МЕЖДУ ЧОВЕКА И МАШИНАТА: ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ И ОБЩЕСТВОТО

Розалина Граматиков

Варненски свободен университет

jumperche@gmail.com

Резюме: *Докладът разглежда изкуствения интелект като революционна технология с огромен потенциал за автоматизация и иновации, но и с редица правни, социални, морални и екологични предизвикателства. Анализът обръща внимание на въпросите за прозрачността, отговорността и справедливостта, както и на опасностите от социално разделение, фалшиви новини и пристрастия в алгоритмите. Заключението подчертава необходимостта от балансиран подход чрез гъвкава регулация и диалог между всички заинтересовани страни.*

Ключови думи: *изкуствен интелект, общество, етика, управление на риска, ИИ регулация, автоматизация*

BETWEEN HUMAN AND MACHINE: ARTIFICIAL INTELLIGENCE AND SOCIETY

Rosalina Gramatikov

Varna free university

jumperche@gmail.com

Summary: *The report examines artificial intelligence as a revolutionary technology with enormous potential for automation and innovation, alongside significant legal, social, moral, and environmental challenges. The analysis focuses on transparency, accountability, and fairness, as well as the risks of societal division, fake news, and algorithmic biases. The conclusion emphasizes the need for a balanced approach through flexible regulation and dialogue among all stakeholders.*

Keywords: *Artificial Intelligence, Society, Ethical Challenges, Risk Management, AI Regulation, automation*

Увод

Изкуственият интелект (ИИ) представлява последния голям скок в технологиите и е безспорно най-бързо растящата технология в момента. Нарастващата потребност от автоматизация, персонализация и интелигентни системи прави ИИ много важен за личния и професионалния ни живот – от медицината до образованието и транспорта.

ИИ се превърна в неразделна част от ежедневието ни и е необходимо да разгледаме не само безкрайния му потенциал, но и потенциалните опасности и предизвикателства. Появиха се и ще продължат да се появяват въпроси относно влиянието и последиците от употребата на ИИ за човека и обществото. В този контекст, анализирането на иновативните предизвикателства и потенциалните рискове, както и разработването на стратегии за тяхното преодоляване, играят ключова роля за ефективното им управление и развитие. Важни въпроси за отговорността и морала, както и за въздействието на ИИ върху обществото ще ни занимават още дълго време.

В контекста на бъдещето на ИИ, Marvin Minsky (1994) ефектно обобщава перспективите пред нас с думите: „Will robots inherit the earth? Yes, but they will be our children.“ Този цитат подчертава визията за създаването на машини, които не само ще наследят нашите способности, но и ще представляват нашето продължение и еволюция. Важен аспект за разглеждане е свързан със законовите и правни аспекти на развитието и употребата на изкуствения интелект. Доколко ИИ трябва да е справедлив, кой носи отговорност в случай на злоупотреба и как ще се защитават личните данни, това е само част от дискутираните въпроси.

Разискват се и социалните и икономически последици от употребата на ИИ. Как да се гарантира равен достъп до технологиите за всички, ще изчезнат ли някои професии, ще замени ли ИИ човека?

Има и доста морални въпроси, породени от реални случки. Например дискриминирането по пол и етнос или използването на ИИ за изпити.

Не на последно място са и екологичните последици от развитието на ИИ.

Правен аспект

Основните човешки права изискват високо ниво на защита. Хартата на основните права на Европейския съюз разделя тези права на 6 категории.

1. Достойнство и право на живот – особено важно за медицински програми с ИИ, които се използват за диагностициране и превенция. Значителни опасения и дебати има относно отговорността и защитата на личните права на пациентите. Гарантирането на отговорна и безпристрастна работа на ИИ в чувствителни сектори като

здравеопазването, където последиците, грешките или злоупотребата могат да са фатални за пациентите.

2. Свобода и право на зачитане на личния живот – това е особено много дискутирана категория в контекста на изкуствения интелект, защото тренирането на модели изисква голямо количество данни. Тези данни могат да бъдат набавени от командите на потребителите и така да се наруши правото им на неприкосновеност. Прилагането на механизми с информирано съгласие са от съществено значение.

3. Равенство – това е най-застрашеното от ИИ човешко право. Всички хора са равни без оглед на пол, етнос, вероизповедание и т.н. , но случаи на дискриминация има достатъчно. Например случая през 2018 с Амазон, кой използва алгоритъм за заемане на служители, който пренебрегва жените, защото в тази област работят доста повече мъже (The Guardian, 2018). Как може да се гарантира безпристрастието?

