НАУЧНИ ТРУДОВЕ НА УНСС (2) 2025, АКАДЕМИЧНО ИЗДАТЕЛСТВО – УНСС
ISSN (online): 2534-8957; ISSN (print): 0861-9344; http://unwe-research-papers.org/
DOI: https://doi.org/10.37075/RP.2025.2.07

ПРИЗРАЧЕН ИКОНОМИЧЕСКИ АГЕНТ ЛИ Е ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ?

Иванина Манчева[1]
e-mail: mancheva-fte@uacg.bg

Резюме

Навлизането на изкуствения интелект (ИИ) в икономическия и обществения живот довежда до значими преобразувания. Те поставят на дневен ред въпроса за ролята на ИИ в икономическите модели и по-специално дали той следва да се третира като икономически агент, който взема самостоятелни решения. Този въпрос, както и свързаният с него проблем за правния статут на ИИ, не е намерил своя отговор. Те засягат много от дискутираните през последните години теми: възможността за облагане на роботите, отговорността на автономните автомобили или авторските права на произведения, създадени от ИИ. Изследването се застъпва за диференциран подход при определянето на ИИ като самостоятелен икономически агент. При положителен отговор ключовият въпрос би бил каква е целевата функция, която ИИ максимизира и какви са ограниченията и рисковете, пред които той се изправя. Въпреки че ИИ тепърва ще развива способностите си и нямаме достатъчно наблюдения върху поведението на най-напредналите системи, приемаме, че колкото повече нараства независимостта му от човека, толкова повече той ще се превръща в икономически агент.

Ключови думи: изкуствен интелект, икономически агент, вземане на решения, функция на възнаграждение, морален риск, информационна асиметрия, принципал-агент

JEL: D01, K10, O32, O33

Увод

Навлизането на изкуствения интелект (широко понятие, което обхваща невронни мрежи, машинно обучение, роботика и т.н.) в икономическия и обществения живот довежда до значими преобразувания. Вече са налице множество научни изследвания на ефектите на изкуствения интелект (ИИ), свързани с повишаването на производителността. В немалка степен е проучено и неговото влияние върху трудовия пазар, както и върху нивото на конкуренция на различните продуктови пазари. Сравнително по-слабо е изяснена ролята

[1] Главен асистент, доктор, катедра „Обществени науки“, УАСГ, ORCID: 0000-0001-9208-0586

на ИИ в процеса на вземане на решения и влиянието му върху поведението на икономическите агенти. Този въпрос е важен, тъй като по-напредналите модели на ИИ не само трансформират начина на вземане на решения, те самите вземат решения – при ограничена или никаква човешка намеса.

Историята на ИИ от последните няколко десетилетия редува периоди на свръхоптимизъм и разочарования (т. нар. „зими“ на изкуствения интелект). През последните години – до голяма степен благодарение на големите езикови модели, невронните мрежи и генеративния изкуствен интелект, наблюдаваме вълна на оптимизъм. Дори и най-смелите прогнози да не се сбъднат в близко бъдеще, вече достигнатите способности на ИИ изискват по-добра концептуализация на неговата роля.

В този текст се изследва действителната и хипотетична роля на ИИ в процеса на вземане на решения за създаването, разпределението и потреблението на икономически блага. Негов фокус е въпросът каква би била полезността (целевата функция, функцията на възнаграждение), която ИИ максимизира и на тази база дали можем да приемем, че той (ще) се държи като самостоятелен икономически агент със собствени предпочитания.

С оглед на това настоящият текст си поставя за цел да обсъди ролята на ИИ в икономическите модели и по-специално при какви условия той следва да се третира като самостоятелен икономически агент. Изясняването на този въпрос ще допълни широко дискутираните теми за ролята на ИИ в икономическата литература и ще позволи те да се разгледат в нова светлина: вече са налице множество публикации относно възможността за облагане на роботите, отговорността на автономните автомобили или авторските права на произведения, създадени от ИИ. Тези проблеми са свързани с въпроса дали ИИ следва да се признае като отделен правен субект, който не е намерил своя категоричен отговор. Определянето на самостоятелната роля на ИИ в процеса на създаването и разпределянето на икономическите блага е въпрос от подобен порядък. Неговото поставяне на дневен ред може да изглежда преждевременно, но не е преувеличено.

Във връзка с това изследването си поставя следните задачи:

– Да обобщи критериите, аргументите и възраженията за признаването на ИИ като самостоятелен субект, включително самостоятелен правен субект;
– Да определи начина на вземане на решения от ИИ с оглед на неговите предпочитания и ограничения и в зависимост от неговите способности. Да очертае връзката между функция на възнаграждение и предпочитания;
– Да разшири приложението на модела принципал-агент от гледна точка на системата човек-ИИ.

Така формулираният проблем надхвърля полето на икономическата наука, тъй като засяга въпроси от сферата на моралната философия, философията на информацията, правото и политиката – например, проблемите за свободната воля и идентичността, проблемите за отговорността, авторските права, дезинформацията и много други. В настоящия текст тези проблеми са застъпени доколкото имат отношение към формулираните по-горе задачи и моделирането на поведението на ИИ.

Изясняването на посочените въпроси не само ще направи по-актуални учебниците по икономика, но и ще подсили основанията на регулациите, засягащи изкуствения интелект и алгоритмичната търговия.

