ШТУЧНЫ ІНТЭЛЕКТ І ПРАВЫ ЧАЛАВЕКА: ПОШУК ЭТЫЧНЫХ АРЫЕНТЫРАЎ

Human constanta

Гэты артыкул быў напісаны Любоўю Ланінай, адрэдагаваны Кацярынай Пушкаровай і Таццянай Зіняковай, апублікаваны арганізацыяй Human Constanta, з дызайнам вокладкі ад Ксеніі Дзяругі.

Студзень 2026

Email: info@humanconstanta.org
Website: https://humanconstanta.org/
Instagram: https://www.instagram.com/humanconstanta/

Штучны інтэлект і правы чалавека: пошук этычных арыентыраў

Гэты тэкст з'яўляецца вынікам інтэлектуальнай і аналітычнай працы аўтаркі і рэдактаркі. Ідэі, структура і змест артыкула былі цалкам распрацаваныя, напісаныя і адрэдагаваныя чалавекам. Падчас працы выкарыстоўваліся ChatGPT і Perplexity для паляпшэння яснасці фармулёвак, дапрацоўкі асобных сказаў і праверкі лагічнай сувязі паміж думкамі. Усе штучнастварэныя кавалкі тэксту былі праверэныя, вычытаныя і апрацаваныя чалавекам перад выкарыстаннем.

ЗМЕСТ

Пра што гэты артыкул?

Становіцца складана адрозніць, калі ты маеш справу з рэальным чалавекам, а калі з штучным інтэлектам. Фільтры ў сацыяльных сетках, аўтавыпраўленне тэкстаў, беспілотныя аўтамабілі, пераклады і генерацыя карцінак – папулярныя функцыі ШІ, якія прагрэсуюць у якасці з кожным днём. Чат-боты становяцца ўжо не проста функцыяй сайтаў, а сапраўднымі кампан'ёнамі.[1] Перспектывы кіравання дзяржавай ШІ міністрам і запаўненне інтэрнету артыкуламі і карцінкамі, згенераванымі ШІ, і становяцца часткай штодзённага дыскурса. Уплыў на чалавека машынай узрастае, і гэта толькі падкрэслівае важнасць захавання правоў чалавека цэнтральным фокусам пры ўкараненні новых тэхналогій.[2]

Штучны інтэлект цяжка вызначыць адназначна, бо ў сваім развіцці ён хутчэй, чым вызначэнне яго межаў. У гэтым артыкуле пад тэрмінам "штучны інтэлект" мы будзем разумець спектр сістэм з розным узроўнем аўтаномнасці, здольных самастойна навучацца, генераваць вынікі ці прымаць рашэнні,[3] і імітаваць інтэлектуальныя паводзіны.[4]

Хуткае развіццё новых тэхналогій непазбежна ўзнімае пытанні этыкі, бяспекі і правоў чалавека. На тэрыторыі Эўропы ўступіліу моц акт па рэгуляванні штучнага інтэлекту і асобны кодэкс практыкі для штучнага інтэлекту агульнага прызначэння, ААН сфарміраваў Кансультатыўны орган высокага ўзроўню. Узмацнаецца заканадаўчае рэгуляванне ШІ у Кітаі, ЗША, і ўжо ёсць прыклады распрацаваных дзяржавай сістэм штучнага інтэлекту, як "этычны ШІ" пад назвай "Apertus" ад Швейцарскага ўраду.

Разам з бізнесам і дзяржавай да абмеркавання новых тэхналогій далучаюцца недзяржаўныя арганізацыі (НДА) і грамадзянская супольнасць, якія ўзмацняюць гучанне тэмаў этыкі, інклюзіі, роўнасці і правоў чалавека ў эпоху штучнага інтэлекту. Міжнародныя НДА, напрыклад такія як Amnesty International, Human Rights Watch, Privacy International, AlgorithmWatch, Access Now і іншыя актыўна маніторацъ, як тэхналагічныя распрацоўкі і штучны інтэлект ўплываюць на правы чалавека, дапамагаюць аналізаваць і прадухіляць рызыкі.

У той жа час НДА не толькі маніторацъ працэсы, але і актыўна ў іх ўключаюцца, бо адаптуюць штучны інтэлект для дасягнення сваіх мэтаў і

[1] "My Boyfriend is AI": A Computational Analysis of Human-AI Companionship in Reddit's AI Community, 2025
[2] The human line: safegoarding rights and democracy in the AI era, The Council Of Europe, Commissioner for Human Rights, 2025
[3] Рамачная канвенцыя Савета Еўропы аб штучным інтэлекце (CETS №225), 2024
[4] Рэкамендацыі ЮНЭСКА па этыцы штучнага інтэлекту, 2021

выканання працоўных задач. Працэс адаптацыі ШІ ў працу неразрыўна ідзе з спробамі прымаць рашэнні этычна. НДА спрабуюць наладжваць аўтаматызацыю руцінай працы з захаваннем чалавечага кантролю, аналізаваць вялікія масівы звестак з павагай да прыватнай інфармацыі, і імплементаваць іншыя рашэнні без рызыкі для правоў чалавека.

У гэтым артыкуле мы будзем уздымаць этычныя пытанні на мяжы правоў чалавека і новых тэхналогій:

- ➔ Як правы чалавека і штучны інтэлект перасякаюцца і ўзаемаўплываюць адзін на аднога;
- ➔ Якія магчымасці стварае штучны інтэлект, і якія схаваныя рызыкі ён стварае для правоў чалавека і НДА;
- ➔ Як этычна карыстаць ШІ, асабліва ў НДА. **У канцы артыкула вы знойдзеце чэк-ліст па этычным выкарыстанні штучнага інтэлекту.**

Артыкул будзе карысны тым, хто цікавіцца этычна-філасоўкім бокам развіцця ШІ, праваабаронцам і людзям, хто працуе ў падыходзе заснаваным на правах чалавека, а таксама актывістам і актывісткам, прадстаўнікам і прадстаўніцам НДА.

