Projekt: CRP V2-24065 – NOSI-AI

Rezultat 1.3: Seznam standardov za preverjanje skladnosti rešitev UI

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

Datum: 30.6.2025

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

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)

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)

Uvod in namen

Spodnji dokument povzema ključne evropske in mednarodne akterje, ki postavljajo pravila, smernice in tehnične temelje za odgovorno uvajanje umetne inteligence. Vsaka od navedenih organizacij prispeva drugačen, a komplementaren del sestavljenke: od normativnih zahtev in merljivih metrik do arhitekturnih vzorcev, podatkovnih praks in upravljavskih procesov. Za Slovenijo je ta ekosistem še posebej pomemben, ker omogoča usklajen prehod od načel do dokazljivih praks skladnosti – kar je bistveno tako za izvajanje evropske Uredbe o umetni inteligenci (AI Akt) kot za konkurenčnost gospodarstva in zaupanje javnosti.

ISO/IEC JTC 1/SC 42 je osrednje globalno telo za standardizacijo AI. Njegovi dokumenti, kot so ISO/IEC 42001 (sistem vodenja za AI), ISO/IEC 23894 (upravljanje tveganj), ISO/IEC 22989 (terminologija), ISO/IEC 23053 (okvir za ML sisteme), ISO/IEC 5259-1/-3 (kakovost podatkov), ISO/IEC TS 8200 (obvladljivost/človeški nadzor) in TR 24028 (zaupanja vredna AI), določajo skupen jezik, procese in dokazila, ki jih lahko neposredno uporabimo pri pripravi tehnične dokumentacije in presoj skladnosti. Ker so ti standardi pogosto prevzeti tudi kot evropske izdaje (EN) in nacionalne izdaje SIST, so za slovenske deležnike praktično izhodišče pri razpisih, pogodbah in dialogu z nadzornimi organi.

CEN/CENELEC JTC 21 je evropski forum, ki usklajuje in prevzema mednarodne standarde v evropski (EN) obliki ter pripravlja standarde, relevantne za evropsko zakonodajo. Ko so standardi objavljeni kot harmonizirani, omogočajo “domnevo skladnosti” z zahtevami AI Akt. Za Slovenijo to pomeni, da lahko prek SIST uporabljamo enake referenčne dokumente kot druge države članice, s tem pa zmanjšamo regulatorno negotovost in olajšamo čezmejno poslovanje.

ETSI (zlasti skupini SAI in ENI) zagotavlja tehnične specifikacije in poročila za kibernetsko varnost, robustnost in uporabo AI v telekomunikacijskih omrežjih. Dokumenti, kot so ETSI TS 104 223 (osnovne varnostne zahteve za AI sisteme) in 104 050 (ontologija groženj), pretvorijo zahteve robustnosti in varnosti v preskusne in operativne kontrole. Za slovenske ponudnike kritičnih storitev (telekomunikacije, energetika, mobilnost) je to neposredna podpora pri presojah in integracijah, kjer je zanesljivost ključna.

IEEE SA s serijo 7000 dopolnjuje tehnično plat z etičnimi, družbenimi in uporabniškimi vidiki. Standardi, kot so IEEE 7000 (vključevanje etičnih zahtev v zasnovo), 7001 (transparentnost), 7002 (zasebnost) in 7003 (pristranskost), pomagajo prevesti načela v operativne procese, vloge, evidence in merila. To je posebej pomembno za slovenska podjetja in javne institucije, kjer je treba pokazati, da sistemi AI varujejo pravice posameznikov in omogočajo razumljivo rabo.

ITU-T (npr. priporočila Y.3172, Y.3174, Y.Sup55, F.748.32, Y.3176, F.748.45) prinaša domensko natančnost za omrežja in multimedijo: referenčne arhitekture, ravnanje s podatki v omrežjih, konkretne primere rabe, zahteve za detekcijske module in sledljivo integracijo modelov. Ta priporočila so javno dostopna in zato izjemno uporabna kot referenca v tehničnih specifikacijah, preskusnih načrtih in operativnih postopkih slovenskih ponudnikov komunikacijskih in digitalnih storitev.

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)

BDVA/DAIRO in DSBA/DSSC krepita “ekosistemske” temelje: strategijo (SRIDA), interpretativne usmeritve (Position Paper), ter referenčne arhitekture in profile za podatkovne prostore (Data Spaces Blueprint v2.0, Technical Convergence v2.0). Podatkovni prostori, semantika, identitete, politike deljenja in revizijske evidence so v praksi tisto, kar omogoča zakonito, sledljivo in kakovostno rabo podatkov v večorganizacijskem okolju – torej pogoje, ki jih AI Akt neposredno zahteva. Za Slovenijo to pomeni jasnejšo pot do interoperabilnih rešitev, evropskih pilotov in dostopa do referenčnih podatkov ter preskusnih okolij.

Ključno vlogo ima tudi SIST kot slovenski nacionalni organ za standarde. Prek SIST je zagotovljen dostop do evropskih (EN) in mednarodnih (ISO/IEC, ETSI, IEEE) izdaj, nacionalne objave in usklajena terminologija. To omogoča, da slovenske organizacije enako in pravočasno sklicujejo standarde v razpisih, pogodbah, certifikacijah in notranjih pravilnikih, ter da se učinkovito pripravljajo na presoje skladnosti po AI Akt.

Skupaj ta telesa in dokumenti ustvarjajo skladen okvir: ISO/IEC in CEN/CENELEC postavita hrbtenico upravljanja, tveganj, terminologije in arhitekture; ETSI in ITU-T dodata podrobne tehnične in varnostne zahteve za omrežja in storitve; IEEE SA opredeli preglednost, zasebnost in pristranskost na način, ki ga je mogoče uvesti v razvojne procese; BDVA/DAIRO in DSBA/DSSC uredita podatkovne tokove in infrastrukturo na ravni ekosistema. Za slovensko uvajanje UI to pomeni hitrejši prehod od načel do preverljivih dokazil: jasne procese, merljive metrike in sledljive evidence, ki podpirajo inovacije, zmanjšujejo operativna tveganja in krepijo zaupanje uporabnikov ter regulatorjev.

Metodologija in kriteriji vključitve

Standardi so razvrščeni glede na obseg (mednarodni, EU/EN, nacionalni/SIST) ter filtrirani po vsebinskih kriterijih, ki podpirajo skladnost: upravljanje in QMS, upravljanje tveganj, kakovost podatkov, preglednost in razkritja, človeški nadzor/obvladljivost, natančnost/robustnost/kibernetska varnost, ter dokumentacija in sledljivost. Kjer je bilo na voljo, so upoštevane opombe o relevantnosti za AI Akt in za visoko-tvegane sisteme.

ISO/IEC JTC 1/SC 42 za skladno in varno rabo UI: upravljanje, tveganja, arhitektura, podatki in človeški nadzor

V tem sklopu se naslanjamo na nabor ključnih dokumentov ISO/IEC JTC 1/SC 42, ki neposredno uokvirjajo vodenje, tveganja, arhitekturo, podatke in nadzor nad avtomatiziranimi sistemi UI: ISO/IEC 42001 (AIMS), ISO/IEC 23894 (upravljanje tveganj), ISO/IEC 22989 (terminologija), ISO/IEC 23053 (okvir za ML sisteme), ISO/IEC TR 24028 (zaupanja vredna UI), ISO/IEC 5259-1 in 5259-3 (kakovost podatkov) ter ISO/IEC TS 8200 (obvladljivost/človeški nadzor). Gre za dokumente z različnim normativnim statusom: “International Standard” postavlja obvezne zahteve ali priporočila na ravni standarda, “Technical Report (TR)” je pregledna, ne-normativna razlaga konceptov, “Technical Specification (TS)” pa je vmesni, pragmatičen korak z natančnimi smernicami, ki jih je možno neposredno operativizirati. V praksi se to pozna tako, da standard določa “kaj mora biti zagotovljeno”, TS pokaže “kako to smiselno izvesti”, TR pa razjasni ozadje in skupni jezik.

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 Slovenijo imajo ti dokumenti dvojno težo. Prvič, mnogi imajo tudi evropske izdaje (SIST EN ISO/IEC), kar olajša sklicevanje v razpisih, presojah in nadzornih postopkih ter omogoča enoten dialog z drugimi državami članicami. Drugič, skupaj tvorijo celovit upravljavski komplet: 42001 postavi sistem vodenja (QMS) za AI, 23894 prinese metodo upravljanja tveganj čez življenjski cikel, 22989 zagotovi skupni slovar, 23053 strukturira opis arhitekture in artefaktov, TR 24028 prevaja načela "zaupanja vredne UI" v razumljive lastnosti, 5259-1/-3 data-ekipam dasta jezik in procese za kakovost podatkov, TS 8200 pa konkretizira mehanizme človeškega nadzora in obvladljivosti. Ko jih uporabimo skupaj, dobimo sledljiv, preverljiv in ponovljiv način dela, s katerim lahko organizacija hitro pokaže, da tehnične prakse in pravne obveznosti delujejo usklajeno.

Povezava z AI Akt je neposredna. Standard 42001 je organizacijski "hrbtenjačnik" zahtev iz člena 17 (QMS) in praviloma drži niti proti členu 11 (tehnična dokumentacija), členu 9 (tveganja) ter členu 15 (robustnost/kibernetska varnost, kjer se nasloni na druge norme). 23894 je praktičen odgovor na člen 9: opiše, kako določimo kontekst, identificiramo nevarnosti (tudi tiste, ki so specifične za UI), izberemo in spremljamo omilitve, ter kako vse to evidenčno vzdržujemo. 22989 je temelj za člena 11 in 13, saj brez enotne terminologije hitro nastanejo nesporazumi v dokumentaciji in navodilih za uporabnike. 23053 podpira član 11 z jasnimi opisi meja sistema, podatkovnih tokov, komponent in kontrolnih točk; hkrati ponudi zasnovo, kam vpeti varnostne in nadzorne funkcije iz člena 15 in člena 14. TR 24028 ne nalaga zahtev, a razjasni, katere lastnosti zaupanja (robustnost, varnost, preglednost, pravičnost, zasebnost, odgovornost) morajo biti prepoznavno prisotne, da bo kasnejše dokazovanje skladnosti sploh možno. 5259-1 in 5259-3 sta neposreden prevod člena 10 v prakso (merila kakovosti in organizacijsko upravljanje s podatki), TS 8200 pa operacionalizira člen 14: kakšne vloge, vmesnike, alarmne pragove, "override" in "safe-stop" potrebujemo ter kako to testirati in dokumentirati.

Vloga SIST je pri tem zelo praktična: evropske oziroma nacionalne izdaje (npr. SIST EN ISO/IEC 23894:2024, SIST EN ISO/IEC 23053:2023, SIST EN ISO/IEC 22989:2023) so za slovenske uporabnike najlažji način sklicevanja, nabave in skladenjega dela s priglašenimi organi. Tam, kjer evropska izdaja še ni sprejeta, je smiselno slediti ISO/IEC verzijam in spremljati, kdaj bo CEN/CENELEC izdal EN-prevzem, da se bo vzpostavila tudi domneva skladnosti, ko in če bo tak standard kasneje harmoniziran na ravni EU.

Če vse skupaj postavimo v operativno sliko: 42001 v organizaciji definira politike, odgovornosti, presoje in CAPA; 23894 napolni register tveganj in cikle pregledov; 22989 poskrbi, da vsa dokumentacija "govori isti jezik"; 23053 prinese diagrame sistema, artefakte in "gates"; TR 24028 je kompas, po katerem ocenjujemo, ali sistem res deluje "zaupanja vredno"; 5259-1/-3 standardizirata, kako merimo in upravljamo kakovost podatkov; TS 8200 pa zagotovi, da ljudje lahko učinkovito nadzirajo AI, ga po potrebi varno zaustavijo in to tudi dokazljivo vadijo. Za slovenske deležnike – od javne uprave in kritične infrastrukture do zdravstva, financ in industrije – je to najhitrejša pot od načel do dokazov: z navzkrižnim sklicevanjem med navedenimi dokumenti zgradimo celoto, ki je hkrati razumljiva ne-AI strokovnjakom in sprejemljiva za presoje v EU prostoru.

ISO/IEC 42001:2023 – Artificial intelligence — Management system (AIMS)

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42

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)

Organizacija	ISO/IEC
Dokument	ISO/IEC 42001:2023 – Artificial intelligence — Management system (AIMS)
Leto	2023
Tip	International Standard
Uradni URL	https://www.iso.org/standard/42001
Javno dostopen	Ne (povzetek javno)
Cena	CHF 199 (ISO Store)
SIST oznaka	ISO/IEC 42001:2023 (nakup prek SIST e-trgovine)
SIST prodajalec	https://members.sist.si/norm/documentsearch.aspx
AI Akt (zakaj)	Vzpostavitev in upravljanje sistema vodenja za AI (QMS); podpira skladnost in sledljivost.
High-risk (zakaj)	Podpira zahteve glede QMS za ponudnike HR-AI ter postopke PMS/CE.

Za koga je standard in kaj rešuje

ISO/IEC 42001 je standard upravljanja (angl. management system) posebej prilagojen organizacijam, ki razvijajo, uvajajo ali uporabljajo rešitve umetne inteligence. Čeprav je nastal v domeni IT, je namenjen širšemu krogu deležnikov: vodstvom podjetij, skrbnikom tveganj, oddelkom kakovosti, skladnosti in notranje revizije, prav tako pa tudi oddelkom za varnost informacij in varstvo podatkov. Standard odgovarja na praktično vprašanje: kako zagotoviti, da je AI rešitev načrtovana, izvajana in nadzirana na ponovljiv, sledljiv in preverljiv način – podobno kot to počne ISO 9001 za kakovost ali ISO/IEC 27001 za varnost informacij.

V praksi to pomeni, da 42001 določa, katere politike, procese, vloge in evidence (dokaze) moramo imeti, da lahko verodostojno dokažemo odgovorno rabo AI in skladnost s pravili. Standard ne predpisuje, katero tehnologijo uporabiti, temveč vzpostavi 'ogrodje upravljanja': od politike AI, preko upravljanja sprememb modelov, do notranjih presoj in ukrepov za stalne izboljšave (CAPA).

Kaj vsebuje (ključne zahteve in artefakti)

• Voditeljstvo in politika AI: vodstvo določi namene in meje uporabe AI, načela (npr. preglednost, pravičnost), vloge in pristojnosti. Dokumentirane morajo biti politike, postopki in merila uspešnosti.

• Načrtovanje in ocena tveganj: organizacija sistematično prepoznava tveganja in priložnosti, določi cilje upravljanja AI ter načrt za njihovo doseganje. AIMS se povezuje z metodologijo upravljanja tveganj iz ISO/IEC 23894.

• Podpora in kompetence: zahtevana so ustrezna znanja in usposabljanja, ozaveščanje deležnikov, razpoložljivost virov ter komunikacijski kanali.

• Operativno upravljanje AI: procesi za razvoj, validacijo, objavo, spremljanje in umik AI rešitev; sledljivost artefaktov (podatkovni sklopi, modeli, konfiguracije) in sprememb (change control).

• Vrednotenje uspešnosti: notranje presoje, nadzor kazalnikov (KPI), vodstveni pregledi, obravnava neskladnosti in korektivni ukrepi (CAPA).

• Nenehne izboljšave: strukturiran cikel PDCA (Plan-Do-Check-Akt) – izboljšanje procesov na podlagi najdb presoj, incidentov in povratnih informacij.

Povezava z AI Aktom

EU AI Akt predvideva, da morajo ponudniki visokorizičnih sistemov vzpostaviti sistem vodenja kakovosti (QMS) ter zagotoviti sledljivost, tehnično dokumentacijo, upravljanje tveganj, nadzor po dajanju na trg in

podobno. ISO/IEC 42001 je neposredna podpora tem obveznostim: povezuje voditeljstvo, procese, podatke in nadzor v enoten okvir. Za bralce, ki niso AI specialisti, je smiselno razmišljati o 42001 kot o 'poslovno-organizacijskem lepilu', ki omogoča, da tehnični ukrepi (npr. testiranje robustnosti) in pravne obveznosti (npr. obvestila uporabnikom) delujejo usklajeno in dokazljivo.

V členskih sklopih AI Akt se standard največ navezuje na: Art. 17 (QMS), Art. 11 (tehnična dokumentacija), Art. 9 (upravljanje tveganj) in Art. 15 (robustnost/kibernetska varnost – v povezavi z drugimi standardi).

Posebnosti za visoko-tvegane (HR) sisteme

Pri HR sistemih je dokazljivost ključna: kdo je sprejel odločitev, s katerimi podatki, pod katerimi pogoji, in kako se sistem obnaša v mejnih primerih. 42001 zahteva, da so ti elementi urejeni procesno – to pomeni, da obstajajo jasne odgovornosti, evidence in kontrolne točke pred objavo ter v obdobju uporabe (PMS).

Standard spodbuja integracijo z drugimi normami: npr. varnost informacij (ISO/IEC 27001) za zaščito podatkov in modelov, ter 23894 za metodiko tveganj. Za HR primere je pomembno, da AIMS vključuje merila sprejemljivosti (kdaj je model dovolj dober), protokole za 'stop gumb' in postopke ob spremembah verzij.

Primeri uporabe v praksi

• Vodja kakovosti: vzpostavi register procesov AI, imenuje odgovorne osebe in uvede notranje presoje; ob odkriti neskladnosti se sproži CAPA z roki in dokazili.

• IT/OT vodja: uskladi politike dostopa in varnosti za podatkovne sklade in modele; potrdi postopke za objavo nove verzije modela in za povrnitev (rollback).

• Pravna/s skladnost: vključi zahteve AI Akt v QMS, pripravi standardizirano kazalo tehnične dokumentacije in 'transparency profile' za uporabnike.

• Vodstvo: periodično pregleduje KPI (npr. delež zaključenih presoj, odprte CAPA) in odloča o prioritetah (npr. kje povečati naložbe ali prilagoditi politiko AI).

Metapodatki in dostop

Uradni dokument je na voljo na ISO strani (povzetek javno, celotno besedilo proti plačilu). V Sloveniji je nakup mogoč prek SIST e-trgovine. Cena ISO izdaje je orientacijska (CHF 199). Za evropsko uskladitev (EN) bo treba spremljati sprejem/nadomestitve pri CEN/CENELEC/SIST.

ISO/IEC 23894:2023 – Artificial intelligence — Risk management

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC
Dokument	ISO/IEC 23894:2023 – Artificial intelligence — Risk management
Leto	2023
Tip	International Standard
Uradni URL	https://www.iso.org/standard/55645.html

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)

Javno dostopen	Ne (povzetek javno)
Cena	CHF 132 (ISO Store)
SIST oznaka	SIST EN ISO/IEC 23894:2024
SIST prodajalec	https://www.sist.si/image/catalog/DOWNLOAD/Izvlecki/Objave/2024/Mese%C4%8Dni%20seznam%20SIST%202024-05.pdf
AI Akt (zakaj)	Metodologija celovitega upravljanja tveganj za AI skozi življenjski cikel.
High-risk (zakaj)	Neposredno podpira zahteve za sistematično upravljanje tveganj (identifikacija, ocena, ublažitve).

Zakaj je upravljanje tveganj pri AI posebno

AI sistemi se učijo iz podatkov in njihov izhod je odvisen od konteksta uporabe. To pomeni, da tveganja niso le klasične 'IT napake', temveč tudi napačne odločitve ali pristranskosti, ki imajo lahko vpliv na ljudi (npr. nepravična obravnava pri zaposlovanju) ali varnost (npr. napačna zaznava v industriji). ISO/IEC 23894 ponuja strukturiran način, kako ta tveganja prepoznati, oceniti in obvladovati skozi celoten življenjski cikel rešitve – od načrtovanja do umika.

Standard je primerljiv z okvirji, ki jih poznajo tudi ne-AI specialisti (npr. ISO 31000), vendar vsebuje AI-specifične primere: upravljanje podatkovnih tveganj, modelnih napak, drift-a podatkov, adversarnih napadov in človeškega nadzora.

Glavne faze in artefakti

• Označitev konteksta: opredelitev namena sistema, deležnikov, omejitev in meril uspeha; to je osnova za razumno oceno tveganj.

• Identifikacija tveganj: sistematičen popis možnih škodljivih dogodkov – od kakovosti podatkov do kibernetskih groženj in operativnih napak.

• Analiza in vrednotenje: ocena verjetnosti in vpliva; določitev meril sprejemljivosti ter prioritet za ukrepanje.

• Zdravila (mitigation) in kontrole: izbira ukrepov – tehničnih (npr. robustnostni testi, kontrole dostopa) in organizacijskih (npr. procesi odobritve).

• Spremljanje in pregled: periodična ponovna ocena (npr. ob spremembi modela ali podatkov), merjenje učinkovitosti mitigacij in incidentni proces.

• Sledljivost in evidence: dokumenti, ki jih potrebujemo za dokaz (register tveganj, matrika odločitev, 'assumption log').

Povezava z AI Akt in drugimi standardi

AI Akt (Art. 9) zahteva sistematično upravljanje tveganj. ISO/IEC 23894 je 'kako to narediti' – določa korake, vloge in evidence. Povezuje se z ISO/IEC 42001 (QMS), saj tveganja postanejo del poslovnega cikla PDCA, ter z ETSI TS 104 223 (kibernetska varnost) in ISO/IEC 5259 (kakovost podatkov).

