ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ
INTERNATIONAL UNIVERSITY IITU SINCE 2009
МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (24) 4

қазан- желтоқсан

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).
Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігіне мерзімді баспасөз басылымын есепке қою туралы куәлік № KZ82VPY00020475, 20.02.2020 ж. берілген
Тақырып бағыты: ақпараттық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология.
Мерзімділігі: жылына 4 рет.
Тираж: 100 дана.
Редакция мекенжайы: 050040 Алматы қ., Манас к., 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz
Журнал сайты: https://journal.iitu.edu.kz

Журнал сайты: https://journal.iitu.edu.kz

Международный журнал информационных и коммуникационных технологий

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**

Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.

E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 4. Number 24 (2025). Pp. 288–303
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.24.4.017

UDC 004.852

DEVELOPMENT OF A HYBRID HUMAN-AI SYSTEM FOR EDUCATIONAL PROGRAM DESIGN

A.E. Serikov[1*], A.A. Mukhatayev [1], A.A. Biloshchytskyi [2]

[1]Astana IT University, Astana, Kazakhstan;
[2]Kyiv National University of Construction and Architecture, Kyiv, Ukraine.
E-mail: ayanbek.as@gmail.com

Ayanbek Serikov — doctoral student, Astana IT University, Astana, Kazakhstan
E-mail: ayanbek.as@gmail.com, https://orcid.org/0009-0005-2413-5051;
Andrii Biloshchytskyi — Doctor of Technical Sciences, Kiev National University of Civil Engineering and Architecture, Kiev, Ukraine
https://orcid.org/0000-0001-9548-1959;
Aidos Mukhatayev — candidate of Political Sciences, associate professor, Director of the Department of Strategy and Corporate Governance, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0002-8667-3200.

Abstract. Artificial intelligence in educational program design automates learning processes and aligns curricula with labor market demands, enhancing education quality and learner experiences. We present a novel hybrid human-AI system to address limitations in traditional methods, which struggle to adapt to rapidly evolving job markets and diverse needs. Our approach leverages Llama 3 to improve framework flexibility. Key components include: 1) program objectives, 2) relevant job positions, 3) requisite skills, and 4) corresponding courses, refining recommendations. While large language models provide high-quality outputs, they lack deep university discipline knowledge. We evaluated Llama 3 alongside GPT-3.5 and GPT-4. Our system excels in skill extraction (F1 score: 77.6%) and outperforms competitors in content generation (F1: 0.35±0.10, precision: 0.37±0.12, semantic similarity: 0.84±0.10). Llama 3's extensive parameters and knowledge base better integrate university contexts, generating robust program structures with abundant educational entities. Despite promise, challenges persist: technological infrastructure deficits, dataset biases, and institutional resistance to AI adoption. This study aims to advance human-AI collaboration in education, inspiring further research.

Keywords: Hybrid Human-AI System, Educational Program Design, Skill Extraction, Artificial Intelligence in Education, Adaptive Learning, Natural Language Processing (NLP).

For citation: A.E. Serikov*, A.A. Mukhatayev, A.A. Biloshchytskyi. Development of a hybrid human-AI system for educational program design // International journal of information and communication technologies. 2025. Vol. 6. No. 24. Pp. 288–303. (In Eng.). https://doi.org/10.54309/IJICT.2025.24.4.017.

Conflict of interest: The authors declare that there is no conflict of interest.

БІЛІМ БЕРУ БАҒДАРЛАМАЛАРЫН ЖОБАЛАУҒА АРНАЛҒАН ГИБРИДТІ АДАМ-ЖАСАНДЫ ИНТЕЛЛЕКТ ЖҮЙЕСІН ДАМЫТУ

А.Е. Серіков [1*]., А.А. Мұхатаев[1], А.А. Билощицкий [2]

[1]Astana IT University, Астана, Қазақстан;

[2]Киев ұлттық құрылыс және сәулет университеті, Киев, Украина.

E-mail: ayanbek.as@gmail.com

Серіков Аянбек — докторант, Astana IT University, Астана, Қазақстан
E-mail: ayanbek.as@gmail.com, https://orcid.org/0009-0005-2413-5051;

Белощицкий Андрей — техника ғылымдарының докторы, Киев ұлттық құрылыс және сәулет университеті, Киев, Украина
https://orcid.org/0000-0001-9548-1959;

Мұхатаев Айдос — п.ғ.к., доцент, стратегия және корпоративтік басқару департаментінің директоры, Astana IT University, Астана қ., Қазақстан
https://orcid.org/0000-0002-8667-3200.

