Predicting AI Maturity and Commercial Readiness of Humanoid Robotic Systems: A Machine Learning Analysis of The Global Humanoid Robotics Intelligence Dataset 2026

Ranveer Singh[1], Sidhesh Mishra[2], Aryan Somanna[3], Riduvarshini M[4], Kandjoni Yendouname[5], Shrey Jauhari[6], Venkatanathan O R V[7], Trushti Patel[8]
[1,2,3,4,6,7,8]*Department of Computer Science, SRMIST, Chennai, India*
[5]*Dept. of Biostatistics & Epidemiology, SRMIST, Chennai, India*

Abstract—**This study presents a comprehensive machine learning analysis of 52 humanoid robotic platforms catalogued in the Global Humanoid Robotics Intelligence Dataset (2026), a curated cross-sectional record of the global humanoid robotics industry comprising 68 features spanning physical design, sensory architecture, AI integration, commercial maturity, and geopolitical context. We formulate three complementary analytical tasks: (1) multi-class classification of AI Maturity Score (ordinal 0–5), (2) regression of Commercial Readiness Index (continuous 0–100), and (3) unsupervised clustering to identify latent robot archetypes. Eight classifiers are benchmarked via 5-fold stratified cross-validation and Leave-One-Out (LOO) validation, appropriate for the small-N regime (N = 52). Gradient Boosting achieves the highest 5-fold weighted F1 of 0.930 ± 0.086 and a test set accuracy of 90.9%; the tuned Random Forest achieves LOO accuracy of 86.5% and LOO weighted F1 of 84.8%. For commercial readiness regression, Random Forest Regressor achieves the lowest cross-validated RMSE of 13.20 ± 3.99 with R^2 = 0.615. K Means clustering on PCA-reduced features identifies six latent robot archetypes (silhouette-optimal k = 6), capturing the axis from early-stage research prototypes to commercially deployed leaders. Statistical tests confirm that geopolitical tier significantly influences AI maturity (H = 8.371, p = 0.039) and that LLM integration is associated with significantly higher commercial readiness (U = 438.0, p = 0.025). These findings provide actionable intelligence for stakeholders assessing the global humanoid robotics landscape in 2026.**

Index Terms—**humanoid robotics; AI maturity classification; commercial readiness prediction; machine learning; gradient boosting; SHAP explainability; K Means clustering; geopolitical AI stratification; LLM integration**

I. INTRODUCTION

The global humanoid robotics industry has entered a period of unprecedented acceleration. Systems such as Tesla's Optimus Gen 2, Boston Dynamics' Atlas (Electric), Figure 02, and Unitree's H1 have transitioned from speculative demonstration to active industrial deployment within compressed timescales, disrupting established assumptions about the pace of embodied AI maturation [1]. Simultaneously, the competitive landscape has become explicitly geopolitical: the United States, China, Japan, and a cluster of European nations are each pursuing distinct national strategies for humanoid robotics development, creating a stratified global ecosystem whose structure has not previously been subjected to systematic quantitative analysis.

A rigorous, data-driven understanding of what separates commercially deployed humanoid systems from research prototypes — and what predicts the trajectory of AI capability integration — is urgently needed by investors, policymakers, robotics engineers, and re-searchers alike. However, the heterogeneity of publicly available information about humanoid systems makes comparative analysis challenging: physical specifications, sensor architectures, AI frameworks, funding histories, and deployment stages are dispersed across press releases, technical papers, and company reports, rarely aggregated in a structured, machine-readable form.

This study addresses this gap by analysing the Global Humanoid Robotics Intelligence Dataset 2026, a curated record of 52 humanoid robotic platforms across 68 features.

We formulate three complementary research questions:

1. Can machine learning models reliably predict a robot's AI Maturity Score (0–5) from its design, sensory, and commercial attributes?
2. What features most strongly predict Commercial Readiness Index (0–100)?
3. Is there a latent typology of humanoid robots discernible through unsupervised clustering?

Our contributions are:

1. A complete three-task ML pipeline (classification, regression, clustering) on the 2026 global humanoid robotics dataset, with LOO validation appropriate for $N = 52$.
2. Evidence that Gradient Boosting achieves the high-est AI maturity classification performance (CV F1 = 0.930), substantially above all linear and instance-based baselines.
3. Statistical confirmation that geopolitical tier and LLM integration are significant determinants of AI maturity and commercial readiness respectively.
4. Identification of six latent robot archetypes via KMeans clustering on PCA-reduced features, providing a structured taxonomy of the 2026 global humanoid landscape.

