ETHICAL AI
MACEDONIA

ПРИНЦИПИ ЗА ОДГОВОРНА УПОТРЕБА НА ВИ

ВО СЕВЕРНА МАКЕДОНИЈА

Доброволна рамка за подигнување на свеста за одговорна примена на ВИ

Подготвени од Ethical AI Macedonia

Прво издание · Верзија 1.0

© 2026 Ethical AI Macedonia. Сите права се задржани.

Содржина

За Ethical AI Macedonia . . . 3

Обраќање на основачите . . . 4

Вовед: цел и контекст . . . 6

Дефиниции . . . 8

Принципите . . . 9

1. Човечко достоинство и фундаментални права . . . 10

2. Правичност, еднаквост и инклузија . . . 11

3. Транспарентност, објаснивост и следливост . . . 12

4. Управување со податоци и интегритет . . . 13

5. Безбедност, робустност и доверливост . . . 14

6. Учество на луѓето, надзор и компетентност . . . 15

7. Одговорност, управување и правни лекови . . . 16

8. Партиципација, локален контекст и општествена легитимност . . . 17

Благодарност . . . 19

За Ethical AI Macedonia

Поддршка на примена на ВИ преку одговорно користење.

Ethical AI Macedonia ги поврзува бизнисите и експертите за да разговараат како ВИ може пошироко и поодговорно да се користи во Северна Македонија. Во фокусот се одлуките со кои организациите веќе се соочуваат: на кои алатки да им веруваат, кои податоци да ги споделуваат, кога да го проверат одговорот и како да ги зачуваат стручната проценка и вештините.

Принципите се нашата појдовна точка. Преку истражувања меѓу компаниите и дискусии сакаме да учиме од вистинските искуства и да ги унапредуваме насоките во согласност со локалните услови. Особено се насочуваме кон помалите бизниси и професионалните канцеларии, кои често ги носат овие одлуки без посебни технички тимови или тимови за усогласеност со прописите.



Владимир Бошњаковски

Коосновач

Владимир е адвокат чија работа е насочена кон заштита на личните податоци, вклучувајќи здравствени податоци, како и советување софтверски компании и работа со дигитални услуги. Дипломирал и магистрирал право на Универзитетот „Св. Кирил и Методиј“ во Скопје, а завршил и Executive MBA на Wirtschaftsuniversität Wien, со фокус на науката за податоци и дигиталната трансформација. Неговото истражување ја обработува објаснивоста според Актот на ЕУ за вештачка интелигенција (EU AI Act).



Ангела Анѓеловска

Коосновачка

Ангела е адвокатка со искуство во онлајн платформи, дигитален идентитет и граѓански организации. Извршувала раководни функции во ELSA Македонија и во Националниот младински совет на Македонија. Дипломирала и магистрирала право на Универзитетот „Св. Кирил и Методиј“ во Скопје. Во иницијативата го вградува своето искуство во граѓанскиот ангажман и организациската работа.

Обраќање на основачите

ВИ се развива побрзо од јавната дискусија за нејзината употреба. Нè инспирираат можностите што им ги отвора на бизнисите, професионалците и општеството: поголема продуктивност, поширок пристап до знаење и нови можности за иновации и развој на нови деловни модели. Истовремено, користењето ВИ без доволно разбирање на нејзините ограничувања и ризици може негативно да влијае врз основните права, личната автономија, професионалните вештини и угледот на организацијата. Развивањето способности за разбирање и ублажување на овие ризици ќе биде предизвик споредлив со технолошкиот напор потребен за создавање сè поспособни модели на ВИ. Во тие околности ја основавме Ethical AI Macedonia, со цел да ја подигнеме свеста кај бизнисите во Северна Македонија за важноста на одговорната употреба на ВИ и за тоа што таа значи во практика.

Иницијативата ја основавме со поддршка за технологијата, а не како спротивставување на неа. Нè води надежта дека поголемата одговорност при користењето ВИ ќе овозможи нејзина поширока примена со поголема доверба. Сепак, веруваме дека широката и одржлива употреба зависи од тоа организациите да ги разбираат ризиците за приватноста, автономијата, еднаквоста, доверливоста, заштитата и професионалните вештини и соодветно да постапуваат со нив. Употребата без доволно познавање или без соодветно управување може да предизвика штета што можела да се избегне, да им наштети на организациите и да поттикне поширока јавна загриженост. Со тоа се намалува довербата на потрошувачите, а бизнисите и институциите можат да станат уште повоздржани кон користењето ВИ или вложувањето во неа. Особено претпазливите организации можеби нема да сакаат ни да експериментираат без појасно разбирање на ризиците и начините за нивно управување. Така можат да пропуштат вредни можности и придобивки во продуктивноста.