Аннотация. Білім беру бағдарламаларын жобалаудағы жасанды интеллект оқу процестерін автоматтандырады және оқу бағдарламаларын еңбек нарығының талаптарына сәйкестендіреді, білім беру сапасы мен білім алушылардың тәжірибесін жақсартады. Біз тез дамып келе жатқан еңбек нарықтары мен әртүрлі қажеттіліктерге бейімделу қиынға соғатын дәстүрлі әдістердегі шектеулерді жою үшін жаңа гибридті адам-жасанды интеллект жүйесін ұсынамыз. Біздің тәсіліміз құрылымдық икемділікті жақсарту үшін Llama 3-ті пайдаланады. Негізгі компоненттерге мыналар кіреді: 1) бағдарлама мақсаттары, 2) тиісті жұмыс орындары, 3) қажетті дағдылар және 4) ұсыныстарды нақтылайтын сәйкес курстар. Үлкен тілдік модельдер жоғары сапалы нәтижелер бергенімен, оларда университеттік пәндер бойынша терең білім жетіспейді. Біз Llama 3-ті GPT-3.5 және GPT-4-пен бірге бағаладық. Біздің жүйеміз дағдыларды игеруде (F1 ұпайы: 77,6%) және мазмұнды генерациялауда бәсекелестерден асып түседі (F1: 0,35±0,10, дәлдік: 0,37±0,12, семантикалық ұқсастық: 0,84±0,10). Llama 3-тің кең параметрлері мен білім базасы университеттік контексттерді жақсырақ біріктіреді, көптеген білім беру нысандары бар сенімді бағдарлама құрылымдарын жасайды. Уәдеге қарамастан, қиындықтар әлі де бар: технологиялық инфрақұрылымның тапшылығы, деректер жиынтығының бұрмалануы және жасанды интеллектті енгізуге институционалдық қарсылық. Бұл зерттеу білім берудегі адам мен жасанды интеллекттің ынтымақтастығын дамытуға бағытталған, әрі қарай зерттеулерге шабыт береді.

Түйін сөздер: Гибридті адам-ЖИ жүйесі, Білім беру бағдарламасының дизайны, Дағдыларды шығару, Білім берудегі жасанды интеллект, Бейімді оқыту, Табиғи тілді өңдеу (NLP)

Дәйексөздер үшін: А.Е. Серіков, А.А. Мұхатаев, А.А. Билощицкий. Білім беру бағдарламаларын жобалауға арналған гибридті адам-жасанды интеллект жүйесін дамыту // Халықаралық ақпараттық және коммуникациялық техноло-гиялар журналы. 2025. Том. 6. № 24. 288–303 бет. (Ағыл). https://doi.org/10.54309/IJICT.2025.24.4.017.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

РАЗРАБОТКА ГИБРИДНОЙ СИСТЕМЫ ЧЕЛОВЕКА И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РАЗРАБОТКИ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ

А.Е. Сериков [1*], А.А. Мухатаев [1], А.А. Билощицкий [2]

[1]Астана ИТ университет, Астана, Казахстан;

[2]Киевский национальный университет строительства и архитектуры, Киев, Украина.

E-mail: ayanbek.as@gmail.com

Сериков Аянбек Ермекулы — докторант, Астана ИТ университет, Астана, Казахстан

E-mail: ayanbek.as@gmail.com, https://orcid.org/0009-0005-2413-5051;

Белощицкий Андрей Александрович — доктор технических наук, Киевский национальный университет строительства и архитектуры, Киев, Украина

https://orcid.org/0000-0001-9548-1959;

Мухатаев Айдос Агдарбекович — кандидат политических наук, доцент, директор департамента стратегии и корпоративного управления, Астана ИТ университет, Астана, Казахстан

https://orcid.org/0000-0002-8667-3200.

Аннотация. Искусственный интеллект в разработке образовательных программ автоматизирует процессы обучения и согласует учебные программы с требованиями рынка труда, повышая качество образования и опыт учащихся. Мы представляем новую гибридную систему человек-ИИ для устранения ограничений традиционных методов, которые с трудом адаптируются к быстро меняющимся рынкам труда и разнообразным потребностям. Наш подход использует Llama 3 для повышения гибкости структуры. Ключевые компоненты включают в себя: 1) цели программы, 2) соответствующие должности, 3) требуемые навыки и 4) соответствующие курсы, уточняющие рекомендации. Хотя большие языковые модели обеспечивают высококачественные результаты, им не хватает глубоких знаний университетских дисциплин. Мы оценивали Llama 3 вместе с GPT-3.5 и GPT-4. Наша система превосходит конкурентов в извлечении навыков (оценка F1: 77,6 %) и превосходит конкурентов в генерации контента (F1: 0,35 ± 0,10, точность: 0,37 ± 0,12, семантическое сходство: 0,84 ± 0,10). Обширные параметры и база знаний Llama 3 лучше интегрируются с университетским контекстом, создавая надежные структуры программ с большим количеством образовательных объектов. Несмотря на обещания, проблемы сохраняются: дефицит технологической инфраструктуры, предвзятость наборов данных и институциональное сопротивление внедрению ИИ. Данное исследование направлено на развитие взаимодействия человека и ИИ в образовании, вдохновляя на дальнейшие исследования.

