

ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА И ISO СТАНДАРДИ

ARTIFICIAL INTELLIGENCE AND ISO STANDARDS

РАДОСЛАВ РАКОВИЋ¹

Стручни рад
DOI: 10.5937/VI24019R

Резиме: Практична примена вештачке интелигенције тесно је повезана са успостављањем одговарајућег оквира који се односи на организациона питања. Зато је у оквиру Међународне организације за стандардизацију (ISO) донет низ менаџмент стандарда везаних за вештачку интелигенцију. У овом раду дат је сажет преглед најзначајнијих стандарда, пре свега оних који се односе на концепте и терминологију, систем менаџмента вештачком интелигенцијом, смернице за управљање ризицима у области вештачке интелигенције, импликације вештачке интелигенције на управљање у организацијама и оцењивање утицаја вештачке интелигенције.

Кључне речи: вештачка интелигенција (ВИ), Систем менаџмента вештачком интелигенцијом (СМВИ), управљање ризицима, оцењивање утицаја

Abstract: The practical application of artificial intelligence is closely related to the establishment of an appropriate framework related to organizational issues. That is why the International Organization for Standardization (ISO) adopted a series of management standards related to artificial intelligence. This paper provides a concise overview of the most important standards, primarily those related to concepts and terminology, artificial intelligence management system, guidelines for risk management in the field of artificial intelligence, artificial intelligence governance implications for organizations and impact assessment of artificial intelligence

Key Words: Artificial Intelligence (AI), Artificial Intelligence Management System (AIMS), Risk management, Impact Assessment

1. Увод

Вештачка интелигенција обухвата широк спектар области као што су машинско учење (енгл. *Machine Learning*), неуронске мреже (енгл. *Neural*

¹ Радослав Раковић, Инжењерска академија Србије, Кнеза Милоша 9, Београд, rakovic@gmail.com, ORCID 0000-0002-8067-6582

Networks), експертни системи (енгл. *Expert Systems*), обрада великих количина података (енгл. *Big Data*), обрада природног језика (енгл. *Natural Language Processing*), препознавање говора (енгл. *Speech Recognition*), препознавање лица (енгл. *Face Recognition*), рачунарско виђење (енгл. *Computer vision*) итд. које су у својој основи техничко-технолошке природе. Међутим, као и у другим технолошким областима, практична примена вештачке интелигенције тесно је повезана са успостављањем одговарајућег оквира који се односи на организациона питања. Зато је у оквиру Међународне организације за стандардизацију (ISO), у сарадњи са Међународном електротехничком комисијом (IEC), у периоду 2022.-2024. године донет низ менаџмент стандарда везаних за вештачку интелигенцију.

У овом раду дат је сажет преглед најзначајнијих стандарда, пре свега оних који се односе на концепте и терминологију (ISO/IEC 22989), систем менаџмента вештачком интелигенцијом (ISO/IEC 42001), смернице за управљање ризицима у области вештачке интелигенције (ISO/IEC 23894), импликације вештачке интелигенције на управљање у организацијама (ISO/IEC 38507) и оцењивање утицаја вештачке интелигенције (ISO/IEC CD 42005).

2. Стандард ISO/IEC 22989:2022

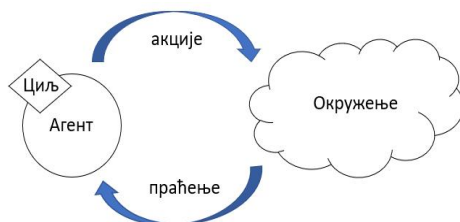
Предметни стандард систематизује терминологију и концепте вештачке интелигенције. Дефинисано је укупно 117 термина груписаних у 7 области, табела 1.

Табела 1. Термини дефинисани у стандарду ISO/IEC 22989, Извор: Аутор

Обл.	Назив области - Термини везани за (област)	
3.1	Вештачка интелигенција – <i>Artificial Intelligence</i>	35
3.2	Подаци - <i>Data</i>	15
3.3	Машинско учење – <i>Machine Learning</i>	17
3.4	Неуронске мреже – <i>Neural networks</i>	10
3.5	Поузданост / Поверење - <i>Trustworthiness</i>	18
3.6	Обрада природних језика – <i>Natural language processing</i>	18
3.7	Рачунарско виђење – <i>Computer vision</i>	4
		117

Од наведених дефиниција, издвајамо појам „Систем вештачке интелигенције“ (ВИ), дефиниција 3.1.4, где се наводи да је ВИ систем „инже-

њерски систем који генерише излазе као што су садржај, прогнозе, препоруке или одлуке за дати скуп циљева које је дефинисао човек". У напоменама уз дефиницију стоји да „инжењерски систем може да користи различите поступке који се односе на ВИ да би се развио модел за представљање података, знања, процеса итд. који се могу користити у извршавању задатака" и да се системи ВИ пројектују да раде са различитим нивоима аутоматизације.



