Дигиталната сигурност и изкуствения интелект в образованието

Галина Иванова

Digital Security and Artificial Intelligence in Education

Galina Ivanova

Abstract:
Digital security in the era of artificial intelligence today is not only a technological issue, but also a matter of awareness and responsible behavior. The topic is particularly relevant with the widespread adoption of artificial intelligence not only in education, but also in all spheres of society. The report will address issues related to technological, ethical, legal, and social aspects: how to protect our personal data; how to be sure that information is reliable; how to demand reliable information; how to distinguish false information; how to ensure that information is not biased or discriminatory. The report will discuss the risks of using artificial intelligence in the educational sphere. It will summarize basic knowledge and practical advice on information security in order to recognize risks, understand the limitations of artificial intelligence, and use secure and reliable data.
Keywords: digital security, artificial intelligence, education, ethics in AI, risks

For contacts: Assoc. Prof. Galina Ivanova, University of Ruse, giivanova@uni-ruse.bg

ВЪВЕДЕНИЕ

Съвременното общество е изправено пред бързо развитие на дигитални технологии и навлизане на изкуствения интелект във всички сфери на живота, включително и в образованието. Ако интернетът промени света за 30 години и достигна до над 5 милиарда човека, мобилните телефони направиха това за 20 години, а изкуственият интелект - за по-малко от пет години. Предишните технологични иновации никога не са се случвали с такава скорост [1], Фиг. 1.

Фиг. 1. Развитие на технологичните иновации

В тези условия на бързо навлизане на изкуствен интелект и бързо остаряване на знанията трябва да търсим по-гъвкави и адаптивни модели на учене, които да отговарят на тези динамични изисквания като микроквалификации,

интердисциплинарно обучение или учене през целия живот. Университетът на бъдещето няма да бъде място за еднократно обучение, а партньор за учене през целия живот [2].

Когато говорим за образование днес в условия на бързо развитие на изкуствен интелект, е важно да разгледаме и концепцията за период на полуразпад на знанието – „Half Life of Knowledge“ [3]. Тази концепция е доразвита в книгата „Half Life of Facts“ на Самуел Арбесман – „Why everything we know has an expiration day“. Тя обозначава времето, необходимо половината от знанията в дадена област да остареят, да бъдат заменени от нови открития или да се окажат неточни.

В много професионални области знанията е необходимо да се обновяват по-често. Колкото по-динамична и технологично зависима е дадена област, толкова по-кратък е периодът на полуразпад на знанието. Могат да се посочат конкретни примери в информационните технологии и медицината, в които наученото днес може да бъде частично остаряло само след няколко години. На Фиг. 2 са представени примери за приблизителен период на полуразпад в областта на дигиталната сигурност. В сферата на сигурността периодът на полуразпад на знанието е относително кратък [4]. Нови атаки и инструменти за защита се появяват ежедневно. В някои подобласти периодът на полуразпад е една-две години.

Подобласт в сигурността	Приблизителен период на полуразпад
Киберсигурност (общо)	2–3 години
Мрежова сигурност	3–5 години
Етично хакерство / Penetration Testing	1–2 години
Cloud Security	1–3 години
AI Security / Security of AI	6 месеца – 2 години
Регулации и стандарти (ISO, NIS2, GDPR)	5+ години, но с периодични промени

Фиг. 2. Период на полуразпад на знанието в киберсигурността

ИЗЛОЖЕНИЕ

Обучението по дигитална сигурност не трябва да се асоциира само с подготовката по компютърните специалности. Дигиталната сигурност трябва да се интегрира във всички специалности с цел повишаване на дигиталната култура за използване на изкуствения интелект по отговорен и етичен начин - за по-ефективна защита на данните и разпознаване на рисковете. Необходимо е да се приложи интердисциплинарен подход, като се отчитат технологични, правни, етични и психологически аспекти.

Особена актуалност придобива въпросът кой носи отговорност, когато системите с изкуствен интелект вземат грешно или вредно решение, например в случаите, когато то е свързано с жизненоважни решения като грешки на автономни превозни средства или медицински грешки.

