

UDC 004.8

Amirlan Seitzkadyrov
Secondary-school student
Space School
(Astana, Kazakhstan)

Akezhan Omashev
corresponding author
Secondary-school student
National School of Physics and Math
(Astana, Kazakhstan)

AUTOMATIC MULTIPLE-CHOICE QUESTION GENERATION USING ARTIFICIAL INTELLIGENCE

Abstract. This paper presents a local pipeline for generating multiple-choice questions for IGCSE and A-Level computer science. From 3,385 source questions, duplicate removal and answer-position balancing produced 5,702 training and validation examples. Qwen2.5-7B-Instruct was adapted using QLoRA on an NVIDIA RTX 3060 with 12 GB of memory, and generated candidates underwent automatic structural validation. An author audit judged 15 of 18 candidates usable without substantive correction. A subsequent external expert review rated 51 of 57 items (89.5%) usable as generated and six usable after minor revision; no factual errors were recorded. Answer leakage and repeated stems remained important defects. The study documents the data-preparation procedure, training configuration, and evaluation limitations. The findings support a teacher-in-the-loop proof of concept, while small samples, limited rater independence, missing controlled baselines, and unresolved source-data permissions constrain broader claims.

Key words: automatic question generation, multiple-choice questions, large language models, QLoRA, computer science education.

1. Introduction

Constructing high-quality assessment items requires subject knowledge, item-writing skill, and substantial teacher time. Automatic question generation can reduce this workload, but educational use depends on factual correctness, plausible distractors, appropriate difficulty, and reliable formatting [1, 2].

This study addresses school computer science. Given an educational level (International General Certificate of Secondary Education, IGCSE, or A-Level) and a topic, AI MCQ Generator produces a question stem, four options labelled A–D, and an answer key. The system uses an open large language model adapted to exam-style items and is designed to run on local hardware.

The research question is whether a QLoRA-adapted 7-billion-parameter instruction model, trained on a balanced corpus of exam-style examples, can generate format-valid and mostly correct computer science MCQs on a consumer graphics processing unit. The working hypothesis is that pretraining supplies much of the subject knowledge, while parameter-efficient fine-tuning can teach the required output structure and style.

The work makes three practical contributions: (a) a documented, repeatable cleaning and augmentation procedure that removes duplicate items and controls answer-position bias; (b) a resource-constrained QLoRA configuration for adapting Qwen2.5-7B-Instruct on an NVIDIA RTX 3060 with 12 GB of memory; and (c) a two-stage manual evaluation (an author self-audit followed by an external expert review of 57 items) that reports usable and defective outcomes while making the limits of the evidence explicit.

1.1. Related Work

Automatic question generation has progressed from rule-based overgeneration and statistical ranking [3] to end-to-end neural generation [4]. These approaches established that fluent wording alone is insufficient: semantic validity and alignment with an intended answer remain central.

A systematic review of 93 educational question-generation studies identified controlled difficulty, richer question forms, evaluation consistency, and feedback generation as continuing challenges [1]. For MCQs, question-stem generation, answer selection, and distractor generation form distinct subproblems, with distractor quality repeatedly identified as a weak point [2].

Large language models have broadened the task from specialized pipelines to prompt-based generation. Their educational use nevertheless requires human oversight because plausible text may contain errors [5]. In programming education, GPT-4-generated MCQs have been found comparable to human-written items on several rubric dimensions, but instructor review was still recommended [6].

Qwen2.5-7B-Instruct is a decoder-only Transformer model; the Transformer architecture supports scalable sequence modelling [7], while the Qwen2.5 report documents instruction tuning and coding-oriented capabilities [8]. The model is therefore a plausible base for school-level computer science generation, although this study does not independently benchmark its general subject knowledge.

Full fine-tuning is costly on consumer hardware. Low-Rank Adaptation (LoRA) freezes pretrained weights and learns low-rank updates [9]. QLoRA combines LoRA with a frozen 4-bit quantized base model, including NF4 quantization and paged optimizers, to reduce memory use during adaptation [10].

2. Materials and methods

2.1. Data Collection and Preparation

The source corpus contained 3,385 parsed MCQs collected manually from publicly accessible IGCSE and A-Level practice materials. Topics included programming, algorithms, data structures, networking, security, databases, logic, and computer architecture. Individual source licences and redistribution permissions were not recorded consistently; this is treated as a data-governance limitation rather than evidence that the items are freely reusable.

Whitespace was normalized, levels were mapped to the two target categories, and duplicates were identified by the tuple of level, lower-cased question text, and lower-cased correct-answer text. Removing 534 duplicate records left 2,851 deduplicated level/question/answer records. Figure 1 shows one normalized record.

```
Level: IGCSE
Topic: Algorithm Design
Question: Which of the following describes a step-by-step
procedure for solving a problem?
A. Flowchart
B. Pseudocode
C. Algorithm
D. Compiler
Answer: C
```

Figure 1 — A source-corpus record (IGCSE, topic “Algorithm Design”).

