VENDIM
Nr. 479, datë 24.7.2024

PËR MIRATIMIN E DOKUMENTIT TË METODOLOGJISË DHE STANDARDET TEKNIKE PËR PËRDORIMIN E INTELIGJENCËS ARTIFICIALE NË REPUBLIKËN E SHQIPËRISË

Në mbështetje të nenit 100 të Kushtetutës dhe të pikës 2, të nenit 39, të ligjit nr. 43/2023, "Për qeverisjen elektronike", me propozimin e Kryeministrit, Këshilli i Ministrave

VENDOSI:

1. Miratimin e dokumentit të metodologjisë dhe standardet teknike për përdorimin e inteligjencës artificiale në Republikën e Shqipërisë, sipas tekstit që i bashkëlidhet këtij vendimi dhe është pjesë përbërëse e tij.
2. Ngarkohen Agjencia Kombëtare e Shoqërisë së Informacionit, ministritë dhe institucionet e tjera publike për zbatimin e këtij vendimi.

Ky vendim hyn në fuqi menjëherë dhe botohet në Fletoren Zyrtare.

ZËVENDËSKRYEMINISTËR
Belinda Balluku

METODOLOGJIA DHE STANDARDET TEKNIKE TË PËRDORIMIT TË TEKNOLOGJISË SË INTELIGJENCËS ARTIFICIALE NË REPUBLIKËN E SHQIPËRISË

Hyrje

Kjo metodologji është konceptuar dhe hartuar duke u bazuar në praktikat dhe standardet më të mira të fushës, ndër të cilat mund të përmenden ato të përcaktuara nga Instituti Kombëtar i Standardeve dhe Teknologjisë (NIST), publikimi IA RMF 1.0[1], OECD[2], Bashkimi Evropian (BE), të cilat theksojnë rëndësinë që ka dokumentimi dhe vlerësimi i ndikimit që kanë në sistemet e informacionit ndryshimet e propozuara mbi proceset operacionale, infrastrukturat teknologjike, personelin dhe përdoruesit e sistemeve, që bazohen në teknologjinë e inteligjencës artificiale etj.

Në kornizën e zhvillimit të sistemeve të inteligjencës artificiale (IA), një metodologji e qartë dhe e strukturuar shërben për implementimin me sukses të saj. Metodologjitë tradicionale përfshijnë hulumtimin dhe zhvillimin e një sistemi, duke përdorur teknika të njohura, si programimi i strukturuar dhe inxhinierimi i *softwer*-ve. Në kontrast, metodologjia e inteligjencës artificiale përdor mënyra të avancuara të mësimit dhe modelimit për të maturuar sistemet kompjuterike. Sistemet IA përdorin mënyra të avancuara të modelimit për të trajtuar informacionin dhe marrëdhënie komplekse në mes të të dhënave.

Megjithatë, ka disa kufizime: të dhënat duhet të etiketohen dhe trajtohen me kujdes para se të procesohen nga sistemi.

Pra, të dhënat duhet t'i jepen të mirëtrajtuara sistemeve me IA dhe të etiketuara me kujdes për të parandaluar sjellje të papritura dhe të paparashikueshme.

Metodologjia e IA-së është dinamike dhe mund të përshtatet në mjedisin e saj për të zbërthyer modele dhe marrëdhënie, duke shfrytëzuar njohuritë e veta. Kjo u mundëson sistemeve IA aftësinë për të adaptuar ndryshimet, duke rritur performancën e tyre me kohën. Ndërkohë që metodologjia

---

[1] https://csrc.nist.gov/pubs/sp/800/53/r5/final
[2] https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

tradicionale është statike dhe kërkon një plan të qartë paraprak, IA-ja është dinamike dhe mund të përshtatet në mjedisin e saj.

KREU I
QËLLIMI DHE OBJEKTI

1. Kjo metodologji është e planifikuar dhe strukturuar në qasjen e një politike makromenaxhuese dhe makroorientuese për implementimin e teknologjive dhe të zgjidhjeve të reja të inteligjencës artificiale në Republikën e Shqipërisë, veçanërisht atyre që prekin institucionet publike nën varësinë e Këshillit të Ministrave.

2. Metodologjia siguron një udhëzues të qartë dhe efektiv për përcaktimin e veprimeve dhe të masave që do të ndërmerren për të garantuar një implementim të suksesshëm dhe të sigurt të teknologjive të reja të IA-së.

3. Objekti i një metodologjie për implementimin e inteligjencës artificiale (IA) është të ofrojë një qasje sistematike dhe të strukturuar për zhvillimin, zbatimin dhe menaxhimin e teknologjive të IA-së.

4. Përcaktimin e hapave dhe të proceseve specifike që duhen ndjekur për të garantuar një implementim të suksesshëm dhe efektiv.

KREU II
STANDARDET

1. Standardet teknike në zhvillimin e inteligjencës artificiale

Standardet teknike për IA-në në këtë dokument referohen si kritere dhe specifikime që kanë për qëllim të sigurojnë besueshmërinë, interoperabilitetin dhe përdorimin etik të teknologjive të inteligjencës artificiale.

1.1. Disa aspekte kyçe të mbuluara nga standardet teknike për IA-në janë:

A. Menaxhimi i riskut[3]: Ky standard ofron kritere për identifikimin dhe menaxhimin e riskut të lidhura me teknologjitë e inteligjencës artificiale. Këto standarde teknike ofrojnë korniza specifike për zbatimin e praktikave të menaxhimit të riskut në fushat ku përfshihen teknologjia dhe inteligjenca artificiale.

Politikat, proceset, procedurat dhe praktikat në institucionet publike lidhur me përcaktimin, matjen dhe menaxhimin e riskut të inteligjencës artificiale janë të mirëpërcaktuara, transparente dhe të zbatuara në mënyrë efektive. Është shumë e rëndësishme që kërkesat ligjore dhe nënligjore që lidhen me inteligjencën artificiale të kuptohen, të menaxhohen dhe të dokumentohen në mënyrën e duhur.

Veprimet që rekomandohen për të realizuar të përcaktuar sa më sipër janë:

a) Në çdo planifikim dhe implementim që do të kryhet apo do të parashikohet për t'u kryer kërkesat ligjore dhe rregullatore të zbatueshme të jenë specifike për industrinë, sektorin dhe qëllimin e biznesit, si dhe të kenë të përfshirë kontekstin e aplikimit të sistemit të inteligjencës artificiale të implementuar;

b) Plani për menaxhimin e riskut dhe veprimet e kryera në funksion të sistemit duhet të jenë në përputhje me përcaktimet ligjore të zbatueshme;

c) Të hartohen dhe të miradministrohen politika për trajnimin (dhe ritrajnimin), pilotimin për punonjësit e institucioneve publike, por edhe për personelin e bizneseve private apo operatorëve që do të zhvillojnë sisteme apo mjete të inteligjencës artificiale lidhur me elemente të rëndësishme të bazës ligjore dhe nënligjore në fuqi, të cilat mund të ndikojnë në dizenjimin dhe zhvillimin e aktiviteteve apo mjeteve/sistemeve që lidhen me IA-në;

ç) Klasifikimi i sistemeve IA të ndara në 4 kategori, sipas nivelit të riskut:

---

[3] *https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook*

I. Rrezik i papranueshëm: Kjo kategori përfshin sistemet e IA-së që shkelin në thelb vlerat dhe të drejtat themelore të BE-së. Shembujt përfshijnë sistemet e vlerësimit social që gjykojnë individët dhe manipulojnë IA, të krijuar për të mashtruar ose dëmtuar njerëzit.