Овие Принципи се првиот чекор на тој пат. Тие се потпираат на заедничките принципи застапени во воспоставени меѓународни и европски рамки, како и во национални рамки усвоени во други јурисдикции. Нивниот посебен придонес е во тоа што ги разгледуваат овие широко признати стандарди во околностите на Северна Македонија. Затоа се подготвени со особено внимание кон: i. работењето на ВИ на јазици и дијалектни варијанти со поограничени дигитални ресурси; ii. степенот на дигитализација, достапноста на податоци, пристапот до технологија и писменоста за ВИ во Северна Македонија; iii. техничките, финансиските и организациските капацитети на помалите бизниси; и iv. ризикот системите развиени за поголеми пазари да не ги одразуваат соодветно локалните закони, административните практики, културите или општествените услови.

Сметаме дека овие Принципи доаѓаат во вистинско време. Северна Македонија сè уште нема општоприменлив збир етички принципи или насоки за одговорна употреба на ВИ од бизнисите и другите организации, иако земји од регионот со во голема мера споредливи услови усвоиле слични насоки или стандарди. Воедно ги цениме напорите на Министерството за дигитална трансформација, иницијативата Везилка и другите чинители што работат во оваа област. Овие Принципи треба да ги надополнат тие напори и да внесат практична деловна перспектива во националната дебата за тоа како ВИ може одговорно да се користи во околностите на Северна Македонија.

Овие Принципи немаат регулаторни овластувања и не се обидуваат да дадат конечен одговор на секое прашање. Тие се практична појдовна точка, а не завршен производ. Ги повикуваме организациите, експертите, институциите, граѓанските организации и поширокаата заедница да ги испробаат и критички да ги разгледаат. Така идните верзии ќе можат да се менуваат и унапредуваат врз основа на практичното искуство, конструктивната дебата и натамошниот технолошки и правен развој.

V. Boshnjakovski

Владимир Бошњаковски
Коосновач, Ethical AI Macedonia

A. Angjelovska

Ангела Анѓеловска
Коосновачка, Ethical AI Macedonia

Вовед: цел и контекст

Нашата мисија. Овие Принципи имаат цел да им помогнат на бизнисите и другите организации во Северна Македонија да ја применуваат вештачката интелигенција со поголема доверба, со потребното знаење и со човекот во центарот на вниманието. Тие нудат заедничка етичка рамка за разбирање на одговорната употреба на ВИ во практика и за разгледување на нејзините придобивки, ограничувања и ризици пред да се донесат одлуки со значајни последици.

Меѓународни основи. Принципите делумно се инспирирани од постојни, воспоставени меѓународни и национални документи. Се потпираат на заедничките теми застапени во Препораката на УНЕСКО за етика на вештачката интелигенција, Принципите на ОЕЦД за ВИ, Рамковната конвенција на Советот на Европа за вештачка интелигенција и човекови права, демократија и владеење на правото, Актот на Европската Унија за вештачка интелигенција (EU AI Act) и практичните рамки за управување со ризици на организациите за стандардизација.

Овие инструменти се разликуваат по својот опсег и правна природа. Некои утврдуваат правно обврзувачки должности, а други нудат етички принципи, определби во јавните политики или практични методи за управување со ризиците поврзани со ВИ. Сепак, се засноваат на заедничко разбирање: ВИ треба да се развива и користи со почитување на човечкото достоинство и основните права, да ја зачува вистинската способност на луѓето самостојно да одлучуваат, да ги земе предвид предвидливите ризици и да обезбеди одговорност за нејзината употреба и последици.

Принципите немаат цел да ги заменат овие рамки или да им конкурираат. Ниту, пак, тврдат дека воведуваат целосно нов етички модел. Се надоврзуваат на нивните широко признати основи и ги изразуваат во облик што е разбирлив, релевантен и практично применлив за организациите што работат во Северна Македонија.

Регионален контекст. Земји со различна големина и степен на технолошки развој развиваат сопствени пристапи кон одговорната ВИ. Србија усвои етички насоки за развој, примена и употреба на доверлива и одговорна ВИ. Албанија воведе методологија и технички стандарди за користење ВИ во јавните институции. Словенија ја вклучи доверливата ВИ, насочена кон човекот, во својата национална политика.

Овие иницијативи покажуваат дека прашањата за одговорната ВИ не можат да се разгледуваат само преку документи подготвени за најголемите светски економии, пазари или технолошки компании. Секоја земја треба да разгледа и како ВИ се вклопува во нејзините институции, јазици, економија, технолошки капацитет и општествени услови.