Слика 1. Парадигма агента (на бази слике 1 ISO/IEC 22989)

На слици 1 приказана је суштина функционисања ВИ система кроз „парадигму агента“. ВИ агент дефинисан је као „аутоматски ентитет који осећа своје окружење и реагује на њега и предузима акције да достигне своје циљеве“ (дефиниција 3.1.1). Као што се са слике може видети, агент прати окружење преко сензора, а делује на њега преко актуатора, у складу са остварењем својих циљева. Агент на одређени начин симулира људску интелигенцију и начин понашања.

Поред парадигме агента, у оквиру овог стандарда дефинишу се и други концепти повезани са ВИ – знање, алгоритми, подаци, машинско учење, неуронске мреже, аутономија, хетерономија и аутоматизација, Интернет ствари (енгл. Internet of Things - IoT) и сајбер физички системи, поузданост/поверење (енгл. Trustworthiness), у коју је укључена пристрасност & поштење (енгл. Bias & Fairness), верификација и валидација, правна питања, друштвени утицај и улоге заинтересованих страна.



Слика 2. Животни циклус система ВИ (на бази слике 3 ISO/IEC 22989)

У оквиру главе 6 предметног стандарда дефинисане су етапе животног циклуса (енгл. *Lifecycle*) система ВИ, слика 2.

У почетној фази (енгл. *Inception*) успостављају се циљеви, дефинишу се захтеви, спроводи се оцењивање ризика и формира план поступања по њима, утврђују се елементи везани за одговорности, трошкове, ресурсе, изводљивост итд. У фази пројектовања и развоја (енгл. *Design and Development*) дефинишу се приступ, архитектура, кодовање, тренинг подаци, поступање по ризицима итд. У фази верификације и валидације спроводи се оцењивање оствареног из угла захтева и намене система и спроводи се праћење и поступање по ризицима. У фази распоређивања (енгл. *Deployment*) систем се пушта у рад и спровode се планиране актив-ности за поступање по ризицима. У фази функционисања система његов рад се прати и вреднује, спровode се ажурирања, као и преиспитивање ризика. У фази поновног вредновања, сагледавају се остварени резултати, преиспитују се и редефинишу (ако је то потребно) циљеви и захтеви као и старегије поступања по ризицима. На крају је етапа гашења (енгл. *Retirement*) када систем ВИ престаје са радом.

3. Стандард ISO/IEC 42001:2023

Овај стандард дефинише захтеве и смернице за систем менаџмента вештачком интелигенцијом (енгл. *Artificial Intelligence Management System – AIMS*) према коме организације које обезбеђују или користе производе или услуге које користе ВИ могу да се сертифицију.

Структура овог менаџмент стандарда идентична је са структуром која је дефинисана за све стандарде тог типа у оквиру ISO (у документу познатом као Annex SL) - стандард садржи 10 глава, од којих прве три (1-Предмет и подручје примене, 2-Нормативне референце и 3-Термини и дефиниције) не садрже захтеве као осталих 7 глава (4-Контекст организације, 5-Лидерство, 6-Планирање, 7-Подршка, 8-Функционисање, 9-Вредновање перформанси и 10-Побољшања). За разлику од других менаџмент стандарда, а слично стандарду са захтевима за безбедност информација ISO/IEC 27001, овај стандард садржи посебан Анекс А у коме су дефинисани циљеви контрола и контроле. Овај Анекс А је нормативан, дакле обавезан, и у њему су садржани захтеви за AIMS као и у основном делу стандарда.

У глави 3-Термини и дефиниције садржано је 26 појмова, али су они углавном везани за генералне појмове који се односе на менаџмент системе сем последња три које дефинишу оцењивање утицаја система ВИ, квалитет података (који може бити вема важан за систем ВИ) и изјаву о

примењивости (енгл. Statement of Applicability – SoA) која, као и код стандарда за безбедност информација ISO/IEC 27001, представља приказ примењених контрола и образложење уколико су неке од њих изостављене због природе делатности организације.