II. Rrezik i lartë: Sistemet IA të identifikuara si me rrezik të lartë përfshijnë teknologjinë IA të përdorur në:

- Infrastruktura kritike (p.sh. transport), që mund të rrezikojë jetën dhe shëndetin e qytetarëve;
- Trajnim arsimor ose profesional, që mund të përcaktojë qasjen në arsim dhe karrierë;
- Komponentët e sigurisë së produkteve (p.sh. kirurgji me asistencë robotike);
- Punësim dhe menaxhim të punonjësve;
- Shërbime private dhe publike thelbësore;
- Zbatimi i ligjit;
- Menaxhimi i migracionit, azilit dhe kontrollit kufitar;
- Administrimi i drejtësisë dhe proceseve demokratike.

III. Rrezik i kufizuar: Këto sisteme të IA-së paraqesin kërcënime minimale, por ende kërkojnë njëfarë niveli të ndërgjegjësimit të përdoruesit. Për shembull, kur përdoren sisteme IA, si *chatbot*-et, njerëzit duhet të jenë të vetëdijshëm që po ndërveprojnë me një makinë.

IV. Rrezik minimal ose asnjë rrezik: Kjo kategori përfshin sistemet e IA-së me praktikisht asnjë rrezik për sigurinë, privatësinë ose të drejtat. Përfshirja e IA-së në aplikacione si videolojërat, filtrat e postës elektronike dhe sisteme të ngjashme.

B. Karakteristikat e besueshmërisë mbi inteligjencën artificiale duhet të jenë të integruara me politikat, proceset dhe procedurat e institucioneve publike apo bizneseve private. Politikat e menaxhimit të riskut të inteligjencës artificiale në këta të fundit duhet të jenë të projektuara:

a) për të përcaktuar termat dhe konceptet kyçe lidhur me sistemet e inteligjencës artificiale dhe shtrirjen e synuar të qëllimeve dhe përdorimeve të tyre;

b) për të lidhur menaxhimin e inteligjencës artificiale me menaxhimin e brendshëm organizativ ekzistues dhe kontrollet e riskut;

c) përputhje me politikat e menaxhimit të riskut të IA-së, me politikat dhe praktikat e menaxhimit të të dhënave, në veçanti menaxhimit të të dhënave sensitive;

ç) detajuar standarde për dizajnin eksperimental, cilësinë e të dhënave dhe trajnimin e sistemeve të IA-së;

d) detajuar proceset e testimit dhe validimit të sistemeve;

dh) detajuar dhe testuar planet e përgjigjes ndaj incidenteve;

e) verifikuar se politikat formale të menaxhimit të riskut të inteligjencës artificiale përputhen me bazën ligjore ekzistuese, dhe me praktikat më të mira ndërkombëtare të fushës;

ë) verifikuar se politikat e menaxhimit të riskut të inteligjencës artificiale përfshijnë sistemet që mund të jenë të implementuar aktualisht C. Procesi i menaxhimit të riskut dhe rezultatet e tij duhet të administrohen nëpërmjet politikave dhe procedurave transparente, si dhe kontrolleve të tjera bazuar në prioritetet e institucioneve publike në Republikën e Shqipërisë.

Veprimet që rekomandohen për të realizuar të përcaktuara sa më sipër janë:

a) hartimi dhe rishikimi rregullisht i politikave të dokumentimit që, midis të tjerave, adresojnë informacionin lidhur me:

i. informacion mbi palët e interesit të IA-së;

ii. zbatimin dhe përdorimin et IA-së;

iii. risqet potenciale dhe ndikimin e mundshëm;

iv. kushtet paraprake dhe kufizimet;

v. metodologjinë algoritmike;

vi. planet për monitorimin dhe menaxhimin e ndryshimeve.

2. Standardet që do të zbatohen gjatë ciklit jetësor të sistemeve të IA-së:

Sistemet e inteligjencës artificiale kanë potencialin të sjellin shumë avantazhe për shoqërinë. Për këtë arsye, kriteret janë ndarë në katër fusha kyçe, brenda ciklit të zhvillimit të sistemit të

inteligjencës artificiale: dizenjimi i sigurt, zhvillimi i sigurt, implementimi i sigurt dhe operimi e mirëmbajtja e siguruar. Këto elemente ndihmojnë në zvogëlimin e riskut të përgjithshëm për procesin e zhvillimit të sistemit të inteligjencës artificiale në institucionet publike në Republikën e Shqipërisë.

A. Dizenjimi i sigurt: Përfshin kuptimin e riskut dhe modelimin e kërcënimeve, si dhe temat specifike për t'u konsideruar në dizajnin e sistemit dhe modelit.

1. Elementet e dizenjimit të sigurt:

i. Ndërgjegjësimi i gjithë stafit të institucioneve publike mbi kërcënimet dhe riskun;

ii. Modelimi i mbrojtjes nga kërcënimet për sistemin;

iii. Në dizenjim të përfshihet funksionimi optimal, siguria dhe performanca e sistemit;

iv. Përfshirja e avantazheve dhe kompromentimeve të sigurisë kur zgjidhet sistemi i IA-së për sistemin.

B. Zhvillimi i sigurt: Kriteret që aplikohen në fazën e zhvillimit gjatë ciklit jetësor të sistemit të inteligjencës artificiale, përfshinë sigurinë e zinxhirit të furnizimit, dokumentacionin dhe menaxhimin e aseteve.

1. Elementet e zhvillimit të sigurt:

i. të sigurohet ai që quhet "zinxhiri i furnizimit";

ii. të identifikohen, monitorohen dhe mbrohen asetet që disponohen;

iii. të dokumentohen të dhënat dhe modelet e IA-së.

C. Implementimi i sigurt: Kriteret që aplikohen në fazën e implementimit gjatë ciklit jetësor të zhvillimit të sistemit të inteligjencës artificiale përfshijnë, mbrojtjes e infrastrukturës dhe sistemeve IA nga kompromentimet, kërcënimet ose humbjet, zhvillimin e proceseve të menaxhimit të incidenteve etj.

1. Elementet e implementimit të sigurt:

i. duhet të merren masa për sigurimin e infrastrukturës;

ii. duhet të zhvillohen dhe testohen procedura të menaxhimit të incidenteve;

iii. përdorimi i sistemeve të IA-së duhet të jetë sa më i thjeshtuar për përdoruesin fundor;

iv. duhet të kryhet vlerësimi i ndikimit në mbrojtjen e të dhënave.

Ç. Operimi dhe mirëmbajtja e sigurt[4]: Kriteret që aplikohen në fazën e operimit dhe mirëmbajtjes gjatë ciklit jetësor të zhvillimit të sistemit të inteligjencës artificiale.

1. Elementet e operimit dhe mirëmbajtjes së sigurt:

i. të monitorohet vazhdimisht sjellja e sistemit;

ii. të monitorohen gjurmët e hyrjeve në sistem;

iii. sistemet e IA-së duhet të mbahen në çdo kohë të përditësuara;

iv. të mblidhen të dhëna dhe të përdoren si "mësime të nxjerra";

v. trajtimi i sistemit përtej ciklit jetësor të tij (arkivim, asgjësim).

D. Testimi i produkteve dhe performancës: institucioni duhet të sigurojë testimet e nevojshme për përcaktimin e pragut të saktësisë të sistemit të IA-së.

