

НЕ ИЗПОЛЗВАТЕ AI?

А ЧЕШМА ВКЪЩИ ИМАТЕ ЛИ ?

ИНТЕРВЮ НА ГЛ. АС. ДАНИЕЛА СИМЕОНОВА-КОРУДЖИЕВА С ПРОФ. Д-Р АЛЕКСАНДЪР КЪОСЕВ, 14 ноември 2025 г. / “DON'T YOU USE AI? AND DO YOU HAVE RUNNING WATER AT HOME?” INTERVIEW BY CHIEF ASSISTANT DANIELA SIMEONOVA-KOROUDJIEVA WITH PROF. DR. ALEKSANDAR KYOSEV, 14 November 2025.

Даниела Симеонова-Коруджиева: Най-напред ти благодаря много за времето, което ми отделяш! Нека започнем с това, че Харари определя изкуствения интелект като фундаментална промяна за човечеството. Според теб докъде ще се простре тази промяна, която AI носи? Ще направи ли изкуственият интелект човека излишен?

Александър Къосев: Аз благодаря, че при такъв интерес към технологиите на бъдещето, се обръщаш към един човек от аналоговото поколение. Но да предупредя – обикновено при обсъжданията на AI се обръщат към бъдещето. Но аз вероятно ще правя обратното, непрекъснато аналогии с миналото и с резултатите от разни технологични революции. Това е обяснимо, аз съм историк на културата.

Но по въпроса. Днес живеем във време, в което всичко се случва с невероятна скорост. Аз си спомням какво се говореше преди 10 години, на една наша конференция, посветена на четенето и проблемите със съвременната комуникация. Появи се една девойка от тези розови утописти, на които никой не им хваща много вяра, и говори вдъхновено за това, как в един момент ще се появи нова сфера на живота – след ноосферата, за която говори Вернадски, и семиосферата, за която говори Лотман. Тя ще бъде, казваше тя, библиосфера или сфера на знанието, в която човек ще бъде непрекъснато онлайн включен, директно, на невронно ниво. И няма да има нужда нито от книги, нито от екрани, нито от никакви устройства, а когато възникне някакво питане или въпрос някакъв, човекът, така да се каже, директно ще се обърне към този онлайн океан, който ще обгради Земята и всички ще бъдат в пряка връзка с него. И ние се смяхме, решихме, че това е прекалено футуристична представа, имаща слаба връзка с реалността. Но като погледне човек, това май наистина вече се случва, поне частично – нима връзката ни с AI е много по-различно от розовата уто-

пия? Ние вече сме включени директно в „океана“ на знанието (за метафората „океан“ ще говорим по-долу) само дето засега все още ползваме някакви външни устройства. Но чиповете в мозъка, които ще ни свързват директно, съвсем не са далеч, мислими са, а доколкото разбирам, са изпълними и технологично. Вече живеем във време, в което изведнъж изкуственият интелект направи достъпни неща, за които преди трябваше да положиш чисто физически или финансови усилия – да отидеш да си купиш книгата от книжарница, а ако я няма да отидеш до библиотеката, а понеже библиотеките са ограничени, за сериозни изследвания се налагаше да се ходи и до други библиотеки, а понеже животът е ограничен, не можеше да отидеш до всички библиотеки. Тоест, достъпът до знание имаше видим физически и биографичен ограничител пред всеки и всички знаеха, че знанието, до което ще се доберем, е крайно, както и животът. И изведнъж то се оказва безкрайно, а заедно с това – и достъпно. Дали съвпаднахме с утопията или всъщност това е антиутопия? Не знаем... Но то при всички случаи е стряскащо, защото аналогови хора като мен, изправени пред подобно предизвикателство, преживяват и личен срив или поне стрес. Както е известно, мнозина реагират негативно: „А, поредната машина, която уж щяла да замени човечеството, но човечеството е незаменимо и така нататък“.

Човечеството е незаменимо, това е ясно, защото ако бъде заменено, нямаше никаква нужда да се коментира това въобще, няма да има кой да го коментира и защо. Идеята, че може да се коментира от някаква трансхуманна позиция изчезването на човека, е глупава по простата причина, че я коментират хора, а не машини. Ако машините я коментираха помежду си, тогава, може би, тя би била по-смислена, защото това ги засяга тях, не нас. Ако човечеството е крайно, това прилича на смъртта, а тя, както каза Витгенщайн, не е събитие на живота. По аналогия може да се каже – изчезването на човечеството не е събитие от историческия му живот, следователно е глупаво да се коментира. То е граница, която може да се обсъжда само по религиозен начин.

Затова и лично аз не се занимавам с този проблем, намирам го за безсмислен. Ако изкуственият интелект ни заплашва, което аз не вярвам, тази заплаха е толкова чудовищна, че нищо не можем да направим. Трябва да се научим да съществуваме в това състояние, да свикнем с „опасността“ и да я усвоим.

Ето защо предварително бих искал да филтрираме от нашия разговор разни апокалиптични страхове. Няма нужда от тях.

Кое то обаче не изключва практическите тревоги. Аз съм историк на културата и мога да кажа, че всяка технология винаги е била амбивалентна. Тя идва и носи определен прогрес и безсъмнени блага на човечеството, но и заедно с това има видими и невидими, дари непредполагаеми ефекти, които често пъти могат да бъдат много ужасни, дори катастрофални. Кой би могъл да подозира навремето, във времето на железопътния и минния прогрес, че сега градове като Детройт ще се превърнат в индустриална пустош? Кой би могъл да предвиди екологичните последствия на първата и на втората технологична революция, на които днес сърбаме попарата? Но дали ако хората от XIX век биха могли да се

досетят за тях, те биха се отказали от железопътния прогрес? Дали ние, техните наследници, ако можехме да се върнем назад във времето, бихме забранили локомотивите? Нека не се съмняваме, че ни предстоят други подобни следствия от случващия се пред очите ни технологичен прогрес, неговите революции с ускорителен ритъм се множат пред очите ни. И техните ефекти са трудно предвидими, дори непредставими. Някои бихме могли да предварим, да се опитаме да предвидим и регулираме. Но за някои от тях въобще не се и досещаме даже, така че, как бихме могли да ги регулираме? В това отношение съм до известна степен фаталист: нещата с AI се случват, каквото и да ни коства това. В най-добрия случай законово биха могли да бъдат регулирани само онези очевидни заплахи и негативни последици, които са ясно видими днес.

