

„СВЕТЛАТА“ И „ТЪМНАТА“ СТРАНА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Любомира Парижкова, Марина Илиева, Светослав Славов,
Юрий Годоров, Марио Тренчев

Резюме: Предмет на този анализ са възможностите и заплахите, които произлизат от употребата на изкуствен интелект, конкретно що се отнася до системите, генериращи съдържание, формулирани в Законодателния акт за изкуствения интелект на ЕС от август 2024. Текстът представя експеримент, проведен в часовете по дисциплината „Нови форми на управление на знанието“ с титуляр доц. д-р Любомира Парижкова от студенти от специалност „Библиотечен и информационен мениджмънт“, втори курс. Демонстрират резултатите от поставената задача студентите да изследват двете страни на влиянието на изкуствения интелект върху живота на хората. В експеримента е включен ChatGPT, а студентите дават свои контрааргументи на представените от него тези.

Ключови думи: експеримент, студенти, ChatGPT, възможности, заплахи, дискусия.

THE "LIGHT" AND "DARK" SIDE OF ARTIFICIAL INTELLIGENCE

Lubomira Parikova, Marina Ilieva, Svetoslav Slavov, Yuri Todorov, Mario Trenchev

Abstract: The subject of this analysis is the opportunities and threats that arise from the use of artificial intelligence, specifically as it relates to content-generating systems, as articulated in the EU Artificial Intelligence Legislative Act of August 2024. The text presents an experiment carried out in the class "New Forms of Knowledge Management" with the holder Assoc. Prof. Dr. Lyubomira Parikova by students of Library and Information Management, second year. Demonstrate the results of the task set students to explore both sides of the impact of artificial intelligence on people's lives. ChatGPT is included in the experiment, and students provide their own counterarguments to the theses it presents.

Key words: experiment, students, ChatGPT, opportunities, threats, discussion.

Въведение

Изкуственият интелект е специалност в областта на компютърните науки, която се занимава със създаването на системи, които могат да възпроизвеждат човешкия интелект и способностите за решаване на проблеми (ChatGPT и AI от нулата, с. 13).

Като всяка иновация и изкуствения интелект поражда поляризирани мнения. От една страна са неговите апологети, които го славославят и преекпонират неговите достойнства и предимства. От друга, са песимистите, които са готови за всички злини и катаклизми по света да обвинят изкуствения интелект. В този контекст има и отговорни ползватели, които подхождат интелигентно и с аналитичност към тази иновация. Изкуственият интелект и по-конкретно онова негово проявление като ChatGPT, Gemini (Google), Copilot (Microsoft) (системи с общо предназначение – The EU AI Act, 2024) се влияе от начина на формулиране на въпросите към него и от инструкцията, която му се

дава. Затова и коректното питане е в основата на коректните отговори. Така наречената рамкова инструкция (prompt) е от важно значение за крайния резултат.

Историята на изкуствения интелект (ИИ) започва още през 50-те години на 20-ти век. Терминът „изкуствен интелект“ е въведен от Джон Маккарти през 1955 г. по време на научен семинар в САЩ и оттогава се възприема като начало и развитие в научна дисциплина, която се изучава в университетите. През 50-те години на 20-ти век се разработват софтуерни програми, които могат да изиграят партия шах с „човек и машина“. ИИ може да решава елементарни математически задачи и уравнения. През 60-70 години на същия век се разработват първите експериментални системи, които могат да вземат решения в области със специфични проблеми като медицина и инженерство. С всяко десетилетие ИИ се усъвършенства, подобряват се неговите качества, бързина и се задълбочава способността му за решаването на все по-сложни задачи, алгоритми и специфични проблеми. Предвид бързината, с която се развива всяка година ИИ, никой не може да предвиди докъде ще стигне неговото развитие и усъвършенстване и дали ще достигне интелекта на човек. Често се задава въпроса дали ще поеме умствените му дейности. Тук вече, според песимистите, от една партия на шах може да се обърне в нещо много по-страшно. Човек трябва да се замисли и да осъзнае, че с това бързо темпо на развитие на ИИ, освен добра страна, има и негативни последици или трябва да се изследват „светлата и тъмната страна“ на изкуствения интелект.

