

**Изкуственият интелект като инструмент  
за автоматизирано създаване на бази данни от тестови въпроси**  
Дилян Димитранов, Георги Димитров

**The Artificial Intelligence as a Tool  
for the Automated Creation of Test Question Databases**  
Dilyan Dimitranov, Georgi Dimitrov

**Abstract:**

The paper examines the practical application of artificial intelligence in the creation of structured databases of multiple-choice test questions within the context of higher maritime education. It presents QuestionForge, a web-based system developed by the authors, which implements a structured eight-step workflow for the automated generation, configuration, verification, and organization of test questions using artificial intelligence. The system was applied in practice within the training course “Maritime Environmental Awareness” During the conducted experiment, more than 656,000 characters of Bulgarian and English text have been processed, resulting in the generation of a total of 553 test questions. The results indicate that between 95% and 99% of the questions could be used directly for examination or training purposes.

The proposed approach demonstrates considerable potential for accelerating the preparation of examination materials and supporting the digital transformation of assessment practices.

**Keywords:** Artificial intelligence, automated question generation, test question databases, digital transformation, higher education, maritime education.

**For contacts:** Capt. Dilyan Dimitranov, PhD, Nikola Vaptsarov Naval Academy, Varna, d.dimitranov@nvna.eu

**ВЪВЕДЕНИЕ**

Навлизането на цифровите технологии в образованието поставя пред академичния състав изисквания, надхвърлящи традиционните рамки на преподаване. Сред тях особено място заема изграждането на разнообразни и актуализирани бази данни от тестови въпроси — задача, която при класическите подходи ангажира значителни времеви и интелектуални ресурси, без задължително да дава пропорционален педагогически резултат.

Развитието на големите езикови модели (Large Language Models — LLMs) в последните години предоставя реалистична основа за частична автоматизация на този процес. Инструменти като ChatGPT, Claude и Gemini демонстрират способност да генерират свързан и контекстуално обоснован текст въз основа на предоставено съдържание — включително тестови въпроси с множествен избор [1]. От съществено значение е, обаче, тези инструменти да се разглеждат като помощно средство в ръцете на преподавателя, а не като негов заместник: педагогическата преценка остава незаменима при оценката и отбора на генерираното съдържание.

Настоящият доклад представя системата QuestionForge — уеб-базирано приложение, разработено от авторите с цел структурирано и контролирано създаване на тестови въпроси от учебен текст. Системата е приложена практически по дисциплината „Познание по опазване на околната среда“ и резултатите от нейното използване са подложени на емпирична оценка.

## ТЕОРЕТИЧНА РАМКА НА ИЗСЛЕДВАНЕТО

Автоматичното генериране на въпроси (Automatic Question Generation — AQG) се утвърди като активна изследователска област в образователните технологии, чиято основна цел е извличането на разнообразни задачи за оценяване от текстово съдържание. Kurdi et al. установяват, че AQG системите могат да постигат значително качество при работа със структурирани учебни текстове, особено когато се прилагат съвременни методи за обработка на естествен език [2]. С появата на LLMs като GPT, Claude и Gemini обхватът на AQG се разшири съществено: изследователите и преподавателите получиха достъп до многоезични, гъвкави системи, способни да генерират разнородно тестово съдържание [1]. В образователната литература тези технологии последователно се разглеждат като инструменти в подкрепа на преподавателя, а не като негов заместник [3] — принцип, залегнал в основата и на настоящото изследване.

Наред с предимствата, редица автори обръщат внимание на съществени рискове: склонност на езиковите модели към генериране на фактически неточности („халюцинации“) и опасността автоматично произведеното съдържание да бъде прието без необходимото критично осмисляне [4]. UNESCO препоръчва интегрирането на инструменти с изкуствен интелект в образованието да бъде съпроводено от специализирани рамки и механизми за контрол [5]. В разработената система тези препоръки са намерили конкретно приложение: генерираните въпроси задължително преминават ръчна проверка от преподавателя, преди да бъдат включени в база данни за оценяване.

Haladyna et al. формулират принципи за конструиране на качествени въпроси с множествен избор, като особено внимание обръщат на педагогически обоснованото разграничаване на дистракторите [6]. Тези принципи са заложили непосредствено в архитектурата на QuestionForge.

## КОНЦЕПЦИЯ И АРХИТЕКТУРА НА СИСТЕМАТА

QuestionForge е уеб-базирано приложение, разработено на платформата Base44 и достъпно на адрес <https://questionforge.base44.app/>. Системата осъществява структуриран осемстъпков работен процес за създаване на тестови въпроси с множествен избор, изграден върху три основни принципа: преподавателят запазва пълен контрол върху крайното съдържание; AI моделът функционира като помощен инструмент; работният процес е достатъчно интуитивен, за да не изисква специализирани технически познания от страна на потребителя. Работният процес на системата обхваща следните осем стъпки, представени на Фиг. 1.

СТЪПКА 1. Преподавателят поставя произволен учебен текст в полето за текст в приложението.

