ПЛОВДИВСКИ УНИВЕРСИТЕТ „ПАИСИЙ ХИЛЕНДАРСКИ“
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
КАТЕДРА „КОМПЮТЪРНА ИНФОРМАТИКА“

Иван Марианов Иванов

МОДЕЛ И СИСТЕМА ЗА АВТОМАТИЗИРАНО ГЕНЕРИРАНЕ И ВЕРИФИКАЦИЯ НА ТЕСТОВИ ВЪПРОСИ С ИЗКУСТВЕН ИНТЕЛЕКТ

АВТОРЕФЕРАТ

на дисертационен труд

за присъждане на образователна и научна степен „доктор“

Област на висше образование
4. Природни науки, математика и информатика

Професионално направление
4.6. Информатика и компютърни науки
докторска програма „Информатика“

Научни ръководители:
проф. д-р Станка Хаджиколева
проф. д-р Георги Димитров

Пловдив, 2026 г.

Дисертационният труд е обсъден и насочен за защита пред научно жури, на заседание на катедра „Компютърна информатика“ при Факултета по математика и информатика на Пловдивския университет „Паисий Хилендарски“, на 09.07.2026 г.

Дисертационният труд „Модел и система за автоматизирано генериране и верификация на тестови въпроси с изкуствен интелект“ съдържа 144 страници. Списъкът на използваната литература включва 89 (осемдесет и девет) източника. Списъкът на авторските публикации по темата се състои от 6 заглавия.

Защитата на дисертационния труд ще се състои на ..... г. от ..... часа в Заседателната зала на ПУ „Паисий Хилендарски“ (Нова сграда, бул. „България“ №236).

Материалите по защитата са на разположение на интересуващите се в Деканата на Факултета по математика и информатика, Нова сграда на ПУ „Паисий Хилендарски“ (бул. „България“ №236), всеки работен ден от 8:30 до 17 часа.

Автор: Иван Марианов Иванов

Заглавие: Модел и система за автоматизирано генериране и верификация на тестови въпроси с изкуствен интелект

Съдържание

УВОД

Бързото развитие на изкуствения интелект през последните години оказва съществено влияние върху различни сфери на обществения живот, включително върху образованието. Съвременните технологии, базирани на изкуствен интелект, разкриват нови възможности за анализ, автоматизация, персонализация и оптимизиране на образователните дейности. Особено значим е техният потенциал за усъвършенстване на разработването на учебно съдържание, организирането и провеждането на обучението, както и на оценяването на знанията и уменията на обучаемите. Генеративните модели и чатботовете, базирани на големи езикови модели, подпомагат създаването на тестови въпроси, анализа на резултатите и предоставянето на персонализирана обратна връзка към обучаемите. Тези възможности създават предпоставки за разработването на ново поколение интелигентни системи за електронно тестване, които повишават ефективността, адаптивността и качеството на процеса на оценяване.

Основната цел на дисертационното изследване е да се разработи модел, архитектура и софтуерна реализация на интелигентна платформа за електронно тестване, използваща генеративен изкуствен интелект за автоматизирано създаване, верификация, управление и оценяване на тестово съдържание.

За постигане на основната цел са поставени следните ***задачи***:

Задача 1: Проучване на съществуващите системи за електронно тестване и идентифициране на основните функционални и нефункционални изисквания към интелигентна система за електронно тестване;

Задача 2: Разработване на концептуален модел и архитектура на интелигентна платформа за електронно тестване, включваща механизми за автоматизирано генериране и верификация на тестови въпроси чрез големи езикови модели.

Задача 3: Разработване на модул за автоматизирано адаптивно оценяване на обучаемите, базиран на съвременни психометрични подходи и изкуствен интелект;

Задача 4: Реализиране на софтуерен прототип, базиран на предложения модел;

Задача 5: Провеждане на експериментални изследвания за оценка на качеството и приложимостта на разработеното решение.

Основна хипотезата на настоящото изследване е, че използването на генеративен изкуствен интелект в рамките на специализирана платформа за електронно тестване ще подпомогне преподавателите при създаването и управлението на тестово съдържание, ще повиши качеството и надеждността на тестовите въпроси чрез автоматизирана верификация и ще осигури по-ефективно, обективно и персонализирано оценяване на обучаемите.

Проблемът на дисертационното изследване е мотивиран от необходимостта от разработване на интелигентни решения за електронно оценяване, които да преодоляват ограниченията на традиционните системи за създаване на тестове. Съществуващите платформи често разчитат на предварително дефинирани правила, шаблони и значителна ръчна работа при подготовката на тестовото съдържание. Въпреки широкото навлизане на генеративните езикови модели, липсват интегрирани решения, които да осигуряват едновременно автоматизирано генериране на въпроси, контрол на качеството чрез независима верификация, адаптивно тестване и анализ на резултатите.

Това налага разработването на нов подход към електронното оценяване, който съчетава възможностите на генеративния изкуствен интелект с педагогическите принципи за валидност, надеждност и адаптивност на оценяването.

Структурата на дисертационния труд отразява проведените теоретични изследвания, проектирането на концептуалния модел, разработването на софтуерния прототип и резултатите от проведените експерименти. Трудът се състои от увод, четири глави, заключение, списък на използваната литература, списък на публикациите по темата и едно приложение, илюстриращо създаването на въпроси от един генеративен модел и верификацията им от втори.

В ***Глава 1. Въведение*** е представено проучване на възможностите за използване на изкуствения интелект в образованието и приложението на генеративните езикови модели и чатботове за електронно тестване. Разгледана е еволюцията на системите за създаване на тестове и класическите подходи за автоматизирано генериране на тестови въпроси. Формулирани са основните изисквания към съвременна интелигентна платформа за електронно тестване.

В ***Глава 2. Концептуален модел на платформа за е-тестове*** е предложен концептуален модел на интелигентна система за електронно тестване. Представени са основните потребителски роли и функционалности, архитектурата на платформата, механизмите за генериране и верификация на тестови въпроси чрез генеративен изкуствен интелект, както и някои случаи на употреба. Разгледани са взаимодействията между отделните компоненти на системата и процесите, свързани със създаването, управлението и оценяването на тестово съдържание в интелигентна образователна среда.

В ***Глава 3. Софтуерна реализация*** е представена реализацията на предложената платформа, използваните технологии, структурата на системата и основните програмни модули. Описани са механизмите за интеграция с големи езикови модели, управлението на тестово съдържание и реализираните средства за автоматизирано оценяване. Разгледани са архитектурните решения, процесът на разработка и взаимодействието между отделните компоненти на системата, както и реализираните функционалности за генериране, верификация и администриране на тестове в електронна среда.

В ***Глава 4. Експерименти*** са представени резултатите от проведените експериментални изследвания и оценка на разработеното решение. Анализирани са качеството на генерираните въпроси, ефективността на верификационните механизми и приложимостта на платформата в реална образователна среда. Представени са и резултатите от проведените експерименти, които потвърждават ефективността на предложените подходи и възможностите за практическото им приложение в процеса на електронно оценяване.

В ***Заключението*** е направен преглед на извършената работа и на постигнатите резултати по поставените цели и задачи. Представени са основните научни, научно-приложни и приложни приноси на изследването, които потвърждават неговата актуалност и практическа значимост. Очертани са също така перспективите за бъдещо развитие на проведеното изследване.

Благодарности

Изказвам своята искрена благодарност към научните ми ръководители – проф. д-р Станка Хаджиколева и проф. д-р Георги Димитров, за предоставената възможност да проведа настоящото дисертационно изследване. Благодаря за получените ценни съвети, конструктивни препоръки, критични бележки и насоки, които бяха от съществено значение за реализирането на поставените цели и успешното завършване на дисертационния труд.

Изразявам своята дълбока признателност към д-р инж. Тодор Рачовски и проф. д-р Емил Хаджиколев за оказаната помощ, полезните идеи, професионалните консултации и подкрепата по време на провеждането на изследването. Тяхното съдействие, споделен опит и градивни предложения допринесоха значително за развитието на настоящата работа и за преодоляването на редица научни и практически предизвикателства.

ГЛАВА 1. ВЪВЕДЕНИЕ

1.1. Изкуствен интелект в образованието

През последните години бързото развитие на софтуерните технологии, използващи изкуствен интелект, създаде нови възможности за тяхното интегриране в образователния процес (Zhang & Aslan, 2021; Rojas & Chiappe, 2024). Това включва използването на технологии за обработка на естествен език чрез големи езикови модели (Kasneci, Sessler et al., 2023; Jeon & Lee, 2023), технологии за разпознаване на образи чрез компютърно зрение (Chen, Samuel et al., 2021; Agbo et al., 2020), за анализ на данни и прогнозиране с помощта на невронни мрежи (Kucak, Juricic et al., 2018; Hadzhikolev, Hadzhikoleva et al., 2021) и др. Резултатите от множество проучвания показват, че ИИ има потенциала да повиши качеството на обучението, да го направи по-достъпно и по-добре съобразено с индивидуалните потребности на обучаемите (Sofianos, Stefanidakis et al., 2024; Harry & Sayudin, 2023). Изкуственият интелект намира приложение в различни области на образованието – социални науки (Nurhayati & Halimah, 2024), инженерно образование (Nunez & Lantada, 2020), природни науки (Darayseh, 2023; Hadzhikoleva, Borisova et al., 2025), науки за живота (Kandlhofer, Steinbauer et al., 2016), медицинско образование (Briganti & Le Moine, 2020), изучаване на чужди езици (Edmett, Ichaporia et al., 2023) и много други образователни направления (Hajkowicz, Sanderson et al., 2023). Това показва, че ИИ постепенно се превръща в универсален инструмент за подпомагане на обучението и оценяването, независимо от конкретната предметна област.

Особено широко приложение ИИ намира в STEM образованието. Използва се за изграждане на интелигентни обучаващи системи, системи за профилиране и прогнозиране, както и за адаптивно обучение и персонализация на учебното съдържание (Xu & Ouyang, 2022). Проведени изследвания доказват положителната роля на ИИ за подобряване на персонализираните учебни преживявания, идентифициране на обучаеми в риск и автоматизиране на административни дейности (Rahiman & Kodikal, 2024). В допълнение различни алгоритми с изкуствен интелект се използват за анализ

на данни и осигуряване качеството на висшето образование чрез ранно откриване на проблеми и отклонения в ключови показатели за ефективност (Mishra, 2019).

Развитието на големите езикови модели (Large Language Models – LLMs) поставя началото на нов етап в развитието на изкуствения интелект и неговото приложение в образованието. Особено внимание през последните години привличат чатботовете, базирани на големи езикови модели. Сред най-популярните представители на този клас системи са ChatGPT, Gemini, Claude, Copilot и други подобни решения. Те предоставят възможност за интерактивно взаимодействие с потребителите чрез естествен език и могат да изпълняват разнообразни задачи, свързани с обучението, преподаването и научноизследователската дейност (Dempere, Modugu et al., 2023). Различни изследвания потвърждават че генеративните чатботове, използвани по подходящ начин, могат успешно да подпомагат обучаемите при усвояването на нови знания, решаването на задачи, подготовката на проекти и провеждането на самостоятелни изследвания (Chaudhry, Sarwary et al., 2023; Pradana, Elisa et al., 2023). Също така могат да се използват за създаване на геймифицрани ресурси (Borisova, Hadzhikoleva et al., 2024; Hadzhikoleva, Gorgorova et al., 2024).

1.2. Автоматизирано оценяване и електронно тестване

Използването на изкуствен интелект при оценяването на знанията позволява автоматизация на редица дейности, които традиционно се извършват от преподавателите. Това включва създаване и проверка на тестови въпроси, анализ на свободни текстови отговори, оценяване на учебни задания и предоставяне на обратна връзка към обучаемите. Изследвания показват, че AI технологиите могат успешно да подпомагат процесите по оценяване както на знания от по-ниско когнитивно ниво, така и на умения, свързани с анализ, интерпретация и решаване на проблеми (Hadzhikolev et al., 2021).

Въпреки значителните предимства на автоматизираното оценяване, неговото прилагане е свързано с редица предизвикателства. Сред тях са гарантирането на надеждност, обективност и прозрачност на процесите по оценяване, както и осигуряването на ефективен контрол върху качеството на генерираното съдържание. Използването на генеративни модели поставя и предизвикателства, свързани с възможността за генериране на фактически неточности и необходимост от експертна верификация и валидация на автоматично създадените тестови въпроси.

