Poročilo 2.2: Seznam relevantnih EU in mednarodnih standardov in predlog za pripravo nacionalnih standardov

Pripravili: Maja Škrjanc (IJS), Maruša Škrjanc (IJS)

Datum: 30.12.2025

Dopolnjeno 27.2.2026

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Kazalo

1. Namen, status in vsebinski okvir dokumenta 5
2. Metodologija in kriteriji vključitve 7
2.1 Metodološko izhodišče 8
2.2 Obseg vključene dokumentacije 9
2.3 Kriteriji vključitve in razvrščanja 10
2.4 Metoda obdelave gradiva 12
2.5 Metodološki doseg in namen uporabe 13
3. Shema standardizacijskega ekosistema za umetno inteligenco 14
3.1 Mednarodni sloj 15
3.2 Evropski sloj 17
3.3 Nacionalni sloj 18
3.4 Sklepna funkcija sheme 20
4. Vsebinska razvrstitev ključnih standardov in referenčnih dokumentov 20
4.1 Standardi za upravljanje in sisteme vodenja umetne inteligence 21
Funkcija sklopa 21
Ključni standardi in dokumenti 22
Praktična vrednost za slovenski prostor 22
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta 23
Pomen za nadaljnje nacionalne tehnične razmisleke 24
4.2 Standardi za upravljanje tveganj in presojo vplivov 25
Funkcija sklopa 25
Ključni standardi in dokumenti 25
Praktična vrednost za slovenski prostor 26
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta 27
Pomen za nadaljnje nacionalne tehnične razmisleke 27
4.3 Standardi za terminologijo, koncepte in arhitekturne okvire 28
Funkcija sklopa 28
Ključni standardi in dokumenti 29
Praktična vrednost za slovenski prostor 29
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta 30
Pomen za nadaljnje nacionalne tehnične razmisleke 31

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

4.4 Standardi za kakovost podatkov in podatkovno upravljanje ... 32
Funkcija sklopa ... 32
Ključni standardi in dokumenti ... 32
Praktična vrednost za slovenski prostor ... 33
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta ... 34
Pomen za nadaljnje nacionalne tehnične razmisleke ... 35
4.5 Standardi za človeški nadzor in obvladljivost ... 35
Funkcija sklopa ... 35
Ključni standardi in dokumenti ... 36
Praktična vrednost za slovenski prostor ... 37
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta ... 38
Pomen za nadaljnje nacionalne tehnične razmisleke ... 39
4.6 Standardi za robustnost, kibernetsko varnost in odpornost sistemov ... 40
Funkcija sklopa ... 40
Ključni standardi in dokumenti ... 40
Praktična vrednost za slovenski prostor ... 41
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta ... 42
Pomen za nadaljnje nacionalne tehnične razmisleke ... 43
4.7 Standardi in dokumenti za transparentnost, zasebnost, pristranskost in odgovorno zasnovo ... 44
Funkcija sklopa ... 44
Ključni standardi in dokumenti ... 45
Praktična vrednost za slovenski prostor ... 46
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta ... 47
Pomen za nadaljnje nacionalne tehnične razmisleke ... 47
4.8 Podporni dokumenti za interoperabilnost, podatkovne prostore in infrastrukturne primere uporabe ... 48
Funkcija sklopa ... 48
Ključni standardi in dokumenti ... 49
Praktična vrednost za slovenski prostor ... 50
Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta ... 51
Pomen za nadaljnje nacionalne tehnične razmisleke ... 51

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

4.9 Operativni povzetek ključnih ugotovitev 52
5. Presoja prednostnih standardnih sklopov in potreb po nacionalni tehnični konkretizaciji 54
5.1 Namen in logika presoje 54
5.2 Nosilne reference za slovenski prostor 55
5.3 Podporne in razvojno pomembne reference 56
5.4 Področja, na katerih se kaže potreba po dodatni nacionalni tehnični konkretizaciji 57
5.5 Operativna prioritizacija za nadaljnje delo 58
6. Prednostna področja za pripravo nacionalnih tehničnih dopolnitev in standardizacijskih predlogov 60
6.1 Namen 60
6.2 Merila za izbor prednostnih področij 60
6.3 Prednostno področje 1: minimalni nabor artefaktov in dokazil 61
6.4 Prednostno področje 2: profil za človeški nadzor in obvladljivost 61
6.5 Prednostno področje 3: kakovost podatkov in podatkovna dokumentacija 62
6.6 Prednostno področje 4: baseline varnostni profil za sisteme UI 63
6.7 Prednostno področje 5: transparentnost, zasebnost, pristranskost in etični artefakti 64
6.8 Prednostno področje 6: interoperabilnostni in infrastrukturni profili 64
6.9 Operativni sklep poglavja 65
7. Predlagani tipi nacionalnih tehničnih izhodov 66
7.1 Namen 66
7.2 Tip izhoda 1: minimalni artefaktni paket in standardizirane predloge dokumentacije 67
7.3 Tip izhoda 2: nacionalni profil človeškega nadzora in obvladljivosti 67
7.4 Tip izhoda 3: smernica za kakovost podatkov in evidenčni profil za dataset dokumentacijo 68
7.5 Tip izhoda 4: baseline varnostna smernica in razpisno-pogodbeni profil za AI sisteme 69
7.6 Tip izhoda 5: transparency profile, privacy evidenca in predloga za analizo pristranskosti (bias) 70
7.7 Tip izhoda 6: interoperabilnostni profil in sektorske tehnične dopolnitve 70
7.8 Operativni povzetek tipov izhodov 71
8. Zaključek 73

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

1. Namen, status in vsebinski okvir dokumenta

Ta dokument je namenjen temu, da bralcu ponudi jasno, strokovno utemeljeno in praktično uporabno sliko standardizacijskega prostora na področju umetne inteligence ter da na tej podlagi odpre vprašanje, katere standardizacijske usmeritve so za slovenski prostor smiselne na nacionalni ravni. Njegov neposredni namen ni zgolj identificirati relevantne evropske in mednarodne standarde, temveč iz širšega standardizacijskega ekosistema izluščiti tiste dokumente, ki so za slovenske deležnike dejansko najpomembnejši, ter na tej podlagi oblikovati strokovno podlago za nadaljnjo pripravo nacionalnih standardov oziroma sorodnih tehničnih dokumentov. V tem smislu gre za dokument, ki standardizacijskega prostora ne obravnava zgolj deskriptivno, temveč selektivno, razvrščevalno in interpretativno, z izrazito usmeritvijo v njegovo nadaljnjo praktično uporabo.

Za pravilno razumevanje tega dokumenta je bistveno njegovo razmerje do že pripravljenih vsebinskih podlag. Rezultat 1.3 je vzpostavil obsežen katalog mednarodnih, evropskih in nacionalnih standardov ter drugih referenčnih dokumentov, pri čemer jih je sistematiziral po standardizacijskih telesih in jih pri posameznem dokumentu obravnaval z vidika namena, vsebine, povezave z Aktom o umetni inteligenci, pomena za visokotvegane sisteme ter praktične uporabnosti. Njegova temeljna funkcija je bila kartiranje standardizacijskega prostora in oblikovanje referenčne baze. Ta dokument izhaja prav iz takšne podlage, vendar ima drugačen poudarek: iz širšega in heterogenega nabora referenc izpeljati zgoščeno, vsebinsko razvrščeno in strokovno ovrednoteno podlago, ki je že usmerjena v vprašanje njihove prednostne uporabnosti za slovensko okolje. V tem pogledu gre za prehod od pregleda k presoji.

Vsebinsko je dokument umeščen med začetno kartiranje standardov in kasnejšo tehnično artikulacijo nacionalnih rešitev. Njegova glavna vloga je, da iz že identificiranega nabora standardov in drugih referenčnih dokumentov oblikuje bolj zgoščeno in za slovenski prostor operativno relevantno podlago: katere standarde je smiselno obravnavati kot temeljne reference, katere kot podporne ali sektorske dokumente, katere je treba spremljati predvsem zaradi njihovega razvojnega pomena ter na katerih področjih obstoječi mednarodni ali evropski okvir še ne daje dovolj neposredne opore za enotno in praktično uporabo v slovenskem okolju. Takšna usmeritev je posebej pomembna zato, ker standardizacijski prostor umetne inteligence ni statičen, temveč se oblikuje na presečišču regulativnih zahtev, evropske harmonizacije, tehničnih specifikacij, varnostnih zahtev in sektorskih praks. Za deležnike zato ni ključno le, da imajo na voljo seznam dokumentov, temveč predvsem to, da imajo na voljo njihovo strokovno presojo, razvrstitev in praktično umeščenost.

Dokument je zasnovan kot strokovno-tehnična podlaga za nadaljnje delo z deležniki, standardizacijskimi organi in drugimi uporabniki projektnih rezultatov. Njegova bistvena

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

vrednost je v tem, da omogoča prehod od širokega pregleda standardov k bolj operativnemu razmisleku o tem, katere reference so za slovenski prostor zares nosilne, kako jih je smiselno razumeti v razmerju do Akta o umetni inteligenci ter kje se v praksi kaže potreba po dodatni nacionalni tehnični obdelavi, profiliranju ali dopolnitvi. Dokument je zato treba razumeti kot pripravljalno in usmerjevalno strokovno podlago, ki podpira nadaljnjo pripravo tehničnih predlogov za nacionalne standarde, usklajenih z mednarodnimi standardi, ter prispeva k bolj strukturiranemu dialogu s SIST-om in drugimi relevantnimi deležniki.

Za bralca je zato najpomembnejše naslednje: dokument pojasnjuje, na katere standarde se je v slovenskem prostoru smiselno opirati, kako te standarde razumeti v razmerju do evropskega regulativnega razvoja ter na katerih področjih je smiselno nadalje razvijati nacionalne tehnične dopolnitve, profile, smernice ali druge podporne dokumente. Njegov namen ni poenostaviti standardizacijskega prostora do mere, da bi prikril njegovo večplastnost, temveč ga narediti dovolj razumljivega, preglednega in operativno uporabnega, da lahko različni deležniki na njegovi podlagi sprejemajo informirane odločitve. S tem želi prispevati k temu, da bodo slovenski ponudniki, uporabniki, presojevalci skladnosti, regulatorji, javni naročniki in drugi deležniki lažje presodili, kateri standardi imajo zanje neposredno praktično vrednost, katere standarde je treba spremljati zaradi prihajajoče evropske harmonizacije in na katerih področjih so potrebni dodatni nacionalni tehnični napori.

Dokument je hkrati vsebinsko tesno povezan z drugimi ključnimi gradniki dosedanjega dela. Če rezultat 1.1 ponuja praktičen in normativno utemeljen pregled zahtev Akta o umetni inteligenci ter jih prevaja v procese, dokumentacijo, kontrolne sezname in spremljajoče artefakte, rezultat 1.3 vzpostavlja strukturiran katalog standardov, rezultat 2.3 pa iz izbranih primerov dobrih praks izlušči prenosljive organizacijske, procesne in tehnične elemente, ki jih je mogoče prevesti v konkretne zahteve, smernice in operativne kontrolne sezname. Ta dokument stoji prav na presečišču teh treh ravni: med regulativnim okvirom, standardizacijskim inventarjem in primeri praktične izvedbe. Njegova funkcija je zato izrazito sintezna: iz že pripravljenih podlag oblikovati bolj usmerjeno in za nadaljnjo uporabo uporabno strokovno celoto.

V tem okviru je treba poudariti še eno metodološko pomembno okoliščino. Standardizacijski prostor umetne inteligence je večnivojski in notranje raznolik. Ne obsega le mednarodnih standardov v ožjem pomenu besede, temveč tudi evropske prevzeme, tehnične specifikacije, tehnična poročila, referenčne arhitekture, sektorske smernice in druge podporne dokumente. Že rezultat 1.3 je pokazal, da za slovenski prostor niso pomembni zgolj dokumenti ISO/IEC JTC 1/SC 42, temveč tudi evropske izdaje CEN/CENELEC, dokumenti ETSI s področja kibernetske varnosti in omrežij, dokumenti IEEE s področja transparentnosti, zasebnosti in pristranskosti ter širši podatkovni in infrastrukturni dokumenti. Nacionalni standardizacijski razmislek zato ne more temeljiti na enem samem osrednjem standardu, temveč na premišljeni kombinaciji referenc, ki skupaj pokrivajo organizacijske, tehnične, podatkovne, varnostne in upravljavske vidike umetne inteligence.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Na tej podlagi je mogoče tudi natančneje opredeliti status in funkcijo tega dokumenta. Gre za besedilo, ki:

- povzema in selektivno nadgrajuje že opravljeno kartiranje standardizacijskega prostora,
- vzpostavlja strokovno podlago za nadaljnjo pripravo nacionalnih standardizacijskih predlogov,
- podpira nadaljnje delo z deležniki in standardizacijskimi organi ter
- je izrazito usmerjeno v praktično uporabo rezultatov v slovenskem okolju.

Njegova ključna dodana vrednost je v tem, da standardov ne obravnava kot seznama naslovov, temveč kot instrumente za vzpostavitev preverljive, razumljive in izvedljive prakse. Za slovenski prostor to pomeni predvsem podporo pri prehodu od abstraktnih zahtev Akta o umetni inteligenci in drugih regulativnih pričakovanj k bolj konkretnim tehničnim in procesnim rešitvam, ki jih bo mogoče uporabiti pri razvoju, uvajanju, presojanju in nadzoru sistemov umetne inteligence.

2. Metodologija in kriteriji vključitve

Pri pripravi tega dokumenta je bilo izhodišče, da standardizacijskega prostora umetne inteligence ni mogoče metodološko ustrezno obravnavati kot linearnega seznama dokumentov ali zgolj kot pregleda standardizacijskih organizacij. Gre za večnivojski normativno-tehnični ekosistem, v katerem sobivajo mednarodni standardi, evropski prevzemi, tehnične specifikacije, tehnična poročila, referenčni arhitekturni dokumenti, sektorske smernice in drugi podporni materiali, ki imajo različno funkcijo, različno stopnjo normativnosti in različno operativno vrednost. Že rezultat 1.3 je tak prostor obravnaval razvrščevalno, ne zgolj evidenčno: standarde je razdelil glede na obseg — mednarodni, EU/EN in nacionalni/SIST — ter jih filtriral po vsebinskih kriterijih, ki neposredno podpirajo skladnost, zlasti po upravljanju, upravljanju tveganj, kakovosti podatkov, preglednosti in razkritjih, človeškem nadzoru oziroma obvladljivosti, natančnosti, robustnosti in kibernetski varnosti ter dokumentaciji in sledljivosti. Kjer je bilo mogoče, je bila dodatno upoštevana tudi njihova relevantnost za Akt o umetni inteligenci in za visokotvegane sisteme.

Ta dokument to metodološko izhodišče ohranja, vendar ga uporablja na bolj selektiven in odločevalsko usmerjen način. Njegov namen ni ponovno enakomerno popisati celotnega standardizacijskega prostora, temveč iz že vzpostavljenega referenčnega kataloga izluščiti tiste dokumente in sklope dokumentov, ki imajo za slovenski prostor največjo strokovno, izvedbeno in standardizacijsko težo. Metodološki poudarek se zato premika od identifikacije k presoji, od popisa k razvrstitvi in od bibliografskega zajema k operativni interpretaciji. V tem smislu poglavje ne predstavlja zgolj tehničnega opisa metode, temveč pojasnjuje tudi,

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

kako je bil obstoječi standardizacijski material preveden v podlago za nadaljnje nacionalne tehnične razmisleke. Takšen pristop je skladen tudi z zaporedjem projektnih rezultatov, saj po začetnem pregledu regulativnega okvira in seznamu standardov sledita analiza primerov in začetek priprave predlogov nacionalnih standardov, kasneje pa še predlogi prilagoditev nacionalnih standardov s tehničnega vidika in dokumentacija tehničnih predlogov.

2.1 Metodološko izhodišče

Osnovna metodološka odločitev je bila, da se kot izhodiščna podlaga uporabi že pripravljeni rezultat 1.3, vendar ne kot zaključen in sam po sebi zadosten seznam, temveč kot referenčni katalog, iz katerega je treba izpeljati nadaljnjo vsebinsko selekcijo. Rezultat 1.3 je namreč že vzpostavil strukturiran pregled standardizacijskega prostora, pri katerem standardi niso bili zgolj našteti, temveč so bili umeščeni po standardizacijskih telesih in obravnavani glede na namen, vsebino, povezavo z Aktom o umetni inteligenci, pomen za visokotvegane sisteme in praktično uporabnost. V tem dokumentu je bila ta osnova uporabljena kot primarni korpus, vendar z drugačnim analitičnim ciljem: ne odgovoriti na vprašanje, kateri standardi obstajajo, temveč na vprašanje, kateri standardi, tehnične specifikacije in podporni dokumenti so za slovenski prostor nosilni, kateri delujejo predvsem podporno in na katerih mestih obstoječi mednarodni oziroma evropski okvir še ne daje dovolj neposredne opore za poenoteno nacionalno rabo.

Pri tem je bilo upoštevano, da standardi in sorodni dokumenti v prostoru umetne inteligence ne opravljajo enake funkcije. Del dokumentov določa organizacijski okvir ali zahteve sistema vodenja; drugi uvajajo terminološko in konceptualno jedro; tretji strukturirajo arhitekturne elemente sistemov; četrti operacionalizirajo kakovost podatkov, obvladljivost, kibernetsko varnost ali transparentnost; peti pa delujejo predvsem kot razlagalni in povezovalni instrumenti, ki omogočajo enotnejšo uporabo strožje normativnih dokumentov. To razlikovanje ni zgolj teoretično, temveč je neposredno pomembno za njihovo praktično uporabo. Že rezultat 1.3 izrecno poudari razliko med “International Standard”, “Technical Specification” in “Technical Report”, pri čemer standard določa, kaj mora biti zagotovljeno, tehnična specifikacija pokaže, kako to smiselno operativizirati, tehnično poročilo pa razjasni ozadje, logiko in skupni jezik. To razlikovanje je bilo zato ohranjeno kot eno izmed osrednjih metodoloških vodil tudi v tem dokumentu.

Dodatna metodološka orientacija izhaja iz rezultata 1.1, ki je pravni okvir Akta o umetni inteligenci obravnaval z izrazito izvedbenega vidika. Njegova metodologija je temeljila na trostopenjskem pristopu: najprej sistematizacija zakonskih zahtev in njihova preslikava v jasne odločilne točke, nato operativni prevod v procese, kontrolne sezname in dokazila, ter nazadnje priprava predlog dokumentacije in spremljajočih artefaktov za sledljivost. Za to poglavje je ta pristop pomemben zato, ker potrjuje osnovno projektno logiko: standardi niso obravnavani kot abstraktne reference, temveč kot most med pravnimi zahtevami in preverljivimi tehničnimi oziroma organizacijskimi rešitvami. To pomeni, da je tudi pri izboru

in razvrstitvi standardov relevantno predvsem vprašanje, kako posamezen dokument podpira prevod regulativnih obveznosti v procese, kontrolne točke, dokumentacijo, evidence in preverljiva dokazila.

Metodološki okvir je bil dodatno dopolnjen z logiko rezultata 2.3, ki je pri izboru dobrih praks uporabil ciljno usmerjen pristop: prednost so imeli primeri z visoko kredibilnimi primarnimi viri, z dovolj podrobno tehnično in organizacijsko dokumentacijo ter z možnostjo prenosa ugotovitev v konkretne zahteve, smernice in kontrolne sezname. Čeprav je predmet obravnave tam drugačen — primeri uporabe in ne standardi — je ta metodološka logika pomembna tudi za obravnavani dokument. Potrjuje namreč, da je bila v projektu že dosledno uporabljena selekcijska metoda, ki ne privilegira formalnega statusa vira samega po sebi, temveč njegovo dejansko uporabnost, prenosljivost in sposobnost, da podpira operativno izvajanje zahtev v praksi.

2.2 Obseg vključene dokumentacije

V obseg obravnave so bili vključeni predvsem tisti mednarodni, evropski in za slovenski prostor relevantni dokumenti, ki lahko neposredno ali posredno podpirajo preverjanje skladnosti sistemov umetne inteligence, njihovo varno in sledljivo uvajanje ter pripravo dokazljive tehnične in organizacijske dokumentacije. Izhodišče vključitve je bilo torej vsebinsko in funkcionalno, ne zgolj formalno. To pomeni, da v obravnavo niso bili zajeti le dokumenti, ki imajo status klasičnega mednarodnega ali evropskega standarda, temveč tudi dokumenti, ki v praksi opravljajo pomembno podporno funkcijo pri razlagi zahtev, tehnični implementaciji, terminološki uskladitvi ali oblikovanju enotnega strokovnega jezika. Takšno izhodišče je na področju umetne inteligence posebej pomembno, ker je standardizacijski prostor še vedno v intenzivnem razvoju in ker praktična uporabnost posameznega dokumenta pogosto presega njegovo formalno kategorizacijo.

Korpus izhodiščne dokumentacije je sledil širini, že vzpostavljeni v rezultatu 1.3. Ta je poleg dokumentov ISO/IEC JTC 1/SC 42 zajel tudi evropske izdaje standardov CEN/CENELEC, dokumente ETSI, IEEE SA in ITU-T ter širše referenčne dokumente s področja podatkovnih prostorov, interoperabilnosti in infrastrukture. V njem so bili kot jedrni standardi posebej izpostavljeni ISO/IEC 42001 za sistem vodenja UI, ISO/IEC 23894 za upravljanje tveganj, ISO/IEC 22989 za terminologijo, ISO/IEC 23053 za arhitekturni okvir sistemov strojnega učenja, ISO/IEC 5259-1 in 5259-3 za kakovost podatkov, ISO/IEC TS 8200 za obvladljivost oziroma človeški nadzor ter sorodni evropski EN prevzemi teh dokumentov. Dopolnilno so bili vključeni tudi ETSI dokumenti s področja kibernetske varnosti UI, med drugim ETSI TS 104 050 in ETSI TS 104 223, ter izbrani IEEE dokumenti s področja etike, transparentnosti, zasebnosti in pristranskosti. Ta širina ni bila naključna, temveč je odražala metodološko oceno, da nacionalni standardizacijski razmislek ne more temeljiti le na enem “osrednjem” standardu, temveč na kombinaciji dokumentov, ki skupaj pokrivajo organizacijski, podatkovni, tehnični, varnostni in upravljavski vidik umetne inteligence.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Obseg vključitve je bil vsebinsko omejen na tiste dokumente, ki so lahko relevantni za praktično uporabo v življenjskem ciklu sistemov umetne inteligence: pri njihovem razvoju, uvedbi, upravljanju, presoji skladnosti, nadzoru, pripravi notranjih pravil, javnih naročilih, pogodbenih mehanizmih, tehnični dokumentaciji in vzpostavitvi dokazljive sledljivosti. V tem smislu je bil dokumentarno prednostno obravnavan tisti del standardizacijskega prostora, ki omogoča prehod od regulativnih in organizacijskih načel k preverljivim artefaktom in tehničnim postopkom. Prav to je v rezultatu 1.1 izrecno poudarjeno kot eden od osrednjih ciljev: povezovanje zakonskih zahtev s procesi v organizacijah, s tehnično dokumentacijo, vodenjem evidenc, človeškim nadzorom, robustnostjo in kibernetsko varnostjo ter s spremljajočimi artefakti za sledljivost.

Vključitev dokumentov je bila torej zasnovana široko, vendar ne nekritično. Dokumenti, ki bi bili za obravnavano temo le posredno ali marginalno relevantni, niso bili v ospredju, tudi če so formalno umeščeni v isti standardizacijski ekosistem. Prednost so imeli tisti dokumenti, ki podpirajo eno ali več od naslednjih funkcij: organizacijsko upravljanje sistemov UI, upravljanje tveganj, terminološko poenotenje, arhitekturno razumljivost, kakovost podatkov, človeški nadzor in obvladljivost, kibernetsko varnost, transparentnost, sledljivost, pripravo tehnične dokumentacije in možnost dokazovanja skladnosti. Takšna opredelitev obsega je hkrati zagotavljala, da končni nabor ostane zvest praktični usmeritvi projekta in ne zdrsne v širši, vendar manj uporaben enciklopedični pregled.

2.3 Kriteriji vključitve in razvrščanja

Pri izboru in razvrščanju dokumentov so bili uporabljeni medsebojno povezani kriteriji, katerih namen ni bil zgolj zagotoviti strokovno korektnost nabora, temveč predvsem njegovo metodološko doslednost in uporabnost za slovenski prostor. Ti kriteriji niso delovali izolirano, temveč kumulativno: posamezni dokument je bil praviloma pomembnejši, če je hkrati izpolnjeval več kriterijev.

Prvi kriterij je bila **vsebinska relevantnost za skladnost in upravljanje sistemov umetne inteligence**. Izhodiščna referenca za ta filter je bil že rezultat 1.3, ki je standarde razvrstil po ključnih vsebinskih sklopih, ki neposredno podpirajo skladnost: upravljanje in sistemi vodenja kakovosti, upravljanje tveganj, kakovost podatkov, preglednost in razkritja, človeški nadzor oziroma obvladljivost, natančnost, robustnost in kibernetska varnost ter dokumentacija in sledljivost. Ta kriterij je bil v tem dokumentu ohranjen kot primarni filter, saj omogoča razlikovanje med dokumenti, ki odgovarjajo na osrednja praktična vprašanja uvajanja UI, in tistimi, ki so za obravnavani namen le postransko relevantni.

Drugi kriterij je bil **normativni in izvedbeni status dokumenta**. Pri obravnavi standardov je bilo ključno razlikovati med mednarodnimi standardi, evropskimi prevzemi, tehničnimi specifikacijami, tehničnimi poročili in drugimi referenčnimi dokumenti. To razlikovanje je metodološko pomembno zato, ker neposredno vpliva na način njihove uporabe v praksi. Standard praviloma predstavlja normativno jedro; evropski prevzem je pomemben zaradi

tesnejše povezave z evropskim regulatornim in nadzornim prostorom; tehnična specifikacija pogosto deluje kot praktični most med zahtevami in izvedbo; tehnično poročilo pa razjasni pojmovni okvir, ozadje in logiko. Za slovenski prostor je takšna razmejitev zlasti pomembna pri odločanju, ali je dokument primeren za neposredno sklicevanje, za spremljanje, za strokovno orientacijo ali kot osnova za morebitno nadaljnjo nacionalno tehnično obdelavo. To razlikovanje je v rezultatu 1.3 jasno izpostavljeno prav pri razlagi različnih tipov dokumentov znotraj sklopa ISO/IEC JTC 1/SC 42.

