

DOI 10.54596/2958-0048-2025-2-220-230

УДК 004.89

МРНТИ 16.21.00

CHATGPT, DEEPSEEK, GROK СИЯҚТЫ LLM-ДІ СТУДЕНТТЕРДІҢ ЖҰМЫСТАРЫН БАҒАЛАУҒА ҚОЛДАНУ

Мунтинов К.Д.^{1*}

^{1*} «Манаш Қозыбаев атындағы Солтүстік Қазақстан университеті» КеАҚ
Петропавл, Қазақстан

*Хат-хабар үшін автор: Kairatmuntinov@gmail.com

Аңдатпа

Бұл мақалада ChatGPT, Deep Seek және Grok сияқты үлкен тілдік модельдерді (LM) студенттердің жұмысын бағалау тапсырмаларында қолдану мүмкіндіктері қарастырылады. Автор автоматтандырылған тәсілдің күшті және әлсіз жақтарын анықтауға баса назар аудара отырып, оқытушылық бағалаумен салыстырғанда ChatGPT арқылы алынған нәтижелерге сапалы талдау жасайды. LLM пайдаланудың ықтимал артықшылықтары – өңдеу жылдамдығы, сәйкестік критерийлері, масштабталу – және шығармашылық пен талдау тереңдігін бағалауға байланысты шектеулер талқыланады. Тапсырма түріне (мәтін, код) және пәннің ерекшелігіне байланысты әртүрлі модельдердің қолданылуына ерекше назар аударылады. Жұмыс шолу-аналитикалық сипатқа ие және білім беруді бағалауды цифрландыру және LLM-ді оқу процесіне біріктіру саласындағы әрі қарайғы зерттеулердің бастапқы нүктесі бола алады.

Кілт сөздер: LLM, ChatGPT, бағалау, студенттік жұмыстар, жасанды интеллект, білім беру, автоматтандыру.

ПРИМЕНЕНИЕ LLM НА ПРИМЕРЕ CHATGPT, DEEPSEEK, GROK ДЛЯ ОЦЕНИВАНИЯ РАБОТ СТУДЕНТОВ

Мунтинов К.Д.^{1*}

^{1*} НАО «Северо-Казахстанский университет им. М. Козыбаева»
Петропавловск, Казахстан

*Автор для корреспонденции: Kairatmuntinov@gmail.com

Аннотация

В статье рассматриваются возможности применения больших языковых моделей (LLM), таких как ChatGPT, DeepSeek и Grok, в задачах оценивания студенческих работ. Автор проводит качественный анализ результатов, полученных с помощью ChatGPT, в сравнении с преподавательскими оценками, с акцентом на выявление сильных и слабых сторон автоматизированного подхода. Обсуждаются потенциальные преимущества использования LLM – скорость обработки, соблюдение критериев, масштабируемость – а также ограничения, связанные с оценкой креативности и глубины анализа. Отдельное внимание уделено применимости различных моделей в зависимости от типа задания (текст, код) и специфики дисциплины. Работа носит обзорно-аналитический характер и может послужить отправной точкой для дальнейших исследований в области цифровизации образовательной оценки и интеграции LLM в учебный процесс.

Ключевые слова: LLM, ChatGPT, оценивание, студенческие работы, искусственный интеллект, образование, автоматизация.

APPLYING LLMS LIKE CHATGPT, DEEPSEEK, GROK FOR STUDENT WORK EVALUATION

Muntinov Kairat^{1*}

^{1*}*Manash Kozybayev North Kazakhstan University NPLC, Petropavlovsk, Kazakhstan*

**Corresponding author: Kairatmuntinov@gmail.com*

Abstract

The article discusses the possibilities of using large language models (LM), such as ChatGPT, Deep Seek and Grok, in the tasks of evaluating student papers. The author conducts a qualitative analysis of the results obtained using ChatGPT, in comparison with teaching assessments, with an emphasis on identifying the strengths and weaknesses of the automated approach. The potential advantages of using LLM are discussed – processing speed, compliance with criteria, scalability – as well as limitations associated with evaluating creativity and depth of analysis. Special attention is paid to the applicability of various models depending on the type of assignment (text, code) and the specifics of the discipline. The work is of a review and analytical nature and can serve as a starting point for further research in the field of digitalization of educational assessment and integration of LLM into the educational process.

Keywords: LLM, ChatGPT, evaluation, student assignments, artificial intelligence, education, automation.

