DOI:10.37708/jsca.swu.v14i1.9

Хенри Кисинджър, Крейг Мънди, Ерик Шмит, 2025. *Генезис. Изкуственият интелект, надеждата и човешкият дух.* София: ИК Труд. 216 стр.

Последната книга на Кисинджър (издадена посмъртно), в която е съавтор с учените технолози Крейг Мънди и Ерик Шмит, е поглед към бъдещото чрез изследване на взаимоотношенията между изкуствения интелект (ИИ) и човешкото знание, власт и политика, както е изведено в отзивите на цитираните Бил Гейтс, Уолтър Айзъксън, Сам Олтман, Фарид Закариас и др.

В „Генезис“ са изследвани множество въпроси, свързани с появата на изкуствен интелект, както и с политическите, етичните и цивилизационните измерения на ИИ и е структурирано в три части: *Първа част: В началото; Втора част: Четирите клона (политика, сигурност, благосъстояние, наука); Трета част: Дървото на живота.*

В предговора на ГЕНЕЗИС от 2024 г. се подчертава, че Кисинджър през цялата си кариера е наясно, че стабилните на пръв поглед модели на великодържавната политика могат да бъдат периодично нарушавани от **технологични промени**. Кисинджър е държавникът, който в есето си от 1957 г. „Стратегия и организация“ извежда ясното заключение, че „*от тук нататък резултатът от пълна война ще бъде, че и двамата участници ще загубят*“, няма победител в подобен конфликт, защото дори по-слабата страна може да нанесе степен на разрушение, която никое общество не може да понесе (Предговор. ГЕНЕЗИС, 2024). Тоест, подчертава се, че Кисинджър разбира огромното значение на подобна технологична революция за политическите отношения между държавите.

Изразено е очакване за „еволюцията на Homo technicus – човешки вид, който в тази нова епоха може да съществува в симбиоза с машинните технологии“ (Предговор. ГЕНЕЗИС, 2024). Неслучайно, заявяват съавторите Ерик Шмит и Крейг Мънди, най-подходящото заглавие на книгата е „Генезис“ – ново начало за всички хора. Така аналогията с Библията и главата Сътворение (на латински: Genesis) е невъзможна за пренебрегване.

Въвеждайки в темата авторите недвусмислено заявяват, че макар ИИ да е тема от първа страница на обществения дебат, все още много важни аспекти на

тази нова ера са пренебрегвани (Въведение. ГЕНЕЗИС, 2025: 33). Сегашните възможности на ИИ вероятно ще изглеждат ограничени в ретроспекция, тъй като те се увеличават с неимоверни темпове, заедно с рисковете, които все още не са напълно известни. Например – машините със способност да определят свои собствени цели не са далеч.

По отношение на партньорствата човек-машина, авторите отбелязват важността на определянето на граници на властта на ИИ над истината – т.е. дали ИИ да стане посредник между хората и реалността. Поставя се и въпросът с ролята на биологично ограничения човешки ум – в полза на потенциално партньорство с ИИ. Но, може би нечовешките скорости на напредък на ИИ ще направят традиционното регулиране безполезно – „*ще ни е необходима фундаментално нова форма на контрол*“.

Днес преобладаващата част от ползите и почти изключителният контрол над напредналите ИИ системи принадлежат на изключително малък брой хора. Налага се въпросът ще се откажат ли те от своите предимства? И дали ако го направят, няма веднага да последват призиви за глобализация. Ще превърне ли някоя нация своя суверенен капитал в общо благо?

Авторите прогнозират, цитирайки Олтман, че двата актива, „които ще съставляват основната част от стойността в света“, а именно – компаниите, които създават и поддържат AI, както и земята (Кисинджър, Мънди, Шмит, 2025: 139). Но в същото време, бъдещите лидери, вземащи решения, трябва да внимават да не закрепят отново социалните и икономическите неравенства, появили се след индустриалната революция и които са били коригирани твърде бавно чрез по-хуманни структури за контрол (op. cit: 140).

