness rankings over time.

**Task 4: Optimal Fund Allocation Strategy**

Under a fixed budget constraint of a 1 trillion yuan special fund, a multi-objective optimization model is developed to determine the optimal allocation of resources across AI development factors, with the goal of maximizing China's comprehensive AI competitiveness by 2035.

By sequentially accomplishing these tasks, this study provides a coherent framework for factor identification, comparative evaluation, future trend analysis, and strategic decision support in the global AI competition landscape.

## 1.3 Our Work

这是我们的工作介绍

# 2 Basic Assumption

To ensure the feasibility, consistency, and interpretability of the proposed models, the following basic assumptions are made.

▸ **Hypothesis 1: Assume that national AI development capability is a latent attribute that can be approximated by a finite set of observable and quantifiable indicators.**

**Legitimacy:** At the national level, AI development is manifested through measurable outcomes and resource inputs recorded in public statistics. Although the true capability cannot be observed directly, its major characteristics can be reasonably inferred from aggregated, quantifiable indicators.

▸ **Hypothesis 2: Assume that all indicators within the same evaluation year are cross-sectionally consistent.**

**Legitimacy:** Although data may be collected from slightly different release years, AI development is a long-term process. Minor temporal discrepancies do not significantly affect national-level competitiveness comparisons and help simplify the modeling process.

▸ **Hypothesis 3: Assume that the indicators are independent of each other in the weighting and evaluation stages.**

**Justification :**While interactions among indicators exist, explicitly modeling such dependencies would increase complexity and reduce interpretability. Treating indicators as independent avoids double counting and ensures the applicability of entropy-based and multi-criteria evaluation methods.

▸ **Hypothesis 4: Assume that the fundamental mechanisms of AI development remain stable during the forecasting and optimization period.**

**Justification :**National AI strategies, infrastructure construction, and talent cultivation generally evolve gradually. This stability makes trend-based prediction and investment optimization reasonable and analytically tractable.

# 3 Symbols

# 4 Data Explanation

## 4.1 Data Description and Sources

The dataset covers ten representative countries and consists of multiple quantitative indicators describing national AI development capability. For organizational clarity, the indicators were grouped into six dimensions reflecting different aspects of AI development. All data corresponded to the same evaluation period and were obtained from publicly available and widely recognized sources, ensuring cross-country comparability.

## 4.2 Data Preprocessing

Basic preprocessing was conducted prior to analysis. Minor missing or abnormal values were handled through reasonable estimation and consistency checks. All indicators were defined as benefit-type variables and were normalized to eliminate dimensional differences before being used in subsequent models.

# 5 Task 1: Identification and Structural Analysis of AI Development Factors

## 5.1 Factor System and Quantification

To quantitatively describe national AI development capability, a multi-factor indicator system is constructed and organized into a standardized data matrix. Let

$$X = [x_{ij}] \in \mathbb{R}^{n\times p} \tag{1}$$

denote the standardized indicator matrix, where $x_{ij}$ represents the normalized value of factor $j$ for country $i$, with $n = 10$ countries and $p = 24$ indicators.

To eliminate scale effects and ensure cross-country comparability, min–max normalization is applied:

$$x_{ij} = \frac{x_{ij}^{\text{raw}} - \min(x_j)}{\max(x_j) - \min(x_j)} \in [0, 1]. \tag{2}$$

The standardized matrix $X$ is directly used as the input for all subsequent structural analyses.

## 5.2 Correlation Structure among AI Development Factors

Correlation analysis is a fundamental tool for exploring linear relationships among multiple factors and is widely used in exploratory structural analysis. In this task, Pearson correlation coefficients are employed to examine the statistical dependence between AI development factors.

The correlation coefficient between factor $j$ and factor $k$ is defined as

$$r_{jk} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x} * j)(x * ik - \bar{x} * k)}{\sqrt{\sum *i = 1^n (x_{ij} - \bar{x} * j)^2 \sum *i = 1^n (x_{ik} - \bar{x} * k)^2}}, \tag{3}$$

forming the correlation matrix

$$R = [r * jk] \in \mathbb{R}^{p\times p}. \tag{4}$$

To highlight the most significant relationships, a strong-correlation edge set is defined as

$$\mathcal{E} = (j, k) \mid |r_{jk}| > \tau,\ j < k, \tag{5}$$

where $\tau$ is a predefined threshold.