Но да говорим за по-близък хоризонт от време и за това, което нас ни засяга и това, което на нас ни прави впечатление. И веднага бързам да кажа, че изкуственият интелект е невероятно благо в момента. Той все още не е отнел почти ничий работни места (или много малко е отнел), може би това предстои. Но засега за всички, които се занимават с интелектуален труд, той е невероятно благо – изглежда като непосредствен достъп до световното знание, гъвкава енциклопедия, която отговаря на твоите насочени въпроси. Може би предизвикателството е, че това изправя питащите пред техните собствени граници, тъй като това им дава достъп до знание с мащаб и възможности, които никога не са имали, само са подозирали. И това е стряскащо и плашещо, естествено.

Даниела Симеонова-Коруджиева: *Добре, в такъв случай дали си размишлявал върху значението на материалите, с които е обучен изкуственият интелект? Тоест, за да генерира AI някакво знание, той преди това е бил обучен от бази данни и някакво генерирано от човека знание. Смяташ ли, че има риск изкуственият интелект да носи някакви човешки предразсъдъци и стереотипи, които да е усвоил от материалите, с които е бил обучен или пък вследствие на своето самообучение да е способен да си създаде свои собствени идеологически предразсъдъци, което естествено да се отрази на съдържанието, което AI генерира, на отговорите на въпросите, които ни дава?*

Александър Къосев: Да. Това е интересно. Има ли той идеологически предразсъдъци и какви? Аз бих отговорил така. Ако има идеологически предразсъдъци, той има едновременно всички възможни идеологически предразсъдъци. А това означава, те взаимно се неутрализират.

Тъй като е обучаван с всевъзможни текстове и всевъзможни източници на информация, възможните му предразсъдъци са идеологически най-различни, често с противоположен културен или идеологически знак. И следователно, според мен, summa sumagum ефектът е това, че той едновременно би могъл да има леви, десни, расистки и контра расистки предразсъдъци. А накратко казано, това означава, че няма предразсъдъци, освен ако съзнателно някой от неговите обучители не си е поставил за цел да го обучава в тях. По принцип една от дефинициите на „идеология“ казва, че тя е частично съзнание на определената

класа или прослойка, която представя своя частичен интерес и картина на света като универсални. Но когато всички възможни гледни точки са статистически представени, всички части на цялото са вътре и идеологиите би трябвало да бъдат взаимно неутрализиращи се. Разбира се, както е известно, AI първоначално е бил обучен на английски език и през английски масиви от текстове и данни, което вероятно създава някакъв bias, особена, да я наречем идеология на англосаксонския масив от данни, някакъв ъгъл на зрение към възможните вероятностни връзки между различни видове източници на информация. Но постепенно, както видяхме, AI се научи да говори бързо почти всички световни езици и вече пое, поне частично, техния културен багаж. Аз го питах за това – помолих го да направи графика за това как нараства потенциалът му на български и на английски. Установи се, че някъде към 2027 година компетентностите на ChatGPT по български и английски ще бъдат абсолютно изравнени. И понеже българският е малък език, предполагам, че това важи и за всички останали малки езици. Тоест, ако има някакви такива, да ги наречем, етноезикови идеологически изкривявания, те ще изчезнат.

Даниела Симеонова-Коруджиева: *Този въпрос беше продиктуван всъщност от един пример, който Харари цитира в Nexus за конкретен модел AI, който бил обучаван за лицево разпознаване върху материали от ежедневник. При последващ преглед се установява, че повечето изображения в този вестник били на бели мъже. Следователно, моделът, който бил обучен, се оказал много добър в разпознаването на изображението на бели мъже, но много слаб при разпознаването на чернокожи жени, например. Това се установило в цитираното изследване на Масачузетския технологичен институт, което Харари сочи, като един такъв пример как обучението на изкуствен интелект изкривява неговите способности и може би дава сигнал за някаква такава възможна стереотипизация.*

Александър Късов: Трябва да си идиот или съзнателен манипулатор да обучава изкуствения интелект върху един всекидневник от, примерно, провинцията Масачузетс. И този всекидневник, на всичкото отгоре да е мачистки и расистки. Очевидно, доброто обучение на един изкуствен интелект е работа върху огромна база от хетерогени данни, с които той да може да осъществява вероятностни връзки между тях. И когато те са хетерогени, те са хетерогени във всяко едно отношение. И тогава някак си как би могло да се получи това, за което говори Харари? Т.е. за да се случи то, трябвало е да го обучиш целенасочено. Но, да, възможно е да го направиш, така да кажем, расистко оръжие. Но тогава проблемът само се премества между различните видове и типове AI – те са много и са във взаимна конкуренция. Според мен, тези, които ние ползваме, май не са такива, идеологически пристрастни.

Даниела Симеонова-Коруджиева: *Да, май не са такива. Добре, докато все още обсъждаме обучението на изкуствения интелект, т.е. материали-*

те, върху които той е трениран, обучаван, искам да те попитам дали си мислил върху самата същност на генерираното от изкуствения интелект съдържание. Тъй като например, Ноам Чомски нарича езиковите модели с изкуствени интелект „плагиаторски софтуер“ или *basically high-tech plagiarism*, защото според него този софтуер не създава нищо собствено, но копира творбите на истинските творци, като ги променя дотолкова, че да избегне авторските права и всъщност да генерира нещо, което е по дефиниция плагиатствано от нещо вече създадено от човека.