Tretji kriterij je bila **evropska in nacionalna uporabnost**. Pri posameznem dokumentu ni bilo odločilno le to, da obstaja, temveč tudi, ali ima evropsko izdajo oziroma ali ga je mogoče učinkovito uporabljati prek nacionalnega standardizacijskega sistema. Rezultat 1.3 izrecno poudarja praktični pomen evropskih in nacionalnih izdaj za slovenske organizacije, saj te olajšajo enotno sklicevanje v razpisih, pogodbah, certifikacijah in notranjih pravilnikih ter pripomorejo k bolj usklajenemu dialogu z drugimi državami članicami. Ta kriterij je bil posebej pomemben zato, ker rezultat 2.2 ni namenjen abstraktni mednarodni primerjavi, temveč pripravi uporabne podlage za slovensko okolje, ki mora biti po eni strani vsebinsko usklajeno z evropskim razvojem, po drugi strani pa institucionalno in tehnično izvedljivo.

Četrti kriterij je bila **povezanost z zahtevami Akta o umetni inteligenci in s preverjanjem skladnosti**. Kjer je bilo mogoče, so bile pri dokumentih upoštevane opombe o njihovi relevantnosti za AIA, zlasti za visokotvegane sisteme. Ta kriterij je pomemben, ker je eden od praktičnih ciljev projekta prav podpreti slovenske deležnike pri razumevanju, kako posamezne standarde uporabljati v okolju, kjer bo dokazovanje skladnosti z evropskim regulativnim okvirom vse pomembnejše. Rezultat 1.1 pri tem izrecno pokaže, da je treba pravne zahteve prevesti v procese, dokumentacijo, kontrolne sezname in spremljajoča dokazila; standardi pa v tem prevodu delujejo kot tehnični in organizacijski nosilci izvedbe.

Peti kriterij je bila **uporabnost za slovenske deležnike**. To pomeni, da so imeli prednost tisti dokumenti, ki jih je mogoče smiselno uporabiti pri pripravi internih pravil, tehnične dokumentacije, organizacijskih procesov, presoj skladnosti, javnih naročil, pogodbenih mehanizmov, strokovnih smernic ali nacionalnih profilov. To merilo je v skladu tudi s širšim projektnih izhodiščem, po katerem so odsotnost razumljivih smernic, pomanjkljivo tehnično razumevanje, pomanjkanje informacij za certifikacijske organe in podvajanje dela med ključnimi praktičnimi problemi, ki jih želi projekt nasloviti. Če naj dokument k temu prispeva, mora biti metodologija vključitve nujno uporabnostno naravnana.

Šesti kriterij je bila **prenosljivost v nadaljnje nacionalne tehnične razmisleke**. Posamezen dokument je bil toliko pomembnejši, kolikor bolj je lahko služil kot osnova za oblikovanje nacionalnega profila, tehnične dopolnitve, priporočila ali drugega podpornega dokumenta. Ta kriterij je posebej relevanten zato, ker je pomembno razločevati med podlago za pripravo nacionalnih standardov in kasnejšim rezultatom 4.1, ki je že izrecno opredeljen kot “Predlog prilagoditev nacionalnih standardov (tehnični vidik)”, ter rezultatom 5.2, ki predvideva dokumentacijo tehničnih predlogov za nacionalne standarde. Dokumenti so bili zato

presojani tudi z vidika tega, ali omogočajo nadaljnje tehnično profiliranje v slovenskem okolju.

Sedmi kriterij je bila **dokazljivost in operativna granularnost vsebine**. Tu je metodološka analogija z rezultatom 2.3 posebej uporabna. Tako kot so bili pri dobrih praksah izbrani tisti primeri, katerih uradna dokumentacija vsebuje dovolj podrobnosti o arhitekturi, upravljanju, varnostnih kontrolah, podatkih, človeškem nadzoru in izboljševanju delovanja, so bili tudi pri standardih višje vrednoteni tisti dokumenti, ki omogočajo dovolj konkretno preslikavo v procese, kontrole, evidence in dokazila. Dokument, ki ostane na ravni splošne programske usmeritve, ima seveda lahko pomembno orientacijsko vrednost, vendar je za namen nadaljnje tehnične operacionalizacije manj nosilen kot dokument, ki omogoča določitev jasne povezave "zahteva – kontrola – dokazilo". Prav takšna povezava je bila v rezultatih 1.1 in 1.3 večkrat poudarjena kot ključ do praktične uporabnosti standardov.

2.4 Metoda obdelave gradiva

Na podlagi navedenih kriterijev je bila uporabljena večstopenjska metoda obdelave gradiva. Prvi korak je predstavljal pregled obstoječega nabora dokumentov iz rezultata 1.3 kot osnovnega referenčnega kataloga. V tej fazi je bilo ključno razlikovati med jedrnimi dokumenti in širšim podpornim okoljem. Med jedrne dokumente so bili uvrščeni tisti, ki neposredno urejajo upravljanje, tveganja, terminologijo, kakovost podatkov, človeški nadzor, transparentnost, kibernetsko varnost in dokazljivost sistemov UI. Podporno okolje pa sestavljajo dokumenti, ki prispevajo pomembne dodatne usmeritve na ravni sektorskih arhitektur, podatkovne interoperabilnosti, posebnih tehničnih profilov ali operativnih varnostnih kontrol. Takšno razlikovanje je omogočilo, da se standardizacijski prostor ne obravnava kot homogena masa, temveč kot strukturiran sistem z različnimi plastmi uporabnosti.

Drugi korak je bil **funkcionalna razvrstitev dokumentov** glede na vlogo, ki jo opravljajo v praksi. Posamezni dokumenti so bili zato obravnavani glede na to, ali primarno vzpostavljajo organizacijski okvir, opredeljujejo terminologijo, strukturirajo arhitekturne elemente, urejajo podatkovne vidike, naslavljajo varnostne kontrole, prispevajo k transparentnosti in etičnemu oblikovanju ali pa delujejo predvsem kot razlagalni in povezovalni instrumenti. Takšna funkcionalna razvrstitev je bila nujna zato, ker standardizacijski prostor umetne inteligence vključuje dokumente, ki rešujejo zelo različne probleme. Šele ko so ti problemi metodološko razločeni, je mogoče smiselno presojati, kje mednarodni oziroma evropski okvir že daje dovolj trdno osnovo in kje se odpirajo vrzeli, ki bi lahko upravičevale nadaljnjo nacionalno tehnično obdelavo.

Tretji korak je obsegal **presojo razmerja med posameznim dokumentom in regulativnim okvirom**. Tu je bil ključni referenčni okvir rezultat 1.1, ki pravne zahteve AIA poveže z organizacijskimi procesi, tehnično dokumentacijo, vodenjem evidenc, človeškim nadzorom, robustnostjo in kibernetsko varnostjo. Posamezni standardi in drugi dokumenti so bili zato

presojani tudi z vidika tega, katere od teh zahtev lahko podpirajo, konkretizirajo ali olajšajo. Ta korak ni pomenil pravne kvalifikacije standardov, temveč operativno oceno njihove uporabnosti pri oblikovanju skladne prakse. Na tej ravni je bilo mogoče razločiti med dokumenti, ki so za slovenski prostor predvsem orientacijsko pomembni, in tistimi, ki bi jih bilo smiselno obravnavati kot nosilne tehnične reference.

Četrti korak je bila **presoja prenosljivosti v slovenski standardizacijski in institucionalni prostor**. Pri tem je bilo ključno vprašanje, ali je dokument primeren za neposredno referenciranje, ali ga je smiselno predvsem spremljati, ali bi ga bilo mogoče priporočiti kot podlago za oblikovanje nacionalnih profilov oziroma tehničnih dopolnitev ali pa njegova vrednost leži predvsem v strokovni orientaciji in terminološki uskladitvi. Ta presoja ni bila zamišljena kot formalna odločitev o nacionalnem prevzemu standarda, temveč kot tehnična ocena njegove potencialne funkcije v slovenskem okolju. Prav zaradi tega je bila v obravnavo zasnovana širše, z vključevanjem SIST-a, delavnice z deležniki, pripravo dobrih praks, nadgradnjo Observatorija in kasnejšo pripravo konkretnejših tehničnih predlogov.

Peti korak je bila **sinteza s primerjalnim in izvedbenim gradivom iz rezultata 2.3**. Namen tega koraka ni bil ponovno analizirati dobrih praks kot takih, temveč preveriti, katere vrste tehničnih, organizacijskih in upravljavskih vprašanj se v verodostojnih primerih uporabe dejansko ponavljajo in v kolikšni meri so ta vprašanja že naslovljena z obstoječimi standardi. Rezultat 2.3 je pri tem posebej uporaben, ker izhaja iz primerov z visoko kredibilnimi viri in z dovolj tehnično-operativnimi podrobnostmi, da jih je mogoče prevesti v konkretne zahteve, smernice in kontrolne sezname. Zato so bili njegovi metodološki poudarki — kredibilnost vira, operativna uporabnost, raznolikost in možnost prenosa v zahteve — uporabljeni kot dodatna validacijska osnova pri presoji, kateri standardi imajo največjo izvedbeno vrednost.

2.5 Metodološki doseg in namen uporabe

Uporabljena metodologija omogoča predvsem troje. Prvič, omogoča bolj pregledno in hierarhično razumevanje obstoječega standardizacijskega prostora, saj ločuje med jedrnimi in podpornimi dokumenti, med normativnimi in razlagalnimi viri ter med dokumenti z visoko neposredno operativno vrednostjo in tistimi, ki imajo predvsem orientacijsko ali razvojno funkcijo. Drugič, omogoča povezavo med standardi in dejanskimi potrebami uporabnikov: standardi so obravnavani skozi njihovo zmožnost, da podprejo procese, dokumentacijo, dokazila, kontrolne sezname in druge praktične gradnike skladnega uvajanja UI. Tretjič, vzpostavlja dovolj trdno podlago za nadaljnjo strokovno razpravo o tem, katere vrste nacionalnih standardov, nacionalnih profilov, tehničnih smernic ali drugih podpornih dokumentov bi bilo v slovenskem prostoru smiselno pripravljati v nadaljevanju.

Metodološki doseg dokumenta je zato analitičen, selekcijski in usmerjevalen. Njegov cilj ni formalno odločanje o tem, kateri standardi bodo v Sloveniji prevzeti, niti ne nadomešča standardizacijskih postopkov, ki sodijo v pristojnost SIST-a in pristojnih tehničnih teles. Njegova funkcija je drugačna in za projekt bistvena: standardizacijski prostor strokovno

uredi, ga razvrsti po dejanski praktični teži in pripravi teren za nadaljnje tehnično delo. V tem je njegova posebna uporabna vrednost. Ker ni vezan na formalno strukturo standarda samega, lahko jasneje poveže mednarodne in evropske reference z življenjskim ciklom sistemov umetne inteligence ter s potrebami ponudnikov, uporabnikov, presojevalcev skladnosti, regulatorjev in drugih deležnikov.

Za uporabnika tega dokumenta to pomeni, da posameznih standardov ne obravnava zgolj kot referenc v formalnem smislu, temveč kot gradnike širšega sistema skladnosti, kakovosti, varnosti in zaupanja. Metodologija je zato namenjena ne le strokovni sistematizaciji, temveč tudi praktični orientaciji: omogoča razumevanje, kateri dokumenti so primerni kot nosilne reference, kateri kot podporni instrumenti, katere je treba spremljati zaradi njihovega razvojnega pomena in kje obstoječi okvir še odpira prostor za nadaljnjo nacionalno tehnično artikulacijo. Prav v tem smislu to poglavje opravlja svojo osrednjo funkcijo: ne postavlja še nacionalnih rešitev, temveč natančno pojasni, kako je bila pripravljena strokovna podlaga, na kateri bo tak razmislek sploh mogoče nadaljevati.

3. Shema standardizacijskega ekosistema za umetno inteligenco

Za razumevanje nadaljnjega pregleda standardov je standardizacijski prostor umetne inteligence smiselno predstaviti kot **večslojni in funkcionalno diferenciran ekosistem**, ne kot enoten seznam dokumentov. Standardi s področja umetne inteligence namreč ne nastajajo znotraj ene same organizacije, niti ne opravljajo enake vloge. Nekateri dokumenti vzpostavljajo temeljne pojme, referenčne modele in skupni strokovni jezik; drugi urejajo upravljanje, tveganja, kakovost podatkov, preglednost, človeški nadzor, robustnost ali kibernetsko varnost; tretji pa odgovarjajo na bolj specializirane sektorske, infrastrukturne ali izvedbene potrebe. Prav zato je za uporabnika pomembno predvsem to, da razume, **na kateri ravni posamezen dokument nastaja, kakšen je njegov normativni status in kakšno praktično funkcijo opravlja**. Tak pristop je skladen tudi z rezultatom 1.3, ki standardov ne obravnava zgolj evidenčno, temveč jih sistematizira po akterjih, vsebini in uporabnosti za preverjanje skladnosti rešitev UI.

Za orientacijo je posebej pomembno tudi razlikovanje med **vrstami standardizacijskih dokumentov**. V mednarodnem prostoru imajo osrednjo težo mednarodni standardi, vendar poleg njih pomembno vlogo opravljajo tudi tehnične specifikacije in tehnična poročila. Mednarodni standard praviloma predstavlja zrelejši konsenz in stabilnejšo referenco; tehnična specifikacija je praviloma primerna tam, kjer se področje še razvija ali kjer je potrebna hitrejša operacionalizacija določenega vprašanja; tehnično poročilo pa deluje predvsem kot razlagalni in sistematizacijski dokument. V evropskem prostoru imajo jedrno normativno funkcijo evropski standardi (EN), poleg njih pa nastajajo tudi tehnične specifikacije, tehnična poročila, vodniki in drugi podporni dokumenti. To razlikovanje ni le

formalno, temveč neposredno vpliva na to, ali je posamezen dokument primeren kot **stabilna referenca za sklicevanje**, kot **operativna izvedbena opora** ali predvsem kot **interpretativni in razvojni dokument**. Tudi rezultat 1.3 prav na tej točki poudarja, da je za učinkovito uporabo standardov nujno razumeti razliko med dokumenti, ki določajo zahteve, dokumenti, ki pomagajo pri njihovi izvedbi, in dokumenti, ki razlagajo širši konceptualni okvir.

V nadaljevanju je zato standardizacijski ekosistem predstavljen v **treh medsebojno povezanih slojih**: mednarodnem, evropskem in nacionalnem. Takšna delitev je za to poročilo posebej uporabna, ker pokaže dejanski tok standardizacije na področju umetne inteligence. Horizontalni tehnični in konceptualni temelji praviloma nastajajo na mednarodni ravni; evropska raven jih prevzema, dopolnjuje in povezuje z regulativnimi potrebami enotnega trga; nacionalna raven pa omogoča njihovo institucionalno uporabo, strokovno usklajevanje, formalni prevzem in praktično vključitev v domače organizacijsko in nadzorno okolje. Takšna večslojna struktura je v neposredni zvezi tudi s cilji projekta NOSI-AI, ki predvideva spremljanje, kategorizacijo in analizo standardov, sodelovanje s SIST in evropskimi telesi, vzpostavitev Observatorija standardov UI ter pripravo tehničnih podlag za nadaljnje nacionalne standardizacijske razmisleke.

3.1 Mednarodni sloj

Mednarodni sloj je izhodiščna raven standardizacijskega ekosistema umetne inteligence. Na tej ravni nastajajo dokumenti, katerih namen je vzpostaviti **široko uporabne, čezsektorske in čezjurisdikcijske reference**. Poudarek je zato praviloma na terminologiji, referenčnih arhitekturah, upravljanju, obvladovanju tveganj, podatkovnih vidikih, zaupanja vredni umetni inteligenci ter na splošnih tehničnih pristopih, ki so lahko uporabni v različnih panogah in pravnih okoljih. Njihova osrednja vrednost je v tem, da vzpostavijo skupni jezik in osnovne metodološke modele, na katerih lahko nato gradijo sektorski, evropski in nacionalni pristopi. Projektna dokumentacija in rezultat 1.3 oba potrjujeta, da je prav ta raven tista, na kateri se oblikuje temeljni referenčni okvir za nadaljnje standardizacijske aktivnosti.

V tem sloju ima osrednje mesto **ISO/IEC JTC 1/SC 42 Artificial Intelligence**, ki predstavlja glavno horizontalno telo mednarodne standardizacije umetne inteligence. V rezultatu 1.3 je prav SC 42 obravnavan kot osrednji vir standardov, ki neposredno podpirajo skladnost, upravljanje, tveganja, arhitekturo sistemov, kakovost podatkov in človeški nadzor. V ta okvir sodijo zlasti standardi, kot so ISO/IEC 42001 za sistem vodenja umetne inteligence, ISO/IEC 23894 za upravljanje tveganj, ISO/IEC 22989 za pojme in terminologijo, ISO/IEC 23053 za arhitekturni okvir sistemov strojnega učenja, ISO/IEC 5259-1 in 5259-3 za kakovost podatkov ter ISO/IEC TS 8200 za obvladljivost avtomatiziranih sistemov UI. Skupaj ti dokumenti ne predstavljajo seznama nepovezanih standardov, temveč relativno koherentno horizontalno ogrodje: 42001 ureja sistem vodenja, 23894 metodologijo tveganj, 22989

terminološko jedro, 23053 arhitekturno strukturiranje, 5259 podatkovne zahteve, TS 8200 pa praktične mehanizme človeškega nadzora. Mednarodni sloj zato na tej ravni ni zgolj konceptualen, temveč že izrazito izvedbeno relevanten.

Pomembno dopolnilno vlogo v mednarodnem sloju opravlja tudi **IEEE Standards Association (IEEE SA)**. Če je ISO/IEC primarni horizontalni nosilec tehnično-upravljavskega okvira, IEEE pogosteje prispeva bolj granularne dokumente s področja etike, preglednosti, zasebnosti, pristranskosti in odgovorne zasnove inteligentnih sistemov. Rezultat 1.3 izrecno poudarja serijo IEEE 7000 kot pomemben vir za operacionalizacijo etičnih in družbenih vidikov UI, zlasti tam, kjer organizacije potrebujejo bolj konkretne usmeritve za transparentnost, vključevanje vrednot v načrtovanje sistema, obvladovanje pristranskosti ali procesno varovanje zasebnosti. IEEE tako v standardizacijskem ekosistemu ne deluje kot konkurent ISO/IEC, temveč kot pomemben dopolnilni vir za vprašanja, ki presegajo zgolj arhitekturo in upravljanje ter segajo v domene odgovornega oblikovanja in družbenih učinkov.

Tretji pomemben akter mednarodnega sloja je **ITU-T**, zlasti tam, kjer se umetna inteligenca prepleta z omrežji, digitalnimi storitvami, multimedijo in komunikacijsko infrastrukturo. V rezultatu 1.3 so priporočila ITU-T predstavljena predvsem kot domensko natančne reference za arhitekturne okvire, ravnanje s podatki v omrežjih, primere uporabe in preverjanje skladnosti v kontekstu prihodnjih omrežij in digitalnih storitev. Njihova posebna vrednost je v tem, da horizontalne standardizacijske koncepte prevajajo v infrastrukturno specifična okolja, kjer so interoperabilnost, kakovost storitev, operativna zanesljivost in podatkovni tokovi še posebej pomembni. ITU-T je zato v tej shemi relevanten predvsem kot most med splošno standardizacijo umetne inteligence in tehnično-operativnimi potrebami komunikacijskih sistemov.

Na meji med mednarodnim in evropskim prostorom je smiselno omeniti še **ETSI**. Formalno ETSI sodi med evropske standardizacijske organizacije, vendar ima zaradi svoje vloge v digitalni infrastrukturi, telekomunikacijah, IoT, kibernetski varnosti in interoperabilnosti izrazit širši vpliv. Rezultat 1.3 kaže, da so ETSI-jevi dokumenti posebej pomembni tam, kjer umetna inteligenca nastopa v povezavi z varnostjo modelov in sistemov, ontologijo groženj, robnim računalništvom, omrežnimi funkcijami in drugimi infrastrukturnimi vprašanji. V tej shemi ga je zato smiselno razumeti kot organizacijo, ki sicer formalno spada v evropski sloj, vendar vsebinsko pogosto deluje kot tehnično-specializirani akter z izrazito mednarodno relevantnostjo.

Če mednarodni sloj povzamemo z vidika njegove funkcije v tem poročilu, ga lahko razumemo kot **temeljni konceptualni in metodološki horizont**. Na tej ravni se oblikujejo osnovni pojmi, ključni procesni modeli in splošni tehnični pristopi, brez katerih kasnejši evropski in nacionalni standardizacijski razvoj ne bi imel dovolj stabilne osnove. Hkrati pa je treba poudariti, da mednarodni sloj sam po sebi še ne rešuje vseh vprašanj praktične uporabe v evropskem pravnem in institucionalnem prostoru. Prav zato postane odločilen naslednji sloj.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

3.2 Evropski sloj

Evropski sloj standardizacijskega ekosistema ni zgolj regionalni odsev mednarodnega sloja, temveč raven, na kateri se mednarodni dokumenti **prevajajo v evropski regulatorni, institucionalni in tržni kontekst**. Njegova posebna funkcija je, da mednarodne tehnične rešitve prevzema, dopolnjuje ali po potrebi prilagaja tako, da postanejo uporabne v okviru prava EU, enotnega trga in evropskih postopkov ugotavljanja skladnosti. Na področju umetne inteligence je ta raven posebej pomembna, ker standardi vse bolj nastopajo kot most med abstraktnimi pravnimi zahtevami in preverljivimi tehnično-organizacijskimi praksami.

Na evropski ravni so ključni akterji **CEN, CENELEC in ETSI**. Za področje umetne inteligence je osrednji forum predvsem **CEN-CENELEC JTC 21 Artificial Intelligence**, ki v projektni dokumentaciji nastopa kot glavni evropski sogovornik pri oblikovanju relevantnih standardov, medtem ko rezultat 1.3 potrjuje praktični pomen evropskih izdaj standardov kot mostu med mednarodnimi vsebinami in evropskim pravnim okoljem. Evropski standardi imajo v tem okviru posebno težo zato, ker predstavljajo skupni referenčni jezik za organe držav članic, nadzorne postopke, javna naročila, presoje skladnosti in druge institucionalne procese v EU. V praksi to pomeni, da evropski sloj ni pomemben le zaradi vsebine posameznega dokumenta, temveč tudi zaradi njegove **formalno-institucionalne umeščenosti** v evropski standardizacijski sistem.

Posebej pomembna značilnost evropskega sloja je, da omogoča **uporabo evropskih izdaj mednarodnih standardov**. Rezultat 1.3 pri tem navaja več konkretnih primerov, med drugim EN ISO/IEC 23894:2024, EN ISO/IEC 22989:2023 in EN ISO/IEC 23053:2023, ki so za slovenski prostor še posebej relevantni prav zato, ker obstajajo tudi kot nacionalne izdaje SIST EN. To bistveno olajša sklicevanje v razpisih, pogodbah, tehnični dokumentaciji, presojah in komunikaciji z nadzornimi organi. Evropski sloj zato ni le vsebinska plast, temveč tudi operativni mehanizem, prek katerega standardi postanejo v praksi lažje dostopni, bolj primerljivi in za deležnike znotraj EU tudi bolj pričakovani. Prav v tem smislu je evropska raven za slovenski prostor pogosto pomembnejša od zgolj neposrednega sklicevanja na izvirno mednarodno izdajo.

Evropski sloj je hkrati tesno povezan z regulatornim razvojem EU, zlasti tam, kjer standardi podpirajo izvajanje zahtev Akta o umetni inteligenci. Tudi če posamezen evropski standard še nima posebnega statusa v okviru prihodnjih postopkov skladnosti, je njegova vrednost v tem, da deluje kot **prevedbeni mehanizem med globalnimi tehničnimi pristopi in evropskimi pravnimi pričakovanji**. Rezultat 1.3 v tem pogledu posebej poudarja, da evropske izdaje standardov živijo znotraj evropskega regulativnega in nadzornega okolja ter so zato za deležnike v Sloveniji neposredno uporabne pri oblikovanju tehnične dokumentacije, procesov in dokazil. Evropski sloj je torej operativna raven, na kateri se

mednarodni standardizacijski material prilagodi okolju, v katerem bodo morale organizacije dejansko izkazovati svojo skladnost in zrelost.

V okviru evropskega sloja je treba posebej izpostaviti tudi **ETSI**, vendar tukaj z nekoliko drugačnim poudarkom kot v mednarodnem sloju. Če je bilo prej pomembno njegovo širše infrastrukturno delovanje, je na tej ravni ključno to, da je ETSI eden od treh formalno priznanih evropskih standardizacijskih organov. To pomeni, da so ETSI-jevi dokumenti pomembni ne le zaradi svoje tehnične vsebine, temveč tudi zaradi njihove umeščenosti v evropski standardizacijski prostor. Za področje umetne inteligence je ETSI posebej relevanten tam, kjer se UI povezuje z omrežji, kibernetsko varnostjo, preskušanjem, interoperabilnostjo in operativno infrastrukturo. V rezultatih 1.3 so prav ETSI-jeve tehnične specifikacije in poročila predstavljeni kot pomembna opora za operacionalizacijo robustnosti, varnosti in infrastrukturne odpornosti sistemov UI.

Če evropski sloj povzamemo z vidika funkcije v tem dokumentu, ga je mogoče razumeti kot **najpomembnejšo operativno raven za slovenski prostor**. Mednarodni dokumenti zagotavljajo osnovno strukturo, terminologijo in metodološko ogrodje, vendar je prav evropska raven tista, ki te vsebine umešča v postopke, prakse in pričakovanja, relevantna za organizacije, ki delujejo znotraj EU. Prav zato bo v nadaljnjih tehničnih razmislekih o slovenskem prostoru evropski sloj praviloma deloval kot primarna referenčna ravnina.

3.3 Nacionalni sloj

Nacionalni sloj standardizacijskega ekosistema ni ločen in samozadosten standardizacijski svet, temveč raven, na kateri se mednarodni in evropski dokumenti **prevzamejo, obravnavajo, uskladijo in uporabijo v konkretnem institucionalnem okolju države**. Za slovenski prostor je ta sloj posebej pomemben zato, ker prav na njem poteka dejanska vključitev domačih deležnikov v standardizacijske procese, oblikovanje nacionalnih stališč, javna obravnava osnutkov, formalni prevzem standardov in njihova uporaba v organizacijski, gospodarski in nadzorni praksi. Tudi projekt NOSI-AI je od začetka zasnovan prav okoli te potrebe: ne kot ustvarjanje avtonomnega slovenskega standardizacijskega sistema od začetka, temveč kot priprava strokovne podlage, Observatorija, seznama standardov, dobrih praks in tehničnih predlogov, ki bi slovenskim deležnikom olajšali vključevanje v širši evropski in mednarodni okvir.