Кіріспе

Қазіргі заманғы білім беру тиімді және объективті бағалауды қажет ететін Оқу материалдары мен міндеттерінің өсіп келе жатқан көлеміне тап болады. Жұмысты қолмен тексеруге негізделген дәстүрлі бағалау әдістері көбінесе көп еңбекті қажет етеді, субъективті және студенттерге уақтылы және жекелендірілген кері байланыс беруге мүмкіндік бермейді. Осыған байланысты бағалау процесін оңтайландыруға және оның сапасын арттыруға қабілетті инновациялық тәсілдерге қажеттілік туындайды.

Жасанды интеллектке негізделген автоматтандырылған жүйелер тексеру уақытын айтарлықтай қысқартуы мүмкін, бірақ контекстті есепке алу үшін мұқият калибрлеуді қажет етеді [1; 3].

Соңғы жылдары ChatGPT, DeepSeek және Grok сияқты үлкен тілдік модельдердің (LLM) қарқынды дамуы білім беру тәжірибесін автоматтандыру мен жетілдірудің жаңа перспективаларын ашты. Бұл модельдер мәтінді түсіну, талдау және генерациялаудың әсерлі мүмкіндіктерін көрсетеді, бұл оларды әртүрлі типтегі студенттердің жұмысын бағалаудың пайдалы құралы етеді. Бұл зерттеудің мақсаты студенттердің жұмысын бағалау үшін ChatGPT, DeepSeek және Grok ұсынған LLM пайдалану мүмкіндіктері мен шектеулерін зерттеу болып табылады. LM мәтіндердің құрылымы мен стилі сияқты формальды аспектілерін тиімді талдай алады, бірақ оларды талдаудың креативтілігі мен тереңдігін бағалауда қолдану адамның бақылауын қажет етеді [2; 45].

Мақсатқа жету үшін келесі міндеттерді шешу қажет:

- Студенттік жұмыстарды бағалаудың қолданыстағы тәсілдерін талдау және олардың негізгі кемшіліктерін анықтау;
- LLM жұмыс принциптерін қарастырыңыз және олардың мәтіндерді бағалау міндеттеріне қолданылуын бағалаңыз;
- Студенттік жұмыстардың әртүрлі түрлерін (эсселер, есептер, бағдарламалық код және т. б.) автоматты түрде бағалау үшін ChatGPT, DeepSeek және Grok қолдану мүмкіндіктерін зерттеу;
- Бағалау процесінде, атап айтқанда объективтілік, тиімділік және кері байланысты жекелендіру контекстінде LLM пайдаланудың артықшылықтары мен кемшіліктерін бағалау;

- Білім беру процесіне LLM енгізуге байланысты этикалық және практикалық мәселелерді анықтаңыз.

Зерттеудің өзектілігі қазіргі білім беру жүйесінің өсіп келе жатқан қажеттіліктерін қанағаттандыра алатын тиімді және масштабталатын бағалау әдістерін іздеу қажеттілігіне байланысты. LLM-ді бағалау процесіне енгізу оқытушылардың жұмысты тексеруге кететін уақытын едәуір қысқартуы және студенттерге уақтылы және жекелендірілген кері байланыс беруі мүмкін, бұл өз кезегінде оқу сапасын жақсартуға ықпал етеді. CHATGPT сияқты LLM білім беру процестерін түрлендірудің әлеуетін көрсетеді, бірақ олардың күрделі контексттерді түсіндірудегі шектеулері қосымша зерттеуді қажет етеді [3; 12].

LLM көмегімен жұмысты бағалау туралы екі гипотеза ұсынылды:

- Гипотеза №1: ChatGPT және оқытушының жеке критерийлер бойынша бағалары (мазмұны, құрылымы, тілі, шығармашылығы) статистикалық тұрғыдан маңызды емес. Балама: кем дегенде бір критерий бойынша айтарлықтай сәйкессіздік бар

- Гипотеза №2: ChatGPT оқытушымен салыстырғанда эссенің қорытынды бағаларын төмендетуге бейім. Балама: айырмашылықтардың таралуы симметриялы (жүйелі орын ауыстыру жоқ).

Зерттеудің жаңалығы студенттік жұмыстардың әртүрлі түрлерін бағалау үшін DeepSeek және Grok (бұрыннан танымал ChatGPT-тен басқа) сияқты ең жаңа LLM-ді пайдалану мүмкіндіктері мен шектеулерін жан-жақты талдау болып табылады.

Зерттеу әдістері

Студенттік жұмысты бағалау тек бағалау үшін ғана емес, сонымен қатар студенттерге кері байланыс беру, білім алшақтықтарын анықтау және оқу процесін түзету үшін де қызмет етеді. Бағалаудың көптеген тәсілдері бар, олардың әрқайсысының өзіндік артықшылықтары мен кемшіліктері бар. Дәстүр бойынша мұғалімнің жұмысын қолмен тексеру негізгі әдіс болып қала береді. Бұл әдіс студенттің контекстін, шығармашылығын және жеке стилін ескеруге мүмкіндік береді, бірақ ол айтарлықтай уақытты қажет етеді және субъективті әсерге ұшырайды. Субъективтілік, өз кезегінде, біркелкі емес бағалауға әкелуі мүмкін және процестің әділдігіне теріс әсер етуі мүмкін [4; 45].