Пры чым тут этыка?

Карыстанне штучнага інтэлекту не абмяжоўваецца адным канкрэтным кірункам, яно закранае палітыку, эканоміку, медыцыну, ахову здароўя, адукацыю і іншыя сферы. **Правы чалавека маюць універсальны характар**, і таксама прысутнічаюць ва ўсіх сферах дзейнасці чалавека. Менавіта гэты агульны і ўзаемапранікальны характар робіць выкарыстанне штучнага інтэлекту непазбежна значным для правоў чалавека. Прывядзем тры прыклады карыстання штучнага інтэлекту ў кантэксце правоў чалавека:

1. **Права на здароўе**[5]

Карыстанне штучнага інтэлекту ў сферы аховы здароўя адкрывае новыя магчымасці для лекавання і даследванняў, ранняга выяўлення і прагназавання эпідэмій, і сваечасовых рэакцый. Напрыклад, у карыстанні з'яўляюцца прыкладанні, заснаваныя на працы штучнага інтэлекту[6], напрыклад трэкінг здароўя, якія дапамагаюць аналізаваць сімптомы хваробаў і прапаноўваць

[5] Артыкул 25 Усеагульнай дэкларацыі правоў чалавека; артыкул 12 Міжнароднага пакту аб эканамічных, сацыяльных і культурных правах.
[6] Artificial intelligence (AI) in healthcare and research, Nuffield Council on Bioethics, 2018

дыягназы, ацэньваць агульны стан, даваць парады па здароваму ладу жыцця. Гэтыя і іншыя сістэмы ШІ адкрываюць новыя магчымасці, напрыклад доступ да медычнай інфармацыі ў рэгіёнах, дзе ёсць абмежаванне на атрыманне прафесійнай дапамогі. Але ў той жа час, карыстанне штучным інтэлектам у такой сферы як ахова здароўя ўздымае пытанні правоў чалавека:

- ➔ ці дастаткова разнастайнай была база звестак, на якой навучалася ШІ прыкладанне, каб даваць парады шырокаму колу людзей, ці яна мае ўбудаваную дыскрымінацыю?
- ➔ Хто мае доступ[7] да настолькі чуллівай інфармацыі, як хранічныя хваробы і стан здароўя?
- ➔ І хто нясе адказнасць за ненадзейныя вынікі, калі яны нанеслі шкоду чалавеку?

2. Права на працу[8]

Штучны інтэлект стаў неад'емнай часткай амаль кожнага працоўнага месца: ён уключаны ў штодзённыя працэсы, уплывае на хуткасць і вынікі працы, а таксама істотна змяняе адносіны паміж працадаўцамі і супрацоўнікамі.

Сутыкненне з ШІ пачынаецца ўжо на этапе падачы резюмэ на працу, бо заяўкі праходзяць праз сістэмы аўтаматызаванага адбору, ці ATS. Такія сістэмы навучаюцца на інфармацыі, якая ўжо існуе у кампаніі, таму яны схільныя да ўзнаўлення сістэмнай дыскрымінацыі, якая ўжо прысутнічае ў сферы працаўладкавання, а таксама да памылак пры падборы персаналу. Напрыклад, аўтаматызаваныя сістэмы скрынінгу могуць адхіляць заяўку нават мэнэджара кампаніі, ці ацэньваюць заяўкі жанчын як менш адпаведныя, калі ў кампаніі існуе перавага супрацоўнікаў-мужчын, якіх мадэль ШІ лічыць "лепшымі" кандыдатамі, бо бачыць іх часцей падчас свайго навучання. ШІ стварае значныя рызыкі для супрацоўнікаў і супрацоўніц, калі выкарыстоўваецца для выдаленнага бесперапыннага назірання, распазнавання твараў, мімікі і эмоцый, а таксама ў некантраляваных ці прадузятых аўтаматызаваных сістэмах прыняцця рашэнняў, па тыпу ATS.

- ➔ Наколькі этычна рабіць высновы па аналізу эмоцый калег падчас відэа званкоў, і далей карыстаць атрыманную інфармацыю для прыняцця рашэнняў?
- ➔ Калі мы дэлегуем ШІ збор інфармацыі з інтэрв'ю, наколькі аб'ектыўным будзе яго абагульненне для фінальнага рашэння, нават калі яго будзе прымаць чалавек?

[7] Health data governance in the age of AI: policy imperatives, WHO European Region, 2025
[8] Артыкул 23 Усеагульнай дэкларацыі правоў чалавека; артыкулы 6 і 7 Міжнароднага пакту аб эканамічных, сацыяльных і культурных правах.

➔ Дзе праходзіць мяжа ўплыву ШІ на чалавека на працоўным месцы?

3. Права на справядлівы суд[9]

Карыстанне штучным інтэлектам у судовай сістэме ўжо не выглядае як антыўтапічная ідэя. Выкарыстанне ШІ мае вялікія перспектывы для сістэмы правасуддзя: яно павышае эфектыўнасць працы, паскарае апрацоўку звестак, дазваляе падрыхтаваць абарону за лічаныя хвіліны. ШІ выкарыстоўваецца для ацэнкі і аналізу дакументаў і пастаноў, паскарае доступ да юрыдычнай інфармацыі, транскрыбуе судовыя працэсы і нават ужываецца для прадказання вынікаў судовых рашэнняў на падставе аналізу папярэдніх спраў.

