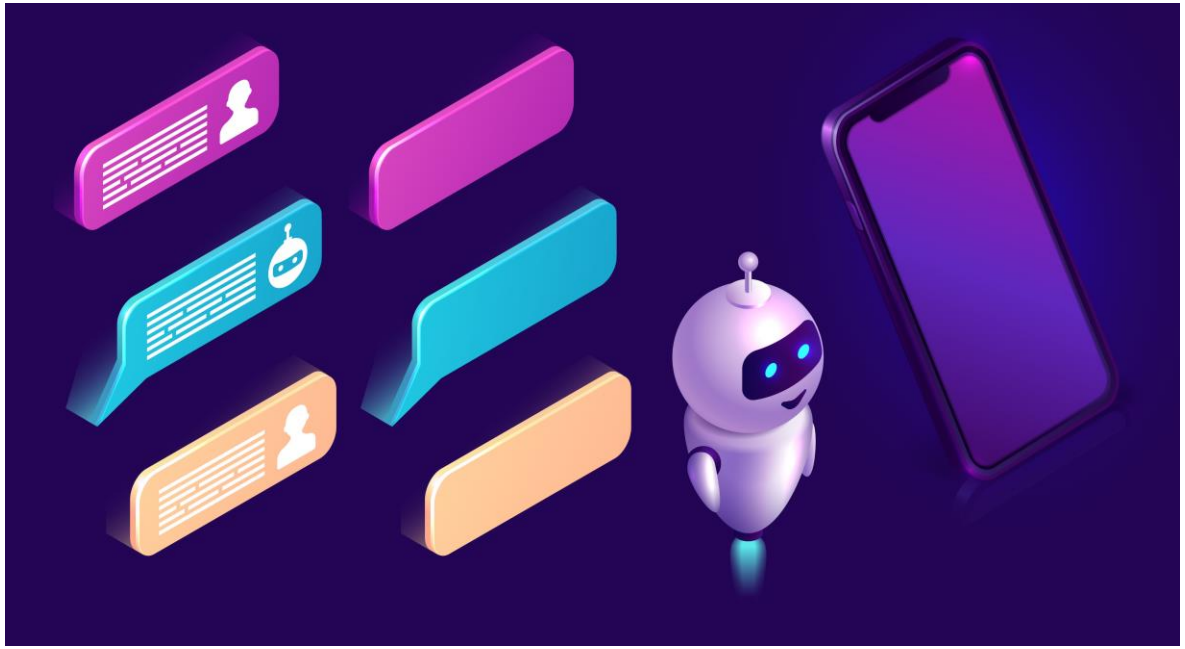
Tehisintellekt ja turvalisus



upklyak / Freepik

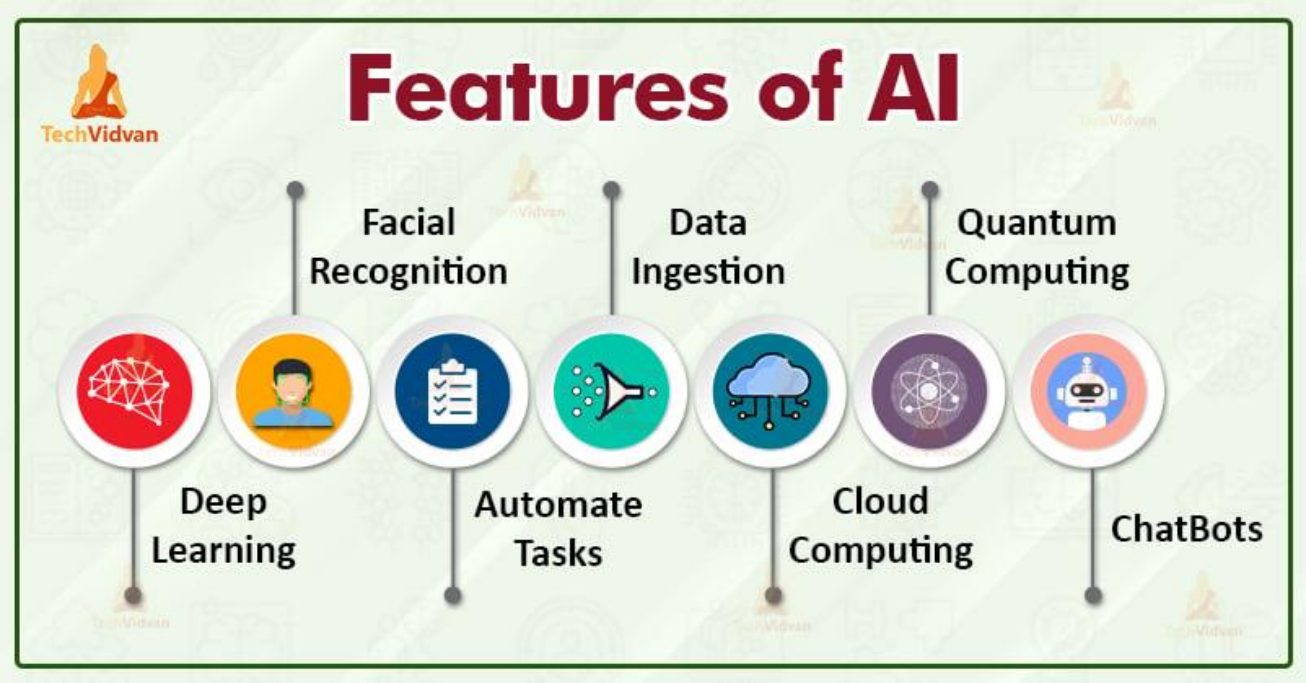
Targalt internetis TAL TECH Lastekaitse Liit Estonian Union for Child Welfare Kaasrahastanud Euroopa Liit

Eesmärgid

- Õpime tundma tehisintellekti turvalisuse põhimõtteid.
- Saame teada tehisintellekti tehnoloogiaga seotud võimalikest ohtudest ja väljakutsetest.
- Oskame tehisintellekti tehnoloogiast tulenevaid ohte märgata ja ennast kaitsta.