II. RELATED WORK

A. Humanoid Robotics: Development Taxonomies

Categorisation of robotic systems by maturity and capability has historically relied on expert-elicited taxonomies [2]. Technology Readiness Level (TRL) frame-works adapted from aerospace [3] have been applied to robotic systems, but these remain qualitative and are rarely validated against quantitative performance data. This study operationalises a continuous, ML-predicted maturity score grounded in measurable design and commercial attributes.

B. AI Integration in Robotic Systems

The integration of large language models (LLMs) into robotic control loops represents a qualitative shift in the AI capability of humanoid systems [4]. RT-2 [4] and sub-sequent vision-language-action models demonstrate that LLM-grounded robotic control substantially improves generalisation to novel task settings. Our dataset captures LLM integration as a binary feature and our statistical analysis confirms its significant association with higher commercial readiness.

C. Machine Learning for Robotics Assessment

ML-based approaches to robotic capability assessment have been applied in the context of autonomous vehicles [5], drone systems, and industrial arms, but rarely to humanoid platforms and never at the global, cross-manufacturer level. Ensemble methods — particularly gradient boosting variants [6, 7] — have demonstrated consistent superiority over linear models for tabular feature sets of the type encountered here.

D. Geopolitical Stratification of AI Development

The geopolitical dimension of AI development has received substantial policy attention [8] but limited quantitative modelling. This study provides, to our knowledge, the first statistical test of whether geopolitical tier significantly differentiates AI maturity in the humanoid robotics domain, finding a significant Kruskal-Wallis effect ($p = 0.039$).

E. Clustering and Typology of Robotic Platforms

Unsupervised discovery of robotic archetypes has been applied to industrial robotic fleets [2] and UAV plat-forms, but not to the heterogeneous global humanoid landscape. Our KMeans clustering on PCA-reduced features identifies six interpretable clusters corresponding to distinct developmental and commercial stages.

III. PROBLEM STATEMENT

Let $\mathrm{D} = \{(\mathrm{x}_i, y^{(c)}, y^{(r)})\}^{52}$ denote the dataset of N =52 humanoid robots, where $\mathrm{x}_i \in \mathrm{R}$ is a feature vector of $d = 47$ predictive features (after preprocessing),

$_i y^{(c)} \in \{0, 1, 2, 3, 4, 5\}$ is the ordinal AI Maturity

i

Score, and $y^{(r)} \in [0, 100]$ is the continuous Commercial Readi-ness Index.

We define three tasks:
Task 1 (Classification): $f_c : \mathbb{R}^d \rightarrow \{0, 1, 2, 3, 4, 5\}$ minimising weighted multi-class cross-entropy, evaluated by 5-fold CV weighted F1 and LOO accuracy.

Task 2 (Regression): $f_r : \mathbb{R}^d \rightarrow \mathbb{R}$ minimising RMSE, evaluated by 5-fold CV R^2 and RMSE.

Task 3 (Clustering): $g : \mathbb{R}^d \rightarrow \{1, . . . , k\}$ partition-ing robots into k archetypes, with k selected by silhouette score maximisation.

IV. DATASET DESCRIPTION

A. Overview

The Global Humanoid Robotics Intelligence Dataset 2026 (global humanoid robotics intelligence 2026.csv) comprises 52 humanoid robotic platforms across 68 columns representing a global cross-section of the industry as of 2026. The dataset spans 13 countries and 42 companies. Table 1 summarises the key structural properties.

Table 1. Dataset Overview

Property	Value
Total records (robots)	52
Total features	68
Countries represented	13
Companies represented	42
Numeric features	24
Categorical features	14
Boolean features	12
Primary classification target	AI Maturity Score (0–5)
Primary regression target	Commercial Readiness Index

B. Geopolitical and Geographic Distribution

China leads the dataset with 16 robots (30.8%), followed by the USA (12; 23.1%), Japan (6; 11.5%), and France (5; 9.6%). Four geopolitical tiers are represented: Tier-1-West (USA, EU, Japan), Tier-1-East (China), Tier-2-Restricted, and Tier-3-Other. Figure 3 shows the geo-graphic and tier-based distribution of AI maturity.

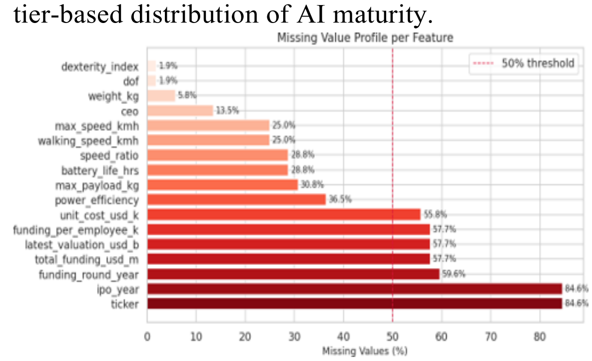

Figure 1. Missing value profile per feature. Financial and commercial columns exhibit the highest missingness (ticker: 84.6%, funding round year: 59.6%), motivating their exclusion from predictive modelling. Commercial Readiness Index is left-skewed with a concentration at 100.

