

3. modulis

Privātums un
datu aizsardzība
MI projektēšanā



Co-funded by
the European Union



Ethical Engineer

www.ethicalengineer.eu

SATURS

- 01 Privātuma riska novērtēšana sistēmā
- 02 Privātuma iekļaušana arhitektūrā
- 03 Pārvaldība, novērtēšana un tiesību nodrošināšana
- 04 Privātuma lēmuma pieņemšana un aizstāvēšana



Co-funded by
the European Union

Šis projekts ir finansēts ar Eiropas Komisijas atbalstu. Par šo publikāciju (paziņojumu) atbild vienīgi autors, un Komisija neuzņemas nekādu atbildību par tajā ietvertās informācijas izmantošanu. Saskaņā ar jauno Vispārīgās datu aizsardzības regulas (GDPR) regulējumu lūdzu ņemiet vērā, ka partnerība apstrādās jūsu personas datus vienīgi projekta interesēs un nolūkos, neierobežojot jūsu tiesības.



Ētiskā inženiera kursa pārskats

Katras moduļa galvenais saturs un mācību rezultāti

Modulis 01	Modulis 02	Modulis 03	Modulis 04	Modulis 05	Modulis 06
Ētika un pārredzamība	Priekšstati, taisnīgums un iekļaušana	Privātums un datu aizsardzība	Atbildība un drošība	Uz vērtībām balstīts dizains	Ilgspēja un ietekme
Ētikas principi un ES vērtības Pārredzamība un izskaidrojamība AI un robotikas gadījumu analīze	Priekšstati visā AI dzīves ciklā Taisnīguma mērķi un pierādījumi Iekļaujošas inženierijas gadījumi	Privātums jau projektēšanas posmā un Vispārīgā datu aizsardzības regula (GDPR) Datu pārvaldība un PET Novērtēšana un apstrīdējamība	Lomas un atbildība Kļūdas, noturība un nodošana Reģistrēšana un mācīšanās no incidentiem	Ētiskās dizaina prasības Datu, modeļu un saskarnes izvēle Kompromisi un ātrās pārbaudes	M Energija, ūdens un aparatūra Ietekme uz sabiedrību un dzīves ciklu Rādītāji, pārvaldība un pārveidošana

Moduļa ievads

Šis modulis pārvērš privātuma principus arhitektūrā, pierādījumos un atbildīgos inženierijas lēmumos.

Mērķis

Sniegt inženierzinātņu studentiem iespēju identificēt privātuma riskus un izstrādāt samērīgus tehniskos un organizatoriskos kontroles pasākumus.

Mērķauditorija

Augstskolu studenti, kas studē inženierzinātnes, datorzinātnes, robotiku un ar tām saistītās disciplīnas.

Pamatojas uz

1. un 2. modulis: ētikas principi, pārredzamība, aizspriedumu analīze un inženierijas lēmumu dokumentēšana.

[Norādījumi pasniedzējam]

Laiks: 3 minūtes. Pirms nodarbības sākuma izskaidrojiet moduļa secību un mācīšanās rezultātus.

Pieejamība: iepriekš nodrošiniet mācību iznākumus un darba lapu.

Moduļa ievads

Šis modulis pārvērš privātuma principus arhitektūrā, pierādījumos un atbildīgos inženierijas lēmumos.

Mācību secība

Noteikt privātuma riskus, izstrādāt kontroles pasākumus, novērtēt pienākumus un pamatot inženierijas lēmumu par privātumu.

Mācību rezultāti

Datu aprites cikla karte, privātuma arhitektūras apraksts, DPIA pārbaude un inženierijas lēmuma protokols.

[Norādījumi pasniedzējam]

Laiks: 3 minūtes. Pirms nodarbības sākuma izskaidrojiet moduļu secību un mācīšanās rezultātus.

Pieejamība: iepriekš nodrošiniet mācību iznākumus un darba lapu.

Mācību rezultāts

Līdz moduļa
beigām studenti
spēs:

1. Kartēt personas datus, dalībniekus, mērķus un privātuma riskus visā AI sistēmas dzīves ciklā.
2. Piemērot „privātuma ievērošanu jau projektēšanas posmā”, „privātuma ievērošanu pēc noklusējuma” un Vispārīgās datu aizsardzības regulas (GDPR) principus inženierijas prasībām.
3. Izvēlēties drošības un privātuma uzlabošanas tehnoloģijas konkrētai apdraudējuma situācijai un lietošanas kontekstam.
4. Veikt projekta pārbaudi, lai veiktu datu aizsardzības ietekmes novērtējumu un nodrošinātu lēmumu apstrīdējamību.
5. Dokumentējiet kontroles pasākumus, atlikušo risku, uzraudzību un atbildīgu inženierijas lēmumu

Privātums kļūst par inženierijas jautājumu, kad sistēmas dizaina izmaiņas rada risku

AI sistēmas var novērot uzvedību, savienot ierakstus, secināt jutīgas īpašības un ietekmēt nozīmīgus lēmumus. Tādējādi privātuma risks ir atkarīgs no arhitektūras, mērķa un lietošanas, kā arī no konfidencialitātes.

Inženierijas jautājums ir konkrēts: kādi dati vai secinājumi ir nepieciešami norādītajai funkcijai, kam ir piekļuve šiem datiem vai tiesības rīkoties ar tiem, kāds kaitējums var rasties un kādi pierādījumi pamato šo dizainu?

Pamatota atbilde saista ietekmētās personas ar prasībām, datu plūsmām, tehniskajiem kontroles pasākumiem, tiesībām, datu dzēšanu, reaģēšanu uz incidentiem un atbildīgu apstiprināšanu.

Pamatnostādnes saista privātuma pienākumus ar darbības pierādījumiem

Eiropas tiesību akti un vadlīnijas nosaka, kāpēc ir jāaizsargā privātums. Darbības pamatnostādnes palīdz komandām pārvērst šos pienākumus atkārtamos inženierijas pierādījumos.

ES prasības:

ES ES ētikas vadlīnijas privātumu un datu pārvaldību nosaka kā uzticamas AI priekšnosacījumu. GDPR nosaka principus, tiesības un atbildību. AI likums papildina attiecīgo AI sistēmu datu pārvaldību, dokumentēšanu un uzraudzības pienākumus

Operatīvais atbalsts:

NIST AI RMF organizē riska pārvaldību, izmantojot pārvaldību, kartēšanu, novērtēšanu un vadību. ISO/IEC 27701:2025 nosaka privātuma informācijas pārvaldības sistēmu personas identifikācijas informācijas (PII) pārziņiem un apstrādātājiem.

Paskaidrojiet atšķirību starp prasībām un īstenošanas atbalstu. Eiropas tiesību akti un vadlīnijas nosaka pienākumus un gaidas. NIST AI RMF un ISO/IEC 27701 var palīdzēt komandām organizēt riska un privātuma pārvaldības pierādījumus. Jautāji apmācāmajiem, kurš moduļa artefakts ilustrē katru prasību.

[Avoti]

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
<https://www.nist.gov/itl/ai-risk-management-framework>
<https://www.iso.org/standard/27701>

Pamatnostādnes saista privātuma pienākumus ar darbības pierādījumiem

Eiropas tiesību akti un vadlīnijas nosaka, kāpēc ir jāaizsargā privātums. Darbības pamatnostādnes palīdz komandām pārvērst šos pienākumus atkārtojamos inženierijas pierādījumos.

Pierādījumi šajā modulī:

Datu kartes, draudu modeļi, arhitektūras lēmumi, DPIA pārbaude, tiesību izmantošanas ceļi, testi, atbildīgie un pārskatīšanas izraisītāji.

Atvērts:

- ES ētikas vadlīnijas
- GDPR
- ES AI likums
- NIST AI RMF
- ISO/IEC 27701:2025

Izskaidrojiet atšķirību starp prasībām un īstenošanas atbalstu. Eiropas tiesību akti un vadlīnijas nosaka pienākumus un gaidas. NIST AI RMF un ISO/IEC 27701 var palīdzēt komandām organizēt riska un privātuma pārvaldības pierādījumus. Jautājiem apmācāmajiem, kurš moduļa artefakts ilustrē katru prasību.

[Avoti]

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>

<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

<https://www.nist.gov/itl/ai-risk-management-framework>

<https://www.iso.org/standard/27701>

01

Privātuma riska novērtēšana sistēmā

Kartējiet mērķi, personas datus, dalībniekus, secinājumus, datu koplietošanu, glabāšanu un ietekmētās personas.

