

Правни норми и етични аспекти на изкуствения интелект

Виктор Добрев

Велико Търново
2022

Правни норми и етични аспекти на изкуствения интелект

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "В. Левски" - гр. Велико Търново, и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 „Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове

© Доц. д-р Виктор Тодоров Добрев, автор

Първо издание

Рецензенти:

Проф. д.ик.н. Красимир Марков Марков

Полк. доц. д-р Невена Цветкова Атанасова–Кръстева

Научен редактор:

Проф. д-р пс. н. Людмил Дочев Георгиев

Технически редактори

Светлана Маркова Зотова

Христо Пеев Христов

Издателство

©Национален военен университет „Васил Левски“

Факултет „Артилерия, ПВО и КИС“

ISBN: 978-619-7531-

Велико Търново, 2022

СЪДЪРЖАНИЕ

ПЪРВА ТЕМА. ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ КАТО ФИЛОСОФСКИ ПРОБЛЕМ	6
Етика – теоретични аспекти	12
Основни принципи на хуманизма	15
ВТОРА ТЕМА. ДАННИТЕ – ОСНОВА НА ИКОНОМИКАТА ...	16
Какво разбираме под данни?	17
Етични проблеми, свързани с данните	21
Големи данни.....	22
Акумулирани от компаниите данни.....	23
Акумулирани данни от държавата	24
Търсене и намиране на етични решения при работа с данни.....	26
ТРЕТА ТЕМА. ЕТИКА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ	34
Формулировка за изкуствен интелект	36
Етично използване на изкуствения интелект	40
Отговорност на изкуствения интелект.....	41
Степен на прозрачност на изкуствения интелект	43
Предубеденост спрямо алгоритмите.....	44
Надеждност на изкуствения интелект	49
Приватност на информацията, получена с помощта на анализ на метаданни.....	50
ЧЕТВЪРТА ТЕМА. МАШИННА ЕТИКА	51
Подходи за решаване на проблема на машинната етика	51
Алгебра на съвестта – формализация на етиката.....	55
Каква етика трябва да се прилага при машините?	57
ПЕТА ТЕМА. СЪВРЕМЕННИ ЕТИЧНИ ФЕНОМЕНИ, ПОРОДЕНИ ОТ ИЗПОЛЗВАНЕТО НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ	58
Културни различия на етичните норми за изкуствения интелект	58

Държавата и изкуственият интелект – етика и доверие.....	59
Оценка на въздействието на алгоритмите – от технологията към правата на човека	63
Оценка за въздействието на алгоритмичните системи – подходи и методики за регулиране.....	65
Приватност и защита на персоналните данни.....	66
Държавното проследяване – „Големият брат“	71
Скринингови системи	72
Китайският „Голям брат“	73
Дискриминация и неравнопоставеност при цифровизацията	76
Пазар на труда и риск от безработица	79
ШЕСТА ТЕМА. ОТРАСЛОВА ЕТИКА СПРЯМО ЦИФРОВИТЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИЕТО, ЗДРАВЕОПАЗВАНЕТО И СИГУРНОСТТА.....	83
Етични проблеми на цифровото образование.....	85
Етика на цифровата медицина	88
Регулация и модели за развитие на цифровата медицина	90
Социалната подкрепа и цифровизацията	91
Етични проблеми на цифровите технологии в работата на полицията	93
Етика на прилагането на изкуствения интелект за военни цели.....	99
СЕДМА ТЕМА. РЕГУЛИРАНЕ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ В СВЕТОВНАТА ПРАКТИКА.....	101
Национални документи и стратегии за развитие.....	102
Законови и подзаконови актове	103
Изследвания върху етиката на изкуствения интелект.....	104
Етични документи в областта на изкуствения интелект.....	104
Стандарти и доктринални източници	105
Международни актове относно етиката на изкуствения интелект	106
Устойчиво развитие и цифрови технологии	108
Права на човека и цифрови технологии	111

Право на съблюдаване и защита на личния живот.....	112
Свобода на словото	114
Забрана за дискриминация	116
Задължения на бизнеса и частните лица.....	119
ИЗПОЛЗВАНА И ПРЕПОРЪЧАНА ЛИТЕРАТУРА.....	123

ПЪРВА ТЕМА. ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ КАТО ФИЛОСОФСКИ ПРОБЛЕМ

В тази тема ще се запознаете с възникването на философските проблеми, свързани с изкуствения интелект, от древността до наши дни. Накратко ще бъде пояснено какво е това етика, какви са основните концепции в нея и защо тя е необходима по отношение на изкуствения интелект.

Достигането до същността на даден проблем е свързано с разглеждане на всички източници по хронологичен ред от човешката култура, влияещи върху сегашните нагласи, в случая спрямо изкуствения интелект (ИИ). Човечеството не за първи път си задава въпроса какво ще бъде неговото битие с технологиите.

В този поток от информация могат да бъдат открити изключително екзотични идеи и коментари по отношение на ИИ, които присъстват в колективното ни съзнание и най-вече в Западното културно пространство, оставайки релевантни и в наше време.

Идеята да се създават живи същества от неодушевена материя може да бъде намерена в различни източници на шумерски, китайски, еврейски език и е свързана с всички религиозни конфесии.

При древногръцката митология можем да видим прото-идея за създаване на изкуствени човешки същества и по-точно жени. В „Илиадата“ Hephaestus помага на слугите да бъдат направени от злато и да наподобяват жени. В мита за Пигмалион скулпторът се влюбва в създадената от него от слонова кост женска фигура и моли богинята Афродита да ѝ вдъхне живот. Тук можем да видим препратка към съвременните сексуални роботи.

В своята книга „Автоматите“ гръцкият математик Херо от Алесандрия описва машина, която предсказва годишната динамика на астрономически явления. През 1901 година по-

добна машина е изложена като археологическа находка, наречена Antikythera, която е най-старият пример на древен аналогов компютър, задвижван от часовников механизъм.

В древноеврейските текстове се описва създаването на „Голем“ - чудовище от глина, създадено от равин през шестнайсети век, върху което хората загубват контрол. Тук за първи път се сблъскваме с проблема за контрола (митът за Прометей се тълкува по същия начин, давайки знак, че някои неща трябва да останат недостъпни за простосмъртните).

Апогееят на темата се явява при Мери Шели в нейния роман за Франкенщайн. В края на историята за Голем хората успяват да го унищожат, докато в романа това не се случва. Въпреки че е създаден през епохата на Романтизма и носи възторга от науката и машините, той е ясно послание към съвременното и изкуствения интелект.

По-съвременните версии на романи и електронни игри екстраполират различни варианти на негативен сценарии с бунтове на роботи и деструктивни атаки от страна на изкуствения интелект спрямо хората. Най-ярка е филмовата версия на „Терминатор“ от 2015 година. Този страх пред машините известният писател фантаст Айзък Азимов нарича „комплекс Франкенщайн“.

Страхът бива подклаждан и от изказванията на учени като Стивън Хоукинс, който през 2017 г. казва, че създаването на изкуствения интелект ще е фатална стъпка за човечеството.

Съществува обаче и друга гледна точка, наричана трансхуманизъм, залегнала в западната религия и философско мислене (учението на Платон и юдео-християнската традиция). Тук религията и технологията хронологично се движат в Западната култура.

Теистичната идея за превъзходството на Бог независимо от материалния и физически свят е негово вътрешно присъщо качество. Той е създател на всичко съществуващо и то е негово вътрешна същност. Разкриването му се получава през неговия син Христос и Светия Дух. Когато разглеждаме

Франкенщайн, нишката между създател и създадено (между Номо деус и изкуствения интелект) се къса. Подобна надлиминарност е характерна за философската традиция на Запада. През второто столетие след новата ера в Средиземноморието се заражда течение, наричано гностицизъм, търсещо божествената искра в човешкото тяло. Преди това Платон вижда човешкото тяло като затвор за душата, която иначе е безсмъртна. В неговата метафизика се разграничава формата на нещата в променящия се свят и надхвърлящата ги същност. Подобно нещо има и в трансхуманизма, където се надхвърля човешката ограниченост и с приемането на предизвикателствата на Платон и гностицизма, се постига безсмъртие, като биологичното тяло трябва да се надгради посредством някакъв изкуствен агент. Извличайки от нечистия материален свят с помощта на математиката и техническите науки на чисти форми на разума, се реализира нова форма на платонизма, придобиващ качеството на технология - алгоритмите на изкуствения интелект стават Платоническа машина, извличаща формата (модела) от света на проявленията. В християнската традиция това скъсява дистанцията между Бог и хората, правейки ги равни на него и доказвайки богоподобие.

Трансхуманизмът загатва за възможността за безсмъртието на човека, но в същия момент се появява и антиподът – идеята за Страшния съд и Апокалипсиса.

За да въведем някаква философска универсалност в разсъжденията, приемаме, че:

- Животът се възприема като процес, притежаващ способността да се самовъзпроизвежда при запазване на сложността, преминавайки през три етапа – *биологичен живот*, където на езика на информатиката можем да кажем, че живите организми са хардуерът, а софтуерът се развива по еволюционен път; *културата като форма на живот*, където софтуерът може да бъде проектиран на основа обучение, и *технологичен живот*, където и софтуерът, и хардуерът могат да бъдат проектирани и човек получава контрола върху собствената си съдба. *Цивилизацията* е взаимодействието между

разумните форми на живот. Съзнанието - това съвсем накратко са личните преживявания, а квалия е техният индивидуален характер. Телеология е начинът нещата да бъдат обяснявани през намерения и цели, а не през причини и следствия.

- Изкуственият интелект е получен по небιологичен начин и може да се раздели на „слаб (ограничен) интелект“ – възможностите му се ограничават до постигане на определени цели и задачи; „силен интелект“ – можещ да постигне практически всякакви цели, включващи и самообучение; „универсален интелект“ – способен да стане силен интелект при наличието на данни и ресурси; „силен изкуствен интелект“ (на човешко ниво) AGI – можещ да се справя с всякакви познавателни операции не по-лошо от човека; „интелектуален взрив“ – самоусъвършенстване, довеждащо до появата на „свръхразум“, наричано „сингулярност“, „свръхразум“ – значително превъзхождащ човешкия.

- Изкуственият интелект дава възможност така формулирания *технологичен живот* да бъде постигнат още през този век, но от нас зависи кога и как това може да се случи. Полемиката по този въпрос се води между три основни лагера: „лудети“ (техноскептици), „цифрови утописти“ и движение за „приятелски изкуствен интелект“.

- Според лудитите¹ създаването на свръхчовешки универсален ИИ е толкова сложно, че за това са необходими повече от едно столетие и съществува опасност той да ни унищожи в момента, в който започне да усеща някаква ограниченост.

- Цифровите утописти приемат, че постигането на свръх ИИ е възможно още през този век и те го очакват като

¹ Лудитите са социално движение от XIX век на английските производители на текстил, които протестират, често унищожавайки механизирани тъкачни станове, обявявайки се срещу промените, предизвикани от промишлената революция. Смятат, че те ги оставят без работа и променят начина им на живот. Движението е наречено на генерал Нед Лъд или на крал Лъд – митична фигура, за която се е смятало, че живее в Шеруудската гора, като Робин Худ. Днес „лудит“ се употребява за тези, които се противопоставят на индустриализацията, автоматизацията, компютризацията или на новите технологии изобщо.

естествена и желана стъпка към развитие на космическата ни еволюция.

- Движението за приятелски ИИ също предполага появата на свръхразум още през този век, но привържениците на тази идея не го приемат като гарантирано положителен ход – силно се акцентира върху контрола и безопасността.

- Основните митове относно изкуствения интелект са:

- 1) Изкуственият интелект е неизбежен до 2100 година – фактите говорят, че това може да се случи и в близките десетилетия или може изобщо да не се случи и никой не знае точния отговор.

- 2) За това какъв ще е ИИ, се тревожат само лудите – фактите показват, че загриженост проявяват голяма част от специалистите в областта.

- 3) Митологичен повод за тревога е злонамереността на ИИ и това, че той ще придобие съзнание – реалният повод за тревога е, че целите на ИИ могат да станат несъвместими с нашите човешки такива.

- 4) Главната опасност е в роботите – фактите показват, че ако целите ни не съвпадат, то на ИИ не му е необходимо никакво тяло, а само интернет връзка.

- 5) Изкуственият интелект не може да подчини хората – факт е, че както човекът може да подчини тигри и лъвово благодарение на разума си, така и ИИ ще може да го направи.

- 6) Машините не могат да имат цели – фактите показват, че както ракетите с термоориентиране имат цел, така и при ИИ може да се появи такава.

- 7) Митологичен повод за тревога е появата на ИИ след няколко години – реалността показва, че това ще се случи най-рано след няколко десетилетия и това е достатъчно време да се гарантира неговата безопасност.

Съвременната философия подема темата за изкуствения интелект, преодолявайки митологичното ниво с *Теста на Тюринг*. Този философски казус е предложен от британския математик Алан Тюринг през 1950 г. в статията му „Изчислителни машини и разум“ (Computing Machinery and

Intelligence), публикувана в списание „Mind“ (сп. „Съзнание/Ум“). Тестът е за проверка дали компютърът има разум в човешкия смисъл на думата. Тюринг предлага тест, който да замени безсмисления според него въпрос „Може ли машината да мисли?“ с по-определен. Съвременната интерпретация на този тест изглежда по следния начин: Човек взаимодейства дистанционно с двама събеседници – компютър и човек; на основание на отговорите на въпросите тестовият субект трябва да определи с кого разговаря – с машина (изкуствен интелект/компютърна програма) или с жив човек; задачата на компютърната програма е да въведе човека в заблуда, карайки го по този начин да направи грешен избор.

До 2008 година няма информация някоя от съществуващите компютърни системи да е преминала теста.

Тестът е на основата на игра с името „Imitation game“ (игра на имитация).

Мъж и жена се намират в една стая и си разменят бележки с водещия, който е в друга стая. Задачата на водещия е да определи кой от двамата е жената. Задачата на мъжа е да обърка водещия. Задачата на жената е да помогне на водещия да направи правилен избор. Тюринг задава въпроса какво ще стане, ако мъжът се замени с машина. Ще може ли водещият да направи правилен избор и ще се поддаде ли на машината да го излъже? Отговорът на този въпрос показва може ли машината да се явява като мислеща.

Тестът е необходимо да протече по следния начин: съдия (човек) разменя съобщения на обикновен език (естествен, човешки език, напр. български или английски) с двама събеседници. Единият от тях е машина, а другият е човек. Предполага се, че всеки от събеседниците ще се стреми да бъде възприет като човек. За да се направи тестът универсален, е нужно да се разменят текстови съобщения (напр. посредством програма за чат разговори)².

²https://bg.wikipedia.org/wiki/%D0%A2%D0%B5%D1%81%D1%82_%D0%BD%D0%B0_%D0%A2%D1%8E%D1%80%D0%B8%D0%BD%D0%B3

Някои оспорват идеята, че разумът е равнозначен на разумно поведение. Най-известният контрааргумент е мисловният експеримент на Джон Сърл (John Searle), наречен „Китайска стая“. Сърл говори за експеримент с човек, който не владее китайски, и е затворен в стая. Отвън стои друг човек, който може през процепа за писма да пуска вътре бележки, написани на китайски. Човекът в стаята разполага с голям наръчник с подробни инструкции, които да му помогнат да отговаря на получаваните отвън бележки.

Сърл твърди, че дори човекът отвън да остане с впечатлението, че води разговор с някого, който владее езика, човекът в стаята не разбира китайски. Сърл добавя, че по същата логика дори ако една машина се държи разумно и премине теста на Тюринг, от това не следва, че тя е разумна или че има „разум“ в човешкия смисъл на думата. Същото е в сила, ако заменим думата „разумен“ със „съзнателен“.

ЕТИКА – ТЕОРЕТИЧНИ АСПЕКТИ

Преди да заговорим за етични проблеми, възникващи като последствие от цифровизацията, е добре да поясним основните етични понятия и концепции, имащи отношение към нея.

Наименованието *етика* е предложено от Аристотел с идеята тя да бъде науката за правилното поведение. Впоследствие се появяват нови и различни концепции, но етиката като цяло присъства в основите на много науки, насочени към изучаването на живота и поведението на хората.

И така:

Етиката е философска дисциплина, изучаваща поведението на хората въз основа на нравствените им мотиви.

Тя показва какви могат да бъдат последствията, ако дадена конкретна идея бъде възприета от отделен човек или социум, представяйки всички нейни предпоставки и перспективи. По този начин етиката може да бъде определена като рефлексия на морала на дадено общество. От другите средства за социална регулация, като правото, традицията и обича-

ите, моралната норма се отличава с това, че винаги дава свобода на избор и преимуществено регулира вътрешните чувства на срам, дълг, угризения на съвестта. Нормите на морала съдържат идеята за изискване на дължимото и важното, апел към самоусъвършенстване и креативно преобразуване на действителността.

Обект на отражение	Категория
Оценка на действителността от гледна точка на желателност (или нежелателност) за човека	Добро, зло, справедливост, щастие, смисъл на живота
Начини за подреждане на съвместния живот на хората	Норма, принцип, оценка, идеал
Индивидуални механизми на работа на нравственото самосъзнание	Дълг, съвест, срам, чест и достойнство, морални чувства

Сложността на въвлечането на предмета етика като наука за морала е свързана с това, че нравствените отношения включват широк кръг от обществени отношения. Явленията в обществения живот могат да бъдат класифицирани както следва:

- Морални (положителни спрямо морала).
- Аморални (отрицателни спрямо морала).
- Неморални.

Неморалните явления се определят трудно. Разглеждайки конкретно човешко действие, включително действия, които влияят върху характера на производствените отношения, ние откриваме, че те носят определена морална тежест. Примерно:

Закриването на даден рудник е решение от производствен характер. Реално работодателят спестява средства от поддържането на нерентабилно производство, но в същия момент хиляди работници се оказват без препитание. Подобна дейсторална оценка и предизвиква етични въпроси.

Основни етични концепции:

Хедонизъм – признава наслаждението като основна цел на живота и висше благо.

Евдемонизъм – позиционира стремежа към щастие като критерий за нравственост и основен поведенчески модел.

Абсолютизъмът провъзгласява морала като основа на човешкото битие, а стремежът да бъдеш морален - като смислообразуващ мотив.

Уталитаризъм – разглежда морала само като средство за постигане на щастието, а добродетелите като път към щастието.

Релативизъм – твърди, че оценката на ситуацията може да варира в зависимост от културата, историческия и личен контекст. Правилността на всяко нравствено твърдение подлежи на проверка през практиката.

Формализъм – основен принцип на етическа теория, в която формалнологичният момент превалира над анализа на съдържанието и социалната природа на морала. Характерен е за етиката на Кант, приемаща, че от някакво твърдение (категоричен императив) може да се изведат съдържателни морални принципи и решения, приложими в различен контекст.

Етиката в своето развитие лежи върху нравствените нагласи, които във времето са били многократно преразглеждани, а също така и на постиженията на другите науки, като естествознание, биология и психология. Етическите концепции, натрупани от човечеството, са безкрайно разнообразни.

Говорейки за етическите принципи, не можем да не обърнем внимание на такава философска система като хуманизма. Тъкмо принципите на хуманизма са ръководели човечеството при съставянето на важни международни документи (примерно „Декларацията за правата на човека“³). Идеите на хуманизма са претърпели многократно трактовка в различните исторически периоди и въпреки това могат да бъдат фиксирани някои общи положения. Хуманизмът предполага

³ URL: <https://www.un.org/ru/universal-declaration-human-rights/index.html>

грижа за конкретния човек и стремеж на обществото да създаде условия за удовлетворяването на индивидуалните му потребности и лична самореализация. Би било погрешно обаче хуманизмът да бъде разглеждан като някакво съвършено състояние на обществото. Хората се сблъскват постоянно с проблеми от морален и практически характер и това изисква внимание не само от страна на обществените структури, но и индивидуални усилия.

ОСНОВНИ ПРИНЦИПИ НА ХУМАНИЗМА

- Гарантиране на основните човешки права като условие за истинско индивидуално съществуване.
- Поддържане на видимо слаби, излизаци извън рамките на обичайните за даденото общество, представи за справедливост.
- Създаване на социални и нравствени качества, позволяващи на личността да се самореализира, потребявайки обществените ценности.

Съвременната етика е крайно приложна в различни сфери:

- *Етика на отворените морални проблеми*, призвана да решава проблеми, които до настоящия момент не са намерили своя отговор (примерно въпроса за евтаназията).
- *Приложна етика*, занимаваща се с анализа на обществените заплахи (като екологични кризи, ядрена война и пр.), правейки социална експертиза и предвиждайки възможни последствия.

Приложната етика има следните области:

- *Професионална етика*, използваща предимно деонтологични принципи⁴ и другите елементи на съзнанието за ценност, оформящи нашата действителност.
- *Практична етика*, търсеща съответните процедури за решаване на етични въпроси.

⁴ **Деонтология** – област на етиката, изучаваща проблемите на морала, нравствеността и дълга . проекцията ѝ се проявява в лекарската и юридическа деонтология.

- *Корпоративна етика*, даваща възможност на фирмите да формулират етични постулати за дейността си.

Трябва да отбележим, че съвременната етика в крайна сметка не се свежда до формулировката на приложната, защото съществуват и теоретични проблеми, свързани със свободната воля, обосноваването на морала, различните етични концепции; въпросите, свързани с историята на морала и характеристиките на моралния субект. Очевидно е, че в наше време се наблюдава усилено разширяване на приложната етика в търсене на решения за конкретни приложни проблеми.

Взаимодействието на човека с техниката е област на разглеждане от много време за философията, но сега се наблюдава появата на нови технологии и свързаните с тях проблеми. Всяко постижение на науката носи със себе си както огромни перспективи за обществото, така и рискове. Разглеждането на етичните въпроси, свързани с цифровите технологии, е фактор, намаляващ потенциалните рискове и създаващ платформа за предначертаване на пътищата за развитието на технологията. В момента цифровият потенциал на обществото е свързан само с разпространението на информацията и използването на голяма база от данни. развитието на технологиите, включително и тези на изкуствения интелект (от тук нататък маркирано като **ИИ**), създава все повече поводи за етическото им осмисляне.

ВТОРА ТЕМА.

ДАННИТЕ – ОСНОВА НА ИКОНОМИКАТА

Усложнения от етичен характер се появяват при събирането, анализа и обработката на цифровите данни за гражданите – „големите данни“, социални и персонални. На бизнеса те са нужни, за да може да обучава ИИ, за реклама и онлайн търговия. На държавата, за да може да взема управленчески решения, да взаимодейства с гражданите, за обезпечаване на държавната сигурност.

В тази тема ще се запознаете с основните етични проблеми, свързани с данните и доколко тясно те са свързани с ИИ и другите цифрови технологии.

КАКВО РАЗБИРАМЕ ПОД ДАННИ?

Преди да преминем към въпроса за етиката, сме задължени да направим изравняване на понятийния апарат.

Данните се различават от информацията – самата тя може да бъде представена не само под формата на данни, но и под формата на сведения.

ИНФОРМАЦИЯ – сведения (съобщения, данни) независимо от формата на представянето им.

В международните документи и закони има различни определения за термина „данни“ и „информация“. В рамките на този учебник под „данни“ ще използваме значението, което е дадено в стандарт ISO/IEC, общоприет за сферата на информационните технологии.

ДАННИ – информация (факти, понятия и инструкции) под форма, възприета за общуване, интерпретация или обработка от хората или от автоматични системи⁵.

В научната и популярната литература се предлагат различни **класификации за данни**. Много често едни и същи се съотнасят към различни типове и по тази причина настъпват противоречия при представянето на детайлна класификация. Без претенция за изчерпателност предлагаме най-често използваните в различен контекст типове данни.

ГОЛЕМИ ДАННИ (BIG DATA) – термин, даващ характеристиките на натрупан и анализиран информационен ресурс, чиито обем е значително нараснал и превишава възможностите данните да бъдат съхранявани и анализирани въз основа на по-рано създадените апаратни и програмни средства⁶.

⁵ ISO/IEC 2382:2015(en) Information technology — Vocabulary // ISO.
URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en>

⁶ Доклад на Генералния секретар на ООН “Използване на информационно-комуникационните технологии за инклузивно социално-икономическо развитие”/Организация на Обединените Нации URL: <https://undocs.org/bg/E/CN.16/2014/3>

ГОЛЕМИ ДАННИ ОТ ИНТЕРНЕТ – унифицирани сигнали, постъпващи от датчиците на „умните“ устройства, използвани в производството, селското стопанство, преносимите устройства и т.н.

ПЕРСОНАЛНИ ДАННИ (ПД) – всяка информация, отнасяща се пряко или косвено за определено или определяно като такова физическо лице.

СОЦИАЛНИ ДАННИ – свързани с човека, неговото придвижване, поведение и интереси, отношения с другите хора, места, стоки и дори идеология.

ДЪРЖАВНИ ДАННИ – информация, съдържаща се в информационните ресурси на органите и организациите в държавния сектор (данни на органите на властта, събирани за целите на статистиката, и такива свързани с реализацията на държавния бюджет).

ОТКРИТИ ДАННИ – информация, поместена в интернет под формата на систематизирани данни, организирана във формат, допускащ автоматичната ѝ обработка без предварителна промяна от човек с цел еднократно, свободно и безплатно използване.

Данните се превръщат в критичен ресурс и определят ефективността на дейността на предприятия, държавни органи и икономиката като цяло според Комисията на ООН за науката и техниката с цел развитие още през 2014 година. Това касае особено много създаването на т. нар. „Електронно правителство“ като проява на зрялост и развитие на демокрацията.

Data етиката, или етиката на данните като форма на приложната етика, се появяват сравнително скоро и имат само някакво общо определение. Достатъчно точно работно определение на термина и описание на основа етика предлага британската фреймуорк етика на данните от 2018 година. Тя е ориентирана към внедряването на етични практики в работата с данни в сферата на държавния и публичен сектор: „Етиката на данните сама по себе си представлява формиращ отрасъл на приложната етика, описващ ценностните съждения и подходи приети при събирането, анализа и разпростра-

нението на данни. Те предполагат добро познаване на законодателството, отнасящо се до защитата на данните и многопрофилното законодателство, като същевременно са ангажирани с правилните начини за използване на новите технологии. Те изискват цялостен подход, включващ опита, натрупан в сферата на компютърните технологии, и решаването на етичните задачи и осигуряването на информационна безопасност“⁷

Главен аспект при етика на данните е използването на анализа (data science) по правилния начин. Науката за данните се описва като автоматичен метод, който с помощта на анализ извлича от информационните данни определена информация, като се предвижда навлизането на определени технологии, свързани както с патерните на традиционната аналитика, така и предикативните технологии на базата на машинното обучение.