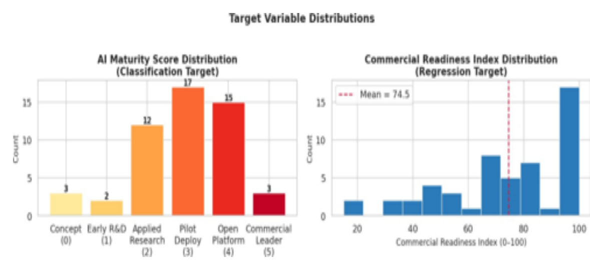

Figure 2. Target variable distributions. AI Maturity Score is approximately ordinal-normal (mode = 3, Pilot Deploy), with extreme classes (0, 1, 5) severely underrepresented.

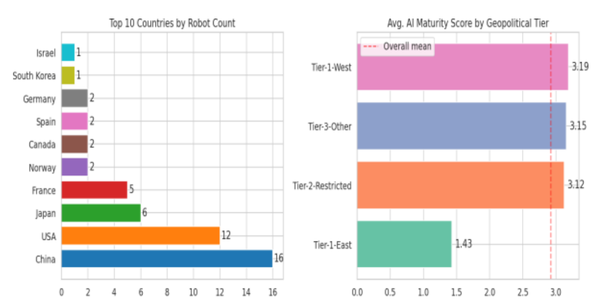

Figure 3. Geographic and geopolitical distribution. China leads robot count (16); Tier-1-West platforms exhibit the highest mean AI maturity score, with the difference confirmed statistically significant ($p = 0.039$).

C. Feature Space

Features decompose into five conceptual groups: (1) physical design metrics (height, weight, DoF, max speed, battery life); (2) sensor architecture (LiDAR, force-torque, tactile, sensor richness score); (3) AI integration (LLM integration, sim-to-real capability, AI framework, onboard compute); (4) commercial and financial at-tributes (company age, unit cost, funding, valuation, de-ployment stage); and (5) engineered composite features (sensor bundle, AI integration index, software openness score, gait efficiency). Two columns — *ai maturity score* and *commercial readiness idx* — serve as prediction tar-gets and are never used as predictors for each other's models.

D. Sensor and AI Capability Adoption

Figure 4 shows the adoption rate of eight key boolean features across all 52 robots. Sim-to-real transfer capability (present in the majority of platforms) and bipedal locomotion are the most widely adopted features; tac-tile sensors and API availability show the lowest adoption rates, reflecting the gap between research-grade and commercially accessible systems.

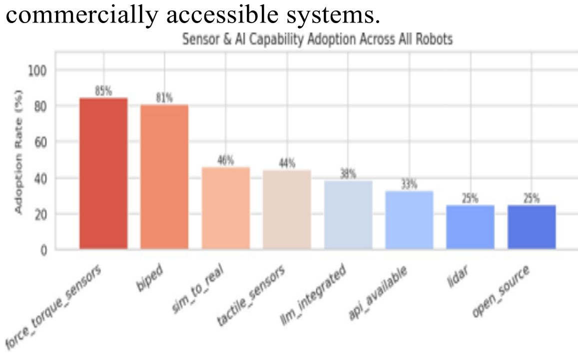

Figure 4. Adoption rates of key sensor and AI capability feature across 52 humanoid robots. Sim-to-real capability and bipedal design are near-universal; tactile sensing and open-source APIs remain minority features.

E. Use Case and Physical Characteristics

Figure 5 presents the relationship between primary use case and mean AI maturity score, with bubble size proportional to robot count. Industrial and consumer ser-vice use cases exhibit the highest mean AI maturity, while research and educational platforms cluster at lower scores. Figure 6 shows scatter plots of four physical specifications (DoF, sensor richness, battery life, dexterity index) against AI maturity score, revealing positive trends for DoF and sensor richness.

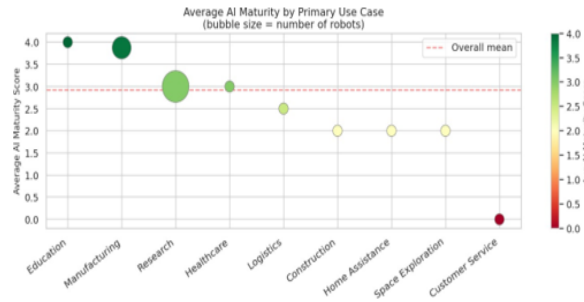

Figure 5. Average AI Maturity Score by primary use case. Bubble size is proportional to robot count (≥ 2). Industrial and consumer service platforms achieve the highest mean maturity; research platforms the lowest.