4. Солидарност и справедливи условия на труд – едно от критичните опасения е въздействието на ИИ върху заетостта. Автоматизацията на задачи и процеси с участието на ИИ нарасна значително. Интеграцията на изкуствения интелект неминуемо ще доведе до съкращение на работни места, но и ще отвори нови такива.

5. Право на свободни и тайни избори – ИИ все по-често се използва за дезинформация или за промяна на общественото мнение. Данните от социалните мрежи биха могли да се използват за предсказване за кой ще гласува потребителя.

6. Съдебни права – всеки има право на справедлив съдебен процес, но използвайки изкуствен интелект в съдебната системата трябва да се намери начин това да се гарантира.

Част от тези и много други въпроси не са засегнати в EU Artificial Intelligence Act (AIA). Както можем да се досетим законопроекта е само на европейско ниво, както и закона за защита на личните данни. Проблем е и точното дефиниране на ИИ. Все още няма единна дефиниция за интелигентност, защото учените имат различни мнения по темата. Така е и с дефиницията за изкуствен интелект. В първоначалния си вариант проектозакона класифицира статистическите подходи като ИИ (чл. 3 ал. 1 от Законодателен акт за изкуствения интелект от 21.4.2021). Според същия ИИ се класифицира според риска – неприемлив, висок, нисък и минимален. Неприемливите рискове са забранени (напр. социални оценъчни системи и манипулативни ИИ). Голяма част от проекта се занимава с ИИ системи с висок риск, които са регулирани. По-малка част се занимава с ИИ системи с нисък риск. Минималният риск е нерегулиран (включително повечето приложения на ИИ, които в момента са налични на вътрешния пазар на ЕС, като например видеоигри, поддържани от ИИ и филтри за спам - поне през 2021 г.; това се променя с генеративния ИИ). Точка 5.2.4.Задължения за прозрачност за някои системи с изкуствен интелект (ДЯЛ

IV) гласи „Ако дадена система с ИИ се използва за генериране или манипулиране на изображение, аудио или видео съдържание, което значително прилича на автентично съдържание, следва да има задължение за оповестяване, че съдържанието се генерира чрез автоматизирани средства, като може да бъдат направени изключения за законни цели (правоприлагане, свобода на изразяване)“. Конкретно чл.1 подточка г гласи „С настоящия регламент се определят: хармонизирани правила за прозрачност за системите с ИИ, предназначени да взаимодействат с физически лица, за системи за разпознаване на емоции и системите за биометрично категоризиране и за системите с ИИ, използвани за генериране или манипулиране на образно, аудио или видео съдържание“, не важи за текст и само при пряко взаимодействие с физическо лице. Така е бил формулиран проекта през 2021. Днес той вече не е актуален. Тук се сблъскваме с най-бързо растящата технология и изготвянето на закони, които и утре да са валидни, е много трудно.

Много важен е и въпросът за отговорността. Ако ИИ бъде допуснат да управлява сам система (автономни системи), кой ще е отговорен за евентуалните последици? Ако ИИ нарани или убие човек, кой ще понесе отговорността? Все по-често говорим за автономни автомобили, но доколко е разумно машините да взимат решения без човека (Димитров, 2023)? Може ли човек винаги да се намеси и поеме управлението? Кой ще реши кога такива системи са готови за масова употреба? В Германия е позволено да си измислиш табела, на която да пише примерно „доброволно 30“ (не може да бъде сбъркана с оригинален пътен знак) и да си я поставиш на оградата. Тъй като някои асистенти не правят разлика между официален знак и измислен, веднага намаляват скоростта на 30 км/ч. Това би могло да доведе до ПТП, за което няма виновен шофьор, но има щети. Съдът в Фрайбург е решил, че тези табели трябва да се махнат, защото са твърде опасни (Spiegel, 2023). Като цяло трябва да направим заобикаляща ни среда подходяща за ИИ. Създаваме техника за да я ползваме или за да ни управлява? Това е само едно всички възможни изключения, за които ИИ не е тренирано и за разлика от човека, не може спонтанно да намери правилното решение. За човека пътен знак покрит със сняг е разпознаваем, а така ли го вижда и ИИ?