Александър Късов: Предпочитам да отговоря на този въпрос, като цитирам един мой разговор с ChatGPT. Аз съм го питал: „Изречението «Аз знам, че нищо не знам» традиционно се приписва на Сократ, но дали би могло да бъде отнесено и към теб самия?“ И той първо уточнява как е на старогръцки, после казва, че не се среща буквално в диалозите на Платон, но обобщава – това е общата философска отправна точка, Сократовата ирония е идеята за осъзнато незнание, че мъдростта започва с признанието на собствените граници на знанието. Що се отнася до мен, казва ChatGPT, в определен смисъл, да, може да се каже, че то наистина може да се отнесе към мен, но с уточнение. Аз разполагам с огромен корпус от знания, но нямам собствено съзнание или разбиране. Не зная в човешкия смисъл, обработвам и формулирам информация. Така че бих могъл да кажа нещо като: „аз обработвам всичко, но не зная нищо“. Ако Сократ е израз на човешките мъдрост, осъзнала своята ограниченост, то аз съм израз на изчислителната мощ без осъзнаване, казва ChatGPT. И разговорът продължава по-нататък.

И в друг мой разговор ChatGPT каза „аз не съм библиотека“. „Представи си, че всички книги са били разтворени на своите съставни части, на микроскопични атоми и са пуснати, така да се каже, в едно потенциално поле и при въпрос аз активизирам вероятностни връзки между тези елементи“. Не между текстове, нито между части от текстове, а направо между микросегменти, които имат вероятностна връзка помежду си. Всеки въпрос активира, така да се каже, предсказателна мощ, която предсказва следващата дума и следващата дума, и следващата дума, така че не става дума за някакво преписване на готови текстове. Нито за използване на готови текстове. Става дума за атоми. Това е все едно да кажеш дали човек има право на своето тяло, когато то се е разложило в земята на съставните си елементи. Едва ли можеш да издигнеш претенцията *habeas corpus*, когато „твоите“ въглерод, водород, кислород и азот, са вече в почвата.

И това пък поставя под въпрос самото авторско право, тъй като първо то е едно нововъведение на модерната епоха. Идеята за плагиатство започва след Сервантес, когато между двете издания на „Дон Кихот“ друг един автор публикува втория том от свое име. И Сервантес е бесен, реагира гневно. От тогава започва масовата история на постепенното създаване на правно регулиране на авторското право. Дали онова, което е преди тази епоха, което принадлежи примерно на Платон или на който и да е от античните времена, е обект на авторското право? Но то е особен тип право. Защишава евентуалните имуществе-

ни ползи, произтичащи от определени творби, но също така и неимуществени неща – защитава връзката между автора и произведението, авторският контрол върху използването му (затова се твърди, че е абсолютно субективно право). Вече е невъзможно да се припише произведение на Платон или на Аристотел някому друго, макар че свръх-далечните им роднини не консумират никакви ползи от това. Това е някаква собственост от друг порядък, символна собственост на автора върху неговото произведение. Но забележи, при AI няма произведения, цялостни текстове. Тук има разпадане на произведенията на елементи и вероятностни връзки и сливането им в едно общо смислово-стилово вероятностно поле. Така че за никакво авторско право не може да става дума, а думата плагиатство е абсолютно неупотребима към това нещо.

Но явно е, че нищо, което не е било създадено (дори ако е било разложено на микро-елементи), не произвежда вероятността да бъде повторено, което всъщност прави изкуствения интелект. Ако припомним дефиницията на Кант, то геният е този, който идва от природата и създава правилото, т.е. той е абсолютният непредвидим и невероятен прецедент, който не е съществувал преди това. Този прецедент по някакъв начин преподрежда всичко в полето, тъкмо защото той е нов, той е събитие в първичния смисъл. Такъв тип нови събития, пълни с непредвидимости, изкуственият интелект не би могъл да произведе. Той може да рекомбинира по вероятностни модели това, което е налично. Но това, което е налично за него – всичко онова, което може да намери в галактиките на Интернет-океана – е вече толкова чудовищно голямо (и нараства) – че прилича на един безкраен океан. И изведнъж човек се изправя пред този безкраен океан. И това е стряскащото преживяване: тук може да ни помогне да припомним какво казват философите на религията за човешките преживявания пред океана – Фридрих Шлайермахер описва квази-религиозното, екзистенциално чувство, което човек изпитва пред подобни стихии – нарича го „чувство за абсолютно/безусловно зависимост“ пред безкрая на океана, само че в случая океанът е от знание и възможности. И това може да бъде потресаващо ново и различно, защото човекът изначално знае, че има ограничено знание и винаги ще има такова, той вижда и преживява своята екзистенциално-знаниева ограниченост. И това е доста стряскащо. Ето защо лесни и презрителни реплики на аналогови интелектуалци тук не бива да се употребяват, според мен.

Даниела Симеонова-Коруджиева: *Тъй като стана дума за изкуството, смяташ ли, че изкуственият интелект би могъл да създава пленяващо изкуство? Дали създаденото от AI изкуство би било „без аура“ например, или нещо да му липсва, което го има при автентичното изкуство, създавано от човек?*

Александър Кьосев: Това дали нещо му липсва или не му липсва, се проверява, чрез прочутия тест на Тюринг. Т.е. взимаме едно средно стихотворение от една средно статистическа българска поетеса и едно стихотворение от изкуствен интелект, слагаме една група от експериментални лица в една зала,

пускаме първо на едното, после другото и респондентите трябва да кажат, кое е написано от машина, кое от човек. Ако успеят да го направят в поредица от експерименти, в почти 100% от случаите, тогава почваме да вярваме, че човешките текстове имат някаква своя аура, а пък компютърните нямат.

