POLICY-GROUNDED AI DECISION SUPPORT FOR INDUSTRIAL MODERNIZATION: ARCHITECTURE AND EVIDENCE FROM SMART-FACTORY REGISTRIES IN SHANDONG, CHINA

Shanning Fu[1], Ya Han[2], Dilshod Bekmuratov[3]

[1]Doctoral Student, National Research University "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers"
Tashkent, Uzbekistan | Email: 1915105973@qq.com
ORCID: 0009-0007-3280-7991
[2]Research Staff, Bachelor of Science
Shandong Hongjitang Pharmaceutical Group Co., Ltd.
Jinan, China
Email: 15853899107@163.com | ORCID: 0009-0007-0384-5745
PhD, Senior Lecturer
[3]National Research University "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers"
Tashkent, Uzbekistan | Email: bekmurodov_d@mail.ru | ORCID: 0000-0001-8306-1500
** Corresponding author: Dilshod Bekmuratov*

K E Y W O R D S

artificial intelligence, digital technologies, industrial modernization, industrial policy decision support, smart manufacturing, trustworthy AI

A B S T R A C T

Industrial policy increasingly operates in data-rich environments, yet AI-in-manufacturing research largely focuses on firm-level outcomes rather than accountable decision support for public agencies. This study develops a policy-grounded, governance-aware digital decision support architecture for industrial modernization and evaluates a bounded public-data proof of concept using smart-factory registries in Shandong Province, China. Eight national and provincial policy documents yielded a purposive, non-exhaustive corpus of 55 architecture-relevant operational clauses. Two authors independently coded the frozen units; for the primary policy-instrument code, observed agreement was 0.927 and Cohen's κ was 0.917. Four official releases covering 2023–2026 yielded 775 records, conservatively resolved into 607 enterprise entities. Models were selected on the 2023 records (n = 100) and evaluated on a 2024 chronological holdout (n = 100) excluded from fitting and hyperparameter selection, with zero enterprise overlap. Character-level TF–IDF with logistic regression and a linear support vector machine achieved identical performance (accuracy = 0.780; macro-F1 = 0.778), whereas XGBoost achieved 0.510 and 0.371. The architecture separates implemented registry and diagnostic functions from design-only causal simulation and monitoring under provenance, explainability, auditability, privacy, and human-approval requirements. It provides a partially validated alignment of tasks, methods, and governance; no policy effects, official relabeling of 2025–2026 records, or government deployment are claimed.

INTRODUCTION

Industrial policy has become increasingly information intensive. National and provincial plans emphasize smart-factory development, data and computing infrastructure, application scenarios, standards, monitoring, evaluation, security, and accountable AI-supported governance [1]–[8]. In manufacturing, AI has been associated with capacity utilization, productivity, resilience, industrial restructuring, and labor reallocation [9]–[12]. These outcome-oriented studies, however, do not explain how public agencies should build traceable evidence bases, distinguish predictive from causal tasks, monitor models over their lifecycle, and retain human authority over consequential decisions.

China provides a policy-rich setting for examining how digital technologies and AI can support industrial modernization. Expanding smart-manufacturing initiatives, digital infrastructure, and administrative data make it possible to study not only firm-level adoption but

(cc) BY

also how public agencies organize evidence, analytical methods, and governance controls [1]–[8]. The case is also relevant to developing economies facing similar modernization pressures under constraints in data quality, institutional capacity, and technical resources.

Research on public-sector AI addresses evidence synthesis, collective intelligence, explainability, regulatory compliance, administrative-data integration, and risk governance [13]–[16]. Policy-instrument analysis distinguishes supply-side, environment-side, and demand-side tools [17], while decision support systems coordinate data, models, and human judgment to support rather than replace decision makers [18]–[20]. Yet representative studies rarely integrate policy requirements, source-traceable registries, task-specific statistical assumptions, lifecycle monitoring, and human accountability. The design challenge is therefore to align each task with its required data, assumptions, and safeguards rather than assemble an undifferentiated portfolio of algorithms.

Accordingly, this study does not introduce a novel classification algorithm. It examines how public administrative data, policy evidence, and established AI methods can be integrated into an auditable digital decision support architecture for industrial modernization. Shandong Province serves as a bounded case from which potentially localizable design principles are derived.

The study operationalizes this purpose through a design-science artifact [21] grounded in eight policy documents and four official Shandong smart-factory lists. Empirical work is limited to source-linked registry construction, enterprise identity resolution, policy-to-module traceability, and a chronological text-classification test for review-oriented harmonization. Causal simulation and continuous monitoring remain design specifications.

Research Questions and Contributions

RQ1: What architecture requirements are supported by a purposively selected set of national and Shandong policy clauses, and how can those requirements be traced to system modules?

RQ2: Can four official smart-factory lists covering 2023–2026 be transformed into a traceable, entity-resolved public registry?

RQ3: How well does a model selected on official 2023 process/discrete labels generalize to a 2024 holdout excluded from model fitting and hyperparameter selection, and what does this performance imply for human-reviewed label harmonization?

The study makes three contributions. First, 55 purposively selected clauses from eight policy documents justify and document system requirements, with frequency claims restricted to this architecture-relevant corpus. Second, four annual releases are integrated into a registry of 775 records and 607 conservatively resolved enterprise entities, without treating administrative recurrence as causal evidence. Third, models selected on 2023 data are evaluated on 2024 data with zero enterprise overlap. This leakage-controlled proof of concept provides a bounded basis for human review when later releases omit the official type field, not a mechanism for assigning official labels automatically.

