

RETRIEVAL-AUGMENTED GENERATION AND KNOWLEDGE -
AUGMENTED GENERATION FOR HUMAN-SUPERVISED
CURRICULUM DESIGN

Kozhanov M. , Mosavi A. , Kozhukhova M. 

Abstract Universities design curricula from programme goals, learning outcomes, courses, credits, and prerequisites. Regulatory requirements are also checked during curriculum design. We present Curriculum-KAG, a human-supervised decision-support prototype for curriculum design. It supports bachelor's, master's, and doctoral programmes. The proposed method combines Retrieval-Augmented Generation (RAG) and Knowledge-Augmented Generation (KAG). It uses Kazakhstan's historical educational-programme corpus and multilingual retrieval. Contextual expert evidence is also incorporated into the process. A curriculum graph represents courses, prerequisites, and programme structures. Constrained semester planning is used to generate feasible curriculum plans. Archived pair-classification and ranking results are retained as separate experiments. Results from six programmes are also reported as a separate pilot study. An earlier engineering exercise produces 80 curriculum plans. Consultations with 28 experts provide additional contextual evidence. These results are reported with their respective evidentiary limitations. Neither study is treated as independent proof of academic quality. Course provenance is exposed for committee inspection and revision. Post-generation structural checks are also applied to proposed plans. The current evidence supports supervised institutional piloting of Curriculum-KAG.

Keywords: Retrieval-Augmented Generation, Knowledge-Augmented Generation, Curriculum Design, Curriculum Planning, Educational Programmes, Knowledge Graphs, Multilingual Retrieval, Human Supervision, Evidence-Aware AI, Decision Support Systems, Higher Education, Academic Programme Development, Artificial intelligence, machine learning, Data Science, Big Data.

AMS Mathematics Subject Classification: 68T50, 68T30, 90C10.

DOI: 10.32523/2306-6172-2026-14-3-85-103

1 Introduction

Educational programme design follows several connected academic stages. Programme purpose and professional profiles are defined first. Programme learning outcomes (LOs) are then formulated to specify graduate knowledge and abilities [1]. Courses are selected for outcome instruction and assessment. Credits are allocated across semesters, and prerequisites are ordered. Academic and regulatory requirements are then checked. Each course must support at least one programme outcome [2]. These activities require evidence from existing programmes and expert judgement. Historical programme structures provide valuable evidence for curriculum design. However, such evidence requires appropriate retrieval and contextual interpretation. Course titles alone provide insufficient evidence for curriculum decisions. Course-to-LO

*Corresponding Author.

relations also require semantic and expert evidence. Curriculum planning therefore requires evidence retrieval and structural verification. Earlier platform and syllabus research is extended in e.g., [3], [4]. Prior curriculum studies are incorporated in [5], [6]. Future evaluation may use hybrid attainment models [7]. Kazakhstan provides a suitable national context for this study. Educational programme records are maintained within the Unified Higher Education Platform. The abbreviation EPVO is used throughout this article. Programme records have been submitted for approximately ten years. Bachelor’s, master’s, and doctoral programmes are represented. External specialists assess declared course-to-LO relations. Authentic programme structures and expert judgements are therefore retained. EPVO is not treated as a simple course catalogue. Programme identities, education levels, field codes, and group codes are stored. Programme goals, LOs, courses, credits, and recommended semesters are also stored. Course descriptions and course-to-outcome mappings are included. The public Registry is used as the authoritative interface [8]. This longitudinal evidence provides a basis for curriculum recommendation and comparison. Prior studies address digital education platforms and AI-assisted syllabus design [3, 4]. Curriculum graphs, sequence models, and educational prediction are also examined [5, 7]. The initial Curriculum-KAG design is previously introduced [6]. This study combines these directions within one curriculum-design process. Curriculum-KAG is proposed as a human-supervised decision-support approach. Retrieval-Augmented Generation (RAG) and Knowledge-Augmented Generation (KAG) are combined. Multilingual retrieval identifies programme evidence, while historical judgements provide course-to-LO evidence. A knowledge graph represents courses, outcomes, and prerequisites. Constrained planning then allocates courses across semesters. Users specify programme level, field, language, duration, credits, goals, and LOs. Approved EPVO courses are retrieved, and course-to-LO relations are estimated. Comparable historical judgements are transferred under defined conditions. One or three explainable alternatives are returned for review. The method avoids unrestricted text generation. Each course retains its source and LO evidence. Semester rationale and structural roles are also recorded. A deterministic verifier checks each plan before review. Credits, workload, prerequisites, level, domain relevance, and LO coverage are examined. Regulatory components are also checked. Human confirmation remains required before institutional use. Several limitations affect the evaluation. The historical 80-plan exercise contains only an archived success set. Failed attempts and repair records are unavailable. Later planner results belong to a separate evaluation branch. Proxy negatives are used in the frozen classification test. Independent negative expert judgements are unavailable. The 28-expert consultation also lacks blinded common samples. Therefore, structural passes are not treated as quality or accreditation evidence. Historical EPVO judgements are treated as expert memory. They reflect different institutions, periods, and local conditions. Thus, historical decisions are not treated as universal rules. Programme-disjoint evaluation and conservative transfer rules reduce evidence transfer. Post-generation checks are also applied, but complete leakage auditing is not claimed. The institutional design workflow also follows national methodological recommendations for universities[9]. Regulatory requirements are represented through SCES RK and GOSO RK [10]. Protected components are specified for each education level. These components are required under Kazakhstan-compliant mode. Some are not ordinary EPVO profile courses. Therefore, both complete-plan and profile-only results are reported. Four evaluation questions are examined in this study. H1 concerns domain-adapted embedding discrimination. Declared and sampled unlinked pairs are compared with an unadapted encoder. H2 concerns shortlist improvements from train-only expert memory. A fixed retrieval split is used for this evaluation. H3 concerns structural compliance under bounded planning. Prospective success rates are not estimated from historical plans.

H4 concerns classical and interdisciplinary programme modes. No causal regulatory effect is inferred from the six-programme pilot. Also five contributions are provided by the study. First, a reproducible national data process is established. Raw records, normalized entities, approved courses, and candidates are separated. Second, a two-stage recommendation process is provided. Semantic classification retrieves plausible links. Historical expert evidence then reorders the shortlist. Third, a graph-based curriculum planner is implemented. Credits, levels, domains, prerequisites, and statutory components are constrained. Fourth, an explanation and feedback process is provided for committee review. Course sources and linked outcomes are displayed. Model evidence and expert evidence remain separate. Uncertain bridge modules are also identified. Expert decisions are versioned for later inspection. Fifth, generated plans receive post-generation structural checks. Internal optimisation scores are not treated as independent validation. Curriculum-KAG uses a bounded repository of approved courses. Plausible course names are not freely generated. Whole-course credit values are preserved. Generative AI is restricted to marked assistance functions. Core plans remain reproducible without external APIs. The proposed approach supports supervised curriculum design and institutional review. Its current evidence does not support autonomous programme approval. Further evaluation is required across institutions and programme types.

