# Етични аспекти в приложението на изкуствен интелект в обучението

Теодора Данева

# Ethical aspects in the application of artificial intelligence in education

Teodora Daneva

**Abstract:**
In education, the application of artificial intelligence (AI) is the source of a number of benefits related to the personalization of educational content, the automation of administrative activities, as well as the effective integration of children with special educational needs (SEN). Despite all the advantages, the implementation of AI requires the development of a unified framework for regulation and ethics of use. The adaptation of AI should not contradict basic human rights, based on the principles of transparency of information, non-discrimination and protection of personal data. The report accepts the following limitation – it does not seek to analyze all aspects related to ethics in the application of AI in education, but aims to synthesize the main ethical principles of AI through a comparative analysis of the regulatory framework developed so far in the countries of the EU, USA and China. In this way, scientific research enables the follow-up of the generally accepted normative restrictions to date and creates prerequisites for the development of new ones at national level in line with world practices.
**Keywords:** applied AI, education, ethical principles of AI

**For contacts:** as. prof. Teodora Daneva, University of Economics - Varna, teodora.daneva@ue-varna.bg

## ВЪВЕДЕНИЕ

Възможностите на AI да изпълнява човекоподобни задачи чрез преминаване на т.нар. Тюринг тест и интеракцията машина – човек (в определени случаи – дете) поставят един от най-важните въпроси, свързани с етиката и приложение му за добросъвестни цели, като налага необходимостта от по-детайлно проучване на регулативната рамка, свързана с употребата на ИИ в световен мащаб. Предмет на научния доклад са етичните принципи на приложния ИИ в обучението, а обект – регулацията в страните от ЕС, САЩ и Китай при употреба за образователни цели. Докладът приема следното ограничение – не се стреми да анализира всички аспекти, свързани с етиката при приложение на ИИ в обучението, а има за цел да синтезира основните етични принципи на ИИ чрез сравнителен анализ на разработената до момента регулативна рамка в страните от ЕС, САЩ и Китай. По този начин научното изследване дава възможност за мониторинг и проследяване на общоприетите нормативни ограничения до момента и създава предпоставки за разработване на национално ниво на нови такива в унисон с възприетите световни практики.

## ИЗЛОЖЕНИЕ

### Приложение на ИИ в обучението

На базата на бихейвиористичната теория (Holland, Skinner, 1961) и чрез алгоритмите за машинно обучение с потвърждение[35] ИИ е инструмент за превод на текст и реч при изучаване на чужди езици, автоматизирано управление на автомобили в шофьорски курсове, но най-важната му характеристика е, че има способността да подобрява педагогическите подходи, използвайки придобития опит с предходни ученици и/или студенти (Iglesias et al, 2009). В парадигмата за мултимедийното учене (Mayer, 2020) и теорията за конструктивизма (Dewey, 1938), употребата на ИИ подпомага експертите в прилагането на т.нар. техника „скеле“, придобила популярност в работата на Wood, Bruner, Ross (1976) за ролята на тютора по време на обучението на ученика чрез създаване на виртуални асистенти, чатботове като SmarterChild, Siri, Alexa, изпълняващи различни дейности и подпомагащи обучаваните в изпълнението на множество задания (Okonkowo, Ibijola, 2021). Чрез отворените модели[36] за обучение в дигитална среда ИИ поема ролята на своеобразен експерт, осигуряващ персонализирани насоки, подходящи учебни материали и постоянна обратна връзка относно прогреса в обучението (Zhang, Aslan,2021; Crompton, Burke, 2023). Като част от интелигентни класни стаи – основно чрез IoT[37], приложният ИИ увеличава ангажираността на учениците, помага за наваксване с учебния материал и адаптирането му съгласно индивидуалните потребности на обучаващите се (Heineman, Uskov, 2017). Тези предимства на приложния ИИ допринасят за успешно интегриране на деца със СОП в учебния процес и подобряване на техните резултати. Основавайки се на NLP[38] модели, ИИ оптимизира процеса по планиране на занятията от страна на учителите, като предоставя възможности за автоматизирано оценяване на тестове и задания. По този начин учителите са в състояние да се фокусират върху намирането на по-ефективни методи на преподаване, подобряване на интеракцията учител-ученик и внедряването на иновации в различните дисциплини. Не на последно място, част от технологиите, базирани на ИИ, прогнозират броя на отпадналите студенти от образователния процес. Тези характеристики на приложния ИИ предоставят множество опции за мониторинг на индивидуалните резултати от обучението и формулиране на стратегии за предотвратяване на отпадането от него (Kamalov, 2023).