Za ne-AI strokovnjake: 23894 je analogija oceni tveganj, ki jo poznate iz varnosti pri delu ali informacijske varnosti – le da je vsebina prilagojena učnim modelom, podatkovnim tokovom in odločitvam, ki jih AI samodejno generira.

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)

Posebnosti za HR sisteme

• Strožja merila sprejemljivosti: določite jasne prage (npr. najnižjo sprejemljivo točnost po podskupinah).

• Preskusi robustnosti in bias: načrtni testi za najhujše scenarije (worst-case) in analize po podskupinah.

• Post-market monitoring: spremljanje drift-a in incidentov; ponovna ocena ob vsaki večji spremembi podatkov ali modela.

• Vključitev človeškega nadzora: definirana pravila za 'override' in 'safe-stop'.

Primeri uporabe v praksi

• Banka: pri kreditnem ocenjevanju se vzpostavi register tveganj, kjer so jasno navedeni viri pristranskosti, pragi za odobritev in postopki za človeški pregled v mejnih primerih.

• Industrija: pri vizualnem nadzoru kakovosti se določijo testi na motečih vzorcih (svetloba, rotacija, obraba), redni 'drift check' in plan za re-trening ob preseženih pragih.

• Zdravstvo: pri triaži se definira nabor varnostnih metrik (npr. 'false-negative' prag), protokol za eskalacijo zdravniku in redno poročanje etičnemu odboru.

Metapodatki in dostop

Dokument je na voljo pri ISO (povzetek javno, celotno besedilo plačljivo) ter kot evropska izdaja "SIST EN ISO/IEC 23894:202" . Slovenske informacije in nakup so navedeni v mesečnih seznamih SIST; priporočljivo je preveriti tudi, ali obstaja prevod/ nacionalne prilagoditve.

ISO/IEC 22989:2022 – Artificial intelligence — Concepts and terminology

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC
Dokument	ISO/IEC 22989:2022 – Artificial intelligence — Concepts and terminology
Leto	2022
Tip	International Standard
Uradni URL	https://www.iso.org/standard/74296.html
Javno dostopen	Da (brezplačen prenos)
Cena	CHF 0
SIST oznaka	SIST EN ISO/IEC 22989:2023
SIST prodajalec	https://www.sist.si/Sporocila/Sporocila_2_2023/files/basic-html/page22.html
AI Akt (zakaj)	Skupna terminologija; olajša tehnično dokumentacijo, ocenjevanje in komunikacijo.
High-risk (zakaj)	Zagotavlja konsistentne pojme v tehnični dokumentaciji in presoji skladnosti HR-AI.

Zakaj je terminologija ključna

V reguliranih okoljih dvoumnost povzroča neskladja in napačne odločitve. ISO/IEC 22989 to naslavlja tako, da jasno opredeli pojme, kot so 'model', 'algoritem', 'usposabljanje', 'inferenčni čas', 'podatkovni sklop za

validacijo', 'pojasnljivost', 'pristranskost' itd. S tem omogoča, da projektne ekipe, vodstva, pravniki in presojevalci govorijo isti jezik.

Standard služi kot 'slovar in priročnik' hkrati: poleg definicij ponuja razlage in odnose med pojmi, kar olajša pripravo dokumentacije, zahtev in pogodb.

Kaj dokument vsebuje

• Strukturiran nabor definicij z normativnimi sklici na druge standarde SC 42 in širše (npr. podatkovne in varnostne norme).

• Pojasnila razlik med sorodnimi pojmi (npr. 'poštenost' vs. 'pristranskost', 'zanesljivost' vs. 'robustnost').

• Priporočila za dosledno rabo terminov v dokumentaciji, uporabniških navodilih in komunikaciji s strankami/regulatorji.

• Križne sklice na terminologijo v drugih serijah (npr. ISO/IEC 23053, 23894), kar olajša integracijo v večje okvire.

Povezava z AI Akt in korist za ne-AI strokovnjake

AI Akt zahteva jasna navodila, tehnično dokumentacijo in informacije za uporabnike. Dosledna terminologija je predpogoj, da so ta gradiva razumljiva in enoznačna. V presoji skladnosti terminološke nejasnosti sicer vodijo do napačnih interpretacij in zavrnitev.

Za vodstva in pravnike 22989 predstavlja 'standardni slovar', na katerega se lahko sklicujejo v internem pravilniku ali pogodbah. To zmanjša tveganje nesporazumov in pospeši komunikacijo z nadzornimi organi.

Posebnosti za HR sisteme

V HR sistemih so pojmi, kot so 'človeški nadzor', 'sledljivost', 'evidence', 'validacija po podskupinah', kritični. 22989 omogoča, da organizacija te pojme dosledno definira in uporablja v celotni tehnični dokumentaciji, manualih in poročilih o testiranju.

Primeri uporabe v praksi

• Tehnična dokumentacija: uvedba standardiziranih definicij na začetku dokumenta in sklici skozi celotno poročilo.

• Pogodbe in SLA: natančne definicije, kaj pomeni 'natančnost', 'razpoložljivost' ali 'incident AI'.

• Navodila za uporabnike: poenoteni opisi omejitev sistema in 'transparency profile' (skladno z IEEE 7001).

• Notranje izobraževanje: delovni list s ključnimi pojmi, ki ga uporabljajo vse ekipe.

Metapodatki in dostop

Dokument je brezplačno dostopen (PDF) na strani ISO – kar je redkost med standardi in dodatna prednost za širšo uporabo. Na voljo je tudi evropska izdaja **SIST EN ISO/IEC 22989:2023**, uporabna za sklicevanje v regionalnih procesih.

ISO/IEC 23053:2022 – Framework for AI systems using machine learning

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC

Dokument	ISO/IEC 23053:2022 – Framework for AI systems using machine learning
Leto	2022
Tip	International Standard
Uradni URL	https://www.iso.org/standard/74438.html
Javno dostopen	Ne (povzetek javno)
Cena	CHF 177 (ISO Store)
SIST oznaka	SIST EN ISO/IEC 23053:2023
SIST prodajalec	https://www.sist.si/Sporocila/Sporocila_2_2023/files/basic-html/page22.html
AI Akt (zakaj)	Okvir za komponente/življenjski cikel ML sistemov; podpira tehnično dokumentacijo.
High-risk (zakaj)	Olajša razmejitev meja sistema, podkomponent in nadzora za oceno tveganj HR-AI.

Zakaj potrebujemo arhitekturni okvir za ML

AI rešitve niso zgolj modeli; sestavljajo jih podatkovne cevi (pipeline), orodja za učenje, storitve za uvajanje, nadzorni sistemi in operativne procedure. ISO/IEC 23053 poda skupni 'zemljevid' teh komponent in odnosov med njimi, da lahko različne ekipe (IT, razvoj, varnost, pravna služba) koordinirano delujejo in dokumentirajo rešitve na primerljiv način.

Za ne-AI strokovnjake je 23053 uporaben kot arhitekturna referenca: kako je sistem sestavljen, kje tečejo podatki, kdo je odgovoren za posamezne dele, in kje se izvajajo kontrole (npr. pred objavo ali pri spremembah).

Ključni elementi okvira

• Komponente: zbiranje/ priprava podatkov, učenje, validacija, uvajanje (deployment), spremljanje in umik; podporne storitve (katalogi modelov, registri).

• Artefakti: podatkovni sklopi (train/validation/test), modelni artefakti, konfiguracije, dokumenti (npr. datasheet, eval poročila).

• Procesi: pretok od ideje do uporabe; mejniki za odobritev; postopki za spremembe in obvladovanje incidentov.

• Vloge in odgovornosti: kdo odobri objavo, kdo upravlja dostop do podatkov, kdo skrbi za spremljanje (PMS).

• Integracije z drugimi normami: povezava z 23894 (tveganja), 42001 (QMS), 5259 (kakovost podatkov) in ETSI (varnost/robustnost).

Povezava z AI Akt in praktične koristi

AI Akt pričakuje tehnično dokumentacijo, ki jasno opredeli meje sistema, podatkovne tokove, testne postopke in nadzorne mehanizme. ISO/IEC 23053 je 'predloga arhitekture', ki te zahteve prevede v razumljive diagrame, sezname artefaktov in kontrolne točke.

V presoji skladnosti se pogosto zahteva dokaz, kje nastajajo podatki, kako so urejene različice modela in kdo odobrava spremembe. 23053 omogoča, da ta dokaz pripravimo konsistentno in pregledno.

Posebnosti za HR sisteme

• Jasna meja sistema: natančno določene vhodne/izhodne točke in odgovornosti, kar je ključno za varnost in odgovornost.

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)

• Sledljivost: vezava modela na dataset in konfiguracijo (kdo, kdaj, s čim je treniral), kar omogoča rekonstrukcijo odločitev.

• Kontrolne točke: formalni 'gates' za odobritev pred objavo, spremembami in ob večjih incidentih.

• Operativna pripravljenost: načrt spremljanja (PMS), alarmiranje in postopki 'safe-stop'/rollback.

Primeri uporabe v praksi

• Javno upravo zanima, kako sistem sprejema odločitve: 23053 nudi diagrame in opise, ki pokažejo, kje so človeški pregledi in kako je zagotovljena sledljivost.

• Industrija: pri proizvodnem nadzoru 23053 pomaga določiti, kateri del sistema je 'kritičen' in zahteva dodatne kontrole ter testiranja pred uvajanjem.

• Fintech: jasno razmeji dele, ki potrebujejo poglobljeno varnostno presojo (npr. API za odločanje), in dele, ki so manj tvegani.

Metapodatki in dostop

Dokument je na voljo pri ISO (plačljivo) in kot evropska izdaja "SIST EN ISO/IEC 23053:2023". Za potrebe Observatorija je priporočljivo hraniti povezave na uradne strani in nacionalne prodajalce (SIST).

ISO/IEC TR 24028:2020 – Overview of trustworthiness in AI

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC
Dokument	ISO/IEC TR 24028:2020 – Overview of trustworthiness in AI
Leto	2020
Tip	Technical Report (TR) – pregled in smernice
Uradni URL	https://www.iso.org/standard/77608.html
Javno dostopen	Ne (povzetek javno)
Cena	CHF 177 (ISO Store)
SIST oznaka	—
SIST prodajalec	https://members.sist.si/norm/documentsearch.aspx
AI Akt (zakaj)	Pojmi/atributi zaupanja vredne AI (robustnost, varnost, preglednost).
High-risk (zakaj)	Napeljuje na načela za izpolnitev zahtev o robustnosti, nadzoru in kibernetski varnosti HR-AI.

Namen in komu je namenjen

TR 24028 je pregledni tehnični dokument, ki opredeli, kaj pomeni 'zaupanja vredna AI'. Ne predpisuje obveznih zahtev, temveč sistematično opisuje lastnosti, ki jih morajo rešitve izpolnjevati, da jim lahko uporabniki, regulatorji in družba zaupajo. Dokument je razumljiv tudi ne-AI strokovnjakom, saj izhaja iz splošnih načel, kot so zanesljivost, varnost, preglednost, zasebnost in odgovornost.

Uporabniki: vodstva organizacij, pravniki in skrbniki skladnosti, varnostni strokovnjaki, produktni vodje, kakovost in tudi razvijalci. TR 24028 je pogosto 'uvodna referenca', na katero se sklicujejo drugi standardi (npr. 23894, 23053, ETSI TS 104 223).

Kaj dokument vsebuje (osnovni atributi zaupanja)

• Robustnost in zanesljivost: sistem dosledno deluje v pričakovanih pogojih in predvidljivo v mejnih primerih; ob napakah se obnaša varno.

• Varnost in kibernetska varnost: zaščita pred zlonamernimi posegi (npr. zastrupljanje podatkov, 'model extrAktion'), zaščita podatkov in modelov.

• Preglednost in razložljivost: razumljiv opis delovanja, omejitev in negotovosti; navodila, kaj lahko in česa ne.

• Pravičnost in pristranskost: prepoznava in zmanjševanje sistemskih nepravičnosti; spremljanje razlik med skupinami uporabnikov.

• Zasebnost in varstvo podatkov: spoštovanje pravnih podlag, minimizacija podatkov, sledljivost obdelav.

• Odgovornost in upravljavska načela: jasne vloge, procesi, evidence in možnost revizije odločitev.

Povezava z AI Aktom

AI Akt ne uporablja vedno enake terminologije, vendar zahteve glede robustnosti, kibernetske varnosti, človeškega nadzora, preglednosti in dokumentacije neposredno odražajo atribute iz TR 24028. Zato je dokument uporaben kot 'most' med splošnimi načeli in konkretno izvedbo: pokaže, katere lastnosti naj ima sistem, da bomo kasneje lažje dokazali skladnost po čl. 9, 10, 13, 14 in 15.

Za vodstva in pravnike: 24028 pomaga oblikovati politiko 'zaupanja vredne AI', ki jo nato QMS (ISO/IEC 42001) prevede v procese in kontrole.

Posebnosti za HR sisteme

• Zahtevana višja stopnja dokazljivosti: robustnost (odpornost na napake in zlonamerne posege), varnost (kontrole dostopa, integriteta modelov), preglednost (jasna navodila in opozorila) ter človeški nadzor (možnost zaustavitve/override).

• Povezava na PMS: atributi zaupanja zahtevajo stalno spremljanje – npr. trend poslabšanja točnosti, porast incidentov, napadi na model. TR 24028 je koristen seznam 'kaj spremljati' v fazi uporabe.

Primeri uporabe v praksi

• Politika zaupanja vredne AI: organizacija sprejme okvir (na osnovi 24028), ki določa načela, vloge, kazalnike in obvezne preglede pred objavo.

• Razvoj produkta: produktni vodja pripravi 'transparency profile' in jasno opiše omejitve; varnostna ekipa definira scenarije napadov, ki jih bo preverjala.

• Odnosi z javnostmi in regulatorjem: dosledna, ne-tehnična razlaga, zakaj sistemu lahko zaupamo in kako je to zavarovano v praksi.

Metapodatki in dostop

TR 24028 je na voljo pri ISO (povzetek javno, celotno besedilo plačljivo). V Sloveniji je mogoče iskanje/nakup preko SIST portala.

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)

ISO/IEC 5259-1:2024 – Data quality for analytics and ML — Part 1

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC
Dokument	ISO/IEC 5259-1:2024 – Data quality for analytics and ML — Part 1
Leto	2024
Tip	International Standard
Uradni URL	https://webstore.iec.ch/en/publication/96735
Javno dostopen	Ne
Cena	CHF 132 (IEC Webstore)
SIST oznaka	— (ISO/IEC izdaja)
SIST prodajalec	https://members.sist.si/norm/documentsearch.aspx
AI Akt (zakaj)	Merila in pojmi za kakovost podatkov (treniranje/validacija/test).
High-risk (zakaj)	Praktična opora za izpolnitev obveznosti glede podatkov in podatkovnega upravljanja HR-AI.

Namen in osnovne definicije

5259-1 vzpostavi temeljno terminologijo in merila za 'kakovost podatkov' v analitiki in stroj-nem učenju. Kakovost podatkov ni enodimenzionalna: vključuje popolnost (manjkajoče vrednosti), točnost, skladnost, pravočasnost, enoličnost, reprezentativnost in še druge lastnosti. Dokument opredeli, kako o teh lastnostih razmišljamo in jih merimo na podatkovnih sklopih za učenje, validacijo in test.

Za ne-AI strokovnjake: kot pri proizvodnji tudi pri AI 'surovina' (podatki) določa kakovost izdelka (model). 5259-1 je standardni jezik za oceno te surovine.

Kaj konkretno določa del 1

• Pojme in taksonomijo kakovosti podatkov: nabor atributov in način, kako jih razmejimo (npr. popolnost vs. točnost).

• Tipične vrste napak: neuravnoteženost razredov, šum v oznakah, neustrezna porazdelitev, pomanjkanje določenih primerov (edge cases).

• Osnovne metode merjenja: stopnja manjkajočih vrednosti, razmerja razredov, stabilnost porazdelitev; kako interpretirati rezultate.

• Povezavo na življenjski cikel: od zbiranja do priprave in pregledov pred učenjem ter ponovne ocene po spremembah.

Povezava z AI Akt

AI Akt v členu 10 zahteva ustrezen nadzor nad podatki: kakovost, ustreznost in dokumentacija. 5259-1 omogoča, da ti pojmi niso le načelni, temveč operativni – s skupnim naborom meril in metod ocenjevanja, ki jih lahko organizacija uvede v politiko in prakse.

Posebnosti za HR sisteme

• Dosledno merjenje po podskupinah: pri HR sistemih je ključno, da kakovost podatkov preverjamo ločeno za relevantne skupine (npr. starost, regija).

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)

• Pragi sprejemljivosti in odobritev: definicija ‘go/no-go’ meril za objavo modela (npr. min. pokritost redkih primerov).

• Povezava s PMS: redno spremljanje drift-a porazdelitev (npr. PSI/KS) in ponovna ocena kakovosti po pomembnih spremembah virov.

Primeri uporabe v praksi

• ‘Data quality checklist’: obvezni pregledi pred učenjem (manjkajoče, neuravnoteženost, anomalije), s shranjenimi poročili kot evidence.

• Javno naročanje ali dobaviteljski odnos: zahteva po skladnosti z 5259-1 kot merilo minimuma kakovosti vhodnih podatkov.

Metapodatki in dostop

Dokument je na voljo v IEC Webstore. Priporočljivo je spremljati nacionalne objave (SIST), če bo izdan EN sprejem.

ISO/IEC 5259-3:2024 – Data quality for analytics and ML — Part 3: Management requirements

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC
Dokument	ISO/IEC 5259-3:2024 – Data quality for analytics and ML — Part 3: Management requirements
Leto	2024
Tip	International Standard
Uradni URL	https://www.iso.org/standard/81092.html
Javno dostopen	Ne
Cena	CHF 155 (ISO Store)
SIST oznaka	— (ISO/IEC izdaja)
SIST prodajalec	https://members.sist.si/norm/documentsearch.aspx
AI Akt (zakaj)	Procesne zahteve za upravljanje kakovosti podatkov v AI projektih.
High-risk (zakaj)	Dokazljiv nadzor nad podatki podpira skladnost HR-AI (npr. ustreznost/ločljivost/napake).

Namen: iz konceptov v upravljanje

Medtem ko 5259-1 opredeljuje ‘kaj merimo’, 5259-3 pove ‘kako to upravljamo’ v organizaciji. Določa procese, vloge, dokumente in nadzorne točke za sistematično obvladovanje kakovosti podatkov, od pogodb z dobavitelji do revizijske sledljivosti sprememb.

Glavne procesne zahteve

• Politike in odgovornosti: formalna politika kakovosti podatkov, imenovani lastniki podatkov, odgovorne osebe za odobritev 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)

• Načrtovanje in ocenjevanje: plan kakovosti podatkov za vsak projekt AI; določeni KPI in pragi; periodično poročanje vodstvu.

• Operativne kontrole: pregledi pred učnim ciklom (pre-training checks), kontrole označevanja, validacije referenčnih resnic (ground truth).

• Sledljivost in dokumentacija: evidence o izvoru podatkov, transformacijah, dovoljenjih; revizijski dnevnik sprememb datasetov.

• Obvladovanje incidentov: proces za prijavo napak v podatkih, popravljalne ukrepe in učenje na podlagi incidentov.

Povezava z AI Aktom in Quality Management Sistemi

Z vidika AI Akt (čl. 10 in 11) 5259-3 zagotavlja 'organizacijski dokaz': ne le da znamo izmeriti kakovost, temveč da jo upravljamo v ponovljivem procesu. Učinkovito se poveže z ISO/IEC 42001 (QMS) – kazalniki, notranje presoje, CAPA – in s tem tvori del celotnega sistema vodenja AI.

Posebnosti za HR sisteme

• Eskalacijske poti: jasna pravila, kdaj je potrebna ustavitev objave ali rollback modela zaradi kakovosti podatkov.

• Segmentni KPI: spremljanje kakovosti posebej za ranljive skupine ali kritične funkcije.

• Pogodbeni mehanizmi: zahteve do dobaviteljev podatkov (npr. SLA za kakovost in prenos evidenc).

Primeri uporabe v praksi

• Razpisna dokumentacija: kot pogoj za dobavitelje podatkov določimo skladnost z 5259-3 in minimalni nabor evidenc.

• Notranji 'data steward' program: določimo vloge, usposabljanja in redne preglede datasetov; poročila so del PMS.

Metapodatki in dostop

Na voljo pri ISO in IEC (plačljivo); spremljajte nacionalne prodajalce (SIST).

ISO/IEC TS 8200:2024 – Controllability of automated AI systems

Polje	Vrednost
Kategorija	ISO/IEC JTC 1/SC 42
Organizacija	ISO/IEC
Dokument	ISO/IEC TS 8200:2024 – Controllability of automated AI systems
Leto	2024
Tip	Technical Specification (TS)
Uradni URL	https://www.iso.org/standard/83012.html
Javno dostopen	Ne
Cena	CHF 155 (ISO Store)

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)

SIST oznaka	— (ISO/IEC izdaja)
SIST prodajalec	https://members.sist.si/norm/documentsearch.aspx
AI Akt (zakaj)	Načela obvladljivosti in posegov v delovanje; povezano s človeškim nadzorom.
High-risk (zakaj)	Podpira vzpostavitev učinkovitih ukrepov človeškega nadzora in varnostnih stikal.