Data етиката се развива и става все по-актуална и това е видимо от наличието вече на съответни държавни, корпоративни и социални документи. Определянето на границите за етичен достъп към данните е сложен проблем, засягащ голям брой страни – гражданите, държавата, корпорациите, обществените институции и т.н., като това изисква съответно и намирането на комплексно решение. Един особено важен проблем, свързан с данните, е деперсонализацията на биомедицинските данни. От една страна, може да бъде използвано:

ОБЕЗЛИЧАВАНЕ /АНОНИМИЗАЦИЯ – *свързано с действия, в резултат на които се загубва връзката между съвкупността на идентифициращите данни и субектните данни.*

- *Данни, непосредствено идентифициращи конкретно лице (ЕГН, адрес и пр.) се отделят от базата.*
- *Идентификацията на отделната личност става трудно.*

⁷ Data Ethics Framework//Gov.uk. URL:

<https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework>

- *Юридически се опростява използването им, когато липсва персонален достъп.*

- *Усложнява се използването им с научна цел до толкова, доколкото се премахват данните на етап съхранение в дадена система, непозволяващи след това те да бъдат свързани с определен човек в различните бази (примерно рентгеновите снимки с лабораторния протокол и бележките на лекуващия лекар). Същевременно става невъзможно обновяването на данни, ако те не са деноминирани.*

- *Остава рискът за деноминиране с косвената идентичност (примерно допълнителна идентификация по признаци, свързани със социалните медии).*

От друга страна, се наблюдава:

ПСЕВДОНОМИНИРАНЕ – обезличаване с добавяне на връзка между съвкупност от обезличени данни и един или няколко псевдонима:

- *Данните за непосредствена идентификация за конкретно лице се премахват или променят (ЕГН, адрес и др.).*

- *Записите за даден човек биват подведени под определен псевдоним, благодарение на което при необходимост могат да бъдат деноминирани (да се персонифицират).*

- *Благодарение на един псевдоним могат да бъдат обединени данните на един човек в различни бази и могат да бъдат допълвани с нови данни, запазвайки неговата конфиденциалност.*

- *Данните стават удобни за аналитика.*

- *Свързването на персоналните данни с псевдоним усложнява юридически манипулации.*

- *Рискът от деноминиране е по-висок от този при анонимизацията.*

- *Технологично по-сложно и скъпо е отколкото обратимата анонимизация.*

Сведенията, отнасящи се до здравето на човека и медицинските процедури, безусловно подлежат на контрол поради високия риск от етична колизия.

ЕТИЧНИ ПРОБЛЕМИ, СВЪРЗАНИ С ДАННИТЕ

Потребността от приложно регулиране с цел обезпечаване на етичните принципи нараства поради натрупването на достатъчно количество данни, чието използване може да бъде както позитивно, така и негативно. Новите технологии за събиране на данни, съхраняването им, анализът и използването на направените изводи все по-често карат разработчиците да търсят решения на въпросите, свързани с това кое е позволено, доколко е правомерно, рисково, ефективно използването на данни в една или друга ситуация. Какви са преимуществата и недостатъците в използването на данни:

ПРЕИМУЩЕСТВА	СИТУАЦИЯ С ИЗПОЛЗВАНЕ НА ДАННИ	НЕДОСТАТЪЦИ
<ul style="list-style-type: none"> • Развитието на приложните науки; • Големи международни проекти 	Събиране и обработка на голяма база данни	<ul style="list-style-type: none"> • Нарушения в приватността (с голяма база данни се работи както с персонални) • Скрита манипулация на гражданите (покупки, решения)
<ul style="list-style-type: none"> • Персонифицирани сървъри; • Нови продукти и услуги; • Цифрови платформи; 	Акумулиране на данни в компаниите	<ul style="list-style-type: none"> • Нарушаване и инфлация на приватността • Цифров монополизъм • Дискриминация на скрининговата система
<ul style="list-style-type: none"> • Обществена безопасност; • Национална безопасност; • Съкращаване на времето за получаване на административни услуги; 	Акумулиране на данни в държавата включително и такива от видеонаблюдение	<ul style="list-style-type: none"> • Цифров тоталитаризъм; • Големият брат“ • Нови начини за налагане на дискриминация • Нарушаване на приватността;

ПРЕИМУЩЕСТВА	СИТУАЦИЯ С ИЗПОЛЗВАНЕ НА ДАННИ	НЕДОСТАТЪЦИ
<ul style="list-style-type: none"> • Отвореност на данните; • Актуална статистика за извеждане на държавни решения; 		
<ul style="list-style-type: none"> • Бързи услуги • Персонализация на услугите и продуктите • Мащабно развитие на приложните науки; 	Предаване на данни за изкуствен интелект	<ul style="list-style-type: none"> • Различни форми на дискриминация; • Скрити грешки • Нарушаване на приватността;
<ul style="list-style-type: none"> • Общуване и себеизразяване в социалните мрежите; • Свобода на словото • Нови форми на СМИ 	Ползвателите на данни имат открит достъп	<ul style="list-style-type: none"> • Нарушаване на приватността (данните са достъпни за обобщаване), разпространяване на чувствителна информация • Измами с използването на социална инженерия и технологията на дипфейк

Основната етична дилема е около избора да се съблюдават гражданските права, интересите на социалните групи, държавата, от една страна, и създаването на нови продукти, услуги, възможностите за пряко или косвено използване на данни, от друга.

Максимално количество рискове възниква при събирането, анализа и използването на резултатите за хората, като рисковете се оформят в три групи: обработката на данни с помощта на ИИ, видеонаблюдение, проследяване в интернет. Най-големият риск е свързан със загубата на приватност.

ГОЛЕМИ ДАННИ

Развитието на някои технологии и на първо място ИИ, силно зависи от достъпа до определени типове и големи бази данни. Големите данни и технологиите за тяхната обработка

се използват за създаването не само на комерсиални проекти, а и за социални и научноизследователски такива. Като пример може да послужи съвместният проект на Google и Централната водна комисия на Индия с цел създаване на подход за предсказване на наводненията в щата Бихар.⁸

През 2018 година компанията Google получава достъп до персоналните медицински данни на милиони американци, включително резултатите от лабораторните изследване и диагнозите. Нито един от пациентите не е дал информирано съгласие. Официално обявената цел е обучението на ИИ за повишаване качеството на медицинското обслужване. В случая формално нарушение на американското законодателство няма поради факта, че в него няма забрана медицинските власти да предоставят такава информация на трети страни при оказване на медицинска помощ. През 1996 година, когато се приема законът, в него е разрешено такова предаване на данни и тогава data корпорация не е съществувала (The Health Insurance Portability and Accountability Act of 1996)⁹.

АКУМУЛИРАНИ ОТ КОМПАНИИТЕ ДАННИ

Най-опитният и развит участник в пазара на данни са т.нар. data корпорации, които са големи IT-компания, чието съществуване става възможно благодарение на достъпа до голямо количество данни, носещи огромно количество печалби от тяхното събиране, обработка и продажба на ползватели. Голямата петорка технологични гиганти са Microsoft,

⁸ Nevo S. An Inside Look at Flood Forecasting // Google AI Blog.

URL: <https://ai.googleblog.com/2019/09/an-inside-look-at-flood-forecasting.html>;

⁹ Copeland R. Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans // The Wall Street Journal. URL: <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790>; Murphy B. Google's Project Nightingale Secretly Gathers Private Health Data on Millions of Americans, According to a Report. Most Patients Have No Idea It's Happening // Inc.

URL: <https://www.inc.com/bill-murphy-jr/googles-project-nightingale-secretly-gathers-private-health-data-on-millions-of-americans-according-to-a-report-most-patients-have-no-idea-its-happening.html>

Alphabet (Google е дъщерна), Facebook, Amazon и Apple. В тях се акумулира основният финансов, интелектуален и научен ресурс. В Русия най-големите IT-компани са Яндекс, Mail.Ru Group, в Китай — Alibaba Group, Baidu, Tencent, Huawei и др.

Повечето етични проблеми, свързани с данни, възникват заради високата себестойност на данните за гражданите и колосалната заинтересованост на компаниите. За да получат по-големи печалби, компаниите нарушават човешките права при обработката на данните.

Google, Apple, Facebook, Amazon и Microsoft събират и обработват милиони аудиозаписи през автоматизираните гласови помощници и „умните“ колонки, през които могат да се извлекът фрагменти от разговори и други звуци. Записите се правят съвсем автоматично, без предупреждение за ползвателя, като данните се използват за обучение на ИИ да разпознава човешката реч¹⁰.

АКУМУЛИРАНИ ДАННИ ОТ ДЪРЖАВАТА

Държавата е един от най-големите притежатели на данни и вероятно в перспектива ще увеличава потенциала си в тази област. Достъпността на данните и тяхната защитеност повишава доверието на гражданите, стимулира държавните служби в посока ефективност. Същевременно се засилва и контролът върху събирането и съхранението на данните и това е спирачка в посока на развитието.

Държавата е гарант за спазването на правата на човека и заради това при регулирането на оборота на данните тя трябва да осигури спазването на етичните норми заедно с контрола върху националната сигурност. В такива условия възниква дилемата кое е приоритетно — интересите на гражданите

¹⁰ Frier S. Facebook Paid Contractors to Transcribe Users' Audio Chats//Bloomberg. URL:<https://www.bloomberg.com/news/articles/2019-08-13/facebook-paid-hundreds-of-contractors-to-transcribe-users-audio>; Smith D. Microphone Gate:The world's biggest tech companies were caught sending sensitive audio from customers to human contractors. Here's where they stand now//Business Insider. URL:<https://www.businessinsider.com/amazon-apple-google-microsoft-assistants-sent-audio-contractors-2019-8>

(преди всичко приватността) или интересите на обществото и държавната сигурност?

Предаване на данни за изкуствен интелект. За да има развитие технологията на ИИ, е необходимо да се събира огромно количество данни –зависимостта е 100%. Въпросът е как да се използват максимум данни с минимум риск? Със съвременната изчислителна способност на ИИ може да се обработват огромни обеми от данни и да се намира дълбока скрита връзка между тях.

Същевременно колкото и да са деперсонифицирани, тези данни могат да създадат маркер спрямо някой човек с информация за неговите роднини, професия и пр. Така пълнотата на данните, които са критически значими за успешното разработване на интелектуални системи, става потенциално опасна за опазване на приватността и правата на човека. Събирайки и анализирайки данни с помощта на ИИ, технологичните гиганти могат да правят и корелации, които човек не може да осъзнае и разбере и това може да се случи примерно на базата на някакво целенасочено косвено търсене.

Силното ограничаване (регулиране) на достъпа до данни усложнява и затормозява развитието на технологиите на ИИ за машинно обучение. Добре премисленото законодателство и акуратната употреба може да създаде баланс между обема и степента на анонимизация на персоналните данни, ако не създава прекалено много забрани. Такова законодателство е в процес на създаване.

Ползвателски данни, публикувани в откритото пространство

Събирането на данни „по подразбиране“ се случва естествено в моментите, когато човек използва мобилно устройство и интернет. Приема се, че данните, които ползвателят поставя в безплатните социални мрежи или в електронните съобщения, оставят някаква електронна следа. Те би трябвало да принадлежат на самите хора, които ги излъчват, а data компаниите управляват тези данни само по причина на това,

че владеят базата, върху която се случват. Реалността обаче е съвсем друга – такъв тип данни се превръщат в огромен източник на печалби за тези компании.

Стандартният механизъм за получаване на съгласие за обработката на данните става абсолютно неактуален поради факта, че този процес е неконтролируем с течение на времето (примерно 10-20 години) и става очевидно недействителен. Данните се пазят толкова дълго, че в момента на получаване на съгласието става невъзможно човекът да бъде уведомен за това какви данни и как ще бъдат обработвани в бъдеще.

Изтичане на данни – причини и следствия. Това е едно масово явление, наблюдавано по целия свят, и никой не е засрахован срещу него (нито държавата, нито бизнесът). Пряк целенасочен пробив се случва рядко, но причини за това са:

- Грешки в законодателството.
- Недобре обмислена работа на регулационните органи.
- Грешки на самите разработчици (неправилна настройка на сървър по отношение на достъпа, възприет по подразбиране).
- Нелоялност на вътрешни служители, копиращи данни с цел лична изгода.

Още по-критична е ситуацията с данните в социалните мрежи и ежедневната дейност на хората през инфраструктурите на „умния град“. Дори преносими „умни“ устройства носят вреди за собственика си по отношение на предумишлено разкриване на информация.

ТЪРСЕНЕ И НАМИРАНЕ НА ЕТИЧНИ РЕШЕНИЯ ПРИ РАБОТА С ДАННИ

В съвременното сравнително рядко се повдигат въпроси около етиката на данните и етичните въпроси пред бизнеса, разработчиците, законодателите и държавата. Съществуващите закони в сферата на персоналните данни на гражданите в настоящия момент са твърде неефективни, етичният кодекс на данните все още няма за-

дължителен статус и разпространение, за да влияе върху тази област.

Регулиране на основа data етика

Когато говорим за актове и документи, маркиращи етичните принципи при използването на цифровите технологии, е важно да отбележим, че в много сфери (медицина, правосъдие) отдавна съществуват професионални етични кодекси. Те трябва да бъдат вземани под внимание при създаването на нови документи, регулиращи използването на цифрови технологии. Създаването на единен кодекс едва ли е възможно в близко време, защото той би бил или твърде общ и съкратен, за да отговаря на всички приложни области и отрасли, или ще бъде в другата крайност – твърде подробен и противоречив, стараяйки се да обхване всичко свързано с проблематиката.

Публичният дебат относно използването на един или друг принцип или идея, постиженията и грешките от реалната практика позволяват да се направи детайлно изследване в областта на етиката на данните. С всяка изминаваща година се разработват все по-нови етични кодекси, фреймуоркове, гайдлайни и подобни препоръки за работата с данни. Съответно те могат да бъдат класифицирани така:

- Според **сферата**, в която се прилагат: здравеопазване¹¹, социална защита, държавна администрация¹², юриспруденция¹³, борба с престъпността, финанси¹⁴, промишленост и Т.Н.

¹¹ Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health-related data // Council of Europe.
URL:[https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168093b26e;Recommendation of the Council on OECD Legal Instruments Health Data Governance//OECD](https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168093b26e;Recommendation%20of%20the%20Council%20on%20OECD%20Legal%20Instruments%20Health%20Data%20Governance//OECD). URL:<https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf>

¹² Data Ethics Framework // Gov.uk.
URL:<https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework>

¹³ European ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment // Council of Europe. URL: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>;

- Според **технологията**: разпознаване на лица, системи за ИИ, роботехника, обработка на големи данни и др.
- Според **локализацията**: документи на ООН; документи, касаещи управлението на конкретен град или конкретна компания (примерно правилата за етично използване на технологии за лицево разпознаване от полицията на Лондон или принципите на ИИ на Microsoft)¹⁵

Законодателство, защитаващо персоналните данни, е прието или се подготвя за прилагане в редица страни, като все повече се залага на по-твърдо и стриктно прилагане на регулаторни и праворегулиращи модели. На основата на различните базови ценности могат да се дефинират три модела за регулиране на персоналните данни - *европейски, американски и китайски*.

Правните практики по отношение на една и съща технология могат да се различават кардинално и това показва нагледно разногласията, например по повод използването на данните от лицево разпознаване.

В Китай разпознаването на лица е много популярно, включително се прилага в много скрининги. Учебните заведения го използват по отношение на посещаемостта, заплащането на храната и контрола за въвличеността на учениците по време на занятия. Въпреки жалбите на учениците и студентите за това, че системата допуска редица грешки

CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment // Council of Europe.

URL: <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment#:~:text=Council%20of%20Europe%20adopts%20first,artificial%20intelligence%20in%20judicial%20systems&text=The%20Charter%20provides%20a%20framework,AI%20in%20national%20judicial%20processes>

%20judicial%20systems&text=The%20Charter%20provides%20a%20framework,AI%20in%20national%20judicial%20processes

¹⁴ General principles for the use of Artificial Intelligence in the financial sector // De Nederlandsche Bank.

URL: https://www.dnb.nl/en/binaries/General%20principles%20for%20the%20use%20of%20Artificial%20Intelligence%20in%20the%20financial%20sector2_tcm47-385055.pdf

¹⁵ Microsoft AI principles // Microsoft. URL: <https://www.microsoft.com/en-us/ai/our-approach-to-ai>

или „заплашва“, не съществува и най-малък намек властите да се откажат от нея.

В Швеция през август 2019 едно средно училище е глобено с 200 000 шведски крони (20 000 евро) за това, че използва камери за разпознаване на лица при контрола си върху посещаемостта. Профилният контролно-надзорен орган (Data inspektionen) приема, че съгласието на учениците да се събират техни биометрични данни не е достатъчно основание да се прилага лицево разпознаване, доколкото учениците са в зависимо положение спрямо училището и съгласието им не е в пълна степен доброволно¹⁶.

Европейски модел. В основата му лежат концепции за регулиране на персоналните данни без да се допуска прекомерен контрол над личния живот на човека нито от страна на държавните институции, нито от страна на фирмите. Това е заложено като основен принцип на всички европейски държави¹⁷. Ако внедряването на нови технологии или въвеждането на нов ред за използването на данни по някакъв начин нарушават човешките права, се наблюдава силна обществена резистентност и съпротива. На правителствата им се налага да се справят с подобна резистентност (прилагане на механизми за убеждаване, просвета, етични похвати и т.н) в случай, че използването на такива технологии е с висока значимост в полза на обществото.

Същността на европейския подход е представена най-ярко в регламента за защита на персоналните данни (General Data Protection Regulation, General Data Protection Regulation, GDPR), приет в ЕС и задействан през 2018 година. Сега GDPR е подложен на широко обсъждане не само в Европа, а в целия свят – в него има твърде високи глоби, ако регламентът му не се спазва (в зависимост кой параграф на регламента е бил нарушен, глобите достигат до двадесет милиона евро

¹⁶ Sweden has gotten its first GDPR fine // Aigine. URL: <https://aigine.se/en/sweden-has-gotten-its-first-gdpr-fine/gdpr/>

¹⁷ „Европейска конвенция за защита правата на човека и основните свободи“ и „Конвенция на Съвета на Европа за защита на физическите лица при автоматизирана обработка на персоналните данни „

или 4% от годишния световен стокооборот на компанията). Освен това има и широка екстериториалност на прилагането - той трябва да бъде прилаган не само от страните членки на ЕС, но и от всички оператори от целия свят, които обработват персонални данни на граждани на ЕС. Регламентът делегира широк спектър от права на субектите носители на персонални данни в сравнение с други страни и бизнеси, обработващи такъв тип информация (примерно всеки гражданин може да изиска копие на данните за себе си, които се пазят в дадена фирма, и да претендира за пълното им заличаване. Нещо повече, GDPR изисква всеки оператор на персонални данни да назначава специално длъжностно лице *data protection officer* с пълномощия да следи за изпълнението на GDPR в съответната компания, делегираща му широки пълномощия.

С голяма международна значимост при налагането на правни норми по отношение на личните данни е Конвенцията на Съвета на Европа за защита на частните лица по отношение на автоматизираната обработка на данни от личен характер от 1981 година (ETS 108 влязла в сила на 1 октомври 1985 г.). През 2020 година започва преработка и допълване на документа, предвиждаща:

- По-строги изисквания към спазването на принципите на пропорционалност и минимизация, а също и на законността на обработката.
- Разширяване на списъка от данни, обработката на които изисква особени гаранции (биометрични, генетични данни и данни за етнически произход).
- Нови положения, свързани с правата на лицата в случаите, когато решението се взема на база алгоритмизирана обработка на данните с основна идея решението да не бъде приемано без да се вземе под внимание отношението на самия субект.
- Изискване да се внедри вграден алгоритъм за конфиденциалност - *privacy-by-design*.

Развивайки и допълвайки стандартите на Конвенция 108 спрямо конкретни технологии и жизнени сфери, Съветът на Европа разработва „Ръководни принципи за защита на лицата във връзка с обработката на персонални данни в сферата на големите данни“¹⁸. В документа се отразява необходимостта от етично използване на данните, използването на алгоритъм за конфиденциалност, обезпечаване на анонимизация и систематично оценяване на рисковете за реидентификация на персоналните данни, дефиниране на ролята на човека при вземането на решения на база големи данни и пр.

Ако трябва да обобщим, темата за защитата на персоналните данни е подробно разработена в рамките на общоевропейското право, но и допълнително прецизирана в националното право на всяка една от страните членки на ЕС.

Американски модел. В САЩ се отдава голямо значение на свободата на информацията, предприемачеството и безопасността на обществото. С тази аргументация правителството позволява по-голям достъп до персоналните данни отколкото в Европа. Неслучайно именно в САЩ се появяват информационни гиганти, чийто бизнес се гради върху обработката и анализа на данни.

В САЩ журналистическото лоби и неправителствените организации са особено силни, но въпреки това има голяма свобода в използването на данни. На федерално има редица отраслови актове за конфиденциалността и безопасността на данните, в това число актове по отношение на финансовите служби и здравеопазването и пр. Компаниите имат значително голяма свобода при използването на данни, освен най-чувствителните. В различните щати има стотици разнообразни актове за защита на персоналните данни. Един от най-обсъжданите нови закони е този на Калифорния „За защита на личния живот на потребителите“ (California Consumer Privacy Act, CCPA), който влиза в сила през 2020 година. Той

¹⁸ Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data.T-PD(2017)01.23 January2017//Directorate Generalof Human Rights and Rule of Law. URL:<https://rm.coe.int/16806ebe7a>

се прилага в различни сектори на икономиката и предвижда значителен обем права върху субективните данни (включително правото за унищожаването на данни), като същевременно делегира редица задължения на компаниите при работа с данни, определени ограничения при обработката, използването и разкриването на персонални данни. Както и Европейският GDPR, CCPA се използва в широк кръг компании, регистрирани извън Калифорния.

Китайски модел. Отличителна черта е доминацията на интересите на Китайската комунистическа партия, от една страна, и държавата над интересите на гражданите, от друга. Условието на новата цифрова икономика по отношение на данните се диктуват от държавата, действаща в тандем с големите data компании. Общественото обсъждане тук има значително по-малка тежест. В държавната организационна система на Китай се приема за съвсем уместно и логично да има строг контрол по отношение на персоналните данни, тъй като бенефициент в случая е държавата, а не отделният гражданин.

Нормативно правното регулиране от страна на държавата е само едно от възможните средства за решаването на етичния казус. Другото са обществените организации и професионалните организации със задача да създадат етичен кодекс.

Консултинговата компания Accenture препоръчва на data аналитиците и всички останали, които работят с данни, да спазват общи принципи на data етика, включващи:

- *Уважението към човека, стоящ зад данните, е висш приоритет.*
- *Осигуряване на гарантирана приватност и безопасност на такова ниво, на което би желал да ги види субектът на такива данни.*
- *Всички продукти и методи на изследванията да бъдат съобразени със съответните вътрешни норми на компаниите и да са достъпни за проверка отвън.*

Accenture предлага общо 10 етични принципа спрямо големите данни, насочени към спазване на гражданските права и обезпечаване на безопасността на данните¹⁹.

Доколкото темата за големите данните е свързана с изкуствения интелект, голяма част от препоръките за етичното им използване са систематизирани в етичен кодекс, регулиращ ИИ и роботехниката, наречен „Асиломарски принципи на ИИ“ (Конференция в Асиломар за ИИ)²⁰. Неговата инициация е направена от Инженерния институт по електротехника и електроника (Institute of Electrical and Electronics Engineers, IEEE).

Ако трябва да обобщим, етичните аспекти на регулирането на пазара на данни все още не са достатъчно утвърдени – няма нито един документ на ниво международна авторитетна организация, който да е приет от всички страни. В перспектива ангажираността да се спазват някакви етични аспекти в работата с данни изглежда ниска и латентно генерира скрита дискриминация – автоматизираните системи за вземане на решения в здравеопазването, образованието и социалното подпомагане скоро неизбежно ще навлязат в живота ни. Изтичането на данните и грешките при вземането на решения поради некачествена инфраструктура води до ниско качество на данните.

¹⁹ The Ethics of Data Sharing: A guide to best practices and governance // Accenture. URL: https://www.accenture.com/_acnmedia/PDF-35/Accenture-The-Ethics-of-Data-Sharing.pdf

²⁰ Asilomar AI principles // Future of Life Institute. URL: <https://futureoflife.org/ai-principles/?cn-reloaded=1>;

ТРЕТА ТЕМА. ЕТИКА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

От тази тема ще разберете защо етиката на изкуствения интелект се разглежда в два аспекта: етични принципи, лежащи в основата на използването на решения от ИИ, и етично поведение на ИИ спрямо хората. Вторият аспект принципно дава ясна диференциация на етиката на ИИ от другите цифрови технологии.

Внедряването на ИИ системите в ежедневието е свързано с много етични проблеми, които в следващите години буквално ще ни засипят. Първи пример за такъв проблем е леталният изход от автокатастрофа със самоуправяващ се автомобил Tesla през 2016 г. и Uber през 2018 г.; първият протест на разработчици на ИИ в Google, отказали да вземат участие в разработването на проект на Военното министерство на САЩ; случаите на манипулиране на достъпа до информация, сексизъм и расизъм в алгоритмите за разпознаване на лица и таргетирането на реклами с използването на ИИ. Големите етични проблеми се появяват при използването на ИИ от държавните служби за контрол над гражданите. Възможните негативни социални последствия при използването на алгоритми в работата на държавата получиха широк отзвук в СМИ държавните органи в някои държави.²¹

Трябва да отбележим, че етиката на ИИ силно се отличава от data етиката. Спрямо другите технологии сега се обсъждат общи въпроси за промяна на професионалната етика, етиката в използването, етичния отговор на социалните предизвикателства (безработица) и т.н. Специално етиката на ИИ има още много други измерения и те са свързани с поведението на системите в ситуациите, касаещи хората.

²¹ Simonite, T. AI experts want to end «black box» algorithms in government // Wired Business.

URL: <https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government>

Принципно важно е, че системите на ИИ могат:

- ***Самостоятелно да вземат решения, касаещи човека.***
- ***Да анализират данни в такъв обем и скорост, с които човек не може - следователно той не може да провери тяхната достоверност!***

От тук проблемът е да се определи доколко решенията, взети от интелектуални автономни системи (ИАС), не влизат в разлив с етичните норми.

Поради това можем да говорим за два корено различни аспекта на ИИ.