**Сравнителен анализ на регулативната рамка в страните от ЕС, САЩ и Китай**

Създаването на технологии с ИИ предполага разработването на насоки, съобразени с основните човешки права и на които да отговаря надеждният ИИ. ЕС, Китай и САЩ създават препоръки и регулаторни рамки с различен фокус и механизми за прилагане. Преобладаващата част от разглежданите нормативни документи в Китай и ЕС ограничават приложния ИИ в резултат от нарастващата осведоменост в обществото и потенциалните вреди при употреба. Законът за ИИ на ЕС (ЕР, 2024), приет от Европейския парламент през 2024 г., е фокусиран върху

[35] reinforcement learning от англ. език – подсилено обучение
[36] Open Learning Models от англ. език – отворени модели за обучение
[37] Internet of Things от англ. език – Интернет на нещата
[38] Natural Language Processing от англ. език – естествени езикови модели

основните права на хората, като класифицира системите с ИИ според степента на риск. Забранени са технологиите за социално оценяване, за лицево разпознаване и системите, оказващи манипулативно въздействие върху потребителите. Регулирани са ИИ технологии със способност да създават фалшиво/нереално съдържание[39] като образи, клипове и е поставен фокус върху прозрачността и проследимостта на наличната информация. Друг документ, насочен към разработчиците на системи с ИИ в страните от ЕС, са препоръките на ЕС за надежден ИИ[40] (EU, 2019). В САЩ регулативната рамка, свързана с приложния ИИ и разработена от Службата за управление и бюджет към Белия дом, разчита на доброволни ангажименти, част от които е Изпълнителната заповед за безопасно, сигурно и надеждно развитие и използване на изкуствен интелект (The White House, 2023). Тя дава конкретни насоки за оценяване на надеждността на ИИ от правителствените агенции. Позицията на САЩ дава предимство на иновациите, като се стреми да едновременно да елиминира рисковете при употреба чрез създаването на Плана на закона за правата на ИИ (The White House, 2023), Рамката за управление на риска, свързана с ИИ (Tabassi, 2023), докато ЕС дава приоритет на безопасността и правата на потребителите чрез реални законови ограничения.

В Китай наредба от 2021 г. относно препоръчителните алгоритми, правилата от 2022г. за дълбок синтез (синтетично генерирано съдържание) и проектоправилата от 2023 г. относно генеративния ИИ представляват основна регулаторна рамка, свързана с приложния ИИ, но за разлика от ЕС и САЩ и предвид особеностите на политическата система, контролът на разпространяваното съдържание е основен фокус и на трите разпоредби. Правилата за разработване на алгоритми забраняват ценовата дискриминация и защитават правата на работниците, а регламентът за дълбок синтез[41] изисква върху съдържанието, генерирано с ИИ, да се поставят водни знаци. Внесеният през април 2023 проектозакон за ИИ, който предвижда строги мерки относно дейността и отговорността на доставчиците на технологии с ИИ, е видоизменен, като голяма част от първоначалните мерки са смекчени и е поставен акцент върху желанието на КНР за превръщане в основен иновационен център за развитие на ИИ чрез насърчаване на чуждестранните инвестиции и изключване на част от правилата генеративния ИИ, използван за изследователски цели (PwC, 2023).

**Етични аспекти в приложението на ИИ в обучението**

Системите с ИИ следва да бъдат разработвани с цел подпомагане благополучието на хората, но алгоритмите за машинно обучение се основават на достъп до голям обем от публични данни, в които са вградени стойностите на техните създатели и които включват съществуването на исторически, системни пристрастия и културни стереотипи. Множество изследвания по темата констатират полови, възрастови и расови пристрастия (Akgun, Greenhow, 2021). Stypinska (2021) въвежда концепцията за AI ageism - практики, идеологии в