**Results and Interpretation.** The Pearson correlation analysis shows that AI development factors exhibit a highly interconnected structure. A large number of indicator pairs display strong positive correlations, indicating substantial co-movement among factors across countries.

Core indicators related to talent, market scale, policy intensity, investment, and infrastructure form dense correlation patterns. This overall correlation structure suggests that AI development capability is characterized by strong system-level linkage rather than independent factor variation.

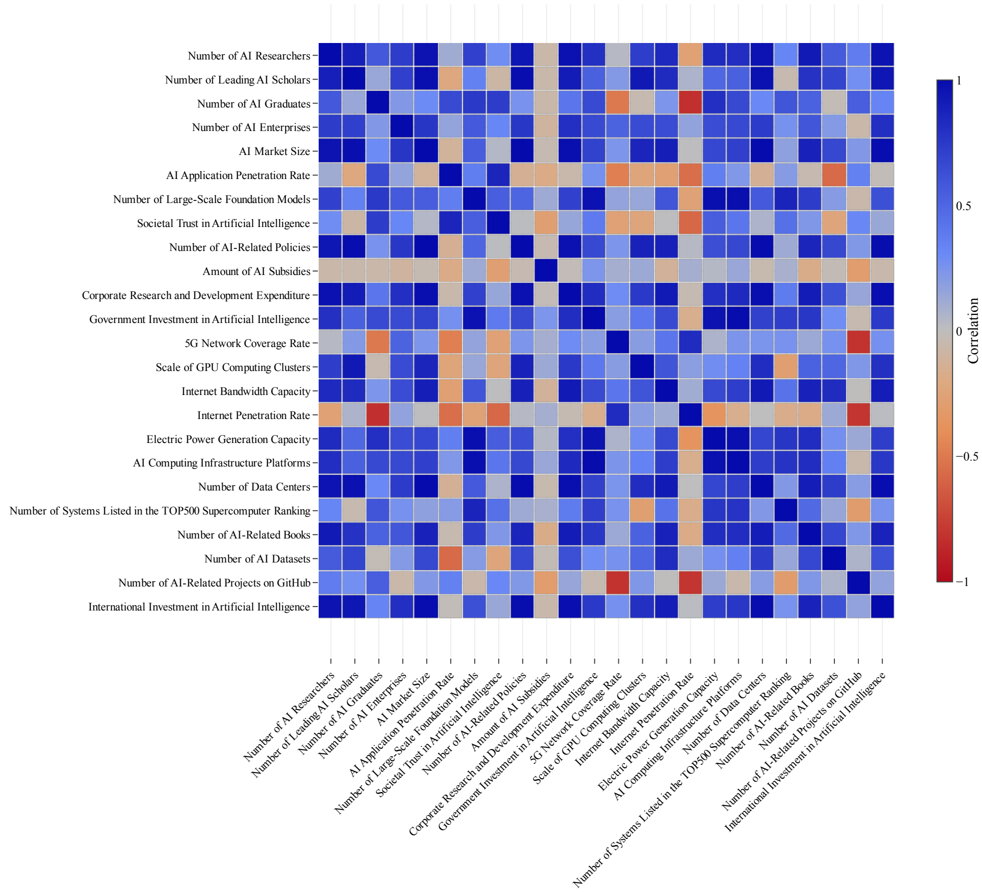


Figure 1: Correlation Heatmap of AI Development Factors

## 5.3 Structural Grouping via Hierarchical Clustering

While correlation analysis reveals pairwise relationships, clustering methods are commonly used to uncover higher-level structural organization among factors. To this end, correlation-based distances are introduced to reflect similarity in factor behavior.

The distance between factor $j$ and factor $k$ is defined as

$$d_{jk} = 1 - |r_{jk}|, \tag{6}$$

yielding the distance matrix $D = [d_{jk}]$.

Using the average linkage criterion, the distance between two clusters $C_a$ and $C_b$ is computed as

$$D(C_a, C_b) = \frac{1}{|C_a||C_b|} \sum_{j \in C_a} \sum_{k \in C_b} d_{jk}. \tag{7}$$

Hierarchical clustering is then applied to identify data-driven groupings within the AI development factor system.