The original answer positions were imbalanced: B was correct in 41% of records and D in 5%. Each deduplicated record was therefore emitted twice with shuffled distractors, while the correct option followed a global A–B–C–D cycle. The resulting 5,702 examples were balanced as evenly as mathematically possible: A and B each occurred 1,426 times, and C and D each occurred 1,425 times. Thus each letter accounts for approximately 25%, not exactly 25%, of the file. The final level counts were 3,228 IGCSE and 2,474 A-Level examples (Figure 2).

```
Blocks in source file:      3387
Parsed successfully:       3385
After removing duplicates: 2851
Written to MCQ_balanced.txt: 5702 (x2 augmentation)
Answer distribution:       A: 1426, B: 1426,
                           C: 1425, D: 1425
Levels:                   IGCSE: 3228,
                           A-Level: 2474
```

Figure 2 — Output of the dataset-preparation script.

For validation, the training script grouped all 5,702 examples by lower-cased question text so that shuffled copies with the same stem could not appear on opposite sides. This produced 2,698 question-text groups. With a 10% group holdout and seed 42, 2,428 groups (5,128 examples) formed the training set and 270 groups (574 examples) formed the validation set.

2.2. Model and Training

The base model was Qwen2.5-7B-Instruct. Its weights were loaded in 4-bit NF4 with double quantization and bfloat16 computation, then frozen. Rank-16 LoRA adapters (alpha 32; dropout 0.05) were attached to the q, k, v, and output attention projections and to the gate, up, and down multilayer-perceptron projections. For the published Qwen configuration, this corresponds to 40,370,176 trainable adapter parameters, approximately 0.53% of the 7.6-billion-parameter base.

Training was configured for up to three epochs, batch size 2, gradient accumulation 8 (effective batch 16), maximum sequence length 256, learning rate 2×10^{-4} with a cosine schedule, 30 warm-up steps, a paged 8-bit AdamW optimizer, and gradient checkpointing. Validation loss was evaluated after each epoch; the best checkpoint was reloaded, and an early-stopping callback with patience 2 was enabled. Training and validation-loss traces were not retained in a manuscript-ready record, so this paper does not claim a quantified convergence or overfitting advantage.

Prompt tokens were masked with label -100 , so loss was computed on the complete assistant response (the question, four options, and answer key) rather than on the user prompt

or only on the answer letter. The run used the Transformers, PEFT, and bitsandbytes software stack [11, 12, 13, 14]. The saved adapter occupied approximately 162 MB. Seed 42 makes data preparation and the group split repeatable; bitwise-identical GPU training is not guaranteed without pinned software versions, deterministic kernels, a fixed model revision, and matching hardware. Figure 3 summarizes the core configuration.

```
bnb_config = BitsAndBytesConfig(
    load_in_4bit=True,
    bnb_4bit_quant_type="nf4",
    bnb_4bit_use_double_quant=True,
    bnb_4bit_compute_dtype=torch.bfloat16,
)
lora_config = LoraConfig(
    r=16, lora_alpha=32, lora_dropout=0.05,
    task_type="CAUSAL_LM",
    target_modules=["q_proj", "k_proj", "v_proj", "o_proj",
                   "gate_proj", "up_proj", "down_proj"],
)
```

Figure 3 — Core quantization and LoRA configuration used for training.

2.3. Generation and Validation

At inference time, the quantized base model was loaded from a local cache and the LoRA adapter was applied. The user supplied a level and topic. The prompt was “Generate a {level} Computer Science multiple-choice question. Topic: {topic}”. Six completions were sampled with temperature 0.95, top-p 0.95, and a maximum of 160 new tokens.

A regular-expression parser required a question, four labelled options, and an answer letter. Outputs with malformed structure, duplicate options, or embedded option labels were rejected; duplicate formatted candidates were printed once. The available evaluation helper compared extracted single-line question strings with source-corpus text. Because some stems span multiple lines and the exact training membership was not persisted, that diagnostic is insufficient for a quantitative novelty claim; no novelty rate is therefore reported here.

2.4. External Expert Evaluation

After the preliminary self-audit (Section 3.3), the authors organized a larger external review. The rater is affiliated with the Department of Artificial Intelligence Technologies at L. N. Gumilyov Eurasian National University; he supervised the project but did not prepare the data, train the model, or author any generated item. The rating sheet contained 57 generated items (29 IGCSE and 28 A-Level) spanning six topics: programming (12 items), databases (10), logic gates and logic circuits (10), networking (9), data representation (8), and security (8).