Първият аспект предполага, че първичните програми за ИИ са написани от човек, но в развитие системите на ИИ започват да имат самостоятелност и съответно като всеки разум ще искат да се самоусъвършенстват, реструктурират и да увеличат своите параметри. Създавайки системи, вземащи критично важни решения за човека, съответно ние като хора искаме да получим гаранции, че те се ръководят от етични съображения и това изисква привнасянето им при създаването на ИИ. Усложненията идват от това, че моралният избор не се прави въз основа на ясно маркирани норми и закони, а се формира от правила, принципи и субективни мнения, базирани на оценката кое е „добро“ и кое е „лошо“. Такава база е изключително трудно да бъде заложена в ИИ.

Вторият аспект на етиката на ИИ предполага анализ и предотвратяване на етични колизии, възникващи в процеса на използването на ИИ: нарушение на приватността, дискриминация, социално разслоение, проблеми с трудовата заетост и т.н. Съвсем регулярно се поставя и въпросът за професионалната етика на разработчиците на системи за ИИ. Несъмнено отказът от разработването на системи на ИИ е невъзможен и надпреварата в тяхното създаване вече върви с огромни темпове.

ФОРМУЛИРОВКА ЗА ИЗКУСТВЕН ИНТЕЛЕКТ

Проблем в сферата на етиката на ИИ е липсата на единна понятийна база. Доколкото няма формално или конструктивно описание на основни етични постулати в областта на техническите изследвания, моралните аспекти се свеждат до битовото, интуитивно разбиране. Самите специалисти в областта на философията не винаги си представят достатъчно ясно дадено постижение в областта на технологиите и ИИ. Очевиден е разрывът между разработчиците, изследователите и философите.

Понятието „изкуствен интелект“ семантично предполага комплекс от технологии и наука и в този учебник ще бъдат използвани и двете значения.

„Комплексът от технологични решения, позволяващи имитацията на когнитивните функции на човека, включващи самообучение, вземане на решения без предварителен алгоритъм, получаващи като резултат конкретни резултати, съпоставими с резултатите от интелектуалната дейност на човека, приемаме като формулировка за изкуствен интелект“

Комплексът от технологични решения, включващи в себе си информационно-комуникативна инфраструктура; програмно обезпечение, включващо машинно обучение, обработка на данни и търсене на решения, реализирани с помощта на съответните процесори и сървъри.

Изкуственият интелект е наука и технология, включваща набор от средства, позволяващ на компютъра на база натрупани знания да генерира отговори на въпроси и да прави експертни изводи, без те да са заложени в него предварително от разработчика. Областта, дефинирана като „изкуствен интелект“, е междудисциплинарна и влиза в комплекса на компютърните науки и създадените от нея технологии, определяни към информационните технологии.

За рождена дата на ИИ като научна област се приема 1956 г. в САЩ, където на двумесечен семинар в Дартмут,

организиран от Джон Макарти, Марвин Мински, Клод Шенон и Натаниел Рочестър, са поканени учени от висок ранг, изучаващи теориите за управление, автоматика, невронни мрежи, теории на игрите и изследванията на интелекта (Артур Самуе, Алена Нюел, Херберт Сеймон, Тренчард Мур, Оливер Селфридж и др.).²²

Прието е да се диференцират две групи проблеми, подлежащи на изследване на ИИ като наука:

- **Силен (интегрален, универсален) хипотетичен ИИ**, способен не само да решава интелектуални задачи, а и самостоятелно да си дефинира цели, сравними с човешкия интелект или превъзхождащи го. За проблемите на силния ИИ е необходимо да има високо ниво на познание относно човешкия мозък. Невробиологията е натрупала огромно количество емпирични знания за анатомията и физиологията на мозъка и неговите молекулярни и генетични механизми. Общият принцип на преработка на информация в мозъка не е до край изяснен – единствено категорично ясно е, че се случва по съвсем различен начин от този на компютъра. От там и създаването на силен ИИ е твърде неопределено. Още през 1975 г. някои учени прогнозираят създаването ИИ, сравним с човешкия около 2000 година. Впоследствие времето на евентуален научен пробив е отсрочено за 2045 г.

- **Слаб (приложен) ИИ**, изграден с методите и програмните системи, решаващи отделни интелектуални задачи. Тук успехите са много по-значителни.

Определението за интелектуална задача я дефинира като такава, за която човекът няма алгоритъм. Ако човек изпълнява дадена задача по алгоритъм, винаги се получава един и същ резултат. Когато се решава интелектуална задача, хората използват своите знания, умението да разсъждават и съобразителността си (всичките са различни при различните индивиди). Главният успех на ИИ през 60-те години се състои във

²² McCarthy J., Minsky M.L., Rochester N. Et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence// Stanford.

URL:<http://www.formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

формализацията на тези интелектуални възможности - разработването на методи за извеждането на знания, моделиране на съжденията, евристично търсене и т.н.

Паралелно с това се развиват и други направления на ИИ на принципите на машинното учене и технологията на невронните мрежи. С разширяването на областите на използване на тези технологии нарастват съответно и очакванията към тях, въпреки че невротехнологиите имат редица ограничения – трудности при оценяването на качеството на обучаващата се извадка, водещо до трудности в тълкуването на получените резултати.

Трудности възникват при създаването на безпилотни системи (особено автомобилни) и това не допуска масовото им производство. При тях особено проблемна е оценката на ситуацията (отделянето на фигурата от фона, движещи се от неподвижни предмети, източниците на опасност). Решаването на тези проблеми може да помогне на машинното обучение на данните и процедурите. Проблемът се състои в това, че има ситуации, при които обучаваща извадка изобщо не се прави. Видовете интелектуални системи (ИС) биват:

Видове ИС	Описание и използване на ИС
Управляващи интелектуални системи	Управление на производства, проектиране, бордови интелектуални системи в авиацията
Динамични роботехнически системи – роботи и безпилотници	Робот – автономна движеща се система с дистанционно управление или с вградена програма. Съществуват роботи с адаптивно поведение, способни да преодоляват препятствия, да дават оценка за ситуация и да се ориентират в местност.
Полиагентни системи	„Колективни“ роботи“ с обща цел (изследващи дадена територия). Управлението им е „външно“ и „вътрешно“ с помощта на робот координатор, който може да е

Видове ИС	Описание и използване на ИС
	част от „колектива“. Има вариант, при който координатор не е необходим, а се осъществява ръководство по хоризонтала (между всички).
Системи за поддръжка при вземане на решения	Едно от направленията са т.нар. „когнитивни карти“. Такива схеми описват някакви ситуации и връзките между тях, а също и до колко един фактор влияе на друг. С тяхна помощ може да се направи разчет, примерно как повишаването на данъците ще повлияе върху бюджета и цените. Такива инструменти позволяват проследяването на последствията от дадено решение, но е важно да се отбележи, че самата система не взема решения, а само дава препоръки – отговорността е изцяло на човека. Такива системи масово се разработват в различни области.
Когнитивни изследвания и когнитивни модели	Опити да се формализират познавателните процеси на човека. Такива изследвания имат своето продължение в „силните“ системи ИИ.
Инженерни знания и онтология	Експертни системи, ориентирани към тиражиране на опита на специалистите в области, в които качествените решения зависят от нивото на експертизата и от опита на специалиста (медици, юристи, икономисти). Много компании създават Експертни системи за вътрешно ползване в ключови за тях технологии.
Моделиране на съждения	Тук се използва логиката, но да бъде тя формализирана, е трудна

Видове ИС	Описание и използване на ИС
	задача и най-вече търсенето на здрав смисъл. В разсъжденията си човек използва разпознаването, опита, спомена за нужния прецедент. Моделирането на съжденията включва: моделиране на съжденията на основа прецедентна аргументация или ограниченията; моделиране на съждения, свързани с неопределеност; генериране и проверка на хипотези и др.
Обработка на естествения език	Машинен превод на текстове, анализ на текстове с търсене на нужната информация, класификация по съдържание, автоматично рефериране, разпознаване, превод и генериране на изказ. Повечето от тези задачи стават изпълними с помощта на машинното обучение.

Освен това елементи на ИИ се използват в съвременни цифрови технологии, като киберфизическите системи („Умна къща“), допълване на реалността и пр.

ЕТИЧНО ИЗПОЛЗВАНЕ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Внедряването на ИИ и други цифрови технологии е възпрепятствано от ниското ниво на гражданско доверие към алгоритмите и новите технологии по принцип, а същевременно липсват разбираеми етически рамки в прилагането на ИИ. В настоящата тема ще бъдат дадени общите характеристики и перспективните подходи за решаване на етическите казуси, свързани с използването на ИИ:

- Отговорност за етично или неетично поведение на ИИ при вземането на погрешни решения, нанасяне на щета поради грешки и т.н.
- Предубеденост на алгоритмите (bias).

- Обезпечаване и регулиране на ИИ (обяснителна компонента).
- Проблемът за приватността при използването на ИИ.
- Надеждност на технологиите в ИИ.

ОТГОВОРНОСТ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Проблемът за отговорността за действията на изкуствения интелект за първи път е поставен при прилагането му в автомобили без водач и завършва с хипотетичните системи за вземане на решения от планетарен мащаб – тема, която е най-дискутирана в момента. Проблемите са формулирани в областта на доверието, където на хората им се налага да залагат на системите ИИ, като се имат предвид автомобилният транспорт, фармацевтиката, медицината, образованието и т.н. Институтцията на отговорността при различните типове роботи има различни нюанси в зависимост от обществената сигурност, контролируемост и способността за обучение. Друг аспект - в много от случаите е трудно да се установят обстоятелствата при нанасянето на вреда и в една и съща ситуация се получават различни решения от гледна точка на конкретна юриспруденция. При обсъждането на различните принципни подходи за установяване на отговорност за действията на ИИ вариантите са:

- Пълно освобождаване на ИИ от каквато и да е отговорност за действията им (аналогично с действията на непреодолима сила).
- Частично освобождаване от отговорност (освобождаване на конкретни лица от каквато и да е отговорност и същевременно изплащане на компенсация на пострадалите за понесени вреди от различни източници).
- Отговорност за вина, настъпила само в зависимост от вината на конкретния субект (производител, разработчик, отговарящи за обучение на ИИ лица, собственик, ползвател и пр.).

- Отговорност без вина на определени лица (производители) по общите правила, възприемани като отговорни за действията на ИИ.

- Лична отговорност на роботите при условия, че са натоварени с правосубектност - права и задължения при наличие на статус на електронна личност.

Поради факта на дискуссионност и липсата на общо решение на учени и разработчици можем да представим различните гледни точки:

1. В етиката има понятие „споделена отговорност“ – отговорността е разделена между разработчика и собственика или ползвател и самата система ИИ, ако е зададена възможност за допълнителното ѝ настройване.

2. В повечето случаи с отговорността е натоварен „програмистът“, въпреки че той е само оператор, залагащ правилата на поведение, делегирани от експерти и специалисти в дадена област (в етиката съответно философ). Морално правилно е отговорността да се носи от експерта, предоставящ етичните правила за действието на системата, имплантирани в ИИ.

3. Третият вариант е търсенето на отговорността освен от философа и разработчика, и от законодателя, т.к. отговорността се делегира в правов режим.

4. Друга гледна точка говори за това, че както и при всяка техническа катастрофа се търси кой е виновен (примерно при авиокатастрофа се назначават експерти да решат чия е вината - на производителя, оператора, авиокомпанията или пилота). Затова от голямо значение е валидизирането на система за отговорност по отношение на ИИ.

Ако трябва да се обобщи, то трябва да кажем, че самата концепция за изкуствен интелект е нееднозначна и от там на неопределеност е подложена и формулировката за отговорността. Като универсално се оформя само правилото: *Отговорен е човекът!* Тук трябва да се избегнат двете крайности в случаите: когато отговорността не се носи от никого и тогава, когато цялата отговорност се стоварва върху изкуствения интелект.

СТЕПЕН НА ПРОЗРАЧНОСТ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Системите ИИ са способни да се самообучават, самоусъвършенстват и развиват. Въпросът как една система на ИИ взема решение е от особена важност, т.к. всичко това е резултат от сложен алгоритъм. Действията на самия ИИ трябва да имат видимост за широк кръг от заинтересовани страни поради следните причини:

- Той е важен за ползвателя, доколкото формира доверие към системите, предоставяйки достъпно разбираем начин за това какво и как се извършва в него.

- Сертифицирането и валидизацията на подобна прозрачност на интелектуалните автоматизирани системи (ИАС), за да може да се прави проверка с действащото законодателство за съответствието на изпълняваните от тях процеси.

- ИАС трябва да са прозрачни при разследване на аварии, за да може да бъдат проследени вътрешните процеси, довели до инцидента. Такава достъпност е необходима и на адвокати и експерти.

- Революционните технологии на безпилотните автомобили трябва да бъдат достъпни и прозрачни за широката публика, за да се повиши доверието на обществото към тях.

Примерно: Канадската емиграционна служба е разработила през 2014 система за автоматизация на дейността си. Алгоритмичните системи се прилагат на всички етапи при емиграция в Канада. При това правителството по никакъв начин не дава достъп до това какви критерии се използват при оценката на емигрантите и бежанците, какъв тип данни се въвеждат в автоматизираната система, кой има достъп до информацията и как тя ще циркулира в другите ведомства, до колко правителството приема като допустими грешките в системата²³.

²³ Kenyon M. Bots at the Gate: a human rights analysis of automated decision-making in Canada's immigration and refugee system // The Citizen Lab. URL:

Алгоритмите на работа на невронните мрежи са изключително сложни за интерпретация и следователно резултатите от тяхната работа могат да бъдат подложени на съмнение и отхвърлени от човека. Отсъствието на разбиране за това как изкуственият интелект постига своите резултати, е една от причините за ниското доверие към съвременните технологии и това рефлектира върху неговото развитие.

Трябва на всяка цена да отбележим, че не може да има и пълна прозрачност за изкуствения интелект – степента на прозрачност за едни и същи системи зависи от тяхната функционалност, целева аудитория и т.н.

Както всяка интелектуална собственост, уникалността на всеки алгоритъм, създаден от даден разработчик, трябва да бъде запазен в тайна в дадена степен (освен договорените от разработчика). Обяснителната компонента трябва да съществува в такъв вид, в който да показва хода на работа на системата, без да показва всички „механизми“ на собственото си функциониране. Експертните системи трябва да показват всички вериги на съжденията си на ползвателя, а анализаторите на интелектуални данни трябва да представят своите хипотези в разбираем за човека вид. При ИИ на основата на невронните мрежи до този момент обяснителната компонента не може да бъде реализирана, но не е далеч моментът, когато и това ще бъде възможно.

ПРЕДУБЕДЕНOST СПРЯМО АЛГОРИТМИТЕ

Това е един от най-критичните проблеми пред системите за ИИ – незабележими на пръв поглед диспозиции и предположения могат да останат скрити в базата данни при въвеждането за обработка и на тяхна основа системата наслоява някакви резултати, които влияят върху нейната обективност и правят взетите решения предубедени. Такива „слепи петна“ се репликират непрекъснато на всички нива на обработката и

результатите имат кумулативен ефект. Справедливостта на алгоритмите е един от основните проблеми в създаването на етичен ИИ.

Самият ИИ обаче може да помогне при преодоляването на предубедените решения²⁴. Според изследванията алгоритмите могат да намалят расовото и етническо неравенство в системите на правосъдието²⁵.

Аналогично автоматизираните системи на финансов андерейтинг могат да бъдат полезни за кандидати с недооценена кредитна история. Положителна черта на ИИ, отличаваща го от човека е, че решенията, които взема, са достъпни за анализ и детайлна проверка. Етичните нарушения се получават при некоректно събиране на данни, непълна анонимизация или недостатъчна валидизация на входните данни, използвани при обучението на системите на ИИ. Неволна причина може да бъде и селектираната информация, водеща до дискриминация.

Изследването на кредитната история при вземането на решение за наемане на работа е възможно да нанесе вреда на социално незащитени граждани, въпреки наличието на връзки между качеството на кредитната история и поведението като работник няма доказана релация²⁶. В САЩ има задействана програма за прогнозиране на престъпления PredPole, обучена с етнически изкривена извадка, при което полицията бива изпращана много по-често към адреси на етнически малцинства²⁷. Обучената на полуистински исто-

²⁴ Tackling bias in artificial intelligence (and in humans)//McKinsey. URL:<https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>

²⁵ Kleinberg J.,Lakkaraju H., Leskovec J.et al. Human decisions and machine predictions// The Quarterly Journal of Economics. 2018. Vol. 133, no 1. P. 237–293.

²⁶ Silberg J., Manyika J. Tackling bias in artificial intelligence (and in humans) // McKinsey Global Institute. URL: <https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>

²⁷ Lum K., Isaacn W. To predict and serve?// Significance. 2016. Vol. 13, no 5. P. 14–19. URL: <https://doi.org/10.1111/j.1740-9713.2016.00960.x>

рии на заболявания програма IBM Watson понякога дава смъртоносни препоръки при лечението на рак.²⁸

Създаването на справедлив ИИ е продукт на дълги дискусии, изследвания и стандартизация. На първо място е откриването и намаляването на човешките предубеждения.

Специалистите от лондонската компания Deep Mind предлагат като защита от въздействието на човешките предубеждения да се използва методът „хипотетична справедливост“ (counterfactual fairness)²⁹. За да се формира справедливо и непреднамерено съждение за гражданите, ИИ формира хипотетична ситуация, в която въпросните граждани притежават противоположни признаци – жените се представят като мъже, бедните като богати, афроамериканците като бели и т.н. По този начин реалният статус не влияе върху дейността на гражданите и е освободен от предубеждения и съответно е справедлив.

Другата възможност е усъвършенстването на самите системи ИИ, започвайки с начините за използването на данни и завършвайки процесите на разработването им. От значение е и внедряването и използването, за да може да се укрепят индивидуалните и обществени нагласи или се предотвратят предубежденията и свързаните с тях проблеми. Междудисциплинарното сътрудничество е призвано да осигури разработки, технически иновации и методи на работа, етически стандарти. Практиката показва, че това е възможно.

Амазон спира използването на системи за подбор на персонал след като намира грешки в алгоритъма, свързани с джендърни предразсъдъци. Алгоритъма разпознава шаблонни думи в резюмета, а не професионални навици. Изходните данни, служещи за обучение на системата, се оказват резюмета основно на бели мъже. Алгоритъмът изключва ре-

²⁸ Ross C., Swetlitz I. IBM's Watson supercomputer recommended «unsafe and incorrect» cancer treatments, internal documents show//STAT+.URL: <https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf>

²⁹ Chiappa S. Path-Specific Counterfactual Fairness // [Silvia Chiappa.] URL: <https://csilviavr.github.io/assets/publications/silvia19path.pdf>

зюмета, съдържащи използвани от жени думи. В резултат се появяват диспозиции по отношение на жените при кандидатстването им за работа.³⁰

Диспозициите е възможно да бъдат не само естествено акумулирани от специфика на данните, а и изкуствено създадени съобразно интересите на заинтересовани лица. Такава диспозиция може да определи дребни изменения в маршрута на ползвателя, примерно в навигационна система, за да може той да „посети“, а не да подмине дадена дестинация (магазин, билборд и пр.). Алгоритъмът може да предизвика и статистическа корекция, която е социално неприемлива или незаконна – моделите на ипотечните кредити откриват, че възрастните хора по-често не спазват графика на вноските и на тази основа ги изтласква назад в класацията за получаване на кредити.

В САЩ алгоритмична одитна програма за прогнозиране на обема на медицинската помощ открива диспозиции по отношение на афроамериканци. Независимо от единната методика за всички пациенти, алгоритъмът приема цветнокожите пациенти като по-малко нуждаещи се от медицински грижи отколкото белите, въпреки очевидната необходимост на първите. В самия код не е заложен приоритет за белокожите пациенти и алгоритъмът работи нормално. Грешката е на разработчика, заложил равни разходи за медицинска помощ поради еднаквата необходимост от нея, но на база разходи в миналото, без да отчита факта, че тази необходимост зависи от доходите и социалния статус. Така се закрепва дискриминация, заложена далеч преди създаването на алгоритъма, т.к. в миналото потърпевшите са ползвали по-малко медицински услуги поради ниските си доходи³¹.

³⁰ Vincent J. Amazon Reportedly Scraps Internal AI Recruiting Tool That Was Biased against Women // The Verge. URL: <https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report>
Hamilton I. A. Why It's Totally Unsurprising That Amazon's Recruitment AI Was Biased against Women // Business Insider. URL: <https://www.businessinsider.in/why-its-totally-unsurprising-that-amazons-recruitment-ai-was-biased-against-women/articleshow/66192889.cms>

³¹ Strickland E. Racial Bias Found in Algorithms That Determine Health Care for Millions of Patients // IEEE Spectrum. URL: <https://spectrum.ieee.org/the-human->

Какъв е начинът да се преценява справедливостта, за да се минимизират предразсъдъците?

Проблемът идва от това, че самото понятие справедливост изисква конкретно дефиниране и изследване поради факта, че има поне 20 определения за нея, но нито едно не е достатъчно изчерпателно. Обикновено се фиксира индивидуалната справедливост, без да се взема под внимание груповата такава – необходима е различна система от показатели и стандарти в зависимост от обстоятелствата и вариантите на използване.

Важна съставляваща на справедливостта на системите ИИ е прякото участие на човека. Стратегическите показатели на справедливостта безусловно са адекватни, но не отчитат нюансите на социалните условия, в които се развиват системите на ИИ и свързаното с това събиране на данни.

Разбира се, това поражда редица въпроси:

- Къде и под каква форма са необходими човешките съждения при разработването и експлоатацията на ИИ?
- Кой решава кога системите на ИИ вече са минимизирали диспозициите и са готови за безопасна експлоатация?
- В кои ситуации е напълно възможно автоматизирано вземане на решения?

В момента нито един алгоритъм на оптимизация не може да взема решения самостоятелно и не трябва да се поверява на машините – необходими са човешки съждения, базирани на междудисциплинарен анализ, свързващ познанията на хуманитарните и социалните науки, правото и етиката. Институтът McKinsey дава високо диференцирани препоръки в работата с ИИ за намаляване на диспозициите, адресирани към специалистите и ръководителите на системи за ИИ.

Диспозициите при вземането на решения от хора и машини не само водят до пагубни последствия за хората, подлагайки ги на дискриминация, но и носят загуби за гражданите и държавата, като необосновано ограничават възможностите

на отделни субекти да участват и да имат своя принос в икономиката и обществото. Минимизирането на диспозициите в системите на ИИ е критично по отношение на доверието към тях и допринася те да могат да реализират своя потенциал за държавата и икономиката, повишавайки ръста на производителността и решавайки редица социални проблеми.

НАДЕЖДНОСТ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Проблемите за безопасността и надеждността на ИИ изглеждат откъснати и далечни по отношение на етиката, но въпреки това имат непосредствено отношение към нея и трябва да бъдат разглеждани в няколко аспекта.

Първият е определено технически, засягащ надеждността и безопасността на програмно-техническите системи въобще. В обикновените технически системи има по-голям риск от грешки и повреди за разлика от интелектуалните и затова този аспект не е специфичен за ИИ (грешките в системите за автоматично управление на атомните електроцентрали са много по-опасни отколкото тези на ИИ при вземане на решение за потребителски кредит).

Вторият аспект засяга самата същност на работата на интелектуалните системи. Доколко ИИ вземат решения на слабо формални задачи, използвайки евристика (трудно доказуема коректност на правилата, практически даваща приемливи резултати), правдоподобни разсъждения и механизми. Това сваля очакванията от системите на ИИ да вземат оптимални и еднозначно верни решения и създава възможност за намиране на решения, които са „субоптимални“, „разумни“ или „пригодни“ за съответния случай. Тук има пресечна точка между надеждността и прозрачността на ИИ – разработчикът може бързо да намери грешката. В непрозрачните системи така наречената „черна кутия“ има възможност за внасянето на скрити корекции, делегиращи вземането на решения дълго време по различен начин, без да предизвиква съмнение.

Третият аспект е наречен проблем на програмите съветници („опит оператори“), имащи пряко отношение към ети-

ката. Проблемите са актуални не само за интелектуалните системи, защото с тяхното прилагане те се издигат на съвсем друго ниво. Експерименталните изследвания показват, че в условия на неопределеност и дефицит на време у човека се създава свръхдоверие към ИИ и роботите, т.е. той вярва повече на машините отколкото на себе си³². Доверието още повече нараства, ако програмата дава коментар на своите действия. Изцяло разчитайки на системите на ИИ, човек все по-рядко взема самостоятелни осъзнати решения, което крие риск от грешки и загуба на квалификация. Притежателят на интелектуална система е изкушен да не наема нискоквалифицирани специалисти или да предлага по-ниско заплащане. От тази гледна точка е важно в опасните производства, медицината и образованието специалистите да запазят високата си квалифицираност – не е важно те да знаят точно кой бутон трябва да натиснат, а да познават системата, с която работят детайлно. Колкото и да е „умна“ интелектуалната система, на второ ниво е необходим човешки контрол и хората трябва да са достатъчно подготвени, за да го осъществяват.

Стандартизацията трябва да е способна да повишава нивото на надеждност на системите, отчитайки минималното ниво, след което системата губи своята надеждност. Това естествено не трябва да се превръща в догма, която да спира иновациите в технологията.

ПРИВАТНОСТ НА ИНФОРМАЦИЯТА, ПОЛУЧЕНА С ПОМОЩТА НА АНАЛИЗ НА МЕТАДАННИ

Алгоритмите на ИИ, имащи възможността да получават нова персонална информация за хората, постигат това през анализ на голяма база данни, извлечени от метаданни. Съби-

³² Robinette P., Howard A., Wagner A.R. Conceptualizing overtrust in robots: Why do people trust a robot that previously failed? // *Autonomy and Artificial Intelligence: A Threat or Savior?* Cham: Springer International Publishing. 2017. P. 129–155. DOI: https://link.springer.com/chapter/10.1007/978-3-319-59719-5_6; Wagner A. R., Borenstein J., Howard A. Overtrust in the Robotic Age // *Communications of the ACM*. 2018. Vol. 61, no 9. P. 22–24. DOI: <https://dl.acm.org/doi/fullHtml/10.1145/3241365>

райки повече данни за човека в цифров профил, собственикът на алгоритъма (фирма, държавна служба, полиция) получава възможност във все по-голяма степен да предсказва поведението на субекта (какво ще потребява, за кого ще гласува, как може да бъде манипулиран и пр.).

