

Evidence-Constrained Construction and Evolution Analysis of Job Competency Graphs

Ge Tian¹, Anhong Chen¹, Hanxiao Cao¹, Zhen Hao^{2,*}

¹Henan University of Technology, Zhengzhou, Henan, China

²iFLYTEK Co., Ltd., Hefei, Anhui, China

**Corresponding Author.*

Abstract: To address hallucinated facts, missing evidence chains, unauditible versions, and unquantified uncertainty in job competency graphs, this study develops an integrated framework using 15,709 job postings from Zhaopin, ByteDance Careers, and BOSS Zhipin. The framework combines data denoising, evidence binding, review and release, versioned storage, and trend forecasting. It uses a three-component weighted similarity measure to control textual noise, nine-dimensional evidence metadata and a draft-review-release-discard state machine to govern graph publication, and an ensemble of weighted linear trend regression and exponentially weighted moving average for skill-demand forecasting. Experiments produced a versioned graph with 15,766 entities and 29,667 relations. Compared with the baseline, duplicate-detection F1 increased by 11.4%; 78% of published relations were traceable to source evidence; and the rate of erroneous forecast delivery decreased by 62% in the three-month minimum-sample setting. The framework supports traceability, version auditing, and interpretable trend analysis for talent inventory and vocational education planning.

Keywords: Job Competency Graph; Large Language Model Hallucination; Evidence Constraints; Version Management; Uncertainty-Aware Time-Series Forecasting

1. Introduction

The US O*NET[1] and the EU ESCO [2] provide mature occupational classifications, but they are difficult to timely reflect emerging positions in the domestic market. Existing research has combined the classification system with large-scale model to generate job maps [3], while pointing out that the generated model may

produce unverifiable facts [4]; the extraction of job skills still requires domain constraints and manual verification [5]. Therefore, the source evidence and review status should become the conditions for knowledge release.

Technological changes will reorganize job tasks and skill requirements [6]. Retrieval-enhanced generation can introduce external knowledge [7], but it is difficult to audit when there is a lack of evidence location, sample size, and version records. There is a linkage between technology, tasks, and skills [8], but the existing static job maps and point-value predictions are difficult to explain this transmission. Repeated publishing, template reuse, and local rewriting will increase the frequency of skills. SimHash is suitable for near-duplicate recall [9], but it is difficult to simultaneously explain the full text, sentence order, and fragment overlap. This paper places composite duplicate checking at the front end of the map construction and integrates evidence, version, and confidence level access throughout the subsequent analysis.

2. Multi-Source Recruitment Data Preprocessing and Composite Noise Detection

The data comes from Zhaopin, ByteRecruitment and BOSS Direct Recruitment. There are issues such as field heterogeneity, duplicate postings and invalid short texts. This chapter introduces data normalization and three-layer weighted JD similarity detection.

2.1 Unified Normalization Processing of Multi-Source Heterogeneous Data

The three platforms differ in terms of fields, formats, dimensions and quality. The system achieves standardization in three layers.

1) Data collection layer

The system uses Playwright and DrissionPage in collaboration for data collection: the former handles dynamic pages and interactions, while

the latter maintains the session; the reuse of login cookies is used to ensure stable incremental collection.

2) Field Adaptation Layer

The system maps the same-sounding fields through field aliases, combines regular expressions and entity extraction to achieve standardization: salary is split and the units are unified, work experience and educational background are converted to enumerated values, job titles have redundant suffixes removed, skills are transformed into unique labels, and finally a unified Schema is output, which is compatible with five types of files such as JSON and CSV.

3) Text noise reduction and quality verification

The preprocessing process first filters out invalid short texts, removes HTML tags, garbled characters and special symbols, then performs full-text deduplication using SHA-256, outputs the standard JSONL format, and conducts graded verification according to 13 quality rules.

2.2 Three-Layer Weighted Composite JD Similarity Detection Algorithm

For the cases of full text copying, partial copying and reordering of sentence structure, the system calculates the weighted similarity based on three indicators:

$$(Sim_{JD} = 0.55 \cdot Jaccard_{3gram} + 0.25 \cdot SeqSim + 0.20 \cdot OverlapRatio) \quad (1)$$

The three sub-indicators in the formula respectively depict the degree of overlap in different dimensions. Their definitions and calculation procedures are as follows.