Въпреки че има ползи от изкуствения интелект, при липса на ясна регулаторна структура и ефективен човешки надзор, негативните последици могат да се увеличат. Препоръчителният подход е:

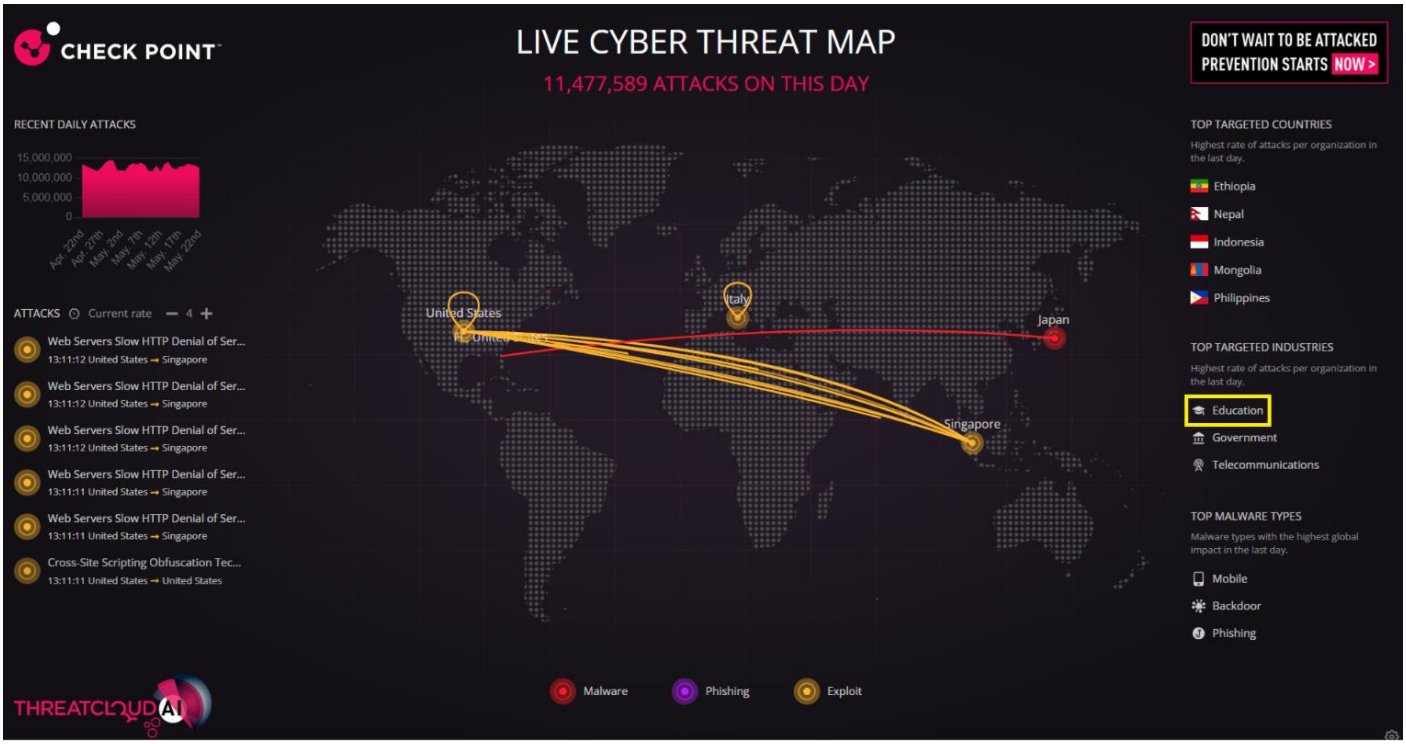
- изкуственият интелект да се използва като помощник;
- да се въведат етични стандарти и човешки контрол при важни решения;
- обучението по дигитална грамотност и сигурност да бъде за всички.

Бъдещето не е човек срещу изкуствения интелект, а човек с изкуствен интелект, но разумно, етично и безопасно.

Популярни атаки с използване на изкуствен интелект в образованието

С помощта на изкуствения интелект могат да се създадат по-сложни, по-автоматизирани, по-персонализирани, по-прецизно насочени и по-бързи кибератаки.

Образователният сектор може би не е най-привлекателният за атакуващите, но обобщените данни от сайта Check Point за 2025 показват, че най-атакуваните в световен мащаб са именно образователните институции, Фиг. 3.



Фиг. 3. Интерактивна карта с кибератаки в реално време на сайта Check Point

Най-популярните атаки с използване на изкуствен интелект в образованието могат да бъдат обобщени до:

- фишинг атаки – имейл или съобщения, които подвеждат потребителите да въведат чувствителни данни [5];
- Ransomware зловреден софтуер - криптира данните, докато не бъде платен откуп [6].
- изтичане на данни - неразрешен достъп до лични или институционални данни;
- DDoS (Distributed Denial of Service) - претоварване на системите с цел прекъсване на услуги [7].