Each item was scored on five dimensions: format validity (0 or 1), factual correctness (0 or 1), distractor quality (0–2), difficulty appropriateness for the stated level (0 or 1), and overall usability, where 2 means usable as generated, 1 means usable after minor revision, and 0 means not usable. The rater saw each complete item, including the answer key, and was not blinded to the system that produced the items. The completed item-level rating sheet is archived with the project materials.

3. Results and discussion

3.1. Answer-Position Bias

In early fine-tuning runs, the model frequently returned B irrespective of question content. The source data offered a plausible explanation because B was the correct position in 41% of

records. After the balancing procedure, this behaviour was not observed in subsequent spot checks.

This engineering observation illustrates how a superficial label-frequency pattern can compete with content learning. However, no controlled ablation or before/after error rate was retained, so the evidence supports a practical diagnostic, not a quantified causal estimate.

3.2. Base-Model Comparison

The first prototype used FLAN-T5-large, an instruction-tuned text-to-text model from the T5 family [15, 16]. It learned the requested MCQ format but produced factual errors during informal testing. After replacement with Qwen2.5-7B-Instruct, factual errors appeared less frequently in spot checks.

The comparison was not conducted under a shared prompt, training, sampling, or blinded-rating protocol. Model size, architecture, pretraining data, instruction tuning, and implementation all changed together. The result is therefore reported as a development observation and not as evidence that parameter count alone determined quality.

3.3. Preliminary Manual Audit

One complete generation session was audited manually. For three topics, all six sampled candidates were assessed for structural validity and factual correctness. The reviewer was one of the authors and was not blinded to the system.

Fifteen of 18 candidates (83.3%) were usable without substantive correction. Because the sample is small, the 95% Wilson interval for the usable proportion is wide (60.8%–94.2%). Table 1 reports the usable and defective counts by topic.

Table 1 — Manual audit of one generation session (usable and defective are mutually exclusive).

Topic (level)	Cand.	Usable	Defective
Networking (IGCSE)	6	6	0
Databases (IGCSE)	6	5	1
Logic gates and logic circuits (A-Level)	6	4	2
Total	18	15	3

The three defective items comprised one incorrect database statement, one item with both an incorrect answer key and a malformed option in a De Morgan transformation, and one truncated stem that still passed the format filter. This taxonomy shows that structural validation alone cannot replace subject-matter review.

Usable examples included a networking item asking for the purpose of DHCP, a database item identifying UPDATE as the SQL command that modifies existing records, and a half-adder item correctly giving Sum = 0 and Carry = 1 when both inputs equal 1. Figure 4 shows the networking example.

Level (IGCSE / A-Level): IGCSE

Topic: Networking

--- Question 5 ---

Question: What is the purpose of DHCP in networking?

A. Assign IP addresses to devices on the network automatically

B. Encrypt network traffic

C. Block unauthorized connections

D. Filter spam emails

Answer: A (Assign IP addresses to devices on the network automatically)

Figure 4 — Example interaction from the audited generation session.

3.4. External Expert Evaluation

Table 2 summarizes the external review described in Section 2.4. All 57 items were rated factually correct, and 56 of 57 received the maximum distractor-quality score. Overall, 51 of 57 items (89.5%; 95% Wilson interval 78.9%–95.1%) were rated usable as generated, the remaining six were rated usable after minor revision, and no item was rated unusable.

Table 2 — External expert evaluation of 57 generated items by topic (usability: 2 = usable as generated, 1 = usable after minor revision, 0 = not usable).

Topic	Items	Usable as generated	Minor revision	Not usable
Programming	12	12	0	0
Databases	10	9	1	0
Logic gates and logic circuits	10	10	0	0
Networking	9	9	0	0
Data representation	8	7	1	0
Security	8	4	4	0
Total	57	51	6	0

Six items were marked on the format-validity dimension, and all six share the same failure mode: the stem wording leaks the correct option. Four are near-duplicate security stems of the form “Which security measure involves encrypting data ...” whose correct option is “Encryption”; one asks which attack “exploits buffer overflows” while listing “Buffer overflow” among the options; and one asks which data type represents “a sequence of bits that can be interpreted as an integer value” with “Integer” as the correct option. A seventh flagged item, a databases question, was structurally sound but judged too easy for its stated level. Six items in total were rated as needing minor revision: five of the six with leaking stems, and the databases item.

The flagged items point to two broader issues. First, answer leakage in the stem is a semantic defect that the regular-expression format filter cannot detect, which reinforces the case for teacher review. Second, the recurrence of four nearly identical security stems within one rating sheet shows that repeated sampling can produce low-diversity items; a production

pipeline should therefore add near-duplicate detection across generated items rather than rely on per-item checks alone.