Аднак ужыванне штучнага інтэлекту ў судовай сістэме адносіцца да вобласці высокай рызыкі, бо мае наўпроставы ўплыў на права на справядлівае судовае разбіральніцтва. Такія сістэмы нясуць рызыку празмерна жорсткага прымянення закона, алгарытмічнай дыскрымінацыі, могуць паставіць пад сумнеў прынцып непрадузятасці суда і пагоршыць лічбавую ізаляцыю, што ў сваю чаргу ўплывае на роўнасць і ўспрыманне судовай сістэмы як справядлівай і незалежнай. Акрамя таго, узрастае рызыка памылак падчас судовых працэсаў: паводле сусветнай практыкі, на канец кастрычніка 2025 года зафіксавана больш за 480 выпадкаў, калі ў суд падавалі дакументы з выдуманымі прэцэдэнтамі або законамі, згенераванымі штучным інтэлектам.

Рэалізацыя права на справядлівы суд патрабуе эмпатыі, маральных разважанняў і разумення кантэксту, з чым штучны інтэлект пакуль спраўляецца з цяжкасцю[10]. У той жа час, ШІ можа істотна павялічыць эфектыўнасць судовай працы і скараціць час на выкананне тэхнічных задач, пры ўмове яго адказнай і этычнай імплементацыі. Аднак застаюцца пытанні, напрыклад:

➔ Хто нясе адказнасць за несправядлівыя судовыя рашэнні і памылкі ў пастанаўленнях, зробленыя ШІ?

➔ З развіццём ШІ, ці зможа ён калі-небудзь дасягнуць амаль чалавечага ўзроўню эмпатыі?

Для правоў чалавека, штучны інтэлект стварае ўнікальныя магчымасці аўтаматызацыі, паскарэння працэсаў, узмацнення ўплыву. У той жа час, ШІ робіць пагрозы больш вострымі, а іх распаўсюд хутчэйшым. Акрамя таго, з укараненнем тэхналогій у сферы, якія маюць значны ўплыў на жыццё чалавека, уздымаюцца пытанні этыкі, на якія пакуль няма адназначнага адказу.

[9] Артыкул 10 Усеагульнай дэкларацыі правоў чалавека; артыкул 14 Міжнароднага пакту аб грамадзянскіх і палітычных правах

[10] AI Judges in Courts of the Future and the Right to a Fair Trial in the ECHR, 2025

Такая шматбаковасць і вострая значнасць патрабуе ведання, з якімі рызыкамі можна сутыкнуцца, калі працаваць з ШІ. Сярод рызыкаў адзначаюць дызінфармацыю, схаваную дыскрымінацыю, прадузятасць і ўзнаўленне ілжывага кантэнту. Яны не з'яўляюцца ўнікальнымі, але здабываюць новы сэнс у кантэксце новых тэхналогій

Якія рызыкі для правоў чалавека?

Дызінфармацыя. Карыстанне і развіццё сістэм штучнага інтэлекту ўплываюць на хуткасць распаўсюду[11] ілжывай, недакладнай, прадузятай, фэйкавай ці абразлівай інфармацыі сярод шырокіх аўдыторый. Акрамя таго, якасць такога кантэнту становіцца ўсё вышэйшай, і яго ўсё цяжэй адрозніць ад сапраўднай інфармацыі. Ужо на верасень 2025 года колькасць артыкулаў, створаных з дапамогай ШІ, перавысіла колькасць артыкулаў, напісаных людзьмі і апублікаваных у Інтэрнеце. Пры гэтым, штучны інтэлект схільны да галюцынацый (генерацыі выдуманых фактаў, звестак і іншай інфармацыі), што без факт-чэкінга і чалавечага маніторынгу можа прывесці да сур'ёзных парушэнняў правоў чалавека, памылак у працы і ўплыву як на персанальнае жыццё, так і на жыццё іншых. Напрыклад, амерыканскі адвакат атрымаў вялікі штраф за прадстаўленне ў суд спасылак на кейсы, якія насамрэч не існавалі, бо былі выдумкай ШІ.

Адной з найбуйнейшых рызыкаў з'яўляецца **дыпфэйкі**, якія могуць выкарыстоўвацца ад агітацыі падчас палітычных кампаній да распаўсюду ў dark net у выглядзе дыпфэйк-порна. Такі распаўсюд стварае асаблівую пагрозу для правоў жанчын[12] і дзяўчынак, якія найбольш часта з'яўляюцца ахвярамі такога тыпу кантэнта.

➔ *Прыклад барацьбы з дыпфэйкамі прадэманстравала Данія. Данія стала першай краінай Еўрапейскага саюза, якая прызнала за сваімі грамадзянамі і грамадзянкамі права валодаць уласнымі тварамі, целамі і голасам. Такія меры накіраваныя на супрацьдзеянне дыпфэйкам і забяспечваюць прававыя механізмы абароны.*

Дыскрымінацыя і алгарытмічная прадузятасць ("algorithmic biases"). Сістэмы штучнага інтэлекту адлюстроўваюць прадузятасці і стэрэатыпы,

[11] AI and Democracy: The Impact of Disinformation, Social Bots and Political Targeting, 2019
[12] Unveiling the Threat: AI and Deepfakes' Impact on Women, Emily Chapman, 2024

уласцівыя людзям. Гэта адбываецца ў выніку якасці, паўнаты і разнастайнасці звестак, на якіх яны навучаюцца. У выніку алгарытмы могуць узнаўляць дыскрымінацыйныя і несправядлівыя мадэлі прыняцця рашэнняў.