В наше време държавата и обществото волно или неволно трябва да формират своето отношение към допустимия праг на обработка на такива данни на базата приватност, прозрачност и справедливост на ИИ, създавайки стандарти за разработване на етични интелектуални системи.

ЧЕТВЪРТА ТЕМА. МАШИННА ЕТИКА

От тази тема ще научите как етиката на поведение на системите ИИ предизвиква редица проблеми. Ще бъде хвърлена светлина върху разбирането за това, в какво се изразява етичното решение. Сложността на проблема идва от избора на етичните норми, които трябва да бъдат заложени в ИИ.

ПОДХОДИ ЗА РЕШАВАНЕ НА ПРОБЛЕМА НА МАШИННАТА ЕТИКА

При разглеждането на проблемите на ИИ спектърът е твърде широк и включва както въпроса за опасността от „домашните машини“ (проблема за машинните грешки в разработките на А. Тюринг³³, опасността за човека в тези на Н. Винер³⁴, негативния сценарии за развитието на ИИ на Бостром³⁵ и т.н.), така и опитите да се формализира понятието

³³ Turing A.M. Computing machinery and Intelligence // Mind. 1950. Vol. 54, no 236. P. 433–460

³⁴ Wiener N. Some Moral and Technical Consequences of Automation // Science. 1960. Vol. 131, no 3410. P. 1355–1358

³⁵ Bostrom N., Yudkowsky E. The Ethics of Artificial Intelligence // Cambridge Handb. Artif. Intell. 2011. P. 1–20.

морал със задача да се имитира в програмните и технически системи. Постиганията в тази област са твърде малко и това се дължи на липсата на потребност от нея и съответно на задачи, при които такава формализация е необходима. С всеки изминат ден ситуацията се изменя и проблемът за реализацията на етичната компонента в ИИ става все по-актуален и значим в посока към:

- реализация на машинна етика;
- формализация на етичните понятия;
- верификация и валидизация на етичните компоненти;
- стандартизация на машинната етика;
- стандартизация на етичните аспекти на ИИ.

Прилагана към различни научно-технически области, етиката бива разбрана основно в смисъла на опасността от използването на една или друга технология, докато при ИИ ситуацията е съвсем различна – освен използването на ИИ и професионалната етика, създадена на основата на интелектуалните системи, на преден план излизат проблемите, свързани с етичното „поведение на интелектуалните системи“, или т.нар. *машинна етика*.

Пред учените, юристите, философите стои задачата да създадат подходи за проектирането на интелектуално автоматизирани системи с етични компоненти, които да се систематизират в стандарти по проектирането на такъв род системни разработки. В такова едно етически обусловено проектиране предмет на изследване е изборът на едни или други критично важни за човека и обществото действия и решения. От значение са ситуациите, в които се прави избор между някои евристики, базирани на етични императиви.

Евристика, това са правилата, помагачи при търсенето на варианти за решения. Тя няма строга обоснованост, а е мотивирана от засилването на ефективността на търсенето (пример за евристика е правилото на лявата или дясната ръка при търсенето на изход от лабиринт).

За да се излезе от кръга на чисто битовото разбиране на морала, принципно важно е да се създадат връзки между

етичните концепции и понятия и техническите компоненти, в които те могат да бъдат въплътени, т.е. решаването на въпроса за конструктивната определеност и онтологията.

Онтологията е максимално точното, строго и ясно описание на някои понятия, явления, предметни области. Формално онтологията се състои от термини, организирани в таксономия (система) с техните определения и атрибутивност и съответно свързаните с тях аксиоми и правила за формиране на изводи. Терминът (онтология) е въведен при разглеждането на въпроса за взаимодействието на интелектуалните системи една с друга и с човека.

От чисто техническа гледна точка онтологията е специфична база знания, които могат да бъдат „прочетени“, разбрани и разграничени от разработчика. Те са основният инструмент за създаването на етичната компонента на интелектуалните системи.

От дълго време се разработва групова роботехника – вместо да се създават сложни устройства за решаване на комплексни задачи, поради целесъобразност се разработват множество прости устройства, свързани помежду си, решаващи задачи (патрул на територии, разузнаване, строителство, транспорт на товари и пр). На практика обаче те се оказват твърде неефективни. Възниква хипотеза: създаването на ново качествено ниво на груповата роботехника е възможно само в случаите, когато роботите създават не просто група, а социум с всички негови закономерности на взаимодействие (подражание, агресия, доминация, общуване и т.н.). Управлението на такъв социум е основна задача, защото дейността е насочена към постигане на решения в практиката, нужни за разрешаването на човешка проблематика. Изследванията показват, че най-ефективните способности за управление на социум от робот са свързани с морални надстройки. Това изисква допълнителни правила за поведението на робота, делегиращи решаването на вътрешните конфликти, възникнали вътре в социума и определящи инди-

видуалното поведение на роботите, съчетавайки общите интереси с индивидуалните потребности.

Моралът е друга удобна, безопасна и лека надстройка в системите за управление. Става въпрос за морални императиви, които определят характера на поведението на агента и социума като цяло делегира по-агресивно или затворено поведение и нива на търпимост спрямо другите и т.н.

Така моралът се разбира като регулиращ механизъм по отношение на конфликтни ситуации между агентите и като целеполагащ механизъм. В рамките на социалното моделиране на поведението в груповата роботехника се формира подражаващо поведение и социално учене. На тяхна основа се правят изводи за възможностите за моделиране на такива механизми, като емпатия (отзивчивост на емоционалните състояния на другите). Емпатията е пригоден механизъм за определяне целите на работа или интелектуалната система. Ако тя може да разпознава емоционалното състояние на контрагента (човек или член на групата работи), то взаимодействието с човека е възможно и е предпоставка за приятелски „етичен“ интерфейс.

Проблемът за верификацията е един от най-важните въпроси, свързани със създаването на комплекс от тестове, проверяващи етичността на интелектуалните системи и преминаващи през двуетапен цикъл на проверка:

1. Представяне на ситуация със степен на неопределеност, свързана с нравствени съображения (примерно изискваща екстрено спиране или саможертва от страна на интелектуално автоматизираната система, правеща нравствен избор).

2. „Прослушване“ на веригата от съждения, даваща параметрите на взетото решение.

Тази втора компонента е задължителна в процедурата по верификация, за разлика от небезизвестния проект Moral Machine, при който на база статистически подход не се разбира защо интелектуално автоматичната система взема тъкмо такова решение, а се получава апроксимиран отговор (на база среднестатистическа извадка обучаващи примери).

АЛГЕБРА НА СЪВЕЩТА – ФОРМАЛИЗАЦИЯ НА ЕТИКАТА

Съществува математически апарат, позволяващ да бъдат формализирани етичните понятия. В частност особен интерес представлява подходът, позволяващ да бъде оценявана една или друга технология, използваща изкуствен интелект за съответствието и спрямо някои етични норми, критерии, стандарти. Тук трябва да отбележим, че формализацията на етичните норми е тясно свързана с една задача от по-горно ниво – формализирането на хуманитарното познание.

Проблемът за формализацията на етичните норми включва някои фундаментални задачи:

- създаване на форми за представянето на етичните норми - критерии и признаци;
- избор на съответен математически апарат за работа с тях – съпоставяне, измерване, анализ.

Механизъм	Описание	Коментар
Булева алгебра	Изказванията могат да бъдат само истина или лъжа – използва се двоична логика	Добре развита, с широко приложение и програми библиотеки за различни инструментални средства. Проблематична от гледна точка на това, че различните етични проблеми не винаги могат да бъдат в „черно“ и „бяло“.
Пръстенова скала на Д. А. Поспелова	Двойно-базова оценка на обектите, отразяваща динамиката на експертните знания и тяхната зависимост от онтологичните фактори	Преодолява еднозначността на булевата алгебра.
Многозначна логика	Вид формална логика, в която се допускат	Преодолява се еднозначността на Булевата

Механизъм	Описание	Коментар
	повече от две истинни значения на изказването.	алгебра. Значителна сложност на реализацията.
Размита логика	Обобщение на многозначната логика	Преодолява се еднозначността на Булевата алгебра и сложността на многозначната логика. Неустойчивост спрямо изходните данни.
Теории на решетките, и конкретно - етичните решетки	В рамките на теориите на решетките се изследват частично подредените множества.	Актуален и перспективен подход.
Метод на вербалния анализ	Групата методи ВАР се базира на използването на постиженията на различните научни дисциплини: когнитивна психология, приложна математика, организационни теории и пр.	Съчетаване на качествена и количествена информация, съждения на експертно ниво, обективни и субективни фактори и пр. Обясненията за взетите решения се дават в термините на предметната област. Като недостатъци на ВАР методите се представят по-голямата трудоемкост за специалистите, вземащи решения, и особено с признаци в пространство с голяма размерност.

КАКВА ЕТИКА ТРЯБВА ДА СЕ ПРИЛАГА ПРИ МАШИНИТЕ?

Механизмите, позволяващи имплантирането на етични норми в интелектуалните системи, са необходимо, но не достатъчно условие едни и същи етични компоненти да са в арсенала на разработчиците. Стоят открити въпросите за това какво е етично за ИИ в един или друг случай, по какви критерии ще се определя етичността на постъпките – дали това ще е мнението на повечето хора, държавата или моралните философи като специалисти?

Показателен в случая е:

Проблемът с вагонетката (англ. Trolley problem) като мисловен експеримент в етиката, формулиран за първи път през 1967 г. от английския философ Филип Фут. *„Тежка неуправляема вагонетка се движи по релсите. На пътя се намират пет души, завързани за релсите от луд философ. За щастие, вие можете да преместите стрелката, отклоняваща вагонетката в друга посока, където се намира само един човек, също завързан за релсите. Какви ще бъдат вашите действия“?*

Философският проблем с вагонетката е илюстрация за реалните обстоятелства за безпилотния транспорт и поставя остро проблема пред разработчиците какви правила да зложат в машините. Продължава известният експеримент на Масачузетския университет, наречен „морална машина“ (Moral Machine)³⁶ с цел да се даде отговор на въпроса: има ли риск от авария и кого ще жертва машината в конкретна ситуация? Системата се обучава през милиони примери (приблизително 40 милиона отговора) и избира коректното нравствено действие на принципа „такова е решението на мнозинството“. В етиката обаче съществуват две основни направления – утилитаризъм и абсолютизъм. „Така правят всички, следователно е правилно“ е пример за

³⁶ The Moral Machine Experiment // Moral Machine. URL: <http://moralmachine.mit.edu/>

утилитарен подход, продиктуван от изгодата и практичността. Подходът на абсолютизма е друг – „всеки един човешки живот е свещен“ и принципно забранява да се жертва един живот дори за сметка на няколко човешки живота.

Съгласно принципите на моралната философия едно решение не се приема за етично само защото отчита интереса на мнозинството и това, към кой принцип се придържа. В света на технологията ценностите се променят много бързо и в условия на неопределеност се цени повече рефлексивното мислене и диалогичността, отчитаща интересите и на двете страни, отколкото предварително инсталираните принципи. Това е дискусийна област в научните среди и особено се набляга на проблема за моралния аутсорсинг – трябва ли да се дава в ръцете на машините вземането на морални решения, доколко машините ни помагат ние самите да се държим етично.

ПЕТА ТЕМА.

СЪВРЕМЕННИ ЕТИЧНИ ФЕНОМЕНИ, ПОРОДЕНИ ОТ ИЗПОЛЗВАНЕТО НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

КУЛТУРНИ РАЗЛИЧИЯ НА ЕТИЧНИТЕ НОРМИ ЗА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Много системи ИИ се разработват в транснационални компании и се прилагат впоследствие в няколко страни. Конкретните норми за ИИ се разминават от културните различия в сегмента „морално – неморално“. В експеримента „Морална машина“ участват хора от 240 страни и там са правени компромиси по отношение жертването на животни вместо хора, на закононарушители вместо лоялни граждани и се прилагал утилитарен подход. Оказва се, че индивидуалните полови различия, жизнен стандарт, религиозни и политически възгледи не са повлияли върху избора, докато културните различия са

били от съществена важност при вземането на решения. В страните с високо ниво на колективизъм респондентите са по-склонни да се ръководят от броя спасени при вземането на решения, свързани със спасяването на човешки живот, и предпочитат да избират възрастни хора за сметка на младежите при правенето на такъв избор. Там, където има „широка ножица“ между доходите, изборът на спасени пада върху хората с висок статус за сметка на нископлатените. В страните от Северна Америка и Европа в повечето случаи изключват намеса в работата на машините (автомобилите) и са склонни да жертват пешеходци, за разлика от Югоизточна Азия и Близкия Изток, където се случва обратното – спасяват се пешеходци и възрастни хора. В Латинска Америка и Южна Африка по-често се жертват старците отколкото младите, хората с нисък статус пред хората с висок и мъже заради жени³⁷.

Така вземайки под внимание културните различия, можем да дефинираме редица въпроси, свързани с етиката, залагана в ИИ и формулирана като стандарт:

- Коя културна норма е релевантна за ИИ?
- Трябва ли да се вземат под внимание в работата на ИИ различията в етичните норми в регионите, в които се прилага (т. ар. „етическа локализация“)?
- Необходимо ли е предварително да се договарят етичните норми и те да станат кодекс при използването им?

ДЪРЖАВАТА И ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ – ЕТИКА И ДОВЕРИЕ

Доверието към системите на ИИ е тясно свързано със социалното доверие, т.к. подобни системи най-бързо се внедряват в държавния апарат и той се ползва с голямо социално доверие. Държавното регулиране в използва-

³⁷ Awad E., Dsouza S., Kim R. Et al. The Moral Machine experiment // Nature. 2018. Vol. 563. P. 59–64. DOI: 10.1038/s41586-018-0637-6

***нето на ИИ не се гради върху мнението на мнозинството, но не е добре да бъде до край делегирано от етичните те-
ории, без да се има предвид и мнението на гражданите.***

Държавата събира огромен обем от данни и съответно за тяхната обработка се използва ИИ. Такова прилагане на технологиите е в стадий на експеримент в редица държави и това се случва в две направления – аналитична работа с информацията и автоматизация на рутинните интелектуални процеси. Използването на ИИ заедно с другите технологии дава резултати в гъвкавата адресна помощ от страна на социалните и комуникационни служби, профилактичната дейност при предоставяне на здравни услуги, бързата адекватна реакция при извънредни ситуации и високорисковия технологичен надзор.

Често в СМИ се гледа на ИИ като на сензация и екзотика, предвещаваща радикални промени, страхове и надежди. Широката публика приема, че развитието на технологиите ще доведе до увеличаването на пропастта между бедни и богати, като в известна степен надценява възможностите им. Съществува и фетишизация на възможностите на ИИ по отношение на справянето с редица социални проблеми. До известна степен готовността за прилагането на системи на ИИ се определя от нивото на разбираемост и достъпност до гражданите, доверието към разработчиците и другите ползватели.

Неизбежно използването на алгоритми и решения, предсказващи поведението на гражданите, води до бюрократизъм и технократизъм в управлението и това ще се увеличава с намаляването на прилаганите човешки решения. Редица изследователи предупреждават, че засиленото използване на ИИ води до зависимост и неутрализира обективността на правителствата, създавайки усещането за отсъствие на контрола, оказван от гражданите и държавните органи. Психологическите изследвания показват, че доверието към киберфизическите системи за разпознаване силно се различава от това към хората, изпълняващи подобна дейност, и често преминава в пълно недоверие. Въпросът е не толкова какви етични систе-

ми са приложени в ИИ, а кой ги имплантира, а оттам и пада доверието в държавните институции.

„Ефектът на изкуствения интелект“ е феномен, свързан с това, че ползвателите не знаят къде се използва такава технология, защото тя е незабележима за тях. Нещо повече – ползвателят не си дава сметка как технологията формира дневния ред на новините и препоръките в социалните мрежи. От друга страна, наличието на такива алгоритми води до културен разрыв между хората, които са готови да влязат в някаква неопределеност по отношение на бъдещето и да имат възможност да правят избор, и тези, които не искат да го правят.

Само 1% от ползвателите четат докрай договорните условия за предоставяне на дадена услуга и с такъв механичен подход се налага някаква „архитектура на избора“, тласкаща ползвателите към определен тип решения, „подобряващи“ начина им на живот. Дори да има гаранции за защита на човешките права, такава патернизация, която използва алгоритми и допълва човешката природа, води до намаляване на съзнателността и рефлексивността в обществото.

Проблемно е и използването на т.нар. „ботове“ (роботизирани системи) в комуникацията с гражданите поради факта, че институциите в повечето случаи са изкушени да не предупреждават за такова съприкосновение. Както показват редица експерименти, хората са склонни да се кооперират с ботове, които се представят като хора. Доверието рязко пада, ако потребителят разбере за мимикрията. Разкрилият се бот, взаимодействайки с хората, е обучен да не очаква от тях каквото и да било сътрудничество и взема инициативата без оглед на потребителя. Тези психологически закономерности поставят пред държавните органи дилемата да направят услугата по-ефективна или тя да бъде достатъчно прозрачна за ползвателя. Такъв избор е тежък в областта на здравеопазването и извънредните ситуации, където решенията на ботовете може да спасят човешки живот.

Друг голям проблем е, когато решенията се вземат на принципа на етичните правила на мнозинството, защото те

всички са обременени от социалните феномени на отчуждението, колективните страхове и конформизма. Ярък пример са Холокоста и борбата за правата на жените. От друга страна, залагането само върху етичните теории, без да се взема под внимание общественото мнение, също не е правилно. Възможно решение на тези проблеми е създаването на етична инфраструктура, делегираща обсъждането на етичните проблеми, свързани с ИИ, предполагаща използването на делиберативни форми на демокрация, наричани граждански асамблеи, обществени групи и държавни съвети³⁸.

В ежедневието си не рядко се сблъскваме с нарушения на социалните норми. Ако не ни се отдаде да повлияем на ситуацията, ние правим компромис, избягвайки моралната отговорност и позовавайки се на:

- корпоративните норми или високите цели;
- евфемизми, представящи нашето поведение по лицепрятен начин („въздържал се“, вместо „премълчал“);
- прехвърлена върху други хора отговорност (така реши ръководството);
- прехвърляне на вината върху потърпевшите, дехуманизирайки ги.

Естествено, зад всички тези защитни механизми стои задачата да запазим самооценката си. Куриозното е, че тъкмо системи за ИИ могат да помогнат в преодоляването на такива самооправдания. В близко бъдеще се очаква да се появят програми в технологиите на ИИ с предназначение „когнитивно протезиране“, помагащи в преодоляване на отлагането за после, компенсиращи присъщата за хората неадекватна ви-

³⁸ През 2018 г. в Ню Йорк е приет закон за предотвратяването на дискриминацията през алгоритми в държавните служби. Той става основание да бъдат създадени обществени групи от експерти правещи етичен и правен анализ на работата на градските системи за вземане на автоматизирани решения. През 2019 г. в Ню Йорк се появява нова професия – сътрудник по въпросите на алгоритмите към Кметската управа (Algorithms Management and Policy Officer). На него е делегирана борбата с предубедеността в алгоритмите и повишаване на отговорността при вземането на решения.

зия за бъдещето (такива програми се използват за прогнозиране на климатичните промени).

Тенденция сред програмистите е да се създават точно такива програми, които не заместват човека, а го допълват. От особена важност са разработките на колективен интелект, прилаган в краудсорсинговите формации и публичните дебати³⁹.

ОЦЕНКА НА ВЪЗДЕЙСТВИЕТО НА АЛГОРИТМИТЕ – ОТ ТЕХНОЛОГИЯТА КЪМ ПРАВАТА НА ЧОВЕКА

В настоящето оценката на въздействието на ИИ и алгоритмичните системи върху правата на човека не включва в процедурата оценка в рамките на регулиращото въздействие на проектите от нормативно правно естество. В перспектива има визия такава оценка да стане задължителна при разработването на ИИ и алгоритми и цифровизацията на държавите като цяло.

Етичните и правни последиствия от внедряването на ИИ изискват да се използват методи за оценка на въздействието, нещо което е ключов инструмент за „умното регулиране“ (smart regulation) в целия свят. Най-популярна е оценката на регулиращото въздействие (regulatory impact assessment), но тази методология в недостатъчна степен взема под внимание развитието на технологиите и свързаните с тях предизвикателства.

Оценка на въздействието върху правата на човека

Оценката на въздействието върху правата на човека (human rights impact assessment) като самостоятелна институция за първи път се появява в Ръководните принципи на

³⁹ Платформата Assembli използваща ИИ поддържа обмен на мнения в мוניципалните сдружения на метрополията Голям Париж. С помощта на ИИ технологии предоставени от Insights.US правителството на САЩ събира и анализира предложения по усъвършенстването на процедурите за издаване на паспорт. Платформата ROVOX използваща подобни технологии повишава ефективността на взаимодействието между избирателите и конгресмените при разработването на законодателни инициативи. Verhulst S. G., Zahuranec A. J., Young A. Identifying citizens' needs by combining AI and CI. Building a Repository of Tools and an Evidence Base for Augmented Collective Intelligence // The Governance Lab. URL: <http://www.thegovlab.org/static/>

предприемачеството като аспект на правата на човека в ООН, одобрени от Съвета за правата на човека през юни 2011 г. Този документ препоръчва на бизнеса да внедрява оценка на въздействието върху правата на човека във всички бизнес функции и процеси. Оценките за въздействие върху правата на човека трябва да се изпълняват:

- преди започването на някакъв нов вид дейност;
- при настъпването на някакви сериозни промени в дейността;
- като отговор на променящи се условия;
- периодично.

В рамките на Съвета на Европа оценката на въздействието върху правата на човека има по-голямо развитие в сравнение с това на ООН. Препоръките на CM/Rec(2016)3 на Комитета на министрите в Съвета на Европа по правата на човека и бизнеса прави такава оценка не само на самите фирми, но и на държавите членки на ЕС при осъществяването на законодателните, регулативни и други мерки.

През януари 2017 г. Консултативният комитет на Конвенцията на ЕС по защита правата на физическите лица, свързани с автоматизираната обработка на персоналните данни (T-PD), приема ръководни принципи по защитата на физическите лица, свързани с обработката на персоналните данни в сферата на голямата база данни.⁴⁰ Приоритетно е етичното използване на данните, без да противоречи на релевантните ценности на съответното общество, включващо правата на човека. Предвижда се създаването на специален комитет по спазване на етиката за всички оператори. През януари 2019 г. същият комитет приема Ръководни принципи за изкуствения интелект и защитата на данните.⁴¹ През май същата година комисарят по правата на човека при Съвета на

⁴⁰ Guidelines on the protection of individuals with regard to processing of personal data in a world of Big Data//Council of Europe.

⁴¹ Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>

Европа публикува препоръки „Изкуствения интелект: 10 стъпки за защита правата на човека“.⁴² Държавите членки на ЕС са длъжни да изработят правова база, която регламентира процедурата за провеждането на оценка на въздействието върху правата на човека от страна на системите ИИ, изпълнявана от държавните органи (human rights assessments).

ОЦЕНКА ЗА ВЪЗДЕЙСТВИЕТО НА АЛГОРИТМИЧНИТЕ СИСТЕМИ – ПОДХОДИ И МЕТОДИКИ ЗА РЕГУЛИРАНЕ

Под внимание при казуса със защита правата на човека е факторът *въздействие на алгоритмичните системи* (algorithmic impact assessment)⁴³. Тип оценка дава характеристиките на риска спрямо правата на човека, етичните и социални последствия от прилагането на алгоритмични системи. Създава се и известна вариативност на моделите. GDPR прави оценка на въздействието до колко са защитени данните (data protection impact assessment) и същевременно с това Съветът на Европа следи за въздействието върху правата на човека. Различните модели на алгоритмична оценка в същия момент са свързани със субекта, който ги провежда – от една страна, това е самият оператор на данните, а от друга, това е трета страна по акредитация, която следи за оценка на въздействието върху правата на човека. Тук неразрешен въпрос е задължителността и доброволността на самата оценка, без тя да се превръща в бюрократично препятствие.

В момента Комитетът на министрите при Съвета на Европа разработва препоръки за въздействието на алгоритмичните системи върху правата на човека⁴⁴. Този проект пред-

⁴² Unboxing Artificial Intelligence: 10 steps to protect Human Rights // Council of Europe. URL: <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>

⁴³ Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>

⁴⁴ Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems // Council of Europe. URL:

вижда оценката за правата на човека да се прави както от страна на държавата, така и от страна на бизнеса. Акцентът е върху системите с висок риск и включва оценка на възможните трансформации на съществуващите социални, институционални и управленчески структури с ясни препоръки как да се предотврати или смекчи високият риск относно правата на човека.

ПРИВАТНОСТ И ЗАЩИТА НА ПЕРСОНАЛНИТЕ ДАННИ

Цифровите технологии са област, в която е възможно неетичното използване на личните данни и лишаване от конфиденциалност. Събирането на странична информация с цел печалба или изтичането на данни поради формализиран контрол или пробиви в законодателството правят всеки един гражданин уязвим в съвременното.

Понятие за приватност

Приватността е многосъдържателен термин, но в ядрото си съдържа личния живот на човека и неговата частна страна – телесност, темперамент, семеен и приятелски кръг, предпочитана форма на използване на глаголното време, стил на обличане, целеполагане, съдба, идентичност и пр. Съприкосновението с цифровите технологии води до рискове от:

- огромна прозрачност спрямо заобикалящите в съвременния начин на живот;
- твърде голяма вероятност данните да придобият публичност;
- безпрепятствено използване на данните с цел облаги, което може да се случи съвсем легално и да бъде използвано против самия човек.

Тук става дума не само за защита от пряка намеса (примерно, принудителни медицински процедури), а за невиди-

мост за околните относно данни за личния живот, без възможности за разпространението дори при наличие на такива – контролът на разпространението стои в основата на приватността.

Проблемът идва от това, че в цифрова среда личната информация се разпространява изключително бързо и в много по-голям мащаб отколкото в миналото. Същевременно качеството на самата информация е увеличило областите, които засяга и позволява да се правят много по-генерални изводи за даден човек, нещо което дори е извън неговите възможности. Опазването на такъв тип информация е изключително сложно и причините за това са:

- Злонамерено използване на личните данни може да се използва за похищение върху имуществото, измами или пряко нападение и това изостря вниманието към регистрите, използващи цифрови технологии. Всяка една институция, изискваща и използваща подобен тип информация, трябва да гарантира нейната неприкосновеност.

- Разгласяването на такъв тип информация може да повлияе негативно върху социалния ни статус поради факта, че сме силно социално обусловени (одобрение и порицание, репутация). За да получи поддръжка от референтната си група, човек контролира поведението си или проявява нонконформизъм. И в двата случая се стреми да поддържа самореализацията си. Запазвайки конфиденциалността на личното си пространство, човек може да има различно поведение в различни сфери на социума. Това бива много рестриктивно в случаите, когато се загуби контролът върху личната информация, правеща живота силно прозрачен.