V Sloveniji to vlogo opravlja Slovenski inštitut za standardizacijo (SIST), ki je osrednji nacionalni kanal za prenos mednarodnih in evropskih standardov v slovenski sistem ter za vključevanje slovenskih strokovnjakov v standardizacijske aktivnosti. Praktična vrednost SIST-a je večplastna. Prvič, omogoča dostop do evropskih in mednarodnih izdaj standardov in njihovo nacionalno objavo. Drugič, zagotavlja institucionalni okvir, znotraj katerega se standardi ne le formalno prevzemajo, temveč tudi strokovno razlagajo, usklajujejo in

uporabljajo. Tretjič, deluje kot forum, prek katerega lahko slovenski deležniki vplivajo na razvoj prihodnjih standardov in se vključujejo v mednarodno ter evropsko standardizacijsko delo. SIST je pomemben deležnik pri pripravi seznama standardov in pri opredeljevanju potreb po nadaljnjih nacionalnih tehničnih rešitvah.

Nacionalni sloj je v praksi pomemben predvsem zaradi tega, ker omogoča prevod standardov v dejansko slovensko uporabo. To pomeni uporabo v razpisih, pogodbah, notranjih pravilih, tehnični dokumentaciji, presojah skladnosti, usposabljanjih in strokovni komunikaciji med organizacijami, regulatorji in presojevalci. Rezultat 1.3 posebej opozarja, da imajo evropske in nacionalne izdaje standardov za slovenske deležnike praktično prednost prav zato, ker olajšajo enotno sklicevanje in ustvarjajo primerljive pogoje za dialog z drugimi državami članicami, nadzornimi organi in strokovno javnostjo. Nacionalni sloj je torej ključen z vidika uporabnosti, dostopnosti in institucionalne operativnosti.

Za področje umetne inteligence je ta sloj dodatno pomemben tudi zato, ker so zaznane vrzeli: pomanjkanja jasnih smernic, omejenega tehničnega razumevanja, pomanjkanja informacij za certifikacijske organe, potrebe po razumljivih in uporabnih standardih ter nevarnosti podvajanja dela in stroškov. Nacionalna raven je tista, na kateri je te vrzeli sploh mogoče praktično nasloviti. Ne s tem, da bi Slovenija za vsako vprašanje umetne inteligence razvijala popolnoma samostojne standarde, temveč s tem, da zna obstoječe mednarodne in evropske dokumente pregledno prevzeti, pravilno uporabiti, po potrebi profilirati in dopolniti z dodatnimi tehničnimi usmeritvami, kadar se v domačem okolju pokaže potreba po večji konkretizaciji. Prav tu se standardizacijski ekosistem neposredno poveže z namenom tega poročila.

V večini primerov bo za slovenski prostor strokovno najprimernejša pot prevzem in dosledna uporaba mednarodnih ter evropskih standardov, ne razvoj velikega števila povsem novih nacionalnih standardov. Vendar to ne pomeni, da nacionalni sloj nima razvojne funkcije. Nasprotno: prav na tej ravni se lahko pokaže potreba po bolj konkretnih nacionalnih profilih, priporočilih, terminoloških poenotenjih ali tehničnih dopolnitvah, kadar obstoječi mednarodni in evropski dokumenti še ne dajejo dovolj neposrednih odgovorov za domače institucionalne in sektorske okoliščine. V nadaljnjih fazah standardizacijskega procesa se lahko pokaže potreba po tehničnih predlogih za nove oziroma izboljšane nacionalne standarde, vendar v smislu nadaljnje strokovne nadgradnje in prilagoditve, ne kot alternativa obstoječemu širšemu standardizacijskemu okviru.

Če nacionalni sloj povzamemo z vidika njegove funkcije v tej shemi, ga lahko razumemo kot raven, na kateri standardizacija umetne inteligence **postane dejansko operativna za slovenske organizacije in institucije**. Mednarodna raven daje temeljni jezik in horizontalne modele; evropska raven jih poveže z regulatornim in institucionalnim okvirom EU; nacionalna raven pa omogoči, da ta večslojni okvir postane uporaben v domačem strokovnem, upravljavskem in nadzornem okolju. Prav v tem smislu je nacionalni sloj ključen za nadaljnje razmisleke tega poročila: ne kot izhodišče standardizacije, temveč kot raven, na kateri se odloča o njeni dejanski uporabnosti za slovenski prostor.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

3.4 Sklepna funkcija sheme

Predstavljena večslojna shema standardizacijskega ekosistema omogoča predvsem dvoje. Prvič, pokaže, da standardizacija umetne inteligence ne nastaja na eni sami ravni in ne v okviru enega samega tipa dokumentov, temveč skozi preplet mednarodnih, evropskih in nacionalnih standardizacijskih procesov. Drugič, omogoča razlikovanje med dokumenti, ki vzpostavljajo temeljni horizontalni okvir, dokumenti, ki ta okvir prevajajo v evropski regulatorni in institucionalni kontekst, ter dokumenti oziroma pristopi, s katerimi se standardi vključujejo v nacionalno strokovno in organizacijsko prakso.

Na tej podlagi je mogoče preiti od shematskega prikaza standardizacijskega prostora k njegovi vsebinski obravnavi. Če je bilo v tem poglavju bistveno predvsem vprašanje, kako je standardizacijski ekosistem notranje strukturiran, je v nadaljevanju v ospredju vprašanje, **katere skupine standardov in referenčnih dokumentov so za slovenski prostor vsebinsko in operativno najpomembnejše**. Naslednje poglavje zato ne izhaja več zgolj iz ravni posameznih standardizacijskih organizacij, temveč iz funkcije, ki jo posamezni standardi opravljajo pri upravljanju, razvoju, uvajanju, presojanju in nadzoru sistemov umetne inteligence.

Tak prehod je metodološko pomemben tudi zato, ker omogoča, da se standardi v nadaljevanju ne obravnavajo kot nepovezan katalog naslovov, temveč kot medsebojno povezani gradniki širšega sistema skladnosti, kakovosti, varnosti, sledljivosti in zaupanja. Prav v tem smislu se shematski prikaz standardizacijskega ekosistema v naslednjem koraku prevede v podrobnejši pregled tistih vsebinskih sklopov in referenčnih dokumentov, ki imajo za slovenski prostor največjo praktično težo.

4. Vsebinska razvrstitev ključnih standardov in referenčnih dokumentov

Na podlagi predstavljene sheme standardizacijskega ekosistema je mogoče v nadaljevanju preiti od splošne strukturne predstavitve k vsebinski razvrstitvi tistih standardov in referenčnih dokumentov, ki imajo za slovenski prostor največjo praktično in strokovno težo. Če je bilo v prejšnjem poglavju bistveno predvsem vprašanje, na katerih ravneh standardizacijski dokumenti nastajajo in kako se umeščajo v mednarodni, evropski in nacionalni prostor, je v tem poglavju v ospredju njihova funkcionalna vloga pri razvoju, uvajanju, upravljanju, presojanju in nadzoru sistemov umetne inteligence. Tak pristop je metodološko skladen tudi z rezultatom 1.3, ki standarde ne obravnava le po standardizacijskih telesih, temveč jih filtrira po vsebinskih kriterijih, kot so upravljanje in

sistemi vodenja, upravljanje tveganj, kakovost podatkov, preglednost in razkritja, človeški nadzor, robustnost, kibernetska varnost ter dokumentacija in sledljivost.

Namen tega poglavja zato ni ponovno enakomerno povzeti celotnega standardizacijskega kataloga, temveč iz širšega nabora izluščiti tiste sklope dokumentov, ki tvorijo jedro operativno uporabnega standardizacijskega okvira. Rezultat 1.3 je pokazal, da se takšno jedro oblikuje na presečišču več komplementarnih virov: ISO/IEC in CEN/CENELEC postavljata hrbtenico upravljanja, tveganj, terminologije, arhitekture, kakovosti podatkov in človeškega nadzora; ETSI dodaja preverljive varnostne in robustnostne kontrole; IEEE operacionalizira transparentnost, zasebnost, pristranskost in druge družbeno pomembne vidike; ITU-T pa prispeva domensko natančne arhitekture, podatkovne prakse in izvedbene modele v omrežjih ter digitalnih storitvah. Prav zaradi te večplastnosti bo nadaljnji pregled organiziran po vsebinskih sklopih in praktični funkciji dokumentov, ne zgolj po formalni pripadnosti posameznemu standardizacijskemu telesu.

Za slovenski prostor je takšna razvrstitev posebej pomembna, ker omogoča prehod od splošne orientacije v standardizacijskem ekosistemu k bolj neposredno uporabnemu vprašanju, na katere dokumente se je smiselno opirati pri pripravi tehnične dokumentacije, internih pravil, kontrolnih seznamov, presoj skladnosti in nadaljnjih nacionalnih tehničnih dopolnitev. Tudi drugi projektni rezultati sledijo isti logiki: rezultat 1.1 povezuje regulativne zahteve z dokumentacijo, procesi in dokazili, rezultat 2.3 pa iz primerov dobrih praks izlušči prenosljive organizacijske, tehnične in upravljavske elemente, ki jih je mogoče prevesti v konkretne zahteve in operativne kontrolne sezname. V tem smislu je namen tega poglavja predvsem identificirati nosilne standardne sklope, ki bodo v nadaljevanju dokumenta služili kot strokovna osnova za razmislek o njihovi prednostni uporabi in o morebitnih potrebah po nadaljnji nacionalni tehnični artikulaciji.

4.1 Standardi za upravljanje in sisteme vodenja umetne inteligence

Funkcija sklopa

Prvi vsebinski sklop tvorijo standardi, ki vzpostavljajo organizacijski in upravljavski okvir za umetno inteligenco. Njihova funkcija ni v tem, da bi samostojno urejali posamezna tehnična vprašanja, kot so kakovost podatkov, robustnost modelov ali človeški nadzor, temveč v tem, da določijo **strukturo upravljanja**, znotraj katere je te zahteve mogoče dosledno, sledljivo in preverljivo izvajati. V tem smislu gre za dokumente, ki organizaciji pomagajo urediti politike, odgovornosti, postopke, presoje, nadzor nad spremembami, korektivne ukrepe in mehanizme stalnega izboljševanja. Prav zato imajo standardi za sisteme vodenja v celotnem standardizacijskem okviru UI vlogo organizacijske hrbtenice: brez njih tudi kakovostni tehnični ukrepi ostanejo parcialni in nepovezani. Rezultat 1.3 tak sklop upravičeno obravnava kot jedrni del horizontalnega ogrodja za skladno in varno rabo UI.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Ključni standardi in dokumenti

V tem sklopu ima osrednje mesto **ISO/IEC 42001:2023 – Artificial intelligence — Management system (AIMS)**, ki je v rezultatu 1.3 predstavljen kot temeljni standard upravljanja za organizacije, ki razvijajo, uvajajo ali uporabljajo rešitve umetne inteligence. Gre za mednarodni standard, namenjen širšemu krogu deležnikov: vodstvom organizacij, skrbnikom tveganj, funkcijam kakovosti, skladnosti in notranje revizije ter tudi oddelkom za varnost informacij in varstvo podatkov. Njegovo bistvo je v tem, da odgovarja na praktično vprašanje, kako zagotoviti, da je AI rešitev načrtovana, izvajana in nadzirana na ponovljiv, sledljiv in preverljiv način. Standard v tem smislu ne predpisuje konkretne tehnologije ali modelne arhitekture, temveč vzpostavlja ogrodje upravljanja, primerljivo z vlogo standardov ISO 9001 na področju kakovosti ali ISO/IEC 27001 na področju informacijske varnosti.

Rezultat 1.3 posebej poudari naslednje vsebinske elemente standarda ISO/IEC 42001:

- voditeljstvo in politika AI, vključno z določanjem namena, meja uporabe, načel, vlog in pristojnosti;
- načrtovanje in ocena tveganj, povezano z metodologijo iz ISO/IEC 23894;
- podpora in kompetence, vključno z znanji, usposabljanji, viri in komunikacijskimi kanali;
- operativno upravljanje AI, zlasti procesi razvoja, validacije, objave, spremljanja in umika rešitev;
- vrednotenje uspešnosti, notranje presoje, KPI, vodstveni pregledi, obravnava neskladnosti in CAPA;
- nenehne izboljšave po logiki cikla PDCA.

Zaradi takšne vsebine je ISO/IEC 42001 smiselno razumeti kot **organizacijsko vstopno točko** v celoten standardizacijski okvir UI. Ne deluje izolirano, temveč v povezavi z drugimi dokumenti istega sklopa: z ISO/IEC 23894 za upravljanje tveganj, z ISO/IEC 22989 za terminologijo, z ISO/IEC 23053 za arhitekturni okvir, z ISO/IEC 5259-1 in 5259-3 za kakovost podatkov ter z ISO/IEC TS 8200 za obvladljivost oziroma človeški nadzor. Rezultat 1.3 to medstandardno povezanost izrecno izpostavi kot operativno celoto, v kateri 42001 definira politike, odgovornosti, presoje in CAPA, drugi standardi pa posamezne upravljavske zahteve napolnijo s specifično metodologijo, artefakti in kontrolami.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker omogoča prehod od splošnih zahtev in razpršenih dobrih praks k **ponovljivemu sistemu organizacijskega upravljanja UI**. Organizacije, ki bodo morale izkazovati skladnost, ne potrebujejo le posameznih tehničnih dokumentov, temveč zlasti dokaz, da obstaja notranji sistem, v katerem so vloge razdeljene, postopki dokumentirani, odločitve sledljive, kontrole periodično preverjane, neskladnosti obravnavane in izboljšave sistematično vodene. Prav to raven vzpostavlja ISO/IEC 42001. Rezultat 1.3 pri tem posebej poudarja njegovo uporabnost za:

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

- pripravo in vzdrževanje politik AI,
- razdelitev odgovornosti med poslovnimi, tehničnimi in nadzornimi funkcijami,
- vzpostavitev notranjih presoj in vodstvenih pregledov,
- upravljanje sprememb modelov, podatkovnih sklopov in konfiguracij,
- pripravo dokumentiranih evidenc za potrebe presoj skladnosti in nadzora po dajanju sistema v uporabo.

Praktična vrednost standarda je dodatno razvidna iz primerov uporabe, ki jih navaja rezultat 1.3. Vodja kakovosti lahko na njegovi podlagi vzpostavi register AI procesov, imenuje odgovorne osebe ter uvede notranje presoje in CAPA. IT oziroma OT funkcija lahko uskladi politike dostopa, varnosti in sprememb modelov ter definira postopke za objavo nove verzije in rollback. Pravna oziroma skladnostna funkcija lahko na isti podlagi vključi zahteve AI Akta v sistem vodenja, pripravi standardizirano kazalo tehnične dokumentacije in pomaga oblikovati pregledne informacije za uporabnike. Vodstvo pa dobi okvir za periodično spremljanje KPI in odločanje o prioritetah nadaljnjih izboljšav. Takšne uporabe so za slovenski prostor posebej pomembne, ker pokažejo, da standard ni abstraktna upravljavska norma, temveč neposredno uporaben instrument za organizacijsko operativizacijo zahtev.

Dodatna praktična vrednost izhaja tudi iz dejstva, da je dokument dostopen prek SIST e-trgovine, rezultat 1.3 pa posebej opozarja, da imajo evropske oziroma nacionalne izdaje standardov za slovenske uporabnike praktično prednost pri sklicevanju v razpisih, pogodbah in nadzornih postopkih. Kjer evropska izdaja še ni sprejeta, ostaja mednarodna izdaja vseeno uporabna kot primarna strokovna referenca, zlasti če je vključena v širši sistem medsebojno povezanih standardov.

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je neposredna in večplastna. Rezultat 1.3 izrecno poudari, da ISO/IEC 42001 podpira vzpostavitev in upravljanje sistema vodenja za AI ter s tem neposredno prispeva k skladnosti in sledljivosti. Najtesneje se navezuje na obveznosti, ki jih rezultat 1.3 poveže z:

- členom 17 AI Akta, ki zadeva sistem vodenja kakovosti,
- členom 11, ki zadeva tehnično dokumentacijo,
- členom 9, ki zadeva upravljanje tveganj,
- členom 15, ki zadeva robustnost in kibernetsko varnost v povezavi z drugimi standardi.

Posebej za visokotvegane sisteme rezultat 1.3 poudari dokazljivost: kdo je sprejel odločitev, s katerimi podatki, pod katerimi pogoji in kako se sistem obnaša v mejnih primerih. ISO/IEC 42001 je pri tem pomemben zato, ker te elemente ne obravnava posamično, temveč jih procesno uredi: zahteva jasne odgovornosti, evidence, kontrolne točke pred objavo in v obdobju uporabe ter povezavo z nadzorom po dajanju na trg oziroma v uporabo. Prav zato

ga rezultat 1.3 utemeljeno opisuje kot “poslovno-organizacijsko lepilo”, ki omogoča, da tehnični ukrepi in pravne obveznosti delujejo usklajeno in dokazljivo.

Ta sklop je tesno povezan tudi z drugimi dosedanjimi rezultati projekta. Rezultat 1.1 je pravne zahteve AI Akta sistematično prevajal v procese, kontrolne točke, dokumentacijo in artefakte, kar pomeni, da je prav sistem vodenja tista raven, na kateri ti elementi postanejo organizacijsko izvršljivi. Rezultat 2.3 pa je na ravni dobrih praks pokazal, da se vprašanja governance, upravljanja sprememb, logiranja, nadzora, varnostnih postopkov in dokazljivosti v praksi pojavljajo kot ponovljivi operativni vzorci. Standardi za upravljanje in sisteme vodenja so zato naravna povezovalna točka med regulativnim okvirom, katalogom standardov in izvedbenimi dobrimi praksami. Rezultat 2.2 ni končna točka, temveč vmesna strokovna podlaga pred kasnejšimi tehničnimi prilagoditvami nacionalnih standardov in dokumentacijo tehničnih predlogov.

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop pomemben predvsem zato, ker ponuja **najbolj stabilno organizacijsko izhodišče** za vse nadaljnje dopolnitve, profile ali tehnične usmeritve. Če naj se v slovenskem prostoru v kasnejših fazah projekta pripravljajo tehnični predlogi, ti po svoji naravi ne bodo mogli uspešno delovati brez jasnega upravljavskega konteksta. Standardi za sisteme vodenja zato ne predstavljajo zgolj ene izmed vsebinskih tem, temveč raven, na katero se morajo navezovati tudi vsi nadaljnji sklopi, zlasti upravljanje tveganj, kakovost podatkov, dokumentacija, človeški nadzor in varnostne kontrole.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru poenotiti minimalni nabor upravljavskih artefaktov za organizacije, ki razvijajo ali uvajajo UI;
- kako strukturirati povezavo med AIMS, tehnično dokumentacijo in kontrolnimi seznami za potrebe presoj skladnosti;
- kako opredeliti pričakovane vloge, odgovornosti in pregledne točke za različne tipe organizacij;
- kako zagotoviti, da je sistem vodenja uporaben tudi za organizacije, ki niso velike ali visoko specializirane, vendar vseeno uporabljajo ali uvajajo sisteme UI.

Prav tu se pokaže razvojna vrednost tega sklopa: ne kot podlaga za podvajanje obstoječega mednarodnega standarda, temveč kot podlaga za razmislek, ali in kje bi slovenski prostor potreboval dodatne profile, pojasnila, minimalne organizacijske sheme ali druge podporne dokumente, ki bi obstoječi standard naredili še neposredneje uporaben v domačem institucionalnem in strokovnem okolju. To je hkrati tudi smer, ki je skladna s projektno logiko: rezultat 2.2 pripravi strokovno podlago, kasnejše faze pa jo nadgradijo v tehnične predloge in dokumentacijo za nadaljnje standardizacijske korake.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

4.2 Standardi za upravljanje tveganj in presojo vplivov

Funkcija sklopa

Drugi vsebinski sklop zajema standarde in sorodne referenčne podlage, katerih osrednja funkcija je metodološka ureditev prepoznavanja, vrednotenja, obravnave in spremljanja tveganj, povezanih z uporabo umetne inteligence. Če standardi za sisteme vodenja vzpostavljajo organizacijski okvir upravljanja, standardi za upravljanje tveganj to ogrodje napolnijo z operativno vsebino: določajo, kako se opredeli kontekst sistema, kako se identificirajo škodljivi scenariji, kako se ocenjujeta verjetnost in vpliv, kako se izberejo ukrepi za zmanjšanje tveganja ter kako se skozi življenjski cikel sistema vzdržujejo evidence, pregledi in ponovne ocene. Posebnost tega sklopa je v tem, da tveganj ne obravnava kot statičnega seznama nevarnosti, temveč kot iterativen proces, ki mora slediti spremembam podatkov, modelov, konteksta uporabe in učinkov na posameznike ali javni interes. Prav zato rezultat 1.3 standard ISO/IEC 23894 utemeljeno obravnava kot osrednjo metodološko referenco za člen 9 AI Akta.

Ključni standardi in dokumenti

V tem sklopu ima osrednje mesto **ISO/IEC 23894:2023 – Artificial intelligence — Risk management**, ki je v rezultatu 1.3 predstavljen kot mednarodni standard za celovito upravljanje tveganj pri sistemih umetne inteligence skozi celotni življenjski cikel. Dokument izhaja iz logike splošnih okvirov upravljanja tveganj, vendar jo prilagodi posebnostim UI, zlasti podatkovnim tveganjem, modelnim napakam, spremembam porazdelitev, adversarialnim posegom in vlogi človeškega nadzora. Njegova vrednost ni v abstraktni razpravi o tveganju, temveč v tem, da določa konkretne faze in artefakte: opredelitev konteksta, identifikacijo tveganj, analizo in vrednotenje, izbiro mitigacij in kontrol, spremljanje, pregled ter vzdrževanje sledljivih evidenc, kot so register tveganj, matrika odločitev in assumption log.

Za slovenski in širši evropski prostor je posebej pomembna tudi **EN ISO/IEC 23894:2024**, torej evropska izdaja istega standarda. Rezultat 1.3 izrecno poudarja, da EN izdaja ohranja isto jedrno vsebino kot izvirni ISO/IEC dokument, vendar omogoča sklicevanje na standard v obliki, ki je v evropskih postopkih javnega naročanja, nadzora in presoj skladnosti praviloma pričakovana ali vsaj izrazito praktičnejša. Posebej pomembna je njena institucionalna funkcija: ker se standard v slovenskem prostoru uporablja kot **SIST EN ISO/IEC 23894:2024**, postane bistveno lažje njegovo vključevanje v razpisne pogoje, notranje pravilnike, presoje ter komunikacijo z nadzornimi in morebitnimi priglašenimi organi. Rezultat 1.3 ob tem opozarja tudi na možnost prihodnje harmonizacije, pri kateri bi standard lahko podpiral domnevo skladnosti glede člena 9 AI Akta, če bi bil objavljen kot harmoniziran standard na ravni EU; te možnosti pa dokument ne predstavlja kot že vzpostavljenega dejstva, temveč kot potencialen nadaljnji razvoj.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Poleg samega standarda 23894 je za ta sklop pomembno tudi njegovo sistemsko umeščanje med druge dokumente. Rezultat 1.3 ga neposredno povezuje z ISO/IEC 42001, ker tveganja vstopajo v širši cikel sistema vodenja, z ISO/IEC 5259-1 in 5259-3, ker kakovost podatkov pomembno vpliva na tveganost rešitev, ter z ETSI TS 104 223, kjer se določena varnostna tveganja prevedejo v konkretne kontrolne zahteve. Na pravno-regulativni ravni pa rezultat 1.1 posebej poudari, da člen 9 AI Akta zahteva kontinuiran, iterativen in dokumentiran proces obvladovanja tveganj skozi celotni življenjski cikel sistema. V tem smislu ISO/IEC 23894 ne predstavlja vzporednega ali ločenega pristopa, temveč tehnično-operativni prevod tega zakonskega standarda v procesno izvedbo.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker organizacijam omogoča prehod od splošnega sklicevanja na "upravljanje tveganj" k preverljivemu in ponovljivemu procesu, ki ga je mogoče dejansko izvajati, dokumentirati in presojati. V praksi to pomeni, da organizacija ne ostane pri splošni izjavi, da tveganja spremlja, temveč mora biti sposobna pokazati, kako je določila namen sistema, katere škodljive scenarije je prepoznala, katere pragove sprejemljivosti je postavila, katere ukrepe je izbrala, kdo je bil za posamezne odločitve odgovoren in kako bo ob spremembah modela, podatkov ali konteksta uporabe izvedla ponovno oceno. Rezultat 1.3 prav v tem vidi ključno praktično vrednost standarda 23894: kot metodološko osnovo za strukturirano evidenčno delo, ki je razumljivo tudi ne-AI strokovnjakom, saj uporablja logiko, sorodno upravljanju tveganj na drugih reguliranih področjih.

Posebna uporabnost za slovenske deležnike izhaja tudi iz možnosti neposredne uporabe v različnih sektorjih. Rezultat 1.3 kot praktične primere navaja kreditno ocenjevanje v bančnem okolju, industrijski vizualni nadzor kakovosti in medicinsko oziroma zdravstveno triažo. Skupna značilnost teh primerov ni njihova sektorska podobnost, temveč dejstvo, da v vseh treh kontekstih standard omogoča enako strukturo dela: register tveganj, določitev meril sprejemljivosti, predhodne teste na kritičnih ali mejnih scenarijih, pravila za človeški pregled in ponovno ocenjevanje po večjih spremembah. Takšna prenosljivost je za slovenski prostor posebej dragocena, ker omogoča, da se enotna metodološka logika uporablja v javni upravi, finančnem sektorju, zdravstvu, industriji in drugih reguliranih okoljih, ne da bi bilo treba za vsak sektor znova izumljati osnovni proces upravljanja tveganj.