Ключевые слова: гибридная система человек-ИИ, разработка образовательных программ, извлечение навыков, искусственный интеллект в образовании, адаптивное обучение, обработка естественного языка (NLP)

Для цитирования: А.Е. Серіков, А.А. Мұхатаев, А.А. Билощицкий. Разработка гибридной системы человека и искусственного интеллекта для разработки образовательных программ//Международный журнал информационных и коммуникационных технологий. 2025. Т. 6. No. 24. Стр. 288–303. (На анг.). https://doi.org/10.54309/IJICT.2025.24.4.017.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction

Traditionally, educational programs (EP) must tell students what they will be taught, based on a curriculum designed by teachers. As a rule, the course author starts working on these issues by searching for similar or related programs in the university database. Even if he already has a clear structure of the future course in his head, he will still try to brainstorm to make his idea better. And this can take up to a week. But there are many different artificial intelligence-driven solutions people began to use for EP design, and they yield flexible, adaptive personalization based on student's and teacher's behavior (Chang et al., 2025: 50). Another reason people prefer AI is student satisfaction and motivation despite challenges in content delivery (Mancin et al., 2025: e485–e493). That doesn't mean technology with AI and information technology will overlap education entirely, but it would provide a bridge for collaboration.

To assume that AI won't have an influence on education would be foolish. Educational institutions prefer AI not just because it is used for educational content generation, but it is a more personalized and optimized learning approach. For example, this study writes a good analysis about several AI-based educational program platforms, such as eDoer, NextEra, Edmentum, XuetangX, and Manaba LMS, as you can see in Table 1.

This will come as a surprise to a lot of people, but eDoer makes it possible to bond technological and pedagogical aspects. Which is exciting because, among skill taxonomies and competency construction, it means teachers and interested parties can use these machine learning algorithms to design educational programs despite knowledge constraints. You can use this eDoer system to break knowledge into measurable skills development rather than deeper understanding and ethics, as stated in the term "learnification" by philosopher Gert Biesta (Biesta, 2015: 75–87). Standardized knowledge is prioritized, at the expense of alternative viewpoints, raising questions about the validity (Cope & Kalantzis, 2016: 2332858416641907). And over in Egypt, NextEra Education startup supply their students with free AI teachers. For example, many international universities perform personalized learning in some programs like IT, AI, Business and Cybersecurity to modernize education.

The somewhat more surprising educational platforms were ITMO Constructor and Edmentum, with their divergent philosophical approaches. The average ontological model of ITMO Constructor has a pretty much centered around semantic web technologies to fit education domains in certain knowledge structures, mostly on traditional knowledge (Shnaider et al., 2023: 1761–1768; Govorov et al., 2022: 203–208). But they shouldn't rigidly shape knowledge into structure, considering its flexibility and dynamic nature (Jonassen, 1991: 5–14). Now in Edmentum, programs employ assessments and Bayesian knowledge tracing. Which means measurement and data are prioritized over holistic learning (Corbett & Anderson, 1994: 253–278). They are both technologically advanced platforms, yet face limited knowledge acquirement, ignoring diverse learning environments and real-world context.

XuetangX and Manaba LMS already seem, grudgingly, to show key sociotechnical challenges in educational program design. But XuetangX is still dragging serious questions about the lack of transparency in content creation. It means XuetangX's black-box architecture is harder for teachers and educators to assess accuracy and bias (Bender et al., 2021: 610–623). While Manaba LMS combines AI with human input, yet recommendations are based on pre-built assumptions (Macgilchrist, 2019: 77–86). They'd be in a better position if they'd be interdisciplinary collaboration between experts in tech, education and ethics to make sure that AI serves real learning needs rather than efficiency and commercial needs.

Table 1. Global use of AI technologies in EP development

Country	Germany	Egypt	USA	China
Platform	eDoer	NextEra Education	Edmentum	XuetangX
Purpose & Scope	Structured program creation (university focus)	Modernizes Egypt's education with AI personalized learning	AI-driven curriculum & adaptive learning (K-12 & higher ed)	AI-powered university course structuring
Use of Structured Data	Limited; relies on pre-defined structures	Structured & unstructured data + feedback	Processes student behavior, performance trends, feedback	Processes student behavior & engagement data
AI Capability	AI-assisted curriculum planning	Adaptive learning, progress tracking	AI-driven adaptive learning paths & performance tracking	Advanced AI course curation & perform. prediction.
Integration with Other Systems	Moodle, LMS platforms.	Egypt's system (no LMS integration)	Blackboard, Canvas, Google Classroom	Fully integrated with Chinese university systems

AI has fundamentally been supporting educational programs advancement through machine learning and text mining techniques. Instep of traditional methods, Chang et al. (2025) shows how curricula are adjusted based on student performance and preferences in real-time using AI. Generative AI (GenAI) and Natural Language Processing models uphold content creation, lesson planning, flexibility to different learning styles and equity in content difficulty to student's respective proficiency levels (Msafiri et al., 2023: 44). That never works unless you have no framework limitations and educators' endorsement of the new technologies adoption.