2 Related Work

Sentence-BERT places sentences within a shared vector space [11]. Efficient semantic retrieval is thereby supported. Multilingual representations are important in Kazakhstan. Course texts appear in Kazakh, Russian, and English. Exact technical expressions are preserved through BM25 [12]. Such expressions may be weakened within sentence embeddings. Retrieval-Augmented Generation (RAG) is introduced by Lewis et al. [13]. Knowledge-Augmented Generation (KAG) adds structured entities and relations [14]. Transformer foundations are described in [15]. Dense retrieval is described in [16]. Curriculum construction is treated as multi-objective optimisation. Coverage, relevance, diversity, and interdisciplinarity are considered. Credits, duplication, prerequisites, and domain balance are constrained. An experimental Pareto branch is supplied by NSGA-II [17]. Bounded planning and repair are used operationally. Identical inputs can be rerun and checked. However, determinism is not independent academic verification. A shared checker is not independent verification either. The reranker is related to learning-to-rank objectives [18]. Curriculum dependencies can be represented by neural graph models. They can also be represented by recurrent architectures. Greater complexity is not automatically beneficial. Controlled GNN and LSTM pilots are compared with the baseline. No improvement is obtained on the frozen splits. Therefore, SBERT retrieval is retained. Graph reasoning is used for prerequisite closure [19]-[21]. AI support for syllabus design is reviewed in the 2024 study. A national operational generator is not implemented there. One IT-forensics example is emphasized in the initial 2026 paper. The present implementation is extended across educational levels. Graded evidence and multilingual records are included. Exact-credit planning and prerequisite graphs are included. Conditional SCES RK components are also included. PostgreSQL persistence and alternative plans are provided. Post-generation checks and multilingual interfaces are provided. Matched evidence for every component is not claimed.

2.1 KAG

Relevant text fragments are usually retrieved through RAG. Those fragments are supplied to a generative model. Better factual grounding can then be achieved. However, curricular rela-

tions are not expressed automatically. Course order is not established by retrieved passages. Outcome support and expert provenance are not established either. Complete credit and regulatory compliance are not guaranteed. Questions can be answered from documents through RAG. Valid curricula are not automatically constructed. KAG is discussed separately from RAG [14]. Curriculum-KAG is inspired by KAG. Every general-purpose KAG mechanism is not implemented. Canonical courses, levels, and fields are represented. Graded course-to-LO evidence is retained. Prerequisite and postrequisite edges are retained. Domain membership and regulatory status are represented. Credits, semester evidence, and provenance are represented. Candidates are filtered through these relations. Constrained planning is then applied. The stored plan receives post-generation checks. Structured relations provide additional academic functions. Course selection reasons can be exposed. Candidate prerequisite relations can be shown. Approved courses can be distinguished from synthetic bridges. Structurally invalid plans can be rejected. Universal superiority over RAG applications is not claimed. The earlier Curriculum-KAG concept is extended [6]. Classical and two-domain programmes are supported across levels. Figure 1 contrasts document retrieval with structured curriculum construction and verification.

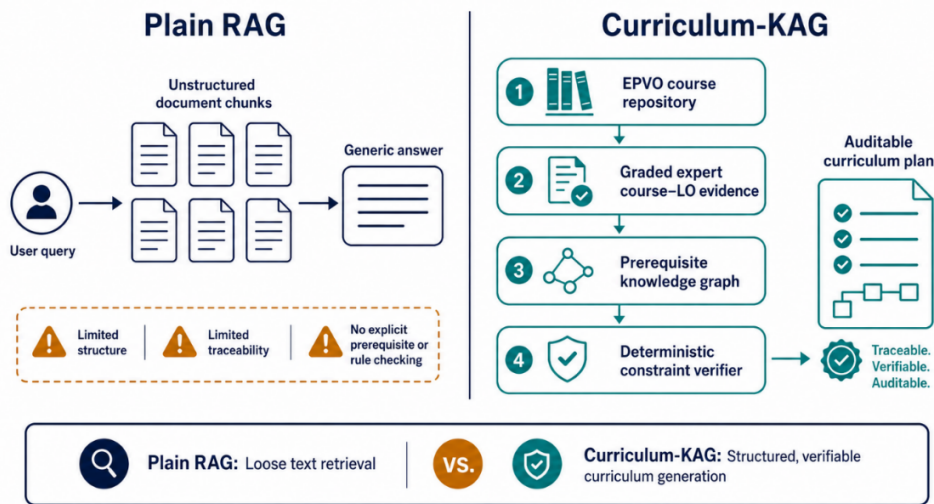


Figure 1: Conceptual comparison of plain RAG and Curriculum-KAG. The deterministic constraint check is a separate software stage, not independent academic validation.

3 Dataset and Evidence Formation

Data are obtained from the public EPVO Registry. Programme mappings are submitted by university departments. Those mappings are assessed by external experts. Values of 0, 0.5, and 1 are assigned. Expert strength is represented by these values. Model confidence is not represented. Four data layers are preserved. Raw records remain immutable. Normalized entities and links are stored separately. Approved multilingual courses form the operational repository. Project candidates are scoped from that repository. Every selected course remains traceable. PostgreSQL and pgvector are used for storage. Texts, embeddings, edges, scores, feedback, and plans are retained. Complete programmes are separated across data splits. Thresholds and weights are frozen before testing. Train-only memory is derived from 309 programmes. Repeated texts may still occur across programmes. All memorisation risk

is therefore not removed. Each programme is represented by level, field, group, goal, outcomes, and courses. Raw courses include texts, credits, components, semesters, and outcome links. Canonical courses may reference several source records. Duplicate plan positions are thereby prevented. Source provenance is preserved. ECTS workload principles are followed [22]. Computing scope can be checked against CC2020 [23]. Each expert observation includes an ordinal value. Different programmes may receive different judgements. Ordinal codes are not treated as validated intervals. Their mean is used only as an engineering summary. Raw records are normalized before approval. Whitespace, codes, credits, and multilingual fingerprints are standardized. Candidates are then filtered by level and scope. A dataset passport is maintained. Checksums, schemas, languages, splits, and limitations are recorded. Graded judgements are preserved during normalization. Binary collapse is therefore avoided. A total of 876,842 votes is aggregated. These votes produce 800,793 contextual pairs. The pairs provide evidence and evaluation targets. Curriculum admission is not determined automatically. Complete programmes remain within single splits. Shared programme vocabulary cannot cross those boundaries. Validation data are used for thresholds and early stopping. Test data are scored after those choices are frozen. Training negatives are sampled from unlinked outcomes. Most negatives receive high lexical overlap. Validation and test subsets each contain 6,000 rows. Declared and sampled classes are equally represented. Unlinked pairs are not confirmed expert rejections. Thus, proxy-label discrimination is measured. Direct expert agreement is not measured. Subgroup stratification is not claimed. Held-out votes are excluded from train-only memory. The permitted boundary is defined below. Repeated texts may still cross programme boundaries. Further lineage audits remain necessary. Figure 2 shows evidence formation, normalization, and programme-disjoint data splitting. Table 1 summarizes corpus entities and approved repository counts.

$$M_{\text{val}} = M_{\text{te}} = f(D_{\text{tr}}), \quad D_{\text{tr}} \cap D_{\text{val}} = D_{\text{tr}} \cap D_{\text{te}} = D_{\text{val}} \cap D_{\text{te}} = \emptyset. \quad (1)$$

$$y_{pcl} = \frac{1}{|E_{pcl}|} \sum_{e \in E_{pcl}} e, \quad |E_{pcl}| > 0. \quad (2)$$

$$P_a \cap P_b = \emptyset, \quad \forall a, b \in \{\text{tr}, \text{val}, \text{te}\}, \quad a \neq b. \quad (3)$$

Table 1: EPVO corpus and operational repository.