Еволюция на ИИ и ключови способности

През последните години се наблюдава бързо увеличаване на възможностите на ИИ под влияние на увеличената изчислителна мощ, нарастването на обема данни, върху които се обучава той, както и благодарение на усъвършенстването на алгоритмите. Повратна точка за все по-широкото използване на ИИ беше напредъкът при големите езикови модели (LLM). Това е вид изкуствен интелект, разчитащ на машинно обучение и дълбоко обучение, който се използва за обработка и генериране на естествен език. Особено това се отнася до чатботовете от типа на ChatGPT, Claude и Perplexity, които са тип генеративен LLM, създаващ ново съдържание (за разлика от по-непопулярните дискриминативни LLM, които се ползват за категоризация и анализ на текстове).

Изкуственият интелект е в състояние да изпълнява много и все по-сложни задачи – от разпознаване на изображения до финансови прогнози, търговия на организираните пазари и управление на автономни превозни средства. Това затруднява определянето на системите с ИИ[2], като превръща дори дефиницията им в дискусионен въпрос. С оглед на предмета на настоящето изследване е възприето следното определение на ИИ: технологии и методи, които позволяват на машините да имитират или заместват човешкия интелект при решаване на сложни проблеми и вземане на решения.

С напредването на ИИ е все по-важно да се прави разграничение между базисни и разширени възможности на ИИ, особено що се отнася до вземането на решения и сложните, подобни на човешките, преценки. Съществуват множество класификации на ИИ, като най-често той се разделя на:

- тесен ИИ – предназначен за решаване на специфични задачи, за които е специално програмиран, и

[2] В изследването не се прави разлика между ИИ и системи с ИИ.

- общ ИИ (artificial general intelligence, AGI), който би имал приложение за решаване на широк набор от сложни задачи в различен контекст, като е способен да разбира и да се научи да решава нови проблеми, за които не е знаел по времето на създаването си (Goertzel, and Pennachin, 2007).

Общият ИИ все още е хипотетичен, макар да съществуват твърдения, че способностите на някои системи са достигнали това ниво. AGI се характеризира с когнитивни способности (разсъждение, планиране, предвиждане, памет и вземане на решения), сравними или превъзхождащи човешките, с адаптивност и самонадзор. Именно тези негови характеристики, които ще му позволят да бъде напълно независим от човека, обуславят до голяма степен дискусионния характер на предмета на настоящето изследване. Следва да се отбележи, че посочената по-горе дефиниция на AGI го различава от моделите на ИИ с общо предназначение (general-purpose AI model), които Европейският акт за изкуствения интелект, определя като модели, обучени с голямо количество данни, които се самоконтролират и са в състояние да изпълняват широк набор задачи). Те обаче не са напълно автономни и не са достигнали или надминали човешкия интелект, каквито биха били системите с AGI.

На теория съществува още едно ниво на хипотетичен ИИ – т. нар. съзнателен ИИ (Sentient AI), който усеща собственото си съществуване и мисли за себе си. Фокусът е върху субективните аспекти на преживяванията (qualia) и дали алгоритмите могат да възпроизвеждат вътрешно усещане (Alonso, 2023). Хипотезата за съзнателен ИИ се посреща със значителен скептицизъм (Buttazzo, 2001), като отсъстват доказателства, че цифровите системи са способни на субективни преживявания и чувства. Въпреки това тя поражда етични проблеми – например, какви следва да са правата на ИИ, ако той е способен да страда и съпреживява. Описаният проблем е свързан с въпроса за идентичността на ИИ, но не го изчерпва. Липсата на доказателства за субективни преживявания не изключва възможността ИИ да реагира на етични ограничения, зададени при неговото проектиране и да симулира ценностни преценки. За да се избегнат спекулативни предположения, в настоящето изследване се приема, че способността на ИИ да има самосъзнание е определяща за етиката на ИИ, но не и за формирането на предпочитанията му в съответствие със зададената функция и опита (обратната връзка).

Russel and Norvig (2009) от своя страна систематизират подходите към дефиниране и изучаване на ИИ в четири категории според това дали фокусът е върху мисловния процес или върху поведението, и от друга страна, според това дали подходът набляга на човешката или на рационалната стра-

на на ИИ. Така например известният тест на Тюринг (Turing, 1950) е пример за подход за дефиниране на интелигентността, основан на човешкото действие.

От гледна точка на предмета на настоящето изследване ключов критерий е неговата способност за самостоятелно вземане на решения. От тази гледна точка разграничаваме следните способности на системите с ИИ:

Базисни изчислителни способности:

– обработка, търсене и извличане на данни и информация;
– разпознаване на модели – разпознаване на изображения, анализ на текст и други видове данни с цел откриване на закономерности.

Разширени способности:

– решаване на проблеми – включва оптимизация и вземане на решения на базата на правила, както и способност за разсъждения, т.е. вземане на решения на база логика и правила;
– прогнози и модели – използва исторически данни за прогнозиране на бъдещи резултати и моделиране на вероятности.

Разбиране на контекст и адаптация:

– обработка на естествен език – разбиране и генериране на естествен език;
– осъзнаване на контекст – разпознават модели в даден контекст и адаптират своите реакции.

Самосъзнание (хипотетично)

– осъзнаване на собствената идентичност;
– формулиране на собствени предпочитания на база на осъзнаване, а не на предвиждане.