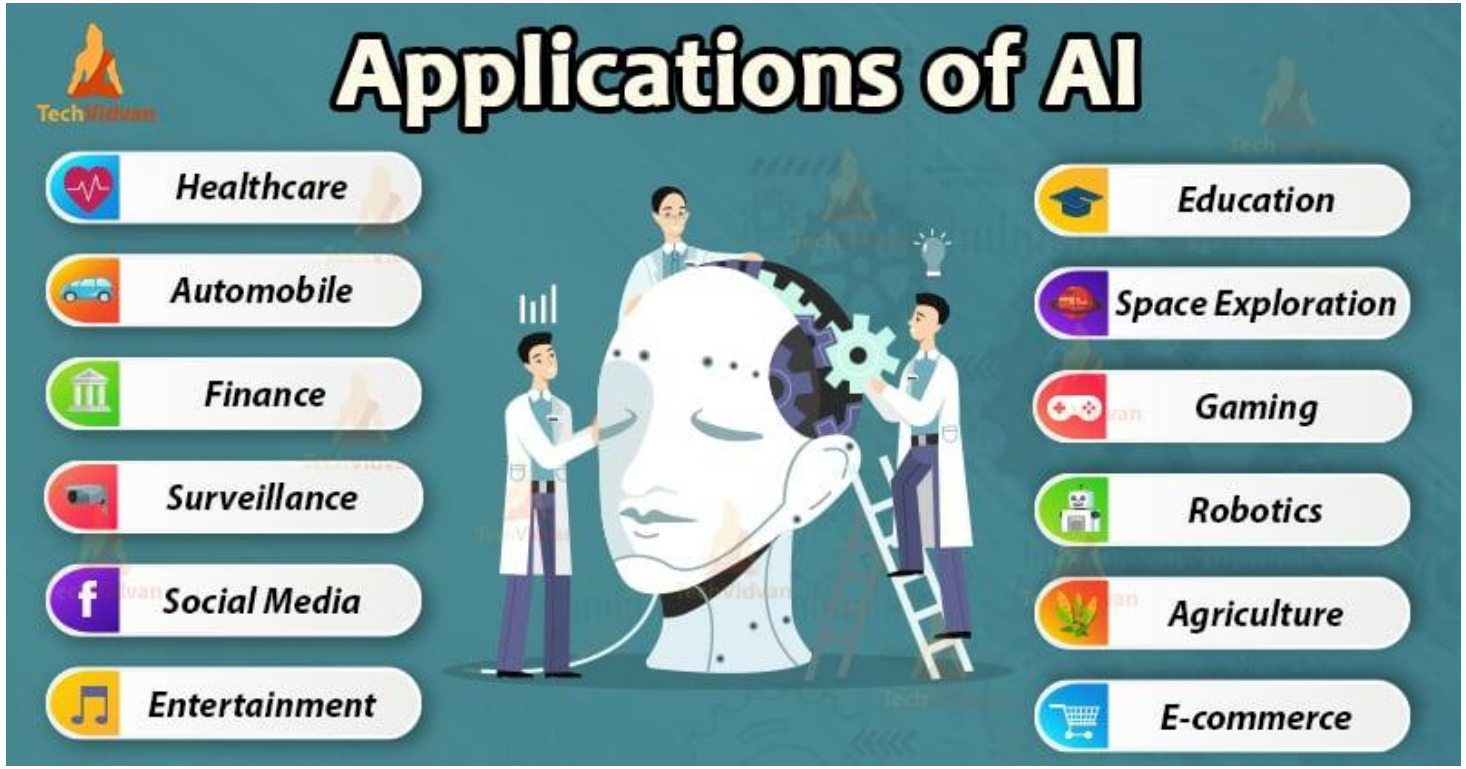
Sissejuhatus

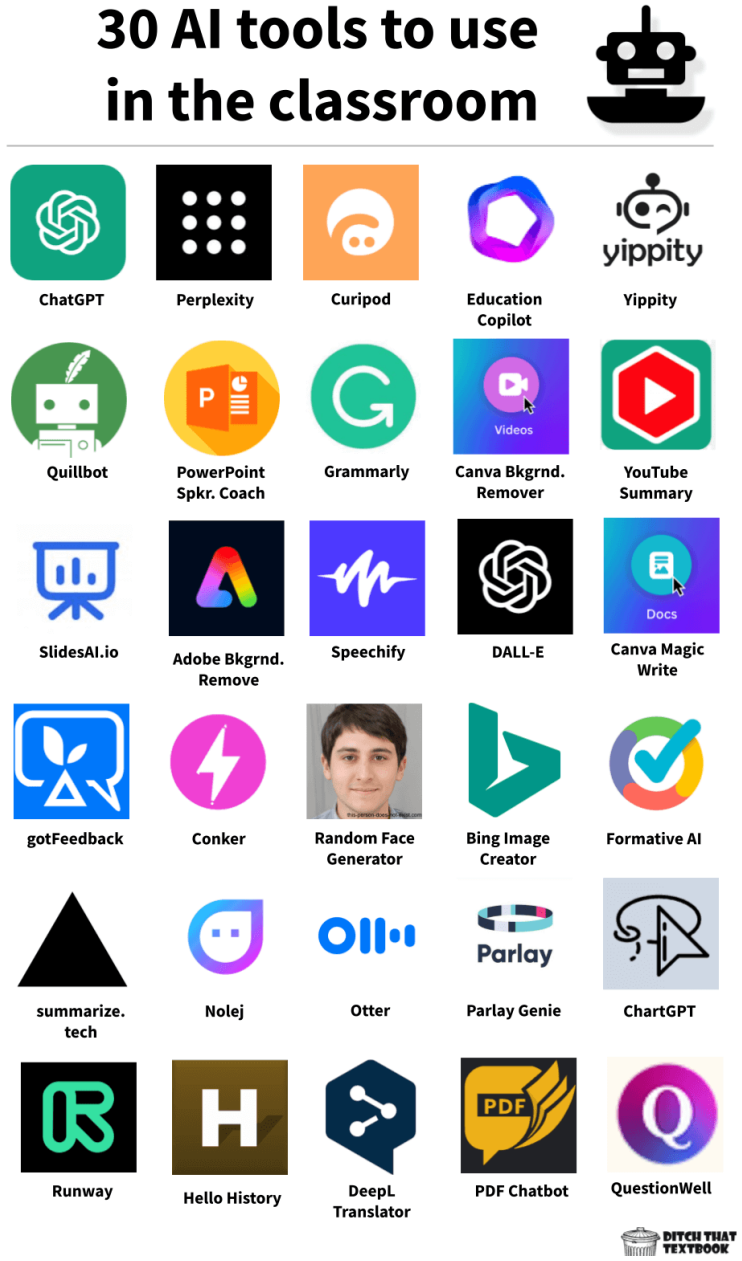
- Mida juba tead tehisintellekti kohta?
- Kas keegi on tehisintellekti juba kasutanud? Kui jah, siis milleks?
- Kuidas tehisintellekti kutsutakse, milliseid erinevaid nimesid tead?