In the human-AI systems there is a progressing battle between computerization and human oversight (Cross et al., 2014: 1167–1175). Everyone knows how crowdsourcing platforms enable collaborative curriculum design, enriched by assorted human input (Weld et al., 2012: 159–163). The expanded AI combines automation of quality evaluations, keyword extraction and human imagination. For case, the crowdsourced models produce curriculum and learning materials based on student's input, what appears to be promising in adoption potential (Farasat et al., 2017: 221–224). We can see this promise with task complexity and time challenges require AI's part to streamline tasks without losing educational esteem.

Suppose, for illustration, you require to design curriculum, and for that you can use emerging GenAI technologies. For example, generative adversarial networks could be used to generate relevant content (Khosravi et al., 2022: 100074). Once you begin considering this question, studies show that in this case ChatGPT can draft learning objectives, subject matter or human experts are still needed to refine them for accuracy and pedagogical standards (Tupper et al., 2025: 512–526). Balance between technology and humans would ensure both quality and ethics in education.

We are aware that AI has both societal and technological boundaries. Tech typically could make education accessible, algorithms could be biased. What I mean is the potential scrutiny, data raises privacy risks and stronger regulations. Yes, tools play a role in spreading educational resources, and then they may accommodate learner needs, yet poor-quality metadata commonly suppress personalized features (Molavi et al., 2020: 455–460). Whether educators like it or not, big changes are coming, to balance these human, industry needs and skill taxonomies, we propose a curriculum design framework.

Materials and methods of the research

We present one way to develop an educational program by the chosen learning goals and skills. Our framework is (1) to define the learning goals of the educational program, then (2) select relevant job positions from the labour market, and then (3) once vacancy descriptions are parsed, collect all associated skills, and then (4) choose courses based on required skills. Educational program design is hard in both senses: hard like coming up with a solution to a puzzle, and hard like a physically overwhelming process. But much of the difficulty could be eliminated by AI-based specialized services, such as learning goals, learning outcomes and course suggestion and recommendation service, that give a review on skill taxonomies and competencies from Kazakhstani labour market platforms (HeadHunter and Enbek.kz). The resulting four parts define our conceptual model shown in Figure 1.

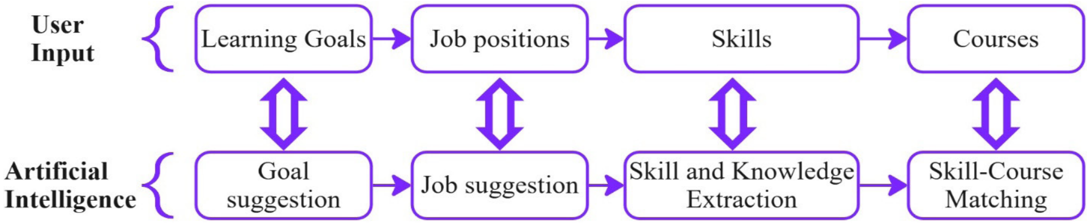

Fig.1. EP development framework

The solution is an optimization problem. For each piece of the educational puzzle e_i, including learning goal, learning outcome and course content, we figure out a relevance score $R(e_i)$:

$$R(e_i) = \gamma_1 S_{job}(e_i) + \gamma_2 S_{coh}(e_i) + \gamma_3 S_{human}(e_i) + \gamma_4 S_{bloom}(e_i) \tag{1}$$

where:

- $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ are weights that add up to 1.
- $S_{job}(e_i)$ is how well e_i matches job market needs, using a similarity score between its embeddings:

$$S_{job}(e_i) = \cos\left(E(e_i), \frac{1}{|J|}\sum_{j \in J} E(j)\right) \tag{2}$$

- $S_{coh}(e_i)$ is a coherence:

$$S_{coh}(e_i) = \frac{1}{|E|-1} \sum_{e_j \in E, j \neq i} \cos\left(E(e_i), E(e_j)\right) \quad (3)$$

- $S_{human}(e_i)$ is expert feedback:

$$S_{human}(e_i) = \lambda \cdot S_{initial}(e_i) + (1-\lambda) \cdot S_{feedback}(e_i) \quad (4)$$

- $S_{bloom}(e_i)$ is how well it aligns with Bloom's taxonomy, using through a classification system.

What these formulas would like to do, if the criteria are met, is adjust educa- $R(e_i)$ tional content determined by AI and human feedback, with enough refining for recommendations convergence.

We used (1) PyTorch to handle tensor calculations and GPU power, then (2) Unsloth is used to train the model 30x faster and with 90% less memory usage, then (3) huggingface/transformers and huggingface/datasets prepared models and datasets, then (4) also used Transformer Reinforcement Learning to fine-tune with super-vision (SFT) and last (5) Bitsandbytes and PEFT to keep efficient.