1.3. Еволюция на системите за създаване на тестове

Създаването на тестове обикновено изисква значителни времеви и човешки ресурси. Освен това, тестовете трябва да бъдат оценени за надеждност и валидност, като от съществено значение е провеждането на пилотно тестване с цел идентифициране на слабости и пропуски и тяхното отстраняване (Osterlind, 1998). Това мотивира преподавателите активно да търсят начини за автоматизиране на този процес, включително генериране на тестови въпроси, конструиране и проверка на тестове (Bugbee, 1996). В началото се разработват приложения, които използват традиционни методи и технологии.

Класическите подходи за създаване на тестове имат някои съществени ***ограничения***:

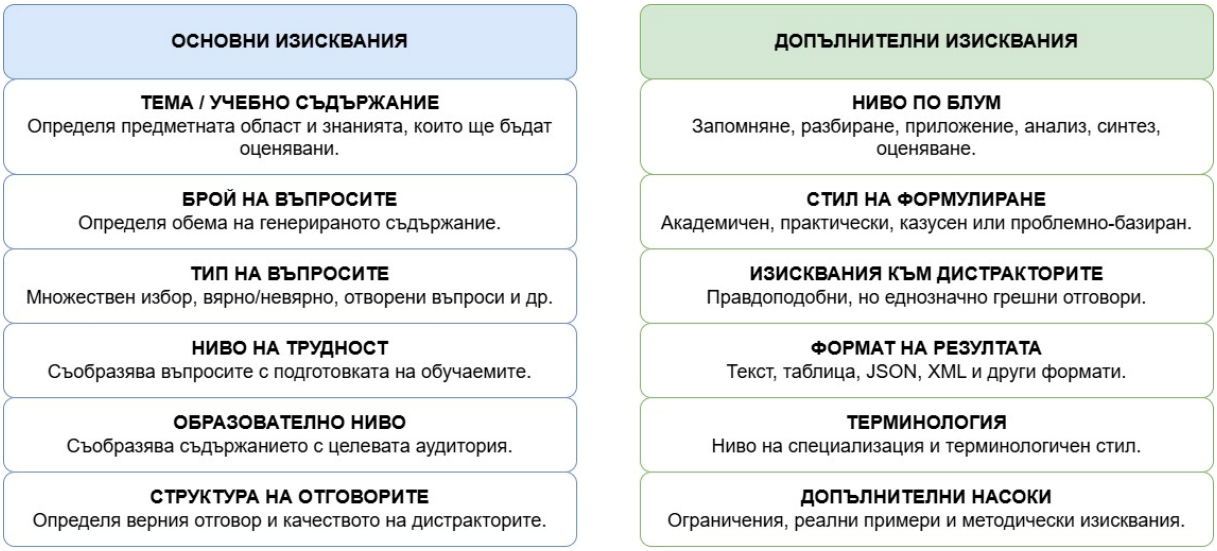
- Традиционните системи са силно ***зависими от предварително зададени правила и шаблони***, което ограничава възможността за генериране на разнообразни въпроси, особено когато се изискват по-сложни или адаптивни тестови формати.
- Тези системи изискват ***значителна ръчна настройка*** както при създаването на правила, така и при избора на подходящи ключове и разсейващи отговори, което ги прави времеемки и трудоемки.
- ***Адаптацията е ограничена – обикновено е базирана на предварителни отговори*** и не може лесно да се приспособи към динамично променящи се условия.
- ***Приложенията използват ограничена обработка на семантична информация***, което намалява способността им да разбират и интерпретират по-дълбоки контексти и връзки в текста.

Генеративните модели с изкуствен интелект предоставят възможност за бързо създаване на голям брой разнообразни въпроси с различно ниво на сложност и намаляване на времето, необходимо за подготовка на тестове. Въпреки големите възможности на съвременните генеративни модели, качеството на генерираните въпроси зависи в значителна степен от качеството на подадените инструкции. За разлика от традиционните софтуерни системи, при които потребителят работи чрез предварително дефиниран интерфейс, взаимодействието с големите езикови модели се осъществява чрез текстови заявки (промптове). За да бъдат получени коректни, педагогически обосновани и съобразени с учебните цели тестови въпроси, преподавателят трябва да формулира достатъчно подробен и структуриран промпт.

Базовата информация, която следва да бъде включена в такъв промпт, включва:

- ***Тема или учебно съдържание*** **–** определя предметната област и знанията, които ще бъдат оценявани. Колкото по-конкретно е зададена, толкова по-релевантни са генерираните въпроси.
- ***Брой на въпросите*** **–** определя обема на генерираното съдържание.
- ***Тип на въпросите*** **–** необходимо е да бъде уточнен видът на въпросите – с един верен отговор, множествен избор, вярно/невярно, отворени въпроси и др., тъй като различните формати оценяват различни аспекти на знанията.
- ***Ниво на трудност*** **–** задаването на ниво позволява въпросите да бъдат съобразени с подготовката на обучаемите и предотвратява генерирането на прекалено лесни или прекалено сложни задачи.
- ***Образователно ниво или целева аудитория*** **–** съдържанието и терминологията на въпросите трябва да бъдат съобразени с аудиторията – ученици, студенти, специалисти и др.
- ***Структура на отговорите и верен отговор.*** При въпросите от затворен тип е необходимо моделът да генерира не само условието на въпроса, но и набор от възможни отговори, сред които ясно да бъде определен верният. Добре конструираните отговори са ключов фактор за надеждността и валидността на тестовото оценяване.

От гледна точка на педагогическия дизайн най-критични са *темата, типът на въпроса, нивото на трудност и образователното ниво*, защото те определят дали генерираните въпроси изобщо ще бъдат подходящи за конкретната учебна цел (фиг. 1).



Фигура 1. Изисквания за добър промпт за генериране на тестови въпроси

В промпта могат да бъдат включени и допълнителни изисквания, с които да се подобрят качеството и педагогическата стойност на тестовите въпроси. Такива може да бъдат например:

- ***Когнитивно ниво по таксономията на Блум.*** Задаването на когнитивно ниво позволява въпросите да бъдат насочени към определен тип познавателна дейност. В зависимост от целите на обучението въпросите могат да проверяват запомняне на факти, разбиране на концепции, приложение на знания в нови ситуации, анализ на зависимости, оценяване на решения или създаване на нови идеи и подходи.
- ***Стил на формулиране.*** Въпросите могат да бъдат генерирани в различен стил според спецификата на учебната дисциплина. Възможни са академичен стил, практически ориентирани въпроси, въпроси под формата на казуси, проблемно-базирани задачи или въпроси, свързани с реални ситуации от професионалната практика.
- ***Изисквания към дистракторите.*** Качеството на грешните отговори е от съществено значение за надеждността на теста. Добре формулираните дистрактори трябва да бъдат правдоподобни и тематично свързани с въпроса, без да бъдат подвеждащо двусмислени или очевидно неправилни.
- ***Формат на резултата.*** Генерираното съдържание може да бъде представено в различни формати в зависимост от начина на последваща обработка – обикновен текст, таблица, JSON, XML или друг структуриран формат, подходящ за употреба.
- ***Използвана терминология.*** Може да бъде зададено предпочитано ниво на специализация и терминологичен стил. Например въпросите могат да

използват общодостъпна терминология за начинаещи обучаеми или специализирана професионална терминология за напреднали студенти и експерти.

- ***Допълнителни ограничения и насоки.*** В промпта могат да бъдат включени специфични изисквания към формулировката на въпросите, като избягване на отрицателни конструкции, недопускане на подвеждащи или двусмислени твърдения, използване на реални примери, включване на практически ситуации или съобразяване с конкретни образователни стандарти и методически препоръки.

1.4. Основни изисквания към интелигентна платформа за електронно тестване

Една модерна платформа за електронно тестване следва да отговаря на набор от функционални и нефункционални изисквания, които гарантират ефективност, надеждност и педагогическа стойност.

На първо място, системата трябва да предоставя възможности за ***създаване и управление на тестово съдържание***. Това включва поддръжка на различни типове тестови въпроси, като въпроси с множествен избор, отворени въпроси, въпроси за съвпадения и други. Съвременните решения следва да поддържат възможност за автоматизирано генериране на въпроси чрез използване на генеративни модели, като се задават критерии като тематика, ниво на трудност, когнитивно ниво и др.

Друго важно изискване е наличието на ***механизми за валидация и редактиране на генерираното съдържание***. Преподавателят трябва да има възможност да преглежда, коригира и одобрява тестовите въпроси преди използването им, за да осъществява контрол върху качеството им. Това е особено важно при използването на AI-базирани генеративни системи, където съществува риск от некоректно или неподходящо съдържание.

Друга ключова функционалност е ***управлението на тестове и тестови сесии***. Системата трябва да позволява конфигуриране на параметри като брой въпроси, ред на представяне на въпросите, времеви ограничения, условия за достъп и др. В този контекст е важно да се поддържат както статични тестове, така и адаптивни тестове, при които последователността от въпроси се определя динамично в зависимост от представянето на обучаемия.

Съвременните системи следва да осигуряват ***автоматично оценяване на отговорите***, както и възможност за ръчна проверка при необходимост. При въпроси със затворен отговор оценяването може да бъде напълно автоматизирано, докато при отворени въпроси е необходимо да се предвидят механизми за подпомагане на оценяването, включително чрез използване на изкуствен интелект.

Особено важно изискване е ***поддръжката на адаптивно тестване***. Това предполага използването на психометрични модели, като Item Response Theory, за динамично определяне на следващия въпрос въз основа на оценената способност на обучаемия. По този начин се постига по-прецизно и персонализирано оценяване с по-малък брой въпроси.

Друг важен аспект е надеждността на използваните AI модели. Различните модели могат да интерпретират една и съща задача по различен начин и да генерират съдържание с различно качество. Поради тази причина е целесъобразно в системата да бъде реализиран механизъм за ***многомоделна верификция*** (multi-model verification), при който въпросите, генерирани от един езиков модел, се оценяват от друг независим модел.

1.5. Изводи

Проведените изследвания показват, че генеративните модели могат успешно да се използват за създаване на тестови въпроси, генериране на учебни материали и предоставяне на обратна връзка към обучаемите. Въпреки това съществуват редица предизвикателства, свързани с качеството, надеждността и педагогическата пригодност на генерираното съдържание.

Въз основа на направения анализ са формулирани целта и задачите на настоящата дисертация, представени в увода.

Изпълнението на поставените задачи ще създаде предпоставки за изграждане на надеждна и интелигентна среда за електронно тестване, която подпомага преподавателите при създаването на тестово съдържание и осигурява по-ефективно, обективно и персонализирано оценяване на обучаемите.

ГЛАВА 2. КОНЦЕПТУАЛЕН МОДЕЛ НА ПЛАТФОРМА ЗА Е-ТЕСТОВЕ

2.1. Основни роли и функционалности

Концептуалният модел на платформата за електронно тестване дефинира ***три основни потребителски роли***: администратор, преподавател и обучаем. Всяка роля има достъп до определени функционалности.

Администраторът отговаря за цялостното управление и поддръжка на системата. Неговите дейности включват управление на потребители, контрол на достъпа, конфигуриране на основни системни параметри, наблюдение на активността и анализ на логове. Той има ключова роля и при управлението на AI функционалностите – конфигурира използваните модели, задава политики и ограничения за тяхното използване, следи натоварването, ефективността и разходите, свързани с AI услугите. Освен това администраторът отговаря за сигурността, генерирането на отчети, както и за архивирането и възстановяването на данни.

Преподавателят управлява основните процеси, свързани със създаването, провеждането и анализа на тестовете. Той създава тестови въпроси ръчно или чрез AI-базирано генериране, след което ги преглежда, коригира, валидира и одобрява. В рамките на тестовия процес преподавателят конфигурира тестове, определя тяхната структура и съдържание, назначава ги на обучаемите и следи изпълнението им. След приключване на тестовете той анализира резултатите чрез инструментите за оценка и статистика и предоставя обратна връзка. Изкуственият интелект подпомага генерирането и проверката на въпроси, анализа на резултатите и формулирането на персонализирана обратна връзка.