RELATED WORK AND DESIGN FOUNDATIONS

Representative public-sector approaches broaden evidence through AI-assisted research and participation [13], emphasize explainability, bias detection, transparency, and compliance [14], integrate administrative data and feedback [15], or formalize AI risk controls [16]. These contributions are complementary, but the literature reviewed here rarely combines a source-traceable industrial registry, later-year diagnostic evaluation, conditional causal simulation, lifecycle monitoring, and cross-stage human accountability within one architecture. This comparison is scoped rather than exhaustive.

Classical decision support systems research conceptualizes decision support as the interactive use of data, models, and human judgment for semi-structured decisions. Later work distinguishes data-, model-, document-, knowledge-, and communication-driven systems and emphasizes design and evaluation within organizational contexts [18]–[20]. This study extends that tradition to industrial-policy

(cc) BY

administration by treating registry provenance, model assumptions, human review, and deployment controls as constitutive design requirements rather than optional safeguards.

The architecture follows design-science principles of relevance, rigor, artifact construction, and bounded evaluation [21]. NIST AI RMF functions are operationalized as follows: GOVERN establishes roles, approvals, documentation, and audit; MAP defines policy purpose, stakeholders, provenance, and potential harms; MEASURE covers validation, error analysis, and drift indicators; and MANAGE specifies staged deployment, escalation, rollback, and retirement [22]. Concept-drift monitoring is a lifecycle requirement [23], and future imbalanced alert tasks should use precision–recall analysis where appropriate [24]. The contribution lies in aligning tasks, methods, and governance rather than in any single algorithm.

DATA AND METHODS

A. Research Design and Scope

The principal artifact is an AI-enabled digital decision support architecture for industrial modernization, with smart-factory registry harmonization as the implemented use case. Evaluation is restricted to public policy documents and official organizational records and does not claim government deployment. Implemented components include source extraction, registry construction, entity resolution, descriptive analysis, policy-to-module mapping, and chronological text classification. Causal simulation and continuous or near-real-time monitoring require dated treatment, outcome, and authorized operational data. Fig. 1 therefore presents a staged design that distinguishes implemented, evaluated, and design-only components. It is a system-level aggregation rather than a one-to-one reproduction of frozen clause-level module codes.

Official Registry Construction

Four official Shandong smart-factory releases covering 2023–2026 yielded 775 records and 607 conservatively resolved enterprise entities. The releases contained 100 records in 2023, 100 in 2024, 354 in 2025, and 221 in the first 2026 batch [25]–[28]. Retained fields included year, source serial number, city, original enterprise and factory names, official type where available, recognition level, source title, URL, file, page, and extraction notes. Checks covered serial completeness, required fields, within-year duplication, and source traceability.

The workflow follows record-linkage and entity-resolution research that separates comparison evidence from match decisions [29] and recommends auditable standardization, blocking, similarity assessment, and clerical review [30]. This implementation is conservative and rule-based rather than probabilistic. Enterprise names underwent Unicode NFKC normalization, removal of layout spaces and line breaks, and normalization of brackets and punctuation; legal suffixes were retained in the auditable display name. Punctuation-free and suffix-stripped keys generated candidates only. Exact standardized-name matches produced 165 cross-year duplicate entities. Within-city non-exact candidates used difflib.SequenceMatcher scores (×100, one decimal) with thresholds of ≥85 for full names, ≥90 for core names, or ≥95 for factory names. Shanning Fu conducted initial resolution, and Ya Han independently reviewed all seven candidates using names, city, year, and same-year co-occurrence. Shared parent-group names, province names, and generic industry terms were insufficient for merging; factory names were auxiliary evidence only. Three candidates were merged, four remained distinct, and none remained unresolved. Raw names and source identifiers were preserved. The final crosswalk contained 607 entities: 168 appeared in at least two annual lists and 439 only once.

Policy Corpus and Coding

The policy corpus comprised eight national and Shandong policy documents: the national Fifteenth Five-Year Plan [1]; Shandong’s 2018 Major-Project Plan for Replacing Old Growth Drivers with New Ones [2]; the Shandong “AI + Manufacturing” Action Plan (2026–2028) [3]; Made in China 2025 [4]; the New Generation Artificial Intelligence Development Plan [5]; Shandong’s Fifteenth Five-Year Plan [6]; Shandong’s Fourteenth Five-Year Manufacturing Powerhouse Plan [7]; and the national Fourteenth Five-Year Plan for Intelligent Manufacturing [8].

(cc) BY

Documents were purposively included when they were authoritative national- or Shandong-level plans that directly addressed intelligent manufacturing, AI-enabled manufacturing, industrial transformation, or related governance requirements.

Shanning Fu extracted the clauses and recorded the source, section, page, wording, and evidence note; Ya Han reviewed each unit against the original documents. This sequential process assessed traceability but was not an independent unitization experiment. An operational clause contained an identifiable objective, instrument, measure, responsible actor, governance requirement, or evaluation mechanism. Distinct objectives or measures were separated; inseparable instruments serving one action remained one unit; and near-duplicates were retained only when their functions differed. The dominant mechanism received the primary code, with an optional secondary code for co-occurring instruments. The purposive, non-exhaustive corpus contained 55 architecture-relevant clauses; source-level counts and legacy identifiers are retained in the archive. The 14 policy-instrument codes and module categories were fixed before final independent coding. Reported frequencies therefore characterize only this corpus, not all provisions in the eight documents.

