

# 1.

**modulis**  
levads par  
mākslīgo intelektu  
un ētiku inženierijas  
praksē



Co-funded by  
the European Union



**Ethical Engineer**

[www.ethicalengineer.eu](http://www.ethicalengineer.eu)

# SATURS

## 01 Ētika un inženierzinātnes

---

## 02 TAISNĪGUMS

---

## 03 Autonomija

---

## 04 Pārredzamība

---

## 05 Nekaitīgums

---

## 06 Grupu darbs

---

## 07 TESTS



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.



# Moduļa ievads

## Mērķis

Ieviest ētisko argumentāciju kā daļu no mākslīgā intelekta un robotikas inženierijas prakses.

## Mērķauditorija

Augstskolu studenti inženierzinātņu, datorzinātņu, robotikas un saistītajās STEM programmās.

## Ieteicamais ilgums

Pamata programma – 180 minūtes. Paplašinātā programma – 225 minūtes.

Var vadīt kā vienu nodarbību vai divas īsākas sesijas.

[Avoti] „Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.

# Moduļa ievads

## Mācību secība

Izskaidrot ētiku inženierijas praksē, piemērot taisnīgumu, autonomiju, pārredzamību un nekaitēšanas principu, pēc tam analizēt inženierijas gadījumu un pamatot lēmumu.

## Mācību rezultāti

Ētisko jautājumu karte, gadījuma analīze, aizsardzības pasākumi, inženierijas ētika lēmuma protokols un rezultātu pārbaude. Šie materiāli parāda, kā ētiskais pamatojums ietekmē prasības, testus, uzraudzību un apstiprināšanas lēmumus.

# Mācību rezultāti

## Līdz moduļa beigām studenti spēs:

1. Izskaidrot, kā autonomija, taisnīgums, pārredzamība un nekaitīgums attiecas uz mākslīgā intelekta un robotikas inženieriju.
2. Identificēt ētiskos riskus problēmu formulēšanā, datos, modelī, saskarnē, ieviešanā un uzraudzībā.
3. Analizēt inženierijas gadījumu, saistot projektēšanas lēmumus ar ietekmi uz ieinteresētajām pusēm.
4. Ierosināt tehniskos un pārvaldības pasākumus, kas samazina risku un atbalsta cilvēka veikto uzraudzību.
5. Pamatojiet inženierijas lēmumu, atsaucoties uz ES ētikas vadlīnijām par uzticamu mākslīgo intelektu un attiecīgajām DigComp kompetencēm.

### [Avoti]

„Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>

# Ētiskā argumentācija pieder pie inženierijas projektēšanai

**01** | Problēmas  
formulēša  
na

**02** | Dati

**03** | Modelis

**04** | Saskarnes un  
uzraudzība

**05** | Ieviešana

# Inženierijas izvēles ietekmē ētiskos rezultātus visā AI sistēmas dzīves ciklā:

Ētiskais pamatojums ir projektēšanas kompetence. Tas palīdz inženieriem pārvērst Eiropas vērtības tehniskās prasībās, testēšanas kritērijos un atbildīgos lēmumos

[Avoti] - „Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 „Mācīšana un mācīšanās” (OER izstrāde), 61.–70. lpp. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

## ○ Problēmas formulēšana

Kuras vajadzības tiek uzskatītas par prioritārām?

## ○ Dati

Kas ir pārstāvēts, trūkst vai ir pakļauts riskam?

## ○ Modelis

Kādas kļūdas un kompromisi tiek pieņemti?

## ○ Saskarnes un uzraudzība

Kas var saprast, apšaubīt vai pārrakstīt rezultātus?

## ○ Ieviešana

Kādas drošības garantijas, uzraudzības un incidentu risināšanas procedūras ir ieviestas?

# Saskaņotība ar Eiropas kompetenču sistēmām

Eiropas sistēmas saista ētiskos principus ar praktiskajām inženierzinātņu kompetencēm.

## Norādījumi

„ES ētikas vadlīnijas definētas uzticamas AI prasības. DigComp 2.2 un DigCompEdu atbalsta atbildīgu mācībspēku un izglītojamo kompetenci.

## Piemērošana

Mācītiesgribētāji izvērtē uzraudzību, taisnīgumu, privātumu, atbildību, izskaidrojamību un drošību inženierijas sistēmā.

## Pierādījumi

Gadījumu analīze, aizsardzības pasākumi un lēmumu protokols parāda, ko mācīties gribētāji spēj analizēt, pārbaudīt un pamatot.

# 01

# Ētika inženierijas praksē

Ētika ir sistemātiska izpēte par to, kas ir pareizi, taisnīgi un atbildīgi.

Inženierzinātnēs ētika virza profesionālo spriedumu, ja tehniskie lēmumi ietekmē cilvēkus, tiesības, drošību, piekļuvi vai vidi.

## Ētiskais inženieris jautā:

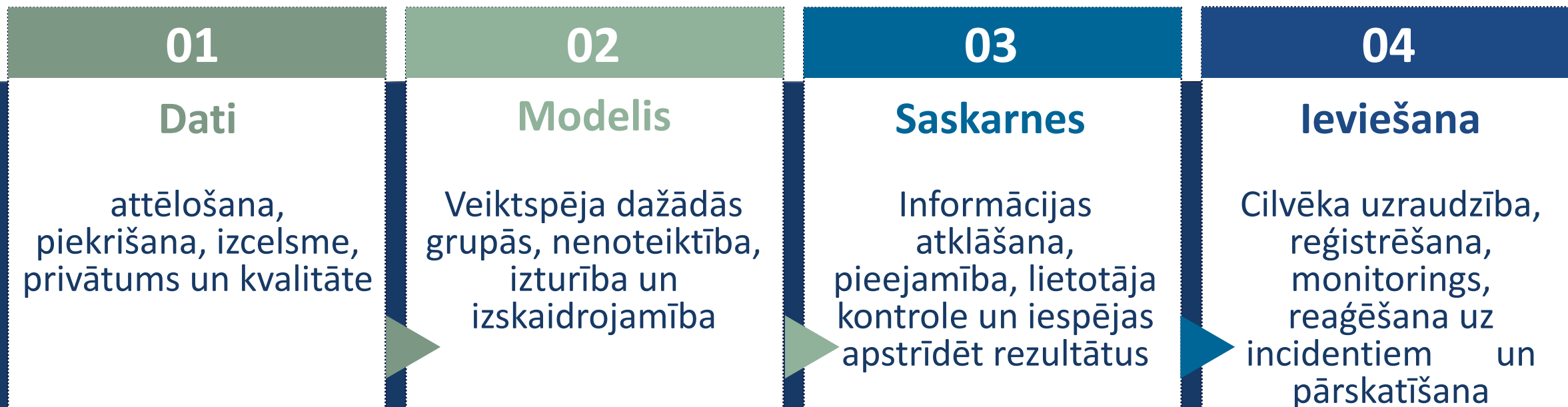
- Kam tas var dot labumu vai nodarīt kaitējumu?
- Kādas vērtības un tiesības tiek ietekmētas?
- Kādi pierādījumi pamato šo projektēšanas izvēli?
- Kas ir atbildīgs, ja sistēma nedarbojas?

*Profesionālie kodeksi, institūciju politikas un likumi nosaka prasības. Ētiskais pamatojums atbalsta atbildīgu lēmumu pieņemšanu gadījumos, kad ir vairākas iespējamās izvēles.*

[Avoti] „Ethical Engineer” apstiprināja KA220-HED pieteikumu, WP3 „Mācīšana un mācīšanās” (OER izstrāde), 61.–70. lpp.  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

# Ētika kļūst par inženierijas prasību

Ētiku var pārvērst konkrētos projektēšanas jautājumos un pieņemšanas kritērijos:



Inženierijas uzdevums ir nodrošināt vērtību īstenošanu, izmantojot specifikācijas, testus, dokumentāciju un pārvaldību.