DH. Ndërfaqja dhe rrjetëzimi: institucioni duhet të sigurohet që produktet dhe sistemet IA të mund të punojnë së bashku, duke rritur shtrirjen dhe përdorimin e tyre pa shpërdorim.[5]

E. Integriteti: sigurimi i saktësisë dhe plotësisë së informacionit dhe metodave të përpunimit.

Ë. Burimet e të dhënave: në një sistem IA, të dhënat mund të merren nga disa burime, duke përfshirë sensorë (të tilla si kamerat CCTV, telefonat inteligjentë ose pajisjet mjekësore) dhe asetet digjitale (të tilla si të dhënat nga platformat e tregtimit, ekstraktet e dokumenteve ose skedarët e regjistrit). Të dhënat mund të jenë në shumë forma të ndryshme, duke përfshirë tekstin, imazhet, videon e audion dhe mund të jenë të strukturuara ose të pastrukturuara. Të dhënat mund të jenë të disponueshme nga burime të hapura, për shembull në internet ose me burime të mbyllur, si ato komerciale ose private, në atë masë që është e nevojshme dhe proporcionale për realizimin e qëllimit dhe kur subjekti i të dhënave ka dhënë pëlqimin e tij ose në çdo rast tjetër që parashikon

---

[4] *https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf*

[5] *https://aistandardshub.org/resource/main-training-page-example/2-different-types-of-standards/*

legjislacioni në fuqi për mbrojtjen e të dhënave personale. Të dhënat mund të merren nga bota reale, ose sintetike.

Grupet e të dhënave që përdoren për trajnimin e sistemeve IA mund të jenë shumë të mëdha dhe shpesh vijnë nga burime të shumta të shpërndara, duke krijuar grup të dhënash "trajnues". Asnjë kategori apo burim i të dhënave nuk është plotësisht imun ndaj risqeve të sigurisë prandaj ky dokument përcakton se në çdo rast duhet të tregohet vëmendje e shtuar që në zhvillimin dhe implementimin e sistemeve të IA-së të garantohet maksimalisht integriteti i të dhënave.

F. "Kurimi" i të dhënave: faza e kurimit ose e përpunimit përfshin një numër hapash të grumbullimit dhe transformimit, duke përfshirë ruajtjen e të dhënave, parapërpunimin, pasurimin e seteve të të dhënave dhe etiketimin.

G. Trajnimi dhe testimi i sistemeve IA: gjatë trajnimit, parametrat e modelit përshtaten në mënyrë që të arrihet një performancë më e saktë dhe kjo përsëritet derisa të arrihet një nivel i pranueshëm saktësie.

Të dhënat e trajnimit, validimit dhe testimit të sistemit duhet të jenë të përshtatshme, gjithëpërfshirëse, pa gabime dhe të plota. Duhet të kenë vetitë e duhura stastistikore, ku përshihen sipas nevojës, personat apo grupet e personave, për të cilët sistemi IA do të përdoret. Të dhënat duhen trajtuar sipas:

a) zgjidhjes përkatëse të projektimit të të dhënave;

b) grumbullimit të të dhënave;

c) përgatitja e procesimit, si: etiketimi, pastrimi, pasurimi dhe grumbullimi;

ç) vlerësimi paraprak për sasinë dhe përshtatshmërinë e grupit të të dhënave të nevojitura;

d) ekzaminimi për anësi të të dhënave;

dh) identifikimi i çdo boshllëku ose mangësie të mundshme të të dhënave dhe si mund të adresohen ato boshllëqe dhe mangësi.[6]

GJ. Ambienti i pastër kibernetik: integrimi i sigurisë kibernetike në të gjitha hapat e ciklit jetësor të IA-së është shumë i rëndësishëm; disa elemente bazë të sigurisë janë, janë si më poshtë:

- sulmet kibernetike;
- përditësimi i niveleve të rregullimit të sistemit;
- çelësa dhe fjalëkalimeve të mbrojtura;
- kontrolle të forta aksesi;
- plan reagimi ndaj incidenteve kibernetike.

H. Zinxhiri i furnizimit: përveç parimeve të gjera të higjienës kibernetike të mësipërme, poseduesi dhe përdoruesit e sistemit IA duhet të aplikojnë, gjithashtu, zinxhirin e furnizimit të burimeve që shërbejnë për funksionimin e këtij sistemi.

3. Standardet etike në zhvillimin e inteligjencës artificiale[7]

A. Përputhshmëria dhe minimizimi i dëmeve: përdorimi i sistemeve të inteligjencës artificiale nuk duhet të tejkalojë atë që është e nevojshme për të arritur një qëllim të ligjshëm.

B. Siguria: dëmtimet (risqet e sigurisë), si dhe vulnerabilitetet ndaj sulmeve të evitohen dhe adresohen nga aktorët e inteligjencës artificiale.

C. Të drejtat për privatësi dhe mbrojtje të të dhënave: privatësia të mbrohet dhe të promovohet gjatë gjithë ciklit të inteligjencës artificiale nëpërmjet strukturave të përshtatshme për mbrojtjen e të dhënave. Përdorimi i sistemeve të inteligjencës artificiale në Republikën e Shqipërisë duhet të jenë konform dhe të zbatojë rregullat dhe parimet e parashikuara në legjislacionin për mbrojtjen e të dhënave personale.

Ç. Qeverisja dhe bashkëpunimi: pjesëmarrja e të interesuarve të ndryshëm është e nevojshme për qasje përfshirëse në qeverisjen e inteligjencës artificiale.

D. Përgjegjësia dhe llogaridhënia: sistemet e inteligjencës artificiale të jenë të audituehsme dhe të gjurmuara nëpërmjet mekanizmave të mbikëqyrjes, vlerësimit të ndikimit, auditimit dhe

---

[6] *https://artificialintelligenceact.eu/wp-content/uploads/2024/02/AIA-Trilogue-Coreper.pdf*

[7] *https://www.unesco.org/en/artificial-intelligence/recommendation-ethics*

mekanizmave të kujdesit të duhur për të evituar konflikte me normat e të drejtave të njeriut dhe kërcënime ndaj mirëqenies mjedisore.

DH. Transparenca dhe shpjegueshmëria: Zhvillimi etik i sistemeve të inteligjencës artificiale varet nga transparenca dhe shpjegueshmëria e tyre. Niveli i këtyre të dyjave të jetë i përshtatshëm për kontekstin.

E. Mbikëqyrja njerëzore dhe vendosmëria: institucionet të sigurojnë që sistemet e inteligjencës artificiale nuk zëvendësojnë përgjegjësinë dhe llogaridhënien përfundimtare njerëzore.

Ë. Ndërgjegjësimi: edukimit gjithëpërfshirës, aftësimit digjital dhe trajnimit në përdorimin etikë të inteligjencës artificiale, si dhe publikimit në media.

F. Drejtësia dhe mosdiskriminimi: avantazhet e inteligjencës artificiale janë të arritshme për të gjithë, janë të drejta dhe mosdiskriminuese.

KREU III
PARIMET[8]

Në një epokë ku teknologjia po avancon me shpejtësi dhe inteligjenca artificiale po bëhet një pjesë e pandarë e jetës së përditshme të qytetarëve, është e rëndësishme të kemi udhëzime dhe rregulla të qarta që të sigurojmë përdorimin etik dhe drejtë të kësaj teknologjie të fuqishme. Për këtë arsye, duhet të hartohen parimet e inteligjencës artificiale, që synojnë të rregullojnë përdorimin e IA-së në mënyrë etike dhe të drejtë.

1. Parimet e vlerave

Parimi 1.1: Rritja e qëndrueshme, gjithëpërfshirëse dhe mirëqenia e qytetarëve

Ky parim thekson potencialin e inteligjencës artificiale për të kontribuuar në rritjen dhe prosperitetin e përgjithshëm, për të avancuar objektivat e zhvillimit kombëtar dhe global.

Ky parim njeh se udhëzimi i zhvillimit dhe përdorimit të inteligjencës artificiale drejt mirëqenies dhe rezultateve të favorshme për njerëzit dhe vendin është një prioritet. Inteligjenca artificiale e besueshme luan një rol të rëndësishëm në avancimin e rritjes inkluzive, zhvillimin e qëndrueshëm të mirëqenies dhe në objektivat e zhvillimit kombëtar. Ndër të tjera, në arritjen e qëllimeve të zhvillimit të qëndrueshëm në fusha, si arsimi, shëndetësia, transporti, bujqësia, mjedisi dhe qytetet e qëndrueshme.