1) Character triplet Jaccard similarity $Jaccard_{3gram}(G_3(\cdot))$.

First, normalize the text - remove all characters other than Chinese, English, and numbers, and convert all to lowercase. Then, extract the set of consecutive 3-character N-grams, and calculate the intersection-over-union ratio of the two sets:

$$Jaccard_{3gram}(A, B) = \frac{|G_3(A) \cap G_3(B)|}{|G_3(A) \cup G_3(B)|} \quad (2)$$

This indicator can identify sentence order adjustments and minor rewrites. It is the main indicator with a weight of 0.55.

2) Full text sequence similarity ($SeqSim(A, B)$).

Using the SequenceMatcher algorithm based on the longest common subsequence, the normalized text sequence consistency ratio is calculated to quantify the similarity of the JD's writing structure. This indicator has high sensitivity to copying by simply copying the full text or making minor additions or deletions, but

it is susceptible to interference from reordering of words. Therefore, a weight of 0.25 is set as an auxiliary verification basis for the text structure dimension.

3) Repetition paragraph coverage rate ($OverlapRatio(A, B)$).

Based on the punctuation at the end of the sentence, the JD is segmented into text fragments. Invalid fragmented segments with a length less than 8 characters are filtered out. The optimal similarity between the fragments is calculated using SequenceMatcher. If the similarity is greater than 0.80, it is determined as a duplicated fragment, and the coverage rate is calculated as the ratio of the total length of the duplicated fragments to the total length of the source text.

$$OverlapRatio = \frac{\sum_{s \in S_{rep}} len(s)}{\sum_{s \in S} len(s)} \quad (3)$$

Coverage is used to identify the copied responsibilities or job requirements, with a weight of 0.20, and outputs no more than 5 segments for review.

Based on the comprehensive similarity, three levels of risks are classified: ≥ 0.75 is high plagiarism risk and is directly filtered; 0.45 to 0.75 is medium risk, marked for manual review; < 0.45 is low noise, and directly enters entity extraction. Unit testing verification: the similarity of completely duplicated text is 1.0, the duplicate coverage is greater than 0.9, and the similarity of JDs unrelated to the topic is lower than 0.45.

2.3 Large-Scale Retrieval Optimization Architecture

1. Recall stage: Conduct approximate text retrieval for the entire database based on Elasticsearch, and filter out the candidate duplicate job descriptions set;

2. Fine sorting stage: Calculate the comprehensive similarity according to formula (1), and output duplicate segments, proportion and risk level.

3. Evidence Constrains the Knowledge Production and Evolution Analysis Framework

The quality of the knowledge graph is determined by the traceability and credibility of the knowledge. This chapter proposes an evidence-constrained knowledge production paradigm: The entities and relationships in the officially entered knowledge graph must be

bound to the original evidence. They are then subject to dual verification through automated access control and manual review, and the graph evolution analysis with uncertainty awareness is carried out based on versioned storage.

3.1 Six-Layer Overall Framework Structure

The framework consists of six layers: the multi-source signal acquisition layer, the data preprocessing and noise reduction layer, the evidence-constrained knowledge generation layer, the versioned graph storage layer, the uncertainty evolution analysis layer, and the upper-level business application layer. The entire framework is divided into six layers from top to bottom, and the specific functions are shown in Figure 1:

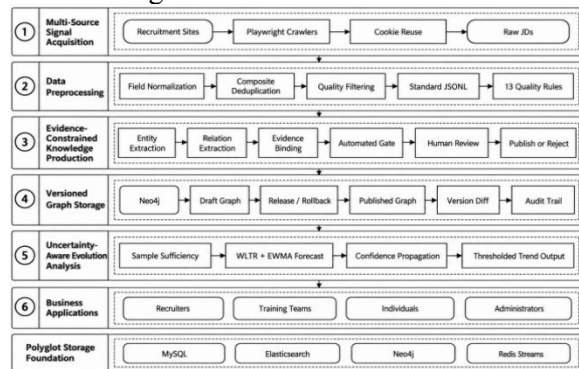


Figure 1. Evidence-Constrained Framework for Knowledge Production and Evolution Analysis

3.2 The Core Mechanism of Position-Specific Knowledge Production under Evidence Constraints

3.2.1 Eight-stage closed-loop knowledge production pipeline

The eight-stage pipeline sequentially completes multi-source JD collection, standardization noise reduction, composite duplicate checking, entity relationship extraction, evidence binding, rule access control, manual review, and version release or deletion. Each stage can intercept anomalies, as shown in Figure 2.

