Tekoälyjärjestelmien haavoittuvuuksien testausalusta

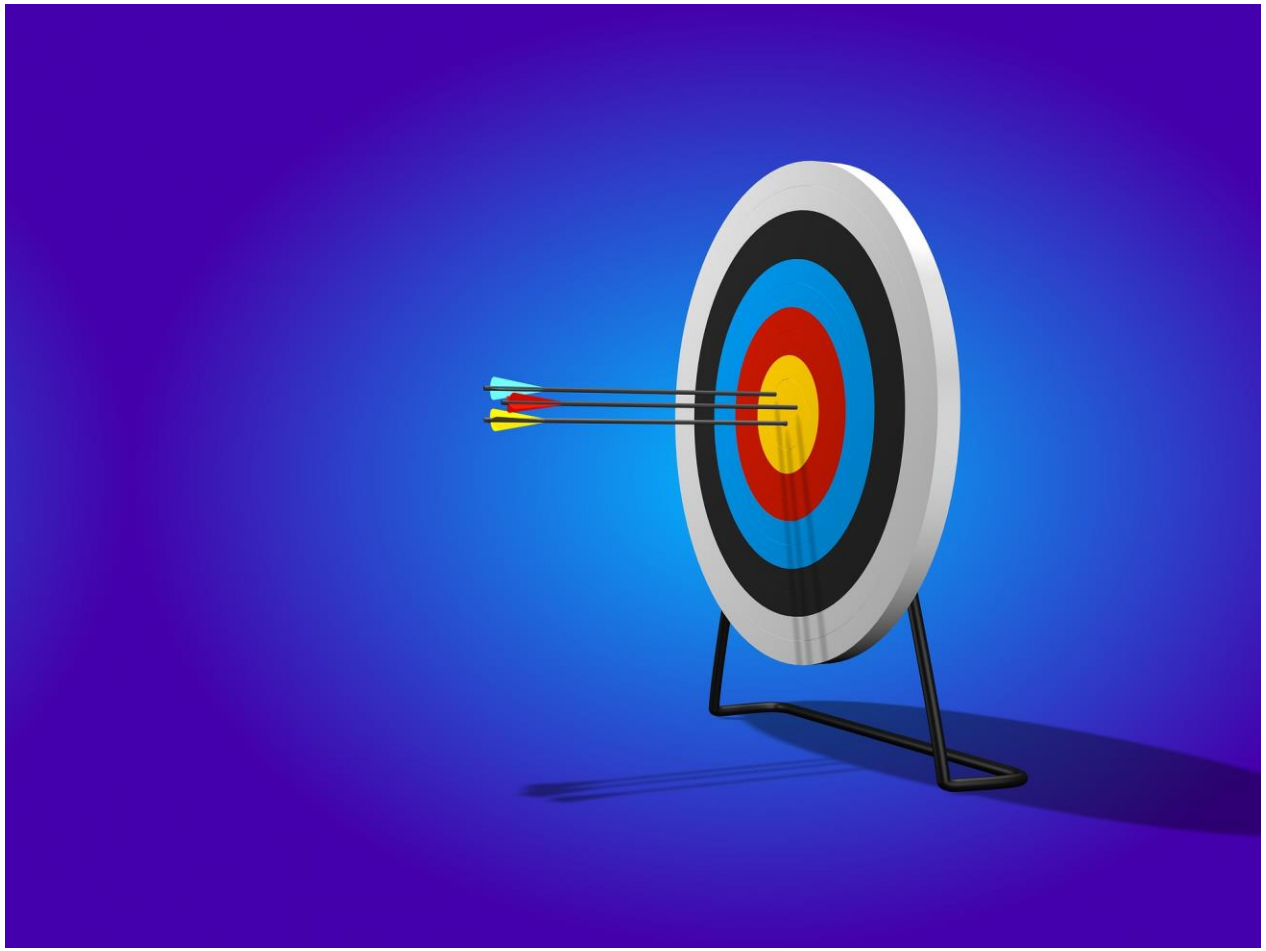
Matinen rahoitus: 158 960 euroa vuosille 2023-2024
1. vuoden tulokset