Баламалы тәсілдерге автоматтандырылған жауаптарды тексерумен стандартталған тесттер, peer-review (студенттердің бір-бірінің жұмысын өзара бағалауы) және мәтінді автоматты түрде талдау үшін онлайн құралдарды пайдалану кіреді. Алайда, стандартталған тестілер, әдетте, дағдыларды емес, білімді тексерумен шектеледі, ал peer-review студенттерді оқытуды және тиімді модерация жүйесін қажет етеді [5; 122]. Мәтінді автоматты түрде талдауға арналған құралдар жетілдірілгенімен, көбінесе эссе жазу немесе бағдарламалық кодты әзірлеу сияқты күрделі оқу тапсырмаларына тән нюанстарды есепке ала алмайды [6; 96].

Соңғы жылдары үлкен тілдік модельдердің (LLM) қарқынды дамуы бағалау процесін автоматтандыру және жақсарту үшін жаңа көкжиектер ашады. LLM-бұл мәтіндік деректердің үлкен массивтерінде оқытылған терең нейрондық желілер. Олар мәтінді үлкен дәлдікпен түсінуге, құруға және талдауға қабілетті. LLM-бұл мұғалімді алмастыру емес, құрал екенін түсіну маңызды. Оларды бағалау процесінде қолдану автоматтандыру мен адам бақылауы арасындағы ақылға қонымды тепе-теңдікке негізделуі керек [7; 51]. LLM грамматика мен емлені тексеру сияқты күнделікті тапсырмаларды тиімді орындай алады және жұмысты алдын ала бағалауды қамтамасыз етеді. Мұғалім, өз кезегінде, шығармашылық компонентті талдауға, өзіндік ерекшелігін бағалауға және жеке Кері байланыс беруге шоғырлана алады [8; 103]. Айырмашылықты

түсіну үшін мұғалімнің қатысуымен және нейрондық байланыстарды қолданумен шағын эксперимент жасайық.

Мысал ретінде "жасанды интеллекттің еңбек нарығына әсері" тақырыбындағы эссені бағалауды қарастырыңыз. Студентке оң және теріс аспектілерді, сондай-ақ даму перспективаларын ескере отырып, тақырыпты ашатын 500-700 сөзден тұратын эссе жазу ұсынылды [9; 23].

АИ бағалау

Бағалау үшін келесі элементтерді қамтитын prompt көмегімен конфигурацияланған chatgpt моделі қолданылды:

Жұмыс түрі: Эссе.

Тақырыбы: "жасанды интеллекттің еңбек нарығына әсері".

Бағалау критерийлері:

о мазмұны (40%): тақырыпты ашудың толықтығы, талдау тереңдігі, дәлелдер мен мысалдардың болуы.

о құрылымы (20%): кіріспенің, негізгі бөліктің, қорытындының болуы, презентацияның қисындылығы.

о Тіл және стиль (20%): сауаттылық, экспозицияның анықтығы, академиялық стильге сәйкестік.

о Шығармашылық (20%): идеялардың өзіндік ерекшелігі, талдау тереңдігі.

* Нұсқаулық: "эссені көрсетілген критерийлер бойынша бағалаңыз, әр критерий бойынша балл қойыңыз (0-ден 10 баллға дейін) және өз бағаңыздың негіздемесін беріңіз." Chatgpt Ұпайы:

* Мазмұны (35/40): эссе тақырыпты жақсы ашады, мысалдар келтірілген, бірақ талдау әрдайым терең бола бермейді.

* Құрылым (18/20): құрылым логикалық, бірақ қорытынды неғұрлым кең болуы мүмкін. * Тіл және стиль (18/20): Мәтін сауатты, стиль академиялық, бірақ шамалы стилистикалық қателіктер бар.

* Шығармашылық (15/20): идеялар әдетте стандартты, өзіндік ерекшелігі жоқ.

* Жалпы балл: 86/100

ChatGPT негіздемесі: Эссе тақырыпты жақсы түсінетіндігін көрсетеді, бірақ талдау тереңдігі мен өзіндік ерекшелігі жоқ. Құрылым талаптарға сәйкес келеді, ал тіл мен стиль негізінен дұрыс.