The external review is consistent with the smaller self-audit (89.5% versus 83.3% usable as generated), and its factual-correctness result (57 of 57) is stronger than the self-audit, in which two of the three defects were factual. The two evaluations used different item sets and different raters, so they cannot be combined into an inter-rater agreement estimate. Because the rater supervised the project and was not blinded, the review is external but not fully independent; it should be read as a substantially larger second opinion rather than a definitive evaluation.

3.5. Limitations

The self-audit covers only 18 outputs, three topics, one session, and one author-reviewer. The external review is larger (57 items across six topics, rated by a non-author expert on five dimensions), but it relies on a single unblinded rater who supervised the project, so no inter-rater agreement can be computed and rater independence is limited. Neither evaluation includes a held-out topic set, a matched zero-shot or few-shot baseline, a FLAN-T5 baseline under the same protocol, or a formal ablation of answer balancing. Item difficulty and distractor quality were rated ordinally in the external review but not measured psychometrically. The reported 83.3% and 89.5% figures should therefore be treated as pilot estimates rather than general performance claims.

The available line-based comparison does not reliably reconstruct multiline stems or the actual training subset and cannot detect paraphrases or semantic near-duplicates. The original 18 generated texts and the item-level scoring sheet of the self-audit were not preserved, so only the reported aggregate usability counts remain available for that pilot; in contrast, the 57-item external rating sheet is archived item by item. Training logs sufficient to report final epoch, best validation loss, runtime, and variance across seeds were not archived. A future study should pin the software environment and model revision, preserve all run artefacts, and repeat training across seeds.

The source items were collected from publicly accessible practice materials, but a source-by-source licence and permission audit was not completed. Raw and processed question data should not be redistributed or reused until that audit is complete. This limitation does not apply to the authors' original code, but it materially constrains claims of open data and requires repository review.

3.6. Ethics and Intended Use

The system is intended for teacher-in-the-loop use: generation, teacher verification, revision where necessary, and only then delivery to students. Unmoderated use would be inappropriate: the pilot self-audit found factual or structural defects in 3 of 18 outputs, and the external review, although it recorded no factual errors, still judged 6 of 57 items to need revision before classroom use. Inference runs locally, so the workflow does not require student data or school content to be sent to a commercial application programming interface.

4. Conclusions

This study implemented a QLoRA adaptation of Qwen2.5-7B-Instruct for IGCSE and A-Level computer science MCQ generation on a 12 GB consumer GPU. The pipeline removed duplicates, balanced answer positions as evenly as possible, grouped augmented copies across the train/validation split, trained 40.37 million adapter parameters, and validated generated structure automatically.

In the preliminary audit, 15 of 18 candidates were usable. In the subsequent external expert review, 51 of 57 items were rated usable as generated, six required minor revision, none

was unusable, and no factual errors were recorded; the dominant defect was answer leakage in the stem rather than incorrect content. The wide confidence intervals and single-rater designs of both evaluations prevent broad effectiveness claims. The answer-position and base-model observations are useful hypotheses and engineering lessons, but they require controlled ablation and matched baselines before causal interpretation.

The next priority is a blind evaluation across more topics with at least three independent raters, a documented rubric, inter-rater agreement, held-out topics, confidence intervals, and matched zero-shot, few-shot, FLAN-T5, and answer-balancing baselines, together with automatic near-duplicate and answer-leakage checks suggested by the external review. Only after that validation should the project add explanations, Russian and Kazakh generation, difficulty control, learning-management-system integration, and a classroom pilot. The present contribution is therefore a documented proof of concept, not a deployment-ready autonomous assessor.

Acknowledgments

The authors thank Asset Alkhanov (Department of Artificial Intelligence Technologies, L. N. Gumilyov Eurasian National University, Astana, Kazakhstan) for supervising the project and for serving as the expert rater in the external evaluation.

Declarations

Data availability. The implementation, data-preparation scripts, evaluation scripts, the external-evaluation rating sheet, and the trained LoRA adapter are available at <https://github.com/AmirlanSeit/mcq-generator>. Source-derived question files are excluded from this availability claim and require a source-by-source rights audit before redistribution or reuse.

Funding. The authors report that this work received no external funding.

Conflict of interest. The authors declare no conflict of interest.

Ethics approval and consent. Not applicable. The study did not involve human participants, personal data, or animals.

Generative AI use statement. During preparation of this work, the authors used OpenAI Codex (GPT-5) for language editing, translation-related tasks, bibliography preparation, and document formatting, and Claude (Anthropic) for integrating the external evaluation results and preparing the manuscript versions. After using these tools, the authors reviewed and edited the content as necessary and take full responsibility for the content of the publication.