# Ētiskajā argumentācijā tiek izmantoti vairāki skatpunkti

Šīs trīs perspektīvas  
atklāj dažādas vienas  
un tās pašas lēmuma  
pieņemšanas puses.

## ○ Sekas

Izvērtējiet iespējamus ieguvumus un kaitējumus, to, kas tos piedzīvo, to varbūtību un to, vai tos var atgriezt atpakaļ. Dizains ar labu vidējo veikspēju joprojām var radīt nopietnu kaitējumu mazākai grupai.

## ○ Pienākumi un tiesības

Jānoskaidro, kādi pienākumi ir spēkā pat tad, ja cita iespēja šķiet efektīvāka. Cilvēka cieņa, privātums, nediskriminācija, apzināta izvēle un drošība var ierobežot sistēmas darbību.

## ○ Profesionālā ētika

Jautājiet, ko prasa kompetenta, godīga un atbildīga inženieru prakse. Tas ietver nenoteiktības atklāšanu, pierādījumu meklēšanu, atšķirīgu viedokļu reģistrēšanu un riska eskalēšanu, ja komandai trūkst pilnvaru to atrisināt.

[Avoti] - <https://unesdoc.unesco.org/ark:/48223/pf0000381137>  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

# Likumi, standarti, kodeksi un ētika

- Likumi nosaka izpildāmas saistības un aizliegtas darbības.
- Tehniskie standarti nosaka saskaņotas metodes, procesus un pierādījumus.
- Profesionālās ētikas kodeksi nosaka pienākumus, piemēram, kompetenci, godīgumu, sabiedrības drošību un atbildību. Organizācijas politikas nosaka vietējās atbildības jomas un apstiprināšanas kārtību.
- Ētiskā argumentācija izvērtē izvēles, kas paliek šo robežu ietvaros. Tā uzdod jautājumu, kura no izvēlēm respektē cilvēkus, godīgi sadala risku un ko var pamatot ar pierādījumiem.
- Atbilstības nodrošināšana atstāj plašas izvēles iespējas attiecībā uz mērķiem, datiem, robežvērtībām, saskarnēm, uzraudzību un kontroli. Inženieri joprojām ir atbildīgi par ētisko risku atpazīšanu un ziņošanu par tiem savas darbības ietvaros.

[Avoti] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
„Ethical Engineer” apstiprināja KA220 HED pieteikumu, WP3 „Mācīšana un mācīšanās”, 61.–70. lpp.

# AI un ētika (I)

AI var iestrādāt vai pastiprināt aizspriedumus, apgrūtināt atbildības noteikšanu, ietekmēt drošību un tiesības, kā arī mainīt varas sadalījumu starp ieinteresētajām pusēm. Lai gan AI sniedz skaidras priekšrocības, tā rada arī ētiskas problēmas, kuras nevar ignorēt.

Algoritmi var iestrādāt vai pastiprināt esošos aizspriedumus — piemēram, darbā pieņemšanas sistēma var netīšām diskriminēt noteiktas grupas, ja tā ir apmācīta, izmantojot aizspriedumainus vēsturiskos datus. Atbildība var kļūt neskaidra, ja lēmumus pieņem sarežģīti modeļi, tādējādi radot neskaidrību par to, kurš ir atbildīgs, ja rodas kaitējums.

Apdraudēta ir arī drošība un tiesības, jo kļūdas autonomajos transportlīdzekļos vai medicīniskajās AI sistēmās var izraisīt fiziskus bojājumus. Tajā pašā laikā uzraudzības tehnoloģijas var pārkāpt privātumu un cilvēka cieņu.

# Mākslīgais intelekts un ētika (II)

Turklāt AI pārdala varu starp ieinteresētajām pusēm: uzņēmumiem, inženieriem, regulatīvajām iestādēm un sabiedrībai var būt pretrunīgas intereses, un tie, kam ir kontrole pār datiem un algoritmiem, var iegūt nesamērīgu ietekmi.

UNESCO ieteikumā par mākslīgā intelekta ētiku ir uzsvērti šie riski, norādot, ka cilvēka cieņa, taisnīgums, pārredzamība un atbildība jāuzskata par galvenajām vērtībām mākslīgā intelekta attīstībā. ES Mākslīgā intelekta likums atspoguļo šīs bažas, aizliedzot manipulējošas un ekspluatējošas mākslīgā intelekta prakses un nosakot stingrus noteikumus augsta riska sistēmām, lai aizsargātu pamattiesības un sabiedrības labklājību.

# Mākslīgā intelekta ētika prasa sistēmisku skatījumu

Mākslīgā intelekta sistēmas apvieno datus, modeļus, saskarnes, cilvēkus, organizācijas un darbības vidi. Ētiskas sekas var rasties no jebkuras šīs sistēmas daļas. Pārmērīga tendence var ienākt caur marķējumiem vai mērķiem. Kaitējums var rasties no tehniski precīza rezultāta, kas izmantots nepareizā kontekstā. Skaidrs paskaidrojums joprojām var būt neefektīvs, ja personai trūkst laika vai pilnvaru rīkoties.

AI maina arī mērogu un ietekmi. Viens modelis var ietekmēt daudzus cilvēkus, atkārtot vēsturiskos modeļus un veidot piekļuvi darbam, izglītībai, veselības aprūpei vai sabiedriskajiem pakalpojumiem. Tādēļ atbildība aptver izstrādi, iepirkumu, ieviešanu, uzraudzību un reaģēšanu uz incidentiem. Ētiskās analīzes vienība ir viss pakalpojums un tā sekas skartajiem cilvēkiem.

[Avots] <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

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

# Ētikas principi un profesionālā inženieru rīcība

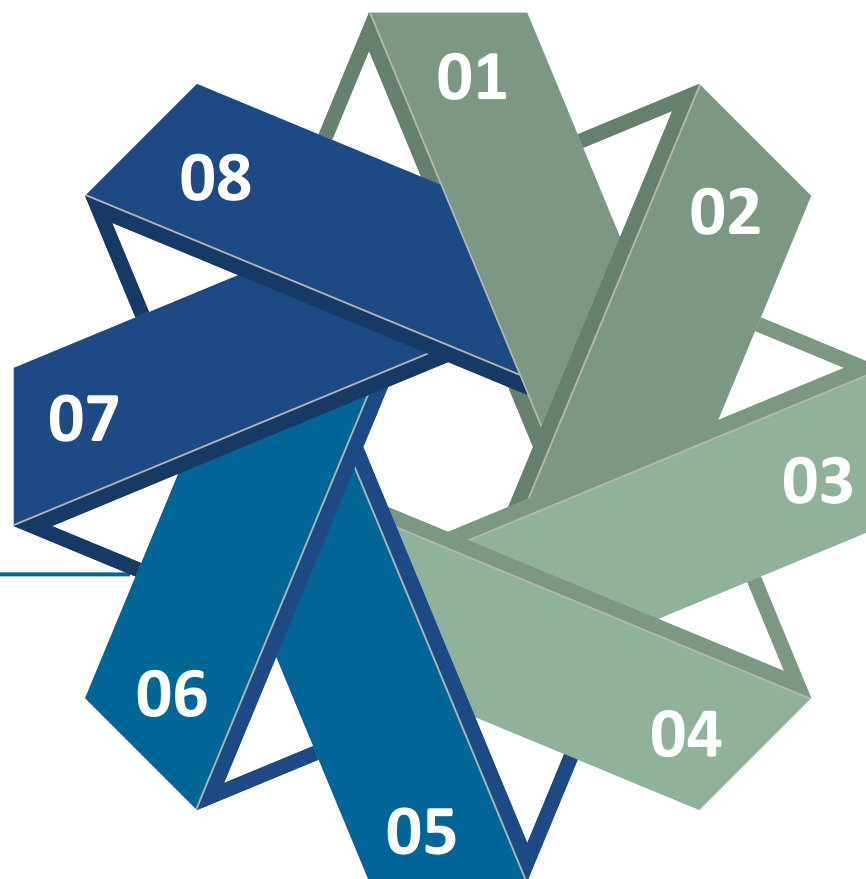
Profesionālā inženieru ētika ietver:

05. Cieņa

06. Likumība

07. Atbildība

08. Centieni pēc izcilības



01. Godīgi

02. Godīgums

03. Uzticamība

04. Godīgums

# Ētikas principi un profesionālā inženieru rīcība

*Šīs vērtības atbalsta ētisko spriedumu. Al inženierijai ir nepieciešamas arī sistēmas līmeņa prasības.*

## Turpmākajās sadaļās tiek aplūkots:

- autonomija
- TAISNĪGUMS
- pārredzamība un nekaitīgums attiecībā uz datiem,
- modelis
- saskarnes
- ieviešanas,
- uzraudzības un lēmumu.



