



Eesti Teadusagentuuri seisukoht suurte keelemudelite kasutamise osas teadusraha jagamisel

Veebruar 2024

Tehisintellektisüsteemid on Eesti teadussüsteemis kasutusele võetud ning jätkub uute rakendusala otsimine. Koos tehisintellekti laialdase kasutusele võtuga kerkivad üles uued küsimused selliste rakenduste kasutamise lubatavuse osas.

Käesoleva seisukohagaⁱ määratleb Eesti Teadusagentuur üldised piirid tehisintellekti rakenduste, eelkõige suurte keelemudelite kasutamisele teadusraha taotlemise ja taotluste hindamise kontekstis. Seisukoht ei puuduta muid tehisintellekti kasutamise ja võimalike ühiskondlike mõjudega seotud küsimusi, sh generatiivsete mudelite potentsiaalset väärkasutust teaduses. Eesti Teadusagentuur soosib Eesti teadlaste püüdlusi tehisintellekti meetodite ja rakenduste arendamisel eeldusel, et seda tehakse ühiskondlikes huvides ja eetilisi põhimõtteid järgides.

Arvestades tehisintellekti kasutamise potentsiaali, ei pea Eesti Teadusagentuur vajalikuks piirata selle kasutamist tegevustes, kus võib mõistlikult eeldada, et see teadlaste tööd lihtsustab, ilma et see kahjustaks ühiskonna usaldust teaduse suhtes. Samas arenevad tehisintellekti tehnoloogiad kiiresti ning avalduda võivad uued probleemid ja väljakutsed, mistõttu tuleb siinseid põhimõtteid uuendada või muuta.

Üldised põhimõtted

- 1. Vastutus:** vastutus tehisintellekti rakenduse kasutamise eest lasub seda kasutaval teadlasel, kes peab järgima üldisi teaduseetilisi põhimõtteid, austama inimeste õigusi ning vältima nende kahjustamist.
- 2. Läbipaistvus ja selgitatavus:** teadlane annab alati teada tehisintellekti kasutamisest ning täpsustab, millisteks tegevusteks on millist rakendust kasutatud.
- 3. Inimese järelevalve:** kõik olulised teadustöö etapid peavad olema inimese poolt kontrollitud ja kinnitatud, isegi kui selleks on kasutatud tehisintellekti abi. Minimaalselt hõlmab see järgnevaid tegevusi: hüpoteeside sõnastamine, teadustöö eesmärkide sõnastamine, puuduva teadmise tuvastamine, andmete analüüs ja tõlgendus, käsikirjade kirjutamine ning retsenseerimine.

Tehisintellekti kasutamine Eesti Teadusagentuuris

- 1. Taotluste koostamine:** Eesti Teadusagentuur lubab tehisintellekti, eelkõige suurte keelemudelite kasutamist taotluste ja muude materjalide koostamisel, eeldusel, et sellisest kasutamisest on teavitatud ja kasutamist on piisavalt selgitatud.
- 2. Retensioonide koostamine:** retsensioonide ja hinnangute koostamisel ei tohi kasutada generatiivse tehisintellekti abi. Keeld tuleneb eelkõige vajadusest kaitsta hindamise konfidentsiaalsust ja vältida konfidentsiaalse teabe keelemudelite treeningmaterjali sattumist.
- 3. Hindamisprotsessi toetamine:** Eesti Teadusagentuur ei kasuta praegu tehisintellekti abi hindamisprotsessi üheski etapis, kuid kaalub võimalusi hindamisprotsessi tõhustamiseks tehisintellekti tööriistade abil, ilma et see kahjustaks hindamise usaldusväärsust. Nii taotlemisel kui ka hindamisel jääb iga protsessi etapi eest vastutajaks inimene ning kogu protsess peab olema läbipaistev.

ⁱ ETAGi seisukoha koostamisel on eeskujuks võetud: Bockting, C. L., van Dis, E. A. M., van Rooij, R., Zuidema, W., & Bollen, J. (2023). Living guidelines for generative AI - why scientists must oversee its use. *Nature*, 622(7984), 693–696. <https://doi.org/10.1038/d41586-023-03266-1>

Deutsche Forschungsgemeinschaft (2023). *Statement by the Executive Committee of the Deutsche Forschungsgemeinschaft on the Influence of Generative Models of Text and Image Creation on Science and the Humanities and on the DFG's Funding Activities*. <https://www.dfg.de/resource/blob/289676/89c03e7a7a8a024093602995974832f9/230921-statement-executive-committee-ki-ai-data.pdf>

European Commission, Directorate-General for Communications Networks, Content and Technology, (2019). *Ethics guidelines for trustworthy AI*, Publications Office. <https://data.europa.eu/doi/10.2759/346720>