У табели 2 приказана је структура Анекса А стандарда ISO/IEC 42001:2023. Укупан број контрола износи 38 и оне су груписане у 9 области А.2 до А.10. У оквиру Аппех В предметног стандарда дате су детаљније смернице за примену наведених контрола.

Табела 2. Структура Анекса А стандарда ISO/IEC 42001, Извор: Аутор

Обл.	Назив области		
A.2	Политике које се односе на ВИ	3	A.2.2-A.2.4
A.3	Интерна организација	2	A.3.2-A.3.3
A.4	Ресурси за системе ВИ	5	A.4.2-A.4.6
A.5	Оцењивање утицаја ВИ система	4	A.5.2-A.5.5
A.6	Животни циклус ВИ система		
A.6.1	Менаџмент смернице за развој ВИ система	2	A.6.1.2-A.6.1.3
A.6.2	Животни циклус ВИ система	7	A.6.2.2 -A.6.2.8
A.7	Подаци за ВИ	5	A.7.2-A.7.6
A.8	Информације за заинтересоване стране ВИ	4	A.8.2-A.8.5
A.9	Коришћење система ВИ	3	A.9.2-A.9.4
A.10	Односи са трећим странама и корисницима	3	A.10.2-A.10.4

Аппех С предметног стандарда је информативног карактера, али у себи садржи потенцијалне организационе циљеве и изворе ризика повезаних са вештачком интелигенцијом. Потенцијални организациони циљеви везани за ВИ могу бити крајња одговорност (енгл. Accountability), ВИ експертиза, расположивост и квалитет података за обуку и тестирање система ВИ, утицај на окружење, поштење (енгл. Fairness), погодност за одржавање (енгл. Maintainability), приватност, робусност, сигурност (енгл. Safety) и безбедност (енгл. Security). Извори ризика повезани са ВИ могу бити комплексност окружења, недостатак транспарентности и објашњивости, ниво аутоматизације, извори ризика везани за машинско учење, системски хардвер, питања животног циклуса система и спремност технологије.

4. Стандард ISO/IEC 23894:2023

Овај стандард дефинише смернице како организација која развија, производи, дистрибуира или користи производе, системе и услуге које користе

ВИ може да управља ризицима који су специфично повезани са вештачком интелигенцијом. Овај стандард ослања се на базични стандард за менаџмент ризицима ISO 31000:2018 исказујући специфичности везане за ВИ. У овом стандарду се, као и у базичном стандарду, ризик разматра у контексту његове дуалне природе тј. да ризици могу имати и позитивне и негативне утицаје, а не само негативне, како се то обично подразумева. Позитивни утицаји ризика се у неким другим стандардима називају прилике (енгл. *Opportunities*).

Базични стандард прокламује осам принципа везаних за систем менаџмента ризицима, који треба да буде:

- Интегрисан тј. саставни део активности организације
- Структуриран и разумљив
- Прилагођен кориснику
- Инклузиван тј. да благовремено укључи све заинтересоване стране
- Динамичан тј. да прати промене које се дешавају у окружењу
- Заснован на најбољим расположивим информацијама
- Да узима у обзир људско понашање и културу
- Да се стално побољшава.

Стандардом ISO/IEC 23894 је указано да за прва три принципа нема специфичних смерница када су у питању системи ВИ, док код осталих има специфичности.

Ове специфичности могу се, у сажетом облику, сумирати на следећи начин:

- Заинтересоване стране могу да помогну да квалитет прикупљених података буде бољи (што је посебно важно код машинског учења), могу помоћи да се дефинишу циљеви и опишу средства за повећање транспарентности и објашњивости система и могу исказати своје перцепције и ставове о ситуацијама када систем ВИ за аутоматизовано доношење одлука треба да има људски надзор. Исто тако, заинтересоване стране могу да помогну да се дефинишу критеријуми правичности и елиминацију пристрасности у раду система ВИ.
- Систем ВИ мора бити динамичан како би испратио све промене које се дешавају у екстерном и интерном контексту, у законским и регулаторним захтевима, очекивањима клијената итд. Неки од система ВИ и експлицитно функционишу на принципима непрекидног учења, усавршавања, вредновања и валидације и морају се са свим тим променама усклађивати „у ходу“.