**Results and Interpretation.** The hierarchical clustering dendrogram reveals that AI development factors are grouped into several coherent clusters. Indicators within each cluster exhibit similar statistical patterns, indicating consistent structural grouping across countries.

Clusters broadly correspond to scale and investment factors, computing and infrastructure factors, and talent-related factors, providing a data-driven partition of the AI development indicator system.

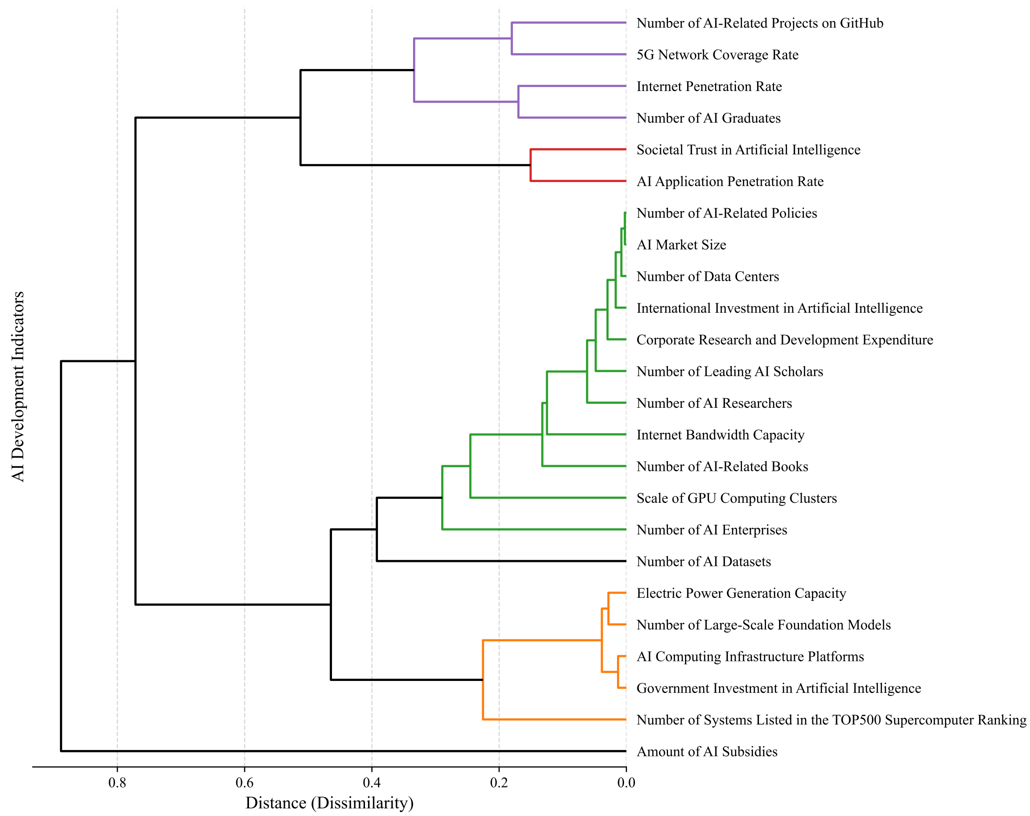


Figure 2: Hierarchical Clustering Dendrogram of AI Development Factors

## 5.4 Principal Component Analysis of Factor Structure

Given the high dimensionality and potential redundancy among factors, principal component analysis (PCA) is employed to extract a reduced set of comprehensive components. PCA is widely used for dimensionality reduction while preserving the dominant structural information in multivariate data.

The centered data matrix is given by

$$\tilde{X} = X - \mathbf{1}\bar{X}^T, \tag{8}$$

and the covariance matrix is

$$C = \frac{1}{n-1}\tilde{X}^T\tilde{X}. \tag{9}$$

Eigen-decomposition of $C$ yields

$$C = V\Lambda V^T, \tag{10}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. The number of retained components $m$ is determined by the cumulative variance criterion

$$\sum_{k=1}^{m} \frac{\lambda_k}{\sum_{j=1}^{p} \lambda_j} \geq \eta, \tag{11}$$

where $\eta$ denotes the predefined threshold.