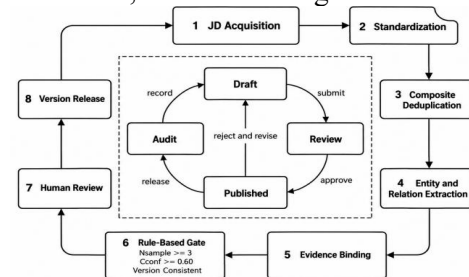


Figure 2. Eight-Stage Closed-Loop Knowledge Production and State Machine

Sample size sampling: When the total number of samples supporting a certain piece of knowledge is less than 3 ($N_{\text{sample}} < 3$), This relationship was determined to be insufficiently supported by the data, and was flagged for inclusion in the queue for manual review.

Confidence threshold interception: If the relationship score is below 0.60, the "review_required" flag is set and manual review is initiated;

Release access control: If the number of position capability relationships is less than 3, the release is denied to prevent incomplete graphs from entering the official database;

Rejection and backflow: The rejected content along with the reason is returned to the draft library. After modification, it can be resubmitted;

Release traceability: The system generates change diffs and audit records, and saves the additions, deletions, and modifications of the position-skill relationships.

3.2.2 Evidence metadata model

Each job position, skill, and job-skill association knowledge is uniformly bound with evidence records. The formal definition is as follows:

$$\text{EvidenceRecord} = \{S_{\text{doc}}, N_{\text{sample}}, C_{\text{conf}}, V_{\text{model}}, V_{\text{data}}, V_{\text{graph}}, S_{\text{audit}}\} \quad (4)$$

The engineering constraints for each field are as follows:

(S_{doc}) (Source Document Collection): The set of original JD document numbers that support this piece of knowledge. After duplicates are removed and combined, they are stored and support the ability to trace back to the original recruitment text with a single click.

(N_{sample}) (Sample total quantity): Sum the sample quantities during the aggregation process and set the threshold ($N_{\text{sample}} < 3$) as the recommended review boundary line.

(C_{conf}) (Comprehensive Confidence): Multi-source data consistency comprehensive confidence. The default benchmark is 0.8. It is dynamically adjusted based on data quality scores, cross-source consistency verification, and manual review with addition or subtraction of points. The upper limit is 0.95. When there is no confidence source, it is marked as "uncalibrated".

($V_{\text{model}}, V_{\text{data}}, V_{\text{graph}}$) (Version3)Record the model version, dataset version, and graph version separately.

1). Combined prediction formula.

The system combines long-term trends and short-term fluctuations through a weighted linear trend (WLTR) and an exponentially weighted moving average (EWMA):

$$TrendScore = \alpha \cdot WLTR + (1 - \alpha) \cdot EWMA \quad (5)$$

In the formula, WLTR employs weighted least squares to fit the long-term demand trend. The weights are determined jointly by the logarithm of the sample size of each period and the data confidence level, so that periods with larger samples and higher confidence levels contribute more to the trend fitting; EWMA smooths short-term data noise in a recursive manner. The combination coefficient as α is adaptively determined by the historical fitting errors (reciprocal of RMSE) of the two models:

$$\alpha = \frac{1/RMSE_{linear}}{1/RMSE_{linear} + 1/RMSE_{ewma}} \quad (6)$$

That is, the model with the smaller fitting error will have a higher weight in the combination, achieving automatic selection among models. At the same time, the trend direction is determined based on the linear slope and the tolerance threshold, and the uncertainty interval for each predicted point ($1.96, \sigma\sqrt{t}$) is relaxed as the prediction step length increases, explicitly quantifying the uncertainty of the prediction.