Контекстот во Северна Македонија. Северна Македонија не почнува од нула. SMART/MK 2030 ја вклучува ВИ во поширокатa агенда за дигитален развој на земјата. Иницијативата Везилка е насочена кон зајакнување на домашните капацитети за ВИ, а продолжува и работата на национална стратегија за ВИ и на други иницијативи во јавните политики. Северна Македонија ја потпиша и Рамковната конвенција на Советот на Европа за вештачка интелигенција. Во рамките на процесот на европска интеграција се очекува својата правна и институционална рамка да ја усогласи со Актот на ЕУ за вештачка интелигенција (EU AI Act). Домашните чинители придонесуваат и преку истражување, образование, развој на капацитети и работа во одделни сектори, вклучувајќи насоки за одговорна употреба на ВИ во новинарството.

Овие иницијативи се вредни, но имаат различна намена. Национална стратегија, инфраструктурен проект, меѓународна определба или документ за политики во одделен сектор не им обезбедува нужно на бизнисите и другите организации општоприменлива етичка рамка за практичните одлуки што веќе ги носат. Во времето на објавувањето, во прегледаните официјални извори не беше најден споредлив и пристапен збир принципи што опфаќа повеќе сектори и е насочен конкретно кон одговорната употреба на ВИ во бизнисите и другите приватни организации во Северна Македонија.

Оттука, она што недостига не е технолошка активност или амбиција, туку заедничка етичка рамка заснована врз локалните услови. Организациите можат да ја користат при избор на алатки за ВИ,

набавка на услуги за ВИ, развој или приспособување системи, обработка на информации, потпирање на резултати создадени со ВИ или процена на влијанието на нивната употреба на ВИ врз вработените, клиентите и другите луѓе.

Посебниот придонес на Принципите. Принципите ги разгледуваат воспоставените меѓународни стандарди во условите во кои работат организациите во Северна Македонија. Затоа посебно внимание им посветуваат на:

- работењето на ВИ на македонски, албански и други јазици и дијалектни варијанти со поограничени дигитални ресурси;
- различните степени на дигитализација, поврзаност, достапност на податоци, пристап до технологија и писменост за ВИ;
- техничките, финансиските и организациските капацитети на помалите компании и институции; и
- можноста системите развиени за поголеми странски пазари да не ги одразуваат соодветно локалните закони, административни практики, професионални стандарди, култури или општествени услови.

Оваа локална перспектива не ја менува универзалната важност на човечкото достоинство, еднаквоста, автономијата, безбедноста и одговорноста. Нејзината цел е овие вредности да добијат поконкретно значење, со земање предвид на практичните услови во кои треба да се применуваат.

Природа и статус. Овие Принципи се доброволна рамка за подигнување на свеста и појдовна точка за натамошна дискусија. Тие не се законски прописи, официјална владина политика, сертификациска шема, прирачник за усогласување со прописите, технички стандард или план за спроведување. Самото следење на Принципите не докажува усогласеност со законите. Тие не ги заменуваат применливите прописи, професионалната одговорност, договорните обврски или стручните совети.

Принципите треба да им помогнат на организациите да ги утврдат релевантните прашања, да препознаат кој може да биде засегнат и да одредат кога се потребни заштитни мерки, човечки надзор или дополнително стручно знаење. Нивната примена треба да биде сразмерна со ризикот, со поголемо внимание кога ВИ може значително да влијае врз основните права, безбедноста, егзистенцијата, пристапот до можности или други важни интереси.

Понатамошен развој. Оваа прва верзија намерно е оставена отворена за критичко разгледување и подобрување. По објавувањето, Принципите ќе бидат претставени на консултации, тркалезни маси, работилници и други јавни и стручни дискусии. Бизнисите, развивачите и давателите на услуги за ВИ, стопанските комори, професионалните здруженија, јавните институции, универзитетите, истражувачите, граѓанските организации, работниците и нивните претставници, како и експертите од релевантните области, ќе бидат поканети да ги разгледаат и да дадат критички коментари.

Целта на овој процес е да се провери дали Принципите се разбирливи, релевантни и реално применливи. Во идните верзии тие можат да бидат појаснети, изменети, проширени или поинаку организирани врз основа на практичното искуство, конструктивната критика и натамошниот технолошки, правен и општествен развој.

Цел на Принципите. Принципите ги имаат следниве цели:

- да ја подигнат свеста за можностите, ограничувањата и ризиците поврзани со ВИ;
- да воспостават заедничка терминологија за дискусија за одговорната употреба на ВИ;
- да им помогнат на организациите да ги постават соодветните прашања пред и за време на користењето ВИ;
- да поттикнат примена на заштитни мерки сразмерни со можните влијанија врз луѓето и организациите;
- да овозможат поширока примена со поголема доверба преку намалување на неизвесноста и штетите што можат да се избегнат; и
- да внесат во националната дискусија за ВИ практична перспектива од работењето на организациите.

Дефиниции

Овие работни дефиниции служат само за полесно разбирање на Принципите. Тие не ги заменуваат дефинициите утврдени во применливите прописи.

