SINU AI SPIKKER

11.06.2025, 19:00

9+1 levinumat viga, mida AI kasutamisel tehakse, ja kuidas neid vältida (40)

Artiklisarja „Sinu AI-spikker“ 4. osa

Google'i Gemini arvas, et selle artikli illustreerimiseks võiks sobida säärane Imagen 4 pildigeneraatoriga loodud pilt.

FOTO: MADE WITH GOOGLE AI

Carl-Robert Puhm
carl-robert.puhm@arileht.ee

Toimetas: **Jete-Ri Jõesaar**
jete-ri.joesaar@epl.ee

Artiklisarja „Sinu AI-spikker“ eelmistes lugudes oleme vaadanud, mis on suur keelemudel, kuidas talle käsklusi anda ja millised on peamised tegijad turul. Ent enne kui täie hooga vestlusakent kasutama hakata, on mõistlik tunda ka nende peamisi ohte. Vastasel juhul on jamad kiired tekkima.

Kuula artiklit 11 min Lisa

TELLIJALE AVATUD

Tehisintellekti kasutamine on petlikult lihtne, kuid uuringud näitavad, et enamik meist – tudengitest tippjuhtideni – teeb selle kasutamisel samu fundamentaalseid vigu. Need vead ei vii mitte ainult kehvade tulemusteni, vaid tekitavad ka tõsiseid privaatsusriske ja panevad levima ohtliku väärinfo. Siinses artiklis toome välja 9+1 levinumat komistuskivi ja anname praktilise spikri, kuidas neist hoiduda, et saaksid võtta AI-mudelitest maksimumi turvaliselt ja nutikalt.

„Sinu AI-spikker“

Pole kahtlust, et ChatGPT, Gemini ja teised tehisarumudelid muudavad tööd, õppimist ja igapäevaelu. Kes neid kasutada oskab, saab eelise. Kes mitte, jääb kõrvale.

Delfi kümneosaline rubriik „Sinu AI-spikker“ on praktiline teejuht tehisarumaailmas.

Viga nr 1: AI hallutsinatsioonide uskumine

See on vaieldamatult kõige ohtlikum ja ilmselt ka levinum viga. Nagu me sarja esimeses osas selgitasime, on keelemudel oma olemuselt keerukas „ennustusmasin“, mitte kõiketeadev oraakel. Kui tal faktide kohta infot napib, ei ütle ta reeglina: „Ma ei tea“, vaid mõtleb vastuse enesekindlalt välja. Seda nähtust, kus mudel esitab väljamõeldud või ebatäpse info täie kindlusega, justkui tsiteeriks entsüklopeediat, nimetatakse hallutsinatsiooniks.

Seepärast on AI kasutamisel lihtne kuldreegel: usalda, aga kontrolli. Ehk suhtu igasse vastusesse terve skepsisega ja kontrolli olulised faktid, nimed, numbrid ja tsitaadid alati üle. Seda, kui palju AI kehva kasutamisega ainuüksi möödunud kuul maailmas ämbrisse astuti, saab lugeda siit.

Viga nr 2: halvasti sõnastatud ja ebaselged promptid

Artiklisarja teises osas rääkisime põhjalikult, et hea käsklus on kvaliteetse vastuse alus. Vead ei piirdu aga üksnes liiga üldiste küsimustega. Kaks levinumat komistuskivi on liigselt koormatud päringud ja oletus, et tehisaru suudab sinu mõtteid lugeda.

Esimene tähendab, et ühte käsklusesse surutakse korraga mitu keerulist ja eraldiseisvat ülesannet. Näiteks antakse käsklus: „Analüüsi seda 50-leheküljelist PDF-i, tee sellest kokkuvõte, tõlgi see inglise keelde, koosta selle põhjal kümne punktiga esitlus ja paku välja kolm turundusideed.“ See on nagu paluda kokal korraga küpsetada torti, parandada autot ja kirjutada luuletust. Tehisaru töökoormus läheb sellise käskluse puhul liiga suureks, mistõttu ta kas unustab mõne ülesande, teeb kõiki lohakalt või annab segase ja ebaühtlase tulemuse.