2). Comprehensive calculation of confidence level.

$$Confidence = 0.30 \cdot C_{cov} + 0.25 \cdot C_{sample} + 0.25 \cdot C_{fit} + 0.10 \cdot C_{source} + 0.10 \cdot C_{review} \quad (7)$$

The five indicators are weighted and normalized to the range of 0 to 1: C_{cov} is the timing coverage, C_{sample} is the sample size, C_{fit} is the goodness of fit.

$$\frac{I}{I + RMSE/\bar{y}} \quad (8)$$

In the formula, C_{source} represents the number of independent data sources, and C_{review} represents the review status.

3). When the confidence level is lower than 0.60, mark "review_required"; when there is only a single observation or a short-term sample, the confidence level is capped at 0.39 and 0.59 respectively.

3.3.3 Technical - Position - Skill Three-level Impact Transmission Chain

Based on Neo4j graph query, a three-level hierarchical association of "technology trend - position - skill" is established: The technology trend entity is linked to the corresponding position node through the IMPACTS influence

edge, and the position is then bound to the detailed skill nodes through the REQUIRES requirement edge. Each transmission chain records the complete path triple (trend ID, position ID, skill ID) and the relationship type, and binds the supporting document source, allowing traceability to industry reports and original job descriptions. At the same time, on the chain, the influence weight, requirement weight, and respective confidence levels are recorded, supporting filtering by the minimum confidence threshold, and only retaining the reliable transmission relationships, providing a quantifiable evidence chain for "how technology trends are transmitted to specific skills requirements through specific positions".

4. Algorithm Refinement

This chapter elaborates on the algorithms of deduplication, extraction, prediction and confidence propagation. All inputs retain their source and version information. Results with low confidence or incomplete evidence are only placed in the draft section; predictions also return the number of samples, intervals, and explanation paths.

4.1 Algorithm for Duplicate Removal and Similarity Calculation of JD Texts

The system normalizes the job title, responsibilities, requirements, salary and location, and then blocks the recall by job family, region and release time window to avoid full-scale comparison and misjudgment by the general template. The process is shown in Figure 6.

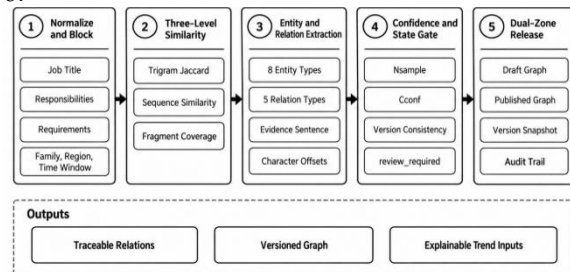


Figure 6. Algorithm Flow for JD De-Duplication and Evidence Constraint Extraction

For each candidate JD pair, the Jaccard similarity of character triplets, the normalized sequence similarity, and the repeat fragment coverage rate were calculated: these three metrics successively reflect the local lexical form, the overall structure, and the fragment reuse, and the comprehensive score is shown in

Equation (9).

$$Sim_{JD} = 0.55 \times Jaccard_{3gram} + 0.25 \times SeqSim + 0.20 \times OverlapRatio \quad (9)$$

Equation (9) uses the weights from Chapter 2. When Sim_{JD} is greater than or equal to 0.75, it is marked as a high-risk duplicate; when it is between 0.45 and 0.75, it is classified as a suspected duplicate; and when it is less than 0.45, it is considered an independent sample. The threshold only determines the review path; the system saves the components, threshold version, document identifier, and timestamp.

The system calculates the scores for "Job Title and Responsibilities" and "Job Requirements" separately. Only when both areas reach the threshold will they be automatically merged; if only one area meets the threshold, it will be retained as the same source version. The original JD is not deleted; instead, it is retained using the `DUPLICATE_OF` or `DERIVED_FROM` relationship.

4.2 Algorithms for Extracting Multiple Types of Entities and Relationships

Entity extraction adopts the method of "rule recall - model recognition - constraint verification" to identify eight types of entities such as positions, skills, enterprises, industries, educational qualifications, certificates, knowledge points and tools[10], and then verifies them based on type, dictionary version, context and character position.