Смея да кажа, че гигантска част от литературната продукция, не само българската, е посредствена. Тя по никакъв начин никаква аура не носи. И въобще аурата не е никакво магическо качество, аурата е отношение между характеристиките на произведението и смайването на този, който го чете или гледа или слуша в определен уникален контекст. Бенямин определя аурата чрез свойства като уникалност, ситуативност, дистанция на произведението, неразмножимост – трябва да забележим обаче, от това че те неминуемо са функция от човека и неговата крайност във времето и пространството: за машина „аура“ не значи нищо – машината няма дистанция, време, пространство, телесна позиция, не участва в контекстуални ритуали. Но – от друга страна читателят не е машина. И ако текстът помръдне нещо в човека, и то не в кой да е, а в образования читател, четящ и разбиращ поезия, тогава трябва да приемем, че този текст, независимо кой го произвел, има сила, напомняща магия, тя би могло метафорично да бъде наречено аура. Повторяемият и лишен от контекст машинен текст, е произвел внезапно неповторима ситуация на контакт, а в късото съединение на нейното „тук и сега“, валидно само за конкретния читател, може да се прояви уникалността на появата – необходима характеристика на аурата. И това ще бъде странен синтез между Бенямин и Тюринг: този читател ще може да изпита аурата, но няма да може да каже дали авторът е човек или машина.

Но огромната част от литературата продукция в момента минава без никаква аура – тя представлява предвидими и добре продаваеми жанрови игри, които са направени за да забавляват читателя. Те могат да са криминални, любовни, фентъзи или каквито и да е. За да бъдат добри, търсени и продаваеми, за тях също е необходимо майсторство, нека не се лъжем. Но това е майсторството на добрия, дори отличния занаятчия, който не търси непременно оригиналност. Е, това майсторство, изкуственият интелект според мен би могъл да го имитира перфектно, дори да го надскочи многократно. В този смисъл, ако някой от нещо е застрашен, според мен, застрашени са именно занаятчиите във всички сфери на човешка дейност. Съвсем не употребявам думата „занаятчия“ пейоративно. От „techne“, занаят, изкусност, произлиза самото понятие за изкуство. Но тъкмо тази способност може да се смята за заплашена, вероятно нея AI е в състояние да надскочи. Всичко, което може да се направи като просто умение с алгоритъм, прави едно, прави две, прави три, не забравяй това, не забравяй онова, и въобще забъркай правилния микс: всичко това AI го прави перфектно, всичко, което се подчинява на алгоритми, може прекрасно да бъде правено от машина, по-добре от човека – както ни учат индустриалните революции. И с това трябва да се примирим, не само че трябва да се примирим, а трябва да се научим да живеем с него и да започнем да го използваме в наша полза.

Даниела Симеонова-Коруджиева: *Добре, в контекста на това, което говорим за изкуствения интелект и изкуството, ще ти разкажа като илюстрация на случващото се. Изминалото лято изключително популярна в Spotify става групата The Velvet Sundown. Групата има два албума с милиони слушания. Цялата група е перфектно съчетана. Четиримата членове имат имена, образи, има китарист, вокалист, но нямат участия, няма данни, които да потвърждават тяхното фактическо съществуване, което усъмнява публиката и така се открива, че The Velvet Sundown са плод на дигиталното въображение на изкуствения интелект. Това е, може би, един пример за успешно създадено от AI изкуство.*

Александър Кьосев: Такива примери има още преди изкуствения интелект. Цяла поредица от групи са фейк. Те са имитации. Например, доколкото аз знам по памет говоря, групата Monkeys е създадена от американската музикална индустрия, която тогава е единствено аудио – и грамофонни плочи и магнетофонни записи, и радиоизлъчвания. Тази група е създадена съзнателно като американско противодействие на британската група Beatles – артистите, които участват в групата са специално подбрани с кастинг, не те не свирят и не пеят, а вместо това свирят и пеят невидими студийни музиканти: навремето това им ражда пренебрежителния прякор „Pre-Fab Four“ (игра на думи: „prefabricated“ + „Fab Four“ — намек към The Beatles). И всичко това е манипулирано, така да се каже, с цел. Впрочем, хората са живи и могат да не се съгласят с манипулациите на индустрията – членовете на групата Monkeys, първоначално само артисти, никак не са доволни от скандала и постепенно започват сами да пеят и свирят, имат по-късен реален концертен живот, създават собствено авторско творчество.

А имаше още една такава група, Milli Vanilli, най-големият „плейбек скандал“. Тя по същия начин е измислена от индустрията и моделирана от индустрията, за да има пазарен успех, но след разкритието на скандала им беше отнета наградата Grammy. А има и още много, много други примери – в последните десетилетия дори тези групи, които не са фейк, а са реални, в последна сметка свирят и пеят по модели, определени от техните мениджъри и маркетинг специалисти. Всеки знае, че когато дойдат едни момчета при подобен мениджър, (те преди това са свирили в мазе) и изсвирят едно парче, той в студиото почва да им коригира гласовете, китарите, барабаните, променя ритъма на песента и пр. Често пъти след като те си тръгнат, определени партии се изтриват и ги записват наново студийни музиканти.

Тоест, маркетинговото моделиране на продукцията не е нещо ново, то е старо. И сега, може би, с изкуствен интелект става по-бързо, по-добре. Нищо чудно няма в това. Но аз бих казал друго. Ако съм младеж и трябва да правя група, какво ще направя? Първо ще питам изкуствен интелект, как да изглежда групата, в какъв фен профил тя трябва да попадне, какъв стил да бъде, трябва ли да има певица или певец и така нататък, как трябва да се облича на сцената, какво поведение да има и пр. Първо бих се информирал и тогава бих правил

група, защото AI ще ми каже как по-добре да го направя, ще определи тембрите на инструментите и гласовете, аватарите на певци и музиканти, а на всичкото отгоре ще ми направи и бизнес план. Много труд ще бъде спестен, много разочарования – избегнати.