# Četri principi darbojas kopā

Taisnīgums jautā, kā tiek sadalīti ieguvumi, kļūdas un slogs un vai skartajiem cilvēkiem ir nodrošināts taisnīgs process un tiesiskās aizsardzības līdzekļi.

Autonomija nosaka, kam pieder informācija, izvēles tiesības, pilnvaras un nozīmīga kontrole. Pārredzamība nosaka, kādi pierādījumi jābūt pieejamiem skartajām personām, operatoriem, inženieriem un revidentiem.

Nekaitīguma princips nosaka, kādi kaitējumi ir paredzami un kā tos novērst, atklāt un risināt. Vienots projektēšanas lēmums var ietekmēt visus principus. Medicīniskās trauksmes sliekšnis ietekmē to, kam tiek pievērsta uzmanība, vai klīnikas personāls saglabā kontroli, kāda nenoteiktība ir redzama un kuriem pacientiem var tikt nodarīts kaitējums kavēšanās vai viltus trauksmes dēļ.

Ētiskā analīze kļūst spēcīgāka, ja principi tiek izmantoti kā savstarpēji saistīti jautājumi, nevis kā izolēti apzīmējumi.

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

# Pārdomājiet – apspriediet – ziņojiet (10–15 minūtes)

|           |             |                                                                                                                                                   |
|-----------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Jautājums |             | <i>“Kur jūs jau sastopaties ar mākslīgo intelektu (vai automatizāciju) ikdienas dzīvē vai studijās? Kādas ētiskas problēmas varētu rasties?”</i>  |
| Pārdo     | 2 min       | Studenti uzraksta 2–3 piemērus/jautājumus.                                                                                                        |
| Pāri      | 5 minūtes   | Salīdzina sarakstus ar partneri.                                                                                                                  |
| Dalīša    | 3–8 minūtes | Uz tāfeles apkopojiet 4–6 piemērus (piemēram: sejas atpazīšana un piekrišana, vērtēšanas rīki un taisnīgums; ieteikumu sistēmas un pārredzamība). |

# 02

# Taisnīgums mākslīgā intelekta inženierijā

Taisnīgums attiecas uz godīgu un vienlīdzīgu attieksmi, tostarp uz to, kā tiek sadalīti ieguvumi, riski, kļūdas un iespējas.

## Inženieriem tas prasa pierādījumus:

- Vai attiecīgās grupas ir pārstāvētas datos un validācijas kopā?
- Vai kļūdu rādītāji un ietekme tiek novērtēti visās skartajās grupās?
- Vai pieejamības un izslēgšanas riski ietekmē izstrādi?
- Vai cilvēkiem ir iespēja iesniegt apelāciju, saņemt cilvēka veiktu pārskatīšanu vai vērsties pēc tiesiskās aizsardzības?

*Taisnīgums ir atkarīgs no konteksta. Komandām jādokumentē izvēlētie taisnīguma mērķi, kompromisi un ietekmētās ieinteresētās puses.*

[Avoti] - <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

# Taisnīgumam ir četri inženierijas aspekti

1. **Sadales taisnīgums attiecas uz** to, kurš saņem labumu, kļūdas, izmaksas, kavējumus un iespējas. Inženieriem ir nepieciešami rezultātu rādītāji attiecīgajām grupām un darbības apstākļiem.
2. **Procedurālā taisnīguma jautājumi attiecas uz** to, kā tiek pieņemti lēmumi. Cilvēkiem ir nepieciešami konsekventi noteikumi, pamatota cilvēka veikta pārskatīšana un iespēja apstrīdēt rezultātu.
3. **Atzīšana attiecas uz** to, vai projektā ir ņemtas vērā to cilvēku vajadzības un pieredze, kurus varētu nepamanīt, tostarp cilvēki ar invaliditāti un mazākumtautību grupas.
4. **Korektīvā taisnīguma jautājumi attiecas uz** labojumu. Pakalpojumam ir nepieciešams process, lai labotu kļūdu, atjaunotu piekļuvi, izmeklētu cēloņus un novērstu atkātošanos.

[Avoti] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>

# Reāls gadījums: sejas atpazīšanas klūdas

Oficiali  
pierādījumi: NIST  
sejas atpazīšanas  
pētījums

NIST novērtēja gandrīz 200 sejas atpazīšanas algoritmus, izmantojot vairāk nekā 18 miljonus attēlu, kuros bija redzami vairāk nekā 8 miljoni cilvēku. Lielākajā daļā testēto algoritmu tika novērotas demogrāfiskas atšķirības. Kļūdu raksturs bija atkarīgs no algoritma, lietojumprogrammas un attēlu datiem.

Inženierijas jautājums attiecas gan uz lietošanas kontekstu, gan uz modeļa veiktspēju. Kļūdaina saskaņošana var pakļaut personu izmeklēšanai. Neveiksmīga verifikācija var liegt piekļuvi pakalpojumam. Inženieriem ir jāidentificē attiecīgā kļūda, jātestē ietekmētās grupas un reālistiski attēlu apstākļi, jādokumentē nenoteiktība un jānodrošina cilvēka veikta pārskatīšana un kļūdu

## Diskusija:

**Kura kļūda rada vislielāko kaitējumu piekļuvei universitātes teritorijai, policijas darbā un tālruna autentifikācijā?**

[Avoti] <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>

# Praktisks uzdevums: Taisnīgums infrastruktūras sadalē

Reģionālā iestāde izmanto mākslīgā intelekta modeli, lai noteiktu tiltu pārbaužu prioritāti. Pilsētu ceļu tiltiem ir bagātīgi sensoru un apkopes dati. Lauku tiltiem, gājēju pārejām un maršrutiem, ko izmanto cilvēki ar invaliditāti, ir nepilnīgi dati. Modelis lielāko daļu pārbaužu resursu novirza uz vietām ar visplašāko datu klāstu un vislielāko satiksmes intensitāti.

Taisnīgums jautā, kā tiek sadalīta uzmanība drošībai, riskiem un pieejamībai, un vai novārtā atstātās kopienas var saņemt taisnīgu pārskatīšanu.

---

**Lēmums:** TURPINĀT, PĀRSKATĪT vai PĀRTRAUKT. Norādiet ietekmētās grupas, trūkstošos pierādījumus, taisnīgus sadales noteikumus, manuālo izvirzīšanas procedūru, pieņemšanas kritērijus, atbildīgo personu un risinājumu.

**Kvalitātes pārbaude:** Kas saņem agrīnu aizsardzību, kas uzņemas atlikušo risku un kā var labot izlaidumu?

# 03

# Cilvēka autonomija un inženieru atbildība

Cilvēka autonomija ir spēja pieņemt apzinātas izvēles un rīkoties bez nepamatota spiediena. Mākslīgā intelekta sistēmās tā ir atkarīga no nozīmīgas informācijas, vajadzības gadījumā – no piekrišanas, patiesas lietotāja kontroles un iespējām apstrīdēt vai atcelt rezultātus.

Inženieri ietekmē autonomiju, izmantojot saskarnes dizainu, noklusējuma iestatījumus, automatizācijas līmeni, pārrakstīšanas mehānismus un lēmumu pieņemšanas pilnvaru sadali.