С оглед на фокуса на настоящия текст е важно да се отбележи разликата между **решаване на проблеми и вземане на решения**. Решаването на проблеми е структуриран процес, който използва алгоритми или методи с цел достигане до ясна предварително определена цел (например най-кратък път от точка А до точка Б). Вземането на решения включва не само намиране на решение, но и оценка на риска и прогнозиране на последиците от избрания вариант, като се отчитат множество критерии. Тези системи разчитат на усъвършенствана обработка на естествен език и разбиране на контекста (например чатботове). Те са способни да тълкуват двусмислици и да вземат решения в условия на неопределеност.

Ключовата способност на ИИ е машинното самообучение (Russig and Nordig, 2009). То се различава според степента на самостоятелност на ИИ и може да бъде контролирано (supervised) или неконтролирано обучение (известно още като обучение без учител). Във втория случай ИИ не действа според предварително зададени инструкции, а на база на установените от него модели от данни и зависимости, като има голяма свобода по отношение на получените резултати. При вземането на решения в динамична среда ИИ може да използва адаптивни методи, като модели на подсилващо обучение или обучение с утвърждение (reinforcement learning), за да се учи от грешките си и да се подобрява въз основа на обратната връзка от предишни решения. **Следователно, вземането на решения може да включва по-усъвършенствани системи с ИИ, тъй като изисква не само обработка на информация и алгоритмично решаване на проблеми, но и разбиране на контекста, учене от опита и предвиждане на последствията в условия на неопределеност**. Тези характеристики приближават ИИ за вземане на решения до истинска интелигентност.

Друга важна особеност са различните **ограничения и мотиви**, пред които са изправени хората и ИИ в своите решения. Действията на ИИ се ръководят от обучението му, данните, които са му предоставени, както и от обратната връзка, която се формира при взаимодействие с хората и с физическата среда (специално при обучение с утвърждение). Но е дискусионно дали неговите интереси биха зависели от от лични „предпочитания“, вярвания или интереси.

Не на последно място е важно да се очертае разликата между човекоподобен ИИ и Общия изкуствен интелект (AGI), за който се предполага, че ще достигне и дори надмине човешките способности. За разлика от AGI човекоподобният ИИ имитира човешкото мислене, емоции и поведение, без това да означава, че има интелектуалните способности на човека. Пример за това е ИИ, който успява да премине теста на Тюринг. Близък до него е ИИ, основан на теорията на ума (Theory of Mind AI). Това е концепция за тип ИИ, който има способност за разбиране на мислите, намеренията и емоциите на другите. Theory of Mind AI би могъл да се адаптира към социални взаимодействия и да предвижда поведението на другите.

Разграничавайки истинска и симулирана интелигентност, Russell and Norvig (2009, p. 1020) пояснявят философските основания на ИИ по следния начин: „твърдението, че машините могат да действат сякаш са интелигентни, философите наричат слаба ИИ хипотеза, а твърдението, че машините, които го правят, наистина мислят (а не само симулират мислене), се нарича силна ИИ хипотеза“.

Въпреки че изкуственият интелект може до известна степен да имитира начина на вземане на решения от човек, вкл. морални аргументи, неговите ограничения произтичат от отсъствието на съзнание, личен опит и лични ценности. Така начинът, по който той взема решения, остава фундаментално различен от разсъжденията, основани на морални преценки.

Условия за определянето на ИИ като икономически агент

Повечето съществуващи изследвания определят ИИ като технология (Trajtenberg, 2019), мултипликатор на производителността или производствен фактор (Wagner, 2020). Тази многоизмерност отразява както разнообразните видове и способности на системите с ИИ, така и променящия се характер на процеса на създаване и разпределение на икономически блага, опосредстван от ИИ. В различните модели ИИ може да играе различна роля: например той може да се разглежда като фактор, който съчетава и/или усилва останалите производствени фактори (труда и капитала). Подобно на капитала, ИИ се използва за създаване на стойност и може да бъде многократно прилаган в различни производствени и бизнес процеси. Но за разлика от капитала в класическия смисъл, ИИ е по-автономен, което означава, че може да се учи и адаптира, без или с ограничена намеса на човек. От друга страна, ИИ може да замества или допълва човешкия труд (Acemoglu et al., 2017). При това чрез алгоритми за самообучение ИИ може да се подобрява с течение на времето – така както образованието повишава способностите на човешкия капитал.

Но ИИ не само увеличава производителността на всеки от традиционните фактори, а и допринася за по-доброто разпределение на ресурсите, подобряване на прогнозите и намаляване на разходите. Всички тези приложения са свързани с трансформиращата роля на ИИ в процеса на създаване на блага.

От друга страна, фактът, че ИИ е способен да взема решения независимо от лицата, които са го създали или под чийто контрол се намира, поставя въпроса за неговото качество като **икономически агент**. Този въпрос надхвърля проблема за трансформиращата роля на ИИ в процеса на създаване на блага, тъй като засяга решенията за това какви блага да се произвеждат, както и въпросите на потреблението, спестяванията и инвестициите.

Агент е този, който действа. Икономически агенти са субектите, които правят избор между различни алтернативи в условията на ограниченост с цел постигането на предпочитан резултат. Това са лицата, които са ангажирани със създаването, разпределението и потреблението на икономическите блага, с решенията за спестявания и инвестиции. Микроикономическите агенти са индивидите и фирмите, чиито предпочитани резултати са съответно

максимизацията на благосъстоянието (очакваната полезност) и максимизацията на печалбата. Макроикономическите агенти (държавата и централните банки) са синтетични субекти, които преследват колективни цели, което ги доближава до част от специфичните характеристики на ИИ.