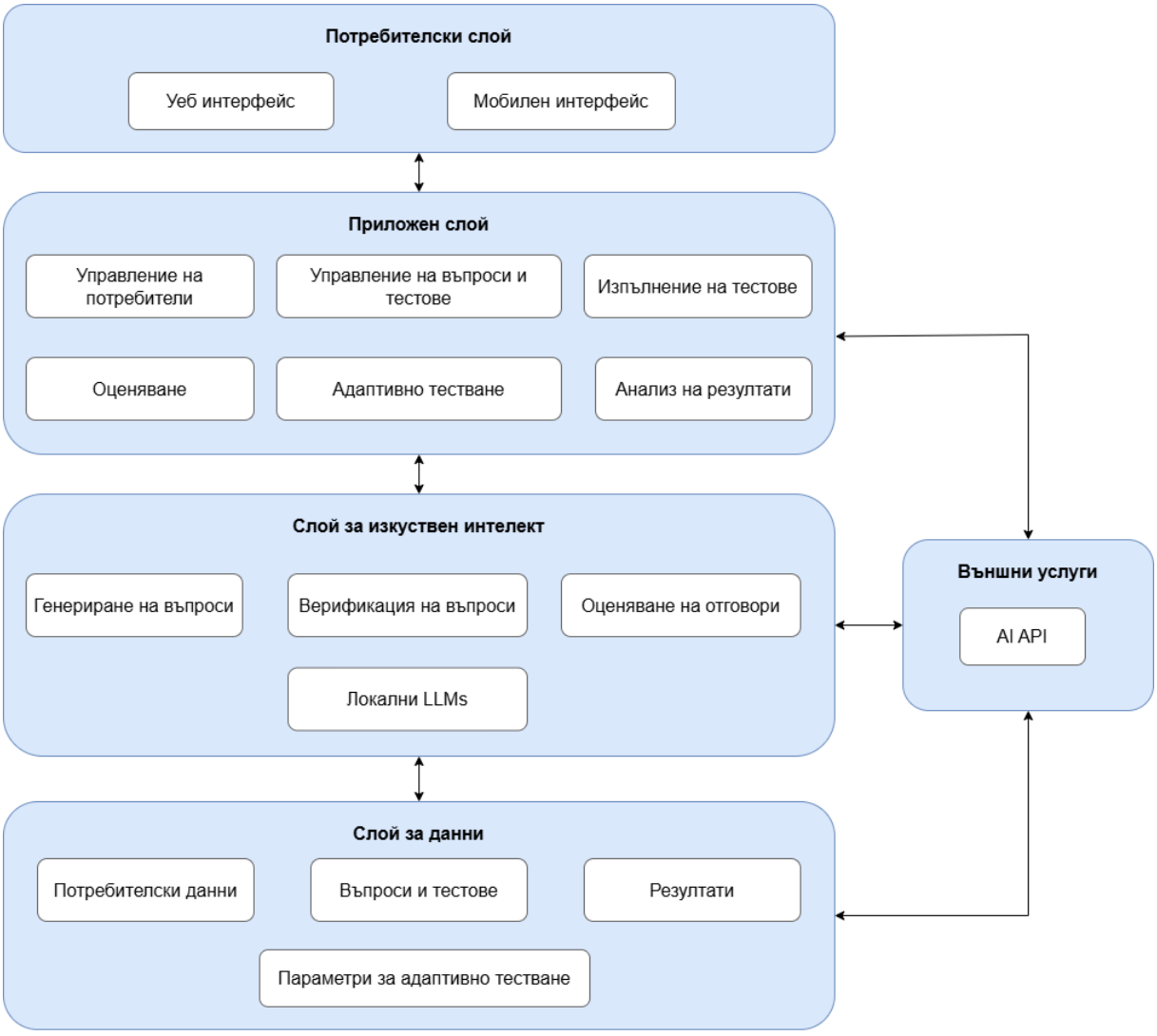
Обучаемият е основният потребител на системата за електронно тестване. Той участва в тестови сесии, назначени от преподавателя, както и в адаптивни тестове за

самообучение. При адаптивното тестване системата генерира последователност от въпроси, съобразена с текущото ниво на знания на обучаемия, което подпомага по-прецизното оценяване. След приключване на тестовете обучаемият има достъп до своите резултати, история на тестовете и статистики за представянето. Системата може да предоставя AI-базирана обратна връзка, включваща обяснения на правилните отговори, анализ на грешките и препоръки за подобрение.

2.2. Обща архитектура

Платформата за електронно тестване се базира на многослойна архитектура. Тя осигурява ясно разделение на отговорностите, гъвкавост и възможност за мащабиране. В общия случай архитектурата може да бъде разгледана като съставена от четири основни слоя (фиг. 2). Това включва:

- потребителски слой (User Layer);
- приложен слой (Application Layer);
- слой за изкуствен интелект (AI Layer);
- слой за данни (Data Layer).



Фигура 2. Архитектура на системата

Потребителският слой осигурява интерфейс за взаимодействие със системата чрез уеб или мобилни приложения. Той предоставя достъп до функционалностите на платформата в зависимост от ролята на потребителя, като включва визуализация на тестове, инструменти за създаване и конфигуриране на въпроси, представяне на резултати и анализи и др. Основна цел на този слой е да осигури интуитивен и ефективен достъп до системата, независимо от техническата подготовка на потребителите.

Приложният слой реализира основната бизнес логика на системата и координира взаимодействието между останалите слоеве. Той включва модули за управление на потребители, управление на въпроси и тестове, изпълнение на тестови сесии, оценяване и анализ на резултатите. В този слой се реализират и механизмите за адаптивно тестване, които използват текущите резултати на обучаемия за динамично определяне на следващите въпроси. Приложният слой играе ролята на посредник между потребителския интерфейс и интелигентните услуги, като осигурява последователност и контрол върху процесите.

Слоят за изкуствен интелект представлява ключов компонент в архитектурата, който осигурява интелигентността на системата. Той включва интеграция с големи езикови модели, които се използват за автоматизирано генериране на тестови въпроси, както и за тяхната верификация и оценка. В този слой се реализират механизми за управление на заявки (prompt engineering), обработка на резултатите в структуриран формат (например JSON) и комбиниране на резултатите от различни модели с цел повишаване на надеждността. Чрез абстракция на AI функционалностите се осигурява възможност за лесна подмяна или добавяне на нови модели, без да е необходимо да се променят останалите части на системата.

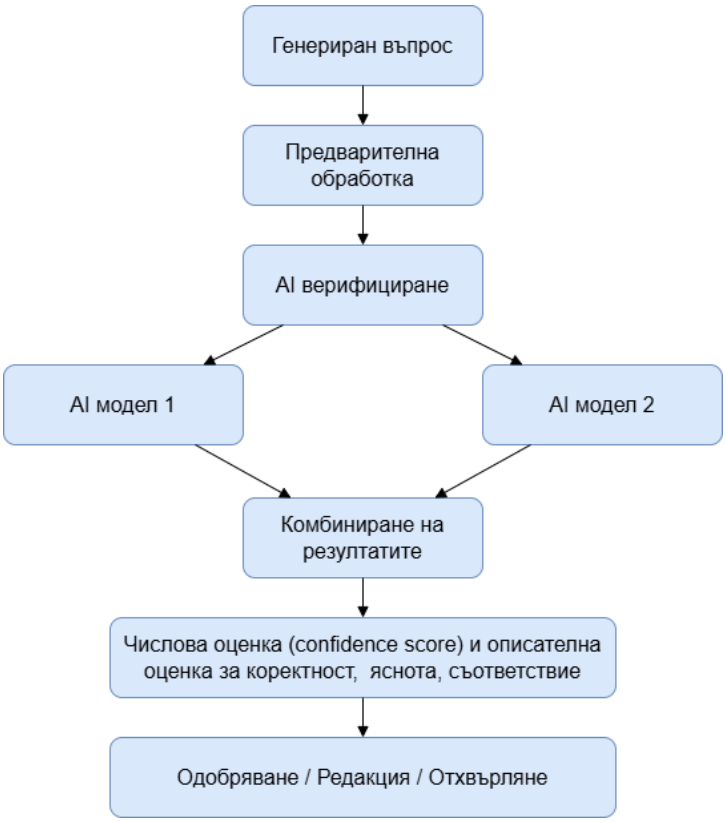
Слоят за данни осигурява съхранението и управлението на всички данни в системата. Той включва база данни за потребители, въпроси, тестове, резултати от тестове и параметри, свързани с адаптивното оценяване. Данните се използват както за текущото функциониране на системата, така и за последващ анализ и подобряване на качеството на тестовете. В допълнение, този слой може да поддържа съхранение на исторически данни, които са необходими за калибриране на параметрите при адаптивното тестване.

2.3. Основни модули

Модулът за генериране на въпроси използва големи езикови модели за автоматизирано създаване на тестово съдържание. Входните данни към модула включват критерии, зададени от преподавателя, като тематика, ниво на трудност, когнитивно ниво по таксономията на Bloom, тип на въпросите и допълнителни инструкции. На базата на тези параметри се конструира заявка към езиковия модел, който генерира структурирани въпроси и отговори. Резултатът се представя в стандартизиран формат, което позволява последваща обработка и интеграция със системата.

С оглед на повишаване на надеждността на генерираното съдържание, системата включва ***модул за верификация***, който използва допълнителен AI модел за оценка на качеството на въпросите. Този модул анализира коректността, яснотата, съответствието с поставените критерии и потенциални проблеми като двусмисленост или грешки.

Възможно е използването на повече от един модел (многомоделна верификация), като резултатите се комбинират за постигане на по-висока точност. Модулът може да предоставя и оценка на доверие (confidence score), която подпомага преподавателя при вземане на решения (фиг. 3).



Фигура 3. Процес на верификация и оценка на тестови въпроси

Модулът за управление на въпроси и тестове осигурява функционалности за създаване, редакция, съхранение и организиране на въпроси и тестове. Преподавателят може да избира въпроси от база данни или да използва генерираните от AI въпроси, като ги комбинира в тестове по различни начини. Модулът поддържа конфигуриране на параметри като времеви ограничения, ниво на сложност, брой въпроси, критерии за оценяване. Осигурява се и възможност за повторна употреба на въпроси, което повишава ефективността на процеса.

Модулът за адаптивно тестване реализира механизми за динамичен подбор на въпроси, съобразени с текущото ниво на знания на обучаемия. За реализиране на адаптивността се използва ***Item Response Theory (IRT)*** – психометричен модел, описващ зависимостта между латентната способност на обучаемия и вероятността за правилен отговор на конкретен тестов въпрос.

Всеки въпрос се характеризира с набор от психометрични параметри, като ***дискриминация, трудност*** и ***вероятност за случайно познаване***. На базата на тези параметри и текущата оценка на способността на обучаемия (θ) се изчисляват вероятността за правилен отговор и информационната стойност на всеки въпрос от тестовата банка. Следващият въпрос се избира така, че да предостави максимална

информация за текущата оценка на способността, което позволява постепенно намаляване на стандартната грешка на измерването (SE) и повишаване на точността на крайната оценка (Embretson & Reise, 2025; Baker & Kim, 2004; Wyse, 2010; Van der Linden & Glas, 2010; Wainer et al., 2000).

Съществуват различни IRT модели – 1PL, 2PL и 3PL. Основната разлика между тях е в броя на параметрите (характеристиките на въпроса), които моделът отчита.

Вероятността $P(X = 1 \mid \theta)$ даден обучаем с латентна способност θ да отговори правилно на въпрос, при различните IRT модели се изчислява чрез следните функции:

1PL Модел: $P(X = 1 \mid \theta) = \frac{1}{1+e^{-(\theta-b)}}$

2PL Модел: $P(X = 1 \mid \theta) = \frac{1}{1+e^{-a(\theta-b)}}$

3PL Модел: $P(X = 1 \mid \theta) = c + \frac{1-c}{1+e^{-a(\theta-b)}}$,

където ***a*** е ***дискриминация***, **b** е ***трудност***, и ***c*** е ***вероятност за случайно познаване*** на въпроса.

Адаптивни тестове е подходящо да се използват за самооценка, персонализирано обучение и експериментални изследвания.

Модулът за оценяване обработва отговорите на обучаемите и изчислява резултатите. При въпроси със затворен отговор оценяването се извършва автоматично, докато при отворени въпроси могат да се използват AI-базирани методи за подпомагане на оценяването или ръчна проверка от преподавателя. В контекста на адаптивното тестване, модулът изчислява и актуализира стойността на латентната способност на обучаемия.

Модулът за анализ предоставя инструменти за обработка и визуализация на резултатите от тестовете. Той генерира статистики за представянето на обучаемите, аналитични параметри на въпросите (например трудност и дискриминация), както и агрегирани отчети за преподаватели и администратори. Данните могат да се използват за подобряване на качеството на тестовете и оптимизация на учебния процес.