Асаблівая рызыка заключаецца ў тым, што сістэмы ШІ часта працуюць як «чорная скрынка»: карыстальнік не бачыць, якім чынам быў прыняты той ці іншы вынік і што адбываецца «ўнутры» мадэлі. Гэта зніжае тлумачальнасць выніковай інфармацыі, падрывае давер да тэхналогіі ШІ ў цэлым і можа ўзмацняць уразлівае становішча асобных груп людзей. Такія праблемы асабліва выразныя ў сферах, дзе выкарыстоўваюцца алгарытмы адбору персаналу, крэдытавання, праваахоўнай дзейнасці або рэкамендацыйных сістэм.

Якасць і аб'ектыўнасць вынікаў працы сістэм штучнага інтэлекту ў значнай ступені залежаць ад паўнаты, дакладнасці і разнастайнасці даных, на якіх яны навучаюцца. Чым больш інклюзіўнай з'яўляецца інфармацыя ў базах (у тым ліку пра людзей рознага ўзросту, культурнай і этнічнай прыналежнасці, тэрытарыяльнага паходжання і сацыяльнага статусу), тым вышэйшая верагоднасць таго, што мадэль будзе забяспечваць больш дакладныя, справядлівыя і збалансаваныя вынікі.

➔ *Неканвенцыйны падыход у 2023 годзе прадэманстравала кампанія Anthropic, чый штучны інтэлект Claude працуе на падставе так званай "канстытуцыі". Падчас навучання Claude выкарыстоўваў шэраг прынцыпаў і выразна сфармуляваных правілаў, праз якія "фільтраваліся" яго адказы (сярод іх ёсць артыкулы Усеагульнай дэкларацыі правоў чалавека). Асноўная мэта гэтай "канстытуцыі" забяспечыць, каб адказы не ўтрымлівалі дыскрымінацыйных, абразлівых або шкодных выказванняў. Хаця такі падыход не гарантуе поўнай абароны ад прадузятасці, ён з'яўляецца важным крокам да распрацоўкі этычнага штучнага інтэлекту.*

Пагроза праву інтэлектуальнай уласнасці. Сістэмы ШІ навучаюцца на той інфармацыі, якая існуе ў інтэрнеце, у тым ліку на чужых творах мастацтва, музыцы, ілюстрацыях і нават асобных персанажах мульцікаў і фільмаў. Гэты кантэнт ахоўваецца правамі ўласнасці (copyright), і за іх карыстанне неабходна плаціць ці пра яго дамаўляцца. Штучны інтэлект, які можа "азнаёміцца" з канкрэтнымі выявамі, пачынае запраграмавана ці выпадкова імітаваць іх стыль у сваіх адказах. Прыкладам з'яўляецца Ghibli AI Image Generator, які маляваў фатаграфіі ў стылі японскага аніматара Міядзакі.

У гэтай праблемы ёсць некалькі адценяў: з улікам таго, што большасць сістэм ШІ з'яўляецца чорнай скрынкай, унутры якой адбываюцца не заўсёды падкантрольныя і празрыстыя працэсы, як можна даказаць выкарыстанне ШІ

першакрыніцы, якая ахоўвалася правам уласнасці? І як тады карыстацца функцыяй генерацыі этычна?

Гэтая пагроза становіцца прадметам дэбатаў і судовых спрэчак. Пры гэтым, судовыя спрэчкі адбываюцца як аўтарамі кантэнту супраць ШІ (Disney і Universal супраць Midjourney), так і тэхналагічнымі кампаніямі супраць творцаў (напрыклад, Мета выйграў судовую спрэчку з групай пісьменнікаў). І пакуль вялікія кампаніі змагаюцца за тое, ці можа згенераваны штучным інтэлектам кантэнт быць абароненым аўтарскім правам, і хто адказны за ўзнаўленне абароненага гэтым правам кантэнту, мы можам дзейнічаць з улікам рэкамендацый EU AI Act па аўтарскім праве, і пазначаць кантэнт, згэнераваны ШІ.[13]

Парушэнне прыватнасці. Сістэмы штучнага інтэлекту, якія выкарыстоўваюцца для відэасачэння, распазнавання эмоцый, а таксама для выстаўлення “сацыяльнай адзнакі” аналізуюць асабістую інфармацыю людзей, часта без іх ведама і згоды. І хаця сістэмы камер відэаназірання існуюць доўгі час, інфармацыя з іх аналізуецца з дапамогай ШІ ў невядомых раней пласкасцях і маштабах. Негледзячы на тое, што ў Еўрасаюзе дзейнічае забарона на падобныя сістэмы ШІ (EU AIAct), існуюць выключэнні для паліцыі і міграцыйных службаў, што дазваляе выкарыстоўваць гэтыя тэхналогіі як зброю рэпрэсій, напрыклад, для пераследу ўдзельнікаў і ўдзельніц мірных пратэстаў.

Іншы бок гэтай праблемы - бескантрольная апрацоўка асабістых звестак, якія маглі быць выкарыстаны для навучання ШІ, у тым ліку звесткі, загружаныя самімі людзьмі. Загружаныя з адной мэтай, нашыя звесткі могуць быць апрацаваныя для навучання алгарытмаў ці для аналітыкі. Нягледзячы на абяцанні захоўваць нашу прыватнасць, непразрыстасць алгарытмаў працы ШІ ставіць пад пагрозу бяспекі асабістай інфармацыі. Напрыклад, ChatGPT у 2025 годзе часова спыняў выдаленне інфармацыі з сервераў, што падкрэслівае праблемы кіравання звесткамі і прыватнасцю ў галіне ШІ.