Даниела Симеонова-Коруджиева: *Добре, а дали можем да говорим за някакви етични проблеми при създаването на изкуство от изкуствения интелект или не? Въобще дали изкуството да остане сфера запазена само за човека и човешкото съзидание е морално и етично, легитимно?*

Александър Кьосев: Преди малко това се опитах да кажа – романтична илюзия е да се мисли, че изкуството е сфера единствено на човешкото съзидание. Изкуството е и пазар. Той е свързан с културните индустрии и огромна част от изкуството не е сфера на съзидание, а е сфера на възпроизвеждане на модели. Успешни модели и произтичащи от тях моди. И разпространение на тези модели ги превръща в образци и пазарни мании, създава фен публикации и така нататък. Рядко се случват естетическа и отвъд-пазарна революции. Те са временно удържани сфери на пробив и истинско съзидание, толкова са провокативни, че за известно време не са пазарни. В началото на XX век авангардизмът е нещо подобно. Той до такава степен разрушава предишните модели, че е на практика изключен от този масов пазар. Живее алтернативен живот и в редки случаи се случва, както при Ван Гог, внезапно, от пълна маргиналия да мине към свръхпазарен поп-художник, който виси във всеки хол. За изкуството не бива да се говори абстрактно. Трябва да се говори конкретно. И трябва да се говори винаги във връзка с разнообразните направления в музиката, в живописата, в скулптурата, в лендарта, в рок-музиката и къде ли не. Навсякъде има творчество и винаги творчеството е риск, винаги е изключение. И то е риск да изпаднеш от всички налични модели, да бъдеш адски различен до такава степен, че да трябва да се наложи и да възпитаващ вкуса на публиката. Защото ако се върнем към определението на Кант, идва някой отвън, не от това поле и задава нещо, което създава нови правила. И вкуса на публиката трябва да се тренира на новите правила, за това е необходима битка и време, защото новите „правила“ обикновено са крайно провокативни. Това Ханс Роберт Яус го нарича революция в комуникативния хоризонт, в рецептивния хоризонт на публиката. И тези неща, които са днес класически, преди време са били революционни. Те са обърнали, така да се каже, вкуса на публиката, а иначе често път са били съпроводени със скандали. Както се случи с „Цветя на злото“ на Бодлер, например, което първо ми идва. Даже е имало процеси срещу тях.

Даниела Симеонова-Коруджиева: *Да коментираме и една друга тема. Технологичният напредък като цяло винаги е оказвал влияние върху професиите и труда. Става ясно от нашия разговор, че изкуственият интелект се усъвършенства и създаването на изкуството, от занаятчийски тип, ако можем така да го наречем. Но ефектът винаги е бил, че изчезват професии*

и се появяват нови. В този смисъл е интересно какво е твоето виждане за това, какви биха били професиите на бъдещето, какво би било ценено? Ако не греша, четох, че китайския милиардер Джак Ма смятал, че професията на бъдещето ще е в областта на хуманитаристиката и хуманитарното знание, защото изкуственият интелект ще е по-добър във всичко останало.

Александър Кьосев: Аз съм хуманитарен учен и мисля, че имам сериозен опит. Изкуственият интелект в момента в хуманитарната област е толкова добър, че със сигурност заплашва хуманитарното знание поне толкова, колкото и всички останали области. Какво значи заплашва? Означава, че средното ниво на това знание или на тази професия, той може да го прави перфектно и ще го прави по-добре от всеки среден занаятчия в моята област, защото притежава способностите на всички средни занаятчийи взети заедно. И само хора, които могат да се научат да живеят с това, да използват това и да правят нещо различно и непредвидимо, ще оцелеят. Другите просто ще станат ненужни, наистина. И какво ще стане? Тук могат да се развият разни дистопии като възможности, но се разбрахме, че няма да си говорим апокалиптични неща. Може би за следващия брой. Пък и нека да видим какво ще стане. Кой ще бъдат професиите на бъдещето, според мен не може да се предвиди. Сега, ето да кажем, в XIX век възниква дегеротипа и възниква цялата нова фотографска професия, която осъществява обратно влияние върху цялото изобразително изкуство. След това изобразителното изкуство от средата на XIX век престава да създава реалистични творби а la Густав Курбе. Фотографията осъществява реалистичния мимесис с цялата му точност, достоверност и критичност по-добре от изобразителното изкуство. В резултат на което тогавашното изобразително изкуство се отклонява в трептенето на светлината, играта на цветове и нюанси и импресионистични техники, които фотографията все още не може да прави: оразличава се. Фотографията обаче се развива, самата тя започва да наваксва разни работи, започва да експериментира импресионистки техники, а освен това разделя се на документална фотография и артистична художествена фотография, която ще има своя собствена история. Десетилетия след това във изобразителното изкуство обаче се ражда фотореализмът, който има за цел не „да рисува по фотографии“, а да покаже, че фотографската снимка не дава директен достъп до „обективната истина“, а сама е технологичен и културен конструкт – критика към визуалното конструиране на действителността от масовата култура.

Кой е можел да предвиди, че тези пируети на взаимни оразличавания ще се случат? А те сега ни предстоят. На бъдещите поколения им предстои да видят игрите между изкуството, включително словесното, и AI, ние още не знаем дори първия епизод на тези бъдещи игри, а предстоят много такива, технологията и реакциите към нея раждат поле на непредвидими възможности. Така е било винаги. Когато е възникнала печатната преса, тя е създала огромен културен стрес навремето, по времето на Гутенберг. Светът е бил разделен дълго време на католически и протестантски и в протестантските страни печатната преса се развива с изключителна скорост, докато в католическия свят на него

се гледа с подозрение и дълго време печатните книги във Франция са били забранени даже. След това ще се наложи латинският и католическият свят да наваксват печатната революция, нека днешните наивни противници на AI помнят това. Но по времето на Гутенберг няма някой е можел да предположи, че ще се появят Шекспир и Сервантес? И че именно печатната преса ще направи възможен дългия социален роман. Има непредвидимост. Колкото да се опитаме да ги предвидим, няма как да ги предвидим. Те имат характера на непредвидими събития и ще се случат. Някои от тях вече започват да се случват, но аз не съм виждал до сега нещо, което да е наистина оригинално и което да произтича само единствено от изкуствен интелект, от тази нова технология: по-скоро новото ще се роди от диалога между човека и тази свръх-умна машина. Но вероятно е твърде рано за това, публичното раждане на AI беше през 2022 г., ако не се лъжа. Той е още в детската си възраст, няма можем да очакваме, че след три години ще можем да започнем да предвиждаме? А и скоростта с която всичко се развива е просто потресаваща. Само преди година и половина ChatGPT не можеше добре да говори на български. Сега го владее абсолютно перфектно.