Оқытушыны бағалау

* Мазмұны (38/40): тақырып жеткілікті түрде ашылды, дәлелдер келтірілді, бірақ кейбір аспектілерге сыни көзқарас жетіспейді.

* Құрылым (19/20): құрылым логикалық, жақсы ұйымдастырылған, бірақ қорытынды сәл Үстірт.

* Тіл және стиль (19/20): Мәтін сауатты, түсінікті, стиль академиялық талаптарға сәйкес келеді. * Шығармашылық (16/20): қызықты ойлар бар, бірақ жалпы эссе өзіндік ерекшелігімен ерекшеленбейді.

* Жалпы балл: 92/100

Мұғалімнің негіздемесі: эссе жақсы, тақырыпты жақсы түсінуді көрсетеді, бірақ терең талдау мен кейбір аспектілерге сыни көзқарас жетіспейді. Құрылым мен стиль талаптарға сәйкес келеді.

Салыстырмалы талдау 1 кестеде келтірілген:

Кесте 1. Салыстырмалы талдау

Критерий	ChatGPT Оценка	ChatGPT Обоснование	Преподаватель Оценка	Преподаватель Обоснование
Мазмұны	35/40	Эссе тақырыпты жақсы ашады, мысалдар келтірілген, бірақ талдау әрдайым терең бола бермейді.	38/40	Тақырып жеткілікті түрде ашылды, дәлелдер келтірілді, бірақ сыни көзқарас жоқ.
Құрылымы	18/20	Құрылым қисынды, бірақ қорытынды неғұрлым кең болуы мүмкін.	19/20	Құрылым логикалық, жақсы ұйымдастырылған, бірақ қорытынды сәл Үстірт.
Тіл және стиль	18/20	тін сауатты, стиль академиялық, бірақ шамалы стилистикалық қателіктер бар.	19/20	Мәтін сауатты, түсінікті, стилі академиялық талаптарға сәйкес келеді.
Шығармашылық	15/20	Идеялар әдетте стандартты, өзіндік ерекшелігі жоқ.	16/20	Қызықты ойлар бар, бірақ жалпы эссе өзіндік ерекшелігімен ерекшеленбейді.
Жалпы балл	86/100		92/100	

ChatGPT және оқытушы ұсынған бағалаудың салыстырмалы талдауы бағалау процесінде LLM қолданудың күшті және әлсіз жақтарын көрсетеді. ChatGPT берілген критерийлерге сәйкес келетін Жылдам және объективті бағалау қабілетін көрсетті. Алайда, мұғалімнің бағалауымен салыстырғанда, ChatGPT талдаудың төмен тереңдігін көрсетті, әсіресе шығармашылық пен нюанстарды есепке алуға қатысты. Мұғалім кең контекстік түсінікке ие бола отырып, Аргументтің қыр-сырын және идеялардың өзіндік ерекшелігін ескере отырып, жұмысты жан-жақты бағалай алды.

Гипотезаларды растау және/немесе жоққа шығару мақсатында екі түрлі тақырып бойынша "деректерді өндіру және анықтау" пәні бойынша 13 студенттің эсселері алынды, онда студенттер деректердің үлкен көлемін талдау үшін заманауи құралдарды қолдану туралы өз ойларын, сондай-ақ деректермен жұмыс істеу алгоритмінің бейімделуін пайдаланудың маңыздылығын түсіндірді. Барлық жұмыстарды оқытушы да, ChatGPT де тексерді, «Деректердің үлкен көлемін талдау үшін заманауи құралдарды қолдану» тақырыбындағы эссені бағалау нәтижелері 2-кестеде келтірілген.

Кесте 2. Эссені бағалау нәтижелері

№	ChatGPT	Оқытушы	Айырмашылық (Δ)
1	85	91	-6
2	83	89	-6
3	78	85	-7

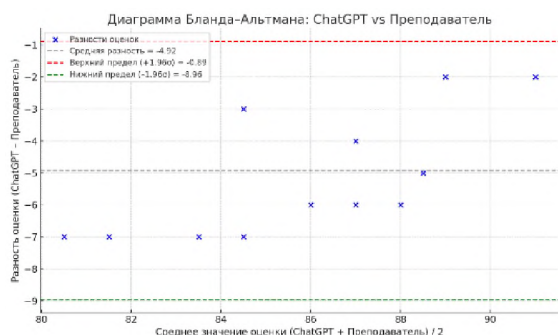
4	88	90	-2
5	80	87	-7
6	86	91	-5
7	90	92	-2
8	84	90	-6
9	77	84	-7
10	85	89	-4
11	81	88	-7
12	88	90	-2
13	83	86	-3

«Деректермен жұмыс істеу алгоритмінің бейімделуін пайдалану» тақырыбындағы "деректерді өндіру және оларды анықтау" пәні бойынша эссені бағалау нәтижелері 3-кестеде келтірілген.