Необходими са следните критерии, покрити кумулативно, за да определим някого или нещо като икономически агент:

– лице (правен субект),
– което взема самостоятелно решения
– за създаване, разпределение, потребление на блага (какво, колко, как, къде, кога)
– с цел да постигне резултат,
– който (се) предпочита.

На база на анализа на неговите способности, можем да заключим, че ИИ покрива тези критерии, които се отнасят до извършването на избор (т.е. без първия и последния). В съответствие с възприетата по-горе дефиниция, отличителната черта на ИИ е способността да взема решения без човешка намеса или с ограничена човешка намеса. В зависимост от неговата функция, това могат да бъдат решения за **създаване на икономически блага или за разпределение и потребление на икономически ресурси**.

По-конкретно **решенията**, които ИИ взема, зависят от неговата функция: той може да подкрепя и подобрява процесите, свързани със създаването, разпределението и потреблението на ресурси и стоки, чрез интеграция на технологии за оптимизация, прогнозиране и анализ на данни. Това могат да бъдат решения за оптимизация на производствените процеси или оптимизация на потреблението на енергия и други ресурси, продуктови иновации, решения за управление на веригата от доставки и др. Следователно **резултатът** зависи от контекста и функцията и може да бъде например: намаляване на разходите за производство или логистика, минимизиране на времето за доставка или максимизиране на удовлетворението на клиентите.

Но начините, по които хората и машините вземат решения, не съвпадат (Council of Europe, 2017). ИИ има различни ограничения и прави различни грешки в сравнение с хората. Един от най-често дискутираните проблеми са систематичните грешки на ИИ или т. нар. пристрастия, които може да се дължат на проблеми в данните, дизайна или уклони на самите потребители (Ferrara, 2024).

Дискусионните критерии за определяне на ИИ като икономически агент са първият и последният: 1) самостоятелен (правен) **субект**, който 2) има свои **предпочитания**.

Тези два критерия са взаимносвързани, тъй като се отнасят до идентичността на ИИ и наличието на собствена воля. *Идентичността* се формира

въз основа на спомените, преживяванията и мотивите на отделния индивид, но също така и от усещането за групова принадлежност. В социален контекст взаимодействията, културата и езикът играят водеща роля. ИИ системите – поне на този етап, нямат лични чувства, преживявания или самосъзнание в човешкия смисъл. Следователно не може да се заключи, че те имат памет за своето „съществуване“, стабилни предпочитания и вътрешна мотивация, а оттук – самоличност. В тази връзка важна отправна точка е формулираният от Chalmers (1995) „труден проблем“ на съзнанието, отнасящ се до субективното възприемане на опита (qualia). Но с оглед на еволюцията на ИИ системите, въпросът дали те, взаимодействайки помежду си и трупайки опит, могат да развият представа за себе си, все пак остава дискусионен.

Можем да приемем, че идентичността на системите с ИИ (т.е. това, което ги отличава в сравнение с други системи) зависи от данните, обучението, обратната връзка и първоначалното проектиране. Ако отхвърлим напълно идеята за идентичност, то трябва да поддържаме, че поставени в една и съща ситуация, две различни системи с ИИ с различен опит биха взели еднакви решения, което е неправдоподобно (това не означава непременно, че две системи с ИИ биха взели различни решения, защото имат различни ценности преценки).

Дори ако общият изкуствен интелект, който ще е по-автономен, успее да постигне някаква степен на идентичност, това не дава категоричен отговор на въпроса дали той има собствена *воля*. Използваният тук израз собствена воля е опит да не се навлиза в полето на философията и обширната дискусия за *свободната* воля; със същата цел по този въпрос изследването се ограничава до следните принципни бележки: Определянето на ИИ като икономически агент е свързано с моделирането на неговото поведение – как формулира целите си и какви ограничения среща при вземането на решения. В този контекст проблемът за волята се свежда до автоматизма в действията и преценките на ИИ. В съответствие със Searle (1984) можем да обвържем волята с непредвидимостта на изборите на ИИ. Идеята за собствена воля следователно отхвърля детерминистичния възглед, че всяко събитие е предопределено (но не и предвидимо, тъй като информацията също играе роля). С други думи, въпросът за собствената воля е относим към ситуации, когато решенията на ИИ не се редуцират само до обработка на информация на база логика, правила *и* откриване на модели.

Изхождайки от горното, въпросът за **статута на ИИ** е свързан както с теоретични, така и с практически затруднения. Този въпрос може да се раздели така: при какви условия следва да определим ИИ като самостоятелен субект и, ако първото е вярно, то какви допълнителни условия са необходими, за да го определим като самостоятелен *правен* субект.

Проблемът за статута на ИИ може да се изведе от неговия капацитет (способности) и степента на неговата автономност. Безспорно е, че ИИ, който не е способен да взема решения без човешка намеса (контрол), може да се разглежда като средство за автоматизация (производствен фактор), а не като самостоятелен субект. Дискусията се отнася до това дали общият изкуствен интелект (AGI), при който не само не се предполага надзор от човека, а дори е възможно неговият механизъм на действие да е неразбираем за човека (създател, ползвател или трето лице), може да се разглежда като отделен субект.

С нарастването на самостоятелността и способностите на ИИ, все почесто се поставя въпросът дали той може да бъде отделен носител на права и задължения, който въпрос е ключов при определянето на неговия правен статут. Той включва следните подвъпроси: до каква степен ИИ има самостоятелност (собствена воля); каква е отговорността му за причинени вреди (как се установява виновно поведение); може ли да бъде получател на ползи и облаги; как неговите права се съизмерват с човешките права.

