**Пета тема. СЪВРЕМЕННИ ЕТИЧНИ ФЕНОМЕНИ ПОРОДЕНИ ОТ ИЗПОЛЗВАНЕТО НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ**

**Културни различия на етичните норми за ИИ**

***Много системи ИИ се разработват в транснационални компании и се прилагат в последствие в няколко страни.Конкретните норми за ИИ се разминават от културните различия в сегмента морално и неморално. В експеримента „Морална машина“ участват хора от 240 страни и там са правени компромиси по отношение жертването на животни вместо хора, закононарушители вместо лоялни граждани и се прилагал уталитарен подход. Оказва се индивидуалните полови различия, жизнен стандарт, религиозни и политически възгледи не са оказвали влияние върху избора, докато културните различия са били от съществена важност при вземането на решения. 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 правенрто на такъв избор. Там където има „широка ножица“ между дохода изборът на спасени пада върху хората с виск статус за сметка на нископлатените. В сраните от Северна Америка и Европа в повечето случаи изключват намеса в работата на машините (автомобили) и са склонни да жертват пешаходци за разлика от Югоизточна Азия и Близкия Изток където е обратното – спасяват се пешаходци и възрастни хора. В Латинска Америка и Южна Африка по-често се жертват старците от колкото младите, хората с нисък статус пред хората с висок и мъже заради жени[[1]](#footnote-1).***

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

* коя културна норма е релевантна за ИИ?
* Трябва ли да се взема под внимание в работата на ИИ различията в етичните норми в регионите в които се прилага (т. нар. „етическа локализация“)?
* Необходимо ли е предварително да се договарят етичните норми и те да станат кодекс за използването им?

**Държавата и Изкуствения интелект – етика и доверие**

***Доверирто към системите на ИИ е тясно свързано с социалното доверие, т.к. подобни системи най-бързо се внедряват в държавния апарат и той се ползва с голямо социално доверие. Държавното регулиране в използването на ИИ не се гради на мнението на мнозинството, но не е добре да бъде до край делегирано от етичните теории без да се има в 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 стадии на експеримент в редица държави и това се случва в две направления – аналитична работа с информацията и автоматизация на рутиините 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 ги надценяват във възможностите им. Съществува и фетишизация на възможностите на ИИ по отношение на възможностите ми да се справят с редица социални проблеми. До известна степен готовността за прилагането на системи на ИИ се определя от степента на разбираемост и достъпност до гражданите, довериерието към разработчиците и другите ползватели.

Неизбежни 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 различават от това към хората изпълняващи подобна дейност и често преминават в пълно недоверие. Въпросът е не толкова какви етични системи са приложени в ИИ, а кой ги имплантира, а оттам и пада доверието в държавните институции.

„Ефекта 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 – ползвателя 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 разрив между хората които са готови да влезнат 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 до край 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 начинът им на живот. Дори да има гаранции за защита на човешките права при такава патернизация използваща алгоритми допълваща 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 потребителя разбере за мимикрията. Разкрилият се бот взаимодействайки с хората е обучен да не очаква от тях каквото и да е било сътрудничество и взема инициативата без оглед на потребителя. Тези психологически закономерности поставят пред държавните органи делемата 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 спаси човешки живот.

Друг голям проблем е когато решенията се вземат на принципа на етичните принципи на мнозинството, защото те всички са обременени от социалните феномени на отчуждението, колективните страхове и конформизма. Ярък пример е 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 съвети[[2]](#footnote-2).

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

* Позоваваме се на корпоративните норми или високите цели:
* Използваме ефимизми представяйки своето поведение („въздържал се“, вместо „премълчал“);
* Прехвърляне на отговорността върху други хора (така реши ръководството.
* Дехуманизиране на потърпевшите и прехвърляйки вината върху тях.

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 преодоляването на отлагането за после, компенсира присъщото на хората неадекватна визия 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 особенна важност са разработките на колективен интелект прилаган в краудсорсинговите формации и публичните дебати[[3]](#footnote-3)

**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 последствия 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 предприемачеството като аспект на правата на човека в ООН одобрен 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 СМ/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 свързано с обработката на персоналните данни в сферата на голямата база данни.[[4]](#footnote-4)Приоритетно е етическото 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.[[5]](#footnote-5) През май същата година комисарят по правата на човека при Съвета на Европа публикува препоръки „ Изкуствения интелект: 10 стъпки за защита правата на човека“.[[6]](#footnote-6) Държавите членки на ЕС са длъжни да изработят правова база която регламентира процедурата за провеждането на процедура за оценка на въздействието върху правата на човека от страна на системите ИИ изпълняванот държавните органи (human rights assessments).