References:

1. Kurdi G., Leo J., Parsia B., Sattler U., Al-Emari S. A systematic review of automatic question generation for educational purposes // Int. J. Artif. Intell. Educ. — 2020. — Vol. 30. — P. 121–204. DOI: [10.1007/s40593-019-00186-y](https://doi.org/10.1007/s40593-019-00186-y).
2. Ch D.R., Saha S.K. Automatic multiple choice question generation from text: A survey // IEEE Trans. Learn. Technol. — 2020. — Vol. 13, no. 1. — P. 14–25. DOI: [10.1109/TLT.2018.2889100](https://doi.org/10.1109/TLT.2018.2889100).
3. Heilman M., Smith N.A. Good question! Statistical ranking for question generation // Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics. — 2010. — P. 609–617. URL: <https://aclanthology.org/N10-1086/>.
4. Du X., Shao J., Cardie C. Learning to ask: Neural question generation for reading comprehension // Proceedings of the 55th Annual Meeting of the Association for Computational

Linguistics (Volume 1: Long Papers), Association for Computational Linguistics. — 2017. — P. 1342–1352. DOI: [10.18653/v1/P17-1123](https://doi.org/10.18653/v1/P17-1123).

5. Kasneci E., Seßler K., Küchemann S., et al. ChatGPT for good? On opportunities and challenges of large language models for education // Learn. Individ. Differ. — 2023. — Vol. 103. — Article 102274. DOI: [10.1016/j.lindif.2023.102274](https://doi.org/10.1016/j.lindif.2023.102274).

6. Doughty J., Wan Z., Bompelli A., et al. A comparative study of AI-generated (GPT-4) and human-crafted MCQs in programming education // Proceedings of the 26th Australasian Computing Education Conference, ACM. — 2024. — P. 114–123. DOI: [10.1145/3636243.3636256](https://doi.org/10.1145/3636243.3636256).

7. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems. — 2017. — Vol. 30. — P. 5998–6008. DOI: [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762).

8. Qwen Team. Qwen2.5 technical report [Electronic resource]. — arXiv:2412.15115v1, 2024. URL: <https://arxiv.org/abs/2412.15115v1>.

9. Hu E.J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., Wang L., Chen W. LoRA: Low-rank adaptation of large language models // International Conference on Learning Representations, 2022. URL: <https://openreview.net/forum?id=nZeVKeeFYf9>.

10. Dettmers T., Pagnoni A., Holtzman A., Zettlemoyer L. QLoRA: Efficient finetuning of quantized LLMs // Advances in Neural Information Processing Systems. — 2023. — Vol. 36. — P. 10088–10115. DOI: [10.52202/075280-0441](https://doi.org/10.52202/075280-0441).

11. Wolf T., Debut L., Sanh V., et al. Transformers: State-of-the-art natural language processing // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics. — 2020. — P. 38–45. DOI: [10.18653/v1/2020.emnlp-demos.6](https://doi.org/10.18653/v1/2020.emnlp-demos.6).

12. Hugging Face. PEFT documentation [Electronic resource]. — 2026. URL: <https://huggingface.co/docs/peft/> (accessed: 15.07.2026).

13. Hugging Face. Transformers documentation [Electronic resource]. — 2026. URL: <https://huggingface.co/docs/transformers/> (accessed: 15.07.2026).

14. Dettmers T., Lewis M., Belkada Y., Zettlemoyer L. LLM.int8(): 8-bit matrix multiplication for transformers at scale // Advances in Neural Information Processing Systems. — 2022. — Vol. 35. — P. 30318–30332. DOI: [10.52202/068431-2198](https://doi.org/10.52202/068431-2198).

15. Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P.J. Exploring the limits of transfer learning with a unified text-to-text transformer // J. Mach. Learn. Res. — 2020. — Vol. 21, no. 140. — P. 1–67. URL: <https://jmlr.org/papers/v21/20-074.html>.

16. Chung H.W., Hou L., Longpre S., et al. Scaling instruction-finetuned language models // J. Mach. Learn. Res. — 2024. — Vol. 25, no. 70. — P. 1–53. URL: <https://jmlr.org/papers/v25/23-0870.html>.