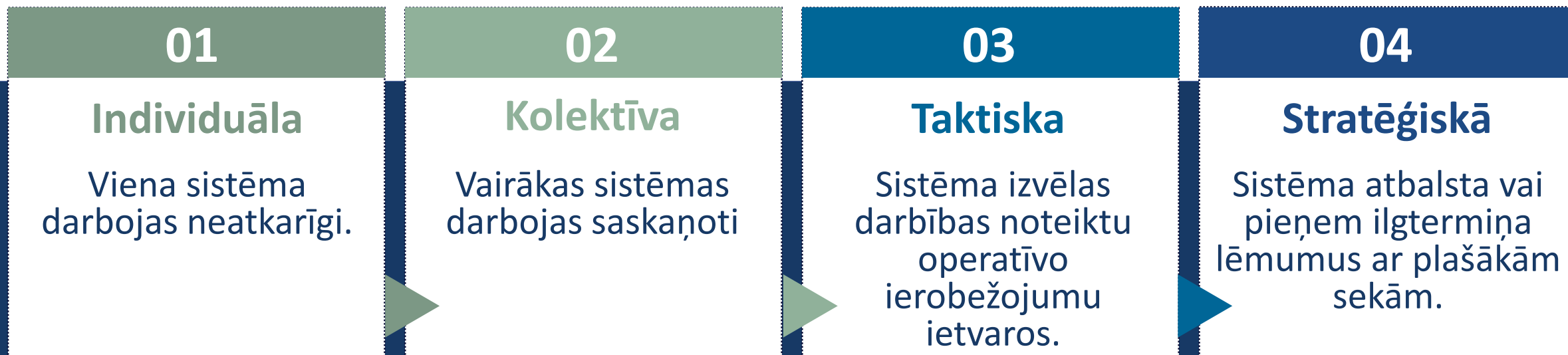
## Galvenā atšķirība:

Sistēmas autonomija apraksta to, ko mašīna var darīt patstāvīgi. Cilvēka autonomija attiecas uz sistēmas ietekmēto cilvēku rīcības brīvību un tiesībām

[Avoti] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

# Sistēmas autonomijas līmeņi

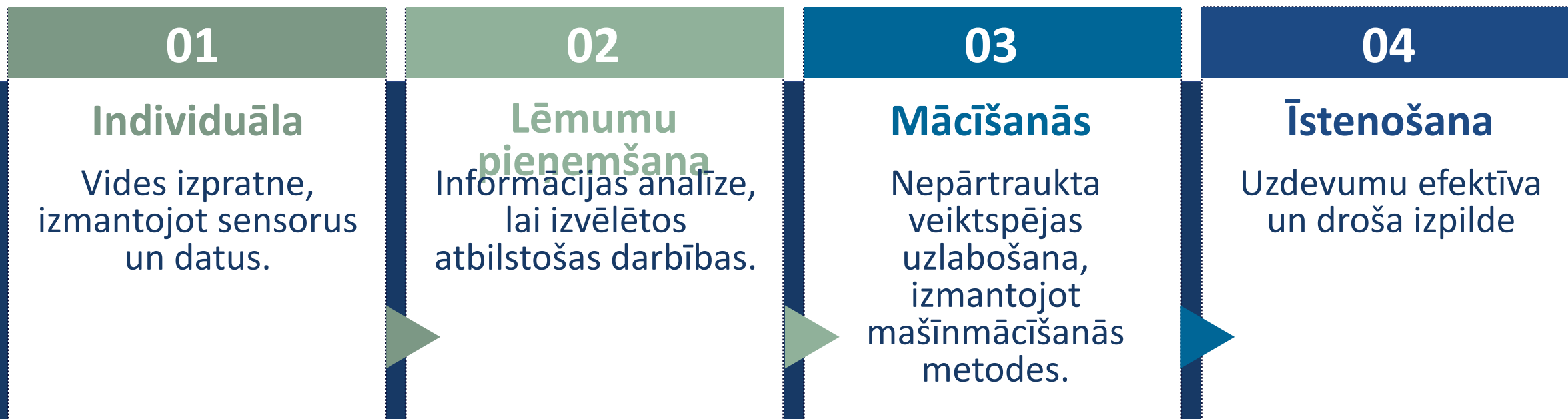
Inženieri var aprakstīt autonomiju dažādos līmeņos:



Katram līmenim jānosaka darbības robežas, cilvēka uzraudzība, drošības pasākumi, atbildība un pierādījumi, kas nepieciešami pirms ieviešanas. Katram līmenim ir savs riska profils un drošības prasības.

# Autonomais AI (II)

Autonomās AI galvenās iezīmes ietver:



Autonomā AI pārveido dažādas nozares, piemēram, transportu, ražošanu, veselības aprūpi un loģistiku, nodrošinot lielāku efektivitāti, drošību un jaunas iespējas. Tomēr tā rada arī ētiskus, drošības un regulatīvus apsvērumus, jo īpaši attiecībā uz atbildību par lēmumiem un ietekmi uz nodarbinātību.

# Autonomija maina lēmumu pieņemšanas pilnvaras

Katrai automatizētajai funkcijai inženieriem jānorāda atļautā darbība, darbības robežas, cilvēka loma, informācija, kas nepieciešama ieviešanai, reaģēšanas laiks, rezerves risinājums un atbildīgā persona. Darba plūsmā norādītajai personai ir nozīmīga kontrole tikai tad, ja šī persona spēj atpazīt problēmu, izprast sistēmas stāvokli un rīkoties, pirms rodas kaitējums.

Sistēmas autonomija apraksta funkcijas, kas tiek veiktas bez tūlītējas cilvēka rīcības. Cilvēka autonomija attiecas uz cilvēku spēju saprast, izvēlēties, atteikties, iesaistīties un apstrīdēt lēmumus. Tādējādi ētiskais jautājums ir par pilnvarām: kurš drīkst lemt, kurš drīkst rīkoties un kurš var apturēt procesu.

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

# Reāls gadījums: automatizētās braukšanas uzraudzība

2018. gada Tempe avārijā Uber testa automašīna notrieca un nogalināja gājēju. NTSB konstatēja nepietiekamu drošības riska novērtējumu, neefektīvu uzraudzību un vāju drošības kultūru. Sadursmes brīdinājums un automātiskā avārijas bremzēšana bija atslēgti. Oficiālā izmeklēšana: NTSB pieņemtais

zinojums.

## Diskusija:

*Cilvēka uzraudzība ir paredzēta kā drošības funkcija. Tai nepieciešama skaidra informācija par sistēmas stāvokli, reālas modrības prasības, pietiekams reaģēšanas laiks, pārbaudīta iejaukšanās, droša rezerves sistēma un organizatoriska uzraudzība.*

Kādi lēmumi par sistēmu, saskarnēm un pārvaldību ietekmēja rezultātu?

# Praktiskais uzdevums: Autonomija procesa vadībā

Mākslīgā intelekta kontrolieris regulē dzesēšanas ūdens vārstus ūdeņraža elektrolizatorā. Tas var mainīt vārsta stāvokli katru sekundi. Sensora kļūmes gadījumā astoņu sekunžu laikā var izveidoties nedrošs spiediens.

Operatoram ir nepieciešamas aptuveni desmit sekundes, lai izprastu trauksmes signālu, un vēl piecas sekundes, lai rīkotos. Piegādātājs sistēmu apraksta kā cilvēka uzraudzītu, jo tajā ir pieejama avārijas apstādināšanas poga.

---

**Izlemiet:** TURPINĀT, PĀRSKATĪT vai APSTĀTĪTIES. Precizējiet kontroliera atļautās darbības, darbības robežas, cilvēka pilnvaras, trauksmes informāciju, reaģēšanas laiku, automātisko drošo stāvokli, neatkarīgo aizsardzības slāni, pārbaudes testu, pierādījumu īpašnieku un pārskatīšanas iemeslu.