Отделено е съществено внимание на въпроса за ИИ в икономически план. Особено внимание е отделено и на труда, особено на бъдещето на труда в контекста на ИИ. Доколкото съвременната икономика все още се формира от производството на стоки и услуги, а не от техните абстрактни представяния, парите не могат да съществуват без пазар. Те се превръщат в запис в база данни за разпределение на ресурси, при който стойността постепенно престава да има смисъл. Разгледани през призмата на информационната теория, парите днес имат много общо с интернет връзката, която за всяка разпознаваема употреба също изисква релационен контекст. Но дори ако успеем да изобретим валути, системи, пазари и политики, които отговарят разумно на появата на този мащабен потенциал за генериране на стойност, как на практика би могъл ИИ да реши или да премахне бедността например? Как би могъл да установи абсолютен глобален минимум за качеството на живот? (op. cit: 142)

В „Генезис“, авторите приемат, че разпределителни системи с изкуствен интелект могат да бъдат проектирани така, че чрез масово произведени роботи и инфраструктури, оптимизирани отново чрез ИИ, да осъществят връзка с онези, които в момента са изключени от глобалната мрежа. Тоест изкуственият интелект, според тях, би могъл да намали дефицита на таланти и да изравни разпределението на ресурси, което би направило по-малко значими онези фун-

даментални линии на разделение, върху които дълго време се е осъществявала международната търговия и по които световният просперитет е стратифициран.

Но докато се инвестира в тези идеи, следва да се държи сметка и за потенциалните ползи: дълбока трансформация в общия стандарт на колективния човешки живот и в равенството на индивидуалното съществуване – независимо от раса, пол, националност, място на раждане или семеен произход. Изравняването на „цената“ на интелигентността, разпределена по цялата планета, би могло да създаде равнопоставеност, каквато никога досега не е съществувала (op. cit: 144). Но тук възниква въпросът какво ще се случи, ако ИИ, действайки като икономически изравнител, доведе до рязко спадане на ценността на интелигентността, а от там – и на труда? Това би било края на краткия, но изключително продуктивен период от човешката история, в който индивидите в свободни общества подобряват жизнените си обстоятелства чрез собствени усилия.

Тоест прави се изводът, че *елиминирането на селектиращата функция на труда*, ще изправи човечеството и *пред елиминирането на професиите* – а с тях и на *социалния статус, идентичността и смисъла,* които традиционно са били неразривно свързани с тях. Един коренно различен свят (op. cit: 145). Предвид това, че личното преживяване на някаква трудност често поражда усещане за гордост след преодоляването ѝ, усилието когато е съчетано с отдаденост, придава собствен разказен ритъм на разбирането ни за времето, за самите нас и за човешката способност за овладяване, това би било наистина напълно различен за човека свят. Затова, подчертава се, изправени пред възприеманата заплаха от автоматизацията на човешкия труд, много анализатори се фокусират върху настъпването на нова духовна криза: В свят на споделено изобилие, твърдят те, ние бихме станали като безотговорни победители от лотарията, обзети от хедонизма на излишествата.

Все пак, предвид особеностите на човешката психика, авторите остават на мнение, че много източници на радост и удовлетворение вероятно ще останат непроменени в ерата на ИИ – въпреки множеството промени, които неминуемо ще преминат през нашия живот. Например изкуството може да преживее разцвет, защото истинността – звънът на автентичното – вероятно ще запази своята притегателна сила (op. cit: 147).

Особено внимание е отделено на науката, доколкото ИИ е на път да преформулира границите на възможното във всяко творческо начинание и да открие нови хоризонти във всяка научна област.

Може би съвсем скоро ще притежаваме способността да определяме темпото и посоката на собствената си еволюция. С тази винаги противоречива идея, се появява и един очевиден, но коварен въпрос: *Как всъщност изглежда съвършеният човек?* Този въпрос е задаван многократно в различни общества, като в някои случаи е служил за основа на „научни“ и политически проекти, довели до трагични последици за човечеството. Ето защо, смятат авторите, непременно трябва да се запитаме с известна предпазливост: *Трябва ли изобщо да се опитваме да разберем?* (op. cit: 156)

И тъй като прекалено често изкуственият интелект бива възприеман като поредния продукт на разрушителната философия, надеждата е, както е заявил Уинстън Чърчил през 1940 г., да станем свидетели как: „*Новият свят, с цялата си сила и мощ, пристъпва в помощ и за освобождението на Стария*“ (op. cit: 160).

Обмислени са въпроси, свързани с възможността ИИ да надхвърли пределите, които човек досега не е могъл да премине, защото винаги сме били ограничени както от възможностите на нашите инструменти, така и от границите на собственото си въображение. Системите с ИИ може да се окажат не просто пасивни наблюдатели или получатели на сигнали – а и преводачи, изследователи или разузнавачи. Те могат да поемат ролята на астронавти, способни да отиват по-далече, отколкото е било възможно за човека някога (op. cit: 161–163).