The relationship extraction identifies five types of relationships: `REQUIRES`, `CO_OCCURS`, `BELONGS_TO`, `USED_IN` and `EVOLVES_TO`. Each relationship is bound with evidence sentences, character positions, model versions and states; those with incomplete evidence are not allowed to enter the formal graph or downstream computations.

Let the rule recall score, model confidence score and knowledge base matching score be Re , Me , Ke respectively, and the type conflict penalty be $Ce_{conflict}$. The comprehensive score is:

$$C_e = w_r Re + w_m Me + w_k Ke - w_c C_e^{conflict} \quad (10)$$

The weights are calibrated based on the F1 score of the validation set and the review cost. If the score is lower than 0.70, it is marked as "review_required"; when it reaches the threshold and the evidence is complete, it is placed in the pending release queue. For a formal release, the requirements of accessible source, valid evidence location, no type conflicts, and consistent version must be met, as shown in

Figure 7.

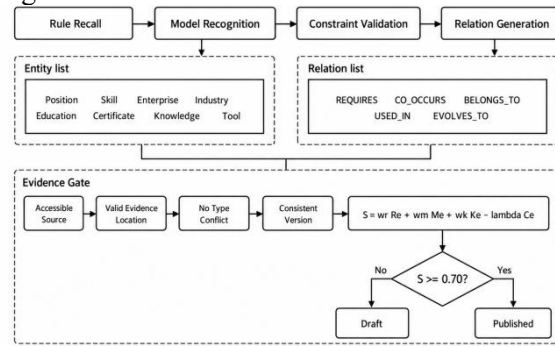


Figure 7. Extraction of Multiple Types of Entity Relationships and Evidence Access Control

The relationship is based on the logical key of "head entity - type - tail entity - valid time". Evidence with the same key is aggregated. Cross-platform conflicts retain the branches and reduce the confidence level. Entities retain the original mention and aliases to facilitate the retrieval of JD.

4.3 Time Prediction and Confidence Propagation Algorithm

4.3.1 Combination of WLTR weighted local trend and EWMA

Skill trend prediction is based on the skill occurrence rate in the monthly valid job descriptions as the basic time series. The system first groups by industry, job family, and region, and then divides the number of valid job descriptions where the skill occurred by the total number of valid job descriptions in that group for the month, to reduce the direct impact of changes in collection scale on the frequency; subsequently, it estimates the recent changes within a local window and combines them with the exponentially weighted moving average[11]. Months with insufficient sample size or abnormal collection are not padded with zeros, but marked as missing and included in the confidence calculation to prevent data interruptions from being interpreted as a sudden drop in skill demand.

$$TrendScore_t = \alpha \cdot WLTR_t + (1-\alpha) \cdot EWMA_t \quad (11)$$

$$WLTR_t = \sum_{i=1}^k \lambda_i (x_{t-i+1} - x_{t-i}), \sum_{i=1}^k \lambda_i = 1 \quad (12)$$

The combination coefficients of Equation (11) are determined by the rolling validation error, and the time weights of Equation (12) are normalized by the attenuation coefficient. The output includes point estimates, intervals, effective sample size and version; when there is

insufficient observation or the intervals are too wide, only the message "Insufficient evidence" is given.

4.3.2 Multi-dimensional confidence normalization and graph propagation

The prediction confidence is composed of five dimensions: data sufficiency D , evidence coverage E , time stability T , prediction error P , and source reliability S [12]. Each dimension is first mapped to the interval $[0,1]$ according to the pre-defined range, and then weighted and summed. The initial weights are derived from the business priority in the system implementation and calibrated on the manual review samples; the calibration only updates the weight version, without overwriting or covering the original evidence.

$$Confidence = 0.30D + 0.25E + 0.20T + 0.15P + 0.10S \quad (13)$$

When the confidence level is lower than 0.60, the "review_required" status is entered. The upper limit is 0.39 when there is only one valid cycle; for a 12-month window with valid observations, if the period is less than 52 weeks, the upper limit is 0.59. The trigger dimension and the reason for the absence are simultaneously displayed on the business end. On the "Technology Trends - Position - Skill" path, with the path length denoted as L , the edge confidence as c_i , the node evidence completeness coefficient as q_v , and the attenuation coefficient as γ , the derived path confidence is defined as:

$$C_{path} = q_v \gamma^{L-1} \prod_{i=1}^L c_i \quad (14)$$

The derived scores are written in batches and do not overwrite the original edges. The presentation layer returns the main path with complete evidence, along with the original JD, evidence sentences, relationship versions and status, to clearly define the transmission process and evidence boundaries. See Figure 8.