Ky parim, gjithashtu, njeh se sistemet e inteligjencës artificiale mund të konsolidojnë paragjykimet ekzistuese dhe të kenë një ndikim të ndryshëm në popullatat me risk më të lartë për t'u paragjykuar, si minoritetet etnike, gratë, personat me aftësi ndryshe, fëmijët, të moshuarit dhe ata me arsim më të ulët.

Parimi 1.2: Vlerat e orientuara nga njeriu dhe drejtësia

Sistemet e inteligjencës artificiale duhet të jenë të projektuara në mënyrë që të respektojnë sundimin e ligjit, të drejtat e njeriut, vlerat demokratike dhe diversitetin dhe duhet të përfshijnë mbrojtjet e duhura për të siguruar një shoqëri të drejtë dhe të ndershme.

Aktorët e IA-së duhet të respektojnë sundimin e ligjit, të drejtat e njeriut dhe vlerat demokratike gjatë gjithë ciklit jetësor të sistemit të IA-së. Këto vlera përfshijnë lirinë, dinjitetin dhe autonominë, privatësinë dhe mbrojtjen e të dhënave, të drejtat e qytetarëve, pronësinë intelektuale dhe të drejtat tregtare, mosdiskriminimin dhe barazinë, diversitetin, drejtësinë shoqërore dhe të drejtat ligjore të punës. Për këtë qëllim, aktorët e IA-së duhet të zbatojnë mekanizma dhe mbrojtje, të garantojnë më çdo rast mundësinë për vendimmarrje njerëzore dhe të përshtaten sipas kontekstit të sektorit ku ato zbatohen ose funksionojnë.

Duke i qëndruar besnik vlerave të përbashkëta demokratike, ky parim do të ndihmojë në forcimin e besimit të publikut në IA dhe do të garantojë përdorimi i IA-së për të mbrojtur të drejtat e njeriut dhe për të reduktuar diskriminimin ose rezultatet tjera të padrejta dhe/ose të pabarabarta.

---

[8] *https://oecd.ai/en/ai-principles*

Ky parim, gjithashtu, njeh rolin e masave, si dhe vlerësimeve të ndikimit të të drejtave të njeriut, vendimmarrjes njerëzore (pra, "njeriu mbi çdo gjë"), kodeve të sjelljes etike ose shënimeve dhe certifikatave, të cilat synojnë të promovojnë vlerat e orientuara nga njeriu dhe drejtësinë.

Parimi 1.3: Transparenca dhe shpjegueshmëria

Ky parim ka të bëjë me transparencën dhe dhënien e informacionit në lidhje me sistemet e IA-së për të siguruar që përdoruesit të interpretojnë të dhënat e prodhuara të saj, si dhe të kuptojnë kur janë duke komunikuar me sistemet IA dhe të kenë mundësinë të "sfidojnë" rezultatet.

Aktorët e inteligjencës artificiale duhet të angazhohen në transparencë dhe përgjegjshmëri në lidhje me sistemet e inteligjencës artificiale. Për këtë qëllim, ata duhet të sigurojnë informacion të kuptueshëm, të përshtatshëm për kontekstin ku ato zbatohen dhe në përputhje me zhvillimet më të fundit të teknologjisë:

- për të nxitur një kuptim të përgjithshëm të sistemeve të IA-së;
- për të bërë përdoruesit e interesuar të kenë vetëdije për ndërveprimet e tyre me sistemet e IA-së, përfshirë edhe në vendin e punës;
- për të mundësuar ata që preken nga një sistem IA të kuptojnë rezultatin; dhe
- për të mundësuar ata që preken negativisht nga një sistem IA të sfidojnë rezultatin e tij, bazuar në informacion të qartë dhe të lehtë perceptueshëm mbi faktorët dhe logjikën që shërbyen si bazë për parashikimin, rekomandimin ose vendimin.

Parimi 1.4: Qëndrueshmëria, siguria dhe mbrojtja

Sistemet e IA-së duhet të funksionojnë në një mënyrë të qëndrueshme, të sigurt dhe të mbrojtur gjatë gjithë jetëgjatësisë së tyre dhe rreziqet potenciale duhet që vazhdimisht të vlerësohen dhe menaxhohen.

Sistemet e IA-së duhet të jenë të qëndrueshme, të sigurta dhe të mbrojtura gjatë gjithë ciklit të tyre të jetës, duke parashikuar që në kushte të përdorimit normal, keqpërdorimit ose kushteve të tjera jo të përshtatshme, ato funksionojnë në mënyrë adekuate dhe nuk sjellin rrezik të papërballueshëm për sigurinë.

Aktorët e IA-së duhet të aplikojnë, në bazë të rolit të tyre, situatës dhe aftësisë së tyre për të vepruar, një qasje sistematike të menaxhimit të rrezikut për çdo fazë të ciklit të sistemit të IA-së në një bazë të vazhdueshme për të adresuar rreziqet lidhur me sistemet e IA-së, duke përfshirë privatësinë, sigurinë digjitale dhe mbrojtjen.

Ky parim, gjithashtu, përcakton se sistemet e IA-së nuk duhet të sjellin rrezik të papërballueshëm për sigurinë, përfshirë sigurinë fizike, në kushte të përdorimit normal ose keqpërdorimit gjatë gjithë ciklit të tyre të jetës.

Aktorët e IA-së mund të përdorin një qasje të menaxhimit të rrezikut për të identifikuar dhe mbrojtur nga keqpërdorimi i parashikueshëm, si dhe nga rreziqet e lidhura me përdorimin e sistemeve të IA-së për qëllime të tjera, përveç atyre për të cilat janë projektuar fillimisht. Çështjet e qëndrueshmërisë, sigurisë dhe mbrojtjes së IA-së janë të lidhura për të mbajtur sistemet e IA-së të qëndrueshme, të sigurta dhe të mbrojtura. Ato zbatohen në dy principe:

1. Gjurmimi, analiza dhe hetimi i mëtejshëm; dhe
2. Aplikimi i një qasjeje të menaxhimit të rrezikut.

Ashtu sikurse shpjegueshmëria (shih parimin 1.3), gjurmimi mund të ndihmojë analizën dhe hetimin e rezultateve të një sistemi IA dhe është një mënyrë për të promovuar përgjegjshmërinë. Gjurmimi ndryshon nga shpjegueshmëria në faktin se fokusi është në mbajtjen e regjistrave të tipareve të të dhënave, si metadata, burimet e të dhënave dhe pastrimi i të dhënave, por jo domosdoshmërisht të dhënat vetë. Në këtë mënyrë, gjurmimi mund të ndihmojë në kuptimin e rezultateve, në parandalimin e gabimeve të ardhshme dhe në përmirësimin e besueshmërisë së sistemit të IA-së.

Qasja e menaxhimit të rrezikut:

Një qasje e menaxhimit të rrezikut do të aplikohet gjatë gjithë ciklit të sistemit të IA-së për të ndihmuar në identifikimin, vlerësimin, përdorimin e prioriteteve dhe zbutjen e rreziqeve potenciale, që mund të ndikojnë negativisht në sjelljen dhe rezultatet e një sistemi.