- Системи ВИ морају да раде са најбољим расположивим информацијама у датом тренутку. Део тих информација је историјски, будућа очекивања се могу променити. Све то ствара услове неизвесности, који се морају узети у обзир. Исто тако, управљање ризицима мора узети у обзир и ограничења у праћењу употребе система ВИ која могу проистацати из законских оквира, тржишта, уговора, права интелектуалне својине итд.
- Приликом пројектовања, развоја или примене система ВИ неопходно је уважити људско и културолошко окружење, у којој мери ти системи одговарају постојећим друштвеним обрасцима који могу довести до утицаја на правичност резултата, поштовање приватности, слободу изражавања, животну средину итд.
- Системи ВИ морају да имају уграђене механизме за праћење достигнутих перформанси и недостатака које ће, кроз механизам „научених лекција“, применити за унапређења и побољшања.

Стандардом су прецизирани елементи које је потребно узети у обзир при разматрању екстерног и интерног контекста организације као и заинтересованих страна које се ту појављују, а све из угла примене система ВИ. У конципирању система менаџмента ризицима од примене ВИ посебно је неопходно размотрити да ли системи ВИ могу да нашkode људима, да им ускрате нека основна права, да ли и на који начин могу да угрозе животну средину, да ли је производња, дистрибуција или примена система ВИ у складу са друштвеном одговорношћу организације (смернице из одговарајућег стандарда ISO 26000) итд.

Сам процес оцењивања ризика (енгл. Risk Assessment) у примени система ВИ у суштини се одвија на исти начин као и у другим областима – прво се идентификују ризици, затим се они анализирају и вреднују по значају, како би се највећи део времена и ресурса могао посветити онима који су најзначајнији. Слична је ситуација и са поступањем по ризицима – ризици се могу избећи (нпр. реорганизовањем активности или отклањањем узрока), умањити или подстицати применом контрола (зависи да ли је ризик негативан или позитиван по организацију) како би се утицало на вероватноћу догађаја и /или на његове последице, могу се пренети на неку трећу страну (нпр. осигурање) или прихватити ако нису превелики.

5. Стандард ISO/IEC 38507:2022

Овај стандард бави се импликацијама које системи ВИ имају на управљање (енгл. Governance) у организацијама. Базиран је на смерницама за управљање организацијама датим у основном стандарду ISO 37000, као

и специфичном стандарду за управљање ИТ сегментом ISO 38500, с обзиром да су системи ВИ у суштини један подсистем ИТ. Оно што системе ВИ разликује од других ИТ може се свести на следећа три елемента:

- Аутоматизовано одлучивање. Систем ВИ формира излаз на основу историјских и тренутних података, предвиђања формира са одговарајућим нивоом вероватноће и на основу тога усмерава своје деловање, и то великом брзином.
- Решавање проблема засновано на подацима. Систем ВИ ради са великим количинама података, уочавајући у њима обрасце. Уједно, са променом у том сету података остварује се утицај на решавање сваког од следећих проблема, што је тешко изводљиво класичним системима програмирања.
- Адаптивни системи. Системи ВИ у највећем броју случајева користе улазне податке за побољшање интерних модела и из тог разлога брзо се прилагођавају тј. могуће је да се за исти улаз добије различит резултат, јер је систем у међувремену напредовао у процесу сталног учења.

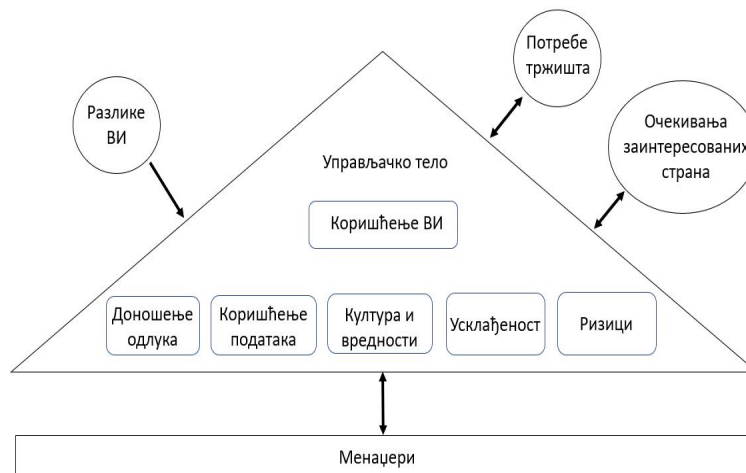
Нагласак предметног стандарда је на управљању, које спроводе људи, употребом ВИ, а не на самим системима ВИ. Кључни појам је надзор (енгл. *Oversight*) као „праћење имплементације организационих и управљачких политика и управљање повезаним задацима, услугама и производима које је поставила организација, у циљу прилагођавања променама у интерном и екстерном окружењу“ (дефиниција 3.2.1 стандарда ISO/IEC 38507).