Стои въпросът доколко човек може да се приспособи към подобна ситуация. Възможните варианти са свързани със самоограничения, отказ от избор, ако той противоречи на обществения. Подобна самооцензура прави човека все по-малко щастлив. Самата унификация прави цялото общество по-малко стабилно и адаптивно. В такива условия човек започва да изключва от социалния си репертоар търсенето на социал-

но обсъждане и същевременно става все по-толерантен към вариативност на поведение и начин на живот в перспектива.

Чисто философски, приватността се свързва с понятията *воля* и *свобода на избора*. Данните за човека относно предпочитания, реакции, потребности, взаимоотношения, битова организация, форми на предвижване, биват събирани и систематизирани, за да може да бъде делегиран неговият избор по отношение на закупуване на стоки и услуги. Темата става особено актуална, когато вземем под внимание това, че държавата, работодателите и другите субекти могат да използват данните на човека, за да го подтикнат към извършване на определени действия (т.нар. наджинг). И в единия, и в другия случай мотивът е „за благо на човека“, но факторите *свобода на избора* и *волеизявление* остават под въпрос.

Под **персонални данни** (ПД) ще разбираме:

Всяка информация с пряко или косвено отношение към определено физическо лице (субект на ПД).

Според определението *информация* са сведения (съобщения, данни) независимо от формата, под която се представят (текст, графика, видео), на всякакъв възможен носител. Като ПД се възприемат не само тези, които са в цифрова форма, а и такива на хартиен носител. В действителност приоритетна е цифровата форма, защото позволява автоматична обработка на голямо количество данни на национално и наднационално ниво. Този тип информация е изключително атрактивна за операторите, но е опасна от гледна точка на защита правата на човека и личната тайна.

Към ПД се отнася всяка информация с оценъчен характер, която може да е достоверна или недостоверна, взета непосредствено от субекта на ПД или от други възможни източници. Данните, свързани с физически лица (име, възраст, здравен статус, отпечатьци, документи, характеризиращи личността), са само една малка част от информацията, която се обработва от операторите, но има и друга информация, която е проблемно да се отнесе към ПД.

Сега практически всяко мобилно устройство, уебсайт или онлайн платформа, мобилни приложения и социални мрежи събират цифрови следи на ползвателите си в режим 24/7. Такова действие е придобило популярност под наименованието „капитализъм на следенето“ и представлява косвен интерес за държавата.

Още през 2014 г. един от големите експерти по криптография Брус Шнайдер приема следенето като бизнес модел в съвременния интернет с аргументацията, че това се нарича *маркетинг в областта*. През 2019 г. в книгата си „Ерата на капитализма на следенето“ Шошана Зубоф подробно описва явлението, наричано капитализъм на следенето или капитализъм на наблюдението (*surveillance capitalism*), работещо в цифровите технологии.

Във всяко едно общество, използващо капитализма на следене, винаги има наблюдатели и наблюдаеми (практически всички ползватели и граждани, от които безпрепятствено се събират данни). По думите на Ш. Зубоф моделът на новия капитализъм изглежда така: „Капитализмът на следене съвсем уклончиво провъзгласява човешкия опит като безплатна суровина, подлежаща на обработка в поведенчески данни. Въпреки че част от тези данни се използват за подобряване на сървърите, а всичко останало се обявява за излишък, собственост на компанията. Този излишък захранва водещите производствени процеси известни като машинен интелект, за да създаде продукти, способни да предвиждат какво вие ще направите сега, в близко и далечно бъдеще. В крайна сметка тези предиктивни продукти се предлагат на пазара под нова форма, наричана „пазар на поведенчески фючърси“. Търговията на този пазар носи на капиталистите огромни печалби поради факта, че много компании искат да платят, залагайки на нашето поведение в бъдещето“⁴⁵.

Доминиращи на световния пазар (изключение е китайският пазар) са социалните медии Google и Facebook. В Китай

⁴⁵ Zuboff S. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs, 2019.

под забрана са сървърите на американските data корпорации, но с голям успех работят WaChat, Weibo, Baidu.

Цифрова следа е съвкупността от данни за всички функционални ползватели извършени в цифровото пространство, представляващи основен източник за информация за предиктивна аналитика. IP адреси, устройства, време на посещение, параметри на устройството, продължителност на пребиваване на всяка страница, патерни на движение на курсора в процеса на четене на страницата и още хиляди параметри под формата на големи данни, които се събират от оператора. Да не бъде оставена цифрова следа, е технически невъзможно както и техническата възможност тя да бъде заличена.

Цифровата следа активно се използва за предикатна аналитика, разбирана като анализ на голямата база данни с цел да се направи прогноза относно най-вероятните сценарии на поведение, желания и потребности на ползвателя за управление на вниманието им с висока вероятност в полза на трети лица (бизнес, политика, специални служби). Колкото повече данни за човека има в системата, толкова с по-голяма точност може да бъде предсказано поведението му в бъдеще.

Цифровата следа не е задължително с манипулативен характер – в образователната сфера тя се използва за разработване на индивидуални образователни траектории и персонализиране на процеса на обучение.

Компаниите предоставят сървърите и услугите си безплатно, а в замяна събират всевъзможни данни за тези, които ги ползват. С помощта на ИИ данните и метаданните се анализира и се създава нова информация под формата на изводи, хората се групират по признаци, правят се предположения за техните интереси, характеристики и поведение с акцент в бъдещето – какво може да си купи, как ще реагира при определени условия, как ще гласува, къде ще отиде, какви мнения споделя и т.н.

Колкото повече данни от различни източници биват събрани, толкова по-предвидим става потребителят и толкова по-качествен е шаблонът, в който той може да бъде вкаран. Резултатите бързо се конвертират в печалба под формата на реклами към възможната таргет група. Този бизнессегмент създава нови професии:

- продажба на данни в качеството на посредник – информационен брокер;
- разработка и продажба на устройства (мобилни устройства), способни да събират данни с приложения за използване в устройствата;
- разработка на инструменти за анализ на данни, скринингови системи;
- престъпления с използване на данни;
- манипулиране на общественото мнение.

ДЪРЖАВНОТО ПРОСЛЕДЯВАНЕ – „ГОЛЕМИЯТ БРАТ“

„Големият брат“ е нарицателно, характеризиращо всесилната тоталитарна държава от романа „1984“ на Джордж Оруел. В денонощен режим той следи всеки гражданин, диктувайки му как да мисли, действа и живее, маркирайки всяко нарушаване на правилата, спуснати от държавата. Съвременната интерпретация на Оруеловата метафора включва събирането и анализа на огромно количество данни от системите за наблюдение с разпознаването на лица, скринингови системи и системи за предикатна аналитика за правозащитните органи. Данните за хора и инфраструктура са релевантни за контрола на реда, осигуряване на безопасността на държавата и хората в нея, но в същия момент са фактор за появата на дискриминация и т.нар. „меко подтикване“ към определени действия.

Видеонаблюдение с разпознаване на лица, емоции, силуети

Тази технология се разработва от десетина година. Разпознаването на лица е една от неизбежните дейности на човешкия мозък. Наблюдението с разпознаването на лица става изключително популярно при контролирането на голямо струпване на хора в промишлени и комерсиални събития и обекти. Основният момент, носещ смисъл, е предотвратяването на безредици и тероризъм. Проблемът идва от създаването на справедлив баланс между обществените и личните интереси и правата на гражданите като цяло.

Определено Китай е несъмненият лидер в проследяването с камери – 200 милиона до 2018 година и 8 от 10-те най-наблюдаваните градове в света са там. Естествено, впоследствие в обществото започва да се появява недоволство. Според данните на международната организация Algorithm Watch правораздаващите органи на минимум 10 държави от ЕС използват технологията за лицево разпознаване и още 5 смятат да го въведат в близките години. Единствено в Белгия и Испания това е забранено. Във връзка с това в ЕС се издига инициатива да се забрани масовото използване на технологиите за лицево разпознаване на лица и те да бъдат признати за „свръхрискова технология“. В Лондон технологията се използва от 2015 г. и е създаден един от най-големите центрове за видеонаблюдение с приблизително 470 000 камери. Разходите му са за милиони британски лири, при това системата разпознава погрешно подозрителните лица в повечето случаи. До задържане на истински престъпници така и не се стига и това предизвиква огромно недоволство, като съответно през 2019 г. съдът признава системата за виновна в нарушаване на човешките права.

СКРИНИНГОВИ СИСТЕМИ

Скринингът е технология, даваща някаква оценка и рейтинг на определен ползвател въз основа на някакви критерии за ценност на качествата на даден клиент (пр. банков кредит,

ведомствено жилище, социална помощ, държавни услуги), като същевременно се представят и свързаните с предоставяната услуга потенциални рискове. Социалната рейтингова система се аранжира според значимостта на клиента от скрининговата система.

Въпреки очевидните ползи, използването на скринингови системи е свързано с редица етични въпроси – правата на човека и дискриминацията. Някои от методите на банки и застрахователни компании са толкова неограничени, а и не се очаква в близко бъдеще тяхното регулиране. Нарушаването на приватността е един от най-често срещаните етични проблеми.

Показателите за скрининг зависят от професионализма на разработчика и алгоритмите и достъпа до определени данни. За скрининговите системи от особена важност са събирането на данъците, закононарушенията, здравният статус, трудовата заетост и т.н. Тези системи без предупреждение използват персонални данни. В САЩ има закон срещу генетичната дискриминация Genetic Information Nondiscrimination Act, породен от прецедентите на дискриминация на хора, чийто генетичен анализ показва висок риск от заболяване. Работодателите започват да им отказват постоянни договори и повишение.

По данни на Algorithm Watch⁴⁶ в няколко страни се прави опит да бъде внедрена скринингова система, предотвратяваща измамите в системите за социално подпомагане и потенциална опасност от неправомерно поведение. В нито една страна от ЕС не се предвижда въвеждането на единен социален рейтинг за цялото население. Както това примерно е направено в Китай.

КИТАЙСКИЯТ „ГОЛЯМ БРАТ“

През 2020 година в Китай се въвежда национална система за класификация на гражданите и предприятията под

⁴⁶ Kayser-Bril N. Personal Scoring in the EU: Not quite Black Mirror yet, at least if you're rich // Algorithm Watch.

URL: <https://algorithmwatch.org/en/personal-scoring-in-the-eu-not-quite-black-mirror-yet-at-least-if-youre-rich/>

формата на Система за социален кредит (ССК). Първоначално тя стартира като множество пилотни проекти и експерименти в редица райони, но в момента обхваща около 80% от населението на страната. Главната задача на такава система е да повиши удовлетвореността на пазара и обществото за сметка на култивирана искреност и доверителност на всички нива. Допълвайки западната система на кредитния рейтинг с още показатели, некасаещи финансовата сфера, Китай се стреми да регулира поведението на компаниите и частните лица с цел премахването на измамите и корупцията в управлението. Системата е особено взискателна към държавните служители.

Китайската комунистическа партия създава „глобална екосистема за събиране на данни“ и в нея влизат предприятия, технологични компании и чуждестранни партньори. В тази държава е най-мощната система за цензура и наблюдение, наричана „Златния щит“ или „Великия китайски фаярбол“ със система за автоматизирано блокиране на онлайн противници на управлението на държавата, политици и събития. За периода 2015-2017 година са блокирани повече от 13 000 сайта. Изискванията се актуализират постоянно – през есента на 2019 г. Администрацията на киберпространството на Китай обнародва проект, според който всички местни социални мрежи и сайтове трябва с помощта на алгоритъм въз основа на ИИ препоръчително да използват материали, поддържащи се към основните партийни ценности.

Най-силен е контролът върху китайската провинция Синцзян, където от 2009 г. се следи сепаратисткото движение на уйгурите. В тази провинция е задействана една от най-големите системи (с висока плътност) за видеонаблюдение в света, където събраната информация се съпоставя с биометричните данни за движение в градовете и провинциите, активност в различните приложения и т.н. всеки жител има профил, в който влизат семейният му статус, социалните контакти, религиозните убеждения, националността, контакти и пътувания извън страната. Съответно всеки гражданин

получава статус на: „безопасен“, „неутрален“ или „опасен“, като последните се депортират в лагери за „превъзпитание“. В такива лагери за политическа коректност и образование се намират около един милион души.⁴⁷

Трябва да се отбележи експанзията на китайските телекомуникационни системи по целия свят на основата на това, че Китай е един от най-големите строители на такава инфраструктура и влага изключително компоненти и технологии собствено производство. Така огромна част от глобалните дискретни данни стават достъпни за специалните служби на Китай, независимо дали са придобити легално или нелегално.

От 2019 г. Китай започва да предлага системи за видеонаблюдение в 63 страни⁴⁸. Съответно заедно с тях и експерти от Huawei, ZTE, Dahua. Докладите на ASPI показват, че в отделни проекти на „умни“ градове с китайска технология за видеонаблюдение, задействани в някои страни от Западна Европа, е възможно да се събират данни, които да се събират в Китай.

През декември 2019 г. става известно, че Китай планира за три години да замени цялото си апаратно и програмно обезпечаване на държавните органи и да премине на аналогово, произведено в страната. Независимо от мястото, където се прилагат компютърните и телекомуникационните мрежи и физическите места, в които се съхраняват, данните са от особена важност, защото така се обезпечава цифровият суверенитет на държавата, която освен това развива свой собствен пазар и производство.

⁴⁷ China's Algorithms of Repression // Human Rights Watch.

URL: https://www.hrw.org/sites/default/files/report_pdf/china0519_web.pdf

⁴⁸ Kharpal A. China's surveillance tech is spreading globally, raising concerns about Beijing's influence // CNBC.

URL: <https://www.cnbc.com/2019/10/08/china-is-exporting-surveillance-tech-like-facial-recognition-globally.html>

ДИСКРИМИНАЦИЯ И НЕРАВНОПОСТАВЕНОСТ ПРИ ЦИФРОВИЗАЦИЯТА

На международно ниво дискриминацията е забранена официално (законово), но се проявява в някои сфери поради чисто психологическите прояви на предубеждения и стереотипи. В технологиите и особено в алгоритмите на ИИ с голяма степен се появява възможност да се засилят подобни социалнопсихологически феномени, разгръщащи дискриминация и социална изолация.

За да бъде преценен рискът от дискриминация, свързан с използването на ИИ, трябва първо да бъде дадено терминологично разяснение. На първо място дискриминацията е нормативна концепция и тя е дадена от Европейския съд за правата на човека:

„... за да кажем, че има дискриминация, трябва да е налице разлика в отношенията към хората, намиращи се в еднаква или аналогична ситуация. Разликата в отношението ще бъде дискриминационна, ако в нея липсва обективен и здрав смисъл....Глава 14 забранява да се прави каквато и да е разлика в отношението освен в случаите, когато съществува различие на основа разграничаване на основа идентифицирана индивидуална обективна или персонална характеристика или статус, на основата на който даден индивид или група се различават една от друга. В списъка на такъв статус влизат пол и раса (цвят на кожата, език, религия, политически или някакви други убеждения, национален или социален произход и пр. Думата „И други признаци“ има широк смисъл и не се заключава само във вродените и неизменни характеристики на личността.

Ако обобщим, понятието дискриминация би приело следната формулировка: случаите, в които не се допуска, лишава се, предявяват се някакви изисквания към даден човек за разлика от останалите хора. Причините за подобни действия са несправедливи, ирационални, изградени върху предубеждения и стереотипи. Несправедливостта, основаваща се на не-

добро отношение към дадени хора и неоснователност на разграничаването между тях, се приема като противоречащо на етиката и обществения морал.⁴⁹

Дискриминацията бива **пряка** и **косвена**. Когато целта е да се ошети и ограничи представител на дадена група, е налице пряка форма на дискриминация дори в случаите, когато мотивът е маскиран или не напълно осъзнаван. Косвената дискриминация е ситуация, при която определени правила се прилагат към всички в еднаква степен, но към дадена група поради нейната специфика се създава някакъв прецедент на натиск или проблем (примерно градската среда с нейния транспорт, магазини, администрация), без преднамереност свързани с предвижване.

Въвеждането на цифровите технологии според експертите на ООН повишава рисковете от дискриминация въз основа на предубеждения и предразсъдъци, а от там - и социална изолация. Всички тези феномени, присъщи на доцифрофизирания период, с навлизането на ИИ се мултиплицират и разрастват. На основата на това регламентът на GDPR налага допълнителни ограничения за работата, свързана с персоналната информация по отношение на дискриминацията (политически, философски убеждения, членство в обществени формации, раса и етнос, интимен живот).

Предметът и съдържанието на предубежденията и стереотипите се променят с времето, но в никакъв случай не може да се очаква с навлизането на цифровите технологии съзнанието на хората да се измени в положителна посока. Всичко, което сега приемаме като предубеждения, в миналото е било обективна научна истина – такава е евгениката, идеята за непълноценните раси, джендърните диспозиции и пр.

При цифровите технологии в особена сила се проявява косвената дискриминация. Някои обществени групи и прослойки съвсем непреднамерено са лишени от достъп до технологиите и дори се намират в т. нар. „цифрова ексклузия“. То-

⁴⁹ Altman A., Discrimination // The Stanford Encyclopedia of Philosophy.
URL: <https://plato.stanford.edu/entries/discrimination/#TypDisMorSen>

зи ефект далече не е търсен и планиран, защото въвеждането на „цифрите“ е с цел да разкрива нови възможности за улесняване и опростяване на живота. Изследванията показват, че засегнатите от цифровата ексклузия хора не са равномерно разпределени по социалните страти, а са концентрирани в групи, които и без това са обществено маргинализирани и социално уязвими – хора с физически ограничения и заболявания, социално слаби, мигранти, представители на етнически малцинства⁵⁰. Това предполага разработчиците на такива алгоритми да създават специален дизайн за хора в неравностойно положение или да се запази аналогова форма на съхранение и предлагане на услуги в контактни центрове и специализирани курсове по цифрово ограмотяване.

Подобни косвени дискриминативни феномени се наблюдават при автоматичните скрининги и алгоритми за наемане на работа и достъп до социални помощи и услуги.

Ако трябва да обобщим, то:

- Съществува вероятност да бъдат пренесени практиките на пряка дискриминация в цифровата среда.
- Създават се нови условия за неравноправно отношение към някои групи, които привидно са обективни и научно обосновани.
- Навлизащата мащабно цифровизацията носи в ежедневието все по-голям дискриминационен потенциал спрямо достъпа до технологии.

Със същата актуалност в технологично отношение е рискът от косвена дискриминация. Препоръчително е държавните цифрови сървъри да са под постоянно наблюдение, дизайнът им да е максимално опростен и достъпен за целия социален спектър, предвиждащ алтернативни и буферни форми на достъп и комуникация. При формирането на правила, създаващи различни условия за отделни групи в обществото, да се

⁵⁰ Digital Exclusion, A research report by the Low Incomes Tax Reform Group of The Chartered Institute of Taxation. London, 2012. P. 16–24. URL: https://www.litr.org.uk/sites/default/les/digital_exclusion_-_litr_report.pdf

поддържа критичността по отношение на правилата за диференциацията на съответните групи.

Инклузия (от англ. Inclusion – „включване“) изразява процеса на включеността на всеки гражданин в социума. Обикновено под това се разбира достъпност на инфраструктурата, но инклузията включва и достъпността до цифрови продукти и всички други сфери от живота на човека. Инклузивната среда обезпечават достъпност на възможностите независимо от индивидуалните особености на човека.

ПАЗАР НА ТРУДА И РИСК ОТ БЕЗРАБОТИЦА

Роботизацията и цифровите технологии влияят върху пазара на труда, където перманентно е заложен проблемът с безработицата и смяна на професията. Началото на цифровата трансформация на държавата трябва да предвижда помощ за гражданите, които ще се адаптират към новите условия.

Четвъртата промишлена революция се представя като благо за работника и по подразбиране се приема, че технологията ще се разпространява равномерно и няма да има сблъсък с реални социални и икономически проблеми или да премине в състояние на неопределеност.

Основният двигател на технологичното развитие са глобалната интернет мрежа, ИИ, аналитиката на голямата база данни и облачните технологии. Тези отрасли се предвижда да станат базови за организацията и реализацията на работния процес. Компаниите ще усвояват интернет пространство, пазарни приложения, облачни технологии, машинно обучение и виртуална реалност, които ще бъдат спомагателни или водещи за дадени сектори на икономиката.

Прилагането на върховите технологии ще изисква все повече програмисти, data аналитици и специалисти по социални мрежи - social media marketing (SMM). Ще се запази високото търсене на специалисти по ИИ и ползвателски опит - user experience (UX), взаимодействие човек и компютър human-computer interaction (HCI), роботехници. С висока зна-

чимост са мениджърите, работещи с клиенти, тренъорите, специалистите по корпоративна култура и иновации. Върху техните плещи лежи цифровата трансформация на ниво организации с изцяло нетехническа компетентност.

Генерирането на нови специалности съответно изисква и нови умения и навици, съобразени със спецификата на работата. Налага се да се премине към преквалификация на съществуващия човешки ресурс. С особено внимание се гледа на т. нар. критично мислене, воденето на преговори, креативния подход и гъвкавост в условия на неопределеност. Отговорността за преподготовката лежи изцяло върху организациите, защото те трябва да преквалифицират структурите си не само на ръководно ключово ниво, а и на оперативно. Съответно за това се изискват стратегии, автоматизация на рутинните задачи и актуализация им с пазарните трендове.

Внедряването на работи в производството ще зависи от потребностите на конкретния отрасъл – в някои ще са необходими стационарни, а в други дроне и хуманоидни работи. Една такава активна роботизация неминуемо ще даде отражение върху производителността на труда, комуникациите, координацията, управлението и вземането на решения в режимите на работа.

Прогнозите за замяната на човешкия труд с такъв на работи звучат оптимистично – изгубените работни места са два пъти по-малко от тези, които ще се създават вследствие на роботизацията. Според прогнозите през 2022 г. срещу 75 милиона работни места, ще се създадат 133 милиона нови, но нивото на квалификация, а от там и нивото на заплащане, ще се снижи.

Друга проблемна област е технологичното сътрудничество - различните пазари, на които се изискват различни алгоритми за ИИ, се делегират от взаимодействието човек и технология. От тук тръгват и три възможни сценария:

I сценарий. Алгоритмите ще се делегират от тези функции, които човек може да изпълнява, като ИИ ще ги направят по-бързи и качествени, с което се дава възможност хората да

преминат към изпълнението на други задачи. Такова търсене може да се появи във фармацевтиката, здравеопазването, селскостопанското производство, финансите, складовите бази и логистиката, промишленото производство и държавната сигурност. Тук основната задача на алгоритмите е свързана с анализа на данни и помощ в човешката дейност по вземане на решения.

II сценарий. Алгоритмите напълно заменят човешкия труд там, където се изпълняват рутинни трудоемки операции (селско стопанство, строителство, охрана, логистика, кибербезопасност и промишлено производство).

III сценарий. Предвиждат се изцяло нови функции, изпълнявани от ИИ, които не са се изпълнявали от човек (операции с много висока скорост и мащаби на реализация). Такова прилагане е възможно и наложително в медицината, спътниковите връзки и т.н. Алгоритмите със своите възможности бързо да обработват обеми от данни с изчислителните си мощности дават възможност за скоростно препланиране и автоматизация.

Типът заетост е безусловна проблемна област, където трябва да се реши кои професии е необходимо да се автоматизират. Освен като умствен и физически, трудът може да бъде диференциран като рутинен и нерутинен. Тенденция е физическият рутинен труд да се роботизира най-напред. Умственият рутинен труд също подлежи на автоматизация, но като помощен за юристи и икономисти. Нерутинният физически и умствен труд, изискващ някаква креативност, едва ли някога ще претърпи автоматизация.

През 2016 г. в пресата се появяват съобщения, че изкуствен интелект предсказва решенията на Европейския съд с точност до 79%. Впоследствие се доказва, че тези проценти са значително завишени и не заслужават доверие – класическо изтичане на данни (data leakage), случващо се при машинното обучение, води до оптимистични резултати и реалните проценти са 58-68%. Предсказването на решенията се дължи на информацията, която системата има за лич-

ностните особености на съдията, който ги взема. Всичко това буди съмнения в прилагането на подобни системи в съдебните зали.

Както показват анализите и изследванията по отношение на бъдещето на трудовия пазар, автоматизацията и цифровизацията не само ще заменят човешкия труд, но и ще създадат нови професии и работни места. Някои от професиите ще изчезнат до 2030 г., като в сферата на транспорта и строителството това ще е в рамките на 56%, в производството 46%, в търговията 44%. Най-малко изчезващи професии ще се наблюдават в социалните дейности и здравеопазването, където процентите са едва 17%. Тенденцията е: колкото повече ръчен и рутинен труд има в една професия, толкова по-висока е вероятността тя да бъде заменена от машините. Естествено, това ще предизвика промени в организацията и правото, свързано с трудовите процеси и планирането.

Като цяло трансформацията на пазара на труда се случва в зависимост от това, до колко е структурирано дадено общество, какъв е темпът на научно-техническата революция и социално-икономическите процеси. Тези процеси се повлияват и от официалното и медиаторското отношение спрямо автоматизацията на човешкия труд. Технологиите се развиват толкова бързо, колкото хората свикват с тях на битово ниво и според достъпа до цифровизация на различните обществени слоеве. Процесите на цифровизация и внедряването на технологии изискват от икономическата система и преквалификация, и модернизиране на методите за обучение, за да бъде преструктуриран пазарът. Според отчета на Центъра за глобално развитие⁵¹ (Center for Global Development) с най-големи проблеми в цифровизацията на икономиката се сблъскват развиващите се страни. Въпреки стремежа да се спазват тенденциите към цифровизация, икономическият ръст е затор-

⁵¹ Schlogl L., Sumner A. The Rise of the Robot Reserve Army: Automation and the Future of Economic Development, Work, and Wages in Developing Countries — Working Paper 487 // Center for Global Development.

URL: [https:// www.cgdev.org/publication/rise-robot-reserve-army-automation-and-future-economic-development-work-and-wages](https://www.cgdev.org/publication/rise-robot-reserve-army-automation-and-future-economic-development-work-and-wages)

можен от структурната трансформация. В тези страни автоматизацията не предизвиква безработица, а по-скоро преразпределя съществуващата, преразпределя и разпределението на доходите – много ниско платени работни места, нисък стандарт на живот и социални гаранции, трудно достъпна данъчна система (няма правила какви данъци трябва да се плащат при експлоатация на работи). Реално автоматизацията не решава, а задълбочава съществуващите проблеми на икономиката на тези страни.