В часовете по дисциплината „Нови форми на управление на знанието“ с титуляр доц. д-р Любомира Парижкова на студенти от специалност „Библиотечен и информационен мениджмънт“, втори курс бе поставена задачата да изследват двете страни на влиянието на изкуствения интелект върху живота на хората. Студентите Марина Илиева и Марио Тренчев бяха отговорни за разкриване само на позитивите, а Светослав Славов и Юрий Тодоров съответно на негативите. Експериментът се проведе на 17 септември 2024 г. и продължен в следващите седмици за доработка от двата екипа. Първоначално двете двойки написаха своите предложения, а след това бяха поканени да се допитат до свободно избран от тях ИИ за мнение по същите въпроси. Важно е да се спомене, че първата група получи резултат 80% съответствие с отговорите на ИИ, а втората (заплахите) – 60 % се покриват предложенията на студентите и отговорите на самия ChatGPT.

Въз основа на прочетените материали от блогове, посветени на изкуствения интелект, студентите дават следното определение за ИИ:

Изкуственият интелект (ИИ) е област на компютърните науки, която се фокусира върху създаването на системи, способни да изпълняват конкретни задачи за много кратко време за изпълнение. Тези задачи включват разсъждение, обучение, решаване на даден проблем, възприемане и разбиране който обикновено изискват човешки ресурс като умствена дейност.

Първата група написа за ползите от изкуствения интелект следното:

1. Може да предоставя персонализирани препоръки/съвети на потребителя в различни сфери (TikTok, Instagram, YouTube);
2. Подпомага производството;
3. Предоставяне на информация бързо и лесно (ChatGPT);
4. Революция в автомобилостроенето (автопилоти);

5. Подпомага за точността на изследванията в медицината;
6. Екология – намаляване на замърсяването на околната среда;
7. Клиентско обслужване;
8. Сигурността в интернет.

А втората – за рисковете/вредата от изкуствения интелект:

1. Може да бъде манипулиран – грешна информация;
2. Може да се използва за извършването на зловредни действия;
3. Да доведе до закърняване на способността на хората да мислят/насърчаване на шаблонното мислене;
4. Кой пише алгоритъма? – кой интерес за него е водещ;
5. Не може самостоятелно да взема решения за процеси;
6. Не машините да служат на хората – можем да изгубим контрол;
7. Човек споделя лична информация.

След като естествения интелект изброи ползите и вредите, студентите се допитаха до ChatGPT сам да оцени «светлата» и «тъмна» страна на ИИ. Съвместимостта на групата за ползите е 80%, а на тази за вредите е 60% – общо 70% съвместимост между отговорите и тези на ChatGPT.

В следващия параграф се представя обобщение от самите студенти, направили експеримента, както и техните контрааргументи, чрез които всяка група опровергава отговорите на ChatGPT.

Ползите от ИИ според самия изкуствен интелект (ChatGPT) и контрааргументите на авторите

1. Автоматизация на процеси. ИИ може да автоматизира рутинни и повтарящи се задачи, което спомага за повишаване на ефективността и намаляване човешката грешка. Това е особено полезно в производството, където роботите с ИИ може да извършват операции по-бързо и с по-голяма точност.

Контрааргумент: Автоматизацията елиминира ролята на човека в много дейности, в които има заети лица и ще повиши безработицата, която ще произтича от закриване на редица длъжности и професии.

2. Подобро обслужване на клиенти. Чатботовете и виртуалните асистенти, задвижвани от ИИ, могат да отговарят на клиентски запитвания 24/7, като осигуряват по-бързо и ефективно обслужване, без да се налага човешка намеса.

Контрааргумент: Чатботовете и виртуалните асистенти не са оптимизирани за отговори на въпроси и съдействие по конкретен проблем на клиента, а са разработени за тематично представяне на потребителски заявки за търсене. Те дават отговори “по шаблон” и не заместват пълноценно поддръжката от жив оператор.