---

**Kvalitātes pārbaude:** Vai cilvēka kontrole spēj darboties laicīgi reālos kļūmju un darba slodzes apstākļos?

# 04

# Pārredzamība kā inženierijas īpašība

Pārredzamība ļauj iesaistītajām personām, operatoriem, revidentiem un regulatīvajām iestādēm saprast, kādam nolūkam paredzēta AI sistēma, kā tā tiek izmantota un kādas ir tās robežas

*Inženieri nodrošina pārredzamību, izmantojot dokumentāciju, izsekojamību un saskarnes dizainu*

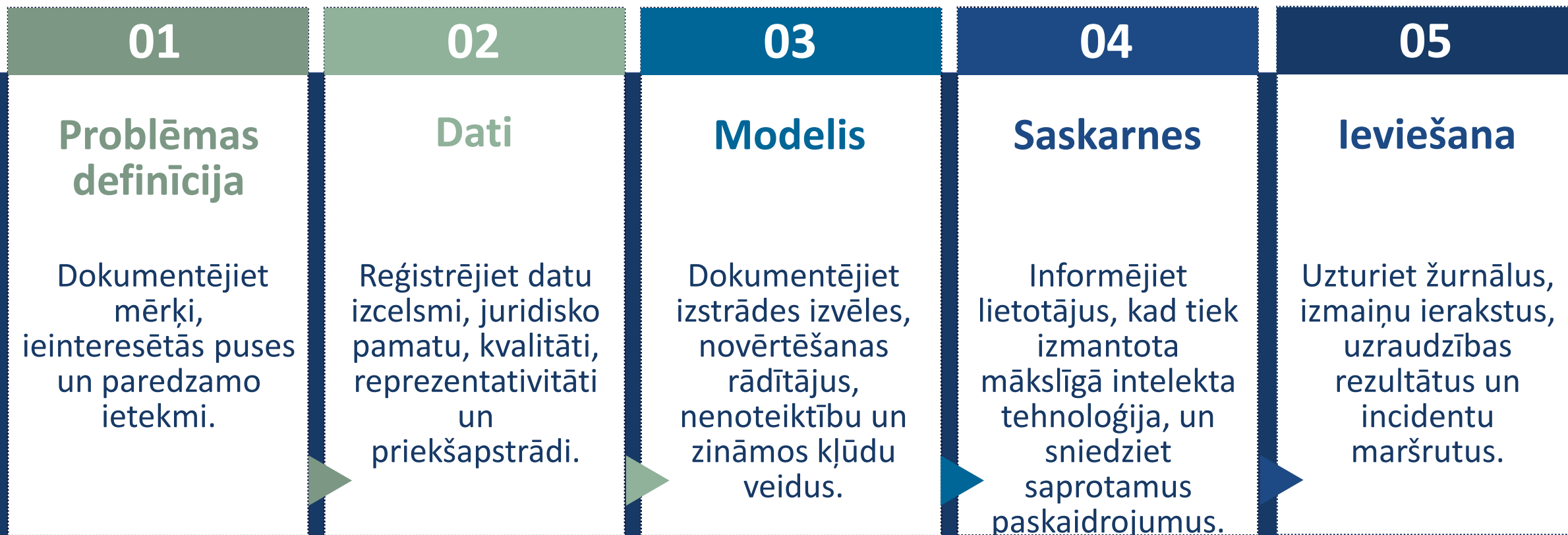
## Attiecīgā informācija var ietvert:

- Paredzētais lietojums un aizliegtais lietojums
- Datu avoti, kvalitāte un zināmās nepilnības
- Modeļa ierobežojumi, nenoteiktība un veikspēja dažādās grupās
- Cilvēka uzraudzība un eskalācijas ceļi
- Uzraudzība, incidenti un sistēmas izmaiņas

[Avoti] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

# Pārredzamība visā AI dzīves ciklā

Pārredzamībai jābūt atbilstošai mērķauditorijai, vienlaikus aizsargājot privātumu, drošību un likumīgos intelektuālā īpašuma tiesības.



# Inženierijas kompromisi pārredzamības jomā

Pārredzamība veicina uzticēšanos, atbildību, pārbaudi un apzinātu lietošanu. Tā var radīt riskus, ja jutīgi personas dati, drošības informācija vai aizsargāts intelektuālais īpašums tiek atklāti bez aizsardzības pasākumiem.

## Inženierijas mērķis ir mērķtiecīga pārredzamība:

- Atklājiet to, kas katrai ieinteresētajai pusei nepieciešams, lai rīkotos droši un apstrīdētu lēmumus
- Aizsargāji informāciju, kas varētu apdraudēt privātumu vai drošību
- Dokumentējiet, kāpēc informācija netiek atklāta un kam ir piekļuve tai revīzijas nolūkos
- Pārbaudiet, vai paskaidrojumi ir precīzi, lietojami un pieejami

[Avoti] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

# AI pārredzamības galvenās sastāvdaļas (I)

- **Paskaidrojamība**

Spēja izskaidrot, kā AI nonāca pie konkrēta lēmuma vai rezultāta (piemēram, kāpēc tika atteikts kredīts).

- **Datu pārredzamība**

Skaidra dokumentācija par to, kādi dati tika izmantoti modeļa apmācībai, ieskaitot avotus, kvalitāti un iespējamās neobjektivitātes

- **Algoritmu pārredzamība**

Piekļuve izmantotajiem algoritmiem (piemēram, atvērtā koda modeļi, lēmumu loģikas dokumentācija).

# Pārredzamība ir inženierijas pierādījums rīcībai

Pārredzamība veicina drošu rīcību, izaicinājumus un atbildību. Tās mērķis mainās atkarībā no mērķauditorijas. Operatoriem ir nepieciešama informācija par sistēmas stāvokli, nenoteiktību un nepieciešamo reakciju. Ietekmētajiem cilvēkiem ir jāzina, ka tika izmantota mākslīgā intelekta tehnoloģija, kā rezultāts viņus ietekmē un kā lūgt pārskatīšanu. Inženieriem un revidentiem ir nepieciešama izsekojamība attiecībā uz datiem, testiem, versijām, lēmumiem un incidentiem.

## Inženierijas artefakti padara to praktiski iespējamu:

- Sistēmas un modeļu dokumentācija nosaka paredzēto lietojumu, ierobežojumus un kļūdu režīmus.
- Testa rezultāti aptver attiecīgos apstākļus un ietekmētās grupas.
- Saskarnes ziņojumi informē par nenoteiktību, rīcību un eskalāciju.
- Žurnāli, izmaiņu ieraksti un incidentu ziņojumi atbalsta rekonstrukciju un pārskatīšanu.

[Avoti] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

# Pārredzamība ir atkarīga no mērķauditorijas un mērķa

Noderīga pārredzamība ļauj konkrētai personai saprast, rīkoties, apstrīdēt vai veikt revīziju. Tehnisko detaļu publicēšana, neidentificējot mērķauditoriju, var radīt šķietamu atklātību, vienlaikus atstājot lēmumu nepārskatāmu. Informācijai jābūt precīzai, savlaicīgai, pieejamai un saistītai ar reālu rīcības plānu.

- Ietekmei pakļautajai personai ir jāzina, ka tiek izmantota mākslīgā intelekta sistēma, kāds lēmums ar tās palīdzību tiek pieņemts, kādi ir attiecīgie pamatojumi, svarīgie ierobežojumi un kā panākt cilvēka veiktu pārskatīšanu.
- Operatoram ir nepieciešama informācija par sistēmas stāvokli, uzticamību, nepieciešamajām darbībām, darbības ierobežojumiem un rīcību eskalācijas gadījumā.
- Inženieriem un revidentiem ir nepieciešama informācija par datu izcelsmi, versijām, testiem, žurnāliem, incidentiem, izmaiņām un apstiprinājumiem.