Training Process:

We used a 4-bit version of standard Llama 3, tweaked out with LoRA adapters. As seen from the Figure 2, the training ran for 200 steps, with warm up steps at

1) Data Processing:

The dataset used for training was from the Hugging Face SkillSpan dataset and custom builded EP dataset, which were manually downloaded from Kazakhstani universities (Astana IT University, East Kazakhstan University named after S. Amanzholov, Yessenov University, Altynsarin University, Academy of Physical Culture and Mass Sports) websites. The data was formatted using an Alpaca-style instruction template and saved in Fralet/skillspan_prepared with 4800 train, 3174 validation, and 3569 test datasets, while the EP dataset has 944 unique rows, saved in Fralet/eduEP on 472 educational programs. This paper's skill extraction focuses on the jjzha/skillspan dataset, which includes diverse job postings, which has explicit and implicit skills. The dataset is split into training, validation, and test. You can see instructions in Figure 3 and Figure 4.

```
trainer = SFTTrainer(
    model=model,
    tokenizer=tokenizer,
    train_dataset=dataset,
    dataset_text_field="text",
    max_seq_length=max_seq_length,
    dataset_num_proc=2,
    packing=False,
    args=TrainingArguments(
        per_device_train_batch_size=2,
        gradient_accumulation_steps=4,
        warmup_steps=5,
        max_steps=200,
        num_train_epochs=4,
        learning_rate=2e-4,
        fp16=not torch.cuda.is_bf16_supported(),
        bf16=torch.cuda.is_bf16_supported(),
        logging_steps=1,
        optim="adamw_8bit",
        weight_decay=0.01,
        lr_scheduler_type="linear",
        seed=3407,
        output_dir="outputs",
    ),
)
```

Fig.2. Model's training parameters

instruction
string

Your job is to extract skills and knowledge from given text.

You are a helpful information extraction system. Your job is to extract skill entities and knowledge entities from the given sentence.

*Fig.3.*Prompts for Fralet/skillspan_prepared

instruction
string

```
Write an educational program goal for this educational program in
Russian, Kazakh, and English. The goal should be as concise and
specific as possible within the context of the professional
field.
```

```
Write 10-12 learning outcomes for this educational program in
Russian, Kazakh, and English. The outcomes should be formulated
based on Bloom's taxonomy: knowledge, comprehension, application,
analysis, synthesis, and evaluation.
```

Fig.4 – Prompts for Fralet/eduEP

1) **Model Training and Architecture:**

The architecture used a Llama model using LoRA for efficient fine-tuning, focusing attention mechanisms like q_{proj}, k_{proj} and v_{proj}, to pick up new tasks without changing much. We also used 4-bit quantization to save memory and boost performance.

2) **Experimental Configurations:**

We trained the model using the Hugging Face SFTTrainer (TRL library), with a batch size of 2 and with a gradient accumulation of 4 steps per batch. The AdamW optimizer (8-bit, LR=2e-4, weight decay=0.01) was used over 4 epochs, with a 2048-token sequence length.

Results

The progression of a Llama 3 model is over 200 steps. Train/loss demonstrates a steady rise in the training epochs, starting from approximately 0.02 and steadily going to 0.32. While train/epoch shows the loss from a peak of 3.0 to around 0.5 with some fluctuations. Fortunately, it proves effective optimization without significant overfitting or instability.

The left graph depicts the training loss over global steps, modelled as:

$$\mathcal{L}(t) = \mathcal{L}_0 \cdot e^{-\alpha t} + \mathcal{L}_{plateau} + \epsilon(t) \quad (5)$$

where:

- $\mathcal{L}(t)$ is t loss

- $\mathcal{L}_0 \approx 3.0$ (initial loss)

- α is a decay rate
- $\mathcal{L}_{plateau} \approx 0.5$ (stabilized loss)
- $\epsilon(t)$ represents a noise

While, right graph shows linear relationship between steps and epochs, expressed as:

$$E(t) = E_0 + \beta \cdot t \tag{6}$$

where:

- $E(t)$ is t epoch value
- $E_0 \approx 0$ (initial epoch value)
- $\beta \approx 0.0016$ calculated as $\frac{0.32}{200}$ (slope)

Our system digs into a fine-tuned sequence labeling approach. Given an input text sequence $X = \{x_1, x_2, \dots, x_n\}$ representing job descriptions or curriculum materials, we define a skill extraction $f_{(\theta)}: X \rightarrow Y$ where $Y = \{y_1, y_2, \dots, y_n\}$ and $y_i \in \{\text{B-SKILL,I-SKILL,B-KNOWLEDGE,I-KNOWLEDGE}, O\}$ following the BIO tagging scheme. The θ are optimized by minimizing the cross-entropy loss:

$$\mathcal{L}_{CE}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{c=1}^{C} y_{i,j,c}\log\left(\widehat{y_{i,j,c}}\right) \tag{7}$$

where N is the No of training examples, n is the sequence length, C is the number of classes, $y_{i,j,c}$ is a binary label where $y_{i,j,c} = 1$ if class c is the correct for token j in example i, and 0 otherwise, and $\widehat{y_{i,j,c}}$ is the predicted probability. Performance is evaluated using entity-level F1 score:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{8}$$

where TP, FP and FN are true positives, false positives, and false negatives.