Dataset entity	Count
Educational programmes	12,215
Raw course records	408,638
Raw learning outcomes	124,521
Normalized courses	191,292
Normalized expert links	932,483
Pairs with numerical expert labels	800,793
Approved operational courses	21,525

4 Method

Course and outcome texts are encoded into normalized multilingual vectors. Cosine similarity provides semantic evidence. BM25 preserves exact professional terminology. Both scores are

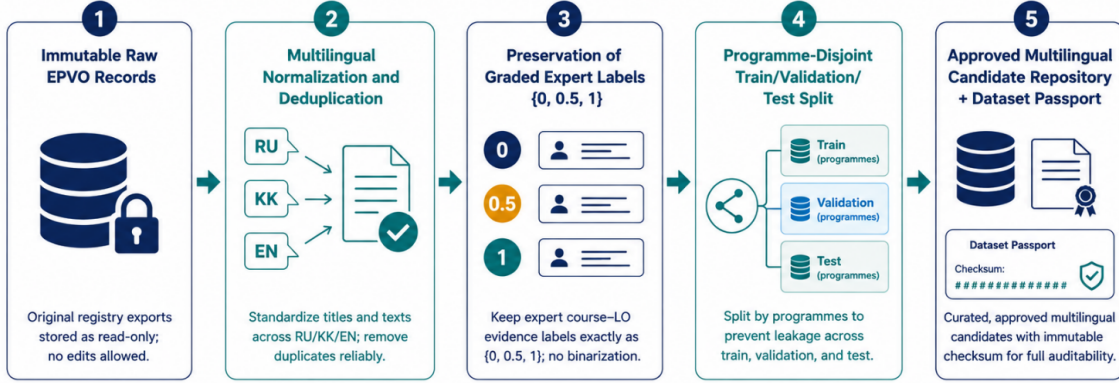


Figure 2: EPVO evidence formation, normalization, and programme-disjoint splitting. Dis-joint IDs do not alone rule out feature leakage or repeated text across programmes.

combined through hybrid retrieval. A fixed weight of 0.75 is used. Validation selection is not claimed for that weight. Historical evidence is transferred conservatively. Sufficient outcome similarity is required. Several votes are aggregated without binary collapse. Model and EPVO values are displayed separately. Admission requires appropriate level and scope. Outcome support and professional context are also required. Feasible prerequisite closure must remain possible. Approved courses receive preference. Bridge modules are used only for unresolved gaps. Every bridge requires expert confirmation. Semester assignment uses historical evidence and prerequisite depth. Semantic complexity and semester capacity are also considered. Any adjustment must preserve load and prerequisite rules. Figure 3 separates model evidence from transferred historical expert evidence.

$$\mathbf{u}_i = \frac{g_{\text{enc}}(t_i^{\text{course}})}{\|g_{\text{enc}}(t_i^{\text{course}})\|_2}, \quad \mathbf{v}_j = \frac{g_{\text{enc}}(t_j^{\text{outcome}})}{\|g_{\text{enc}}(t_j^{\text{outcome}})\|_2}, \quad \|g_{\text{enc}}(t_i^{\text{course}})\|_2, \|g_{\text{enc}}(t_j^{\text{outcome}})\|_2 > 0. \quad (4)$$

Here, t_i^{course} and t_j^{outcome} denote both texts.

$$s_{ij}^{\text{sem}} = \mathbf{u}_i^\top \mathbf{v}_j \quad (5)$$

$$r_{ij} = a s_{ij}^{\text{sem}} + (1 - a) s_{ij}^{\text{lex}}, \quad 0 \leq a \leq 1, \quad s_{ij}^{\text{lex}} \in [0, 1]. \quad (6)$$

$$p_{ij} = \text{logistic}\left(\frac{r_{ij} - r_0}{T}\right), \quad T > 0. \quad (7)$$

Scoring channels: evidence, not calibrated probability

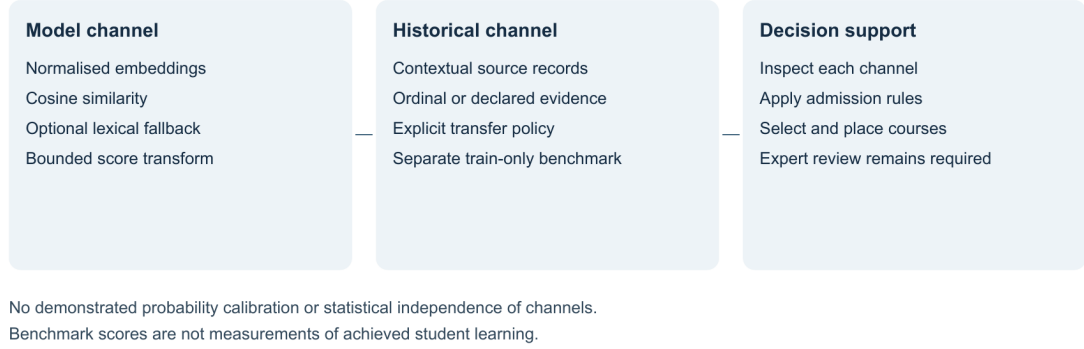


Figure 3: Model and historical evidence channels; neither percentage-formatted model score nor their fusion is shown to be calibrated probability.

4.1 Auditable Admission and Domain Evidence

Nearest-neighbour retrieval does not guarantee admission. Four binary safeguards are applied. Level, scope, outcome evidence, and context are checked. Every safeguard must be satisfied. Deduplication is completed before domain accounting. Evidence from all source records is aggregated. Exact group evidence receives priority. Direction evidence is used when required. Dual-domain support can satisfy interdisciplinary quotas. Database order cannot determine domain assignment. Each course must support at least one outcome. Strong model evidence can establish a relation. Comparable EPVO evidence can confirm a relation. Weak evidence cannot satisfy coverage. Protected components receive dedicated outcomes. Professional evidence is not falsely assigned.

$$a_{jl} = \begin{cases} \frac{|K(o_j) \cap K(l)|}{|K(o_j) \cup K(l)|}, & K(o_j) \cup K(l) \neq \emptyset, \\ 0, & K(o_j) \cup K(l) = \emptyset. \end{cases} \quad (8)$$

Here, $\mathcal{H}_i$ denotes historical programme–outcome pairs for course i .

$$e_{ij} = \max(\{0\} \cup \{y_{pil} a_{jl} : (p, l) \in \mathcal{H}_i, a_{jl} \geq a_{\min}\}). \quad (9)$$

$$A_i = I_i^{\text{level}} I_i^{\text{scope}} I_i^{\text{LO}} I_i^{\text{context}}, \quad I_i^r \in \{0, 1\}, \quad r \in \{\text{level}, \text{scope}, \text{LO}, \text{context}\}. \quad (10)$$

$$w_{ij} = \max(p_{ij}, e_{ij}, f_{ij}), \quad p_{ij}, e_{ij}, f_{ij} \in [0, 1]. \quad (11)$$

4.2 Exact Credits, Bridge Control, and Regulatory Closure

Each approved course is treated as an indivisible credit unit. Original credits cannot be reduced or divided. Protected components are inserted first. The remaining professional total is solved exactly. Dynamic programming is used for credit combinations. Evidence-bearing combinations receive preference. Synthetic fillers are considered last. Bridges represent new interdisciplinary content or unresolved outcome gaps. Generic placeholders are not treated

as disciplines. Three real alternatives are offered during replacement. Their credits, evidence, levels, and prerequisites are exposed. The expert selection is stored. Regulatory closure depends on mode and level. International plans omit Kazakhstan-specific components. SCES RK plans receive applicable protected components. Practice, research, and attestation requirements are retained. Protected items are excluded from replacement and variation. Figure 4 presents plan construction, verification, and final human approval.

$$C^* - d \leq \sum_i c_i x_i \leq C^* + d, \quad x_i \in \{0, 1\}, \quad c_i > 0, \quad d \geq 0. \quad (12)$$

For dynamic programming, credits are expressed in a common integer unit.