Пералічаныя вышэй рызыкі не з’явіліся зараз, аднак развіццё ШІ ўзмацняе іх маштаб і патэнцыйную шкоду. Штучны інтэлект стварае значныя пагрозы правам чалавека, калі карыстаецца несвядома, без чалавечага кантролю, ці супраць чалавека. Этычны падыход да штучнага інтэлекта дазваляе мінімізаваць уплыў рызыкаў на вынікі чалавечай працы. Менавіта такі падыход абіраюць недзяржаўныя арганізацыі, каб уключыць новыя тэхналогіі ў сваю дзейнасць.

[13] Generative AI and watermarking, European Parliament, 2023

Як НДА выкарыстоўваюць штучны інтэлект?

З развіццём штучнага інтэлекту недзяржаўныя арганізацыі таксама адаптуюцца і трансфармуюцца, каб заставацца эфектыўнымі, актуальнымі і здольнымі адказваць на выклікі сучаснасці. НДА карыстаюць новыя тэхналогіі для выканання вялікага спектра задач, ад камунікацыі з мэтавай аўдыторыяй да прадказальнай аналітыкі. Функцыі ШІ дапамагаюць рэагаваць на парушэнне правоў чалавека у больш комплексным падыходзе, асабліва з улікам таго, што ўсё часцей правы чалавека парушаюцца з выкарыстаннем самога ШІ.[14]

Мы аб'ядналі магчымыя спосабы карыстання ШІ ў НДА ў тры вялікія групы:[15] апрацоўка вялікіх масіваў інфармацыі, аўтаматызацыя і генерацыя кантэнту.

1. **Апрацоўка вялікіх масіваў інфармацыі,** у тым ліку расшыфроўка відэа ці аўдыё ў тэкст, аналіз апытанняў, навінаў, спадарожнікавых здымкаў і іншых крыніц інфармацыі. Апрацаваная інфармацыя выкарыстоўваецца для прыняцця рэкамендацый ці рашэнняў. Без удзелу ШІ апрацоўка вялікіх масіваў інфармацыі займае значна больш часу, а таксама важная інфармацыя, якая можа мець крытычнае значанне, магла б застацца незаўважнай чалавечым вокам. Штучны інтэлект бачыць тэндэнцыі і трэнды ў вялікім наборы звестак. На падставе вялікіх масіваў інфармацыі і іх аналізу, штучны інтэлект дапамагае НДА ў задачах:

- **Прагназаванне трэндаў і маніторынг рызыкаў,** у тым ліку змянення клімату, прагназаванне гуманітарных катастроф і надзвычайных сітуацый для своечасовай рэакцыі.
    - ➔ *World Wildlife Fund з дапамогай ШІ распрацавалі інструмент Wildlife Insights. Гэта інструмент які дапамагае аўтаматычна ідэнтыфікаваць і маркіраваць жывёлаў на здымках з схаваных камер. У выніку, сістэма апрацоўвае і фільтруе тысячы выяваў за хвіліны, якія аналізуюцца для выяўлення трэндаў у перамяшчэнні жывёлаў і прыняцця мераў па ахове біяразнастайнасці.*[16]

[14] The Role of Artificial Intelligence in Revolutionizing NGOs' Work, 2023

[15] Спіс падрыхтаваны на падставе наступных крыніц::
1) AI Unlocking the Potential of NGOs, 2025
2) AI Handbook for NGOs
3) How nonprofits use AI to find and keep good donors
4) Mapping the Landscape of AI-Powered Nonprofits, 2024

[16] Тут і далей, кейсы бяруцца з наступных сайтаў:
- Top 50 NGOs Leveraging AI to Drive Social Impact and Address Global Challenges.
- Tech to rescue, case study

➔ *International Rescue Committee карыстаецца предыктыўнай аналіткай для прагназавання надзвычайных сітуацый, якія звязаныя з кліматам. Гэтыя сістэмы ранняга апавяшчэння дапамагаюць выяўляць групы людзей, якія найбольш верагодна знаходзяцца пад рызыкай паводак, што дазваляе хутчэй размяркоўваць грашовую дапамогу і эфектыўней рыхтавацца да паводак.*

- **Ацэнка вынікаў і ўплыву дзейнасці НДА** праз збор і аналіз зваротнай сувязі, маніторынг індыкатараў дзейнасці і справаздачнасці.
    ➔ *УВКБ ААН прымяняе генератыўны штучны інтэлект для аналізу зваротнай сувязі ад уцекачоў і ўцякачак. Яны могуць пакінуць інфармацыю праз званок ці тэкст у WhatsApp, і сістэма пасля расшыфроўвае і класіфікуе гэтыя паведамленні, а таксама візуалізуе іх на мапе, каб вызначыць месцазнаходжанне праблемных месцаў.*
- **Вывучэнне мэтавай аўдыторыі і праца з уразлівымі групамі,** распрацоўка сацыяльных праграм, вывучэнне патрэбаў і аналіз уплыву. ШІ можа значна павялічыць інклюзіўнасць працы НДА, калі выкарыстоўваецца для аналізу ўключанасці ўразлівых груп у праект, паляпшэння камунікацыі, забяспячэння доступу да паслуг і прадуктаў аўдыторый з розных рэгіёнаў.
    ➔ *The Lily Project выступае за адукацыю ў галіне сексуальнага здароўя жанчын па ўсёй Лацінскай Амерыцы, і пры падтрымцы Tech to Rescue распрацавала аплікацыю пад назвай Chava, якая ўтрымлівае інфармацыю пра сэксуальнае здароўе для лацінаамерыканак. Доступ да аплікацыі забяспечвае персаналізаваную і канфідэнцыйную адукацыю і медычную падтрымку ў галіне сэксуальнага здароўя.*

2. **Аўтаматызацыя руцінных задач,** як адна з найбольш “папулярных” функцый ШІ, дапамагае паскорыць працу і вызваліць час на выкананне іншых задач.