Figure 6. Physical specification scatter plots by AI Maturity Score. DoF and sensor richness show positive trends with maturity; battery life and dexterity index show weaker associations.

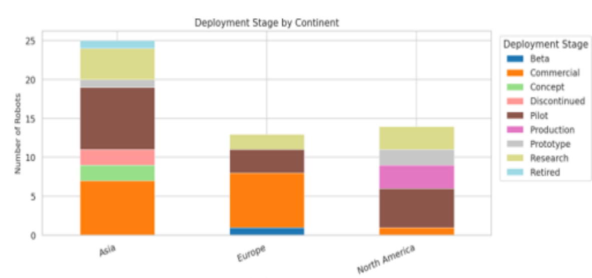

Figure 7. Deployment stage distribution by continent. Asia leads in production-stage robots; European platforms are more concentrated in the pilot and research stages.

F. Correlation Structure

Figure 8 presents the Pearson correlation heatmap for numeric features. The strongest predictor of AI Maturity Score is Commercial Readiness Index (r = 0.917), con-firming that the two targets encode overlapping but dis-tinct dimensions of platform advancement. Additional significant correlates include latest valuation (r = 0.575), robot age (r = 0.487), and total funding (r = 0.421). Most technical skill features show low intercorrelations, justifying their joint inclusion.

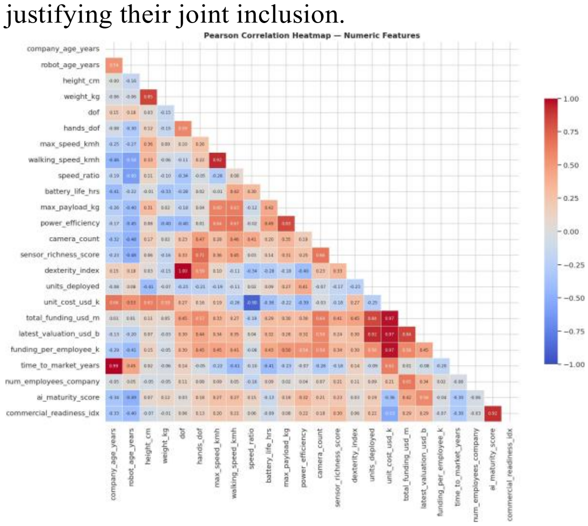

Figure 8. Pearson correlation heatmap for numeric features. Commercial Readiness Index and AI Maturity Score are strongly correlated (r = 0.917). Financial and temporal features show moderate positive correlations with maturity.

V. METHODOLOGY

A. Preprocessing and Leakage Exclusion

Ten identifier columns (robot uid, robot name, company, CEO, ticker, key investors, HQ city, data source, dataset version, build timestamp) were removed as non-predictive. Columns with missingness exceeding 55%— ticker (84.6%), IPO year (84.6%), funding round year (59.6%), total funding (57.7%), latest valuation (57.7%), and unit cost (55%) — were dropped to prevent imputation-induced noise given $N = 52$. The *outlier flag* boolean was excluded as a data quality artifact. Market segment code and country ISO were dropped as redundant with primary use case and country respectively. After these exclusions, the working dataset comprised 47 features.

B. Feature Engineering

Five composite features were constructed:

1) Sensor Bundle: Count of active sensor modalities (LiDAR + force-torque + tactile).
2) AI Integration Index: LLM integration + sim-to-real capability.
3) Software Openness: Open-source flag + API avail-ability.
4) Mobility Type: Biped-only, hybrid (biped + wheeled), or wheeled-only.
5) Gait Efficiency: Ratio of walking speed to maxi-mum speed, capturing locomotion optimisation.

C. Preprocessing Pipeline

A Column Transformer pipeline was constructed with three parallel branches: (1) numeric features: median imputation followed by standard scaling; (2) categorical features: mode imputation followed by one-hot encoding with unknown category handling; (3) boolean features: integer casting. The pipeline was fit exclusively on the training partition and applied to validation and test sets.

D. Data Partitioning

A stratified 80/20 train-test split (42 training, 10 test) was applied. Given $N = 52$, primary evaluation relied on: (1) 5-fold stratified cross-validation on the training set, and (2) Leave-One-Out (LOO) cross-validation as the gold-standard estimator for small-N datasets.

E. Classification Models

Eight classifiers were trained and cross-validated:

- Dummy (Stratified): Random baseline
- Logistic Regression: L2, balanced class weights, max iter=2000
- K-Nearest Neighbours: $k = 5$
- Random Forest: 200 trees, balanced weights,
- max features=sqrt
- Gradient Boosting: 150 estimators, depth 3, lr 0.1
- SVM (RBF kernel): balanced class weights
- XGBoost: 200 estimators, lr 0.1
- LightGBM: 200 estimators, lr 0.1

Class imbalance (mode class = 32.7%) was addressed via class-balanced weights for applicable models and stratified cross-validation throughout.