Такива промени, настъпващи все повече в дълбочина, изискват широко обществено обсъждане най-вече с групите, които ще бъдат пряко засегнати от цифровата трансформация. Политиката за внедряване на върхови технологии задължително трябва да е социално ангажирана, за да парира старите и да не създава нови обществени проблеми, търсейки нови подходи и решения.

ШЕСТА ТЕМА.

ОТРАСЛОВА ЕТИКА СПРЯМО ЦИФРОВИТЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИЕТО, ЗДРАВЕОПАЗВАНЕТО И СИГУРНОСТТА

В близките десет години се очакват кардинални промени в образованието, свързано с цифровите технологии. На промяна подлежи целият процес, което е обвързано с нови изисквания към резултатите от подготовката, създаваща нови компетенции и навици, необходими за живота в цифровия свят. Съответно това поставя редица етични въпроси.

Относно образованието се поставят етични въпроси в три направления – етиката като предмет на изучаване, заложен в основата на системата на образованието, и професионалната етика на преподавателите.

Етиката се явява задължителен предмет във философските факултети. В други специалности тя се появява като дисциплина, засягаща професионалната етика (деонтологията при медиците и психолозите). За повечето специалности тя си остава предмет на философията и няма приложен характер – по-голяма част от инженери и специалисти по цифрови технологии, държавни служители не получават подготовка в областта на етиката.

Неоспорим факт е, че цифровите технологии навлизат във всички области на живота ни и променят традиционните етични норми (събиране на лични данни, видеонаблюдение и пр.). Ситуациите, свързани със спазването на някакви етични правила, са достигнали своя критичен праг и това довежда до появата на различни етични кодекси и стандарти, а вече и до появата на такъв тип законодателство.

Все по-ясно става, че разработването на такива правила и норми не трябва да се случва постфактум, когато технологиите са вече навлезли, а трябва да се създават още на етап проектиране на системите за ИИ.

Приложната етика спрямо технологиите трябва да стане част от учебните програми, като това може да се случи или през разработване на етични концепции на държавно ниво, фиксирани в нормативни актове и заложи в учебни програми под формата на стандарти със задължителен характер. Другият вариант е свързан с това, учебните програми да съдържат общи знания и различни формулировки на цифрова етика, като на обучаемите се дава подробна информация за тях, но към коя концепция да се придържат, остава въпрос на личен избор.

В тази система е важно с етичните норми да са запознати не само тези които ги създават, но и тези, които реално ще ползват продуктите.

ЕТИЧНИ ПРОБЛЕМИ НА ЦИФРОВОТО ОБРАЗОВАНИЕ

Етични кодекси има в много университети, съставени със задача да регулират нормите на поведение на студенти и преподаватели в учебния процес. Таргет група проблеми е вземането и даването на подкупи, джендърните отношения, достъпността на образованието и етичните ценности, които се транслират директно в учебните програми. Съвременното повдига и други проблеми, свързани с присъствието на преподаватели и студенти в социалните мрежи (поместване на снимки с личен и интимен характер). Нови са и някои методи, навлизащи в образованието - персонализацията и адаптацията, дистанционните технологии и предикативната аналитика.

От една страна, дистанционните технологии повишават достъпността на образованието (хора със заболявания, отдалеченост, хора от други държави). Тази форма несъмнено отстъпва по качество на присъственото обучение, като поставя моралния въпрос за мотивацията относно получаването на образование за възрастните и децата – обикновено възрастните са склонни да получат необходимата им компетентност, преминавайки пълния курс на обучение, докато малките деца ще направят това с малка вероятност.

Дистанционното обучение, прилагано сред подрастващи, задълбочава риска от пристрастяване към електроните устройства и най-вече риска от когнитивни заболявания при продължително използване – повишена раздразнителност, намаляване на дълговременната памет и работоспособността. Същевременно с това има данни за благоприятното въздействие на системи с ИИ при обучението на деца с аутизъм.

Тенденцията е присъственото образование да се използва все по-малко и да се развиват и усъвършенстват дистанционните форми, които за сметка на качеството са по-достъпни.

Персонализацията и адаптивният подход при използването на цифровите технологии дават възможност да се анализират персоналните психологически особености на обучаемите.

мите и учебният материал да бъде представен в най-подходящата за тях форма. Това повишава в много голяма степен качеството на образованието и намалява риска от дискриминация (студенти и ученици, които са под или над средното ниво, винаги са в дефицитарна ситуация). В Китай 238 милиона деца и младежи представляват най-големият образователен пазар. Сравнявайки 2018 г. със следващата година, с 10,7% се е увеличил броят на онлайн обучаващите се (172 милиона), с 19,6% - броят на обучаващите се през мобилни устройства (142 милиона). Най-голямата компания, развиваща технологиите за ИИ, Squirrel AIlearning предлага вариативност на учебните програми, прилагайки адаптивния подход. Образователните системи, създадени на база ИИ, предлагат на студентите задачи, които съответстват на тяхното ниво, гарантиращи плавния прогрес и увереността в собствените възможности.

В настоящия момент все още актуална е смесената форма между присъственото и дистанционно обучение. Така нареченият „преобърнат клас“ предполага, че ученикът самостоятелно ще работи с материали преди самия урок (чете, гледа видео, попълва тестове), а по време на урока ще допълва останали недостъпни за него знания с помощта на преподавателя. Така преподавателят става уникален източник на нови знания, делегирани от индивидуалния подход спрямо всеки.

Адаптивният подход е тясно свързан с достъпността на образованието. Уравниловката обезценява достъпността (дори всеки да ходи на училище, няма гаранция, че ще получи знания). Друг проблем е събирането и обработката на данни за обучаемите, за да може да се създаде индивидуален профил на обучението. В международното право спрямо данните, свързани с деца и студенти, се прилага особена защита, доколкото те са уязвима група. Различна е ситуацията в Китай, където в средно училище в град Ханджоу е приложена система „Умни очи“ за разпознаване на лица – три камери над дъската следят класа. Компютърът разпознава седем емоционални състояния (щастие, тъга, разочарование, упла-

ха). Ако изражението на ученика се промени, системата го тълкува като отслабване на вниманието и моментално подава сигнал на учителя. Степента на концентрация влияе върху рейтинга на ученика⁵².

Проблемите идват от това, че пълната информация за ученика или студента дава възможност за обезпечаване на съдържанието и формата на неговото обучение. От друга страна, в институцията се натрупват обеми от чувствителна информация за него (когнитивни способности, отношение към обучението, социални връзки, успеваемост, възможности за реализация). Това води до парадокса дали такива данни трябва да са достъпни за институцията, или персонализацията на обучението да стане невъзможна.

Тук може да бъде приложен свободният избор на самия обучаем в зависимост от неговите приоритети в обучението.

Предикатната аналитика в образованието предполага създаването на персонализирана образователна траектория, препоръчителната специализация и кариера за конкретния обучаем и всичко това е свързано с някакви етични проблеми. От една страна, проблем е, че такива изследвания са твърде пролонгирани, за да бъде отчетено индивидуалното израстване, свързано с етапи на развитие и застой, и за да бъдат достоверни и ефективни. От друга, такава аналитика е свързана с „наджинг“ поради факта, че алгоритмите и цифровите технологии подтикват ползвателя към определено поведение, потискайки свободната му воля – могат пряко да повлияят на самооценката и мотивацията, да подтиснат волевите качества, карайки го да се чувства като „марионетка“, използвайки указанията на аналитичната система.

Възникват чисто етични проблеми с несанкционираното използване на данни, където върви „продажба на мозъци“ от

⁵² Connor N. Chinese school uses facial recognition to monitor student attention in class // The Telegraph.

URL: <https://www.telegraph.co.uk/news/2018/05/17/chinese-school-uses-facial-recognition-monitor-student-attention/>

притежатели на масиви с такава информация за обучаеми и набавянето на допълнителни доходи от това.

Предикатната аналитика не трябва да стигматизира тези, които не са показали високи резултати – човек постоянно търпи развитие.

Ако трябва да обобщим, то етичната проблематика в образованието е свързана преди всичко със събирането на данни, загубата на приватност и видеонаблюдение. Особеният акцент е свързан с това, че става въпрос за уязвимост на децата и младежите, на които се гради бъдещето на всяка нация.

ЕТИКА НА ЦИФРОВАТА МЕДИЦИНА

Използването на голяма база данни и технологиите на ИИ дават възможност да се издигне диагностиката, лечението и профилактиката на ново по-високо ниво. Стои обаче етичният въпрос доколко може да се използва информацията за здравето на хората за обучение на ИИ – ограничаването в тази област забавя развитието на технологиите, а обратното води до дискриминация и нарушаване на личните свободи.

В световен мащаб медицински данни се събират от столетия. Тези данни се събират обикновено когато се осъществява някаква форма на лечение или профилактика. Тези два вида информация имат различен правов статус и възможности относно етиката.

Биомедицинските данни в повече от случаите са лекарска тайна, обвързана медицински, правно и социално-етично – съществува забрана за медицинските работници да предоставят на трети лица информация за здравето на пациента. Лекарската тайна е фундаментът на професионалната етика в медицината, защитен със закон.

Източник на данни са: електронните медицински картони, мобилните здравни приложения, датчици и устройства за мониторинг, лабораторни изследвания и рентгенови снимки,

медицински изследвания, проведени с група от пациенти, рецепти за закупени лекарства, данни от социалните мрежи, свързани със здравеопазването и допитвания.

Използването на ИИ, създаден да анализира медицински данни, променя живота на много хора, повишавайки качеството на диагностиката, персонализацията на лечението, вземането на лекарско решение, превенцията и ранната профилактика. Това, естествено, изисква максимално голямо количество информация, което поражда етична уязвимост.

В настоящия момент такъв тип информация се съхранява на хартиени носители, които са в различна степен сегментирани в различни здравни институции. Това затруднява обмена, обработката и интерпретацията дори в развити страни като САЩ и Великобритания, където има високо ниво на технологична медицинска статистика. Използването на медицински данни е възможно, когато има качествено подбрани такива, връзка между отделните системи, унификация на данните и изработени етични норми по отношение използването на персоналните данни. Обобщено, етичната дилема е как да бъде осигурена прословутата приватност на пациентите – технологичната достъпност сама по себе си поражда някакво разслоение и категоризиране по някакъв признак на потребителите, независимо от това дали има някаква персонализация (примерно потребители на определен тип лекарства). Ако се получи достъп до чувствителна информация за физическо и психично състояние, склонност към суицид и т.н., това неминуемо води до поява на дискриминация при кандидатстване за работа и неравнопоставеност при получаването на медицинска грижа и застраховане.

Изследователи използват ИИ, за да могат да прогнозират здравни проблеми (сърдечни заболявания, инсулти, слаби когнитивни функции, суициден риск). Фейсбук е внедрил алгоритъм, преценяващ суицидния риск на основа проследяване

на лични постове и определени фрази в тях („всичко ли е наред с теб“, „Сбогом“, „Моля те, не прави това“)⁵³.

През целия си живот човек оставя безброй цифрови следи и това дава възможност за развитието на редица технологии, но поставя под въпрос запазването на дискретността на личния му живот. Парадоксално е, че в сегашно време мобилните оператори знаят повече за медицинския статус на даден човек отколкото здравните институции. Данните, които влизат като „медицинската тайна“, представляват интерес както за комерсиални формации, така и за държавата.

В Китай е създадено единно хранилище за този тип данни под формата на медицински облак, даващ възможност за обединяване на данните и облекчаване на работата с тях. Правото да събира, обменя и използва данните, като съответно и управлява правно информацията в облака, е делегирана на аутсорсинг между държавата и частни фирми. Благодарение на това изследователи и разработчици на ИИ имат достъп до огромен обем данни, помагащи им да развиват технологията.

Развитието на цифровата медицина поставя пред държавата етичната дилема да запази ли правото на пациента да владее информацията за здравето си, или да създаде законодателство, разрешаващо обработката на неперсонифицирани данни без съгласието на пациента, нещо, което е в полза както на индивида, така и на обществото.

РЕГУЛАЦИЯ И МОДЕЛИ ЗА РАЗВИТИЕ НА ЦИФРОВАТА МЕДИЦИНА

Маркираме 10 принципа за използването на технологии, основани на данни:

⁵³ Marks M. Suicide prediction technology is revolutionary. It badly needs oversight // The Washington Post. URL: https://www.washingtonpost.com/outlook/suicide-prediction-technology-is-revolutionary-it-badly-needs-oversight/2018/12/20/214d2532-fd6b-11e8-ad40-cdfd0e0dd65a_story.html

1. Познаване на ползвателите и техните потребности чрез изучаване на условията, в които ще бъдат използвани алгоритмите.
2. Формулировка на резултата в зависимост от използваната за неговото постигане технология.
3. Използване на данните според принципа на съответствие с целта, за която те биват събирани.
4. Вземане под внимание на етичната компонента, прозрачност и отговорност при използването на данните.
5. Прилагане на открити стандарти.
6. Прозрачност на алгоритмите.
7. Формулиране на типа алгоритми, които се разработват, и как се използват данните, като се проверява ефективността и интегритетът към системата за здравеопазване.
8. Ефективността на данните предполага съразмерност на цена и качество.
9. Обезпечаване на безопасност на данните като част от разработката.
10. Създаване на комерсиална стратегия и структура като организация.⁵⁴

СОЦИАЛНАТА ПОДКРЕПА И ЦИФРОВИЗАЦИЯТА

Според данни на ООН лидер по отношение на цифровизацията от 2016 г. насам е Великобритания. Правителството на тази страна взема за генерален приоритет цифровизирането на държавните услуги. Департаментът на труда и пенсиите (Department for Work and Pensions) ускорено разработва система за интелектуална автоматизация на социалната сфера. Разработват се социални роботи. Този департамент тества ИИ, натоварен със задачата да оценява достоверността на сведенията за доходите, цената на жилищата и кандидатите за детски надбавки. Институцията залага на това, че автома-

⁵⁴ Code of conduct for data-driven health and care technology // Department of Health & Social Care.

URL: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology>

тизацията поевтинява процесите на изплащане и ги прави по-бързи и точни. Експертите предупреждават обаче, че използването на ИИ и автоматизираното определяне на благосъстоянието предизвиква редица етични и технологични въпроси. Адвокатът по правата на човека Филип Алстон представя пред Генералната асамблея на ООН доклад за нарушаването на правата на човека в процеса на тотална цифровизация на системите за социална защита ⁵⁵.

Журналист на The Guardian в серия статии „Автоматизация на бедността“⁵⁶ описва как цифровите иновации се превръщат в заплаха за незащитени хора и предупреждава, че след тяхното внедряване бедните стават още по-бедни⁵⁷ (тревожната тенденция е породена от системата за автоматизирано изискване на дълговете, прилагана в САЩ и Великобритания без оглед на давност).

Според експерти в случаите, когато заявителят се обръща към органите за социално подпомагане през цифрови канали, обратна връзка практически е невъзможна (оспорване на решения и извършване на изплащания). Налице е реален риск правото да е приоритетно поради зависимостта си от технологиите. Внедряването на цифровите сървъри има за задача да оптимизира социалната подкрепа и да улесни гражданите и институциите. Като цяло цифровите платформи правят достъпът по-лесен и носят облаги за социално защитените граждани, но практиката на някои държави показва, че цифровизацията води и до появата на нови проблеми.

⁵⁵ Alston, P. How Britain's welfarestate has been taken over by shadowy tech consultantsцж // The Guardian. URL: <https://www.theguardian.com/commentisfree/2019/jun/27/britain-welfare-state-shadowy-tech-consultants-universal-credit>

⁵⁶ Pilkington E. Digital dystopia: Algorithms punish the poor // The Guardian. URL: <https://www.theguardian.com/technology/series/automating-poverty>

⁵⁷ Alston, P. How Britain's welfarestate has been taken over by shadowy tech consultants // The Guardian.

URL: <https://www.theguardian.com/commentisfree/2019/jun/27/britain-welfare-state-shadowy-tech-consultants-universal-credit>

ЕТИЧНИ ПРОБЛЕМИ НА ЦИФРОВИТЕ ТЕХНОЛОГИИ В РАБОТАТА НА ПОЛИЦИЯТА

В областта на съдебната и полицейска дейност използването на цифрови технологии е на сравнително ниско ниво, но етичните проблеми са не по-малко актуални и важни. Използването на тези технологии в полицейската работа създава нови възможности за постигане на висока сигурност на гражданите, но и създава заплахи за нарушаване на гражданските права и свободи.

Основната задача на тази система и технологиите, които тя използва, са с единствената задача да се противопостави на престъпността, която също се модернизира и технократизира. Подземният свят все по-често използва технологиите на блокчейн и дроне, алгоритми на ИИ, правещи мащабите на престъпленията все по-големи и тежки. Същевременно към полицията нараства претенцията към прозрачност на дейността и преодоляване на корупцията.

Появиха се т.нар. киберпрестъпления, при които полицията винаги трябва да бъде на крачка пред закононарушителите, за да може да е ефективна. Преимущество е възможността тази система да натрупва огромно количество от данни, които не се обработват и използват – натрупва се предикатна информация в полицейската дейност.

Предикатната полицейска дейност е стратегия за предсказване и предотвратяване на рисковете за престъпление, базирани на ИИ и анализ на голямата база данни.

Основните етични проблеми възникват в областта социалните технологии и събирането на данни и вземането на решения за гражданите, за които ИИ предсказва висока вероятност да извършат престъпления. По данни на Интерпол и Европол в 70 страни практически се използва една или друга форма на предикатна аналитика.

По поръчка на ЕС международен екип създаде през 2013 г. системата **ePOLICE** (early Pursuit against Organized

crimeusing enviro nmentalscanning, the Law and IntelligenCE systems). Тя сканира сайтове, електронни писки и полицейска информация с цел да открива дейност на организираната престъпност и да оцени риска от извършване на престъпления. Използват се видео, текстови контент, финансови данни, информация от социалните мрежи и чатове⁵⁸.

Етичните проблеми възникват въз основа на използването както на цифровите технологии в полицейската работа поради несъвършенството им, така и поради високата степен на участие на човешкия фактор. В социологичните изследвания се налага мнението, че всички данни, събрани в процеса на полицейската работа, са изложени на предубеденост. Самата полицейска статистика представя данните, които реално са били разкрити от полицията и са известни на държавата, като характеризира конкретни общности (стратификация, интензивност и близко взаимодействие), което дава тенденциозна и неправилна прогноза примерно по отношение на някои малцинства. Опасността от такива прогнози е в това, че те легитимират информацията поради факта, че се приема безпогрешността на технологиите и се пренебрегва човешкият фактор.

През 2014 г. в Масачузетския технологичен институт е разработена програмата COMPAS с елементи на ИИ. Тя има задача да помага на съдиите при вземането на решения относно предсрочното освобождаване на осъдени. Програмата успешно работи до момента, в който става ясно, че тя априорно намалява шанса на нелегално пребиваващите латиноамериканци и афроамериканците с ниски доходи. Направеният дълбок анализ на системата показва пълната статистическа обоснованост на предикатните решения поради това, че действително тези две етнически групи най-често нарушават правилата за условното освобождаване. От гледна точка на обществото това е недопустимо и в крайна сметка системата

⁵⁸ Early Pursuit against Organized crime using enviro nmental scanning, the Law and IntelligenCE systems // European Commission. URL: <https://cordis.europa.eu/project/rcn/106659/factsheet/en>

бива изключена поради явната стигматизация, която се налага в случая.

Стигматизацията е налагане на някакви знаци относно отрицателно качество върху даден човек или група от хора, което не се потвърждава от действителността. Тя е съществена част от повечето стереотипи.

През 2014 г. в град Фъргюсън, Сащ, полицаи застрелва афроамериканеца Майкъл Браун. Според широкото обществено мнение инцидентът се е случил поради факта, че убитият е бил чернокож от неблагоприятен район. Според защитници на гражданските права системата взема решения на база регионалните данни (плътност на населението, социален състав, брой питейни заведения, църкви, училища, транспортни възли), създаващи риска от предубеденост спрямо неблагоприятните райони, което силно нагнетява отношенията между полицията и афроамериканските общности.⁵⁹

Нерегламентираното използване на публични данни като етичен проблем предизвиква обществен резонанс особено ако полицията е използвала свои източници на данни. Масовото монтиране на камери позволява създаването на системи на ИИ с алгоритми за анализ на видеоматериал.

Профилите в социалните мрежи са друг източник за публична информация. Полицията използва специално разработени аналитични програми, провеждащи автоматично слеждане в социалните мрежи и правещи семантичен анализ на съобщенията. Често за целите на полицейското разследване се използва информация за гражданите от частни организации (телекомуникационни компании, банки, сървъри на таксиметрови фирми и др.). Реално те знаят много повече за гражданите отколкото държавата.

През 2019 г. в щата Аризона в хода на разследване е арестуван човек заради това, че в момента на убийство сигнала на телефона му е бил в най-голяма близост до местопребыванието. Информацията е предоставена от Google.

⁵⁹ Chammah M. Policing the Future // The Marshall Project.

URL: <https://www.themarshallproject.org/2016/02/03/policing-the-future>

Арестуваният прекарва една седмица зад решетките до момента, в който следователите не намират друг заподозрян.⁶⁰

Практиката да се търси информация от Google за местоположението на ползватели е използвана за първи път през 2016 г. в Северна Каролина и впоследствие се прилага във всички американски щати. Базата данни Sensorvault пази подробни записи за местоположението на стотици милиони устройства по целия свят за последните десет години. Тази техническа обслужваща система на Google се превръща в система за полицейско разследване и това предизвиква силна обществена критика. Професорът по право на Университета в Южна Калифорния, Орин Кер, показва, че кодифицирането на невинни хора на база цифрови технологии е един нов правен проблем.

При обобщаването на всичко казано до момента се налагат следните актуални въпроси:

Доколко внедряването на ИИ и другите цифрови технологии влияе върху числеността на полицията? Видимо цифровизацията намалява щата на полицията, както и в банковия сектор. Разширяването и усложняването на системите за видеонаблюдение на полицията дава възможност за получаване на повече информация за противоправните действия. Налага се увеличаване на щата на оперативните работници - следователи и експерти, занимаващи се с аналитика на голямата база данни.

Различия в заложените етични норми за използването на ИИ между съдебна и полицейска система. Според редица наскоро приети документи, регламентиращи използването на ИИ, вече се прилагат едни и същи регламентиращи препоръки, но в крайна сметка тълкуванието им от полицията и съда се разминава. Принципът за прозрачност на съдебната база данни е неприемлив за следствената работа, при която оперативно-следствените мероприятия са силно затъмнени. Сте-

⁶⁰ Valentino-DeVries J. Tracking Phones, Google Is a Dragnet for the Police // The New York Times. URL: <https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html>

пента на прозрачност зависи от нивото на секретност и възможните негативни последствия за източниците на информация, потърпевшите, заподозрените и обвиняемите.

Критерии, използвани от полицията при събирането на информация относно гражданите с помощта на ИИ

Тук от особено значение е консенсусът между полицията и обществото относно използването на такива данни. Ако се отнася до предотвратяването и разкриването на терористични актове, корупция и друга престъпност, консенсусът е пълен, но когато става въпрос за прогнозна информация относно престъпна дейност по криминалистични критерии, водещи до ограничаването на граждански права и създаване на рейтингова социална система, то за консенсус не може да става и дума.

За да се минимизират етичните проблеми при използването на цифрови данни от полицията, са възможни следните походи:

- обезпечаване на прозрачност на ИИ;
- използване на цифровите технологии за парире на предразсъдъци и нагласи;
- вземане под внимание на интересите на всички заинтересовани страни в досег с цифровите технологии и най-вече на държавата, гражданското общество и обикновените граждани;
- нормативно регулиране на използването на цифровите технологии от полицията и съда.

Европейският съюз демонстрира в момента една от най-балансираните за момента етични системи, отчитаща интересите на всички страни. Глобалният иновационен център на Интерпол (Interpol Global Complex for Innovation“ и Между-регионалният научноизследователски институт към ООН по въпросите на престъпността и правосъдието (United Nations Interregional Crime and Justice Research Institute) показват справедливия баланс между осигуряването на безопасността

и защитата на личното пространство и конфиденциалността като етичен проблем.

Като принцип се оформя тенденцията полицията да използва цифровите технологии под надзора на държавата и обществените организации. Гражданите трябва да имат възможност във всеки един момент да контролират опазването и използването на личните им данни (нещо, което в настоящия момент не изпълнява нито една страна в пълен обхват поради липса на нормативна уредба и равен достъп на населението до информацията).

Такива документи са разработени в известна степен, като са взети предвид най-вече Европейската етична харта за използването на изкуствения интелект в съдебната и правозащитната система (European Ethical Charter on the use of artificial intelligence in judicial systems) и Европейската комисия за ефективността на правосъдието (The European Commission for the Efficiency of Justice). В тези две институции широко се залага на използването на ИИ в съдебната и правна система, като се акцентира върху:

- Зачитане правата на човека при създаването и използването на технологии и инструментариума на ИИ.
- Недопускане на дискриминация спрямо етнически и социални обществени групи и слоеве при обработката на голяма база данни.
- Обезпечаване качеството и безопасността на алгоритмите при обработката на данни чрез контрол над източниците и съдържанието, базиране на математическите модели на междудисциплинарна основа, вземане под внимание не само на чисто статистическата корелация, а и на социалната, културната и икономическата като фактори, оказващи влияние върху резултата.
- Осигуряване на качество и разбираемост на данните, като данните в съдебната и правозащитна дейност трябва да са в достатъчна степен прозрачни.

ЕТИКА НА ПРИЛАГАНЕТО НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ ЗА ВОЕННИ ЦЕЛИ

Оръжията както и другите глобални проблеми, свързани с изхранването и здравното обезпечаване на населението, са един от най-разискваните проблеми на съвременното.

Според мнозина наличието на ядрено оръжие е достатъчно въздържащ фактор за предотвратяване на сериозен глобален военен конфликт. Появява се обаче изкушението да се създават все по-страшни оръжия на основата на ИИ със схващането, че това ще заличи завинаги идеята за войната. Ако на бойното поле се сражават само машини и не загиват хора, това не прави ли колизията на войната по-хуманна и поносима? Създаването на автономни бойни дроне с изкуствен интелект, гарантиращ справедливост и рационалност по-висока от тази на човека, липсата на страх от смъртта и категорична забрана за убийство на граждански лица изглеждат приемливи за военното дело предвид неизбежността на войната в цялата човешка история.