Kimmo Halunen
Kimmo.halunen@oulu.fi
Oulun yliopisto ja Maanpuolustuskorkeakoulu

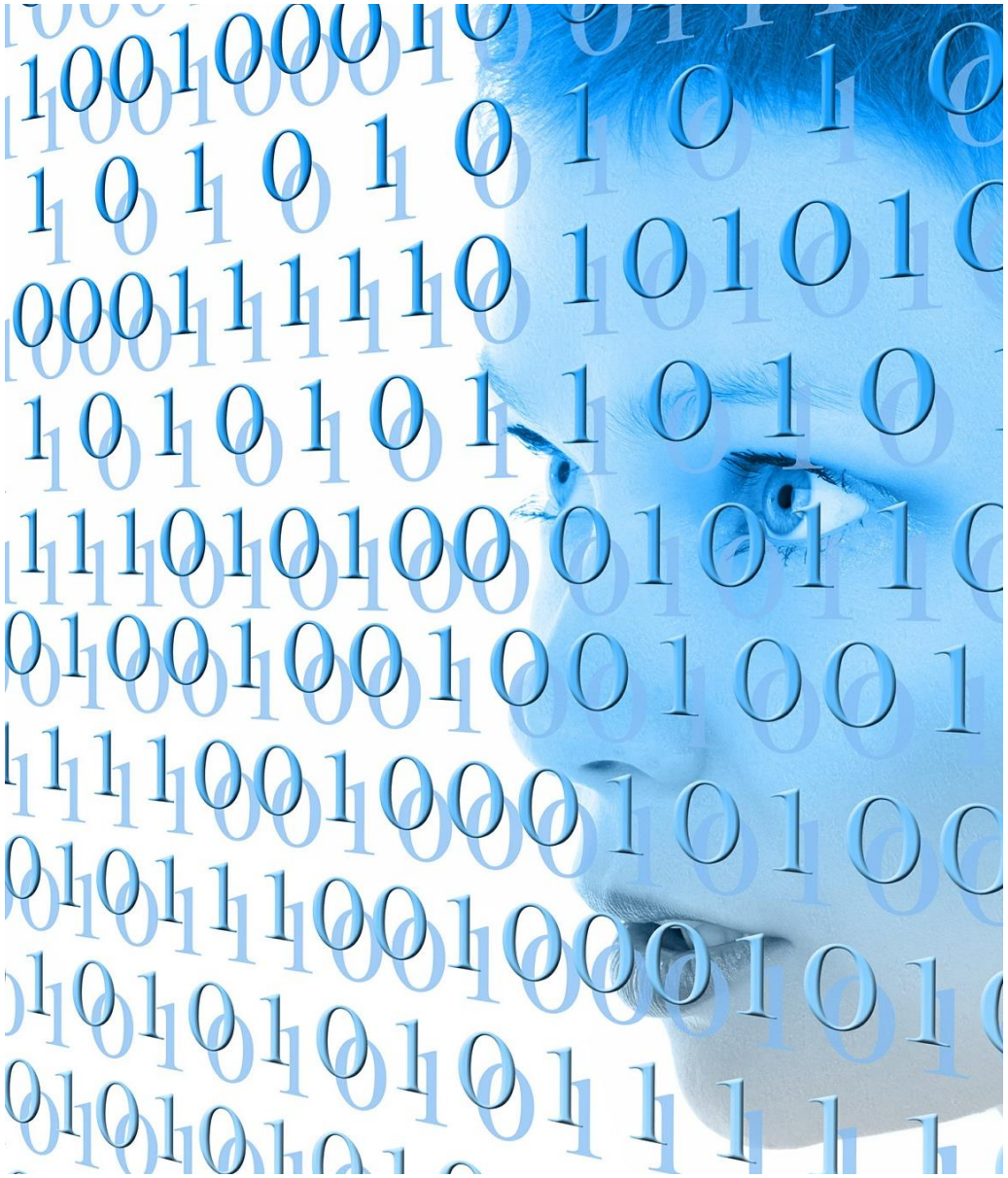
Mistä puhutaan?