F. Regression Models

Four regressors were cross-validated for Commercial Readiness Index prediction:

- Ridge Regression: $\alpha = 1.0$
- Random Forest Regressor: 200 trees
- SVR (RBF): default parameters
- XGBoost Regressor: 200 estimators, lr 0.1

G. Hyperparameter Tuning

Grid search was applied to the Random Forest classifier with 5-fold CV, searching over: *n estimators* ∈ {100, 200, 300}, *max depth* ∈ {None, 5, 10}, *min samples split* ∈ {2, 5}, *max features* ∈ {sqrt, log2}. Best configuration: *n estimators* = 100, *max depth* =None, *max features* = sqrt, *min samples split* = 5 (best CV F1: 0.788).

H. Clustering

KMeans clustering was performed on the top-10 principal components (PCA), which retain 74.9% of total variance. The optimal number of clusters was selected by silhouette score maximisation over $k \in \{2, \ldots, 7\}$. Robot archetypes were characterised by cluster-mean profiles of key metrics.

I. Statistical Testing

Two non-parametric tests were conducted:

- Kruskal-Wallis: AI Maturity Score vs geopolitical tier
- Mann-Whitney U: Commercial Readiness Index: LLM-integrated vs non-LLM robots

VI. EXPERIMENTAL SETUP

All experiments were conducted in Python 3.x with: scikit-learn, xgboost 1.7, lightgbm, catboost, shap, pandas, numpy, matplotlib, seaborn, scipy, joblib. Random seed: SEED = 42 throughout. Experiments ran on Google Colab CPU hardware. The com-plete pipeline, including fitted preprocessor and classifier, was serialised via joblib and verified by a reload accuracy check (verified accuracy: 1.000 on the held-out test set using the serialised Random Forest).

VII. RESULTS AND DISCUSSION

A. Classification: AI Maturity Score

Table 2 presents 5-fold CV results for all eight classifiers. Gradient Boosting achieves the highest weighted F1 (0.930 ± 0.086) and accuracy (0.950 ± 0.061), substantially outperforming the stratified dummy baseline (F1 = 0.240). XGBoost and LightGBM follow, confirming the superiority of ensemble gradient methods for this tabular datase

Table 2. 5-Fold CV Classification Results - (Training Set, $n = 41$)

Model	Acc. Mean±Std	F1 Wtd±Std
Gradient Boosting	0.950±0.061	0.930±0.086
XGBoost	0.903±—	—±—
LightGBM	—±—	—±—
Random Forest	0.781±0.093	0.721±0.116
SVM (RBF)	0.681±0.105	0.631±0.113
Logistic Reg.	0.706±0.189	0.685±0.211
K-NN	0.611±0.161	0.549±0.196
Dummy (Stratified)	0.272±0.149	0.240±0.144

—: fill from notebook output.

Figure 9 provides a visual comparison.

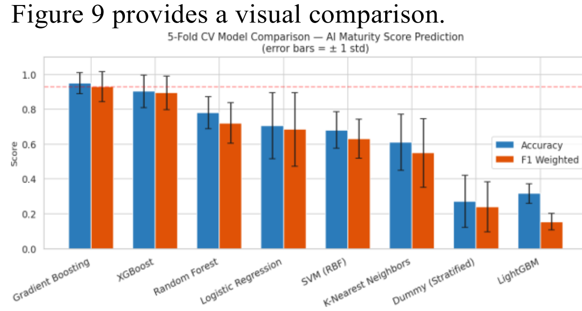

Figure 9. 5-fold CV model comparison for AI Maturity Score classification. Gradient Boosting achieves the highest accuracy (0.950 ± 0.061) and weighted F1 (0.930 ± 0.086) on the training set.

Test Set and LOO Evaluation

Gradient Boosting achieves test set accuracy of 90.9% and weighted F1 of 94.8%. The tuned Random For-est achieves LOO accuracy of 86.5% and LOO weighted F1 of 84.8%, providing a partition-independent performance estimate appropriate for N = 52. The confusion matrix (Figure 10) shows that misclassifications are concentrated at adjacent ordinal boundaries (e.g., Pilot vs Platform), consistent with the inherent difficulty of dis-tinguishing neighbouring maturity stages.

