

TARTU ÜLIKOOL
ÕIGUSTEADUSKOND
Eraõiguse osakond

Hegle Pärna

**AUTORIÕIGUSEGA KAITSTUD TEOSE GENERATIIVSE TEHISINTELLEKTI
TREENING- JA SISEDANDMETENA KASUTAMISE ÕIGUSPÄRASUS**

Magistritöö

Juhendaja
Mag. iur. Gea Lepik
Kaasjuhendaja
Mag. iur. Henrik Trasberg

Tartu
2024

SISUKORD

SISSEJUHATUS	3
1. AUTORIÕIGUSEGA KAITSTUD TEOSE KASUTAMINE LÄHTUVALT GENERATIIVSE TEHISINTELLEKTI TEHNOLOOGILISEST OLEMUSEST	9
1.1. Generatiivse tehisintellekti tehnoloogiline olemus	9
1.2. Treening- ja sisendandmed kui autoriõigusega kaitstud teosed	13
1.3. Autori ainuõigused kaitstud teoste treening- ja sisendandmetena kasutamisel	16
1.3.1. Autori varaliste õiguste kasutamine	16
1.3.1.1. Treening- ja sisendandmete reprodutseerimine ning töötlemine	16
1.3.1.2. Treeningandmete üldsusele kättesaadavaks tegemine	21
1.3.2. Autori isiklike õiguste kasutamine	24
1.3.2.1. Treening- ja sisendandmete autorsus	24
1.3.2.2. Au ja väärkuse kaitse treeningandmete kasutamisel	27
2. TREENING- JA SISENDANDMETE KASUTAMINE KEHTIVA NING KAVANDATAVA ÕIGUSE KOHASELT	30
2.1. Treening- ja sisendandmete kasutamine vaba kasutuse erandite alusel	30
2.1.1. Treening- ja sisendandmete ajutine või juhuslik reprodutseerimine	30
2.1.2. Treeningandmete reprodutseerimine teadustöös teksti- ja andmekaeve eesmärkidel	34
2.1.3. Treeningandmete reprodutseerimine väljaspool teadustööd teksti- ja andmekaeve eesmärkidel	40
2.1.4. Sisendandmete reprodutseerimine isikliku kasutamise eesmärkidel	45
2.2. Treening- ja sisendandmete kasutamise koostõla kolmeastmelise testiga	46
2.3. Tehisintellekti määrusest tulenevad nõuded treeningandmete kasutamisele	50
KOKKUVÕTE	54
THE LEGALITY OF USING A COPYRIGHTED WORK AS TRAINING AND INPUT DATA IN GENERATIVE ARTIFICIAL INTELLIGENCE. Abstract	58
KASUTATUD LÜHENDID	63
KASUTATUD ALLIKAD	64

SISSEJUHATUS

2023. aasta turu-uuringute järgi kasutavad Eesti ettevõtted viimastel aastatel järjest rohkem tehisintellekti tehnoloogiaid.¹ Samuti näitavad turuanalüüsid, et tehisintellekti tüüpidest moodustab generatiivne ehk iseseisvalt uut sisu luua suutev tehisintellekt 2025. aastaks 30% kogu tehisintellekti globaalse turu osakaalust.² Tehisaru hoogsast arengust ja innovaatsilisusest tingituna on hakatud üha enam käsitlema tehisintellekti, sh generatiivse tehisintellekti ning selle arendamisega seonduvaid autoriõiguslikke küsimusi. Arutelusid on tekitanud küsimused alates tehisintellekti autorina määratlemisest kuni tehisintellekti loodud väljundandmete autoriõigusliku kaitse andmiseni.³ Mõnevõrra vähem on seni käsitletud teemasid, mis on seotud autoriõigusega kaitstud teoste kasutamisega generatiivse tehisintellekti treening- ning sisendandmetena.⁴

Generatiivne tehisintellekt (edaspidi ka mudel) ning selles kasutatavate treening- ja sisendandmete autoriõiguste rikkumise küsimus kerkis aktuaalsemalt esile pärast 2022. aasta lõpu võimekate kõne- ja keelemudelite valmimist ning laiemasse kasutusse jõudmist.⁵ Läbimurre mudelite võimekuses tekkis suures osas tänu sellele, et mudelite arendamisel oli uudsetest tehnilistest omadustest tulenevalt võimalik varasemast veelgi tõhusamalt läbi töötada internetis leiduvaid andmeid, sh autoriõigusega kaitstud teoseid. See on tõstatanud küsimuse autoriõiguse võimalikust rikkumisest, mille tõttu algatati 2023. aastal ka mitmeid kohtuvaidlusi, eelkõige Ameerika Ühendriikides.⁶ Generatiivset tehisintellekti on võimalik pea

¹ Eesti Statistikaamet. Tehisintellekti tehnoloogiate kasutamine ettevõtetes on tõusutrendis. 15.09.2023. Kättesaadav: <https://www.stat.ee/et/uudised/tehisintellekti-tehnoloogiate-kasutamine-ettevotetes-tousutrendis> (17.01.2024). Vt lisaks tehisintellekti, sh generatiivse tehisintellekti legaldefiniitsiooni kohta peatükk 1.1. ja viide 27.

² AlphaSense. The Future of Generative AI: An Analysis of the Leaders, Opportunities and Threats. 01/2023. Kättesaadav: <https://www.alpha-sense.com/resource/report/the-future-of-generative-ai-an-analysis-of-the-leaders-opportunities-and-threats/> (17.01.2024). Tulenevalt generatiivse tehisintellekti kasvunumbritest, keskendub käesolev magistritöö generatiivse tehisintellektiga seotud autoriõiguslikele probleemidele.

³ Vt näiteks tehisintellekti arendamise ja autoriõiguste kuuluvuse küsimuse kohta Margoni, T. Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI? CREATE Working Paper 12/2018.

⁴ Generatiivse tehisintellekti puhul on oluline eristada üldisi mudeli treenimiseks ja loomiseks kasutatavaid andmeid ehk treeningandmeid (*training data*) ning mudeli kasutaja käskluse (*prompt*) kaudu sisestatud andmeid ehk sisendandmeid (*input data*). Vt ka käesoleva töö peatükk 1.1.

⁵ 2023. aastal jõudis laiemasse kasutusse näiteks OpenAI arendatud ChatGPT keelemudel. Ka Eestis on hakatud viimastel aastatel arendama erinevaid tehisintellekti tehnoloogiaid, näiteks avaliku sektori virtuaalassistenti Bürokratt. Selle kohta vt lähemalt Vilbaste, K. Keeletehnoloogia ja Bürokratt. Õiguskeel 2/2023; Majandus- ja Kommunikatsiooniministeerium jt. Tehisintellekti tegevuskava 2024–2026. 2024, lk 8.

⁶ 2023. aastast pärinevatest kohtuvaidlustest on markantsemad nt Sarah Silverman, Christopher Golden, Richard Kadrey v. OpenAI ja OpenAI v. New York Times kohtuasjad. Avaldatud kohtuasjade järgi on Ameerika Ühendriikides autoriõiguse võimaliku rikkumisega seoses käesoleva töö kirjutamise hetkel menetluses 20 kohtuasja mudeli arendajate vastu.

igaihel kasutada abilisena nii tööülesannete täitmisel kui ka meelelahutuseks. Lihtsustatult öeldes võimaldab generatiivne tehistaip paljudel inimestel vähese vaevaga luua erinevaid pilte, tekste ja tasapisi ka videoid. Selline võimalus on aga tekkinud seetõttu, et mudelid on arendatud miljardite internetis leiduvate materjalide abil. Mudelite levikust tingitud vähenenud nõudluse tagajärjel võivad loomevaldkondade esindajad jääda ilma oma sissetulekuallikatest, kuigi nende loomingut jäljendavat tehnoloogiat ei oleks olnud võimalik ilma nende teosteta arendada.⁷ Juhul, kui mudelite arendamine kujutab endast autoriõiguste rikkumist, s.o treening- ja sisendandmetena kasutatakse autoriõigusega kaitstud teoseid ning nende kasutamine on õigusvastane, siis võib see tähendada mudelid arendavate juriidiliste isikute jaoks ulatusliku majandusliku kahju tekkimist, kui autorid hakkavad oma ainuõigusi teostama.⁸ Eeltoodud põhjustel on käesolevas magistritöös lahendatavad küsimused olulise ühiskondliku mõjuga, sh nii tehnoloogia arengu kui ka kultuurivaldkonna seisukohalt.

Tehisintellekti etteaimatavast tõusulainest tulenevalt esitas Euroopa Komisjon juba 2021. aastal ettepaneku Euroopa Liidu harmoniseeritud tehisintellekti käsitleva õigusakti loomiseks (edaspidi ka tehisintellekti määrus).⁹ Tehisintellekti määruks üheks laiemaks eesmärgiks on luua stiimul, mis aitaks kaasa õiguste omajate ning tehisintellekti arendajate õiguste ja huvide kaitsele, kuid samas säilitada innovatsiooni soosiv keskkond. Euroopa Komisjon on juba käesoleva sajandi alguses rõhutanud, et intellektuaalse omandi õiguskaitse on innovatsiooni jaoks *sine qua non*, mistõttu on oluline leida tasakaal intellektuaalse omandi õiguste ja innovatsiooni vahel.¹⁰ Käesoleva töö kirjutamise hetkel on selgusetu, kas olemasolev ja kavandatav õigusraamistik on piisavad selleks, et tagada ühest küljest autorite huvide ja õiguste kaitse, kuid teisalt aidata kaasa ka tehnoloogia arengule ning innovatsioonile.

⁷ Kõnealune küsimus on tekitanud ühiskonda mitmeid vastasseise erinevate teoste autorite ja mudelite arendajate vahel. Näiteks toimus 2023. aastal pea viis kuud kestnud seriaali- ja filmistsenaristide streik (tuntud ka kui „2023 Writers Guild of America Strike“) muuhulgas nõude tõttu, et tehisintellekti kasutamine oleks rohkem piiratud. Vt selle kohta näiteks Anguiano, D. Hollywood writers strike ends studio deal. – The Guardian, 27.09.2023. Kättesaadav: <https://www.theguardian.com/culture/2023/sep/26/hollywood-writers-strike-ends-studio-deal> (08.12.2023).

⁸ Näiteks on tehnoloogiaettevõtte OpenAI, Inc. Ühendkuningriigi parlamendile esitatud avalduses märkinud, et võimekad mudelid ei saaks ilma autoriõigusega kaitstud teoste kasutamisetä eksisteerida. Vt Reisner, Alex. The Flaw That Could Ruin Generative AI. – The Atlantic, 11.01.2024. Kättesaadav: <https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/> (08.03.2024).

⁹ Euroopa Komisjon. Ettepanek: Euroopa Parlamendi ja nõukogu määrus, millega nähakse ette tehisintellekti käsitlevad ühtlustatud õigusnormid (tehisintellekti käsitlev õigusakt) ja muudetakse teatavaid liidu õigusakte. 21.04.2021. Kättesaadav: <https://eur-lex.europa.eu/legal-content/EN-ET/TXT/?from=EN&uri=CELEX%3A52021PC0206> (06.09.2023).

¹⁰ Commission of the European Communities. Putting knowledge into practice: A broad-based innovation strategy for the EU. 13.9.2006, lk 6. Kättesaadav: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52006DC0502> (07.01.2024).