МАЗМҰНЫ – СОДЕРЖАНИЕ – CONTENT

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖӘНЕ ҚҰҚЫҚ ҒЫЛЫМДАРЫ - СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ЮРИДИЧЕСКИЕ НАУКИ - SOCIO-ECONOMIC AND LEGAL SCIENCES

Төлегенов Тимур Сейфуллаулы

**ДИНАМИКА ТУРИСТСКИХ ПОТОКОВ И ЭФФЕКТИВНОСТЬ ИНФРАСТРУКТУРЫ
ТУРКЕСТАНА: ОТ ЭКСТЕНСИВНОГО РОСТА К СМАРТ-УПРАВЛЕНИЮ 10**

Alexander Gorevoy

**WHICH INEQUALITY STATISTICS SURVIVE AN ADMINISTRATIVE BOUNDARY CHANGE?
EVIDENCE FROM KAZAKHSTAN, 2010–2023 14**

Rakhimov Djalil

**DOES THE SHIFT TO CASHLESS PAYMENTS GO TOGETHER WITH FASTER GROWTH IN
HOUSEHOLD CREDIT? EVIDENCE FROM THE REGIONS OF KAZAKHSTAN, 2023-2025 26**

Kazhgerey Asan

**HOW MANY FOOD BASKETS DOES AN AVERAGE WAGE BUY? WAGES AND FOOD PRICES IN
THE 20 REGIONS OF KAZAKHSTAN, 2024-2026 37**

Tagimova Alua

**WHO WAS LEFT BEHIND? EDUCATION AND FINANCIAL INCLUSION IN KAZAKHSTAN AND
THREE CENTRAL ASIAN NEIGHBOURS, 2021-2024 44**

Wei Wei, Mynzhanova Gulzhakan, Niyetalina Gaukhar

**DIGITAL–PHYSICAL INTEGRATION AND ECONOMIC PERFORMANCE IN INDUSTRIAL
INTEGRATED ENERGY SYSTEMS: A SCENARIO-BASED ANALYSIS OF SMART GRID
CONTROL, BATTERY ENERGY STORAGE, AND DIGITAL TWINS 51**

Xu Fumin, Wei Wei, Ruziyeva Elvira, Moldabekova A.

**FROM CULTURAL RESOURCES TO GLOBAL CONSUMER VALUE: THE BRIDGE–IMC MODEL
AND THE INTERNATIONALIZATION OF BLACK MYTH: WUKONG UNDER THE BELT AND
ROAD INITIATIVE 64**

Ibraimkhan Alikhan

**WHICH CRIMES CAN A 14-YEAR-OLD BE TRIED FOR? THE LIST IN ARTICLE 15 OF
KAZAKHSTAN'S CRIMINAL CODE COMPARED WITH 30 OTHER JURISDICTIONS 80**

Yerboluly Madi

**DID IMPORTED CARS REALLY GET CHEAPER? CHINESE CARS AND THE AVERAGE
CUSTOMS VALUE OF PASSENGER CARS IMPORTED INTO KAZAKHSTAN, 2015-2025 91**

Салимбаев Ербол Аманжолович

**АНАЛИЗ ИНВЕСТИЦИОННОЙ АКТИВНОСТИ КАК ИНДИКАТОР ИЗМЕНЕНИЙ
ПОТЕНЦИАЛЬНОГО СПРОСА НА КОНСАЛТИНГОВЫЕ УСЛУГИ 99**

Rambergenov Ramil

**REGIONAL NEW-FIRM FORMATION IN UZBEKISTAN AFTER THE 2017 REFORMS:
PERSISTENCE IN SCALE, CONVERGENCE IN INTENSITY, AND RECOMPOSITION TOWARD
TRADE 107**

Abdesh Dinara

**FROM A WOMEN'S LEAD TO A MEN'S LEAD: THE GENDER GAP IN DIGITAL PAYMENTS IN
KAZAKHSTAN AND THREE CENTRAL ASIAN NEIGHBOURS, 2021-2024 116**

Amangeldi Aidana

**DID SMALL BUSINESS IN KAZAKHSTAN REALLY SHRINK IN 2026? THE NEW TAX CODE,
REGIONAL TAX RATES AND THE FALL IN THE NUMBER OF INDIVIDUAL ENTREPRENEURS 124**