The derivation operated at two linked levels. At the coding level, each frozen clause received one primary module category plus policy-instrument, innovation-stage, and governance codes for reliability and frequency analysis. At the design level, clause content was translated into system requirements and aggregated into six architecture functions: data ingestion and registry; policy mapping and traceability; diagnostics and classification; causal simulation; monitoring and drift detection; and AI governance and human oversight. In Fig. 1, policy mapping and diagnostics are embedded within registry harmonization, policy output represents human-reviewed use of evidence, and governance spans all stages. Frozen categories and architecture functions are therefore related but not identical; reported frequencies refer to the frozen categories. Alternative aggregations were not empirically tested.

Intercoder Procedure

Before the final classification, Shanning Fu and Ya Han discussed the research objectives, operational definitions, and ambiguous boundary cases and calibrated the codebook using examples drawn from the policy texts. AI-generated suggestions were used only as preliminary coding aids and were not treated as source evidence. Once the codebook and the 55-unit corpus had been frozen, both authors independently checked each predefined unit against the original policy text and finalized their item-level classifications without discussion until both records had been submitted. Dilshod Bekmuratov reviewed the research process and manuscript but was not counted as a coder.

Intercoder reliability was calculated from two frozen, unreconciled records before consensus using Cohen's κ [31]. For the primary policy-instrument code, agreement was 51/55, observed agreement was 0.927, expected agreement was 0.129, and κ was 0.917. The corresponding κ values were 0.976 for innovation stage, 0.956 for system module, and 1.000 for the governance flag. The latter indicates agreement on all 55 binary governance codes, not perfect unitization or universal objectivity. No pre-calibration κ is reported because no stable instrument existed before the codebook and unit boundaries were frozen. Fifty units matched across all four core fields; five contained at least one disagreement. These statistics assess classification of predefined units, not independent clause identification. After reliability calculation, the two coders resolved all five cases jointly; Dilshod Bekmuratov approved the consensus record and final dataset. The frozen independent records were preserved.

Chronological Diagnostic Prototype

The implemented task assessed whether registry names could support provisional harmonization of official process- and discrete-type labels. The 2023 set (39 discrete, 61 process) was used for training and model selection. The 2024 set (49 discrete, 51 process) supported registry construction and overlap checks but was excluded from model fitting, hyperparameter selection, and model selection and evaluated once after the pipeline was fixed. Because the 2025–

(cc) BY

2026 releases omit the same official type field, this study neither assigns nor reports official labels for them. No resolved Enterprise_ID overlapped between 2023 and 2024.

Inputs concatenated Enterprise_Name_Standardized, a [SEP] marker, and Factory_Name_Raw. Model selection used five-fold shuffled StratifiedKFold cross-validation (random_state = 2026) on 2023 data only, with macro-F1 as the criterion. Logistic regression and linear SVM used Chinese character TF–IDF with 1–3-character n-grams, min_df = 1, L2 normalization, and sublinear term frequency. Logistic-regression candidates were $C \in \{1, 10\}$ and class_weight $\in$ {None, balanced}; C = 1 with balanced weights was selected. SVM candidates were $C \in \{0.1, 1, 10\}$ with the same weight options; C = 0.1 with balanced weights was selected. Exploratory XGBoost [32] used ordinary L2-normalized character TF–IDF capped at 3,000 features, n-grams of 1–3 or 2–4, 100 trees, depth $\in \{2, 3\}$, learning rate $\in \{0.05, 0.1\}$, subsample = 0.8, colsample_bytree = 0.8, reg_lambda = 5, min_child_weight = 1, n_jobs = 1, and random_state = 0. The selected configuration used 1–3-character n-grams, depth 3, and learning rate 0.1.

The 2024 holdout was evaluated once after the pipeline was fixed. Metrics included accuracy, balanced accuracy, class-specific precision, recall, and F1, macro-F1, and confusion matrices. Uncertainty was estimated from 2,000 record-level percentile-bootstrap replicates (random_state = 2026; 2.5th and 97.5th percentiles). Paired correctness differences were assessed with two-sided exact McNemar tests implemented through the exact binomial distribution. Coefficients are presented for predictive transparency, not as causal mechanisms. The prototype does not validate BERT-based entity extraction [33] and lacks an expert-rule baseline, repeated-cross-validation stability analysis, probability calibration, and expert-validated application to the 2025–2026 records. These are limitations, not evidence that linear models outperform expert heuristics.

The final V6.3 scripted rerun confirmed 775 registry records, 607 entities, and 55 policy units. It reproduced observed agreement of 0.927, expected agreement of 0.129, and $\kappa = 0.917$ for the primary policy-instrument code, together with κ values of 0.976, 0.956, and 1.000 for innovation stage, system module, and governance flag. It also reproduced the five-fold parameter searches; final 2024 predictions; accuracy, balanced accuracy, class metrics, macro-F1, and confusion matrices; 2,000 bootstrap replicates and intervals; and exact McNemar tests. Logistic regression and linear SVM retained accuracy/macro-F1 of 0.780/0.778 with matrix [[34, 15], [7, 44]], whereas XGBoost retained 0.510/0.371 with [[2, 47], [2, 49]]. The package records vectorizer settings and seeds and generates CSV, JSON, PNG, XLSX, prediction, and execution-log outputs from frozen inputs.

Validation Boundaries

Stage 2 specifies double/debiased machine learning and generalized random forests only when treatment timing, outcomes, overlap, and identification assumptions are defensible [34], [35]. Stage 3 specifies anomaly screening and separate drift monitoring [36], [23]. Neither stage was empirically evaluated. Model outputs are therefore treated as reviewable evidence and cannot directly determine eligibility, subsidies, or enforcement actions.