- **Систем за ВИ** - систем заснован на машинска обработка што користи влезни податоци за да создаде резултати, како предвидувања, содржини, препораки или одлуки, кои можат да влијаат врз физичката или дигиталната средина.
- **Одлука или процес со помош на ВИ** - одлука или процес во кој ВИ препорачува, рангира, филтрира, генерира или на друг начин значително влијае врз некое дејствие, дури и кога конечната одлука формално ја носи човек.
- **Засегнато лице или група** - секое лице или група врз чии права, интереси, можности, работни услови, безбедност или пристап до услуги може директно или индиректно да влијае систем за ВИ.
- **Писменост за ВИ** - знаењето и практичното разбирање потребни за соодветно користење, оценување и надзор на ВИ, вклучувајќи познавање на нејзините способности, ограничувања и ризици.
- **Употреба или одлука со значајни последици** - употреба или одлука што може значително да влијае врз нечии права, безбедност, егзистенција, вработување, образование, пристап до основни услуги, правна положба или други слично важни интереси.
- **Суштински човечки надзор** - надзор што го врши лице со доволно информации, компетентност, време, овластувања и практична можност да ја преиспита употребата на систем за ВИ, да ја преземе контролата врз неговото автоматизирано работење, привремено да ја запре или да ја прекине неговата употреба.
- **Лични податоци** - информации што се однесуваат на идентификувано лице или лице што може да се идентификува. Правно обврзувачкото значење на поимот и барањата за заштита се утврдени со Законот за заштита на личните податоци и другите применливи прописи.
- **Давател** - внатрешен или надворешен субјект што развива или испорачува систем за ВИ, модел на ВИ или поврзана услуга.
- **Организација-корисник** - организација што избира, набавува, интегрира, управува со систем за ВИ или го користи во професионална или институционална дејност.

Принципите

Осум групи на принципи

Осумте групи принципи подолу треба да се читаат заедно. Тие нудат етичка рамка за размислување и дискусија и не ги заменуваат применливите законски или професионални должности.

1 Човечко достоинство и фундаментални права

2 Правичност, еднаквост и инклузија

3 Транспарентност, објаснивост и следливост

4 Управување со податоци и интегритет

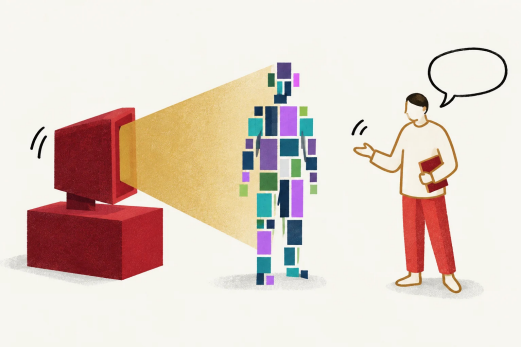
5 Безбедност, робустност и доверливост

6 Учество на луѓето, надзор и компетентност

7 Одговорност, управување и правни лекови

8 Партиципација, локален контекст и општествена легитимност

1. Човечко достоинство и фундаментални права



1.1 Почитување на човечкото достоинство

Вештачката интелигенција треба да се развива и користи на начин што ја признава вродената вредност на секој поединец. Ниту едно лице не смее да се третира само како податок или како средство за постигнување организациска цел. Системите со вештачка интелигенција не смеат да ги објективизираат, искористуваат или дехуманизираат луѓето, ниту да им припишуваат морална или општествена вредност врз основа на изведени особини, однесување или припадност на група.

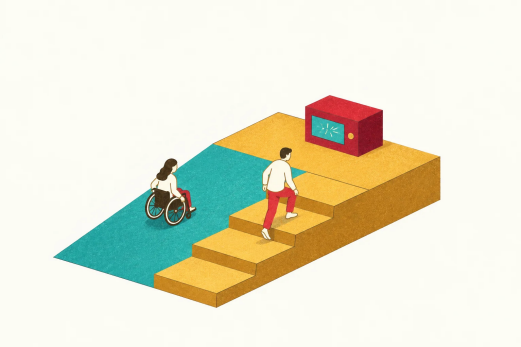
1.2 Заштита на фундаменталните права и слободи

Развојот и употребата на вештачката интелигенција треба да бидат во согласност со фундаменталните права и слободи. Организациите треба да ги идентификуваат правата што може да бидат засегнати, да ги проценат директните и индиректните влијанија врз нив и да усвојат заштитни мерки за спречување или намалување на негативните последици. Оваа одговорност важи во текот на целиот животен циклус на системот и не може да се исполни само со договорни одредби, согласност од корисникот или потпирање на давател на технолошки услуги.

1.3 Човечка автономија и самоопределување

Вештачката интелигенција треба да ја поддржува, а не да ја заменува способноста на луѓето самостојно да ги формираат, изразуваат и менуваат своите избори. Системите не смеат да користат прикриена манипулација, искористувачко убедување или намерно искористување на нечија ранливост со цел да го насочат лицето кон одлуки што првенствено им служат на интересите на организацијата или на трето лице.