3. Анализ на големи количества данни. ИИ има способността да обработва и анализира огромни количества данни много по-бързо от човек. Това позволява откриването на скрити тенденции и модели, което може да помогне бизнеса в процеса на вземане на решения.

Контрааргумент: Интерпретацията и практическото приложение на този анализ, както и вземането на решения, свързани с бизнес преобразувания, иновации и сътрудничество с партньори си остава изключително предимство на човека, тъй като ИИ не може самостоятелно да формулира решения.

4. Прогнозиране и оптимизация. Чрез машинно самообучение ИИ системите могат да правят точни прогнози за бъдещи събития, като пазарни тенденции или потребителско поведение. Това е от съществено значение за подобряване на производствени процеси, логистиката и маркетинговите стратегии.

Контрааргумент: Машинното самообучение не е толкова развито и може да прави погрешни изводи за причинно-следствените връзки в процеси. При грешно разработени програмни алгоритми за анализ на данните, поради едностранно, несъщинско познаване на процесите, ИИ може да създаде неверни прогнози и да предложи несполучливи оптимизации.

5. Здравеопазване. ИИ се използва за диагностика на заболявания, анализ на медицински изображения и разработване на медицински лечения. Алгоритмите могат да идентифицират модели, които биха били трудни за откриване от лекарите, което помага за по-ранна и по-точна диагноза.

Контрааргумент: Не трябва да се абсолютизират предложенията на ИИ за анализ на медицински изображения и предложенията му за медицински лечения. Силната диагностика, която може да се постигне чрез ИИ и предложеното от него лечение трябва да се потвърдят от доктори, защото отговорността в медицината е на човека (лекаря).

6. Образование. Чрез адаптивни платформи за обучение ИИ може да персонализира образователния процес спрямо нуждите на всеки ученик, предоставяйки по-ефективни и подходящи методи на обучение.

Контрааргумент: Преминава се от непосредствен (преподавател–учащ) към опосредствен образователен процес (преподавател–платформа–учащ), което прави критично важно дали платформите за обучение са качествени, в т.ч. адаптирани ли са към целите и методите на обучението и резултатни ли са за придобиване на знания. Освен това ИИ не може да оценява на този етап отговори на отворени въпроси от тестове.

7. Сигурност и киберсигурност. ИИ може да открива кибератаки и аномалии в реално време, което спомага за защита на информацията и предотвратяване на заплахи. Също така, използването на ИИ в лицевото разпознаване и биометричните технологии подобрява физическата сигурност.

Контрааргумент: Фактът, че ИИ може да борави с личната информация на клиентите си, в частен интерес на създателите си, без ефективен контрол на тази дейност. Практиките, по които програмистите събират и споделят данните ни, може да бъдат определени от тях по занижени мерки за сигурност и поверителност, което да доведе до изтичане на чувствителни лични данни на хората, които може да се обработват, разпространяват и ползват по користен начин, без оглед на законната им защита. Нужно е ползването на лични данни от ИИ да бъде стриктно регламентирано с международен регламент.

8. Подобро вземане на решения. Чрез събиране и анализ на данни ИИ може да предостави подкрепа за вземане на по-информирани и прецизни решения в различни сфери, включително бизнес, финанси и публичния сектор.

Контрааргумент: В зависимост от алгоритъма ИИ може да предложи неправилен ракурс за решения, вследствие на едностранен анализ на данни, преувеличаване на несъществени фактори или омаловажаване на важни фактори за управленски решения.

9. Опазване на околната среда. ИИ може да помогне за оптимизиране на ресурсите, управление на енергийни системи и проследяване на изменението на климата, като предоставя данни за по-устойчиво развитие.

Контрааргумент: В ИИ може да бъдат заложени приоритети, които са в противоречие с усилията на световната общност за енергийна ефективност, възстановяване и опазване на околната среда, и защита на климатичното равновесие. Причина за това може да бъде частен интерес на компаниите - създатели на ИИ.

10. Иновации и нови продукти. Изкуственият интелект спомага за създаване на нови технологии и иновации, които водят до разработването на нови продукти и услуги, които по-рано не са били възможни. Ползите от ИИ са значителни, като допринасят за ускорено развитие и повишаване на качеството на живот.