Zakaj 'obvladljivost' in ne le 'preglednost'

Obvladljivost pomeni, da lahko ljudje sistem AI učinkovito nadzorujejo in, ko je treba, posežejo v njegovo delovanje – preprečijo škodo ali zmanjšajo tveganje. TS 8200 se osredotoča na praktične mehanizme: od 'safe-stop' in 'override' do zasnove vmesnikov, ki omogočajo pravočasne in informirane odločitve.

Kaj določa TS 8200 (ključni gradniki)

• Vloge in odgovornosti: kdo nadzoruje, kdo odobrava posege, kakšne kompetence so potrebne (usposabljanje).

• Načela zasnove: 'human-in-the-loop/on-the-loop' glede na kritičnost naloge; načela ergonomije in kognitivne obremenitve operaterjev.

• Operativni mehanizmi nadzora: safe-stop, ročni override, eskalacije; protokoli za testiranje teh mehanizmov.

• Nadzorni podatki in alarmiranje: katere informacije mora sistem posredovati operaterju (zaupanje, negotovost, stanje modela), pragovi in prioritetizacija alarmov.

• Preizkusi obvladljivosti: scenariji, ki preverjajo odzivnost in zanesljivost nadzornih funkcij v praksi (tudi v 'worst-case' pogojih).

Povezava z AI Aktom

Člen 14 AI Akta zahteva učinkovite ukrepe človeškega nadzora. TS 8200 je neposredni 'priročnik' za implementacijo: pomaga določiti, katere nadzorne funkcije so primerne, kako jih načrtovati, testirati in dokumentirati ter kako to vključiti v QMS in PMS.

Posebnosti za HR sisteme

• Merila za 'ustavitev': vnaprej določeni pogoji, kdaj mora sistem preiti v varno stanje (npr. nizko zaupanje modela, kritičen kontekst).

• Vaja odziva (drills): periodični preizkusi ekip, z merjenjem MTTR (čas do varne zaustavitve) in 'override success rate'.

• Dokumentacija za presojo: opis vmesnikov, dokazila o testih, usposabljanja operaterjev, matrika odgovornosti.

Primeri uporabe v praksi

• Klicni center z AI asistenco: operater lahko začasno izključi samodejne predloge, ko zazna neprimernost; sistem beleži razloge in izboljša modele.

• Industrijski vizualni nadzor: ob nenavadnih vzorcih sistem sam predlaga zaustavitev linije; operater potrdi ali zavrne s pojasnilom.

• Javni sektor: pri avtomatiziranih odločitvah je zagotovljen 'human-in-the-loop' za mejne primere in možnost pritožbe.

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)

Metapodatki in dostop

TS 8200 je na voljo na ISO (plačljivo). V nacionalnih okoljih (SIST) spremljajte objave o morebitnih sprejemih/prevodih.

CEN/CENELEC (EN) standardi

V slovenskem okolju (SIST) imajo evropske izdaje standardov (EN ISO/IEC 23894, EN ISO/IEC 22989, EN ISO/IEC 23053) praktično prednost pred zgolj mednarodnimi ISO/IEC različicami, ker neposredno "živijo" v evropskem regulativnem in nadzornem okviru ter se lažje vključijo v javne politike in postopke preverjanja skladnosti.

Možna domneva skladnosti (harmonizacija): ko/če bo kateri izmed EN standardov objavljen kot harmoniziran v UL EU, njegova uporaba omogoča domnevo skladnosti z ustreznimi členi AI Akt — to močno olajša dokazovanje skladnosti, še posebej pri HR sistemih.

Nacionalna veljava prek SIST: EN izdaje se prenašajo kot SIST EN (npr. SIST EN ISO/IEC 23894:2024), zato jih slovenski organi in priglašeni organi neposredno poznajo in pričakujejo; hkrati so dostopnejše za nabavo in sklicevanje v postopkih.

Javna naročila in nadzor: razpisi in presoje v EU pogosto zahtevajo EN sklice; uporaba teh treh EN standardov zmanjšuje tveganje zavrnitev zaradi »neustreznega sklica« (ISO vs. EN) in poenostavi komunikacijo z nadzornimi organi.

Operativno jedro skladnosti: ti trije EN pokrijejo tri nosilne "plasti" skladnosti AI:

- EN ISO/IEC 23894 – upravljanje tveganj (Art. 9),
- EN ISO/IEC 22989 – skupna terminologija (temelj za teh. dokumentacijo in Art. 11/13),
- EN ISO/IEC 23053 – arhitekturni okvir in sledljivost (Art. 11/14/15). Skupaj tvorijo hrbtenico za tehnično dokumentacijo, procese in dokazila.

Usklajenost po EU in čezmejna sprejetost: EN izdaje zagotavljajo, da so slovenska pravila, prakse in poročila primerljiva in sprejeta v celotni EU — koristno pri čezmejnih projektih, auditih in tržni inšpekciji.

V nadaljevanju sledi opis posameznih CEN/CENELEC standardov (EN ISO/IEC 23894, EN ISO/IEC 22989, EN ISO/IEC 23053) z vidika njihove vloge pri skladnosti z AI Akt in v HR kontekstih.

EN ISO/IEC 23894:2024 – AI risk management (EU izdaja)

Polje	Vrednost
Kategorija	CEN/CENELEC JTC 21
Organizacija	CEN/CENELEC (EN izdaja)
Dokument	EN ISO/IEC 23894:2024 – Artificial intelligence — Risk management (EU)

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)

Leto	2024
Tip	EN (evropski standard)
Uradni URL	https://shop.bsigroup.com/products/information-technology-artificial-intelligence-risk-management/standard
Javno dostopen	Ne (plačljivo)
Cena	Po prodajalcu (BSI/EVS/SIS/SIST ipd.)
SIST oznaka	SIST EN ISO/IEC 23894:2024
SIST (informacija)	https://www.sist.si/image/catalog/DOWNLOAD/Izvlecki/Objave/2024/Mese%C4%8Dni%20seznam%20SIST%202024-05.pdf
Relevanca (AI Akt)	EU izdaja metodologije za upravljanje tveganj; potencialno za harmonizacijo.
Relevanca (HR)	Ko/če bo harmoniziran standard, lahko omogoča domnevo skladnosti za čl. 9.

Namen in kontekst EU izdaje

Evropska izdaja 23894 prinaša isto jedrno vsebino kot izvirni ISO/IEC standard, a omogoča sklicevanje na EN različico, ki je v EU postopkih pogosto zahtevana ali vsaj pričakovana. To je pomembno v postopkih javnega naročanja, pri nacionalnih nadzornih organih ter v razmerju s priglašenimi organi, kjer EN dokumenti predstavljajo 'skupni jezik' za presojo. EU izdaja olajša tudi prevode in nacionalne prilagoditve (SIST EN), kar znižuje prag uporabe v praksi.

Vsebina standarda obsega celoten ciklus upravljanja tveganj: od opisa konteksta in deležnikov, prek sistematične identifikacije tveganj, do njihove analize (verjetnost × vpliv), obravnave (mitigacije), spremljanja ter poročanja. Posebna teža je na sledljivosti odločitev: vsaka pomembna odločitev pri tveganjih mora imeti evidenco (kdo, kdaj, na podlagi česa), da je presoja pregledna in ponovljiva.

Pomen za različne deležnike

Za vodstva in oddelke kakovosti je 23894 upravljavski dokument: namesto tehničnih podrobnosti o modelih govori o postopkih, odgovornostih in kontrolnih točkah. V vsakdanjem jeziku: uvede 'mapo' tveganj za AI projekte, s cilji, pragi in ukrepi, ki jih je mogoče preveriti v presoji.

Za pravne in skladnostne službe standard ponuja podlago za vključitev obveznih pregledov v notranje politike: npr. 'brez opravljene analize tveganj in podpisanega načrta mitigacij se model ne sme objaviti'. Za varnost in IT pomeni integracijo z obstoječimi procesi (incidenti, spremembe, dostopi).

Povezava z AI Aktom in harmonizacija

AI Akt (čl. 9) zahteva, da je upravljanje tveganj iterativno, čez celotni življenjski cikel. 23894 to operacionalizira s ciklom načrtovanja, izvajanja, preverjanja in ukrepanja (PDCA). Če bo EN 23894 uvrščen med harmonizirane standarde (objava v UL EU), bo njegova uporaba ustvarjala domnevo skladnosti: organizacija bo z dokazili po tem standardu lažje dokazala, da izpolnjuje zahteve člena 9.

Pomembno: domneva skladnosti ne nadomesti razmisleka. Vloga standarda je, da zagotovi preverjen 'recept', ki ga priglašeni organi in regulatorji prepoznajo.

Posebnosti za HR sisteme

• Merila sprejemljivosti: določite kvantitativne in kvalitativne prage, kdaj je model sprejemljiv (tudi po podskupinah).

• ‘Worst-case’ scenariji: poleg povprečne točnosti preverite skrajne primere in robne pogoje (slabe vhodne slike, nenavadni vnosi).

• PMS: nadzor po dajanju na trg – spremljanje drift-a, incidentov, degradacije metrik; sprožilci za ponovno učenje ali ‘rollback’.

• Vloge: jasno opredeljeni lastniki tveganj, odločevalci in odobritve sprememb (change control).

Primeri uporabe

• Kadrovski postopek: pri AI selekciji kandidatov je potrebno izdelati register tveganj (npr. pristranskost po spolu/starosti), določiti meritve po podskupinah in uvesti ‘človeški pregled’ za mejne primere.

• Industrija: pri vizualnem nadzoru kakovosti je potrebno planirati teste robustnosti (svetloba, rotacija, umazanija), določitev prage za alarm in protokol za zaustavitev.

• Finančne storitve: za model ocene kreditnega tveganja je potrebno definirate pragove, kdaj mora odločitev potrditi človek, in shranite razloge pri ‘override’.

Dostop in nakup

Standard je na voljo pri evropskih prodajalcih (BSI, EVS, SIS, SIST). Priporočljivo je preveriti nacionalno izdajo SIST EN zaradi lažje nabave in morebitnih jezikovnih prilagoditev.

EN ISO/IEC 22989:2023 – AI concepts and terminology (EU)

Polje	Vrednost
Kategorija	CEN/CENELEC JTC 21
Organizacija	CEN/CENELEC (EN izdaja)
Dokument	EN ISO/IEC 22989:2023 – Artificial intelligence — Concepts and terminology (EU)
Leto	2023
Tip	EN (evropski standard)
Uradni URL	https://www.sis.se/en/produkter/information-technology-office-machines/information-technology-general-topics/350204/sis-en-isoiec-229892023/
Javno dostopen	Ne (plačljivo)
Cena	Po prodajalcu
SIST oznaka	SIST EN ISO/IEC 22989:2023
SIST (informacija)	https://www.sist.si/Sporocila/Sporocila_2_2023/files/basic-html/page22.html
Relevanca (AI Akt)	Terminologija EU izdaje; podpira skladnost in razlago tehničnih zahtev.
Relevanca (HR)	Olajša pripravo tehnične dokumentacije in komunikacijo z nadzornimi organi.

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)

Zakaj je EU terminologija ključna

Če se deležniki ne strinjajo o pomenih besed, hitro pride do zastojev. EN 22989 omogoča, da se celotna organizacija – od pravnikov do razvojnikov – sklicuje na isti 'slovar'. To je še posebej pomembno v sporih ali pri razpisih, kjer natančnost izrazov pomeni manj nejasnosti in manj pravnih tveganj.

V praksi standard pomaga pri pripravi priročnikov, razpisov, internih politik in zahtevnikov: vsi se lahko sklicujejo na iste definicije in razmerja pojmov.

Kaj vključuje (operativni vidik)

• Normativne definicije: pojmi, ki jih lahko dobesedno citirate v dokumentih (npr. 'model', 'usposabljanje', 'inferenčni čas').

• Razmerja med pojmi: razlaga sorodnosti (npr. robustnost vs. zanesljivost; poštenost vs. pristranskost).

• Navodila za uporabo: priporočila, kako dosledno uporabljati izraze v tehnični dokumentaciji, navodilih za uporabnike in pogodbah.

Povezava z AI Akt

AI Akt terja razumljive informacije za uporabnike (čl. 13) in strukturirano tehnično dokumentacijo (čl. 11). EN 22989 je temelj, na katerem ta gradiva nastanejo: če je terminologija enotna, je manj prostora za napačne interpretacije in ugotovitve neskladnosti.

Posebnosti za HR sisteme

Pri HR sistemih so nekateri pojmi še posebej občutljivi (npr. 'človeški nadzor', 'pristranskost', 'sledljivost'). EN 22989 omogoča, da so ti izrazi konzistentno uporabljeni v vseh poročilih, protokolih in zabeležkah o odločanju – kar bistveno olajša presojo skladnosti.

Primeri uporabe

• Glosar v tehnični dokumentaciji: prva priloga z izbranimi definicijami in sklici na EN 22989.

• Razpisi in pogodbe: standardizirane definicije meril ('natančnost', 'incident AI', 'razpoložljivost') zmanjšajo spore.

• Komunikacija z regulatorji: sklicevanje na EN terminologijo pospeši razumevanje brez dodatnih pojasnil.

Dostop in nakup

EN izdaja je dostopna pri evropskih prodajalcih (SIS, BSI, EVS, SIST). Nacionalna izdaja SIST EN je primerna za sklicevanje v Sloveniji.

EN ISO/IEC 23053:2023 – Framework for AI systems using ML (EU)

Polje	Vrednost
Kategorija	CEN/CENELEC JTC 21
Organizacija	CEN/CENELEC (EN izdaja)
Dokument	EN ISO/IEC 23053:2023 – Framework for AI systems using machine learning (EU)
Leto	2023
Tip	EN (evropski standard)
Uradni URL	https://www.evs.ee/en/evs-en-iso-iec-23053-2023
Javno dostopen	Ne (plačljivo)
Cena	Po prodajalcu
SIST oznaka	SIST EN ISO/IEC 23053:2023

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)

SIST (informacija)	https://www.sist.si/Sporocila/Sporocila_2_2023/files/basic-html/page22.html
Relevanca (AI Akt)	Okvir EU izdaje za strukturo AI/ML sistemov; podpira preskušanje in dokumentiranje.
Relevanca (HR)	Poveča sledljivost komponent in kontrol (zajem, učenje, validacija, nadzor).

Zakaj je arhitekturni okvir 'EN' koristen

Arhitektura AI/ML sistemov vključuje več ekip in orodij. EN 23053 omogoča, da nastane enoten opis, razumljiv v pravnem, tehničnem in poslovnem kontekstu. To zmanjša nesporazume pri presojah in pri komunikaciji z dobavitelji, saj vsi gledajo na sistem z iste 'karte'.

Kaj standard zahteva in kako to izgleda v dokumentaciji

• Komponente: podatkovni tokovi, učenje, validacija, uvajanje, spremljanje; podpora (registri, katalogi, varnostne storitve).

• Artefakti: datasheeti, evaluacijska poročila, modeli in konfiguracije s sledljivostjo verzij; obvezne 'evidence links' v dokumentaciji.

• Kontrolne točke: formalni 'gates' pred objavo in pri spremembah; kdo odobrava prehode, kateri testi so obvezni, kje se hranijo dokazi.

• Vloge: odgovorne osebe za podatke, modele, varnost, skladnost; matrika odgovornosti (RACI).

Povezava z AI Aktom in HR

• Člen 11: jasen opis arhitekture, meja sistema in sledljivosti; dokazljivost vira podatkov in nastavitev modelov.

• Člen 14: v arhitekturi označene točke človeškega nadzora (npr. 'human-in-the-loop'); povezava s TS 8200 za testiranje obvladljivosti.

• Člen 15: lokacije varnostnih in robustnostnih kontrol (npr. pred napadi, pri spremembah), kar olajša presojo in PMS.

Primeri uporabe

• System Description: diagrami komponent in tokov, opremljeni s sklici na evidence; priloga tehnične dokumentacije pri prijavi HR sistema.

• Change control: obrazec spremembe natančno navede, katere komponente so prizadete, kateri testi so obvezni in kje se shranijo rezultati.

Dostop in nakup

Standard je dostopen pri evropskih prodajalcih (npr. EVS) in kot SIST EN ISO/IEC 23053:2023 v Sloveniji.

ETSI dokumenti

V tem sklopu obravnavamo šest dokumentov organizacije ETSI, ki izhajajo iz delovnih skupin SAI (Securing AI) in ENI (Experiential Networked Intelligence). Gre za dva tipa gradiv. Prvi so **Group Report (GR)**, informativni dokumenti, v katerih delovne skupine zberejo strokovni konsenz o pojmih, grožnjah, scenarijih in dobrih praksah. GR ne postavljajo zavezujočih zahtev, temveč delujejo kot uvod v problematiko in kot referenčni okvir za politike, "threat modeling" in načrtovanje kontrol. Drugi tip so **Technical Specification**

(TS), ki stopijo korak bliže praksi: opredelijo konkretne zahteve, taksonomije, postopke ali kontrolne mehanizme, ki jih je mogoče neposredno uvesti v procese razvoja, preskušanja, uvajanja in nadzora AI sistemov.

Razlika do **standardov** (npr. EN ISO/IEC) je pomembna. EN/ISO/IEC standardi so normativni in, ko so objavljeni kot **harmonizirani**, omogočajo domnevo skladnosti z ustreznimi določbami evropske zakonodaje (npr. AI Akt). ETSI GR in TS praviloma nimajo tega formalnega statusa, vendar prinašajo hitro dostopne, praktične in javno objavljene smernice, s katerimi organizacija pokaže, *kako* zahteve iz zakonodaje in standardov dejansko izvaja. V praksi to pomeni, da z GR poimenujemo probleme in grožnje, s TS pa jih operacionaliziramo v kontrolah, testih in poročilih.

Za Slovenijo so ti dokumenti strateško in operativno pomembni iz več razlogov. Prvič, vsi so javno dostopni in brezplačni, kar olajša uporabo v javnem sektorju, pri malih in srednjih podjetjih ter v raziskovalnih projektih, kot je CRP NOSI. Drugič, delujejo kot most med zahtevami AI Akt (npr. robustnost in kibernetska varnost po členu 15, upravljanje tveganj po členu 9, tehnična dokumentacija po členu 11) in vsakdanjim delom ekip. Grožnje in napadalne vektorje, opredeljene v GR SAI 004, lahko neposredno vnesemo v register tveganj po EN ISO/IEC 23894; temeljne koncepte kibernetske varnosti iz GR SAI 001 pretvorimo v interne politike in postopke v okviru ISO/IEC 42001; vloge strojne opreme, kot jih razloži GR SAI 006, povežemo z zahtevami za varno infrastrukturo in dobavno verigo. TS 104 050 nam da enotno ontologijo groženj, ki omogoča konsistentno označevanje testov in incidentov v orodjih in poročilih, TS 104 223 pa nabor osnovnih varnostnih zahtev, iz katerih sestavimo pred-objavne “gates”, kontrolne sezname in merila uspešnosti v post-market nadzoru. Končno, GR ENI 007 doda domeni telekomunikacij in kritične infrastrukture jasen kontekst uporabe AI v omrežjih, kar olajša oceno tveganj in postavitev nadzornih točk.

Z vidika javnih naročil, nadzora in auditov je prednost tudi v tem, da so dokumenti ETSI dobro prepoznani v evropski telco/ICT skupnosti. V razpisih in pogodbah jih je smiselno navajati kot referenco za poimenovanje groženj, opis kontrol in zahtevana dokazila, še posebej kadar želimo doseči primerljivost in ponovljivost poročanja med dobavitelji. Kjer EN/ISO/IEC standardi določajo “kaj mora biti zagotovljeno”, ETSI dokumenti pomagajo pokazati “kako je bilo to izvedeno” – z jasnimi preslikavami na procese, teste in evidence.

V Observatoriju standardov je sinergija posebej očitna. Ontologija iz TS 104 050 omogoča, da se grožnje in incidenti dosledno označujejo skozi različne projekte in čas, kar poenostavi agregacijo na dashboardih in analizo trendov; zahteve iz TS 104 223 se naravno pretvorijo v kontrolne sezname in metrike, ki jih spremljamo v PMS; povezave na EN ISO/IEC 23894 (upravljanje tveganj), 23053 (arhitektura in sledljivost), 5259-1/-3 (kakovost podatkov) ter TS 8200 (obvladljivost) pa zaprejo krog dokazljivosti.

V obravnavanem svežnju so vključeni naslednji dokumenti: ETSI GR SAI 004 »Securing AI; Problem Statement«, ki opiše izhodišča groženj in napadov; ETSI GR SAI 001 »AI cybersecurity; general concepts«, ki predstavi temeljne koncepte kibernetske varnosti za AI; ETSI GR SAI 006 »Role of hardware in security of AI«, ki obravnava strojno opremo in dobavno verigo; ETSI TS 104 050 »AI Threat Ontology«, ki uvaja standardizirano klasifikacijo groženj; ETSI TS 104 223 »Baseline Cyber Security Requirements for AI Models and Systems«, ki določa osnovne varnostne zahteve; ter ETSI GR ENI 007 »Categories for AI application to networks«, ki sistematizira področja uporabe AI v omrežjih. Skupaj ti dokumenti tvorijo praktičen, javno dostopen in z evropskimi praksami usklajen paket, s katerim slovenske organizacije hitreje prevedejo zahteve AI Akt v operativne kontrole, teste in dokazila.