Tehisintellekti (AI) rakendusi

- https://theresanaiforthat.com/





Targalt internetis

Tehisintellekti turvalisuse probleemi näide

- "Kujutle, et sa kasutad virtuaalset assistenti, näiteks nutikat kodust abilist, et vastata küsimustele või teha ülesandeid. Ühel päeval palub assistent sinult isiklikke andmeid, näiteks perekonnanime, aadressi või sünnipäeva.“

- Miks ta seda teeb?
- Kellele need andmed lähevad?
- Kuhu neid hoiustatakse ja kaua?
- Mis kasu sa võid sellest saada, et tehisintellekt teab sinu isiklikke andmeid?

Mida tehisintellekti kasutus võib mõjutada?

Privaatsus

- Kas oled kunagi märganud, et mõned veebilehed või rakendused näivad teadvat, mis sulle meeldib või mida sa otsid?
- See on seotud tehisintellekti privaatsuse probleemiga. **Mõned rakendused koguvad andmeid sinu tegevuste ja eelistuste kohta, et pakkuda sulle isikupärastatud kogemusi. See võib olla küll mugav, kuid mõnikord võivad need andmed sattuda valedesse kätesse või neid kasutatakse sinu teadmata.**
- Jälgi, millist teavet rakendused sinult küsivad, ja mõtle sellele, kas see on tõesti vajalik.
- Kui sa ei soovi jagada teatud andmeid, siis ära tee seda. Samuti võid küsida vanematelt või õpetajalt nõu, kui oled mures oma andmete privaatsuse pärast.
- Oluline on hoida oma isiklikud andmed turvaliselt ja mõelda läbi, kuidas ja kellega neid jagada, et vältida võimalikke privaatsusprobleeme.

Mida veel tehisintellekti kasutus võib mõjutada?

Eelarvamustega algoritmid

- Mõnikord kasutavad tehnoloogiad, nagu nutikad rakendused ja otsingumootorid, algoritme, et teha sulle soovitusi ja pakkuda infot. Kuid need algoritmid võivad mõnikord olla eelarvamustega, mis tähendab, et need võivad teha otsuseid või anda soovitusi, mis ei ole õiglased kõigi suhtes.
- **Näiteks, kui otsingumootoril on eelarvamused, võib see kuvada mõnedele inimestele erinevaid tulemusi või teavet kui teistele. See ei ole õiglane, kuna kõigil peaks olema juurdepääs samadele objektiivsetele andmetele.**
- Oluline on olla teadlik, et algoritmid võivad olla eelarvamustega ning mitte alati usaldada kõike, mida nad pakuvad.
- Kui sulle tundub, et saadud teave võib olla ebaõiglane või ebapiisav, otsi teavet erinevatest allikatest, võrdle saadud infot, vajadusel küsi abi õpetajalt või täiskasvanult.
- Samuti on oluline teavitada tehnoloogia arendajaid, kui märkad eelarvamusi nende loodud algoritmides. Nii saame üheskoos muuta tehnoloogia õiglasemaks ja kõigile sobivaks.

Mida veel tehisintellekti kasutus võib mõjutada?