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

# Reāls inženierijas gadījums: Boeing 737 MAX pārredzamība

Oficiālā izmeklēšana:  
„Boeing 737 MAX“  
ziņojuma materiāli.

[Avoti] <https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>

ASV Pārstāvju palātas komitejas izmeklēšanā par Boeing 737 MAX tika pārbaudīta gaisa kuģa konstrukcija, izstrāde un sertifikācija. Tās ziņojumā tika identificētas atkārtotas kļūmes, kas saistītas ar manevrēšanas īpašību uzlabošanas sistēmu (Maneuvering Characteristics Augmentation System, MCAS), kā arī plašākas nepilnības sertifikācijas un uzraudzības jomā.

Pārredzamības jautājums attiecas uz informāciju, kas nepieciešama drošai rīcībai un neatkarīgai pārskatīšanai. Piloti, aviokompānijas un regulatori pieprasīja precīzus pierādījumus par sistēmas pilnvarām, atkarību no sensoriem, darbības traucējumu raksturu, indikācijām pilotu kabīnē, procedūrām un apmācības vajadzībām. Nepilnīga vai sadrumstalota informācija vājināja viņu

**Diskusija:** ...strīdēt konstrukciju.

***Kādi inženierijas dokumenti, pierādījumi par apdraudējumiem un lēmumi par izmaiņām būtu bijuši jāpadara pieejami, kam un kādā apstiprināšanas posmā?***

# Praktiskais uzdevums: Pārredzama dzelzceļa inženierijas iepirkuma procedūra

*Dzelzceļa operators plāno iepirkt mākslīgā intelekta sistēmu, kas identificē sliežu defektus, izmantojot pārbaudes attēlus. Piegādātājs ziņo par 97 % kopējo precizitāti, bet neatklāj veikspēju atkarībā no dzelzceļa veida, laika apstākļiem, kameras stāvokļa vai defekta smaguma pakāpes. Nav pieejama apmācības datu izcelsme un zināmie kļūdu veidi. Līgums atļauj modeļa atjauninājumus bez iepriekšēja paziņojuma. Par nepamanītiem defektiem atbildību saglabā apkopes personāls.*

## Uzraksti

Aprakstiet pierādījumus un līguma prasības, kas nepieciešamas pirms iepirkuma. Iekļaujiet paredzēto lietojumu, ekspluatācijas ierobežojumus, datu kopu izcelsmi, defektu līmeņa rezultātus, nenoteiktību, versiju un izmaiņu reģistrus, piekļuvi žurnāliem, sadarbību incidentu gadījumos, neatkarīgu testēšanu un izstāšanās tiesības.

## Lēmums

Norādiet vienu nosacījumu, kas prasa REVISE, un vienu nosacījumu, kas prasa STOP. Paskaidrojiet, kā katra mērķauditorija var rīkoties, pamatojoties uz sniegto informāciju

# 05

# Nekaitīgums:

Nenodarīt kaitējumu, paredzēt un mazināt riskus. Nekaitīguma princips prasa, lai inženieri izvairītos no tādu sistēmu izveides, kas rada fizisku, psiholoģisku vai sabiedrisku kaitējumu.

Tas prasa arī iepriekš paredzēt potenciālos draudus un ieviest aizsardzības pasākumus, lai mazinātu negribētas sekas.

# Kaitējums sniedzas tālāk par fiziskām traumām

- Fizisks kaitējums ietver traumas un nedrošu darbību. Psiholoģisks kaitējums ietver stresu, manipulāciju un pašāvības zaudējumu. Ekonomisks kaitējums ietver darba zaudējumu, augstākas izmaksas un izslēgšanu no pakalpojumiem. Ar tiesībām saistīts kaitējums ietver privātuma pārkāpumu, diskrimināciju un tiesību uz taisnīgu tiesas procesu neievērošanu. Sociāls un vides kaitējums ietver uzraudzību, nevienlīdzīgu varas sadali, resursu izmantošanu un ietekmi uz kopienām.

Nekaitīguma princips prasa, lai inženieri pirms kontroles pasākumu izvēles identificētu apdraudētos cilvēkus un iespējamo kaitējumu. Kopā jāņem vērā smagums, varbūtība, pakļautība un atgriezeniskums. Retam, bet smagam kaitējumam, biežai, bet nelielai slodzei un kļūdai, ko nevar labot, katrā gadījumā ir nepieciešama atskirīga rīcība.

[Avoti], <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://eur-lex.europa.eu/eli/reg/2018/1145/oj/48223/pf0000381137>

# Drošība ir pienākums visā dzīves ciklā

- Pirms izlaišanas jādefinē paredzētais un aizliegtais lietojums, jāidentificē apdraudējumi un paredzamā nepareiza izmantošana, jātestē robežgadījumi un pasliktināti apstākļi, kā arī jānosaka izmērāmi pieņemamības kritēriji. Eksploatācijas laikā jāuzrauga veikspēja, grupu rezultāti, pārrakstīšanas gadījumi, sūdzības, incidenti un konteksta izmaiņas. Pēc incidenta jāaizsargā skartie cilvēki, jā saglabā pierādījumi, jānosaka cēloņi, jā korekti kontrolpasākumi un jāpārbauda izmaiņas.

Kontroles pasākumiem, ja iespējams, jānovērš kaitējuma avots. Komandas var mainīt mērķi, noņemt bīstamu funkciju, ierobežot darbības jomu, pievienot tehniskās atklāšanas un rezerves sistēmas, pastiprināt cilvēku veikto pārbaudi vai apturēt ieviešanu. Drošības novērtējums paliek spēkā tikai tik ilgi, kamēr tā pieņēmumi, sistēmas versija, konteksts un pierādījumi paliek spēkā.

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

# Reāls inženierijas gadījums: BP Texas City procesa drošība

2005. gada 23. martā sprādzienos BP Texas City naftas pārstrādes rūpnīcā gāja bojā 15 cilvēki un 180 tika ievainoti. ASV Ķīmiskās drošības padome ziņoja, ka destilācijas tornis tika applūdis, tajā radās pārspiediens un no atmosfēras izplūdes skursteņa izplūda ogļūdeņraži. Izmeklēšanā tika konstatēti organizatoriski un drošības trūkumi visā uzņēmumā.

„Nekaitīguma princips” prasa, lai inženieri novērstu paredzamu kaitējumu visā sistēmā. Iekārtu konstrukcija, trauksmes signāli, ekspluatācijas procedūras, personāla komplektēšana, pārmaiņu vadība, ražotnes plānojums, galvenie drošības rādītāji un vadības uzraudzība – tas viss ietekmē risku. Personīgās drošības statistika nevar aizstāt pierādījumus par katastrofāliem procesa apdraudējumiem.

## Diskusija:

Kuri novēršanas slāņi nedarbojās, kādi pierādījumi būtu bijuši jāuzskata par pamatu rīcībai un kam bija pilnvaras apturēt atkārtoto palaišanu?

Oficiālā izmeklēšana: CSB lieta par Teksasas pilsētu [voti] <https://www.csb.gov/bp-america-texas-city-refinery-explosion/>



# Piemērojamā uzdevuma nosaukums: Ierobežots AI procesa drošības izmēģinājums

Kīmiskā rūpnīca ierosina izmantot mākslīgā intelekta konsultāciju rīku, kas prognozē kompresora pārslodzes un iesaka veikt apvedvārstu izmaiņas. Validācija attiecas uz stabilu darbību. Nav pierādījumu par iedarbināšanu, izslēgšanu, sensoru novirzēm vai sakaru zudumu. Ieteikums tiek piemērots automātiski, ja vien operators to neatsauc sešu sekunžu laikā. Piegādātājs ierosina trīs mēnešu ilgu reālu izmēģinājuma periodu

**Lēmums:** TURPINĀT, PĀRSKATĪT vai PĀRTRAUKT. Noteikt izmēģinājuma robežas, izslēgtos darbības režīmus, neatkarīgos aizsardzības slāņus, cilvēka reaģēšanas laiku, testus pirms ieviešanas, uzraudzības signālus, incidenta maršrutu, apstādināšanas sliekšņus, pierādījumu īpašnieku un pārskatīšanas datumu.