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)

ETSI GR SAI 004 V1.1.1 – Securing AI; Problem Statement

Polje	Vrednost
Kategorija	ETSI (SAI/ENI)
Organizacija	ETSI
Dokument	ETSI GR SAI 004 V1.1.1 – Securing AI; Problem Statement
Leto	2020
Tip	Group Report (GR)
Uradni URL	https://www.etsi.org/deliver/etsi_gr/SAI/001_099/004/01.01.01_60/gr_SAI004v010101p.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ETSI poročilo)
SIST prodajalec	—
AI Akt (zakaj)	Izhodišča groženj in vektorjev napadov za AI.
High-risk (zakaj)	Podlaga za načrtovanje varnostnih kontrol in oceno robustnosti HR-AI.

Namen in kontekst

SAI 004 je uvodno poročilo, ki 'mapira' problem kibernetske varnosti pri AI: opredeli, zakaj so AI sistemi drugačni od klasičnih IT sistemov, kateri novi vektorji napadov se pojavljajo ter zakaj tradicionalni pristopi (npr. zgolj perimetrska varnost) niso dovolj. Poročilo služi kot temelj za kasnejše tehnične smernice ETSI in je koristno za vodstva, varnostne ekipe in arhitekte rešitev.

Kaj zajema (grožnje in scenariji)

• Viri groženj: zlonamerni akterji (notranji in zunanji), nenamerne napake, napake v podatkovnih tokovih, pomanjkljivosti v orodjih in knjižnicah.

• Napadi na podatke: zastrupljanje učnih podatkov (poisoning), manipulacije pri označevanju, napadi s ponarejenimi vzorci pri uporabi (evasion).

• Napadi na modele: 'model extrAktion', 'model inversion' in 'membership inference' – odvzem znanja, razkrivanje občutljivih informacij ali ugotavljanje, ali je bil določen zapis uporabljen pri učenju.

• Operativni scenariji: napadi na infrastrukturo za učenje/uvajanje, manipulacija konfiguracij in odvisnosti (supply-chain).

Povezava z AI Aktom in standardi

AI Akt zahteva robustnost in kibernetsko varnost (čl. 15). SAI 004 ponuja 'slovar groženj', ki ga lahko neposredno uporabimo v registru tveganj (po EN ISO/IEC 23894) in pri načrtovanju testov robustnosti (povezava z ETSI TS 104 223 / TS 103 909, kjer obstaja).

Posebnosti za HR sisteme

• Izbor prioritetnih groženj glede na vpliv na ljudi (npr. napačna odločitev v zdravstvu ali javni upravi).

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)

• Evidenca scenarijev: za HR sisteme je ključno zapisati, katere napade se preverja in s kakšnimi rezultati; to je del tehnične dokumentacije.

• Povezava s PMS: ponavljajoči se testi in spremljanje indikatorjev (npr. nenadni padec točnosti lahko kaže na evasion ali drift).

Primeri uporabe v praksi

• Threat modeling delavnica: varnostna ekipa skupaj z razvijalci izdela grožnjni model AI po seznamu iz SAI 004 in določi 'abuse cases'.

• Testni načrt: izbor minimalnega sklopa napadov, ki se izvajajo pred objavo modela (npr. evasion na ključnih razredih, poisoning simulacije).

• Dokumentacija: povzetek groženj postane priloga tehnični dokumentaciji in podlaga za odobritev objave.

Dostop in licenciranje

Dokument je prosto dostopen na ETSI spletni strani; uporaben kot referenca v razpisih in notranjih pravilnikih.

ETSI GR SAI 001 V1.2.1 – AI cybersecurity; general concepts

Polje	Vrednost
Kategorija	ETSI (SAI/ENI)
Organizacija	ETSI
Dokument	ETSI GR SAI 001 V1.2.1 – AI cybersecurity; general concepts
Leto	2022
Tip	Group Report (GR)
Uradni URL	https://www.etsi.org/deliver/etsi_gr/SAI/001_099/001/01.02.01_60/gr_sai001v010201p.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ETSI poročilo)
SIST prodajalec	—
AI Akt (zakaj)	Pregled konceptov kibernetske varnosti za AI sisteme.
High-risk (zakaj)	Usmerja na zahteve robustnosti in kibernetske varnosti za HR-AI.

Namen in komu je namenjen

SAI 001 je 'krovni' dokument, ki predstavi osnovne pojme kibernetske varnosti v kontekstu AI. Primeren je za uvod v tematiko za vodstva, pravne/s skladnostne službe ter IT/varnostne ekipe, ki se prvič srečujejo s specifičnostmi AI. Dokument izpostavi, kako AI spreminja napadalno površino in zakaj potrebujemo dodatne kontrole poleg klasičnih varnostnih praks.

Ključne vsebine

• Terminologija in koncepti: kaj pomeni robustnost, zanesljivost, odpornost na napade in kako se razlikujejo od klasičnih IT pojmov.

• Referenčna arhitektura varnosti: pregled točk, kjer so potrebne kontrole – podatki, modeli, orodja, infrastruktura, dostopi in operativni nadzor.

• Metode ocenjevanja: uvod v testiranje evasion/poisoning, preverjanje integritete modelov in spremljanje anomalij pri uporabi.

• Upravljanje tveganj: povezava s procesi po EN ISO/IEC 23894 in QMS po ISO/IEC 42001.

Povezava z AI Aktom

SAI 001 je uporaben kot temeljni vir pri izpolnjevanju čl. 15 (robustnost/kibernetska varnost) in čl. 11 (tehnična dokumentacija). Kot 'katalog konceptov' pomaga strukturirati tehnične kontrole, ki jih nato organizacija operacionalizira v svojih politikah in testnih načrtih.

Posebnosti za HR sisteme

• Minimalni varnostni nabor: zaščita podatkovnih tokov, preverjanje integritete modelov, spremljanje napadov med uporabo in odzivni načrti.

• Sodelovanje ekip: jasna delitev odgovornosti med MLOps, varnostjo in produktom; povezava na incidente in CAPA v QMS.

Primeri uporabe

• Baseline kontrolni seznam: organizacija uvede nabor kontrol, ki se preverjajo pri vsakem AI projektu (npr. preverjanje odvisnosti, SBoM, preskus evasion).

• PMS: spremljanje indikatorjev (npr. nenadna sprememba distribucije vhodov, padec točnosti) kot možen znak napadov.

Dostop

Dokument je prosto dostopen na ETSI spletni strani in je primeren za interni 'onboarding' ekip.

ETSI GR SAI 006 V1.1.1 – Role of hardware in security of AI

Polje	Vrednost
Kategorija	ETSI (SAI/ENI)
Organizacija	ETSI
Dokument	ETSI GR SAI 006 V1.1.1 – Role of hardware in security of AI
Leto	2022
Tip	Group Report (GR)
Uradni URL	https://www.etsi.org/deliver/etsi_gr/SAI/001_099/006/01.01.01_60/gr_SAI006v010101p.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ETSI poročilo)

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)

SIST prodajalec	—
AI Akt (zakaj)	Varnostne značilnosti in grožnje na ravni strojne opreme.
High-risk (zakaj)	Koristno za HR-AI v varnostno kritičnih okoljih (zdravstvo, industrija).

Zakaj je strojna oprema pomembna pri varnosti AI

Medtem ko se večina razprav osredotoča na modele in podatke, je strojna oprema temelj, na katerem teče AI. Pomanjkljivosti mikroarhitektur, kriptopospeševalnikov, GPU/DPU/NPU enot ali varnostnih modulov lahko omogočijo obvode varnostnih kontrol in razkritje občutljivih podatkov. SAI 006 sistematično predstavi vlogo strojne opreme v verigi zaupanja AI sistemov.

Ključne teme in grožnje

• Zaupanje v strojno opremo: 'root of trust', varni zagoni (secure boot), modul TPM/HSM, zaščita ključev in integriteta firmware-a.

• Napadi in pomanjkljivosti: stranski kanali (side-channel), napadi na pomnilnik, injekcije napak (fault injection), ranljivosti v pospeševalnikih.

• SBoM in odvisnosti: pregled komponent in firmware-a; pomen posodobitev in nadzor nad dobavno verigo (supply-chain).

• Izolacija in posredovanje: varne enklave (TEEs), virtualizacija in orkestracija – kako preprečiti uhajanje podatkov med procesi.

Povezava z AI Aktom

AI Akt pričakuje robustnost in kibernetsko varnost tudi na ravni infrastrukture. SAI 006 pomaga določiti, katere strojne lastnosti so potrebne za doseganje teh ciljev, in kako jih dokumentirati ter testirati v tehnični dokumentaciji (čl. 11) in pri robustnosti (čl. 15).

Posebnosti za HR sisteme

• Kritična okolja: zdravstvo, industrija, promet – kjer napaka lahko ogrozi varnost ljudi ali povzroči večjo gospodarsko škodo.

• Politike posodobitev: nadzor nad firmware-om, varne posodobitve in sled revizij; dokazila o skladnosti pospeševalnikov in modulov.

• Preizkusi: vključenost strojnih testov v načrt robustnosti (npr. test izolacije TEEs, scenariji ob okvari GPU).

Primeri uporabe

• Specifikacije infrastrukture: zahteve za 'secure boot', TPM/HSM, TEEs in SBoM kot del naročila za strojno opremo AI.

• Operativni nadzor: spremljanje CVE in vzdrževanje nastavitev varnosti (npr. onemogočanje neuporabljenih funkcij pospeševalnikov).

Dostop

Dokument je brezplačno dostopen na ETSI; primeren je kot priloga varnostnim politikam in razpisom.

ETSI GR SAI 004 V1.1.1 – Securing AI; Problem Statement — dodatki

Dodatna pojasnila: povezave in uporaba v procesih

Povezava z ISO/IEC 27001: kontrola tveganj in obravnava incidentov; SAI 004 dopolni katalog groženj s specifičnimi AI napadi, kar omogoča, da varnostni register zajame scenarije, ki jih klasični ISMS pogosto spregleda (npr. evasion in poisoning).

Povezava z EN ISO/IEC 23894: pri identifikaciji tveganj se grožnje iz SAI 004 preslikajo v register tveganj in povežejo z mitigacijami (npr. dodatni preskusi robustnosti, kontrole nad označevanjem).

Povezava z ISO/IEC 5259-1/-3: ob grožnjah, ki izvirajo iz podatkov, se uvede rutinsko merjenje kakovosti podatkov in sledljivost sprememb (audit trail datasetov), da je mogoče analizirati vpliv incidentov na učenje modelov.

Razširjeni primeri uporabe

Telekomunikacije: pri zaznavanju anomalij v omrežju se v testni fazi uvede obvezni evasion test na vzorcih z nizkim SNR in simuliranimi napadi ‘label-flip’; rezultati in ponovitve testov so del ‘release gates’.

Javna uprava: pri klasifikaciji vlog se definira ‘abuse case’ s ciljnim spreminjanjem vhodnih podatkov; uvede se človeški pregled za vse primere z nizkim zaupanjem modela in razloge za ‘override’ se beleži kot evidence.

Industrija: v vizualnem nadzoru kakovosti se načrtujejo ‘adversarial’ motnje (odsev, prah, delna zakritost); ob preseženih pragih se sproži varno zaustavljanje linije in analiza incidenta.

ETSI GR SAI 001 V1.2.1 – AI cybersecurity; general concepts — dodatki

Dodatna pojasnila: povezave in uporaba v procesih

Povezava z ISO/IEC 42001: koncepti iz SAI 001 se vgradijo v AI-QMS kot obvezne politike (npr. varnost modelov, SBoM za modele in podatkovne pipeline).

Povezava z ETSI TS 104 223 in sorodnimi tehničnimi specifikacijami: SAI 001 daje ‘kaj’ in ‘zakaj’, TS-ji pa ‘kako testirati’ (testne baterije, metrike).

Povezava z DPIA/PIA: pri obdelavi osebnih podatkov se koncepti kibernetske varnosti dopolnijo z ocenami zasebnosti – skupaj tvorijo celovit pogled.

Razširjeni primeri uporabe

Finančne storitve: uvede se ‘baseline’ kontrolni seznam (integriteta modela, spremljanje drift-a, alarmi za nenavadne vzorce); preskusi pred objavo so pogoj za ‘go-live’.

Zdravstvo: za triažne modele se določijo minimalne varnostne nastavitve (TEEs za občutljive podatke, kontrola dostopa do modelnih artefaktov, verifikacija odvisnosti); dokumentira se načrt odziva na incidente.

E-trgovina: pri priporočilnih sistemih se spremlja zlonamerna manipulacija ocen in klikov; uvede se detekcija ‘bot’ aktivnosti in zaščita API-jev.

ETSI GR SAI 006 V1.1.1 – Role of hardware in security of AI — dodatki

Dodatna pojasnila: povezave in uporaba v procesih

Povezava z IEC 62443: strojne varnostne kontrole (secure boot, TEEs) se vgradijo v zahteve za industrijske sisteme in preverijo pri FAT/SAT preskusih.

Povezava z ISO/IEC 27036: upravljanje dobavne verige; zahteva SBoM in dokazila o posodobitvah firmware-a ter preverjanje podpisov proizvajalca.

Povezava z AI PMS: nadzor CVE in stanj varnostnih modulov je del rednega poročanja; preseganje pragov sproži korektivne ukrepe in reteste.

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)

Razširjeni primeri uporabe

Medicinske naprave: zahteva TEEs za obdelavo občutljivih podatkov, 'secure boot' ter kriptografsko zaščito modelnih uteži; predvideni so testi odpornosti na 'fault injection'.

Avtomobilska industrija: preverjanje izolacije med procesi v ECU-jih in trdnost modelov ADAS; načrt za varne posodobitve OTA z možnostjo 'rollback'.

Kritična infrastruktura: za modele, ki nadzorujejo procese, se vključi HSM/TPM in strogo beleženje sprememb firmware-a; redni 'drills' ob izpadu pospeševalnika.

ETSI TS 104 050 V1.1.1 (2025-03) – AI Threat Ontology

Polje	Vrednost
Kategorija	ETSI (SAI)
Organizacija	ETSI
Dokument	ETSI TS 104 050 V1.1.1 (2025-03) – AI Threat Ontology
Leto	2025
Tip	Technical Specification (TS)
Uradni URL	https://www.etsi.org/deliver/etsi_ts/104000_104099/104050/01.01.01_60/ts_104050v010101p.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ETSI specifikacija)
SIST prodajalec	—
AI Akt (zakaj)	Ontologija groženj za AI – pomoč pri enotni klasifikaciji in poročanju.
High-risk (zakaj)	Olajša sistematičen RM in tehnično dokumentacijo o grožnjah HR-AI.

Namen in kontekst

TS 104 050 predstavlja ontologijo groženj za AI: to je standardiziran slovar in taksonomija, s katerimi lahko ekipe dosledno označujejo in poročajo o grožnjah v vseh fazah življenjskega cikla AI (podatki, modeli, orodja, infrastruktura). Ontologija odpravi zmedo okoli izrazov in podvojenih pomenov (npr. različna poimenovanja za isti tip napada), kar olajša sodelovanje med ekipami in zunanjimi presojevalci.

Kaj dokument vsebuje

• Nabor pojmov (entitet) in odnosov med njimi: npr. 'napad evasion' je vrsta 'napada na uporabo modela', ki vpliva na 'robustnost' in je lahko povezan z 'metrikami odkrivanja anomalij'.

• Klasifikacijsko shemo za grožnje: hierarhija in oznake (ID-ji), ki jih je mogoče uporabiti v orodjih (register tveganj, orodja za testiranje).

• Smernice za uporabo pri poročanju: kako grožnje zapisovati, povezovati z izidi testov in incidenti ter jih povezati z mitigacijami.

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 AI Aktom in drugimi standardi

AI Akt (čl. 15) zahteva robustnost in kibernetsko varnost; TS 104 050 omogoča, da evidence o grožnjah in incidentih uporabljajo enotne oznake, ki jih lahko priglašeni organi hitro razumejo in primerjajo. V praksi se povezuje s SAI 004 (seznam groženj in scenarijev) ter SAI 001 (koncepti varnosti) in z EN ISO/IEC 23894 (register tveganj).

Posebnosti za HR sisteme

• Enoten jezik v presoji: ko HR sistem poroča o evasion testih ali incidentih, uporabi označevanje iz ontologije, kar omogoča primerljivost med projekti.

• Avtomatizacija: ontološke oznake se lahko vgradijo v orodja (issue tracker, eval pipeline), zato je poročanje konsistentno in sledljivo.

• 'Traceability': grožnjo lahko povežete z mitigacijo (kontrola/test), metrikami in odgovorno osebo.

Primeri uporabe

• Register tveganj: vsak vnos dobi ID iz TS 104 050; poročila o testih referencirajo ta ID, zato je sledljivost enostavna.

• Incidentni postopki: ob dogodku se grožnja označi po ontologiji; poročila PMS zbirajo trend po kategorijah groženj.

• Javno naročanje: zahteva 'uporabo standardizirane AI Threat Ontology (ETSI TS 104 050)' pri poročanju dobaviteljev.

Dostop

Specifikacija je brezplačno dostopna na ETSI. Priporočljivo: ontološke oznake vključite v predloge poročil in orodja Observatorija.

ETSI TS 104 223 V1.1.1 (2025-04) – Baseline Cyber Security Requirements for AI Models and Systems

Polje	Vrednost
Kategorija	ETSI (SAI)
Organizacija	ETSI
Dokument	ETSI TS 104 223 V1.1.1 (2025-04) – Baseline Cyber Security Requirements for AI Models and Systems
Leto	2025
Tip	Technical Specification (TS)
Uradni URL	https://www.etsi.org/deliver/etsi_ts/104200_104299/104223/01.01.01_60/ts_104223v010101p.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ETSI specifikacija)
SIST prodajalec	—

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)

AI Akt (zakaj)	Osnovne varnostne zahteve za AI sisteme; neposredna navezava na robustnost/kibernetsko varnost.
High-risk (zakaj)	Praktična izpeljava zahtev natančnosti/robustnosti/kibernetske varnosti za HR-AI.

Namen in vloga TS 104 223

TS 104 223 določa osnovni nabor varnostnih zahtev za AI modele in sisteme. Ne govori zgolj o IT varnosti, temveč naslovlja posebnosti AI: zaščito učnih podatkov, integriteto modelov, obrambo pred evasion/poisoning, sledljivost sprememb in obvladovanje 'model supply chain'. Standard je zasnovan kot 'operativni kontrolni seznam' – kaj mora biti implementirano in preverjeno pred objavo ter kako nadzirati med uporabo.

Kaj dokument vsebuje

• Zahteve za podatke: zaščita virov, kontrola označevanja, sledljivost transformacij, preverjanje licenc in dovoljenj za uporabo.

• Zahteve za modele: integriteta in izvor (provenance), kontrola dostopa, verzioniranje, zaščita pred izvlekom (extrAktion) in razkritjem (inversion).

• Zahteve za procese: varno učenje in uvajanje, 'change control', obvezni testi robustnosti, dokumentacija in evidence.

• Zahteve za PMS: spremljanje drift-a, detekcija napadov in incidentni postopki, redni 'drills' za nadzorne funkcije.

Povezava z AI Aktom in drugimi standardi

TS 104 223 je neposredna podpora čl. 15 (robustnost/kibernetska varnost) in podpira čl. 11 (tehnična dokumentacija) ter čl. 9/10 (tveganja/podatki). Dobro se povezuje z EN ISO/IEC 23894 (proces tveganj), ISO/IEC 5259-1/-3 (kakovost podatkov), ISO/IEC TS 8200 (obvladljivost) in EN ISO/IEC 23053 (arhitekturna sledljivost).

Posebnosti za HR sisteme

• 'Minimum mandatory' kontrole za kritične domene (zdravstvo, javna uprava, industrija); možnost razširjenih testov za visoko občutljive primere.

• Dokumentirani pragi za robustnost: v naprej določeni scenariji evasion/poisoning in sprejemljivi izidi; 'no-go' pogoji.

• Zapisniki o nadzoru in incidentih: poročanje z enotnimi oznakami (TS 104 050) ter povezava na CAPA v QMS.

Primeri uporabe

• Pred-objavni 'gate': seznam kontrol po TS 104 223 mora biti 'green' pred 'go-live'; izjemam mora odobriti odgovorna oseba in pripraviti CAPA.

• Dobaviteljske zahteve: v razpisih je zahteva po skladnosti s TS 104 223, vključno z dostavo poročil o testih in opisom kontrol.

• PMS dashboard: redne metrike (drift, napadi, uspešnost override) in revizijska sled sprememb modelov.

Dostop

Specifikacija je brezplačno dostopna na ETSI; priporočilo: v Observatorij vključite matriko 'zahteva → kontrola/test → evidence'.