Autoriõigusega kaitstud teoste kasutamisel mudeli treening- ja sisendandmetena võidakse teostada nii autorile kuuluvaid isiklikke kui ka varalisi õigusi, mis võib endast kujutada autoriõiguste rikkumist autoriõiguse seaduse (edaspidi ka AutÕS¹¹) tähenduses.¹² AutÕS-s on kehtestatud mitmed vaba kasutuse erandid, mille järgi on autori varaliste õiguste teostamist võimalik teatud juhtudel piirata, näiteks Euroopa Parlamendi ja nõukogu direktiivist EL 2019/790 (edaspidi ka DSM-direktiiv¹³) tulenevad teksti- ja andmekaeve erandid, kuid ka Euroopa Parlamendi ja nõukogu direktiivist 2001/29/EÜ (edaspidi ka infoühiskonna direktiiv¹⁴) tulenevad vaba kasutuse erandid.¹⁵ Need AutÕS-s sätestatud vaba kasutamise erandid võimaldavad välistada autoriõiguse rikkumise.¹⁶ Praegusel ajal puudub aga õigusselgus selles osas, kas ja kuidas kehtiv õigusraamistik tehisintellekti treening- ja sisendandmete kasutamisele kohaldub.¹⁷

Kirjeldatud ebaselgus õiguslikus olukorras on toonud kaasa riikide erinevaid praktikaid. Näiteks on erasektori ettevõtted hakanud ära kasutama partnersuhteid ülikoolidega, mida ettevõtted ka ise tihti rahaliselt toetavad. Ettevõtted lasevad ülikoolidel koguda tehisintellekti treenimiseks autoriõigusega kaitstud teoseid, kasutades seejärel nende teoste abil treenitud tehisintellekti oma äritegevuses. Seda praktikat on õiguskirjanduses kutsutud ka „andmepesuks“, kuna mudeli loomiseks kasutatakse ära ülikoolidele rakenduvaid vaba kasutuse erandeid, mis võimaldavad autoriõigusega kaitstud teoste õiguspärast kasutamist teksti- ja andmekaeve eesmärkidel ilma autori nõusolekuta ning autorile tasu maksmata.¹⁸ Näiteks rahastas tehisintellektiettevõtte Stability AI, Ltd. Müncheni ülikoolis uurimisrühma,

¹¹ Autoriõiguse seadus. – RT I, 29.06.2022, 16.

¹² Isiklikest õigustest AutÕS § 12 lg 1 p 1 (õigus autorsusele), § 12 lg 1 p 5 (õigus au ja väärikuse kaitsele) ning varalistest õigustest AutÕS § 13 lg 1 p 1 (õigus teose reprodutseerimisele), § 13 lg 1 p 5 (õigus teose töötlemisele) ning § 13 lg 1 p 9¹ (õigus teose üldsusele kättesaadavaks tegemisele).

¹³ Euroopa Parlamendi ja nõukogu 17. aprilli 2019. aasta direktiiv (EL) 2019/790, mis käsitleb autoriõigust ja autoriõigusega kaasnevaid õigusi digitaalsel ühtsel turul ning millega muudetakse direktiive 96/9/EÜ ja 2001/29/EÜ. – ELT L 130.

¹⁴ Euroopa Parlamendi ja nõukogu 22. mai 2001. aasta direktiiv 2001/29/EÜ autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ühtlustamise kohta infoühiskonnas. – ELT L 167.

¹⁵ Mudeli treenimise ja kasutamise seisukohalt eelkõige infoühiskonna direktiivist üle võetud AutÕS § 18 ja § 18¹ ning DSM-direktiivist üle võetud AutÕS § 19¹ ja § 19².

¹⁶ Autoriõiguse rikkumine võib tekitada täiendavaid õiguslikke küsimusi autoriõigusliku vastutuse osas, kuid vastutust puudutavaid küsimusi magistritöö piiratud mahust tulenevalt käesolevas töös ei käsitleta.

¹⁷ Ka Maailma Intellektuaalse Omandi Organisatsioon on märkinud oma avalikustatud intellektuaalomandi poliitikale ja tehisintellektile keskenduvast ülevaatest, et enamik tehisintellekti autoriõiguslikest probleemidest on seotud tehisintellekti treening- ja sisendandmetega. Vt WIPO. Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence. 21.05.2020, lk 8. Kättesaadav: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf (09.09.2023).

¹⁸ Vt täpsemalt selle kohta nt Guadamuz, A. A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs. GRUR International 2/2024, lk 22.

mis lõi nende jaoks masinõppe mudeli ning kasutas selle treenimiseks andmeid, mida kogus teine ettevõttega seotud mittetulundusühingust uurimisinstituut. Kahtlemata on eeltoodud näide võimalik praeguse ebaselge olukorra tõttu kehtivas õiguses.

Käesoleva magistritöö eesmärk on selgitada välja, kas ja millistel tingimustel on autoriõigusega kaitstud teose kasutamine generatiivse tehisintellekti treening- ja sisendandmetena õiguspärane. Samuti on eesmärgiks analüüsida, kas ja milliseid muudatusi oleks õiguskorda vaja, et tagada tasakaal autori huvide kaitse ja generatiivse tehisintellekti innovatsiooni vahel, sh kas kavandatavas tehisintellekti määruuses olevad autoriõigusi käsitlevad sätted aitavad neid eesmärke saavutada. Kirjeldatud eesmärkidest tulenevalt analüüsib autor kolme uurimisküsimust:

- i) Millistel juhtudel ning missuguses ulatuses kasutatakse autoriõigusega kaitstud teose kasutamisel generatiivse tehisintellekti treening- ja sisendandmetena autorile kuuluvaid ainuõigusi?
- ii) Kas ja missugustel tingimustel võib autorile kuuluva ainuõiguse rikkumise välistada mõni kehtivas õiguses sätestatud vaba kasutamise erand?
- iii) Missugune mõju on kavandataval tehisintellekti määruusel autoriõigusega kaitstud teose treeningandmetena kasutamisele?

Magistritöö autor püstitab eeltoodud uurimisküsimuste põhjal hüpoteesi, et autoriõigusega kaitstud teose kasutamisel generatiivse tehisintellekti treening- ja sisendandmetena kasutatakse autori ainuõigusi, kuid olemasolevad vaba kasutuse erandid võivad teatud juhtudel välistada autoriõiguse rikkumise. Samuti on autor esialgsel seisukohal, et nii kehtiv AutÕS kui ka kavandatav tehisintellekti määruus tagavad tasakaalu autori huvide kaitse ja generatiivse tehisintellekti innovatsiooni vahel. Juhul, kui püstitatud hüpoteesid ei leia kinnitust, võib olla vajalik täiendada või muuta kehtivat õigust, et oleks tagatud nii autorile kuuluvate ainuõiguste kaitse kui ka tehnoloogia areng ning innovatsioon kaitstud teose kasutamisel generatiivse tehisintellekti treening- ja sisendandmetena. Kuivõrd asjassepuutuvad küsimused on reguleeritud Euroopa Liidu õiguses, võib sellest tulenevalt olla vajalik muudatuste tegemine Euroopa Liidu tasandil.¹⁹

¹⁹ Vaba kasutuse erandite kehtestamise piirangute kohta vt nt infoühiskonna direktiivi põhjenduspunkt 32.

Magistritöö uurimisküsimustele vastamiseks ja hüpoteeside kontrollimiseks on töö jaotatud kaheks peatükiks. Esimeses peatükis hindab autor esmalt generatiivse tehisintellekti tehnoloogilist olemust. Generatiivse tehisintellekti toimimise ja tehnoloogilise olemuse mõistmine on oluline eelkõige selleks, et välja selgitada, millistel juhtudel kasutatakse generatiivse tehisintellekti treening- ja sisendandmetena mõnda autorile kuuluvat ainuõigust. Enne autorile kuuluvate ainuõiguste analüüsi analüüsitakse seda, millised autoriõigusega kaitstud teosed võivad olla mudelis kasutatavateks treening- ja sisendandmeteks ning seejärel, kas ja missuguseid autori ainuõigusi mudeli treenimisel kasutatakse. Teine peatükk keskendub olemasolevate vaba kasutamise erandite hindamisele, selgitades välja, kas ja missugustel tingimustel võib autorile kuuluva ainuõiguse rikkumise välistada mõni kehtivas õiguses sätestatud vaba kasutamise erand. Lisaks analüüsitakse tehisintellekti määruse eelnõud, eelkõige seda, kas määrusest tulenevad nõuded aitavad tagada treening- ja sisendandmete õiguspärase kasutamise ning tasakaalu autorile kuuluvate ainuõiguste ning innovatsiooni vahel.

Magistritöös esitatud eesmärkide saavutamiseks kasutab autor võrdlevat ja analüütilist uurimismeetodit. Analüütiline uurimismeetod hõlmab asjakohaste õigusnormide, õiguskirjanduse ning kohtupraktika analüüsi. Võrdleva uurimismeetodi kasutamiseks on autor valinud võrdlusriikideks angloameerika õigussüsteemi kuuluvad Ameerika Ühendriigid ja Ühendkuningriigi, kuivõrd tegemist on riikidega, mis on juhtival positsioonil nii kõnealuse magistritöö teemat puudutavas akadeemilises arutelus kui ka mudelite arendamisel. Põhjuseks on ka see, et Ameerika Ühendriikides on käesoleva töö kirjutamise hetkel menetluses hulgaliselt kohtuasju, mis on seotud generatiivse tehisintellekti arendamise ning treeningandmete kasutamise kontekstis võimaliku autoriõiguse rikkumisega. Lisaks on Ühendkuningriigis olnud pärast Euroopa Liidust välja astumist üheks võtmeteemaks vajadus kohandada intellektuaalse omandi seadust tehisintellekti valguses.²⁰ Võrdlevat uurimismeetodit kasutab autor läbivaldt asjakohastes peatükkides.

Varasemalt on Eestis tehisintellekti arendamist ja sellega seonduvaid autoriõiguslikke probleeme puudutatud eestikeelsetes teadustöodes tehisintellekti loodavate väljundandmete vaatest.²¹ Ühes teadusartiklis on varasemalt käsitletud vestlusrobotite loomisega seonduvaid

²⁰ Vt lisaks Bently, L. jt. *Intellectual Property Law* (6th edition). Oxford University Press: 2022, lk 60-61; Ameerika Ühendriikides menetluses olevate kohtuasjade kohta viide 6.

²¹ Vt Vahlberg, J.-A. *Tehisintellekti loomingu autoriõiguslik kaitse*. Tartu 2017; Arvisto, Ä. *Tehisintellekti loomingu autorsus*. Tartu 2022.

autoriõiguslikke probleeme.²² Lisaks on ühes magistritöös analüüsitud DSM-direktiivist tulenevat teksti- ja andmekaeve erandit, kuid seda treening- ja sisendandmeid käsitlemata.²³

Magistritöö esmasteks allikateks on Eesti ja Euroopa Liidu õigus koos asjakohase kohtupraktikaga, mis aitab seda sisustada.²⁴ Kuna töö eesmärgiks on hinnata kavandatava tehisintellekti määruse mõju treeningandmete kasutamisele, on töö üheks allikaks ka tehisintellekti määruse eelnõu.²⁵ Samuti on allikateks võrdlusriikide õigusaktid ja kohtupraktika, menetluses olevad kohtuasjad, eelkõige Ameerika Ühendriikides, ning võrdlusriikide tehisintellekti käsitlevate õigusaktide eelnõud. Teised allikad on seaduste seletuskirjad, õiguskirjandus ning muu tehisintellekti ja autoriõiguste valdkonda puudutav kirjandus.

EMS märksõnad: intellektuaalne omand, autoriõigus, tehisintellekt²⁶

²² Kelli, A jt. Vestlusrobotid ja autoriõigus. *Juridica* 5/2020. Kõnealune teadusartikkel käsitleb erinevalt käesolevast magistritööst vestlusrobotite loomisega seonduvaid õiguslikke kitsaskohti. Võrreldes vestlusrobotitega, ei märgistata generatiivse tehisintellekti treenimisel andmestikke. Vt selle kohta ka peatükk 1.1.

²³ Vt ka Sõrmus, S. Digiühiskonna direktiivis sätestatud teksti- ja andmekaeve erandi autoriõiguslikud väljakutsed. Tallinn 2020.

²⁴ Kuigi tegemist on globaalselt kõneainet pakkuva temaga, on Eesti õiguse analüüsi pinnalt võimalik teha järeldusi ka Euroopa Liidu õiguse kohta laiemalt, kuivõrd ELi liikmesriikides on autoriõigus osaliselt harmoniseeritud.

²⁵ Euroopa Parlament võttis 13. märtsil 2024. aastal Euroopa Komisjoni esitatud tehisintellekti määruse ettepaneku vastu, kuid Euroopa Liidu Nõukogu ei ole veel käesoleva lõputöö kirjutamise hetkel Euroopa Parlamendi seisukohta heaks kiitnud. Edaspidi on töös läbivalt kasutatud tehisintellekti määruse eelnõu kohta vastet „tehisintellekti määruse eelnõu“ või „kavandatav tehisintellekti määrus“. Vt Euroopa Parlamendi 13. märtsi 2024. aasta otsus tehisintellekti määruse kohta. Kättesaadav: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (18.03.2024).