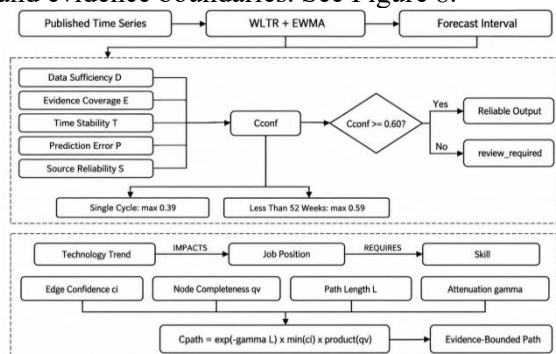


Figure 8. Time Prediction Confidence Propagation and Access Control Decision

4.4 Algorithm Complexity and Reproducibility Control

When the number of candidate pairs after blocking is M and the average length is L , the calculation complexity is approximately $O(ML)$; the full comparison of N JDs is $O(N^2)$. Blocking ensures scalability, and the graph writing is merged stably with the ID and logical key in a consistent manner.

Table 1. Key Algorithm Modules and Evidence Constraints Output

Module	Core Input	Evidence constraints on output
Remove duplicates (JD)	Standardization of JD, time window, candidate comparison	Similarity component, threshold version, repeated fragment position
Entity relationship extraction	Dictionary, model, evidence sentence	Entity type, character position, source document, model version
Confidence -level access control	Sample size, source consistency, prediction error	Comprehensive confidence level, review status, triggering reason
Release of the chart	Stable ID, version snapshot, review result	Formal relationship, change log, traceable path

Rules, dictionaries, models, thresholds, weights and snapshots are all numbered. Task records include input hash, random seed and batch. The results can be traced back along the path of "original document - cleaning - extraction - review - graph - prediction". The inputs, outputs, and evidence constraints of the key algorithm modules are summarized in Table 1.

5. Experimental Verification and Case Analysis

5.1 Experimental Data, Environment and Evaluation Methods

5.1.1 Data Scope and Time Division

The data includes 8,653 job descriptions from Zhaopin, 5,648 from Boss Zhipin, and 1,408 from Byte Recruitment, totaling 15,709 original job descriptions. After preprocessing, 10,485 valid records were obtained, and then samples with missing time, irrecoverable evidence, or insufficient continuity were excluded to form 9,928 experimental job descriptions. The graph contains 15,766 entities and 29,667 relationships. The trend experiment selected 31 consecutive sequences. The data was divided into training,

validation, and test sets according to time, and the threshold was determined only for the former two; near-duplicate documents were isolated by source group.

5.1.2 Experimental Environment and Replication Settings

The backend uses Python, FastAPI, MySQL 8.0, Elasticsearch, Neo4j and Redis Streams. The models, dictionaries, rules and snapshots are all in fixed versions, and random seeds, thresholds and input hashes are also saved.

The duplicate removal experiment replaced the similarity method, the evidence experiment disabled the different constraints, and the trend experiment compared the WLTR, EWMA, and combined models. The manual review part only counted the completed samples with accessible evidence.

5.1.3 Evaluation Indicators

Entity relationship extraction and repetition identification employ accuracy rate P , recall rate R and $F1$, and report false positive rate, false negative rate and the proportion of review_required.

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2PR}{P + R} \quad (16)$$

The trend prediction uses MAE and RMSE; the quality of the graph is evaluated by evidence coverage rate, traceability rate, manual review pass rate, and the review duration for every thousand JDs, taking into account both effectiveness and cost.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

5.2 Knowledge Extraction and Evidence Constraint Experiment

The ablation study shows that when using the model alone, there are more type confusions and no evidence relationships; after adding the dictionary and constraints, low-frequency entities become more stable, and low-confidence results are transferred for re-examination. The evidence constraints aim to control the publishable knowledge.