Контрааргумент: Новите технологии и иновации трябва да бъдат поставени в служба на човечеството, да се запази “в човешки ръце” контрола върху процесите на производство и обслужване, и да се насърчава развитието на световната икономика и на заетостта. ИИ не дава гаранции за това, напротив. ИИ може да накърнява интересите на обществото, защото се развива от частни компании, изцяло автоматизираните процеси, основани на ИИ може да доведат до лоши резултати и информационни сринове, чатботовете засега не дават пълна поддръжка на клиентите си, при бъдещото широко навлизане на ИИ редица човешки дейности ще станат излишни, което ще увеличи безработицата и ще породи социално недоволство. Новите технологии и иновации, предложени от ИИ трябва да се изпробват за надеждност и ефективност в реални условия/ в лабораторни условия, сходни с реалните, а крайната преценка за внедряването им трябва да бъде на хора-разработчици.

Вредите от ИИ според самия него и контрааргументите на авторите

1. Загуба на работни места. ИИ и автоматизацията могат да заменят хора в много професии, особено тези, които включват рутинни и повтарящи се задачи. Това може да доведе до масова безработица в определени сектори, особено за нискоквалифицирани работници. Примерите включват фабрики, транспорт, обслужване на клиенти и други.

Контрааргумент: Технологиите водят до трансформация, а не до масова безработица. С появата на ИИ се създават нови работни места свързани с неговата поддръжка и развитие. Работниците ще трябва да се адаптират към новите технологии и да получат необходимата квалификация.

2. Пристрастност и дискриминация. Алгоритмите за ИИ могат да отразяват съществуващи предразсъдъци и дискриминация в данните, на които са обучени. Това може да доведе до несправедливи решения в области като наемане на работа, отпускане на кредити или правосъдие. Ако данните, с които се обучава ИИ, съдържат предразсъдъци, системата може да ги възпроизведе и усилва.

Контрааргумент: Дискриминацията и предразсъдъците при избора и вземането на решения не се създават от ИИ, а се възпроизвеждат в него от разработчиците му. Същите може да се премахнат чрез възприемане на морално-етичните норми на националното общество и международната общност чрез етичните кодекси на професионалните колегии. Този процес трябва да се насърчава от държавите и/или от професионалните ИТ сдружения.

3. Нарушение на личната неприкосновеност. ИИ се използва за събиране, обработка и анализ на огромни количества лични данни. Това може да доведе до злоупотреби и нарушения на правата на неприкосновеност на личния живот, особено ако тези данни попаднат в неправилни ръце или се използват без съгласието на индивидите.

Контрааргумент: Вярваме, че ИИ може всъщност да подобри киберсигурността и защитата на личните данни, стига да бъде използван правилно:

- като средство за засилен контрол върху права за информационен достъп и за нива на валидиране на управленско решение (за повишена сигурност на идентификацията);
- като се регулира употребата на ИИ със задължителен за спазване от ИТ компаниите международен регламент, който да защитава по надежден начин личните данни от користо, неразрешено използване и от злоупотреба с тях.

4. Дезинформация и манипулация. Системите на ИИ могат да се използват за създаване на фалшива информация, манипулиране на медийни съдържания или разпространение на дезинформация (например deepfakes). Това може да доведе до обществено недоверие, социална поляризация и подкопаване на демократичните процеси.

Контрааргумент: ИИ може да бъде използван и за борба с дезинформацията, като разпознава фалшиви новини, открива манипулирани медийни съдържания и съдейства за проверка на фактите.

5. Автономни оръжия и военни приложения. ИИ се разработва и за военни цели, включително за създаване на автономни оръжия, които могат да вземат решения без човешка намеса. Това повдига сериозни етични въпроси относно контрола, отговорността и възможните непредвидени последствия от такива системи.

Контрааргумент: Напредъкът на технологиите в отбраната са неизбежни, но това не означава, че не може да бъде контролиран. Използването на ИИ може да бъде ограничено, а и може да се използва за предотвратяване на войни чрез по-добра комуникация и анализ на риска.