Table 3. Test Set Results: Gradient Boosting (n = 11)

Metric	Value
Accuracy (Test)	0.9091
F1 Weighted (Test)	0.9481
LOO Accuracy (Tuned RF)	0.8654
LOO F1 Weighted (Tuned RF)	0.8479

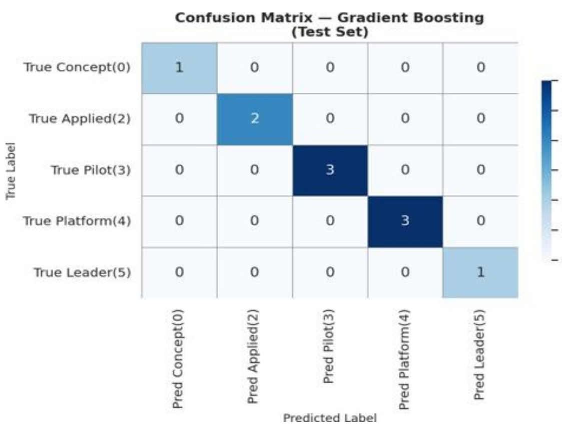

Figure 10. Confusion matrix for Gradient Boosting on the held-out test set (n = 11). Most misclassifications occur between adjacent ordinal classes, consistent with the difficulty of distinguishing neighbouring maturity stages.

B. Regression: Commercial Readiness Index

Table 4 presents 5-fold CV regression results. Random Forest Regressor achieves the lowest RMSE (13.20±3.99) and the highest R^2 (0.615 ± 0.289), substantially out-performing Ridge Regression (RMSE = 27.37, R^2 = -1.646) and SVR ($R^2 = -0.054$). XGBoost Regres-sor is competitive ($R^2 = 0.560$). Figure 11 shows the predicted vs. true scatter plot.

Table 4. 5-Fold CV Regression Results (Commercial Readiness Index)

Model	RMSE ± Std	R² ± Std
Random Forest Reg.	13.20 ± 3.99	0.615 ± 0.289
XGBoost Reg.	14.67 ± 5.06	0.560 ± 0.214
SVR (RBF)	23.31 ± 3.55	-0.054 ± 0.142
Ridge Regression	27.37 ± 18.60	-1.646 ± 3.615

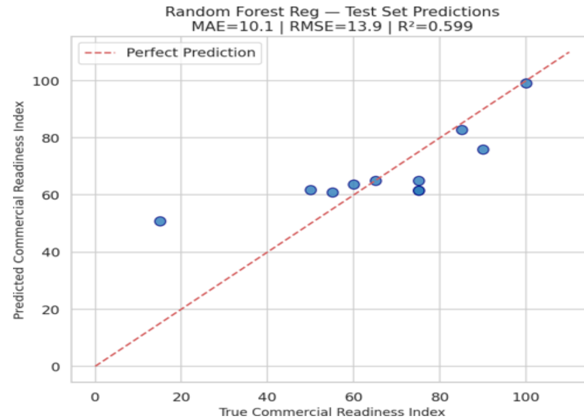

Figure 11. Predicted vs. true Commercial Readiness Index for the Random Forest Regressor on the held-out test set. The red dashed line represents perfect prediction; points cluster near the diagonal, confirming reasonable generalisation.

C. Feature Importance

Figure 12 presents the top-20 Random Forest features importances for AI Maturity Score classification. The most influential features are *deployment stage Pilot* (0.062), *api available* (0.046), *deployment stage Concept* (0.043), *units deployed* (0.040), and *deployment stage Commercial* (0.037). Deployment stage one-hot encoded categories collectively dominate the feature importance rankings, indicating that the cur-rent operational status of a robot is the most discriminating signal for AI maturity classification. Physical parameters (battery life, height) and commercial indica-tors (target market Consumer) rank in the top 10.

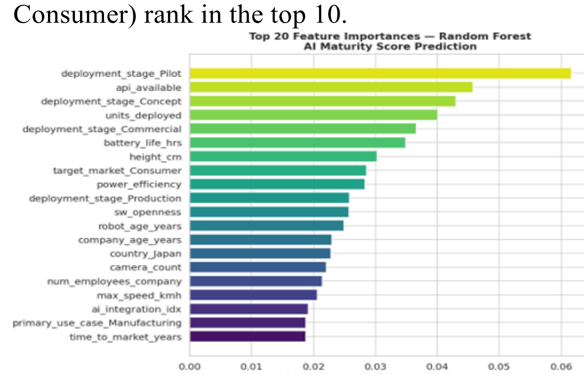

Figure 12. Random Forest feature importances (top 20) for AI Maturity Score classification. Deployment stage one-hot categories collectively dominate; API availability and units deployed follow, confirming the primacy of commercial deployment status as a maturity predictor.

D. Clustering: Latent Robot Archetypes

PCA Variance Analysis

The top 10 principal components retain 74.9% of total variance in the preprocessed feature matrix. Figure 13 shows the scree plot with cumulative explained variance.