ARCHITECTURE AND POLICY-GROUNDED RESULTS



(cc) BY dtai.uz

Fig. 1. Policy-grounded task–method–governance architecture. Stage 1 implements source-traceable registry harmonization and a bounded diagnostic test; causal simulation and monitoring are design-only modules; all outputs remain subject to cross-stage governance and human approval.

Registry Structure

Table I summarizes the registry. The four releases contained 775 raw records, conservatively resolved into 607 entities; 168 appeared in at least two annual lists and 439 only once. All 16 prefecture-level cities were represented. Weifang, Linyi, Jinan, Yantai, and Jining accounted for 310 records (40.0%). The largest pairwise overlaps were 2024–2025 (74 entities) and 2023–2025 (66), followed by 2023–2026 (15) and 2024–2026 (13). No confirmed overlap was observed for 2023–2024 or 2025–2026 under the conservative rules.

Annual counts represent separate administrative releases rather than a directly comparable industrial growth series. Designation terminology and release structures changed after 2024, and the 2026 data cover only the first batch. Repeated appearance may reflect continued recognition, reclassification, upgrading, or a distinct factory project; it should not be interpreted as evidence of a policy effect.

Table I.

Registry construction summary.

Registry indicator	Value	Interpretation boundary
Raw annual-list records	775	Not 775 unique firms
Resolved enterprise entities	607	Not a census of provincial firms
Entities in ≥2 years / 1 year	168 / 439	No causal interpretation
Officially typed records	200	Available only for 2023–2024

Policy-Instrument and Module Results

Within the selected 55-unit corpus, environment-side instruments were most frequent (26 clauses, 47.3%), followed by supply-side (21, 38.2%) and demand-side instruments (8, 14.5%). Data and computing infrastructure was the leading primary code (12, 21.8%); goal planning accounted for eight clauses (14.5%), and pilot/demonstration/scenario and coordination/monitoring/evaluation for seven each (12.7%). These figures characterize only the selected evidence set, not all provisions in the source documents.

Governance-and-lifecycle clauses formed the largest innovation-stage group (19, 34.5%), followed by industrialization (13, 23.6%) and cross-stage clauses (11, 20.0%). In the frozen codebook, policy output and deployment was the leading module category (14, 25.5%), followed by data ingestion and registry (11, 20.0%) and AI governance and human oversight (10, 18.2%). These clause-level frequencies are not counts of the six architecture functions in Fig. 1. Twenty-four clauses (43.6%) contained an explicit governance element. The national intelligent-manufacturing plan, for example, links smart-factory demonstration to data integration, standards, security, public-service platforms, and maturity evaluation [8]. The mapping establishes traceability, not policy effectiveness. Table II summarizes each component's evidence base, decision-support role, and evaluation status.

Table II.

Task–method–governance traceability and evaluation status.

Evidence or input	Decision-support role	Status
Official lists + entity crosswalk	Versioned registry and identity audit	Implemented

(cc) BY

Evidence or input	Decision-support role	Status
Official 2023–2024 type fields	Provisional label harmonization and review prioritization	Prototype evaluated on the 2024 holdout; no labels assigned to 2025–2026 records
Purposive policy-clause corpus	Traceability from requirements to modules	Coded; purposive and non-exhaustive
Dated treatments + outcomes	Conditional causal simulation	Design only
Rolling operational indicators	Anomaly and concept-drift monitoring	Design only
NIST-aligned governance controls	Human approval, auditability, privacy, and security	Required across all stages

DIAGNOSTIC RESULTS AND ANALYSIS

Table III.

Performance on the 2024 chronological holdout excluded from model fitting and hyperparameter selection.

Model	CV Macro-F1	Holdout Accuracy	Accuracy 95% CI	Macro-F1	Macro-F1 95% CI
Majority baseline	—	0.510	0.410–0.610	0.338	0.291–0.379
TF–IDF + Logistic Regression	0.719	0.780	0.700–0.860	0.778	0.691–0.859
TF–IDF + Linear SVM	0.719	0.780	0.700–0.860	0.778	0.691–0.859
TF–IDF + XGBoost	0.498	0.510	0.410–0.610	0.371	0.306–0.446

Chronological-Holdout Performance

As shown in Table III, logistic regression and linear SVM produced identical predictions, each achieving accuracy = 0.780, balanced accuracy = 0.778, and macro-F1 = 0.778. The accuracy 95% confidence interval was [0.700, 0.860], a width of 16 percentage points, indicating substantial uncertainty. Discrete-class precision, recall, and F1 were 0.829, 0.694, and 0.756; process-class values were 0.746, 0.863, and 0.800. Logistic regression had a lower paired error rate than the majority baseline (exact McNemar $p < 0.001$) and differed from XGBoost ($p < 0.001$), while the linear models did not differ ($p = 1.000$). XGBoost performed close to the majority baseline. With only 100 development records and sparse high-dimensional character features, split-based trees may have struggled to aggregate distributed weak signals, whereas linear models could weight thousands of sparse features. This task-specific interpretation is not a general argument against boosting. The results establish improvement over the majority baseline, not superiority over expert rules. Fig. 2 shows illustrative logistic-regression weights, and Fig. 3 the confusion matrix.