Јасно је да системи ВИ могу донети значајне стратешке користи за организацију, али уједно могу условити додатне ризике за њу, у смислу потенцијалне штете која може настати за заинтересоване стране. То у значајној мери утиче на рад управљачког тела у организацији, укључујући и повећану одговорност за доношење одлука. Ово тело треба да усмерава рад организације кроз политике и стратегије, да спроводи надзор коришћења ВИ у организацији, да вреднује интерне и екстерне утицаје на организацију и да извештава заинтересоване стране о томе да се ефикасно управља коришћењем ВИ.

Наравно, чланови управљачког тела не морају бити експерти за ВИ, али морају да познају природу тих технологија и да буду свесни потенцијалних последица у њиховој примени, да би могли да доносе одговарајуће одлуке.

На слици 3 приказан је модел импликација које ВИ има на процес управљања у организацијама у складу са предметним стандардом. Тај процес одвија се у другачијим околностима услед разлика које ВИ уноси у цео амбијент, морају се пратити потребе тржишта и очекивања заинтересованих

страна, коришћење ВИ посматра се као прилика коју треба искористити да би се унапредило пословање организације, а то се спроводи преко пет политика, у тесној сарадњи са менаџментом организације.



Слика 3. Импликације ВИ на управљање (на бази слике 3 ISO/IEC 38507)

Прва политика односи се на управљање доношењем одлука. То је део целокупног управљања организацијом, а заснива се на делегирању овлашћења и одговорности у целој организацији, при чему крајња одговорност за све што се у организацији одвија остаје на управљачком телу. Кључни елементи доношења одлука укључују усклађеност са циљевима организације, да сваки ниво доношења одлука одговара додељеном овлашћењу и одговорностима, да они који доносе одлуке поседују адекватне вештине и способности за то о чему одлучују, да процес доношења одлука буде транспарентан за заинтересоване стране и да постоји адекватан надзор целог процеса одлучивања.

Неки системи ВИ се у значајној мери ослањају на податке за формирање и обучавање модела, па је управљање коришћењем података кључно за одговорну употребу ВИ. То подразумева да ти подаци буду адекватни за пројектовани систем ВИ, да се прикупљају на контролисан начин, да се води рачуна о њиховој безбедности, чувању, доступности и квалитету тих података, укључујући и испитивање тих података у погледу елемената пристрасности и других недостатака у подацима. Аспекти управљања употребом ВИ, када је реч о подацима, обухватају вредност података (квалитет, количина, правовременост, контекст и цена њихове употребе), ризике (безбедност, могуће злоупотребе, угрожавање приватности, конкурентски ризици услед неискоришћења употребе података) као

и ограничења (законски захтеви, уговорне обавезе, ауторска права, комерцијални интереси).

Управљачко тело је одговорно за дефинисање и успостављање жељене културе организације и њеног система вредности, у односу на заинтересоване стране. Велики део те културе и система вредности је имплицитно уграђен у понашање људи и процесе. Системи ВИ немају ниво разумевања контекста као људи, ослањају се на моделе, алгоритме и податке који их обучавају. Зато управљачко тело треба да води рачуна да у неким случајевима системи ВИ морају да буду ограничени људским деловањем, управо да би се обезбедило да донете одлуке буду усклађене са културом организације и системом вредности.

Организација има своје обавезе усклађености према окружењу и управљачко тело мора да предузме све како би се те обавезе испуњавале. У том смислу, ово тело треба да прати у којој мери системи ВИ могу да утичу на испуњавање обавеза усклађености и да интервенише када год постоји опасност да обавеза не буде испуњена. То се може постићи кроз додатне контроле, преиспитивање одлука донетих од стране система ВИ, обезбеђивањем да се подаци користе у складу са политиком организације и надгледањем рада система ВИ.

О ризицима је већ било речи у делу који се односио на стандард ISO/IEC 23894. У суштини, управљачко тело кроз дефинисање сврхе и циљева организације као и стратегија за остваривање тих циљева разматра и све оно што би могло да угрози остваривање тих циљева на глобалном нивоу, а детаље везане за идентификацију, анализу, вредновање и поступање са ризицима делегира унутар организације. О циљевима и изворима ризика када је реч о системима ВИ већ је било речи у склопу горе наведеног стандарда.