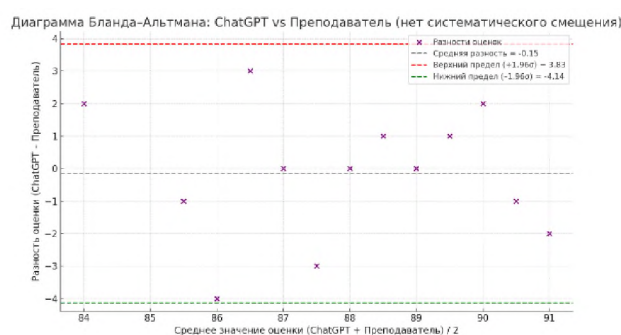
Кесте 3. Эссені бағалау нәтижелері

№	ChatGPT	Оқытушы	Айырмашылық (Δ)
1	90	89	+1
2	86	89	-3
3	85	86	-1
4	88	88	0
5	91	89	+2
6	84	88	-4
7	89	88	+1
8	87	87	0
9	90	92	-2
10	88	85	+3
11	90	91	-1
12	89	89	0
13	85	83	+2

ChatGPT көмегімен орындалған автоматтандырылған бағалау мен оқытушының сараптамалық бағасы арасындағы сәйкестік дәрежесін бағалау үшін олардың орташа мәніне қатысты жұптастырылған бағалаулардың ($\Delta = \text{ChatGPT} - \text{оқытушы}$) айырмашылықтарын көрсететін Бланд–Альтман диаграммалары салынды. Диаграммалар 1-суретке сәйкес ұсынылған.



а



б

Сурет 1. Бланд–Альтман диаграммалар

1а-суретте көрсетілгендей, студенттердің 13 эссесінің деректері келтірілген, онда ChatGPT моделі әр жағдайда мұғалімге карағанда төмен баға берді. Орташа айырмашылық $d^- = -5.15$ болды, бұл тұрақты теріс мәннің болуын көрсетеді. 95% сенімділік аралығындағы келісім шектері -7.38-ден -2.92-ге дейін болды және барлық бақылаулар осы диапазонға кірді, бұл сәйкессіздіктердің тұрақтылығын және бағалауда LLM-ді одан әрі пайдалану кезінде олардың ықтимал болжамдылығын көрсетеді.

1б-суретте ChatGPT мен оқытушының ұпайлары арасындағы айырмашылықтар екі бағытта өзгертін 13 эссенің балама үлгісінің деректері берілген. Орташа айырмашылық $d^- = 0.15$ болды, бұл іс жүзінде нөлге тең. Бағалаудың таралуы симметриялы болып шықты, айқын төмендеу немесе асыра сілтеу үрдісі жоқ. Барлық нүктелер келісімнің шекарасына түсіп, қорытынды жасауға мүмкіндік береді: бұл деректер жиынтығында жүйелі ауытқу жоқ және LLM нәтижелері көбінесе тапсырманың сипатына, prompt ' а тұжырымдамасына және пәннің ерекшелігіне байланысты.

Ұсынылған диаграммалар LLM мен оқытушы арасындағы ықтимал сәйкессіздіктер жүйелі және кездейсоқ болуы мүмкін екенін көрсетеді. Бұл қосымша бақылау, prompt нұсқауларын бейімдеу және автоматтандыруды сараптамалық интерпретациямен біріктіретін гибриді бағалау тәсілінің қажеттілігін көрсетеді.

Нәтижелер CHATGPT сияқты LLM бағалау процесінің бір бөлігін автоматтандырудың пайдалы құралы болуы мүмкін екенін көрсетеді, бірақ студенттердің жұмысын жан-жақты және сапалы бағалауды қамтамасыз ету үшін оқытушының қатысуы қажет, әсіресе шығармашылық және аналитикалық аспектілерді бағалау кезінде. Оқытушының уақытын үнемдей отырып және студенттер үшін жоғары сапалы кері байланысты қамтамасыз ете отырып, процесті оңтайландыруға мүмкіндік беретін LLM автоматтандырылған бағалауы мен адам сараптамасының үйлесімі тамаша тәсіл болып көрінеді.

Жұмыстарды бағалау үшін LLM пайдалану туралы қорытынды 4-кестеде келтірілген.