Модулът за управление на изкуствен интелект играе ролята на посредник между приложния слой и външните AI услуги. Той управлява комуникацията с езиковите модели, обработва входните и изходните данни и осигурява стандартизация на заявките чрез конструиране на унифициран промпт с помощта на указаните от преподавателя параметри. Освен това, модулът може да реализира механизми за избор на модел и проследяване на качеството на генерираните резултати.

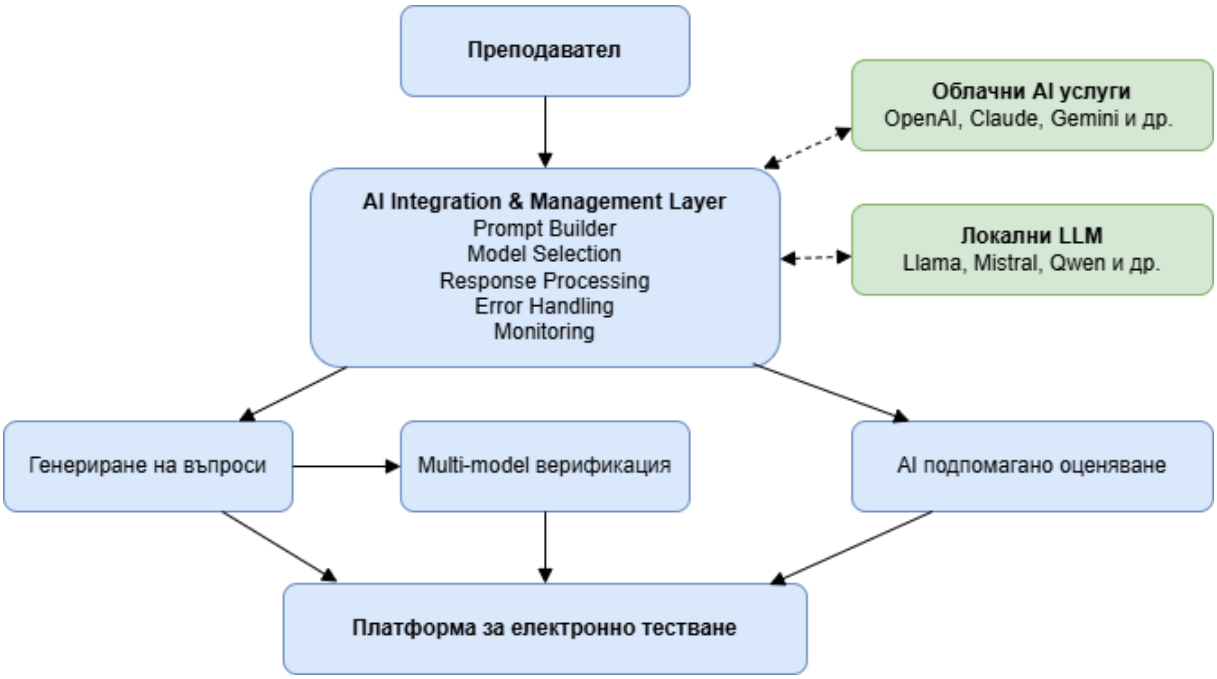
2.4. Случаи на употреба

Основни случаи на употреба, които илюстрират взаимодействието между различните роли и модули в системата, са:

- Създаване (генериране и верификация) на тестови въпроси;
- Създаване и конфигуриране на тест;
- Изпълнение на тест от обучаем;
- Оценяване и предоставяне на обратна връзка;
- Анализ на резултатите и оптимизация.

2.5. Интеграция на изкуствен интелект

В предложения модел (фиг. 4) взаимодействието с AI услуги се реализира чрез специализиран междинен слой (AI Integration & Management Layer), който осигурява унифициран интерфейс между приложната логика на системата и използваните езикови модели. Този слой отговаря за конструирането на заявки (prompts) въз основа на зададените от потребителя параметри, избора на подходящ модел, обработката и стандартизирането на получените резултати, както и за управлението на комуникацията с външните AI услуги.



Фигура 4. Архитектура на интеграцията на изкуствен интелект

Чрез тази архитектура различните функционалности, свързани с изкуствения интелект – генериране на тестови въпроси, верификация чрез множество модели и подпомагане на оценяването – използват общ механизъм за взаимодействие с LLM, без да е необходимо приложните модули да бъдат обвързани с конкретен доставчик или технология.

За генериране на тестови въпроси, LLMs създават съдържание въз основа на предварително дефинирани параметри, зададени от преподавателя. Те могат да включват тема, ниво на трудност, когнитивно ниво по таксономията на Bloom, тип на въпросите и допълнителни инструкции относно начина на формулиране.

Взаимодействието с LLM се осъществява чрез специализирани заявки (prompts), които се конструират динамично от AI Integration & Management Layer. Получените резултати се обработват, верифицират и преобразуват в унифициран структуриран формат (например JSON), което позволява тяхното автоматично интегриране в останалите компоненти на платформата.

Получените резултати се предоставят на преподавателя под формата на препоръки и индикатори за надеждност, като окончателното решение за приемане, редактиране или отхвърляне на даден въпрос е изцяло отговорност на преподавателя.

Изкуственият интелект може да бъде използван и за подпомагане на процеса на оценяване, особено при въпроси с отворен отговор. LLM могат да анализират текстови отговори, да ги сравняват с очаквани решения и да предоставят предварителна оценка или препоръки. Това позволява частична автоматизация на оценяването и намалява времето, необходимо за проверка на тестове.

Въпреки това, прилагането на AI в оценяването изисква внимателен контрол, като се предвижда възможност за ръчна намеса и окончателна оценка от преподавателя. По този начин се съчетава автоматизацията с човешка експертиза, което гарантира по-висока надеждност.

2.6. Изводи

Предложеният концептуален модел на платформа за електронно тестване от ново поколение има за цел да интегрира възможностите на генеративния изкуствен интелект и адаптивните методи за оценяване в единна, модулна и разширяема архитектура. Моделът надгражда традиционните системи за електронно тестване, като осигурява по-висока степен на автоматизация, гъвкавост и персонализация на процеса на оценяване.

ГЛАВА 3. СОФТУЕРНА РЕАЛИЗАЦИЯ

3.1. Технологичен стек

За разработката на платформа са използвани множество технологии, които според употребата си са групирани в пет категории: backend, бази данни, frontend, доставчици на изкуствен интелект и инфраструктура. Използваният фреймуърк е .NET 10.0. Технологичният стек е илюстриран на фиг. 5.

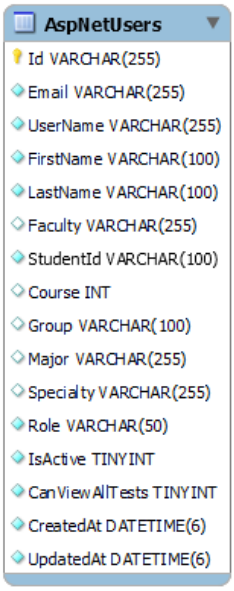
Фигура 5. Технологичен стек на приложението

3.2. Модел на данните

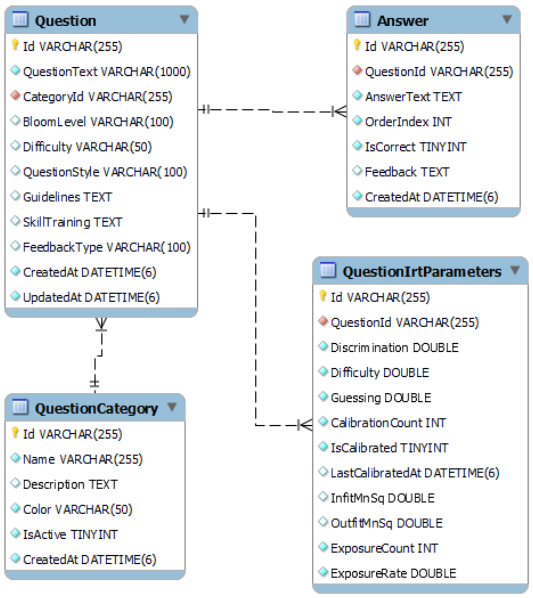
Базата данни съдържа набор от основни и служебни таблици, които поддържат управлението на потребители, въпроси, тестове, адаптивно оценяване, AI генериране, верификация и системна конфигурация. За визуализиране на структурата на базата данни и връзките между отделните таблици е използван инструментът MySQL Workbench. Поради големия брой таблици и референции моделът е разделен на няколко логически групи. С цел по-добра четимост и яснота на представянето, връзките между отделните групи в следващите диаграми не са показани.

User (фиг. 6) е централен модел за потребителите на системата, вкл. администратори, преподаватели и обучаеми. Той разширява стандартния *IdentityUser*, като добавя полета за име, фамилия, факултет, факултетен номер (уникален), курс,

група, специалност, роля, статус на активност и допълнително право за преглед на всички тестове.



Фигура 6. Модел User



Фигура 7. Моделът Question

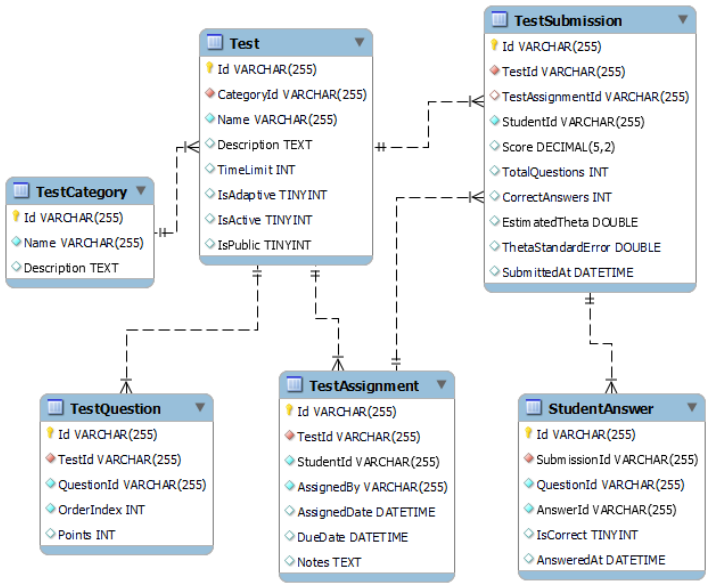
Моделът *Question* (фиг. 7) представлява тестови въпрос. Той има следните характеристики: текст (до 1000 символа), категория, ниво по таксономията на Блум, трудност, стил и допълнителни насоки. Всеки въпрос има множество *Answer* записи с текст, ред и маркер за верен отговор, както и един *QuestionIrtParameters* запис с IRT параметри: дискриминация (a), трудност (b), отгатване (c), *Infit* и *Outfit MnSq* статистики и флаг за калибриране.

Моделът *Test* (фиг. 8) описва тест. Класическите и адаптивните тестове имат следните характеристики – заглавие, описание, категория, създател, времево ограничение, флагове за активност и публичност. За адаптивните тестове с допълнителни полета се указват – максимален брой въпроси, праг на стандартната грешка за спиране на теста и категория на пула от въпроси. Връзката между тест и въпроси се осъществява чрез *TestQuestion* с подредба.

TestAssignment (фиг. 8) моделира назначаването на тест. Характеристиките определят назначаващия теста преподавател, обучаемите, към които е насочен теста, опционален краен срок за изпълнение на теста и бележки. *TestSubmission* записва решението на студента, вкл. следната информация: оценка (процент), брой общи и брой верни отговори, статус, време за което теста е изпълнен, оценка, въведена ръчно от преподавателя (по шестобалната скала), коментар, θ оценка от IRT и стандартна грешка. *StudentAnswer* съхранява отговорите, които е дал обучаемия за всеки отделен въпрос.

Моделът *AdaptiveTestSession* следи състоянието на адаптивен тест в процеса на решаването му от обучаемия. Моделът *SystemSetting* реализира механизъм за съхранение на конфигурационни параметри по схемата ключ-стойност. Чрез него се управляват глобални настройки на платформата, като конфигурация на AI доставчиците, параметри на адаптивното тестване, прагове за валидация и др.

Моделите *GenerationHistory*, *QuestionValidationResult* и *ErrorLog* се използват за съхранение на информация за AI генерирането и верификацията на въпроси, както и за регистриране на възникнали системни грешки и диагностични събития, подпомагащи мониторинга и поддръжката на платформата.



Фигура 8. Модели Test и TestAssignment

3.3. Основни функционалности

Преподавателите и администраторите могат да създават тестови въпроси по два начина – чрез ръчно въвеждане или генериране с помощта на ИИ.