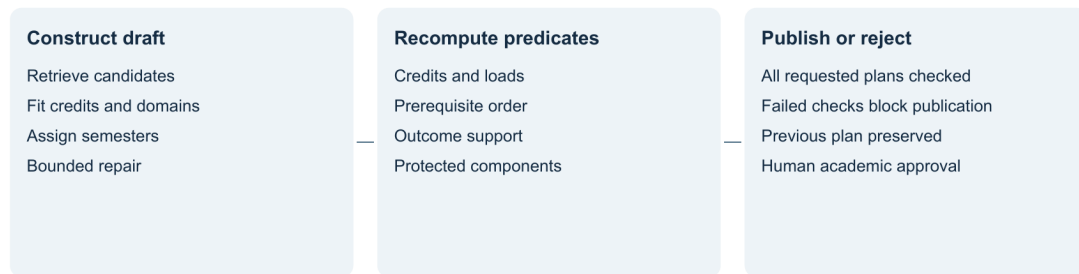
$$D_i(k) = \begin{cases} \max\{D_{i-1}(k), D_{i-1}(k - c_i) + q_i\}, & k \geq c_i, \\ D_{i-1}(k), & 0 \leq k < c_i, \end{cases} \quad (13)$$

$$D_0(0) = 0, \quad D_0(k) = -\infty \quad (k > 0).$$

$$k \in \mathbb{Z}_{\geq 0}, \quad c_i \in \mathbb{Z}_{> 0}.$$

$$B(P) = \sum_i x_i I_i^{\text{bridge}}, \quad x_i, I_i^{\text{bridge}} \in \{0, 1\}. \quad (14)$$

Plan construction and post-generation constraint checks



Construction and checks share implementation assumptions.
A structural pass is not independent validation or accreditation.

Figure 4: Plan construction, post-generation structural checks, and human approval. Shared implementation assumptions limit the independence of the software checker.

4.3 Knowledge Graph, Semester Repair, and A/B/C Transactions

Course and outcome nodes are represented in the graph. Prerequisites, outcomes, domains, and semesters are linked. Documented edges are separated from inferred suggestions. Inferred edges require expert review. Their precision and recall are not reported. Cycles are checked after each repair. Ordinary prerequisites must precede dependent courses. Concurrent regulatory activities are classified separately. Acyclic ordering indicates structural consistency only. Whole-course moves and swaps are used for repair. Hard violations receive priority. Pedagogical order is addressed afterward. Historical proximity is also considered. Advanced and foundation courses receive semester safeguards. Alternative plans vary weights and course selections. A shared regulatory core is preserved. Fingerprints and Jaccard similarity measure plan distinction. All alternatives are checked before commitment. Previous

plans remain intact after failed transactions. Figure 5 distinguishes the four evidence streams used in evaluation.

$$G = (V_C \cup V_O, E_{\text{pre}} \cup E_{\text{LO}} \cup E_{\text{sim}}) \quad (15)$$

$$\sum_{s=1}^S z_{is} = x_i, \quad z_{is}, x_i \in \{0, 1\}. \quad (16)$$

$$L_s \leq \sum_i c_i z_{is} \leq U_s, \quad 0 \leq L_s \leq U_s. \quad (17)$$

$$x_j = 1 \implies x_i = 1 \text{ and } \sum_{s=1}^S s z_{is} + 1 \leq \sum_{s=1}^S s z_{js}, \quad \forall (i, j) \in E_{\text{pre}}. \quad (18)$$

$$s_i^* \in \arg \min_{s \in F_i} |s - t_i|, \quad F_i \neq \emptyset. \quad (19)$$

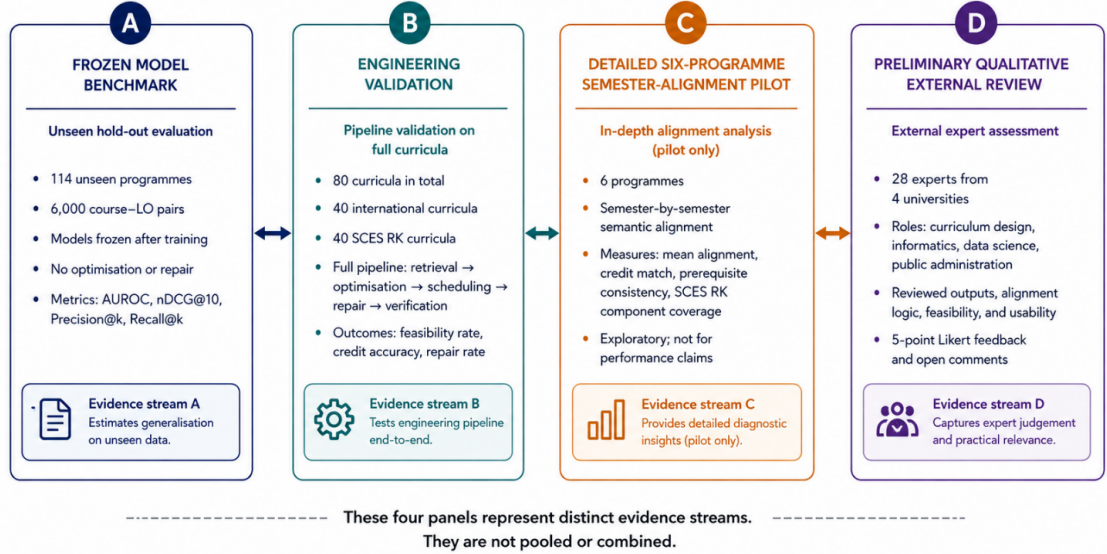


Figure 5: Distinct historical evidence streams: frozen pair benchmark, archived selected-plan engineering exercise, six-programme descriptive pilot, and preliminary qualitative expert consultation. The 80-plan panel does not show the number of initial attempts or repairs.