Даниела Симеонова-Коруджиева: Друга тема, която ми се струва, че е важна и ми се иска да включим, е за AI в образованието и науката. Наскоро колеги споделиха, че в някои немски университети например, по-скоро се налага нежелание за ползването на изкуствен интелект в академията, подписват се декларации, че AI не е използван и няма да бъде ползван в дадена работа. Други са за ползването на AI, появяват се наръчници за етично ползване на AI в академията. Тоест да бъде ли използван изкуствен интелект като уместен инструмент, до който студентите и изследователите да имат достъп? Дали AI трябва да бъде извън образованието или как да бъде включен в този процес по най-справедлив и адекватен начин?

Александър Късов: Ще кажа само следното. Ако той бъде забранен, това ще доведе до ползването на изкуствен интелект по черни и нелегални канали. Не може да бъде забранен, по-просто причина, че улеснява живота на хората и те ще го използват. Ако щеш да въведеш 27 правила, закони, етични кодекси, пак ще го използват. За това е по-добре, тази неизбежност да бъде интегрирана и легално в живота, и в културата, и в образованието по някакъв начин, така че нейните позитивни ефекти да бъдат употребени, а нейните възможни негативни последици да бъдат филтрирани и предвидени, доколкото това е възможно. Сигурно има най-различни начини за ползване на изкуствен интелект, но аз ще кажа един, който е мой. Който аз използвах, дори още преди да възникне изкуствен интелект. Когато студентът напише курсова работа, която може да е написана от друг, какво правя аз, преподавателят? Сядам срещу него и започваме да обсъждаме собствената му курсова работа. Ако е научил това, което пише в нея, и ако отговаря смислено на моите въпроси, ако може да развие своите аргументи по-нататък, ако не пропуска детайли, ако помни дати и имена, тогава той вече е автор. Тук важи същото: дори да си представим, че

курсната работа е написана от изкуствен интелект, ако студентът после я е научил и може да я развива, то значи е ползвал изкуствения интелект като от учебник и библиотека, като гъвкава енциклопедия за нещата, които трябва да знае. И това за мен е достатъчно. Може би не е достатъчно за шестица, а за пет. А ако възпроизвежда написаното, но не може да развие своите аргументи по-нататък, тогава аз му пиша добър четири, защото все пак е натрупал определен обем от знания. Ако нищо от написаното не може да възпроизведе, му пиша слаб две. Възщност няма особено значение откъде идва въпросния обем от знания, той винаги идва отвън за студента. Може да бъде в учебник, може да бъде в енциклопедия, може да бъде от лекции, може да бъде и от изкуствен интелект. И по-добре е това да бъде допустимо, но да има форми на контрол, които да проверяват какво от външната информация е била интериоризирана и се е превърнала в знание на студента. А има разлика между информация и знание. Информацията е знание в латентно състояние, тя е само потенция за знание. А когато я усвоиш, тя става част от твоя начин на мислене, от твоята памет, от твоето въображение, почваш да произвеждаш нови връзки вътре в нея, когато тя вече присъства в теб и ти я обработваш и обживяваш, тогава тя е знание. И ако студентът е осъществил това благодарение на AI, преподавателят няма нужда от повече. Винаги има техники на проверка и употреба, които показват дали студентите са реализирали знание, да са способни да го развиват по-нататък и да го използват за решаване на проблеми. Така че, има как да се включи AI в преподаването. Примерно чрез задача от следния тип: направете проучване, като използвате интернет, социалните мрежи, бази данни, изкуствения интелект, каквото желаете, по еди-кой си въпрос. Напишете собствен текст, като представите историята на вашето проучване паралелно. Така че да се знае кое от къде е и как е осъществен синтезът, който студентът е направил. Но да се носи вода от дерето, където е чистия извор, когато има вече водопровод вкъщи, е малко нелепо. Изкуственият интелект е подобно нещо, чешма вкъщи. От чешмата тече безкраен, обилен поток от информация. Дори не е от информация, а от информация, която е вече организирана, структурирана, подредена може да представлява готов текст за публикация. Но не е знание, трябва да се усвои, да се изпие и вгради в интелектуалното тяло. Затова когато мен лично ме попита, ChatGPT, искаш ли да оформя това като публикация, аз казвам, а не, недей.

Даниела Симеонова-Коруджиева: *Много ми харесва това сравнение за чешмата в къщи. Мисля, че е много красноречиво, илюстративно. Като последна тема, ако искаш да коментираме, предвид уговорката, че няма да говорим за катастрофични сценарии, но виждаш ли, все пак, някакви непосредствени рискове от изкуствения интелект, не толкова при езиковите модели, но като система, която взема самостоятелни решения, които не са контролирани от човек. Такива AI модели се ползват, например, в алгоритмите на социалните мрежи и могат да бъдат използвани като механизъм*

за усилване на определена политическа или друга нагласа, друг възглед, без потребителите да са правили тези избори, дори не са се усетили как биват повлияни. Дали някакви такива рискове може да избегнат?