Privātuma risks ir sistēmas īpašība

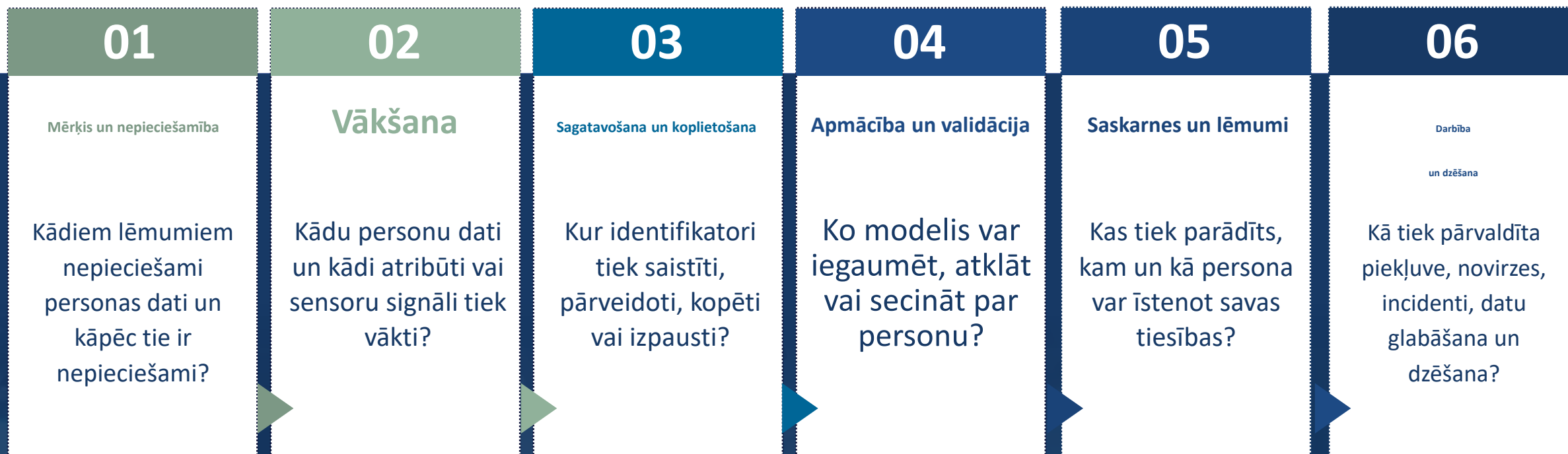
Privātums attiecas uz personas spēju saprast un ietekmēt to, kā informācija par viņu tiek vākta, saistīta, secināta, izmantota un uzglabāta. Sistēma var būt droša pret neatļautu piekļuvi, bet tomēr radīt uzmācīgu uzraudzību, negaidītu atkārtotu izmantošanu vai netaisnīgu autonomijas zaudējumu.

Risks rodas saistībā ar mērķi, sensoru izvēli, identifikatoriem, datu saistīšanu, modeļu darbību, saskarnēm, piekļuves noteikumiem, attiecībām ar piegādātājiem un datu glabāšanu.

Tāpēc inženieri izpēta visu sociotehnisko sistēmu, tās datu plūsmas un lēmumus, kurus tā atbalsta.

Privātuma kartēšana visā AI sistēmas dzīves ciklā

Izvēlieties vienu inženierijas lēmumu. Katrā posmā identificējiet datus, skartās personas, privātuma risku, pierādījumus un kontroles pasākumus.



CAPSTONE: Datu dzīves cikla karte, kas katru posmu saista ar mērķi, iesaistītajām personām, pierādījumiem, privātuma riskam un praktisku kontroli.

Pabeigtais privātuma ceļvedis: Caresecure kritiena noteikšana

Pētījuma gadījums:
„Ethical Engineer CareSecure“
kritiena noteikšana.

Lēmums:
Apstiprināt mājas uzraudzību
gados vecam iedzīvotājam.
Katrā posmā tiek reģistrēti
dati, iesaistītās personas,
privātuma risks, pierādījumi
un kontroles pasākumi.

01

Mērķis un nepieciešamība

- **Dati:** kustība, kas nepieciešama kritiena noteikšanai.
Ietekmētās personas: iemītnis un apmeklētāji.
- **Risks/kontrole:** aizliegt uzvedības novērošanu;
dokumentēt drošības mērķi un lietotāja iestatījumus.

02

Vākšana

- **Dati:** kustības telpā un īss avārijas buferis.
Ietekmētās personas: iedzīvotāji, aprūpētāji
un apmeklētāji.
- **Pierādījumi/kontrole:** sensoru uzskaite;
apstrādāt datus lokāli un izvairīties no
nepārtrauktas videomateriāla uzglabāšanas.

[Norādījumi pasniedzējam] Laiks: 10 minūtes. Atklājiet pa vienam posmam un lūdziet apmācāmos prognozēt nākamo kontroles pasākumu. Uzsver, ka pabeigtais piemērs ir pamatots risinājums, nevis universāls juridisks atbildes variants.

Pabeigtais privātuma ceļvedis: Caresecure kritiena noteikšana

Pētījuma piemērs:
Ethical Engineer CareSecure
kritienu noteikšana.

Lēmums:

Apstiprināt mājas uzraudzību gados vecam iedzīvotājam. Katrā posmā tiek reģistrēti dati, iesaistītās personas, privātuma risks, pierādījumi un kontrole.

03

Sagatavošana un dalīšanās

- **Dati:** ierīces ID, brīdinājuma laiks un avārijas klips. Ietekmētās personas: iemītnis un apstiprinātie aprūpētāji.
- **Pierādījumi/kontrole:** atsevišķa identifikācija; datu pārraides šifrēšana un uz lomu balstītas piekļuves nodrošināšana.

04

Apmācība un validācija

- **Dati:** apstiprināti, simulēti un reprezentatīvi kritienu modeļi. Risks: iegaumēšana vai rutīnas secinājumi.
- **Pierādījumi/kontrole:** novērst neapstrādātu datu eksportu; pārbaudīt datu noplūdi, kļūdainus brīdinājumus un darbības ierobežojumus.

[Norādījumi pasniedzējam] Laiks: 10 minūtes. Atklājiet pa vienam posmam un lūdziet apmācāmos prognozēt nākamo kontroles pasākumu. Uzsvērt, ka pabeigtais piemērs ir pamatots risinājums, nevis universāla juridiska atbilde.

Pabeigtais privātuma ceļvedis: Caresecure kritiena noteikšana

Praktiskais piemērs:
Ethical Engineer CareSecure
kritiena noteikšana.

Lēmums:

Apstiprināt mājas uzraudzību
gados vecam iedzīvotājam.
Katrā posmā tiek reģistrēti
dati, iesaistītās personas,
privātuma risks, pierādījumi
un kontroles pasākumi.

05

Saskarnes un lēmumi

- **Dati:** brīdinājums un pārliecība. Risks: brīdinājums var atklāt informāciju par veselības stāvokli vai klātbūtni.
- **Pierādījumi/kontrole:** samazināt ziņojumu skaitu; pārbaudīt pirms eskalācijas un nodrošināt pieejamus kontroles līdzekļus.

06

Darbība un dzēšana

- **Dati:** revīzijas žurnāls un pārskatīts incidentu reģistrs. Risks: datu uzkrāšanās, piegādātāju piekļuve vai datu noplūde
- **Pierādījumi/kontrole:** automātiska dzēšanas grafika, piekļuves pārskatīšana, incidentu maršruts un apturēšanas izraisītājs.

[Norādījumi pasniedzējam] Laiks: 10 minūtes. Atklājiet pa vienam posmam un lūdziet apmācāmos prognozēt nākamo kontroles pasākumu. Uzsvērt, ka pabeigtais piemērs ir pamatots risinājums, nevis universāls juridisks atbildes variants.

Pabeigtais privātuma ceļvedis: Caresecure kritiena noteikšana

Pētījuma gadījums:
Ethical Engineer
CareSecure kritiena
noteikšana.

Lēmums

Apstiprināt ierobežotu izmēģinājuma projektu tikai pēc datu aizsardzības ietekmes novērtējuma (DPIA) veikšanas, skaidri noteiktas darbības robežas, datu apstrādes vietējā līmenī, uz lomu balstītas piekļuves, pārbaudītiem brīdinājumiem, definētas datu dzēšanas kārtības un iespējas pārtraukt vai atsaukt apstrādi.

[Norādījumi pasniedzējam] Laiks: 10 minūtes. Atklājiēt vienu posmu pēc otra un lūdziet apmācāmos prognozēt nākamo kontroles pasākumu. Uzsvērt, ka pabeigtais piemērs ir pamatots risinājums, nevis universāla juridiska atbilde.

Personas dati ietver identifikatoru s, signālus un pamatotus secinājumus

- Personas dati ir informācija, kas attiecas uz identificētu vai identificējamu personu. Vārdi un konta numuri ir skaidri piemēri. Arī atrašanās vietas pēdas, ierīču identifikatori, balss, attēli, darba paradumi un sensoru datu kombinācijas var identificēt vai izcelt personu.
- Īpašās kategorijas saskaņā ar GDPR ietver veselības datus un biometriskos datus, ko izmanto, lai unikāli identificētu personu. AI var radīt jutīgus secinājumus pat tad, ja ievadītā informācija šķiet parasta.

Klasificējiet katru lauku un secinājumu, reģistrējiet, kāpēc tas ir nepieciešams, un uzskatiet identificējamību par konteksta atkarīgu risku.

Mērķis, datu minimizācija un tiesiskais pamats nosaka tehnisko dizainu

- Mērķa ierobežojums prasa konkrētu un likumīgu izmantošanu. Datu minimizācija ierobežo vākšanu un apstrādi līdz tam, kas ir atbilstošs, būtisks un nepieciešams šai izmantošanai. Pirms apstrādes sākšanas ir jānosaka likumīgs pamats. Piekrišana ir viens no iespējamajiem pamatiem, taču tā var būt nepiemērota gadījumos, kad personai nav patiesas izvēles.
- Pārvērtiet šos principus prasībās: izņemiet neizmantotos laukus, nošķiriet fakultatīvās funkcijas, ierobežojiet sekundāro izmantošanu, ierobežojiet saskarnes un noteiktiet datu glabāšanas un dzēšanas noteikumus.

Pat lietderīgam datu laukam joprojām ir nepieciešams likumīgs mērķis, derīgs pamats, kontrolēta piekļuve un nepieciešamības pierādījums.