Kufizime[9]:

a) vendosja në treg e një sistemi IA, i cili përdor teknika autonome të cilat manipulative ose teknika mashtruese dhe që synojnë efektin e ndryshimit të sjelljes së individëve, duke i dëmtuar gjykimin për një vendimmarrje të informuar dhe duke i shkaktuar dëm të konsiderueshëm.

b) vendosja në treg ose vënia në shërbim ose përdorimi i një sistemi IA, që shfrytëzon ndonjë nga dobësitë e individëve, për shkak të moshës së tyre, paaftësisë ose një situatë specifike sociale ose ekonomike, me qëllim ose efekt të shtrembërimit material të sjelljes që i përket këtyre individëve, e cila mund të shkaktojë dëm të konsiderueshëm tek ata ose të tjerët.

c) vendosja në treg ose vënia në shërbim e sistemit IA për qëllim specifik, duke përdorur karakteristika fizike të individëve për të kategorizuar këta individë sipas racës, gjinisë, opinionet politike, besimin fetar, jetës seksuale ose orientimit seksual.

ç) vendosja në treg ose vënia në shërbim ose përdorimi i sistemeve të IA-së për vlerësimin ose klasifikimin e individëve për një periudhë të caktuar të kohës, bazuar në sjelljen e tyre shoqërore ose personale të njohura ose karakteristika të personalitetit të tyre, të cilat çojnë në:

i. trajtimin e dëmshëm ose të pafavorshëm të individëve në aspektin social, që nuk kanë lidhje me të dhënat, janë gjeneruar ose mbledhur fillimisht;

ii. trajtimin e dëmshëm ose të pafavorshëm të individëve në mënyrë të pajustifikuar ose në disproporcion me sjelljen e tyre shoqërore.

Parimi 1.5: Përgjegjshmëria

Institucionet publike, organizatat, bizneset dhe individët që zhvillojnë, implementojnë ose operojnë sisteme IA duhet të mbajnë përgjegjësi për funksionimin e tyre të duhur në përputhje me parimet e kësaj metodologjie, bazën ligjore përkatëse dhe vlerat njerëzore e shoqërore.

Aktorët e IA-së duhet të jenë përgjegjës për funksionimin e duhur të sistemeve të IA-së dhe për respektimin e parimeve të mësipërme, bazuar në rolin e tyre, situatën dhe përgjegjshmërinë për zhvillimet më të fundit të teknologjisë.

Në këtë rast "përgjegjësia" ka të bëjë me prirjen që institucionet publike, bizneset, organizatat ose individët do të sigurojnë funksionimin e duhur, gjatë gjithë jetëgjatësisë së tyre, të sistemeve IA që ata projektuan, zhvilluan, operuan ose implementuan, në përputhje me rolin e tyre dhe kornizat ligjore të zbatueshme, dhe për t'i dëshmuar këto gjëra nëpërmjet veprimeve dhe proceseve të tyre të vendimmarrjes (për shembull, duke ofruar dokumentacion mbi vendimet kyçe gjatë ciklit të sistemit të IA-së ose duke lejuar auditimin kur është i arsyetueshëm).

2. Parimet organizative dhe teknike

Parimi 2.1: Investimi në hulumtimin dhe zhvillimin e inteligjencës artificiale (IA)

Do të konsiderohen investimet publike afatgjata dhe investimet private në hulumtim dhe zhvillim, duke përfshirë përpjekjet ndërdisiplinore për të nxitur inovacionin në fushën e inteligjencës artificiale të besueshme, duke u fokusuar në sfidat teknike dhe në çështjet sociale, ligjore dhe etike të lidhura me inteligjencën artificiale dhe çështjet e politikave. Gjithashtu, do të konsiderohen investimet publike dhe inkurajohen investimet private në të dhënat e hapura, duke respektuar privatësinë dhe mbrojtjen e të dhënave, për të mbështetur një mjedis për hulumtim dhe zhvillim në fushën e inteligjencës artificiale, që zbaton parimet etike dhe përmirëson interoperabilitetin dhe përdorimin e standardeve.

Zbulimet shkencore të mundësuara nga inteligjenca artificiale mund të ndihmojnë në zgjidhjen e sfidave shoqërore dhe të krijojnë industri të reja në mënyrë të plotë. Këto mundësi theksojnë rëndësinë e kërkimit bazik dhe të konsiderimit të horizonteve të gjata kohore në politikën e kërkimit. Ky lloj investimi është i domosdoshëm për të shtyrë dhe përcaktuar inovacionin e inteligjencës artificiale të besueshme dhe për të siguruar rezultate të dobishme për të gjithë, veçanërisht për zonat që kanë vend për përmirësim në aspektin e investimeve të drejtuara nga kërkesa e tregut. Kërkimi në fushën e inteligjencës artificiale përfshin studime mbi aplikacionet e inteligjencës artificiale, si: procesimi i gjuhës natyrore; teknikat për mësimin e sistemeve të

[9] *https://artificialintelligenceact.eu/documents/*

inteligjencës artificiale, si rrjetet neurale; optimizimi, në veçanti për të reduktuar sasinë e të dhënave të nevojshme për zhvillimin e inteligjencës artificiale, siç është parimi i mbledhjes së të dhënave dhe studimet që adresojnë konsideratat shoqërore, si transparenca dhe shpjegueshmëria, si dhe teknologjitë për të mbrojtur integritetin e të dhënave.

Përveç kësaj, për shkak të rëndësisë së të dhënave për ciklin e jetës së sistemeve të inteligjencës artificiale, një element kyç në sigurimin e hulumtimit të mëtejshëm dhe më të mirë të inteligjencës artificiale është disponueshmëria e seteve të të dhënave të hapura, të arritshme dhe të përfaqësueshme që nuk kompromentojnë privatësinë dhe mbrojtjen e të dhënave personale dhe të konsumatorëve, të drejtat e pronësisë intelektuale dhe të drejtat e tjera të rëndësishme. Veçanërisht, ndërsa është e pamundur të arrihet një mjedis krejtësisht "me të dhëna etike", duke ofruar (dhe duke dhënë incentiva) për grupe të dhënash që janë përfaqësuese dhe të disponueshme publikisht, qeveria mund të ndihmojë në zvogëlimin e rreziqeve të përmbajtjeve joetike, të padobishme në sistemet e inteligjencës artificiale.

Parimi 2.2: Nxitja e një ekosistemi digjital për inteligjencën artificiale

Ky lloj ekosistemi përfshin në mënyrë të veçantë teknologjitë digjitale dhe infrastrukturën, si dhe mekanizmat për ndarjen e njohurive të inteligjencës artificiale, sipas nevojës. Në këtë kontekst, aktorët e IA-së duhet të marrin në konsideratë promovimin e mekanizmave, siç janë "*data trusts*", për të mbështetur ndarjen e të dhënave në mënyrë të sigurt, të drejtë, ligjore dhe etike.

Zhvillimi i inteligjencës artificiale të besueshme kërkon një ekosistem lehtësues.

Teknologjitë digjitale dhe infrastruktura e nevojshme përfshijnë qasjen në rrjetet dhe shërbimet e internetit të shpejtësisë së lartë, fuqinë e kompjuterizimit dhe ruajtjen e të dhënave – si dhe teknologjitë që gjenerojnë të dhëna si interneti i gjërave (IoT). Gjithashtu, janë të nevojshme mekanizmat e përshtatshëm për ndarjen e njohurive të inteligjencës artificiale, duke përfshirë të dhënat, kodin, algoritmet, modelet, studimet dhe njohuritë, për të kuptuar dhe marrë pjesë në ciklin e jetës së sistemit të inteligjencës artificiale. Këto mekanizma duhet të respektojnë privatësinë, pronën intelektuale dhe të drejtat e tjera. Mjetet e ardhura nga burime të hapura për menaxhimin dhe përdorimin e inteligjencës artificiale, të cilat lejojnë përhapjen e teknologjisë së inteligjencës artificiale dhe gjetjen e zgjidhjeve të *software*-ve, përmes të ashtuquajturit "*crowdsourcing*", luajnë një rol kyç në zhvillimin e inteligjencës artificiale.