The comparison of repeated recognition between SimHash, single Jaccard and the method in this paper shows that the composite score simultaneously utilizes local semantics, overall order, and fragment coverage information, and the F1 score is 11.4% higher than the baseline, and it can locate similar fragments for manual re-examination.

Closing the evidence binding, access control or auditing will weaken the source and change recovery capabilities. The complete system isolates evidence-free or low-confidence relationships in the draft area, and the proportion of formal relationships that can be traced is 78%.

5.3 Evolutionary Prediction and Confidence Threshold Experiment

The results indicate that input deduplication, evidence release, and prediction access control form a continuous control chain, which respectively improve data quality, knowledge traceability, and output reliability.

Table 2 .Summary of Key Experimental Results

Verification object	Comparison/evaluation criteria	Result	Explanatory conclusion
JD Repeated Recognition	Repetition recognition F1 compared with multiple baseline methods	Increase by 11.4%	Multi-feature combination improves detection of local copying and word-order changes.
Evidence Constraint Graph	Proportion of release relationships traceable to source evidence	Reach 78%	Evidence binding and release access control strengthen relationship traceability.
Trend Prediction Access Control	Erroneous prediction push rate under the scenario with a minimum sample span of three months	Decrease d by 62%	Low-confidence predictions are submitted for review, and conclusions with insufficient evidence are prevented from entering business workflows.

5.4 Typical Position Cases and Business Value

The AI engineer cases are sequentially completed with tasks such as normalization,

duplication check, extraction, evidence binding, access control, review, and release. Users can view skills, source job descriptions, evidence sentences, confidence levels, and version differences, enabling the job profile to be

updated and auditable. When the occurrence rate of skills increases, the system displays the number of job descriptions, time range, prediction interval, and transmission path; if the sample size is insufficient or the source is single, it marks for review to avoid over-explanation.

The system can assist in recruitment, training and personal skills analysis, but it does not replace recruitment, assessment or career choice; in the application, an entry for manual review and appeal should be retained, as shown in Figure 9.

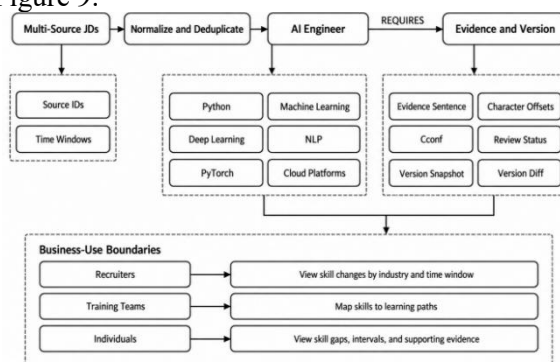


Figure 9. Evidence Chain for AI Engineer Position Case

5.5 Validity and Reproducibility Boundaries of Results

Time segmentation, document group isolation, and version locking reduce the risk of leakage, but the review criteria still affect the threshold. The samples mainly come from public platforms and technical positions and cannot be directly extrapolated; the occurrence rate of JDs is not equivalent to employment, salary, or productivity.

11.4%, 78%, and 62% correspond to specific baselines, versions, and small sample scenarios. Replication requires the use of the same snapshots, thresholds, and filtering conditions; recalibration should be performed after changing the data or model.

6. Limitations and Future Prospects

The data in this article is sourced from an open recruitment platform and may have deviations in terms of enterprise, job position, and regional distribution. Some records lack complete information on the time and enterprise attributes. The current entity relationship system also has limited coverage of soft skills and cross-job transfer. The confidence weight needs to be recalibrated according to the industry. WLTR and EWMA have not yet incorporated

exogenous variables such as macroeconomics and policies. Therefore, the prediction is only used for trend judgment and cannot explain causality. Cross-language processing, real-time updates, cross-platform disambiguation, and external testing across the entire industry still need to be verified. In the application, personal information should be protected to avoid the influence of sensitive attributes on job matching or modeling of historical biases [13]. Subsequently, a cross-platform and cross-industry validation set will be constructed, external events and causal analysis will be introduced, and mechanisms for skill evolution, job transfer, conflict evidence integration, and versioned updates will be improved.