Privātuma aizskārumš ietver informācijas noplūdi, secinājumus, novērošanu un rīcības brīvības zaudējumu

- Datu pārkāpums var izpaust personas informāciju, taču privātuma aizskārumš var notikt arī bez pārkāpuma. Pastāvīga novērošana var mainīt uzvedību. Datu saistīšana var atklāt identitāti. Modeļi var izdarīt secinājumus par veselību, uzskatiem, atrašanās vietu vai neaizsargātību. Automatizēti lēmumi var padarīt personu redzamu iestādēm, kamēr sistēma šai personai paliek neskaidra.
- Novērtējiet, kas ko var uzzināt, cik droša ir secinājuma pareizība, kādas darbības var sekot un vai persona to var saprast, labot vai apstrīdēt.

Drošība aizsargā datus. Privātuma inženierija ierobežo arī novērošanu, secinājumus, mērķi, pilnvaras un datu glabāšanu.

Līdzdalība uzlabo privātuma prasības

- Ietekmētās personas var atklāt privātuma gaidas, kuras tehniskā komanda varbūt neredz. Iedzīvotāji, darbinieki, pacienti, apmeklētāji, operatori, apkopes darbinieki un garāmgājēji var vienu un to pašu sensoru vai modeli uztvert atšķirīgi.
- Izmantojiet intervijas, kopīgu projektēšanu, reprezentatīvu testēšanu vai konsultatīvo padomi, lai izvērtētu pieņemamos mērķus, privātās telpas, redzamos indikatorus, kontroles mehānismus, pieejamos paziņojumus, atteikšanos, labojumus un kompensācijas pasākumus. Līdzdalībai jābūt skaidri definētai ceļam uz prasībām un apstiprināšanu.

Reģistrējiet, kuru viedokļi ir ietekmējuši dizainu un kādas neatrisinātas bažas joprojām pastāv.

Praktisks mehāniskās inženierijas piemērs: privātumu ievērojošs eksoskelets

Rūpnīca ierosina izmantot ar sensoriem aprīkotu rūpniecisko eksoskeletu, lai uzlabotu pielāgojumu, atklātu nedrošu slodzi un plānotu apkopi. Privātuma aizsardzības risinājums nosaka, vai sistēma atbalsta darbiniekus vai kļūst par uzraudzības rīku.

[Norādījumi pasniedzējam]

Laiks: 12 minūtes. Izmantojiet šo piemēru, lai privātuma jautājums paliktu mehāniskās inženierijas uzmanības centrā. Jautāji, kura prasība tiktu iekļauta ierīces specifikācijā un iepirkuma līgumā.

01

Sākotnējais uzdevums

Vāciet datus par locītavu leņķiem, spēkiem, ierīces stāvokli un sirdsdarbības ātrumu, lai inženieri varētu uzlabot pielāgojumu un prognozēt komponentu bojājumus.

02

Privātuma risks

Biomehānika var atklāt informāciju par veselību, invaliditāti, nogurumu un tempu. Darbinieki, apakšuzņēmēji un apmeklētāji var būt identificējami.

03

Pārformulēts mērķis

Atbalstīt aprīkojuma drošību un individuālo pielāgojumu. Izslēgt ražīguma novērtēšanu, apmeklējuma kontroli un disciplināro izmantošanu.

Praktisks mehāniskās inženierijas piemērs: privātumu ievērojošs eksoskelets

Rūpnīca ierosina izmantot ar sensoriem aprīkotu rūpniecisko eksoskeletu, lai uzlabotu pielāgojumu, atklātu nedrošu slodzi un plānotu apkopi. Privātuma dizains nosaka, vai sistēma atbalsta darbiniekus vai kļūst par uzraudzības rīku.

[Norādījumi pasniedzējam]

Laiks: 12 minūtes. Izmantojiet šo piemēru, lai privātuma jautājums paliktu mehāniskās inženierijas uzmanības centrā. Jautāji, kura prasība tiktu iekļauta ierīces specifikācijā un iepirkuma līgumā.

04

Arhitektūra

Izvelciet ierīces drošības funkcijas, nodaliet darbinieka identitāti, pārrakstiet neapstrādātos sensoru buferus un šifrējiet apkopes notikumus.

05

Piekļuve un tiesības

Piešķiriet darbiniekiem viņu pašu datu ierakstus. Nodrošini inženieriem apkopotas tendences. Ierobežojiet individuālo veselības datu interpretāciju tikai pilnvarotajam personālam.

06

Lēmumu pieņemšana un pārskatīšana

Izmēģinājuma projektu īstenot tikai ar dokumentētu likumīgu pamatojumu, darbinieku pārstāvību, DPIA pārbaudi, dzēšanas grafiku, sūdzību iesniegšanas kārtību un pārtraukšanas mehānismu.

Praktisks mehāniskās inženierijas piemērs: privātumu ievērojošs eksoskelets

Rūpnīca ierosina izmantot ar sensoriem aprīkotu rūpniecisko eksoskeletu, lai uzlabotu pielāgojumu, atklātu nedrošu slodzi un plānotu apkopi. Privātuma aizsardzības dizains nosaka, vai sistēma atbalsta darbiniekus vai kļūst par uzraudzības rīku.

[Norādījumi pasniedzējam]

Laiks: 12 minūtes. Izmantojiet šo piemēru, lai privātuma jautājums paliktu mehāniskās inženierijas uzmanības centrā. Jautājiet, kura prasība tiktu iekļauta ierīces specifikācijā un iepirkuma līgumā.

Privātums mainīja dizainu

Stāvokļa uzraudzības uzdevums kļuva par ierobežotu drošības sistēmu ar lokālu datu apstrādi, darbinieku tiesībām un specifikāciju par aizliegto lietošanu.

02

Iekļaujiet privātumu arhitektūrā

Piemērojiet privātuma ievērošanu jau projektēšanas posmā, drošību visā dzīves ciklā un atbilstošas privātuma uzlabošanas tehnoloģijas.

Privātuma integrēšana projektēšanas posmā ir inženierijas metode

Privātums jau projektēšanas posmā iekļauj datu aizsardzību sistēmā un tās darbības modeli visagrīnākajā posmā. Privātums pēc noklusējuma nodrošina, ka aizsardzības konfigurācija tiek piemērota, nepieprasot lietotājam meklēt un mainīt iestatījumus.

Septiņi privātuma integrēšanas projektēšanas posmā principi sniedz noderīgas vadlīnijas. Inženierijas darbs tos pārvērš par prasībām, saskarnēm, testiem, ierakstiem un kontroles mehānismiem visā dzīves ciklā.

[Norādījumi pasniedzējam]

Ilgums: 8 minūtes. Lūdziet grupām katru no septiņiem „privātuma iebūvēšanas” principiem saistīt ar vienu inženierijas artefaktu.

Privātums jau projektēšanas posmā ir inženierijas metode

01

Mērķis

Noteikt funkciju, iesaistītās personas, nepieciešamos datus un aizliegtos izmantošanas veidus. Apstrīdēt jebkādu datu vākšanu, kurai nav konkrēta nepieciešamība.

02

Arhitektūra

Sistēmā iekļaujiet datu minimizāciju, nošķiršanu, lokālu apstrādi, piekļuves kontroli, drošus noklusējumus, datu glabāšanu un dzēšanu.

03

Atbildība

Pārbaudiet kontroles mehānismus, dokumentējiet lēmumus, piešķiriet tiesības un nodrošiniet tiesiskās aizsardzības līdzekļus, uzraudzījiet izmantošanu un reaģējiet uz incidentiem un izmaiņām.

Privātums jau projektēšanas posmā ir nepārtraukta inženierijas metode. Proaktīva novēršana, aizsardzība pēc noklusējuma, iebūvētas kontroles, pozitīvās summas dizains, drošība visā dzīves ciklā, pārredzamība un cieņa pret personu ir jābūt redzamām gan dokumentācijā, gan darbībā.

Privātums pēc noklusējuma nosaka drošu pamata līmeni

- Noklusējuma iestatījumi ietekmē reālos rezultātus, jo daudzi cilvēki tos nekad nemaina. Privātumu aizsargājošs noklusējums vāc minimālo datu apjomu, atklāj vismazāko auditoriju, nodrošina vismazāk uzmācīgas funkcijas un saglabā informāciju uz īsāko pamatoto laiku.
- Papildu analītikai, datu koplietošanai vai sekundārai izmantošanai ir jābūt skaidri un atsevišķi izvēlējamai opcijai, kuras pamatā ir atbilstoša piekrišana. Funkcijas izslēgšanai ir jāpārtrauc datu vākšana un jāaktivizē norādītais datu glabāšanas noteikums.

Jāpārbauda pirmās lietošanas pieredze, iestatījumu atjaunošana, programmatūras atjauninājumi un administratora konfigurācija. Noklusējuma iestatījumiem jādarbojas normālos ekspluatācijas apstākļos.

Datu plūsmas karte atklāj uzticības robežas un slēptās kopijas

- Sāciet ar fizisko un digitālo sistēmu. Reģistrējiet katru sensoru, lietotāja ievadi, atvasināto funkciju, modeli, žurnālu, kešatmiņu, vadības paneli, eksportu, piegādātāja saskarni un dublējumu. Katrai plūsmai identificējiet sūtītāju, saņēmēju, mērķi, identifikatoru, atrašanās vietu, piekļuves noteikumu un glabāšanas termiņu.
- Atzīmējiet uzticamības robežas, kur dati pārvietojas starp ierīci un mākonī, vienu organizāciju un citu vai darbības sistēmu un pētniecības vidi. Iekļaujiet kļūdu ceļus, piemēram, bezsaistes rindas, atklūdošanas žurnālus un atbalsta piekļuvi.