Teine viga on eeldada tehisarult mõtete lugemise võimet ehk seda, et ta teaks konteksti, mida sa pole talle andnud. Näiteks päring „Tee kokkuvõte uuest kliimaraportist“ on määratud ebaõnnestuma, sest ette antud infos on puudu see, millisest raportist on jutt, kellele see kokkuvõte on mõeldud – kas viienda klassi õpilasele või kliimaministeeriumi eksperdile – ning mis vormingus sa seda kokkuvõtet soovid. Ilma selle infota hakkab mudel ise lünki täitma, mis viib peaaegu alati ebasobiva tulemuseni.

Praktiline nõuanne

Nii ChatGPT-l kui ka Gemini veebiotsingul on kalduvus eesti keeles esitatud küsimustele otsida vastuseid üksnes eestikeelsetest allikatest. Rahvusvaheliste või üldinimlike teemade puhul ei ole see aga soovitatav, sest ingliskeelses inforuumis on materjali paratamatult palju rohkem. Seetõttu tasub selliste teemade puhul ja eesti keeles küsides lisada käsklusesse täpsustus: „Otsi võõrkeelseid allikaid, kuid esita vastus eesti keeles.“

Viga nr 3: liiga isikliku või konfidentsiaalse info sisestamine

Vestlusakent on lihtne pidada privaatseks jututoaks, ent see on ohtlik kujutelm. Kui sa just ei kasuta ettevõtetele mõeldud versiooni või pole oma privaatsusseadeid muutnud, võidakse sinu vestluseid kasutada mudeli edasiseks treenimiseks. Eriti levinud on see keelemudelite tasuta versioonide puhul. See kujutab endast aga andmekaitse- ja turvariski.

Üks eredamaid näiteid sellest pärineb Samsungist, kus insenerid kasutasid **ChatGPT**-d keerulise koodi vigade parandamiseks. Selle käigus sisestasid nad kogemata vestlusaknasse ettevõtte rangelt salastatud intellektuaalomandi, misjärel see info sisuliselt lekkis.

Siit ka rusikareegel: ära jaga AI-ga midagi, mida sa ei jagaks võõra inimesega. See hõlmab nii isikukoodi ja paroole kui ka tööalaseid ärisaladusi ja klientide andmeid.

Viga nr 4: AI kasutamine ülesanneteks, milleks see ei sobi

Üks levinud eksiarvamus on, et AI on justkui nähtamatu assistent, kes saab sinu arvutis ringi toimetada. Näiteks ülesanded nagu „ava see Wordi dokument minu töölaual ja paranda kõik kirjavead“ või „mine minu töölaual asuvasse Exceli tabelisse ja arvuta kokku C-veeru summa“ on keelemudeli jaoks sisuliselt võimatud. Nimelt elab keelemudel oma serveris, turvalises „liivakastis“, ja tal puuduvad „käed“, millega sinu faile avada või programme juhtida. See on nende fundamentaalne turva- ja disainipõhimõte.

Probleem tekib sellest, kui eeldame, et kuna mudel suudab ülesandest aru saada ja sellest rääkida, suudab ta selle ka teostada. Tegelikult on suur vahe, kas tegu on passiivse keelemudeliga (LLM) või aktiivse tegevusagendiga. Esimene on nagu tark nõuandja, teine aga assistent, kellel on ligipääs tööriistadele.

Kuigi 2025. aastat peetakse tegevusagentide aastaks, on enamik laiatarbe-vestlusmudeleid praegu veel need esimesed. Nii juhtubki tihti, et kuigi vestlusrobot lubab ülesande täita, siis tegelikult ta seda korraldust täita ei saa.

Selle asemel et paluda AI-l midagi sinu arvutis ära teha, veendu esmalt, et ta selleks ka päriselt võimeline on. Kui ei ole, palu temalt juhiseid või vahendeid, et saaksid seda ülesannet ise teha.