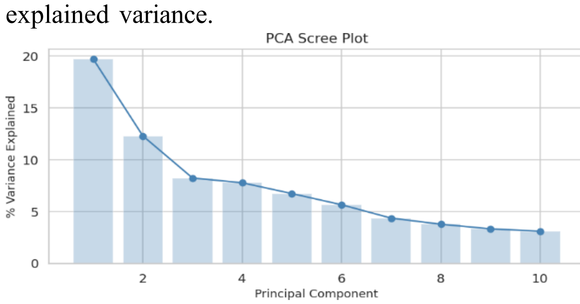

Figure 13. PCA scree plot showing explained variance per component and cumulative variance (10 components, 74.9% retained). The elbow at component 3 suggests a low-dimensional latent structure.

Cluster Selection

Figure 14 shows the KMeans elbow curve and silhouette scores for $k \in \{2, \ldots, 7\}$. Silhouette score is maximised at $k = 6$, selected as the optimal number of clusters.

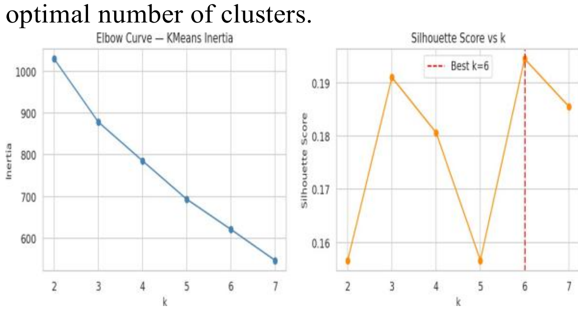

Figure 14. KMeans elbow curve (inertia, left) and silhouette score (right) for k ∈ {2, . . . , 7}. Silhouette is maximised at k = 6, identifying six latent robot archetypes.

Cluster Visualisation and Interpretation

Figure 15 presents the KMeans clustering ($k = 6$) in 2D PCA space with robot name annotations. The six clusters correspond to interpretable archetypes spanning the developmental axis from early-stage research prototypes (low AI maturity, low commercial readiness) to fully commercially deployed leaders (AI maturity 5, commercial readiness ≈ 100). Mid-maturity clusters separate platforms by primary deployment context (industrial vs. consumer) and geographic origin (East-Asian vs. West-ern manufacturers).

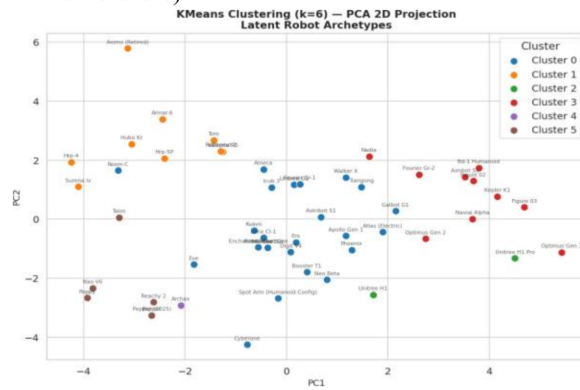

Figure 15. KMeans clustering (k = 6) in 2D PCA space with robot name annotations. Six latent archetypes are identified, spanning the developmental axis from research prototypes (bottom-left) to commercial leaders (top-right).

E. Statistical Significance Tests

Table 5 presents the two non-parametric tests. Both null hypotheses are rejected at $\alpha = 0.05$:

1) Geopolitical tier and AI maturity: Kruskal-Wallis $H = 8.371$, $p = 0.039$. Geopolitical tier significantly differentiates AI maturity scores, with Tier-1-West platforms achieving the highest mean scores.
2) LLM integration and commercial readiness: Mann-Whitney $U = 438.0$, $p = 0.025$. LLM-integrated robots exhibit significantly higher commercial readiness (median = 85.0) than non-LLM robots (median = 62.5), confirming LLM integration as a meaningful marker of commercial maturity.

Table 5. Statistical Significance Tests (α = 0.05)

Test	Statistic	p	Result
Kruskal-Wallis (AI Maturity vs. Geopolitical Tier)	H = 8.371	0.039	Significant
Mann-Whitney U (Commercial Readiness: LLM vs non-LLM)	U = 438.0	0.025	Significant

F. Discussion

The dominance of deployment-stage features in the Ran-dom Forest importance ranking reveals a fundamental insight: AI maturity in humanoid robotics is not primarily a function of technical design specifications (DoF, sensor count, onboard compute) but of *operational deployment status*. This suggests that AI maturity scores in the dataset encode achieved deployment milestones rather than latent capability potential, a finding with direct implications for how investors and policymakers should interpret published maturity ratings. The poor performance of Ridge Regression for commercial readiness prediction ($R^2 = -1.646$) underscores the non-linear, high-dimensional nature of the feature-target relationship in this domain. Random Forest's superior performance ($R^2 = 0.615$) confirms that ensemble tree methods are better suited to capture the interaction effects between deployment stage, funding, and technical specifications that jointly determine commercial readiness.