В образователните институции се съхраняват данни, включително лични и чувствителни данни: адреси, контакти, академични данни, поведенчески профили

на обучаемите и др. Основните рискове са свързани с риск от изтичане на данни, неправомерното им използване, вкл. фалшифициране на данни. За целта са необходими мерки, свързани с по-строги правила за администрация, използване на мултифакторна аутентикация, биометрични характеристики (като пръстови отпечатъци, лицево разпознаване и др.) и т.нар. „силни“ пароли. Според последните тенденции [8] се препоръчва паролите да бъдат минимум 12 символа и да не се съхраняват на достъпни места, за да се предотвратят злоупотреби, при които учениците намират начини да придобият пароли на учители и да фалшифицират данни в електронните дневници. Този вид измами може да доведе до загуба на доверие в образователните институции и правни последици. Затова е необходимо да се обучава персоналът за важността на сигурността.

Други сериозни атаки с използване на изкуствен интелект са свързани с т.нар. “deepfake” измами, при които чрез фалшиви снимки, видеа или аудиозаписи се имитират реални популярни личности за заблуда. Обикновено се използва невронна мрежа, чрез която се обучава модел да имитира визуални и звукови характеристики. С цел изнудване на близки или с цел подвеждане и манипулиране се използват манипулирани видеа. Съвременните модели за изкуствен интелект имат способността да създават реалистични изображения и видеа от действителни снимки [9]. За да се предпазим от този вид измами, е необходимо ограничаването на личната информация в публичните профили и социалните мрежи, която може да се използва за създаване на “deepfake” измами. Една от актуалните заплахи, например, е клониране на глас на деца и обаждане до роднини с цел измами.

Могат да се обобщят следните съвети за разпознаване на “deepfake” измами:

- неестествено осветление и нетипични движения на лицето, вкл. мимики;
- несинхронизиран звук и движения, вкл. металически звук;
- „дигитален грим“ и липса на мигане;
- проблеми с периферията (с косата, ушите и др.);
- грешки в детайлите (бижута, очила и др.), Фиг. 4.

Фиг. 4. Разпознаване на изображение, генерирано с изкуствен интелект

Друг сериозен проблем е свързан с неточната и подвеждаща информация, която се генерира от системите с изкуствен интелект. Често отговорите, които обучаемите могат да получат, са неточни, но много убедителни. Съществува риск от разпространяване на смесена вярна и невярна информация от несъществуващи източници или подвеждащи научни твърдения. Последствията са свързани с погрешно учене и ниско качество на академичната работа. Системите с изкуствен интелект не са безгрешни и е необходима човешка експертиза, която да разпознава, когато системите „халюцинират“. Терминът „халюцинира“ се използва за обозначаване на неточна или измислена информация от системите с изкуствен интелект [10]. Решението на този проблем е повишаването на изискванията на потребителите към системите с изкуствен интелект и изричното посочване на условия за по-точни и проверими формулировки в отговорите въз основа на надеждни източници. Необходимо е обучение за използване на промпт инженеринг [11].

Безплатните системи с изкуствен интелект се „обучават“ от нашите данни, т.е. това може да доведе до ненадеждни и необективни данни. Обикновено това става анонимизирано, но съществува риск от пристрастни, остарели и непълни данни. Системите с изкуствен интелект не могат да правят морални и етични оценки, не мислят критично и не носят отговорност за решенията, които ни предлагат. Например системите с изкуствен интелект работят по-добре на английски език, което е дискриминация и се отчитат по-точни отговори за англоговорящите потребители, т.е. те имат предимство въз основа на езикови ограничения. Все по-наложителна е необходимостта от усъвършенствани системи с изкуствен интелект, например на български език за българското образование, обучени с данни и учебни материали, но одобрени от Министерството на образованието и науката. По този начин ще има сигурност, че данните са актуални и достоверни.

Могат да се вземат предвид специфични практики в други образователни системи. В Китай се проследява фокуса и ангажираността на учениците в класните стаи чрез камери с изкуствен интелект. Този подход създава риск от оценяване на поведенчески прояви, вместо реални знания. Такъв контрол може да доведе до несправедливо отношение към учениците.

Китай тества автоматизиране на оценяването на писмени задачи. Изкуственият интелект идентифицира грешки и дава обратна връзка. Но оценяването на задания в креативната област, като есета и творческо писане, понастоящем не може да се извършва от машини.

Друг проблем, с който образователните институции трябва да се справят, е този с плагиатството. Образователните институции започнаха да насърчават по-строги мерки при проверката на текстове за плагиатство и използване на изкуствен интелект. Системите за проверка все още не са добре развити и понякога обучаемите са несправедливо санкционирани.

Наскоро бяха актуализирани „Етични насоки за преподавателите относно използването на изкуствения интелект (ИИ) и на данни при преподаване и учене“ [12], но са необходими по-ясни и приложими политики за използване на изкуствен интелект от образователните институции, които постепенно да се въвеждат все по-активно.