²⁶ Eesti märksõnastiku (EMS) kohaselt on tehisintellekt infotehnoloogia ja informaatika valdkonda kuuluv märksõna, kuid arvestades valdkonna seost käesoleva tööga, on märksõna loetlemine töö seisukohast asjakohane.

1. AUTORIÕIGUSEGA KAITSTUD TEOSE KASUTAMINE LÄHTUVALT GENERATIIVSE TEHISINTELLEKTI TEHNOLOOGILISEST OLEMUSEST

1.1. Generatiivse tehisintellekti tehnoloogiline olemus

Käesoleva peatüki eesmärgiks on hinnata, millistel juhtudel ning missuguses ulatuses on autoriõigusega kaitstud teose kasutamine mudeli treening- ja sisendandmetena käsitatav mõne autorile kuuluva ainuõiguse rikkumisena. Selleks, et hinnata, missuguses ulatuses ja missugustel juhtudel autoriõigusega kaitstud teoseid mudeli treening- ja sisendandmetena kasutatakse, on oluline esmalt mõista, milline on generatiivse tehisintellekti ülesehitus. Eelkõige on oluline mõista, kuidas generatiivse tehisintellekti treenimiseks kasutatavad andmed saadakse, töödeldakse ja säilitatakse. Pärast generatiivse tehisintellekti tehnoloogiliste etappide väljaselgitamist hindab autor, missuguseid treening- ja sisendandmeid võib käsitada autoriõigusega kaitstud teosena. Seejärel on võimalik hinnata, missugusel määral kasutatakse generatiivse tehisintellekti treenimisel ja hilisemal kasutamisel autorile kuuluvaid ainuõigusi.

Enne täpsema generatiivse tehisintellekti tehnoloogilise olemuse juurde jõudmist käsitleb autor generatiivse tehisintellektiga seonduvat terminoloogiat. Ühest kehtivat tehisintellekti legaaldefinitsiooni käesoleva töö kirjutamise hetkel veel ei ole.²⁷ Generatiivse tehisintellektina mõistetakse igasugust tehisintellekti, mis on võimeline looma sisu, sealhulgas pildi-, heli-, video- või tekstisisu.²⁸ Tehisintellekti määruse eelnõus on defineeritud üldine (üldotstarbeline) tehisintellekti mudel tehisintellekti mudelina, mida on treenitud suure andmemahuga ning mis suudab täita mitmesuguseid erinevaid ülesandeid.²⁹ Ühe täiendava kriteeriumina on välja toodud võimekus integreerida mudelit mitmesugustesse süsteemidesse või rakendustesse.³⁰ Sisult vastab üldise tehisintellekti mudeli määratlus generatiivsele tehisintellektile, mistõttu on

²⁷ Tehisintellekti määruse eelnõus on defineeritud tehisintellekti süsteem (*AI system*) „kui masinapõhine süsteem (*a machine-based system*), mis on loodud töötama erinevate autonoomiasemetega ja mis võib pärast kasutuselevõtmist olla kohanisvõimeline ning mis otseste või kaudsete käskluste puhul suudab saadud sisendi põhjal luua väljundeid, nagu ennustusi, sisu, soovitusi või otsuseid, mis võivad mõjutada füüsilist või virtuaalset keskkonda“. Määruses olev definitsioon lähtub Majanduskoostöö ja Arengu Organisatsiooni (OECD) varasemalt välja töötatud definitsioonist. Sellest järelduvalt võib asuda seiskohale, et määruse jõustumisega saavutatakse teatud üksmeel rahvusvahelisel tasandil tehisintellekti puudutavate definitsioonide osas. Vt Euroopa Komisjoni tehisintellekti määruse eelnõu, artikkel 3 p 1 (viide 25); OECD. *AI Principles*. 2019. Kättesaadav: <https://oecd.ai/en/ai-principles> (02.02.2024).

²⁸ Feuerriegel, S. jt. *Generative AI. Business & Information Systems Engineering 2024*, lk 112.

²⁹ Üldine (üldotstarbeline) tehisintellekti mudeli (*general purpose AI model*) definitsioon ei hõlma tehisintellekti mudeleid, mida kasutatakse enne turule toomist uurimis-, arendus- või prototüüpimise tegevuses. Vt tehisintellekti määruse eelnõu artikkel 3 p 63 (viide 25).

³⁰ Samas.

määruses nimetatud üldise tehisintellekti mudeli puhul tegemist muuhulgas generatiivse tehisintellektiga.

Kavandatavas tehisintellekti määruses leiduvad ka treening- ja sisendandmete õiguslikud määratlused. Treeningandmete näol on tegemist andmetega, mida kasutatakse mudeli koolitamiseks ning mudeli kaalude loomiseks.³¹ Üheks levinumaks treeningandmete kogumise meetodiks mudeli treenimisel on andmete kraapimine.³² Üsna sageli hoiavad mudelite arendajad andmekraape abil saadud treeningandmeid ning nendest moodustuvaid andmekogusid konfidentsiaalsena.³³ Samuti jäävad täpsed treeningmeetodid ettevõtte ärisaladusega seonduvatel põhjustel avalikkuse eest varjatuks.³⁴ Lisaks treeningandmetele kasutatakse mudelis ka sisendandmeid, mille esitab mudeli kasutaja valmis mudelile ning nende abil genereerib mudel väljundi.³⁵ Nii treening- kui ka sisendandmed muudetakse mudelites väiksemateks tükkideks ehk tookeniteks, mis võimaldavad nende üheaegset tervikuna töötlemist. Näiteks kasutati OpenAI arendatavas ChatGPT-s kasutatava GPT3-tüüpi mudeli treenimiseks ligikaudu 499 miljardit tookenit.³⁶ See ilmestab, et andmekogud, mille peal generatiivne tehisintellekt on treenitud, hõlmavad endas väga suurtes kogustes andmeid, mistõttu on ka tõenäoline, et treeningandmed sisaldavad autoriõigusega kaitstud teoseid, mida on teksti- ja andmekaeve abil töödeldud.

Generatiivse tehisintellekti aluseks on masinõpe, mis on tehisintellekti alamliik ning mis kasutab väljundväärtuste ennustamiseks varasemaid andmeid.³⁷ Tänu viimase aastakümne

³¹ Tehisintellekti määruse eelnõu põhjenduspunkt 102 ja artikkel 3 p 29 (viide 25). Mudeli kaaludest on lähemalt rääkinud nt USA autoriõiguste amet. Vt United States Copyright Office. Artificial Intelligence and Copyright, 30.08.2023. Kättesaadav: <https://www.copyright.gov/ai/docs/Federal-Register-Document-Artificial-Intelligence-and-Copyright-NOI.pdf> (28.03.2024).

³² Teksti- ja andmekaeve kohta vt täpsemalt käesoleva töö peatükk 2.1. Teksti- ja andmekaevet (*text and data mining*) tuleb eristada andmete kraapimisest, andmekraapest (inglisekeelsena *web scraping* või *web mining*). Vt nt Krotov, V., Johnson, L., Silva, L. Tutorial: Legality and Ethics of Web Scraping. Communications of the Association for Information Systems 2020, lk 555-581.

³³ Matulionyte, R. Reconciling Trade Secrets and AI Explainability: Face Recognition Technologies as a Case Study. European Intellectual Property Review 2022, lk 10-13.

³⁴ Seda, kas mudelite treeningmeetodid ka Eesti õiguses ärisaladusena kaitstud oleks, käesolev töö ei käsitle. Vt ebaausa konkurentsi takistamise ja ärisaladuse kaitse seadus. – RT I, 07.12.2018, 2.

³⁵ Tehisintellekti määruse eelnõu artikkel 3 p 33 (viide 25).

³⁶ Vt Guadamuz, A. (viide 18), lk 113. GPT-tüüpi mudelite puhul on tegemist eeltreenitud generatiivsete transformerimudelitega (*generative pre-trained transformer*), mida on nimetatud ka suurteks keelemudeliteks (*large language models*), kuna need on võimelised mõistma inimkeelseid käskluseid ning looma ka arusaadavat väljundit.

³⁷ Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science 2021, lk 160.

jooksul toimunud erakordsele hüppele kiipide, eriti graafikaprotsessorite (edaspidi ka GPU³⁸) võimsuses, on esmakordselt suudetud luua mitmekihilisi närvivõrke.³⁹ Mitmekihilised ehk süvanärvivõrgud on masinõppe meetodite edasiarendus, mis suudavad täita suuremas koguses ülesandeid ning õppida keerukamaid protsesse. Eelnevalt mainitud arengud on kiirendanud mudeli treenimise protsesse ligi sajakordselt ning seeläbi on saanud alguse ka generatiivse tehisintellekti eksponentsiaalne kasv.⁴⁰ Generatiivse tehisintellekti treenimisel õpetatakse mudelit treeningandmete abil läbi mitme erineva kihi, millest igaüks tunneb ära mõne kindla omaduse treeningandmetest.⁴¹ Näiteks näotuvastusmudeli loomisel võib üks kiht õppida tundma ära silmi, teine kiht õpib tundma ära suud ning kolmas kiht ninasid. Kokkupanudult suudavad aga kõik kihid koos tunda ära inimeste nägusid. Lisades veel kihte, saavutavad mudelid oskuse täita mitte ainult sisendandmete kirjeldamise, vaid ka nende loomise ülesannet.⁴²

Uute loomeoskuste treenimiseks on aga vajalik suur kogus andmeid, mille abil närvivõrke luua ja välja õpetada.⁴³ Treeningandmetena kasutatakse internetist kokku kogutud tekste ja muud sisu, mida on tehtud näiteks ka ChatGPTs kasutatavate GPT-tüüpi mudelite puhul, mida eeltreenitakse suurte märgistamata andmete ja tekstikogumike põhjal.⁴⁴ Seda tüüpi mudelid on suutelised õppima just märgistamata ja korrastamata andmetest, mistõttu on mudeleid võimalik treenida niivõrd suure hulga andmestikega, mida ükski inimene ei oleks suuteline oma eluajal läbi lugema. Autoriõiguste kontekstis võib see tähendada aga taaskord suuremat võimalust, et treeningandmetena kasutatakse autoriõigusega kaitstud teoseid.

Mudelite kasutamise ja treenimise võib jagada kolme etappi. Esimesed kaks, mudeli (eel)treenimine ning peenhäälestamine, kuuluvad mudeli loomise alla.⁴⁵ Treenimisel kasutatakse suurt kogust kogutud andmeid, et luua üldine arusaam keelest ja sisust ning

³⁸ Vt graafikaprotsessorite (*graphical processing unit*) kohta veel ka Pandey, M. jt. The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence* 2022, lk 211–221.

³⁹ Samas, lk 220–221.

⁴⁰ Alzubaidi, L. jt. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* 8/2021, lk 49.

⁴¹ Samas.

⁴² Harshvardhan, G. M. jt. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 11/2020, lk 22.

⁴³ Samas.

⁴⁴ Uszkoreit, J. Transformer: A Novel Neural Network Architecture for Language Understanding. 31st Conference on Neural Information Processing Systems NIPS 2017, lk 2-3.

⁴⁵ Samas.

tekitada mudelile n-ö baasoskused.⁴⁶ Mudelit õpetatakse treenimise abil mõistma andmekogus leiduvaid mustreid, korrapärasusi ja reegleid.⁴⁷ Pärast mudeli treenimist võib andmekogu kustutada, kuna nendes leiduvad teadmised on juba talletatud mudeli parameetrite õpitavasse osasse ehk mudeli kaaludesse. Siiski on autori arvates treenimiseks kasutatavate andmekogude kustutamine vähetõenäoline, kuna andmekogud on iseseisva majandusliku väärtusega ning vajalikud ka järgmiste mudelite loomisel. Mudeli loomisel on teiseks, valikuliseks sammuks mudeli peenhäälestus.⁴⁸ Peenhäälestus võimaldab mudelit suunata mõnda spetsiifilisemat ülesannet või valdkonda paremini mõistma.⁴⁹ Erinevalt mudeli eeltreenimisest, võib peenhäälestamine toimuda ka mõne kolmanda isiku poolt, kes soovib mõnda olemasolevat mudelit enda tarbeks kohandada. Ka peenhäälestuseks on vaja luua esmalt andmekogu, kuid selle protsessi tarbeks on kogud üldiselt väiksemad.⁵⁰ Seega ilmneb eeltoodust, et mudeli treenimise ja peenhäälestamise etappide puhul on tegemist erinevate tegevustega, mida võivad teha erinevad isikud. Mudeli treenimisel on selleks mudeli algne arendaja, kuid peenhäälestamisel võib lisaks mudeli arendajale selleks olla ka mõni kolmas isik. Peenhäälestamine võib seega toimuda nii veel valmimise järgus oleva kui ka juba valminud mudeliga.