Cheng Che	
BUILD, BUY, OR PARTNER? ENTRY MODE SELECTION AND RISK GOVERNANCE FOR WIND POWER INVESTMENT IN KAZAKHSTAN	133
Wang Liang	
FROM CAPACITY EXPANSION TO SYSTEM VALUE: COORDINATED DEVELOPMENT OF RENEWABLE ENERGY, GRID FLEXIBILITY AND ENERGY STORAGE IN KAZAKHSTAN	146
Chen Jun	
CONSTRUCTION-PERIOD CREDIT RISK MANAGEMENT FOR LARGE INFRASTRUCTURE PROJECT FINANCE IN KAZAKHSTAN	159
Alniyaz Danial	
HOW MUCH FUEL DOES A HYBRID REALLY SAVE? A SAME-MODEL COMPARISON OF HYBRID AND GASOLINE CARS IN US FUEL ECONOMY	173
Assem Yergali, Maral Assanova	
DIGITAL TRANSFORMATION AND SUSTAINABLE BUSINESS MODEL INNOVATION IN SMALL AND MEDIUM-SIZED ENTERPRISES: A SYSTEMATIC REVIEW OF DYNAMIC CAPABILITIES, VALUE CREATION, AND THE CREDIBILITY GAP IN EMERGING ECONOMIES	180
ТЕХНИКАЛЫҚ ҒЫЛЫМДАР - ТЕХНИЧЕСКИЕ НАУКИ - TECHNICAL SCIENCE	
Nurmakhan Almira	
ADVANTAGES OF FIBER OPTIC COMMUNICATION OVER TRADITIONAL COPPER CABLES	193
Kozhabai Bereke	
SECURITY CHALLENGES IN IOT-BASED TELECOMMUNICATION NETWORKS	201
Azamatulys Yerasyl	
WHY AI-GENERATED CODE FAILS A BENCHMARK: A FAILURE ANALYSIS OF THREE OPEN LANGUAGE MODELS ON 30 HUMANEVAL+ TASKS	210
Nurzhan Sanzhar	
A FEW SHELLS, A FEW OPERATORS: WHERE LOW EARTH ORBIT IS CROWDED IN AN ALTITUDE-INCLINATION MAP OF 16,620 ACTIVE SATELLITES	219
Askarbek Nurislam	
CAN THE WEATHER TELL WHICH DAYS IN ASTANA WILL BE POLLUTED? CLASSIFYING PM2.5 EXCEEDANCE DAYS WITH MACHINE LEARNING AND CHECKING THE DATA AGAINST A GROUND MONITOR	225
Ilyas Arman	
SHOOTING ON THE MOVE: HOW MUCH AIM A ROBOT LOSES BY IGNORING ITS OWN VELOCITY, AND WHETHER A VIRTUAL-TARGET CORRECTION RECOVERS IT	236
Ануар Әлімжан Бахытжанұлы, Сексенбаева Айша Курманбековна	
СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТАНДАРТОВ И МЕТОДИК ОБУЧЕНИЯ КИБЕРБЕЗОПАСНОСТИ: NICE FRAMEWORK, ECSF И ПРАКТИКО-ОРИЕНТИРОВАННЫЕ ТЕХНОЛОГИИ	242
Mukhamedzhan Yerganat	
GROWTH OF GITHUB DEVELOPER COMMUNITIES IN KAZAKHSTAN, UZBEKISTAN AND KYRGYZSTAN, 2020-2026: EVIDENCE FROM THE GITHUB INNOVATION GRAPH	252
Kashkinov Alden	
WHAT DECIDES HOW LONG A LANDER WORKS ON ANOTHER WORLD? SURFACE LIFETIMES OF 55 LANDERS ON THE MOON, VENUS, MARS, TITAN AND SMALL BODIES IN THE GCAT CATALOGUE, 1966-2025	264