- З дапамогай аўтаматызацыі **можна збіраць і аналізаваць фотаздымкі, навіны, інфармацыю** з сацыяльных сетак і апрацоўваць яе значна хутчэй.
    ➔ *Amnesty International у супрацоўніцтве з арганізацыяй Element AI заснавалі праект Troll Patrol, у рамках якога з дапамогай машыннага навучання было сабрана і апрацавана больш за 288 000 твітаў, якія ўтрымліваюць абразы і пагрозы ў адносінах да жанчын-палітыкаў і журналістак з ЗША і Вялікабрытаніі. У тым ліку, была распрацавана аўтаматызацыя выяўлення абразлівых паведамленняў для яе выкарыстання ў будучых праектах.*
- **Чат-боты** дапамагаюць выбудаваць кругласутачную камунікацыю з мэтавай аўдыторыяй, прапаноўваць навучальную інфармацыю і падтрымку

➔ *ЮНІСЕФ выкарыстоўвае чат-бот U-Report для ўзаемадзеяння з моладдзю па ўсім свеце па такіх пытаннях, як адукацыя, ахова здароўя і правы чалавека.*

3. **Генерацыя кантэнту і візуальных матэрыялаў** актыўна выкарыстоўваецца для напаўнення сацыяльных сетак, суправаджэння медыя-матэрыялаў, працы з кантэнтам.

- **Стварэнне візуала для медыя-платформаў, а таксама генерацыя ШІ-аватараў,** для візуалізацыі праблем і персаналізацыі камунікацыі, асабліва ў сітуацыі асабістай пагрозы рэальным прадстаўнікам арганізацый ці для ўздымання пытанняў парушэння правоў чалавека.
  - ➔ *Беларуская праваабарончая арганізацыя Human Constanta стварыла ШІ-аватар праваабаронцы і палітычнай зняволенай Насты Лойкі, каб яе голасам гучалі навіны ў сферы правоў чалавека.*
  - ➔ *Polish Smog Alert разам з Brainhub распрацавала інтэрактыўны інструмент "Віртуальныя лёгкія", які аналізуе інфармацыю пра забруджванне паветра ў рэальным часе і візуалізуе яе праз мадэляванне ўплыву забруджвання паветра на здароўе лёгкіх. Візуальны інструмент суправаджаецца адукацыйнай інфармацыяй.*
  - ➔ *Amnesty International у 2021 годзе падчас пратэстаў у Калумбіі суправаджала навіны выявай, якая была згенераваная ШІ, замест рэальнага фота пратэстоўцаў, каб абараніць іх ад ідэнтыфікацыі і пераследу за ўдзел у мітынгах.*
- **Пераклады і рэдакцыя тэкстаў з дапамогай ШІ**, што дае магчымасць перакладаць матэрыялы на рэдкія мовы, і перакладаць тэксты, нават калі няма носьбітаў мовы ў камандзе. Гэта робіць інфармацыю даступнай шырэйшай аўдыторыі, дапамагае ствараць інклюзіўныя сервісы і паслугі.
  - ➔ *Lenovo разам з НДА CESAR прэзентавала тэхналогію на базе ШІ, якая можа перакладаць бразільскую мову жэстаў у тэкст і гукавую мову ў рэальным часе.*
  - ➔ *Lenovo таксама падтрымалі некамерцыйны фонд Scott-Morgan Foundation, і з выкарыстаннем DeepBrain AI стварылі ШІ-аватар, які захоўвае голас, асобу і асаблівасці паводзінаў людзей з інваліднасцю, што стварае для іх новыя магчымасці камунікацыі.*
- **Персаналізаванае навучанне і праца з мэтавымі аўдыторыямі** праз кантэнт, які часткова ці цалкам згенераваны ШІ.
  - ➔ *Rocket Learning (НДА ў Індыі) разам з Google.org распрацавалі ШІ-настаўніка ў WhatsApp, які генеруе кантэнт на ключавых індыйскіх мовах, выкарыстоўвае прадказальнае мадэляванне для выяўлення дзяцей з рызыкай адставання і прадастаўляе аналітыку для*

адсочвання прагрэсу. ШІ-прыкладанне выкарыстоўваецца для паляпшэння навучання і павелічэння доступу да адукацыі дзяцей і іх бацькоў, асабліва з уразлівых грамадскіх груп.

Прыклады вышэй падкрэсліваюць, што штучны інтэлект стварае новыя магчымасці для НДА, робіць працу больш хуткай, дапамагае аўтаматызаваць працэсы, дазваляе вызваліць час на стратэгічнае планаванне і прадумванне новай дзейнасці. Пры гэтым, кожная магчымасць і яе ўкараненне павінны ісці побач з ацэнкай уплыву канкрэтнага тэхналагічнага рашэння на правы чалавека, прагназаваннем рызыкаў, і планаваннем змяньшэння іх уплыву, каб не стварыць дадатковыя пагрозы дыскрымінацыі, выключэння, парушэння прыватнасці і іншых. У тым ліку, НДА рызыкуюць страціць давер, калі будуць несвядома распаўсюджваць памылковыя даныя, парушаць прыватнасць сваёй аўдыторыі ці сваёй інфармацыі, а таксама калі будуць прымаць прадузятыя рашэння праз неапазнаную дыскрымінацыю ў працы ШІ. Таму асабліва важна захоўваць этычныя падыходы да імплементацыі ШІ.

