

ЕТИЧНИ И ПРАВНИ ПРОБЛЕМИ, СВЪРЗАНИ СЪС СУБЕКТНОСТТА И ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Доц. д-р Веселина Славова

Медицински университет „Проф. д-р Параскев Стоянов“ – Варна

Доц. д-р Дарина Димитрова

Икономически университет – Варна

Резюме. С навлизането във все повече сфери на личния и обществения живот темата за ИИ става обект на много дискусии. Разширяването на възможностите на тези технологии поражда необходимостта от въвеждането на етичен и правен регламент с цел контрол. Оттук възниква въпросът доколко може да се смята една система за автономен субект подобно на личността, притежаваща тази способност.

Статията изследва въпроса дали една система с изкуствен интелект може да се смята за субект в етичен и в правен аспект. Целта е да се докаже разликата между човешкото мислене и поведение и дейностите, извършвани от системите с ИИ, за да се очертае рамка с етични и правни норми, регламентиращи създаването и използването на ИИ.

С прилагането на мултидисциплинарен подход, съчетаващ анализа на философски и юридически понятия за субекта, се изследват понятия като интелект, автономност, морална и правна отговорност.

Изводът е, че колкото и развити да са тези системи, те не могат да бъдат определени за субекти, тъй като не могат да извършват напълно съзнателни действия и не могат да носят нито морална, нито правна отговорност.

Ключови думи: субект; интелигентност; изкуствен интелект; етични норми; правни норми

Въведение

Още с изобретяването си изкуственият интелект (ИИ) става тема на много дискусии, третиращи различни въпроси от сферата на научното знание и технологиите.

ИИ се смята за феномен на четвъртата индустриална революция. За разлика от предходните настоящата се отличава със своя мащаб и комплексност, както и със скоростта, с която се развиват новите технологии. Това поражда

много въпроси за начина, по който технологиите и човекът ще съществуват заедно в едно бъдещо общество. Както посочва Клаус Шваб, четвъртата индустриална революция не се изчерпва само с интелигентните и свързани помежду си машини и системи. Нейният обхват е много по-широк, тъй като ще се стигне до сливането на технологиите и тяхното взаимодействие във всички сфери – физическа, цифрова и биологична (Schwab 2016, p. 21).

Системите с ИИ безспорно са технологично и конкурентно предимство в редица стопански дейности – маркетинг, продажби, управление на бизнеса, производство, застрахователно дело, финансов и банков сектор, медицински изследвания. Според някои автори същността на ИИ са научният анализ и автоматизацията на интелектуалните функции на човека (Atanasova & Boncheva 2000, p. 13). Но на този етап не е изведено единно определение на понятието „изкуствен интелект“, а само се посочват трите възможни форми на проявлението му: 1) тесен/първичен/слаб ИИ; 2) силен/общ/генерален ИИ и 3) супер/превъзхождащ човека ИИ (Todorova 2020, p. 50).

Първият вид ИИ вече е реалност. Той служи за събиране и обработка на данни, статистически анализи, маркетингови проучвания, езикови преводи, лицево разпознаване, чатботове, самоуправляеми автомобили. Неговото приложение и употреба са напълно под контрола на човека. Другите два вида ИИ все още не са разработени и не се използват на практика. Но предвид бурното развитие на технологиите и стремежа към икономическо превъзходство може би не е далече денят, когато това ще стане. Вторият вид ИИ се очаква да се изравни по ментални способности с човека, а суперизкуственият – да превъзхожда човешките умствени възможности, което поражда опасения от загубата на контрол над технологиите от страна на човека.

В този смисъл, рискове биха възникнали след създаването на генералния и суперизкуствения интелект, а това поражда редица етични и правни проблеми, за които трябва да се намерят адекватни решения. Въпросите, които възникват по повод използването на ИИ, са свързани най-вече с това какви ще са последиците за обществото, какви ще са взаимоотношенията и правоотношенията между хората и ИИ.

Целта на статията е да се изследва социалната функция на системите с ИИ от гледна точка на идеята за субектността. Чрез поставянето на граници между субекта човек и агента ИИ се определят и моралните и правните ограничения, които са задължителни при използването на системите, прилагащи ИИ.

В първата част се анализира понятието за субект във философията, като се прави съпоставка между естествен и изкуствен интелект и опитите на някои учени да докажат познавателния и моралния капацитет на последния.

Втората част разглежда въпроса за правосубектността на ИИ. Направен е преглед и критичен анализ както на литературата и доктриналните схващания по този въпрос, така и на нормативната уредба и законодателните тенденции относно правния режим на ИИ.

1. Субект и изкуственият интелект – философско-етични аспекти на проблема

Етичната дискусията относно ИИ произтича от по-фундаментален въпрос, а именно трябва ли да считаме системите, притежаващи ИИ, за субекти подобно на хората. Тъй като субектът се характеризира с две основни способности – мислене и действие, то този въпрос поставя два подвъпроса: първо – какво разбираме под мислещ субект и такива ли са тези системи; и второ – можем ли да ги третираме като автономни морални агенти, носещи отговорност за действията си?