7. Conclusion

This paper connects JD noise reduction, evidence binding, hierarchical review, version storage, and uncertainty prediction into a closed loop. The graph constructed based on 15,709 original JDs contains 15,766 entities and 29,667 relationships; the F1 score of repeated recognition is 11.4% higher than the baseline, the proportion of traceable relationships is 78%, and the error prediction push rate in the minimum sample scenario within three months has decreased by 62%. Within the given data and evaluation standards, this framework enhances the verifiability, auditability, and trend interpretation ability of the job capability graph, and can provide auxiliary evidence for talent management and vocational education.

References

- [1] Zhang Wen, Lai Kee-Hung, Gong Qiguo. The Future of the Labor Force: Higher Cognition and More Skills. *Humanities and Social Sciences Communications*, 2024, 11(1):Article479.DOI:10.1057/s41599-024-02962-1.
- [2] Vrolijk J, Poslavsky V, Bijl T, et al. Ontology Learning for ESCO: Leveraging LLMs to Navigate Labor Dynamics. *Joint Proceedings of the 2nd Workshop on Knowledge Base Construction from Pre-Trained Language Models and the 3rd Challenge on Language Models for Knowledge Base Construction (KBC-LM 2024 and LM-KBC 2024)*, Baltimore, USA. *CEUR Workshop Proceedings*, 2024, 3853: 1-16.
- [3] Karakeben K, Sorotos H E, Milchevskaya A,

- et al. Rapid Graph Generation from Job Descriptions: Combining Taxonomies and LLMs//The Semantic Web: ESWC 2024 Satellite Events. Lecture Notes in Computer Science, Vol. 15344. Cham: Springer, 2025: 143-146.DOI:10.1007/978-3-031-78952-6_15.
- [4] Wang Y, Wang M, Manzoor M A, et al. Factuality of Large Language Models: A Survey//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024). Miami, USA: Association for Computational Linguistics, 2024: 19519-19529.DOI:10.18653/v1/2024.emnlp-main.1088.
- [5] Nguyen K, Zhang M, Montariol S, et al. Rethinking Skill Extraction in the Job Market Domain Using Large Language Models//Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024). St. Julian's, Malta: Association for Computational Linguistics, 2024: 27-42.DOI:10.18653/v1/2024.nlp4hr-1.3.
- [6] Autor David, Levy Frank, Murnane Richard. The Skill Content of Recent Technological Change: An Empirical Exploration. The Quarterly Journal of Economics, 2003, 118(4):1279-1333.DOI:10.1162/003355303322552801.
- [7] Cuconasu F, Trappolini G, Siciliano F, et al. The Power of Noise: Redefining Retrieval for RAG Systems//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2024). Washington, D.C., USA: Association for Computing Machinery, 2024: 719-729.DOI:10.1145/3626772.3657834.
- [8] Acemoglu D, Autor D. Skills, Tasks and Technologies: Implications for Employment and Earnings//Ashenfelter O, Card D, eds. Handbook of Labor Economics. Vol. 4B. Amsterdam: Elsevier, 2011: 1043-1171.DOI:10.1016/S0169-7218(11)02410-5.
- [9] Sood S, Loguinov D. Probabilistic Near-Duplicate Detection Using Simhash//Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11). Glasgow, UK: Association for Computing Machinery, 2011:1117-1126.DOI:10.1145/2063576.2063737.
- [10] Devlin J, Chang M W, Lee K, et al. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding//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). Minneapolis, USA: Association for Computational Linguistics, 2019: 4171-4186.DOI:10.18653/v1/N19-1423.
- [11] Hyndman R J, Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne: OTexts, 2021.
- [12] Dong X L, Gabrilovich E, Heitz G, et al. Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14). New York, USA: Association for Computing Machinery, 2014: 601-610.DOI:10.1145/2623330.2623623.
- [13] De-Arteaga M, Romanov A, Wallach H, et al. Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting//Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). Atlanta, USA: Association for Computing Machinery, 2019:120-128.DOI:10.1145/3287560.3287572.