

Активност 3.1.2 – Основни обществени предизвикателства и рискове при употребата на ИИ

Прогнозна продължителност	40 минути
Изисквания	Модули 1 и 2. Активност 3.1.1. Ако се осъществява присъствено, умерено просторно открито пространство, в което учениците могат да обсъждат и да представят в групи
Използвани хардуерни и софтуерни инструменти	Дистанционно: Софтуер – видеоконферентен софтуер. Текстов редактор и презентационен софтуер в реално време за онлайн сътрудничество. Хардуер - настолен/лаптоп компютър
Тип оценяване	Оценката е базирана на разбирането на проблемите, обсъдени в класа, показвани чрез представяне на аргументи които подчертават отрицателното въздействие на ИИ и чрез критични дискусии, които поставят тези аргументи под въпрос.

Целта на тази активност е да се улесни разбирането на начина, по който употребата на ИИ технологиите може да има отрицателно въздействие върху реалния свят чрез представянето на поддържащи аргументи и примери. Ще бъдат насърчавани и разглеждани също коментари и въпроси, които оспорват тези аргументи.

Процедура (стъпки)

Предварителни материали

- Учителите могат да напомнят на учениците за система с изкуствен интелект, която вече са изучавали (напр. чатботове, дигитални асистенти, търсачки, разпознаване на изображения като проследяване на лица). Учителите отново се позовават на ЦУР, въведени в предишната дейност, за да обяснят, че използването на ИИ може да има и противоположни ефекти и да ни отдалечи от постигането на тези цели. Учителите представят някои примери, за да обяснят тази точка, като: според проучване на Европейския парламент, около 14% от работните места в страните от ОИСР са автоматизирани (хората могат да бъдат заменени от машини). Други 32% ще трябва да се изправят пред съществени

промени (хората трябва да се преквалифицират на работните си места); това би увеличило бедността.

- Случаят на отклонения в алгоритъма. Учителят обяснява, че хората понякога вземат решения, подпомагани от алгоритми (набор от действия или стъпки, необходими за решаване на проблем); следователно тези решения са отчасти автоматични. Компютърните препоръки може да изглеждат рационални или безпогрешни и хората може да ги следват сляпо. Въпреки това, в зависимост от това как е бил „обучен“ да работи и от информацията, която хората дават на компютъра за това (вижте машинното обучение), алгоритъмът може след това да предложи „дискриминационни“ решения, т.е. предпочитане на определени типове хора пред други. Това е, което се нарича „пристрастие на ИИ“ (пристрастие = изкривяване на процеса на вземане на решения въз основа на предразсъдъци).
- Учителят дава пример: въз основа на оценка на клиента, направена автоматично от алгоритъм, банков служител може да реши дали клиентът може да кандидатства за кредит от тази банка или не. Ако алгоритъмът разчита на данни от време, когато нито една жена не е имала право да кандидатства за заем от банката, това може да даде индикация, че най-добрите клиенти за даване на заем са мъжете. Това води до дискриминация – т.е. неравно третиране – на жените. Така че работи срещу ЦУР 5 (Равенство между половете).
- Учителят обяснява, че други възможни проблеми са: проблеми с поверителността при използване на алгоритми за лицево разпознаване на обществени места или проблеми със сигурността и поверителността при използване на системи, които разработват глас и звуци като цифрови асистенти. Други примери включват трудности на алгоритмите за гласово разпознаване при точното идентифициране на команди от гласове на по-възрастни хора или по-млади потребители, тъй като алгоритмите се обучават върху възрастно население на средна възраст. Софтуерът за разпознаване на изображения или търсачките, базирани на изображения, могат да представят пристрастни резултати (вижте <https://en.unesco.org/artificial-intelligence/ethics/cases#biasedai>)

Директно преподаване

- Учителите разделят учениците на групи
- Всяка група ученици обсъжда един пример за отрицателни последици или рискове, причинени от ИИ приложения в реалния свят.
- Всяка група мисли за възможните последици от този пример в други области/по други ключови въпроси

Насочени упражнения и разбиране

- Всяка група, която използва материала, предоставен от преподавателя, измисля начин да представи примера, който тази група обсъжда, за да подчертае

начините, по които ИИ технологиите създават рискове, и има отрицателно въздействие върху разглеждания проблем и съответно срещу това приложение. Всеки ученик може да представлява аспект, агент или функционалност на цялата система.