Што такое этычны штучны інтэлект?

Дэзынфармацыя, алгарытмічная прадузятасць, пагроза ўцечкі персанальных звестак і іншыя рызыкі сталі падмуркам для распрацоўкі міжнародных стандартаў і прынцыпаў «этычнага штучнага інтэлекту». Калі параўнаць прынцыпы, якія пералічаныя ў гайдах Еўрапейскай камісіі па надзейным ШІ, і рэкамендацыі ЮНЭСКА па этычным карыстанні ШІ, можна пабачыць падабенства ў падыходах. Сярод асноўных прынцыпаў называюцца **празрыстасць, адказнасць, прыватнасць (абарона персанальных звестак), недыскрымінацыя і чалавечы надзор.**

Гэтыя прынцыпы этычнага выкарыстання ШІ маюць асаблівае значэнне для недзяржаўных арганізацый, паколькі дапамагаюць захоўваць этычныя межы пры дасягненні сваіх мэтаў і місіі, асабліва пры карыстанні новымі тэхналогіямі. На падставе агульных міжнародных рэкамендацый і прынцыпаў, можна вылучыць **практычныя рэкамендацыі для НДА:**[17]

[17] Рэкамендацыі падрыхтаваныя на падставе наступных крыніц:
- Principles for the ethical use of artificial intelligence in the United Nations system, 2022
- Considerations for AI Use in International Non-Governmental Organizations, 2025
- AI and NGOs: How to Use Artificial Intelligence Responsibly in Nonprofit Organizations, 2025
- Principles for trustworthy AI, OECD AI Principles overview

- **Чалавечы кантроль.** Штучны інтэлект з’яўляецца інструментам, і менавіта чалавек застаецца адказным за яго прымяненне. Менавіта людзі ажыццяўляюць крытычную ацэнку вынікаў ШІ, праводзяць пераправерку і маніторынг, а таксама маюць магчымасць кіраваць працэсамі працы сістэм ШІ: вырашаць, калі і як іх выкарыстоўваць, адмяняць або пераглядаць іх рашэнні. Такі падыход забяспечвае захаванне чалавечай аўтаноміі і адказнасці ў адносінах да тэхналогій.
- **Маніторынг і зніжэнне алгарытмічнай прадузятасці.** Неабходна быць асвядомленымі, што ШІ можа быць прадузятым і схільным да дыскрымінацыі, таму яго трэба выкарыстоўваць з разуменнем яго абмежаванняў. Варта ведаць пра яго магчымыя культурныя, моўныя або ідэалагічныя прадузятасці, якія могуць адлюстроўвацца, ці наадварот ігнаравацца ў выніках, і быць асабліва ўважлівымі да сацыяльнага і эмацыйнага кантэксту, у якім працуе арганізацыя, бо яны могуць патрабаваць дадатковай працы і праверкі па забеспячэнні інклюзіўнасці і нейтральнасці.
- **Празрыстасць і тлумачальнасць**. Пра выкарыстанні штучнага інтэлекта для стварэння, рэдагавання, праверкі ці іншай працы над кантэнтам варта паведамляць адкрыта. Варта пазначыць не толькі сам факт выкарыстання ШІ, але і паспрабаваць прапісаць для чаго ён быў выкарастаны, яго ролю ў працэсе. Празрыстасць забяспечвае давер да крыніцы інфармацыі і дазваляе зразумець межы чалавечага кантролю. Акрамя таго, калі нейкія рашэнні былі прынятыя машынай, а не чалавекам, пра гэта таксама варта паведамляць адкрыта, у тым ліку патлумачыць, якія былі выкарыстаны алгарытмы, як яны былі праверaны, хто быў адказны за імплементацыю рашэнняў. Тлумачальнасць дазваляе атрымальнікам паслуг і людзям, якіх датычны рашэнні, захоўваць кантроль над сітуацыяй, сваімі звесткамі, і ў выніку памылкі звярнуцца за тлумачэннем.
- **Абарона прыватнасці.** Персанальная інфармацыя строга абараняецца, і праца з штучным інтэлектам будуецца па прынцыпу “нулявога даверу” (zero trust) - адсутнасць даверу да сістэм ШІ, якая патрабуе не перадаваць прыватныя або канфідэнцыйныя звесткі інструментам ШІ. У працы з ШІ трэба ведаць, што любая інфармацыя можа быць захаваная супраць згоды, узноўленая ў будучыні ў працы ШІ і карыстацца для навучання мадэляў. Аўдыторыя НДА павінна ведаць, як і з якой мэтай іх персанальная інфармацыя можа быць выкарыстана, і мець права адклікаць згоду на яе апрацоўку.

- Explained: Generative AI’s environmental impact, 2025

- **Падсправаздачнасць і адказнасць за вынікі.** Штучны інтэлект не нясе адказнасці за вынікі сваёй працы, таму менавіта чалавек застаецца адказным за ўсе рашэнні, прынятыя з выкарыстаннем сістэм ШІ, за маніторынгам іх тэставання і імплементацыі, рэалізацыяй рэкамендацый і ацэнкай вынікаў. Асобная ўвага надаецца забеспячэнню выканання этычных прынцыпаў і стварэнню празрыстых і зразумелых механізмаў прававой абароны, якія дазваляюць гарантаваць справядлівасць пры выкарыстанні ШІ.
- **Захоўваць экалагічны фокус.** Навучанне, апрацоўка звестак і вывад выніка патрабуюць значнага энергаспажывання. Навучанне і выкарыстанне мадэляў ШІ, асабліва глыбокага навучання і генератыўных мадэляў, суправаджаюцца значнымі экалагічнымі выдаткамі вады і энергіі. Таму варта аптымізаваць карыстанне штучным інтэлектам, скарачаць непатрэбныя запыты і ўкараняць экалагічныя падыходы.