Na tej ravni je pomembna tudi evropska oziroma nacionalna izdaja standarda. Rezultat 1.3 izrecno poudarja, da so EN oziroma SIST EN izdaje za slovenske uporabnike praviloma praktično primernejše od zgolj mednarodnih različic, ker "živijo" v evropskem regulativnem in nadzornem okolju ter so lažje vključljive v javna naročila, presoje in dialog z nadzornimi organi. Pri sklicevanju v domačem okolju je zato smiselno poudariti prav **SIST EN ISO/IEC 23894:2024**, ne zato, ker bi ta vsebinsko spreminjal mednarodni standard, temveč zato, ker zmanjšuje operativno negotovost pri njegovi uporabi.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je neposredna in izrazito tesna. Rezultat 1.1 pri členu 9 izrecno poudari, da mora biti upravljanje tveganj pri visokotveganih sistemih kontinuiran, iterativen in dokumentiran proces skozi celoten življenjski cikel sistema. Rezultat 1.3 pa natančno pokaže, da ISO/IEC 23894 to zakonsko zahtevo operacionalizira v zaporedje korakov, vlog in evidenc. V tem pogledu gre za enega najbolj neposrednih primerov razmerja med regulativno zahtevo in tehnično-standardizacijsko podlago v celotnem dokumentu: AI Akt določa, da mora tveganje biti sistematično obvladovano, standard 23894 pa določa, kako naj bo tak proces zasnovan, dokumentiran in vzdrževan.

Ta sklop ni pomemben le za člen 9 v ožjem smislu, temveč tudi za druge zahteve AI Akta. Tveganja se v praksi ne ločujejo od podatkov, tehnične dokumentacije, človeškega nadzora ali robustnosti. Zato rezultat 1.3 standard 23894 umešča v neposredno povezavo z ISO/IEC 42001, z dokumenti o kakovosti podatkov iz serije 5259 ter z varnostnimi in robustnostnimi dokumenti ETSI. Enako logiko potrjuje tudi rezultat 1.1, kjer se med razširjenimi dokazili za visokotvegane sisteme posebej omenjajo register tveganj po ISO/IEC 23894, povezava z opisi arhitekture, podatkovnimi metrikami, robustnostnimi testi in ukrepi človeškega nadzora. To pomeni, da tveganostna presoja ni samostojen dokument, temveč povezovalni mehanizem, skozi katerega se usklajujejo druge sestavine skladnosti.

S projektnega vidika je ta sklop tesno povezan tudi z drugimi rezultati. Rezultat 1.1 je vzpostavil pravni prevod zahtev Akta o UI v procese, dokumentacijo in kontrolne točke. Rezultat 2.3 pa skozi primere dobrih praks pokaže, da se v operativnem okolju vprašanja tveganj praviloma ne pojavljajo ločeno, temveč v povezavi z governance, omejitvijo namena uporabe, nadzorom nad podatki, človeškim posegom in spremljanjem po uvedbi. Tako se potrjuje, da je upravljanje tveganj ena izmed osrednjih povezovalnih ravni med pravnim okvirom, tehničnimi standardi in dejanskimi implementacijskimi praksami.

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop pomemben predvsem zato, ker ponuja eno izmed najbolj neposredno prenosljivih podlag za prihodnje slovenske profile, smernice ali podporne tehnične dokumente. Mednarodni oziroma evropski standard že določa temeljno metodologijo, vendar se v praksi hitro odpre vprašanje, kako to metodologijo prevedemo v minimalne operativne zahteve za slovenske organizacije: kakšna naj bo osnovna struktura registra tveganj, kateri pragovi in kontrolne točke so pričakovani v posameznih tipih organizacij, kako naj bodo povezani z dokumentacijo za notranje presoje, kako naj se vodijo ponovne ocene ob spremembah in kako naj se povežejo z morebitnimi ocenami vplivov na temeljne pravice ali varstvo podatkov. Prav na tej ravni se pokaže prostor za nadaljnjo nacionalno tehnično konkretizacijo.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Za nadaljnjo obravnavo so posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru poenotiti minimalno strukturo registra tveganj za sisteme UI;
- kako povezati upravljanje tveganj z dokumentacijo po členu 11 AI Akta in z notranjimi kontrolnimi seznami;
- kako metodološko uskladiti tveganostno presojo z ocenami vplivov na temeljne pravice, varstvo podatkov in kibernetsko varnost;
- kako pripraviti poenostavljene, vendar še vedno strokovno ustrezne profile za organizacije, ki uporabljajo UI, vendar nimajo visoke notranje specializacije na področju standardizacije ali skladnosti.

Razvojna vrednost tega sklopa zato ni v tem, da bi slovenski prostor podvajal vsebino ISO/IEC 23894, temveč v tem, da na njegovi podlagi razvije bolj neposredno uporabne podporne instrumente za domače institucionalno, javnoupravno in gospodarsko okolje. Tak pristop je tudi v skladu s projektno logiko NOSI-AI, po kateri se v poznejših fazah predvidevajo tehnični predlogi za nacionalne standarde in pripadajoča dokumentacija.

4.3 Standardi za terminologijo, koncepte in arhitekturne okvire

Funkcija sklopa

Tretji vsebinski sklop zajema standarde in sorodne referenčne dokumente, katerih osrednja funkcija je vzpostavitev **enotnega pojmovnega okvira**, **konceptualne konsistentnosti** in **arhitekturne preglednosti** sistemov umetne inteligence. Če standardi za sisteme vodenja določajo organizacijski okvir, standardi za upravljanje tveganj pa proces presoje nevarnosti in omilitev, dokumenti iz tega sklopa skrbijo za nekaj še bolj temeljnega: da vsi deležniki sploh govorijo o istem sistemu, z istimi izrazi, z istim razumevanjem meja sistema, njegovih komponent, artefaktov, vhodov, izhodov in kontrolnih točk. Brez takšne osnove postanejo tudi sicer kakovostna dokumentacija, presoja skladnosti in komunikacija z regulatorji ali presojevalci notranje nekonsistentne.

V praksi ta sklop rešuje dva tipa problemov. Prvi je **terminološki problem**: različne ekipe, funkcije in deležniki pogosto uporabljajo iste izraze v različnih pomenih ali pa različne izraze za iste pojave. Drugi je **arhitekturni problem**: AI sistem se v praksi skoraj nikoli ne izčrpa v enem samem modelu, temveč vključuje podatkovne cevovode, učne in validacijske postopke, podporne storitve, registre, nadzorne mehanizme, vmesnike in operativne procese. Če te celote niso ustrezno razmejene in opisane, se hitro zabriše, kaj je dejanski predmet presoje, kje nastajajo tveganja, kdo je odgovoren za posamezen del sistema in katera dokazila je sploh treba pripraviti. Prav zato rezultat 1.3 standarda ISO/IEC 22989 in ISO/IEC 23053 utemeljeno obravnava kot konceptualno infrastrukturo celotnega standardizacijskega okvira.

Ključni standardi in dokumenti

V tem sklopu imata osrednje mesto dva dokumenta. Prvi je **ISO/IEC 22989:2022 – Artificial intelligence — Concepts and terminology**, ki vzpostavlja osnovne pojme in terminologijo za področje umetne inteligence. Rezultat 1.3 ga obravnava kot temeljni terminološki dokument, saj omogoča, da organizacija ključne izraze dosledno uporablja v tehnični dokumentaciji, navodilih za uporabnike, pogodbah, poročilih o testiranju in v komunikaciji z regulatorji. Posebej pomembno je, da standard ne vsebuje le slovarskih definicij, temveč tudi razmerja med pojmi in priporočila za njihovo dosledno uporabo, kar zmanjšuje prostor za napačne interpretacije pri pripravi dokumentacije in pri kasnejših presojah skladnosti. Rezultat 1.3 posebej poudari tudi praktično vrednost evropske izdaje **SIST EN ISO/IEC 22989:2023**, ki je za slovenski prostor primerna referenca za regionalno in institucionalno sklicevanje.

Drugi ključni dokument je **ISO/IEC 23053:2022 – Framework for AI systems using machine learning**, katerega funkcija je arhitekturna in sistemska. Rezultat 1.3 ga predstavlja kot okvir za komponente in življenjski cikel AI/ML sistemov ter ga neposredno povezuje s tehnično dokumentacijo in z razmejitvijo meja sistema pri presoji tveganj in skladnosti. Standard izhaja iz dejstva, da AI rešitve niso zgolj modeli, temveč vključujejo zbiranje in pripravo podatkov, učenje, validacijo, uvajanje, spremljanje, umik ter podporne artefakte in storitve. Njegova glavna vrednost je v tem, da takšno sestavljenost pretvori v primerljiv in razumljiv opis: definira komponente, artefakte, procese, vloge in kontrolne točke, s čimer postane možna jasna tehnična dokumentacija, sledljivost sprememb in pregledno določanje odgovornosti. Tudi tu rezultat 1.3 posebej poudarja pomen evropske izdaje **SIST EN ISO/IEC 23053:2023**, ki je za slovenski prostor posebej uporabna pri sklicevanju v dokumentaciji, presojah in tehničnih zahtevah.

Kot širši dopolnilni dokument tega sklopa je smiselno omeniti tudi **ISO/IEC TR 24028:2020 – Overview of trustworthiness in AI**. Čeprav po svojem normativnem statusu ne gre za standard, temveč za tehnično poročilo, rezultat 1.3 jasno pokaže, da ima pomembno razlagalno vlogo: sistematizira osnovne atribute zaupanja vredne UI, kot so robustnost, varnost, preglednost, zasebnost, pravičnost in odgovornost, ter s tem pomaga konceptualno povezati terminologijo, arhitekturo in kasnejše standarde za tveganja, podatke, nadzor in varnost. Njegova vloga je torej predvsem interpretativna in povezovalna, ne pa neposredno normativna.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker omogoča prehod od nejasnih, pogosto neenotnih opisov AI rešitev k **primerljivi, konsistentni in presojljivo uporabni tehnični dokumentaciji**. V domačem okolju se bodo z istimi sistemi soočali razvojne ekipe, kakovostne in skladnostne funkcije, pravne službe, vodstva organizacij, javni naročniki, regulatorji in morebitni organi ugotavljanja skladnosti. Če med njimi ni enotnega slovarja in

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

če ni jasno razvidno, kje se sistem začne in konča, kateri deli spadajo v modelni sloj, kateri v podatkovni sloj, kateri v operativno infrastrukturo in kateri v nadzorne mehanizme, postane praktična uporaba tudi sicer dobrih standardov bistveno težja.

Rezultat 1.3 to praktično vrednost pri ISO/IEC 22989 ponazori zelo jasno. Standard omogoča, da organizacija uvede enoten glosar v tehnični dokumentaciji, uporablja standardizirane definicije v pogodbah in SLA-jih, poenoti izraze v navodilih za uporabnike ter izobraževanjih in s tem zmanjša spore, nesporazume ter regulatorno nejasnost. V slovenskem prostoru je to posebej pomembno zato, ker se bo velik del nadaljnje uporabe UI odvijal v meddisciplinarnem okolju, kjer vsi deležniki niso specialisti za AI, vendar morajo kljub temu delati z istim naborom pojmov in razlikovati med izrazi, kot so natančnost, robustnost, incident, nadzor, sledljivost ali namen uporabe.

Praktična vrednost ISO/IEC 23053 pa je v tem, da organizacijam ponudi **predlogo arhitekture** za opis sistema. Rezultat 1.3 posebej poudari, da standard pomaga pri prikazu podatkovnih tokov, faz učenja in validacije, uvajanja, spremljanja in umika ter podpornih artefaktov, kot so datasheeti, evalvacijska poročila, konfiguracije in povezave med evidencami. V praksi to pomeni, da lahko organizacija bolj jasno pokaže, kje nastajajo podatki, kako so urejene različice modela, kdo odobri objavo, kje potekajo spremembe in kateri deli sistema zahtevajo dodatne kontrole. Tak okvir je v slovenskem prostoru posebej uporaben pri pripravi tehnične dokumentacije za javni sektor, pri notranjih presojah, pri komunikaciji z dobavitelji in pri vzpostavitvi minimalnih artefaktnih standardov, kakršne rezultat 1.1 že nakazuje z modelom “artefact passport”.

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je nekoliko manj neposredno “kontrolna” kot pri standardih za sisteme vodenja ali tveganja, vendar je vsebinsko izjemno pomembna. Rezultat 1.3 pri ISO/IEC 22989 izrecno poudari, da Akt o UI zahteva razumljive informacije za uporabnike po členu 13 ter strukturirano tehnično dokumentacijo po členu 11, pri čemer je enotna terminologija pogoj za oboje. Če so pojmi v dokumentaciji uporabljeni nekonsistentno, se poveča možnost napačnih interpretacij, neskladnosti in težav pri presoji. Standard za terminologijo zato ni zgolj “slovar”, temveč osnovni instrument za terminološko disciplino, brez katere tehnična dokumentacija ne more postati zanesljivo dokazilo.

Pri ISO/IEC 23053 je povezava z AI Aktom še tesneje vezana na tehnično dokumentacijo in sledljivost. Rezultat 1.3 izrecno navaja, da standard podpira opis meja sistema, podatkovnih tokov, komponent, nadzornih mehanizmov in kontrolnih točk, kar je ključno za zahteve po členu 11, posredno pa tudi za razumevanje člena 9 o tveganjih, člena 14 o človeškem nadzoru in člena 15 o robustnosti in varnosti. Če ni jasno, kako je sistem sestavljen, kateri artefakti sodijo vanj in kateri deli so kritični, je tudi upravljanje tveganj metodološko šibkejše, človeški nadzor slabše umeščen, varnostne zahteve pa težje dokazljive. Standard 23053 zato deluje kot “strukturna podlaga” za več različnih sklopov zahtev hkrati.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Ta sklop je tesno povezan tudi z drugimi rezultati projekta. Rezultat 1.1 je prav pri tehnični dokumentaciji, vodenju evidenc in prilogah z artefact passport obrazci pokazal, da bo za slovenski prostor ključna sposobnost jasnega opisovanja modelov, podatkovnih naborov, konfiguracij, odgovornosti, omejitev uporabe, metrik, robustnostnih testov in mehanizmov človeškega nadzora. Takšna dokumentacija pa ne more nastati brez dosledne terminologije in brez arhitekturnega okvira, ki povezuje vse te artefakte v razumljivo celoto. Rezultat 1.3 ta most že nakazuje, rezultat 2.2 pa ga zdaj smiselno prevaja v bolj sistematično standardizacijsko presojo.

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop posebej pomemben zato, ker ponuja **najprimernejšo osnovo za poenotenje jezika in minimalne strukture tehnične dokumentacije** v slovenskem prostoru. Če bo projekt v poznejših fazah pripravljal tehnične predloge, se bo zelo verjetno izkazalo, da slovensko okolje ne potrebuje novih temeljnih terminoloških ali arhitekturnih standardov, temveč predvsem bolj neposredno uporabne profile, predloge in priporočila, ki bodo obstoječa mednarodna in evropska izhodišča prevedli v domačo prakso.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru poenotiti rabo ključnih pojmov v tehnični dokumentaciji, pogodbah, razpisih in presojah;
- kako določiti minimalni nabor arhitekturnih prikazov, artefaktov in sledljivostnih povezav, ki jih je smiselno zahtevati pri različnih tipih sistemov UI;
- kako povezati terminološki in arhitekturni okvir z obrazci, kontrolnimi seznami in artefact passport pristopom, ki ga nakazuje rezultat 1.1;
- kako zagotoviti, da bo tehnična dokumentacija razumljiva ne le razvojni ekipi, temveč tudi pravnim, upravljavskim in nadzornim deležnikom.

Razvojna vrednost tega sklopa zato ni v podvajanju vsebine ISO/IEC 22989 ali ISO/IEC 23053, temveč v pripravi bolj uporabnih nacionalnih podpornih instrumentov: glosarjev, predlog za arhitekturne opise, minimalnih zahtev za artefakte, vzorčnih diagramov in povezav med dokumenti, ki bodo olajšali pripravo dokazljive, primerljive in institucionalno uporabne dokumentacije. Prav v tem se ta sklop neposredno navezuje na nadaljnje faze projekta, zlasti na poznejše tehnične prilagoditve nacionalnih standardov in na dokumentacijo tehničnih predlogov za nacionalni prostor.

4.4 Standardi za kakovost podatkov in podatkovno upravljanje

Funkcija sklopa

Četrti vsebinski sklop zajema standarde in sorodne referenčne dokumente, katerih osrednja funkcija je opredelitev **meril kakovosti podatkov**, **načinov njihovega preverjanja** ter **procesov upravljanja podatkov skozi življenjski cikel sistema umetne inteligence**. Če standardi za sisteme vodenja vzpostavljajo organizacijski okvir, standardi za upravljanje tveganj pa metodologijo obravnave nevarnosti, dokumenti iz tega sklopa odgovarjajo na vprašanje, kako zagotoviti, da so podatki, na katerih sistem temelji, dovolj kakovostni, ustrezni, sledljivi in upravljani na način, ki omogoča zanesljivo ter dokazljivo uporabo sistema.

Na področju umetne inteligence je ta sklop posebej občutljiv zato, ker kakovost modela ni odvisna zgolj od arhitekture ali optimizacijskih postopkov, temveč v veliki meri od kakovosti vhodnih podatkov, njihove reprezentativnosti, uravnoteženosti, označenosti, aktualnosti in sledljivosti sprememb. Rezultat 1.3 to jasno poudari pri standardih serije ISO/IEC 5259, kjer se kakovost podatkov ne obravnava kot abstraktno načelo, temveč kot nabor merljivih atributov in upravljavskih procesov, ki jih je mogoče vključiti v politiko, procese, evidence in presoje. Prav zato je ta sklop ključen most med zahtevami člena 10 AI Akta in dejansko operativno prakso podatkovnega dela.

Ključni standardi in dokumenti

V tem sklopu sta osrednji referenci **ISO/IEC 5259-1:2024 – Data quality for analytics and ML — Part 1** in **ISO/IEC 5259-3:2024 – Data quality for analytics and ML — Part 3: Management requirements**. Rezultat 1.3 ju obravnava komplementarno: prvi del določa, **kaj merimo**, tretji del pa, **kako to upravljamo v organizaciji**. Takšna delitev je metodološko pomembna, ker loči med merilnim in upravljavskim vidikom kakovosti podatkov. ISO/IEC 5259-1 vzpostavlja osnovno terminologijo, atribute kakovosti in metode ocenjevanja za podatkovne sklope za učenje, validacijo in test, medtem ko ISO/IEC 5259-3 določa procese, vloge, dokumente in nadzorne točke za sistematično obvladovanje kakovosti podatkov v AI projektih.

Rezultat 1.3 pri **ISO/IEC 5259-1** posebej izpostavi naslednje elemente:

- pojme in taksonomijo kakovosti podatkov, vključno z atributi, kot so popolnost, točnost, skladnost, pravočasnost, enoličnost in reprezentativnost;
- tipične vrste napak, kot so neuravnoteženost razredov, šum v oznakah, neustrezna porazdelitev in pomanjkanje redkih ali mejnih primerov;
- osnovne metode merjenja, na primer stopnjo manjkajočih vrednosti, razmerja razredov in stabilnost porazdelitev;

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

- povezavo s celotnim življenjskim ciklom podatkov, od zbiranja in priprave do pregledov pred učenjem in ponovnih ocen po pomembnih spremembah virov.

Pri **ISO/IEC 5259-3** pa rezultat 1.3 poudari predvsem procesni vidik:

- formalne politike kakovosti podatkov in odgovornosti, vključno z imenovanimi lastniki podatkov in odobritvami sprememb;
- načrtovanje in ocenjevanje kakovosti na ravni posameznega AI projekta, z določenimi KPI, pragovi in periodičnim poročanjem;
- operativne kontrole pred učnimi cikli, vključno s pregledi, validacijo označevanja in preverjanjem ground truth;
- sledljivost in dokumentacijo, z evidencami o izvoru podatkov, transformacijah, dovoljenjih in revizijskimi dnevniki sprememb datasetov;
- procese obvladovanja incidentov, kadar se odkrijejo napake v podatkih ali druge pomembne pomanjkljivosti.

Pomembno je tudi, da rezultat 1.3 ta sklop jasno umešča v širšo standardizacijsko celoto. Serija 5259 je neposredno povezana z ISO/IEC 42001, ker se kazalniki, presoje in CAPA ukrepi glede podatkov vključujejo v širši sistem vodenja, z ISO/IEC 23894, ker kakovost podatkov vpliva na profil tveganj sistema, ter z ISO/IEC 23053, ker so podatkovni artefakti, podatkovni tokovi in njihove spremembe sestavni del arhitekturnega opisa sistema. Ta medstandardna povezanost je za to poglavje pomembna zato, ker pokaže, da kakovost podatkov ni ločen "tehnični detajl", temveč samostojen, a hkrati prečno povezan sloj standardizacijskega okvira.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker omogoča prehod od splošnih zahtev o "ustreznih podatkih" k **merljivim kriterijem, dokumentiranim pregledom in ponovljivim upravljavskim procesom**. V praksi to pomeni, da organizacija ne ostane pri izjavi, da uporablja kakovostne podatke, temveč mora biti sposobna pokazati, kako to kakovost določa, meri, odobrava in spremlja skozi čas. Prav na tej ravni sta standarda 5259-1 in 5259-3 posebej uporabna: prvi zagotavlja standardni jezik in nabor metrik, drugi pa organizacijski proces, znotraj katerega te metrike postanejo del dejanske prakse.

Rezultat 1.3 praktično vrednost tega sklopa pokaže zelo konkretno. Pri **ISO/IEC 5259-1** navaja možnost uvedbe obveznih "data quality checklist" pregledov pred učenjem, pri katerih se preverjajo manjkajoče vrednosti, neuravnoteženost, anomalije in druge bistvene lastnosti podatkovnih sklopov, poročila o teh pregledih pa se hranijo kot evidence. Pri **ISO/IEC 5259-3** pa izpostavi razpisno dokumentacijo za dobavitelje podatkov, notranje programe "data

steward" vlog, redne preglede datasetov ter vključevanje poročil o kakovosti podatkov v širši post-market monitoring ali PMS okvir. To pomeni, da je ta sklop neposredno uporaben pri javnem naročanju, pri dobaviteljskih razmerjih, pri notranjih kontrolah ter pri pripravi dokazil za presoje skladnosti.

Posebna praktična vrednost se pokaže tudi v kontekstu slovenskih organizacij, ki niso nujno visoko specializirane za AI standardizacijo, vendar uporabljajo sisteme umetne inteligence v reguliranih ali občutljivih okoljih. Zanje je pomembno, da imajo na voljo dovolj jasen minimalni okvir: katere atribute podatkov je treba spremljati, katere evidence hraniti, katere prage določiti pred objavo modela in kako urediti odgovornosti med podatkovnimi ekipami, razvojem, skladnostjo in vodstvom. Serija 5259 omogoča prav to: standardizira tako merilni kot organizacijski minimum za odgovorno ravnanje s podatki v AI projektih.

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je neposredna in vsebinsko zelo močna. Rezultat 1.3 izrecno poudari, da **ISO/IEC 5259-1** daje merila in pojme za kakovost podatkov pri učenju, validaciji in testiranju, s čimer neposredno operacionalizira zahteve člena 10 AI Akta glede kakovosti, ustreznosti in dokumentiranosti podatkov. Standard omogoča, da pojmi iz člena 10 niso le načelni, temveč postanejo merljivi in vključljivi v politiko, postopke in tehnično dokumentacijo organizacije.

Pri **ISO/IEC 5259-3** je povezava še bolj procesna. Rezultat 1.3 jasno navaja, da standard z vidika člena 10 in 11 zagotavlja "organizacijski dokaz": ne le, da organizacija zna meriti kakovost podatkov, temveč da jo upravlja v ponovljivem procesu, z vlogami, KPI, notranjimi pregledi, CAPA ukrepi in revizijsko sledljivostjo sprememb. Prav zato je standard neposredno povezan z ISO/IEC 42001 in širšim sistemom vodenja AI. V tem smislu ta sklop ni zgolj tehnična dopolnitev člena 10, temveč eden ključnih operativnih mehanizmov za dokazovanje, da so podatkovne zahteve dejansko vključene v organizacijsko prakso.

Ta sklop je tesno povezan tudi z drugimi rezultati projekta. Rezultat 1.1 pri prevodu AI Akta v operativne zahteve poudarja pomen tehnične dokumentacije, evidenc, sledljivosti in preverljivih artefaktov, kar se pri podatkih neposredno odrazi v zahtevah glede izvora, transformacij, dovoljenj, kontrol pred učenjem in spremljanja sprememb. Rezultat 2.3 pa na ravni dobrih praks potrjuje, da so ravnanje z vhodnimi podatki, minimizacija, kakovost virov, omejitve rabe in nadzor nad spremembami med najbolj ponovljivimi varovalkami v dejanski uporabi sistemov UI. Skupaj to pomeni, da serija 5259 deluje kot standardizacijska osnova za enega najbolj praktično občutljivih delov celotnega AI življenjskega cikla.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop posebej pomemben zato, ker ponuja **zelo neposredno uporabno osnovo za pripravo slovenskih podpornih profilov, smernic in minimalnih evidenčnih zahtev** na področju podatkov za sisteme UI. Temeljna mednarodna standarda že določata, kaj je treba meriti in kako je treba to upravljati, vendar se v praksi hitro odpre vprašanje, kako te zahteve prevesti v konkretne obrazce, kontrole in minimalne organizacijske zahteve, ki bodo za slovenske deležnike dovolj jasne in uporabne.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru poenotiti minimalni nabor metrik in pregledov kakovosti podatkov pred učenjem, validacijo in objavo modela;
- kako oblikovati standardiziran nabor evidenc o izvoru podatkov, transformacijah, dovoljenjih in spremembah datasetov;
- kako povezati kakovost podatkov z dokumentacijo po členu 11 AI Akta ter z registri tveganj, notranjimi presojami in CAPA ukrepi;
- kako določiti minimalne zahteve za dobavitelje podatkov, zlasti v javnem sektorju in drugih reguliranih okoljih;
- kako pripraviti poenostavljene, a še vedno strokovno ustrezne profile za organizacije, ki nimajo lastnih velikih podatkovnih ekip, vendar uporabljajo ali naročajo sisteme UI.

Razvojna vrednost tega sklopa zato ni v podvajanju vsebine standardov 5259-1 in 5259-3, temveč v pripravi bolj neposredno uporabnih slovenskih instrumentov: kontrolnih seznamov za kakovost podatkov, predlog evidenc, minimalnih pogodbenih zahtev za dobavitelje, smernic za data steward vloge ter povezav med podatkovno dokumentacijo in širšim sistemom skladnosti. Prav na tej ravni se odpira prostor za nadaljnjo nacionalno tehnično konkretizacijo, ki je skladna z logiko projekta NOSI-AI in z njegovimi kasnejšimi fazami tehničnih predlogov.