The geopolitical stratification finding ($p = 0.039$) is notable given the small sample size and suggests that the structural AI governance and investment differences be-tween Tier-1-West and Tier-1-East (China) are already measurably reflected in the realized AI maturity of deployed platforms in 2026.

The LOO accuracy of 86.5% provides the most honest performance estimate for the $N = 52$ setting, confirming that Gradient Boosting and Random Forest generalise reliably beyond the 5-fold CV partition.

VIII. LIMITATIONS

L1 — Sample size:
$N = 52$ is a small sample for six-class classification. LOO validation mitigates this but does not eliminate it. Confident generalization requires broader data collection as the industry expands.

L2 — Curated/synthetic provenance:
The dataset appears to be curated or partially synthetic (the notebook notes this explicitly). Performance metrics may be optimistic relative to a fully empirical survey.

L3 — High missingness in financial features:
Columns with > 55% missingness were excluded, removing potentially valuable financial and valuation signals from the commercial readiness regression.

L4 — SHAP computation failure:
The SHAP TreeExplainer encountered a dimensionality error with the Column Transformer output ("Per-column arrays must each be 1-dimensional"), precluding class-specific SHAP attribution. Feature importances were used as a fallback.

L5 — Temporal cross-section:
The dataset is a 2026 cross-section. Longitudinal tracking of the same platforms would enable trajectory modelling.

L6 — No inter-rater reliability:
The ordinal AI Maturity Score may reflect subjective assessments by dataset curators; inter-rater reliability was not reported.

IX. CONCLUSION AND FUTURE WORK

This study presents the first comprehensive machine learning analysis of the global humanoid robotics land-scape in 2026, addressing three complementary tasks across 52 robotic platforms and 68 features. Gradient Boosting achieves the highest AI Maturity Score classification performance (5-fold CV F1 = 0.930; test accuracy = 90.9%); Random Forest Regressor achieves the best Commercial Readiness prediction (R2 = 0.615); and KMeans clustering identifies six interpretable robot archetypes from PCA-reduced features.

Three substantive findings emerge: (1) deployment stage is the dominant predictor of AI maturity, out-weighing raw technical specifications; (2) geopolitical tier significantly stratifies AI maturity (p = 0.039), con-firming measurable geopolitical inequality in the humanoid robotics landscape; and (3) LLM integration is significantly associated with higher commercial readiness ($p = 0.025$), establishing LLM capability as a meaningful commercial maturity marker beyond its technical significance.

Future work should: (a) expand the dataset to include more robots as the industry grows,

enabling more powerful inference; (b) resolve the SHAP dimensional-ity issue to provide class-specific feature attribution; (c) collect longitudinal data to model the trajectory of in-dividual platforms through maturity stages over time;
(d) extend the geopolitical analysis to test specific hypotheses about policy interventions and their effects on deployment timelines; and (e) develop a public-facing dashboard for real-time humanoid robotics intelligence monitoring.

DATA AND CODE AVAILABILITY
intelligence 2026.ipynb) are provided as supplementary materials. All model artifacts are serialised via joblib and included in the supplementary package for reproducible inference.

AUTHOR CONTRIBUTIONS
Ranveer Singh, Sidhesh Mishra: Conceptualization, Methodology, Formal Analysis, Writing — Original Draft. *Aryan Somanna, Riduvarshini M:* Data Cu-ration, Visualization. *Kandjoni Yendouname:* Sta-tistical Analysis, Writing — Review. *Shrey Jauhari, Venkatanathan O R V, Karthik H:* Software, Writing — Review & Editing.

CONFLICT OF INTEREST
The authors declare no conflicts of interest.

REFERENCES

[1] M. J. Matarić, The Robotics Primer. Cambridge, MA, USA: MIT Press, 2007.
[2] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. Cambridge, MA, USA: MIT Press, 2005.
[3] NASA, "Technology Readiness Levels," NASA Technical Report, Washington, DC, USA, 2012.
[4] A. Brohan et al., "RT-2: Vision-language-action models transfer web knowledge to robotic control," in Proceedings of the Conference on Robot Learning (CoRL), 2023, pp. 1–25.
[5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[7] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 3146–3154, 2017.
[8] OECD, OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market. Paris, France: OECD Publishing, 2023.
[9] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[10] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 4765–4774, 2017.
[11] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[12] McKinsey Global Institute, The State of AI in 2024: Humanoid Robotics and the Physical AI Frontier. New York, NY, USA: McKinsey & Company, 2024.