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)

ETSI GR ENI 007 V1.1.1 – Categories for AI application to networks

Polje	Vrednost
Kategorija	ETSI (ENI)
Organizacija	ETSI
Dokument	ETSI GR ENI 007 V1.1.1 – Categories for AI application to networks
Leto	2019
Tip	Group Report (GR)
Uradni URL	https://www.etsi.org/deliver/etsi_gr/ENI/001_099/007/01.01.01_60/gr_ENI007v010101p.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ETSI poročilo)
SIST prodajalec	—
AI Akt (zakaj)	Kontekst uporabe AI v omrežjih (telco); prispeva k oceni tveganj.
High-risk (zakaj)	Uporabno za HR-AI v omrežnih storitvah/kritični infrastrukturi.

Namen in kontekst

ENI 007 kategorizira načine, kako se AI uporablja v telekomunikacijskih in omrežnih okoljih (npr. vodenje prometa, optimizacija virov, detekcija anomalij). Poročilo je koristno kot kontekstni katalog: pomaga razumeti, kje in kako AI posega v omrežja ter katere so tipične točke tveganja in nadzora.

Kaj dokument vsebuje

• Opis kategorij uporabe: od načrtovanja in optimizacije do samodejnega obratovanja in samoprilagodljivih funkcij (self-organizing networks).

• Vpliv na zahteve: latenca, zanesljivost, varnost, zasebnost; kaj to pomeni za arhitekturo, testiranje in nadzor.

• Primeri scenarijev: npr. optimizacija porabe energije, prediktivno vzdrževanje, upravljanje QoS/QoE.

Povezava z AI Aktom in standardi

Kategorije iz ENI 007 pomagajo pri oceni tveganj (EN ISO/IEC 23894): za vsako kategorijo je mogoče določiti specifične grožnje (po SAI 004/TS 104 050) in zahteve za robustnost (TS 104 223). Pri omrežjih, ki so del kritične infrastrukture, je presoja posebej pomembna.

Posebnosti za HR sisteme

• Kritičnost storitev: vpliv na komunikacije uporabnikov ali varnostne sisteme; potreba po strožjih pragih in hitrih ‘safe-stop’ mehanizmih.

• Skladnost in nadzor: jasne vloge NOC/SOC, evidence sprememb modelov in konfiguracij, ter redne ‘drills’ za incidente.

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)

Primeri uporabe

• Operater: pri detekciji anomalij se uvede evasion test z realističnimi motnjami in scenariji ‘data spoofing’; PMS spremlja false-positive/negative razmerja.

• Kritična infrastruktura: za AI, ki upravlja QoS v javnih storitvah, se določi človeški nadzor in pragi za ročni ‘override’.

• Dobaviteljske zahteve: razpisi zahtevajo skladnost s TS 104 223 in sklic na kategorije ENI 007 za opis funkcionalnega konteksta.

Dostop

Poročilo je brezplačno dostopno; priporočilo: uporabljajte ga kot prilogo za ‘use-case’ opis v tehnični dokumentaciji.

ETSI TS 104 050 V1.1.1 (2025-03) – AI Threat Ontology — dodatna razširjena pojasnila

Zakaj ontologija odpravi zmedo v praksi

V realnih projektih različne ekipe iste grožnje poimenujejo različno (npr. ‘adversarial noise’, ‘evasion’, ‘perturbation attack’). Ontologija uvede enoten nabor oznak z opisi in odnosi, zato lahko tehnična poročila, incidentni zapisi in upravljavska poročila govorijo isti jezik. To neposredno izboljša kakovost evidence v tehnični dokumentaciji in skrajša čas razumevanja pri presojah.

Ontološki ID-ji omogočajo tudi napredno iskanje po zgodovini incidentov in testov: vodje lahko analizirajo trende (npr. porast evasion napadov) in povežejo rezultate z mitigacijami ter odgovornostmi.

Operativna vključitev v orodja in procese

Ontološke oznake vključimo v predloge ‘risk register’ v Observatoriju in v testna poročila. V CI/CD lahko orodje za evalvacijo avtomatsko označi rezultate z ustreznimi grožnjami, kar zagotovi konsistentno poročanje skozi čas in med projekti.

Pri PMS dashboardu se grožnje agregirajo po ontoloških kategorijah; to omogoča ciljno upravljanje (npr. usposabljanje ekipe za najpogostejšo kategorijo).

Primer razširjene uporabe (telekom/omrežja)

Pri detekciji anomalij v omrežju so ‘false positives’ dragi. Ontologija se uporabi za označevanje napadov, ki zavajajo model (evasion), in za beleženje mitigacij, kot so odločitvena pravila in nadzor zaupanja modela; poročila so primerljiva med izdajami sistema.

ETSI TS 104 223 V1.1.1 (2025-04) – Baseline Cyber Security Requirements for AI Models and Systems — dodatna razširjena pojasnila

Povezava ‘zahteva → kontrola → dokazilo’ (end-to-end)

TS 104 223 je zasnovan kot most med regulatornimi zahtevami in konkretnimi ukrepi. Za vsako zahtevo je smiselno določiti merljive metrike in obvezna dokazila (evidence): poročila o robustnosti, dnevniki integritete modelov, SBoM za modele in pipeline, zapisniki o odobritvah sprememb.

Z vidika audit-readiness se vzpostavi ‘pre-release gate’, kjer odgovorna oseba pregleda izpolnjen kontrolni seznam in pripadajoča dokazila. V Observatoriju se to odrazi kot status kontrol (Planned/In Progress/Done) in povezave na evidence.

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)

Razširitev primerov uporabe (zdravstvo in finance)

Zdravstvo: TS 104 223 vodi uvedbo varnostnih kontrol za modele triaže – TEEs za občutljive podatke, preverjanje integritete modelnih artefaktov, najhujši scenariji (worst-case) za robustnost in obvezni 'override' z merjenjem MTTR v drills.

Finance: pri odkrivanju prevar se uvede zaščita pred 'model extrAktion' (rate-limiting, watermarking, monitoring neobičajnih poizvedb) in poenotena poročila o incidentih, povezana z ontološkimi oznakami TS 104 050.

Navezave na ISO/EN

EN ISO/IEC 23894: TS 104 223 napolni korake obravnave tveganj s konkretnimi kontrolami in testi. ISO/IEC 5259-1/-3: definira merjenje in upravljanje kakovosti podatkov kot predpogoj robustnosti. EN ISO/IEC 23053: določa, kje v arhitekturi se izvajajo kontrole (gate točke) in kako zagotoviti sledljivost artefaktov.

ETSI GR ENI 007 V1.1.1 – Categories for AI application to networks — dodatna razširjena pojasnila

Kako ENI 007 pomaga pri razvrstitvi tveganj

Z razvrstitvijo scenarijev v omrežjih (npr. samodejna optimizacija, odpravljanje nepravilnosti, upravljanje QoS) lahko organizacija vnaprej določi tipične grožnje, poslovne vplive in zahteve po nadzoru. To olajša pripravo 'risk appetite' in odločanje o tem, kje uvesti dodatne kontrole.

Usklajevanje med NOC/SOC in MLOps postane preprostejše, saj kategorije omogočajo skupni pogled na funkcije in odgovornosti.

Razširjeni primeri (kritična infrastruktura)

Pri AI, ki vpliva na razpoložljivost javnih storitev, so ključni 'safe-stop' pogoji in jasni 'fallback' mehanizmi; ENI 007 pomaga opisati operativni kontekst, medtem ko TS 104 223 določa konkretne kontrole in testi robustnosti pred objavo.

Pri oblačnih 'edge' arhitekturah se doda zahteve za izolacijo in sinhronizacijo modelov ter za poročanje incidentov v realnem času.

IEEE SA 7000 – Etika & Transparentnost

Paket IEEE SA (7000-serija) združuje vodila, standarde in priporočene prakse, ki prevajajo etična načela, zasebnost, pristranskosti, transparentnost in vplive na dobrobit v konkretne, izvedljive korake pri zasnovi, razvoju, uvajanju in nadzoru sistemov umetne inteligence. V nasprotju z "čistimi" tehničnimi standardi (npr. arhitektura ali metrike robustnosti) ta sklop naslavlja človeški vidik UI: kako zagotavljati preglednost in informacije za uporabnike, kako sistematično obravnavati etične zahteve in vplive na temeljne pravice, kako dokazovati skrb za zasebnost ter kako meriti družbene učinke, ki jih uvajanje UI prinaša.

Dokumenti imajo različen normativni status. IEEE standardi (7000, 7001, 7002, 7003) so normativni in opisujejo postopke ali zahteve, ki jih je mogoče neposredno vgraditi v organizacijske procese in kontrolne sezname. IEEE 7010 je priporočena praksa (Recommended PrAktice), ki daje strukturiran pristop k merjenju vplivov na dobrobit. Ethically Aligned Design – First Edition (EAD1e) je obsežno vodilo (white paper), ki nudi

konceptualni kompas in nabor praktičnih smernic, čeprav samo po sebi ni normativni standard. V praksi to pomeni, da EAD1e opredeli načela in cilje, standardi pa pokažejo kako jih pretvoriti v konkretne postopke, evidence in merila.

Za Slovenijo je sklop IEEE 7000 posebej uporaben iz treh razlogov. Prvič, večina dokumentov je javno dostopna prek programa IEEE GET, zato so takoj uporabni v javnem sektorju, MSP in raziskovalnih projektih brez dodatnih stroškov. Drugič, odlično dopolnjujejo evropske in mednarodne standarde (EN ISO/IEC 23894, 23053, 42001, 5259): IEEE dokumenti zagotovijo operativne korake in predloge, ki jih lahko neposredno vključimo v QMS, upravljanje tveganj, tehnično dokumentacijo in post-market nadzor. Tretjič, vsebine so utemeljene na praksi (zdravstvo, javna uprava, finance, industrija), zato jih je mogoče hitro vključiti v javna naročila, pogodbe in interne politike kot referenčne zahteve za preglednost, pristranskosti, zasebnost ali vplive.

Povezava z AI Akt je jasna in neposredna. Člen 13 (informacije in navodila za uporabnike) ter člen 11 (tehnična dokumentacija) dobita operativno podporo v IEEE 7001 – Transparency of Autonomous Systems, ki uvaja ravni transparentnosti in določa, katere informacije zagotoviti različnim deležnikom ter kako slediti odločitvam. Člen 14 (človeški nadzor) in člen 9 (upravljanje tveganj) sta podprta s IEEE 7000 – Model Process for Addressing Ethical Concerns, ki etične zahteve pretvori v sledljive korake, evidence in "gates" v razvojnih ciklih. Člen 10 (upravljanje podatkov) dopolnjuje IEEE 7002 – Data Privacy Process z metodiko PIA, minimizacijo podatkov in evidencami obdelav. Vplive na temeljne pravice in pravičnost naslavlja IEEE 7003 – Algorithmic Bias Considerations z izborom metrik, validacijo po podskupinah in tehnikami ublažitve. Nazadnje IEEE 7010 – Well-being Metrics for A/IS pomaga meriti širše družbene in uporabniške učinke, kar je dragoceno pri utemeljevanju sorazmernosti in pri poročanju v post-market nadzoru. Skupaj ti dokumenti tvorijo most med načeli in dokazljivimi praksami, zato so za slovenske organizacije najhitrejša pot, kako skladnost AI Akt podpreti z merljivimi koraki, kontrolami in evidencami.

V tem sklopu so zajeti naslednji dokumenti in vloge: EAD1e (konceptualni kompas etike in transparentnosti), IEEE 7000 (procesno vodenje etičnih zahtev in sledljivost odločitev), IEEE 7001 (ravni transparentnosti in informacije za uporabnike/operaterje/revizorje), IEEE 7002 (proces zasebnosti in PIA kot del upravljanja podatkov), IEEE 7003 (metodologije obravnave pristranskosti in vplivov na temeljne pravice) ter IEEE 7010 (kazalniki dobrobiti in spremljanje učinkov rabe UI). V navezavi na SIST/EN ISO/IEC standarde ti dokumenti zagotovijo celovit, javno dostopen in operativen okvir, s katerim lahko slovenske institucije – od javne uprave in kritične infrastrukture do zdravstva, financ in industrije – dosledno prevedejo zahteve AI Akt v vsakdanje delo, nadzor in poročanje.

IEEE SA – Ethically Aligned Design (EAD1e)

Polje	Vrednost
Kategorija	IEEE SA (7000-serija)
Organizacija	IEEE SA
Dokument	Ethically Aligned Design – First Edition (EAD1e)
Leto	2019
Tip	Guidance (White Paper)
Uradni URL	https://ethicsinAktion.ieee.org/
Javno dostopen	Da (brezplačno)
Cena	Brezplačno

SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Načela etičnosti/transparentnosti; dopolnjuje zahteve o nadzoru in informacijah za uporabnike.
High-risk (zakaj)	Podpira pripravo navodil za uporabo in nadzorne ukrepe pri HR-AI.

Namen in kontekst

EAD1e je obsežen smerni dokument IEEE, ki sistematično predstavi etična načela in priporočila za zasnovo in uporabo inteligentnih in avtonomnih sistemov. Ni normativni standard, temveč 'kompas', ki ekipam pomaga razmišljati o vplivih tehnologije na ljudi in družbo ter to prevesti v praktične smernice.

Kaj dokument vsebuje

EAD1e obravnava teme, kot so človeški nadzor, preglednost, odgovornost, zasebnost, varnost, pravičnost in vključevanje deležnikov. Za vsako temo nudi razlago, zakaj je pomembna, ter niza priporočila, kako ta načela prenesti v procese razvoja, testiranja, uvajanja in komuniciranja z uporabniki.

Povezava z AI Akt in drugimi standardi

EAD1e je koristen 'most' med visokoravninskimi načeli in normativnimi zahtevami AI Akt. Pomaga pri oblikovanju politik, 'transparency profile' in navodil za uporabnike (čl. 13), podpira razmislek o človeškem nadzoru (čl. 14) in o obvladovanju tveganj (čl. 9). Uporaben je v kombinaciji z ISO/IEC 42001, ISO/IEC 23894 in ISO/IEC TS 8200.

Posebnosti za HR sisteme

Pri visoko-tveganih sistemih EAD1e pomaga že v fazi opredelitve namena uporabe, prizadetih deležnikov in možnih negativnih vplivov. Na podlagi tega lahko ekipe načrtujejo mehanizme človeškega nadzora, opozorila in navodila za varno rabo ter postopke za obvladovanje incidentov.

Primeri uporabe

V javni upravi se EAD1e uporabi kot referenca za oblikovanje načel 'zaupanja vredne UI' in za pripravo pravilnika o transparentnosti odločanja. V zdravstvu pomaga strukturirati komunikacijo omejitev modela in navodil za operaterje ter za pritožbene poti.

Dostop

Dokument je javno dostopen na portalu IEEE 'Ethics in Aktion' in ga je mogoče brezplačno prenesti ter vključiti v interne politike in usposabljanja.

Dodatna razširjena pojasnila

Kako EAD1e prevede načela v prakso

EAD1e vsebuje več kot deset tematskih sklopov, ki razčlenijo etična načela v operativne smernice. Za 'človeški nadzor' npr. navaja razloge, kdaj je 'human-in-the-loop' nujen, kako določiti meje avtomatizacije, kako komunicirati tveganja uporabnikom in kako dokazati, da so varnostna stikala resnično izvedljiva in preizkušena.

Pri 'preglednosti' EAD1e poudari, da zgolj odprtje kode ni dovolj. Preglednost mora zajeti namen sistema, omejitve, negotovosti, merila uspešnosti, znane scenarije napak in pritožbene poti. V praksi se to prevede v standardizirane 'transparency profile' dokumente, ki jih organizacija vzdržuje in redno posodablja.

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)

Integracija z organizacijskimi procesi

EAD1e priporočila imajo največji učinek, ko so vezana na obstoječe procese upravljanja: politike v AIMS (ISO/IEC 42001), registre tveganj (ISO/IEC 23894), kontrole zasebnosti (npr. DPIA) in varnostne politike. S tem se načela ne obravnavajo ločeno, temveč kot del celotnega sistema vodenja kakovosti.

Za javni sektor in kritične domene je posebej koristno, da EAD1e izrecno opozarja na komunikacijo z deležniki, preizkuse razumevanja navodil ter na potrebe ranljivih skupin. To olajša usklajevanje pravnih, tehničnih in operativnih zahtev.

Primeri poglobljene rabe

Zdravstvo: razvojni tim pripravi 'transparency profile' z opisom omejitev in zanesljivosti; skupaj s kliniki izvedejo delavnice o razumevanju opozoril. Rezultati delujejo kot evidence za tehnično dokumentacijo in usposabljanje osebja.

Javna uprava: pri avtomatiziranem odločanju se EAD1e uporabi za določitev obsega informacij v obrazložitvah odločb, za oblikovanje pritožbenih poti in za periodične preglede pravičnosti ter vplivov na različne skupine prebivalcev.

Podrobnosti procesnega modela

IEEE 7000 zahteva, da organizacija sistematično identificira deležnike (vključno s sekundarnimi, ki niso neposredni uporabniki), zbere etične zahteve, jih preslika v funkcionalne in nefunkcionalne zahteve ter določi merila sprejemljivosti. Posebej pomembna je sledljivost: vsaka etična zahteva mora imeti lastnika, način verifikacije in dokazila.

Standard predvideva tudi vključevanje raziskav z uporabniki in eksperimentov, kadar ni enoznačnih odgovorov. Izvedeni poskusi, rezultati in odločitve se dokumentirajo, da lahko revizor razume, kako so bile sprejete kompromisne odločitve.

Križne povezave na AIMS/RM

V AIMS (ISO/IEC 42001) se koraki IEEE 7000 vežejo na politike, notranje presoje in CAPA; v RM (ISO/IEC 23894) se uporabljajo kot viri za identifikacijo tveganj in kot podlaga za določitev pragov 'go/no-go'. To omogoča, da etične razmisleke integriramo v 'release gates' in PMS nadzor.

Razširjeni primeri uporabe

Finančne storitve: dokumentiranje odločitev o razlagi modelov in o pragovih za dodatni človeški pregled pri mejnih primerih; sledljivost se zapiše v 'decision logs'.

Industrija/OT: za varnostno-kritične aplikacije standard vodi do uvedbe operativnih vaj ('drills') za override/safe-stop in do zahteve po metrikah, kot sta MTTR in stopnja uspešnih posegov.

Ravni transparentnosti in prilagoditev kontekstu

IEEE 7001 uvaja več ravni transparentnosti, ki jih organizacija izbere glede na tveganja in vlogo deležnika. Za končne uporabnike so v ospredju jasna navodila, opozorila in razlaga omejitev; za operaterje dodatno stanje modela, negotovosti in pragovi alarmov; za revizorje in regulatorje pa sledljivost odločitev, verzije modelov, podatkovni viri in evidence o preizkusih.

Standard poudarja, da transparentnost ni enkratna objava informacij, temveč vzdrževalni proces. Ob spremembi modela ali podatkov se posodobijo tudi informacije za deležnike – z zapisom sprememb in razlogov.

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 izgleda ‘transparency profile’ v praksi

Tipičen profil vključuje namen in meje sistema, opis podatkovnih virov in relevantnih pristranskosti, glavne metrike uspešnosti (tudi po podskupinah), seznam znanih scenarijev, v katerih se sistem lahko zmoti, ter navodila in kontaktne točke za pomoč ali pritožbe.

Za HR sisteme profil vsebuje še protokole za ravnanje v incidentih, opis človeškega nadzora (kdo, kako, kdaj poseže) in pravila za obveščanje uporabnikov.

Poglobljeni primeri

Zdravstvo: poleg navodil za operaterje dokument vključuje pragove zaupanja, pri katerih je obvezna eskalacija zdravniku, in opisuje, kako se beležijo odločitve.

E-uprava: profil določa, katere informacije mora prejeti vlagatelj (povzetek razlogov, opozorila o omejitvah, pritožbeni kanali) in kako se zagotovi skladnost z obveznostmi dokumentiranja po členu 11 AI Akt.

IEEE 7000-2021 – Model Process for Addressing Ethical Concerns During System Design

Polje	Vrednost
Kategorija	IEEE SA (7000-serija)
Organizacija	IEEE SA
Dokument	IEEE 7000-2021 – Model Process for Addressing Ethical Concerns During System Design
Leto	2021
Tip	Standard
Uradni URL	https://standards.ieee.org/standard/7000-2021/
Javno dostopen	Ne (običajno plačljivo)
Cena	Plačljivo (preko IEEE/posrednikov)
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Procesno vodenje vključevanja etičnih zahtev v zasnovo.
High-risk (zakaj)	Krepi procesne kontrole QMS/RM pri HR-AI (stakeholderji, vplivi, dokumentacija).

Namen in komu je namenjen

IEEE 7000 je normativni standard, ki opisuje model procesa za sistematično obravnavo etičnih vprašanj v fazah zasnove in razvoja. Primeren je za organizacije, ki želijo etične zahteve iz načel pretvoriti v ponovljive korake z jasnimi evidencami.

Kaj standard zahteva

Standard pokrije identifikacijo in analizo deležnikov, kartiranje etičnih zahtev, določitev projektnih ciljev in kriterijev sprejemljivosti, načrtovanje raziskav z uporabniki, dokumentiranje odločitev ter integracijo v razvojne ‘gates’. Poudarek je na dokazljivosti: zapisniki, matrike sledenja, pregledi.

Povezava z AI Akt in QMS/RM

IEEE 7000 pomaga organizaciji, da etične zahteve in vplive formalizira v okviru QMS (ISO/IEC 42001) in upravljanja tveganj (ISO/IEC 23894). Nastanejo preverljivi artefakti, ki jih priložimo tehnični dokumentaciji (čl. 11) in uporabimo pri nadzoru po dajanju na trg.