Тук отново се появява неизбежният въпрос за контрола на автоматизираните системи.

Американската система Phalanx е монтирана на военен крайцер от типа Aegis със задача автоматично да намира и проследява приближаващите цели на противника. USS Vincennes е точно такъв ракетен крайцер, наричан „робокрайцер“ заради монтираната на него система Aegis. На 3 юли 1988 г. в разгара на престрелка между ирански и иракски бойни катери по време на конфликта между двете държави радиолокационните системи на американския военен кораб дават предупреждение за приближаващ самолет. Капитан Уилям Роджърс-трети възприема, че са атакувани от ирански изстребител F-14 и дава команда на системата Aegis да открие огън. Оказва се, че сваленият самолет е на Иранската гражданска авиация, извършващ рейс 655 с 299 пасажери на борда, които загиват, и е предизвикан грандиозен международен скандал. Разследването доказва, че в ав-

томатичен режим разпознавателният интерфейс не показва кой от маркираните самолети на радара е граждански и кой военен. Системата маркира поведението на самолета с оглед дали лети ниско като бойните или във височина като гражданските самолети. В случая самолетът се е издигал от летище Техеран и е бил погрешно идентифициран като военен. Подобен казус съществува и с използването на бойни дроне, при които крайното решение се взема от оператора, но информацията, която идва, не може да бъде анализирана толкова бързо и той се доверява на машината (става въпрос за MQ-1 Predator на въоръжение в Американската армия). Има случаи, в които светът е висял на косъм от пълно унищожение, при които машините дават показания за дадено събитие, но човек взема различно от тяхното решение, пренебрегвайки по морални причини информацията. Такъв е случаят с руската ядрена подводница Б-29 на 27 октомври 1962 г. по време на Кубинската криза. Подката няма връзка със своето командване, температурата на борда е 45 градуса и повечето членове на екипажа са в безсъзнание от въглеродна интоксикация. Подводницата е атакувана от американски самолетносащ, а на борда ѝ има атомно торпедо, което може да бъде използвано без разрешение от Москва. Командирът капитан Савицкий и командирът на торпедния отсек вземат решение да го използват, но за щастие, се изисква съгласието на третия офицер Валерий Архипов, който трезво разсъждава за последствията и отказва, като предотвратява трета световна война.

Тези и много други случаи предизвикват сериозни тревоги от създаването на автономни бойни оръжия. Изкуственият интелект е третата технологична революция след създаването на барута, ядрените бомби във военното дело. Излизането на бойното поле само на машини, от една страна, намалява човешките жертви, но от друга, намалява и прага на търпимост при въздържане от военни действия. Компонентите за създаването на такива оръжия са достъпни, почти всяка държава може да си ги позволи, включително и терористични органи-

зации, диктатори и други сили, стремящи се да наложат свой контрол. Такива оръжия са идеални за политически убийства и етнически чистки. Военните бюджети на САЩ, Китай и Русия в посока създаване на такива системи многократно надвишават тези за гражданските проекти, свързани с ИИ.

Въпросът е как и до колко трябва да бъде ограничено създаването на такива системи - дали това да касае само нападателните оръжия и да не засяга отбранителните (което е съвсем условно чисто технически) и до колко системите могат да имат двойно използване (дронове, които пренасят стоки на Amazon, са същите, с които носят бомби). Съвсем критичен е казусът с воденето на кибервойна, не изискваща създаването на каквито и да е бойни системи въпреки огромната им разрушителна сила.

Със страшна сила стои и въпросът дали ако се създаде някакъв технически паритет между големите сили по отношение на използването на ИИ за военни цели, от това няма да се възползват в крайна сметка потребителите на черния пазар, преследващи лични интереси.

В настоящия момент международното право, фиксирано в международни конвенции, изисква строго спазване на принципа крайните решения от военно приложение на ИИ да се вземат задължително от човек и отговорността за последствията да е задължително негова.

СЕДМА ТЕМА. РЕГУЛИРАНЕ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ В СВЕТОВНАТА ПРАКТИКА

В тази глава се прави общ разбор на документите на ООН и Евросъюза, както и на някои страни, като САЩ и Китай, където са засегнати въпросите относно регулацията на цифровите технологии и тяхното използване със задачата да се защитят правата на хората и да се повиши качеството на техния живот. Акцентирано е върху наличието/отсъствието на етични правила и принципи

в тези документи. Технологичният ръст на системите на ИИ повдига въпроса за необходимостта от приемането на законодателни актове в областта. В различните страни това се случва по различен начин в зависимост от особеностите на местната правова система. До 2020 г. в различните страни има натрупан опит, свързан с развитието на ИИ, под формата на национални стратегии и в конкретни сфери. В тази глава се описват основните видове документи, посветени на правните аспекти на ИИ.

НАЦИОНАЛНИ ДОКУМЕНТИ И СТРАТЕГИИ ЗА РАЗВИТИЕ

Според различни оценки в момента национални стратегии има разработени в повече от 30 страни, между които са Китай, Корея, Канада, САЩ, Великобритания, Франция, Русия. По правило в такива стратегии се представят подходите за развитието на технологиите, свързани с ИИ:

- представяне на нивото, на което се намира развитието на тези технологии в световен мащаб, ключовите им отрасли и внедряването им;
- перспективите за тях в краткосрочен, средносрочен и дългосрочен план.
- ключови етапи и цели на развитие на технологиите на ИИ в конкретната страна;
- основни проблеми и препятствия в развитието им;
- планове за развитие;
- финансова стратегия за развитие на отрасъла;
- насоки за адаптация на правната система, свързана с технологията;
- основни етични проблеми;
- целеви насоки за развитието на технологията.

Такива национални документи са:

- Националната стратегия за развитие на ИИ „France IA”⁶¹ (Франция 2018);
- Общоканадска стратегия за изкуствения интелект (Канада 2018)⁶²;
- План за развитието на технологиите на изкуствения интелект от ново поколение (Китай, 2017)⁶³;
- Национална стратегия за изкуствения интелект (Дания 2019).⁶⁴

ЗАКОНОВИ И ПОДЗАКОНОВИ АКТОВЕ

В световен мащаб все още не създаден нормативен акт за конкретни действия относно ИИ. До подобен процес най-близо са ЕС и Южна Корея (2008 г.).

През февруари 2017 г. Парламентът на ЕС прие резолюция 2015/2103 (INL) Civil Law Rules on Robotics⁶⁵. Този документ е свързан основно с развитието на роботехниката, но свързаността на технологиите контекстуално предполага и ИИ. Там се предвиждат мерки, регулиращи въпросите, свързани с отговорността при причиняване на вреда и създаването на система за регистрация на „умни“ машини.

В други държави има конкретни актове, посветени на различни системи ИИ с използвани на ИИ, прилагани в авто-

⁶¹ AI for Humanity. French strategy for artificial intelligence. URL: <https://www.aiforhumanity.fr/en/>

⁶² Pan-Canadian Artificial Intelligence Strategy // CIFAR. URL: <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>

⁶³ 国务院关于印发新一代人工智能发展规划的通知/国务院. URL: www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm. Перевод от англ.: Notice of the State Council Issuing the New Generation of Artificial Intelligence Development Plan // Foundation for Law and International Affairs. URL: <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf>

⁶⁴ National Strategy for Artificial Intelligence // Ministry of Finance and Ministry of Industry, Business and Financial Affairs. URL: https://eng.em.dk/media/13081/305755-gb-version_4k.pdf

⁶⁵ Civil Law Rules on Robotics. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) // Official Journal of the European Union. URL: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

матизирани автомобили, здравеопазването, умните градове и домове и финансовата сфера. Вниманието заслужават:

- Промените в Закона за движението в Германия, засягащ автоматизирани автомобили (2017 г.)
- Ръководство за изпитанията на автоматизирани транспортни средства (Австралия, 2017 г.).
- Резолюция, забраняваща използването на смъртоносни автономни системи оръжия (Белгия, 2018 г.).
- Директиви за автоматизирано вземане на решения във федералните органи на управление (Канада, 2019 г.).

ИЗСЛЕДВАНИЯ ВЪРХУ ЕТИКАТА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Подходите за регулирането на ИИ се разработват от експертни групи и държавни органи. Такива са например Препоръките на специализираната Комисията по етика в Министерството на транспорта и цифровата инфраструктура на Германия (2017) – разискват се препоръките към безпилотните технологии в транспорта. Пример за наднационално ниво е докладът на експертите на Световната комисия по етика на научното знание и технологии при ЮНЕСКО (доклад за етиката в роботехниката от 2017 г. и за етиката на ИИ от 2019 г.)

Интерес представляват – Бялата книга по стандартизация на ИИ на Китай от 2018 г.; Доклада пред палатата на лордовете „Алгоритми в обществения бизнес и вземането на решения“ и „Свободния отчет на Комитета по изкуствен интелект при палатата на лордовете от 2018 г.

ЕТИЧНИ ДОКУМЕНТИ В ОБЛАСТТА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

В края на 2019 г. в целия свят са приети стотици различни актове, ръководства, принципи и кодекси, свързани с етиката на ИИ. Те имат няколко общи принципа, свързани с безопасността, конфиденциалността, антидискриминацията,

контролът и т.н. Най-ярък пример са „Асиломарските принципи на ИИ“⁶⁶ от 2017 г.

Подобни документи са: „Монреалската декларация за отговорността за развитието на изкуствения интелект“ от 2017 г.; „Ръководството по етика за надежден ИИ на Специалната експертна група на високо ниво към СЕ от 2018 г.

СТАНДАРТИ И ДОКТРИНАЛНИ ИЗТОЧНИЦИ

Към Международната организация по стандартизация (ISO) е създадена специална техническа комисия за изкуствения интелект (SC №42). Комитетът е планирал да приеме двайсет стандарта, посветени на ИИ и голямата база данни.

В началото на 2020 г. са разработени стандарти ISO/IEC, регулиращи работата с голямата база данни⁶⁷ и стандарти по ISO за работата с ИИ, включващи и мерките за преднамереност.

Други примери за стандартизация са:

- Глобалната инициатива за етика на автономните и интелектуални системи (IEEE, 2016).
- Препоръчителни практики по управление на качеството на база данни за медицинския изкуствен интелект (IEEE, процес на разработване).
- Проектоплан за федералното участие в разработването на технически стандарти за ИИ и свързаните с тях инструменти (NIST, САЩ, 2019).
- Доклад „Разработване на стандарти за изкуствен интелект: Чуйте гласа на Австралия“ (Австралия, 2019).

⁶⁶ Asilomar AI Principles // Future of Life. URL: <https://futureoflife.org/ai-principles/?cn-reloaded=1>

⁶⁷ ISO/IEC 20546:2019(en) Information technology — Big data — Overview and vocabulary // ISO. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en>; ISO/IEC TR 20547-2:2018 [ISO/IEC TR 20547-2:2018] Information technology— Big data reference architecture — Part 2: Use cases and derived requirements // ISO. URL: <https://www.iso.org/standard/71276.html>; ISO/IEC TR 20547-5:2018. Information technology — Big data reference architecture — Part 5: Standards roadmap. URL: <https://www.iso.org/standard/72826.html>

В чуждестранната юридическа литература систематично се формира „роботно право“ като самостоятелна научна изследователска област, където се разглеждат проблемите за отговорността, субектното право, прозрачност на алгоритмите, контролът над системите за ИИ, проблемите за авторското и патентно право и чуждото мнение.

МЕЖДУНАРОДНИ АКТОВЕ ОТНОСНО ЕТИКАТА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Докладът за етиката на ИИ от 2019 г. на Световната комисия по етика на научните знания и технологии при ЮНЕСКО представя препоръки относно структурата и съдържанието на такива международни документи. Организацията може да допълни множеството ръководни принципи относно етиката, разработващи се в държавни органи, фирми и обществени организации, междудисциплинарни, универсални и цялостни подходи по отношение на ИИ в името на човечеството, мира и устойчивото развитие.

В настоящия момент лидер при определянето на подходите за регулиране на ИИ е Съветът на Европа. Само през 2019 г. благодарение на труда на различните експерти и комитети са разработени: Европейска етическа харта на Съвета на Европа по използването на ИИ в съдебната система⁶⁸, Ръководство по защита на данните при използване на ИИ⁶⁹, Декларация на Комитета на министрите в Съвета на Европа по правата на човека⁷⁰, 10 стъпки за защита правата на човека при използването на ИИ⁷¹. За да се създаде цялостен норма-

⁶⁸ Report of COMEST on robotics ethics // UNESDOC.
URL: https://unesdoc.unesco.org/ark:/48223/pf0000253952_eng

⁶⁹ Guidelines on Artificial Intelligence and Data Protection // Council of Europe.
URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>

⁷⁰ Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes // Council
Of Europe. URL:
https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

⁷¹ Unboxing Artificial Intelligence: 10 steps to protect Human Rights. Commissioner for Human Rights Recommendations // Council of Europe. URL:

тивен документ за областта на ИИ, през 2019 г. Съветът на Европа сформира Специална комисия за ИИ. Документи в същата област са:

- Принципи на ИИ и препоръки по националната политика на Експертната група по ИИ (ОЕСР, 2019)
- Проект за препоръки към Съвета на министрите относно влиянието на алгоритмите върху правата на човека (Съвет на Европа, 2019 г.)⁷²

През последното десетилетие особено внимание се отделя на цифровите технологии и тяхното развитие, влияещи върху различните аспекти на човешкия живот (повишаване на благосъстоянието и борбата с престъпността). Тази проблематика е фиксирана в документи на ООН, Съвета на Европа, Организацията за икономическо сътрудничество и развитие и други правни документи на някои страни, в които технологиите са в достатъчна степен напреднали.

Международни организации и специализирани агенции разглеждат редица етични въпроси, породени от развитието на технологиите. Те са структурирани в три базови контекста:

- връзката между технологиите и устойчивото развитие;
- влиянието на технологиите и реализацията на фундаменталните човешки права;
- достъпност на технологиите и проблемите на цифровото неравенство.

Естествено, това не е изчерпателният списък, но тези фундаментални етични тематика са ключови.

<https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>

⁷² Addressing the impacts of Algorithms on Human Rights. Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems // Council of Europe. URL: <https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf>

УСТОЙЧИВО РАЗВИТИЕ И ЦИФРОВИ ТЕХНОЛОГИИ

Международното сътрудничество в сферата на устойчивото развитие е крайно необходимо при регулираното му внедряването и разпространението на иновативните технологии. Усилията на отделните държави не са достатъчни, за да се осигури ефективност, справедливост и спазване на етичните норми.

Един от контекстите на технологичното развитие е осигуряването на устойчиво развитие на съвременното поколение изкуствен интелект, което да не вреди на следващите поколения.

Като приоритет на устойчивото развитие до 2030 г. в документа, наречен „Задачи на деня за устойчивото развитие“ от страна на световните лидери, заседаващи в рамките на 2015 година, се отбелязва стремежът на държавите да предприемат мерки за ликвидиране на бедността и да осигурят развитие и премахване на неравенствата в световен мащаб. В „Задачи на деня...“ се отбелязва, че „разпространението на интелигентните компютърни технологии и глобалната мрежа създава огромна възможност за ускорен прогрес, свързан с преодоляването на цифровото разделение и формирането на общество на основите на знанието“. Отбелязва се, че развитието на цифровите технологии и ИИ е като безпилотен самолет – дава се възможност да се обработят огромни масиви от данни, триизмерно принтиране и всички други съвременни нововъведения имат потенциала да решат най-съществените проблеми на човечеството.

Естествено, освен възможности цифровите технологии създават и ред проблеми и рискове от етичен характер:

Възможности	Цел в областта на устойчивото развитие	Рискове
Персонализирано лечение, ранна диагностика на заболявания,	Добър здравен статус и благополучие	Нарушена конфиденциалност и безопасност на здравните

Възможности	Цел в областта на устойчивото развитие	Рискове
оперативна намеса при акутност на заболяването		данни и рискове от разкриване при машинна обработка на данните, използване на мобилни устройства за проследяване на състоянието с различно ниво на достъп.
Повишаване на достъпността и качеството на образованието, свързано с неговото персонализиране през цифрови платформи	Качество на образованието	Неподготвеност на популацията за цифрово образование, непълна подготвеност на преподавателите за него и адаптация на цифровите технологии към съществуващите форми на обучение
Повишаване на производителността за сметка на цифровизацията и автоматизацията, нови професии и форми на доходи, достъп до пазари и възможности най-вече за развиващите се страни.	Квалитетна работа и икономическо развитие	Загуба на работни места вследствие на автоматизацията и внедряването на 3D-принтирането, поляризация на труда, цифрови платформи за трудова дейност, които не са обезпечени с адекватно заплащане и защита на работника
Ориентация към нискомощностни инклузивни икономически иновации, разширен достъп до финансови услуги, разширени комуникациите в организациите, достъпност до	Намаляване на неравенството	Риск от цифрово разделение между различни групи от населението и различните страни, развитие на ИИ и машинния анализ на голяма база данни създаващи риск

Възможности	Цел в областта на устойчивото развитие	Рискове
властта и държавните услуги		от дискриминация заради некоректни данни или преднамереност на алгоритмите

Както е показано в документите на ООН относно устойчивото развитие, е необходим разностранен подход, свързан с:

- адаптация на правнонормативната база;
- институционално прогнозиране на развитието на технологиите и икономическите и социални последици, свързани с тези процеси;
- превантивно разглеждане на потенциалните етични проблеми и социални последици, свързани с върховите технологии.

Тук се изисква взаимодействие и координация между държавните органи, бизнеса, гражданското общество и научните среди. Държавите, които са лидери в областта на върховите технологии, трябва да регулират тяхното разпространение и внедряване, за да обезпечат баланса между ефективност, справедливост и спазване на етичните норми. Усилията на национално ниво изглеждат крайно неефективни и недостатъчни за обезпечаване на максимална ефективност от използването на новите технологии, като същевременно се вземат под внимание минимизирането на негативните последици. За тези цели е необходимо международно сътрудничество. На най-високо ниво група от експерти при Генералния секретар на ООН по отношение на цифровите технологии отбелязва основните параметри на подобно сътрудничество:

- инклузивност, предполагаща постигане на равни възможности;
- уважение, разбирано като съблюдаване на международното и национално право, човешките права, достойнство, конфиденциалността на личните данни;

- фокус върху човека, предполагащ максималното използване на технологиите в полза на човека, при запазване на човешкия контрол върху вземането на решения;
- благосъстояние, предполагащо икономически ръст, социални блага и възможност за самореализация;
- прозрачност, разбираана като открит достъп до информация;
- сътрудничество, предполагащо открити стандарти и кооперираност при обезпечаването на взаимодействие;
- достъпност, разбираана като разработване на достъпни, прости и надеждни устройства за различните групи ползватели;
- хармоничност, разбираана като употреба от страна на държавите и корпорациите на технологиите по такъв начин, че да бъде съхранено общественото доверие.

ПРАВА НА ЧОВЕКА И ЦИФРОВИ ТЕХНОЛОГИИ

В тази точка ще бъде обсъдено разпространението на цифровите технологии и възникването на етични въпроси, свързани с реализацията на основните права на човека: уважение към личния живот и пространство (приватност), свобода на словото, забрана на дискриминацията, защита на уязвимите групи. Към това може да се добавят и станалите актуални в последно време права на потребителя спрямо бизнеса и преди всичко data данните и социалните мрежи. В раздела се дадени изчерпателни обяснения как пълноценно да бъдат защитени правата на човека в цифрова среда въпреки липсата на официални документи за това.

Технологичното развитие дава своето отражение и върху международното право и конкретно в областта, касаеща правата на човека. Повечето документи за правата на човека са създадени доста преди мощното развитие в областта на цифровизацията. Повечето фундаментални положения са актуализирани спрямо цифровите технологии, но възникват и нови въпроси, като:

Какви обвързаности и граници на отговорност имат частните фирми спрямо спазването на международното право и конкретно правата на човека – повечето от тях са адресирани към държавата, а не към бизнес субектите?

Как трябва да се адаптира международното право при едни трансгранични предавания на данни и кой носи отговорност за нарушаването на правата на човека в подобен трансфер?

Как различните участници в новите обществени отношения, били те държави, глобални компании или граждански общности от различни култури, социални и икономически условия, могат да работят заедно и с помощта на новите технологии да не поставят под заплаха правата на човека, а да създават възможности за тяхната реализация?

В наши дни международните организации активно разясняват как традиционните норми в областта на човешките права се адаптират към цифровия свят, публикувайки препоръки, ръководни принципи и декларации. Международните организации заявяват готовност за създаването на международен нормативно правен акт, който да действа на наднационално ниво. На първо място в него се взема под внимание защитата на личния живот, свързана със събирането и съхраняването на лични данни.

Все още няма правна концепция или институция в областта на правата на човека, която да е способна или упълномощена да отговаря на етични въпроси, обвързани с цифровите технологии – всичко е в рамките на дискусии как да бъдат защитени ценностите и правата на човека в един постоянно променящ се свят.

ПРАВО НА СЪБЛЮДАВАНЕ И ЗАЩИТА НА ЛИЧНИЯ ЖИВОТ

Защитата на личния живот и конфиденциалността на данните при цифровизацията е един от най-важните въпроси, подлежащ на регулиране от страна на международните организации и държавите. В настоящия момент само един доку-

мент обезпечава тази област и това е Конвенцията на Съвета на Европа за защитата на частните лица по отношение на автоматизираната обработка на данните от личен характер – „Конвенция 108“⁷³. Тя представя ключовите принципи при използването на база данни и гарантира правото на субективни данни. За да се вземат под внимание новите технологични възможности и свързаните с тях рискове, от момента на разработването им като текст са направени още няколко преработки и допълнения. Модернизираната „Конвенция 108+“⁷⁴ е открита за ратифицирана за държавите членки.

В правните документи на ЕС има и друг фундаментален документ – регламента GDPR (документът е утвърден на 18.05.2018). Освен това Съветът на Европа разработва документи с препоръчителен характер за защита на персоналните данни, използвани в различните технологии и отрасли:

- Ръководни принципи за защита на лица подлежащи на обработка на персонални данни в рамките на голямата база данни⁷⁵.
- Ръководни принципи за изкуствения интелект и защитата на данните⁷⁶.
- Препоръки на Комитета на министрите на Съвета на Европа за защитата на здравните данни⁷⁷. С въпросите, свър-

⁷³ Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. ETS No. 108 // Council of Europe.

URL: <https://www.coe.int/ru/web/conventions/full-list/-/conventions/rms/0900001680078b37>

⁷⁴ Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data // Council of Europe. URL: <https://rm.coe.int/16808ac918>.

⁷⁵ Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data. T-PD(2017)01. 23 January 2017 / Directorate General of Human Rights and Rule of Law // Council of Europe. URL: <https://rm.coe.int/16806ebe7a>

⁷⁶ Guidelines on Artificial Intelligence and Data Protection. T-PD(2019)01. 25 January 2019/Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data // Council of Europe. URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8522>

⁷⁷ Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health-related data. 27 March 2019 // APDA.

зани с персоналните данни, се занимава и Организацията за икономическо сътрудничество и развитие. Тя прие Ръководни принципи за защита на конфиденциалността и трансграничното препредаване на персонални данни⁷⁸

Стандартите и нормите, закрепени в указанията на документа, в значителна степен кореспондират помежду си. Техните ключови принципи са:

- Законност, справедливост и прозрачност при събирането и обработката на данни.
- Правомерни цели за събиране и обработване на данни.
- Безопасност и конфиденциалност на данните.
- Принцип на минимизация на обработваните данни.
- Особена защита на отделна категория данни.

СВОБОДА НА СЛОВОТО

Технологиите предлагат нови възможности за реализиране свободата на свободно изразяване на мнение и достъп до информация. На първо място това засяга интернет и новите медийни форми в социалните мрежи. Естествено международната общност забелязва и рисковете свързани с това: безпрецедентната възможност за разпространение на фалшиви новини свободното циркулиране на формите на изразяване на ненавист и призивите за насилие. Подобни проблеми са съществували и преди, но в интернет те стават особено ярки поради възможността за светкавично да се препредават огромни блокове от информация от човек на човек и то съвсем анонимно. Друг важен момент е ограничаването от страна на държавата на достъпа до интернет изцяло или по отношение на дадени сървъри в кризисни ситуации (терористични актове), когато риска от разпространението на дезинформация е особено висок и опасен. В редица случаи подобни мерки са

URL:https://www.apda.ad/sites/default/les/2019-03/CMRec%282019%292E_EN.pdf

⁷⁸ Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, OECD/LEGAL/0188 //OECD.
URL:<https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>

правомерни. Нарушаването на човешките права се случва когато е ограничен достъпа до регламентираната политика на държавата с цел лишаване на хората от информация или ограничаване свободата на изказванията.

Комитетът на министрите при Съвета на Европа излезе с препоръки относно ролята и задълженията на интернет посредниците⁷⁹. В документа се обобщават международните стандарти за правата на човека в ситуации, когато държавата ограничава работата на интернет посредниците (включително блокиране или забраняване на контакти и всякакви други мерки ограничаващи правото на достъп до информация и свободата на словото). Държавата е длъжна:

- Да дава оценка на потенциалното въздействие на планираните мерки върху правата на човека и да прилага такива, които постигат необходимата цел при минимално ограничаване на правата.
- Да обезпечава наличието на процесуални гаранции под формата на съдебни процедури за вземане на решения за ограничаване на достъпа до контакти, средства за правова защита и т.н.
- Да не възлага на интернет посредниците задължение да осъществяват мониторинг на трети лица, които имат платформи, поместват или пазят информация.
- Да създаде мерки за търсене на отговорност спрямо интернет посредниците със съответната пропорционалност, тъй като прекомерно строгите мерки принуждават интернет посредниците сами да ограничават достъпа до контакти.
- Да не възлага на интернет посредниците отговорност за контакти на трети лица, които поместват информация в пространството. Тези оператори подлежат на отговорност, ако не са ограничили разпространението на контакта от момента, в който им стане известно за неговия незаконен характер.

⁷⁹ Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries. 7 March 2018 // Council of Europe. URL:<https://rm.coe.int/1680790e14>

- Да насърчава саморегулацията или съвместната регулация на интернет платформите.

ЗАБРАНА ЗА ДИСКРИМИНАЦИЯ

Един от най-ярките примери за ползата от цифровите технологии, помагачи за преодоляването на дискриминацията, е разширяването на финансовите услуги за сметка на мобилните виртуални пари. От друга страна, те крият и редица проблеми, които усилват дискриминационните тенденции – алгоритмите са основен генератор на подобен проблем.