6. Ерозия на отговорността. С използването на ИИ в ключови процеси (като диагностика в медицината или решения в правосъдието), отговорността може да се размие. Ако алгоритъмът направи грешка, може да бъде трудно да се установи кой носи отговорност – разработчикът, потребителят или самата система.

Контрааргумент: Обръща се внимание към въпросите на отговорността и отчетността при използването на ИИ и се въвеждат етични кодекси, които изискват прозрачност в решенията на ИИ, като по този начин позволяват по-ясно проследяване и отговорност.

7. Концентрация на власт. Разработването и внедряването на ИИ често е доминирано от големи технологични компании. Това може да доведе до концентрация на власт и влияние в ръцете на малка група корпорации, което може да увеличи икономическото неравенство и да ограничи конкуренцията.

Контрааргумент: ИИ не е обект на монопол или на олигополна концентрация. ИИ може да бъде разработен и внедрен от нови технологични компании или създаден по проекти на държави, международни съюзи, междууниверситетски и изследователски звена или на публично-частни консорциуми. Наличието на конкурентни софтуери за ИИ от независими създатели ще минимизира този риск.

8. Екологични последици. Обучението на големи модели на ИИ изисква значителни изчислителни ресурси, което води до високо енергопотребление. Това може да има отрицателен ефект върху околната среда, като допринася за увеличаване на въглеродните емисии и изменението на климата.

Контрааргумент: ИИ се развива устремно, което ще позволи да се създадат по-качествени разработки с използване на квантови и други високоскоростни чипове, които ще повишат изчислителната му ефективност за единица време. По такъв начин тези иновации ще намалят енергопотреблението при употребата на ИИ, а с това ще съкратят рисковете за околната среда и за климатичното равновесие.

9. Проблеми с интелектуалната собственост. ИИ може да генерира нови произведения на изкуството, музика, литература и други творчески продукти. Това води до нови въпроси относно правата на интелектуална собственост – кой е собственик на създаденото от ИИ произведение и как се регулират тези права?

Контрааргумент: С практиката и със съдействието на експерти по интелектуална собственост ще се прокара разграничение между произведенията на изкуството, музиката, литературата и др. творчески продукти, създадени с ИИ от тези създадени от човека. Вероятно първите ще бъдат обект само на търговски, но не и на авторски права. Същевременно ИИ може да се използва, предвид големите му възможности, за да възпрепятства неавторските репродукции (от човек или от ИИ) на творчески продукти, без уредени авторски права със създателя им.

10. Липса на прозрачност и обяснимост. Много от съвременните модели на ИИ, като невронните мрежи, са "черни кутии", които вземат решения, които често са трудни за обяснение. Това може да бъде проблематично в критични области като здравеопазване или финанси, където е важно да се разбере логиката зад дадено решение.

Контрааргумент: Тук човекът може да се намеси, за да подобри ИИ. Учените могат да разработят алгоритми, които да предоставят по-ясни и разбираеми обяснения за своите решения, което улеснява тяхната употреба в критични области като медицина и финанси.

Заклучение

И ако американският икономист Питър Дракър предвижда, че преходът от индустриално към знаниево общество настъпва между 2010-2020 г. и основните икономически ресурси ще са знанието, „знаещите“ и „можещи“ хора (Дракър, 2000), то към края на 2024 г. все повече на дневен ред излиза въпросът дали е възможно четвъртата индустриална революция да е „човекоцентрична“, имайки предвид

песимистичните предсказания по отношение на заплахите от употребата на изкуствения интелект. Моликс казва, че досега „никога не сме създавали общоприложима технология, която може да повиши нашата интелигентност“. Моликс казва, че регулациите няма да са достатъчни и вместо това „пътят напред изисква широк обществен отговор“ с „норми и стандарти“ и „информирани граждани“, като всички работят заедно (Mollick, 2024).