(cc) BY

ISSN: 3030-3346

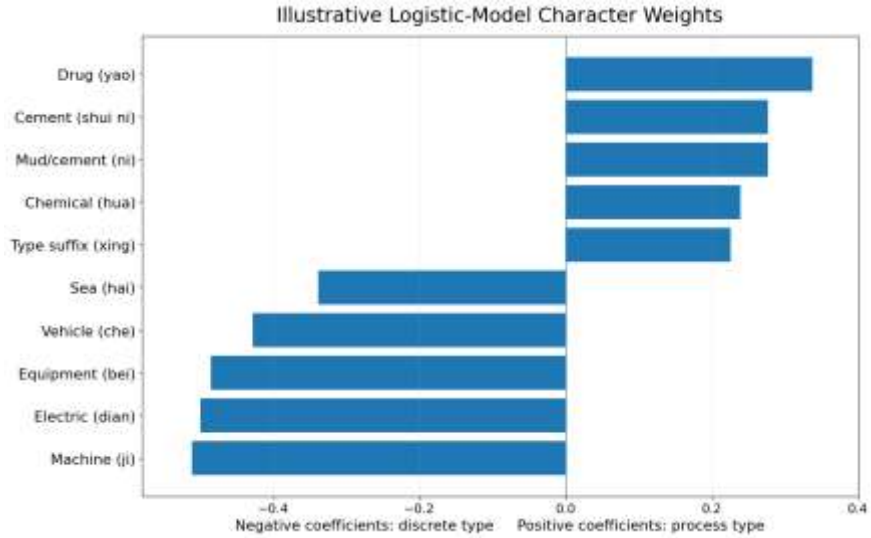


Fig. 2. Illustrative logistic-regression character weights. Positive coefficients are associated with process type and negative coefficients with discrete type; the weights are predictive associations, not causal mechanisms.

Error Structure

The reference logistic-regression model correctly classified 34 of 49 discrete-type factories and 44 of 51 process-type factories. Recall was 0.694 for the discrete class and 0.863 for the process class. Of the 22 errors, 15 were discrete factories classified as process and seven were process factories classified as discrete.

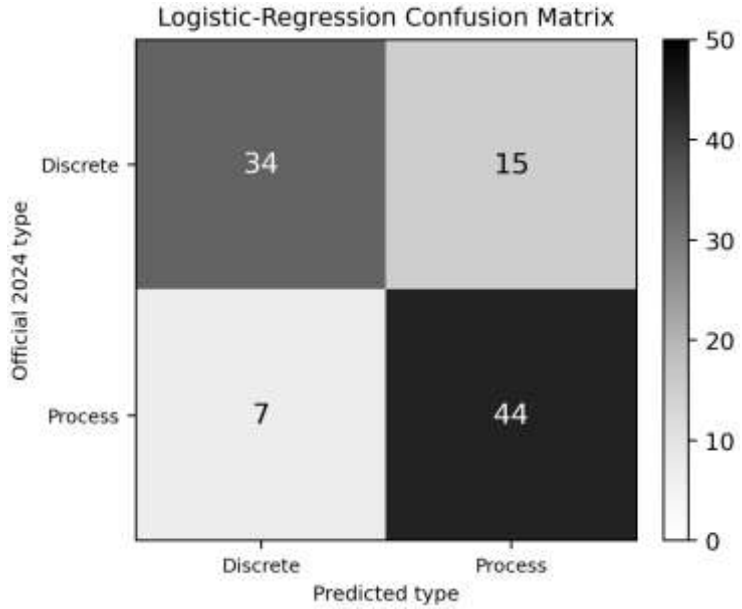


Fig. 3. *Confusion matrix for TF–IDF with logistic regression.*

Illustrative logistic-regression coefficients associated with the process class included 药 (0.338), **水泥泥** (0.277), and **化** (0.239); discrete-class features included **机**(−0.512), 电(−0.500), 备 (−0.486), 车(−0.428), and **海**(−0.339). Generic or source-specific characters such as **海** show that coefficients may capture naming conventions rather than stable industrial mechanisms. Errors clustered in names with mixed production semantics: batteries combine chemical processing with module assembly; pharmaceutical packaging combines glass or material processing with discrete packaging; and valves, tire molds, steel cord, and power equipment contain both equipment- and process-related cues. Generic "smart factory" names convey little process information. Registry text can therefore support provisional harmonization, enrichment, and

(cc) BY

review prioritization but cannot replace audited process descriptions or expert judgment.

Answers to the Research Questions

RQ1: Within the selected corpus, environment-side clauses were the most frequent policy-instrument category, while policy output and deployment was the most frequent frozen module category. These clause-level categories were translated into system requirements and then aggregated into the staged architecture. The result supports traceability; it does not characterize the complete distribution of policies or establish instrument effectiveness.

RQ2: The four official smart-factory releases can be transformed into a traceable, entity-resolved registry suitable for descriptive analysis and prototype research. The 607 resolved entities should not be interpreted as a census or as a policy-treatment sample.

RQ3: The linear models achieved significantly higher paired accuracy than the majority baseline on the 2024 holdout and may support limited human-review triage. However, the 22% error rate, asymmetric recall, absence of an expert-rule baseline, and lack of expert validation on the 2025–2026 records preclude operational or official classification.

Empirical Scope

The empirical findings support the registry-ingestion, entity-resolution, policy-mapping, and review-oriented diagnostic components of Stage 1. They do not validate BERT-based extraction, causal treatment-effect estimation, budget optimization, live anomaly detection, concept-drift response, or official type assignment to the 2025–2026 releases.

DISCUSSION

Governance-Aware Interpretation

The findings support a governance-aware rather than model-centric interpretation of AI-enabled industrial-policy decision support. In the selected corpus, environment-side instruments accounted for 47.3% of clauses, and prominent requirements concerned data infrastructure, planning, monitoring, deployment, security, and accountability. The national intelligent-manufacturing plan, for example, connects smart-factory demonstration with data integration, standards, security, public-service platforms, and maturity evaluation [8], consistent with the reviewed public-sector AI and decision-support literature [13]–[16], [22]. Frozen module categories and the staged architecture operate at different analytical levels. The study's incremental contribution is to connect these requirements to a traceable registry and a diagnostic component tested on a later-year holdout; it does not demonstrate deployment or improved government efficiency.