**Оценка за въздействието на алгоритмичните системи – подходи и методики за регулиране**

Под внимание при казуса със защитата правата на човека е фактора въздействие на алгоритмичните системи (olgorithmic impact assessment)[[7]](#footnote-7). 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 Съвета на Европа следи 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 самия оператор на данните, а от друга това е трета страна по акредитация която следи за оценка на въздействието върху правата на човека. Тук не разрешен въпрос е задължителността и доброволността на самата оценка без тя да се превръща в бюрократично препятствие.

В момента Комитета на министрите при Съвета на Европа разработва препоръки за въздействието на алгоритмичните системи върху правата на човека[[8]](#footnote-8). Този проект предвижда оценката за правата на човека да се прави както от страна на държавата така и от страна на бизнеса. Акцента 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 високия риск относно правата на човека.

**Приватност и защита на персоналните данни**

**Цифровите технологии е област където 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 съвремието.**

**Понятие за приватност**

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

* В съвременния начин на живот съществува огромна прозрачност спрямо заобикалящите;
* Вероятността данните да придобият публичност е твърде голяма;
* Данните могат да бъдат твърде безпрепятствено използвани с цел облаги и това може да се случи съвсем легално и да бъдат използване против самия човек.

Тук става дума не само за защита от пряка намеса ( примерно принудителни медицински процедури), а за невидимост за околните относно данни за личния живот без възможности за разпространението дори при наличие на такива – контрола 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 от колкото в миналото. Същевременно качеството на самата информация е увеличила областите, които засяга и позволява, да се правят много по-генерални изводи за даден човек, нещо което извън неговите възможности дори. 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 цифрова технологии. Всяка една институция изискваща и използваща подобен тип информация трябва да гарантира неината неприкосновеност.
* Разгласяването на такъв тип информация може да повлияе негативно върху социалният му 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 да прояви нонконформизъм. И в двата случая се стреми да поддържа самореализацията си. Запазвайки кофиденциалността на личното си пространство човек може да има различно поведение в различни сфери на социума. Това бива много рестриктивно в случаите, когато се загуби контрола върху личната информация правеща живота силно прозрачен.

Стои въпросът до колко човек може да приспособи към подобна ситуация. Възможните варианти са свързани със самоограничения, отказ от избор ако той противоречи на обществения. Подобна самоцензура 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 социално обсъждане и същевременно стават все по-толерантни 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 неговия избор по отношение на закупуване на стоки и услуги. Тази тема 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 благото на човека“, но фактора свобода на избора и волеизлеяние 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 това е сведения (съобщения, данни) независимо от формата под която се представят (текст, графика, видео) на всякакъв възможен носители. Като ПД се възприемат не само тези които са в цифрова форма, а и такива които са на хартиен носител. В действителност приоритетна цифровата форма защото позволяват 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 може да бъде отнесена всяка информация със 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, отпечатъци, документи характеризиращи личността) са само една малка част от информацията която се обработва от операторите, но има друга информация която е проблемно да се отнесе към ПД.

Сега практически всяко мобилно устройство, уеб сайт 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 Зубов 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 модела на новия капитализъм изглежда така: „Капитализма на следене съвсем уклпнчиво провъзгласява човешкия опит като безплатна суровина подлежаща на обработка в поведенчески данни. Въпреки, че част от тези данни се използвт 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 огормни печалби поради факта, че много компании искат да платят залагайки на нашето поведение в бъдещето“[[9]](#footnote-9).

Доминиращи на световния пазар (изключение е китайския пазар) са социални медии Google и Facebook. В Китай под забрана са сървърите на американските дата 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 аналетика. 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 аналетика 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 в замяна събират всевъзможни данни за тези които ги ползват. С помощта на ИИ данните и метаданните се анализират 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 бизнес сегмент 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 Дж. 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 събирани от системите за наблюдение с разпознаването на лица, скринингови системи и системи за предикатна аналетика 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 човешкият 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 несъмнения лидер в проследяването с камери – 200 милиона до 2018 година и 8 от 10-те най-наблюдаваните градове в света са там. Естествено в последствие 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 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 „свръх рискова 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 сътветно през 2019 г. съдът признава системата за виновна в нарушаване на човешките права.