Parimi 2.3: Krijimi i një mjedisi politik lehtësues për inteligjencën artificiale

Ky rekomandim përqendrohet në mjedisin politik që mundëson inovacionin në fushën e inteligjencës artificiale, pra kornizat institucionale, politike dhe ligjore. Duke marrë parasysh ritmin e shpejtë të zhvillimeve në fushën e inteligjencës artificiale, krijimi i një mjedisi politik që është fleksibël aq sa të mund të ndjekë zhvillimet dhe të promovojë inovacionin, por që mbetet i sigurt dhe siguron mbrojtjen ligjore është një sfidë e rëndësishme. Në nxitjen e përdorimit të inteligjencës artificiale duhet të ndiqet një qasje që është orientuar drejt njeriut.

Aktorët e IA-së duhet të promovojë një mjedis politik që mbështet një kalim të shpejtë nga faza e kërkimit dhe e zhvillimit në fazën e implementimit dhe të operimit të sistemeve të besueshme të inteligjencës artificiale. Për këtë qëllim, do të merret parasysh përdorimi i pilotimit për të ofruar një mjedis të kontrolluar, ku sistemet e inteligjencës artificiale mund të testohen dhe, sipas nevojës, të shtrihen në masë më të madhe.

Qasja e kornizave politike dhe rregullative, si dhe mekanizmat e vlerësimit të tyre që aplikohen në sistemet e inteligjencës artificiale për të nxitur inovacionin dhe konkurrencën në fushën e inteligjencës artificiale të besueshme duhet të jenë objekt përmirësimi dhe përshtatjeje të vazhdueshme.

Theksohet roli i eksperimentimit si një mjet për të krijuar mjedise të kontrolluara dhe transparente, në të cilat sistemet e inteligjencës artificiale mund të testohen dhe në të cilat modelet e biznesit të bazuar në inteligjencën artificiale që mund të promovojnë zgjidhje për sfidat kombëtare mund të zhvillohen. Eksperimentet politike mund të operojnë në "mënyrë *start-up*", ku eksperimentet janë të implementuara, të vlerësuara dhe të modifikohen e pastaj shtrihen ose zvogëlohen apo ndërpriten, në varësi të rezultateve të testit.

Parimi 2.4: Ndërtimi i kapacitetit njerëzor dhe përgatitja për transformimin e tregut të punës

Do të jetë në fokus kryesor fuqizimi i njerëzve për të përdorur dhe për të ndërvepruar efektivisht me sistemet e inteligjencës artificiale nëpër shumëllojshmëri aplikimesh, duke përfshirë pajisjen e tyre me shkathtësitë dhe njohuritë e nevojshme.

Inteligjenca artificiale do të ndryshojë natyrën e shumë aspekteve të jetës, duke u përhapur nëpër sektorë. Kjo është veçanërisht e vërtetë në kontekstin e punës, punësimit dhe vendit të punës – ku inteligjenca artificiale do të përfundojë me rolin e njeriut në disa detyra, do t'i zëvendësojë ato në detyra të tjera dhe do të krijojë lloje të reja të punës dhe organizimit të punës.

Kjo përfshin aftësitë e nevojshme të praktikantëve të inteligjencës artificiale (të cilat aktualisht mungojnë) dhe ato që nevojiten për punonjësit e tjerë (si mjekët etj.) për të marrë përfitim nga inteligjenca artificiale në fushat e tyre të ekspertizës, në mënyrë që IA-ja të shtojë aftësitë njerëzore. Paralelisht, politikat e zhvillimit të aftësive do të duhet të përqendrohen në aspektet e veçanta humane të nevojshme për të shfrytëzuar sistemet e inteligjencës artificiale, si vlerësimi, mendimi krijues e kritik dhe komunikimi.

Gjithashtu, ndryshimet e drejtuara nga inteligjenca artificiale në tregun e punës mund të kërkojnë përshtatjen ose pranimin e standardeve të punës dhe të marrëveshjeve midis drejtuesve dhe punonjësve, kur është e aplikueshme, për të reflektuar këto ndryshime, adresuar sfidat e mundshme për barazi, diversitet dhe drejtësi (siç paraqiten, për shembull, nga mbledhja dhe procesimi i të dhënave) dhe për të përforcuar vendet e punës të besueshme, të sigurta dhe produktive. Kjo mund të arrihet përmes një kombinimi rregullimi, dialogut social dhe bisedave kolektive. Është e rëndësishme të lejohet fleksibiliteti në vendin e punës duke ruajtur autonominë dhe cilësinë e punës për punëtorët.

Parimi 2.5: Bashkëpunimi ndërkombëtar për inteligjencën artificiale të besueshme

Palët e interesuara duhet të bashkëpunojnë aktivisht për të avancuar këto parime dhe për të përparuar në menaxhimin e përgjegjshëm të inteligjencës artificiale të besueshme.

Bashkëpunimi ndërkombëtar përfshin përparimin dhe përhapjen e implementimit të këtyre parimeve dhe politikave me të gjitha vendet partnere dhe jo vetëm, duke përfshirë vendet e zhvilluara dhe ato në zhvillim, si dhe vendet më pak të zhvilluara, dhe me palët e tjera të interesuara.

Bashkëpunimi ndërkombëtar mund të përdorë forume ndërkombëtare dhe rajonale për të ndarë njohuritë mbi inteligjencën artificiale në mënyrë që të ndërtohet një ekspertizë e gjerë mbi këtë fushë; për të zhvilluar standarde teknike për inteligjencën artificiale që janë ndërveprues dhe të besueshëm; dhe për të zhvilluar, përhapur dhe përdorur parametra matës për të vlerësuar performancën e sistemeve të inteligjencës artificiale, si: saktësia, efikasiteti, përparimi i qëllimeve shoqërore, drejtësisë dhe qëndrueshmërisë. Bashkëpunimi ndërkombëtar, gjithashtu, mund të kontribuojë në përpunimin e sasive të mëdha të të dhënave mbi parimin e besueshmërisë, duke ruajtur sigurinë, privatësinë, pronën intelektuale, të drejtat e njeriut, vlerat demokratike dhe që janë të rëndësishme për inovacionin e inteligjencës artificiale.

KREU IV
SFIDAT DHE RISQET

1. Vlefshmëria dhe besueshmëria[10]: Që sistemet e IA-së të jenë të besueshme, ato shpesh duhet t'i përgjigjen një sërë kriteresh që kanë vlerë për palët e interesuara.

Vlefshmëria është "konfirmimi, nëpërmjet ofrimit të provave objektive, se kërkesat për një përdorim apo aplikim specifik të synuar janë përmbushur". Implementimi i sistemeve të IA-së që janë të pasaktë, jo të besueshëm ose të ndërtuar në mënyrë jo të duhur për të mbrojtur të dhënat, rrit risqet e sistemeve të IA-së dhe redukton besueshmërinë. Besueshmëria përcaktohet në të

[10] *https://nvlpubs.nist.gov/nistpubs/IA/NIST.IA.100-1.pdf*

njëjtin standard si "aftësia e një artikulli për të performuar", siç kërkohet, pa dështim, për një interval kohor të caktuar, në kushte të dhëna.

Matjet e saktësisë duhet të shoqërohen gjithmonë me grupe testesh të përcaktuara qartë dhe realiste, që janë përfaqësuese të kushteve të përdorimit të parashikuar që pritet të ndodhë në kushte reale dhe detaje rreth metodologjisë së testimit, këto duhet të përfshihen në dokumentacionin përkatës.

Matjet e saktësisë mund të përfshijnë ndarjen e rezultateve për segmente të ndryshme të të dhënave.