Ръчното създаване се извършва на две стъпки. На *първата стъпка* преподавателят трябва да въведе текст на въпроса и мета информация, вкл. категория на въпроса, сложност (Лесно, Средно, Трудно, Много трудно), тип на въпроса („множествен избор“, „вярно/грешно“, „есе“, „кратък отговор“) и точки. На *втора стъпка* се въвежда допълнителна информация за въпроса, в зависимост от неговия тип и ниво по таксономията на Блум (Знание, Разбиране, Прилагане, Анализ, Синтез, Оценяване). Напр. ако е избран тип „множествен отговор“, на втора стъпка трябва да се добавят възможните отговори, като точно един се маркира като верен.

При *генериране на въпрос с ИИ*, освен категория, ниво на сложност, тип на въпросите и ниво по Блум, преподавателят задължително трябва да укаже:

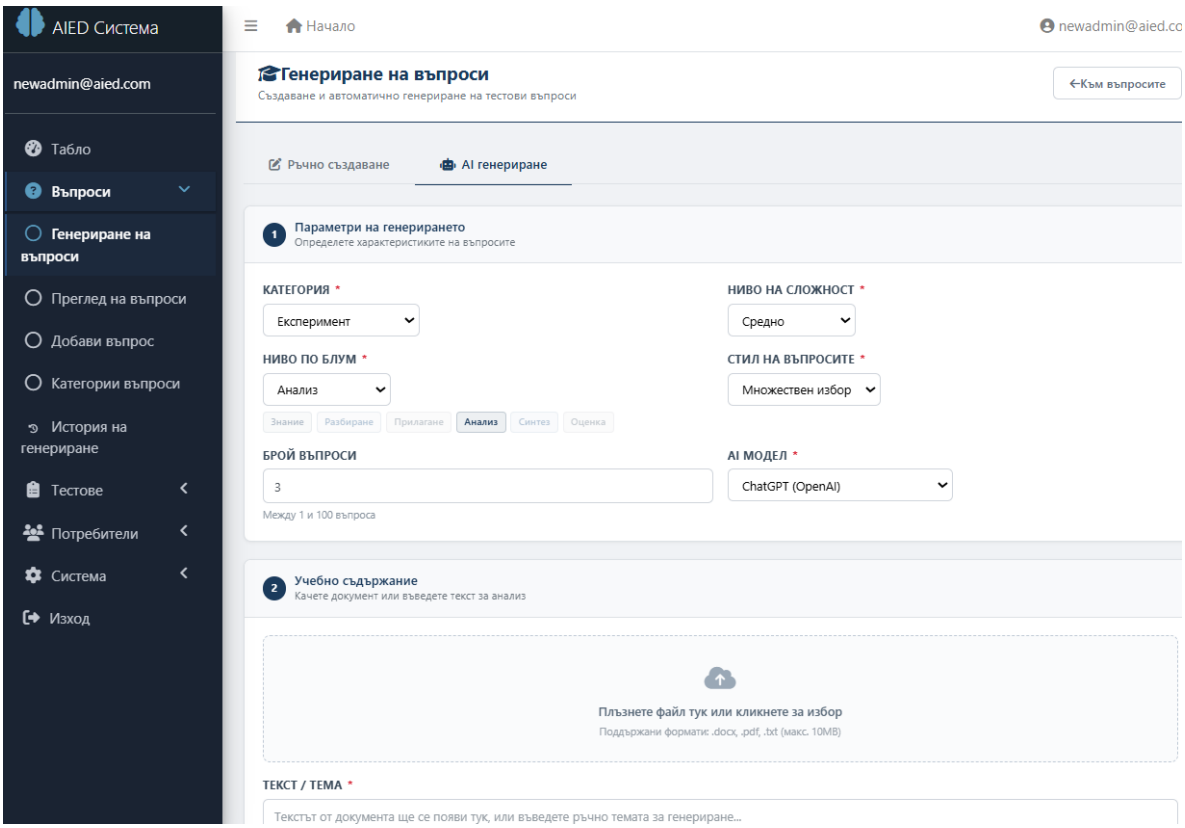
- учебно съдържание (файл, текст или заглавие на тема), на база на което да се генерират въпроси;
- AI модел, който да се използва. Към настоящия момент се поддържат опции за локален модел deepseek-r1:8b и облачните GPT-5.4 mini от Open AI, Gemini 2.5 Pro от Google и Claude Sonnet 4.6 от Anthrophic.

Опционално могат да се зададат допълнителни параметри, вкл.

- допълнителни указания за въпросите, напр. фокус върху практически примери, числови задачи и т. н.

- втори генеративен модел, който да се използва за оценка на генерираните тестови въпроси.

Тези опции са илюстрирани на фиг. 9.



Фигура 9. Екран за автоматично генериране на тестови въпроси от LLM

След натискане на бутон „Генерирай въпроси“, системата изпраща промпт на български език към избрания доставчик на ИИ и обработва JSON отговора.

Ако преподавателят е заявил оценяване на въпросите с генеративен модел, под всеки въпроса има кратка оценка и линк „Анкетна карта за оценяване“, съдържаща съответните оценки.

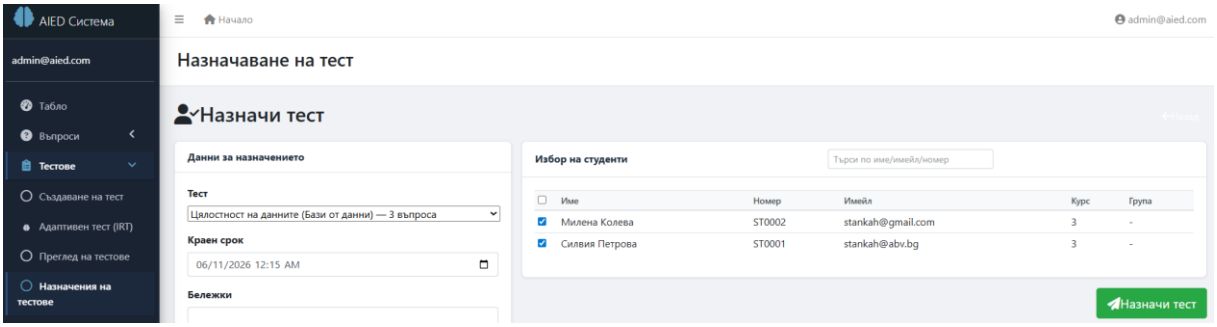
Всеки един от генерираните въпроси може да бъде опционално редактиран от преподавателя и записан в базата данни – в конкретна категория въпроси.

Платформата поддържа ***два типа тестове*** – класически и адаптивни.

Класическите тестове се съставят чрез избор на въпроси от предварително създадените. При конфигурирането им се задават име на теста, описание, категория, ниво на сложност и времетраене на теста. Тестът може да бъде със статут *активен* (за оценяване) или с *публичен достъп* (за самообучение).

Адаптивните тестове са базирани на IRT и работят с динамичен подбор на въпроси от пул. При конфигуриране на такъв тест се задават име, описание, категория на теста, пул от въпроси (категория), максимален брой въпроси (от 5 до 100), времетраене на теста и спиращ праг на стандартната грешка (от 0.1 до 1.0).

Преподавателят може да назначава тест на студенти, които избира от списъка с регистрираните в системата студенти. Опционално може да зададе краен срок за изпълнение на теста и пояснения към него (фиг. 10).



Фигура 10. Назначаване на тест

При ***класически тест***, студентът вижда всички въпроси наведнъж, избира отговори и накрая предава теста. Оценката се изчислява автоматично на база процента на верните отговори. След предаването на теста, чрез SignalR се изпраща уведомление, което актуализира екрана за мониторинг на преподавателя.

При стартиране на ***адаптивен тест*** системата създава нова тестова сесия и инициализира началната оценка на способността на обучаемия (θ). Ако за обучаемия съществува предходна оценка, тя се използва като начална стойност; в противен случай θ се инициализира със стойност 0.0, съответстваща на средното ниво на способност. След всеки даден отговор се актуализират използваните параметри. Тестът завършва при:

- достигане на целевата стандартна грешка (спиращия SE праг);
- достигане на максималния брой въпроси;
- изчерпване на въпросите от пула;
- изтичане на времето за теста.

При финализиране на теста, в базата данни се създава запис за решението, включително стойностите на θ, стандартна грешка и процентна оценка на верните отговори.

Автоматичното оценяване изчислява оценката като процент на верните отговори. Оценяването се извършва по използваната в България шестобална скала, като са приети следните критерии за оценка: Отличен 6 (90–100%), Много добър 5 (75–89%), Добър 4 (60–74%), Среден 3 (45–59%) и Слаб 2 (0–44%).

Преподавателят може да задава ръчно оценки, които имат приоритет над автоматично изчислените, и да добавя индивидуални коментари за обратна връзка.

Платформата предоставя ***няколко вида отчети***: общ отчет с броя тестове, въпроси, студенти и средна оценка; анализ по тест с разпределение на оценки; индивидуално представяне на студент с решени тестове и тенденция; IRT анализ с калибрирани параметри и средно θ; и детайлен IRT отчет по въпрос с дискриминация, трудност, отгатване, Infit/Outfit MnSq и ICC крива. Общите отчети са достъпни за администратори и преподаватели, а IRT анализът – само за администратори.

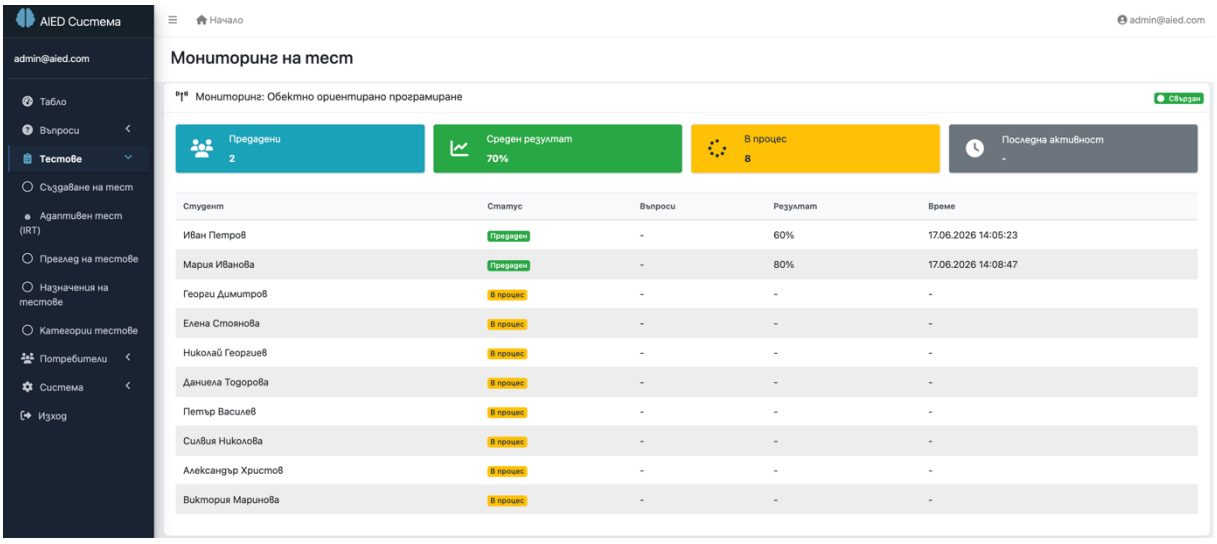
3.4. Real-time мониторинг със SignalR

Платформата поддържа мониторинг в реално време, чрез който преподавателят проследява решаването на тестовете и напредъка на студентите. За двупосочна комуникация между сървъра и клиентските приложения се използва SignalR, който при възможност работи чрез WebSocket, а при необходимост автоматично преминава към алтернативни механизми като Long Polling или Server-Sent Events.

При отваряне на страницата за мониторинг клиентското приложение се свързва със SignalR Hub и се присъединява към група, свързана с конкретния тест. Така събитията за даден тест се изпращат само до преподавателите, които го наблюдават. При предаване на тест или отговор в адаптивен тест сървърът изпраща събития с данни за студента, резултата, момента на предаване, текущата оценка на способността θ, стандартната грешка и броя отговорени въпроси.

Интерфейсът визуализира обобщена и детайлна информация, включително броя предадени тестове, средния резултат, броя активни студенти и данни за всеки участник. Информацията се обновява автоматично без презареждане на страницата. За по-висока надеждност е предвидено автоматично възстановяване на връзката при временни прекъсвания, а текущият ѝ статус се показва чрез цветен индикатор.