СТЪПКА 2. Задава се концептуалната дълбочина на въпроса (единична концепция, интегрирани концепции или комплексен синтез) и нивото на трудност (лесно, средно, трудно). Тази стъпка е въведена в отговор на установена в практиката тенденция: без явно зададена дълбочина AI моделите създават тясно-специализирани въпроси, фокусирани върху отделен детайл от текста, вместо върху по-широката концепция, която той изгражда.

**СТЪПКА 3.** При необходимост AI моделът може автоматично да извлече ключовите понятия от текста. Преподавателят може да постави акцент върху конкретни понятия или да ги изключи изцяло от създаването на въпроси.

**СТЪПКА 4.** Системата генерира въпроса и четири варианта на отговор с педагогически обосновани характеристики:

- верен отговор — изграден директно от текста с точна терминология; фактически верен и недвусмислен,
- правдоподобен грешен отговор — правдоподобен, но фактически неточен: причинно-следствената връзка е обърната, условие е заменено или квалификаторът е изместен; насочен срещу повърхностно усвоеното знание,
- подвеждащ отговор — използва реално, но погрешно схващане от текста: погрешно приложен механизъм или объркване между корелация и причинност, и
- дистрактор — правдоподобна концепция извън обхвата на текста; широко приложим общ принцип, използван в неподходящ контекст; никога абсурден и никога тривиално очевиден.

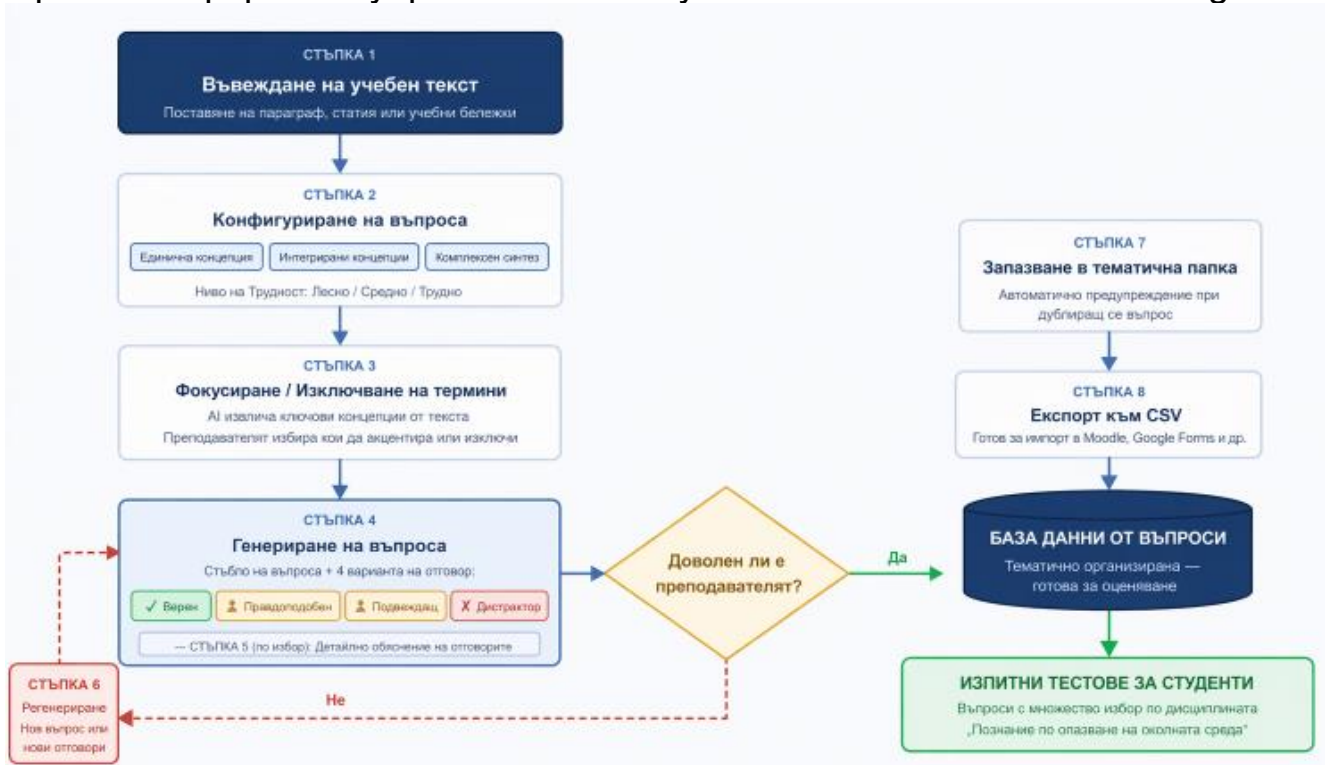
Тази типология е в съответствие с установените принципи за конструиране на качествени MCQ задачи [6].

**СТЪПКА 5** (по избор). Системата може да генерира подробно обяснение защо всеки отговор е верен или грешен, с позоваване на входящия текст.

**СТЪПКА 6.** При неудовлетворителен резултат преподавателят може да регенерира въпроса или само вариантите на отговор, като добави уточняваща бележка за насочване на AI модела.