Šī karte kļūst par pierādījumu datu apjoma samazināšanai, drošības projektēšanai, datu aizsardzības ietekmes novērtējumam (DPIA), līgumiem, dzēšanas testēšanai un incidentu reaģēšanai.

Pseudonimizācija samazina saistīšanas risku, bet neatbrīvo no GDPR pienākumiem

- Pseudonimizācija aizstāj tiešos identifikatorus un nodrošina papildu informācijas atsevišķu aizsardzību. Dati joprojām ir personas dati, ja to atkārtota saistīšana ir pamatoti iespējama. Tā var samazināt risku un veicināt pienākumu nošķiršanu.
- Anonimizācijas mērķis ir novērst identifikāciju ar līdzekļiem, kurus ir saprātīgi iespējams izmantot. Vienīgi vārdu dzēšana nav pietiekama, ja unikālas trajektorijas, reti sastopami apstākļi vai saistīti datu kopumi var identificēt personu.

Dokumentējiet uzbrucēju, papildu datus, saistīšanas ceļu un atkārtotas identifikācijas testu. Veiciet atkārtotu novērtēšanu, kad kļūst pieejami jauni datu kopumi vai iespējas.

Datu aprites cikla drošība nodrošina konfidencialitāti, integritāti un kontrolētu izmantošanu

- Šifrēšana aizsargā datus pārraides laikā un uzglabāšanas laikā, savukārt atslēgu pārvaldība nosaka, vai šī aizsardzība ir reāla. Piekļuve, pamatojoties uz lomām, minimālās privilēģijas, daudzfaktoru autentifikācija, drošas atjaunināšanas, žurnālu reģistrēšana un pārskatīšana samazina neatbilstošu piekļuvi. Integritātes pārbaudes un izcelsmes pārbaudes aizsargā lēmumus pret mainītiem vai nepilnīgiem datiem.
- Piemērojiet kontroles pasākumus apmācības datiem, modeļu artefaktiem, uzvednēm, izvadiem, žurnāliem, dublējumiem un dzēšanas rindām. Testējiet piegādātāju un administratoru piekļuvi, kā arī ārējos uzbrukumus.

Pilnīga kontrole nosaka draudu, komponentu, konfigurāciju, pierādījumus, īpašnieku, reaģēšanu uz kļūmēm un pārskatīšanas iemeslu.

Privātuma uzbrukumu virsmas
sniedzas tālāk par uzglabātajiem
datiem

AI var izpaust personisko
informāciju caur apmācības
ierakstiem, modeļu uzvedību,
atjauninājumiem, saskarnēm
un metadatiem. Šifrēšana
aizsargā svarīgus ceļus,
savukārt citiem ceļiem
nepieciešami modeļiem
specifiski kontroles pasākumi.

Modeļa inversija vai rekonstrukcija:

Vaicājumi, izvades dati, gradienti vai parametri var
atklāt atribūtus vai aptuvenus apmācības piemērus.

Piederības secināšana:

Uzbrucējs pārbauda, vai konkrēta persona vai ieraksts
bija daļa no apmācības datiem.

Gradianta noplūde:

Kopīgotie atjauninājumi izklaidētā mācīšanās var atklāt
informāciju par vietējiem piemēriem.

[Avoti]

- <https://csrc.nist.gov/pubs/ai/100/2/e2025/final>
- <https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Report%20-%20Securing%20Machine%20Learning%20Algorithms.pdf>

Privātuma uzbrukumu virsmas
sniedzas tālāk par uzglabātajiem
datiem

Mākslīgais intelekts var
izpaust personisko
informāciju caur apmācības
ierakstiem, modeļu uzvedību,
atjauninājumiem, saskarnēm
un metadatiem. Šifrēšana
aizsargā svarīgus ceļus,
savukārt citiem ceļiem
nepieciešami modeļiem
specifiski kontroles pasākumi.

Datu ieguve un blakuskanāli:

Atklātie galapunkti, žurnāli, pakešu izmērs, laiks vai
citi metadati var atklāt datus, modeļa uzvedību vai
jutīgas tēmas.

Tehniskā reaģēšana:

Noteikt uzbrucēja piekļuvi, ierobežot izvadi un vaicājumus,
nodrošināt atjauninājumu drošību, pārbaudīt informācijas
noplūdi, uzraudzīt ļaunprātīgu izmantošanu un noteikt, kad
ieviešana ir jāpārtrauc.

[Avoti]

- <https://csrc.nist.gov/pubs/ai/100/2/e2025/final>
- <https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Report%20-%20Securing%20Machine%20Learning%20Algorithms.pdf>

**Federatīvā mācīšanās
saglabā
neapstrādātos
apmācības datus
lokāli, taču tai
nepieciešami papildu
aizsardzības
pasākumi**

- Federatīvā mācīšanās ļauj ierīcēm vai organizācijām veikt modeļu atjauninājumus, neapvienojot savus neapstrādātos datu kopumus. Tas var samazināt centralizētu datu vākšanu un atbalstīt datu lokalizāciju.
- Modeļu atjauninājumi joprojām var atklāt informāciju, izmantojot gradienta noplūdi, piederības secināšanu vai ļaunprātīgu dalībnieku. Centralizēta koordinācija, atjauninājumu autentifikācija un ierīču kompromitēšana joprojām ir daļa no draudu modeļa.

***Ja nepieciešams, izmantojiet drošu apkopojumu,
atjauninājumu apgriešanu, diferenciālo privātumu,
dalībnieku autentifikāciju un noplūdes testēšanu.
Reģistrējiet, kas iziet no katras ierīces un ko koordinators var
novērot.***

Droša daudzpusēja aprēķināšana atbalsta kopīgu analīzi, neizpaužot ievades datus

- Droša daudzpusēja aprēķināšana ļauj pusēm aprēķināt vienošanās funkciju, vienlaikus savstarpēji nesakot savus ievaddatus. Slepības dalīšana vai kriptogrāfiskie protokoli var atbalstīt starporganizāciju statistiku, modeļu novērtēšanu vai saskaņošanu.
- Protokols aizsargā ievadi saskaņā ar norādītajiem pieņēmumiem. Tas nepadara nedrošu izvadi nekaitīgu, neizlabo nekvalitatīvus datus, nenovērš slepenu vienošanos, kas pārsniedz sliekšni, un nelemj, vai aprēķins ir likumīgs. Lietošanu var ierobežot arī veikspēja un darbības sarežģītība.

Norādiet dalībniekus, draudu modeli, saskaņotības pieņēmumu, atklāto rezultātu, kļūdu apstrādi un neatkarīgu pārbaudi.

Diferenciālā privātums nodrošina izmērāmu robežu individuālajam ieguldījumam

- Diferencētā privātums pievieno kalibrētu nejaušību, lai vienas personas datu klātbūtne vai trūkums ierobežoti ietekmētu publicēto rezultātu. Privātuma parametrs un sastāvs atkārtotās publicēšanas gadījumos nosaka garantijas stiprumu.
- Trokšņi ietekmē lietderību, it īpaši nelielām grupām vai retiem notikumiem. Deklarētā garantija ir nepilnīga bez mehānisma, aizsardzības vienības, privātuma budžeta, nogriešanas noteikuma, sastāva un īstenošanas pierādījumiem.

Izmantojiet diferenciālo privātumu noteiktam datu publiskošanas vai mācīšanās procesam, pēc tam pārbaudiet gan privātuma noplūdi, gan tehnisko veiktspēju.

Izvēlieties privātuma uzlabošanas tehnoloģiju atbilstoši norādītajam draudam

- Privātuma uzlabošanas tehnoloģija ir projektēšanas kontrole, nevis vispārīgs apzīmējums. Izvēlieties to pēc datu plūsmas, uzbrucēja, spēju, aizsargātās informācijas un nepieciešamā rezultāta definēšanas.
- Federatīvā mācīšanās samazina neapstrādātu datu centralizāciju. Droša daudzpusēja aprēķināšana slēpj ievadi kopīgos aprēķinos. Diferenciālā privātuma aizsardzība ierobežo informāciju, ko atklāj izvades dati vai atjauninājumi. Papildus tam var būt nepieciešama šifrēšana, piekļuves kontrole, droša aparatūra un lokāla apstrāde.

Salīdziniet privātuma garantijas, pieņēmumus, veikspēju, kļūdu režīmus, darbības slogu un verifikāciju. Apvienojiet kontroles pasākumus tikai tad, ja katrs no tiem novērš dokumentētu risku.

Malas, IoT un federatīvā mākslīgā intelekta ietekme uz privātuma apdraudējuma modeli

01

Malas un IoT pakļautība:

Lokālā apstrāde var samazināt datu pārraidi uz mākonī. Lauka ierīcēm var būt fiziska piekļuve, vāja identitāte, resursu ierobežojumi, neregulāras atjauninājumu instalēšanas, izkliedēti žurnāli un grūti dzēšamie dati

02

Malas un IoT kontroles pasākumi:

Izmantojiet drošu sistēmas uzsākšanu, ierīču identifikāciju, minimālo datu apjomu, īsus lokālos buferus, parakstītus atjauninājumus, tīkla segmentāciju un pilnīgu dzīves cikla inventarizāciju.

03

Ģeneratīvās AI kontroles:

Noteikt apstiprinātos datu ceļus, rediģēšanu, piekļuves nošķiršanu, uzglabāšanu, pakalpojumu sniedzēju noteikumus, izvades pārskatīšanu un noplūdes testus.