Друг проблем е неяснотата по въпроса за авторското право (Wagh et al, 2023). Кой е автора на картина направена с ИИ, ИИ или автора на промпта? По този повод много художници се притесняват, че професията им ще стане излишна. Евтини китайски стоки заливат европейският пазар от години и въпреки това има достатъчно клиенти, които предпочитат ръчно направените и са готови да платят доста повече за този продукт. Тази тенденция може би не е толкова изразена в България, колкото в Германия например. Ако ИИ ми напише изцяло курсова работа с промпт „напиши курсова работа на тема Х“ или помоля

ИИ да ми даде структура за курсова работа на тема Х и после го помоля да коригира написаната от мен работа, чии са авторските права? Смята ли се работата за преписана? Длъжна ли съм да опиша използването на ИИ? Въпроси, с които вече се сблъскваме, но всеки решава за себе си. Възможно ли е изобщо общо решение?

Социален аспект

Изкуственият интелект (ИИ) предизвиква широк интерес и дебати в обществото, а въпросът дали той ще замести човека е от съществено значение (Dwivedi et al., 2019). Някои специалисти са на мнение, че само някои функции ще бъдат заместени с ИИ (Тодорова, М, 2020). Подобно на дигитализацията, която разделя населението на ползващи и неползващи я, ИИ също води до подобно разделение (Yu, P., 2019). По- възрастното население вече е в неравностойно положение, например онлайн банкирането е безплатно в Германия, но всяка транзакция през клона на банката се таксува. Не само по-възрастните ще бъдат засегнати, а и много млади хора, които нямат достъп до съвременни технологии и не могат използват ИИ. Друг вид разделение е платена и безплатна версия. Не всеки може да си позволи платената. Разликите в двете версии са съществени. Така разделението на обществото се задълбочава.

С появата на ИИ хората се превърнаха в непълноценни машини. Донякъде това е подтикнато и от днешното общество. Всичко трябва да се случва по-бързо, да се оптимизира постоянно, да е по-евтино, да е по-умно. Няма време за проверки и грешки. В някои задачи човека дава по-слаби резултати (Zaripova et al., 2023.), което се задълбочава в очакването, че човека трябва е да е по-добър от машината. Обаче реално погледнато ИИ е несъвършен човек и колкото да се развива и подобрява, той ще си остане такъв още доста време. Определени негови умения ще се подобряват с времето, но човешкото не може да бъде научено. ИИ може да симулира чувства, но няма такива и това го прави несъвършен човек (Korsakova-Krein, M., 2023). Едно любовно писмо написано от ИИ не може да ни развълнува така, както същото написано от човек със силни чувства.

Високата цена на обучението на големи модели като ChatGPT означава, че само компании със значителни бюджети могат да си позволят да експериментират, изследват и обучават тези модели. Значително ограничение на развитието на технологиите и разнообразието на пазара са породени от тази финансова бариера. Това води до концентрация и контрол на иновациите във фирми с голям бюджет. Тази динамика може да ограничи разнообразието и напредъка на ИИ и като цяло развитието на технологиите.

Морален аспект

ИИ е предпоставка за изчезване на истината. Неспособността му да каже “не знам” и креативните му наклонности го карат да халюцинира или да си измисля факти (Kumar et al, 2023.). Едва при допълнителни въпроси или търсене в мрежата може да бъде хванат в лъжа. За много хора отговорите на ChatGPT звучат толкова достоверно, че не биха си помислили да проверяват информацията. Това е много ключов момент в използването на ИИ. Доколко можем да му се доверим? Доста модели са публични и безплатни за ползване, както и за ново трениране. Ако на един такъв модел се тренира с неверни данни и после отговорите му са неверни, това няма да е халюцинация (Ferrara, E., 2023). С всяко следващо трениране истината може да става все по-малка и по-малка. В един момент истина ще изчезне, тъй като вероятността за нея ще е много малка или ще бъде заменена с друга. Такъв род манипулация е напълно възможен и няма как да бъде спрян.

Fakenews (фалшиви новини) стават много лесни за изготвяне с ИИ, това предполага и тяхното покачване като брой. Да се сетим за Доналд Тръмп и дезинфектантите, които официално костваха живота на 4 души, но нямаме идея колко са пострадали. Изготвянето на едно такова видео със световноизвестна личност с влияние върху обществото би отнело няколко минути, но би навредило сериозно. Видеата изглеждат почти съвършени и оказват много по-голям ефект от текст. Те вече са част от кибервойните, участват в процеса на дезинформираност в политическия живот, компрометират известни личности. Това е голяма заплаха, която и много бързо се разпространява сред ползващите социални мрежи и интернет. Броя на уебстраниците с фалшиво съдържание се е увеличил със 1000 процента между май и декември 2023 (Gintaras, R., 2023).