Table 2. Model Performance on SkillSpan test dataset (Skills & Knowledge)

Method	Parameters	Skill Span F1	Knowledge Span F1	Combined F1	Training
BERT (Zhang et al., 2022: 4962–4984)	110M	54.2 %	61.7 %	57.7 %	fine tuning
jobSpanBERT (Zhang et al., 2022: 4962-4984)	110M	56.3 %	61.9 %	58.9 %	pretrain + fine tuning
GPT-3.5 NER-Style (Nguyen et al., 2024: 27-42)	175B	-	-	17.8 %	5-shot
GPT-3.5 Extract-Style (Nguyen et al., 2024: 27-42)	175B	-	-	25.0 %	5-shot
GPT-4 (Nguyen et al., 2024: 27-42)	1.7T	-	-	27.8 %	≤ 350 subset
Finetuned Llama 3 (Ours)	8B	76.2 %	79 %	77.6 %	fine tuning

From Table 2, if you had to compare our model on skill extraction, it achieved a solid 77.6 % F1 score, with 76.2% for skills and 79% for knowledge. The impressive thing is, having fewer parameters our fine-tuned model that can be comparable to other larger models. I would like not to ignore errors if output is not labelled correctly. It could be due to the dataset's limitations and inconsistencies caused by human error.

After skill extraction, we use technique to generate educational goals and outcomes called sequence-to sequence framework. In the program description $D = \{d_1, d_2, \dots, d_m\}$ and optional context information $C = \{c_1, c_2, \dots, c_k\}$ (such as industry standards or institutional requirements), we define a generation function

$$g_\phi : (D, C) \to G \tag{9}$$

where $(G = \{g_1, g_2, \dots, g_p\})$ represents the generated learning goals and outcomes. The parameter ϕ are optimized using a mix of maximum likelihood estimation and a semantic similarity:

$$\mathcal{L}_{total}(\phi) = \alpha \mathcal{L}_{MLE}(\phi) + \beta \mathcal{L}_{sem}(\phi) \tag{10}$$

where: $\mathcal{L}_{MLE}(\phi) = -\sum_{t=1}^{p} \log P_\phi(g_t | g_{<t}, D, C)$ and $\mathcal{L}_{sem}(\phi) = 1 - \cos(E(G), E(G^*))$

What we're seeing now, α and β are weighting hyperparameters, P_ϕ is the model's output distribution, E maps text to a semantic space, and G^* represents reference goal statements. This method guarantees both natural language and semantic relevance to recognized educational level.

From Table 3, our Llama 3 model ends up outshining other models in generating content. It has the highest F_1 of 0.35 ± 0.10, surpassing GPT-4 (0.31 ± 0.10) and base Llama 3 (0.25 ± 0.10). The precision of 0.37 ± 0.12 and recall of 0.33 ± 0.10 show casing relevant data while capturing a broader range of outcomes. Not always, but in specifically recall, it is subdued by GPT-4 (0.36 ± 0.10). With a BLEU of 0.25 ± 0.10 and a semantic similarity of 0.84 ± 0.10, the model generates text that is aligned in meaning and relatively close in structure.

Table 3. Performance results for the different models

Candidate	F_1	Precision	Recall	Bleu	Semantic Similarity
GPT-3.5	0.16 ± 0.10	0.25 ± 0.12	0.08 ± 0.10	0.11 ± 0.10	0.63 ± 0.10
GPT-4	0.31 ± 0.10	0.26 ± 0.12	0.36 ± 0.10	0.23 ± 0.10	0.49 ± 0.10
Llama 3	0.25 ± 0.10	0.31 ± 0.12	0.20 ± 0.10	0.18 ± 0.10	0.77 ± 0.10
Finetuned Llama 3 (Ours)	0.35 ± 0.10	0.37 ± 0.12	0.33 ± 0.10	0.25 ± 0.10	0.84 ± 0.10

Of course, GPT-3.5 struggles with a low F_1 of 0.16 ± 0.10, precision of 0.25 ± 0.12, and a recall of just 0.08 ± 0.10. It shows a BLEU of 0.11 ± 0.10 and a semantic similarity of 0.63 ± 0.10. More likely the reason is that GPT-3.5 struggles in generating needed content. GPT-4 does better, with a higher recall (0.36 ± 0.10) and BLEU score (0.23 ± 0.10), but its precision (0.26 ± 0.12) and semantic similarity (0.49 ± 0.10) are still lagging behind our model. The standard Llama 3 shows a solid precision of 0.31 ± 0.12 and recall of 0.20 ± 0.10, but its semantic similarity of 0.77 ± 0.10 and BLEU score of 0.18 ± 0.10 indicate it does not perform well enough.