„Изкуственият интелект е инструмент в нашите ръце. Ако бъде използван по правилния начин и за правилните цели, той може да бъде в полза на развитието и просперитета на човешкото общество. Той е ресурс, който е недоразвит напълно, но с времето може да стане част от нашето ежедневие и да подобри живота ни в различни области. Важно е да се запази човешкият контрол върху изкуствения интелект, особено за сложни технологични процеси и при вземане на стратегически решения от мениджърите. Назря моментът да се изгради във всяка държава Държавен орган или Министерство, който да контролира развитието на ИИ по строги правила и закони. Да следи стриктно развитието на ИИ, спазването на строгите правила и да не дава надмощие на ИИ върху човешката раса. Много важна необходимост е да разработим регулации и етични насоки, които да гарантират, че технологията ще служи на човечеството по безопасен и справедлив начин“ – заключават студентите, участвали в експеримента.

Моликс формулира „Четирите правила за съвместен интелект“: 1. Винаги канете изкуствения интелект на масата. 2. Бъдете човекът, който участва в цикъла. 3. Отнасяйте се с ИИ като с човек (но му кажете какъв човек е той). 4. Предполагайте, че това е най-лошият ИИ, който някога ще използвате. (Mollick, 2024). „Въпреки неговите уговорки за антропоморфизиране, машините не са нищо подобно на хората и не заслужават дори метафоричния статус, който той им дава“ (Ridley. 2024).

Законодателният акт за изкуствения интелект на ЕС е в сила от 1 август 2024 г., така че орган, който да следи за коректността на използването на ИИ съществува. Иницирирана е и Служба за изкуствен интелект към Европейската комисия, в задълженията на която влиза и търсенето на иновации, чрез които да се развива ИИ.

Към настоящия момент, тъй като изкуственият интелект и неговото приложение е топ адженда се провеждат множество форуми, посветени на тази тема. Правят се публикации, които изследват възможностите на изкуствения интелект да повиши уменията и да преквалифицира човешкия капитал (upskilling and reskiling).

Множество въпроси се повдигат, отговорите на които ще бъдат предмет на търсене на учени и специалисти и за в бъдеще – с или без подкрепата на изкуствен интелект – според личната отговорност и разумната мяра на всеки един от тях.

Литература/References

ChatGPT и AI от нулата. Как да станеш майстор на изкуствения интелект, без да губиш месеци в проба-грешка, дори да нямаш опит с технологиите. С., 2024, Commerce Masters. 176 с.

Дракър, Питър. Посткапиталистическото общество. С., Изд. Лик, 2000.

ИИ в изследванията: иновации и приложения. Достъпен на:

<https://julienflorkin.com/bg/%D1%82%D0%B5%D1%85%D0%BD%D0%BE%D0%B%D0%BE%D0%B3%D0%B8%D1%8F/%D0%B8%D0%B7%D0%BA%D1%83%D1%81%D1%82%D0%B2%D0%B5%D0%BD-%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%B5%D0%BA%D1%82/ai-in-research/> /Посетен на 19.09.2024 г./

Изкуствен интелект (AI): Разбиване на мистерията около тази технология. Достъпен на: <https://www.ali.bg/387/> /Посетен на 17.09.2024 г./

Койчева, Д. Какво е изкуствен интелект и как да го използваме? Достъпен на: <https://advanceacademy.bg/blog/kakvo-e-izkustven-intelekt-i-kak-da-go-izpolzvame>. /Посетен на 18.09.2024 г./

Кулина, А. Какво е изкуствен интелект. Достъпен на: <https://digitalmarketing.bg/what-is-artificial-intelligence/> /Посетен на 17.09.2024 г./

Сайт на проект **Information matters**. Достъпен на: <https://informationmatters.eu/bg/> /Посетен на 01.10.2024 г./

Mollick, E. (2024). Co-intelligence: Living and working with AI. Portfolio/Penguin.

Ridley, Michael. (2024). Book review: Ethan Mollick. Co-Intelligence: Living and Working with AI. Portfolio/Penguin, 2024. Canadian journal of information and library science - La Revue canadienne des sciences de l'information et de bibliothéconomie (CJILS-RCSIB) Vol. 47, No. 1 (2024). DOI: 10.5206/cjils-rcsib.v47i1.18653

The EU AI Act, 2024. Достъпен на: <https://artificialintelligenceact.eu/ai-act-explorer/> /Посетен на 08.10.2024 г./