В последната част авторите насочват вниманието си към *възможната стратегия* спрямо изкуствения интелект, поставяйки въпроса: *Ще станем ли повече като тях, или те ще станат повече като нас*? (op. cit: 168)

Въвежда се концепцията за „коеволюция“. До днес развитието на компютърните технологии следва траектория на все по-голяма *интеграция и взаимодействие* между хората и машините. Изграждайки инструментите си така, че да ни пасват все по-добре – в съответствие с хилядолетната човешка практика, никога досега не сме обмисляли създаването на инструменти, които не съответстват на нашата анатомия или интелект. Напротив, *винаги сме се ръководили единствено от ограниченията на биологията ни. Но сега появата на изкуствения интелект може да накара някои от нас да се замислят за обратното:* в случай, когато инструментите ни сякаш надминават нашите способности – както ИИ вече прави (op. cit: 169). Така също, пишат авторите, биологичният инженеринг бива насочен към все по-тясно *сливане на човека с машините* и невроинженерството може да се окаже само междинен етап по пътя към истинска симбиоза (напомня Кърцуайл и the singularity *(Kurzweil, R. 2024. The Singularity is nearer. When we merge with AI. New York: Viking).*

През исторически разкази и митове се насочва вниманието на читателя към невъзможността за управление на сили, които са неприсъщи на хората. Както Харари използваше за илюстрация разказа за „Чиракът магьосник“, Кисинджър, Шмит и Мънди използват мита за цар Мидас, превръщащ всичко до което се докосне в злато и сирийската приказка за Аладин (op. cit: 173; *Ювал Ноа Харари. 2024. Nexus. Кратка история на информационните мрежи от каменната епоха до ерата на изкуствения интелект. София: изд. Изток-Запад*.).

Тоест изследвайки въздействието на изкуствения интелект върху различни области на човешката мисъл и дейност, както се цитират думите на Артър Кларк (Цит. в Предговор. ГЕНЕЗИС, 2024), че „*всяка достатъчно напреднала технология е неразличима от магията“,* точно както тук се прави препратка с цар Мидас, Аладин и вълшебната лампа и аналогията на Харари (Harari, 2024) с „Чиракът магьосник“ на Гьоте. Тоест и двете истории говорят за универсалната трудност при активирането и упражняването на сила, която ние смъртните,

не можем да разберем или контролираме. Тази сила сега се явява изкуственият интелект.

Разбира се, подчертава се това, че ранната социализация (т.е. въвеждането на ИИ за масова употреба) може да намали риска от проблематично поведение, като едновременно с това допринася за по-нататъшното „обучение“ на самия ИИ, а за хората – способства за повишаване на осведомеността, устойчивостта и здравословния скептицизъм. Милиони взаимодействия всеки ден спомагат за тестването дори на най-малко вероятното е сценарии, с които ИИ може да се сблъска. Така, макар и далеч от съвършенството, системите с ИИ, функциониращи в реалния свят, не само ще улесняват нашата адаптация към тях, но тяхната поява дава възможност за формулиране на по-усъвършенствани теории за това как те да се адаптират към нас (op. cit: 174).

До този момент подходите ни към постигането на така нареченото „**съгласуване между ИИ и човека**“ се групират в две основни категории: системи, базирани на правила, и „усилващо обучение“, базирано на обратна връзка от хора.

В този контекст важна е ролята на правните инструменти за контрол и тяхната реализация и изпълнимост в самите системи с ИИ. Според авторите в момента се полагат множество усилия за изграждане на интегрирана система за контрол, предварително внедрена в най-мощните ИИ, която би могла активно да насочва тези системи към законосъобразна, безопасна и дори ползотворна употреба (op. cit: 175).