4.5 Standardi za človeški nadzor in obvladljivost

Funkcija sklopa

Peti vsebinski sklop zajema standarde in sorodne referenčne dokumente, katerih osrednja funkcija je zagotoviti, da umetna inteligenca ostane **dejansko obvladljiva s strani človeka** tudi tedaj, ko deluje avtomatizirano, v realnem času ali v operativno občutljivem okolju. Če standardi za sisteme vodenja določajo organizacijski okvir, standardi za upravljanje tveganj pa metodologijo presoje nevarnosti, dokumenti iz tega sklopa odgovarjajo na vprašanje, kako

mora biti sistem zasnovan, opremljen, dokumentiran in upravljan, da lahko usposobljene osebe njegovo delovanje učinkovito spremljajo, prepoznajo odstopanja, zmanjšujejo tveganje pretiranega zaupanja v avtomatizacijo ter po potrebi ne uporabijo izhoda sistema, ga preglasijo ali varno zaustavijo.

Ta sklop je vsebinsko posebej občutljiv zato, ker človeški nadzor v praksi ni izpolnjen že s tem, da je človek formalno prisoten "v procesu". Učinkovit nadzor predpostavlja kombinacijo tehničnih, organizacijskih in vmesniških pogojev: človek mora razumeti, kaj sistem počne, kdaj je potrebna intervencija, kakšne so posledice posega, katere informacije so mu na voljo in ali organizacijsko okolje sploh omogoča, da svojo presojo dejansko uveljavi. Prav v tem smislu AI Akt v členu 14 zahteva ne le prisotnost človeka, temveč *effective human oversight*, ki je prilagojen tveganjem, stopnji avtonomije in kontekstu uporabe.

Rezultat 1.3 ta sklop upravičeno obravnava kot ločeno in samostojno plast standardizacijskega okvira. Posebej poudari, da obvladljivost ni isto kot preglednost: preglednost človeku posreduje informacije, obvladljivost pa pomeni, da lahko človek na podlagi teh informacij tudi pravočasno in učinkovito poseže v delovanje sistema. Prav zato je ta sklop metodološko bliže operativnemu oblikovanju varovalk, vmesnikov, alarmov, pragov in "safe-stop" mehanizmov kot pa zgolj pripravi pojasnilne dokumentacije.

Ključni standardi in dokumenti

V tem sklopu ima osrednje mesto **ISO/IEC TS 8200:2024 – Controllability of automated AI systems**, ki je po uradnem povzetku ISO/IEC tehnična specifikacija, namenjena določitvi osnovnega okvira s principi, značilnostmi in pristopi za uresničevanje in izboljševanje obvladljivosti avtomatiziranih AI sistemov. Z vidika normativnega statusa ne gre za mednarodni standard v ožjem pomenu, temveč za **Technical Specification**, kar je pomembno: dokument je usmerjen izrazito pragmatično in izvedbeno, zato deluje kot most med splošnimi pravnimi zahtevami po človeškem nadzoru in konkretnimi mehanizmi, s katerimi se ta nadzor tehnično ter organizacijsko zagotavlja.

Rezultat 1.3 pri TS 8200 posebej izpostavi naslednje ključne gradnike:

- **vloge in odgovornosti**, torej kdo sistem nadzira, kdo lahko odobri poseg in kakšne kompetence oziroma usposobljenost so za to potrebne;
- **načela zasnove nadzora**, vključno z razločevanjem med "human-in-the-loop" in "human-on-the-loop" glede na kritičnost naloge;
- **operativne mehanizme nadzora**, kot so ročni override, safe-stop, eskalacije in povezani postopki;
- **nadzorne podatke in alarmiranje**, torej katere informacije o zaupanju, negotovosti, stanju modela ali kritičnih odstopanjih mora sistem posredovati operaterju;

- **preizkuse obvladljivosti**, ki preverjajo, ali nadzorne funkcije delujejo tudi v dejanskih oziroma obremenjenih ali mejnih scenarijih.

Kot pomembna dopolnilna referenca tega sklopa nastopa tudi **IEEE 7001**, vendar z drugačno funkcijo. IEEE ga opredeljuje kot standard, ki določa merljive in preverljive ravni transparentnosti avtonomnih sistemov, da jih je mogoče objektivno ocenjevati in ugotavljati stopnjo skladnosti. To pomeni, da IEEE 7001 ni standard za obvladljivost v strogem smislu, je pa pomemben tam, kjer je učinkovit človeški nadzor odvisen od kakovosti informacij, ki jih sistem posreduje operaterju, revizorju ali drugemu nadzornemu akterju. V kontekstu tega poglavja zato IEEE 7001 deluje kot dopolnilni standard za **informacijsko podporo nadzoru**, ne pa kot osrednji standard za mehanizme posega.

Rezultat 1.1 poleg tega pravilno poveže ta sklop še z **EN ISO/IEC 23053**, saj je treba točke človeškega nadzora, override mehanizme, eskalacijske prage in varne zaustavitve pravilno umestiti že v arhitekturo sistema in v tehnično dokumentacijo. V tem pogledu je ta sklop tesno povezan z arhitekturnimi standardi, vendar od njih vsebinsko ločen: 23053 pomaga pokazati, *kje* v sistemu nadzor nastopi, TS 8200 pa pomaga določiti, *kako* mora tak nadzor delovati.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker omogoča prehod od formalnega sklicevanja na "človeški nadzor" k **dejansko preverljivim mehanizmom operativnega posega**. V praksi to pomeni, da organizacija ne ostane pri splošni trditvi, da je človek vključen v proces, temveč mora biti sposobna pokazati:

- kdo nadzor izvaja,
- kdaj mora intervenirati,
- katere informacije ima pri tem na voljo,
- kako lahko izhod sistema zavrne ali preglasi,
- kako se sistem varno zaustavi,
- ter kako se takšni mehanizmi testirajo, dokumentirajo in periodično preverjajo.

Rezultat 1.3 to praktično vrednost lepo ponazori z izrazom *obvladljivost*: bistvo ni le v tem, da je sistem "transparenten", temveč da je **vodljiv in zaustavljiv** v okoliščinah, kjer avtomatizirano delovanje lahko povzroči škodo ali nesorazmerno tveganje. Posebej pomembni so zato vmesniki, alarmni pragovi, ergonomija prikaza informacij, protokoli za eskalacijo ter redni preizkusi nadzornih scenarijev.

Rezultat 1.3 navaja tudi zelo uporabne praktične primere. V klicnem centru z AI asistenco lahko operater začasno izključi samodejne predloge, kadar zazna neprimernost, sistem pa beleži razloge za poseg in omogoča učenje na podlagi teh dogodkov. V industrijskem vizualnem nadzoru lahko sistem ob nenavadnih vzorcih sam predlaga zaustavitev linije, operater pa odločitev potrdi ali zavrne s pojasnilom. V javnem sektorju so pri avtomatiziranih odločitvah posebej pomembni mejni primeri, kjer je zagotovljen "human-in-the-loop" in

možnost naknadnega izpodbijanja oziroma pritožbe. Ti primeri so za slovenski prostor posebej pomembni zato, ker pokažejo, da je obvladljivost uporabna v zelo različnih okoljih: od upravnih postopkov in javnih storitev do industrije, zdravstva in drugih kritičnih ali občutljivih domen.

Dodatno praktično vrednost potrjuje tudi rezultat 2.3. Ta pri analiziranih dobrih praksah zelo jasno pokaže, da učinkovit človeški nadzor ni zgolj procesna kljukica, temveč kombinacija razumljive predstavitve izhodov, informacij, ki omogočajo strokovno presojo, usposabljanja, organizacijskih komunikacijskih kanalov in realne možnosti, da uporabnik predlogu sistema ne sledi brez negativnih organizacijskih posledic. Posebej pomemben je poudarek na zmanjševanju *automation bias*: če organizacija ne vzpostavi usposabljanja, povratnih kanalov in kulture, v kateri je preglasitev sistema legitimna, je tudi formalno prisoten človek lahko v praksi le pasivni potrjevalec avtomatiziranega izhoda.

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je neposredna. Člen 14 določa, da morajo biti visokotvegani AI sistemi zasnovani in razviti tako, da jih lahko med uporabo učinkovito nadzirajo fizične osebe, vključno z ustreznimi orodji človek–stroj, pri čemer morajo biti ukrepi nadzora sorazmerni tveganjem, ravni avtonomije in kontekstu uporabe. Člen nadalje predvideva, da je namen človeškega nadzora preprečevati ali zmanjševati tveganja za zdravje, varnost ali temeljne pravice, pri čemer morajo nadzorne osebe razumeti zmožnosti in omejitve sistema, pravilno interpretirati njegove izhode in po potrebi odločiti, da izhoda ne uporabijo, ga ignorirajo, razveljavijo ali ustavijo delovanje sistema.

Rezultat 1.1 to logiko pravilno prevaja v operativne zahteve: zmanjšanje automation bias, vloge in odgovornosti, usposabljanja, override/safe-stop “drills”, merjenje uspešnosti posegov, jasno prikazane negotovosti in opozorila ter dokumentirana pravila eskalacij. Posebej dragoceno je, da rezultat 1.1 ta sklop poveže s konkretnimi standardnimi referencami: ISO/IEC TS 8200 za obvladljivost, IEEE 7001 za informacije operaterjem in revizorjem ter EN ISO/IEC 23053 za označitev točk človeškega nadzora v arhitekturi. Tako nastane zelo koherenten most med členom 14 AI Akta in standardizacijskim okvirom, ki ga projekt postopoma gradi skozi rezultate 1.1, 1.3 in 2.3.

Rezultat 1.3 pa ta most še dodatno konkretizira, ko TS 8200 opiše kot neposredni “priročnik” za implementacijo člena 14: pomaga določiti primerne nadzorne funkcije, način njihove zasnove, testiranja, dokumentiranja ter vključevanja v sistem vodenja kakovosti in post-market spremljanje. Ta poudarek je metodološko pomemben, ker pokaže, da je človeški nadzor nekaj, kar je treba ne le načrtovati, temveč tudi **dokazljivo testirati in periodično vaditi**.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop posebej pomemben zato, ker odpira prostor za zelo praktične in za slovenski prostor neposredno uporabne podporne instrumente. Mednarodna tehnična specifikacija že določa osnovni okvir obvladljivosti, vendar se v praksi hitro pokaže potreba po dodatni konkretizaciji:

- kako naj bodo v slovenskem prostoru oblikovani minimalni override in safe-stop postopki;
- kako naj se določajo pragovi za obvezni človeški pregled ali zaustavitev;
- kakšne informacije morajo biti prikazane operaterju v različnih tipih sistemov;
- kako naj se dokumentirajo preizkusi obvladljivosti, usposabljanja in odgovornosti;
- ter kako naj se preprečuje automation bias v dejanskih delovnih procesih.

Rezultat 1.3 tu že nakaže nekaj posebej relevantnih smeri, na primer merila za zaustavitev, vaje odziva z merjenjem MTTR in override success rate ter dokumentacijo za presojo, ki vključuje opis vmesnikov, dokazila o testih, usposabljanjih in matriko odgovornosti. Prav ti elementi so dobra podlaga za nadaljnje nacionalne tehnične profile, smernice ali minimalne zahteve za dokumentacijo človeškega nadzora.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru poenotiti minimalne zahteve za override, safe-stop in eskalacijske postopke;
- kako standardizirati dokazila o testiranju obvladljivosti in o usposobljenosti oseb, ki izvajajo nadzor;
- kako v tehnični dokumentaciji enotno označiti točke človeškega nadzora in meje dopustne avtonomije sistema;
- kako povezati organizacijske ukrepe proti automation bias s tehničnimi in vmesniškimi zahtevami;
- kako pripraviti posebne profile za javni sektor in druge regulirane domene, kjer mora biti človeški nadzor posebej skrbno dokumentiran.

Razvojna vrednost tega sklopa zato ni v podvajanju vsebine TS 8200 ali IEEE 7001, temveč v pripravi bolj neposredno uporabnih nacionalnih podpornih instrumentov: vzorčnih eskalacijskih protokolov, minimalnih seznamov dokazil, smernic za zasnovo operaterskih vmesnikov, predlog usposabljanj in vaj ter povezav med členom 14 AI Akta in konkretnimi tehničnimi artefakti. Prav v tem smislu je ta sklop eden izmed najbolj operativno obetavnih za kasnejše nacionalne tehnične predloge.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

4.6 Standardi za robustnost, kibernetsko varnost in odpornost sistemov

Funkcija sklopa

Šesti vsebinski sklop zajema standarde in tehnične specifikacije, katerih osrednja funkcija je zagotoviti, da sistem umetne inteligence ostane **tehnično zanesljiv, odporen in varovan** tudi v okoliščinah napak, neobičajnih vhodov, zlonamernih posegov, sprememb okolja, degradacije modela ali kompromitacije podatkovnih in infrastrukturnih komponent. Če standardi za upravljanje določajo organizacijsko strukturo, standardi za tveganja metodologijo presoje, standardi za človeški nadzor pa pogoje učinkovitega posega človeka, dokumenti iz tega sklopa odgovarjajo na vprašanje, kako mora biti AI sistem tehnično zasnovan, zaščiten, testiran in spremljan, da njegovo delovanje ostaja dovolj robustno in varno skozi celotni življenjski cikel.

Ta sklop je metodološko pomemben zato, ker pri sistemih UI varnost in robustnost nista zgolj nadaljevanje klasične informacijske varnosti. ETSI pri TS 104 223 izrecno poudari, da AI sistemi predstavljajo posebno varnostno kategorijo, ker vključujejo tveganja, kot so data poisoning, prompt injection, obfuscation, zlorabe podatkovnega upravljanja in druge ranljivosti, povezane s kompleksnim življenjskim ciklom modelov ter podatkov. To pomeni, da robustnost in kibernetska varnost v AI okolju ne pomenita le zaščite programske opreme ali omrežja, temveč tudi zaščito modela, učnih artefaktov, podatkovnih tokov, konfiguracij, integritete izhodov in infrastrukturnih odvisnosti.

V tem smislu ta sklop pokriva tri medsebojno povezane ravni:

- **robustnost delovanja**, torej odpornost proti napakam, nestabilnim vhodom, distribucijskim premikom in degradaciji zanesljivosti;
- **kibernetsko varnost sistema**, torej zaščito modelov, podatkov, API-jev, cevovodov, dnevnikov, integracij in dobavne verige;
- **odpornost operativnega okolja**, torej sposobnost zaznave incidentov, varne zaustavitve, obnove, rollbacka in ponovne vzpostavitve nadzora po varnostnem dogodku.

Prav zato rezultat 1.3 ta sklop obravnava kot nujno tehnično dopolnitev organizacijskih in tveganostnih standardov, ne kot njihov podsklop.

Ključni standardi in dokumenti

V tem sklopu ima osrednje mesto **ETSI TS 104 223 V1.1.1 – Securing Artificial Intelligence (SAI); Baseline Cyber Security Requirements for AI Models and Systems**. ETSI ga opredeljuje kot tehnično specifikacijo, ki določa osnovne varnostne zahteve za AI modele in sisteme, vključno z generativno UI, ter je namenjena deležnikom po celotni AI dobavni verigi. Njegova posebna vrednost je v tem, da ne ostaja pri splošni zahtevi po varnosti, temveč uvaja strukturiran nabor osnovnih zahtev, prilagojenih posebnostim AI sistemov. Rezultat 1.3 ga

zato upravičeno obravnava kot enega najpomembnejših tehničnih dokumentov za varnostno operacionalizacijo sistemov UI.

Rezultat 1.3 pri ETSI TS 104 223 posebej poudari več vsebinskih sklopov, ki so za nadaljnji dokument posebej relevantni:

- zahteve glede **upravljanja modelov in artefaktov**, vključno z integriteto modelnih datotek, verzioniranjem, dokazovanjem izvora in varnim uvajanjem sprememb;
- zahteve glede **podatkovne varnosti**, vključno z zaščito učnih in validacijskih podatkov, spremljanjem zastrupljanja podatkov in nadzorom nad transformacijami;
- zahteve glede **infrastrukturne varnosti**, vključno z zaščito API-jev, dnevnikov, skrivnosti, dostopov in povezav z drugimi sistemi;
- zahteve glede **varnostnega testiranja in spremljanja**, vključno s preskusi odpornosti, validacijo kontrol, incidentnimi postopki, rollbackom in varnim delovanjem ob degradaciji.

Druga ključna referenca tega sklopa je **ETSI TS 104 050 V1.1.1 – Securing Artificial Intelligence (SAI); AI Threat Ontology and definitions**. ETSI v uradnem besedilu navaja, da dokument definira, kaj je AI grožnja, in kako jo je mogoče razlikovati od ne-AI groženj, pri čemer to stori v obliki ontologije groženj. Njegova funkcija ni v neposrednem določanju varnostnih kontrol, temveč v **strukturiranju besedišča in razmerij med grožnjami, akterji, sredstvi in scenariji napada**. Zaradi tega deluje kot konceptualni in analitični temelj za kasnejše preskušanje, ocene tveganj in oblikovanje varnostnih zahtev. Ta dokument je zato metodološko bliže arhitekturnim in terminološkim dokumentom, vendar je v varnostnem sklopu posebej pomemben zato, ker omogoča bolj dosledno klasifikacijo AI-specifičnih groženj.

Kot širši institucionalni okvir tega sklopa je smiselno omeniti tudi **ETSI OCG AI** in **ETSI TC SAI**. ETSI na svoji uradni strani pojasnjuje, da OCG AI koordinira standardizacijske aktivnosti, povezane z UI, ki se izvajajo v različnih tehničnih telesih, medtem ko je cilj TC SAI razvijati tehnične specifikacije, ki naslavljajo grožnje, ki izhajajo iz uporabe UI, ter grožnje AI sistemom samim, bodisi iz drugih AI sistemov bodisi iz konvencionalnih virov. To je za ta dokument pomembno zato, ker potrjuje, da ETSI varovanja UI ne obravnava obrobno, temveč kot samostojen standardizacijski tok z jasno institucionalno podporo.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker omogoča prehod od splošne zahteve, da mora biti AI sistem "varen in robusten", k **preverljivemu naboru tehničnih kontrol, artefaktov in operativnih postopkov**. V praksi to pomeni, da organizacija ne

ostane pri trditvi, da model deluje stabilno ali da je infrastruktura zaščitena, temveč mora biti sposobna pokazati:

- kako varuje modele, podatke in konfiguracije pred nepooblaščenimi spremembami;
- kako preverja odpornost proti znanim grožnjam in mejnim scenarijem;
- kako vodi evidence o verzijah, incidentih, testih in posodobitvah;
- kako obravnava degradacijo ali kompromitacijo modela;
- ter kako zagotavlja varno nadaljevanje ali varno ustavitev delovanja sistema po incidentu.

Rezultat 1.3 zelo jasno pokaže, da je ETSI TS 104 223 v tem pogledu posebej uporaben prav zato, ker uvaja "baseline security requirements", torej varnostni minimum, ki ga je mogoče uporabiti kot referenco v različnih tipih organizacij. To je za slovenski prostor pomembno zlasti v primerih, ko organizacije uvajajo UI v kritičnejša ali regulirana okolja, vendar nimajo lastnih zelo zrelih AI security okvirjev. Standard jim omogoča, da se naslonijo na strukturiran nabor zahtev, namesto da bi morale same graditi varnostne zahteve od začetka.

Praktična vrednost se pokaže tudi pri povezavi z dobavno verigo in javnimi naročili. Kadar organizacija kupuje ali naroča AI rešitev, je posebej pomembno, da lahko od dobavitelja zahteva minimalne varnostne lastnosti: sledljivost modelnih različic, zaščito podatkov, dokumentirane postopke za obvladovanje ranljivosti, varnostne preglede posodobitev in zmogljivost za rollback ali izolacijo sistema ob incidentu. Prav takšne zahteve je lažje oblikovati, če se lahko oprejo na ETSI-jevo tehnično specifikacijo kot na zunanjo referenco. To je v slovenskem prostoru posebej relevantno za javni sektor, kritično infrastrukturo, zdravstvo, finance in druge domene, kjer bo moral biti varnostni nivo AI rešitev razmeroma hitro preveden v tehnično in pogodbeno prakso.

Dodatna praktična vrednost izhaja tudi iz dejstva, da ETSI SAI dokumenti upoštevajo posebnosti sodobnih AI sistemov, vključno z generativno UI. ETSI pri TS 104 223 izrecno navaja, da specifikacija zajema tudi sisteme, ki vključujejo globoke nevronske mreže in generativno UI. To je za slovenski prostor pomembno zato, ker se bo del uvajanja UI odvijal prav v okoljih, kjer tradicionalni varnostni standardi ne zadostujejo za prompt injection, model extraction, poisoning, nepričakovane izhode ali druge AI-specifične oblike zlorab.

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je neposredna predvsem prek **člena 15**, ki zahteva, da so visokotvegani AI sistemi zasnovani in razviti tako, da dosegajo ustrezno raven natančnosti, robustnosti in kibernetske varnosti ter da ostajajo odporni proti napakam, nedoslednostim, nepooblaščenim spremembam in poskusom izkoriščanja ranljivosti, ob upoštevanju namena in konteksta uporabe. V tem pogledu ETSI TS 104 223 deluje kot zelo

pomembna tehnična podlaga za operacionalizacijo te zahteve, saj jo prevede v konkretne varnostne zahteve za modele, podatke, infrastrukturo in operativne procese.

Povezava z drugimi rezultati projekta je prav tako močna. Rezultat 1.1 pri operativnem prevodu AI Akta že poudari pomen varnostnih testov, sledljivosti uvozov in posodobitev, minimalnih kontrol za dobavno verigo, SBoM oziroma sorodnih evidenc ter varnostnih postopkov za safe-stop in rollback. Ta logika je z ETSI TS 104 223 v veliki meri usklajena, saj specifikacija prav na teh ravneh določa varnostne baseline zahteve. Rezultat 1.3 ta most še dodatno utrdi, ko ETSI dokumente umešča kot tehnično dopolnitev ISO/IEC standardom za upravljanje, tveganja, podatke in arhitekturo: organizacijski okvir brez varnostnih kontrol namreč ne zagotavlja zadostne odpornosti sistema.

Ta sklop je smiselno povezati tudi z rezultatom 2.3. Dobri praksi, obravnavani v projektu, kažeta, da se v dejanskem operativnem okolju ponavljajo potreba po omejitvah dostopa, spremljanju izhodov, varnem ravnanju z vhodnimi podatki, logiranju, obvladovanju sprememb, odzivnih postopkih in možnostih varne ustavitve. Čeprav rezultat 2.3 teh elementov ne obravnava primarno kot "cyber security" standarde, potrjuje njihovo operativno nujnost. Ta sklop zato deluje kot tehnični standardizacijski odgovor na varnostne in robustnostne potrebe, ki jih projekt prepoznava tudi prek primerov uporabe.

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop posebej pomemben zato, ker ponuja eno izmed najbolj uporabnih podlag za pripravo **minimalnih varnostnih profilov, kontrolnih seznamov in dokazilnih zahtev** za AI sisteme v slovenskem prostoru. ETSI-jevi dokumenti že določajo osrednje varnostne gradnike, vendar se v praksi hitro pokaže potreba po dodatni konkretizaciji:

- katere baseline varnostne kontrole naj se štejejo za minimalno pričakovane pri različnih tipih AI sistemov;
- kako naj bodo dokumentirani testi robustnosti, varnostni pregledi modelov, zaščita podatkov in upravljanje ranljivosti;
- kako naj se varnostne zahteve vključijo v razpise, pogodbe in presoje skladnosti;
- ter kako naj se uskladijo z organizacijskim okvirom, registrom tveganj in tehnično dokumentacijo.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru poenotiti minimalne baseline zahteve za zaščito modelov, podatkovnih sklopov, API-jev in AI infrastrukture;
- kako standardizirati dokazila o varnostnih testih, robustnostnih preizkusih, obravnavi incidentov in rollback postopkih;
- kako povezati AI-specifične varnostne kontrole z obstoječimi okviri informacijske varnosti, ne da bi se izgubile posebnosti UI;

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

- kako pripraviti profile za javni sektor in druge regulirane domene, kjer mora biti varnost AI rešitev posebej pregledna in dokazljiva;
- kako v razpisni in pogodbeni praksi oblikovati minimalne zahteve za dobavitelje AI rešitev in njihove podizvajalce.

Razvojna vrednost tega sklopa zato ni v podvajanju vsebine ETSI TS 104 223 ali TS 104 050, temveč v pripravi bolj neposredno uporabnih slovenskih podpornih instrumentov: vzorčnih baseline varnostnih seznamov, predlog dokazil, profilov za naročanje in presojo, smernic za beleženje incidentov ter povezav med členom 15 AI Akta in konkretnimi tehničnimi artefakti. Prav v tem smislu je ta sklop eden izmed najbolj perspektivnih za kasnejše nacionalne tehnične predloge, saj omogoča, da se robustnost in kibernetska varnost sistemov UI v slovenskem prostoru obravnavata ne le deklarativno, temveč operativno in dokazljivo.

Če sklop robustnosti, kibernetske varnosti in odpornosti naslavlja predvsem vprašanje, kako sistem UI zaščititi pred tehničnimi napakami, zlorabami in kompromitacijo, je naslednji sklop usmerjen v drugo, vendar enako pomembno razsežnost: kako zagotoviti, da je tak sistem tudi razumljiv, zasebnostno ustrezno zasnovan, procesno občutljiv za tveganja pristranskosti ter etično sledljiv v svojih razvojnih in implementacijskih odločitvah

4.7 Standardi in dokumenti za transparentnost, zasebnost, pristranskost in odgovorno zasnovo

Funkcija sklopa

Sedmi vsebinski sklop zajema standarde in referenčne dokumente, katerih osrednja funkcija ni upravljanje sistema kot celote, ne obvladovanje tveganj v ožjem smislu, ne operativni človeški nadzor in ne kibernetska zaščita, temveč **normativno in procesno strukturiranje tistih lastnosti sistema umetne inteligence, ki odločilno vplivajo na njegovo sprejemljivost, razumljivost in legitimnost v uporabi**. V ta okvir sodijo predvsem transparentnost, zasebnost, obravnava algoritemske pristranskosti in vgradnja etičnih zahtev v samo zasnovo sistema.