Outcome coverage, relevance, diversity, and semester agreement are rewarded. Redundancy and synthetic bridges are penalized. Hard constraints define the feasible set. Utility cannot compensate for a hard violation. The persisted plan receives post-generation checks. Credits, loads, prerequisites, levels, and uniqueness are examined. Domains, outcomes, and protected components are also checked. Every implemented predicate must be satisfied. Independent academic judgement is not provided by this check. Binary variables represent course selection and semester assignment. Each selected course receives one semester. Semester loads remain within defined limits. Prerequisite order is imposed. Bounded coverage is computed for

every outcome. Mean coverage cannot replace per-outcome requirements. The objective combines four positive terms. Redundancy and bridge counts are subtracted. Only hard-feasible plans are compared. Global optimality is not claimed for every branch. Bounded search and repair are used. Complexity depends on candidate size and credit range. Stored-plan checks remain deterministic.

$$C_j(P) = 1 - \prod_{i:x_i=1} (1 - w_{ij}), \quad 0 \leq w_{ij} \leq 1. \quad (20)$$

$$F_{\text{cov}}(P) = \frac{1}{m} \sum_{j=1}^m \min \left\{ \frac{C_j(P)}{c_{\min}}, 1 \right\}, \quad m \geq 1, \quad c_{\min} > 0. \quad (21)$$

$$U(P) = w_1 F_{\text{cov}}(P) + w_2 \text{Rel}(P) + w_3 \text{Div}(P) + w_4 \text{Sem}(P), \quad w_1, \dots, w_4 \geq 0. \quad (22)$$

$$P^* \in \arg \max_{P \in \mathcal{F}} \{U(P) - w_5 \text{Red}(P) - w_6 B(P)\}, \quad w_5, w_6 \geq 0, \quad \mathcal{F} \neq \emptyset. \quad (23)$$

$$V(P) = \prod_{k=1}^K I_k(P), \quad I_k(P) \in \{0, 1\}. \quad (24)$$

5 Experimental Design

Three evaluation levels are separated. Pair classification is examined first. Shortlist order is examined second. Complete-plan validity is examined third. ROC-AUC, PR-AUC, and F1 are reported for classification. Recall@K, MRR, and nDCG@K are reported for retrieval. Plan checks cover credits, loads, prerequisites, levels, and duplicates. Domains, outcomes, and protected components are also checked. An archived exercise includes 80 selected plans. Forty plans are assigned to each mode. A complete attempt ledger is unavailable. Initial failures and repairs are not fully recorded. Therefore, 80/80 is conditional on selected plans. A prospective generation rate is not established. Later planner work is kept separate. SCES RK is established by Kazakhstan's Order No. 2. The order is dated 20 July 2022. Protected requirements vary by education level. International mode omits this national block. Profile measures are therefore reported separately.

5.1 Metrics

Classification and retrieval measures answer different questions. ROC-AUC evaluates pair separation across thresholds. PR-AUC emphasizes sparse supported relations. F1 evaluates one frozen threshold. Recall@K measures shortlist coverage. MRR rewards an early first relevant course. nDCG rewards strong courses near the top. Pair results cannot establish plan validity. Relevant courses may still produce impossible credit totals. Prerequisite order may also be reversed. A staged evaluation hierarchy is therefore used. Classification, retrieval, graph consistency, and plan checks are separated. Qualitative expert review is applied last. The mode comparison is descriptive. Three different programmes are included per group. Random regulatory assignment is not used. Protected objects lack ordinary reference semesters. They are excluded from semester agreement. Planner rules adjust the reference semesters. The resulting measure is not independent. Educational outcomes are not measured.

$$A_p = \frac{1}{|I_p|} \sum_{i \in I_p} \mathbf{1}\{|s_i - \tilde{t}_i| \leq 1\}, \quad \bar{A}_G = \frac{1}{|G|} \sum_{p \in G} A_p, \quad |I_p| > 0, |G| > 0. \quad (25)$$

$$\text{Recall@}K(q) = \frac{|\text{Rel}(q) \cap \text{TopK}(q)|}{|\text{Rel}(q)|}, \quad |\text{Rel}(q)| > 0. \quad (26)$$

$$\text{MRR} = \frac{1}{|Q|} \sum_{q \in Q} \text{RR}(q), \quad \text{RR}(q) = \begin{cases} 1/\text{rank}_q, & \text{if a relevant item is retrieved,} \\ 0, & \text{otherwise,} \end{cases} \quad |Q| > 0. \quad (27)$$

$$\text{DCG@}K = \sum_{r=1}^K \frac{2^{g_r} - 1}{\log_2(r+1)}, \quad g_r \geq 0. \quad (28)$$

$$\text{nDCG@}K = \begin{cases} \frac{\text{DCG@}K}{\text{IDCG@}K}, & \text{IDCG@}K > 0, \\ 0, & \text{IDCG@}K = 0. \end{cases} \quad (29)$$

$$D(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}, \quad A \cup B \neq \emptyset. \quad (30)$$

6 Results

Three baseline values are reported. ROC-AUC equals 0.7031. PR-AUC equals 0.6946. F1 equals 0.6886. A 40,000-example multilingual checkpoint is evaluated. Programme-disjoint source data are used. Cosine-similarity loss and seed 42 are documented. Complete optimizer details are not recovered. A threshold of 0.33737 is selected on validation programmes. The held-out test contains 114 programmes and 6,000 pairs. Declared and sampled classes are equally represented. Test ROC-AUC equals 0.7647. Test PR-AUC equals 0.7673. Test F1 equals 0.7209. Precision equals 0.6212. Recall equals 0.8587. Confidence intervals are computed from programme-level resamples. Curriculum quality is not measured by this proxy task. The clean-v2 memory split is used. Training, validation, and test counts equal 309, 81, and 65. The test contains 531 LO queries. Candidate pools range from 7 to 66 courses. No programme-ID overlap is found. Exact texts still recur across splits. Memory is restricted to training rows. Recall@10 rises from 0.5348 to 0.5657. Recall@20 rises from 0.7476 to 0.7580. MRR rises from 0.2898 to 0.3740. nDCG@10 rises from 0.3163 to 0.3735. Complete-planner effects are not established. Repeated-text memorisation is not fully quantified. Other experiments are not pooled. The production model is not replaced. Six programmes are used for semester comparison. Three programmes are included per mode. Adjusted means equal 86.84% and 81.23%. The difference is descriptive. A causal regulatory effect is not established. One negative control is excluded. The pilot is not pooled with historical plans. All 80 selected plans pass archived checks. Every initial request is not represented. The percentage therefore remains conditional. Curriculum quality is not proved. Later code states are not combined. Table 2 reports the recoverable checkpoint and benchmark configuration. Table 3 lists programme-level semester agreement for both operating modes. Table 4 compares classification results across the evaluated model variants. Figure 6 compares raw and adjusted semester agreement across pilot programmes. Table 5 reports archived structural checks for the selected plans.

Table 2: Recoverable configuration and explicit omissions for the archived 40k checkpoint and frozen classifier rerun. Script defaults are not silently presented as the historical training invocation.

Item	Archived evidence or status
Base model and objective	<code>paraphrase-multilingual-mpnet-base-v2</code> ; checkpoint card: <code>CosineSimilarityLoss</code>
Train checkpoint	40,000 examples; exact batch size, learning rate, optimizer, schedule, weight decay, and gradient accumulation of the original invocation not independently recovered
Training-script defaults	Seed 42; batch 4; one epoch; learning rate 10^{-5} ; maximum 192 tokens; 75% lexically mined proxy negatives (defaults, not proof of the original invocation)
Validation	109 programmes; 6,000 pairs; F1-selected threshold 0.33737 for this rerun
Test	114 disjoint programmes; 6,000 pairs; 3,000 declared links and 3,000 sampled unlinked pairs; 2,000 programme-level bootstrap draws
Unreported subgroup detail	Per-language, level, and field scores and training-class distributions beyond the sampler’s one-to-one design were not retained in this summary

Table 3: Individual descriptive semester-alignment results in the six-programme pilot; approved EPVO courses with a valid reference semester only.