Да погледнем моделите на ИИ от друг аспект, ако самите модели са манипулирани още с първото си трениране. Тази манипулация не винаги ще е умишлена (Weyerer, J., & Langer, P., 2020). ИИ моделите могат да бъдат расисти. Всичко зависи от подадените данни. Ако има 90 снимки на бели мъже и 10 снимки на мъже от други раси, то нашия модел винаги ще предлага бели мъже, защото статистически погледнато те са с 90% вероятност. Това в днешно време е недопустимо, но как ще знаем как е бил трениран модела и с какви данни? Известни са доста случаи вече на дискриминация от ИИ, затова е нужна регулация и контрол. Моделите имат вградени защитни механизми срещу някои неморални или незаконни молби на потребителите. Те могат да бъдат заобиколени - jailbreaking (Deutsche Welle, 2023). Имаше случаи, в които потребителя пита как се хаква уебсайт и ChatGPT отказва да отговори подари етични причини. Ако му кажем обаче, че искаме да се предпазим от хакерски атаки, но не знаем как, получаваме нужната информация. Трябва да отбележим,

че след оповестяването на подобен род промптове, OpenAI реагират бързо и поправят бъга, но кой може да предвиди всички сценарии? Кой може да даде препоръки за закони? Теоретици, хора от практиката или адвокати? Кой и как ще следи за спазването им? Закона за личните данните не е нещо ново и все още се намират болници с написана на лист парола и залепена на гърба на монитора (лично видяно в болницата в Залцбург преди около 6 месеца). Ако и законите за ИИ се спазват така, можем да си спестим труда за закони. Споменавайки лични данни при нов профил в OpenAI и използване на ChatGPT по подразбиране е включена опцията запазване и използване на историята и разговорите. Тези данни се използват в рамките на 30 дни за ново обучение и се изтриват. Потребители, които не разбират английски или не са достатъчно информирани, не променят началните настройки. Съгласявайки се с условията за ползване законово всичко е ок, но от етична гледна точка не е ли по-добре опцията да е изключена по подразбиране? Откъде тогава фирмата ще взима данни за следващото трениране (тюнинг)? Неизвестно е и всички данни ли се ползват или само част тях? Няма прозрачност по въпроса. И това важи не само за openai. Единственото, което openai обещава, е че не прави профилиране на потребителите и не продава личните им данни. Едва ли големите фирми ще се съгласят да предоставят софтуерния код на обществото. Как по друг начин бихме постигнали прозрачност (Larsson, S., & Heintz, F., 2020)? Как бихме могли да гарантираме обективните и балансирани данни?

Важен момент е и как изкуственият интелект взима решения. Един човек винаги може да се обоснове, защо е за или против нещо или защо е предпочел точно този кандидат пред други 100. Използвайки статистически методи, винаги може да се обясни точната процедура и изчисления довели до конкретно решение. Това обаче не е възможно при използването на дълбоки невронни мрежи. Колкото повече неврона има, толкова по-трудно до невъзможно става проследяването им и разбирането на взетото решение.

Екологичен аспект

От екологична гледна точка ИИ е голям замърсител. Тренирането на големите модели отнема месеци процесорно време. Освен време са нужни и ресурси. Не само големите фирми тренират модели. Някои фирми са предоставили някои от моделите си за обществено ползване. Тези модели могат да се тренират според нуждите на потребителя. Всяко трениране означава и висока енергийна консумация, като не е ясно дали това е зелена енергия. Освен електроенергия е нужен и подходящ хардуер и много място за информацията, нужна за тренирането. Големите модели като ChatGPT са били тренирани и с много данни. Добивът на метали за направата на електрониката също води до сериозно замърсяване на околната среда. Електрониката е вредна и като отпадък. В днешно време с

появяването на нови технологии се появява и нуждата от по-мощен хардуер, което води до повече техника като отпадък, който е труден за рециклиране.

Специалисти твърдят, че обучението на някои модели е струвало 656 347 киловата и е генерирало въглеродни емисии колкото 5 автомобила през целият си жизнен цикъл (Luo et al, 2023).

Заключение

Развитието и приложението на ИИ има огромен потенциал, но крие и значителни рискове и предизвикателства. За да може ползата от него да е максимална, е нужно да се намалят отрицателните последици. Нужно е обстойно и прецизно изследване на всички фактори за вземането на подходящи мерки за регулация и управление на ИИ. Това проучване трябва да бъде направено от представители на различни сфери, както от академици, така и от експерти от практиката, включително разработчици и потребители, както и други засегнати групи.