Posebnosti za HR sisteme

Pri HR sistemih standard podpira določitev meril sprejemljivosti, načrtovanje 'human-in-the-loop/on-the-loop', dokumentiranje tehtanj med koristmi in tveganji ter pripravo 'impAkt assessment' materiala, ki je uporaben v presoji skladnosti.

Primeri uporabe

V banki se IEEE 7000 uporabi za strukturirano obravnavo pravičnosti in vplivov na različne skupine komitentov; v industriji za varnostno-kritične primere z utemeljitvijo nadzora in 'safe-stop'; v javnem sektorju za uravnoteženje učinkovitosti z zahtevami po transparentnosti.

Dostop

Standard je plačljiv in dostopen prek IEEE SA ali posrednikov; smiselno ga je navesti kot referenco v razpisih ter vključiti v interne postopke.

IEEE 7001-2021 – Transparency of Autonomous Systems

Polje	Vrednost
Kategorija	IEEE SA (7000-serija)
Organizacija	IEEE SA
Dokument	IEEE 7001-2021 – Transparency of Autonomous Systems
Leto	2021
Tip	Standard
Uradni URL	https://standards.ieee.org/standard/7001-2021.html
Javno dostopen	Da (prek IEEE GET programa)
Cena	Brezplačno (prek IEEE GET)
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Merljive ravni transparentnosti in informacije za uporabnike.
High-risk (zakaj)	Neposredno podpira zahteve o preglednosti in navodilih za HR-AI.

Namen in kontekst

IEEE 7001 določa okvir transparentnosti za avtonomne sisteme ter uvaja ravni transparentnosti z zahtevami za informacije, ki jih je treba zagotoviti različnim deležnikom: uporabnikom, operaterjem, revizorjem ali regulatorjem.

Kaj standard zahteva

Opredeljuje vrste informacij (npr. namen in meje sistema, zanesljivost in negotovost, pravila nadzora, protokoli za incidente), načine predstavitve in minimalne zahteve po sledenju odločitev. Poudari povezavo med kontekstom tveganj in obsegom razkritij.

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 AI Akt
AI Akt (čl. 13, 11, 14) zahteva jasna navodila, tehnično dokumentacijo in človeški nadzor. IEEE 7001 ponuja strukturiran način, kako te informacije zbrati in predstaviti ter kako presojati ustreznost transparentnosti glede na kritičnost uporabe.

Posebnosti za HR sisteme
Za HR primere standard pomaga določiti, katere informacije morajo biti na voljo pred, med in po uporabi (opozorila, omejitve, zaupanje modela, kontaktne točke) ter kako zagotoviti sledljivost in evidence o posredovanih informacijah.

Primeri uporabe
V zdravstvu se IEEE 7001 uporabi za pripravo navodil za operaterje in paciente; v industriji za 'transparency profile' nadzornih sistemov; v javnem sektorju za standardizacijo informacij pri avtonomnih odločanjih.

Dostop
Standard je pogosto na voljo prek programa IEEE GET, kar olajša uporabo v javnih projektih in MSP.

IEEE 7002-2022 – Data Privacy Process

Polje	Vrednost
Kategorija	IEEE SA (7000-serija)
Organizacija	IEEE SA
Dokument	IEEE 7002-2022 – Data Privacy Process
Leto	2022
Tip	Standard
Uradni URL	https://standards.ieee.org/downloads/
Javno dostopen	Da (IEEE GET program)
Cena	Brezplačno (prek IEEE GET)
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Procesi zasebnosti, PIA; dopolnjuje obveznosti glede podatkov.
High-risk (zakaj)	Pomaga zagotavljati skladnost, kadar HR-AI obdeluje osebne podatke.

Namen in kontekst
IEEE 7002 opredeljuje procesne korake za sistematično zaščito zasebnosti v celotnem življenjskem ciklu podatkov in AI sistemov. Standard ni omejen na en sam predpis (npr. GDPR), temveč poda nevtralen postopek, ki ga lahko organizacije prilagodijo različnim pravnim režimom.

Kaj standard zahteva
Vključuje identifikacijo pravnih podlag in namenov obdelave, analizo tveganj za posameznike, načrtovanje minimizacije podatkov, upravljanje privolitev, anonimizacije in psevdonimizacije, določanje politik hrambe, dostopov in delitev, ter pripravo in vodenje **Privacy ImpAkt Assessment (PIA)**.

Definira tudi evidence: register obdelav, zasnove privzete zasebnosti ('privacy by design/default'), zapisnike o testih in pregledih ter postopke za obravnavo zahtev posameznikov (DSAR).

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 AI Akt in drugimi standardi

AI Akt v členu 10 obravnava upravljanje podatkov, vključno z ustreznostjo, kakovostjo in dokumentacijo. IEEE 7002 doda vidik zasebnosti kot proces, ki ga je mogoče zlahka povezati z ISO/IEC 5259-1/-3 (kakovost podatkov), ISO/IEC 42001 (QMS) in ISO/IEC 23894 (tveganja).

Posebnosti za HR sisteme

Pri HR sistemih 7002 podpira določitev **namenov** in **pravnih podlag**, razmejitev občutljivih podatkov, upravljanje posebnih kategorij podatkov in določitev dodatnih kontrol (npr. omejitev samodejnega profiliranja, nadzor dostopov in diferenciacija vlog).

Primeri uporabe

Zdravstvo: PIA oceni vplive na paciente, definira minimizacijo in čas hrambe, ter protokole za anonimizacijo. Javna uprava: standardizira obrazce za obvestila in zahteve posameznikov; v Observatoriju se hranijo povezave na evidence PIA in register obdelav.

Dostop

Standard je praviloma dostopen brezplačno prek programa IEEE GET; priporočljivo je zagotoviti interni dostop in vpeljati ga v predloge projektne dokumentacije.

Dodatna razširjena pojasnila

Za primere z visoko občutljivimi podatki uvedite ‘four-eyes’ pravilo pri dostopih ter revizijske dnevnike izvozov podatkov. V PMS spremljajte kazalnike DSAR (čas odziva, delež uspešno rešenih) in incidente zasebnosti.

IEEE 7003-2024 – Algorithmic Bias Considerations

Polje	Vrednost
Kategorija	IEEE SA (7000-serija)
Organizacija	IEEE SA
Dokument	IEEE 7003-2024 – Algorithmic Bias Considerations
Leto	2024
Tip	Standard
Uradni URL	https://standards.ieee.org/ieee/7003/11357/
Javno dostopen	Da (IEEE GET program)
Cena	Brezplačno (prek IEEE GET)
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Metodologije za obravnavo pristranskosti in diskriminacije.
High-risk (zakaj)	Podpira obveznosti glede obvladovanja tveganj vplivov na temeljne pravice.

Namen in kontekst

IEEE 7003 ponuja strukturiran pristop k prepoznavanju, merjenju in ublažitvi **algoritmične pristranskosti**. Standard naslavlja različne vire pristranskosti: v podatkih (vzorec, označevanje), v modelih (arhitektura, optimizacija) in v uporabi (kontekst, interpretacija).

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)

Kaj dokument vsebuje

Nabor korakov: definiranje prizadetih skupin in atributov, izbira ustreznih metrik pravičnosti, načrt poskusov in validacij po podskupinah, izbor tehnik ublažitve (predobdelava, 'in-processing', 'post-processing'), ter dokumentiranje odločitev in kompromisov.

Vključene so smernice za komunikacijo rezultatov in omejitev, da deležniki razumejo, kako interpretirati metrike pravičnosti in kaj pomenijo za odločanje.

Povezava z AI Akt in drugimi standardi

AI Akt posredno zahteva obvladovanje vplivov na temeljne pravice (zlasti pri HR sistemih). IEEE 7003 je neposreden 'kako' za te ocene: povezuje se z ISO/IEC 23894 (tveganja), ISO/IEC 5259 (kakovost podatkov) in IEEE 7001 (transparentnost rezultatov po podskupinah).

Posebnosti za HR sisteme

Za HR primere je ključno določiti **merila sprejemljivosti** po podskupinah ter protokole za **človeški pregled** v mejnih primerih. Standard predlaga tudi spremljanje pristranskosti v PMS, da spremembe v populaciji ali rabi ne povzročijo ponovne diskriminacije.

Primeri uporabe

Kreditno točkovanje: definiranje skupin, metrik (npr. disparate impAkt, equal opportunity), pragov in učinkovitosti ublažitev; poročila so del 'release gate'. Zaposlovanje: validacija testov po podskupinah, razlaga razlik in objava omejitev v navodilih.

Dostop

Standard je pogosto brezplačno dostopen v IEEE GET; priporočljivo je pripraviti interni vodič z izbranimi metrikami in pragovi za vaše domene.

Dodatna razširjena pojasnila

Za komunikacijo do uporabnikov pripravite jasne razlage metrik v jeziku, ki ga razume ne-strokovno občinstvo; s tem povečate razumljivost in zaupanje.

IEEE 7010-2020 – Well-being Metrics for A/IS

Polje	Vrednost
Kategorija	IEEE SA (7000-serija)
Organizacija	IEEE SA
Dokument	IEEE 7010-2020 – Well-being Metrics for Autonomous and Intelligent Systems
Leto	2020
Tip	Recommended PrAktice
Uradni URL	https://standards.ieee.org/standard/7010-2020.html
Javno dostopen	Da (IEEE GET program)
Cena	Brezplačno (prek IEEE GET)
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Merjenje vplivov na dobrobit; dopolnjuje ocene tveganj in navodila za uporabo.
High-risk (zakaj)	Koristno pri oceni vplivov HR-AI na temeljne pravice/uporabnike.

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)

Namen in kontekst
IEEE 7010 je priporočena praksa za opredelitev in merjenje **vplivov AI sistemov na dobrobit ljudi**. Dopolnjuje tehnične metrike z družbenimi in psihološkimi kazalniki, ki pomagajo razumeti vplive rabe AI na posameznike in skupnosti.

Kaj dokument vsebuje
Okvir za izbor kazalnikov (npr. varnost, psihološko blagostanje, avtonomija, kompetenca, pravičnost), metode zbiranja podatkov, interpretacijo rezultatov in vključevanje deležnikov v vrednotenje. Predlaga cikle merjenja pred in po uvedbi ter v PMS.

Povezava z AI Akt
Čeprav AI Akt neposredno ne predpisuje 'dobrobiti', zahteva obvladovanje tveganj za temeljne pravice in ustrezna navodila za uporabnike. IEEE 7010 omogoča, da organizacija te vplive prepozna, izmeri in poroča na konsistenten način, kar krepi utemeljitev odločitev v presoji skladnosti.

Posebnosti za HR sisteme
V HR kontekstih je pomembno ovrednotiti, ali sistem povzroča škodljive učinke (npr. povečana stresnost, zmanjšanje avtonomije, občutek krivice). 7010 predlaga, da se ti učinki spremljajo z metodami, ki so primerne dani populaciji, in da se uvedejo korektivni ukrepi ob negativnih trendih.

Primeri uporabe
Izobraževanje: spremljanje vpliva AI tutorjev na motivacijo in samostojnost učencev. Zdravstvo: merjenje zadovoljstva in zaupanja pacientov pri AI-podprti triaži. Javna uprava: ocena razumevanja odločitev in občutka pravičnosti pri avtomatiziranih postopkih.

Dostop
Priporočena praksa je na voljo prek IEEE GET; zaradi razumljivosti je primerna tudi za ne-tehnične deležnike in jo lahko vključite v poročila o vplivih.

Dodatna razširjena pojasnila
Pri komuniciranju do uporabnikov vključite rezime vplivov in sprejete ukrepe v navodila za uporabo, skladno z zahtevami preglednosti.

ITU-T priporočila za AI v omrežjih: arhitektura, podatki, primeri rabe in preverjanje skladnosti

Skupina ITU-T – zlasti študijski skupini SG13 (omrežja) in SG16 (multimedija) – razvija priporočila, ki urejajo uporabo umetne inteligence v komunikacijskih omrežjih ter v storitvah, ki tečejo po njih. V primerjavi z ISO/IEC in CEN/CENELEC, ki podajata horizontalne okvirje (upravljanje, tveganja, terminologija), ITU-T prispeva domensko natančnost: kako AI dejansko vgradimo v omrežje, kako ravnamo s podatki, kako ocenjujemo zaznavne module v multimediji in kako sledljivo uvajamo modele prek repozitorijev ali tržnic. Priporočila ITU-T so praviloma javno dostopna, zato so priročna kot neposredni sklic v tehnični dokumentaciji in v presoji skladnosti.

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)

Šesterica dokumentov pokriva celoten lok od zasnove do uporabe. Y.3172 postavi referenčno arhitekturo, ki definira, kje v omrežju živijo podatkovni, učni in inferenčni moduli, kdo jih orkestrira in na katerih točkah se izvajajo nadzor, varna zaustavitev ter povrnitev na stabilno stanje. Y.3174 ta okvir dopolni s procesi ravnanja s podatki: od zajema in priprave do metapodatkov, revizijske sledi in preverjanja kakovosti, s čimer daje operativni odgovor na vprašanje, kako zagotoviti dokazljivo poreklo in ustreznost podatkov za učenje in inferenco. Y.Sup55 doda plast zelo konkretnih primerov rabe v omrežjih – od napovedovanja prometa in razporejanja virov do detekcije anomalij in samo-zdravljenja –, ki pomagajo določiti realistična merila uspešnosti, tipična tveganja in primerne točke človeškega nadzora.

V multimedijskem delu niza priporočilo F.748.32 opredeli zahteve, arhitekturo in vrednotenje AI-podprte detekcije v sporočilih (npr. prepoznavanje škodljivih ali manipuliranih vsebin). Dokument jasno pove, katere metrike in dnevniški podatki so potrebni za naknadno razlago odločitev ter kako preskušati odpornost na manipulacije. F.748.45 pa se posveti AI-generiranju kode v omrežnih in multimedijskih okoljih: določi merila pravilnosti in varnosti generirane kode, postopke ocenjevanja in sledljivost od specifikacije do uvajanja, vključno z zahtevami za hiter preklic in povrnitev. Okvir zaključuje Y.3176, ki ureja varno integracijo modelov in drugih artefaktov iz repozitorijev ali tržnic v produkcijska omrežja. Z registri artefaktov, preverjanjem integritete, licenc in združljivosti ter s formalnimi odobritvami pred uvedbo omogoča popolno sledljivost življenjskega cikla modela.

Za skladnost z AI Akt so ta priporočila neposredno uporabna. Člen 11 zahteva tehnično dokumentacijo z jasno definicijo meja sistema, tokov podatkov, odločilnih točk in evidenc testov – Y.3172, Y.3174, F.748.32 in Y.3176 ponudijo natančno strukturo, kako to zapisati in dokazati. Člen 10 o upravljanju podatkov dobi operativno vsebino v Y.3174, ki je v praksi smiselno povezati z ISO/IEC 5259-1/-3 za merjenje in procesno upravljanje kakovosti. Člen 15 o robustnosti in kibernetski varnosti se nasloni na arhitekturne nadzorne točke in safe-stop mehanizme (Y.3172), na varnostne preglede in preskuse odpornosti (F.748.32), ter na kontrolirano uvajanje prek registrov artefaktov (Y.3176). Kjer AI generira kodo, F.748.45 doda zahteve za varnostne preglede, upravljanje odvisnosti in sledljivost “chain-of-custody”, kar je ključno v kontekstu člena 15 in post-market nadzora.

V visoko-tveganih primerih uporabe imajo ITU-T priporočila dodatno težo, ker omogočajo dokazljiv nadzor nad celotnim življenjskim ciklom. Arhitekturne karte in odobritveni “gates” iz Y.3172 zagotovijo, da se kritične spremembe modelov ne uvajajo brez preverjenih testov in jasne odgovornosti. Procesi Y.3174 poskrbijo, da so datasets verzionirani in da je kakovost ocenjena tudi po podskupinah, kar je bistveno za obravnavo pristranskosti in za utemeljene odločitve “go/no-go”. Y.Sup55 omogoča, da so pragovi odločitev in scenariji človeškega nadzora zasnovani na realističnih primerih, F.748.32 in F.748.45 pa poskrbita za merljive, revizijsko ponovljive preskuse detekcij in generirane kode, vključno z načrti za hitro zaustavitev in povrnitev. Y.3176 zagotovi, da je vsaka uvedba in vsaka vrnitev na prejšnjo različico sledljivo zabeležena.

Priporočena je usklajena raba teh dokumentov z vodilnimi horizontalnimi standardi: z EN ISO/IEC 23053 za konsistentno opisovanje komponent in artefaktov, z ISO/IEC 23894 za sistematično obvladovanje tveganj, z ISO/IEC 42001 za upravljavski okvir (vloge, notranje presoje, CAPA), z ETSI TS 104 223 za osnovne varnostne zahteve ter z ISO/IEC TS 8200 za človeški nadzor in mehanizme safe-stop. Takšna kombinacija omogoča, da so zahteve AI Akt pretvorjene v preverljive procese, metrike in evidence, primerne za notranje in zunanje presoje.

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)

Ker so priporočila ITU-T javno dostopna, so primerna kot referenca v razpisih, pogodbah in internem pravilniku. Organizacije lahko na njihovi osnovi oblikujejo standardizirane priloge tehnični dokumentaciji: arhitekturne sheme, profile primerov rabe z merili uspešnosti in tveganji, evidence kakovosti podatkov, "artefAkt passport" za modele ter protokole testiranja in povrnitve. S tem postanejo zahteve AI Akt in sektorske prakse prevedene v konkretne, ponovljive korake, ki zmanjšujejo operativno tveganje in pospešujejo presojo skladnosti.

ITU-T Y.3172 – Architectural framework for ML in future networks

Polje	Vrednost
Kategorija	ITU-T
Organizacija	ITU-T
Dokument	ITU-T Y.3172 (06/2019) – Architectural framework for ML in future networks incl. IMT-2020
Leto	2019
Tip	Recommendation
Uradni URL	https://www.itu.int/rec/T-REC-Y.3172
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ITU priporočilo)
SIST prodajalec	—
AI Akt (zakaj)	Arhitektura ML v omrežjih; podpira sistematično dokumentiranje in nadzor.
High-risk (zakaj)	Uporabno v telco/kritičnih domenah, kjer se HR-AI izvaja v omrežjih.

Namen in kontekst

Priporočilo ITU-T Y.3172 opredeli referenčno arhitekturo, po kateri je mogoče v omrežja 5G/IMT-2020 na urejen in sledljiv način vgraditi komponente strojnega učenja. Dokument poenoti jezik med operaterji, dobavitelji in ekipami za skladnost: natančno določi, kje v omrežju nastajajo podatki, kje se izvajata učenje in inferenca ter kje potekajo nadzorne in orkestracijske dejavnosti. S tem postavi temelje za primerljivo tehnično dokumentacijo in zanesljivo upravljanje sprememb.

Kaj dokument vsebuje

Dokument razdeli sistem na medsebojno povezane funkcionalne gradnike. Opisuje komponente za pridobivanje in pripravo omrežnih podatkov, učne cevovode, inferenčne funkcije na robu (edge) in v jedru (core), ter nadzorni in orkestracijski sloj, ki skrbi za konfiguracijo, razporejanje virov, spremljanje zdravja in varno izločanje ali zamenjavo modelov. Posebna pozornost je namenjena povezavi z obstoječimi NMS/OSS/BSS sistemi in postopkom uvedbe, ki mora vključevati preverjanje združljivosti, testiranje in možnost povrnitve na stabilno stanje.

Povezava z AI Akt in sorodnimi standardi

Y.3172 neposredno podpira izpolnjevanje zahtev EU AI akta. Člen 11 o tehnični dokumentaciji zahteva, da so meje sistema, tokovi podatkov, odločilne točke in odgovornosti jasno opisani – to je jedro arhitekturnega

okvira. Člen 9 o upravljanju tveganj se v praksi izvaja tako, da so v arhitekturi določene točke, kjer potekajo identifikacija tveganj, validacije in odobritve. Člen 15 o robustnosti in kibernetski varnosti zahteva predvidene mehanizme za varno zaustavitev, izolacijo okvar in kontrolirano vračanje na stabilne verzije. Z vidika standardov je Y.3172 tesno poravnan z EN ISO/IEC 23053 (okvir za ML sisteme), ISO/IEC 23894 (metodika obvladovanja tveganj) in ISO/IEC 42001 (AIMS/QMS za AI). Varnostne kontrole in pred-objavni preskusi se lahko oprejo na ETSI TS 104 223, mehanizmi človeškega nadzora pa na ISO/IEC TS 8200.

Zakaj je pomembno za HR sisteme

Pri visoko-tveganih sistemih je ključna dokazljivost nadzora nad življenjskim ciklom modela. Y.3172 zahteva formalno določene meje sistema, jasno razmejene odgovornosti in sledljivost verzij modelov, konfiguracij in podatkovnih virov. Predvideva odobritvene točke (‘release gates’), v katerih se na podlagi rezultatov testov in ocen tveganj odloči o uvajanju ali zavrnitvi. Dokument tudi določa, kje v arhitekturi morajo biti prisotni mehanizmi za ročni poseg, varno zaustavitev in povratno skladnost (rollback).