Но по-устойчиви и последователни от всяко правило, наложено чрез санкция, са онези базисни, инстинктивни и универсални човешки разбирания. Френският социолог Пиер Бурдийо ги нарича **doxa** (от старогръцки: общоприети вярвания): преплетена мрежа от норми, стимули, както и механизми за възнаграждение и наказание, които в съвкупност обучават човека да различава добро от зло, правилно от неправилно. Докса е кодекса на човешката истина, но неописан чрез конкретни механизми или артефакти. Възприема се не чрез инструкции, а чрез участие в самия живот. *Как тази човешка докса може да бъде предадена на ИИ?* Според авторите за да изградим и осигурим функционирането на тази система от правила и ценности, най-вероятно *трябва да разчитаме на самия ИИ.* Хората не са успели досега да формулират и да се съгласят върху изчерпателен набор от собствени правила (op. cit: 178). Съвършенството на крайния механизъм за съгласуване между ИИ и човека ще бъде от критично значение. И най-важното – нежеланото поведение на машините, трябва не просто да бъде забранено, а напълно предотвратено (op. cit: 179).

При въпроса как да постигнем това, авторите се насочват към възможността частният сектор, с подкрепата на държавата и академичната общност, да създаде така наречените модели „заземени“ в човешката реалност (grounding models). Наред с разработване на тестове за валидиране и сертифициране на моделите като легални (в различни юрисдикции) и безопасни. Т.е. според тях е добре моделите с ИИ да бъдат подлагани на преглед преди публичното им пус-

кане наред с възможността за носене на отговорност, както и необходимостта от налагане на санкции в случаи, когато даден модел умишлено е обучен да заобикаля правни ограничения.

Етичното съгласуване е изведено като особено съществен проблем с ИИ. В тази заключителна част вниманието е насочено към изначални философски въпроси, които не могат да бъдат заобиколени тук. Това е така, защото без общо разбиране за това кои сме, човешкият род рискува да предаде на изкуствения интелект основополагащата задача да дефинира нашата стойност (op. cit: 182).

Затова тук ясно и недвусмислено се твърди, че ако се окаже, че е невъзможно да бъде създаден режим на надежден технически и стратегически контрол, *би било за предпочитане съветът да остане без изкуствен общ интелект, отколкото да съществува свят, в който дори един изкуствен интелект остава без човешките ценности.*

Но кои именно са тези базови човешки ценности? Има ли възможно определение за човечността? (op. cit: 183). При опита за формулиране на етически минимум се обмисля значението на възловото понятие *достойнство и значимостта* на човешкия живот. Изразени са опасенията, че математическата строгост трудно може да обхване някои човешки понятия, например понятието за милост. Поставя се и въпроса може ли ИИ сам по себе си да притежава достойнство? Вероятно не, смятат авторите, доколкото ИИ не се ражда, не умира, не изпитва несигурност или страх и няма естествени наклонности или индивидуалност. Затова, заключава се, той *трябва да бъде третиран философски, като литературен персонаж например.* Той може да въплъщава елементи на човешкото, но той не е реален в морален смисъл.

В заключение можем да се върнем към цитираната в предговора на книгата препратка към есето „Как завършва просвещението“ на Кисинджър (2018, в The Atlantic), че

> „обективната способност на ИИ да достига до нови и точни заключения за света чрез нечовешки методи не само разрушава доверието ни в научния метод, развиван последователно в продължение на пет столетия, но и поставя под въпрос човешката претенция за изключителен или уникален достъп до реалността. Какво означава това? Дали епохата на ИИ не просто няма да изведе човечеството напред, но и ще катализира завръщане към предмодерно приемане на необяснима власт? С други думи: изправени ли сме, или бихме могли да се окажем пред прага на велик обрат в човешкото познание – тъмно Просвещение?“ (Цит. Предговор. ГЕНЕЗИС, 2024).

Категоричен отговор авторите не дават. Но самото ясно извеждане на значимостта на темата за ИИ и необходимостта от задълбочен анализ на възможните посоки и стратегии за бъдещето несъмнено допринася за по-доброто ни

разбиране и ориентация в епохата на изкуствения интелект. Тоест книгата без да предлага категорични решения, ясно формулира мащаба на предизвикателството: изкуствения интелект може едновременно да разшири човешките възможности и да постави под въпрос самите основания на модерното знание и човешката уникалност. Именно това прави „Генезис“ значим принос към разбирането на политическите, етичните и цивилизационните измерения на изкуствения интелект.

Гл. ас. д-р Даниела Симеонова-Коруджиева
Българска академия на науките
Институт за етнология и фолклористика с Етнографски музей
Секция Етнология на социализма и постсоциализма
Ул. „Московска 6А, 1000 София, България
E-mail: daniela.koroudjieva@iefem.bas.bg