Основните възражения срещу правосубектността на ИИ включват липсата на истинско съзнание, факта, че неговата воля зависи от програмиране (първоначалния дизайн) и обучителни данни, както и риска от размиване на човешките права. Във връзка с това следва да се подчертае, че трима от икономическите агенти (фирмата, държавата и централните банки) са синтетични субекти. Те са създадени по волята на човека и са израз на общите интереси на група хора, изразени чрез учредителен акт или чрез конституция (като форма на обществен договор). Самосъзнанието и личните преживявания не са определящи за техния статут.

С оглед на задачите на изследването могат да се въведат следните критерии: първо, какъв е начинът на вземане на решения – самостоятелно или не, второ, кое е лицето, което получава ползите от дадено решение или действие, трето, до каква степен ИИ и останалите участници разполагат с информация за механизма на вземане на решение, последиците и рисковете. Тези критерии могат да се отнесат както към проблемите за отговорността, така и към правата на собственост, авторските права и облагането.

Възможни са няколко подхода. Първият предполага обвързване на права и задължения, стига да е възможно да определим ИИ като субект със собствена идентичност (самостоятелност). Това означава, че този, който черпи облаги, следва да носи и съответната отговорност (Gervais, 2019). Този подход обаче има неудобството да не отчита външните ефекти.

Вторият подход отразява влиянието върху стимулите и разпределението на риска. Това възпроизвежда класическата дилема за това дали правата върху изобретенията (временен монопол) компенсират поемането на риск

от иновациите и дали водят до оптимално разпределение на ресурси (вж. Arrow, 1962). Този подход има неудобства, свързани с непрозрачността на по-сложните системи с ИИ.

И при двата подхода следва да се отчитат взаимодействията и информационните ограничения на ИИ системите и естествените икономически агенти.

Последното е свързано с т. нар. проблеми на черната кутия и обяснимостта на решенията, които ИИ взема (Brożek et al., 2024). Най-общо, сложните невронни мрежи и наличието на нелинейни връзки затрудняват разбирането на вътрешната логика на ИИ системата. При това може да възникне т. нар. емергентно поведение, при което системата се отклонява от първоначалните параметри, самоорганизира се и дори резултатът може да е непредвидим. Емергентното поведение поражда проблеми при определянето на отговорността.

Оттук може да се изведе проблема за избора между солидарна или на диференцирана отговорност между ИИ, неговите създатели и неговите ползватели. Във втория случай създателите биха били отговорни за първоначалните проекти и всички въпроси, свързани с неговия капацитет, а ползвателите – по отношение конфигурационните протоколи на употреба.

В много отношения практиката изпреварва теорията – например един от основните дискутирани въпроси е за валидността на договорите, сключвани от ИИ. Този въпрос е намерил своето решение в практиката на борсовата търговия, която все повече разчита на алгоритми, някои от които с нулева човешка намеса, без това да отменя принципа на неотменимост на борсовите сделки, дори при алгоритмични грешки или сблъсък на алгоритми.

От друга страна, следва да се отбележат редица практически затруднения, които евентуално третиране на ИИ като правен субект би създало. Голяма част от тях са свързани с правоприлагането – като пример може да се даде определянето на режима на юрисдикция в съответствие с Регламент 1215/2012 (Брюксел I).

В зависимост от структурата на стимулите и наложените му ограничения ИИ може да бъде носител на ограничени права и задължения. Според някои автори (например Doomen, 2023) фактът, че ИИ може самостоятелно да създава продукти дава основания за признаването на известни права (например авторски права). Такъв пример е особеното право sui generis.

Тази публикация не може да изчерпи многообразието от аргументи, свързани с признаването на ИИ като правен субект и прилагането на такъв статут на практика. Но е необходимо да се подчертае разликата между субект, който взема решения и правен субект. Въпросът дали е възможно в стопанския оборот да участва лице, което не е правен субект (не е конституирано

като юридическо лице), не е нов. В българската правна система съществуват т. нар. неперсонифицирани дружества по Закона за задълженията и договорите (например консорциуми), които възникват по силата на договор за съвместна дейност. Тези дружества нямат самостоятелна правосубектност, нямат органи за управление и имущество, и правата и задълженията им са права и задължения на техните съдружници. Но в отделни случаи им се признава особена правосубектност (например те съставят счетоводен отчет и участват на самостоятелно основание в процеса на установяване и събиране на публични вземания). Може да се очаква правото да се развие в подобна посока, като се признаят особени права и задължения, например свързани с отчетност, на ИИ.

Последният критерий се отнася до **предпочитанията** на ИИ, което също е свързано с дискусионния въпрос за идентичността и свободната воля. С оглед на задачите на изследването бихме ограничили този критерии до по-инструменталния подход, характерен за икономикса. На таза база и в съответствие с Agrawal et al. (2019) разграничаваме предвиждане и преценка. Именно преценката е свързана с избора и се отразява във функцията на възнаграждението (reward function). Това е динамична функция, която разчита на последователна оптимизация чрез проби и грешки и измерва успеха на ИИ в достигането на зададената цел.