[Avoti]

- <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>
- <https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.1800-36.pdf>

Malas, IoT un federatīvā mākslīgā intelekta ietekme uz privātuma apdraudējuma modeli

04

Generatīvās mākslīgās intelektuālās sistēmas riski:

uzvednes, atgūtie dokumenti, datu precizēšana, rezultāti un žurnāli var saturēt personisku vai konfidenciālu informāciju. Modeļi var iegaumēt, secināt vai radīt neprecīzus apgalvojumus par cilvēkiem.

05

Lēmuma pārbaude:

Reģistrējiet, kas paliek lokāli, kas tiek nosūtīts ārpus sistēmas, ko var secināt, kā tiek pārbaudītas kontroles un kas var apturēt lietošanu.

[Avoti]

- <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>
- <https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.1800-36.pdf>

Projekta gadījums: Faceboard sejas atpazīšana lidostā

Atveriet „Ethical Engineer FaceBoard” gadījuma izpēti un nosakiet, kura arhitektūra var nodrošināt iekāpšanas mērķi ar vismazāko privātuma pārkāpumu.

- „The Ethical Engineer” FaceBoard gadījums ierosina sejas atpazīšanas sistēmas izmantošanu lidostu iekāpšanai. Sistēma var samazināt rindas un krāpšanu, vienlaikus apstrādājot biometriskos datus vidē, kur atteikšanās var ietekmēt ceļošanu. Gadījums izvirza jautājumus par mērķi, nepieciešamību, neobjektivitāti, pārkāpumiem, datu glabāšanu un alternatīviem risinājumiem.
- **Inženierijas fokuss:** veidnes izveide ierīcē, centrālā uzglabāšana, viens pret vienu pārbaude, dzēšana pēc ceļojuma, piegādātāja piekļuve, prezentācijas uzbrukumu atklāšana, cilvēka veikta pārskatīšana un nebiometrisks risinājums.

[Norādījumi pasniedzējam] Laiks: 12 minūtes. Mācībniekiem pirms nodarbības jāizlasa saite uz gadījuma izpēti vai jāizmanto kopsavilkums slaidā. Nodrošiniet teksta transkripciju un ļaujiet dalībniekiem pirms diskusijas individuāli pārdomāt.

Praktisks uzdevums: Izstrādājiet lēmumu par privātuma arhitektūru

[Norādījumi pasniedzējam]

Laiks: 25 minūtes. Novērtējiet mērķa skaidrību, datu plūsmas precizitāti, atbilstību draudu kontrolei, testējamību, tiesības un argumentāciju par atlikušo risku.

- Izmantojiet FaceBoard vai eksoskeleta piemēru. Izstrādājiet projektēšanas uzdevumu, ko cita inženieru komanda varētu īstenot un pārbaudīt.
- noteikt funkciju, iesaistītās personas, personas datus, mērķi, jautājumu par tiesisko pamatu, uzticības robežas un draudu modeli. Salīdzināt divas arhitektūras. Precizēt minimizāciju, nošķiršanu, piekļuvi, šifrēšanu, glabāšanu, dzēšanu, tiesības, uzraudzību un incidentu kontroles pasākumus.

Pievienojiet pieņemšanas testus, atbildīgo īpašnieku, atlikušo risku un nosacījumu, kas prasa REVISE vai STOP.

Rezultāts

Vienas lappuses arhitektūras apraksts un divu minūšu uz pierādījumiem balstīts pamatojums.

03

Pārvaldība, novērtēšana un tiesību nodrošināšana

Saistiet GDPR, DPIA, apstrīdējamību un AI likuma datu pārvaldību ar inženierijas pierādījumiem.

GDPR principi klūst par inženierijas prasībām

- GDPR 5. pants nosaka principus likumīgai, godīgai un pārredzamai apstrādei; mērķa ierobežošanai; datu minimizācijai; precizitātei; uzglabāšanas ierobežojumiem; integritātei un konfidencialitātei; kā arī atbildībai. Šie principi tiek piemēroti kopā.
- Pārveidojiet tos par datu specifikāciju, atļautās izmantošanas politiku, datu kvalitātes pārbaudi, piekļuves modeli, glabāšanas grafiku, dzēšanas pārbaudi, paziņojumu, tiesību īstenošanas ceļu, incidentu apstrādes procesu un lēmumu reģistru. Noteikt pārziņa, apstrādātāja un piegādātāja atbildību.

Atbilstības nodrošināšana prasa organizācijas lēmumu. Inženieri sniedz pierādījumus par to, kā arhitektūra darbojas praksē.

Likumīgais pamats, piekrišana un pārredzamība prasa saskaņotu darbību

- Pārzinim ir jānosaka atbilstošs tiesisks pamats katram apstrādes mērķim. Piekrišanai ir nepieciešama patiesa izvēle, konkrēta informācija un efektīvs atsaukšanas mehānisms. Varas nevienlīdzība var padarīt piekrišanu nepiemērotu, tostarp dažās nodarbinātības un būtisku pakalpojumu situācijās. Īpašas kategorijas datiem ir nepieciešams papildu nosacījums.
- Pārredzamībai jāpaskaidro, kas tiek vākts, kāpēc, kas to saņem, cik ilgi tas tiek glabāts, kādas tiesības attiecas un kā tās īstenot. Saskarnes un faktiskajām datu plūsmām jāatbilst paziņojumam.

Jāpārbauda atsaukšana, labošana un dzēšana no sākuma līdz galam, ieskaitot atvasinātos datus, piegādātājus, dublējumus un modeļu apstrādes procesus.

Automatizētiem lēmumiem nepieciešama pamatota pārskatīšana un apstrīdēšanas iespēja

- GDPR 22. pants attiecas uz noteiktiem lēmumiem, kas balstās vienīgi uz automatizētu apstrādi un rada juridiskas vai līdzvērtīgas nozīmīgas sekas, ievērojot noteiktos nosacījumus un izņēmumus. Saistītie pārredzamības un piekļuves noteikumi var prasīt nozīmīgu informāciju par loģiku un sekām attiecīgajos gadījumos.
- Pārskatīšanas procesam ir nepieciešama pilnvarota un kompetenta persona, piekļuve attiecīgajiem pierādījumiem, laiks, lai mainītu rezultātu, brīvība nepiekrīst sistēmai un reģistrēts pamatojums. Poga, kas nevar mainīt lēmumu, nav efektīva cilvēka iejaukšanās.

Iekļaujiet paziņojumu, paskaidrojumu, labojumu, eskalāciju un tiesiskās aizsardzības līdzekļus kā sistēmas funkcijas, pēc tam pārbaudiet tās kopā ar skartajiem lietotājiem

DPIA saista augsta riska datu apstrādi ar projektēšana s lēmumiem

- A datu aizsardzības ietekmes novērtējums ir nepieciešams gadījumos, kad apstrāde var radīt augstu risku cilvēku tiesībām un brīvībām. Pārbaudē jāņem vērā sistemātiska uzraudzība, inovatīvas tehnoloģijas, neaizsargātas personas, liela apjoma īpašo kategoriju dati, datu saskaņošana un nozīmīgi automatizēti lēmumi.
- DPIA apraksta apstrādi un mērķi, novērtē nepieciešamību un samērīgumu, identificē riskus, reģistrē pasākumus un atlikušo risku, kā arī, ja nepieciešams, iesaista attiecīgos ekspertus un skartās personas.

Sāciet pietiekami agri, lai varētu mainīt prasības un arhitektūru. Pārskatiet novērtējumu, ja mainās mērķis, dati, modelis, apjoms, piegādātājs vai konteksts.

Aizpildītais DPIA skrīnings: CareSecure kritienu noteikšana mājās



ES AI likums, datu pārvaldība un GDPR privātuma pienākumi mijiedarbojas

- Attiecībā uz attiecīgajām augsta riska AI sistēmām AI likums prasa riska pārvaldību, datu pārvaldību, tehnisko dokumentāciju, reģistrēšanu, cilvēka uzraudzību, precizitāti, stabilitāti un kiberdrošību. Apmācības, validācijas un testēšanas datu kopām nepieciešama atbilstoša pārvaldība attiecībā uz atbilstību, reprezentativitāti, kļūdām un trūkumiem.
- GDPR un AI likumam ir atšķirīga darbības joma. Datu kvalitāte neatceļ mērķa ierobežojumu vai datu apjoma samazināšanu. Datu aizsardzības ietekmes novērtējumam (DPIA) un pamattiesību ietekmes novērtējumam ir atšķirīgi juridiskie izraisītāji un saturs, lai gan pierādījumi un kontroles pasākumi var tikt koordinēti.

Uzturiet izsekojamu kontroles sistēmu, kurā tiek reģistrēta katra prasība, pierādījuma vienība, atbildīgā persona, atlikušais risks un pārskatīšanas punkts

HealthTech privātuma pierādījumi ir arī pacientu drošības pierādījumi

Ja AI programmatūrai ir medicīnisks mērķis, lēmumi par privātumu ir saistīti ar MDR, AI likumu un GDPR. Iecerētais mērķis un risks nosaka piemērojamās pienākumus attiecībā uz ierīci.

01

Datu integritāte

Izsekojiet avotu, likumīgu piekļuvi, marķējumu, izņēmumus, pārveidojumus, trūkstošos datus un versiju. Sabojāti vai nereprezentatīvi klīniskie dati var ietekmēt diagnozi, pacientu šķirošanu vai ārstēšanu.