На фиг. 11 е показан екран за мониторинг на тест в реално време чрез SignalR. В горната част се визуализират основните показатели за текущото състояние на теста, а в централната част – списък с обучаемите, техния статус, постигнати точки и последна активност.



Фигура 11. Мониторинг на тест по време на изпълнение чрез SignalR

Основното предимство на този подход е автоматичното актуализиране на данните при настъпване на събития от сървъра, което позволява на преподавателя да следи напредъка на обучаемите и своевременно да открива технически проблеми или необичайно забавяне.

3.5. Изводи

Представената в настоящата глава интелигентна платформа за електронно тестване реализира основните функционалности на предложения в Глава 2 концептуален модел и потвърждава неговата практическа приложимост.

ГЛАВА 4. ЕКСПЕРИМЕНТИ

За оценка на възможностите на генеративния изкуствен интелект за автоматизирано генериране и оценяване на тестови въпроси е проведено пилотно експериментално изследване. В рамките на изследването е анализирано поведението на два големи езикови модела, използвани в две различни роли – като ***генератори на тестови въпроси (GenModel)*** и като ***оценители на тяхното качество (EvalModel)***.

4.1. Цели на експеримента и изследователска рамка

Проведеното експериментално изследване има характер на ***пилотно изследване*** и е насочено към експериментална проверка на предложения модел и разработения софтуерен прототип за автоматизирано генериране и оценяване на тестови въпроси чрез генеративен изкуствен интелект. Конкретните цели на експеримента са:

1. Да се оцени ***способността на GenAI моделите автоматично да генерират коректни и педагогически адекватни тестови въпроси***;
2. Да се анализира ***способността на GenAI моделите автоматично да оценяват качеството на тестови въпроси***, генерирани както от същия, така и от различен генеративен модел;
3. Да се изследва ***влиянието на избора на езиков модел върху качеството на генерираните въпроси*** и надеждността на тяхното оценяване;
4. Да се демонстрира ***приложимостта на предложения модел*** за многомоделно генериране и оценяване като основен компонент на разработената интелигентна платформа за електронно тестване.

Използваният набор от ***72 тестови въпроса*** не е формиран произволно, а е подбран така, че да обхваща различни нива на трудност и когнитивни равнища съгласно таксономията на Блум, като осигурява достатъчна основа за експериментална проверка на предложените механизми за генериране и оценяване.

В съответствие с целите и обхвата на пилотното изследване са формулирани изследователски въпроси, насочени към експериментална оценка на приложимостта на големите езикови модели при автоматизираното генериране и оценяване на тестови въпроси:

- **RQ1:** До каква степен LLM моделите могат автоматично да генерират коректни, педагогически адекватни и релевантни тестови въпроси?
- **RQ2:** До каква степен LLM моделите могат надеждно да оценяват качеството на автоматично генерирани тестови въпроси?
- **RQ3:** Наблюдават ли се различия в представянето на различните LLM модели при генериране и оценяване на тестови въпроси?
- **RQ4:** Как влияе изборът на модел за генериране (GenModel) и модел за оценяване (EvalModel) върху качеството на получените резултати?

4.2. Експериментален дизайн

Експериментът включва четири основни етапа:

- **EP1**: Генериране на тестови въпроси от LLM модели;
- **EP2:** Оценяване на генерираните тестови въпроси от LLM;
- **EP3:** Оценяване на генерираните тестови въпроси от експерти-оценители;
- **EP4:** Обработка, анализ и интерпретиране на получените резултати.

Първият етап на експеримента е свързан с ***генериране на тестови въпроси от 2 избрани LLM модела – GPT-5.4 mini и Claude Sonnet 4.6***.

През втория етап на експеримента се извършва ***оценяване на тестовите въпроси, но от LLM модели в ролята на автоматизирани оценители***. Моделите оценяват същите въпроси по утвърдена оценъчна карта, без информация за произхода на въпросите.

Третият етап има за цел ***оценяване на генерираните тестови въпроси от експерт-оценител***. Получените експертни оценки се използват като референтен стандарт, спрямо който се оценява надеждността и валидността на генерираните въпроси.

Четвъртият етап е насочен към ***обработка, анализа и интерпретиране на данните***, събрани в предходните две експериментални фази.

Експериментът е проведен в рамките на една учебна дисциплина от областта на компютърните науки – „Бази от данни“.

В експеримента са използвани ***два големи езикови модела – GPT-5.4 mini и Claude Sonnet 4.6***, които изпълняват ясно разграничени функционални роли в рамките на експерименталния дизайн. В зависимост от конкретния сценарий, всеки модел може да участва като:

- ***GenModel*** – модел, използван за автоматично генериране на тестови въпроси;
- ***EvalModel*** – модел, използван за оценяване на качеството на тестовите въпроси въз основа на предварително дефинирани критерии.

4.3. Процедури за генериране и оценяване на тестови въпроси

За генерирането на тестовите въпроси се задават ***множество параметри***, които отразяват реални сценарии от практиката на преподаване и оценяване във висшето образование.

С всеки един от избраните LLM модели са генерирани *идентичен брой въпроси* – по 18, чрез един и същи prompt, автоматично генериран от платформата на база избраните за експеримента параметри.

В рамките на този етап от експеримента, в ролята на автоматизирани оценители се използват големи езикови модели. Основната цел е анализ на тяхната способност да оценяват качеството на тестови въпроси по начин, съпоставим с човешката експертна оценка. За всеки тестов въпрос, оценъчният модел (EvalModel) получава следната информация:

- *текста на тестовия въпрос*, без информация за модела, който го е генерирал, с цел предотвратяване на пристрастия;
- *правилния отговор*, предоставен като референтна информация, необходима за коректна оценка на научната вярност и еднозначността на въпроса;

- *оценъчната карта*, използвана при човешката експертна оценка, включваща предварително дефинираните 5 критерии и 10 индикатори;
- *петстепенната ликертова скала*, използвана за оценяване от експерт-оценителите;
- *ясни инструкции да не пренаписват, редактират или подобряват въпроса,* а единствено да извършват оценяване по зададените индикатори.

За оценяване на качеството на генерираните от изкуствен интелект въпроси, е създадена оценъчна карта. Тя съдържа 5 критерия за оценяване, всеки от които има по 2 индикатора (табл. 1).

Таблица 1. Оценъчна карта за генерираните от LLM въпроси

Критерий за оценяване **/** Индикатор	**Тегло**
I. Надеждност и научна коректност	**15%**
1. Въпросът е научно коректен и не съдържа фактологични грешки.	
2. Правилният отговор е еднозначен и не допуска алтернативни интерпретации.	**10%**
II. Съответствие с учебното съдържание	**10%**
3. Въпросът е пряко свързан с предоставеното учебно съдържание (текст или файл).	
4. Въпросът не изисква знания извън обхвата на подадения учебен материал.	**10%**
III. Адекватност на нивото на сложност	**10%**
5. Реалната сложност на въпроса съответства на зададеното ниво (лесно, средно, трудно, много трудно).	
6. Сложността на въпроса произтича от съдържанието, а не от неясна формулировка.	**10%**
IV. Съответствие с когнитивното ниво по Bloom	**10%**
7. Въпросът реално измерва зададеното когнитивно ниво по таксономията на Bloom.	
8. Типът мисловна дейност, изисквана от въпроса, съответства на избраното ниво (напр. анализ, синтез).	**10%**
V. Педагогическа адекватност и приложимост	**5%**
9. Формулировката на въпроса е ясна, недвусмислена и граматически коректна.	
10. Въпросът е педагогически адекватен и може да бъде използван директно в реален тест.	**10%**

Експерименталният дизайн включва ***четири сценария*** (табл. 2), които комбинират различни големи езикови модели в ролите на генератор на тестови въпроси (GenAI) и оценител на тяхното качество (EvalAI). Целта на тези сценарии е да се изследва както индивидуалното поведение на отделните модели, така и ефектът от тяхното взаимодействие в рамките на процеса по генериране и оценяване.

Сценарий S1 комбинира GPT като генератор на тестови въпроси и Claude като оценител. В ***сценарий S2*** един и същ LLM модел – GPT се използва както за генериране, така и за оценяване на тестовите въпроси. В ***сценарий S3*** ролите са разменени, като Claude изпълнява функцията на генератор, а GPT – на оценител. ***Сценарий S4*** включва Claude като генератор и като оценител.

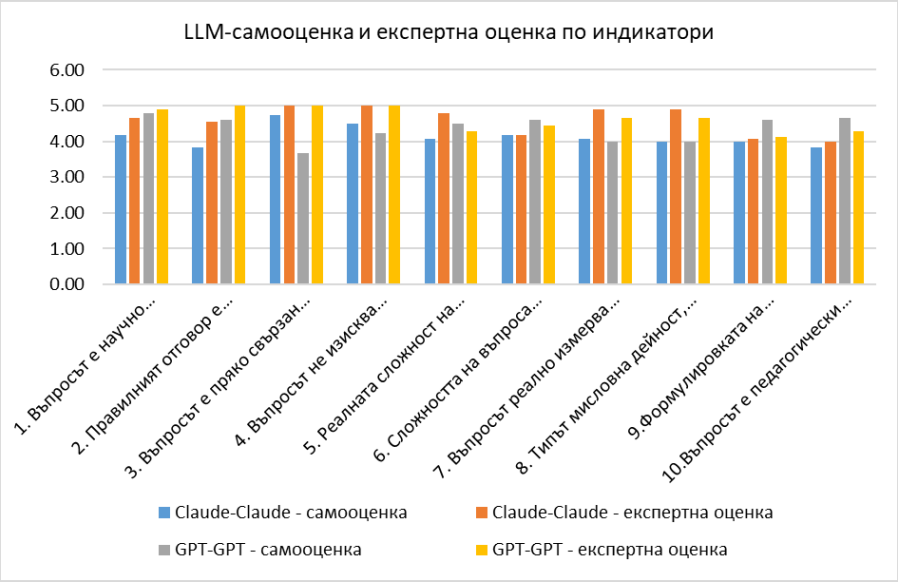
Таблица 2. Сценарии за оценяване LLM-LLM

Сценарий	GenAI	EvalAI
S1	GPT	Claude
S2	GPT	GPT
S3	Claude	GPT
S4	Claude	Claude

4.4. Анализ на резултатите

Резултатите от оценяването на генерираните въпроси по разгледаните сценарии са обобщени и анализирани чрез йерархичен подход, при който оценките се агрегират последователно на три нива. Първоначално се изчисляват оценки по отделните индикатори, после те се използват за формиране на оценки по критерии чрез прилагане на зададените тегла, и накрая се извежда крайната претеглена оценка за всеки сценарий.

На фиг. 12 са представени резултатите за сценариите, при които даден LLM генерира въпроси, а оценяването се извършва от същия модел (самооценка) и от експерт. На фиг. 13 са показани резултатите за сценариите, при които въпросите са оценени от друг LLM и от експерт.