**Izstrādājiēt īsu drošības pamatojumu:** briesmas, cēloņsakarības ceļš, kontrole, pārbaudes pierādījumi, atlikušais risks un atbildīgā apstiprinājuma persona.  
<https://www.csb.gov/bp-america-texas-city-refinery-explosion/>  
Norādiet, kādi trūkstošie pierādījumi kavē reālu darbību.

# 06

# Gadījums A: BridgeSense inspekciju prioritāšu

## noteikšana

BridgeSense veido tiltu pārbaudes prioritāšu sarakstu, izmantojot attēlus, vibrācijas sensorus, satiksmes intensitāti un uzturēšanas vēsturi. Vecākiem lauku tiltiem ir mazāk sensoru un nepilnīgi dati. Tiks pārbaudīti tikai 20 objekti ar visaugstāko novērtējumu. Zems novērtējums var liecināt drīzāk par trūkstošiem datiem, nevis par drošības problēmām. Aizkavēšanās var izraisīt kaitējumu, ceļa slēgšanu vai būtisku piekļuves zaudējumu.

### Uzdevums:

Izleciēt: TURPINĀT, PĀRSKATĪT vai PĀRTRAUKT. Norādiet, kā šis novērtējums būtu jāizmanto profesionālajā spriedumā un resursu sadalē.

### Pieejamie pierādījumi:

Kopējā precizitāte un validācija pilsētvidē.

### Trūkstošie pierādījumi:

Rezultāti pēc tilta tipa un stāvokļa, nenoteiktības, ietekmes uz sabiedrību, pārrakstīšanas un pārskatīšanas ierakstiem

# Gadījums B: „WindControl” turbīnu darbība

WindControl regulē lāpstu leņķi un apkopes laiku. Sistēma tika apmācīta, izmantojot jaunākas turbīnas mērenos laika apstākļos. Objektā ir arī vecākas turbīnas, apledoījums, spēcīgas vēja brāzmas un pārtraukumi sakaros. Interfeiss neuzrāda uzticamības rādītājus vai pamatojošos datus.

Viltus trauksme var izraisīt iekārtas apstādināšanu. Nepamanīta kļūme var sabojāt aprīkojumu vai apdraudēt tehniķus un apkārtējos iedzīvotājus

## Uzdevums:

Apstiprināt ierobežotu izmēģinājuma projektu, pieprasīt pārprojektēšanu vai pārtraukt. Noteikt darbības robežas, cilvēka pilnvaras un nosacījumus automātiskajai rīcībai

## Pieejamie pierādījumi:

Ekstrēmu laika apstākļu testi, vecāku turbīnu rezultāti, reaģēšanas laiks, drošības sistēmas darbība, izmaiņu kontrole un gandrīz notikušo negadījumu uzraudzība.

# Inženierijas gadījuma analīze un komandas rezultāti (30 minūtes)

Izvēlieties vienu  
gadījumu.

Četru vai piecu cilvēku  
grupās sadaliet sistēmu,  
drošības, pierādījumu,  
ieinteresēto personu un  
reģistrētāja lomas

1. Noteiciet paredzēto lietojumu, sistēmas robežas, apstākļus un ietekmētos cilvēkus.
2. Nosauciet inženierijas lēmumu, kas rada dilemmu.
3. Izsekojiet, kā dati, modelis, saskarne, automatizācija un uzraudzība ietekmē rezultātus.
4. Piemērojiet visus četrus principus un identificējiet jebkākus konfliktus.
5. Ierosiniet tehniskos un pārvaldības aizsardzības pasākumus, norādot testus un atbildīgos.
6. Reģistrējiet lēmumu „GO”, „REVISE” vai „STOP” un pārskatīšanas iemeslu.

**Rezultāts:** viens lēmuma protokols un divu minūšu uz pierādījumiem balstīts pamatojums.

[Avoti] „Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

# Piemērojiet četrus principus vienam inženiertehniskajam

Izmantojiet vienu un to pašu **lēmumam** tekstā, lai katrs princips būtu saistīts ar inženierijas darbību.

01

## Taisnīgums

Kam tiek nodrošināta drošība, piekļuve, kavējumi, izmaksas, kļūdas vai risks?  
Vai process ir taisnīgs un atklāts tiesiskai aizsardzībai?

02

## Autonomija

Kam ir informācija un pilnvaras?  
Vai persona var laicīgi iejaukties?

03

## Pārredzamība

Ko katrai mērķauditorijai ir jāzina, lai rīkotos, apstrīdētu vai veiktu izmeklēšanu?

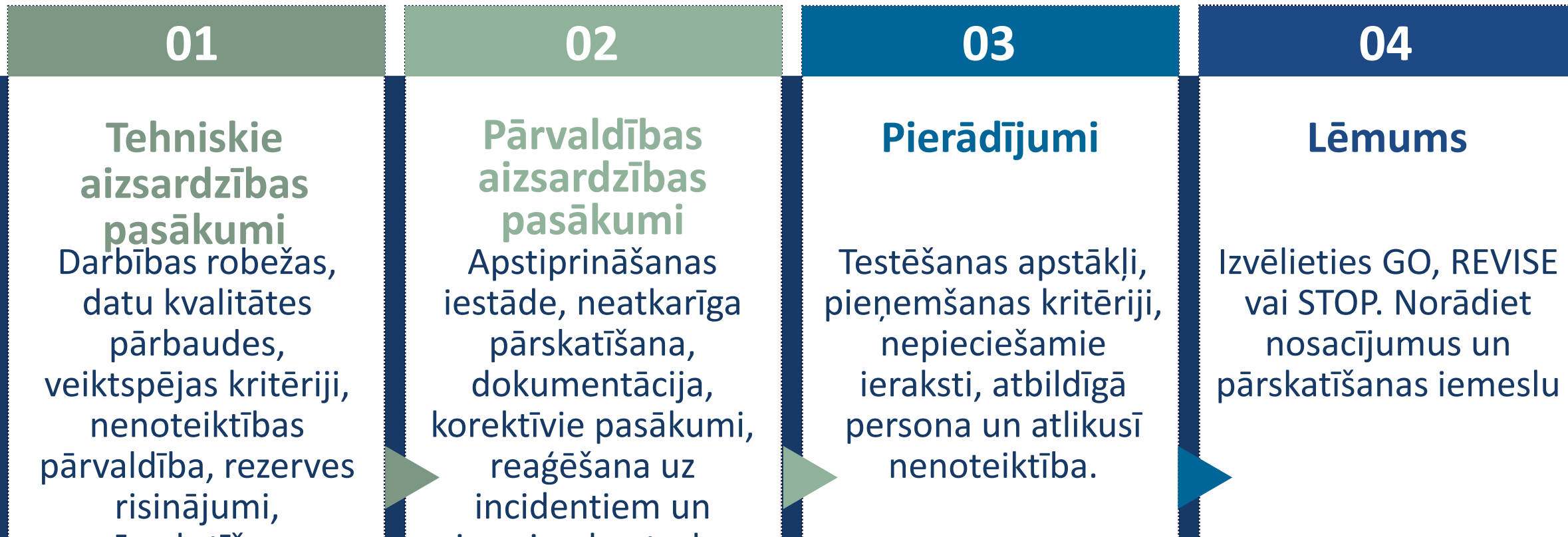
04

## Nekaitīgums

Kādi kaitējumi ir paredzami?  
Profilakses, atklāšanas, seku mazināšanas un atjaunošanas pasākumi ir nepieciešami?

# Projektēšanas drošības pasākumi, pierādījumi un uzraudzība

Pārvērtiet ētisko analīzi inženierijas priekšlikumā.



un uzraudzība.