- Tutkimuksen tavoite
- Tutkimuksen tausta
- Tämän hetkiset tulokset
- Tulevat toimenpiteet
- Yhteenveto
- Kysymyksiä?

Tutkimuksen tavoite



- Hankkeessa kehitetään tekoälyjärjestelmien haavoittuvuuksien testausalusta.
- Tavoitteena on tutkia alustan mahdollisuuksia erityisesti sensorifuusioon perustuvien tekoälymenetelmien harhauttamiseen ja haavoittuvuuksien löytämiseen.
- Hankkeessa tutkitaan myös mahdollisuuksia suojata järjestelmiä vaikuttamiselta.
- Testausjärjestelmää testataan yhdessä MPKK:n kanssa LAYKKA –alustan tekoälyjärjestelmien kanssa.
- Järjestelmä toteutetaan avoimen lähdekoodin alustana, mikä mahdollistaa järjestelmän jatkokehittämisen projektin jälkeen



Tutkimuksen tausta

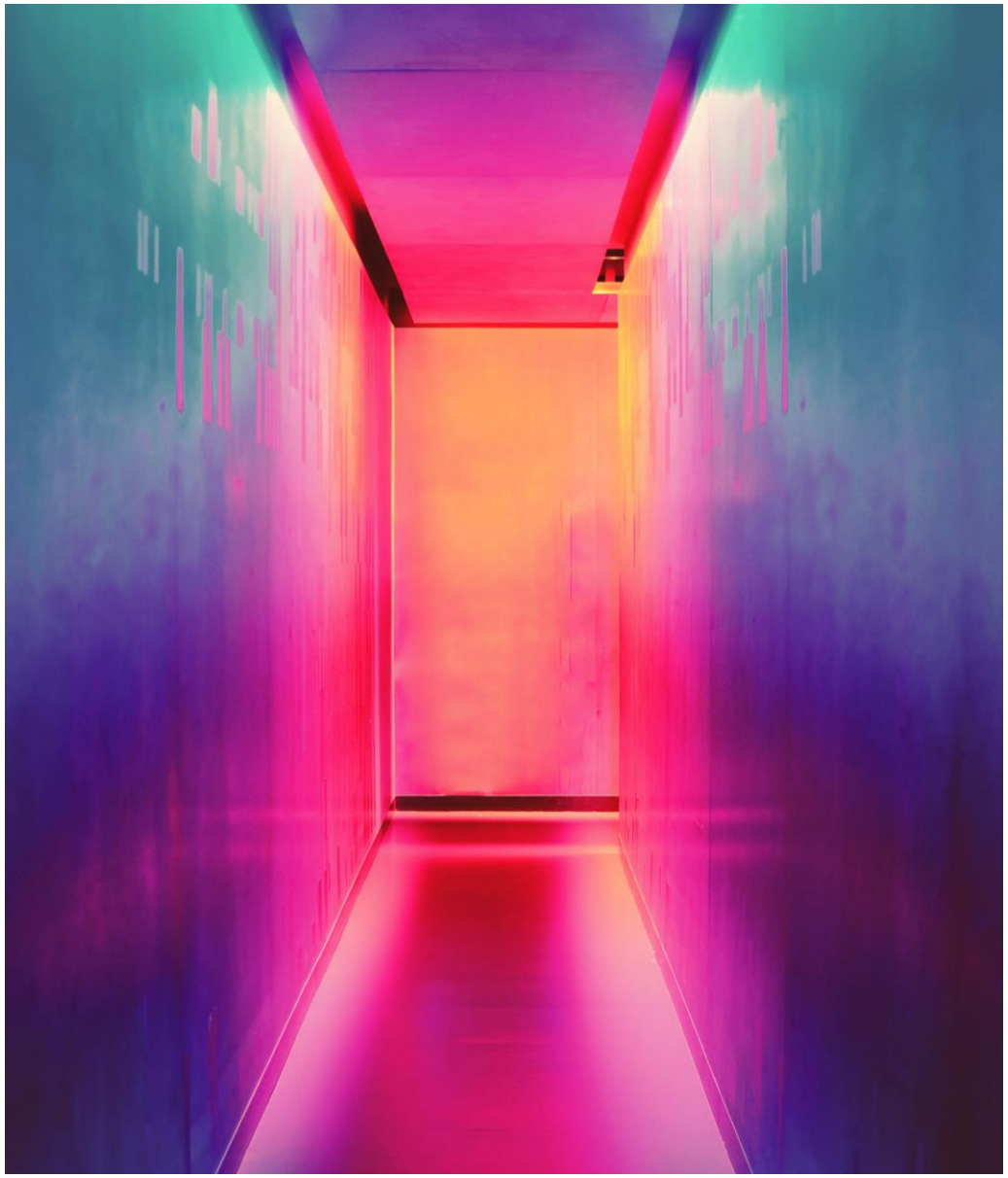
- Tekoälyjärjestelmät ovat yleistyneet yhteiskunnassamme valtavasti
- Monet Puolustusvoimien kannalta merkitykselliset järjestelmät käyttävät nyt tekoälyä ja koneoppimista itseohjautuvuuteen ja päätöksenteon tueksi
- Tekoälyjärjestelmien haavoittuvuuksien tuntemus on vähäistä
- Tekoälyjärjestelmien haavoittuvuuksien tutkimiseen ei ole samankaltaisia työkaluja kuin perinteisiin järjestelmiin
- Tällaisten menetelmien ja järjestelmien kehittäminen palvelee sekä Puolustusvoimia että laajemmin yhteiskuntaa