1.1. Субектът като мислещ

В самото словосъчетание изкуствен интелект е заложена съпоставка с естествения. Първоначалната цел на употребата на ИИ е да подпомага извършването на определени логически операции по предварително заложиени алгоритми. Поради това се прави сравнение с човешкия интелект. С развитието на науката става възможно разработването на разширения, които да надхвърлят способността за справяне с конкретни проблеми, добавяйки нови аспекти на интелигентността, като адаптивност към нови обстоятелства и креативност при разрешаването на проблеми, за които не са програмирани. Оттук се поражда идеята за ИИ, надхвърлящ човешките възможности във всеки възможен аспект – и като решаващ логически задачи, и като правещ независими избори, като подбиращ, обработващ и прилагащ конкретна информация. Дискусиите се пренасят към идеите за независимост на агентите с ИИ и превъзхождащата им роля по отношение на човешките способности (Moor 2006; Vallor 2015). От това зависи доколко може да се говори за сходство между човека като субект и агент с ИИ. Дали трябва да се търси морална и правна отговорност на последния. За целта първо ще изследваме въпроса дали системите с ИИ са аналогични на човешкия интелект.

Щом като естественият интелект служи за основа при разработването на изкуствения, то първоначално трябва да се зададе въпросът „Какво е интелект?“.

Най-често интелигентността се свързва с ума, затова и голяма част от първите тестове за интелигентност изследват способностите на ума за логическо мислене. Има теории, основаващи се на идеята за обща интелигентност, наричана още *general intelligence*, или накратко „g“ (Jensen 1998). Този вид интелигентност най-често се интерпретира като адаптивност при взаимодействие „с външна среда, проблем или ситуация“ (Legg & Hutter 2007, p. 11). В същото време, някои учени оспорват обвързаността между g и адаптивността, като посочват проблемите на съвременния начин на живот – глобалното затопляне, увеличаването на вредните емисии във въздуха, унищожителните войни (Sterberg 2019), които показват, че общата интелигентност не е свързана с адаптивността в един общ и дългосрочен план за човечеството.

Възможна е дефиниция за интелект, ориентирана към рационалността и основаваща се на утилитарната теория, за която е важно постигането на добър резултат (Russell & Norvig 2002). Т.е. интелектът е рационална способност, която

трябва да доведе до успешен избор. Въпросът при тази концепция е кое може да се приеме за добър избор? Възможно е и изборът, който субектът прави, да му изглежда добър от определена гледна точка, но не толкова от друга, или е добър избор към настоящия момент, но не и погледнато в по-дългосрочен план.

Други изследователи разширяват кръга от способности, определяни като необходими за интелигентността. Х. Гарднър създава понятието за плуралистичен интелект или множество интелигентности (Gardner 2008). Освен логическата интелигентност той говори за междуличностна и личностна, музикална, телесна, езикова, логистична, пространствена интелигентност. За него интелигентността е повлияна от много фактори, включително културни и социални. Робърт Сърнбърг прави опит да обедини теориите за общия и плуралистичния интелект и създава т.нар. триархична теория, като определя интелигентността като съвкупност от фактори, надхвърлящи рационалните способности, и групира три вида интелигентност – аналитична, творческа и практическа (Sternberg 1985).

Въз основа на тези определения ще приемем, че интелектът е съвкупност не само от умствени способности, но и от емоционална и социална компетентност. Поради това не е възможно пренебрегването на значимостта на качества като познаване на себе си и другите, разглеждани като съзнание и самосъзнание и допълващи дефиницията за интелигентност.

Могат ли машините да мислят като човека, е въпрос, който се поставя още с появата на ИИ? По този въпрос и до днес се дискутират възможностите за тестване на интелигентността им посредством въведения от А. Тюринг тест (Turing 1950). Идеята на британския математик е да се проведе експеримент, при който човек, влязъл в ролята на съдия (тестващ), води разговор едновременно с машина и с човек, като се опитва да разпознае машината. Целта на системата е да заблуди задаващия въпроси съдия, че е човек. Премаването на подобен тест трябва да докаже, че машината е интелигентна.

След като човешкото мислене е все още в процес на изучаване, няма как да се използва като устойчив модел за тестване интелигентността на машината. Копирането на човешки способности е частично и избирателно, което означава, че е невъзможно да се постигне цялата система от фактори, способстваща за формирането, развитието и тестването на човешките интелектуални заложби. Разликата между изкуствен и естествен интелект трябва да бъде ясно очертана. Машинният интелект се различава от човешкия, на първо място, защото е създаден от човек и имитира определени операции, характерни за човешките мисловни процеси. Но за разлика от човека тази система не може да се ръководи от норми, извън заложените ѝ в процеса на програмиране. Дори и да надхвърли първоначалния проект, не се знае дали ще подобри, или ще влоши замисъла на своя създател. Ако допуснем, че машината преминава теста за интелигентност, това не означава, че е интелигентна, тъй като тя не осъзнава собствените си качества, а осъществява операции, които са ѝ заложили, т.е. липсват ѝ самосъзнание и самокритичност.

Разширяването на кръга от способности, които определят интелекта, води до следващия проблем, свързан с постигането на изкуствено съзнание. Ако приемем, че интелектът предполага и съзнание за себе си и за другите, осмисляне на действията и понасяне на отговорност, то в този смисъл е невъзможно да се говори за съзнание при тези системи.

Според изследователите автономност на ИИ може да се постигне, когато той взема решения независимо от първоначалния дизайн (Manzotti & Chella 2018). Или ако приемем, че алгоритмите, с които се управляват действията на една система, бъдат съобразени с научните закони, то роботите биха имали съзнание и съответно могат да извършват автономни действия. В същото време, ако приемем, че съзнанието е съвкупност от биологични и социални предпоставки, които влияят на развитието на когнитивни способности, емоционална и социална интелигентност у индивида, то съзнанието на една система не е съставено от тези компоненти и не може да развие подобни умения съзнателно и целенасочено.