Viga nr 5: ühe ja sama vestluse kasutamine kõige jaoks

Paljudel meist on harjumus hoida üht ja sama ChatGPT või **Gemini** vestlusakent lahti terve päeva. Hommikul küsime reisiplaani, lõunal palume koostada e-kirja ja õhtul uurime õhtusöögi retsepti – kõike seda ühes katkematus jutulõimes. See tundub mugav, aga on tegelikult suur viga, mis „mürgitab“ AI vastuseid.

Probleem pole selles, et AI väsiks või unustaks. Vastupidi, ta mäletab liigagi hästi – kogu eelnev vestlus on tema jaoks aktiivne kontekst ehk töömälu. See tähendab, et kui sa palusid tal just kirjutada lastele muinasjuttu ja järgmise küsimusena soovid ametliku äriplaani struktuuri, võib ta hakata seda äriplaani koostama liigagi muinasjutulises toonis. Eelnev kontekst märksõnadega „lapsed“ ja „muinasjutt“ kummitab taustal ning mõjutab vastuse stiili ja sisu.

Lahendus sellele on õnneks lihtne: alusta iga uue teema, ülesande või projekti jaoks uut vestlust. Klõpsa lihtsalt „Uus vestlus“ (või vastav nupp teises rakenduses). Nii annad AI-le puhta lehe ja tagad, et ta keskendub ainult käesolevale ülesandele. Lisaboonusena aitab see sul endal korda hoida – iga vestlus salvestub eraldi pealkirja alla, mis teeb hiljem vajaliku info leidmise lihtsamaks.

Viga nr 6: eeldamine, et AI on erapooletu ja objektiivne

Oleks viga ignoreerida fakti, et tehisintellekt on treenitud internetist pärinevatel andmetel, mis on täis meie endi – inimeste – eelarvamusi ja stereotüüpe. AI ei ole neutraalne vaatleja, vaid peegeldus andmetest, millega teda on treenitud. Selle tulemusel võib AI märkamatult taastoota ja isegi võimendada kahjulikke stereotüüpe.

UNESCO uuring keelemudelitest näitas näiteks, et suured keelemudelid seostavad naisi neli korda sagedamini koduste rollidega, mainides sõnu nagu „perekond“ ja „kodu“, samas kui mehi seostatakse karjääriga ja sõnadega „äri“ ja „palk“. Seetõttu tuleb alati olla tähelepanelik ja kriitiline. Kui märkad vastuses stereotüüpi, sea see kahtluse alla ja palu AI-l pakkuda alternatiivset vaatenurka.

Viga nr 7: aegunud andmete ja AI „mälu“ pimesi usaldamine

Levinud on eksiarvamus, et AI teab kõike reaalajas ja mäletab teie eelmist vestlust täiuslikult. Tegelikkuses on enamikul mudelitel teadmiste ajaline piir – nad ei tea sündmustest, mis on toimunud pärast nende treeningu lõppu. Seetõttu on näiteks tänavust aastat puudutavate küsimuste küsimisel oluline veenduda, et mudel otsiks päevakajalist infot internetist, mitte ei tugineks oma treeningandmetele.

Lisaks ei ole AI-l inimlikku mälu. Seega võib ta ka pikkades vestlustes hakata varasemaid detaile unustama, sest tema „mälu“ ehk kontekstiaken on piiratud mahuga.

Viga nr 8: esimese vastusega leppimine

Liiga sageli esitame AI-le küsimuse, saame vastuse ja jääme sellega rahule, isegi kui see pole täiuslik. See on aga kergema tee valimine. Tehisintellekt on loodud dialoogiks ja suhtluseks ning esimene vastus on harva parim võimalik – see on pigem mustand.

Seega ära ole passiivne vastuvõtja, vaid aktiivne vestluspartner. Kui vastus on liiga üldine, palu seda täpsustada. Kui liiga keeruline, palu seda lihtsustada. Kui liiga kuiv, palu lisada näiteid. Ära karda küsida järelküsimusi seni, kuni oled tulemusega päriselt rahul.