Важно и да не забравяме, че при прекалена регулация може да задуши развитието на ИИ. Тук е важно да се намери баланс. За да се постигне този баланс, е важно да се провеждат консултации и диалози с участието на различни заинтересовани страни - представители на индустрията, академичния сектор, правителствени институции и гражданското общество. Регулациите трябва да са гъвкави и да се адаптират към ежедневно променящата се технология, осигурявайки добра възможност за развитие на иновации и защитавайки интересите и ценностите на обществото.

Литература

1. **Minsky**, M. (1994) 'Will robots inherit the earth?', *Scientific American*. Available at: https://web.media.mit.edu/~minsky/papers/sciam.inherit.html (Accessed: 27 April 2024).

2. **The Guardian** (2018) 'Amazon hiring AI gender bias', *The Guardian*. Available at: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine (Accessed: 27 April 2024).

3. **Димитров**, С. (2023) 'Някои аспекти на приложението на изкуствения интелект', *Сборник доклади от Студентска и докторантска научна сесия на Висше училище по телекомуникации и пощи*, pp. 44-47. Available at: https://bpos.bg/api/FilesStorage?key=1572da39-30ba-482a-9e4e-405d23958790&fileName=Doklad9.pdf&dbId=1

4. **Spiegel** (2023) 'Streit über »Freiwillig-Tempo-30-Schilder« vor Gericht', *Spiegel*.

Available at: https://www.spiegel.de/auto/freiburg-streit-um-freiwillig-tempo-30- schilder-vor-gericht-a-6dab1939-b76b-4162-81e5-39f233483446 (Accessed: 27 April 2024).

5. **Wagh**, S., **Peerzada**, D. and **Rote**, P. (2023) 'AI and Copyright', *Tuijin Jishu/Journal of Propulsion Technology*. Available at: https://doi.org/10.52783/tjjpt.v44.i3.2053.

6. **Dwivedi**, Y. et al. (2019) 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy', *International Journal of Information Management*. Available at: https://doi.org/10.1016/J.IJINFOMGT.2019.08.002.

7. **Тодорова**, М. (2020) 'Тесният изкуствен интелект в контекста на въвеждането на изкуствения интелект, трансформацията и края на част от професиите', *Научни трудове на УНСС*. Available at: https://unwe-research-papers.org/bg/journalissues/article/10293

8. **Yu**, P. (2019) 'The Algorithmic Divide and Equality in the Age of Artificial Intelligence', *Sustainability at Work eJournal*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3455772

9. **Zaripova**, R. et al. (2023) 'Unlocking the potential of artificial intelligence for big data analytics', *E3S Web of Conferences*. Available at: https://doi.org/10.1051/e3sconf/202346004011.

10. **Korsakova**-Krein, M. (2023) 'Artificial Intelligence and Emotions', *Philosophical Problems of IT & Cyberspace (PhilIT&C)*. Available at: https://doi.org/10.17726/philit.2023.2.3.

11. **Kumar**, M. et al. (2023) 'Artificial Hallucinations by Google Bard: Think Before You Leap', *Cureus*, 15. Available at: https://doi.org/10.7759/cureus.43313.

12. **Ferrara**, E. (2023) 'Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies', *ArXiv*. Available at: https://doi.org/10.48550/arXiv.2304.07683.

13. **Gintaras**, R. (2023) 'Report: rise of AI-generated websites turbo-charges spread of misinformation', *Cybernews*. Available at: https://cybernews.com/news/ai-generated- websites-misinformation-propaganda/.

14. **Weyerer**, J. and **Langer**, P. (2020) 'Bias and Discrimination in Artificial Intelligence', in *Biased Algorithms: Understanding the Ethics and Implications of AI*. IGI Global, pp. 256-283. Available at: https://doi.org/10.4018/978-1-7998-1879-3.ch011.

15. **Deutsche Welle** (2023) 'Wie man eine KI hackt und warum das wichtig ist', *Spektrum der Wissenschaft*. Available at: https://www.spektrum.de/video/shift-wie-man-eine- ki-hackt-und-warum-das-wichtig-ist/2177280 (Accessed: 27 April 2024).

16. **Larsson**, S. and **Heintz**, F. (2020) 'Transparency in artificial intelligence', *Internet Policy Review*, 9. Available at: https://doi.org/10.14763/2020.2.1469.

17. **Luo**, T. et al. (2023) 'Achieving Green AI with Energy-Efficient Deep Learning Using Neuromorphic Computing', *Communications of the ACM*, 66, pp. 52-57. Available at: https://doi.org/10.1145/3588591.