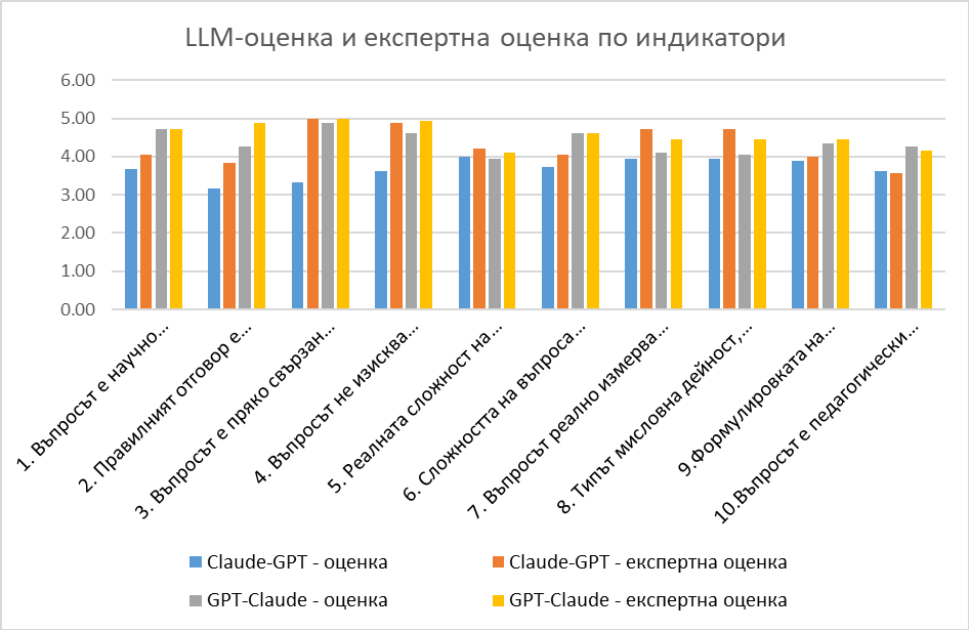


Фигура 12. LLM-самооценка и експертна оценка по индикатори

Като цяло всички средни оценки са над 3.0, а преобладаващата част от тях са в интервала между 4.0 и 5.0. Това показва, че независимо от използвания сценарий генерираните въпроси са с високо качество и удовлетворяват в значителна степен предварително зададените критерии. В повечето случаи експертните оценки са по-

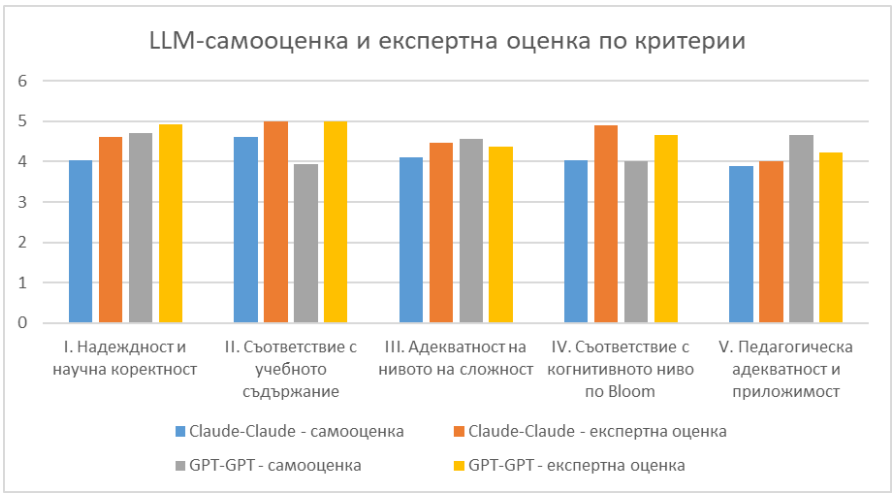
високи от оценките, поставени от LLM моделите. Това показва, че от гледна точка на човешкия експерт, генерираните въпроси са с приемливо високо качество и са подходящи за реална образователна употреба

Интересен резултат се наблюдава и при кръстосаното оценяване. Когато GPT оценява въпроси, генерирани от Claude (сценарий Claude-GPT), оценките са системно по-ниски от експертните. Обратно, когато Claude оценява въпроси, генерирани от GPT (сценарий GPT-Claude), разликите между оценките на модела и експерта са по-малки.



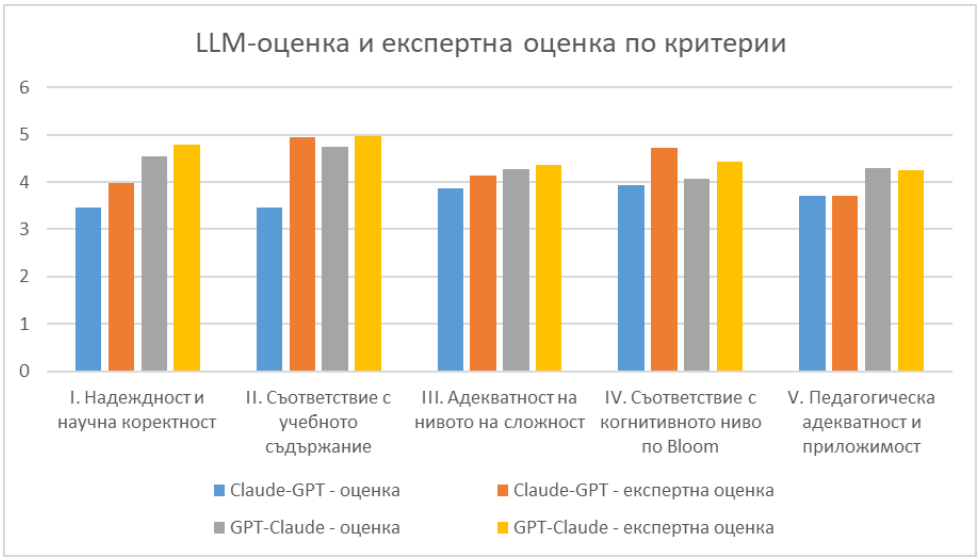
Фигура 13. LLM-оценка и експертна оценка по индикатори

За оценките по критерии се използват зададените тегла на индикаторите, съответните им оценки и нормализация в интервала [1, 5]. Получените резултати са представени на фиг. 14 и фиг. 15.



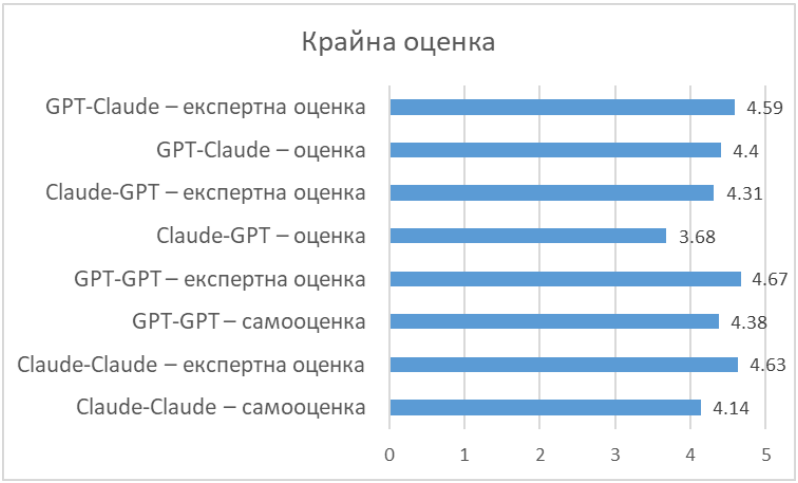
Фигура 14. LLM-самооценка и експертна оценка по критерии

Анализът на резултатите по критерии потвърждава наблюденията, направени при анализа по отделни индикатори. Високите оценки при всички сценарии показват, че генерираните въпроси са с високо качество. Наблюдава се също така тенденция експертните оценки да бъдат по-високи от оценките и самооценките на моделите.



Фигура 15. LLM-оценка и експертна оценка по критерии

Крайните претеглени оценки за всички разгледани сценарии са представени на фиг. 16.



Фигура 16. Крайна претеглена оценка за всички сценарии

Получените резултати показват високо общо качество на генерираните въпроси при всички сценарии като всички експертни оценки са над 4.30.

При самооценките и взаимните оценки на моделите се наблюдават по-големи разлики. Най-висока оценка, поставена от LLM, е отчетена при сценария GPT-Claude (4.40), следван от GPT-GPT – самооценка (4.38). Най-нисък резултат е получен при сценария Claude-GPT (3.68), а при сценария Claude-Claude резултатът е 4.14.

Като цяло резултатите показват, че както GPT, така и Claude са способни да генерират въпроси с високо качество, с леко предимство на GPT.

4.5 Изводи

Проведените експерименти демонстрират приложимостта и ефективността на предложения подход за използване на генеративен изкуствен интелект при създаването, верифицирането и оценяването на тестово съдържание. Получените резултати потвърждават, че големите езикови модели могат успешно да подпомагат процеса на разработване на тестови въпроси, като същевременно намаляват необходимото време и усилия от страна на преподавателите.

ЗАКЛЮЧЕНИЕ

Дигиталната трансформация на образованието и бързото развитие на изкуствения интелект създават нови възможности за усъвършенстване на процесите по обучение и оценяване. Съвременните образователни среди изискват инструменти, които не само автоматизират рутинни дейности, но и подпомагат преподавателите при създаването на качествено учебно съдържание и обективно оценяване на знанията на обучаемите. Генеративният изкуствен интелект и големите езикови модели предоставят значителен потенциал за автоматизирано генериране на тестови въпроси, анализ на резултати и персонализиране на процеса на оценяване. Настоящото дисертационно изследване е насочено към разработването на интелигентна платформа за електронно тестване, която интегрира възможностите на генеративния изкуствен интелект в единна среда за създаване, верифициране, управление и оценяване на тестово съдържание.

В рамките на дисертационното изследване са решени следните основни задачи:

- Проучени са съществуващи системи за електронно тестване и са идентифицирани основните функционални и нефункционални изисквания към интелигентна система за електронно тестване;
- Разработен е концептуален модел и архитектура на интелигентна платформа за електронно тестване, включваща механизми за автоматизирано генериране и верификация на тестови въпроси чрез големи езикови модели.
- Разработен е модул за автоматизирано адаптивно оценяване на обучаемите, базиран на съвременни психометрични подходи и изкуствен интелект;
- Реализиран е софтуерен прототип на предложената платформа;
- Проведени са експериментални изследвания за оценка на качеството и приложимостта на разработеното решение.

С изпълнението на поставените задачи е постигната основната цел на дисертационния труд – разработване на модел, архитектура и софтуерна реализация на интелигентна платформа за електронно тестване, използваща генеративен изкуствен интелект за автоматизирано създаване, верифициране, управление и оценяване на тестово съдържание.

Приноси

Основните приноси на дисертацията могат да бъдат класифицирани като научни, научно-приложни и приложни.

Научни приноси

- Създадена е концептуална рамка на интелигентна платформа за електронно тестване, базирана на генеративен изкуствен интелект;

- Предложен е подход за многомоделна верификация на тестови въпроси чрез използване на независими AI модели.

Научно-приложни приноси

- Разработен е модел на интелигентна платформа за автоматизирано генериране на тестови въпроси чрез използване на генеративни езикови модели и параметризирани промптове;
- Предложен е механизъм за автоматизирана оценка на качеството на генерираните въпроси чрез независима AI верификация;

Приложни приноси

- Реализиран е софтуерен прототип на интелигентна платформа за електронно тестване, вкл. функционалности за автоматизирано генериране, верифициране и управление на тестови въпроси;
- Проведени са експериментални изследвания, доказващи приложимостта на разработеното решение.

Апробация

Участие в научноизследователски проекти

Получените по време на изследването ***резултати са представяни в три научно-изследователски проекта***:

- МУПД25-ФМИ-015, *Школа за ИКТ иновации: AI технологии за трансформация на образованието* (2025-2026).
- СП23-ФМИ-008, *Развитие на творческия потенциал в Школа за ИКТ иновации* (2023-2024).
- МУПД23-ФМИ-021, *Използване на методи на изкуствения интелект в бизнеса* (2023-2024).

С финансовата подкрепа на споменатите проекти, ***резултатите са докладвани на четири научни конференции***.

Устни и постерни доклади на конференции

1. *Етични предизвикателства и стратегии за безопасност при използването на генеративни AI системи в образованието*, Научна конференция "Дни на науката 2025", 21-22 ноември 2025г., гр. Пловдив.
2. *Бъдещето на обучението в облака: AI-базирана архитектура на софтуерно приложение за оценяване*, Научна конференция "Дни на науката 2023", 23-24 ноември 2023г., гр. Пловдив.
3. *Алгоритмично генериране на тестови въпроси от текст: обзор на съвременни AI платформи и тяхното приложение в образователния контекст*, Научна конференция "Дни на науката 2023", 23-24 ноември 2023г., гр. Пловдив.
4. *Системи с изкуствен интелект за управление на обучението*, XXXIII Международна online научна конференция, 1 юни 2023 г., гр. Стара Загора.