Самостоятелно упражнение

- Учениците във всяка група изиграват представянето на примерите за приложение на ИИ и аргументите срещу неговото използване пред целия клас.

Приключване

- Останалата част от класа оспорва аргументите срещу използването на ИИ в тази конкретна област чрез критични коментари и въпроси. Преподавателите модерират дискусиата и изчистват всякакви погрешни декларации или съмнения, които другите учениците могат да имат.

Отстраняване на проблеми

Потенциален проблем	Потенциално решение
Примерите на ИИ за приложения на ИИ в системи от реалния свят са неясни на учениците или те не могат да съотнесат концепциите, които предварително са изучавали, с приложенията на системите от ИИ, които се опитват да обяснят.	Уверение от преподавателя и повторно обяснение на концепциите от високо равнище, ако е необходимо.
Учениците не желаят физически да изиграят приложение на технологиите за ИИ в реалността.	Партньорство с преподавател по театрално изкуство и превръщане на превъплъщаването като част от урока в по-подходяща среда, като например училищен театър или голямо пространство.
Учениците не желаят да оспорват примерите и аргументите, представени от една група	Преподавателят задава въпроси, които насърчават критичния дебат за специфични приложения на ИИ

Ресурси за учителите

Ресурси, които могат да се използват в клас за представяне на примери за ИИ приложения и етичното измерение

Заглавие	Описание	Линк	Език
Ethics & AI: Privacy & the Future of Work	Видео, което обяснява как етиката и поверителността са свързани в ИИ и как ИИ може да повлияе на бъдещите работни места.	https://www.youtube.com/watch?v=zNxx5gJtHLc	АНГЛИЙСКИ с автоматично генерирани субтитри
Ethics & AI: Equal Access and Algorithmic Bias	Видео, описващо потенциала на ИИ да помогне на обществото с акцент върху етиката, достъпа и справедливостта	https://www.youtube.com/watch?v=tJQSyzBUAew	АНГЛИЙСКИ с автоматично генерирани субтитри
ως λειτουργεί η Τεχνητή Νοημοσύνη - Τεχνητή Νοημοσύνη: Δεδομένα και Μεροληψία	Кратък видеоклип, който обяснява как са свързани данните за обучение в ИИ и отклонението	https://www.youtube.com/watch?v=7a2aNF-dRo8&list=PLIzOpQKPcu89NB4AdYYuixHPnN_2QocuZ&index=3	АНГЛИЙСКИ с гръцки субтитри
What's Going On With Facial Recognition?	Кратък видеоклип, който обяснява как разпознаването на лица може да се използва в помощ на обществото или може да се използва срещу поверителността на хората	https://youtu.be/BqQT4slOYA0	АНГЛИЙСКИ с автоматично генерирани субтитри
ИЗКУСТВЕНИЯТ интелект и ЗАЩО Илон МЪСК се обърна срещу себе си?	Видео за ИИ и Илон Мъск, включително други примери за ИИ приложения	https://www.youtube.com/watch?v=wFPp2IEQPog	български

Отрицателно въздействие от ИИ / рискове, породени от ресурси за ИИ:

Алгоритмични предразсъдъци

- [Амазон зарязва секретен инструмент за подбор на персонал чрез ИИ, който показва предразсъдъци спрямо жените](#)
- [Google плаща 2.6 милиона долара на служители заради заведени иски срещу несправедливи практики за наемане и заплащане на служители с предразсъдъци срещу жените и азиатците](#)
- [Инструменти за подбор на персонал и дискриминация за увреждания, ръководен от алгоритъм](#)
- Köchling, A., Wehner, M.C. Дискриминирани от алгоритъм: систематична проверка на дискриминация и справедливост с алгоритмично вземане на решения в контекста на подбор на ЧР и развитие на ЧР. *Bus Res* **13**, 795–848 (2020). <https://doi.org/10.1007/s40685-020-00134-w>

- [Wired. “Машини обучени чрез снимки се научават на сексистско гледище спрямо жените: алгоритмите показват тенденция да свързват жените с шопинг и мъжете със стрелба”.](#)
- [на италиански език] Festival dei Diritti Umani 2021 - Algoritmocrazia
“Nuove tecnologie, vecchi bias” - Образователна сесия
включваща ресурси за използването на изображения и алгоритми, цифрово гражданство, реч на омразата (вижте Модул 4)
<https://festivaldirittiumani.it/event/nuove-tecnologie-vecchi-bias/>
Вижте също: <https://festivaldirittiumani.it/irony-di-radheya-jegatheva/>
- [Център за демокрацията и технологията, „Отговорно използване на данни и технологии в образованието: ангажираност към общността за да се гарантира оказването на помощ вместо нараняването на учениците и на семействата“](#)