While I'm sure the results bring hope for AI in education, some caveats exist. So what's the real reason for these caveats? Curiously enough, it is the reason of lack of broader dataset with various educational areas. Also keep in mind that our results are based on limited data from only 6 Kazakhstani universities as the ground truth and it will be hard to scale to broader applications. But still, the issue is in the dataset, not the way we evaluate or calculate. Our model performed well in that respect, proving they can provide content that hits the standard.

Discussion

The exciting thing is that tuned Llama 3 produces relevant learning goals and outcomes. This is colossal, since it does surpass at different assessment metrics, outflanking GPT-3.5 and GPT-4. For case, our model's F_1 of 0.35 ± 0.10 outperformed the standard Llama 3 (0.25 ± 0.10) and GPT-4 (0.31 ± 0.10). And so, it is in this case, discoveries demonstrate self-evident that fitting the model to specific instructions makes clear and relevant results. Also, semantic similarity of 0.84 ± 0.10 implies the outcome aligns with the referenced content and there is a near relationship between texts. ChatGPT-4 performed well - worse than standard Llama 3 in semantic similarity, but way better than in other measurements and ChatGPT-3.5. So, we can recommend it for budget information sharing in universities and colleges. Models with less parameters produced more shallow results and coped worse. GPT-3.5 often "got lost" when working with the educational program design context. It requires more precise fine-tuning to successfully work in such assignments.

The first configurations I tried were directed to supervised fine-tuning. And yet as it gets adapted to subtleties of EP design, more values of evaluation metrics rise. It was affirmation of its prevalent execution, that it was able to (1) recognize semantic similarity, (2) identify different cognitive skills and (3) distinguish Bloom's taxonomy. This model is not designed just for academics, it has real-world operations. In human resources it could be used to smooth hiring by extracting skills from resumes. This human-AI model can boost decision-making in all sorts of areas, from streamlining corporate workflows to configuring AI super assistants from predefined skill sets.

The other thing we see are limitations. The less sure you are about the dataset's domain areas, the more problems you will collect in generalizability. There's still a need for further research with more diverse datasets and varying real-world contexts. Generally, most limitations are tied to the dataset rather than evaluation or training techniques. There is a lot of future research on exploring the effect of new data sources and nailing configurations and parameters. Most such work would help better understand potential in designing educational settings. And thus the root of great work is in gaining a deeper understanding of the scope of the study.

Conclusion

To sum up, this work pushes the frontier of hybrid human-AI approaches for educational program design, solving old problems of conventional methods, which can't pace to fast evolving world of job market and student needs. Thanks to integration of our fine-tuned Llama 3 with human experience, our structure provides flexible and data-informed decisions on educational program design, including program purpose, job vacancies and their corresponding skill sets. Model's skill extraction performance is evidenced by F1 score of 77.6 % (76.2 % for skills and 79 % for knowledge) and for text generation by F1: 0.35 ± 0.10, precision: 0.37 ± 0.12, recall: 0.33 ± 0.10, BLEU: 0.25 ± 0.10, semantic similarity: 0.84 ± 0.10, that surpasses popular models like ChatGPT 3.5 and 4.0, on SkillSpan and custom-collected Kazakhstani university based datasets. These underscores fine-tuned Llama 3 model's parameters and deep knowledge base in educational environment.

The result of this work is not only academia but fostering collaboration between human and AI. In education sector, AI typically used to automate routine tasks, like grading assessment, attendance tracking and aligning courses, which in turn helps educators and professors to concentrate on creative and ethical improvements of the system. This is also observed in global trend with education and AI, where hybrid systems enhance human capabilities, instead of replacing it. For example, similar hybrid educational settings demonstrated promising opportunities for effective management, increasing student engagement and fostering inclusivity. In Kazakhstan, we have national initiatives as AI Development Concept 2024-2029, which make accent on countries that are developing various policies to strengthen human capital, including through the development of educational programs and courses and the creation of AI innovation centers. Some of the above-mentioned measures, such as the integration of AI into educational programs and funding and grants for AI, are applicable to the current situation in the country.

On other hand, challenges must be addressed to ensure sustainable integration. Need for technological infrastructure, especially regions with limited computerization access, shows scalability problematic. Dataset biases and subjective content from sources such as HeadHunter and Enbek.kz may lead to inequality, which can be solved only through diverse and inclusive data collection methods. Institutional hesitation, often fueled by fears of job loss and human control loss, makes adoption difficult. In Central Asia, ethical and legal concerns, especially with the data confidentiality and algorithmic biases, aggravate existing problems. More general global barriers as lack of funds and AI specialists parallel Kazakhstan's situation, where only a few research centers as ISSAI under Nazarbayev University lead research innovation. These limits underscores need for interdisciplinary collaboration between educators, technical specialists and ethical experts over mere efficiency.

Moving forward, future works ought to center on growing datasets to cover more differing educational sectors and global standards-based settings, subsequently improving generalizability. Research on progressive strategies, such as support learning or generative fine-tuned systems, can advance content generation whereas reducing bias. Integration of new services, such as AI tutors or virtual reality-based simulations, as shown in pilot projects demo in Kazakhstan, can grow the scope for adaptive learning. Policymakers in Kazakhstan and around the world ought to contribute to AI education programs, ethical rules, and foundation to encourage broad selection, steady with UNESCO's accentuation on the responsible use of AI in instruction. Eventually, this research not only advances the part of AI in education but motivates to use hybrid solutions to make imaginative learning environments that will plan understudies for an AI-powered future.