Според Джон Сърл (Searle 1980), ако приемем, че интенционалността е способност на естествения интелект, постигната от биологични мозъчни процеси, то тя е несъществуваща при ИИ и не може да се създаде по изкуствен начин, колкото и сложна да е една програма. С помощта на експеримент, наречен „китайската стая“, той доказва несъпоставимостта между осъзнатите процеси, извършвани от човешкия ум, и процесите, симулиращи мисловна дейност, осъществени от една изкуствена програма.

Дискусиите „за“ и „против“ способността на ИИ да осъзнава своите решения, са насочени в посоката на доказване и отричане възможностите за разрешаване на проблеми, но не и върху способността за поемане на отговорност. Ако разгледаме случаите, в които ИИ ще се използва в контекста на междуличностните отношения и от него ще зависи вземането на важни решения за живота на хората (използването му като социален и здравен асистент, използването му в медицината или при подбор на кадри в предприятия или на студенти в университетите и други подобни), то тогава не е без значение осъзнаването на мотивите за извършване на съзнателен и отговорен избор, от който зависят животът, здравето и благополучието на тези хора.

Съзнанието включва в себе си света, нас самите и другите и по този начин стои в основата на вземането на решения и от гледна точка на добрия резултат, но и от гледна точка на мотивацията за извършване на действие. То е „субективно-преживяване на осъзнаването на определена мисъл или чувство“ (Haladjian & Montemayor 2016, p. 215).

ИИ не е субект и не притежава съзнание, в този смисъл. Той може да изпълнява дейности, свързани с възприемането на околната среда и използването на това знание за вземане на решения, „за да увеличат максимално шансовете за постигане на цел“ (Gabriel 2020, p. 412). Целта на този етап не е генерирана от системите, притежаващи ИИ, а по-скоро от техните дизайнери. Това означава, че

те не могат да обосноват самостоятелно целта си, да се ръководят от ценности и норми извън заложените им.

Така разглеждан, интелектът ще позволи поставянето на граница между изкуствен и естествен интелект и ще определи обхвата от дейности, които може да извършва една система. ИИ не включва всички съставни елементи на естествения интелект и това го определя по-скоро като инструмент, отколкото като самостоятелен фактор. Той няма съзнание и житейски опит, няма саморефлексия и възприемане на другите като такива и това го прави непълноправен участник във вземането на важни решения.

1.2. Субектът като морален агент

Морален агент е този, който прави мотивиран избор според определени морални норми и изпитва удовлетворение или разочарование от резултата на извършените постъпки.

Опасенията, произтичащи от прилагането на разширените форми на ИИ, са свързани с възможността той да може да взема независими решения. Това означава да извършва самостоятелни действия, които не са предвидени от неговия дизайнер и не са под контрола на човека, който го програмира или използва. Последиците от това разширение на способностите на ИИ са непредвидими и биха могли да бъдат опасни за човечеството. Извършването на самостоятелни действия, предхождани от избор, поставя въпроса за автономността на ИИ. Този въпрос е важен при определяне функциите на ИИ дали да бъде третиран като инструмент, или като автономен агент. Втората хипотеза предполага и понасянето на отговорност (морална и правна).

В настоящата статия ще защитаваме тезата, че трябва да се прави разлика между морален субект и автономен агент. Автономността на субекта не е тъждествена с неговата моралност. Един автономен агент може да извършва действия и да реализира правото си на избор в посока, противоречаща на моралните ценности, и обратно – един неавтономен субект може да извършва морални постъпки. Моралният субект съзнателно планира действията си съобразно морални ценности, приложими към конкретни ситуации. Автономният агент има свободата да избира, но може да прави неморални избори. За да се направи морален избор, би трябвало моралният субект да е едновременно и автономен, но и отговорен. Той трябва да мотивира действието съобразно определена цел, подчинена на морални принципи, и да търпи обратното действие на постъпката си. Т.е. моралният субект проектира постъпката си и се чувства ангажиран с възможните последици.

Дебатите, свързани с моралната автономност на ИИ, вървят в две посоки – от една страна, се доказва необходимостта от създаване на изкуствени морални агенти, от друга – се представят аргументи за рисковете от това действие. На аргументите, че създаването на тези системи е неизбежно, нужно за усъвършенстване на моралната аргументация, полезно за вземане на най-добри решения, могат да се противопоставят контрааргументите, че „те могат да засилят социалните

неравенства, да стеснят човешката морална мисъл и да допринесат за нарастващо незачитане на моралността на етичния агент“ (Herzog 2021, pp. 3 – 4).

Моралното отсъждане не може да се възпроизведе от машината, притежаваща силен интелект, тъй като тя не може да мотивира избора си съзнателно и отговорно. Контратезата на това, че човекът превъзхожда машината със способността си да мотивира решенията си за разлика от ИИ, е, че машината превъзхожда човека, тъй като може да преодолее пристрастията при вземането на решения. Тук възниква въпросът за прозрачността на вземане на решения. При човека може да се проследи мотивацията – користна или безкористна, докато при ИИ това би било доста трудно, дори и невъзможно.