Кесте 4. Артықшылықтары мен кемшіліктері:

Артықшылықтары (ChatGPT)	Кемшіліктер (ChatGPT)
Жылдамдық: жұмысты жылдам өңдеу және бағалау.	Талдаудың жеткіліксіз тереңдігі: тақырыпты түсіну үшін маңызды нәзіктіктер мен нюанстарды жіберіп алуы мүмкін.
Объективтілік: адаммен салыстырғанда субъективті әсерге аз әсер ету.	Шығармашылық ойлаудың болмауы: талдаудың өзіндік ерекшелігі мен тереңдігін, әсіресе гуманитарлық ғылымдар саласында бағалау қиын.
Тәртіп: берілген бағалау критерийлерін қатаң сақтайды.	Өнеркәсіптік өнімдердің сапасына тәуелділік: нәтиже сұраныстың дұрыс тұжырымдалуына байланысты.
Жеке Кері байланыс әлеуеті: әр критерий бойынша нақты ескертулер бере алады. (қайта қарауды қажет етеді)	Контекстті түсіну мәселелері: күрделі контекст пен метафораларды түсінуде қиындықтар туындауы мүмкін.

Артықшылықтары (ChatGPT)	Кемшіліктері (ChatGPT)
Масштабтау мүмкіндігі: көптеген жұмыстарды оңай бағалаңыз.	Тексеру қажеттілігі: оқытушы LLM ұсынған бағалар мен негіздемелерді тексеруді талап етеді.
	Біржақтылық ықтималдығы: LLM оқыту деректеріндегі бар әлеуметтік стереотиптер мен жағымсыздықтарды көрсетуі мүмкін.
Терең түсіну: оқушының контекстін, шығармашылығын, жеке стилін ескеруі мүмкін.	Субъективтілік: бағалау мұғалімнің жеке қалауы мен тәжірибесіне байланысты болуы мүмкін.
Сыни тұрғыдан ойлау: талдау тереңдігін, өзіндік ерекшелігін және дәлелдеу сапасын бағалай алады.	Уақыт шығындары: жұмысты тексеру айтарлықтай уақытты қажет етеді.
Нюанстарды ескеру: күрделі жұмыстарға тән нюанстар мен нәзіктіктерді ескеруге қабілетті.	Масштабтау: көптеген жұмыстарды бағалау қиын.
Толық кері байланыс беру: жұмыстың күшті және әлсіз жақтарын ескере отырып, егжей-тегжейлі және жекелендірілген кері байланыс бере алады.	Бағалаудың біркелкі естігі: оқытушының шаршауына немесе критерийлердің өзгеруіне байланысты бағалаулар біркелкі болмауы мүмкін.

DeerSeek-бұл жасанды интеллектке негізделген қуатты құрал, оны студенттердің жұмысын, әсіресе техникалық және математикалық пәндерді тексеру және талдау үшін пайдалануға болады. Оның мүмкіндіктері оқытушылар мен студенттерге тапсырмаларды тиімдірек өңдеуге және бағалауға мүмкіндік береді.

Жұмыстарды бағалау үшін Deer Seek пайдалану туралы қорытынды 5-кестеде келтірілген.

Кесте 5. Артықшылықтары мен кемшіліктері DeerSeek:

Критерий	Артықшылықтары	Кемшіліктері
Кодты тексеру	Кодты тез талдайды, қателерді табады және түзетулер ұсынады.	Стандартты емес немесе шығармашылық шешімдерді ескермеуі мүмкін.
Математикалық есептер	Есептеулерді тексереді, шешім логикасын талдайды.	Студенттің шешімге қатысты түсіндірмелерін әрдайым түсіне бермейді.
Техникалық мәтіндерді талдау	Формулалардың, есептеулердің және дәлелдердің дәлдігін бағалайды.	Күрделі ғылыми мәтіндердің стилистикасы мен логикасын нашар талдайды.
Бағалауды автоматтандыру	Тапсырмаларды тексеру процесін жылдамдатады, мұғалімдерге жүктемені азайтады.	Адам тарапынан түзету мен бақылауды қажет етеді.

Кері байланыс	Қателерді түсіндіріп, балама шешімдерді ұсына алады.	Әрқашан егжей-тегжейлі және түсінікті түсініктеме бермейді.
Плагиатты анықтау	Кодты немесе стандартты шешімдерді көшіруді анықтай алады.	Гуманитарлық жұмыстардағы мәтіндік плагиат үшін тиімсіз.
Тілдерді қолдау	Ағылшын және қытай тілдері үшін жақсы, техникалық терминдерге сәйкес келеді.	Басқа тілдерге, соның ішінде орыс тіліне шектеулі қолдау.
Гуманитарлық мәтіндермен жұмыс	-	Эсселерді, эсселерді және шығармашылық мәтіндерді нашар талдайды.