**Results and Interpretation.** The PCA variance contribution results show that the first principal component explains 52.1% of the total variance. The first four principal components together account for approximately 91.4% of the cumulative variance.

According to the predefined threshold, four principal components are retained to represent the dominant structure of the original indicator set.

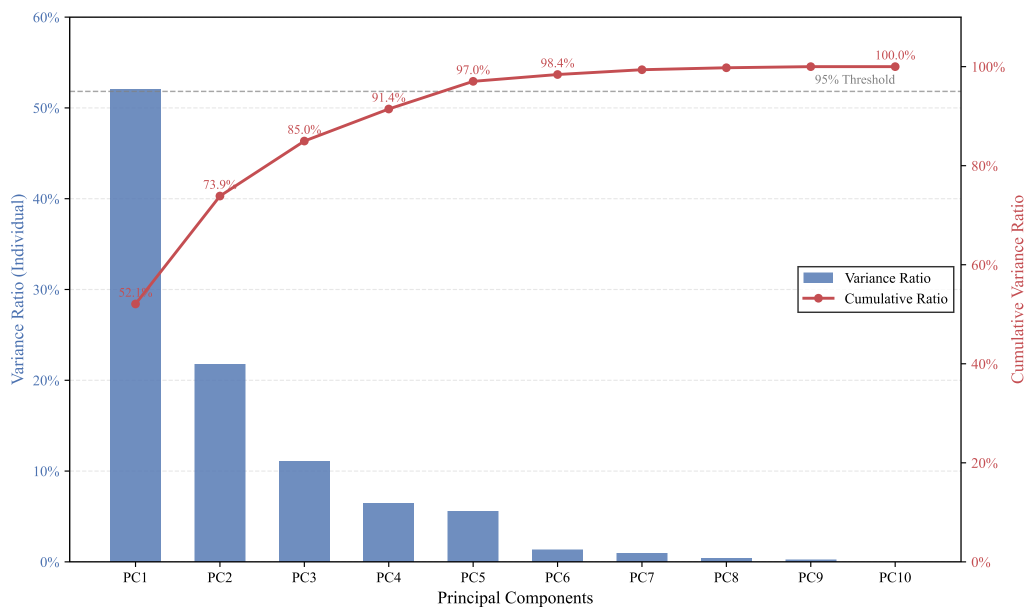


Figure 3: Variance Explained Plot for Principal Components

## 5.5 Relative Importance of AI Development Factors

To identify key driving factors within the system, the relative importance of each factor is quantified based on PCA results. This approach integrates factor loadings and component contributions to reflect the overall influence of each indicator.

The importance of factor $j$ is defined as

$$I_j = \sum_{k=1}^{m} v_{jk}^2 \cdot \frac{\lambda_k}{\sum_{l=1}^{p} \lambda_l}, \tag{12}$$

where $v_{jk}$ denotes the loading of factor $j$ on principal component $k$.

**Results and Interpretation.** The factor importance ranking shows that explanatory power is concentrated in a limited number of indicators. The Top–k factors account for a substantial proportion of the overall structural variance.

High-importance indicators consistently load on the leading principal components, indicating that a small subset of factors dominates the reduced representation of AI development capability.