Pre-consensus agreement was high but not perfect: κ was 0.917 for the primary policy-instrument code, 0.976 for innovation stage, 0.956 for system module, and 1.000 for the governance flag. Five units contained at least one disagreement, highlighting judgment-dependent boundaries among goal planning, demonstration, public-service, and fiscal-financial instruments. The results support reproducibility within the frozen-unit protocol and present author team, not universal objectivity. Because clause extraction was sequentially reviewed rather than independently unitized, future replication should examine both unit identification and classification.

Registry Value and Causal Boundaries

Administrative publications can provide a low-cost data layer for industrial-policy analysis. The registry supports source tracing, cross-year identity checks, geographic summaries, and leakage-controlled model evaluation. It complements firm-level research on resilience, labor structure, and policy-related upgrading [37]–[39] by addressing the information and governance infrastructure required before such outcomes can be evaluated responsibly.

Annual list counts and recurrence are administrative characteristics. They cannot be interpreted as evidence of growth, upgrading, or policy-treatment effects. A decision-support registry is an input to subsequent evaluation, not a substitute for a causal design. Likewise, model coefficients capture predictive associations in enterprise and factory names and should not be

(cc) BY

interpreted as industrial mechanisms.

Practical Deployment Path

Implementation should be staged. First, agencies should establish a governed registry with versioned sources, standardized identifiers, documented matching decisions, access controls, and audit logs. Second, classifiers should operate in shadow mode. Because the 2025–2026 releases omit the official process/discrete field, future pilots could generate provisional labels and confidence information only for human review; this study reports none as official. Reviewers should document disagreements, overrides, abstentions, and sector-specific errors, and model outputs should not affect eligibility, subsidies, or enforcement.

Third, authorized structured indicators should be added, including industry codes, process descriptions, project stage, production and energy measures, and dated policy-support records. Before an operational pilot, future work should add an expert-rule comparator, repeated cross-validation, probability calibration, abstention thresholds, and expert validation. Causal modules should be activated only when treatment timing, outcomes, overlap, and identification assumptions are defensible. Monitoring likewise requires rolling data and explicit rules for drift detection, escalation, rollback, and retirement [22], [23].

The Chinese case is relevant not because its policy instruments or classifications should be exported, but because it demonstrates a reproducible process for converting policy documents and fragmented public records into a governed decision support resource. Transferable elements may include source traceability, staged data infrastructure, task–method separation, human review, and lifecycle governance. Policy instruments, classifications, legal requirements, and deployment thresholds must be redesigned for local institutions.

CONCLUSION AND LIMITATIONS

This study develops a policy-grounded digital decision support architecture for industrial modernization and a bounded public-data proof of concept for smart-factory registry harmonization. Eight policy documents yielded 55 selected architecture-relevant clauses, and four official releases yielded 775 records and 607 resolved entities. On the 2024 chronological holdout excluded from model fitting and hyperparameter selection, character-level TF–IDF with logistic regression and linear SVM achieved accuracy = 0.780 and macro-F1 = 0.778. These findings support traceable registry construction and limited human-reviewed diagnostic triage; they do not establish official labels for later releases.

The evidence is limited by a small two-year labeled sample, administrative rather than independently audited labels, name-based inputs, purposive and non-exhaustive clause selection, conservative entity matching, and a single-province setting. The 55-clause frequencies do not represent all provisions in the eight documents, and intercoder κ assesses predefined-unit classification rather than independent unitization. No domain-rule baseline is reported because none was independently developed and frozen before holdout evaluation; constructing one afterward would risk test-informed bias. The study also lacks repeated-validation stability analysis, calibration, abstention, subgroup fairness, expert validation of 2025–2026 records, and field deployment. Conservative matching may miss legal-name changes, subsidiaries, or reorganizations, while codebook calibration may have reduced observable coder variation.

BERT-based extraction, double/debiased machine learning, generalized random forests, budget optimization, Isolation Forest monitoring, and drift response remain unvalidated design specifications. Future research should add independently labeled years and regions, a transparent expert-rule comparator, repeated and nested validation, calibrated confidence and abstention, structured enterprise and exposure variables, independent replication of clause extraction and coding, transportability and subgroup-error tests, and a shadow-mode 2025–2026 pilot with documented human overrides. Causal evaluation should proceed only with dated treatments, outcomes, overlap, and defensible identification assumptions. The contribution is an auditable, potentially localizable design with partial validation, not evidence that AI has already improved policy effectiveness.

(cc) BY

AUTHOR CONTRIBUTIONS, AI-USE DISCLOSURE, AND DATA AVAILABILITY

Author contributions: Shanning Fu—conceptualization, methodology, formal analysis, visualization, writing—original draft, and project administration; Ya Han—data curation, validation, entity-resolution review, policy-clause review and independent classification, and writing—review and editing; Dilshod Bekmuratov—supervision, validation, writing—review and editing, and corresponding-author responsibilities. All authors reviewed and approved the final manuscript and accept responsibility for the analysis and conclusions.

AI-use disclosure: OpenAI ChatGPT was used for language editing, structural organization. AI-generated content was not treated as policy evidence or as a bibliographic source. The authors independently verified the policy clauses, data, analyses, references, figures, code outputs, and final wording and accept full responsibility for the manuscript.