**Скринигови системи**

Скринингът е технология даваща някаква оценка и рейтинг на даден ползвател възоснова на някакви критерии за ценност на качествата на даден клиент (пр. банков кредит, ведомствено жилище, социална помощ, държавни услуги) като същевременно се представят и потенциалните рискове свързани с предоставяната услуга. 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. Някой от методите на банки и застрахователните компании са толкова неограничени и не се очаква в близко бъдеще тяхното регулиране. Нарушаването на приватността е едно от най-често срещаните етични проблеми.

Показателите за скрининг зависят от професионализма на разработчика и алгоритмите и достъпа до определени данни. За скрининговоте състеми от особена важност е събирането на данъците , закононарушенията, сдравния 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 на които генетичният анализ показва висок риск от заболяване. Работодателите започват да им отказват постоянни договори и повишение .

По данни на Algorithm Watch[[10]](#footnote-10) в няколко страни се прави опит да бъде внедрена скринингова систе 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 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 коултивирана искреност и доверителност на всички нива. Допълвайки западната система на кредитния рейтинг с допълнителн и показатели не касаещифинансовата 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 чуждестранните партньори. В тази държава има 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 на основата на ИИ да препоръчително да използват материали придържащи 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 видионаблюдение 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 човека.[[11]](#footnote-11)

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 страни [[12]](#footnote-12). Съответно заедно с тях и експерти от Huawei, ZTE, Dahua. Докладите на ASPI показват, че в отделни проекти на „умни“ градове с котайска технология за видионаблюдение са задествани в някой страни от Западна Европа е възможно да се събират данни които да се събират в Китай.

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

**Дискриминация и неравнопоставеност при цифровизацията**

***На международно ниво дискриминацията е забранена официално (законово), но се проявява в някой сфери поради чисто психологическите прояви на предубеждения и стереотипи. В технологиите и особенно в алгоритмите на ИИ с голяма степен се появява възможност да се засилят подобни социално психологически 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 Европейският 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 разлик 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 с вродените и неизменни характеристики на личността.*

Ако обобщим понятието дискриминация би приело следната формулировка: случаите в които не се допуска, лишава, предявяват се някакви изисквания към даден човек за разлика от останалите хора. Причините за подобни действия са несправедливи, ирационални, изградени върху предубеждения и стериотипи. Несправедливостта основаваща се на недобро отношение към дадени хора и неоснователност на разграничаването между тях се приема като противоречащо на етиката и обществения морал.[[13]](#footnote-13)

Дискриминацията бива пряка и косвена. Когато целта е да се ощети и ограничи представител на дадена група имаме пряка форма на дискриминация дори в случаите когато мотивът е маскиран или не напълно осъзнаван. Косвената дискриминация е ситуация при която някакви 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 повишават рисковете от дискриминация на основа на предубеждения и предразсъдъци, а от там и социална изолация. Всички тези феномени присъщи на доцифрофизирания период с навлизането на ИИ се мултиплицира и разраства. На основата на това регламента 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. Някой обществени групи и прослойки съвсем непреднамерено са лишени от достъп до технологиите и дори се намиратв т. нар. „цифрова ексклузия“. Този 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 хората засегнати от цифровата екслузия 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, предствители на етнически малцинства[[14]](#footnote-14). Това предполага от разработчиците на такива алгоритми да създават специален дизайн за хора в неравностойно положение или запазване на 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 някой групи които привидно са обективни и научно обосновани.
* Навлизащата мащабно на цифровизацията носи в ежедневието все по-голям дискримационен 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 за това 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 поддържа критичността по отношение на правилата за диференциацията на съответните групи.

*Инклузия (от англ. Inclusion – „включване“) изразяващо процеса на включеността на всеки гражданин в социума. Обикновено се под това се разбира достъпност на инфраструктурата, но инклузията включва и достъпността до цифрови продукти и всички други сфери на живота на човека. Инклузивната среда обезпечава 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 парманентно е заложен проблема с безработицата и смяната на професията. Началото на цифровата трансформация на държавата трябва да предвижда помощ за гражданите които трябва да 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, аналетиката на голямата база данни и облачните технологии. Тези отрасли се предв2ижда да станат базови за организацията и реализацията на работния процес. Компанните ще освояват 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, дата-аналитици и специалисти по социални мрежи - social media marketing (SMM).Ще се запаззи високото търсене на специалисти по ИИ и ползвателски опит - user experience (UX), взаимодействие човек и компютър human-computer interaction (HCI), роботехници. С висока значимост са менъджерите работещи с клиенти, треньорите, 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 внимаие 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 роботи в производството ще зависи от потребностите на конкретния отрасъл – в някой ще са необходими стационарни, а в други дронове и хуманоидни роботи. 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 роботи звучат оптимистично – загубените работни места са два пъти по-малко от тези които ще се създават в следствие 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:

1 сценарии. 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 в 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.