Прозрачността на вземане на решения от ИИ е един от основните етични проблеми. Той възниква при въвеждането на алгоритми за обработка на големи бази данни за твърде кратко време, което не дава възможност за проследяване и контролиране действията на системата. По този начин може да се стигне до необясними и етично неоправдани решения. Това би могло да се превърне в предпоставка за обостряне на проблеми с неравенството, дискриминирането на групи от хора, конфиденциалността при използването на лична информация. Засегнати ще бъдат ключови сфери, като правосъдие, здравеопазване, социални дейности, подбор на кадри при постъпване на работа или в учебни заведения и всички, изискващи подбор и разпределение на социални блага (MacLure 2021). По тази причина не системите с ИИ, а дизайнерите и тези, които ги използват, трябва да бъдат отговорни за проектирането и прилагането им за решения, които касаят горепосочените дейности. В противен случай нараства недоверието в прозрачността на извършваните дейности.

Така стигаме до въпроса какви ценности трябва да регулират действията на ИИ. Трудността да се зададат основни принципи, от които да се ръководят компютърните системи, се основава на разнообразието от етични теории, дефиниращи ценностите (Nussbaum 2001; Drummond 2009). От дискусиите се открояват въпроси относно това какви ценности доминират – тези, от които се ръководи личността, или тези, които защитават обществото. Поради това ще има значение как ще програмираме етични оценъчни способности на машина, изпълняваща функции на социален или здравен асистент. В такъв случай не е възможно да залагаме различни модели на етичен избор в машините с различно предназначение, тъй като ще зададем морално противоречиви предпоставки, водещи до морален релативизъм.

Възможно решение на този проблем са договорените ценности, за да няма доминиращи индивиди или групи, които да определят ценностния модел, прилаган от ИИ. Тук стигаме до следващото предизвикателство – да допуснем неравенство при формулирането на етичните принципи, когато те се дискутират предимно в държавите с добре развита икономика и големите корпорации, които имат интерес да прилагат този вид технологии (Jobin et al. 2019, p. 396). Това означава, че вземането на решения относно етичната рамка на ИИ трябва да се основава на

ценностите на по-широк кръг от хора и общества с различна култура и икономическо състояние.

Ethics guidelines for trustworthy AI от 2019 г., изготвен от Експертна група на Европейската комисия (AI HLEG 2019), има за цел да създаде общи етични принципи, на които да се основава цялата идея за създаването и използването на ИИ. Създаването на такъв документ се налага от прекомерно широкия спектър от разнородни тълкувания, релативизиращи моралните решения (Floridi 2019).

Според ръководството има седем стъпки за постигане на надежден ИИ: човешко представителство и надзор; устойчивост и безопасност; поверителност и управление на данните; прозрачност; разнообразие, недискриминация и справедливост; обществено и екологично благополучие; отчетност.

Пак там се разглеждат три аспекта на доверието в ИИ (AI HLEG 2019, р. 5): в технологията на ИИ; в дизайнерите, които са го създали, и организацията, които го използват, и в социално-техническите системи, включени в жизнения цикъл на ИИ.

Създаващ доверие ИИ ни кара да помислим върху това дали е възможно да обвързваме технологията с подобни термини, характерни по-скоро за човека.

В анализа си на понятието „надеждност“ и прилагането му към ИИ Марк Райън защитава тезата, че доверието не може да бъде в ИИ, а в тези, които го създават (Ryan 2020). Според него не става дума за доверие в чисто човешки смисъл, а по-скоро за „форма на разчитане“. Ако доверието има рационални, емоционални и нормативни основания, то в нито една от изброените форми не може да се вмести отношението към ИИ. Това е причината да се прокара разделителна линия между доверието като морална категория и научната устойчивост на един продукт. В подкрепа на тезата си Райън посочва, че има разлика между това да съм разочарован от някого/нещо, на когото/което разчитам, и да се чувствам предаден от някого, на когото вярвам. Доверието е двустранен морален ангажимент между мен и някой друг. От една страна, аз възлагам очаквания, от друга – той знае, че трябва да ги оправдае.

На този етап на научно и социално развитие не може системите с ИИ да се третираят като независими агенти, доколкото не разполагат с независим интелект, обвързващ ги с извършването на независими морални действия. За целта отговорността за неморални и противоправни действия се носи от дизайнерите и потребителите, не от машините.

Въпросът за правосубектността на изкуствения интелект

1.3. Преглед на литературата и доктринални схващания

Човекът отдавна използва машини и компютри, но ако те притежават интелект на нивото на човешкия или превъзхождащи го, съвместното съжителство с тях ще се реализира по съвършено нов начин, което ще доведе до революционни промени, включително и в правната система. От една страна, могат да възникнат множество проблеми, свързани с инциденти, причинени от новите технологии,

както и във връзка с конфликтни ситуации при нарушаване правата на човека, което поражда въпроса кой ще носи отговорност в тези случаи. От друга страна, разработването на интелигентни машини и роботи поставя въпроса дали правният ред трябва да признае качество на правен субект на ИИ, съответно той какви права и задължения трябва да има. Това поражда необходимост от законови промени, които да защитят правата на човека и да гарантират устойчиво развитие на човешкото общество.

Съществуващите до момента системи с ИИ са все още напълно под контрола на човека, тъй като не са достигнали неговото интелектуално ниво, не го превъзхождат и предвид дейностите, за които се използват, те са обекти на правото. Съгласно действащата нормативна уредба, субекти на правото са отделните индивиди и организации, които по силата на правните норми могат да имат права и задължения. В този смисъл, нашето право познава следните видове правни субекти – физически лица (отделни човешки индивиди), юридически лица (организации) и държавата, като особен субект.

Основните юридически проблеми при бъдещи законодателни изменения, касаещи уредбата на ИИ, са свързани с отговора на въпроса дали да бъде, или да не бъде субект на правото. Ако правният ред признае на ИИ качеството на правен субект, тогава той ще има права и задължения, съответно ще трябва да носи юридическа отговорност за действията си.