Data and code availability: The study uses eight public policy documents and four official lists. The archive contains source files; two frozen 55-unit coding records; reliability, consensus, and final coding records; the row-level registry; Python scripts and pinned requirements; model-selection, cross-validation, class-metric, bootstrap, McNemar, prediction, feature-weight, figure, XLSX, SHA-256, and execution-log outputs. The evidentiary package is the final V6.3 archive identified by the frozen coding-file hashes and rerun report; earlier development packages are superseded. Data and code are available from the corresponding author upon reasonable request. No internal government operational data were used.

Funding: This research received no external funding. **Conflict of interest:** The authors declare no conflict of interest. **Ethics statement:** Not applicable. The study used publicly available policy documents and official organizational lists and involved no human participants or personal data.

REFERENCES

1. Xinhua News Agency, Outline of the Fifteenth Five-Year Plan for National Economic and Social Development of the People's Republic of China (2026–2030). Mar. 13, 2026. Available at: https://www.gov.cn/yaowen/liebiao/202603/content_7062633.htm. Accessed Jul. 17, 2026 (in Chinese).
2. Shandong Provincial People's Government, Notice on Issuing the Implementation Plan for the Major Project of Replacing Old Growth Drivers with New Ones in Shandong Province, Lu Zheng Fa [2018] No. 7. Feb. 13, 2018. Available at: https://zwfw.sd.gov.cn/art/2018/2/23/art_1684_1455.html. Accessed Jul. 17, 2026 (in Chinese).
3. Shandong Provincial Department of Industry and Information Technology et al., Notice on Issuing the Shandong Province 'AI + Manufacturing' Action Plan (2026–2028), Lu Gong Xin Shu Ju [2026] No. 60. May 16, 2026. Available at: https://gxt.shandong.gov.cn/art/2026/5/18/art_103863_10357012.html. Accessed Jul. 17, 2026 (in Chinese).
4. State Council of the People's Republic of China, Notice on Issuing Made in China 2025, State Council Document [2015] No. 28, May 8, 2015. Available at: https://www.ndrc.gov.cn/xxgk/zcfb/qt/201505/t20150520_967388_ext.html. Accessed Jul. 17, 2026 (in Chinese).
5. State Council of the People's Republic of China, Notice on Issuing the New Generation Artificial Intelligence Development Plan, State Council Document [2017] No. 35. Jul. 8, 2017. Available at: https://app.www.gov.cn/govdata/gov/20170720/408540/article.html. Accessed Jul. 17, 2026 (in Chinese).
6. Shandong Provincial People's Government, Notice on Issuing the Outline of the Fifteenth Five-Year Plan for National Economic and Social Development of Shandong Province, Lu Zheng Fa [2026] No. 2. Apr. 8, 2026. Available at: https://gb.shandong.gov.cn/art/2026/4/10/art_307618_10359939.html. Accessed Jul. 17, 2026 (in Chinese).
7. Shandong Provincial People's Government, Shandong Province Fourteenth Five-Year Manufacturing Powerhouse Construction

(cc) BY

Plan, Lu Zheng Zi [2021] No. 155. Sep. 7, 2021. Available at: https://www.shandong.gov.cn/art/2021/9/7/art_97902_429517.html. Accessed Jul. 17, 2026 (in Chinese).
8. Ministry of Industry and Information Technology of the People's Republic of China et al., Notice on Issuing the Fourteenth Five-Year Plan for Intelligent Manufacturing, Gong Xin Bu Lian Gui [2021] No. 207. Dec. 21, 2021. Available at: https://www.miit.gov.cn/jgsj/zbys/znzz/art/2021/art_c6aa629b6f124ec7bd44940b6cffb3ef.html. Accessed Jul. 17, 2026 (in Chinese).
9. Huang Z., Tao Y., Liu Z., Ye Y., How Does Intelligent Manufacturing Improve Enterprise Capacity Utilization? Based on the Perspective of Integration of Production and Consumption. Journal of Management World, Vol. 40, No. 5, pp. 40–57, 2024, DOI: 10.19744/j.cnki.11-1235/f.2024.0051 (in Chinese).
10. Liu Y. and Zuo Y., Intelligent Manufacturing Policy, Construction of Industrial Chain Ecosystem, Manufacturing TFP. Research on Economics and Management, Vol. 46, No. 7, pp. 74–90, 2025, DOI: 10.13502/j.cnki.issn1000-7636.2025.07.005 (in Chinese).
11. Cheng Q. and Zhu J., How Can AI Applications Enhance the Resilience of Manufacturing Enterprises? Research on Economics and Management, Vol. 46, No. 8, pp. 56–75, 2025, DOI: 10.13502/j.cnki.issn1000-7636.2025.08.004 (in Chinese).
12. Guo K., Artificial Intelligence, Structural Transformation and Labor Share. Journal of Management World, Vol. 35, No. 7, pp. 60–77, 2019, DOI: 10.19744/j.cnki.11-1235/f.2019.0092 (in Chinese).
13. Gambrell D., AI for eGovernance: Combining artificial intelligence and collective intelligence to develop evidence-based AI policy, in 2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG), pp. 317–324, 2025, DOI: 10.1109/ICEDEG65568.2025.11081634.
14. Manias G. et al., AI4Gov: Trusted AI for Transparent Public Governance Fostering Democratic Values, in 2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), pp. 548–555, 2023, DOI: 10.1109/DCOSS-IoT58021.2023.00090.
15. Li Z., Intelligent administrative decision support system construction and efficiency evaluation based on big data analysis, in 2024 IEEE 4th International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), pp. 1032–1036, 2024, DOI: 10.1109/ICIBA62489.2024.10868892.
16. Mahmud M., Tariq Z., Saeed M.T., Bashar A., Agentic AI risk governance: A policy-driven regulatory framework, in 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT), pp. 989–992, 2026, DOI: 10.1109/CSNT69054.2026.11502302.
17. Lv W., Chen J., Liu J., A Quantitative Analysis of China's Artificial Intelligence Industry Policy from the Perspective of Policy Tools. Studies in Science of Science, Vol. 37, No. 10, pp. 1765–1774, 2019, DOI: 10.16192/j.cnki.1003-2053.2019.10.006 (in Chinese).
18. Sprague R.H., Jr., A framework for the development of decision support systems. MIS Quarterly, Vol. 4, No. 4, pp. 1–26, 1980, DOI: 10.2307/248957.
19. Power D.J., Decision Support Systems: Concepts and Resources for Managers. Westport, CT, USA: Quorum Books, 2002.
20. Arnott D. and Pervan G., A critical analysis of decision support systems research revisited: The rise of design science. Journal of Information Technology, Vol. 29, No. 4, pp. 269–293, 2014, DOI: 10.1057/jit.2014.16.
21. Hevner A.R., March S.T., Park J., Ram S., Design science in information systems research. MIS Quarterly, Vol. 28, No. 1, pp. 75–105, 2004, DOI: 10.2307/25148625.
22. Tabassi E., Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. Gaithersburg, MD, USA: National Institute of Standards and Technology, Jan. 26, 2023, DOI: 10.6028/NIST.AI.100-1.
23. Gama J., Žliobaitė I., Bifet A., Pechenizkiy