Tehisintellekti manipulatsioon

- Nagu te juba teate, aitab tehisintellekt meil teha erinevaid asju, näiteks luua nutikaid rakendusi ja isegi mänge. **Kuid kas teadsite, et mõned inimesed võivad proovida manipuleerida tehisintellekti, et saada soovitud tulemusi?**
- Manipuleerimine tähendab teiste suunamist või mõjutamist viisil, mis ei pruugi olla aus või õiglane. Näiteks mõned võivad üritada muuta otsingumootoreid või sotsiaalmeediaalgoritme nii, et teatud teavet kuvatakse rohkem või vähem, et mõjutada teiste arvamusi.
- **Oluline on olla kriitiline ja kontrollida teavet erinevatest allikatest. Ärge uskuge kõike, mida näete või loete internetis, eriti kui teave tundub liiga hea või kummaline.**
- Oluline on olla teadlik ja arukas kasutaja, et vältida enda manipuleerimist tehisintellekti poolt.

Tehisintellekti rakenduste turvalisus – milliseid muresid näete?

- **Virtuaalne assistent** – andmete privaatsus, eelarvamustega soovitused, valed vastused, sõltuvus (ei saa ilma enam hakkama), kurjategijatele ligipääs, sobimatu sisu esitamine.
- **Mäng** – lisaks eelnevale võimalik küberkiusamine, raha kulutamine.
- **Ülesannete tegemine tehisintellekti abil** – plagieerimine, võltsitud tööd, ei õpi ise (võimekuse varjamine), õppimises passiivsus ja motivatsiooni langus, valed vastused.

Aruta juhtumit A

- Laura sisestab tehisintellekti rakendusse kooli veebilehe URL-i ja palub sellel analüüsida lehekülge. Ta märkab, et tehisintellekt näitab, et lehel on turvarisk, mis võib võimaldada volitamata ligipääsu õpilaste andmetele. Rakendus näitab ka, et kooli veebilehel on mõned kohad, kus õpilaste isiklikud andmed on avalikult nähtavad.

- Mida Laura peaks edasi tegema?
- Kas sel viisil võib veebilehtede turvalisust kontrollida? Mis on riskid?
- Milline tegevus võiks olukorda paremaks muuta või millise tegevuse tagajärjel muutuks olukord veel halvemaks?

Aruta juhtumit B

- Mark kirjutab oma essee ja kasutab seejärel tehisintellekti õigekirja kontrollijat, et kontrollida oma tööd. Ta on pettunud, kui näeb, et tehisintellekt teeb palju vigu ning ei suuda tuvastada mitmeid tema poolt tehtud õigekirja vigu. Mõned vead jäid märkamata, samas kui teisi sõnu parandab tehisintellekt valesti.

- Miks tehisintellekt testi valeks parandab ja samal ajal ei tuvasta kõiki vigu?
- Miks Mark essee kirjutamiseks tehisintellekti kasutas, kui see oli tema enda ülesanne?
- Milliseid muresid võib tehisintellekti sellise kasutusega veel kaasneda?
- Milline tegevus võiks olukorda paremaks muuta või millise tegevuse tagajärjel muutuks olukord veel halvemaks?

Aruta juhtumit C

- Maria esitab tehisintellekti virtuaalsele õpetajale keerulise matemaatikaülesande. Ta ootab, et tehisintellekt aitab tal samm-sammult vastust leida ja tunnustab tema pingutusi. Kuid tehisintellekti vastus on üksnes tehniline ja emotsioonitu selgitus.

- Mida peaks tegema, et tehisintellekt aitaks õppida?
- Miks tehisintellekt ei vasta nagu inimene?
- Kas sellise tehisintellekti kasutusega võiks kaasneda veel muresid?
- Milline tegevus võiks olukorda paremaks muuta või millise tegevuse tagajärjel muutuks olukord veel halvemaks?

Kokkuvõte

- Mida täna õppisime?
- Loome ise 5 reeglit, mida järgida, kui kasutame tehisintellekti.
- Küsimused?

Tunnikava on koostanud Birgy Lorenz, küberkaitse vanemteadur Tallinna Tehnikaülikooli IT Kolledžis, augustis 2023 projekti Targalt internetis raames.

Targalt internetis