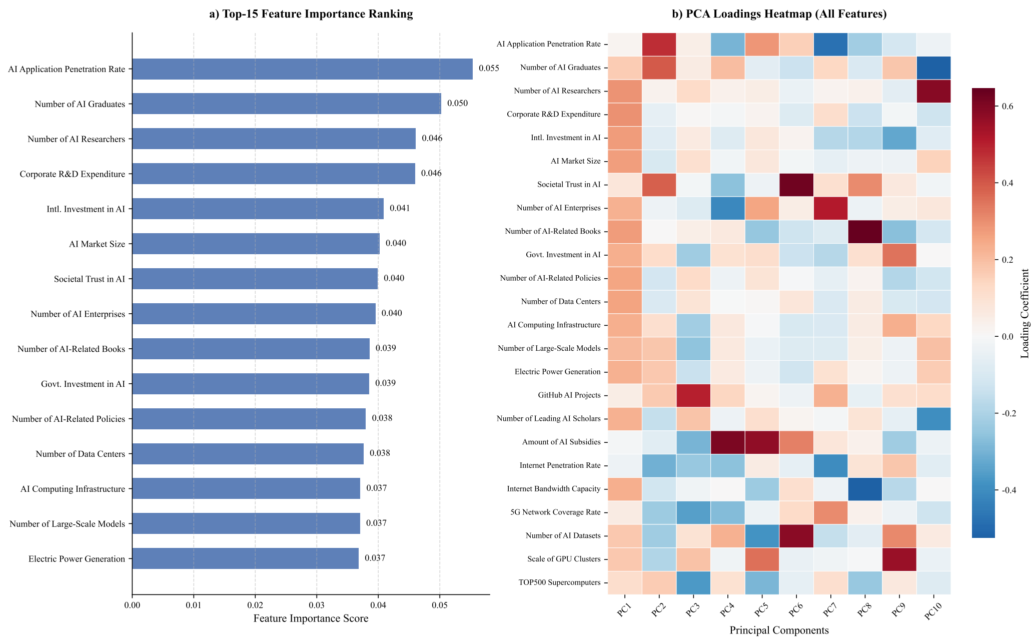


Figure 4: Factor Importance Ranking Bar Chart

## 5.6 Interaction Patterns among AI Development Factors

Based on the strong correlation structure, clustering results, and factor importance rankings, interaction patterns among AI development factors are summarized. Rather than establishing strict causal relationships, this analysis focuses on statistically supported promotion and constraint patterns within the factor system.

**Results and Interpretation.** The strong-correlation network exhibits a centralized structure, with a small number of indicators forming dense hubs connected to multiple other factors.

Combined with clustering results, the network suggests that AI development factors interact through several tightly coupled structural modules rather than isolated pairwise relationships.

## 5.7 Task 1 Summary

Task 1 presents a unified structural view of national AI development capability by integrating correlation analysis, hierarchical clustering, principal component analysis, and interaction network analysis. Rather than interpreting individual results in isolation, these methods jointly characterize the internal organization of the AI development system.

Using a standardized indicator framework, Task 1 identifies strong system-level linkages among AI development factors, showing that capability differences arise from coordinated patterns across talent, investment, policy, and infrastructure dimensions. The results indicate that these factors form coherent structural groups and exhibit substantial interdependence.