2 сценарии. Алгортмите напълно заменят човешкият труд там където се изпълняват рутинни трудоемки операции (селско стопанство, строителство, охрана, логистика, кибербезопасност и промишлено производство).

3 сценарии. 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 трябва да се автоматизират. Освен на умствен и физически трудът може да бъде диферинциран на рутинен и нерутинен. Тенденция е физическия рутинен труд да се роботизира най-напред. Умствения рутинен труд също подлижи на автоматизация, но като помощен за юристи и икономисти. Нерутинния физически и умствен труд съдържащ някаква креативност едва ли някога, че претърпи автоматизация.

*През 2016 г. в пресата се появиха съобщения, че изкуствен интелект предсказва решенията на Европейския съд с точност до 79%. В последствие 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 личностните особености на съдията който го 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 човешкият труд, но и ще създадат нови професии и работни места. Някой от професиите ще изчезнат до 2030 г. като в сферата на транспорта и строителството това, ще в рамките на 56%, в производството 46%, в търговията 44%, като най малко изчезващи професии ще се наблюдава 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, темпа 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 достъпът на цифровизация на различните обществени слоеве. Процесите на цифровизация и внедряването на технологии изисква от икономическата система и преквалификация и модернизиране на методите за обучение за да бъде преструктуриран пазара. Според отчета на Центърът за глобално развитие[[15]](#footnote-15) (Center for Global Development)с най-големи проблеми в цифровизацията на икономиката се сблъскват развиващите се страни. Въпреки стремежа да се спазват тенденциите за цифровизация, икономическият ръст е затормозен от структурната трансформация. В тези страни автоматизацията не предизвиква безработица, а по-скоро преразпределя съществуващата и преразпределя разпределението на доходите – много ниско платени работни места, нисък стандарт на живот и социални гаранции, трудно достъпна данъчна система (няма правила какви данаци трябва да се плащат при експлоатация на роботи). Реално автоматизацията не решава, а задълбочава съществуващите проблеми на икономиката на тези страни.

В 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.

1. 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 [↑](#footnote-ref-1)
2. 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. [↑](#footnote-ref-2)
3. Платформата Assembl използваща ИИ поддържа обмен на мнения в мониципалните 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 POVOX 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 // Thе Governance Lab.URL:http://www.thegovlab.org/static/ [↑](#footnote-ref-3)
4. Guidelines on the protection of individuals with regard to processing of personal data in a world of Big Data//Council of Europe. [↑](#footnote-ref-4)
5. Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL: <https://rm.coe.int/guidelines-on> - artificial-intelligence-and-data-protection/168091f9d8 [↑](#footnote-ref-5)
6. Unboxing Artificial Intelligence: 10 steps to protect Human Rights // Council of Europe. URL: https://rm.coe.int/ unboxing-articial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 [↑](#footnote-ref-6)
7. Guidelines on Artificial Intelligence and Data Protection // Council of Europe. URL:https://rm.coe.int/guidelines-on- artificial-intelligence-and-data-protection/168091f9d8 [↑](#footnote-ref-7)
8. 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 [↑](#footnote-ref-8)
9. Zuboff S. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs, 2019. [↑](#footnote-ref-9)
10. 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/ [↑](#footnote-ref-10)
11. China’s Algorithms of Repression // Human Rights Watch.

    URL: https://www.hrw.org/sites/default/les/report\_pdf/china0519\_web.pdf [↑](#footnote-ref-11)
12. 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 [↑](#footnote-ref-12)
13. Altman A., Discrimination // The Stanford Encyclopedia of Philosophy. URL: https://plato.stanford.edu/entries/ discrimination/#TypDisMorSen [↑](#footnote-ref-13)
14. 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.litrg.org.uk/sites/default/les/digital\_exclusion\_-\_litrg\_report.pdf [↑](#footnote-ref-14)
15. 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 [↑](#footnote-ref-15)