02

Kvalitātes un riska pārvaldība

Saistiet datu kontroles pasākumus ar kvalitātes vadības sistēmu, riska vadības dokumentāciju, klīnisko novērtējumu, kibernetisko drošību, izmaiņu kontroli un pēcpārdošanas uzraudzību.

03

Pacientu privātums

Ierobežojiet īpašas kategorijas datus, nodaliet identitātes, pārbaudiet piekļuvi un dzēšanu, kā arī aizsargājat žurnālus, modeļu artefaktus un piegādātāju saskarnes.

HealthTech privātuma pierādījumi ir arī pacientu drošības pierādījumi

Ja AI programmatūrai ir medicīnisks mērķis, lēmumi par privātumu ir saistīti ar MDR, AI likumu un GDPR. Iecerētais mērķis un risks nosaka piemērojamos pienākumus attiecībā uz ierīci.

04

Pierādījumi pirms izlaišanas

Klīniskā veikspēja atbilstoši attiecīgajai grupai, datu izcelsme, draudu testēšana, cilvēka veikta pārskatīšana, incidentu maršruts, atlikušais risks un apstiprināšanas iestāde.

05

Inženierijas secinājums

Privātuma pārkāpums, kas apdraud datu integritāti, pieejamību vai atbilstošu izmantošanu, var kļūt par ierīces drošības pārkāpumu.

[Avoti]

- <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>
- https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en
- <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Dilemma: drošības izmeklēšana vai funkciju paplašināšanā s?

- Eksoskelets reģistrē gandrīz notikušu negadījumu, kurā darbinieks paceļ smagumu ārpus apstiprinātā diapazona. Drošības komanda pieprasa pēdējo trīs mēnešu individuālos biomehānikas datus, lai izpētītu iekārtas darbību. Līnijas vadītājs pieprasa tos pašus datus, lai identificētu darbiniekus, kuri strādā lēni vai neievēro noteikumus. Sākotnējais paziņojums attiecas uz piemērotību, drošību un apkopi.
- Katram pieprasījumam izlemiet: turpināt, pārskatīt vai pārtraukt. Noteikt mērķi, nepieciešamību, atļauto piekļuvi, datu periodu, darbinieku iesaistīšanos, pierādījumus, dzēšanu un pārskatīšanu.

Paskaidrojiet, kurš tehniskās nošķiršanas un pārvaldības noteikums neļauj likumīgai drošības funkcijai kļūt par darba snieguma uzraudzību.

[Norādījumi pasniedzējam] Laiks: 15 minūtes. Pieprasiet atsevišķus lēmumus par drošības un snieguma pieprasījumiem. Atļaut noteikt atšķirīgu viedokli un reģistrējiet pārvaldības kontroli, kas atrisina konfliktu.

04

Pieņemiet un aizstāviet lēmumu par privātumu

Salīdziniet projekta gadījumus, veiciet projekta pārskatīšanu un aizstāviet lēmumu, izmantojot pierādījumus.

Projekta gadījums A: Mākslīgais intelekts veselības aprūpē un uzraudzība mājās

- CareSecure izmanto vietējos sensorus, lai noteiktu kritienu un brīdinātu aprūpētāju. Drošības ieguvums ir ievērojams, taču nepārtraukta novērošana mājās var ietekmēt privātumu, autonomiju un uzticēšanos. Var tikt fiksēti arī apmeklētāji un aprūpētāji.
- Noteikt minimālo nepieciešamo signālu, privātās telpas, vietējo datu apstrādi, informācijas izpaušanu ārkārtas gadījumos, piekrišanu vai citus ar likumīgo pamatu saistītus jautājumus, pieejamos kontroles mehānismus, kļūdainu trauksmju pārskatīšanu, datu dzēšanu un reaģēšanu uz incidentiem.

Atveriet „Ethical Engineer” veselības aprūpes pētījuma gadījumu. Izlemiet, kādi pierādījumi ir nepieciešami pirms ierobežota izmēģinājuma projekta uzsākšanas un kāda izmantošana ir jāaizliedz.

[Norādījumi pasniedzējam] Laiks: 12 minūtes. Piešķiriet rezidenta, aprūpētāja, inženiera un datu aizsardzības speciālista lomas. Piekluve: atļaujiet rakstiskus lomu aprakstus, nevis izpildīšanu klātienē.

Projekta gadījums B: AI Witness un jutīga liecība

- AI Witness reģistrē liecinieku liecības, izmantojot sarunu sistēmu. Šis pakalpojums var saglabāt detaļas un samazināt kavēšanos, vienlaikus vācot ļoti jutīgus aprakstus, kas var ietekmēt drošību un taisnīgumu. Intervijas process var ietekmēt arī to, ko atklāj cilvēks, kurš atrodas stresa situācijā.
- Noteikt mērķi, datu kategorijas, drošu datu ieguvu, valodas pieejamību, cilvēka veiktu pārskatīšanu, labojumus, juridisku informācijas izpaušanu, datu glabāšanu, datu aizsardzības pārkāpuma ietekmi, piekļuvi modeļa sniedzējam un neautomatizētu alternatīvu.

Atveriet „Ethical Engineer” liecību pētījuma gadījumu. Izlemiet, vai dizainam būtu jāatbalsta, jāierobežo vai jānoraida automatizēta intervēšana.

[Norādījumi pasniedzējam] Laiks: 12 minūtes. Izmantojiet traumas apzinātu pieeju. Mācībnieki var analizēt sistēmas dizainu, neizspēlējot satraukta liecinieka lomu.

Salīdziniet gadījumus, pirms izvēlaties privātuma kontroles pasākumus

- CareSecure attiecas uz pastāvīgu novērošanu privātmājā tūlītējas drošības nodrošināšanas nolūkā. AIWitness attiecas uz laika ziņā ierobežotu, bet ļoti jutīgu mijiedarbību, kuras ieraksts var tikt iekļauts tiesvedības procesā. Abos gadījumos ir nepieciešama drošība, taču to mērķi, varas attiecības, datu glabāšana, tiesības un sekas atšķiras.
- Katram gadījumam norādiet datu subjektu, apkārtējos, pārziņu, mērķi, augsta riska indikatorus, uzticības robežas, vislielāko iespējamo privātuma pārkāpumu un nepieciešamo cilvēka pilnvaru.

Noderīgs secinājums var ietvert atšķirīgas arhitektūras, jautājumus par tiesisko pamatu, pierādījumu sliekšņus un pārtraukšanas nosacījumus. Paskaidrojiet, kāpēc katrs kontroles pasākums atbilst konkrētajam kontekstam.

[Norādījumi pasniedzējam] Laiks: 10 minūtes. Pirms kontroles pasākumu izvēles salīdziniet kontekstus. Novērtējiet labi pamatotas atšķirības, nevis vienu vēlamu atbildi.

Komandas uzdevums: Veiciet privātuma dizaina pārskatīšanu

Izvēlieties CareSecure, FaceBoard, AlWitness vai eksoskeletu. Piešķiriet sistēmu, datu, drošības, tiesību, ieinteresēto personu un protokolētāja lomas.

1. Definējiet funkciju, mērķi un ietekmētās personas.
2. Aizpildiet sešu posmu datu aprites cikla karti.
3. Uzzīmējiet datu plūsmas un uzticamības robežas.
4. Veiciet datu aizsardzības ietekmes novērtējumu (DPIA) un izvēlieties samērīgus kontroles pasākumus.
5. Precizējiet tiesības, pierādījumus, atlikušo risku un uzraudzību.
6. Lemjiet – turpināt, pārskatīt vai pārtraukt.

Rezultāts:

Lēmuma protokols un divu minūšu ilgs pārskats par privātumu.

[Norādījumi pasniedzējam]

Laiks: 30 minūtes. Izveidojiet grupas no četriem līdz sešiem dalībniekiem. Pieprasiet, lai katra loma sniegtu vienu pierādījumu vai izvirzītu vienu neatrisinātu jautājumu.

Kas ir iekļauts spēcīgā privātuma dizainā

- Spēcīgs dizains sākas ar konkrētu funkciju un parāda, kāpēc katrs personas datu elements ir nepieciešams. Tas nošķir personas datus, jutīgus secinājumus un nepersoniskus sistēmas datus. Tas kartē dalībniekus, kopijas, uzticības robežas, mērķus un glabāšanas termiņus.
- Arhitektūra apvieno datu apjoma samazināšanu, datu nošķiršanu, drošību, privātuma uzlabošanas tehnoloģijas, izmantošanas tiesības, cilvēka veiktu pārskatīšanu, datu dzēšanu, uzraudzību un reaģēšanu uz incidentiem. Katram kontroles pasākumam ir norādīts apdraudējums, konfigurācija, pieņemšanas tests un atbildīgā persona.

Lēmuma protokolā ir norādīti jautājumi par tiesisko pamatu un datu aizsardzības ietekmes novērtējumu (DPIA), ieinteresēto pušu viedoklis, alternatīvas, atlikušais risks, apstiprināšanas pilnvaras, pārskatīšanas iemesls un aizliegtie izmantošanas veidi

Darba lapa: Privātuma inženierijas lēmumu protokols

Sistēma, funkcija un ietekmētās personas:

Konkrēts mērķis un aizliegtie izmantošanas veidi:

Personas dati, jutīgi secinājumi un nepieciešamība:

Datu plūsmas, saņēmēji un uzticības robežas:

Tiesiskais pamats un DPIA pārbaudes jautājumi:

Tehniskā kontrole un pieņemšanas tests:

[Norādījumi pasniedzējam] Laiks: 20 minūtes. Darba lapu nodrošiniet kā rediģējamu dokumentu. Pieņemiet aizpildītas darba lapas, kas ir drukātas, rakstītas ar roku, ierakstītas audio formātā vai video formātā.