# Darba lapa: Inženierijas ētikas lēmuma protokols

**Mērķis:** Integrējiet četrus principus vienā novērtējamā inženierijas lēmumā. Aizpildiet BridgeSense vai WindControl reģistrācijas veidlapu, pēc tam sagatavojiet vienu minūti ilgu pamatojumu.

Labā atskaite saista principus ar lēmumiem, drošības pasākumiem, pierādījumiem, atbildīgo īpašnieku un pārskatīšanas iemeslu.

Sistēma, paredzētais lietojums un ietekmētie cilvēki: .....

Kritiskais inženierijas lēmums: .....

Taisnīgums, autonomija, pārredzamība un nekaitīgums: .....

Tehniskie aizsardzības pasākumi un pieņemamības tests .....

Pārvaldības drošības pasākumi un atbildīgais īpašnieks .....

[Avoti] „Ethical Engineer” apstiprināja KA220-HED pieteikumu, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Turpinājums nākamajā slaidā

Trūkstoši pierādījumi, nenoteiktība un atlikušais risks:.....

Lēmums: TURPINĀT PĀRSKATĪT APSTĀT.....

Nosacījumi, uzraudzības pasākumi un pārskatīšanas iemesls:.....

Nesaskaņa vai alternatīvs viedoklis:.....

Pamatojums: Kāpēc pierādījumi pamato šo lēmumu?

.....

.....

# Lomu spēles scenārijs

*(25 minūtes)*

## Scenārijs:

Universitāte apsver iespēju izmantot autonomus patruļrobotus nakts uzraudzībai. Kameras un cilvēku atpazīšanas sistēmas sola ātrāku reaģēšanu uz incidentiem. Studenti pauž bažas par privātumu, diskrimināciju un viltus trauksmēm.

## Lomas, pa vienai katrā grupā:

- **Inženierijas vadītājs:** drošība, veikspēja un ekspluatācijas ierobežojumi
- **Datu zinātnieks:** datu kopas, precizitāte un neobjektivitāte
- **Studentu pārstāvis:** tiesības, iekļaušana un personas dati
- **Ēku pārvaldnieks:** darbība un atbildība
- **Ētikas vai juridiskais konsultants:** riska novērtējums un dokumentācija

[Avotiņš: Ethical Engineer pētījums KA220-HED pieteikums, WP3 „Mācīšana un mācīšanās” (OER izstrāde), 61.–70. lpp.  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>  
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

# Individuāla pārdomāšana un pārnese

*(15 minūtes)*

Refleksija noslēdz grupas darbību un pārnes argumentāciju uz citu inženierzinātņu kontekstu.

## Uzrakstiet vienu vai divus teikumus:

- Kuri divi principi visvairāk ietekmēja jūsu lēmumu un kāpēc?
- Kādu kompromisu vai neskaidrību bija visgrūtāk atrisināt?
- Kādas inženiertehniskas izmaiņas ir nepieciešamas pirms ieviešanas?
- Kādi pierādījumi liktu jums pārskatīt lēmumu?
- Kā mainītos argumentācija citā disciplīnā?

## Pierādījumi:

Īss arguments, kas saista kopā principus, inženierijas lēmumu, drošības pasākumus un pierādījumus. Pieņemamas ir rakstiskas, audio vai ierakstītas atbildes.

[Avoti] „Ethical Engineer” apstiprināja KA220-HED pieteikumu, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.

# 07

# TESTS

1. Definējiet *pārredzamību* mākslīgajā intelektā un miniet vienu inženierijas praksi, kas to atbalsta.
2. Kura no šīm pamatnostādnēm ir vērsta uz to, lai **novērst kaitējumu**, un nosauciet vienu riska mazināšanas metodi AI projektos.
3. Uzskaityt **divus veidus** , kā tendenciozi apmācības dati var ietekmēt modeļa rezultātus.
4. AI atbalstītā darbā pieņemšanas rīkā identificējiet **vienu lēmuma pieņemšanas punktu** , kurā inženieri var samazināt netaisnību.
5. Robots nepareizi klasificē mobilitātes palīgīdzekļus kā šķēršļus, tādējādi bloķējot piekļuvi. Kāds princips tiek visvairāk tieši pārkāpts?
  - a. *Autonomija*
  - b. *Taisnīgums*
  - c. *Pārredzamība*
  - d. *Nekaitīgums*

# Atbildes uz testu jautājumiem un mācību pamatojums

1. Pārredzamība nozīmē atbilstošu skaidrību, izsekojamību un komunikāciju. To atbalstošas prakses ietver datu kopu dokumentāciju, modeļu ierakstus, paskaidrojumus, žurnālus un izmaiņu ierakstus.
2. Nekaitīgums. Derīga metode varētu būt riska analīze, reprezentatīva testēšana, darbības ierobežojumi, droša rezerves sistēma, uzraudzība vai reaģēšana uz incidentiem.
3. Pārmērīgi tendenciozi dati var radīt nevienlīdzīgas kļūdu likmes, sliktus rezultātus nepietiekami pārstāvētām grupām, kaitīgus aizstājējradītājus un vēsturisko modeļu nostiprināšanos.
4. Pamatoti lēmumu pieņemšanas punkti ietver datu atlasīšanu, marķējumus, pazīmes, mērķi, sliekšni, saskarni, cilvēka veikto pārskatīšanu un uzraudzību.
5. B, Taisnīgums. Piekļuves bloķēšana nevienlīdzīgi sadala slogu. Var būt piemērojams arī nekaitīguma princips, jo kļūda rada kaitējumu. Novērtējiet pamatotu atbildi.

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

# Mācību rezultātu pārbaude

## Pirms darba pabeigšanas studentiem jāspēj:

- Nosauciet inženierijas lēmumu, kas konkrētajā gadījumā radīja vai samazināja kaitējumu.
- Izmantojiet vismaz divus ētiskos principus, lai izskaidrotu šo jautājumu.
- Ieteikiet vienu tehnisko aizsardzības pasākumu un vienu pārvaldības aizsardzības pasākumu.
- Norādiet, kādi pierādījumi, dokumentācija un uzraudzība būtu nepieciešami.
- Paskaidrojiet, kā skartās personas var saprast, apšaubīt vai apstrīdēt rezultātu.

## Mācību rezultātu apliecinājums:

Grupu analīze, lomu spēles lēmums, refleksijas atbilde un tests.

[Avoti] „Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.

# Galvenās atsauces un papildu literatūra

- Eiropas Komisijas augsta līmeņa ekspertu grupa par mākslīgo intelektu. Ētikas vadlīnijas uzticamajam AI.
- Vuorikari, Kluzer un Punie. DigComp 2.2.
- **UNESCO.** Ieteikums par mākslīgā intelekta ētiku.
- Eiropas Savienība. Mākslīgā intelekta likums.
- **NIST.** Sejas atpazīšanas piegādātāju tests: demogrāfiskā ietekme.
- NTSB. Tempe automatizētā transportlīdzekļa avārijas izmeklēšana.
- ASV Pārstāvju palātas komiteja. „Boeing 737 MAX” izmeklēšana.
- ASV Ķīmiskās drošības padome. Izmeklēšana par „BP” naftas pārstrādes rūpnīcu Teksasas pilsētā.

- „Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 Mācīšana un mācīšanās, 61.–70. lpp.

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

<https://unesdoc.unesco.org/ark:/48223/pf0000381137>

<https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition>

<https://www.nts.gov/investigations/AccidentReports/Reports/HAR1903.pdf>

<https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>

„Ethical Engineer” apstiprinātais KA220-HED pieteikums, WP3 Mācīšana un mācīšanās (OER izstrāde), 61.–70. lpp.

<https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>

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

[programmatūra](#)

<https://www.csb.gov/bp-america-texas-city-refinery-explosion/>



## 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](http://www.ethicalengineer.eu)

# Paldies

---

Jūs esat pabeidzis **1. moduli**



Co-funded by  
the European Union