REFERENCES

Bender E. M. et al. (2021). On the dangers of stochastic parrots: Can language models be too big? // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. — 2021. — Pp. 610–623. DOI: 10.1145/3442188.3445922.

Biesta G. (2015). What is education for? On good education, teacher judgement, and educational professionalism // European Journal of Education. — 2015. — Vol. 50. — No. 1. Pp. 75–87. DOI: 10.1111/ejed.12109.

Chang S. H. et al. (2025). Integrating motivation theory into the AIED curriculum for technical education: Examining the impact on learning outcomes and the moderating role of computer self-efficacy // Information. — 2025. — Vol. 16. — No. 1. Article 50. DOI: 10.3390/info16010050.

Cope B., Kalantzis M. (2016). Big data comes to school: Implications for learning, assessment, and research

// AERA Open. — 2016. — Vol. 2. — No. 2. Article 2332858416641907. DOI: 10.1177/2332858416641907.

Corbett A. T., Anderson J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge // User Modeling and User-Adapted Interaction. — 1994. — Vol. 4. — No. 4. Pp. 253–278. DOI: 10.1007/BF01099821.

Cross A. et al. (2014). VidWiki: Enabling the crowd to improve the legibility of online educational videos // Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. — 2014. — Pp. 1167–1175. DOI: 10.1145/2531602.2531670.

Farasat A. et al. (2017). Crowdlearning: Towards collaborative problem-posing at scale // Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale. — 2017. — Pp. 221–224. DOI: 10.1145/3051457.3053990.

Govorov A., Chernysheva A., Khlopotov M., Derkunskaia S., Arzumanian A. (2022). Target professions based approach for individual learning pathway creation // International Journal of Information and Education Technology. — 2022. — Vol. 12. — No. 3. Article 1605. DOI: 10.18178/IJIET.2022.12.3.1605.

Jonassen D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? // Educational Technology Research and Development. — 1991. — Vol. 39. — No. 3. Pp. 5–14. DOI: 10.1007/BF02296434.

Khosravi H. et al. (2022). Explainable artificial intelligence in education // Computers and Education: Artificial Intelligence. — 2022. — Vol. 3. — No. 1. Article 100074. DOI: 10.1016/j.caeai.2022.100074.

Macgilchrist F. (2019). Cruel optimism in edtech: When the digital data practices of educational technology providers inadvertently hinder educational equity // Learning, Media and Technology. — 2019. — Vol. 44. — No. 1. Pp. 77–86. DOI: 10.1080/17439884.2018.1556217.

Mancin S. et al. (2025). Efficacy of active teaching methods for distance learning in undergraduate nursing education: A systematic review // Teaching and Learning in Nursing. — 2025. — Vol. 20. — No. 2. Pp. e485–e493. DOI: 10.1016/j.teln.2024.12.015.

Molavi M., Tavakoli M., Kismihók G. (2020). Extracting topics from open educational resources // European Conference on Technology Enhanced Learning. — 2020. — Pp. 455–460. DOI: 10.1007/978-3-030-57717-9_44.

Msafiri M. M., Kangwa D., Cai L. (2023). A systematic literature review of ICT integration in secondary education: what works, what does not, and what next? // Discover Education. — 2023. — Vol. 2. — No. 1. Article 44. DOI: 10.1007/s44217-023-00070-x.

Nguyen K. C. et al. (2024). 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). — 2024. — Pp. 27–42. DOI: 10.48550/arXiv.2402.03832.

Shnaider P., Chernysheva A., Govorov A., Khlopotov M. (2023). Composing a syllabus via Educational Program Maker: A web tool for academic documents creation // International Journal of Information and Education Technology. — 2023. — Vol. 13. — No. 11. Pp. 1761–1768. DOI: 10.18178/ijiet.2023.13.11.1987.

Tupper M., Hendy I. W., Shipway J. R. (2025). Field courses for dummies: To what extent can ChatGPT design a higher education field course? // Innovations in Education and Teaching International. — 2025. — Vol. 62. — No. 2. Pp. 512–526. DOI: 10.1080/14703297.2024.2316716.

Weld D. S. et al. (2012). Personalized online education — A crowdsourcing challenge // Proceedings of the AAAI Workshop on Human Computation (HCOMP). — 2012. — Vol. 12. — No. 8. Pp. 159–163. DOI: 10.1145/2531602.2531670.

Zhang M. et al. (2022). SkillSpan: Hard and soft skill extraction from English job postings // Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. — 2022. — Pp. 4962–4984. DOI: 10.18653/v1/2022.naacl-main.366.

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

Подписано в печать 15.12.2025.
Формат 60х881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59