Kolmandas etapis ehk generatiivse tehisintellekti kasutamisel edastab kasutaja käsklused ning sisendandmed valmis treenitud mudelisse. Mudelid muudavad inimkeeles loetavad sisendandmed masinkeelseks, mida lastakse seejärel järjestikuste eelnevalt treeningandmete põhjal loodud närvivõrgu kihtide läbi.⁵¹ Seejuures võivad inimkeeles loetavad sisendandmed sisaldada viimaste mudelite, näiteks Google'i Gemini puhul tuhandete lehekülgede pikkuseid tekste, mis võivad olla kaitstud ka autoriõigustega.⁵²

Lähtudes eeltoodust, võib treening- ja sisendandmete kasutamise jagada järgnevateks etappideks: i) treeningandmete kogumine; ii) treeningandmete abil mudeli treenimine, mille

⁴⁶ Bommasani, R. jt. On the Opportunities and Risks of Foundation Models. ArXiv 2021, lk 48.

⁴⁷ Samas.

⁴⁸ Peenhäälestamise (*fine-tuning*) kohta vt ka tehisintellekti määruse põhjenduspunktid 104 ja 109 (viide 25).

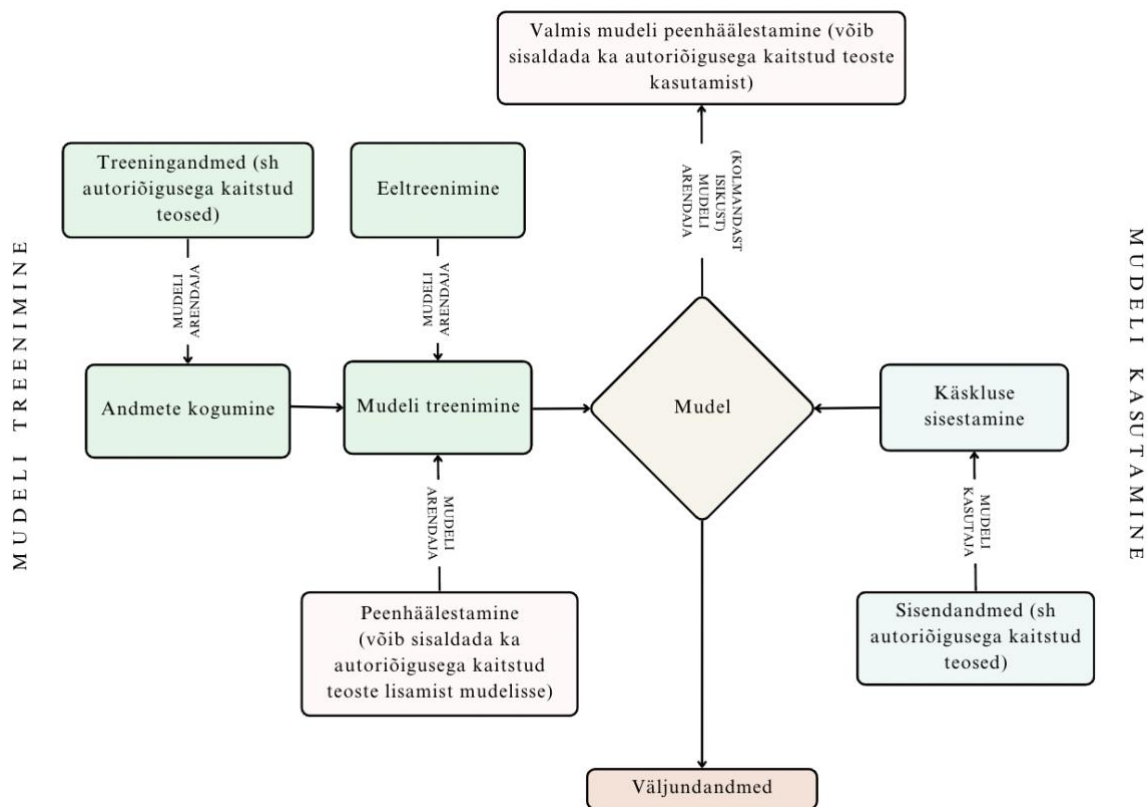
⁴⁹ Dingliwal jt. Prompt Tuning GPT-2 language model for parameter-efficient domain adaptation of ASR systems. InterSpeech. ArXiv 2021, lk 2-3.

⁵⁰ Radiya-Dixit E., Wang, X. How fine can fine-tuning be? Learning efficient language models. International Conference on Artificial Intelligence and Statistics. Italy: PMLR Volume 108 2020, lk 6.

⁵¹ Vaswani, A. jt. Attention is All You Need. 31st Conference on Neural Information Processing Systems. NIPS 2017, lk 5.

⁵² Vt nt Google DeepMindi avaldatud raport. Google DeepMind. Gemini: A Family of Highly Capable Multimodal Models. ArXiv 2024, lk 3. Kättesaadav: <https://arxiv.org/pdf/2312.11805> (12.04.2024).

hulka võib kuuluda lisaks mudeli eeltreenimisele ka mudeli peenhäälestus; ning iii) mudeli kasutamine koos sisendandmetega ja valmis mudeli võimalik hilisem peenhäälestamine. Sellest tulenevalt on võimalik omakorda järeldada, et mudeli treenimise eri etappides toimub treening- ja sisendandmete kasutamine eri viisidel ning erinevate isikute poolt, mis võib ühtlasi mõjutada nii autorile kuuluvate ainuõiguste kasutamist, s.o autoriõiguse rikkumist kui ka rikkumist välistavatele vaba kasutuse eranditele tuginemise võimalikkust ning hilisemat vastutust. Mudeli treenimise, peenhäälestamise ja kasutamise etapid esitab autor kokkuvõtlikult ka alloleval joonisel.



Joonis 1. Generatiivse tehisintellekti treenimine, peenhäälestamine ja kasutamine.

1.2. Treening- ja sisendandmed kui autoriõigusega kaitstud teosed

Eelnevalt selgitatust lähtudes on mudelitele omane võimekus treenida ennast väga suurte andmekogude abil. Üldjuhul valitakse eeltreenimise etapis mudeli treenimiseks mõni olemasolev suur andmekogu või luuakse iseseisev andmekogu, mis koosneb omakorda

mitmetest eri kohtadest pärit andmetest.⁵³ Peamisteks algallikateks andmekogude loomisel võivad olla erinevad teosed, millest osad on kaitstud ka autoriõigusega. Mudelite eeltreenimisel võidakse kasutada muuhulgas kirjanduslikke tekste, fotosid, audio- ja videoteoseid ning muid teoseid. Näiteks selleks, et üks mudel oskaks luua väljundina mõne kirjandusteose kokkuvõtte, on vaja erinevaid treeningandmeid, mis võivad sisaldada hulgaliselt nii luule- kui ka proosateoseid.

Treening- ja sisendandmed on käsitatavad autoriõigusega kaitstud teosena üksnes juhul, kui need vastavad kindlatele kumulatiivsele eeldustele. AutÕS § 4 lg 1 järgi peab teos autoriõigusliku kaitse saamiseks pärinema kirjanduse, kunsti või teaduse valdkonnast.⁵⁴ Samuti peab teos olema originaalne ehk autori enda intellektuaalse loomingu tulemus. Lisaks peab teos olema väljendatud ka objektiivses vormis.⁵⁵ Tulenevalt Euroopa Kohtu (edaspidi ka EK) praktikast tuleb teost tõlgendada Euroopa Liidus autonoomselt ja ühetaoliselt.⁵⁶ EK praktikas on selgitatud, et teos peab olema algupärane, st autori enese intellektuaalne loomingu.⁵⁷ Euroopa Kohtu praktika kohaselt on originaalsuse kriteerium täidetud, kui autor on väljendanud oma loominguilisi võimeid teose loomisel, tehes vabasid ning loominguilisi valikuid.⁵⁸ Mudelite treening- ja sisendandmetena kasutatavate teoste puhul on autoriõigusliku kaitse saamise esmane eeldus täidetud, kuna treenimiseks kasutatakse muuhulgas erinevaid teoseid, mis pärinevad kirjanduse, kunsti ja teaduse valdkonnast. Mudelite treenimise seisukohast on kirjeldatud eeldustest keerulisim originaalsuse hindamine. Kuivõrd mudeli treeningandmetena kasutatakse nii teoseid kui ka teoste osasid, siis tuleb teose puhul igakordselt hinnata, kas teoses väljendub ka autori intellektuaalne loomingu, st tegemist on originaalse teosega. Mudelite treenimiseks kasutatavate andmete puhul peab olema täidetud ka objektiivsuse kriteerium. Euroopa Kohtu praktikast tulenevalt peab autoriõiguse kaitse objekti väljendama selliselt, et “see oleks piisavalt täpselt ja objektiivselt identifitseeritav, olgugi et

⁵³ Alzubaidi, L. jt (viide 40), lk 43.

⁵⁴ Berni kirjandus- ja kunstiteoste kaitse konventsioon – RT II 1994, 16, 49. Vt artikkel 2 lg 1.

⁵⁵ AutÕS § 4 lg 2 kohaselt tuleb eristada ka ideed ja väljendusvormi. Berni konventsiooni artikkel 2 sätestab ka fikseerimise nõude: „fikseeritud mingis materiaalses vormis“. Taolist fikseerimise nõuet Eesti õigusesse üle võetud ei ole.

⁵⁶ EKo C-310/17, *Levola Hengelo*, ECLI:EU:C:2018:899, p-d 33 ja 45.

⁵⁷ EKo C-5/08, *Infopaq International A/S v. Danske Dagblades Forening*, EU:C:2009:465, p-d 37 ja 39. Vt lisaks ka RKTko 2-16-8751, p 22 jj. Selles kohtuasjas pidi kohus hindama, kas meediamonitooringut pakkuv ettevõte on rikkunud infoühiskonna direktiivi artiklis 5 sätestatud reprodutseerimisõigust seeläbi, et kasutas ilma nõusolekuta ning tasu maksmiseta ajakirjandusväljaannetes olevaid artikleid monitoorimisraporti koostamiseks. Monitoorimise tulemusel loodi raport, kus olid 11 sõna pikkused väljavõtted artiklitest.

⁵⁸ EK praktika kohaselt on originaalsuse nõue täidetud üksnes siis, kui autor väljendab oma loominguilist originaalsel viisil, näiteks järjestuse ja elementide valiku kaudu ning kasutab loomevabadust. Vt EKo C-145/10, *Painer*, EU:C:2011:798, p 89-92; EKo C-403/08, *Football Association Premier League*, EU:C:2011:631, p 98.

see väljendus ei pea tingimata olema püsiv”.⁵⁹ Mudelite puhul, seda nii treenimisel kui ka hilisemal peenhäälestamisel, on objektiivse vormi kriteerium täidetud, kuivõrd mudeleid treenitakse reeglina selliste andmekogudesse kogutud teostega, mis on mõnel andmekandjal. See tähendab, et mudeli treenimiseks on vaja teoseid, mis on kirjalikult fikseeritud ja objektiivselt identifitseeritavad. Ka õiguskirjanduses on seda seisukohta jagatud ning leitud, et praktikas on objektiivse vormi tingimus täidetud, kui treenitakse mudeleid.⁶⁰

Neljanda tingimusena on mudelite treenimise ja peenhäälestamise vaatest oluline käsitada seda, et tegemist ei ole sellise intellektuaalse tegevusega, millele autoriõiguslikku kaitset ei anta (AutÕS § 5). Näiteks ei laiene autoriõiguslik kaitse lisaks eelpool nimetatud andmetele muuhulgas ideedele, faktidele ning päevauudistele. Mudelite treenimisel on vaja luua nendes ka arusaam maailmas toimuvatest sündmustest, mistõttu kasutatakse treeningandmetena tihti peale ajaleheartikleid.⁶¹ Päevakajalise sündmuse kirjeldus ei saa seega eeltoodust tulenevalt autoriõiguslikku kaitset, kuid see ei tähenda, et autoriõiguslik kaitse ei laieneks teistele ajaleheartiklitele ega artiklis olevale tekstile.⁶²

Lisaks mudeli treenimiseks ja peenhäälestamiseks kasutatavale treeningandmetele võib hiljem kasutaja poolt sisendandmetena mudelile antav andmestik olla kaitstud teoseks. Sageli võivad kasutaja sisendandmed koosneda paarirealisest lõigust või väljendist. Euroopa Kohtu praktikas on leitud, et ka 11 üksteisele järgnevat sõna on piisav autoriõigusliku kaitse saamiseks.⁶³ Seejuures ei ole tähtis sõnade arv vaid autori intellektuaalne panus.⁶⁴ Seetõttu võivad ka kasutaja sisendandmed kujutada endast autoriõigusega kaitstavat teost.