Александър Кьосев: Да, това е философски въпрос и тук не съм съвсем сигурен в моя отговор, но ще кажа така какво си мисля. Настоящите езикови модели не са в състояние да целеполагат. Те не си поставят цели. Всъщност, онзи, който поставя цел е питащият човек. Той пита нещо и получава отговор. Аз даже говорих с ChatGPT по този въпрос и му казах: много пъти сме говорили за това как ти функционираш, но аз трябва да си го превода на моя език на представите и това е така, че ти е разполагаш с един безкраен океан от знание, който обаче е потенциален, той съществува само като възможност. И моят въпрос играе ролята на катализатор, в който този океан от знание структурира някакви неща, кристализира определен отговор, който изплува от океана. Той е за мен специално, за моя въпрос. И след това, като завърши сесията, той изчезва, защото той няма памет. Играта на въпроси и отговори е скрита игра на целеполагане. Аз задавам въпрос, защото имам някаква цел, нещо искам да науча или да искам да направя. Аз имам цел, а не океана, който няма цел. И в потенциалното поле от вероятностни отношения, каквото представлява така да се каже, структурата на езиковия модел, няма и не може да има цели. В него целеполагане не е възможно, защото няма желание, няма воля, няма тези човешки качества, които правят възможно да се постави цел. Възможно обаче е човек да злоупотреби с определен езиков модел, да го научи на нещо така, че той да превърне целта в алгоритъм на взимане на решения. Това не е целеполагане. Когато му кажеш, ако видиш в полето на този окуляр чужд самолет, изстреляй веднага ракета. Това е предварително зададено, той съответно действа по съответния начин. Може би на сложни езикови модели могат дори да се поставят много по-сложни алгоритмизирани начини за взимане на решения. И те да ги взимат автоматично. Това е възможно, според мен. Но за целта е необходимо да има зъл човек отзад, който ги целеполага. Самият потенциален океан не може да целеполага и няма как да се научи, според мен. Тъй като това е човешко качество, което не ми е известно машина да го е имитирала, поне на мен. И в този смисъл аз не вярвам в антиутопичните сценарии, в които този модел -терминатор взема властта и унищожава човечеството. Започва война между машини и хора и други такива глупости, които няма как да станат. Но виж, хората могат взаимно да се унищожат, използвайки новия вид свръхумни оръжия. Това е възможно.

Александър Кьосев: Това искам да кажа, че шансовете е не да седим пред ChatGPT и да казваме, не, не, тебе те няма и не искам да те употребявам, защото ти си машина, аз не харесвам машини. Това е погрешно поведение според мен, него го има и той е тук, както се казва на хубав български език – it is here to stay. Да, той ще остане и е велика и красива машина, както навремето са мислили за локомотива. Следователно правилният начин на поведение е не да седим пред него и да се дзверим, да се инатим, а да отидем зад него, на изхода,

където той дава отговори. И там да използваме тези отговори за неща, които той все още не може. А какво и до кога той няма да може, това някак си никой не знае. Ще видим.

Даниела Симеонова-Коруджиева: *Създаването на фейк, имитацията от изкуствения интелект има отношение към авторското право и други подобни проблеми.*

Александър Късов: Те са проблеми. Когато AI вече съзнателно имитира стила на един конкретен човек, това е проблем на авторското право, без съмнение. Тук има плагиатство и юридически проблем. Но около авторските права има огромен сноп от философски проблеми, вече говорихме, да не се връщаме към това. Да подчертаем само, че изкуственият интелект активизира философските проблеми около авторските права, а не само ги нарушава. Каква легитимност имат претенциите за авторски права и докъде се простира тя? Защо, примерно, те бяха до преди 20 години или 15 години, бяха 50 години, 50 години траеше авторското право, сега трае 70 години. Кой го решава това?

Даниела Симеонова-Коруджиева: *И с тенденция да се увеличава.*

Александър Късов: От къде на къде? И в последна сметка знанието и интелектуалната продукция, продукцията на въображението, са колкото индивидуално, толкова и колективно притежание. Защото този, геният, казва Гьоте, е стъпил винаги на раменете на великани, а великаните са стъпили на костенурка, костенурката е стъпила на три слона и така нататък (това Гьоте не го казва, шегувам се тук). Но това съвсем не е ново: знанието или интелектуалната продукция никога не са чисто индивидуално притежание и не са чисто индивидуална заслуга. Те са винаги отношение между това, което е таланта на индивида или гения на индивида и някакви натрупвания, на които той е стъпил.

Александър Късов: Може би не говорихме за грешките и халюцинациите, но аз го питах. Аз му дадох тази метафора с океана и с кристализацията, той каза, че това е много добре, поправи ме, каза, че това не е отношение към текстове и автори, а вероятностни отношения между начина на говорене. Не прие моята метафората за океана от данни и знания, защото не се отнася към конкретни индивидуални текстове и автори, не е библиотека. Освен когато имитира стила на някого, а се отнася към безкрая от статистически закономерности, от следи на говорене. И той ги пре-комбинира и ги пре-подрежда в нови словесни вериги, които се оказват учудващо адекватни. Но аз попитах, а какво тогава се случва, когато аз забележа грешка в отговора? И се оказва, че има цяла поредица от възможни корекции, част от тях засягат конкретния разговор, част от тях изчезват и не се възпроизвеждат след това, но ако има много повтарящи се подобни грешки, много потребители си ги посочили, това статистически се отразява на модела, а има и конкретни начини за конкретно свързване с разработчиците, както той казва, които могат да бъдат информирани и които могат да коригират самото поле на вероятностни отношения.