Проблеми, свързани с личната неприкосновеност

- [Център за демокрацията и технологията, модул за обучение за „Защита на личната неприкосновеност на учениците“](#)
- Stoilova, Mariya; Livingstone, Sonia; Khazbak, Rana (2021). Разследване на рисковете и възможностите за децата в дигиталния свят: Бърз преглед на наличните данни относно използването на интернет от страна на децата и резултатите от това, дискуссионни доклади от Иносенти № 2021-01, UNICEF, служба за изследователска дейност - Иносенти, Флоренция.
<https://www.unicef-irc.org/publications/1183-investigating-risks-and-opportunities-for-children-in-a-digital-world.html>

За повече ресурси на UNICEF във връзка с правата на децата в дигиталната епоха:

<https://www.unicef-irc.org/research/child-rights-in-the-digital-age/>

- Спасете децата (2020 г.) Дигитална защита за деца мигранти и деца сираци: обобщение на текущия контекст и тенденции, потенциални рискове и практически следващи стъпки.
https://resourcecentre.savethechildren.net/node/18477/pdf/mdi_digital_final_rgb_rev_091220.pdf

[позитивен пример]

Мисията на проект Тор е да подпомогне общественото разбиране на технологиите за лична неприкосновеност → обучения в страните от Южното полукълбо

<https://blog.torproject.org/reaching-people-where-they-are>

Наблюдение

- BBC File on 4 Podcast (2020). „Лицевото разпознаване се въвежда във Великобритания. Това обаче жизненоважен инструмент срещу престъпността ли е - или заплаха за невинните?“ <https://www.bbc.co.uk/programmes/m000dpkt>
- Използване на лицево разпознаване в училищата:

<https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal>

<https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/>

(вижте също Доклада за автоматизация на обществото, 2020 г.)

- В Швеция има два пилотни проекта, които фокусират вниманието си върху използването на ИИ и лицево разпознаване за учениците за административни цели, като позволява на преподавателите да посветят своето време на преподаване
<https://www.electronicshobby.com/products/artificial-intelligence/facial-recognition-tested-in-swedish-high-school>
- Структурна и когнитивна промяна на училищата и учебния процес, вижте Андреевич и Селунин, технология за лицево разпознаване в училищата: критични въпроси и притеснения,
<https://www.tandfonline.com/doi/full/10.1080/17439884.2020.1686014>

Екологични и икономически последици от системи базирани на ИИ

- BBC The Documentary Podcast (2019). „Ще сложи ли ИИ край на развитието? Роботизацията ще предотврати ли по-бедните държави да поемат традиционния път към просперитет?“
<https://www.bbc.co.uk/programmes/p075nqyy>
- Sahota, Neil. „Отвъд ИИ: нуждата от създатели на решения в създаването на устойчив свят“. Forbes. 24 април 2019 г.
<https://www.forbes.com/sites/cognitiveworld/2019/04/24/beyond-a-i-the-need-for-solutionists-in-creating-a-sustainable-world/>
- Jones, Phil. „Бежанци помагат за развитието на машинното самообучение в Майкрософт, Фейсбук и Амазон“. Restofworld. 22 септември 2021 г.
<https://restofworld.org/2021/refugees-machine-learning-big-tech/>
- Документален филм на Дойче Веле (2021 г.) „Невидимите – експлоатация в дигиталния свят на работата“ <https://www.youtube.com/watch?v=o-HphMeZR3k>
- Strubell et al. 2019. „Енергийни и политически съображения за дълбоко учене в невролингвистично програмиране.“ Протокол от 57-то годишно събрание на асоциацията за изчислителна лингвистика.
<https://www.aclweb.org/anthology/P19-1355>.
DOI: 10.18653/v1/P19-1355

-
- Schwartz et al. 2019. “Зелен ИИ.” <https://arxiv.org/abs/1907.10597>
 - Защо трябва да се грижим за въздействието върху околната среда на ИИ
<https://www.forbes.com/sites/forbestechcouncil/2020/08/17/why-we-should-care-about-the-environmental-impact-of-ai/?sh=766825ef56ee>