2. Правичност, еднаквост и инклузија



2.1 Забрана на дискриминација и спречување пристрасност

Вештачката интелигенција не смее да создава, засилува или легитимира неоправдана неповолност врз основа на заштитени лични својства или споредливи лични и општествени околности. Забраната се однесува и на употребата на варијабли што служат како посредни показатели за својства кои системот не смее непосредно да ги користи.

2.2 Материјална правичност

Вештачката интелигенција треба да се оценува не само според тоа дали го применува истото формално правило кон сите, туку и според тоа дали води до оправдани исходи во реалните услови во кои живеат луѓето. Различните околности може да бараат различни заштитни мерки или поддршка за секој поединец да има вистинска правична можност да оствари право, да добие услуга или корист, или да биде предмет на професионална процена.

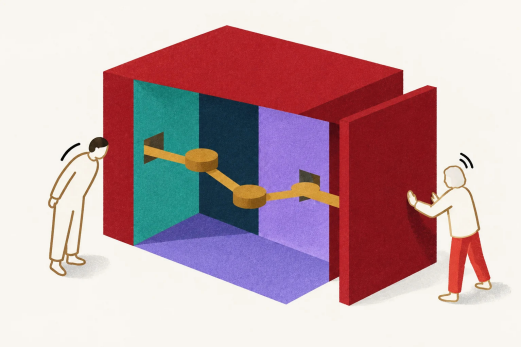
2.3 Процесна правичност

Постапките во кои вештачката интелигенција помага при донесување одлуки со значителни последици треба навремено да обезбедат известување, пристап до релевантни информации, разбирливо образложение и вистинска можност лицето да оспори грешка или да изнесе дополнителни факти пред непристрасно и овластено лице што е надлежно да донесе одлука.

2.4 Посебна заштита и пристапност за децата и ранливите лица

За употребата на вештачка интелигенција што ги засега децата, лицата со попреченост, повозрасните лица, лицата што економски зависат од други или други лица во особено ранлива положба треба да се предвидат посилни заштитни мерки и дизајн што овозможува пристап. Организациите треба да ги земат предвид намалената можност за преговарање, ограничената писменост или техничка способност, зависноста од основни услуги, пречките поврзани со интернет-поврзаноста и опремата, како и можноста согласноста или формалниот избор да не ја одразуваат вистинската слобода на лицето.

3. Транспарентност, објаснивост и следливост



3.1 Известување за употребата на ВИ и за манипулирана содржина

Лицата треба јасно да бидат известени кога комуницираат со систем со вештачка интелигенција или кога вештачката интелигенција значително влијае врз постапка што ги засега. Текст, слики, аудио или видео создадени или значително изменети со вештачка интелигенција треба јасно да се означат кога нивното вештачко потекло е важно за автентичноста, разбирањето во јавноста, правното докажување или утврдувањето одговорност. Треба да се зачуваат и соодветни информации за нивното потекло. Известувањето треба да биде навремено, видливо и формулирано на јазик соодветен за публиката.

3.2 Транспарентност на целта, улогата и опсегот

Организацијата треба јасно да утврди и соопшти зошто се користи системот со вештачка интелигенција, која задача ја извршува, врз кои одлуки или активности може да влијае и во кои граници е овластен да функционира. Транспарентноста на ниво на систем треба да ги опише неговата архитектура, можности, ограничувања, наменети корисници и значајни зависности доволно за одговорните лица да можат да ја разберат и надгледуваат неговата улога. За значително проширување на функциите, кругот на корисници или влијанието на системот врз одлуките треба да се добие ново одобрение. Такво проширување не смее да настане преку неформална употреба или постепено и неконтролирано проширување на функцијата на системот.

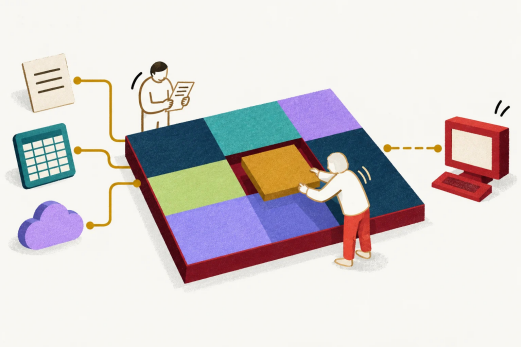
3.3 Објаснување на одлуките

Организациите треба да обезбедат лицата што користат систем со вештачка интелигенција или се потпираат на него да можат да разберат зошто е донесена одредена одлука, дадена препорака или постигнат одреден исход, како и самостојно да проценат дали треба да го прифатат. Прифатливата точност и брзина на објаснувањето треба да бидат соодветни на важноста и итноста на одлуката. Кај одлуките со значителни последици, соодветното и навремено објаснување треба да биде составен дел од процесот на одлучување.