Turpinājums nākamajā slaidā

Tiesības, cilvēka veikta pārskatīšana, dzēšana un tiesiskās aizsardzības līdzekļi:

Atbildīgās personas un pierādījumu īpašnieks:

Lēmums: TURPINĀT PĀRSKATĪT APSTĀTĪTIES

Atlikušais risks, uzraudzība un pārskatīšanas iemesls:

Atšķirīgs viedoklis vai alternatīva arhitektūra:

[Norādījumi pasniedzējam] Laiks: 20 minūtes. Darba lapu nodrošiniet kā rediģējamu dokumentu. Pieņemiet aizpildītu dokumentu, kas ir drukāts, rakstīts ar roku, audio vai ierakstīts.

Lomu spēle: DPIA un iepirkumu pārskatīšanas komisija

- Universitāte plāno iegādāties mākslīgā intelekta kameru sistēmu, kas ieraksta laboratorijas darbību, atklāj nedrošu uzvedību un ģenerē individuālus atbilstības rādītājus. Piegādātājs piedāvā apstrādi mākonī un saglabā videoklipus modeļa uzlabošanai. Priekšlikumā ir norādīta vispārējā precizitāte, bet nav sniegti pierādījumi par apmeklētājiem, invaliditāti, aizsargapģērbu, datu dzēšanu vai personāla piekļuvi.
- Lomas: laboratorijas inženieris, studentu pārstāvis, drošības vadītājs, datu aizsardzības speciālists, piegādātājs, pieejamības konsultants un iepirkumu vadītājs.

Datu plūsmas, DPIA jautājumi, izmēģinājuma robežas, līguma nosacījumi un pieņemšanas pierādījumi. Norādiet vienu nosacījumu „Pārskatīt” un vienu nosacījumu „Pārtraukt”. Reģistrējiet, kurš var apturēt izmantošanu un kā persona var piekļūt ierakstam, to labot vai apstrīdēt.

[Norādījumi pasniedzējam] Laiks: 25 minūtes. Komisijai jānorāda apstiprināšanas nosacījumi, pierādījumi par spēkā lietojamību un atbilstības pierādījumus. Norādiet, kādā veidā komitējam jānodrošina atbilstības pierādījumus.

Individuālā pārdomāšanā & Pārnese

Uzrakstiet īsu inženierijas argumentāciju

1. Kura datu plūsma vai secinājums radīja vislielākās bažas par privātumu?
2. Kura dizaina izvēle visefektīvāk samazināja risku?
3. Kāds jautājums par tiesisko pamatu vai DPIA joprojām ir jāizskata speciālistam?
4. Kādi pierādījumi mainītu jūsu lēmumu?
5. Kāda izmantošana būtu jāaizliedz, pat ja tā ir tehniski iespējama?
6. Kā analīze mainītos citā inženierzinātņu nozarē?

Rakstiet:

Iesniedziet rakstiskus, audio vai ierakstītus pierādījumus. Saistiet skartās personas, inženierijas pasākumus, pierādījumus un pamatotu lēmumu.

[Norādījumi pasniedzējam] Laiks: 10 minūtes. Mācībnieki var iesniegt rakstiskus, audio vai ierakstītus pierādījumus. Pirms jebkādas dalīšanās ar rezultātiem atvēliet klusu sagatavošanās laiku.

Septiņi „Privacy-By-Design” principi kā inženierijas pārbaudes

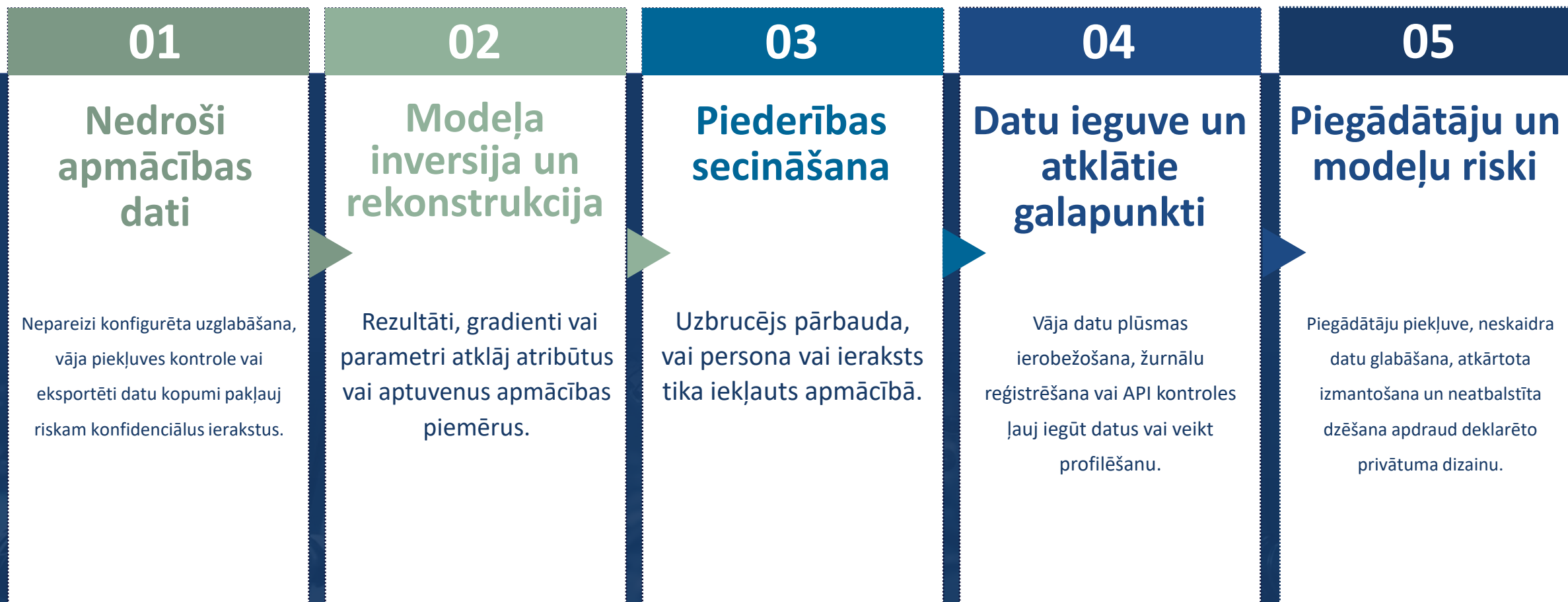
Izmantojiet šos septiņus principus kā ātru diagnostiku, pirms apstiprināt AI dizainu.

- **Proaktīva novēršana:** identificējiet iespējamus privātuma pārkāpumus pirms izstrādes vai ieviešanas.
- **Aizsardzība pēc noklusējuma:** vāciet, atklājiet un saglabāiet pēc iespējas mazāk datu pēc noklusējuma.
- **Iebūvēta privātuma aizsardzība:** iekļaut arhitektūrā datu minimizāciju, nošķiršanu, piekļuvi, dzēšanu un privātuma pastiprinošās tehnoloģijas (PET).
- **Pozitīvās summas dizains:** nodrošināt inženiertehnisko funkcionalitāti, neuzskatot privātumu par kompromisu, no kura jāatsakās.
- **Drošība no gala līdz galam:** aizsargāiet datus, modeļu artefaktus, uzvednes, žurnālus, dublējumus un dzēšanas rindas.
- **Pārredzamība, caurskatāmība un lietotāja orientētība:** dokumentējiet kontroles pasākumus, izskaidrojiet ietekmi un nodrošiniet izmantojamās tiesības.

Sānu kanālu un metadatu noplūde AI sistēmās

- Šifrēšana aizsargā saturu gan pārraides laikā, gan uzglabāšanas laikā, taču AI sistēmas joprojām var nopludināt informāciju caur metadatiem, laiku, vaicājumu uzvedību vai izvades modeļiem.
- **Riska piemērs:** ar TLS aizsargāta LLM saruna joprojām var atklāt jutīgu tēmu modeļus, analizējot pakešu izmēru vai laiku.
- **Tehniskais risinājums:** definējiet metadatus kā daļu no privātuma uzbrukumu virsmas, samaziniet atklātos žurnālus, ierobežojiet vaicājumu uzvedību un uzraudzījiet neparastu piekļuvi.
- Ja risks ir augsts, pārbaudiet secinājumus, balstoties uz izvadi, uzvedības norādēm, gradientiem, pakešu modeļiem un darbības telemetriju.
- Pieņemamības pierādījumiem jāparāda, ka noplūdes testēšana, uzraudzība un reaģēšana uz incidentiem aptver modelim raksturīgos uzbrukuma ceļus.

Testējamas AI privātuma ievainojamības darbības laikā



Pārvaldības triāde un darbības standarti

Pilnībā izstrādāts AI privātuma process apvieno juridiskās, privātuma un kiberdrošības ekspertīzi jau no sistēmas projektēšanas sākuma.

01

Juridiskie

Juridiskais pamats, tiesības, līgumiskā atbildība, atbilstība normatīvajiem aktiem un dokumentācijas pienākumi.

02

Privātums

Datu minimizācija, DPIA pārbaude, pārredzamība, tiesību īstenošanas ceļi, datu glabāšana un privātuma pārvaldības pierādījumi.