Mudeli treenimiseks võidakse kasutada ka juba olemasolevaid andmekogusid. Tulenevalt Euroopa Parlamendi ja nõukogu direktiivist 96/9/EÜ andmebaaside õiguskaitse kohta laieneb

⁵⁹ C-310/17, *Levola Hengelo*, p 40.

⁶⁰ Vt Kelli, A. jt. The Extent of Legal Control over Language Data: the Case of Language Technologies. Proceedings of CLARIN Annual Conference 2019: CLARIN 73 Annual Conference, Leipzig, Germany, 30 September – 2 October 2019. Ed. K. Simov and M. Eskevich. CLARIN, lk 71. Kättesaadav: <http://hdl.handle.net/10138/327736> (09.12.2023). Konkreetne artikkel puudutas küll kitsamalt keelemudeleid, kuid sarnast järeldust on võimalik laiendada ka teistele generatiivse tehisintellekti tüüpidele.

⁶¹ Ray, P. P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems Volume 3, 2023, lk 135.

⁶² Just ajakirjanduslike artiklite kasutamise tõttu on alगतatud ka mitmeid kohtuasju USA-s. Selle kohta vt täpsemalt käesoleva töö peatükk 1.3.

⁶³ C-5/08, *Infopaq International*, p 74.

⁶⁴ Samas, p 45.

andmebaasidele autoriõiguslik kaitse, kui andmebaas vastab teose tunnustele.⁶⁵ Andmebaas peab samuti autoriõigusliku kaitse saamiseks vastama originaalsuse kriteeriumile selles mõttes, et tegemist peab olema andmebaasi sisu valiku või korrastamisega, mis on autori enese intellektuaalne looming. Eeltoodust tulenevalt saab autoriõigusliku kaitse aluste treening- ja sisendandmetena käsitada mitmesuguseid kirjanduslikke tekste, visuaalteoseid, andmebaase ning ka pikemaid väljendeid.⁶⁶ Järgnevalt analüüsitakse, kas ja millistel juhtudel kujutab selliste teoste kasutamine ilma autori nõusolekuta endast autori ainuõiguste teostamist.

1.3. Autori ainuõigused kaitstud teoste treening- ja sisendandmetena kasutamisel

1.3.1 Autori varaliste õiguste kasutamine

1.3.1.1. Treening- ja sisendandmete reprodutseerimine ning töötlemine

Kui mudeli treening- või sisendandmed vastavad eelnevalt nimetatud teoste tunnustele, siis võib selliste andmete kasutamisega kaasneda autorile kuuluvate ainuõiguste kasutamine.⁶⁷ Teose autori nõusolekuta, näiteks ilma litsentsita kasutamine mõnel sellisel viisil, milleks seadus näeb autorile ette ainuõiguse, kujutab üldjuhul mõne AutÕS §-s 13 sätestatud autorile kuuluva õiguse rikkumist. Hindamaks seda, missuguses ulatuses ning juhtudel võib treening- ja sisendandmete kasutamine endast kujutada autori ainuõiguste rikkumist, tuleb esmalt eraldiseisvalt hinnata, missuguseid autorile kuuluvaid varalisi õigusi treening- ja sisendandmete kasutamisel kasutatakse. Varaliste õigustena on vajalik hinnata reprodutseerimisõigust (AutÕS § 13 lg 1 p 1), töötlemisõigust (AutÕS § 13 lg 1 p 5) ja üldsusele kättesaadavaks tegemise õigust (AutÕS § 13 lg 1 p 9¹). Nende puhul on tegemist kõige asjakohasemate autori varaliste õigustega, kuivõrd mudeli treenimisel, peenhäälestamisel ja hilisemal kasutamisel võidakse kõige tõenäolisemalt kasutada nimetatud autorile kuuluvaid varalisi õigusi. Kokkupuude teiste AutÕS §-s 13 sisalduvate varaliste

⁶⁵ Euroopa Parlamendi ja nõukogu 11. märtsi 1996. aasta direktiiv 96/9/EÜ andmebaaside õiguskaitse kohta. – EÜT L 77, 27.3.1996. Vt artikkel 3.

⁶⁶ Võrdlusriikides, Ameerika Ühendriikides ja Ühendkuningriigis teosele autoriõigusliku kaitse tekkimise kohta vt Ameerika Ühendriikide autoriõiguste seadus Copyright Act of 1976, 17. U.S.C § 102(b). Kättesaadav: <https://www.copyright.gov/title17/> (13.12.2023); Ühendkuningriigi autoriõiguse, tööstusdisaini ja patendi seadus Copyright, Designs and Patents Act 1988, §-d 1-3. Kättesaadav: <https://www.legislation.gov.uk/ukpga/1988/48/contents> (13.12.2023).

⁶⁷ AutÕS § 11 lg 2 ja § 46 lg 1 järgi on varalisi õigusi võimalik eelkõige üle anda, kas litsentsi või mõne muu lepingu alusel või seaduse alusel (nt AutÕS § 32 lg 1). Vt ka VÕS § 368 jj litsentsilepinguga õiguste üleandmise kohta. Võlaõigusseadus. – RT I, 06.07.2023, 116.

õigustega on vähetõenäoline arvestades seda, kuidas mudeli treenimine, peenhäälestamine ning kasutamine toimub.

Õigus teose reprodutseerimisele tuleneb infoühiskonna direktiivi artiklist 2. AutÕS-s on õigus teose reprodutseerimisele sätestatud § 13 lg 1 p-s 1 ning see on sarnaselt infoühiskonna direktiivile defineeritud laiaulatuslikult. AutÕS § 13 lg 1 p 1 ja infoühiskonna direktiivi artikli 2 järgi kujutab reprodutseerimine „teosest või teose osast ühe või mitme ajutise või alalise koopia tegemist“. Seejuures võib koopia olla tehtud otseselt või kaudselt ning mis tahes vormis või mis tahes viisil. Reprodutseerimiseks loetakse seega ka seda, kui teosest tehakse väljavõtteid, teost kopeeritakse või teos talletatakse.

Reprodutseerimisõiguse võimaliku kasutamise hindamisel tuleb eristada mudeli treenimist, sh andmete kogumist ja peenhäälestamist ning selle hilisemat kasutamist. Selleks, et mudeli algne arendaja saaks treenida generatiivset tehisintellekti, on vaja koguda erinevaid andmeid, sh autoriõigusega kaitstud teoseid ehk teisisõnu on mudeli treenimiseks vaja teostada andmekraabet.⁶⁸ Teoste kogumise käigus tuleb saadud treeningandmed talletada mõnele välisele andmekandjale või serverisse selleks, et kogutud teoste abil oleks võimalik hiljem mudelit treenida, näiteks teksti- ja andmekaeve meetodit kasutades. Talletamise käigus tuleb seetõttu teha ka kogutud teostest koopiaid. Lähtudes AutÕS § 13 lg 1 p-s 1 olevast reprodutseerimisõiguse sõnastusest, on võimalik järeldada, et sõltumata sellest, kas teosest tehtud koopia säilitamine on ajutine või püsiv, on tegemist reprodutseerimisega. Isegi juhul, kui kogutud treeningandmed kustutatakse pärast mudeli treenimist, on teose reprodutseerimine juba toimunud.⁶⁹ Mudeli treeningandmetena autoriõigusega kaitstud teoste kogumine *per se* kujutab seega endast reprodutseerimisõiguse kasutamist.⁷⁰

Euroopa Kohtu praktikas on autori reprodutseerimisõigust käsitletud korduvalt, kuid käesoleva töö kirjutamise ajal ei ole lahendeid, mis käsitleksid otseselt teksti- ja andmekaevet reprodutseerimisõiguse kontekstis. DSM-direktiivi kommenteerivas väljaandes on leitud kokkuvõtvalt, et teksti- ja andmekaeve teostamist saab käsitada autoriõigusega kaitstud teoste

⁶⁸ Selle kohta vt käesoleva töö alapeatükk 1.2.

⁶⁹ Sarnast seisukohta peegeldab ka võrdlusriigi Ühendkuningriigi kohtupraktika, vt nt *FAPL v. British Communications*, 2017 EWHC 480, p 31. Reprodutseerimisõigus on Ühendkuningriigis sätestatud CDPA §-s 18. Viidatud kohtuasjas käsitati reprodutseerimisõiguse kasutamisenähtena interneti vahendusel striimimist.

⁷⁰ Vt Ørstavik, I. B. *Access to data for training algorithms in machine learning: copyright law and “right-stacking”*. Edward Elgar Publishing 2022, lk 8.

reprodutseerimisena.⁷¹ Sõltumata kohtupraktikas puuduvast otsesest käsitlusest, on võimalik teatud paralleele teksti- ja andmekaeve ning selle käigus teoste reprodutseerimise kohta tuua ka varasemast kohtupraktikast. Euroopa Kohus on märkinud, et kui kasutada autoriõigusega kaitstud teosest mõnda osa, siis selleks, et oleks tegemist reprodutseerimisõiguse kasutamisega tuleb hinnata, kas see teosest kasutatud osa oli ka ilma teose ülejäänud osadeta originaalne, st teose eraldiseisev osa peab samuti väljendama autori enda intellektuaalset loomingut.⁷² Näiteks, kui kasutada mõnda tehnilist osa teosest, siis ei pruugi olla tegemist reprodutseerimisõiguse kasutamisega. Generatiivse tehisintellekti treenimiseks kogutakse andmeid tervikuna, kuid treenimise käigus võidakse need teosed ka osadeks teha.⁷³ Kuna andmete kogumist saab vaadata osana mudeli treenimisest, siis tuleb igakordselt hinnata, kas treeningandmetena kasutatav osa vastab autori enese intellektuaalsele loomingule ja on käsitatav originaalsena. Sama kehtib ka sisendandmete kasutamise ning valminud mudeli hilisema kasutamise puhul.

Mudeli eeltreenimisel on oluline luua mudelites ka arusaam maailmas toimuvatest sündmustest, mistõttu võib mudeli arendaja kasutada treeningandmetena tihtipeale ajalehtedes olevaid artikleid.⁷⁴ Näiteks on Ameerika Ühendriikides pöördunud ajakirjandusväljaanne The New York Times kohtusse põhjusel, et OpenAI on oma arendatavates mudelites ChatGPT ning DALL-E kasutanud ajakirjanike autoriõigusega kaitstud artikleid mudelite treeningandmetena. Hagejate väitel hõlmab võimalik rikkumine miljoneid nende loodud artikleid.⁷⁵ Kõnealuses kohtuasjas ei ole veel lahendini jõutud, kuid see ilmestab taaskord seda, kui laiaulatuslikud võivad olla reprodutseerimisõigusega seotud rikkumised.⁷⁶

Reprodutseerimisõigust puudutavad probleemid on seotud ka mudeli peenhäälestamisega. Lähtudes eelnevalt käsitletust, võib peenhäälestamine toimuda nii treenimise etapis oleva mudeli kui ka juba valmis mudeliga. Peenhäälestamine on vajalik eriti olukordades, kus soovitakse mõne spetsiifilise kunstniku või illustraatori stiili jäljendavat generatiivset

⁷¹ Rosati, E. Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790. Oxford University Press 2021, lk 35-38 ja lk 72-73.