От техническа гледна точка, ИИ се възприема като поредното в човешката еволюция оръдие на труда (Avishay et al. 2017). За икономистите новите условия на труд, свързани с развитието на технологиите, са предпоставка както за значителни ползи, така и за възникването на нови рискове в работната среда. ИИ ще замени трудещите се в конкретни икономически области – напр. ще може да обслужва голяма част от рутинните запитвания за обслужване на клиенти и да изпълнява монотонните задачи. Но същевременно масовото му навлизане може да причини на работещите психологически и социален тормоз, стрес от липсата на сигурен и постоянен статус, страх от безработица (Antonova & Ivanova 2021).

Икономическите интереси стоят в основата на навлизането на ИИ. Цифровите технологии отварят нови възможности за производителност, икономическо развитие и напредък в различни сектори на стопанския живот. Едновременно с това се пораждат множество рискове за правата на човека и за социалното развитие на обществото. Към настоящия момент софтуерът, контролиращ системите с ИИ, е „етично спял“. Като се имат предвид дългосрочните трансформации, които ще причинят автономните интелигентни технологии, социалната отговорност на компаниите изисква ползи не само за заинтересованите лица, но и за бъдещите поколения и за света, в който те ще живеят (Blagoysheva 2021).

Предвид навлизането на ИИ в трудовия процес, в редица юридически публикации той се разглежда предимно като субект на правото на труд (Andreeva et al. 2019; Andreeva & Yolova 2020a), а в други – като технически възможности за наблюдение

ние и контрол на персонала в електронна среда (Aleksandrov 2013). Навлизането и използването на ИИ в трудовите отношения е неизбежна тенденция, същевременно липсва регулаторен механизъм за участието му в трудовоправни връзки. Това налага необходимостта от създаване на адекватна регламентация, която да отговаря на динамиката на социалните процеси главно в насока осигуряване на подходяща етична и правна рамка при употреба на ИИ с оглед защита правото на труд на физическите лица (хората).

Но приложението на системите с ИИ едва ли ще се ограничи само до участие в трудовия процес и произтичащите от него трудови правоотношения. Предвид това, в юридическата литература се разглеждат правни решения, включващи възможността за най-усъвършенстваните автономни роботи да придобият статус на „електронни лица“. С оглед търсенето на отговорност за обезщетяване на вредите, които може да причини ИИ, се предлага прилагането на „електронна правосубектност“ за случаите, в които роботите вземат автономни решения или по друг начин взаимодействат независимо с трети страни (Stavru 2017).

На този етап в европейските държави интелигентните роботи все още се разглеждат като особен вид опасни вещи. Прилага се специален вещноправен режим на собственост върху тях, както и предвидена за собствениците им обективна отговорност за вредите, произлезли от тях. Този подход се обосновава с аргумента, че те все още нямат такава степен на автономност, че да бъдат самостоятелна причина за настъпване на определена вреда. Но потенциалната възможност от постигане на способност за адаптация и взаимодействия със средата, при които поведението става непредвидимо за собственика и за създателя му, прави интелигентния робот опасна вещ и отправя сериозни предизвикателства към съществуващата правна регулация. Това ще наложи прилагането на втория подход при уредбата на правния режим на интелигентните роботи, а именно – признаването им за специфични правни субекти, явяващи се самостоятелни носители на ограничена правосубектност, включваща способността им да носят отговорност за причинени от тях вреди. Предоставянето на дееспособност и правоспособност на интелигентните роботи ще внесе нов смисъл в понятията „субект“ и „обект“ в правото. Системите с ИИ се определят като „хибриди на бъдещето“, които не се вписват в съществуващите правни институти (Stavru 2020).

В редица публикации се изследват различни правни решения на проблема за отговорността на ИИ. Като не само се анализира отговорността при инциденти на работното място, свързани с употребата му (Andreeva & Yolova 2020b), но се стига и по-далеч – да се предлага въвеждане в законодателството на наказателна отговорност за престъпления, извършени от ИИ, като се правят аналогии с юридическите лица (Hallevi 2020). Подобна съпоставка търпи сериозни критики. От една страна, не във всички правни системи е допустимо юридическите лица да носят наказателна отговорност. От друга страна, не могат да се намерят достатъчно общи белези между юридическото лице и ИИ.

Доколкото има някаква прилика между тях, то тя се свежда до това, че и двете са създадени с цел да служат на интересите на човека, за което следва да се уреди законодателно техният правен режим. Разликите между тях обаче са съществени. Юридическото лице е правен субект, създаден по определен от закона ред, за да може самостоятелно да придобива права и да упражнява задължения, както и да отговаря по тях. Възниква от момента на вписването му в съответния регистър. В този смисъл, юридическото лице няма реално битие (съществува само „на хартия“, по документи), за разлика от ИИ, който съществува физически в обективния свят. Юридическите лица участват в правни отношения чрез органите си (състоящи се от физически лица – хора), които имат управителни и представителни функции, докато ИИ в един момент се очаква да може да действа автономно и непредсказуемо.

Прегледът на доктриналните разбирания за същността и правната природа на ИИ показва, че повечето автори подхождат обективно и отчитат както ползите, така и потенциалните рискове от въвеждането и използването му.