6. Стандард ISO/IEC CD 42005: 2023

Овај стандард носи ознаку „CD“ с обзиром да је још у фази *Committee Draft* тј. још увек није одобрен као међународни стандард. У складу са дефиницијом 3.1 из овог документа „оцењивање утицаја ВИ система је формалан документован процес којим се разматрају утицаји на појединце, групе појединаца и друштво од стране организација које развијају, испоручују или користе производе или услуге који користе вештачку интелигенцију“.

Процес оцењивања утицаја ВИ система треба да буде интегрисан са другим менаџмент процесима у организацији, спроводи се периодично у складу са законским оквирима, нивоима ризика, очекивањима заинтере-

сованих страна, мора да има дефинисан предмет и подручје примене, одговорне учесника у оцењивању као и успостављене прагове за осетљиве намене, ограничења и скале утицаја.

Документованост процеса оцењивања утицаја ВИ система, према овом стандарду, подразумева следеће информације:

- Информације о самом систему ВИ (опис, карактеристике, сврха, планирано и непланирано коришћење)
- Информације о подацима који се користе и о квалитету тих података
- Информације о алгоритмима које организација користи и о њиховом развоју, као и информације о моделима који се примењују у систему ВИ и о њиховом развоју
- Информације о окружењу у коме систем ВИ ради (географско подручје, језици, комплексност и ограничења у окружењу)
- Заинтересоване стране (енгл. Interested Parties / Stakeholders)
- Добити (енгл. Benefits) и штетности (енгл. Harms) могу настати као и кварови система ВИ и могуће злоупотребе.

Резултат свих оцењивања утицаја система ВИ треба да буду предлог мера које би требало имплементирати на техничком нивоу (нпр. Криптовање корисничког регистра како би се спречио негативан утицај на приватност, остваривање људског надзора при доношењу одлука и сл), али и препоруке највишем руководству и управљачком органу које би требало да размотре приликом успостављања политика орагнизације везаних за системе ВИ (нпр. подизање свести учесника).

Документ садржи четири информативна анекса који се односе на смернице за примену ISO/IEC 42001 (Annex A), ISO/IEC 23894 (Annex B), таксономију добити и штетности (Annex C) и усклађивање оцењивања утицаја система ВИ са другим оцењивањима (Annex D).

7. Закључак

У раду је дат сажети опис неколико најзначајнијих стандарда који се односе на системе ВИ. Иако је примена стандарда у начелу добровољна, корисно је да постоји оквир за деловање у тако комплексној области као што је вештачка интелигенције. Веома је значајно што су се институције на светском нивоу као што су ISO и IEC благовремено и активно укључиле у ову област како би помогле да се разреши бар неке од мноштва недоумица које постоје и које ће постојати у примени система вештачке интелигенције.

Треба нагласити да је у јулу 2024. године донета Европска уредба која се односи на хармонизацију правила везаних за вештачку интелигенцију, а

њена примена предвиђена је од августа 2026. године. То јасно указује да се регулисање примене вештачке интелигенције из области стандарда, који су у принципу добровољни, премешта у зону законске регулативе која је обавезујућа.

8. Литература

- [1] ISO 26000: Guidance on social responsibility, 2010
- [2] ISO 31000: Risk management – Guidelines, 2018
- [3] ISO Guide 73: Risk Management – Vocabulary, 2009
- [4] ISO/IEC 22989: Information technology-Artificial intelligence-Artificial intelligence concepts and terminology, Ed 1, 2022
- [5] ISO/IEC 23894: Information technology-Artificial intelligence- Guidance on risk management, Ed 1, 2023
- [6] ISO/IEC 37000: Governance of organizations – Guidance, 2021
- [7] ISO/IEC 38500: Corporate governance of information technology, 2024
- [8] ISO/IEC 38507: Information technology- Governance of IT – Governance implications of the use of artificial intelligence by organizations, Ed 1, 2022
- [9] ISO/IEC 42001: Information technology-Artificial intelligence- Management System, Ed 1, 2023
- [10] ISO/IEC CD 42005: Information Technology-Artificial Intelligence-Artificial Intelligence system impact assessment, Ed 1, 2023
- [11] Regulative (EU) 2024/1689: Laying down harmonised rules on artificial intelligence, OJEU 12/07/2024
- [12] Стратегија развоја вештачке интелигенције у Републици Србији за период 2020-2025. година, *СПРС* 96/2019