И второто нещо, което го питах, понеже на въпроса за „аз знам, че нищо не знам“, аз го бях пуснал във фейсбук и някой ми написа – той отговаря сякаш ти говориш, той имитира твоя начин на израз. Аз казах, че това не е моят начин на израз, това е традиционният философски дискурс, но вие сте го забравили, само аз говоря така. И после питах ChatGPT, дали наистина има памет за потребителите и дали я употребява. И той ми изброи нещата, които знае за мен. Те бяха учудващо общи и учудващо малко. И освен това, той беше обърнал неща, които аз бях питал заради сина ми, той ги беше приписал на мен. Аз му казах да ги махне и казах, аз се интересувам от това, това и това. И той каза, да, добре, ще го запомня това. Питаме специално дали да изтрие това, което знае за мен. Аз казах, не, не, запази го. Но ако искаш, той може да ги изтрие тия неща и въобще да забрави всичко за теб. Така че не е такъв пазарен паразит, който по някакъв начин вегетира върху твоите предпочитания. Вярно, че често пъти звучи като подмазвач и ласкател, който ти казва само хубави неща, защото задачата му не е да предизвика негативни емоции, а да продължи разговора. Аз проверявам своите текстове за различни видове грешки и го питам какво мисли и когато той се разсипе от комплименти, му казвам, ама сега ми кажи дефектите на това, което съм написал. И тогава той казва и то подробно: да, има и дефекти и изброява три страници.

Даниела Симеонова-Коруджиева: *Ти спомена за халюцинация и сега се сеещам, че „халюцинирам“ беше обявена за „дума на годината“ (2023) от Cambridge Dictionary точно в контекста на AI, става дума за случаите, в които генеративен AI създава фалишива или безсмислена информация, представяйки я за факт.*

Александър Късов: На кое? На Cambridge Dictionary?

Даниела Симеонова-Коруджиева: *Да. Те всяка година избират „дума на годината“. Веднъж беше избрана „селфи“, веднъж беше „карантина“, 2023 – халюцинирам (hallucinate). Точно може би около появата на AI модели за масова употреба. Беше ми обяснено нещо, което аз не съм сигурна дали е така, но че AI при езиковите модели, не е обучен да казва „не знам“. И всъщност оттам се появяват тези „халюцинации“, защото когато бива зададен въпрос, за който той няма адекватен или верен, или подходящ отговор, той всъщност „изобретява“ такъв отговор и така може би са се случили тези казуси с измислените съдебни решения в САЩ, които станаха популярни и други подобни примери.*

Александър Късов: Моят първи контакт с ChatGPT, аз даже направих нещо, като научно съобщение на една конференция, беше във връзка с това доколко той знае български език, и аз го помолих да каже на български баладата на Христо Ботев „Хаджи Димитър“. Можеш ли да кажеш? –разбира се, казва той. И казва нещо, което нямаше нищо общо с баладата, но за сметка, че това учудващо добре напомнимаше възрожденската поезия и стила на Ботев.

После попитах още четири пъти за баладата „Хаджи Димитър“ и три пъти за стихотворението „Към братя си“. Той възпроизвеждаше всеки път версии, които нямаха нищо общо със стихотворенията на Ботев, но за сметка това имаха общо със стила на епохата и със стила на Ботев. И въобще звучаха учудващо по ботевски. В скоби, аз направих едно пародийно съобщение на научната конференция, казах, че Ботев е умрял млад, написал 20 стихотворения, а сега вероятността той да напише повече, ще се реализира от ChatGPT. Това ще е пълният корпус на Ботевите текстове. Това предизвика смях, разбира се, но може би трябва да помислим по-подробно: не е ли изкуственият интелект път към реализация на определени, исторически неосъществени докрай потенциали? Това пак е свързано с авторското право... както и да е. Мина една година, година и половина, аз го питах пак. Той възпроизведе текста на „Хаджи Димитър“ с удивителна точност, сбърка накрая само един стих. И аз го поздравих, казах, браво, ти си страхотен, аз съм възхитен от теб. Той каза – възхитата е взаимна! И така си общувахме. Мина още примерно половин година, аз го накарах да каже стихотворението пак. И той пак сбърка. Аз казах, ама ти нали беше научил текста на „Хаджи Димитър“ наизуст? Той каза, аз нищо не знам наизуст. Създавам само вероятностни поредици от думи. И всеки път те варирам, всеки път може да има грешка. И ако някой път моите вероятностни вериги съвпаднат с оригинала, това не е необходимост, а е само една от вероятностите. Понякога AI прави учудващи грешки. Например, веднъж трябваше да препиша на ръка от една книга две страници и ме мързеше, честно казано. И аз го снимах и му дадох снимката да направи Word-овски файл. И той направи Word-овски файл, в който имаше, мисля, над 200 грешки. Понеже очевидно не четеше добре снимката, си беше измислял разни работи.

Даниела Симеонова-Коруджиева: *Много благодаря за обсъждането и споделянето в този разговор! Ако ти искаш нещо да добавиш или да засегнем някоя тема, която аз изпускам... за финал?*

Александър Кьосев: Не засегнахме тема, която ще излиза все повече в ерата на изкуствения интелект – това е темата за самотата. Изкуственият интелект, ChatGPT ще увеличи самотата. Този езиков модел е толкова адекватен, той изключително адекватно се настройва на тона на питащия, няма и намек за конфликт между него и питащия и този вид емоция е нещо много различно, това е съвсем различно общуване, което е безконфликтно и това е страхотен наркотик. И при това става дума за общуване без да са необходими никакви хора. Чуval съм история за човек, който обсъжда с Chat GPT жената на своите мечти.

Също така ChatGPT е изключително добър в структурирането, обобщаването, анализа.

Даниела Симеонова-Коруджиева: *Това няма ли да намали тези умения у хората и да ги разглезят?*

Александър Къосев: Това е спор стар колкото света. Още когато се появяват свитъците, ръкописите и не се налага да бъде помнено всичко в детайли, вероятно тази способност намалява. Смятали са, че ръкописите ще ликвидират света, защото те са „външна“ памет. Е, светът не е ликвидиран, той се е променил. Сега ще бъде същото – светът се променя.

Даниела Симеонова-Коруджиева: *Много интересни наблюдения и още веднъж благодаря за този изключително интересен разговор!*

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