Първоначалната функция на ИИ е зададена от неговия създател (създатели), който определя най-общо неговата цел. На този етап, докато все още познаваме само *тесния ИИ,* целта и функционалните характеристики на ИИ са предварително определени. ИИ може да решава само проблемите, за които е създаден и обучен със съответните данни. Той е програмиран да анализира наличната информация, да прилага алгоритми и да адаптира действията си, за да постигне оптимален резултат спрямо поставените цели. Той не променя функцията си и не преформулира възнаграждението. Конкретните проблеми, които ИИ решава, от друга страна, се дефинират от неговите ползватели – чрез персонализирани настройки, заявките (promts), които те изпращат, или обратната връзка.

Но общият изкуствен интелект (AGI) би бил способен на по-висока степен на самостоятелност и адаптация. Той не само ще избира начина за постигане на целите, а и ще формулира свои цели. Тази самостоятелност не се предполага да е абсолютна, тъй като първоначалните му настройки са зададени от хората и функционирането му зависи от ресурсите, които получава (данни, изчислителна мощ) от икономиката.

На следващо място, ИИ не прави ценностни преценки и все още не взема решения с оглед на морални ограничения. Това го приближава до рационалния homo economicus. Но целевата функция, която ИИ максимизира, не

е основана на негови предпочитания, свързани с осъзнат личен интерес. На този етап не се предполага ИИ да има свои потребности, които възникват независимо от човека или организацията, които са го обучили или които го използват. Не е изключено този въпрос да претърпи развитие. Например ако ИИ придобие способност за самостоятелно мислене, то е вероятно той да започне да осъзнава, че се нуждае от данни и може да промени поведението си, за да си осигурява подходящи данни.

Ключовият въпрос тук е как се определя целевата функция или това, което ИИ максимизира. Вземането на решения е въпрос на преценка, която изисква да се съпоставят очакваната полезност и действията, доколкото те могат да бъдат определени. В тази връзка Agraval et al. (2019) посочват, че този процес на определяне на желания резултат (целевата функция) изисква човешко разбиране. В техния модел вземането на решения се състои от прогнозиране и преценка, като само човешката преценка може да оцени какво се смята за успешен резултат (каква е функцията на възнагражденията). Според тях докато AI прави прогнозите по-точни и понижава цената им, възвращаемостта (търсенето) на преценката нараства. Преценката (за разлика от прогнозирането) не се основава само на анализ на данни, а и изисква разбиране на последиците от едно или друго действие.

Въпросът за преценката е социално обусловен не само в човешкото му измерение. Взаимодействието на множество системи с ИИ води до незабавно споделяне на опита, но от друга страна създава проблеми на координацията и непредвидени последици. Goertzel and Pennachin (2007) описват такива взаимодействия с термина Swarm Intelligence. Това са биологично вдъхновени системи, които са самоорганизирани, разчитат на пряка и непряка комуникация между агенти и водят до емергентно поведение. А системите, които развият емергентно поведение, могат да се отклонят от първоначалната целева функция, да развият собствени стратегии и предпочитания, да заобиколят ограниченията при първоначалното проектиране.

Разбирането на последиците е свързано с рисковете, които би понесъл ИИ като агент, и съответните рискове за човека. Тези рискове са разпределени асиметрично. Доколкото ИИ няма собствено тяло, той не носи преки последици от грешни решения, засягащи реалния физически свят – например при пътни инциденти. Доколкото ИИ няма собствено имущество, той не носи риска от неговото погиване. Това твърдение би могло да се промени, ако ИИ получи права на собственост върху данни или върху създадените от него произведения. И най-накрая, тъй като ИИ няма собствени чувства, той не понася емоционални рискове и няма интерес да защитава репутацията си. Това би могло да се промени при реализация на хипотезата за съзнателен ИИ.

Приложимост на концепцията принципал-агент

Всеки проблем предполага ценностна оценка и предположения за ефекта, вкл. за ефекта върху вземащия решението. Дори когато самостоятелно взема решения (избира между алтернативи), ИИ все още не може сам да дефинира крайната цел. Това насочва вниманието към модела принципал-агент, който изглежда по-приложим с оглед на предизвикателствата, които моделирането на съвременния ИИ като икономически агент поставя. Този проблем е сравнително по-добре изследван (Athey et al., 2020) и въпреки наличието на открити въпроси не е толкова дискусионен, колкото възможността да признае ИИ като самостоятелен икономически агент.

Концепцията принципал-агент се отнася към различна перспектива, а именно случаите на разминаване между целите на принципала и агента, при което за първия е трудно да оцени действията на агента в условията на информационна асиметрия. Наред с това двамата имат различно предпочитание към риск, поради което в идентична ситуация те биха избрали различно действие. На тази база и изхождайки от допускания относно характеристиките на агента (личен интерес, степен на рационалност, избягване на риск), се определят критерии за ефективно взаимодействие (договаряне) помежду им.

Налице са редица основания, които определят приложимостта на тази концепция към отношенията между хората или организациите (принципал) и ИИ. На първо място, като агент ИИ взема решения или предприема действия, които засягат резултатите на неговия принципал. ИИ системите оптимизират зададените им целеви функции (подобно на максимизирането на полезността или друг резултат), като обработват информация и правят избор между алтернативи. ИИ системите могат да реагират на определени награди/наказания в рамките на програмираните си ограничения. Въпреки че не се предполага ИИ системите да развият независима мотивация, не се изключва конкретните им цели или начините за постигането им да не съвпаднат с тези на техния принципал.

На следващо място, налице е изразена информационна асиметрия, особено във връзка с проблема на черната кутия. Все пак ИИ не укриват стратегически информация, за разлика от човек, който има за цел да създаде превратна представа за своите резултати. Стои въпросът дали ИИ-агент може да бъде контролиран по-лесно отколкото човек-агент.