Mode	Project ID	Adjusted alignment (%)
International	13	90.91
International	15	83.33
International	16	86.27
SCES RK	19	80.00
SCES RK	136	84.38
SCES RK	137	79.31

Table 4: Course-to-learning-outcome classification.

Model / test	ROC-AUC	PR-AUC	F1	Status
Base multilingual SBERT	0.7031	0.6946	0.6886	baseline
Fine-tuned 40k; frozen	0.7647 [0.7378,	0.7673 [0.7367,	0.7209 [0.7032,	operational
114-programme test	0.7908]	0.7957]	0.7371]	
Cached 12k candidate	0.7748	0.7746	0.7268	experimental

6.1 End-to-End Control

The pilot covers classical and interdisciplinary programmes across all education levels. Plans of 240, 120, and 180 credits are produced. All stored plans pass structural checks. Real EPVO courses are preferred over generic fillers. Selected cases demonstrate individual safeguards. However, matched causal ablations are not provided. Separate component effects therefore cannot be isolated. Alternative plans may share mandatory components. Professional course selections may differ. Identical alternatives are detected by fingerprints. Universal diversity is not guaranteed.

Table 5: Historical author-reported structural checks on 80 selected plans; not a fixed-cohort generation success rate.

Verifier metric	International mode	SCES RK mode	Combined
Successful curricula	40	40	80
Internal checker pass among selected plans	100%	100%	100%
Exact target-credit completion	100%	100%	100%
Minimum LO threshold pass	100%	100%	100%
Mean hard prerequisite violations	0.00	0.00	0.00
Duplicate-course violations	0	0	0
Protected SCES components retained	not applicable	100%	100% of applicable plans

Individual semester agreement: raw and adjusted

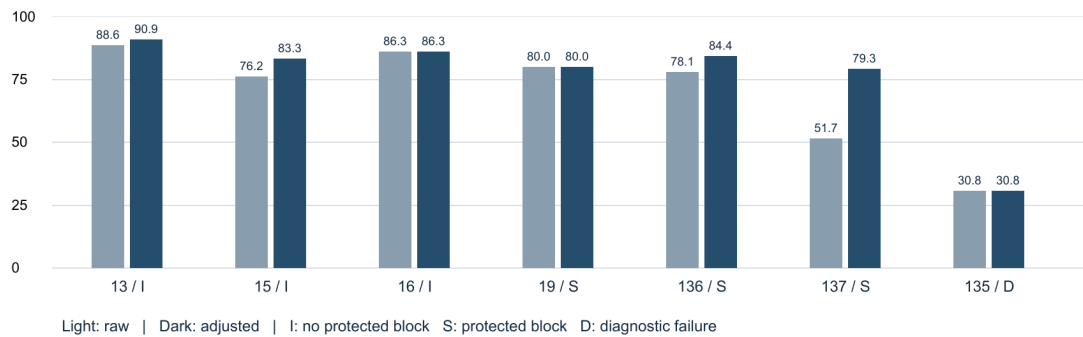


Figure 6: Individual raw and adjusted semester agreement for the six eligible pilot programmes, plus the separately flagged diagnostic control 135. The visual comparison is descriptive, not a causal estimate.

6.2 Efficiency, Failure Handling, and Explainability

Repeated preprocessing is identified as a major latency source. Cache signatures include relevant data and configuration. Unchanged results can therefore be reused. EPVO processing decreases from 28.45 to 0.15 seconds. Scoring decreases from 175.38 to 0.16 seconds. Generation decreases from 320.5 to 138.3 seconds. External AI services are treated as optional assistants. Structured contracts validate every response. Provider failure activates a deterministic fallback. The verifier is never bypassed. Progress is displayed during long operations. Three evidence values are kept separate. Final strength, model score, and EPVO evidence are distinguished. These values are not combined. Percentage formatting does not indicate calibrated probabilities. Course sequences are also listed. No LIME analysis is claimed. An interdisciplinary plan won't copy one historical programme. Several comparison measures are computed. Shared courses, overlap, containment, and recall are reported. High provenance may arise from several source directions. Therefore, multiple neighbours are displayed. The checklist is informed by ABET, OBE, CDIO, and Tuning. Observable outcome, credit, prerequisite, and domain items are checked. Faculty, facilities, support, and formal

review are excluded. A perfect score does not establish accreditation [24]-[26]. Both pilot groups receive a 100% checklist score. Semester agreement remains lower. Neither measure certifies academic quality. Disagreement is presented for committee review. The Learning Trajectory view supports visual audit. Courses, relations, outcomes, and progression are displayed. Documented, inferred, and semantic relations are distinguished. A no-cycles result is only structural. Every proposed edge remains open to expert challenge [6], [19]-[21]. Figure 7 displays semester courses, typed relations, and programme outcomes.

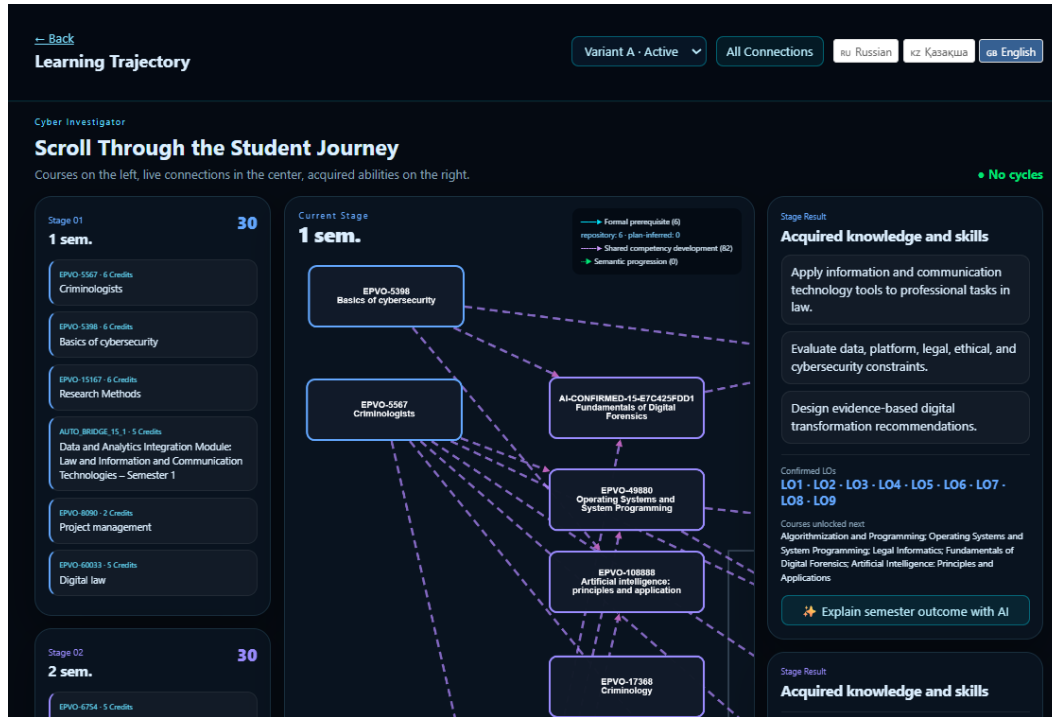


Figure 7: Learning Trajectory interface for visual inspection of semester courses, prerequisite relations, and learning outcomes.