1.4. Правна уредба и законодателни тенденции

Поради очевидните различия между юридическите лица и ИИ някои автори разглеждат възможността в перспектива интелигентните системи, които са достатъчно автономни и независими от човека, да могат да получат правосубектност, като бъдат поставени в самостоятелна правна категория. Това следва да бъде правосубектност от нов тип, различна от тази на физическите и юридическите лица, която да отразява спецификата на тяхната правна същност. Системите с ИИ, като специфични субекти на правото, най-вероятно няма да се нуждаят от правна фикция, за да бъдат разглеждани като правни личности *sui generis* – т.нар. електронна личност. Целта на подобно правно решение се разглежда единствено в посока защита на правата и интересите на хората. По своята същност евентуалната правосубектност на ИИ поне първоначално ще трябва бъде по-ограничена от тази на човека (Kovacheva 2022).

Анализът на въпроса за правосубектността на ИИ ясно показва, че е необходима правна уредба на глобално ниво, тъй като не е възможно да се очаква самостоятелно решение от националното законодателство. Предвид мащабите на дигитализацията и възможността чрез интернет технологиите да се достигне до всяко кътче на земята, ИИ ще премине границите на отделната държава.

Понастоящем позицията на европейските институции и органи е, че системите с ИИ са машини, т.е. обекти на правото и на този етап не се предвижда да им се предостави правосубектност. Но цифровизацията и ИИ са важни приоритети за ЕС, затова вече се предприемат мерки за тяхната регламентация. Основният документ в тази насока е „Бяла книга за изкуствения интелект — Европа в търсене на високи постижения и атмосфера на доверие“ (Европейска комисия, Брюксел, 19.2.2020 г. COM(2020) 65 final). Приети са и две резолюции: 1) Резолюция на Европейския парламент от 20 октомври 2020

г. с препоръки към Комисията относно режим на гражданска отговорност за изкуствения интелект (2020/2014(INL)) и 2) Резолюция на Европейския парламент от 20 октомври 2020 г. с препоръки към Комисията относно рамка за етичните аспекти на изкуствения интелект, роботиката и свързаните с тях технологии (2020/2012(INL)).

От една страна, това не са нормативни актове със задължителен характер, а са стратегически документи, които само очертават рамката на една бъдеща нормативна уредба на ИИ. Като основни източници на производното право на ЕС, характер на законодателни актове имат регламентите и директивите. От друга страна, Бялата книга за ИИ и цитираните резолюции все пак са вид правни актове – други форми на действие за изграждане правния ред на Съюза. Като израз на обща политическа воля, резолюциите улесняват значително постигането на консенсус в работата на европейските органи и институции.

В България, като държава членка на ЕС, е приета с Протокол № 72.4 на МС от 16.12.2020 г. Концепцията за развитието на изкуствения интелект. По своята правна същност концепциите, подобно на стратегиите, не са нормативни актове със задължителен, а с програмен характер и представляват ръководни документи, които дават насоки, перспективи или конкретни задачи за изпълнение. Предвид това, в настоящия етап от развитието на европейското и на българското национално право, че все още ИИ не само че няма правосубектност, но и няма специален правен режим. Независимо че посочените европейски документи и българската Концепция не са нормативни актове със задължителен характер, приемането им трябва да бъде оценено като положителна първа стъпка в посока бъдеща правна уредба на статуса на ИИ. В този смисъл, те не са без правно значение.

Икономическите интереси и стремежът към по-голяма конкурентоспособност са водещ мотив за внедряване на иновации. Правото обаче е консервативна система и е безспорен факт, че нормативната уредба (не само в българското законодателство, но и в общоевропейски план) изостава от техническия прогрес. Ако от оръдие на труда и обект на правото ИИ се трансформира в правен субект, то това ще бъде нов институт за правната система. Какъв ще бъде тогава неговият правен режим, тепърва предстои да се уреди законодателно. Една популярна латинска сентенция гласи: „Правото е изкуство за доброто и справедливото“. Кое е добро и зло, кое е справедливо и несправедливо, се определя от ценностната система на човека, т.е. от етичните му норми. Ако ИИ достигне или дори превъзхожда човешките умствени способности, то тогава е логично той да бъде законодател, а какви ще бъдат етичните ценности, от които ще се ръководи този нов нормотворец, и дали изобщо ще има такива, е все още неизвестно.

Ако на ИИ му се предостави правосубектност и бъде нормативно уреден нов вид правен субект „електронно лице“, това ще доведе до кардинални промени както в правната, така и в обективната действителност. А както е известно, върху правната действителност въздействие оказват редица обществени явления, които

имат оценъчен характер – справедливост, естествени права, нравственост, свобода, принуда, интереси, цел и правосъзнание (Boychev 2006, pp. 5 – 9). Както вече бе посочено, въпросът за правосубектността на ИИ изисква правна уредба на глобално ниво. Проблемът е, че няма ценностна система, общовалидна за всички общества и народи.

Независимо от всички резерви и опасения ИИ е едно от най-големите и значими научни и технологични достижения на нашето време. Прегледът на юридическата литература показва, че повечето автори анализират основно възможните варианти и подходи за признаване правосубектност на ИИ. Но дори и най-добрата юридическа техника не може да създаде адекватна правна конструкция, ако в основата на законодателството не стоят нравствени и етични ценности. Затова е необходим интердисциплинарен подход, изискващ търсене на решения както на правните, така и на етичните, наред със социалните и икономическите предизвикателства, които поставя развитието на новите технологии.