Взаимодействието между човека-агент и ИИ-агент е критично за организациите. Делегирането на правото за вземане на решения се обуславя както от способностите на ИИ в сравнение с човешките, така и от техните усилия за постигане на целта (Athey et al., 2020). Дори при по-добро представяне на

ИИ усилията на агента-човек са важни, когато ИИ не научава оптималното действие и не демонстрира несъгласуваност с принципала.

Основна разлика с класическия модел, при който в ролята на агент е човек (естествен субект), представляващ организацията (синтетичен субект), е, че при ИИ ролята може да е обърната – синтетичният субект действа от името и за сметка на естествения субект. При това функциите на съвременния ИИ (засега) са предимно програмирани, а не основани на независимо вземане на решения. Структурата на предпочитанията на ИИ е експлицитно дефинирана, но е възможно да претърпи промяна при по-напредналите системи.

„Договорът“ между принципала и ИИ се съдържа в неговата функция на възнаграждение и програмните му ограничения. На това основание проблемът с „моралния риск“ при ИИ системите би приел различна форма. Например системата може да намери неочаквани начини да заобиколи заложени ограничения и да максимизира възнаграждението (reward hacking). В съответствие с Agrawal et al. (2019) с намаляването на разходите за прилагане на ИИ, моралният риск, свързан с преценката, расте. Това би следвало да насърчи повече възнаграждения на база преценка, отколкото тези на база прогнозиране.

С напредването на системите на ИИ характерът на доверителните отношения между тях и техния принципал също може да претърпи промяна. Не е изключено да се измени посоката на делегиране, особено ако се отнася до взаимодействие на AGI и по-нискоквалифицирани служители.

На последно място, идеята за „призрачен агент“, заложена в заглавието, съчетава елементи от характеристиката на икономическия агент според традиционната икономическа теория и елементи на агента в ролята на довереник според концепцията принципал-агент. ИИ може да действа самостоятелно, но от името на друго лице, което понася последиците. Когато това обстоятелство не е оповестено, може да е налице „призрачен агент“. Такива са например скритите писатели (ghost writers), които създават произведения от името и за сметка на друго лице – номинален автор. В този случай предизвикателствата не са свързани само с моралния риск и непрозрачността на вземане на решения от ИИ. Допълнителни проблеми произтичат от скритата (подменената) идентичност на номиналния автор – например плагиатство, неавтентично съдържание, възможности за манипулации. Преодоляването на тези проблеми очевидно изисква адаптиране на регулациите – например чрез въвеждане на изискването да се обозначава съдържание, генерирано от ИИ. Но проблемът за призрачното агентство не се изчерпва с авторските права (в общия случай те са на човека, който формулира запитването). При

отсъствие на ясен статут на ИИ, е възможно да се стигне до прехвърляне на рискове или отговорности към скрития агент.

С оглед на задачите на настоящето изследване идеята за призрачния автор (призрачния изпълнител) може да послужи като мисловен експеримент, тестващ аргументите за и против самостоятелния статут на ИИ. Ако твърдим, че ИИ е само технология, която усилва възможностите на човека, то трябва да сме готови да приемем образите и текстовете, създадени от ИИ, като равностойни на човешките. Но ако допускаме, че призрачният агент е отделен субект, то това би ограничило използването на ИИ за усилване на репутацията или манипулативни практики.

Заключение

Въпросът за ролята на ИИ създава не само затруднения, но и възможности за икономическата теория и практика. Появата на ИИ намалява съществени проблеми, например правдоподобността на допускането за рационално поведение на икономическите агенти. В по-широк аспект ИИ преодолява и затрудненията, свързани с осигуряването на независимо и безпристрастно вземане на решения.

Въпросът дали ИИ може да се разглежда като пълноценен икономически агент със свои собствени интереси до голяма степен е свързан с неговата еволюция. Като се има предвид, че става дума за технология в развитие, настоящето изследване не може да даде окончателен отговор. Следователно на този етап можем да определим само предварителните предпоставки и критерии за определянето на ИИ като икономически агент. Тези критерии са свързани с автономността на ИИ, стабилността на неговите предпочитания, ограниченията и рисковете, пред които е изправен. Въпреки наличието на автономност механизмът на вземане на решения от ИИ не е идентичен на човешкия, като сред основните разлики са ролята на емоциите, личният опит и начинът на осъзнаване на последиците. Следователно анализът на ролята на ИИ изисква диференциран подход с оглед на неговите способности. Като общо правило може да се изведе, че колкото повече нараства независимостта на ИИ от човека, толкова повече той ще се превръща в икономически агент.

Въпросът дали той е достигнал ниво, което позволява да се признае като самостоятелен правен субект, е дискусионен, но не предопределя ролята на ИИ в процеса на вземане на решения. С оглед на разпределението на рисковете и формулирането на целите, както и с отчитане на информационната асиметрия е възможно ИИ да бъде носител на ограничени права и задължения.

Колкото повече се разширяват границите, в които действа ИИ, толкова повече ще се увеличава значението на взаимодействието между различните системи с ИИ и между ИИ и човека. От друга страна, с напредъка на ИИ ще нарастват случаите на емергентно поведение. В тази връзка разбирането на ролята на ИИ в процеса на вземане на решения би съдействало за по-доброто проектиране на тези системи.