Grupet e testimit teknik duke përdorur praktikat më të mira të analizës së performancës duhet të jenë të mirëpërcaktuara në çdo fazë të implementimi dhe përditësimi dhe në fazën e mirëmbajtjes së sistemeve IA.

Vlefshmëria dhe besueshmëria për sistemet e inteligjencës artificiale të implementuara shpesh vlerësohen nga testimi ose monitorimi i vazhdueshëm, që konfirmon se një sistem po funksionon, siç është parashikuar dhe ndërtuar që të funksionojë dhe, gjithashtu, në përputhje me bazën ligjore në fuqi. Përpjekjet për menaxhimin e rrezikut të IA-së duhet të kenë prioritet minimizimin e ndikimeve të mundshme negative dhe mund të kenë nevojë të përfshijnë ndërhyrjen njerëzore në rastet kur sistemi i IA-së nuk mund të zbulojë ose korrigjojë automatikisht gabimet e tij.

Qëndrueshmëria ose përgjithësimi përkufizohet, si: "aftësia e një sistemi për të ruajtur nivelin e tij të performancës në rrethana të ndryshme". Qëndrueshmëria është një qëllim për funksionalitetin e duhur të sistemit në një grup të gjerë kushtesh dhe rrethanash, duke përfshirë përdorime të sistemeve të IA-së të paparashikuara që në fillim të dizenjimit të tij. Qëndrueshmëria kërkon jo vetëm që sistemi të funksionojë saktësisht, siç funksionon në përdorimet e parashikuara, por, gjithashtu, duhet të funksionojë në mënyrë të tillë që të minimizojnë efektet negative të mundshme për njerëzit nëse funksionon në një mjedis të paparashikuar që në dizenjimin e sistemit.

Besueshmëria ka për qëllim korrektësinë e përgjithshme të funksionimit të sistemit IA-së në kushtet e përdorimit të pritshëm për një periudhë të caktuar kohe, duke përfshirë të gjithë jetëgjatësinë e sistemit[11]. Saktësia dhe qëndrueshmëria kontribuojnë në vlefshmërinë dhe besueshmërinë e sistemeve të IA-së dhe mund të jenë në tension me njëri-tjetrin në sistemet e IA-së. Saktësia përkufizohet nga ISO/IEC TS 5723:2022, si: "afërsi e rezultateve të vëzhgimeve, llogaritjeve ose vlerësimeve me vlerat e vërteta ose vlerat e pranuara si të vërteta".

Qasjet që rrisin besueshmërinë e IA-së mund të zvogëlojnë risqet negative të IA-së. Në këtë metodologji përcaktohen karakteristikat të besueshme të IA-së, si më poshtë:

- të vlefshme dhe të besueshme;
- të sigurta dhe elastike;
- të përgjegjshme dhe transparente;
- të shpjegueshme dhe të interpretueshme;
- të përmirësuara nga privatësia;
- të drejta.

Karakteristikat e besueshmërisë janë të lidhura ngushtë me sjelljen sociale dhe organizative, grupet e të dhënave të përdorura nga sistemet e IA-së, përzgjedhjen e modeleve dhe algoritmeve të IA-së dhe vendimet e marra nga ata që i ndërtojnë ato.

Krijimi i sistemeve IA-së të besueshme kërkon balancimin e secilës prej këtyre karakteristikave, bazuar në kontekstin e përdorimit të sistemit të IA-së. Ndërsa të gjitha karakteristikat janë atribute të sistemit socioteknik, llogaridhënia dhe transparenca lidhen, gjithashtu, me proceset dhe aktivitetet e brendshme të një sistemi IA dhe mjedisin e tij të jashtëm. Neglizhenca e këtyre karakteristikave mund të rrisë probabilitetin dhe ndikimin e pasojave të mundshme negative.

Gjykimi njerëzor duhet të përdoret kur vendoset për elementet matës specifikë, që lidhen me karakteristikat e besueshmërisë së IA dhe limitet e pranueshme të pragut për ato elemente.

---

[11] *Page 13 NIST IA 100-1 IA RMF 1.0*

2. Siguria dhe elasticiteti: sistemet e IA-së nuk duhet "në kushte të përcaktuara", të çojnë në rezultate në të cilën jeta, shëndeti, prona ose mjedisi i njeriut është i rrezikuar.

A. Funksionimi i sigurt i sistemeve të IA-së përmirësohet nëpërmjet:

- praktikave të përgjegjshme të projektimit, zhvillimit dhe implementimit;
- informacion të qartë për implementuesit mbi përdorimin e përgjegjshëm të sistemit;
- vendimmarrje e përgjegjshme nga implementuesit dhe përdoruesit fundorë;
- shpjegimet dhe dokumentacioni i risqeve bazuar në dëshmi të matshme të incidenteve.

Llojet e ndryshme të risqeve të sigurisë mund të kërkojnë qasje të përshtatura të menaxhimit të riskut bazuar në kontekstin dhe nivelin e kritikalitetit të tyre. Risqet e sigurisë të nivelit të lartë që mund të shkaktojnë lëndime serioze kërkojnë prioritizimin më urgjent dhe procesim më të plotë të menaxhimit të rrezikut.[12]

Qasjet e menaxhimit të rrezikut të sigurisë së IA-së duhet të ndjekin udhëzimet për sigurinë në fusha të tilla, si transporti, kujdesi shëndetësor dhe fusha të tjera të rëndësishme të jetës shoqërore dhe ekonomike, të cilat përputhen me bazën ligjore në fuqi dhe me udhëzimet ose standardet më të mira ndërkombëtare të sektorit.

Sistemet e IA-së, si dhe ekosistemet në të cilat ato vendosen, mund të thuhet se janë elastike nëse mund t'i rezistojnë ngjarjeve të padëshiruara të papritura ose ndryshimeve të papritura në mjedisin ku ato funksionojnë ose në mënyrën e përdorimit të tyre ose nëse mund të ruajnë funksionet dhe strukturën e tyre përballë problemeve të brendshme. Shqetësimet e zakonshme të sigurisë kanë të bëjnë me sulmet kibernetike, dëmtimin e të dhënave, dëmtimin e pronës intelektuale etj.

Sistemet e IA-së duhet të ruajnë konfidencialitetin, integritetin dhe disponueshmërinë përmes mekanizmave mbrojtës që parandalojnë aksesin dhe përdorimin e paautorizuar që të konsiderohen të sigurta. Siguria dhe qëndrueshmëria janë karakteristika të lidhura, por të dallueshme. Ndërsa elasticiteti është aftësia për t'u rikthyer në funksionin normal pas një ngjarjeje të padëshiruar dhe të papritur.

Siguria përfshin elasticitetin, por gjithashtu përfshin protokolle për të shmangur, mbrojtur, për t'iu përgjigjur ose për t'u rikuperuar nga sulmet.

3. Përgjegjshmëria dhe transparenca: Inteligjenca artificiale e besueshme varet nga përgjegjësia.

Llogaridhënia presupozon transparencë. Transparenca pasqyron shkallën në të cilën informacioni për një sistem IA dhe rezultatet e tij është i disponueshëm për institucionet, bizneset dhe qytetarët që ndërveprojnë me një sistem të tillë. Transparenca kuptimplote siguron akses në nivelet e duhura të informacionit, bazuar në fazën e ciklit jetësor të IA-së dhe të përshtatur me rolin ose njohuritë e aktorëve ose individëve të IA-së, që ndërveprojnë ose përdorin sistemin e IA-së. Duke promovuar nivele më të larta transparence rrjedhimisht rritet besimi në sistemet e IA-së. Shtrirja e kësaj karakteristike varet nga mënyra e konceptimit dhe dizenjimit të sistemit përpara implementimit të tij.