Ta sklop je metodološko pomemben zato, ker se številne pravno relevantne lastnosti sistema UI ne izčrpajo v vprašanju, ali sistem deluje tehnično pravilno. Tudi tehnično zelo stabilen sistem je lahko pravno ali organizacijsko problematičen, če uporabniku ne omogoča zadostnega razumevanja njegove funkcije, če ne vključuje strukturiranega ravnanja s osebnimi podatki, če ne vsebuje postopkov za zaznavanje in zmanjševanje neželenih pristranskosti ali če etične zahteve ostanejo zgolj deklarativne in niso prevedene v razvojne artefakte, odločitve in kontrole. Prav tu je posebna vrednost IEEE-jeve serije 7000, ki skuša ta vprašanja prevesti iz ravni splošnih načel v ravnanja, procese, dokazila in preverljive zahteve.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

V tem smislu ta sklop pokriva štiri medsebojno povezane ravni:

- **transparentnost sistema**, torej strukturirano razkritje delovanja, omejitev, namenov in ravni skladnosti;
- **zasebnostno zasnovo**, torej vključitev privacy-oriented zahtev v razvoj in uvajanje sistemov, ki uporabljajo osebne podatke;
- **obravnavo algoritemske pristranskosti**, torej metodološko in procesno ravnanje z neželenimi odstopanji, ki lahko vodijo v sistemske neenakosti ali nepoštene učinke;
- **etično zasnovo sistema**, torej sledljivo vključitev vrednot, etičnih zahtev in njihovih kompromisov v celotni razvojni proces.

Ključni standardi in dokumenti

V tem sklopu ima temeljno mesto **IEEE 7000-2021 – Model Process for Addressing Ethical Concerns During System Design**. IEEE na uradni strani pojasnjuje, da standard opisuje procese, ki omogočajo **sledljivost etičnih vrednot** v konceptu delovanja, etičnih zahtevah in zasnovi sistema, pri čemer je namenjen organizacijam različnih velikosti in ni vezan na en sam razvojni model. Posebna vrednost tega dokumenta je v tem, da etičnih vprašanj ne obravnava kot zunanjega pregleda po koncu razvoja, temveč kot element sistemskega inženirstva že v fazi konceptualizacije, zahtev in oblikovalskih odločitev.

Drugi ključni dokument je **IEEE 7001-2021 – Transparency of Autonomous Systems**. IEEE izrecno navaja, da standard opisuje **merljive in preverljive ravni transparentnosti**, tako da je mogoče avtonomne sisteme objektivno ocenjevati in ugotavljati ravni skladnosti. To je pomembno zato, ker transparentnost v tem okviru ni razumljena le kot splošna "razložljivost", temveč kot strukturiran nabor informacijskih ravni, ki jih je mogoče načrtovati, preskušati in dokazovati. V primerjavi s poglavjem 4.5 je zato poudarek drugačen: tam je bilo vprašanje, ali lahko človek sistem učinkovito nadzira in po potrebi ustavi; tukaj pa je vprašanje, katere informacije morajo biti o sistemu razkrite, v kakšni obliki in na kakšni ravni preverljivosti.

Tretji pomembni dokument je **IEEE 7002-2022 – Data Privacy Process**. IEEE ga opredeljuje kot standard, ki določa zahteve za sistemsko oziroma programsko-inženirski proces za **privacy-oriented considerations** pri proizvodih, storitvah in sistemih, ki uporabljajo osebne podatke zaposlenih, strank ali drugih zunanjih uporabnikov. Njegova funkcija torej ni v ponovitvi pravil varstva osebnih podatkov iz GDPR, temveč v procesnem prevodu teh vprašanj v razvoj, uvedbo in upravljanje sistema. To ga naredi posebej uporabnega v AI okolju, kjer je treba zasebnost obravnavati kot arhitekturno in procesno zahtevo, ne šele kot naknadni pravni pregled.

Četrti osrednji dokument je **IEEE 7003-2024 – Algorithmic Bias Considerations**. IEEE na uradni strani navaja, da standard opisuje procese in metodologije, ki pomagajo uporabnikom obravnavati vprašanja pristranskosti pri ustvarjanju algoritmov. Njegov namen zato ni

obljuba popolne odsotnosti bias, temveč vzpostavitev strukturiranega okvira za njegovo prepoznavanje, presojo, dokumentiranje in zmanjševanje. Prav v tem je njegova posebna uporabnost: omogoča, da je obravnava pristranskosti vključena v metodologijo razvoja in presoje, namesto da ostane zgolj intuitivna ali ad hoc.

Kot širši institucionalni okvir je pomembno omeniti tudi, da IEEE SA ta sklop umešča v širšo pobudo za standarde na področju **Autonomous and Intelligent Systems**, kjer poudarja razvoj globalno prepoznavnih standardov in virov za pragmatično uporabo etičnih in sistemsko-inženirskih načel pri AIS. To potrjuje, da transparentnost, zasebnost, bias in etična zasnova niso obravnavani kot obrobne teme, temveč kot samostojen standardizacijski tok.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker organizacijam omogoča prehod od splošnih zahtev po "odgovorni" ali "zaupanja vredni" UI k **bolj strukturiranim procesnim in dokazilnim oblikam ravnanja**. V praksi namreč pogosto ni največji problem to, da organizacija ne pozna splošnih načel transparentnosti, zasebnosti ali nepristranskosti, temveč to, da teh načel ne zna dosledno prevesti v razvojne postopke, vloge, dokumentacijo, preskuse in operativna pojasnila za uporabnike. IEEE-jevi dokumenti so v tem smislu koristni prav zato, ker vzpostavljajo most med abstraktnim načelom in izvedbenim procesom.

Praktična vrednost se pokaže zlasti v štirih vrstah situacij.

Prvič, kadar mora organizacija pripraviti **uporabniško razumljiva pojasnila** o omejitvah, namenih in načinu uporabe AI sistema.

Drugič, kadar želi v razvojne procese vključiti **privacy-by-design** logiko na način, ki je za tehnične ekipe operativen.

Tretjič, kadar potrebuje strukturiran okvir za **analizo in zmanjševanje pristranskosti** v podatkih, modelnih odločitvah ali učinkih na različne skupine uporabnikov.

Četrtič, kadar želi etične zahteve vključiti že v fazo zasnove sistema, tako da ostanejo sledljive tudi v kasnejših tehničnih in organizacijskih odločitvah.

Za slovenske organizacije je to posebej relevantno v javnem sektorju, zdravstvu, financah, zavarovalništvu, kadrovskih postopkih in drugih okoljih, kjer sistem UI ne sme biti le funkcionalen, ampak mora biti tudi razumljiv, pravno branljiv in družbeno sprejemljiv. V takšnih kontekstih postanejo pomembni ne le tehnični rezultati modela, temveč tudi kakovost razkritij, procesna zaščita osebnih podatkov, dokumentirana obravnava potencialnih pristranskosti in sledljivost etičnih odločitev v razvoju. IEEE-jevi standardi tukaj ponujajo uporabne orientacijske točke, tudi kadar niso neposredno harmonizirani evropski standardi.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je neposredna predvsem prek člena 13, ki za visokotvegane sisteme zahteva, da so zasnovani in razviti tako, da je njihovo delovanje dovolj transparentno, da lahko uvajalci razumejo izhode sistema in jih ustrezno uporabljajo. IEEE 7001 je v tem smislu posebej relevanten, ker transparentnost razume kot večnivojski, merljiv in preverljiv konstrukcijski element sistema, ne kot splošno komunikacijsko željo.

Povezava obstaja tudi s členom 10 AI Akta, saj ta zahteva ustrezne prakse upravljanja podatkov za visokotvegane sisteme, vključno z vidiki, ki se nanašajo na kakovost podatkov in možna pristranska odstopanja v učnih, validacijskih in testnih sklopih. IEEE 7003 je zato relevanten kot procesni okvir za obravnavo algoritemske pristranskosti, medtem ko IEEE 7002 dopolnjuje tisti del organizacijskega in razvojnega ravnanja, ki se nanaša na osebne podatke in zasebnostne zahteve. Pomembno je poudariti, da 7002 ne nadomešča GDPR, temveč pomaga tehnično-operativno strukturirati ravnanje v sistemih, kjer je uporaba osebnih podatkov del zasnove in delovanja.

V širšem smislu je ta sklop usklajen tudi s temeljnim ciljem AI Akta, ki je spodbujanje human-centric and trustworthy AI ob visoki ravni varstva zdravja, varnosti in temeljnih pravic. Prav zato so transparentnost, pristranskost in zasebnost v tem sklopu razumljene kot del operativizacije zaupanja vredne UI, ne kot dodatki po koncu razvoja. V povezavi z dosedanjimi poglavji ta sklop dopolnjuje organizacijski okvir iz 4.1, tveganostno logiko iz 4.2, terminološko in arhitekturno jasnost iz 4.3, podatkovni vidik iz 4.4, človeški nadzor iz 4.5 ter robustnostno-varnostni vidik iz 4.6. Njegova posebnost je prav v tem, da obravnava normativno kakovost sistema v odnosu do uporabnikov, posameznikov in prizadetih skupin.

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop posebej pomemben zato, ker odpira prostor za pripravo bolj konkretnih slovenskih podpornih instrumentov na področjih, kjer sama splošna načela praviloma ne zadostujejo. V praksi se hitro pokaže potreba po vzorčnih transparency profilih, minimalnih seznamih razkritij za uporabnike, standardiziranih pristopih k dokumentiranju bias analiz, razvojnih kontrolah za privacy-by-design ter obrazcih za sledljivost etičnih odločitev v sistemskem načrtovanju. IEEE-jevi dokumenti ponujajo dovolj jasno vsebinsko osnovo, da bi bilo takšne podporne instrumente mogoče začeti oblikovati tudi v nacionalnem prostoru.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako za slovenski prostor opredeliti **minimalni nabor informacij**, ki jih mora organizacija razkriti o delovanju, omejitvah in namenu AI sistema;

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

- kako standardizirati **evidence o obravnavi zasebnosti** v razvojnih in uvajalnih procesih AI rešitev;
- kako pripraviti uporabne smernice za **dokumentiranje in presojo algoritemske pristranskosti** v različnih tipih sistemov;
- kako v razvojne procese vključiti **sledljive etične zahteve**, ne da bi ostale ločene od tehničnih in poslovnih odločitev;
- ter kako te elemente smiselno povezati s tehnično dokumentacijo, PMS postopki, notranjimi presojami in pogodbenimi zahtevami do dobaviteljev.

Razvojna vrednost tega sklopa zato ni v tem, da bi ponavljal splošna etična načela, temveč da omogoča pripravo operativnih nacionalnih predlogov, s katerimi bi bilo mogoče transparentnost, zasebnost, obravnavo bias in odgovorno zasnovo sistemov UI v slovenskem prostoru obravnavati na bolj enoten, preverljiv in izvedben način. Prav v tem smislu je ta sklop eden izmed ključnih kandidatov za kasnejše tehnične profile, smernice in predloge standardizacijskih dopolnitev.

4.8 Podporni dokumenti za interoperabilnost, podatkovne prostore in infrastrukturne primere uporabe

Funkcija sklopa

Osmi vsebinski sklop zajema dokumente, ki praviloma ne tvorijo jedra horizontalnega upravljanja sistemov umetne inteligence, vendar pomembno dopolnjujejo standardizacijski okvir tam, kjer se umetna inteligenca uporablja v večsistemskih, podatkovno povezanih ali infrastrukturno kompleksnih okoljih. Njihova funkcija ni v tem, da bi nadomestili standarde za sisteme vodenja, tveganja, podatke, nadzor ali varnost, temveč da zagotovijo širši tehnični in organizacijski kontekst, v katerem je takšne sisteme sploh mogoče smiselno in interoperabilno uporabljati.

Ta sklop je metodološko pomemben zato, ker se številni sistemi UI ne izvajajo v izolaciji, temveč kot del širših digitalnih ekosistemov: v podatkovnih prostorih, v omrežjih in digitalnih storitvah, v porazdeljenih infrastrukturnih okoljih, v interoperabilnih platformah ali v sektorskih podatkovnih arhitekturah. V takšnih primerih vprašanje ni več le, ali je posamičen model ustrezno dokumentiran ali robusten, temveč tudi, ali sistem lahko deluje z drugimi sistemi, ali so podatkovni tokovi ustrezno opisani, ali obstajajo pravila za suverenost podatkov, sledljivost izvora, upravljanje dostopa in interoperabilno izmenjavo informacij. Data Spaces Support Centre zato svoj Blueprint opredeljuje kot celovit nabor smernic za podporo življenjskemu ciklu podatkovnih prostorov, ki vključuje konceptualni model, gradnike in priporočene standarde ter specifikacije.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

V tem smislu ta sklop pokriva predvsem tri ravni:

- **interoperabilnost in podatkovne prostore**, kjer je pomembna usklajena izmenjava podatkov, pravil dostopa, provenience in sledljivosti;
- **infrastrukturne in omrežne primere uporabe**, kjer je UI vpeta v digitalne storitve, telekomunikacije, 5G in prihodnja omrežja ali druge kompleksne informacijsko-komunikacijske sisteme;
- **povezovalne standardizacijske in referenčne platforme**, ki omogočajo orientacijo v hitro rastočem in razdrobljenem prostoru AI standardov ter sorodnih tehničnih publikacij.

Ključni standardi in dokumenti

V tem sklopu ne gre primarno za en sam "osrednji" standard, temveč za podporno plast referenčnih dokumentov, blueprintov in domensko specifičnih priporočil, ki skupaj dopolnjujejo horizontalni standardizacijski okvir.

Prva pomembna referenca je **Data Spaces Blueprint** Data Spaces Support Centre. DSSC ga predstavlja kot dosleden in celovit sklop smernic za vzpostavitev in razvoj podatkovnih prostorov, ki vključuje konceptualni model, gradnike, priporočene standarde in specifikacije. Posebej pomembno je, da Blueprint ne obravnava podatkovnega prostora le kot tehničnega vmesnika, temveč kot kombinacijo poslovnih, organizacijskih, pravnih in tehničnih gradnikov, med katerimi DSSC izrecno navaja podatkovno interoperabilnost, podatkovne modele, izmenjavo podatkov, provenienco in sledljivost, podatkovno suverenost in zaupanje, upravljanje identitet ter uveljavljanje pravil dostopa in uporabe. To je za AI okolje pomembno zato, ker ravno takšni gradniki omogočajo sledljivo in pravno ter tehnično urejeno uporabo podatkov v večorganizacijskih ekosistemih.

Drugi pomemben sklop referenc izhaja iz **ITU-T**. ITU izrecno navaja, da je razvil standarde za uporabo umetne inteligence pri orkestraciji 5G in prihodnjih omrežij, pri multimedijskih inovacijah, ocenjevanju kakovosti digitalnih storitev in izboljševanju energijske učinkovitosti. To pomeni, da ITU-T v tem dokumentu ni pomemben kot "horizontalni" standardizator za vse vrste UI, temveč kot vir **domensko natančnih priporočil** za infrastrukturne in komunikacijske kontekste, kjer je UI vključena v omrežja, storitvene arhitekture, telekomunikacijske operacije ali druge digitalne platforme. V tem smislu ITU-T dopolnjuje ISO/IEC in ETSI: ne toliko z upravljavskimi ali varnostnimi metanormami, temveč s priporočili za specifične operativne uporabe UI v infrastrukturi in storitvah.

Tretja pomembna referenca je **International AI Standards Exchange / AI Standards Exchange Database**, ki jo AI for Good pod vodstvom ITU, ISO in IEC opisuje kot pobudo, ki združuje vodilna telesa in deležnike pri razvoju odgovornih, varnih in vključujočih AI standardov. FAQ za bazo izrecno navaja, da je baza vzpostavljena pod okriljem World Standards Cooperation in v partnerstvu z IEEE, IETF in AI Standards Hub. Njena funkcija v

tem sklopu ni normativna, temveč orientacijska in povezovalna: omogoča hitrejšo preglednost nad velikim številom standardov in sorodnih publikacij ter olajša razumevanje, kako se posamezna področja standardizacije med seboj prekrivajo ali dopolnjujejo.

Praktična vrednost za slovenski prostor

Za slovenski prostor je ta sklop posebej pomemben zato, ker odpira vprašanja, ki se pojavijo takoj, ko umetna inteligenca preseže raven posamezne aplikacije in postane del širšega ekosistema sodelujočih organizacij, platform ali infrastrukturnih storitev. V takšnih primerih ni več dovolj, da je posamezen AI sistem dobro upravljan in dokumentiran; pomembno postane tudi, ali se lahko podatki varno in sledljivo izmenjujejo, ali je mogoče dokazati njihov izvor, ali so pravila uporabe dovolj jasno določena in ali različni sistemi sploh lahko komunicirajo v skupnem operativnem okolju. DSSC Blueprint je na tej ravni posebej uporaben prav zato, ker te izzive ne obravnava izolirano, temveč jih poveže v skupen model podatkovnega prostora.

Praktična vrednost se pokaže zlasti v okoljih, kjer bo slovenski prostor potreboval večorganizacijsko uporabo podatkov za AI: v javnem sektorju, zdravstvu, raziskovalnih inštitucijah, industrijskih ekosistemih, energetiki, mobilnosti in drugih domenah, kjer se podatki ne zbirajo in obdelujejo znotraj ene same organizacije. V takšnem okolju postanejo vprašanja provenience, sledljivosti, pravil dostopa, identitetnega upravljanja, podatkovnih modelov in interoperabilnosti neposredno uporabna tudi za UI sisteme. Data Spaces Blueprint s tem ponuja strukturirano osnovo, na katero se lahko navezujejo nadaljnji tehnični in organizacijski razmisleki v slovenskem prostoru.

ITU-T priporočila imajo praktično vrednost predvsem v infrastrukturnih kontekstih. Tam, kjer se umetna inteligenca uporablja za upravljanje omrežij, kakovost digitalnih storitev, multimedijske aplikacije ali druge storitvene platforme, je pomembno, da obstajajo domensko natančne reference, ki upoštevajo operativne posebnosti takšnega okolja. Za slovenski prostor je to relevantno zlasti tam, kjer se UI uporablja v telekomunikacijah, digitalnih javnih storitvah, večjih IKT sistemih ali industrijskih okoljih z močno omrežno komponento.

Dodatna praktična vrednost tega sklopa je tudi v boljši orientaciji. Ker je standardizacijski prostor UI hitro rastoč in razdrobljen, je za slovenske deležnike pomembno, da imajo na voljo tudi pregledna povezovalna orodja. AI Standards Exchange Database je na tej ravni uporabna kot spremljevalni vir, ker omogoča širšo sliko razvoja standardov, draftov in povezanih publikacij, kar je posebej dragoceno pri spremljanju trendov, razvoju Observatorija standardov UI in sprotnem dopolnjevanju nacionalnih strokovnih podlag.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Povezava z Aktom o umetni inteligenci in drugimi rezultati projekta

Povezava tega sklopa z Aktom o umetni inteligenci je bolj posredna kot pri standardih za sisteme vodenja, tveganja, podatke, človeški nadzor ali robustnost, vendar je še vedno pomembna. AI Akt visoko tvegane sisteme postavlja v okvir, kjer so tehnična dokumentacija, podatkovno upravljanje, sledljivost, preglednost in skladnost z drugimi pravnimi režimi ključni pogoji za zakonito in odgovorno uporabo. V večorganizacijskih ali infrastrukturnih okoljih teh zahtev ni mogoče učinkovito izvajati brez interoperabilnih pravil za podatke, jasnega upravljanja dostopa, provenience in tehnične povezljivosti med sistemi. Prav na tej ravni ta sklop dopolnjuje operativno uporabo AI Akta, čeprav ni vezan na en sam njegov člen.

Ta sklop je tesno povezan tudi z drugimi rezultati projekta. Rezultat 1.3 je že pokazal, da standardizacijski prostor za umetno inteligenco ni omejen na ISO/IEC SC 42, temveč vključuje tudi ETSI, IEEE, ITU-T ter širše dokumente s področja podatkovnih prostorov in interoperabilnosti. Rezultat 1.1 je pri dokumentaciji, sledljivosti in artefaktih nakazal potrebo po strukturiranem tehničnem okolju, rezultat 2.3 pa je pri dobrih praksah dodatno pokazal, da so nadzor nad vhodnimi podatki, logiranje, sledljivost sprememb in upravljanje deljenih virov med najbolj ponovljivimi operativnimi zahtevami. Ta sklop zato deluje kot podporna plast, ki povezuje AI standardizacijo z realnim podatkovnim in infrastrukturnim okoljem, v katerem bo slovenski prostor sisteme dejansko uvajal.

Pomen za nadaljnje nacionalne tehnične razmisleke

Z vidika nadaljnjih nacionalnih tehničnih razmislekov je ta sklop posebej pomemben zato, ker odpira prostor za sektorsko in infrastrukturno konkretizacijo tam, kjer horizontalni standardi sami po sebi ne zadoščajo. Mednarodni in evropski dokumenti že določajo organizacijske, tveganostne, podatkovne, nadzorne in varnostne okvirje, vendar v praksi hitro nastane potreba po dodatnih usmeritvah za okolja, kjer se UI uporablja v podatkovnih prostorih, omrežjih, večorganizacijskih ekosistemih ali drugih interoperabilnih arhitekturah.

Za nadaljnjo nacionalno obravnavo so na tej ravni posebej relevantna zlasti naslednja vprašanja:

- kako v slovenskem prostoru povezati zahteve za AI dokumentacijo in sledljivost z gradniki podatkovnih prostorov, kot so provenienca, pravila dostopa in podatkovna suverenost;
- kako oblikovati minimalne tehnične profile za AI sisteme, ki delujejo v večorganizacijskih ali infrastrukturnih okoljih;
- kako pri javnih in sektorskih projektih poenotiti zahteve glede interoperabilnosti podatkov, identitetnega upravljanja, izmenjave in spremljanja uporabe;
- kako vključiti infrastrukturne in omrežne posebnosti v nacionalne tehnične razmisleke tam, kjer se UI uporablja v digitalnih storitvah, komunikacijskih sistemih ali kompleksnih platformah;

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

- ter kako Observatorij standardov UI dopolniti z dovolj pregledno orientacijo po podpornih dokumentih, ki ne sodijo v jedro horizontalnih standardov, so pa za prakso kljub temu pomembni.

Razvojna vrednost tega sklopa zato ni v tem, da bi samostojno ustvarjal novo jedro standardizacijskega okvira, temveč v tem, da omogoča pripravo bolj uporabnih slovenskih podpornih instrumentov za interoperabilnost, podatkovne prostore in infrastrukturne primere uporabe. Prav v tem smislu lahko deluje kot most med horizontalno AI standardizacijo in konkretnimi sektorskimi oziroma tehničnimi okolji, v katerih bo treba te standarde dejansko uporabljati.

4.9 Operativni povzetek ključnih ugotovitev

Po izvedeni vsebinski razvrstitvi ključnih standardov in referenčnih dokumentov je mogoče ugotoviti, da standardizacijski okvir za umetno inteligenco za slovenski prostor ni zgrajen okoli enega samega osrednjega dokumenta, temveč okoli medsebojno povezanega sklopa standardov, ki skupaj pokrivajo organizacijski, tveganostni, podatkovni, arhitekturni, nadzorni, varnostni in etično-uporabniški vidik sistemov UI. Rezultat 1.3 to večplastnost povzema zelo jasno: ISO/IEC nudi hrbtenico upravljanja, tveganj, terminologije, arhitekture, kakovosti podatkov in človeškega nadzora; CEN/CENELEC oziroma EN izdaje omogočajo njihovo evropsko operativizacijo; ETSI prevaja robustnost in kibernetsko varnost v konkretne kontrole in preskuse; IEEE serija 7000 operacionalizira transparentnost, zasebnost, pristranskost in etične zahteve; ITU-T in podporni dokumenti za podatkovne prostore pa dodajajo domensko in infrastrukturno natančnost tam, kjer sistemi UI delujejo v širših digitalnih ekosistemih.

Z vidika praktične uporabnosti se kot posebej pomembna kaže naslednja operativna logika. Najprej je treba vzpostaviti organizacijsko jedro sistema skladnosti, pri čemer standardi za sisteme vodenja in upravljanje tveganj omogočajo, da se odgovornosti, evidence, notranje presoje, odobritve in CAPA ukrepi ne izvajajo ad hoc, temveč v ponovljivem okviru. Nato je treba ta okvir napolniti z vsebinskimi tehničnimi gradniki: z enotno terminologijo in arhitekturnim opisom sistema, s preverljivo kakovostjo podatkov, z jasno določenimi mehanizmi človeškega nadzora, z robustnostnimi in kibernetsko-varnostnimi kontrolami ter s strukturiranimi informacijami o transparentnosti, zasebnosti in pristranskosti. Rezultat 1.1 to logiko potrjuje z zelo praktičnim prevodom zahtev AI Akta v dokumentacijo, kontrolne sezname in artefact passport pristop, kjer so ti sklopi že povezani v enotno dokazilno celoto.

Za nadaljnje delo v slovenskem prostoru je zato smiselno izhajati iz treh operativnih sklepov:

1. Prvič, prioritetno je treba obravnavati tiste standarde, ki neposredno omogočajo prehod od načel do dokazil: to so zlasti standardi za sistem vodenja, tveganja,

arhitekturo, podatke, človeški nadzor, transparentnost in varnost, ker se prav prek njih pravne zahteve prevajajo v procese, kontrole, metrike in evidence.

2. Drugič, pri pripravi nadaljnjih tehničnih predlogov ni smiselno podvajati mednarodnih in evropskih standardov, temveč predvsem razvijati nacionalne profile, minimalne artefakte, kontrolne sezname in pojasnjevalne dopolnitve, ki bodo te standarde naredili neposredneje uporabne v slovenskem institucionalnem in gospodarskem okolju.
3. Tretjič, nadaljnji nacionalni tehnični razmislek mora biti izrazito povezan z dejanskimi primeri uporabe, saj rezultat 2.3 pokaže, da se v verodostojnih praksah ponavljajo prav vprašanja namena uporabe, klasifikacije podatkov, infrastrukture, razmejitve odgovornosti, logiranja, kakovosti, človeškega nadzora in upravljanja sprememb.