Списък с публикации по темата на дисертационния труд

По темата на дисертацията са публикувани общо шест научни публикации. Три от публикациите са индексирани в световноизвестните бази от данни Web of science и/или SCOPUS. Едната публикация е индексирана в Web of science с IF=2,5 (Q2) и в Scopus с SJR=0.52 (Q2):

1. Hadzhikoleva, S., T. Rachovski, E. Hadzhikolev, I. Ivanov, *The role of generative artificial intelligence in programmer training: opportunities and challenges*, Proceedings of the 16th International Scientific and Practical Conference “Environment. Technology. Resources”, June 19-20, 2025, Rezekne, Latvia. **(SCOPUS)** https://doi.org/10.17770/etr2025vol2.8608

2. Hadzhikoleva, S., T. Rachovski, I. Ivanov, E. Hadzhikolev, G. Dimitrov, *Automated Test Creation Using Large Language Models: A Practical Application*, Applied Sciences vol. 14, no. 19, 2024. **(Web of science Q2 IF=2,5; SCOPUS Q2 SJR=0.52)** https://doi.org/10.3390/app14199125
3. Rachovski, T., D. Petrova, I. Ivanov, *Automated Creation of Educational Questions: Analysis of Artificial Intelligence Technologies and Their Role in Education*, Proceedings of the 15th International Scientific and Practical Conference "Environment. Technology. Resources", June 27-28, 2024, Veliko Tarnovo, Bulgaria. **(SCOPUS)** https://doi.org/10.17770/etr2024vol2.8101
4. Ivanov, I., T. Rachovski, G. Pashev, *Integration of AI models for question generation and assessment: Application of ChatGPT and Gemini in education*, Scientific researches of the Union of Scientists in Bulgaria-Plovdiv, series B. Natural Sciences and the Humanities, Vol. XXV, ISSN 1311-9192 (Print), ISSN 2534-9376 (On-line), 2024, стр. 154-159. https://usb-plovdiv.org/wp-content/uploads/2024/12/2024_natural_sciences_and_humanities_vol_XXV.pdf
5. Иванов, И., Т. Рачовски, *Алгоритмично генериране на тестови въпроси от текст: обзор на съвременни AI платформи и тяхното приложение в образователния контекст*, Научни трудове на Съюза на учените в България–Пловдив, серия В. Техника и технологии, т. XXI, ISSN 1311-9419 (Print), ISSN 2534-9384 (Online), 2024, стр. 183 – 188. https://usb-plovdiv.org/wp-content/uploads/2024/05/2024_tehnika_i_tehnologii_tom_XXI.pdf
6. Иванов, И., Т. Рачовски, *Системи с изкуствен интелект за управление на обучението*, Science and technologies: Volume XIII, 2023, Number 2: TECHNICAL STUDIES, стр. 1 - 7. https://www.sustz.com/journal/2/2102.pdf

Забелязани са ***над 30 цитирания*** на две от публикациите по дисертационния труд.

БИБЛИОГРАФИЯ

(Agbo et al., 2020) Agbo, G.C.; Agbo, P.A. The role of computer vision in the development of knowledge-based systems for teaching and learning of English language education. *ACCENTS Trans. Image Process. Comput. Vis.* **2020**, *6*, 42–47. https://doi.org/10.19101/TIPCV.2020.618044.

(Baker & Kim, 2004) Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques*. Taylor & Francis Group., R.E.; Bejar, I.I. Validity and automad scoring: It's not only the scoring. *Educ. Meas. Issues Pract.* **1998**, *17*, 9–17.

(Biehl, 2016) Biehl, M. *RESTful API Design: Best Practices in API Design with REST*; ASIN: B01L6STMVW; API-University Press: 2016.

(Borisova, Hadzhikoleva et al., 2024) Borisova, M., S. Hadzhikoleva & E. Hadzhikolev. ChatGPT as a tool for creating educational games for school education in computer technology, International Conference on Virtual Learning, ISSN 2971-9291, ISSN-L 1844-8933, vol. 19, pp. 323-333, 2024. https://doi.org/10.58503/icvl-v19y202427

(Briganti & Le Moine, 2020) Briganti, G.; Le Moine, O. Artificial intelligence in medicine: Today and tomorrow. *Front. Med.* **2020**, *7*, 27. https://doi.org/10.3389/fmed.2020.00027.

(Bugbee, 1996) Bugbee, A.C. The Equivalence of Paper-and-Pencil and Computer-Based Testing. *J. Res. Comput. Educ.* **1996**, 28, 282–299. https://doi.org/10.1080/08886504.1996.10782166.

(Chaudhry, Sarwary et al., 2023) Chaudhry, I.; Sarwary, S.; Refae, G.; Chabchoub, H. Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT–A Case Study. *Cogent Educ.* **2023**, *10*, 2210461. https://doi.org/10.1080/2331186X.2023.2210461.

(Chen, Samuel et al., 2021) Chen, W.; Samuel, R.; Krishnamoorthy, S. Computer Vision for Dynamic Student Data Management in Higher Education Platform. *J. Mult. -Valued Log. Soft Comput.* **2021**, *36*, 5.

(Darayseh, 2023) Darayseh, A.A. Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100132. https://doi.org/10.1016/j.caeai.2023.100132.

(Dempere, Modugu et al., 2023) Dempere, J.; Modugu, K.; Hesham, A.; Ramasamy, L. The impact of ChatGPT on higher education, *Front. Educ.* **2023**, *8*, 1206936. https://doi.org/10.3389/feduc.2023.1206936.

(Edmett, Ichaporia et al., 2023) Edmett, A.; Ichaporia, N.; Crompton, H.; Crichton, R. Artificial Intelligence and English Language Teaching: Preparing for the Future. British Council, 2023. Available online: https://www.teachingenglish.org.uk/sites/teacheng/files/2024-08/AI_and_ELT_Jul_2024.pdf

(Embretson & Reise, 2025) Embretson, S. E., & Reise, S. P. (2025). *Item response theory: Foundations for psychologists and social scientists*. Routledge. https://doi.org/10.4324/9781315726557

(Grassini, 2023) Grassini, S. Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Educ. Sci.* **2023**, *13*, 692. https://doi.org/10.3390/educsci13070692.

(Hadzhikolev, Hadzhikoleva et al., 2021) Hadzhikolev, E.; Hadzhikoleva, S.; Yotov, K.; Borisova, M. Automated Assessment of Lower and Higher-Order Thinking Skills Using Artificial Intelligence Methods. *Commun. Comput. Inf. Sci.* **2021**, *1521*, 13–25.

(Hadzhikoleva, Borisova et al., 2025) Hadzhikoleva, S., Borisova, M., Hadzhikolev, E., & Raykova, Z. (2025, March). Building and Utilizing a Specialized Chatbot in Physics Education–Experiment and Results. In 2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH) (pp. 1-6). IEEE. https://doi.org/10.1109/INFOTEH64129.2025.10959194
(Hadzhikoleva, Gorgorova et al., 2024) Hadzhikoleva, S., M. Gorgorova, E. Hadzhikolev, G. Pashev. (2024). AI-Driven Approach to Educational Game Creation, Proceedings of the 16th ICT Innovations Conference, 28-30 September 2024, Ohrid, Macedonia, ISBN 978-608-65468-4-7. https://ictinnovations.org/wp-content/uploads/2025/01/final_web_proceedings.pdf
(Hajkowicz, Sanderson et al., 2023) Hajkowicz, S.; Sanderson, C.; Karimi, S.; Bratanova, A.; Naughtin, C. Artificial intelligence adoption in the physical sciences, natural sciences, life sciences, social sciences and the arts and humanities: A bibliometric analysis of research publications from 1960–2021. *Technol. Soc.* **2023**, *74*, 102260. https://doi.org/10.1016/j.techsoc.2023.102260.
(Harry & Sayudin, 2023) Harry, A.; Sayudin, S. Role of AI in Education. *Interdiciplinary J. Hummanity* **2023**, *2*, 260–268. https://doi.org/10.58631/injurity.v2i3.52.
(Heilman, 2011) Heilman, M. Automatic Factual Question Generation from Text. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2011.
(Jeon & Lee, 2023) Jeon, J.; Lee, S. Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Educ. Inf. Technol.* **2023**, *28*, 15873–15892. https://doi.org/10.1007/s10639-023-11834-1.
(Kandlhofer, Steinbauer et al., 2016) Kandlhofer, M.; Steinbauer, G.; Hirschmugl-Gaisch, S.; Huber, P. Artificial intelligence and computer science in education: From kindergarten to university. In Proceedings of the 2016 IEEE Frontiers in Education Conference (FIE), Erie, PA, USA, 12–15 October 2016. https://doi.org/10.1109/FIE.2016.7757570.
(Kasneci, Sessler et al., 2023) Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274.
(Kucak, Juricic et al., 2018) Kucak, D.; Juricic, V.; Dambic, G. Machine Learning in Education–A Survey of Current Research Trends. In Proceedings of the 29th DAAAM International Symposium, Vienna, Austria, 24–27 October 2018. https://doi.org/10.2507/29th.daaam.proceedings.059.
(Mishra, 2019) Mishra, R. Usage of Data Analytics and Artificial Intelligence in Ensuring Quality Assurance at Higher Education Institutions. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 4–6 February 2019; pp. 1022–1025. https://doi.org/10.1109/AICAI.2019.8701392.
(Nunez & Lantada, 2020) Nunez, J.M.; Lantada, A.D. Artificial intelligence aided engineering education: State of the art, potentials and challenges. *Int. J. Eng. Educ.* **2020**, *36*, 1740–1751.
(Nurhayati & Halimah, 2024) Nurhayati, T.N.; Halimah, L. The Value and Technology: Maintaining Balance in Social Science Education in the Era of Artificial Intelligence. In Proceedings of the International Conference on Aplied Social Sciences in Education, Bangkok, Thailand, 14–16 November 2024; Volume 1, pp. 28–36. https://doi.org/10.31316/icasse.v1i1.6840.
(O'Connor, 2023) O'Connor, S. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Educ. Pract.* **2023**, *66*, 103537. https://doi.org/10.1016/j.nepr.2022.103537.
(Osterlind, 1998) Osterlind, S.J. What is constructing test items? In *Constructing Test Items. Evaluation in Education and Human Services*; Springer: Dordrecht, The Netherlands, **1998**; Volume 25. https://doi.org/10.1007/978-94-009-1071-3_1.
(Pabitha, Mohana et al., 2014) Pabitha, P.; Mohana, M.; Suganthi, S.; Sivanandhini, B. Automatic Question Generation system. In Proceedings of the 2014 International Conference on Recent Trends in Information Technology, Chennai, India, 10–12 April 2014; pp. 1–5. https://doi.org/10.1109/ICRTIT.2014.6996216.
(Papasalouros, Kanaris et al., 2008) Papasalouros, A.; Kanaris, K.; Kotis, K. Automatic generation of multiple choice questions from domain ontologies. In Proceedings of the International Conference e-Learning 2008, Amsterdam, The Netherlands, 22–25 July 2008; pp. 427–434.
(Pradana, Elisa et al., 2023) Pradana, M.; Elisa, H.; Syarifuddin, S. Discussing ChatGPT in education: A literature review and bibliometric analysis. *Cogent Educ.* **2023**, *10*, 2243134. https://doi.org/10.1080/2331186X.2023.2243134.
(Rahiman & Kodikal, 2024) Rahiman, H.; Kodikal, R. Revolutionizing education: Artificial intelligence empowered learning in higher education. *Cogent Educ.* **2024**, *11*, 2293431. https://doi.org/10.1080/2331186X.2023.2293431.
(Rojas & Chiappe, 2024) Rojas, M.P.; Chiappe, A. Artificial Intelligence and Digital Ecosystems in Education: A Review. *Technol. Knowl. Learn.* **2024**, 1–18. https://doi.org/10.1007/s10758-024-09732-7.
(Sofianos, Stefanidakis et al., 2024) Sofianos, K.C.; Stefanidakis, M.; Kaponis, A.; Bukauskas, L. Assist of AI in a Smart Learning Environment. *IFIP Adv. Inf. Commun. Technol.* **2024**, *714*, 263–275. https://doi.org/10.1007/978-3-031-63223-5_20.
(Stokel-Walker, 2022) Stokel-Walker, C. AI bot ChatGPT writes smart essays-should academics worry? *Nature, 09 December 2022.* https://doi.org/10.1038/d41586-022-04397-7.
(Taecharungroj, 2023) Taecharungroj, V. "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data Cogn. Comput.* **2023**, *7*, 35. https://doi.org/10.3390/bdcc7010035.
(Van der Linden & Glas, 2010) Van der Linden, W. J., & Glas, C. A. (Eds.). (2010). *Elements of adaptive testing* (Vol. 10, pp. 978-0). New York: Springer. https://doi.org/10.1007/978-0-387-85461-8
(Wyse, 2010) Wyse, A. E. (2010). RJ DE AYALA (2009) The Theory and Practice of Item Response Theory. New York: The Guilford Press. ISBN: 978-1-59385-869-8. *Psychometrika*, *75*(4), 778-779. https://doi.org/10.1007/s11336-010-9179-z
(Xu & Ouyang, 2022) Xu, W.; Ouyang, F. The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *Int. J. STEM Educ.* **2022**, *9*, 59. https://doi.org/10.1186/s40594-022-00377-5.
(Zhang & Aslan, 2021) Zhang, K.; Aslan, A.B. AI technologies for education: Recent research & future directions. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100025. https://doi.org/10.1016/j.caeai.2021.100025.