Осылайша, DeepSeek-бағдарламалауды, математиканы және техникалық пәндерді бағалаудың тамаша құралы. Бұл қателерді табуға, кодты жақсартуға және студенттердің жұмысын талдауға көмектеседі. Алайда, гуманитарлық пәндер мен шығармашылық тапсырмалар үшін GPT Жақсы [10; 41].

GROK-бұл XAI тілдік моделі, ол мәтінді құруға, деректерді талдауға және X (Twitter) - ден алынған ақпаратпен жұмыс істеуге бағытталған. Осыған қарамастан, оның студенттік жұмысты бағалау мүмкіндіктері GPT және DeepSeek-пен салыстырғанда әзірге шектеулі. 6-кестеде студенттердің жұмысын бағалау кезінде LLM салыстыруы келтірілген.

Кесте 6. Студенттік жұмысты бағалау кезінде GPT, DeepSeek және GROK салыстыруы

Критерий	GPT (OpenAI)	DeepSeek	GROK (xAI)
Бағдарламалық кодты тексеру	Кодты жақсы талдайды, түзетулер мен оңтайландыруды ұсынады.	Кодты тексеру үшін тамаша, логика мен өнімділікті талдайды.	Шектеулі мүмкіндіктер, бағдарламалауға бағытталған емес.
Математикалық есептерді тексеру	Есептеулерді тексереді, қателерді түсіндіре алады.	Математикалық және логикалық есептерді талдаудағы көшбасшы.	Математикалық есептеулерде әлсіз.
Эсселер мен мәтіндерді талдау	Құрылымды, дәлелді және стильді жақсы бағалайды.	Техникалық есептерді тексере алады, бірақ шығармашылық мәтіндермен нашар жұмыс істейді.	Қысқа мәтіндерді тексере алады, бірақ Үстірт.
Грамматикалық және стилистикалық тексеру	Қателерді түзетеді, мәтін стилін жақсартады.	Техникалық дәлдікке назар аударады, стилистика фонда.	Емлені тексереді, бірақ егжей-тегжейлі талдауда әлсіз.
Плагиатты тексеру	Ашық көздермен сәйкестіктерді таба алады, бірақ толыққанды анти-плагиатты алмастырмайды.	Кодты және стандартты шешімдерді көшіруді анықтай алады.	Шектеулі мүмкіндіктер.
Студенттерге кері байланыс	Жақсарту бойынша егжей-тегжейлі түсініктемелер мен ұсыныстар береді.	Қателерді талдайды және түзетулер ұсынады, бірақ құрғақ.	Қысқаша түсініктеме береді, бірақ әрқашан толық емес.

Талдау тереңдігі	Әмбебап, әртүрлі заттарға жарамды.	Техникалық пәндерде мықты, бірақ гуманитарлық пәндерде әлсіз.	Беттік талдау, қысқа жауаптарға көбірек бағытталған.
Тілдерді қолдау	Көптеген тілдерді, соның ішінде орыс тілін қолдайды.	Қытай және ағылшын тілдеріне баса назар аудару.	Негізінен ағылшын, орыс қолдауы шектеулі.

Осылайша, GPT студенттердің жұмысын (код, математика, мәтіндер) бағалау үшін ең жақсы әмбебап көмекші деп қорытынды жасауға болады. DeepSeek-техникалық және математикалық пәндер үшін өте қолайлы, бірақ гуманитарлық ғылымдарда әлсіз. GROK-қысқа жауаптар мен мәтіндерді жылдам тексеруге жарамды, бірақ терең талдауда басқа модельдерден төмен [11; 94].

Қорытынды

Осы зерттеу барысында студенттердің жұмысын бағалау үшін үлкен тілдік модельдерді (LLM), атап айтқанда ChatGPT, DeepSeek және Grok қолдану мүмкіндіктері қарастырылды. Айта кету керек, LLM жұмыстың формальды аспектілерін (құрылым, грамматика, академиялық стиль), сондай-ақ алдын-ала бағалау мәселелерін бағалауда жоғары тиімділік көрсетті. Оқытушының бағалауымен салыстыру автоматтандыру уақытты айтарлықтай үнемдейтінін растайды, бірақ күрделі элементтерді түсіндіру үшін сарапшының бақылауын қажет етеді.

Алайда, креативтілік пен талдаудың тереңдігін бағалауға қатысты ChatGPT мұғалімге жол берді, бұл күрделі және көп қырлы міндеттерді бағалауда адам сараптамасының қажеттілігін көрсетеді.

CHATGPT сияқты LLM студенттердің жұмысын бағалаудың құнды құралы болып табылады, бірақ оларды пайдалану олардың күшті және әлсіз жақтарын ескере отырып, бар білім беру тәжірибелеріне біріктірілуі керек.