⁷² C-5/08, *Infopaq International*, p 38.

⁷³ Margoni, T., Kretschmer, M. A deeper look into EU Text and Data Mining Exceptions: Harmonisation, Data Ownership and the Future of Technology. GRUR International 8/2022, lk 24-25.

⁷⁴ Vt selle kohta peatükk 1.2.

⁷⁵ *The New York Times v. OpenAI, Inc.* Case 1:23-cv-11195, p 2.

⁷⁶ Laialdase rikkumise illustreerimiseks võib tuua ühe täiendava võrdlusena interneti kasutajate arvu, mis on veidi üle 5 miljardi, mis tähendab, et selles andmekogus kasutati rohkem pilte, kui on maailmas interneti kasutajaid. See näitab ka, et autoriõiguste võimalik rikkumine võib olla toime pandud isegi rohkemate isikute suhtes.

tehistaipe luua. Sellises olukorras luuakse eraldiseisev andmekogu, kuhu kopeeritakse soovitud kunstniku või illustraatori teosed ja peenhäälestuse käigus muudetakse mudel nende teostega tuttavamaks, mille tulemusel suudab mudel täpsemalt jäljendada peenhäälestamiseks kasutatud autori teoseid. See võib aga käesoleva magistritöö autori arvates samuti kujutada reprodutseerimisõiguse kasutamist, kuna andmekogu luuakse sarnaselt mudeli treenimisele, kasutades andmekraabet.

Ka valmis mudeli kasutamisel võidakse kasutada reprodutseerimisõigust. Näiteks on õiguskirjanduses kirjeldatud olukorda, kus mudeli kasutaja kasutab sisendandmetena mõnda autoriõigusega kaitstud teost.⁷⁷ Taolisel juhul on tegemist reprodutseerimisõiguse teostamisega, kuna mudeli kasutaja kopeerib mudelisse teose sisendandmetena ning mudel kasutab sisendandmeid lisaks treeningandmetele väljundi loomisel. Reprodutseerimisõigust kasutab sellises olukorras mudeli kasutaja, kes on mudeli arendajast või eelnimetatud kolmandast isikust eraldiseisev isik. Samuti on töö ühe võrdlusriigi, Ameerika Ühendriikide kohtuasjad tõstatanud küsimuse, kas reprodutseerimisõiguse teostamine võib toimuda ka mudeli väljundi loomise etapis. Üheks selliseks on California osariigi kohtuasi, milles hagejad väitsid, et tehnoloogiaettevõtte OpenAI, Inc. on kopeerinud andmekraabe abil tuhandeid autoriõigusega kaitstud teoseid, sh raamatuid.⁷⁸ Kohtuasjas väitsid hagejad, et ChatGPT suudab luua väljundandmetena autori raamatutest väga täpseid kokkuvõtteid. Eeltoodu puudutab seega eelkõige olukordasid, kus mudel ei loo oma väljundina uut teost, vaid reprodutseerib teose, mille õigused kuuluvad kolmandale isikule. Näiteks on eelnevas peatükis kirjeldatud ajaleheartiklite puhul viimastes võimekamates mudelites esinenud probleem liiga hea meeldejätmisvõimega.⁷⁹ Osad suured mudelid on võimelised mitte ainult leidma mustreid ja korrapärasusi treening- ning sisendandmete seast, vaid ka sõnasõnaliselt meelde jätma ning hiljem väljundina looma treeningandmetes esinenud tekste sisuliselt algsel kujul.⁸⁰ Sõltumata sellest, et selline mudeli eriline meeldejätmise võimekus on pigem harva esinev nähtus, kujutab

⁷⁷ Guadamuz, A. (viide 18), lk 120-124.

⁷⁸ Paul Tremblay jt v. OpenAI, Inc., No. 23-cv-3223, p 41. Selles kohtuasjas tegi kohus 12.02.2024 otsuse, millega jättis osaliselt hagi rahuldamata põhjusel, kuna hagejad ei suutnud tõendada, et ChatGPT loodud väljundandmete ja hagejate teoste vahel oleks oluline sarnasus (*substantial similarity*).

⁷⁹ Vt meeldejätmise (*memorization*) kohta nt Emanuilov, I. Margoni, T. Memorisation in generative models in EU copyright law: an interdisciplinary view. – Kluwer Copyright Blog, 26.03.2024. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2024/03/26/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/> (04.04.2024).

⁸⁰ Rabin, M. R. I. jt. Memorization and Generalization in Neural Code Intelligence Models. Information and Software Technology, Elsevier 2021, lk 4.

selline olukord endast reprodutseerimisõiguse kasutamist.⁸¹ Eraldiseisev küsimus on seotud sellega, kas treeningandmed sisalduvad mudelis ka pärast selle treenimist. Õiguskirjanduses on märgitud eeltoodu osas, et valmis mudel, sealhulgas mudeli kaalud, ei sisalda teoseid, mis vastavad autoriõiguslikult kaitse saava teose eeldustele, muuhulgas objektiivsuse vormi nõudele.⁸² Autor nõustub eeltoodud seisukohaga, kuivõrd mudeli kaalud sisaldavad kirjeldust kasutatud treeningandmetest. Mudeli kaaludes olevate treeningandmete puhul ei ole täidetud AutÕS § 4 lg-s 2 sisalduv objektiivsuse kriteerium. Mudeli kaaludes kirjeldatu ei ole käsitatav „piisavalt täpselt ja objektiivselt identifitseeritava“ väljendusena. Eelnevast järelduvalt saab reprodutseerimisõiguse kasutamisenä AutÕS § 13 lg 1 p 1 tähenduses käsitada seda, kui teoseid kopeeritakse mudeli andmekogusse selle algse arendaja poolt, et kasutada neid treeningandmetena mudeli treenimiseks või siis, kui hiljem kasutab algse mudeli arendaja või kolmandast isikust arendaja treeningandmeid mudeli peenhäälestamiseks. Reprodutseerimisega on tegemist ka juhul, kui mudeli kasutaja sisestab mudelisse sisendandmetena autoriõigusega kaitstud teoseid. Samuti ei ole välistatud, et reprodutseerimine leiab aset olukorras, kus mudel taasloob väljundina autoriõiguslikult kaitstud teose sõltumata sellest, et mudeli kaalud ei sisalda treeningandmetena kasutatud teoseid.

Lisaks AutÕS § 13 lg 1 p-s 1 sätestatud reprodutseerimisõigusele tuleb hinnata, kas treening- ja sisendandmete reprodutseerimine võib endast kujutada ka töötlemisõiguse teostamist AutÕS § 13 lg 1 p 5 tähenduses. AutÕS § 13 lg 1 p 5 järgi on autoril õigus teost töödelda, sh teha teosest erinevaid kohandusi ja teost muuta. Õiguskirjanduses on hinnatud, kas mudeli treenimisel kasutatavad kaalud võiksid kujutada endast tuletatud teost (AutÕS § 35 lg 1), mis tähendaks ühtlasi ka seda, et algset treeningandmetena kasutatavat kaitstud teost on töödeldud.⁸³ Sellisel juhul oleks teose vahetuks töötlejaks eelkõige algse mudeli arendaja, kuid peenhäälestamisel võib selleks olla ka mõni kolmandast isikust arendaja või mudeli kasutaja.⁸⁴ Eelnevast nähtuvalt ei sisalda mudeli kaalud treeningandmetena kasutatavaid teoseid algsel kujul, vaid üksnes juhiseid, kuidas autoriõigusega kaitstud teost taasluua, s.o juhiseid, mille põhjal suudab mudel luua treeningandmetel põhinevaid väljundandmeid. Sellest johtuvalt ei toimu ka treeningandmete kasutamise käigus mudeli arendamisel või peenhäälestamisel treeningandmete

⁸¹ Mudeli meeldejätmise võimekuse kohta vt Emanuilov, I. Margoni, T. (viide 79).

⁸² Vt ka Margoni, T., Kretschmer, M. (viide 73), lk 18-20.

⁸³ Lee, K. jt. Talkin' Bout AI Generation: Copyright and the Generative-AI Supply Chain (July 27, 2023). Forthcoming, *Journal of the Copyright Society* 2024, lk 54.

⁸⁴ Riigikohtu praktikas on selgitatud töötleja olemust, kes võib töötlemisõigust rikkuda. Töötlemisõigust võib RK seisukoha järgi rikkuda isik, kes teost vahetult töötleb. Vt selle kohta RKTKo 2-16-8751/184, p 18.

töötlemist. Seega ei töödelda mudeli treeningandmeteks olevaid kaitstud teoseid ning ei rikuta ka AutÕS § 13 lg 1 p-s 5 nimetatud õigust. Samasugune käsitlus ei kehti aga kasutaja mudelisse lisatud sisendandmete puhul, kuivõrd koos sisendandmete mudelisse lisamisega on võimalik kasutajal anda käsklus, milles seisneb juhis sisendandmetena esitatud teose muutmiseks või ümber kujundamiseks. Samas võidakse töötlemisõigust kasutada sisendandmete kasutamisel, kui mudeli kasutaja annab koos sisendandmete lisamise osana käskluse teost töödelda.⁸⁵

1.3.1.2 Treeningandmete üldsusele kättesaadavaks tegemine

Kuigi mudeli arendamiseks kasutatava treeningandmetest koosneva andmekogu avalikustamist esineb praktikas pigem harva, ei ole sellise olukorra tekkimine välistatud, mistõttu hinnatakse järgnevalt, kas mudeli arendamiseks loodud andmekogu avalikustamine võib kujutada üldsusele kättesaadavaks tegemise õiguse kasutamist. Kuna mudel suudab eelnevast nähtuvalt väljundandmetena taastada ka kasutatud kaitstud teoseid sisuliselt algse kujul, siis analüüsib autor lisaks AutÕS § 13 lg 1 p 9¹ sätestatud õiguse võimalikku teostamist.

Teose üldsusele kättesaadavaks tegemise õigus on sätestatud AutÕS § 13 lg 1 p 9¹ ning selle kohaselt on autoril õigus teha teos üldsusele kättesaadavaks selliselt, et isikud saavad teoseid kasutada enda poolt individuaalselt valitud kohas ja ajal.⁸⁶ Üldsusele kättesaadavaks tegemise õiguse kasutamise küsimus võib tekkida olukorras, kus mudeli arendaja teeb treenimiseks kasutatava andmekogu, näiteks pildi- või tekstiandmekogu, pärast mudeli valmimist kättesaadavaks, et kolmas isik saaks andmekogu mõne teise mudeli treenimisel kasutada. Üldsusele kättesaadavaks tegemise õiguse oluliseks eelduseks on esmalt see, et mõni isik, näiteks mudeli arendaja, teeb teose kättesaadavaks ning teiseks, et teos oleks tehtud kättesaadavaks üldsusele. Riigikohus on oma praktikas varasemalt selgitanud, et AutÕS § 13 lg 1 p-s 9¹ sätestatud üldsusele kättesaadavaks tegemise puhul on esmalt vajalik hinnata seda, millisel viisil teos üldsusele on edastatud ning kas üldsusel on võimalus kaitstud teostele ka ligi

⁸⁵ Kõnealune teema tekitab ka eraldiseisva õigusliku küsimuse, kas mudeli väljundandmeid tuleks käsitada tuletatud teosena AutÕS § 35 lg 1 tähenduses. Kuivõrd käesolev töö keskendub treening- ja sisendandmete kasutamise õiguspärasuse hindamisele, ei käsitla autor väljundandmetega seotud õiguslikku küsimust pikemalt selle töö raames. Mudeli loodud väljundandmete kui võimaliku tuletatud teose kohta vt nt Sobel, B. L. W. Artificial Intelligence's Fair Use Crisis. 41 Colum. J.L. & Arts 45, 2017, lk 16; Guadamuz, A. (viide 18), lk 13-16; Gervais, D. J. AI Derivatives: The Application to the Derivative Work Right to Literary and Artistic Productions of AI Machines. Seton Hall Law Review Vol. 53, 2022, lk 24.