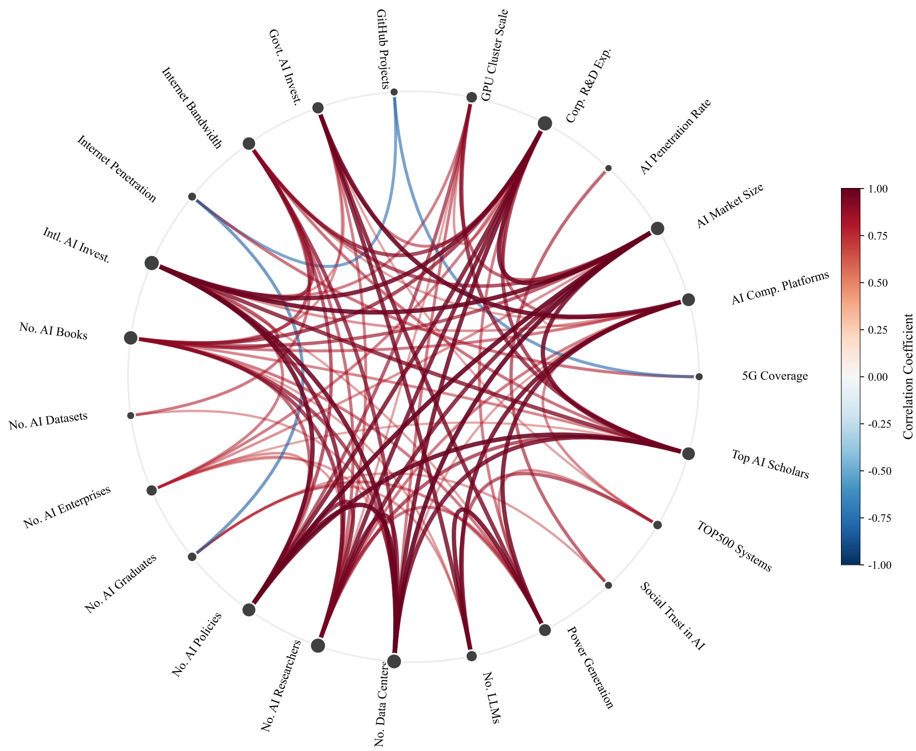


Figure 5: Strong-Correlation Network or Chord Diagram

Dimensionality reduction further demonstrates that the high-dimensional factor space can be effectively represented by a small number of comprehensive components, with explanatory power concentrated in a limited subset of key indicators. Interaction analysis confirms that these dominant factors occupy central positions within the system structure.

Overall, Task 1 clarifies both the composition and structural characteristics of AI development capability, providing a compact and consistent analytical foundation for the subsequent tasks of evaluation, forecasting, and investment optimization.

# 6 Task 2: AI Development Capability Evaluation and 2025 Ranking

## 6.1 Model Overview

Based on the 24 indicators identified in Task 1, let the normalized indicator matrix be

$$X' = \left(x'_{ij}\right)_{n\times p}, \quad n = 10,\ p = 24, \tag{13}$$

where $x'_{ij}$ denotes the normalized value of indicator $j$ for country $i$. All indicators have been unified as benefit-type (larger values indicate stronger AI development capability) and normalized prior to this task.

The objective of Task 2 is to construct an *objective and reproducible* evaluation model to quantify national AI development capability and determine the 2025 competitiveness ranking of ten countries.

To reduce subjective bias and enhance robustness, an integrated evaluation framework combining the Entropy Weight Method (EWM), TOPSIS, and Grey Relational Analysis (GRA) is adopted.

## 6.2 Entropy-Based Weighting

The entropy weight method is derived from information theory and assigns indicator weights according to their dispersion across countries. Indicators with higher variability contain more effective information for discrimination and thus receive larger weights.

Define

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}}, \qquad 0 \ln 0 := 0, \tag{14}$$

where $p_{ij}$ represents the proportion of indicator $j$ contributed by country $i$. The information entropy of indicator $j$ is computed as

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln p_{ij}, \quad k = \frac{1}{\ln n}. \tag{15}$$

The corresponding entropy weight is given by

$$w_j = \frac{1 - e_j}{\sum_{j=1}^{p} (1 - e_j)}. \tag{16}$$

If $\sum_{i=1}^{n} x'_{ij} = 0$ for some indicator $j$, its weight is set to zero and the remaining weights are renormalized.

*(Result observation and interpretation to be added here.)*

## 6.3 TOPSIS-Based Comprehensive Evaluation

TOPSIS aggregates weighted indicators into a single competitiveness score by comparing each country with an ideal reference. The method assumes that the optimal country should be closest to the positive ideal solution and farthest from the negative ideal solution.

The weighted decision matrix is defined as

$$v_{ij} = w_j x'_{ij}. \tag{17}$$

For each indicator $j$, the positive and negative ideal components are defined as

$$A_j^+ = \max_i v_{ij}, \quad A_j^- = \min_i v_{ij}, \qquad j = 1, \dots, p, \tag{18}$$

yielding the ideal vectors $A^+ = (A_1^+, \dots, A_p^+)$ and $A^- = (A_1^-, \dots, A_p^-)$.

The distances from country $i$ to the ideal solutions are computed as

$$D_i^\pm = \sqrt{\sum_{j=1}^{p} (v_{ij} - A_j^\pm)^2}, \tag{19}$$

and the comprehensive competitiveness score is defined as

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}. \tag{20}$$

*(Result observation and interpretation to be added here.)*

## 6.4 Structural Validation by Grey Relational Analysis

Grey Relational Analysis is suitable for small-sample and partially known systems. It evaluates the similarity of development patterns by comparing the geometric proximity of indicator sequences.