[39] Deepfakes от англ. език

[41] Deep synthesis от англ. език – дълбок синтез

областта на ИИ, които изключват, дискриминират или пренебрегват интересите и нуждите на по-възрастното население. Расова дискриминация съществува и по отношение на системите за лицево разпознаване с ИИ на хора с по-тъмна кожа (Werhli, 2021; Lunter, 2020). Създаването на персонализирани материали изисква достъп до поверителна информация, която да е съобразена с правото на анонимност на индивидите и защита на личните им данни съгласно GDPR[42]. В контекста на ИИ управлението на данните е от решаващо значение за гарантиране, че тези, получени при обучение и работа с ИИ системи, са точни, справедливи и използвани отговорно и със съгласие (Stanford University, 2024). Един от законовите нормативи за противодействие в тази посока е въвеждането на протоколи, които да уреждат достъпа до данните, лицата и обстоятелствата за възникване на достъпа (ЕС, 2019). Етичните аспекти на приложния ИИ в обучението изискват пълна прозрачност по отношение на начина на вземане на решения, като се избягва т.нар. “черна кутия” при употреба на LLM, където генерираните данни са трудни за проследяване и обяснение. Потребителите следва да бъдат информирани при комуникация с ИИ и решения, взети изцяло от него, както и връзката на предоставени данни с различни бази. Обяснимостта изисква решенията, взети от ИИ, да бъдат разбирани от потребителите и да се осигурява ефективна намеса във високорискови случаи.

## ЗАКЛЮЧЕНИЕ

Независимо от разработените насоки и препоръки на множество организации, агенции и министерства в разглежданите страни, остава нуждата от единна глобална рамка, която да разпределя отговорността, да контролира риска и подкрепя развитието на иновации като разработва унифицирани правила за създаването на надежден ИИ. Рестрикциите, наложени от САЩ, ЕС и Китай илюстрират необходимостта от намиране на баланс в сложна геополитическа обстановка, в която фокус следва да се поставя върху благополучието на обществото, спазването на основните човешки права и елиминирането на рискове и вреди, независимо от потенциалните печалби.

## ЛИТЕРАТУРА

Akgun, S. and Greenhow, C. (2022) Artificial intelligence in education: Addressing ethical challenges in K-12 settings, *AI and Ethics*, 2(3), pp. 431–440.

Ala-Pietilä, P. *et al.* (2019) Ethics Guidelines For Trustworthy AI, AI High-Level Expert Group on Artificial Intelligence [online]. https://t.ly/7OwQ2 (достъпен на 10.05.24)

Crompton, H., Burke, D. (2023) Artificial intelligence in higher education: the state of the field, *International Journal of Educational Technology in Higher Education*, 20(1).

Dewey, J. (1938) *Experience and Education*. Free Press: Reprint edition [online]. https://t.ly/9eELw (достъпен на 10.05.24)

*EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act* (no date) [online]. https://t.ly/AsuxV (достъпен на 10.05.24)

*Global AI Governance Initiative* [online]. https://t.ly/lvzCn, (достъпен на 10.05.24)

[42] General Data Protection Regulation от англ. език – Закон за защита на личните данни

Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior: A program for self-instruction.* McGraw-Hill.

Iglesias, A. *et al.* (2009) Learning teaching strategies in an Adaptive and Intelligent Educational System through Reinforcement Learning, *Applied Intelligence*, 31(1), pp. 89–106.

Kamalov, F., Santandreu C., D. and Gurrib, I. (2023) New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution, *Sustainability (Switzerland)*, 15(16).

Maslej, N. et al, The AI Index 2024 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024 [online]. https://t.ly/WgXxu (достъпен на 10.05.24)

Mayer, R.E. (2020) *Multimedia Learning*. 3rd edn. Cambridge: Cambridge University Press.

Okonkwo, C.W., Ade-Ibijola, A. (2021) Chatbots applications in education: A systematic review, *Computers and Education: Artificial Intelligence*. Elsevier B.V.

Pricewaterhouse Coopers (2023) *EU AI Act: European AI regulation and its implementation* [online]. https://t.ly/4ldvF (достъпен на 10.05.24)

Roberts, H. *et al.* (2020) The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation,' *AI & Society*, 36(1), pp. 59–77.

*Sheehan, M. Tracing the roots of China's AI regulations* [online]. https://t.ly/T_FOA (достъпен на 10.05.24)

Stypinska, J. (2023) AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies, *AI and Society*, 38(2), pp. 665–677.

Tabassi, E. (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* [online]. https://t.ly/OvOqk (достъпен на 10.05.24)

The White House (2023) *Blueprint for an AI Bill of Rights | OSTP | The White House* [online]. https://t.ly/4wA3v (достъпен на 10.05.24)

The White House (2023) *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* [online]. https://t.ly/F37YG (достъпен на 10.05.24)

Wehrli, S. *et al.* (2022) Bias, awareness, and ignorance in deep-learning-based face recognition, *AI and Ethics*, 2(3), pp. 509–522.

Zhang, K., Aslan, A.B. (2021) AI technologies for education: Recent research & future directions, *Computers and Education: Artificial Intelligence*. Elsevier B.V.