**СТЪПКА 7.** Одобреният въпрос се записва в тематично организирана папка. Системата автоматично предупреждава при наличие на дублиращ се въпрос, което е особено полезно при изграждане на обширни бази данни.

**СТЪПКА 8.** Въпросите се експортират в Excel формат, готов за директен импорт в платформи за управление на обучението като Moodle или Google Forms.



Фиг. 1. Работен процес на системата QuestionForge

## ЕМПИРИЧНА ОЦЕНКА

За оценка на практическата приложимост на системата беше проведен експеримент с реални учебни материали по дисциплината „Познание по опазване на околната среда“. Учебният текст обхваща лекционен материал, международни конвенции, официални документи, научни статии и доклади в областта на морската екология. Въпросите са генерирани последователно — по един — при активно участие на преподавателя в процеса на конфигуриране и одобрение.

Целта на експеримента е да се установи до каква степен системата генерира въпроси, приложими за директна употреба, и какъв обем редакция се налага. Общият обем на обработения текст е приблизително 656 000 знака (без интервали) на два езика. Обобщените резултати са представени в Табл. 1.

Табл.1 Резултати от експерименталното генериране на тестови въпроси

| Език        | Обем текст (знаци без интервали) | Генерирани въпроса | Въпроси изискващи редакция |
|-------------|----------------------------------|--------------------|----------------------------|
| Български   | 324 920                          | 267                | ~4–5%                      |
| Английски   | 331 893                          | 286                | ~1–2%                      |
| <b>Общо</b> | <b>656 813</b>                   | <b>553</b>         | <b>~2–3%</b>               |

От приблизително 1190 знака текст се генерира един тестов въпрос, което позволява бързо изграждане на мащабни бази данни. Необходимата редакция при английските текстове е значително по-малка (~1–2%), отколкото при българските (~4–5%). Анализът на случаите, изискващи редакция, разкрива, че преобладаващият тип нередност е неестествено или несвойствено формулиране на въпроса и/или вариантите на отговор на български. По-рядко срещан проблем е тенденцията на AI модела да се фокусира върху конкретен детайл от текста, вместо върху по-широката концепция. Именно тази наблюдавана слабост е обусловила въвеждането на Стъпка 2 от работния процес — конфигурирането на концептуалната дълбочина на въпроса. Практическото приложение потвърждава, че преобладаващата част от генерираните въпроси са подходящи за оценяване без допълнителна редакция, а необходимите корекции са предимно от езиков и стилистичен, а не от фактически характер.

## ОГРАНИЧЕНИЯ

Настоящото изследване има няколко ограничения, които следва да бъдат взети предвид при интерпретирането на резултатите. Качеството на генерираните въпроси е оценено единствено чрез предварителен редакционен преглед от преподавателя; не е извършен анализ въз основа на резултатите на студентите — например идентифициране на въпроси с необичайно висока или ниска успеваемост. Подобен анализ е планиран като следващ етап на изследването.

Генерирането на въпросите е осъществено последователно — по един въпрос на операция, което предполага значително ангажиране на преподавателя.

Автоматизираното масово създаване на въпроси би могло допълнително да намали необходимото работно време.

Проучването е ограничено до специализирана морска учебна дисциплина, което следва да се има предвид при обобщаването на резултатите. По-ниското качество на генерирания текст на български вероятно отразява по-слабото представяне на езика в обучителните данни на наличните LLMs — тенденция, вероятно характерна и за други езици с ограничено представяне в обучителните корпуси.

## **ЗАКЛЮЧЕНИЕ**

Настоящото изследване показва, че системата QuestionForge може значително да ускори изграждането на бази данни от тестови въпроси във висшето образование. От обработените над 656 000 знака учебен текст по дисциплината „Познание по опазване на околната среда“ са генерирани 553 въпроса с множествен избор, от които между 95% и 99% са готови за употреба с минимална или без редакция.

От особено значение е връзката между практическото използване на системата и нейното усъвършенстване. Установената склонност на AI моделите да създават твърде специализирани въпроси води до въвеждане на механизъм за настройване на концептуалната дълбочина. Това показва, че подобни инструменти могат да се развиват според конкретните педагогически потребности. Резултатите потвърждават, че AI подпомага дигиталната трансформация на оценяването, когато е интегриран в процес, запазващ контрола на преподавателя. Системата не го замества, а намалява рутинната работа и освобождава време за съдържателна и методическа подготовка.

Бъдещото развитие на системата предвижда анализ на качеството на въпросите въз основа на студентски резултати, автоматизирано масово създаване на въпроси, разширяване към допълнителни дисциплини и оптимизиране на езиковото качество на произведеното съдържание на български.

## **ЛИТЕРАТУРА**

1. Brown, T. et al. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
2. Kurdi, G. et al. (2020). A Systematic Review of Automatic Question Generation for Educational Purposes. *International Journal of Artificial Intelligence in Education*, 30, 121–204.
3. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education*. Center for Curriculum Redesign.
4. Bender, E. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of FAccT 2021*, 610–623.
5. UNESCO (2023). *Guidance for Generative AI in Education and Research*. UNESCO Publishing.
6. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309–333.