3.4 Следливост на одлуките

Одлуките донесени со значителна помош од вештачка интелигенција треба да бидат следливи за да може да се реконструираат постапката и исходот. Релевантната евиденција треба да ги содржи системот и верзијата на моделот, значајните влезни податоци и инструкции, резултатот, човечката интервенција или отстапувањето од резултатот, како и конечната преземена мерка.

4. Управување со податоци и интегритет



4.1 Управување со податоците и управување во текот на животниот циклус

Организациите треба да воспостават сразмерен систем за управување со податоците што се користат за развој, проверка, тестирање и работа на системите со вештачка интелигенција. Тој треба да ги опфати собирањето, потеклото, подготовката, означувањето, пристапот, безбедноста, употребата, споделувањето, исправката, чувањето и бришењето на податоците. Значајните податоци треба да останат следливи при нивното менување и обработка, а оригиналните, изведените, синтетичките и информациите создадени од луѓе треба соодветно да се означат.

4.2 Квалитет на податоците

Податоците што се користат за развој, проверка, тестирање или работа на систем со вештачка интелигенција треба да ги исполнуваат техничките стандарди за точност, целосност, доследност, навременост и интегритет. Квалитетот треба да се мери во текот на целиот процес на обработка на податоците. Контролите за собирање, означување, трансформација, проверка, исправка, верзионирање и промени во податоците треба да бидат соодветни на намената на системот и на последиците од грешка.

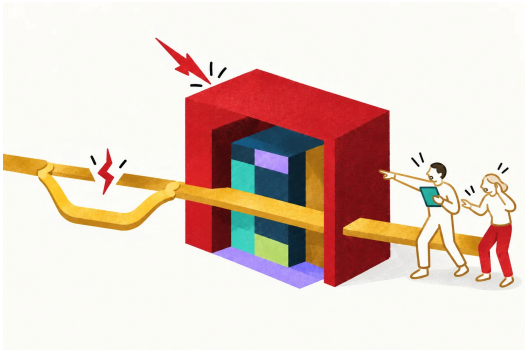
4.3 Приватност, обработувачи и пренос на податоци

Личните податоци треба да се обработуваат во согласност со важечкото законодавство за заштита на личните податоци. Организациите остануваат одговорни кога ангажираат даватели на услуги со вештачка интелигенција или други обработувачи што обработуваат лични податоци во нивно име. Тие треба да обезбедат соодветни инструкции, договорни заштитни мерки, безбедност, контрола врз подизведувачите и транспарентност за тоа каде и од кого се обработуваат податоците. Преносот на податоци во други јурисдикции треба да се врши само кога се обезбедени потребната правна основа и соодветните заштитни мерки.

4.4 Контрола врз личните податоци во моделите

Организациите треба да имаат предвид дека бришењето на изворните податоци не мора да ги отстрани информациите што моделот на вештачка интелигенција ги научил, извел или запомнил. Затоа, личните податоци што се користат за обучување, дополнително обучување или работа на моделот треба да се сведат на неопходниот минимум и да се заштитат со мерки што спречуваат нивно неовластено задржување, извлекување и повторна употреба и, кога е изводливо, овозможуваат нивна исправка, бришење или отстранување од моделот. Обработката на чувствителни податоци треба да се одвива во контролирана средина. Може да се користат модели инсталирани локално или модели со отворени тежински параметри кога овозможуваат поголема контрола врз податоците и можат безбедно да се управуваат и одржуваат.

5. Безбедност, робустност и доверливост



5.1 Безбедност и спречување неприфатлив ризик

Вештачката интелигенција треба да се дизајнира, воведува и користи така што предвидливите опасности ќе се утврдат и контролираат пред да предизвикаат неприфатлива штета. Штетата може да биде физичка, психолошка, економска, правна, штета за угледот или општествена штета, а може да ги засегне и заедниците и институциите. Барањата за безбедност треба да се утврдат имајќи ги предвид реалната средина, засегнатите лица, поврзаните системи и разумно предвидливата злоупотреба, а не само нормалното или предвиденото функционирање.

5.2 Точност и прецизност

Системите со вештачка интелигенција треба да се оценуваат според критериуми за успешност што се соодветни на нивната намена и на последиците од грешка. Кај генеративната вештачка интелигенција, оценувањето треба, кога е релевантно, да ги опфати фактичката точност, поткрепеноста со веродостојни извори, целосноста, релевантноста, доследноста и верноста кон изворите, наместо да се потпира само на една збирна мерка.