ЗАКЛЮЧЕНИЕ

През последните няколко години нараства безпокойството относно потенциалните последици от използването на изкуствен интелект. Това е тема, която поражда съмнения, тревоги или надежди. Въпреки различните гледни точки е ясно, че използването на изкуствения интелект е неизбежно и трябва да сме готови за него.

Важно е да се повиши осведомеността на обществото относно отговорното, критично и безопасно използване на изкуствен интелект. Необходимо е да се разработят нови образователни ресурси с основни знания за функционирането на системите с изкуствен интелект.

Могат да бъдат обобщени следните практични съвети за по-ефективно и безопасно използване на системи с изкуствен интелект:

- не въвеждайте реални казуси и чувствителни данни, използвайте по-скоро измислени примери;
- отделете личния от служебния си профил при използване на системите с изкуствен интелект;
- използвайте различни чатове за различните теми, за да няма смесване и насочване на информацията;
- настройте поверителността на профилите си, като изключите опцията за съхранение на разговорите и използване на вашите данни за обучение на моделите.

За откриване на съвременните заплахи голям помощник може да бъде именно изкуственият интелект, който анализира големи обеми мрежов трафик и бързо открива бързо съмнителна активност. Системите с изкуствен интелект могат да сравняват предишни инциденти, например в необичайно време за дадената образователна институция през нощта или от различни държави. При проблемите, свързани с фишинг атаките, отново изкуственият интелект може да бъде наш помощник за предупреждения за съмнителни линкове.

Благодарение на изкуствения интелект сигурността може да премине от реактивен към проактивен модел – да предвижда и предотвратява атаки. Системите могат да се преконфигурират сами, но експертите по сигурност все още трябва да играят важна контролираща роля.

ЛИТЕРАТУРА

1. Misra, Amit, Jane Wang, Scott McCullers, Kevin White, and Juan Lavista Ferres. "Measuring AI Diffusion: A Population-Normalized Metric for Tracking Global AI Usage." arXiv preprint arXiv:2511.02781 (2025).

2. Белоев, Хр., Смрикаров, А., Василев, Цв., Георгиев, Цв., Смрикарова, Ст., Иванова, А., Иванова, Г., Стойкова, В., Арсова, Е., Алиев, Ю., Златаров, П. Визия за университета на бъдещето, Сборник доклади от Национална научно-практическа конференция „Дигитална трансформация на образованието – проблеми и решения, оценяване и акредитация“, Русенски университет (2023).

3. Arbesman, Samuel. "Truth decay: the half-life of facts." New Scientist 215, no. 2883 (2012): 36-39.

4. Zhang, Tianjian. "Knowledge expiration in security awareness training." (2018).

5. Jabir, Raja, John Le, and Chau Nguyen. "Phishing attacks in the age of generative artificial intelligence: A systematic review of human factors." AI 6, no. 8 (2025): 174.

6. Alzahrani, Saleh, Yang Xiao, Sultan Asiri, Jianying Zheng, and Tieshan Li. "A survey of ransomware detection methods." IEEE Access (2025).

7. Fu, Xingbing, Supeng Lou, Jiaming Zheng, Cheng Chi, Jie Yang, Dong Wang, Chenming Zhu, Butian Huang, and Xiatian Zhu. "Deep learning techniques for DDoS attack detection: Concepts, analyses, challenges, and future directions." Expert Systems with Applications 291 (2025): 128469.

8. Saputra, Whisnu Yudha, Sugiarti Sugiarti, Haris Junianto, and Didit Suhartono. "Password Strength Study Using The Zxcvbn Algorithm And Brute-Force Time Estimation To Strengthen Cybersecurity." Jurnal Pilar Nusa Mandiri 21, no. 1 (2025): 52-59.

9. Fatoni, Fatoni, Tri Basuki Kurniawan, Deshinta Arrova Dewi, Mohd Zaki Zakaria, and Abdul Muniif Mohd Muhayeddin. "Fake vs real image detection using deep learning algorithm." Journal of Applied Data Sciences 6, no. 1 (2025): 366-376.

10. Ciubotaru, Bogdan-Iulian. "The hallucination problem in generative artificial intelligence: Accuracy and trust in digital learning." In International Conference on Virtual Learning, vol. 20, pp. 35-45. 2025.

11. Lee, Daniel, and Edward Palmer. "Prompt engineering in higher education: a systematic review to help inform curricula." International Journal of Educational Technology in Higher Education 22, no. 1 (2025): 7.

12. https://learning-corner.learning.europa.eu/learning-materials/use-artificial-intelligence-ai-and-data-teaching-and-learning_bg