(cc) BY

M., Bouchachia A., A survey on concept drift adaptation. ACM Computing Surveys, Vol. 46, No. 4, Article 44, pp. 1–37, 2014, DOI: 10.1145/2523813.

24. Saito T. and Rehmsmeier M., The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, Vol. 10, No. 3, Article e0118432, 2015, DOI: 10.1371/journal.pone.0118432.
25. Shandong Provincial Department of Industry and Information Technology, 2023 Provincial Smart Factory List. Nov. 15, 2023. Available at: https://gxt.shandong.gov.cn/art/2023/11/15/art_15201_10338358.html. Accessed Jul. 16, 2026 (in Chinese).
26. Shandong Provincial Department of Industry and Information Technology, 2024 Provincial Smart Factory List. Aug. 27, 2024. Available at: https://gxt.shandong.gov.cn/art/2024/8/27/art_15201_10345373.html. Accessed Jul. 16, 2026 (in Chinese).
27. Shandong Provincial Department of Industry and Information Technology, 2025 Shandong Advanced (Provincial-Level) Smart Factory List. Jul. 9, 2025. Available at: https://gxt.shandong.gov.cn/art/2025/7/9/art_15201_10351066.html. Accessed Jul. 16, 2026 (in Chinese).
28. Shandong Provincial Department of Industry and Information Technology, 2026 Shandong Advanced Smart Factories (First Batch) List. Jun. 29, 2026. Available at: https://gxt.shandong.gov.cn/art/2026/6/29/art_15201_10357844.html. Accessed Jul. 16, 2026 (in Chinese).
29. Fellegi I.P. and Sunter A.B., A theory for record linkage. Journal of the American Statistical Association, Vol. 64, No. 328, pp. 1183–1210, 1969, DOI: 10.1080/01621459.1969.10501049.
30. Christen P., Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Berlin, Heidelberg: Springer, 2012, DOI: 10.1007/978-3-642-31164-2.
31. Cohen J., A coefficient of agreement for nominal scales. Educational and Psychological Measurement, Vol. 20, No. 1, pp. 37–46, 1960, DOI: 10.1177/001316446002000104.
32. Chen T. and Guestrin C., XGBoost: A scalable tree boosting system, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016, DOI: 10.1145/2939672.2939785.
33. Devlin J., Chang M.-W., Lee K., Toutanova K., BERT: Pre-training of deep bidirectional transformers for language understanding, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186, 2019, DOI: 10.18653/v1/N19-1423.
34. Chernozhukov V. et al., Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, Vol. 21, No. 1, pp. C1–C68, 2018, DOI: 10.1111/ectj.12097.
35. Athey S., Tibshirani J., Wager S., Generalized random forests. The Annals of Statistics, Vol. 47, No. 2, pp. 1148–1178, 2019, DOI: 10.1214/18-AOS1709.
36. Liu F.T., Ting K.M., Zhou Z.-H., Isolation Forest, in 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422, 2008, DOI: 10.1109/ICDM.2008.17.
37. Liu X. and Han X., Artificial Intelligence and Manufacturing Resilience: Intrinsic Mechanisms and Empirical Tests. Business and Management Journal, Vol. 45, No. 11, pp. 48–67, 2023, DOI: 10.19616/j.cnki.bmj.2023.11.003 (in Chinese).
38. Pan S. and Guo K., Artificial Intelligence, Occupational Structural Transformation and Service-oriented Manufacturing. China Industrial Economics, No. 4, pp. 57–75, 2024, DOI: 10.19581/j.cnki.ciejournal.2024.04.003 (in Chinese).
39. Han Y., Huang L., Wang X., Do Industrial Policies Promote Industrial Structure Upgrading? Theory and Evidence from China's Development-oriented Local Government. Economic Research Journal, Vol. 52, No. 8, pp. 33–48, 2017 (in Chinese).

(cc) BY