Let the reference sequence be the ideal profile

$$v_{0j} = \max_i v_{ij}, \qquad j = 1, \dots, p. \tag{21}$$

Define the absolute deviation

$$\Delta_{ij} = |v_{0j} - v_{ij}|, \tag{22}$$

and let

$$\Delta_{\min} = \min_{i,j} \Delta_{ij}, \qquad \Delta_{\max} = \max_{i,j} \Delta_{ij}. \tag{23}$$

The grey relational coefficient is defined as

$$\xi_{ij} = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{ij} + \rho\Delta_{\max}}, \quad \rho = 0.5, \tag{24}$$

and the grey relational degree of country $i$ is calculated as

$$\gamma_i = \frac{1}{p}\sum_{j=1}^{p} \xi_{ij}. \tag{25}$$

*(Result observation and interpretation to be added here.)*

## 6.5 Fusion Ranking and Reliability Framework

To integrate distance-based performance and structural similarity, an equal-weight fusion score is defined as

$$S_i = \frac{C_i + \gamma_i}{2}. \tag{26}$$

The fusion score $S_i$ is used to determine the final AI competitiveness ranking for 2025.

Ranking consistency is quantified using Spearman's rank correlation coefficient

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{27}$$

where $d_i$ denotes the rank difference of country $i$ between two ranking methods. Robustness is further assessed by perturbing a single indicator weight $w_j$ by $\pm\alpha$ (e.g., $\alpha = 30\%$), followed by weight renormalization and re-ranking.

*(Final ranking results and reliability analysis to be added here.)*

### 6.6 Overall Summary of Task 2

Task 2 establishes a comprehensive evaluation framework for national AI development capability by integrating objective weighting, distance-based evaluation, and structural validation. The framework provides a consistent and reproducible basis for generating the 2025 competitiveness ranking, which serves as the benchmark for subsequent analysis.

*(Overall result summary and interpretation to be added here.)*

## 7 Task 3: AI Competitiveness Ranking Forecast (2026–2035)

### 7.1 Model Objective and Overall Framework

Building upon the indicator system and structural insights identified in Task 1, as well as the objective evaluation framework established in Task 2, Task 3 focused on forecasting the evolution of national AI competitiveness rankings over the period 2026–2035. The primary objective was to construct a *forecast–evaluation integrated pipeline* that remained fully reproducible, interpretable, and strictly consistent with the previously defined evaluation model.

Let $x_{i,j,t}$ denote the observed value of indicator $j$ for country $i$ in year $t$. Using the complete panel data from 2016 to 2025, Task 3 was implemented through three sequential stages:

1. indicator-level time series forecasting for each country and indicator;
2. annual comprehensive evaluation using the fixed weighting and TOPSIS framework from Task 2;
3. ranking evolution and driver analysis across the forecast horizon.

This design ensured that all ranking changes were driven exclusively by data-inferred temporal dynamics, without introducing any exogenous assumptions or subjective adjustments.

### 7.2 Data Structure and Notation

The data used in Task 3 followed a balanced panel structure:

- Countries: $i = 1, \ldots, n$, with $n = 10$;
- Indicators: $j = 1, \ldots, p$, with $p = 24$;
- Historical years: $t \in \{2016, \ldots, 2025\}$;
- Forecast years: $t \in \{2026, \ldots, 2035\}$.

For each forecast year $t$, the predicted indicator matrix was denoted as

$$\hat{X}_t = \left(\hat{x}_{i,j,t}\right)_{n\times p}, \tag{28}$$

which served as the direct input to the annual evaluation procedure.

To preserve physical interpretability and comparability across indicators, indicator-specific constraints were imposed:

- non-negativity constraints for scale and quantity indicators;
- bounded intervals for ratio- or proportion-type indicators.

## 7.3 Indicator Forecasting Methodology

### 7.3.1 Baseline Grey Forecasting Model

Given the limited historical sample size available for each indicator sequence, the Grey Model GM(1,1) was adopted as the baseline forecasting approach. For a fixed country–indicator pair $(i, j)$, the original sequence was defined as

$$X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(T)\right), \quad T = 10. \tag{29}$$

After applying first-order accumulated generation (AGO), a first-order differential equation was fitted and solved to obtain the corresponding time response function. The predicted original sequence was then recovered via inverse AGO. The GM(1,1) model was selected for its robustness in small-sample settings and its suitability for capturing monotonic development trends.