5.3 Доверливост и доследност

Системот со вештачка интелигенција треба да обезбедува доволно стабилни и доследни резултати со текот на времето. Споредливи влезни податоци и околности не треба да доведуваат до значително различни исходи без оправдана причина, особено кога резултатот придонесува за правни, професионални или управни одлуки. Кога променливоста е својствена за системот, организацијата треба да ја контролира и да ја соопшти, како и да воспостави постапки за спречување произволно постапување.

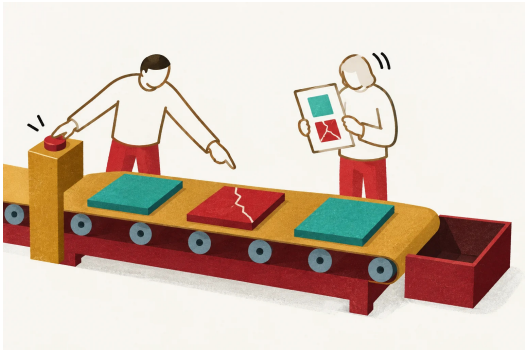
5.4 Робустност и отпорност

Вештачката интелигенција треба да поднесува предвидливи грешки во податоците, невообичаени влезни податоци, промени во средината и делумни откажувања без да предизвика неприфатливи последици. Системот треба да препознава услови што значително се разликуваат од оние во кои бил проверен и да не продолжува да дава самоуверени резултати кога неговите претпоставки веќе не важат. Организациите треба да ја задржат способноста да ги продолжат основните функции кога системот со вештачка интелигенција или надворешната услуга е недостапна, компромитирана или значително изменета.

5.5 Сајбер-безбедност и безбедност на пристапот

Системите со вештачка интелигенција, моделите, збирките на податоци, интерфејсите, пристапните податоци и компонентите во синџирот на снабдување треба да бидат заштитени од неовластен пристап, откривање, извлекување, измена и нарушување. Безбедносните контроли треба да одговараат на чувствителноста на податоците, важноста на функцијата и последиците од компромитирање. Безбедноста треба да ги опфаќа и вообичаените сајбер-закани и ризиците специфични за системите што користат вештачка интелигенција.

6. Учество на луѓето, надзор и компетентност



6.1 Суштинска човечка контрола, надзор и интервенција

Организациите треба да одлучат дали, зошто и под кои услови ќе се користи вештачка интелигенција. Надзорот треба да биде сразмерен на контекстот и ризиците на системот и е суштински само кога одговорното лице е запознаено со неговата намена, ограничувања, неизвесност и предвидливи откажувања; има пристап до релевантните факти и може самостојно да ги процени неговите резултати; и ја има потребната компетентност, овластување и практична можност да го преиспита системот, да ја поништи неговата одлука, да го суспендира или да го прекине неговото работење.

6.2 Зачувување на човечкото расудување и професионалните вештини

Воведувањето на вештачката интелигенција не треба да ги поткопа знаењето и расудувањето потребни за надзор над системот, откривање грешки и извршување на суштинската работа кога технологијата ќе откаже. Организациите треба да ги утврдат компетенциите што мора да останат кај луѓето и да ги организираат обуката и распределбата на работата така што тие компетенции ќе продолжат да се развиваат. На помладите професионалци треба да им се овозможи доволно време да ги стекнат основните вештини пред вештачката интелигенција да ги замени процесите на учење преку кои се формира издржано професионално расудување.

7. Одговорност, управување и правни лекови



7.1 Јасна одговорност во текот на животниот циклус

За секој систем со вештачка интелигенција треба да биде определено одговорно лице во организацијата, кое одговара за неговата намена, одобрување, работа, следење, значајни промени и повлекување од употреба. Одговорностите на раководството, професионалните корисници, техничките тимови, сопствениците на податоците и давателите на услуги треба да бидат јасно утврдени.

7.2 Процена на влијанието и ризиците уште при дизајнирањето

Пред воведувањето на системот, а и по секоја значајна промена, одговорната организација треба да ги процени очекуваните придобивки, засегнатите лица, влијанијата врз правата и еднаквоста, ограничувањата што произлегуваат од податоците, безбедноста и сигурноста, зависностите во организацијата и достапните заштитни мерки. Проценатa и нејзините наоди треба уште од почеток да влијаат врз дизајнот и набавката. Сериозните нерешени ризици треба да ја спречат или ограничат употребата на системот.

7.3 Ефикасен правен лек и обесштетување

Лицето што претрпело штета или било значително ставено во понеповолна положба поради постапка во која помагала вештачка интелигенција треба да има пристап до навремен и ефикасен правен лек. Тој може да опфати објаснување, оспорување, повторно разгледување од човек, исправка, враќање во поранешната состојба и обесштетување, кога тоа е соодветно. Правниот лек треба да го опфати и исходот за конкретното лице и евентуалниот основен недостаток што може да ги засегне и другите.