Primeri rabe

Operater pripravi arhitekturno karto, ki prikazuje agent za zajem telemetrije, mesto učenja in inferenčne točke, repozitorij modelov, nadzorni sloj in postopke posodobitev. V tehnični dokumentaciji so k tej karti priloženi dnevnikov sprememb, rezultati testov združljivosti in opis scenarijev safe-stop. Takšna dokumentacija izpolnjuje zahteve člena 11 in omogoča hitrejšo presojo skladnosti.

Dodatna pojasnila za implementatorje

Za praktično uvedbo priporočamo, da se k arhitekturi priložijo trije artefakti: matrika odgovornosti (RACI) po komponentah, register verzij modelov s povezavo na datasets in konfiguracije v smislu EN ISO/IEC 23053, ter sklop kontrolnih seznamov za pre-deployment in post-deployment. Kontrolni seznami naj vključujejo varnostne in robustnostne teste glede na ETSI TS 104 223 ter preizkuse obvladljivosti in človeškega nadzora po ISO/IEC TS 8200. S tem je zagotovljena neposredna sledljivost med zahtevami AI akta in konkretnimi dokazili.

ITU-T Y.3174 – Data handling framework for ML in future networks

Polje	Vrednost
Kategorija	ITU-T
Organizacija	ITU-T
Dokument	ITU-T Y.3174 (10/2020) – Data handling framework for ML in future networks
Leto	2020
Tip	Recommendation
Uradni URL	https://www.itu.int/rec/T-REC-Y.3174
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ITU priporočilo)
SIST prodajalec	—
AI Akt (zakaj)	Ravnanje s podatki za ML; povezano z obveznostmi glede podatkov.
High-risk (zakaj)	Naslavlja kakovost, poreklo in obdelavo podatkov v HR-AI v omrežjih.

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)

Namen in kontekst
Y.3174 določa, kako v omrežnih okoljih organizirati procese ravnanja s podatki, ki napajajo učenje in inferenco. Osredotoča se na zanesljivost, sledljivost in skladnost obdelav – od trenutka, ko podatki nastanejo v omrežju, do njihove priprave, hrambe, delitev in poznejše uporabe v modelih.

Kaj dokument vsebuje
Dokument uvaja jasne vloge (lastniki virov, skrbniki kakovosti, odobritveni organ), natančno opisuje korake priprave podatkov in predpisuje sklop metapodatkov, s katerimi lahko rekonstruiramo poreklo, časovne kontekste in transformacije. Predvideva revizijsko sled in politiko dostopov, ki določa, kdo in pod kakšnimi pogoji lahko podatke uporablja. Sestavni del so kontrolne točke pred učenjem in pred uvedbo modela, kjer se preverijo porazdelitve, manjkajoče vrednosti, neuravnoteženost in staranje podatkovnih virov.

Povezava z AI Akt in sorodnimi standardi
Priporočilo je neposreden operativni odgovor na zahteve člena 10 AI akta o upravljanju podatkov. Opisuje, kako dokumentirati kakovost in poreklo ter kako določiti pravila hrambe in dostopa. Z vidika standardov se vsebinsko poravna z ISO/IEC 5259-1 (merila kakovosti) in ISO/IEC 5259-3 (procesno upravljanje kakovosti), s čimer omogoča merljivo in revizijsko sledljivo izvajanje. Z IEEE 7002 je usklajeno pri oceni vpliva na zasebnost (PIA), z ISO/IEC 42001 pa pri umeščanju pregledov kakovosti v sistem vodenja. Kadar kakovost podatkov vpliva na varnost ali robustnost modela, je dobrodošla sklicna povezava na ETSI TS 104 050, ki nudi enotno ontologijo groženj.

Zakaj je pomembno za HR sisteme
Visoko-tvegani sistemi zahtevajo segmentirano razumevanje kakovosti podatkov. Y.3174 omogoča, da organizacija oceni kakovost po relevantnih podskupinah in scenarijih ter na tej podlagi postavi prage sprejemljivosti. Obvezno verzioniranje datasetov – skupaj z revizijsko sledjo transformacij – zagotavlja, da je mogoče vsako odločitev modela povezati z natančno različico podatkov in konfiguracije. Če merila niso dosežena, morajo kontrolne točke ustaviti učenje ali uvedbo do odprave vzroka.

Primeri rabe
V praksi se uvedejo ‘data-contrAkti’ med ekipami omrežij in MLOps, kjer so dogovorjeni KPI-ji kakovosti, SLO-ji za dostopnost in protokoli za preverjanje drift-a porazdelitev. Poročila o kakovosti in poreklu postanejo del tehnične dokumentacije in notranjih presoj sistema vodenja, kar olajša presojo skladnosti.

Dodatna pojasnila za implementatorje
Za zanesljivo skladnost priporočamo pripravo standardnega ‘evidence paketa’. V njem naj bodo datasheet z metrikami kakovosti (ISO/IEC 5259-1), procesni zapisnik s sledljivostjo odločitev (ISO/IEC 5259-3 in ISO/IEC 42001), ocena vplivov na zasebnost (IEEE 7002), popolna sled transformacij in izvorov (Y.3174) ter odločitev o sprejemljivosti za učenje ali uvedbo na podlagi ISO/IEC 23894. V post-market nadzoru naj se redno spremlja drift in ponovno ocenjuje tveganja.

ITU-T Y.3176 – ML marketplace integration in future networks

Polje	Vrednost
Kategorija	ITU-T
Organizacija	ITU-T
Dokument	ITU-T Y.3176 (09/2020) – ML marketplace integration in future networks

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)

Leto	2020
Tip	Recommendation
Uradni URL	https://www.itu.int/rec/T-REC-Y.3176
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ITU priporočilo)
SIST prodajalec	—
AI Akt (zakaj)	Integracija modelov/artefaktov; podpora sledljivosti in nadzoru sprememb.
High-risk (zakaj)	Koristno za evidence, registracijo in nadzor nad različicami HR-AI.

Namen in kontekst

Y.3176 obravnava, kako v omrežna okolja varno in sledljivo vključevati modele ter druge artefakte, ki izvirajo iz tržnic ali notranjih repozitorijev. Cilj je vzpostaviti proces, ki zagotovi preverjen izvor, licenčno skladnost, integriteto in združljivost, še preden je artefakt uveden v produkcijo.

Kaj dokument vsebuje

Priporočilo opisuje tipične vloge (ponudnik artefakta, potrošnik in odobritveni organ) ter vsak korak od prejema do uvajanja. Predpisuje registracijo artefaktov z metapodatki o različicah, odvisnostih, licencah, varnostnih oznakah in rezultatih testov. Urejeni so postopki za preverjanje integritete (npr. podpisov), varnostne preglede, preskuse združljivosti in formalizirana odločitev o uvedbi. Poseben poudarek je na pravilih za povratno skladnost (rollback) in varno izločanje artefaktov.

Povezava z AI Akt in sorodnimi standardi

Dokument podpira zahteve člena 11 AI akta, saj omogoča natančno sledljivost artefaktov, testov in odločitev. Vsebinsko se poravna s členom 15 prek pred-objavnih kontrol, preverjanja integritete in postopkov za varno zaustavitev. Spremembe modelov in podatkov sprožijo ponovno vrednotenje tveganj v skladu s členom 9. Z vidika standardov se Y.3176 tesno povezuje z EN ISO/IEC 23053 (povezava model–dataset–konfiguracija–testni rezultati), IEEE 7001 (informacije za operaterje in revizorje), ETSI TS 104 223 (varnostne zahteve) in ISO/IEC TS 8200 (človeški nadzor in override).

Zakaj je pomembno za HR sisteme

Visoko-tvegani sistemi zahtevajo strogo sledljivost in dokazljivo odobritev vsake spremembe. Y.3176 vzpostavi registre artefaktov, v katerih so zbrana vsa ključna dokazila: rezultati robustnostnih in varnostnih testov, licenčni podatki, odvisnosti in varnostne oznake. Vsaka sprememba sproži ponovne teste in formalno odločitev odgovorne osebe, kar preprečuje nenadzorovano uvajanje in olajša presojo skladnosti.

Primeri rabe

Organizacije vzpostavijo notranji registrski sistem za modele in povezane artefakte, ki je integriran z CI/CD. Ob predlogu za uvedbo sistem samodejno preveri podpise, odvisnosti, licence in prisotnost ključnih kontrol, kot so mehanizmi safe-stop in override. Odločitev o uvedbi je zabeležena skupaj s sklici na testna poročila in ocene tveganj.

Dodatna pojasnila za implementatorje

Kot praktično orodje priporočamo ‘artefAkt passport’. Za vsak model naj bodo navedeni izvor in licenca, različica, odvisnosti, povezave na dataset in konfiguracijo (EN ISO/IEC 23053), rezultati testov (ETSI TS 104 223 za varnost/robustnost in IEEE 7003 za pristranskost), dokazila o človeškem nadzoru (ISO/IEC TS 8200), odločitev o sprejemljivosti na podlagi ISO/IEC 23894 in pravila za rollback oziroma izločitev. Tak pasport neposredno podpira zahteve člena 11 in olajša notranje ter zunanje presoje.

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)

ITU-T Y.Sup55 – Use cases for ML in future networks

Polje	Vrednost
Kategorija	ITU-T
Organizacija	ITU-T
Dokument	ITU-T Y.Sup55 (10/2019) – Use cases for ML in future networks
Leto	2019
Tip	Supplement
Uradni URL	https://www.itu.int/rec/T-REC-Y.Sup55-201910-I/en
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ITU dopolnilo)
SIST prodajalec	—
AI Akt (zakaj)	Primeri uporabe; pomoč pri identifikaciji tveganj in meril uspešnosti.
High-risk (zakaj)	Relevantno za dokazovanje ustreznosti nadzora in robustnosti v praksi.

Namen in kontekst

Y.Sup55 je zbirka realistično opisanih primerov uporabe strojnega učenja v omrežjih, ki sega od napovedovanja prometa in optimizacije virov do odkrivanja anomalij, samo-zdravljenja in prilagajanja QoS. Dopolnilo služi kot most med arhitekturo (Y.3172) in praktičnim načrtovanjem sistemov – pokaže, kaj tipično deluje, kateri podatki so potrebni, kje so točke odločanja ter kakšna merila uspešnosti so smiselna.

Kaj dokument vsebuje

Za vsak primer rabe dokument opiše cilj, zahteve glede podatkov, tipične modele, operativna okolja (edge, core, OSS/NMS) in način vrednotenja. Pomemben del so tudi opisani tveganji: napačne odločitve pri upravljanju prometa, lažni alarmi pri detekciji anomalij, degradacija zaradi drift-a, ter napadi, ki ciljajo na vhodne podatke ali model. Ta struktura omogoča, da iz konkretnih scenarijev izpeljemo preverljive zahteve in teste.

Povezava z AI Akt in sorodnimi standardi

Dopolnilo je uporabno pri pripravi tehnične dokumentacije po členu 11 AI akta, ker jasno opredeli namen, meje in merila uspešnosti za posamezen primer rabe. Z vidika člena 9 (upravljanje tveganj) Y.Sup55 nudi katalog tipičnih tveganj, ki jih je smiselno vključiti v register tveganj v skladu z ISO/IEC 23894. Z EN ISO/IEC 23053 se poravna pri opisu komponent in procesov, z ETSI TS 104 223 pa pri izpeljavi varnostnih preskusov in osnovnih zahtev kibernetske varnosti.

Zakaj je pomembno za HR sisteme

Visoko-tvegani sistemi potrebujejo jasen dokaz, da so merila uspešnosti in nadzor prilagojeni konkretni rabi. Y.Sup55 omogoča, da organizacija za vsak primer rabe določi relevantne metrike (npr. zamuda, zanesljivost, 'false positive/negative' profili) in operativne prage, ter jih poveže z odločitvami o odobritvi uvedbe. Ker dopolnilo navaja tudi tipična tveganja, je enostavno prikazati, kako so ta tveganja naslovljena z ukrepi in testi.

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)

Primeri rabe
Pri detekciji anomalij dokument pomaga določiti razrede dogodkov, okna opazovanja, pričakovane lažne alarme in politike eskalacije. Pri upravljanju virov pa usmerja v izbiro metrik (npr. izboljšava spektralne učinkovitosti) in v zahteve po 'fallback' strategijah, če model preseže prag napake.

Dodatna pojasnila za implementatorje
Dobra praksa je, da se za vsak primer rabe izdela 'profil scenarija', ki vsebuje opis ciljev, podatkovne vire, ključne metrike, prage odločitve, tipične napake, varnostne scenarije in protokole safe-stop. Tak profil je neposreden prilog tehnični dokumentaciji in presoji skladnosti.

ITU-T F.748.32 – AI-based detection for multimedia messages

Polje	Vrednost
Kategorija	ITU-T
Organizacija	ITU-T
Dokument	ITU-T F.748.32 (10/2024) – AI-based detection for multimedia messages
Leto	2024
Tip	Recommendation
Uradni URL	https://www.itu.int/rec/T-REC-F.748.32-202410-I
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ITU priporočilo)
SIST prodajalec	—
AI Akt (zakaj)	Zahteve in arhitekture za AI detekcijo; ilustrira state-of-the-art za domnevo skladnosti.
High-risk (zakaj)	V pomoč pri definiranju meril točnosti/robustnosti in preskusov.

Namen in kontekst
F.748.32 obravnava sisteme, ki z uporabo umetne inteligence zaznavajo lastnosti ali vsebine v multimedijskih sporočilih (npr. zlonamerne, neprimerne ali goljufive vsebine). Priporočilo se osredotoča na modularno arhitekturo, zahteve glede zanesljivosti in varnosti ter na jasno opredeljene informacije, ki morajo biti na voljo operaterjem in nadzornim organom.

Kaj dokument vsebuje
Dokument opisuje referenčno arhitekturo detekcijskega sistema, vloge modulov (pridobivanje in predobdelava medijev, ekstrakcija značilk, model za detekcijo, orkestracija, dnevniški in poročevalski modul) ter zahteve za interoperabilnost. Definira tudi sklope metrik, ki pokrivajo točnost, robustnost na manipulacije, latenco obdelave in razpoložljivost storitve, ter zahteva sledljivost modela do uporabljenih podatkov in konfiguracije.

Povezava z AI Akt in sorodnimi standardi
Priporočilo nudi konkretno osnovo za pripravo tehnične dokumentacije po členu 11: arhitektura, podatkovni tokovi, konfiguracija in dnevniški podatki so opredeljeni v standardizirani obliki. Za člen 15 (robustnost in kibernetska varnost) dokument predvideva preskuse odpornosti na manipulacije multimedijskih vsebin ter zahteva zaščito modelov in podatkov v prenosu in mirovanju. Z EN ISO/IEC

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)

23053 se poravna pri definiranju komponent in artefaktov, z ISO/IEC 23894 pri oceni tveganj in načrtovanju mitigacij, z ETSI TS 104 223 pa pri varnostnih kontrolah in pred-objavnih preskusih.

Zakaj je pomembno za HR sisteme

V visoko-tveganih okoljih, kjer detekcijske odločitve vplivajo na varnost, pravice ali dostop do storitev, morajo biti pragi odločitev, metrika učinkovitosti in scenariji testiranja natančno določeni. F.748.32 predlaga sklop metrik (npr. 'false negative rate' v kritičnih razredih, odpornost na tipične napade spremembe medijskih vsebin) in opisuje, kakšni operativni zapisi so potrebni, da se odločitev lahko revidira.

Primeri rabe

V komunikacijskih platformah se standard uporabi za zasnovo modula, ki zaznava škodljive vsebine in sproža človeški pregled v mejnih primerih. V kritičnih sektorjih (npr. javna varnost) se uvede strožji režim testiranja, kjer so definirani stresni scenariji z manipuliranimi mediji in merila odzivnega časa.

Dodatna pojasnila za implementatorje

Priporočljivo je izdelati testni načrt, ki poleg običajnih metrik vključuje tudi preskuse odpornosti na napade (npr. zlonamerne transformacije videa), analizo vplivov na temeljne pravice (posebej pri samodejnem omejevanju vsebin) ter jasno politiko 'human-in-the-loop' za sporne primere.

ITU-T F.748.45 – Requirements & evaluation of AI-based code generation

Polje	Vrednost
Kategorija	ITU-T
Organizacija	ITU-T
Dokument	ITU-T F.748.45 (03/2025) – Requirements & evaluation of AI-based code generation
Leto	2025
Tip	Recommendation
Uradni URL	https://www.itu.int/rec/T-REC-F.748.45-202503-I
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	— (ITU priporočilo)
SIST prodajalec	—
AI Akt (zakaj)	Zahteve in merila evalvacije za generativne modele v multimedijskih aplikacijah.
High-risk (zakaj)	Koristno za dokazovanje varnosti in robustnosti specifičnih funkcij HR-AI.

Namen in kontekst

F.748.45 se osredotoča na sisteme, ki z uporabo AI generirajo kodo ali skripte, zlasti v multimedijskih storitvah in omrežnih platformah. Priporočilo obravnava tako funkcionalno pravilnost kot varnostne vidike in določa, kako naj organizacije načrtujejo vrednotenje takšnih sistemov.

Kaj dokument vsebuje

Dokument opredeli zahteve za podatke za učenje in test, merila kakovosti izhodne kode (pravilnost, učinkovitost, skladnost s specifikacijami), varnostne zahteve (preprečevanje injekcij, preverjanje odvisnosti, spoštovanje licenc) ter postopke ocenjevanja, ki vključujejo avtomatizirane teste 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)

človeški pregled. Posebej navaja sledenje različicam modela in testnim primerom, da je mogoče odločitve reproducirati in revidirati.

Povezava z AI Akt in sorodnimi standardi

Za člen 11 AI akta dokument ponuja jasna navodila, katera dokazila vključiti v tehnično dokumentacijo (pokrivanje testov, sledljivost med specifikacijo, generirano kodo in rezultati preverjanj). Člen 15 se odrazi v robustnostnih preskusih in kibernetski varnosti generirane kode ter v kontrolah za upravljanje odvisnosti. Z ISO/IEC 23894 se dokument ujema pri analizi tveganj (npr. napačno delujoča koda v kritičnih procesih), z EN ISO/IEC 23053 pri sledljivosti artefaktov, z ETSI TS 104 223 pri varnostnih zahtevah in z ISO/IEC TS 8200 pri človeškem nadzoru in možnostih preklica.

Zakaj je pomembno za HR sisteme

V HR kontekstih se lahko z AI generirana koda izvaja v sistemih, ki vplivajo na varnost ali pravice uporabnikov. Zato F.748.45 predvideva dodatne kontrole: strogo ločitev okolij, obvezne statične in dinamične varnostne preglede, sledljivost izvorov kode in licenc, ter obvezne postopke za hiter preklic in povrnitev na varne verzije. Dokument prav tako zahteva merila sprejemljivosti, ki morajo biti dosežena, preden je generirana koda uporabljena v produkciji.

Primeri rabe

V omrežnih platformah se standard uporabi kot podlaga za sistem, ki samodejno generira konfiguracijske skripte. Pred uvedbo skript se izvede komplet avtomatskih testov, preveri skladnost z varnostnimi politikami in licence odvisnosti, nato pa sledi človeški pregled in podpis odgovorne osebe. V multimedijskih aplikacijah, kjer generirana koda upravlja s pretoki, se izvajajo stresni testi in preverja obnašanje v mejnih primerih.

Dodatna pojasnila za implementatorje

Priporočen je ‘chain-of-custody’ zapisnik za generirano kodo: različica modela, poziv/parametri, datum in čas generacije, povezava na specifikacijo, rezultati testov, ugotovitve človeškega pregleda, sprejemna odločitev in načrt povrnitve. Takšna sledljivost izpolnjuje zahteve člena 11 in olajša revizije.

Evropske smernice za podatke in AI ekosistem: BDVA/DAIRO in DSBA/DSSC kot podpora izvedbi AI Akt

Skupnosti BDVA/DAIRO in DSBA/DSSC ne izdajata normativnih standardov, temveč ponujata strateške in tehnične usmeritve, brez katerih zahteve AI Akt težko postanejo operativne. Dokumenti SRIDA, Position Paper, Data Spaces Blueprint v2.0 in Technical Convergence v2.0 opisujejo, kako zgraditi evropski ekosistem za odgovorno uporabo umetne inteligence: kje dobiti zakonite in kakovostne podatke, kako vzpostaviti testna okolja in merila uspešnosti, ter kako zagotoviti interoperabilnost pri izmenjavi podatkov med organizacijami.

SRIDA deluje kot kompas za naložbe in zmogljivosti: od podatkovnih prostorov in preskusnih/eksperimentalnih objektov do referenčnih arhitektur in merilnih skupnosti. Za organizacije to pomeni jasen načrt, kje pridobiti vire in infrastrukturo, ki podpirajo zahteve člena 10 (kakovost in poreklo podatkov), člena 11 (tehnična dokumentacija) in člena 15 (robustnost in varnost). Position Paper dopolni

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)

sliko z vidika izvajalcev: razčleni vloge in odgovornosti ponudnikov, integratorjev in uporabnikov ter nakaže, kako zvezati pravne obveznosti s prakso upravljanja tveganj, kakovosti podatkov in sledljivosti artefaktov.