7 Discussion

The results show that regulatory and professional evidence require separate interpretation. Raw SCES RK provenance accounts for 61.36%. Excluding protected activities raises profile-course provenance to 100%. Thus, statutory requirements are not treated as retrieval failures. Higher adjusted agreement is observed for international programmes. However, different programmes and adjustment rules are involved. A causal programme effect is therefore not established. Committee review remains necessary. External review supports the practical usability of scoped recommendations. Twenty-eight experts from four institutions assess seven domain families. Course relevance, semester order, completeness, and readiness are reviewed. Feedback is generally favourable, with limited manual revision recommended. Selection reasons and course chains remain visible to reviewers. This supports expert correction without replacing academic judgement. Table 6 summarizes reviewer institutions, counts, and programme domains. The system design supports this supervised process through traceable records and controlled transactions. Courses, outcomes, prerequisites, evidence, and plans are stored separately. Planning stages and evidence sources remain visible. Experts can con-

firm, correct, or reject proposed relations. Expert feedback is retained for later model cycles. Language records and text origins are also preserved. These features improve reproducibility and support institutional review. However, the available evidence has important limitations. The 80-plan exercise lacks a complete request denominator. The six-programme pilot lacks random mode assignment. The checker shares assumptions with the planner. Expert review lacks common blinded samples. Programme-disjoint splits reduce leakage risk but do not establish content validity. The reported measures address different system properties. They should therefore not be combined into a single quality measure. Broad effectiveness claims are not supported. These limitations define Curriculum-KAG as an academic co-designer. The system prepares structures and evidence for expert review. It identifies missing coverage and inconsistent prerequisites. Bridge modules remain curriculum hypotheses. Synthetic content is not assigned historical status. Titles, credits, outcomes, prerequisites, and assessments require confirmation. Final academic and legal approval remains institutional. Further evaluation should provide stronger independent evidence. A frozen catalogue and complete request ledger are required. Independent structural recomputation should reduce validator dependence. Harder negatives and graded objectives should be tested. Blind expert studies should use identical samples and shared rubrics. Temporal and subgroup analyses should assess transfer across contexts. Prospective programme observation should also assess longer-term validity. Figures 8 and 9 provide supporting evidence on provenance, expert evidence, and plan-level explanations.

Table 6: Composition and scope of the preliminary external review.

Institution	Experts	Review scope
Abylkas Saginov Karaganda Technical University	20	Engineering and technology programmes
Academy of Public Administration under the President of RK	3	Public administration and interdisciplinary programmes
Karaganda Economic University	3	Management and applied interdisciplinary programmes
Astana IT University	2	Information technology and digital programmes
Total	28	Different programmes across seven domain families

The interface displays three variants of a curriculum plan:

- Variant A (maximum LO coverage)** (Active)
- Variant B (minimum new courses)**
- Variant C (balanced, low redundancy)**

Semester 1 (30 Credits):

- Criminologists** (6 credits): Cycle: BD — basic disciplines (system estimate) - Component: EC — elective component - Source: course repository - Code: EPVO-5567. [Why selected?](#)
- Basics of cybersecurity** (6 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-5398. [Why selected?](#)
- Research Methods** (6 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-15167. [Why selected?](#)
- Integration module data and analytics: Law and Information and communication technologies — semester 1** (5 credits): Component: mandatory component - Source: bridge module. [Why selected?](#)
- Project management** (2 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-8090. [Why selected?](#)
- Digital law** (5 credits): Cycle: BD — basic disciplines (system estimate) - Component: EC — elective component - Source: course repository - Code: EPVO-60023. [Why selected?](#)

Semester 2 (30 Credits):

- Algorithmization and Programming** (5 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-6754. [Why selected?](#)
- Operating Systems and System Programming** (5 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-49880. [Why selected?](#)
- Legal Informatics** (5 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-85273. [Why selected?](#)
- Fundamentals of Digital Forensics** (4 credits): Cycle: BD — basic disciplines (system estimate) - Component: mandatory component - Source: AI-confirmed replacement - Code: AI-CONFIRMED-15-ETC425FD01. [Why selected?](#)
- Artificial Intelligence: principles and application** (4 credits): Cycle: BD — basic disciplines (system estimate) - Component: EC — elective component - Source: course repository - Code: EPVO-108888. [Why selected?](#)
- Criminology** (4 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-17368. [Why selected?](#)

Semester 3 (28 Credits):

- Technologies of Predictive Analytics in the Organization of Maintenance and Repair of Oil and Gas Machinery and Equipment** (5 credits): Cycle: BD — basic disciplines (system estimate) - Component: UC — university component - Source: course repository - Code: EPVO-807090. [Why selected?](#)

Learning-outcome links:

Strongest link — LO9: effective confidence 81%, AI and EPVO scores refer to the same "course → learning outcome" pair and are not additive.

How to read effective is the final plan link strength: AI is the test-model prediction; EPVO is similar expert evidence from EPVO. These are separate signals for one link, not a sum.

LO9: effective: 81% AI 81% EPVO 0%

Effective is the final strength for this course—LO link. AI is the test-model prediction, EPVO is reconstructed expert evidence. They are not added together.

LO9: Prepare expert opinions and investigation materials, present the results of technical analysis in a clear form for investigative authorities, the court, management, and other interested parties.

Final 81% — strength of the "Technologies of Predictive Analytics in the Organization, Maintenance and Repair of Oil and Gas Machinery and Equipment → LO9" link. AI 81% — model prediction from the course description and LO test. EPVO 0% — similar expert annotation from EPVO. These percentages are not added; the system uses the most reliable signal.

[Confirm](#) [Weak link](#) [Incorrect](#)

Pre- and post-requisites in this plan

Only courses that are actually present in the current plan variant are shown.

Before this course: no required earlier courses in the plan

Courses that depend on it: Sem. 6: Business Statistics, Sem. 6: Operating Systems, Sem. 6: Statistics, Sem. 6: Technologies of Defence of Computer Information, Sem. 7: System Software

Figure 8: Curriculum Plan Builder interface for comparing plan alternatives and inspecting course-level evidence.

17.5%
Typical EPVO course completeness

1198
Courses in EPVO reference set

7
Present in plan

95.8%
Plan courses from EPVO

43.5%
Expert support

This shows how many typical courses from the EPVO reference set are already included in the plan.

quality verdict
Below the EPVO 70–75% reference level
Formula: 70% completeness against typical EPVO courses + 30% expert support for course-LO links. **25.3 %**

Selected directions and groups **multi-scope**

1: B049
group: 600 disciplines

2: B057
group: 600 disciplines

Priority missing items [Add selected courses](#)

After adding them, rebuild the A/B/C variants in the plan builder.

all (10) will be added (0) in the repository (10) already in the plan (0) selected: 10

[select new](#) [select visible](#) [clear selection](#)

	TITLE	CREDITS	SEMESTER	SOURCE PROGRAMMES	EXPERT LINKS	STATUS	SIMILARITY
☒	Information security and information protection	4	7	50	103	in repository	14%
☒	International Criminal Law	6	3	36	103	in repository	9%
☒	Minor program	12	6	32	95	in repository	9%

Figure 9: EPVO comparison interface showing programme provenance and expert-evidence indicators.