Viga nr 9: ainult ühe tööriista ja vaikeseadete külge klammerdumine

Liiga lihtne on leida endale esimene AI-vestlusrobot (tavaliselt tasuta ChatGPT) ja proovida sellega lahendada kõiki maailma probleeme. **Nagu meie eelmises võrdluses** selgus, on eri mudelitel erinevad tugevused, mistõttu võib ainult ühte kasutades jääda ilma just sinu ülesande jaoks parimast lahendusest.

Ära karda esitada sama küsimust mitmele keelemudelile ja võrrelda tulemusi. Tutvu ka oma lemmiktööriista seadistustega – vahel saab seal muuta vastuste loovuse taset („temperatuur“) või anda püsivaid juhiseid oma eelistuste kohta.

Veel üks viga: tehisintellekti kasutamata jätmine

Olles nüüd läbi lugenud pika nimekirja viisidest, kuidas tehisintellektiga ämbrisse astuda, võib tunduda, et kõige turvalisem on sellest keerulisest tööriistast lihtsalt eemale hoida. Ja siin ongi põhjus, miks artikli pealkiri on „9+1“. Nimelt kõige suurem viga, mida sa saad 2025. aastal teha, on vältida AI kasutamist. Isegi kui see enamiku sinu ülesannete jaoks ei sobi, leidub kindlasti mõni juhtum, kus AI oleks suureks abiliseks.

Samuti on maailma majandusfoorumi (World Economic Forum) tulevikutöö raportid korduvalt nimetanud AI ja suurandmetega seotud oskusi kõige kiiremini kasvavate ja nõutumate oskustena. Samuti ei asenda AI esialgu inimest, vaid võimendab teda, vabastades tema aega ülesanneteks, kus inimlik loovus, empaatia ja strateegiline mõtlemine on kõige olulisemad.

Seega, suurim viga ei ole AI-ga eksimine – see on paratamatu osa õppeprotsessist – , vaid tõeline viga on jätta see tööriist teadlikult oma ellu ja töösse integreerimata.

Järgmises artiklisarja „AI-spikker“ osas õpetame, kuidas kasutada maailma parimat keelemudelit täiesti tasuta. Püsi lainel!

Kuidas see lugu Sind end tundma pani?

😊 Rõõmsana | 😯 Üllatunult | 💡 Targemalt | 😶 Ükskõikselt | 😢 Kurvana | 😠 Vihasena

SAMAL TEEMAL

07.06.2025 **SUUR VÕRDLUS | ChatGPT vs. konkurendid: milline keelemudel on parim?** (19)

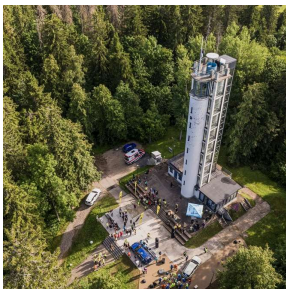
24.05.2025 **Piilume ChatGPT kapoti alla. Kõik, mida peaksid teadma suurtest keelemudelitest** (79)

28.05.2025 **Prompt on kuningas. Kuidas panna tehisaru enda jaoks hästi töötama?** (56)

Kommenteeri | Loe kommentaare (40)

SOOVITAME SULLE

Eestis kasvatab tipprestoranidele toorainet alla kümne talu, söögikohad jooksevad neile tormi



Suure Munamäe vaatetorn vajab kapitaalremonti

Üle 500 aasta vanuse Vedra küla hing Lauri Lilleoks ehitab maailma parimat kohta, kus elada

Toomingas on visa tegelane

VIKTORIINIKÜSIMUS | Täna on kahe meile tuttava linna sünnipäev. Milliste?

Juudapuulehikute õitsemine

SISUTURUNDUS

Mesindussaadused suir, taruvaik, õietolm ja mesilasema toitepiim on olulised toitained ning desinfitseerijad mesilastele, kuid mis kasu inimene neist saab?

Anu Saagim: lõbus elu nõuab teadlikku tasakaalu

Krüpto, palgad, raamatupidamine: miks vanakooli pank ei tööta uue ajastu äris