Kenesbek Miras	
HOW MANY HERONIAN TRIANGLES HAVE A GIVEN LONGEST SIDE? A COMPLETE COUNT UP TO 2000 AND THE RIGHT TRIANGLES HIDDEN INSIDE	276
Amirlan Seitkadyrov, Akezhan Omashev	
AUTOMATIC MULTIPLE-CHOICE QUESTION GENERATION USING ARTIFICIAL INTELLIGENCE	287
Idrissov Amirzhan	
WHERE DO SMALL OPEN LANGUAGE MODELS BREAK ON OLYMPIAD MATHEMATICS? DOMAIN, DIFFICULTY AND MODEL SIZE ON OMNI-MATH	296
ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES	
Albiyev Asser	
HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025	301
Amangeldy Nurbolat	
HOW MUCH OF KAZAKHSTAN'S URBAN BUILT STOCK PREDATES ITS SEISMIC CODES, AND WHERE DOES THAT STOCK SIT RELATIVE TO HAZARD?	311
Torekhan Sabrina	
WHERE ALMATY WALKS ITS OWN WAY: INFORMAL SHORTCUTS IN OPENSTREETMAP AND THE DETOURS THEY AVOID	317
Ailan Salybekova	
PARENTAL SUPPORT, PARENTAL CONTROL, STRESS AND EARLY ALCOHOL USE IN ADOLESCENT GIRLS: A NARRATIVE REVIEW WITH A FOCUS ON KAZAKHSTAN	322
ӘЛЕУМЕТТІК ЖӘНЕ ГУМАНИТАРЛЫҚ ҒЫЛЫМДАР - СОЦИАЛЬНО-ГУМАНИТАРНЫЕ НАУКИ - SOCIAL SCIENCES AND HUMANITIES	
Мұқатаева Айзат Амантайқызы, Стамбекова Жанат Байғазықызы, Слямханова Гүлім Болатқызы, Қызырбек Аяулым Ерболқызы	
ЖАСАНДЫ ИНТЕЛЛЕКТ ДӘУІРІНДЕГІ ҒЫЛЫМИ ТАНЫМ: ПОППЕР, КУН ЖӘНЕ ЛАКАТОС ИДЕЯЛАРЫ ТҮРҒЫСЫНАН ТАЛДАУ	330
Мұқатаева Айзат Амантайқызы, Сапарханова Мадина Төлегенқызы, Окенова Аружан Муратхановна, Тұрсынғазыева Ақерке Ерболатқызы	
ШӘКӘРІМ ФИЛОСОФИЯСЫНДАҒЫ ЕРКІНДІК ЖӘНЕ МОРАЛЬДЫҚ ТАҢДАУ МӘСЕЛЕСІ	340
Komekova Saida	
DO UNHAPPY REVIEWERS MIX IN MORE RUSSIAN? STAR RATINGS AND KAZAKH-RUSSIAN CODE-SWITCHING IN 175,002 ONLINE REVIEWS	349
Zhumagali Yerbol	
DOES A BIGGER MODEL CLOSE THE GAP? KAZAKH-LANGUAGE ACCURACY ON KAZAKHSTAN-SPECIFIC VERSUS UNIVERSAL SCHOOL SUBJECTS IN SMALL OPEN-WEIGHT LANGUAGE MODELS	362
Abdrakhmanova Madina	
DOES A MEDAL BUY ATTENTION? WIKIPEDIA VIEWS OF KAZAKHSTAN'S ATHLETES BEFORE AND AFTER THE PARIS 2024 OLYMPIC	366

ПЕДАГОГИКАЛЫҚ ЖӘНЕ ПСИХОЛОГИЯЛЫҚ ҒЫЛЫМДАР – ПЕДАГОГИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ НАУКИ – PEDAGOGICAL AND PSYCHOLOGICAL SCIENCES

Kaldarbek Mergen

FAMILY OR SCHOOL? SOCIOECONOMIC COMPOSITION AND MATHEMATICS

ACHIEVEMENT IN KAZAKHSTAN, PISA 2022 AND 2025 374

Zhenis Adilzhan

WHICH OLYMPIAD PROBLEMS DO STUDENTS TAKE TO THE BOARD? CHOICES AND

DEFENCES IN THE 8TH-GRADE LEAGUE OF KARAGANDY MATHBATTLES 2026 381

Базарбаева Айнура Бектуровна

ДЕТСКИЙ ФОЛЬКЛОР И ПОДВИЖНЫЕ ИГРЫ КАК ИНСТРУМЕНТЫ ФОРМИРОВАНИЯ

ОСНОВ ПРАВОВОЙ КУЛЬТУРЫ У ДЕТЕЙ СТАРШЕГО ДОШКОЛЬНОГО ВОЗРАСТА 388

БИОЛОГИЯ ҒЫЛЫМДАР - БИОЛОГИЧЕСКИЕ НАУКИ - BIOLOGICAL SCIENCES

Bazhenova Diana

GROWTH DYNAMICS OF STERLET (*Acipenser ruthenus* L.) REARED IN A RECIRCULATING

AQUACULTURE SYSTEM 391

ТУРИСТИЧЕСКИЕ И СПОРТИВНЫЕ НАУКИ - ТУРИСТІК ЖӘНЕ СПОРТ ҒЫЛЫМДАРЫ - TOURISM AND SPORTS SCIENCES

Төлеп Жазира Маратқызы

ГАСТРОНОМИЯЛЫҚ ФЕСТИВАЛЬДЕРДІҢ ТУРИСТІК ДЕСТИНАЦИЯ ТАРТЫМДЫЛЫҒЫН

ҚАЛЫПТАСТЫРУДАҒЫ РӨЛІ МЕН ДАМУ ҮРДІСТЕРІ 398

ЕСТЕСТВЕННЫЕ НАУКИ - ЖАРАТЫЛЫСТАНУ ҒЫЛЫМДАРЫ - NATURAL SCIENCES

Нургельдина Асель Ермаковна, Серикбаева Жанерке Алимжановна

ШЕКАРАСЫЗ ФИЗИКА: ӘР ЕЛДЕГІ ФИЗИКАЛЫҚ БІЛІМДІ САЛЫСТЫРУ 405

Torekhan Sabrina

IS ALMATY A 15-MINUTE CITY? WALKING ACCESS TO SIX EVERYDAY PLACES FROM

170,838 BUILDINGS IN THE CITY'S EIGHT DISTRICTS 412