На падставе рэкамендацый, дзелімся з вамі чэк-лістом этычнага падыходу да ШІ з боку НДА. Ён дапаможа імплементаваць уключэнне штучнага інтэлекту ў дзейнасць арганізацыі ў этычных межах:

Чэк-ліст этычнага ШІ для НДА

- ☐ **Аналіз магчымасцяў:** Прааналізуйце розныя тэхналогіі штучнага інтэлекту *(чат-боты, генератыўныя мадэлі, аналітыка, расшыфроўка і іншыя)* і спытайце сябе: як вашая каманда можа выкарыстоўваць штучны інтэлект? Якія задачы ён можа аўтаматызаваць, паскорыць ці палепшыць? *(пераклады, расшыфроўкі, маніторынг, справаздачнасць)*;
- ☐ **Аналіз рызыкаў:** Пракансультуйцеся па бяспецы, якія ў вас могуць быць рызыкі ў карыстанні ШІ і як іх пазбегнуць, напрыклад з пункту гледжання выканання абавязку захоўваць прыватнасць звестак, бяспечнага доступу да інфармацыі, з боку прадузятасці, недахопу інфармацыі ў вашых базах, нетлумачальнасці вынікаў ШІ;
- ☐ **Тэставанне:** Пратэстуйце розныя тэхналогіі на якасць вынікаў, паспрабуйце эксперыментаваць з роўнымі сістэмамі штучнага інтэлекта. Абярыце тыя сістэмы ШІ, што адказваюць на вашыя патрэбы па функцыяналу і бяспецы, дзе дасягаюцца лепшыя вынікі і лепшая якасць;
- ☐ **Навучанне:** Правядзіце вебінары для вашых каманд па тым, якія ёсць рызыкі ў карыстанні ШІ, як карыстацца бяспечна, правяраць факты і працаваць з выніковай інфармацыяй. Распрацуйце трэнінгі па пісьменасці ў галіне штучнага інтэлекту. Абмяркуйце, як захаваць чалавечы надзор і кантроль над вынікамі працы ШІ;
- ☐ **Вызначэнне адказных:** хто і як правярае вынікі ШІ, хто адказны за памылкі ў выніку працы з ШІ. Забяспечце зразумелую магчымасць звярнуцца да вас за тлумачэннем прынятага рашэння ці ў выпадку памылкі;
- ☐ **Унутраныя палітыкі карыстання:** Распрацуйце ўнутраныя рэкамендацыі ці гайды па працы з ШІ: у якіх кейсах ён можа карыстацца, а ў якіх не, якія персанальныя звесткі ніколі не павінны праходзіць ШІ-апрацоўку, хто адказны і за якія рашэнні, якія сістэмы ШІ бяспечныя ў карыстанні, як адбываецца ацэнка рызыкаў і як часта яна паўтараецца.
- ☐ **Захаванне экалагічнага падыходу:** Дамоўцеся, якія захады можа зрабіць вашая арганізацыя ці вы персанальна, каб захаваць экалагічны падыход, напрыклад, як можна аптымізаваць звароты да ШІ, у якіх выпадках можна да яго не звяртацца, якія сістэмы ШІ прытрымліваюцца найбольш устойлівага падыходу.
- ☐ **Празрыстасць:** Калі вы карыстаецеся штучным інтэлектам для стварэння кантэнту, ці для імплементацыі рашэнняў, якія маюць уплыў на вашую аўдыторыю, паведамляйце пра гэта адкрыта: які і для чаго быў выкарыстаны ШІ, у якой ступені ён меў уплыў, ці быў пасля апрацаваны аўтаматычны вынік чалавекам. Паведамляйце вашай мэтавай аўдыторыі пра маніпуляцыі з ШІ, якія вы робіце ў дачыненні да іх. Гэта захавае давер і пакажа вашую адказнасць за працу з ШІ.

Заключэнне

У гэтым артыкуле мы хацелі паказаць узаемасувязь паміж штучным інтэлектам і правамі чалавека, наколькі ўзаемазвязанымі і непарыўнымі з’яўляюцца рызыкі і магчымасці ў карыстанні штучнага інтэлекта ў НДА, і як этычны падыход да імплементацыі ШІ у дзейнасць НДА можа дапамагчы мінімізаваць пагрозы правам чалавека. Сфера штучнага інтэлекта з’яўляецца адносна новай, і пошук этычных арыентыраў дапамагае захаваць чалавека цэнтрам прыняцця рашэнняў, а сам працэс адаптацыі і карыстання ШІ рабіць адкрытым і празрыстым. Мы заклікаем НДА, актывістаў і актывістак, энтузіастаў і энтузіастак дзяліцца сваім досведам, эксперыментаваць у межах бяспечнага і этычнага карыстання ШІ.

Чэк-лістом этычнага карыстання мы спрабавалі задаць рэкамендацыйныя межы, якія могуць дапамагчы арганізацыям, актывістам і актывісткам рэалізоўваць свае каштоўнасці і місію, і пры гэтым адаптаваць новыя тэхналогіі. Мы арыентаваліся на міжнародныя стандарты і нормы, і мы адкрытыя да дыялёгу і дыскусій у гэтай сферы, і заклікаем да сумеснага пошуку лепшых практык і рашэнняў.