Transparenca është shpesh e nevojshme për korrigjimin e veprimit në lidhje me rezultatet e sistemit të IA-së që janë të pasakta ose çojnë në ndikime negative. Transparenca duhet të marrë në konsideratë ndërveprimin njeri-IA.[13]

Roli i aktorëve të IA-së duhet të merret parasysh kur kërkohet llogaridhënie për rezultatet e sistemeve të IA-së. Marrëdhënia midis riskut dhe llogaridhënies që lidhet me IA dhe sistemet teknologjike ndryshon gjerësisht sipas konteksteve kulturore, ligjore, sektoriale dhe shoqërore. Kur pasojat janë të rënda, si për shembull, kur jeta dhe liria janë në rrezik, zhvilluesit dhe implementuesit e IA-së duhet të konsiderojnë rregullimin proporcional dhe proaktiv të praktikave të tyre të transparencës dhe llogaridhënies.

Zbatimi i praktikave organizative dhe strukturore për reduktimin e dëmtimeve të mundshme, menaxhimin e riskut, do të ndihmojë në përdorimin e sistemeve në mënyrë më të përgjegjshme.

---

[12] *Page 14 NIST IA 100-1 IA RMF 1.0.*

[13] *Faqe 15 NIST IA 100-1 IA RMF 1.0.*

Masat për të rritur transparencën dhe llogaridhënien do të marrin në konsideratë ndikimin e këtyre përpjekjeve në njësinë zbatuese. Mjetet e transparencës janë të gjitha elementet manuale ose automatike, të cilat si pjesë integrale e një sistemi IA ose si pjesë e dokumentimit të këtij të fundit, garantojnë që të gjithë përdoruesit e sistemit të jenë të vetëdijshëm për llojin e të dhënave që përpunohen, mënyrën sesi përpunohen dhe mekanizmat që përdoren për të gjeneruar rezultatet. Këto mjete implementohen nga zhvilluesit e sistemeve IA, të cilët mund të zhvillojnë sisteme të IA-së në interes të tyre ose për përdorim nga palë të treta të cilat mund të jenë subjekte publike ose private. Zhvilluesit e sistemeve IA janë të gjithë personat fizikë ose juridikë, publikë apo privatë që nisur si rezultat i një kërkese apo procedure ose me iniciativën e tyre ndërtojnë (zhvillojnë) një sistem të IA-së.

Zhvilluesit janë një nga aktorët e të ciklit jetësor të sistemeve të IA-së. Ndërkohë, që në tërësi aktorët janë të gjithë subjektet që në një fazë të ciklit jetësor të sistemeve IA ndërveprojnë, ndërtojnë, zhvillojnë apo përdorin këto sisteme.

4. Fokusimi drejt privatësisë: Privatësia u referohet normave dhe praktikave që ndihmojnë në ruajtjen e autonomisë, identitetit dhe dinjitetit njerëzor. Këto norma dhe praktika zakonisht trajtojnë ndërhyrjet e paautorizuara, marrjen e aprovimit të individëve për të dhënë pëlqimin për aksesin ose kontrollin e aspekteve të identitetit të tyre (p.sh., trupi, të dhënat, reputacioni).

Vlerat e privatësisë si anonimiteti, konfidencialiteti dhe kontrolli në përgjithësi duhet të udhëheqin zgjedhjet për hartimin, zhvillimin dhe vendosjen e sistemit të IA-së. Risqet e lidhura me privatësinë mund të ndikojnë në sigurinë dhe paragjykimet. Sistemet e IA-së mund të paraqesin gjithashtu risqe të reja për privatësinë, duke lejuar identifikimin e individëve përmes informacioneve, të cilët më parë ishin private.

KREU V
UDHËZIME

Këto udhëzime synojnë të maksimizojnë përfitimet e IA-së, të minimizojnë rrezikun dhe të sigurojnë mbrojtjen e të drejtave të njeriut. Kjo metodologji përcakton qartë se përgjegjësia kryesore për sistemet e IA-së do të jetë tek ato subjekte publike apo private që financojnë, që zhvillojnë dhe që implementojnë këto sisteme:[14]

E drejta për transparencë. Të gjithë individët kanë të drejtë të informohen nëse sistemi që ata kanë në përdorim ka të implementuar inteligjencën artificiale. Gjithashtu, në rastin e gjenerimit ose të ndryshimit të të dhënave, të cilat ngjasojnë me persona, objekte vende ose entitete ose evente të tjera, dhe për përdoruesin ngjasojnë si të vërteta apo autentike, duhet të informojnë që përmbajtja është e gjeneruar apo manipuluar artificialisht.[15]

E drejta për vendimmarrje njerëzore. Të gjithë individët kanë të drejtën e një vendimi përfundimtar të bërë nga një person human, pavarësisht sugjerimit të bërë nga sistemi i IA-së.

Vlerësimi dhe detyrimi i përgjegjshmërisë. Një sistem IA duhet të implementohet vetëm pas një vlerësimi adekuat të qëllimit dhe objektivave të tij, përfitimeve të tij, si dhe risqeve të tij.

Detyrimet e saktësisë, besueshmërisë dhe vlefshmërisë. Sistemi IA duhet të sigurojnë saktësinë, besueshmërinë dhe vlefshmërinë e vendimeve të propozuara.

Detyrimi i fikjes. Subjekti që ka krijuar një sistem IA ka një detyrim afirmativ për të fikur sistemin, nëse kontrolli njerëzor i sistemit nuk është më i mundur.

KREU VI
MONITORIMI I IMPLEMENTIMIT TË INTELIGJENCËS ARTIFICIALE

Ky monitorim synon të sigurojë një vlerësim të hollësishëm të përputhjes së politikave kombëtare të inteligjencës artificiale (IA) me standardet ndërkombëtare të propozuara nga Shtetet

---

[14] *https://thepublicvoice.org/IA-universal-guidelines/*

[15] *https://artificialintelligenceact.eu/wp-content/uploads/2024/02/AIA-Trilogue-Coreper.pdf*

e Bashkuara të Amerikës, Bashkimi Evropian, OECD-ja dhe ato që mund të miratohen në të ardhmen, si dhe organizma të tjerë në të cilët bën pjesë Republika e Shqipërisë. Procesi i monitorimit do të kryhet nga Agjencia Kombëtare e Shoqërisë së Informacionit (AKSHI), ministritë dhe institucionet e tjera përgjegjëse.

Monitorimi konsiston në mbledhjen dhe analizimin e të dhënave përkatëse mbi performancën e sistemeve të IA-së gjatë gjithë jetëgjatësisë, duke siguruar përshtatje të vazhdueshme me kërkesat rregullatore, në përputhje dhe në zbatim të legjislacionit në fuqi.

Plani i monitorimit, i detajuar në dokumentacionin teknik, do të hartohet me dispozita dhe elemente specifike, në përputhje me përcaktimet në rregullat e Bashkimit Evropian mbi IA-në.

Ofruesit e sistemeve IA mund të integrojnë elemente të këtij plani në sistemet ekzistuese në rrethana të caktuara për të minimizuar dyfishimin dhe ngarkesat, për sa kohë që ruhet një nivel ekuivalent mbrojtjeje.

Metodologjia e IA-së do të monitorohet nga grupi i punës i ngritur nga Agjencia Kombëtare e Shoqërisë së Informacionit, me urdhër të drejtorit të përgjithshëm.

Procesi i monitorimit do të jetë i përshtatshëm dhe i rregullt, me vlerësime periodike dhe vjetore që do të shqyrtojnë progresin dhe sfidat në zbatimin e politikave të IA-së. Kjo përbën një angazhim të vazhdueshëm për sigurimin e zbatimit të përshtatshëm dhe të suksesshëm të politikës së inteligjencës artificiale në Shqipëri, duke ndjekur standardet dhe udhëzimet e ofruara nga organizmat ndërkombëtarë.