7.4 Одговорност на институциите и поединците за одлуките донесени со помош на ВИ

Организацијата останува одговорна за системите со вештачка интелигенција што ги развива, набавува, интегрира или користи, вклучувајќи ги нивните резултати и последици. Кога резултатот од систем со вештачка интелигенција ќе влијае врз конкретна одлука, лицето што ја донесува или усвојува одлуката останува одговорно самостојно да го оцени резултатот и да ја донесе конечната одлука. Тоа лице треба да ги има потребните информации, компетентност и овластување. Ниту организацијата ниту лицето што ја донело одлуката не смеат да ја избегнат одговорноста префрлајќи ја на системот со вештачка интелигенција, продавачот, добавувачот, развивачот или друго трето лице.

8. Партиципација, локален контекст и општествена легитимност



8.1 Учество на засегнатите страни и заедници

Организациите треба да ги утврдат и суштински да ги вклучат лицата, институциите и заедниците што учествуваат во развојот, управувањето или регулирањето на систем со вештачка интелигенција, како и оние што се засегнати од него или се изложени на ризиците поврзани со него, особено засегнатите, маргинализираните и ранливите групи. Учеството треба да започне додека сè уште може да влијае врз намената, дизајнот, заштитните мерки и условите за употреба на системот и да продолжи со појавата на нови последици. Тоа треба да биде пристапно за луѓето без оглед на разликите во јазикот, ресурсите и пристапот до дигитални технологии.

8.2 Приспособување кон локалниот контекст

Системот со вештачка интелигенција треба да се оцени во контекст на законите, институциите, административните практики, инфраструктурата и организациските капацитети на средината во која ќе функционира. Проверката во друга земја или организација сама по себе не потврдува дека системот е соодветен за Северна Македонија. Локалното приспособување треба да ги опфати реалното функционирање на системот, управувањето со него и неговото вклопување во постојните институции и инфраструктура, а не само преводот на неговиот интерфејс.

8.3 Институционален капацитет и зрелост

Организацијата не треба да воведе вештачка интелигенција надвор од своите можности за управување со податоците, надзор над давателите на услуги, оценување на успешноста, спроведување човечки надзор и постапување при грешки или инциденти. Пред воведувањето, треба да ги процени своите кадри, процеси, инфраструктура, овластувања и финансиски можности за одржливо и одговорно функционирање.

8.4 Јазична и културна вклученост

Системите со вештачка интелигенција што се користат во Северна Македонија треба да се оценуваат во однос на македонскиот јазик, неговите дијалектни варијанти и кириличното писмо, како и во однос на албанскиот и другите јазици, писма и културни контексти на заедниците што живеат во земјата. Организациите треба да ги утврдат и отстранат јазичните, културните или податочните недостатоци што доведуваат до послаби резултати, исклучување или погрешно толкување и, кога е потребно, да обезбедат ефикасна човечка алтернатива или алтернатива без вештачка интелигенција.

8.5 Соодветност на податоците за локалниот контекст

Организациите треба да обезбедат податоците што ги користи системот со вештачка интелигенција соодветно да го одразуваат локалниот правен, институционален и општествен контекст во кој функционира системот. Особено внимание треба да се посвети на нерамномерната дигитализација, расцепканата евиденција, непишаните практики и релевантните информации што постојат надвор од формалните дигитални системи. Отсуството на дигитални податоци не треба да се смета за доказ дека некој факт или околност не постои. Значајните празнини во податоците треба да доведат до дополнителна човечка проверка или до ограничување на употребата на системот.

Благодарност

Ethical AI Macedonia им се заблагодарува на организациите што придонесуваат во нејзината работа. Нивната поддршка опфаќа стручни мислења од академската заедница, поврзување со компании, соработка при организирање настани и придонес на адвокатските канцеларии на основачите.

Амбасада на Шведска во Скопје

Финансиска поддршка за ноемвриискиот настан за одговорна употреба на ВИ во бизнисот.

Шведско-македонска стопанска комора

Поврзување со компании, споделување на прашалникот за бизнисите и соработка при организацијата на настанот.

Универзитет во Лунд

Стручно мислење за почетната верзија на принципите и учество на ноемвриискиот настан.

Адвокатска канцеларија Бошњаковски

Адвокатската канцеларија на коосновачот Владимир Бошњаковски, партнер на иницијативата.

Адвокатска канцеларија Ангела Анѓеловска

Адвокатската канцеларија на коосновачката Ангела Анѓеловска, партнер на иницијативата.

Придонесете во понатамошниот развој

Овие принципи ќе продолжат да се развиваат преку коментари и практично искуство. Ги поканuваме бизнисите, истражувачите, стручните лица од практиката и граѓанските организации да споделат примери, критички согледувања и предлози.

aimacedonia.org | info@aimacedonia.org

Адреса за кореспонденција: ДТЦ Палома Бјанка, Даме Груев 16/10
1000 Скопје, Северна Македонија