Ако се приеме, че на ИИ може да бъде предоставена правосубектност, то тогава трябва ясно да се определи обемът от права и задължения, с които ще разполага. Въпросът за правосубектността на ИИ може би все още не е на дневен ред, но тенденцията, която се очертава, е към създаване ако не на глобална правна регламентация, то поне на ниво ЕС. От една страна, правните разпоредби няма как да изпреварят технологичното развитие. От друга страна обаче, има проблеми, които трябва да бъдат предвидени още преди системите с ИИ да са достигнали интелектуалното ниво на човека, т.е. докато все още са под негов контрол.

Изложеното обосновава извода, че е належаща необходимостта от приемане на нормативна уредба по въпросите, касаещи правния режим на ИИ и неговото използване. В Европа вече са предприети стъпки в тази посока. На 25 ноември 2022 г. в официалния сайт на ЕС е публикувано Предложение за Регламент на Европейския парламент и на Съвета за определяне на хармонизирани правила относно изкуствения интелект (законодателен акт за изкуствения интелект) и за изменение на някои законодателни актове на Съюза (COM/2021/206 final).

Заклучение

При разглеждането на проблема за субектността на системите, използващи ИИ, от гледна точка на етиката и правото могат да се направят няколко извода.

Първо, има разлика между субект и ИИ и тя се извежда от дефинициите за интелигентност, съзнателност, морална и правна отговорност. Доказва се разликата между естествен и изкуствен интелект, между автономен агент и морален и правен субект.

Второ, въз основа на направеното разграничение между субект и система се аргументира тезата, че отговорност за неморални и противоправни деяния не

може да се носи от машина с ИИ независимо от нивото на интелигентност и автономност. На този етап отговорността трябва да се търси от дизайнерите и корпорациите, използващи системи с ИИ.

Трето, след анализ на утвърдените към момента етични и юридически стандарти за регулиране на дейностите, свързани с прилагане на ИИ, се прави изводът, че влагането на човешки качества задълбочава, а не решава проблемите от етичен и правен характер. Обсъждат се проблемите, които възникват от приписването на интелигентност, автономност, отговорност при вземането на решения, и се посочват възможните злоупотреби.

Четвърто, обсъждат се проблемите, възникващи с изработването на етична рамка, основаваща се на определени ценности, които да стоят в основата на законодателството. Поставя се въпросът за валидността на тези ценности от гледна точка на личния и обществен интерес, от гледна точка на различните общества, съобразно различни фактори – културни, икономически, социални.

REFERENCES

- ALEKSANDROV, A., 2013. Workplace surveillance as a means of employer control. *Current Issues of Labour and Social Security Law*, Vol. 6, pp. 68 – 86. [in Bulgarian]
- ANDREEVA, A. & YOLOVA, G., 2020a. The labour-law principles – evolution and transformation in the era of digitalization and using of artificial intellect. *Izvestiya Journal of Varna University of Economics, Varna*, 64(1), pp. 22 – 35. [in Bulgarian]
- ANDREEVA, A. & YOLOVA, G., 2020b. The liability in case of incidents on the workplace, connected to the use of artificial intellect – contemporary challenges in front of the legislation. *Yuridicheski sbornik*, Vol. 27, pp. 230 – 239. [in Bulgarian]
- ANDREEVA, A., YOLOVA, G. & DIMITROVA, D., 2019. Artificial intellect: Regulatory Framework and Challenges Facing the Labour Market. *CompSysTech '19: 20-th International Conference on Computer Systems and Technologies*, 21 – 22 June 2019, University of Ruse, Bulgaria: Proceeding, New York: ACM [Association for Computing Machinery] Digital Library, pp. 74 – 77. <https://doi.org/10.1145/3345252.3345261>
- ANTONOVA, K. & IVANOVA, P., 2021. Robotization and artificial intelligence – a new model of cooperation and / or interaction between machines and people in the workplace. *HR & Technologies*, Issue 1, pp. 60 – 76. [in Bulgarian]
- ATANASOVA, T. & BONCHEVA, P., 2000. *Artificial intelligence*. Varna: Bryag Print. [in Bulgarian]

- AVISHAY, D., PAVLOV, V. & PAVLOVA, G., 2017. Fourth industrial revolution – artificial intellect, robots and automation. *Proceedings XXVI MNTK “ADP 2017” – XXVI International Scientific and Technical Conference “Automation of Discrete Manufacturing, Technical University”*, Sofia, pp. 28 – 33. [in Bulgarian]
- BLAGOYCHEVA, H., 2021. Artificial intelligence and corporate social responsibility – collision and new opportunities. *Izvestiya Journal of Varna University of Economics, Varna*, 65(1), pp. 55 – 69. [in Bulgarian]
- BOYCHEV, G., 2006. *Introduction to Law, Part Two. Philosophy of Law*. Sofia: Sv. Kliment Ohridski. [in Bulgarian]
- CHELLA, A., & MANZOTTI, R., 2013. *Artificial consciousness*. Andrews UK Limited.
- DRUMMOND J., 2009. Feelings, emotions, and truly perceiving the valuable. *Mod Schoolman* 86(3 – 4), pp. 363 – 379.
- AI HLEG, 2019. Ethics Guidelines for trustworthy AI. European Commission, high level expert group on AI. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- GABRIEL, I., 2020. Artificial Intelligence, Values, and Alignment. *Minds & Machines* 30, pp. 411 – 437. <https://doi.org/10.1007/s11023-020-09539-2>
- GARDNER, H. E., 2008. *Multiple intelligences: New horizons in theory and practice*. Basic books.
- GOTTFREDSON, L. S., 1997. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24(1), pp. 13 – 23.
- HALADJIAN, H.H. & MONTEMAYOR, C., 2016. Artificial consciousness and the consciousness – attention dissociation, *Consciousness and Cognition*, Volume 45, pp. 210 – 225. <https://doi.org/10.1016/j.concog.2016.08.011>
- HALLEVI, G., 2020. The criminal liability of artificial intelligence entities – from science fiction to legal social control. *Savremenno pravo*, Issue 1, pp. 7 – 37. [in Bulgarian]
- HERZOG, C., 2021. Three Risks That Caution Against a Premature Implementation of Artificial Moral Agents for Practical and Economical Use. *Sci Eng Ethics* 27, 3. <https://doi.org/10.1007/s11948-021-00283>
- FLORIDI, L., 2019. Establishing the rules for building trustworthy AI. *Nat Mach Intell* 1, pp. 261 – 262. <https://doi.org/10.1038/s42256-019-0055-y>
- JENSEN, A.R., 1998. *The G Factor*; Greenwood-Praeger: Westport, CT, USA.
- JOBIN, A., IENCA, M., & VAYENA, E., 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), pp. 389 – 399.
- KOVACHEVA, D., 2022. The new subjects of law: are artificial intelligence systems already among us? *De Jure*, 13(2), pp. 227 – 249. [in Bulgarian]