Data Spaces Blueprint v2.0 prevede načela v konkretno arhitekturo podatkovnih prostorov. Razloži, kako urediti identitete, politike deljenja, semantiko, konektorje in nadzor, da so prenosi podatkov čezorganizacijsko sledljivi in pravno varni. V praksi je to neposreden odgovor na vprašanje, kako izpolniti zahteve AI Akt za zakonito, dokumentirano in ponovljivo uporabo podatkov v učnem in operativnem okolju modelov. Technical Convergence v2.0 pa naslovi ključno izziv: interoperabilnost med različnimi podatkovnimi prostori. S skupnimi profili za identitete, politike, semantiko in revizijske evidence zmanjšuje fragmentacijo, poenostavlja integracije in omogoča, da so dokazi o poreklu podatkov in dovoljenjih primerljivi po celotni poti podatka.

Za visoko-tvegane sisteme je ta četverica dokumentov posebej pomembna, ker ustvarja most med pravnimi zahtevami in izvedbenimi gradniki. Organizacijam pomaga oblikovati verodostojna dokazila: od opisov tokov podatkov in meril uspešnosti do revizijskih sledi, dovoljenj in “chain-of-custody” evidenc. V kombinaciji z horizontalnimi standardi (npr. EN ISO/IEC 23053, ISO/IEC 23894, ISO/IEC 42001 in ISO/IEC 5259-1/-3) omogoča celovit pristop: jasne procese, merljive metrike in preverljive evidence, ki pospešijo presojo skladnosti ter znižajo operativno tveganje.

BDVA/DAIRO – SRIDA: AI, Data & Robotics Partnership (V3.1) (2021)

Polje	Vrednost
Kategorija	BDVA/DAIRO
Organizacija	BDVA/DAIRO (ADRA)
Dokument	SRIDA – AI, Data & Robotics Partnership (V3.1)
Leto	2021
Tip	Strategija (SRIDA)
Uradni URL	https://bdva.eu/wp-content/uploads/pdfs-legacy/AI-Data-Robotics-Partnership-SRIDA%20V3.1.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Strategija R&I; usmeritve za inovacije in infrastrukture (posredno).
High-risk (zakaj)	Posredno – podpira načrtovanje podatkovnih/tehnoloških zmogljivosti HR-AI.

Namen in kontekst

SRIDA je strateški dokument, ki opisuje, kako naj Evropa kot celota gradi zmogljivosti za odgovorno in konkurenčno uporabo umetne inteligence, podatkov in robotike. Ne gre za tehnični standard, temveč za vodilo, ki pomaga državam, podjetjem in raziskovalnim ustanovam razumeti, kam usmeriti vlaganja. Za bralce brez predhodnega poznavanja je koristno vedeti, da SRIDA pojasni, katere skupne komponente (npr.

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)

podatkovni prostori, preskusna okolja in referenčne arhitekture) so potrebne, da AI v praksi deluje zanesljivo, zakonito in primerljivo med različnimi organizacijami.

Kaj dokument vsebuje

Dokument je razdeljen na več tematskih stebrov. V stebru »podatki« opisuje, kako zgraditi vire kakovostnih in zakonitih podatkov ter kako omogočiti varno izmenjavo med organizacijami. V stebru »AI metode« so izpostavljene raziskovalne in razvojne prioritete, kot so robustnost, pojasnljivost in varnost modelov. V stebru »zaupanje in etika« poudarja potrebo po merljivih načelih, ki jih je mogoče preskusiti v praksi. SRIDA hkrati predstavi konkretne objekte, kot so preskusni in eksperimentalni objekti (TEF), ki ponujajo realna okolja za testiranje, ter skupne repozitorije, kjer se lahko delijo rezultati in orodja.

Povezava z AI Akt in standardi

Evropski akt o umetni inteligenci (AI Akt) postavlja obveznosti glede podatkov, dokumentacije, robustnosti in nadzora. SRIDA pomaga razumeti, kje v evropskem ekosistemu najdemo zmogljivosti, ki omogočajo izpolnjevanje teh obveznosti. Na primer, zahteva, da so podatki ustrezni in dokumentirani, je lažje izpolnjena, če organizacija sodeluje v podatkovnem prostoru, kjer so pravila dostopa, metapodatki in evidence že standardizirani. SRIDA je združljiva z uveljavljenimi standardi, kot so EN ISO/IEC 23053 (okvir ML sistemov), ISO/IEC 23894 (upravljanje tveganj), ISO/IEC 42001 (sistem vodenja AI) in ISO/IEC 5259-1/-3 (kakovost podatkov). V praksi to pomeni, da strateške usmeritve SRIDA podpirajo tehnično in procesno izvajanje teh standardov.

Zakaj je pomembno za visoko-tvegane (HR) sisteme

Pri sistemih z visokim tveganjem so zahteve glede dokazil strožje. Potrebujemo zanesljive podatkovne vire, ponovljive preskuse in jasne odgovornosti. SRIDA organizacijam pokaže, kako te elemente zagotoviti na ravni ekosistema: kje pridobiti referenčne nabori podatkov, kako priti do preskusnih okolij in kako vzpostaviti skupne vzorce dokumentacije. S tem se zmanjša tveganje projektnih zastojev in neuspešnih presoj skladnosti.

Primeri rabe

Podjetje, ki načrtuje HR-AI, lahko na podlagi SRIDA pripravi načrt kapacitet: vključitev v podatkovni prostor, dostop do TEF za robustnostne teste, in uvedbo referenčne arhitekture, ki je kompatibilna s standardi. Tako nastane jasna pot od prototipa do sistema, ki je pripravljen za regulirano okolje.

Dodatna pojasnila za implementatorje

Pri pripravi tehnične dokumentacije je smiselno navesti, katera kapaciteta iz SRIDA pokriva posamezno obveznost AI Akt. Na primer: »robustnostni preskusi se izvajajo v TEF X po postopku Y; podatki prihajajo iz podatkovnega prostora Z, kjer so metapodatki in dovoljenja evidentirani«.

BDVA/DAIRO – Position Paper: Response to EC AI Regulation proposal (2021)

Polje	Vrednost
Kategorija	BDVA/DAIRO
Organizacija	BDVA/DAIRO
Dokument	Position Paper – Response to EC AI Regulation proposal (2021)
Leto	2021
Tip	Position Paper

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)

Uradni URL	https://bdva.eu/wp-content/uploads/pdfs-legacy/BDVA_DAIRO%20response-feedback%20AI%20Regulation_Final.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Stališča skupnosti – interpretacija/prioritete (posredno).
High-risk (zakaj)	Posredno – pomaga pri interpretaciji implementacij HR-AI.

Namen in kontekst

Stališče BDVA/DAIRO je nastalo kot odziv na prvi predlog evropske uredbe o AI. Dokument ne uvaja novih pravil, temveč zbere pogled industrije in raziskovalne skupnosti na to, kako razumeti predlagane obveznosti in kako jih v praksi uresničiti. Za bralca to pomeni, da dobi strnjen pregled področij, kjer so potrebna dodatna pojasnila ali usklajevanja.

Kaj dokument vsebuje

Besedilo razlaga ključne pojme, kot so vloge ponudnika, uvoznika, distributerja in uporabnika, ter razmejitev odgovornosti med dobaviteljem modela in integratorjem rešitve. Posebej obravnava vprašanja kakovosti in sledljivosti podatkov, sprememb različic modelov ter pomen merljivih zahtev, ki jih je mogoče preveriti in dokumentirati.

Povezava z AI Akt in standardi

Ker je dokument nastal kot odziv na regulativo, neposredno naslavlja člene o upravljanju tveganj, tehnični dokumentaciji in robustnosti. Hkrati poudarja, da se zahteve lažje izvajajo, če so zvezane na uveljavljene standarde, kot so EN ISO/IEC 23053, ISO/IEC 23894, ISO/IEC 42001 in ISO/IEC 5259. Tak pristop zmanjša možnost različnih interpretacij in pospeši presoje.

Zakaj je pomembno za visoko-tvegane (HR) sisteme

V HR sistemih morajo biti odgovornosti in evidence natančno opredeljene. Stališče pomaga pravnim in skladnostnim ekipam pri vzpostavitvi notranjih pravil, ki jasno določajo, kdo je za kaj odgovoren, katere dokumente je treba pripraviti in kdaj je potrebna ponovna ocena tveganj.

Primeri rabe

Pri pripravi javnega razpisa lahko naročnik ključne točke iz dokumenta pretvori v konkretne zahteve: vodenje sistema upravljanja po ISO/IEC 42001, uporaba metodologije tveganj po ISO/IEC 23894 in poročanje o kakovosti podatkov po ISO/IEC 5259.

Dodatna pojasnila za implementatorje

Priporočljivo je pripraviti kratek notranji »terminološki dodatek« na osnovi dokumenta in ga uskladiti z ISO/IEC 22989. S tem se v projektih izognemo različnemu razumevanju osnovnih pojmov.

DSBA/DSSC – Data Spaces Blueprint v2.0 (portal) (2024)

Polje	Vrednost

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)

Kategorija	DSBA/DSSC
Organizacija	DSBA/DSSC
Dokument	Data Spaces Blueprint v2.0 (portal)
Leto	2024
Tip	Blueprint (portal)
Uradni URL	https://dssc.eu/page/blueprint
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Interoperabilnost in upravljanje podatkov v podatkovnih prostorih (posredno).
High-risk (zakaj)	Posredno – podpora za podatkovno upravljanje in dostopne evidence HR-AI.

Namen in kontekst

Blueprint v2.0 je zbirno mesto znanja o tem, kako zasnovati in upravljati podatkovne prostore v EU. Podatkovni prostor je dogovorjena tehnična in pravna ureditev, v kateri organizacije izmenjujejo podatke na predvidljiv, sledljiv in varen način. Za AI je to pomembno zato, ker kakovost in zakonitost podatkov neposredno vplivata na kakovost modelov.

Kaj dokument vsebuje

Portal predstavi referenčno arhitekturo s plastmi za identitete in zaupanje, politike deljenja, semantiko podatkov, konektorje za prenos in nadzor nad uporabo. Za vsako plast so opisane vloge akterjev (ponudnik, potrošnik, posrednik, operator) ter tipični vzorci, kako podatki potujejo med partnerji. Dodane so povezave do implementacijskih profilov in varnostnih priporočil, ki olajšajo praktično uvedbo.

Povezava z AI Akt in standardi

AI Akt zahteva, da so podatki primerne kakovosti, zakoniti in dokumentirani. Blueprint pokaže, kako v večorganizacijskem okolju vzpostaviti pravila dostopa, metapodatke, revizijske sledi in tehnične zaščite. V praksi se uporablja skupaj s standardi ISO/IEC 5259-1/-3 (merjenje in procesi kakovosti), EN ISO/IEC 23053 (sledljivost artefaktov), ISO/IEC 42001 (odgovornosti in presoje) ter varnostnimi praksami ETSI/ISO.

Zakaj je pomembno za visoko-tvegane (HR) sisteme

HR sistemi pogosto potrebujejo podatke iz več virov. Podatkovni prostor zagotovi, da je za vsak prenos znano, kdo je poslal, kaj je bilo poslano, pod kakšnimi pravili in kdo je to uporabil. Takšna sledljivost zmanjšuje tveganje nezakonitih obdelav in nedoslednosti med treningom in uporabo.

Primeri rabe

Konzorcij industrijskih partnerjev vzpostavi podatkovni prostor za deljenje proizvodnih podatkov, pri čemer so dogovorjeni SLA-ji in politike deljenja. AI ekipe pridobivajo verzionirane nabori podatkov z jasnimi metapodatki, kar poenostavi pripravo dokazil za presojo skladnosti.

Dodatna pojasnila za implementatorje

Smiselno je v sistem vodenja (QMS) vključiti postopek ‘onboardinga’ v podatkovne prostore in zahtevati uporabo certificiranih konektorjev, ki samodejno beležijo metapodatke in dovoljenja. Tako so potrebne evidence na voljo brez dodatnega ročnega dela.

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)

DSBA – Technical Convergence v2.0 (2023)

Polje	Vrednost
Kategorija	DSBA
Organizacija	DSBA
Dokument	DSBA Technical Convergence v2.0 (2023)
Leto	2023
Tip	Discussion Document
Uradni URL	https://data-spaces-business-alliance.eu/wp-content/uploads/dlm_uploads/Data-Spaces-Business-Alliance-Technical-Convergence-V2.pdf
Javno dostopen	Da (brezplačno)
Cena	Brezplačno
SIST oznaka	—
SIST prodajalec	—
AI Akt (zakaj)	Tehnična konvergenca komponent za podatkovne prostore (posredno).
High-risk (zakaj)	Posredno – referenca za varno izmenjavo podatkov pri HR-AI.

Namen in kontekst

Dokument pojasni, zakaj je pomembno, da različni podatkovni prostori v Evropi uporabljajo primerljive tehnične profile. Če vsak prostor uporablja drugačna pravila za identitete, semantiko ali politike deljenja, izmenjava podatkov med prostori postane počasna in tvegana. Konvergenca zato zmanjšuje stroške integracije in povečuje zaupanje med partnerji.

Kaj dokument vsebuje

Opisani so gradniki, ki jih morajo prostori uskladiti: upravljanje identitet in zaupanja, konektorje in politike, semantične modele in ontologije, evidence pogodb in skladnosti ter mehanizme opazovanja in nadzora. Dokument predlaga, kako različne tehnologije preslikati v skupne profile, da so integracije ponovljive in preverljive.

Povezava z AI Akt in standardi

Ko podatki prehajajo med več prostori, je ključna sledljivost. Konvergenčni profili omogočajo, da so metapodatki, pravila in revizijske evidence primerljive v celotni poti podatkov. To neposredno podpira zahteve člena 10 (poreklo in kakovost) in člena 11 (tehnična dokumentacija). Prakse so usklajene s standardi ISO/IEC 5259-1/-3, EN ISO/IEC 23053 in ISO/IEC 42001 ter z varnostnimi smernicami ETSI/ISO.

Zakaj je pomembno za visoko-tvegane (HR) sisteme

HR sistemi pogosto temeljijo na podatkih več partnerjev. Če so pravila in zapisi v vseh vpletenih prostorih konsistentni, je veliko lažje dokazati, da so bili podatki zakonito pridobljeni, ustrezno obdelani in pravilno uporabljeni. S tem se zmanjšajo operativna tveganja in pospešijo odobritve.

Primeri rabe

Podjetje, ki združuje podatke proizvajalcev, servisov in uporabnikov, z uporabo konvergenčnih profilov zagotovi, da identitete, dovoljenja in semantika delujejo enako v vseh prostorih. To omogoča hitrejšo gradnjo modelov in enostavnejše presoje skladnosti.

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)

Dodatna pojasnila za implementatorje

Koristno je pripraviti kontrolni seznam interoperabilnosti: usklajeni profili identitet, politike deljenja, semantična ujemanja in izvoz revizijskih evidenc. V sistem vodenja je smiselno vključiti pravilo, da se pri čezmejni izmenjavi podatkov uporabljajo izključno konvergenčni profili.

Zaključek

Ta dokument je pokazal, da je skladnost rešitev UI večplastna naloga, ki zahteva usklajeno rabo organizacijskih, tehničnih in domenskih standardov. ISO/IEC nudi "hrbtenico" upravljanja, tveganj, terminologije, arhitekture, kakovosti podatkov in človeškega nadzora; CEN/CENELEC (EN) zagotavlja evropsko izdajo istih vsebin in pot do domneve skladnosti; ETSI prevede zahteve robustnosti in kibernetske varnosti v konkretne kontrole in preskuse; IEEE SA (7000-serija) operacionalizira etiko, transparentnost, zasebnost in obravnavo pristranskosti; ITU-T pa prinese natančne arhitekture, procese in primere rabe v omrežjih in multimediji. V Sloveniji je dodatna prednost SIST kot nacionalne točke za dostop do EN/ISO/IEC izdaj in sklicevanje v razpisih, pogodbah ter presojah.

Ključne ugotovitve

- Od načel do dokazil: Vsi obravnavani standardi skupaj tvorijo pot od "kaj mora biti" (zahteve AI Akt) do "kako se to izvede in dokaže" (procesi, kontrole, metrike, evidence).
- Skladnost = organizacija + tehnika: ISO/IEC 42001 in 23894 postavita okvir vodenja in tveganj; 23053 poveže arhitekturo, artefakte in sledljivost; 5259-1/-3 standardizirata kakovost podatkov; TS 8200 zagotovi obvladljivost in človeški nadzor.
- Varnost in robustnost sta preverljivi: ETSI TS 104 223 (osnovne varnostne zahteve) in TS 104 050 (ontologija groženj) omogočita ponovljive preglede, testne baterije in primerljivo poročanje.
- Temeljne pravice in uporabniki: IEEE 7001/7002/7003/7010 prevajajo transparentnost, zasebnost, pristranskost in dobrobit v konkretne korake, merila in dokumentacijo, združljivo z zahtevami o informacijah za uporabnike in PMS.
- Domenska izvedba: ITU-T Y.317x in F.748.x standardizirajo arhitekture, podatkovne prakse, registracijo artefaktov in preskušanje detekcij/generacij v omrežjih in multimediji – tam, kjer običajno nastajajo visoko-tvegani scenariji.

Priporočila za uvedbo

- Vzpostavite minimalni komplet: EN ISO/IEC 23894 (RM), EN ISO/IEC 23053 (arhitektura/sledljivost), EN ISO/IEC 22989 (terminologija), ISO/IEC 42001 (AIMS), ISO/IEC 5259-1/-3 (kakovost podatkov), ISO/IEC TS 8200 (obvladljivost), ETSI TS 104 223/104 050 (varnost/ontologija), IEEE 7001/7002/7003 (transparentnost, zasebnost, pristranskost).
- Standardizirajte dokazila: predloge za (i) tehnično dokumentacijo (čl. 11), (ii) register tveganj z enotnimi oznakami groženj (TS 104 050), (iii) datasheete datasetov in modelov (23053 + 5259), (iv) "transparency profile" (IEEE 7001), (v) PIA/DPIA (IEEE 7002), (vi) protokole nadzora/override/safe-stop (TS 8200).

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)

- Uskladite nabavo: v razpisih in pogodbah zahtevajte sklic na EN izdaje, ETSI TS za varnostne kontrole ter IEEE smernice za pristranskost in transparentnost.
- Načrt PMS: določite metrike drift-a, robustnosti, pravičnosti, varnostnih incidentov in učinkovitosti človeškega nadzora; definirajte sprožilce za retrening, rollback ali izločitev modela.

Spremljanje harmonizacije: redno preverjajte, kateri EN standardi so (ali bodo) harmonizirani za AI Akt, da lahko uveljavljate domnevo skladnosti.

Minimalni nabor dokazil za HR sisteme

- QMS/AIMS (42001): politike, vloge, notranje presoje, CAPA.
- RM (EN 23894): register tveganj, merila sprejemljivosti, "worst-case" scenariji, odločitev go/no-go.
- Arhitektura in sledljivost (EN 23053): meje sistema, tokovi podatkov, artefakti, verzije, "release gates".
- Podatki (5259-1/-3): kakovost po podskupinah, evidence porekla in transformacij, odločitve o primernosti.
- Transparentnost (IEEE 7001): navodila za uporabnike/operaterje/revizorje, sledljivost odločitev.
- Zasebnost (IEEE 7002): PIA, pravne podlage, minimizacija, evidence obdelav.
- Pristranskost (IEEE 7003): metrike po podskupinah, pragovi, rezultati ublažitve, spremljanje v PMS.
- Varnost/robustnost (ETSI TS 104 223 + TS 104 050): preskusi, kontrole, incidentni zapisi z enotnimi oznakami.
- Obvladljivost (TS 8200): protokoli override/safe-stop, vaje odziva, MTTR, usposabljanja.
- Upravljanje in odgovornosti

Upravljanje in odgovornosti

Določite odgovorne lastnike za: podatke (5259-3), modele in spremembe (23053), tveganja (23894), varnost (TS 104 223), transparentnost (IEEE 7001) ter človeški nadzor (TS 8200). Vodstvo naj kvartalno pregleda KPI in stanje CAPA.

Merjenje uspešnosti skladnosti

- Vzpostavite redno poročanje z jedrnimi kazalniki: stopnja pokritosti kontrol (TS 104 223), delež modelov z veljavno PIA (IEEE 7002), delež modelov z validacijo pravičnosti (IEEE 7003), povprečni čas do varne zaustavitve (TS 8200), trendi drift-a in incidentov ter spoštovanje "release gates".

Tveganja in meje

- Standardi ne nadomeščajo strokovne presoje. Potrebni so tudi: domensko znanje, presoja vplivov na temeljne pravice in izrecna politika etičnih omejitev rabe. Kjer so odvisnosti od dobaviteljev, vključite zahteve za evidence, teste in rollback kot pogoj za dobavo.

Sistematična raba tukaj zbranih standardov omogoča slovenskim organizacijam hiter in pregleden prehod od načel AI Akt do preverljivih dokazil. Priporočamo, da se v naslednjih 6 mesecih vzpostavi minimalni komplet procesov in predlog dokumentacije, ki pokrije upravljanje, tveganja, arhitekturo, podatke, varnost, transparentnost, pristranskost in človeški nadzor. S tem pristopom bodo rešitve UI lažje prestale presojo skladnosti, hkrati pa bodo varnejše, bolj pravične in razumljive za uporabnike.

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)