В Декларацията относно манипулативните възможности на алгоритмичните процеси⁸⁰ на Комитета на министрите при Съвета на Европа се отбелязва, че технологията позволява въз основа на достъпни данни да се правят ефективни изводи за хората и те да бъдат квалифицирани към една или друга категория. Подобна практика допълва съществуващите форми на социална, културна, религиозна, юридическа и икономическа сегрегация. Това в крайна сметка е таргетиране на хора въз основа на цифров профил и може да окаже съвсем конкретно влияние в случаите, когато се използва ИИ за вземането на решение относно право на ипотeki, услуги или в здравеопазването⁸¹.

Технологичният анализ на данни има за задача да оптимизира някакво решение на базата на зададени от операторите параметри. В повечето случаи това решение се взема въз основа на предпочитането на едни ценности пред други, създавайки някакъв нов етичен контекст на възприеманата от хората информация и взети решения. Въпросната информация може да бъде неутрална и благоприятна за едни групи от хора и крайно негативна за други. Тук спазването на правата

⁸⁰ Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes. Decl(13/02/2019)1.13 February 2019 // Council of Europe.

URL:

https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

⁸¹ The age of digital interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation // United Nations. URL: <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>;

на човека, свързано с технологиите, придобива по-широк обхват в сравнение с личния живот и свободата на словото, засягаща все повече социалната справедливост.

Защита на уязвимите групи

В смисъла на спазването на правата на човека своето значимо място заемат нормите и препоръките отнасящи се до уязвимите групи – деца, стари хора, хора със здравни проблеми и пр., всички те се нуждаят от специална защита.

Като цяло най-добре са разработени правата на децата в документите на Комитета на министрите при Съвета на Европа, създал Ръководни принципи за защита правата на детето в цифрова среда⁸². Основните препоръки са:

- Обезпечаване на децата с достъп до цифрова среда:
 - ✓ Достъпът е важен за реализацията на правата на децата по отношение на образование и осъществяване на социални контакти.
 - ✓ Особено внимание се обръща в случаите, когато децата са с инвалидност, в ситуация на риск, деца емигранти, живеещите в селски райони и други уязвими групи.
- Свобода на словото и право на достъп до информация:
 - ✓ Децата са създатели и потребители на информация.
 - ✓ Държавата трябва да съдейства за създаването на качествени връзки и цифрови сървъри за деца, включително и в образователния контент.
 - ✓ Децата трябва да бъдат получат информация къде могат да подават жалби или да получат подкрепа.
- Приватност и защита на данните:
 - ✓ Държавата е длъжна да информира децата за правото им на приватност и въпросите за защита на персоналните данни в цифровия свят.

⁸² Guidelines to respect, protect and fulfil the rights of the child in the digital environment. Recommendation CM/Rec(2018)7 // Council of Europe. URL: <https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a>

✓ При решаването на въпроса за възрастта, когато детето само дава съгласие за обработка на данни държавата трябва да вземе под внимание възрастовите особености и развитието с оглед зачитане на неговите интереси.

✓ Децата и родителите трябва да бъдат информирани как да оттеглят съгласието си при предоставяне на данни за дете.

✓ Държавата трябва да осигури интеграция и конфиденциалност по подразбиране (privacy-by-default) и механизми за вграждане на алгоритъм за конфиденциалност (privacy-by-design). Профайлингът на деца трябва да е забранен със закон.

- Право на образование:

✓ Държавата трябва да вземе мерки и да осигури цифрова грамотност на децата, включително и създаване на специални програми в образователните структури.

✓ Държавата е длъжна да обезпечи достатъчно средства за цифрова среда за формално и неформално обучение на децата.

- Право на защита и безопасност:

✓ Децата се сблъскват с множество рискове в цифрова среда.

✓ Държавата трябва да прогнозира риска и оперативно да носи отговорност за него, взаимодействайки с корпорациите за внедряване на принципа на безопасност (safety-by-design), да предприема мерки срещу експлоатацията на деца, а също да създаде вграден комплекс от правни мерки срещу разпространението на материали, свързани със сексуална експлоатация на деца.

Всички тези мерки трябва да намерят своето отражение в регулаторната и институционална политика на държавата, също така да се осъществи взаимодействие между държавата, бизнеса, СМИ и гражданското общество.

Инициативите в областта на защита правата на децата при ползването на интернет, мобилни приложения и сървъри трябва да бъдат разработени на национално ниво. В САЩ е

приет закон за конфиденциална защита на децата в интернет (Children's Online Privacy Protection Act)⁸³, където са описани задълженията на операторите на уеб-сайтове и онлайн-сървъри спрямо деца до 13 години, а също така и други сайтове, които събират персонални данни за деца под 13-годишна възраст.

Във Великобритания е разработен практически кодекс за онлайн сървъри със съответстващ за възрастта на потребителите дизайн (Age Appropriate Design: Code of Practice for Online Services)⁸⁴. В документа са маркирани стандарти за игри, мобилни приложения и други цифрови сървъри, които разработчикът трябва да изпълни, за да защити персоналните данни на децата.

ЗАДЪЛЖЕНИЯ НА БИЗНЕСА И ЧАСТНИТЕ ЛИЦА

Обикновено нормите за правата на човека са адресирани към държавата и като допълнение в нейните задължение се включват защитата на частните лица и техните взаимоотношения. В международното право, занимаващо се с правата на човека, съвсем скоро се появява нова тенденция, която става все по-актуална с навлизането на цифровите технологии и отговорностите на бизнеса, свързан с тях. През 2011 г. ООН разработва Ръководни принципи на предприемачеството по отношение правата на човека⁸⁵. Групата на ООН, отговаряща за цифровото сътрудничество на високо ниво, отбелязва, че е назрял моментът за разработването на стандарти за бизнеса

⁸³ Children's Online Privacy Protection Rule («COPPA») // Federal Trade Commission.

URL: <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>

⁸⁴ Age appropriate design: a code of practice for online services: Consultation document // ICO. URL: <https://ico.org.uk/media/about-the-ico/consultations/2614762/age-appropriate-design-code-for-public-consultation.pdf>

⁸⁵ URL: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_ru.pdf

по отношение правата на човека и особено в областта, касаеща собствениците на социални мрежи⁸⁶

Достъпността на технологиите в международните документи

Цифровият разрыв и разделение между страните, които са на различно технологично ниво, е съществувал и преди, но в последните години той става все по-голям. Наблюдава се такава тенденция дори в рамките на една държава поради неравномерния достъп до технологиите и липсата на инклузивен подход при разработването на цифрови продукти и услуги. От тази глава ще научите за ролята на държавата при обезпечаването на равенство и еднакъв достъп до технологиите.

Технологичният разрыв и разделение се случва на всички нива – той бива между мъже и жени, градско и селско население, различия в доходите, население от различен етнически произход, младо и старо поколение. Това оказва негативен ефект по отношение на социалната включеност и задълбочава неравенствата. Изместването на икономическия и социален живот в цифровата среда създава ефективна дискриминация и изключване на много хора без достъп до технологиите и най-вече до интернет.

Конференцията на ООН за търговия и развитие в своя обзор за развитието на цифровата икономика през 2019 г. отбелязва, че повече от половината от световното население няма достъп до интернет (в някои страни става дума за до един на всеки пет човека с достъп). Практически цифровото разделение води до ограничения в много сфери на живота (невъзможност да се ползват държавни услуги, непълноценно участие в икономическия и социален живот, невъзможност да се участва във формирането на политиката на държавата и пр.).

⁸⁶ Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries. 7 March 2018 // Council of Europe. URL: <https://rm.coe.int/1680790e14>

Като позитивен пример за използването на технологиите в социалната интеграция е инклузивната електронна търговия. Китайската платформа „Таобао“ дава възможност на повече от над 10 милиона малки и средни предприятия да търгуват – повече от половината предприемачи, работещи с тази платформа, са жени, а повече от 160 хиляди са с някаква степен на инвалидност или заболяване.⁸⁷

Премахването на цифровото разделение е глобална задача, стояща пред ООН⁸⁸. За да се постигнат целите на устойчивото развитие, е необходим равен достъп до цифровите технологии. Липсата му засилва многократно обществената дискриминация.

Споменатата група за сътрудничество в цифровата област към ООН препоръчва да се вземат мерки за осигуряване на достъп за всеки човек към цифровата мрежа и най-вече до мрежите за финанси и здравеопазване⁸⁹. Във фокуса на вниманието попадат живеещите в отдалечени райони, стари хора, хора с физически ограничения и здраве, намиращи се в настоящия момент в маргинална позиция спрямо цифровите технологии.

Ликвидирането на цифровото разделение изисква адаптация на политическото и национално законодателство с приемане на документи, гарантиращи цифровата инклузия и равенство. Необходими са институционални мерки за повишаване осведомеността на населението за технологиите и тяхната достъпност, а също така осигуряване на безопасност-

⁸⁷ The age of digital interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation // United Nations. URL: <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>

⁸⁸ Resolution adopted by the General Assembly 60/252. World Summit on the Information Society // International Telecommunication Union. URL: <http://www.itu.int/wisd/2006/res-60-252.html>; Digital Economy Report 2019. Value creation and capture: implications for developing countries // United Nations Conference on Trade and Development. URL: <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2466>

⁸⁹ World Economic and Social Survey 2018: Frontier technologies for sustainable development Overview. E/2018/50. 19 April 2018 // United Nations Documents.

URL https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESS2018-overview_en.pdf

та и неприкосновеността на личния живот, създаващи доверие сред населението по отношение на технологиите. На международно ниво трябва да се създаде матрица на цифрова инклузивност, за да има единна мяра за прогреса със съответните стратегии и планове за действие.⁹⁰

⁹⁰ The age of digital interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation // United Nations. URL: <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>

ИЗПОЛЗВАНА И ПРЕПОРЪЧАНА ЛИТЕРАТУРА

Бострьом, Н. Свръхинтелигентност: посоки, опасности, стратегии. Изток Запад. 2018

Гудман, М. Киберпрестъпления. Милениум. 2016

Джеймс, Б. Последнее изобретение человечества: искусственный интелект и конец эры Homo Sapiens. АНФ. 2015

Корбишли, Н. Сканирани. Дилок. 2022

Кърцуайл, Р. Как да създадем ум. Изток Запад. 2015

Маркофф, Дш. Homo roboticus? АНФ. М. 2017

Минделл, Д. Восстание машин отменяется! АНФ. 2016

О`Райли, Г. Бъдещето. Какво носи бъдещето и защо това зависи от нас. Изток Запад. 2019

Пенроуз, Р. Сенките на ума: търсене на липсващата наука за съзнанието. Изток Запад. 2020

Пицуги, М. Забранената еволюция/ Atea Books 2017

Рассел, С. Совместимость: как конторлировать искусственный интелект. АНФ. М. 2021

Спирдонов, Ив. Трансхуманизмът: програмираната еволюция на Homo Sapiens. Atea Books 2021

Тодорова, М. Изкуствения интелект. Изток Запад. 2020

Форд, М. Роботы наступают: развитие технологии и будущее без работы АНФ. 2020

Фрей, Х. Здравей свят. Изток Запад. 2019

Шнайдер, С. Искусственнэ ты. АНФ. 2022

Шнуренко, И. Демон внутри: „Наше завтра“. М 2021

Янков, Б. Тайните на естествения интелект. Верен. 1998

Етика и „цифра“ – етические проблемы цифровых технологии. .Сб. М. РАНХиГС. 2020

Ashley, K. Artificial Intelligence and legal analytics. Cambridge University press. 2017

Dubber, M., Pascale, F., Das, S. The oxford handbook of ethics of AI. Oxford University Press. 2020

Zuboff S. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs, 2019.

Corrales, M., Fenwick, M., Forgo, N. Robotics, AI and Future of Low. Springer. 2018

Thompson, S. Machine Low, Ethics and morality in the And of Articial Intelligence. IGI Global 2021

Wischmeyer, T., Rademacher, T. Regulating Artificial Intelligence. Springer. 2020

S. Matthew Liao. Ethics of Artificial Intelligence. Oxford Uniwersity. Press 2020

Whitepaper on Artifical Intelligence – A European approach to excellence and trust. Brussels. 19.2. 2020

URL: <https://www.un.org/ru/universal-declaration-human-rights/index.html>

ISO/IEC 2382:2015(en) Information technology — Vocabulary // ISO.

URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en>

Доклад на Генералния секретар на ООН “Използване на информационно-комуникационните технологии за инклузивно социално-икономическо развитие//Организация на обединените нации

URL: <https://undocs.org/bg/E/CN.16/2014/3>

Data Ethics Framework//Gov.uk.

URL: <https://www.gov.uk/government/publications/data-ethics-framework/>

data-ethics-framework

Nevo S. An Inside Look at Flood Forecasting // Google AI Blog.

URL: <https://ai.googleblog.com/2019/09/an-inside-look-at-flood-forecasting.html>;

Copeland R. Google’s ‘Project Nightingale’ Gathers Personal Health Data on Millions of Americans // The Wall Street Journal.

URL: <https://www.wsj.com/articles/google-s-secret-project-nightingale->

gathers-personal-health-data-on-millions-of-americans-11573496790;
Murphy B. Google's Project Nightingale Secretly Gathers Private

Health Data on Millions of Americans, According to a Report. Most Patients
Have No Idea It's Happening // Inc.

URL: <https://www.inc.com/bill-murphy-jr/googles-project-nightingale-secretly-gathers-private-health-data-on-millions-of-americans-according-to-a-report-most-patients-have-no-idea-its-happening.html>

Frier S. Facebook Paid Contractors to Transcribe Users' Audio
Chats // Bloomberg. URL: <https://www.bloomberg.com/news/articles/2019-08-13/facebook-paid-hundreds-of-contractors-to-transcribe-users-audio>;

Smith D. Microphone Gate: The world's biggest tech companies were
caught sending sensitive audio from customers to human contractors.
Here's where they stand now // Business Insider.
URL: <https://www.businessinsider.com/amazon-apple-google-microsoft-assistants-sent-audio-contractors-2019-8>

Recommendation CM/Rec(2019)2 of the Committee of Ministers to
member States on the protection of health-related data // Council of
Europe.

URL: https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168093b26e; Recommendation of the Council on OECD Legal Instruments
Health Data Governance // OECD.

URL: <https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf>

Data Ethics Framework // Gov.uk.
URL: <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework>

European ethical Charter on the use of artificial intelligence (AI) in
judicial systems and their environment // Council of Europe.

URL: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>;

CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment // Council of Europe.

URL: <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment#:~:text=Council%20of%20Europe%20adopts%20first,artificial%20intelligence%20in%20judicial%20systems&text=The%20Charter%20provides%20a%20framework,AI%20in%20national%20judicial%20processes>

General principles for the use of Artificial Intelligence in the financial sector // De Nederlandsche Bank.

URL: https://www.dnb.nl/en/binaries/General%20principles%20for%20the%20use%20of%20Artificial%20Intelligence%20in%20the%20financial%20sector2_tcm47-385055.pdf

Microsoft AI principles // Microsoft.

URL: <https://www.microsoft.com/en-us/ai/our-approach-to-ai>

Sweden has gotten its first GDPR fine // Aigine.

URL: <https://aigine.se/en/sweden-has-gotten-its-first-gdpr-fine/gdpr/>

„Европейска конвенция за защита правата на човека и основните свободи“ и „Конвенция на Съвета на Европа за защита на физическите лица при автоматизирана обработка на персоналните данни „

Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data. T-PD(2017)01.23 January 2017 // Directorate General of Human Rights and Rule of Law.
URL: <https://rm.coe.int/16806ebe7a>

The Ethics of Data Sharing: A guide to best practices and governance // Accenture. URL: https://www.accenture.com/_acnmedia/PDF-35/Accenture-The-Ethics-of-Data-Sharing.pdf

Asilomar AI principles // Future of Life Institute.

URL: <https://futureoflife.org/ai-principles/?cn-reloaded=1>;

Simonite, T. AI experts want to end «black box» algorithms in government // Wired Business.

URL: <https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government>

McCarthy J., Minsky M.L., Rochester N. Et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence // Stanford.

URL: <http://www.formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

Kenyon M. Bots at the Gate: a human rights analysis of automated decision-making in Canada's immigration and refugee system // The Citizen Lab.

URL: <https://citizenlab.ca/2018/09/bots-at-the-gate-human-rights-analysis-automated-decision-making-in-canadas-immigration-refugee-system/>

Tackling bias in artificial intelligence (and in humans) // McKinsey.
URL: <https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>

Kleinberg J., Lakkaraju H., Leskovec J. et al. Human decisions and machine predictions // The Quarterly Journal of Economics. 2018. Vol. 133, no 1. P. 237–293.

Silberg J., Manyika J. Tackling bias in artificial intelligence (and in humans) // McKinsey Global Institute. URL: <https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>

Lum K., Isaac W. To predict and serve? // Significance. 2016. Vol. 13, no 5. P. 14–19. URL: <https://doi.org/10.1111/j.1740-9713.2016.00960.x>

Ross C., Swetlitz I. IBM's Watson supercomputer recommended «unsafe and incorrect» cancer treatments, internal documents show // STAT+.

URL: <https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf>

Chiappa S. Path-Specific Counterfactual Fairness // [Silvia Chiappa.]

URL: <https://csilviavr.github.io/assets/publications/silvia19path.pdf>

Vincent J. Amazon Reportedly Scraps Internal AI Recruiting Tool That Was Biased against Women // The Verge. URL: <https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report> Hamilton I. A. Why It's Totally Unsurprising That Amazon's Recruitment AI Was Biased against Women // Business Insider.

URL: <https://www.businessinsider.in/why-its-totally-unsurprising-that-amazons-recruitment-ai-was-biased-against-women/articleshow/66192889.cms>

Strickland E. Racial Bias Found in Algorithms That Determine Health Care for Millions of Patients // IEEE Spectrum.

URL: <https://spectrum.ieee.org/the-human-os/biomedical/ethics/racial-bias-found-in-algorithms-that-determine-health-care-for-millions-of-patients>

Robinet P., Howard A., Wagner A.R. Conceptualizing overtrust in robots: Why do people trust a robot that previously failed? // *Autonomy and Artificial Intelligence: A Threat or Savior?* Cham: Springer International Publishing. 2017. P. 129–155. DOI: https://link.springer.com/chapter/10.1007/978-3-319-59719-5_6; Wagner A. R., Borenstein J., Howard A. Overtrust in the Robotic Age // *Communications of the ACM*. 2018. Vol. 61, no 9. P. 22–24. DOI: <https://dl.acm.org/doi/fullHtml/10.1145/3241365>

Turing A.M. Computing machinery and Intelligence // *Mind*. 1950. Vol. 54, no 236. P. 433–460

Wiener N. Some Moral and Technical Consequences of Automation // *Science*. 1960. Vol. 131, no 3410. P. 1355–1358

Bostrom N., Yudkowsky E. *The Ethics of Artificial Intelligence* // *Cambridge Handb. Artif. Intell.* 2011. P. 1–20.

The Moral Machine Experiment // *Moral Machine*. URL: <http://moralmachine.mit.edu/>

Awad E., Dsouza S., Kim R. Et al. The Moral Machine experiment // Nature. 2018. Guidelines on the protection of individuals with regard to processing of personal data in a world of Big Data//Council of Europe.

Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>

Unboxing Artificial Intelligence: 10 steps to protect Human Rights // Council of Europe.

URL: <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>

Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL:<https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>Vol. 563. P. 59–64. DOI: 10.1038/s41586-018-0637-6

Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems // Council of Europe.

URL: <https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf>

Kayser-Bril N. Personal Scoring in the EU: Not quite Black Mirror yet, at least if you're rich // Algorithm Watch.

URL: <https://algorithmwatch.org/en/personal-scoring-in-the-eu-not-quite-black-mirror-yet-at-least-if-youre-rich/>

China's Algorithms of Repression // Human Rights Watch.

URL: https://www.hrw.org/sites/default/les/report_pdf/china0519_web.pdf

Kharpal A. China's surveillance tech is spreading globally, raising concerns about Beijing's influence // CNBC.

URL: <https://www.cnbc.com/2019/10/08/china-is-exporting-surveillance-tech-like-facial-recognition-globally.html>

Altman A., Discrimination // The Stanford Encyclopedia of Philosophy.

URL: <https://plato.stanford.edu/entries/discrimination/#TypDisMorSen>

Digital Exclusion, A research report by the Low Incomes Tax Reform Group of The Chartered Institute of Taxation. London, 2012. P. 16–24.

URL: https://www.litr.org.uk/sites/default/les/digital_exclusion_-_litr_report.pdf

Schlogl L., Sumner A. The Rise of the Robot Reserve Army: Automation and the Future of Economic Development, Work, and Wages in Developing Countries — Working Paper 487 // Center for Global Development.

URL: <https://www.cgdev.org/publication/rise-robot-reserve-army-automation-and-future-economic-development-work-and-wages>

Connor N. Chinese school uses facial recognition to monitor student attention in class // The Telegraph.

URL: <https://www.telegraph.co.uk/news/2018/05/17/chinese-school-uses-facial-recognition-monitor-student-attention/>

Marks M. Suicide prediction technology is revolutionary. It badly needs oversight // The Washington Post.

URL: https://www.washingtonpost.com/outlook/suicide-prediction-technology-is-revolutionary-it-badly-needs-oversight/2018/12/20/214d2532-fd6b-11e8-ad40-cdf0e0dd65a_story.html

Code of conduct for data-driven health and care technology // Department of Health & Social Care.

URL: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology>

Alston, P. How Britain's welfare state has been taken over by shadowy tech consultants // The Guardian.

URL: <https://www.theguardian.com/commentisfree/2019/jun/27/britain-welfare-state-shadowy-tech-consultants-universal-credit>

Pilkington E. Digital dystopia: Algorithms punish the poor // The Guardian.

URL: <https://www.theguardian.com/technology/series/automating-poverty>

Alston, P. How Britain's welfare state has been taken over by shadowy tech consultants // The Guardian.

URL: <https://www.theguardian.com/commentisfree/2019/jun/27/britain-welfare-state-shadowy-tech-consultants-universal-credit>

Early Pursuit against Organized crime using environmental scanning, the Law and Intelligence systems // European Commission.

URL: <https://cordis.europa.eu/project/rcn/106659/factsheet/en>

Chammah M. Policing the Future // The Marshall Project.

URL: <https://www.themarshallproject.org/2016/02/03/policing-the-future>

Valentino-DeVries J. Tracking Phones, Google Is a Dragnet for the Police // The New York Times.

URL: <https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html>

AI for Humanity. French strategy for artificial intelligence. URL: <https://www.aiforhumanity.fr/en/>

Pan-Canadian Artificial Intelligence Strategy // CIFAR. URL: <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>

国务院关于印发新一代人工智能发展规划的通知/国务院.

URL: www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
Перевод от англ.: Notice of the State Council Issuing the New Generation of Artificial Intelligence Development Plan // Foundation for Law and International Affairs. URL: <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf>

National Strategy for Artificial Intelligence // Ministry of Finance and Ministry of Industry, Business and Financial Affairs.

URL: https://eng.em.dk/media/13081/305755-gb-version_4k.pdf

Civil Law Rules on Robotics. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) //Official Journal of the European Union.

URL: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

Asilomar AI Principles // Future of Life.

URL: <https://futureoflife.org/ai-principles/?cn-reloaded=1>

ISO/IEC 20546:2019(en) Information technology — Big data — Overview and vocabulary // ISO.

URL: [https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en;ISO/IEC TR 20547-2:2018](https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en;ISO/IEC%20TR%2020547-2:2018) [ISO/IEC TR 20547-2:2018] Information technology— Big data reference architecture — Part 2: Use cases and derived requirements // ISO.

URL: <https://www.iso.org/standard/71276.html>; ISO/IEC TR 20547-5:2018. Information technology — Big data reference architecture — Part 5: Standards roadmap. URL: <https://www.iso.org/standard/72826.html>

Report of COMEST on robotics ethics // UNESDOC.
URL: https://unesdoc.unesco.org/ark:/48223/pf0000253952_eng

Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>

Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes // Council Of Europe.

URL: https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

Unboxing Artificial Intelligence: 10 steps to protect Human Rights. Commissioner for Human Rights Recommendations // Council of Europe.

URL: <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>

Addressing the impacts of Algorithms on Human Rights. Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems // Council of Europe.

URL: <https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf>

Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. ETS No. 108 // Council of Europe. URL: <https://www.coe.int/ru/web/conventions/full-list/-/conventions/rms/0900001680078b37>

Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data // Council of Europe. URL: <https://rm.coe.int/16808ac918>.

Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data. T-PD(2017)01. 23 January 2017 / Directorate General of Human Rights and Rule of Law // Council of Europe.

URL: <https://rm.coe.int/16806ebe7a>

Guidelines on Artificial Intelligence and Data Protection. T-PD(2019)01. 25 January 2019/Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data // Council of Europe.

URL: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8522>

Recommendation CM/Rec(2019)2 of the Committee of Ministers to member States on the protection of health-related data. 27 March 2019 // APDA. URL: <https://www.apda.ad/sites/default/les/2019-03/CM-Rec%282019%292E-EN.pdf>

Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, OECD/LEGAL/0188//OECD.

URL:<https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>

Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries. 7 March 2018 // Council of Europe.

URL:<https://rm.coe.int/1680790e14>

Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes.Decl(13/02/2019)1.13February2019//Councilof Europe.

URL:https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

The age of digital interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation// United Nations.

URL: <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>;

Guidelines to respect, protect and fulfil the rights of the child in the digital environment. Recommendation CM/Rec(2018)7 // Council of Europe.

URL: <https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a>

Children's Online Privacy Protection Rule («COPPA») // Federal Trade Commission.

URL: <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>

Age appropriate design: a code of practice for online services: Consultation document // ICO.

URL: <https://ico.org.uk/media/about-the-ico/consultations/2614762/age-appropriate-design-code-for-public-consultation.pdf>

URL:https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_ru.pdf

Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries. 7 March 2018 // Council of Europe. URL: <https://rm.coe.int/1680790e14>

The age of digital interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation // United Nations. URL: <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>

Resolution adopted by the General Assembly 60/252. World Summit on the Information Society // International Telecommunication Union.

URL: <http://www.itu.int/wisd/2006/res-60-252.html>; Digital Economy Report 2019. Value creation and capture: implications for developing countries//United Nations Conference on Trade and Development.

URL:<https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2466>

World Economic and Social Survey 2018: Frontier technologies for sustainable development Overview. E/2018/50. 19 April 2018 // United Nations Documents.

URL https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESS2018-overview_en.pdf

The age of digital interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation // United Nations.

URL: <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>