- LEGG, S.& HUTTER, M., 2007. Universal Intelligence: A Definition of Machine Intelligence. *Minds & Machines* 17, pp. 391 – 444. <https://doi.org/10.1007/s11023-007-9079-x>
- MACLURE, J., 2021. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. *Minds & Machines* 31, pp. 421 – 438. <https://doi.org/10.1007/s11023-021-09570-x>
- MANZOTTI R. & CHELLA A., 2018. Good Old-Fashioned Artificial Consciousness and the Intermediate Level Fallacy. *Front. Robot. AI* 5:39. doi: 10.3389/frobt.2018.00039
- MOOR, J., 2006. The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, 21(4), pp. 18 – 21. <https://doi.org/10.1109/MIS.2006.80>.
- NUSSBAUM M., 2001. *Upheavals of thought: The intelligence of emotions*. Cambridge University Press, Cambridge.
- RUSSELL, S. J., NORVIG, P., 2002. *Artificial Intelligence: A Modern Approach* (2nd Edition). Prentice Hall. ISBN: 0137903952
- RYAN, M., 2020. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Sci Eng Ethics* 26, pp. 2749 – 2767. <https://doi.org/10.1007/s11948-020-00228-y>
- SEARLE, J., 1980. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), pp. 417 – 424. doi:10.1017/S0140525X00005756
- SCHWAB, K., 2016. The fourth industrial revolution. Sofia: Hermes. [in Bulgarian]
- STAVRU, S., 2017. Electronic persons – new “uses” of legal personality. Scientific readings on “Legal norms and legal principles”. *Proceedings of a scientific conference organized by the Department of Theory and History of the State and Law and the Department of Criminal Justice of the Faculty of Law of the Sofia University “St. Kliment Ohridski”*, pp. 494 – 510. [in Bulgarian]
- STAVRU, S., 2020. The robot – a new “other” between ourselves. *Otherness as a Moral Challenge (The Bulgarian Context)*. V. Tarnovo: Faber, pp. 249 – 267. [in Bulgarian]
- STERNBERG, R. J., 1985. *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press, New York.
- STERNBERG R. J., 2019. A Theory of Adaptive Intelligence and Its Relation to General Intelligence. *Journal of Intelligence*. 7(4):23. <https://doi.org/10.3390/jintelligence7040023>
- TODOROVA, M., 2020. *Artificial Intelligence. A brief history of the development and ethical aspects of the topic*. Sofia: Iztok – Zapad. [in Bulgarian]

VALLOR, S., 2015. Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology*, 28(1), pp. 107 – 124. <https://doi.org/10.1007/s13347-014-0156-9>.

ETHICAL AND LEGAL PROBLEMS RELATED TO SUBJECTIVITY AND ARTIFICIAL INTELLIGENCE

Abstract. AI enters more and more spheres of personal and public life, the topic becomes the object of much discussion. The expanding possibilities of these technologies give rise to the need to introduce ethical and legal regulation in order to control them. Hence the question arises as to how far a system can be considered an autonomous entity, similar to the individual possessing this capability.

This article examines the question of whether a system with artificial intelligence can be considered a subject in ethical and legal aspect. The aim is to demonstrate the difference between human thinking and behaviour and the activities performed by AI systems in order to outline a framework with ethical and legal norms governing the creation and use of AI.

By applying a multidisciplinary approach combining the analysis of philosophical and legal notions of the subject, concepts such as intellect, autonomy, moral and legal responsibility are explored.

The conclusion is that no matter how developed these systems are, they cannot be defined as subjects because they cannot perform fully conscious acts and cannot be held morally or legally responsible.

Keywords: Subject; Intelligence; Artificial intelligence; Ethical norms; Legal norms.

✉ **Dr. Veselina Slavova, Assoc. Prof.**

Researcher ID AAB-8548-2019

Medical University „Prof. Dr. P. Stoyanov“

55, M. Drinov St.

9000 Varna, Bulgaria

E-mail: veselina.slavova@mu-varna.bg

✉ **Dr. Darina Dimitrova, Assoc. Prof.**

Researcher ID AAB-9007-2019

Legal Studies Department

University of Economics – Varna

77, Kniaz Boris, Blvd.

9002 Varna, Bulgaria

E-mail: darina@ue-varna.bg