8 Conclusions

In this study the Curriculum-KAG model is developed for interdisciplinary programmes. The same evidence process is later applied to classical programmes. Bachelor's, master's, and doctoral levels are supported. One or two professional directions can be selected. International and Kazakhstan-regulated modes are provided. Programme type, level, fields, goal, and outcomes are entered. The state expert corpus is then searched. Approved courses are proposed. Outcome support is explained. Prerequisite relations are constructed. Credits are balanced across semesters. One or three alternatives are returned. Retrieval is combined with structured knowledge. Provenance, expert evidence, and hard rules are retained. Plausible but invalid lists are therefore rejected. Improved proxy discrimination is reported after adaptation. Ranking gains are reported separately. Pilot agreement equals 86.84% and 81.23%. No causal regulatory effect is established. Selected plans pass archived checks. A prospective success rate is not established. The application is treated as an academic co-designer. Search, mapping, and structural checks can be assisted. Prospective staff-time reduction is not established. Experts must approve consequential academic decisions. Historical results support supervised institutional pilots. Autonomous readiness and accreditation validity are not demonstrated.

Acknowledgements

Minor use of AI for language editing is acknowledged. This research is supported by the Shakhmardan Yessenov Science and Education Foundation. The foundation was established by Galimzhan Yessenov which is awarded through the 2026 Scientific Internships Programme. The authors also acknowledge the support from project 2020-1.1.2-PIACI-KFI-2020-00129 and BR24993001 B«Creating a Large Language Model (LLM) to support the Kazakh language and technological progressB». This work was also supported by the National Key Research and Development Program of China (2013YFB3406801). An author is among the editorial board of the journal.

References

- [1] J. Biggs, "Enhancing teaching through constructive alignment," *High. Educ.*, vol. 32, no. 3, pp. 347–364, 1996, doi: 10.1007/BF00138871.
- [2] R. M. Harden, "Outcome-based education: The future is today," *Med. Teach.*, vol. 29, no. 7, pp. 625–629, 2007, doi: 10.1080/01421590701729930.
- [3] D. Jantassova, M. Kozhanov, and O. Shebalina, "Digital platform as a tool for internationalization: Model for formation of international competences' database applying the hierarchy analysis method," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 21, pp. 4942–4957, 2021.
- [4] M. Kozhanov *et al.*, "Syllabus design and planning with artificial intelligence and the potential of generative AI," in *Proc. IEEE 24th Int. Symp. Comput. Intell. Informat. (CINTI)*, 2024, pp. 257–264, doi: 10.1109/CINTI63048.2024.10830766.
- [5] M. Kozhanov and A. R. Várkonyi-Kóczy, "Modelling curricula with GNN and LSTM for link and sequence prediction," *Eurasian J. Math. Comput. Appl.*, vol. 13, no. 4, pp. 159–167, 2025, doi: 10.32523/2306-6172-2025-13-4-159-167.
- [6] M. Kozhanov, C. Mako, V. Poser, G. Eigner, and A. Mosavi, "Knowledge augmented generation for curriculum planning," *Eurasian J. Math. Comput. Appl.*, vol. 14, no. 1, pp. 17–34, 2026, doi: 10.32523/2306-6172-2026-14-1-17-34.
- [7] S. Najjar-Ghabel *et al.*, "Deep learning and gradient boosting for predicting educational attainment based on socioeconomic and regional features," *Eurasian J. Math. Comput. Appl.*, vol. 14, no. 2, pp. 95–114, 2026, doi: 10.32523/2306-6172-2026-14-2-95-114.

- [8] Unified Higher Education Platform of the Republic of Kazakhstan, “Registry of educational programs.” Accessed: Sep. 25, 2026. [Online]. Available: <https://epvo.kz/>
- [9] Ministry of Education and Science of the Republic of Kazakhstan, *Metodicheskie rekomendatsii dlya vuzov po proektirovaniyu obrazovatel'nykh programm* [Methodological recommendations for universities on educational-programme design], 2017 (in Russian). Accessed: Sep. 25, 2026. [Online]. Available: <https://info.aues.kz/info/documents/18.pdf>
- [10] Ministry of Science and Higher Education of the Republic of Kazakhstan, “On approval of the state compulsory standards of higher and postgraduate education,” Order No. 2 of Jul. 20, 2022, reg. No. 28916. Accessed: Sep. 25, 2026. [Online]. Available: <https://adilet.zan.kz/rus/docs/V2200028916>
- [11] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 3982–3992, doi: 10.18653/v1/D19-1410.
- [12] S. Robertson and H. Zaragoza, “The probabilistic relevance framework: BM25 and beyond,” *Found. Trends Inf. Retr.*, vol. 3, no. 4, pp. 333–389, 2009, doi: 10.1561/15000000019.
- [13] P. Lewis *et al.*, “Retrieval-augmented generation for knowledge-intensive NLP tasks,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 9459–9474.
- [14] L. Liang *et al.*, “KAG: Boosting LLMs in professional domains via knowledge augmented generation,” 2024, arXiv:2409.13731.
- [15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol. (NAACL-HLT)*, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
- [16] V. Karpukhin *et al.*, “Dense passage retrieval for open-domain question answering,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 6769–6781, doi: 10.18653/v1/2020.emnlp-main.550.
- [17] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002, doi: 10.1109/4235.996017.
- [18] C. Burges *et al.*, “Learning to rank using gradient descent,” in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 89–96, doi: 10.1145/1102351.1102363.
- [19] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017.
- [20] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 1024–1034.
- [21] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [22] European Commission, *ECTS Users’ Guide 2015*. Luxembourg: Publications Office of the European Union, 2015, doi: 10.2766/87192.
- [23] CC2020 Task Force, *Computing Curricula 2020: Paradigms for Global Computing Education*. New York, NY, USA: ACM and IEEE Computer Society, 2020, doi: 10.1145/3467967.
- [24] ABET, “Criteria for accrediting computing programs, 2025–2026.” Accessed: Sep. 25, 2026. [Online]. Available: <https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2025-2026/>
- [25] E. F. Crawley, J. Malmqvist, S. Östlund, D. R. Brodeur, and K. Edström, *Rethinking Engineering Education: The CDIO Approach*, 2nd ed. Cham, Switzerland: Springer, 2014, doi: 10.1007/978-3-319-05561-9.
- [26] J. González and R. Wagenaar, Eds., *Tuning Educational Structures in Europe: Final Report, Phase One*. Bilbao, Spain: Univ. Deusto, 2003.

Murat Kozhanov*

Academy of Public Administration under
the President of the Republic of Kazakhstan,
Astana, Kazakhstan.

Obuda University, Budapest, Hungary.

Email: m.kozhanov@ktu.edu.kz

Madina Kozhukhova

Abylkas Saginov Karaganda Technical Uni-
versity,

Karaganda, Kazakhstan.

Amir Mosavi

Abylkas Saginov Karaganda Technical Uni-
versity,

Karaganda, Kazakhstan.

Univerzita J. Selyeho, Komarom, Slovakia.

Ludovika University of Public Service,
Budapest, Hungary.

Obuda University, Budapest, Hungary.

Farabi University, Kazakhstan.

Received 09.01.2026, Accepted 10.07.2026, Available online 30.09.2026.