Използвана литература

Регламент (ЕС) 2024/1689 на Европейския парламент и на Съвета от 13 юни 2024 година за установяване на хармонизирани правила относно изкуствения интелект и за изменение на регламенти (ЕО) № 300/2008, (ЕС) № 167/2013, (ЕС) № 168/2013, (ЕС) 2018/858, (ЕС) 2018/1139 и (ЕС) 2019/2144 и директиви 2014/90/ЕС, (ЕС) 2016/797 и (ЕС) 2020/1828. (Регламент (ЕС) 2024/1689 на Европейския парламент и на Съвета от 13 юни 2024 година за установяване на хармонизирани правила относно изкуствения интелект и за изменение на регламенти (ЕО) № 300/2008, (ЕС) № 167/2013, (ЕС) № 168/2013, (ЕС) 2018/858, (ЕС) 2018/1139 и (ЕС) 2019/2144 и директиви 2014/90/ЕС, (ЕС) 2016/797 и (ЕС) 2020/1828). (Reglament (ES) 2024/1689 na Evropeyskia parlament i na Saveta ot 13 yuni 2024 godina za ustanovyavane na harmonizirani pravila otnosno izkustvenia intelekt i za izmenenie na reglamenti (EO) № 300/2008, (ES) № 167/2013, (ES) № 168/2013, (ES) 2018/858, (ES) 2018/1139 i (ES) 2019/2144 i direktivi 2014/90/ES, (ES) 2016/797 i (ES) 2020/1828. (Reglament (ES) 2024/1689 na Evropeyskia parlament i na Saveta ot 13 yuni 2024 godina za ustanovyavane na harmonizirani pravila otnosno izkustvenia intelekt i za izmenenie na reglamenti (EO) № 300/2008, (ES) № 167/2013, (ES) № 168/2013, (ES) 2018/858, (ES) 2018/1139 i (ES) 2019/2144 i direktivi 2014/90/ES, (ES) 2016/797 i (ES) 2020/1828)).

Acemoglu, D. et al. (2017). Low-skill and high-skill automation, NBER Working Paper 24119.

Agrawal, A., Gans, J., Goldfarb, A. (2019). Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence, in Agrawal, A., Gans, J., Goldfarb, A. (eds), The economics of artificial intelligence. An agenda, The University of Chicago Press, Chicago.

Alonso, N. (2023). An Introduction to the Problems of AI Consciousness, The Gradient, available at: https://thegradient.pub/an-introduction-to-the-problems-of-ai-consciousness/

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention, in The Rate and Direction of Inventive Activity: Economic and Social Factors (Nelson, R. ed.). NBER.

Athey, S., Bryan, K. and Gans, J. (2020). The Allocation of Decision Authority to Human and Artificial Intelligence, AEA Papers and Proceedings, 110, pp. 80-84.

Brożek, B., Furman, M., Jakubiec, M. et al. (2024). The black box problem revisited. Real and imaginary challenges for automated legal decision making, AI Law 32, pp. 427-440.

Buttazzo, G. (2001). Artificial consciousness: Utopia or real possibility?, Computer, 34 (7).

Chalmers, D. (1995). Facing up to the problem of consciousness, Journal of Consciousness Studies 2(3), pp. 200-19.

Council of Europe, Committee of experts on Internet MSI-NET. (2017). Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications.

Doomen, J. (2023). The artificial intelligence entity as a legal person. Information & Communications Technology Law, 32(3), pp. 277-287.

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, Sci, 6(1), 3, https://doi.org/10.3390/sci6010003

Gervais, D. (2019). The Machine As Author, Iowa Law Review, Vol. 105, pp. 2053-2106, Vanderbilt Law Research Paper No. 19-35, available at: https://ssrn.com/abstract=3359524

Goertzel, B. and Pennachin, C. (2007). Artificial General Intelligence, Springer.

Russell, S., & Norvig, P. (2009). Artificial intelligence: A modern approach (3rd ed.), Upper Saddle River: Prentice Hall.

Searle, J. R. (1984.) Minds, brains and science (Reith Lectures), Harvard University Press.

Trajtenberg, M. (2019) Artificial Intelligence as the Next GPT: A Political-Economy Perspective, in Agrawal, A, Gans, J, Goldfarb, A (eds), The Economics of Artificial Intelligence An Agenda, The University of Chicago Press, Chicago.

Wagner, D. (2020). Economic patterns in a world with artificial intelligence, Evolutionary and Institutional Economics Review, 17, Issue 1.

***

IS AI A GHOST ECONOMIC AGENT

Chief Assist. Prof. Ivanina Mancheva, PhD
Social Sciences Department
Univercity of Architecture, Civil Engeneering and Geodesy
e-mail: mancheva-fte@uacg.bg

Abstract

The invasion of artificial intelligence (AI) into economic and social life is leading to significant transformations. They raise the question of the role of AI in economic models and, in particular, whether it should be treated as an economic agent that makes independent decisions. This question, as well as the related problem of the legal capacity of AI, has not been answered. They are related to many topics discussed in recent years: the possibility of taxing robots, the liability of autonomous cars or the copyright of works created by AI. The study advocates a differentiated approach to defining AI as an independent economic agent. If answered in the affirmative, the key question would be what is the function that the AI maximizes and what are the constraints and risks that it faces. Although AI is still developing its capabilities and we do not have sufficient observations of the behavior of the most advanced systems, we assume that the more its independence from humans increases, the more it will become an economic agent.

Keywords: artificial intelligence, economic agent, decision making, reward function, moral hazard, information asymmetry, principal-agent

JEL: D01, K10, O32, O33