### 7.3.2 Engineering Constraints and Stabilization

To ensure numerical stability and realistic forecasts across heterogeneous indicators, several unified treatments were applied:

- non-negative translation for near-zero or zero-valued series prior to model fitting;
- boundary truncation of forecast values according to indicator type;
- optional logarithmic transformation for indicators with large magnitude dispersion.

All transformations were rule-based and reversible, and no country-specific adjustments were introduced at any stage.

### 7.3.3 Backtesting and Fallback Strategy

Model suitability was evaluated via one-step-ahead backtesting. Specifically, data from 2016–2024 were used for training, and the year 2025 was forecast for validation. Prediction error was measured using the mean absolute percentage error (MAPE).

*Result placeholder.*

**To be added:** summary statistics of backtesting errors across all country–indicator pairs, using data from `forecast_diagnostics.csv`.

**Suggested table:** Three-line table reporting the distribution of MAPE values and the proportion of series exceeding the predefined threshold.

When the GM(1,1) model failed to satisfy accuracy or stability requirements, a fallback predictor (linear trend or Holt exponential smoothing) was automatically activated, subject to the same boundary constraints.

## 7.4 Annual Evaluation Using Fixed Weights and TOPSIS

### 7.4.1 Weight Inheritance Principle

Let $W = (w_1, \ldots, w_p)$ denote the entropy-based indicator weights obtained in Task 2. Throughout Task 3, these weights were held constant across all forecast years to ensure temporal comparability and methodological consistency.

### 7.4.2 Yearly TOPSIS Evaluation

For each forecast year $t$, the predicted matrix $\hat{X}_t$ was normalized and weighted using the same TOPSIS procedure as in Task 2. The relative closeness score for country $i$ in year $t$ was denoted as

$$C_{i,t} \in [0, 1]. \tag{30}$$

*Result placeholder.*
**To be added:** annual TOPSIS scores for all countries during 2026–2035, based on `topsis_scores_2026_2035.cs`
**Suggested table:** Three-line table showing selected years (e.g., 2026, 2030, 2035) with corresponding scores for each country.
**Suggested figure:** Line plot illustrating the temporal evolution of $C_{i,t}$ for major economies.

## 7.5 Ranking Generation and Evolution Analysis

For each year $t$, countries were ranked in descending order of $C_{i,t}$:

$$\mathrm{Rank}_{i,t} = \mathrm{rankdesc}\left(C_{:,t}\right). \tag{31}$$

*Result placeholder.*
**To be added:** complete ranking results for 2026–2035 using `rankings_2026_2035.csv`.
**Suggested table:** Three-line table summarizing average rank, maximum rise/fall, and rank standard deviation for each country.
**Suggested figure:** Ranking evolution plot or heatmap based on `rankings_2026_2035_wide.csv`.

## 7.6 Driving Factor Analysis of Ranking Changes

To interpret ranking dynamics, changes in the TOPSIS score were decomposed with respect to high-weight indicators. For a fixed country, score variations were analyzed in conjunction with changes in weighted normalized indicator values.

*Result placeholder.*
**To be added:** identification of key indicators contributing to score changes for representative countries.
**Data source:** weighted normalized matrices derived from predicted indicators and Task 2 weights.
**Suggested figure:** Contribution bar charts or radar plots highlighting dominant drivers.

## 7.7 Model Validation and Robustness Checks

Model robustness was examined from both the forecasting and evaluation perspectives:

- comparison between predicted and observed 2025 evaluation scores to validate the end-to-end pipeline;
- sensitivity of rankings to fallback thresholds and minor perturbations in high-weight indicators.

*Result placeholder.*
**To be added:** robustness assessment results using sensitivity scenarios.
**Suggested table:** Three-line table reporting rank correlations under different perturbation settings.

### 7.8 Summary of Task 3

Task 3 extended the static evaluation framework into a dynamic forecasting context by integrating indicator-level time series prediction with a fixed and validated evaluation model. The resulting rankings reflected endogenous development trajectories inferred from historical data, rather than externally imposed assumptions.

Overall, the proposed pipeline preserved methodological continuity across tasks, maintained interpretability at both the indicator and system levels, and provided a quantitative basis for analyzing future shifts in global AI competitiveness. These results established a consistent foundation for the subsequent policy-oriented optimization analysis.

## 8 Task 4