⁸⁶ Vt ka infoühiskonna direktiivi artikkel 3 lg 1, mis on AutÕS § 13 lg 1 p 9¹ aluseks.

pääseda.⁸⁷ Riigikohus on selgitanud täiendavalt, et kättesaadavaks tegemise puhul ei oma tähtsust, kas üldsus päriselt ka teostele ligipääsemise võimalust kasutas.⁸⁸ Kuivõrd mudeli treenimiseks kasutatavad andmekogud võidakse teha kättesaadavaks avalikult internetis või mõne muu andmekandja vahendusel, siis nendele võivad isikud ligi pääseda oma valitud kohas ja ajal. Seega kättesaadavaks tegemise hindamisel piisab asjaolust, et mudeli arendaja teeb pärast mudeli valmimist kaitstud teoseid sisaldava andmekogu kättesaadavaks näiteks internetis. Eraldiseisev küsimus on see, kas tegemist on üldsusega.⁸⁹ Eelduslikult on andmekogud kättesaadavaks tehtud suuremale hulgale kolmandatele isikutele, kuivõrd andmekogu avaldamise üheks eesmärgiks võib olla ka teenida täiendavat tulu, mille suurus võib sõltuda muuhulgas sellest, kui paljudele isikutele andmekogu avaldatakse. Seega on üldsuse kriteerium täidetud, kui mudeli arendaja avaldab andmekogu piisavalt suurele hulgale isikutele, mitte üksnes mõnele. Üldsusega ei ole tegemist siis, kui mudeli arendaja teeb andmekogud kättesaadavaks näiteks mudeli kasutajale. Lisaks tuleb hinnata, kas tegemist on uue üldsusega, st teose õiguste omaja ei osanud sellise üldsusega arvestada.⁹⁰ Ka uue üldsuse kriteerium on täidetud, kuivõrd andmekogud sisaldavad eeltoodust nähtuvalt hulgaliselt andmeid, mis on kogutud internetist ja kogumine toimub muuhulgas ka sellistelt lehekülgedelt, millele ei pruugi olla igal isikul juurdepääsu, st andmekogudes sisalduvad kaitstud teosed ei pruugi olla kõik vabalt kättesaadavad. Lisaks ei pruugi õiguste omaja poolt oma teost avaldades antud luba hõlmata olukorda, et tema teost kasutatakse mudeli treenimiseks ning hiljem edastatakse andmekoguna uute mudelite treenimise jaoks. Seega teostatakse õiguste omaja õigust teose üldsusele kättesaadavaks tegemisele, kui andmekogu, mis sisaldab kaitstud teoseid, tehakse mudeli arendaja poolt hiljem internetis kättesaadavaks.

Lisaks sellele, et andmekogud sisaldavad autoriõigusega kaitstud teoseid, võivad need sisaldada ka hüperlinke. Sellisel juhul ei kopeerita mudeli treenimiseks kasutatavaid teoseid mudelisse ega moodustata seeläbi eraldi andmekogu, vaid andmekogu koosneb hüperlinkidest,

⁸⁷ RKTko 2-14-5664, p 17.1.

⁸⁸ Samas. Vt ka viidatud kohtuasi EKO C-306/05, *SGAE*, ECLI:EU:C:2006:764, p 43.

⁸⁹ Üldsuse puhul on tegemist määratlemata arvu isikutega, keda on palju. Vt C-135/10, *SCF*, EU:C:2012:140, p 84 ja EKO C-265/16, *VCAST v. RTI SpA*, EU:C:2017:913, p 47. Lisaks on Euroopa Kohus selgitanud, et üldsuse puhul tuleb arvestada „nii nende isikute arvu, kellele sama teos on üheaegselt kättesaadav, kui ka seda, mitmele neist on see teos üksteise järel kättesaadav“. Viimase kohta vt C-263/18, *Tom Kabinet*, EU:C:2019:111, p 68 ja varasem lahend C-610/15, *Stichting Brein*, EU:C:2006:764, p 37.

⁹⁰ Uue üldsuse kohta vt nt C-597/19, *Mircom*, EU:C:2020:1063, p 57 ja C-392/19, *VG Bild Kunst*, EU:C:2020:696, p 32.

mis viivad autoriõigusega kaitstud teosteni.⁹¹ Küsimus andmekogu üldsusele kättesaadavaks tegemise õiguse kasutamisest tekib olenemata sellest, kas andmekogu sisaldab autoriõigusega kaitstud teoseid või üksnes hüperlinke, mis viivad autoriõigusega kaitstud teoseni. Üldsusele kättesaadavaks tegemise õigust hüperlinkide kontekstis on selgitanud ka Euroopa Kohus. Euroopa Kohus leidis, et hüperlingi teel kasutaja juhatamist veebilehele, kus autoriõigusega kaitstud teos on juba eelnevalt kättesaadavaks tehtud, saab käsitada teose kättesaadavaks tegemisena.⁹² Lisaks teose kättesaadavaks tegemisele on oluline ka see, et teos tehakse kättesaadavaks üldsusele. Eelnevalt viidatud EK lahendis märgiti, et tegemist ei olnud uuele üldsusele kättesaadavaks tegemisega, kuna teos oli juba eelnevalt mujal veebilehel kättesaadavaks tehtud ning seega oleks soovijad saanud külastada ka veebilehte, mis oli juba varasemalt teose üldsusele kättesaadavaks teinud.⁹³ Seega kasutatakse üldsusele kättesaadavaks tegemise õigust hüperlinkide kättesaadavaks tegemisel sarnaselt muudele kaitstud teostele üksnes juhul, kui teos on tehtud kättesaadavaks uuele üldsusele.

Eelkirjeldatud hüperlinkidest koosnevate andmekogude, näiteks LAION andmekogude puhul, on õiguskirjanduses lahknevaid seisukohti seoses sellega, kas sellise andmekogu kasutamine mudeli treeningandmetena võib endast kujutada autori üldsusele kättesaadavaks tegemise õiguse kasutamist. Osad õigusteadlased on leidnud, et hüperlinkide koondamine andmekogusse mudeli treenimise eesmärgil kujutab endast üldsusele kättesaadavaks tegemist.⁹⁴ Töö autor nõustub sellega aga üksnes osaliselt, kuivõrd andmekogu koondab hüperlinkidena ka juba selliseid teoseid, mis on üldsusele varasemalt kättesaadavaks tehtud ning sellisel juhul ei ole tegemist ka üldsusele kättesaadavaks tegemisega. Olukord on erinev, kui teos on tehtud varasemalt üldsusele kättesaadavaks ilma õigusliku aluseta.⁹⁵ Hüperlinkidest koosneva andmekogu kasutamine mudeli treenimiseks võib kujutada endast seega tegevust, mis sarnaneb veebilehekülje sirvimisega, mille puhul ei ole tegemist üldsusele kättesaadavaks

⁹¹ Näiteks LAION 5b andmekogu (edaspidi ka LAION andmekogu) võidakse kasutada mudeli treenimiseks. Vt Beaumont, R. A new era of open large-scale multi-modal datasets. 31.03.2022. Kättesaadav: <https://laion.ai/blog/laion-5b/> (07.01.2024).

⁹² EKO C-466/12 *Nils Svensson jt v Retriever Sverige AB*, ECLI:EU:C:2014:76, p 30-32.

⁹³ Samas, p 39-41.

⁹⁴ Vt selle kohta nt Guadamuz, A. (viide 18), lk 7-8; Lee, K. jt., lk 34-35 (viide 83). Käesoleva töö kirjutamise hetkel on Saksamaal käimas ka LAION andmekoguga seoses autoriõigusi puudutav vaidlus, kus vaidluse all on infoühiskonna direktiivi artiklite 2 ja 3 võimaliku rikkumise hindamine. Selle kohta vt ka Guadamuz, A. Photographer sues LAION for copyright infringement. 05.05.2023. Kättesaadav: <https://www.technollama.co.uk/photographer-sues-laion-for-copyright-infringement> (07.01.2024).

⁹⁵ Sarnasele järeldusele on jõutud ka õiguskirjanduses, vt nt Justin K. *The Right of Communication to the Public in EU Copyright Law*. Bloomsbury Publishing 2020, lk 16-17.

tegemisega.⁹⁶ Kui aga link on lisatud teosele, mis ei ole üldsusele kättesaadavaks tehtud autori nõusolekul, siis on tegemist üldsusele kättesaadavaks tegemise õiguse kasutamisega.⁹⁷

Küsimus üldsusele kättesaadavaks tegemise õiguse kasutamise kohta võib tõusetuda ka juhul, kui mudel suudab meelde jätta sisuliselt algsel kujul treeningandmetena kasutatud autoriõigusega kaitstud teoseid ja neid ka mudeli kasutajale väljundandmetena luua. Isegi kui mudeli kaaludes sisalduvat võiks käsitada autoriõigusega kaitstud teosena, siis ei oleks tegemist üldsusele kättesaadavaks tegemise õiguse kasutamisega. Kättesaadavaks tegemisena on võimalik AutÕS § 13 lg 1 p 9¹ tähenduses käsitada seda, kui mudel loob kasutaja käskluse peale väljundandmed. Kuigi mudeli kasutaja pääseb nendele väljundandmetele ligi sõltumata sellest, kus ta viibib, ei saa mudeli kasutajat käsitada üldsusena, kuivõrd üldsusena saab käsitada suurt hulka inimesi.⁹⁸ Kuna mudeli kasutajat ei saa käsitada sellisel juhul üldsusena, siis ei ole tegemist ka üldsusele kättesaadavaks tegemise õiguse teostamisega ja seega ei teosta mudeli arendaja autori hinnangul ka üldsusele kättesaadavaks tegemise õigust.

Eeltoodu järgi ei too üldjuhul mudeli treenimine ega mudeli enda kättesaadavaks tegemine kaasa mudeli treeningandmetes sisalduvate teoste üldsusele kättesaadavaks tegemist, kuivõrd treeningandmeid ei tehta kättesaadavaks. Üldsusele kättesaadavaks tegemisega oleks tegemist üksnes juhul, kui mudeli arendaja teeb treeningandmed kättesaadavaks pärast mudeli treenimist ning need andmekogud sisaldavad autoriõigusega kaitstud teoseid või hüperlinke, mis viivad teosteni, mis on tehtud kättesaadavaks autori nõusolekuta, mis ei ole aga treenimise seisukohalt vajalik ega tavapärase muuhulgas ärisaladusega seonduvatel põhjustel.

1.3.2. Autori isiklike õiguste kasutamine

1.3.2.1. Treening- ja sisendandmete autorsus

Juhul, kui autoriõigusega kaitstud teost on kasutatud treeningandmetena mudeli treenimiseks mõeldud andmekogus, kuid mudel loob väljundandmeid ilma õiguste omajale viitamata, võib tõusetuda probleem autorsuse kui ühe autorile kuuluva isikliku õiguse võimalikust

⁹⁶ Näiteks kasutab LAION oma andmekogus hüperlinke, mis viivad selliste fotograafiateosteni, mis on varasemalt kättesaadavaks tehtud mõnel muul veebileheküljel, kuhu need on lisatud teoste algsete autorite poolt. Selliste hüperlinkidest koosneva andmekogude kasutamine ei kujuta eelduslikult endast ka üldsusele kättesaadavaks tegemise õiguse rikkumist.

⁹⁷ Selle kohta vt EKo C-160/15 *GS Media BV v. Sanoma Media Netherlands BV and Others*, ECLI:EU:C:2016:644, p 41. Kohus rõhutas selles otsuses, et kui lingi lisamise taga peitub majandusliku kasu eesmärk, siis tuleb eeldada, et isik teadis, et autori nõusolekut ei ole.

⁹⁸ Vt selle kohta käesoleva töö viide 89.

kasutamisest. Kuigi õiguskirjanduses ei ole generatiivse tehisintellekti treenimise käigus isiklike õiguste võimalikku teostamist varasemalt süvitsi käsitletud, siis on erinevad huvigrupid nentunud, et isiklike õigustega, eelkõige õigusega au ja väärkuse kaitsele ning autorsusega seonduvaid probleeme peab rohkem analüüsima.⁹⁹ Seetõttu on mudeli treeningandmete seisukohast põhjendatud esmalt hinnata autori isiklikest õigustest õigust teose autorsusele (AutÕS § 12 lg 1 p 1). Erinevalt autorile kuuluvatest varalistest õigustest on õigus teose autorsusele autori isikust lahutamatu ega ole üle antav, mistõttu on seda võimalik kasutada litsentsilepingu alusel.¹⁰⁰ Autori isiklike õiguste kasutamine ilma litsentsita kujutab üldjuhul mõne AutÕS §-s 12 sätestatud autorile kuuluva õiguse teostamist.¹⁰¹ Eeltoodu tõttu ei sisalda kolmandas peatükis olev vaba kasutuse erandite analüüs autori isiklike õigusi, isegi kui alljärgneva analüüsi tulemusel selgub, et mudelite treenimine, peenhäälestamine või kasutamine võib kujutada endast mõne autori isikliku õiguse teostamist.