Na tej podlagi je mogoče precej jasno opredeliti tudi neposredno uporabno vrednost tega poglavja za nadaljevanje dokumenta. Vsebinska razvrstitev standardov namreč že kaže, na katerih področjih ima slovenski prostor na voljo dovolj trdno mednarodno in evropsko referenčno jedro ter na katerih področjih se ob tem vseeno odpira potreba po dodatni nacionalni konkretizaciji. To velja zlasti za minimalne zahteve glede artefaktov in evidenc, za profile človeškega nadzora, za standardizirane pristope k dokumentiranju podatkovne kakovosti, za baseline varnostne kontrole, za transparency profile ter za interoperabilnostne in infrastrukturne profile v kompleksnejših okoljih. Prav v tem smislu to poglavje ne opravlja le pregledne funkcije, temveč pripravlja strokovno podlago za naslednji korak: za bolj neposredno presojo, katere standarde je treba v slovenskem prostoru obravnavati kot nosilne reference in kje bi bilo smiselno nadaljevati z dodatnimi tehničnimi predlogi.

Izvedena vsebinska razvrstitev je pokazala, da je standardizacijski okvir za umetno inteligenco sestavljen iz več medsebojno povezanih slojev:

- iz organizacijskih standardov za sisteme vodenja,
- metodoloških podlag za upravljanje tveganj,
- dokumentov za terminologijo in arhitekturo,
- standardov za kakovost podatkov, človeški nadzor, robustnost in kibernetsko varnost ter
- dopolnilnih dokumentov za transparentnost, zasebnost, pristranskost, interoperabilnost in infrastrukturne primere uporabe.

Takšna razporeditev potrjuje, da standardizacijski prostor ni nepovezan katalog, temveč funkcionalno povezan sistem referenc, ki skupaj omogočajo prehod od pravnih zahtev k operativnim procesom, tehnični dokumentaciji, kontrolnim točkam in dokazilom. Tudi drugi projektni rezultati sledijo isti logiki: rezultat 1.1 povezuje zahteve AI Akta s procesi, dokumentacijo in artefakti, rezultat 2.3 pa iz primerov dobrih praks izlušči tiste operativne elemente, ki jih je mogoče prevesti v konkretne zahteve in kontrolne sezname.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Na tej podlagi je v naslednjem koraku smiselno preiti od vsebinske razvrstitve k bolj neposredni presoji, katere standarde in standardne sklope je za slovenski prostor smiselno obravnavati kot nosilne reference, katere kot podporne ali razvojno pomembne dokumente ter na katerih področjih se ob obstoječem mednarodnem in evropskem okviru še vedno kaže potreba po dodatni nacionalni tehnični konkretizaciji.

5. Presoja prednostnih standardnih sklopov in potreb po nacionalni tehnični konkretizaciji

5.1 Namen in logika presoje

Po izvedeni vsebinski razvrstitvi ključnih standardov in referenčnih dokumentov je mogoče preiti od pregledne sistematizacije k naslednjemu analitičnemu koraku, ki je za ta dokument osrednji: k presoji, kateri standardni sklopi imajo za slovenski prostor status nosilnih referenc, kateri delujejo predvsem kot podporni ali razvojno pomembni dokumenti ter na katerih področjih se ob obstoječem mednarodnem in evropskem okviru še vedno kaže potreba po dodatni nacionalni tehnični konkretizaciji. Ta korak je vsebinsko bistven, ker dokument s tem preide iz faze kategorizacije v fazo selekcije in prioritetizacije, kar je tudi njegova dejanska funkcija znotraj rezultata 2.2. Rezultat 2.2 je zamišljen kot poročilo, ki povezuje seznam relevantnih EU in mednarodnih standardov z začetkom priprave predlogov nacionalnih standardov, kasnejše faze projekta pa šele nato predvidevajo tehnični vidik prilagoditev in dokumentacijo tehničnih predlogov.

Namen tega poglavja zato ni dokončno odločiti, kateri nacionalni standardi bodo pripravljeni ali prevzeti, temveč strokovno utemeljiti, katere reference in standardne sklope je smiselno obravnavati kot prednostne podlage za nadaljnje delo. Takšna presoja temelji na dosedanji analizi projekta: rezultat 1.3 je vzpostavil obsežen standardizacijski katalog, rezultat 1.1 je pravne zahteve AI Akta prevedel v procese, dokumentacijo, artefakte in dokazila, rezultat 2.3 pa je iz primerov dobrih praks izluščil skupne operativne vzorce, kot so governance, omejitev namena uporabe, klasifikacija podatkov, logiranje, človeški nadzor, upravljanje sprememb in infrastrukturna disciplina. To pomeni, da presoja v tem poglavju ni abstraktna, temveč temelji na vprašanju, kateri standardni sklopi dejansko najbolje podpirajo prehod od regulativnih zahtev k operativni uporabi v slovenskem okolju.

Pri tej presoji je smiselno izhajati iz treh operativnih vprašanj. Prvič, ali posamezen standardni sklop neposredno podpira oblikovanje procesov, dokumentacije, kontrolnih točk in dokazil, ki jih organizacije v praksi potrebujejo. Drugič, ali je sklop za slovenski prostor dostopen in praktično uporaben, zlasti prek evropskih oziroma nacionalnih izdaj ter v kontekstu razpisov, notranjih pravil, presoj skladnosti in sodelovanja s SIST. Tretjič, ali obstoječi mednarodni in evropski dokumenti že dajejo dovolj neposredne opore za uporabo v slovenskem okolju ali pa se kljub temu kaže potreba po dodatnih profilih, pojasnjevalnih

dopolnitvah, minimalnih artefaktih ali drugih podpornih tehničnih rešitvah. Prav ta tri vprašanja določajo notranjo logiko nadaljnje razvrstitve.

5.2 Nosilne reference za slovenski prostor

Kot nosilne reference je za slovenski prostor smiselno obravnavati tiste standardne sklope, ki tvorijo jedro operativno uporabnega okvirja za razvoj, uvajanje, upravljanje, presojo in nadzor sistemov umetne inteligence. To so predvsem sklopi, ki organizacijam omogočajo, da pravne zahteve in širše standardizacijske usmeritve prevedejo v ponovljive procese, odgovornosti, tehnično dokumentacijo, evidence in kontrolne mehanizme. Na podlagi dosedanje analize se kot nosilni sklopi izkazujejo zlasti:

- standardi za sisteme vodenja umetne inteligence,
- standardi za upravljanje tveganj,
- standardi za terminologijo in arhitekturne okvire,
- standardi za kakovost podatkov in podatkovno upravljanje,
- standardi za človeški nadzor in obvladljivost,
- standardi za robustnost in kibernetsko varnost.

Takšna razvrstitev izhaja neposredno iz strukture rezultata 1.3, ki te vsebinske sklope filtrira kot ključne za skladnost, ter iz rezultata 1.1, ki jih poveže z obveznostmi po AI Aktu in z oblikami dokazljive dokumentacije.

Znotraj teh sklopov imajo posebno težo standardi, ki omogočajo prehod od načel k dokazilom. Na ravni sistemov vodenja je to predvsem ISO/IEC 42001 kot organizacijsko jedro upravljanja UI; na ravni tveganj ISO/IEC 23894 oziroma njegova evropska izdaja; na ravni terminologije in arhitekture ISO/IEC 22989 in ISO/IEC 23053; na ravni podatkov serija ISO/IEC 5259; na ravni človeškega nadzora ISO/IEC TS 8200; na ravni robustnosti in varnosti pa ETSI-jevi dokumenti serije SAI, zlasti ETSI TS 104 223 in ETSI TS 104 050. Ti dokumenti imajo skupno lastnost, da ne ostajajo na ravni splošne usmeritve, temveč ponujajo dovolj jasno metodološko ali tehnično podlago za oblikovanje politik, registrov, opisov arhitekture, pregledov podatkov, nadzornih mehanizmov, varnostnih kontrol in spremljajočih evidenc. Prav zaradi te neposredne operativne vrednosti jih je smiselno obravnavati kot osrednje tehnične reference za nadaljnje delo. Ta logika je skladna tudi z rezultatom 2.3, ki med skupnimi vzorci dobrih praks posebej izpostavi governance, infrastrukturno disciplino, klasifikacijo podatkov, človeški nadzor, omejitev namena uporabe, logiranje in upravljanje sprememb.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Za slovenski prostor je dodatno pomembno, da del teh nosilnih referenc obstaja tudi v evropskih oziroma nacionalnih izdajah, kar bistveno olajša njihovo uporabo v praksi. To je posebej relevantno pri standardih, kot so EN ISO/IEC 23894, EN ISO/IEC 22989 in EN ISO/IEC 23053, ki omogočajo enotnejše sklicevanje v razpisih, pogodbah, notranjih pravilnikih in presojah. Kjer evropska izdaja še ni vzpostavljena ali ni v ospredju, pa ostaja mednarodna različica vseeno strokovno nosilna referenca, če je umeščena v širšo, medsebojno povezano celoto drugih standardov. To pomeni, da nosilna funkcija posamezne reference ni odvisna le od njenega formalnega statusa, temveč tudi od njene sposobnosti, da deluje kot dejanska podlaga za izvajanje procesov in pripravo dokazil.

5.3 Podporne in razvojno pomembne reference

Poleg nosilnih sklopov standardizacijski prostor vključuje tudi dokumente, ki jih je za slovenski prostor smiselno obravnavati kot podporne ali razvojno pomembne reference. To niso obrobni dokumenti v smislu nepomembnosti, temveč dokumenti, katerih vrednost je praviloma drugačna: pogosto delujejo kot razlagalni, dopolnilni, specializirani ali domensko natančni viri, ki ne tvorijo sami organizacijskega jedra standardizacijskega okvira, vendar pomembno prispevajo k njegovi vsebinski zrelosti, širini in uporabnosti.

V to skupino sodijo zlasti:

- tehnična poročila in širši razlagalni dokumenti, kot je ISO/IEC TR 24028;
- standardi in dokumenti serije IEEE 7000 za transparentnost, zasebnost, obravnavo pristranskosti in etično zasnovo;
- podporni dokumenti za interoperabilnost, podatkovne prostore in večorganizacijska okolja;
- domensko natančna priporočila ITU-T;
- orientacijske in povezovalne baze oziroma platforme za pregled standardizacijskega razvoja.

Ti dokumenti so pomembni zlasti tam, kjer organizacija potrebuje bolj granularne metodološke usmeritve, bolj razvidno povezavo med tehničnimi in družbenimi lastnostmi sistema ali pa bolj specifične tehnične okvire za posebne infrastrukturne in podatkovne scenarije. Rezultat 1.3 je to širino standardizacijskega prostora že jasno pokazal, rezultat 2.3 pa jo potrdi prek primerov, kjer se vprašanja transparentnosti, minimizacije podatkov, odgovornega ravnanja z uporabniškimi vnosi, omejitve namena uporabe in dostopnosti pojavljajo kot pomembne, vendar pogosto ne povsem pokrite teme.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Za slovenski prostor imajo ti podporni dokumenti dvojno vrednost. Po eni strani širijo razumevanje standardizacijskega ekosistema in preprečujejo, da bi se standardizacija umetne inteligence skrčila zgolj na organizacijske in tehnično-varnostne standarde. Po drugi strani pa delujejo kot **razvojni vir za prihodnje dopolnitve**, zlasti tam, kjer se bodo v praksi pojavila vprašanja transparentnih informacij za uporabnike, privacy-by-design zahtev, dokumentiranja bias analiz, interoperabilnosti v podatkovnih prostorih ali posebnih infrastrukturnih konfiguracij. To pomeni, da podporni dokumenti v tem poročilu niso obravnavani kot sekundarni zato, ker bi bili manj pomembni, temveč zato, ker praviloma ne predstavljajo prvega sloja, na katerem bi bilo smiselno graditi jedro nacionalne tehnične podlage.

5.4 Področja, na katerih se kaže potreba po dodatni nacionalni tehnični konkretizaciji

Osrednji prispevek tega poglavja ni le razlikovanje med nosilnimi in podpornimi referencami, temveč tudi opredelitev tistih področij, kjer obstoječi mednarodni in evropski okvirji kljub svoji vsebinski zrelosti še ne dajejo dovolj neposredne opore za slovenski prostor. Pri tem ne gre za pomanjkljivost standardov kot takih, temveč za dejstvo, da je mednarodni oziroma evropski standard po svoji naravi praviloma širši, bolj generičen in manj prilagojen specifičnim institucionalnim, organizacijskim in sektorskim okoliščinam posamezne države. Rezultat 2.2 služi kot podlaga za kasnejše tehnične prilagoditve nacionalnih standardov in za pripravo dokumentacije tehničnih predlogov.

Na podlagi dosedanje analize se kot posebej pomembna področja za nadaljnjo nacionalno tehnično konkretizacijo kažejo zlasti naslednja:

Prvo področje je **minimalni nabor artefaktov in evidenc**, ki jih je smiselno pričakovati od organizacij, ki razvijajo, uvajajo ali uporabljajo sisteme UI. Rezultat 1.1 je pri tej temi že nakazal pristop z glosarji, artefact passport obrazci, standardiziranimi kazali dokumentacije in kontrolnimi seznami, vendar se prav na tej ravni odpira potreba po bolj poenotenem, slovenskim deležnikom neposredno uporabnem naboru.

Drugo področje je **človeški nadzor in obvladljivost v konkretnih delovnih procesih**. Mednarodni standardi in tehnične specifikacije dajejo dovolj dobro osnovo za razumevanje načel in mehanizmov, vendar se v praksi hitro odpre vprašanje, kako naj bodo override, safe-stop, eskalacijski pragovi, usposabljanja in dokazila o testiranju obvladljivosti strukturirani v domačem okolju, posebej v javnem sektorju in drugih reguliranih domenah.

Tretje področje je **dokumentiranje kakovosti podatkov in povezava med podatkovnimi pregledi, registri tveganj ter tehnično dokumentacijo**. Serija ISO/IEC 5259 že določa, kaj meriti in kako to upravljati, vendar se na nacionalni ravni lahko pokaže potreba po bolj

praktičnih obrazcih, minimalnih metrikah, pogodbenih zahtevah do dobaviteljev podatkov in enotnejših pravilih za sledljivost sprememb datasetov.

Četrto področje je **baseline varnostni minimum za AI sisteme**, zlasti v razpisni, pogodbeni in presojevalni praksi. ETSI-jevi dokumenti že določajo pomemben varnostni minimum, vendar je za slovenski prostor realno pričakovati potrebo po poenostavljenih seznamih dokazil, profilih za različne tipe sistemov ter jasnejši povezavi med varnostnimi kontrolami, registri tveganj, notranjimi presojami in morebitnimi zahtevami organov ugotavljanja skladnosti.

Peto področje je **transparentnost, zasebnost, pristranskost in odgovorna zasnova**, kjer so podporni dokumenti sicer vsebinsko bogati, vendar pogosto potrebujejo dodatni prevod v operativne obrazce: transparency profile, evidence o privacy-by-design ukrepih, vzorčne bias analize, seznam minimalnih razkritij za uporabnike in podobno.

Šesto področje je **interoperabilnost, podatkovni prostori in sektorski oziroma infrastrukturni profili**, kjer horizontalni AI standardi ne zadoščajo za večorganizacijska okolja, omrežne storitve, podatkovne prostore in druge tehnično kompleksnejše scenarije. Tu se zlasti kaže prostor za dopolnilne profile, ne nujno za samostojne nove standarde.

5.5 Operativna prioritizacija za nadaljnje delo

Če se dosedanja presoja prevede v operativno zaporedje nadaljnjega dela, se kot najbolj smiselna kaže trojna prioritizacija.

V **prvi prioriteti** so sklopi, ki neposredno omogočajo oblikovanje jedra sistema skladnosti in dokazljivosti. Sem sodijo sistemi vodenja, upravljanje tveganj, terminologija in arhitekturni okvir, kakovost podatkov, človeški nadzor ter robustnost in kibernetska varnost. To so področja, na katerih se morajo slovenski deležniki najprej opreti na obstoječe mednarodne in evropske reference, saj prav ta jedro najhitreje omogoči prehod od abstraktnih zahtev k organizacijskim postopkom, evidencam in tehničnim dokazilom. Ta prioritetna skupina je najbolj neposredno povezana tudi z verigo rezultatov 1.1–1.3–2.3 in s kasnejšim rezultatom 4.1, ki bo obravnaval tehnični vidik prilagoditev nacionalnih standardov.

V **drugi prioriteti** so področja, na katerih je obstoječi okvir sicer dovolj razvit, vendar potrebuje dodatno operativno pojasnitev za slovenski prostor. Sem sodijo predvsem minimalni artefakti, evidence, transparency profili, baseline varnostni seznami, vzorčne bias analize, pojasnjevalne smernice za privacy-by-design in podobni podporni instrumenti. Tu ni smiselno razmišljati o podvajanju standardov, temveč predvsem o nacionalno prilagojenih oblikah njihove uporabe.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

V **tretji prioriteti** so področja, ki so močno domensko ali infrastrukturno odvisna ter bodo verjetno zahtevala nadaljnjo sektorsko diferenciacijo. Sem sodijo podatkovni prostori, interoperabilnost, infrastrukturni profili, posebnosti javnega sektorja, zdravstva, financ, komunikacijskih omrežij in drugih kompleksnejših okolij. Na tej ravni je priporočljivo izhajati iz horizontalnih standardov in podpornih dokumentov, nato pa po potrebi razvijati bolj specializirane profile ali dodatne tehnične priloge.

Če celotno presojo povzamemo še bolj neposredno: za slovenski prostor je najprej treba identificirati sklope, ki omogočajo sistemsko izvajanje skladnosti, nato sklope, ki ta okvir naredijo neposredno operativen, in šele nato sklope, ki zahtevajo sektorsko ali infrastrukturno nadgradnjo. Prav takšna operativna logika omogoča, da rezultat 2.2 ostane v svoji pravilni vlogi: ne kot formalni predlog standarda, temveč kot strokovno utemeljena podlaga za naslednje faze projekta, sodelovanje s SIST in pripravo kasnejših tehničnih predlogov.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

6. Prednostna področja za pripravo nacionalnih tehničnih dopolnitev in standardizacijskih predlogov

6.1 Namen

Po izvedeni presoji standardnih sklopov je v naslednjem koraku smiselno preiti od vprašanja, kateri standardi in referenčni dokumenti so za slovenski prostor pomembni, k vprašanju, na katerih področjih je smiselno pripraviti nadaljnje nacionalne tehnične dopolnitve, profile ali druge podporne standardizacijske predloge. Namen tega poglavja zato ni oblikovati dokončnega predloga nacionalnih standardov v formalnem smislu, temveč identificirati tista področja, kjer obstoječi mednarodni in evropski dokumenti že zagotavljajo dovolj trdno referenčno osnovo, vendar se v slovenskem prostoru kljub temu kaže potreba po dodatni operativni konkretizaciji.

To poglavje bo predvsem opredelilo prednostna področja, tip pričakovanih nacionalnih izhodov in njihovo operativno utemeljitev.

Dosedanji rezultati projekta za takšen korak že dajejo dovolj konkretno osnovo. Rezultat 1.1 je pravne zahteve AI Akta povezal s procesi, kontrolnimi seznami, dokazili, FRIA/PIA predlogami in artefact passport pristopom, pri čemer je posebej poudaril potrebo po glosarju, standardiziranih predlogah, javno dostopnih obrazcih in vzorčnih razpisnih členih. Rezultat 2.3 pa je pokazal, da se v verodostojnih primerih dobre prakse kot ponavljajoči se operativni vzorci pojavljajo governance, dejanski človeški nadzor, upravljanje vhodov, minimizacija podatkov, logiranje, nadzor nad spremembami in infrastrukturna disciplina. Na tej podlagi je mogoče razmeroma zanesljivo določiti, kje bi nacionalni prostor imel največ koristi od dodatne tehnične konkretizacije.

6.2 Merila za izbor prednostnih področij

Pri opredelitvi prednostnih področij za nadaljnje nacionalne tehnične dopolnitve je smiselno izhajati iz treh medsebojno povezanih meril.

Prvo merilo je **operativna neposrednost potrebe**. Prednost imajo tista področja, na katerih se zahteve iz mednarodnih in evropskih standardov neposredno prevajajo v procese, evidence, obrazce, odobritve, kontrolne sezname ali tehnično dokumentacijo, ki jo morajo organizacije v praksi dejansko uporabljati. Na tej ravni so posebej relevantni rezultati 1.1, kjer so taki gradniki že deloma razviti.

Drugo merilo je **ponovljivost potrebe čez različne primere uporabe**. Prednost imajo tista področja, ki se ne pojavljajo le v enem specifičnem sektorju, temveč se kot potreba ponavljajo v več različnih kontekstih. Rezultat 2.3 pri tem posebej potrjuje ponovljivost vprašanj človeškega nadzora, upravljanja vhodov, odgovornosti, sprememb in kakovosti podatkov.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Tretje merilo je **primernost za nacionalno konkretizacijo brez podvajanja obstoječih standardov**. Prednost imajo tista področja, kjer je mogoče na podlagi že obstoječih mednarodnih in evropskih referenc pripraviti uporabnejše slovenske profile, minimalne artefakte, smernice ali predloge dokumentacije, ne da bi bilo treba vsebino standardov na novo normativno pisati.

6.3 Prednostno področje 1: minimalni nabor artefaktov in dokazil

Prvo prednostno področje predstavlja priprava **minimalnega nabora artefaktov, evidenc in spremljajoče dokumentacije**, ki bi ga bilo smiselno priporočiti za organizacije, ki razvijajo, uvajajo ali uporabljajo sisteme UI v slovenskem prostoru. Gre za področje, kjer je potreba po nacionalni konkretizaciji posebej izrazita, saj mednarodni standardi določajo vsebinske zahteve, ne dajejo pa vedno dovolj neposredno uporabnih predlog, kako naj bodo te zahteve organizirane v dejanskih dokumentih.

Rezultat 1.1 na tej ravni že daje uporabno osnovo, saj vključuje kontrolne sezname za ključne faze, predloge FRIA in PIA/DPIA ter vzorčne obrazce "Artefact Passport" za modele in podatkovne nabore. Ti elementi niso še samostojen standardiziran nacionalni okvir, vendar že nakazujejo praktično smer, v kateri je mogoče povezati arhitekturo sistema, kakovost podatkov, tveganja, transparentnost, pristranskost, varnost in človeški nadzor v bolj enotno dokumentacijsko strukturo.

Prav zaradi te že obstoječe osnove je smiselno to področje obravnavati kot enega najmočnejših kandidatov za nadaljnjo nacionalno tehnično dopolnitev.

Na tej ravni bi bil kot nacionalni izhod smiseln predvsem:

- priporočeni **minimalni artefaktni paket** za sisteme UI,
- poenotene predloge za **FRIA, PIA/DPIA in artefact passport**,
- standardizirano kazalo tehnične dokumentacije,
- ter vzorčni nabor povezav med registri tveganj, arhitekturo, podatkovno dokumentacijo in PMS evidencami.

To področje je posebej prioritetno zato, ker neposredno zmanjšuje nepreglednost, izboljšuje primerljivost dokazil in organizacijam omogoča hitrejši prehod od zahtev standardov k operativni uporabi.

6.4 Prednostno področje 2: profil za človeški nadzor in obvladljivost

Drugo prednostno področje je priprava bolj konkretnega **profila za človeški nadzor in obvladljivost**. Mednarodne tehnične specifikacije, zlasti ISO/IEC TS 8200, in povezani dokumenti za transparentnost že dajejo uporabno osnovo, vendar se v praksi hitro pokaže potreba po dodatnem prevodu v konkretne operativne zahteve: override postopke, safe-stop

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

mehanizme, eskalacijske pragove, dokumentiranje vaj in usposabljanj ter označitev točk nadzora v tehnični dokumentaciji.

Rezultat 2.3 zelo jasno pokaže, da učinkovitega človeškega nadzora ni mogoče reducirati na formalno zahtevo, temveč mora biti zasnovan kot dejanski operativni mehanizem, ki vključuje razumljivost rezultatov, določene odgovornosti, podporo uporabnikom in procese eskalacije. Rezultat 1.1 pa na tej ravni že daje uporabno podlago: v kontrolnih seznamih za ključne faze izrecno omenja dokaze o obvladljivosti oziroma človeškem nadzoru po ISO/IEC TS 8200, v razširjenem seznamu dokumentacije pa posebej navaja pravila za "override" in "safe-stop", pragove zaupanja ter opis vlog pri človeškem nadzoru. Poleg tega priloga z vzorci "Artefact Passport" za model predvideva tudi posebno polje za mehanizme človeškega nadzora.

Na tej ravni bi bil kot nacionalni izhod smiseln predvsem:

- **minimalni profil človeškega nadzora** za različne tipe sistemov UI,
- vzorčni protokoli za override, safe-stop in eskalacijo,
- predloge dokazil o testiranju obvladljivosti,
- ter priporočila za dokumentiranje vmesnikov, vlog in usposabljanj.

To področje je prioritetno zato, ker se ponavlja čez različne primere uporabe in ker je njegova dejanska operativizacija v praksi pogosto bistveno zahtevnejša od same formalne zahteve po prisotnosti človeka.

6.5 Prednostno področje 3: kakovost podatkov in podatkovna dokumentacija

Tretje prednostno področje je priprava bolj uporabnega nacionalnega okvira za **kakovost podatkov, podatkovno dokumentacijo in sledljivost sprememb datasetov**. Mednarodna standarda ISO/IEC 5259-1 in 5259-3 že dovolj jasno določata, kaj je treba meriti in kako to upravljati, vendar se v praksi hitro odpre vprašanje, kako te zahteve prevesti v minimalne metrike, obrazce, evidence in pogodbeno prakso.

Rezultat 1.1 na tej ravni že vsebuje več uporabnih gradnikov, vendar jih je treba razumeti kot predloge in delovne vzorce, ne kot že dokončno standardiziran nacionalni paket. V kontrolnih seznamih za ključne faze so izrecno navedena merila kakovosti podatkov po ISO/IEC 5259-1 ter načrt upravljanja kakovosti podatkov po ISO/IEC 5259-3. V razširjenem seznamu dokumentacije so posebej omenjeni opis podatkov, njihova kakovost, uravnoteženost, postopki čiščenja in metapodatki. V vzorcu "Artefact Passport" za podatkovni nabor pa so vključena polja za kakovostne kazalnike, porazdelitve in neuravnoteženost, označevanje oziroma ground truth, transformacije in pipeline, anonimizacijo ter povezave z GDPR/PIA.

Rezultat 2.3 dodatno potrjuje, da so kakovost vhodov, minimizacija, klasifikacija podatkov in upravljanje sprememb med najbolj ponovljivimi operativnimi vzorci v primerih dobre prakse. Na tej podlagi bi bil kot nacionalni izhod smiseln predvsem:

- priporočeni **minimalni nabor metrik in pregledov kakovosti podatkov**,
- poenoteni obrazci za evidenco izvora, transformacij in sprememb datasetov,
- minimalne zahteve za dobavitelje podatkov,
- ter povezava podatkovne dokumentacije z registri tveganj in tehnično dokumentacijo.