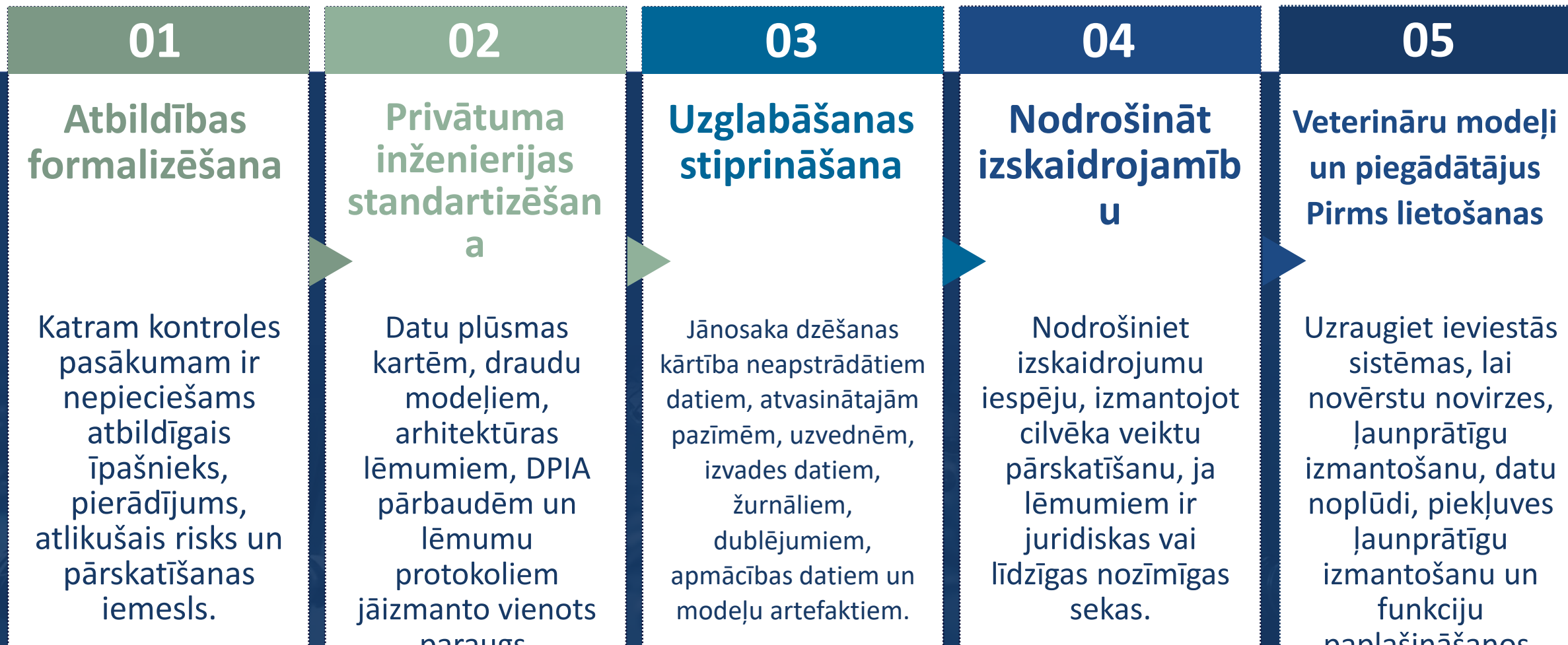
03

Kiberdrošība

Konfidencialitāte, integritāte, pieejamība, pretinieka testēšana, piegādātāju garantijas, uzraudzība un reaģēšana uz incidentiem.

Izmantojiet NIST AI RMF, lai organizētu riska pārvaldības darbu, un ISO/IEC 27701, lai strukturētu privātuma pārvaldības pierādījumus pārziņiem un apstrādātājiem.

Stratēģiskās ieviešanas pārbaudes saraksts



Papildu gatavas mācību aktivitātes

01

Mācību mašīna:

Izpētiet biometrisko privātumu un piekrišanu, izmantojot attēlu vai balss klasifikāciju.

02

"What-If" rīks:

Pārbaudiet apakšgrupu ietekmi, sliekšņus un kontrfaktuālos scenārijus, kas saistīti ar jutīgiem atribūtiem.

03

Faketi:

Pārbaudiet datu kopas, lai atklātu nelīdzsvarotību, anomālijas, trūkstošas vērtības un nevajadzīgus laukus.

Papildu gatavas mācību aktivitātes

04 ○ **ARX:**

Demonstrējiet pseidonimizācijas, k-anonimitātes un atkārtotas identifikācijas kompromisus.

05 ○ **H5P atzarojuma scenāriji:**

Simulējiet lēmumus par piekrišanu, uzraudzību un datu izmantošanu.

06 ○ **DEDA vai ētiskā OS:**

Strukturējiet ieinteresēto pušu dialogu un ilgtermiņa pārdomu procesu par privātuma riskiem.

Viktorīna

1. Kāpēc sistēma var būt droša, bet tomēr radīt privātuma apdraudējumu?
2. Kāda ir atšķirība starp pseidonimizētiem un anonimizētiem datiem?
3. Kāpēc PET ir jāsaista ar draudu modeli?
4. Ko vajadzētu ietvert datu plūsmas kartē?
5. Nosauciet divus rādītājus, kas var pamatot lēmumu par DPIA pārbaudi.
6. Kāpēc cilvēka veikta pārskatīšana ir nozīmīga automatizēta lēmuma pieņemšanā?
7. Kas jāiekļauj privātuma inženierijas lēmuma protokolā?

[Norādījumi pasniedzējam]. Laiks: 8 minūtes. Sākumā izmantojiet individuālu informācijas meklēšanu, pēc tam diskusiju pāros. Nākamajā slaidā sniegtas īsas atbildes un pamatojums.

Atbilžu uz jautājumiem un mācību pamatojums

1. Droša apstrāde joprojām var ļaut veikt uzmācīgu novērošanu, secinājumus, atkārtotu izmantošanu vai rīcības brīvības zaudēšanu.
2. Pseidonimizēti dati joprojām var būt saistāmi ar papildu informāciju; anonimizēti dati jābūt tādiem, ka tos nav iespējams identificēt ar saprātīgi iespējamām metodēm.
3. Katra privātuma inženierijas tehnika (PET) aizsargā konkrētu informāciju saskaņā ar norādītajiem pieņēmumiem un tai ir raksturīgi kļūdu veidi.
4. Avoti, saņēmēji, mērķi, identifikatori, pārveidojumi, kopijas, uzticamības robežas, piekļuve un glabāšana.
5. Sistemātiska uzraudzība, neaizsargātas personas, inovatīvas tehnoloģijas, jutīgi dati, saskaņošana vai nozīmīgi automatizēti lēmumi.
6. Pilnvaras, kompetence, pierādījumi, laiks, brīvība nepiekrīst un pilnvaras mainīt rezultātu.
7. Mērķis, dati, riski, kontroles pasākumi, pierādījumi, lomas, tiesības, atlikušais risks, uzraudzība, lēmuma pieņemšanas un pārskatīšanas iemesls.

Mācību rezultātu pārbaude

Pirms pabeigšanas pārlicinieties, ka jūs varat:

1. Kartējiet personas datus, dalībniekus, mērķus, secinājumus un riskus visā AI dzīves ciklā.
2. Piemērojiet prasībām „privātuma ievērošana jau projektēšanas posmā”, „privātuma ievērošana pēc noklusējuma” un Vispārīgās datu aizsardzības regulas (GDPR) principus.
3. Izvēlieties un pamatojiet privātuma uzlabošanas tehnoloģiju konkrētai apdraudējuma situācijai.
4. veiciet DPIA pārbaudi un izstrādājiet nozīmīgas tiesības un apstrīdējamību.
5. Dokumentu kontrole, pierādījumi, atlikušais risks, uzraudzība

Pierādījumi: mums.

Datu aprites cikla karte, arhitektūras apraksts, DPIA pārbaude un inženiertehnisko lēmumu protokols.

[Norādījumi pasniedzējam]

Laiks: 5 minūtes. Lūdziet apmācāmajiem norādīt uz artefaktu, kas ilustrē katru rezultātu. Reģistrējiet visus rezultātus, kas prasa papildu praksi.

leguldījums kursa dokumentācijā

Pievienojiet šādus 3. moduļa pierādījumus savu ētiskās inženierijas dokumentāciju:

1. Aizpildīta datu aprites cikla un datu plūsmas karte.
2. Mērķis, nepieciešamība un aizliegto izmantošanas veidu specifikācija.
3. Privātuma arhitektūra un PET izvēle.
4. DPIA pārbaudes protokols un neatrisinātie speciālistu jautājumi.
5. Tiesības, dzēšana, incidenti un uzraudzības kontroles.
6. Privātuma inženierijas lēmumu protokols un pārdomas.

Šie pierādījumi saistās ar atbildību, uz vērtībām balstītu dizainu un ilgtspējību turpmākajos moduļos.



Paziņojums par finansējumu un izmantošanu izglītības nolūkos

Šis atklātais izglītības resurss atbalsta mācīšanu un diskusijas. Tas nav uzskatāms par juridisku, normatīvu, drošības vai profesionālu inženierijas konsultāciju. Pielietojot tā saturu, ievērojiet spēkā esošos likumus, standartus, iestāžu politiku un nozarei specifiskos norādījumus.

Līdzfinansē Eiropas Savienības programma „Erasmus+“.





Ethical Engineer

www.ethicalengineer.eu

Paldies

Jūs esat pabeidzis 3. moduli



Co-funded by
the European Union