Салыстырмалы талдау көрсеткендей, GPT сізге көптеген тілдерді, шығармашылық тапсырмаларды және контекстке жақсы есте сақтауды қажет етсе жақсы. Егер сіз математикамен, бағдарламалаумен немесе қытай тілімен жұмыс жасасаңыз, DeepSeek пайдалы болуы мүмкін.

Осылайша, GPT студенттердің жұмысын бағалау үшін ең жақсы әмбебап көмекші деп қорытынды жасауға болады. DeepSeek-техникалық және математикалық пәндер үшін өте қолайлы, бірақ гуманитарлық ғылымдарда әлсіз. GROK-қысқа жауаптар мен мәтіндерді жылдам тексеруге жарамды, бірақ терең талдауда басқа модельдерден төмен. Автоматтандыру уақытты айтарлықтай үнемдей алады, бірақ күрделі элементтерді түсіндіру үшін сарапшының бақылауын қажет етеді. Гибридті тәсіл ұсынылады, атап айтқанда автоматтандырылған LLM бағалаулары мен сараптамалық педагогикалық интерпретацияның үйлесімі.

Әдебиет:

1. Perera Perera, R., Lankathilaka, M. Evaluating the efficacy of ChatGPT in automated essay scoring. Journal of Educational Technology & Society, 2024, 27(1), 1-15.
2. Holmes, W., Bialik, M., Fadel, C. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign, 2023.
3. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 2023, 103, 102274

4. Смирнов, В.А. Использование больших языковых моделей в образовании: обзоры и перспективы // Вестник образования. – 2020. – № 5(12). – С. 45-62.
5. Петрова, О.С. Оценка студенческих эссе: сравнительный анализ традиционных методов и автоматизированных подходов. // Вопросы педагогики. – 2021. – № 28(3). – С. 112-130.
6. Козлов, Д.Ю. Искусственный интеллект в образовании: этические аспекты и вызовы. // Философия образования. – 2022. – № 15(2). – С. 78-95.
7. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. Language models are few-shot learners. *Advances in neural information processing systems*, 33, 2020 1877-1901.
8. Floridi, L., & Chiriatti, M. GPT-3: Its nature, scope, limits, and implications *Minds and Machines*, 30(3), 2020, 683-694.
9. Holmes, W., Bialik, M., & Fadel, C. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. *Center for Curriculum Redesign*. 2023.
10. Баррат, Д. Последнее изобретение человечества: искусственный интеллект и конец эры Homo sapiens / Джеймс Баррат; [пер. с англ. Наталья Лисова]. – 2-е издание. – Москва: АНФ, 2019. – 396 с.
11. Дозэрти, П. Человек + машина. Новые принципы работы в эпоху искусственного интеллекта / П. Дозэрти, Дж. Уилсон; пер. с англ. О. Сивченко, Н. Яцюк. – М.: Манн, Иванов и Фербер, 2019. – 298 с.

References:

1. Perera, R., Lankathilaka, M. Evaluating the efficacy of ChatGPT in automated essay scoring. *Journal of Educational Technology & Society*, 2024, 27(1), 1-15.
2. Holmes, W., Bialik, M., Fadel, C. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. *Center for Curriculum Redesign*, 2023.
3. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 2023, 103, 102274.
4. Smirnov, V.A. (2020). Using Large Language Models in Education: Reviews and Perspectives. // *Bulletin of Education*. – 2020, 5 (12), 45-62. (In Russian)
5. Petrova, O.S. (2021). Assessment of Student Essays: A Comparative Analysis of Traditional Methods and Automated Approaches. // *Problems of Pedagogy*. – 2021, 28(3), 112-130. (In Russian)
6. Kozlov, D.Yu. (2022). Artificial Intelligence in Education: Ethical Aspects and Challenges. // *Philosophy of Education*. – 2022, 15(2), 78-95. (In Russian)
7. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
8. Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and implications. *Minds and Machines*, 30(3), 683-694.
9. Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. *Center for Curriculum Redesign*.
10. Barratt, D. The last invention of mankind: artificial intelligence and the end of the era of Homo sapiens / James Barratt; [trans. from English. Natalia Lisova]. – 2nd edition. – Moscow: ANF, 2019. – 396 p.
11. Doherty, P. Man + machine. New principles of work in the era of artificial intelligence / P. Doherty, J. Wilson; translated from English by O. Sivchenko, N. Yatsyuk. – M.: Mann, Ivanov and Ferber, 2019. – 298 p.

Information about the author:

Muntinov Kairat – corresponding author, Lecturer, department of Information and Communication Technologies, master, Manash Kozybayev North Kazakhstan University NPLC, Petropavlovsk, Kazakhstan; e-mail: Kairatmuntinov@gmail.com.