Tuloksia

- Tekoälyhaavoittuvuuksien taksonomia
  - 7 eri vaikuttavuuden tasoa
  - 1. taso on normaali toiminta ja 7. taso on hyökkääjän hallitsema vaikutus järjestelmään tai sen ulkopuolelle
- Erilaisia ominaisuuksia, joihin tekoälyhaavoittuvuudet voivat vaikuttaa on myös useita

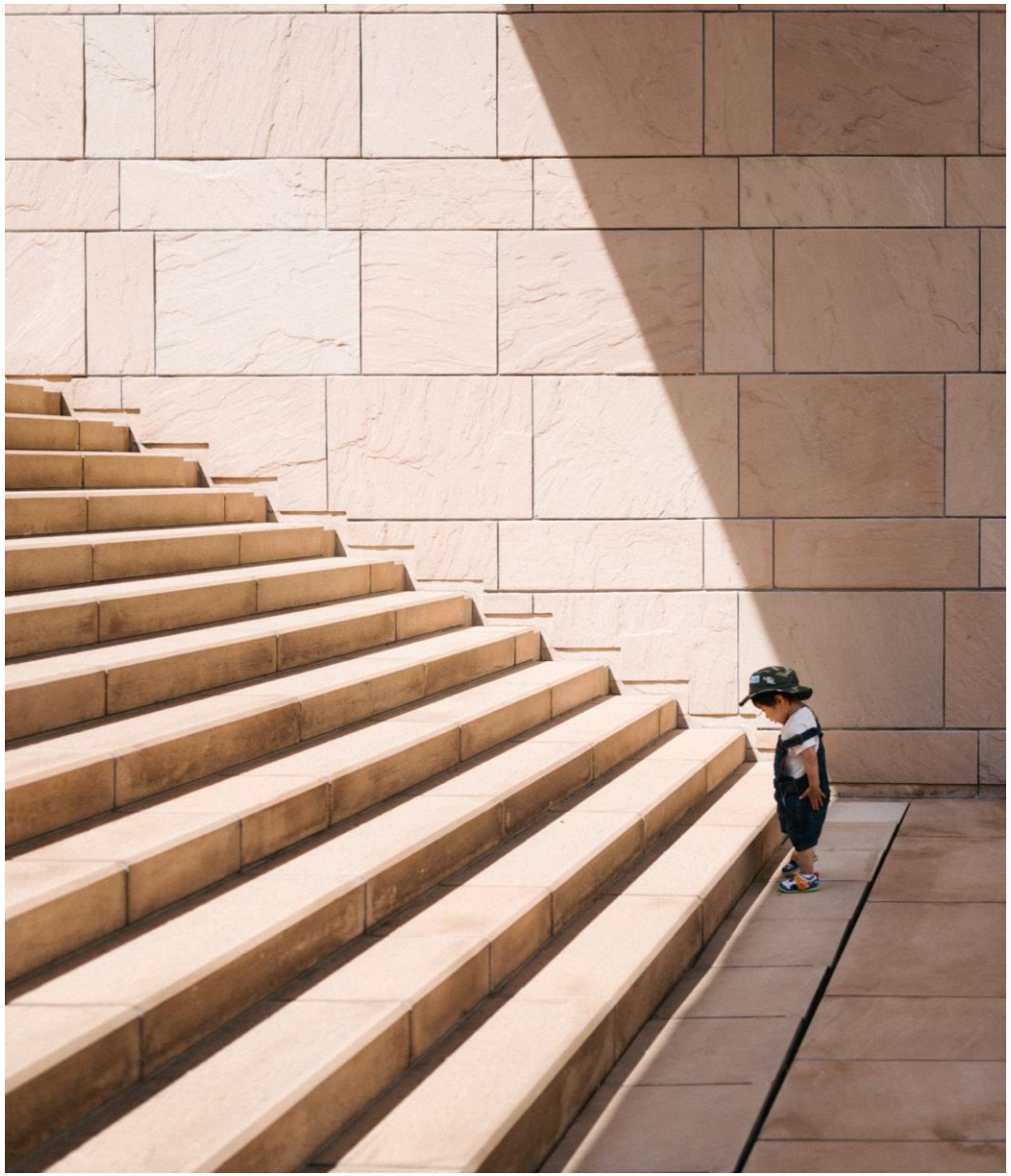
Tuloksia

- Tekoälyalustan kehitys
  - Aluksi vaikutti, että alustoja on useita
  - Alustavat tulokset osoittivat, että vain yksi (ART, Adversarial Robustness Toolbox) on käyttökelpoinen projektin tarpeisiin
  - Tätä testataan ja tarvittaessa jatkokehitetään
- Haavoittuvuuksien ilmeneminen tosielämässä
  - Eri tietokannoista (sekä tekoälyyn liittyvät että geneeriset) löytyy vain vähän raportteja todellisesti vaikuttavista tekoälyhaavoittuvuuksista
  - Tarvitaanko uusi tietokanta ja raportointimalli?



Tulevat toimenpiteet

- Alustan ja hyökkäysten testaus yhteistyössä MPKK:n kanssa
- Alustan jatkokehitys
- Suojautumismenetelmien tutkimus
  - Tätä on jo aloitettu erityisesti sensorifuusioon liittyen



Yhteenveto

- Tekoälyhaavoittuvuudet ovat sekä samanlaisia että erilaisia kuin perinteiset haavoittuvuudet
  - Tarve taksonomialle ja ehkä uusille tavoille raportoida haavoittuvuuksia
- Uutta alustaa ei tarvitse tehdä itse
  - Olemassa olevan alustan käyttöä tulee tehostaa ja hyödyntää tutkimuksessa
- Projektin 2. vuodella käytännön testausta ja sensorifuusion tutkimusta

Kysymyksiä?

Kiitos paljon mielenkiinnosta!


MAKE THIS
WORLD
BETTER