To področje je prioritetno zato, ker je podatkovna kakovost eden najbolj odločilnih in hkrati najtežje enotno dokumentiranih elementov skladnosti.

6.6 Prednostno področje 4: baseline varnostni profil za sisteme UI

Četrto prednostno področje je priprava **baseline varnostnega profila** za sisteme UI, zlasti za potrebe javnih naročil, pogodbenih zahtev in osnovnih presoj skladnosti. ETSI-jeve specifikacije že določajo pomemben varnostni minimum, vendar bo v slovenskem okolju zelo verjetno potrebna bolj neposredno uporabna konkretizacija: katere kontrole se pričakujejo kot minimum, katera dokazila naj se zahtevajo in kako naj se takšne zahteve vključijo v razpisno ter presojevalno prakso.

Tudi tu rezultat 1.1 že daje uporabno podlago, vendar v obliki priporočil, kontrolnih seznamov in vzorčnih artefaktov. V kontrolnem seznamu pred uvedbo izrecno zahteva dokaze o robustnosti in kibernetski varnosti po ETSI TS 104 223. V razširjenem seznamu dokumentacije navaja rezultate testov robustnosti, pristranskosti in kibernetske varnosti ter posebej omenja osnovne evasion/poisoning teste po ETSI TS 104 223. V priporočilih za "hitre korake" in pogodbeno prakso so navedeni tudi minimalni varnostni nabor, politike za "safe-stop", pogoji za "rollback" in spremljanje incidentov. V vzorcu "Artefact Passport" za model pa so vključeni robustnostni testi, varnostne kontrole, odvisnosti oziroma SBoM ter povezava z incidenti in CAPA.

Na tej podlagi bi bil kot nacionalni izhod smiseln predvsem:

- **minimalni baseline seznam varnostnih kontrol** za sisteme UI,
- predloga za dokazila o varnostnih testih, incidentih in rollback postopkih,
- ter vzorčne pogodbene oziroma razpisne zahteve za dobavitelje.

To področje je prioritetno zato, ker varnost in robustnost v praksi pogosto zahtevata hitrejšo operacionalizacijo kot širša standardizacijska harmonizacija.

6.7 Prednostno področje 5: transparentnost, zasebnost, pristranskost in etični artefakti

Peto prednostno področje obsega pripravo podpornih nacionalnih instrumentov za **transparentnost, zasebnost, obravnavo pristranskosti in etično sledljivost zasnove sistema**. Mednarodni IEEE dokumenti na teh področjih že dajejo bogato vsebinsko osnovo, vendar se v praksi kaže potreba po bolj uporabnih profilih in predlogah: katere informacije mora organizacija razkriti uporabnikom, kako dokumentirati privacy-by-design ukrepe, kako strukturirati bias analizo in kako v razvoj vključiti sledljive etične zahteve.

Tudi tukaj je treba rezultat 1.1 referencirati previdno in natančno. Dokument ne vzpostavi še zaključene nacionalne metodologije za ta vprašanja, vendar vključuje več pomembnih gradnikov. V razširjenem seznamu dokumentacije izrecno omenja transparency profile po IEEE 7001, PIA/DPIA ter dokazila o minimizaciji in zakonitosti obdelave osebnih podatkov. V prilogi s PIA predlogo so vključena polja za transparentnost in pravice posameznikov (DSAR), operativne kazalnike, zahteve glede minimizacije, anonimizacije in tehničnih ter organizacijskih kontrol. V "Artefact Passport" za model so posebej vključena polja za metrike pravičnosti po IEEE 7003, transparentnost po IEEE 7001 ter znane omejitve uporabe; pri podatkovnem naboru pa tudi vprašanja pravičnosti. Poleg tega rezultat 1.1 med "hitrimi koraki" izrecno predlaga objavo osnutkov obrazcev in vzpostavitev glosarja ter standardiziranih predlog.

Na tej ravni bi bil kot nacionalni izhod smiseln predvsem:

- **transparency profile** za organizacije,
- vzorčne evidence za privacy-by-design ukrepe,
- osnovna predloga za bias analizo,
- ter pojasnjevalne smernice za vključevanje teh elementov v tehnično dokumentacijo.

To področje je prioritetno zato, ker so ti vidiki izjemno pomembni za legitimnost in sprejemljivost sistemov UI, vendar pogosto ostanejo preveč načelni, če niso prevedeni v konkretne obrazce in procese.

6.8 Prednostno področje 6: interoperabilnostni in infrastrukturni profili

Šesto prednostno področje je priprava dopolnilnih profilov za **interoperabilnost, večorganizacijska okolja, podatkovne prostore in infrastrukturne primere uporabe**. To področje ni v jedru horizontalne AI standardizacije, vendar postaja pomembno tam, kjer sistemi UI delujejo v deljenih podatkovnih in storitvenih okoljih.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Rezultat 2.3 prek analiziranih primerov kaže, da so pravila uporabe, klasifikacija podatkov, odgovornosti uporabniške administracije, logiranje, nadzor vhodov in infrastrukturne odvisnosti med najpomembnejšimi prenosljivimi vzorci. Rezultat 1.1 pa to smer podpira posredno: v priporočilih za integracijo z obstoječimi sistemi vodenja, v razširjenem seznamu dokumentacije in v "Artefact Passport" poudarja povezave na repozitorije, registre modelov, CI/CD, evidence testov ter sledljivost sprememb, kar že nakazuje potrebo po bolj strukturiranem interoperabilnostnem in infrastrukturnem okviru. Vendar je treba poudariti, da rezultat 1.1 na tem področju ne vsebuje še samostojnega ali podrobno razdelanega interoperabilnostnega profila, temveč predvsem uporabne nastavke za nadaljnjo konkretizacijo.

Zaradi tega bi bil kot nacionalni izhod smiseln predvsem:

- osnovni **interoperabilnostni profil** za AI sisteme v večorganizacijskih okoljih,
- minimalne smernice za provenienco, dostop in sledljivost podatkov,
- ter sektorsko prilagodljivi infrastrukturni profili za zahtevnejša okolja.

To področje je smiselno obravnavati kot razvojno, ne kot prvo prioritetno jedro, vendar ga je pomembno vključiti že zdaj, ker odpira prostor za kasnejše specializirane tehnične predloge.

6.9 Operativni sklep poglavja

Na podlagi dosedanje analize je mogoče precej jasno ugotoviti, da slovenski prostor za nadaljnje delo ne potrebuje najprej novih splošnih standardov, temveč predvsem **uporabne nacionalne tehnične dopolnitve**, ki bodo obstoječe mednarodne in evropske standarde naredile neposredneje uporabne v domači praksi. Rezultat 2.2 je namenjen pripravi podlage za nadaljnje korake, medtem ko bodo šele kasnejše faze razvijale tehnični vidik prilagoditev in dokumentacijo tehničnih predlogov.

Kot najbolj prioritetna se na tej ravni kažejo zlasti področja minimalnih artefaktov in dokazil, človeškega nadzora, kakovosti podatkov, baseline varnostnih kontrol ter podpornih profilov za transparentnost, zasebnost in pristranskost. Pri tem je pomembno natančno razumeti vlogo rezultata 1.1: ta že vsebuje uporabne kontrolne sezname, predloge FRIA in DPIA/PIA ter vzorčne obrazce "Artefact Passport", vendar kot neobvezujoče predloge, delovne vzorce in spremljajoče artefakte za sledljivost, ne pa kot že zaključeno nacionalno standardizacijsko rešitev. Prav zato je njegova vrednost v tem dokumentu predvsem v tem, da pokaže smer, v kateri je mogoče obstoječe standarde prevesti v bolj poenotene, pregledne in neposredno uporabne slovenske instrumente.

V vseh navedenih primerih je zato smiselno nadaljnjo nacionalno konkretizacijo razumeti ne kot podvajanje mednarodnih standardov, temveč kot **pripravo profilov, obrazcev, kontrolnih seznamov, minimalnih evidenčnih zahtev in pojasnjevalnih tehničnih**

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

instrumentov, ki jih bo mogoče neposredno vključiti v organizacijsko, presojevalno, pogodbeno in upravljavsko prakso.

7. Predlagani tipi nacionalnih tehničnih izhodov

7.1 Namen

Po opredelitvi prednostnih področij za nadaljnjo nacionalno konkretizacijo je v naslednjem koraku treba odgovoriti še na vprašanje, v kakšni obliki naj se ta konkretizacija sploh izvede. Ta dokument je namenjen oblikovanju strokovne podlage za nadaljnje delo.

Za večino ugotovljenih potreb v tej fazi še ni najbolj smiseln "nov nacionalni standard" v formalnem smislu, temveč predvsem tak tip tehničnega izhoda, ki omogoča hitro, pregledno in operativno uporabo obstoječih mednarodnih in evropskih standardov v slovenskem prostoru. Rezultat 1.1 to logiko že potrjuje.

To pomeni, da je za nadaljnje delo smiselno razlikovati med več tipi možnih nacionalnih izhodov, ki se med seboj razlikujejo po funkciji, stopnji normativnosti in neposredni uporabnosti. Med najpomembnejšimi so zlasti:

- **nacionalni profil uporabe standarda**, kadar je treba obstoječi mednarodni ali evropski standard prevesti v bolj konkreten operativni kontekst;
- **tehnična smernica ali priporočilo**, kadar je treba pojasniti pričakovane minimalne prakse, ne da bi se vzpostavljala formalna nova norma;
- **kontrolni seznam ali minimalni artefaktni paket**, kadar je cilj neposredna uporabnost v organizacijah;
- **predloga obrazca ali dokumentacijski vzorec**, kadar je treba poenotiti način evidentiranja, dokazovanja ali presoje;
- **terminološki glosar ali pojasnjevalni dokument**, kadar je glavni problem neenoten jezik in razhajanje v razumevanju pojmov.

Takšna razdelitev je primernejša od neposrednega skoka v "nacionalne standarde", ker omogoča sorazmernejše in realnejše nadaljevanje projekta. Hkrati je skladna z ugotovitvami rezultata 2.3, ki kaže, da so mnoge najbolj prenosljive dobre prakse pravzaprav povezane z governance, dokumentacijo, logiranjem, omejitvami uporabe, nadzorom, spremembami in operativnimi pravili — torej z elementi, ki jih je pogosto bolje zajeti v profile, smernice, obrazce in kontrolne sezname kot pa takoj v nov formalni standard.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

7.2 Tip izhoda 1: minimalni artefaktni paket in standardizirane predloge dokumentacije

Za prvo prednostno področje — minimalni nabor artefaktov in dokazil — je kot nacionalni tehnični izhod najbolj smiseln **minimalni artefaktni paket**, dopolnjen s standardiziranimi predlogami dokumentacije. To je najbližje že obstoječim rezultatom projekta in hkrati eden najbolj neposredno uporabnih izhodov.

Rezultat 1.1 na tej ravni že vsebuje uporabne gradnike: kontrolne sezname za ključne faze, predloge FRIA in PIA/DPIA ter vzorčne obrazce »Artefact Passport« za modele in podatkovne nabore.

Zato je za nadaljnje delo smiselno predvideti naslednji tip nacionalnega izhoda:

- **osnovni artefaktni paket za sisteme UI**, ki vključuje minimalni nabor obveznih ali priporočenih dokumentov;
- **poenotene predloge** za FRIA, PIA/DPIA, model passport, dataset passport in change log;
- **standardizirano kazalo tehnične dokumentacije**, ki organizacijam pokaže, kako naj povežejo arhitekturo, podatke, tveganja, nadzor, varnost in PMS.

Ta tip izhoda je posebej primeren zato, ker ne podvaja mednarodnih standardov, temveč olajša njihovo uporabo. Hkrati je zelo konkreten: organizacijam, presojevalcem, naročnikom in pravnim oziroma skladnostnim funkcijam omogoča, da delajo z istimi dokumentacijskimi gradniki in s primerljivejšimi dokazili. Prav zato je ta izhod eden najprimernejših kandidatov za prvo fazo nacionalne konkretizacije.

7.3 Tip izhoda 2: nacionalni profil človeškega nadzora in obvladljivosti

Za področje človeškega nadzora in obvladljivosti je kot nacionalni tehnični izhod najprimernejši **nacionalni profil človeškega nadzora**, dopolnjen z vzorčnimi eskalacijskimi in override postopki. Na tej ravni namreč mednarodni standardi in tehnične specifikacije že dajejo dovolj močno osnovo, vendar se v praksi potreba po konkretizaciji pojavi predvsem pri operativni izvedbi.

Rezultat 1.1 v kontrolnih seznamih za fazo pred uvedbo izrecno zahteva dokaze o obvladljivosti oziroma človeškem nadzoru po ISO/IEC TS 8200, v širšem seznamu dokumentacije pa navaja pravila za override in safe-stop, pragove zaupanja in opis

človeškega nadzora. V vzorcu »Artefact Passport« za model je človeški nadzor umeščen kot samostojen dokumentacijski element.

Rezultat 2.3 to operativno potrebo dodatno potrjuje skozi primere dobrih praks, kjer se kot ključni prenosljivi elementi ponavljajo:

- smiselna človeška intervencija,
- omejitev namena uporabe,
- razumljivost izhodov,
- pravila za poseg v delovanje sistema,
- ter povezava med uporabniško odgovornostjo in tehnično podporo sistema.

Na tej podlagi bi bil kot primeren nacionalni izhod smiseln:

- **profil za človeški nadzor po tipih sistemov**,
- **vzorčni override in safe-stop protokoli**,
- **minimalni seznam dokazil o testiranju obvladljivosti**,
- **predloga za dokumentiranje vlog, praga intervencije in usposabljanj**.

Ta tip izhoda je primernejši od formalnega standarda zato, ker omogoča dovolj konkretno uporabo v javnem sektorju, zdravstvu, industriji ali drugih občutljivih okoljih, ne da bi bilo treba najprej razviti popolnoma samostojen normativni dokument.

7.4 Tip izhoda 3: smernica za kakovost podatkov in evidenčni profil za dataset dokumentacijo

Za področje kakovosti podatkov in podatkovnega upravljanja je kot nacionalni tehnični izhod najprimernejša **tehnična smernica**, dopolnjena z evidenčnim profilom za podatkovno dokumentacijo. Mednarodna standarda ISO/IEC 5259-1 in 5259-3 že določata merila in upravljavske zahteve, vendar se v praksi hitro pojavi potreba po poenotenju tega, kaj naj organizacije dejansko merijo, beležijo in hranijo.

Rezultat 1.1 pri kontrolnih seznamih izrecno zahteva merila kakovosti podatkov po ISO/IEC 5259-1 in načrt upravljanja kakovosti podatkov po ISO/IEC 5259-3. V vzorcu »Artefact Passport« za podatkovni nabor pa so vključena polja za kakovostne kazalnike, porazdelitve, neuravnoteženost, označevanje, transformacije, anonimizacijo, dostop in GDPR/PIA povezave.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

Na tej podlagi je kot nacionalni izhod posebej smiseln:

- **smernica za minimalne metrike kakovosti podatkov**,
- **predloga za evidenco izvora, transformacij in sprememb datasetov**,
- **minimalne zahteve za dobavitelje podatkov**,
- ter **poenotena struktura podatkovne dokumentacije** za potrebe notranjih presoj in tehnične dokumentacije.

Ta tip izhoda je zelo uporaben, ker omogoča primerljivejše podatkovne evidence v slovenskem prostoru in hkrati neposredno podpira organizacije, ki nimajo lastnih visoko razvitih podatkovnih metodologij.

7.5 Tip izhoda 4: baseline varnostna smernica in razpisno-pogodbeni profil za AI sisteme

Za področje robustnosti, kibernetske varnosti in odpornosti je najprimernejši nacionalni izhod baseline varnostna smernica, dopolnjena z razpisno-pogodbenim profilom za AI sisteme. ETSI-jeve tehnične specifikacije že določajo dovolj jasen varnostni minimum, vendar je za slovenski prostor posebej pomembno, da se ta minimum prevede v bolj neposredno uporabne kontrolne sezname, dokazila in zahteve do dobaviteljev.

Rezultat 1.1 pri fazi pred uvedbo izrecno zahteva dokaze o robustnosti in kibernetski varnosti po ETSI TS 104 223, v razširjenem seznamu dokumentacije pa posebej navaja rezultate robustnostnih in varnostnih testov. V model passport vzorcu so vključeni tudi robustnostni testi, varnostne kontrole, odvisnosti oziroma SBoM in povezava s PMS, incidenti ter CAPA.

Na tej podlagi bi bil kot nacionalni izhod smiseln predvsem:

- **baseline varnostni seznam za AI sisteme**,
- **predloga dokazil o varnostnih testih, incidentih in rollback postopkih**,
- **minimalni varnostni profil za razpisno in pogodbeno prakso**,
- ter **pojasnjevalni dokument za povezavo med AI security zahtevami in širšim ISMS okoljem**.

Ta tip izhoda je posebej primeren zato, ker se lahko hitro uporabi pri naročanju in presoji sistemov, ne da bi bilo treba čakati na širšo formalno harmonizacijo vseh standardizacijskih tokov.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

7.6 Tip izhoda 5: transparency profile, privacy evidenca in predloga za analizo pristranskosti (bias)

Za področje transparentnosti, zasebnosti, pristranskosti in etične zasnove je kot nacionalni izhod najprimernejša kombinacija **tematskih profilov in predlog dokumentacije**, ne pa en sam "nov standard". To področje je po svoji naravi večplastno in ga je praviloma bolje operacionalizirati prek jasno strukturiranih, a prilagodljivih podpornih instrumentov.

Rezultat 1.1 v širšem seznamu dokumentacije posebej omenja transparency profile po IEEE 7001, PIA/DPIA, dokazila o minimizaciji in zakonitosti obdelave osebnih podatkov ter metrike pravičnosti po IEEE 7003. Ti elementi se pojavijo tudi v vzorčnih obrazcih »Artefact Passport«. Poleg tega rezultat 1.1 med hitrimi koraki izrecno predlaga glosar in standardizirane predloge, kar je posebej relevantno prav na področjih, kjer se pogosto pojavljajo terminološke in procesne neenotnosti.

Na tej podlagi bi bil kot nacionalni izhod smiseln predvsem:

- **transparency profile** za organizacije in sisteme,
- **predloga za dokumentiranje privacy-by-design ukrepov**,
- **osnovna predloga za bias analizo**,
- **kratka pojasnjevalna smernica za etične zahteve v razvojnem ciklu**.

Ta tip izhoda je posebej primeren zato, ker organizacijam omogoča, da razmeroma hitro uvedejo bolj enotne in preverljive prakse na področjih, kjer splošna načela pogosto ostanejo preveč abstraktna.

7.7 Tip izhoda 6: interoperabilnostni profil in sektorske tehnične dopolnitve

Za področje interoperabilnosti, podatkovnih prostorov in infrastrukturnih primerov uporabe je kot nacionalni izhod najprimernejši **interoperabilnostni profil**, po potrebi dopolnjen s **sektorskimi tehničnimi prilogami**. To področje po dosedanjih deliverablih ni v samem jedru horizontalne standardizacije, vendar postaja pomembno tam, kjer sistemi UI delujejo v večorganizacijskih ali infrastrukturno kompleksnih okoljih.

Rezultat 2.3 pri dobrih praksah zelo jasno pokaže, da se v teh okoljih ponavljajo vprašanja klasifikacije podatkov, logiranja, upravljanja vhodov, odgovornosti, varnostnega nadzora in

infrastrukturne discipline. Rezultat 1.1 pa v artefaktnih vzorcih in priporočilih nakazuje pomen povezljivosti z repozitoriji, registri modelov, CI/CD okolji in evidencami sprememb.

Zato bi bil kot nacionalni izhod smiseln predvsem:

- **osnovni interoperabilnostni profil za AI sisteme v večorganizacijskih okoljih**,
- **minimalne smernice za provenienco, dostop in sledljivost podatkov**,
- ter **sektorsko prilagodljive tehnične priloge** za javni sektor, zdravstvo, finance ali druge zahtevnejše domene.

Ta tip izhoda je primeren kot razvojna druga stopnja: ni nujno prvi prioritetni rezultat, je pa pomembno, da je že zdaj konceptualno predviden, ker odpira prostor za kasnejše specializirane tehnične predloge.

7.8 Operativni povzetek tipov izhodov

Če se ugotovitve tega poglavja povzamejo operativno, se pokaže razmeroma jasna slika. Za prednostna področja, ki jih je dokument identificiral v prejšnjem poglavju, v tej fazi niso najbolj smiselni novi formalni standardi v strogem pomenu, temveč predvsem:

- **profili uporabe obstoječih standardov**,
- **tehnične smernice**,
- **kontrolni seznami**,
- **predloge dokumentacije in obrazcev**,
- **minimalni artefaktni paketi**,
- **razpisno-pogodbeni profili**,
- ter po potrebi **sektorske priloge**.

Takšna razporeditev je skladna z dejanskimi rezultati projekta. Rezultat 1.1 že vsebuje več delovnih artefaktov in predlog, ki kažejo, da je ravno ta tip uporabnih izhodov najbolj primeren za zgodnjo nacionalno konkretizacijo. Šele kasnejše faze projekta so namenjene tehničnim prilagoditvam nacionalnih standardov in pripadajoči dokumentaciji.

Zato je za nadaljnje delo smiselno izhajati iz naslednje logike:

- kjer obstaja dovolj zrelo mednarodno in evropsko jedro, je treba pripraviti **profil ali smernico**;

- kjer je glavni problem neenotna dokumentacija, je treba pripraviti **predlogo ali artefaktni paket**;
- kjer je glavni problem praktična uporaba pri naročanju, pogodbah ali presojah, je treba pripraviti **razpisno-pogodbeni profil ali kontrolni seznam**;
- kjer pa gre za kompleksna sektorska okolja, je smiselno predvideti **dopolnilne tehnične priloge**.

Ta dokument je namenjen temu da dovolj jasno in konkretno pokaže, **kateri tipi nacionalnih tehničnih izhodov** so glede na obstoječe standarde, projektne ugotovitve in potrebe slovenskega prostora najbolj smiselni za nadaljnje delo.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)

8. Zaključek

To poročilo je bilo pripravljeno z namenom, da iz širšega standardizacijskega prostora umetne inteligence izlušči tiste mednarodne in evropske reference, ki imajo za slovenski prostor največjo praktično, strokovno in izvedbeno vrednost, ter da na tej podlagi opredeli prednostna področja za nadaljnjo nacionalno tehnično konkretizacijo. V tem smislu dokument ne deluje kot nov katalog standardov, temveč kot strokovno utemeljena sinteza med že pripravljenim pregledom regulativnega okvira, obstoječim naborom standardov in operativnimi ugotovitvami iz dobrih praks. Takšna vloga je skladna tudi z dejanskim mestom rezultata 2.2 v programu dela, kjer je umeščen med seznam relevantnih standardov, analizo primerov in kasnejše tehnične prilagoditve nacionalnih standardov.

Izvedena analiza je pokazala, da standardizacijski okvir za umetno inteligenco ni zgrajen okoli enega samega osrednjega standarda, temveč okoli medsebojno povezanega sklopa dokumentov, ki skupaj pokrivajo sisteme vodenja, upravljanje tveganj, terminologijo, arhitekturo, kakovost podatkov, človeški nadzor, robustnost, kibernetsko varnost, transparentnost, zasebnost, obravnavo pristranskosti ter interoperabilnost v kompleksnejših okoljih. Rezultat 1.3 ta ekosistem že prikaže kot komplementarno celoto, rezultat 1.1 pa pokaže, kako se pravne zahteve AI Akta lahko prevajajo v procese, kontrolne sezname, dokumentacijo, FRIA in PIA/DPIA predloge ter spremljajoče artefakte za sledljivost, vključno z vzorci »Artefact Passport« za modele in podatkovne nabore.

Na tej podlagi je mogoče precej jasno ugotoviti, da slovenski prostor v tej fazi ne potrebuje najprej novih splošnih standardov v formalnem smislu, temveč predvsem bolj neposredno uporabne nacionalne tehnične dopolnitve, ki bodo obstoječe mednarodne in evropske standarde prevedle v preglednejšo in operativno uporabnejšo prakso. Kot posebej pomembna področja za takšno konkretizacijo so se izkazali minimalni nabor artefaktov in dokazil, človeški nadzor in obvladljivost, kakovost podatkov in podatkovna dokumentacija, baseline varnostni profil za sisteme UI, podporni profili za transparentnost, zasebnost in pristranskost ter interoperabilnostni oziroma infrastrukturni profili za zahtevnejša okolja. Ta področja niso bila opredeljena abstraktno, temveč na podlagi ponavljajočih se operativnih potreb, ki jih potrjujejo tudi analizirani primeri dobrih praks, zlasti governance, klasifikacija podatkov, logiranje, razmejitev odgovornosti, človeški nadzor in upravljanje sprememb.

Pomemben sklep tega poročila je tudi, da so za nadaljnje delo v slovenskem prostoru na tej stopnji praviloma primernejši profili, tehnične smernice, kontrolni seznami, predloge dokumentacije, minimalni artefaktni paketi in pojasnjevalni dokumenti kot pa neposreden razvoj povsem novih nacionalnih standardov. Takšna usmeritev je sorazmerna dejanskemu dosegu rezultata 2.2 in hkrati skladna z že razvitimi gradniki iz rezultata 1.1, ki so izrecno

predstavljeni kot uporabni predlogi, delovni vzorci in spremljajoči artefakti za sledljivost, ne kot že zaključena standardizacijska rešitev.

V tem smislu je mogoče zaključiti, da ta deliverable opravlja svojo osrednjo funkcijo: pripravi strokovno in dovolj konkretno podlago za naslednje faze projekta, v katerih bo mogoče iz identificiranih prednostnih področij razvijati bolj določene tehnične predloge, profile, dokumentacijske vzorce in druge podporne instrumente za slovenski prostor. Njegova dodana vrednost ni v tem, da bi že dokončno normativno odločal o vsebini nacionalnih standardov, temveč v tem, da razmeroma jasno pokaže, na katerih področjih, na podlagi katerih referenc in v kakšnih tipih tehničnih izhodov je takšno nadaljnje delo najbolj smiselno nadaljevati.

CRP NOSI-AI (št. V2-24065), ki ga (so)financirata Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost RS (ARIS) in Ministrstvo za digitalno preobrazbo (MDP)