Õigus teose autorsusele on sätestatud AutÕS § 12 lg 1 p-s 1. Selle järgi on teose autoril õigus üldsuse ees esineda ning ühtlasi nõuda tunnustamist teose loojana. Nimetatud säte annab õiguste omajale õiguse esineda teose autorina. Teose kasutamisel on vajalik viidata nii autori nimele kui ka isikule, et üldsusel oleks võimalik tuvastada, kes on teose loonud. Eeltoodu kontekstis tuleb hinnata, kas olukord, kus treeningandmetena kasutatakse kaitstud teoseid, kuid ei tooda välja, missuguseid teoseid on mudeli arendamisel kasutatud ning samuti ei viidata väljundandmete loomisel kasutatud teostele, võib kujutada endast AutÕS § 12 lg 1 p-s 1 sätestatud õiguse teostamist.¹⁰² Sarnaselt mudeli treenimiseks ja peenhäälestamiseks kasutatavate andmekogude konfidentsiaalsena hoidmisele, on ka andmekogudes sisalduvad teosed hoitud saladuses ja mudeli arendajad või ka kolmandad isikud, kes kasutavad mudeli peenhäälestamiseks mõnda andmekogu, jätavad treeningandmetena kasutatavate teoste autorid

⁹⁹ Vt nt Max Planck Institute for Innovation and Competition. Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate. Max Planck Institute for Innovation and Competition Research Paper No. 21-10, 2021, lk 12.

¹⁰⁰ Vt VÕS § 368. Erinevalt varalistest õigustest, on isiklike õigusi võimalik nende olemusest tulenevalt üle anda üksnes lihtlitsentsiga, mitte aga ainulitsentsiga. Vt Varul, P. jt. Võlaõigusseadus II. Komm vlj. Tallinn: Juura 2019, § 368, p 3.3.2.

¹⁰¹ Õiguskirjanduses on leitud, et üksnes mõned isiklikud õigused on litsentsitavad, sh ei ole näiteks õigus autorsusele käsitatav õigusena, mida oleks võimalik litsentseerida. Praktikas on aga tavaline, et autori isiklikud õigused litsentseeritakse. RK praktikas on ka viimast seisukohta toetav lahend nt RKTko 3-2-1-67-07, p 17. Vt ka Rosentau, M. Intellektuaalse omandi õigused infotehnoloogias. Autori varalised õigused. Juridica 10/2010, lk 751; Rosentau, M. Intellektuaalse omandi õigused infotehnoloogias. Autori isiklikud õigused. Juridica 9/2007, lk 660.

¹⁰² Ka õiguskirjanduses on käsitletud just nimetatud õiguse võimalikku teostamist treeningandmete kasutamisel. Vt Matulionyte, R. Can AI Infringe Moral Rights of Authors and Should We Do Anything About it: An Australian Perspective. Law, Innovation and Technology 2023, lk 7-19. Selguse huvides märgib autor, et väljundandmete loomisel teosele viitamine on vajalik üksnes juhul, kui treeningandmetena on kasutatud teost piisaval määral.

mainimata.¹⁰³ Sinna alla ei lange juhtumid, kus andmekogu tehakse pärast treenimist kättesaadavaks.¹⁰⁴ Seega võib mudelite treenimisel ja peenhäälestamisel toimuda AutÕS § 12 lg 1 p-s 1 nimetatud õiguse kasutamine. Kuigi autorite isiklike õiguste rikkumist ei ole õiguskirjanduses laialdaselt käsitletud, on ühes artiklis jõutud sarnasele seisukohale, mille järgi on võimalik tuvastada autori autorsuse õiguse kasutamine ja seeläbi võimalik õiguse *prima facie* rikkumine ainuüksi seetõttu, et mudelite treenimiseks kasutatavaid teoseid ei ole võimalik reeglina kindlaks teha, sest mudelite arendajad ei avalikusta, missuguseid teoseid on mudeli treenimiseks kasutatud.¹⁰⁵ Olukorras, kus mudel loob vastuse (väljundi) statistiliste tõenäosuste ja andmete mustrite põhjal, ei ole võimalik allikaviiteta kontrollida ka mudeli usaldusväärsust ja läbipaistvust. Läbipaistvuse probleemi esinemisele mudeli treeningandmete kasutamisel on viidatud ka õiguskirjanduses.¹⁰⁶ Viidatud probleemist tulenevalt sisalduvad ka kavandatavas tehisintellekti määruses täiendavad nõuded autoriõigusega kaitstud teose kasutamisele treeningandmetena.¹⁰⁷

Lisaks autori mainimata jätmisele võib tähendada AutÕS § 12 lg 1 p-s 1 nimetatud õiguse kasutamist ka see, kui mudeli hilisemal kasutamisel autorile ei viidata.¹⁰⁸ Autorile viitamata jätmise võib toimuda eelkõige juhul, kui mudel loob kasutaja käskluse põhjal väljundandmed, kasutades seejuures mudeli kasutaja lisatud sisendandmeid. Autorile viitamata jätmise tähendaks juba AutÕS § 12 lg 1 p-s 1 nimetatud õiguse absoluutset rikkumist. Kuigi olemasolev kohtupraktika ning õiguskirjandus ei võimalda võtta ühest seisukohta, siis tulenevalt AutÕS § 12 lg 1 p 1 sõnastusest ning selle sätte eesmärgist võidakse magistritöö autori hinnangul vähemalt esmapilgul mudeli treenimise ja peenhäälestamise etapis ning hilisemal kasutamisel kasutada teose autori õigust autorsusele, kuna mudeli arendajad ja kolmandad isikud jätavad mudeli arendamisel kaitstud autorid nimetamata ning autoritele viitamata.

¹⁰³ Selle kohta vt käesoleva töö peatükk 1.1.

¹⁰⁴ Seda olukorda on täpsemalt käsitletud käesoleva töö peatükis 1.3.

¹⁰⁵ Matulionyte, R. (viide 102). Läbipaistvusega seonduvat analüüsib autor lähemalt töö alapeatükis 2.3.

¹⁰⁶ Lucchi, N. ChatGPT: A Case Study on Copyright Challenges for Generative AI Systems. *European Journal of Risk Regulation* 2023, lk 16-17. Sageli kutsutakse seda ka mudeli nn mustaks kastiks (*black box*), kus mudeli treeningandmetena on kasutatud hulgaliselt andmeid, millest mudeli loojal puudub täpsem ülevaade. Vt nt Wischmeyer, T., Rademacher, T. *Regulating Artificial Intelligence*. Springer 2020, lk 375-379.

¹⁰⁷ Vt tehisintellekti määruse eelnõu artikkel 53.

¹⁰⁸ Viitamata jätmist on käsitletud ka Ameerika Ühendriikide kohtuasjades. Vt nt *The Intercept Media, Inc. v. OpenAI, Inc. et al.* Case 1:24-cv-01515, p 40.

1.3.2.2. Au ja väärkuse kaitse treeningandmete kasutamisel

Treeningandmete kasutamise seisukohast on õiguskirjanduses arutletud juhtumi üle, kus mudel suudab luua algupärase teosega sarnaseid väljundeid, sisaldades algteose autori loomingule iseloomulikke tunnuseid, mis võimaldavad hilisema mudeli loodud väljundi seostamist originaalse teose autoriga ka au ja väärkuse kaitse kontekstis.¹⁰⁹ AutÕS § 12 lg 1 p-s 5 sätestab autori õiguse au ja väärkuse kaitsele, mille järgi on autoril õigus vaidlustada muuhulgas erinevaid moonutusi ja teisi ebatäpsusi teoses, mis võivad kahjustada autori au ja väärkust.¹¹⁰ Autori au ja väärkuse hindamisel tuleb hinnata eelkõige autori subjektiivset taju, mis esineb, kui autor tunneb, et tema au ja väärkust, sh mainet, kahjustatakse. Lisaks peab subjektiivne taju olema ka objektiivselt põhjendatud.¹¹¹

Kuigi küsimus autori au ja väärkuse kaitse vaatepunktist võib tõusetuda ka juhul, kui kasutaja käskluse tulemusena loob mudel väljundandmed, mis on seostatavad algse teose autoriga ning mida võidakse kasutada mõnes muus kontekstis, siis käesoleva töö vaatest on põhjendatud analüüsida eelkõige võimalikku AutÕS § 12 lg 1 p-s 5 sätestatud õiguse kasutamist treeningandmete kontekstis. Õiguskirjanduses on au ja väärkuse kaitse näitlikustamiseks viidatud olukorrale, kus algupärast teost kasutatakse mõnes poliitilises kampaanias, mis ei kattu originaalteose autori vaadetega.¹¹² Teine levinud näide puudutab juhtumeid, kus teosest eemaldatakse mõni teose oluline osa. Kuna mudelit õpetatakse treeningandmete abil läbi mitme erineva kihi, millest igaüks tunneb ära mõne kindla omaduse treeningandmetest, siis kasutatakse mudeli treenimisel ja peenhäälestamisel sageli ka üksnes selliseid osasid teostest, mis on vajalikud konkreetse kihi jaoks.¹¹³ Näiteks ainulaadse ja silmapaistva taustaga portreefotost võidakse eemaldada taust selleks, et mudel õpiks tundma mõnda fotol leiduvat konkreetset objekti või kujutist. Selline kihiti mudeli treenimine võib olla vajalik näiteks juhul, kui mudelit treenitakse näotuvastuse otstarbel. Õiguskirjanduse kohaselt võib teosest olulise

¹⁰⁹ Autori au ja väärkuse kaitset seoses mudeli arendamisega on õiguskirjanduses analüüsinud nt Matulionyte, R. (viide 102), lk 12-13. Vt ka Matulionyte, R. The (forgotten) moral rights of the age of AI. – Kluwer Copyright Blog, 07.02.2022. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2022/02/07/the-forgotten-moral-rights-in-the-age-of-ai/> (30.03.2024).

¹¹⁰ Teose moonutamise alla kuuluvad tegevused, mis hõlmavad endast muuhulgas teose struktuurilist muutmist. Vt selle kohta Cheng-Davies, T. Honour in UK Copyright Law is is Not „A Trim Reckoning“ – Its Impact on the Integrity Right and the Destruction of Works of Art. Oxford Journal of Legal Studies Volume 36 Issue 2 2016, lk 272–303.

¹¹¹ Samas.

¹¹² Matulionyte, R. (viide 102), lk 12.

¹¹³ Vt selle kohta ka käesoleva töö peatükk 1.1.

osa eemaldamine mudeli treenimiseks ja peenhäälestamiseks tähendada algupärase kaitstud teose moonutamist.¹¹⁴ Autori hinnangul ei kujuta aga eelkirjeldatud juhtumid endast AutÕS § 12 lg 1 p-s 5 sätestatud õiguse kasutamist. Teose kasutamine mudeli treeningandmetena mõnes teises kontekstis ei muuda otseselt algupärast teost, mistõttu ei ole tegemist ka teose moonutamisega. Samuti ei saa au ja väärikut kahjustavaks pidada seda, kui teosest eemaldatakse oluline osa ning teist osa kasutatakse mudeli treenimiseks või peenhäälestamiseks.¹¹⁵ Kuigi eeltoodust tulenevalt võib olla küsitav, kas mõni konkreetne osa teosest väljendab autori enese intellektuaalset loomingut ja on käsitatav autoriõigusega kaitstud teosena, ei saa taoline kaitstud teose treeningandmetena kasutamine kujutada endast autori au ja väärikuse kahjustamist isegi siis, kui tegemist on kaitstud teosest osa kasutamisega.

Magistritöös püstitatud esimesele uurimisküsimusele, missugustel juhtudel ning ulatuses kasutatakse autoriõigusega kaitstud teoste kasutamisel treening- ja sisendandmetena autorile kuuluvaid ainuõigusi, on võimalik eeltoodu pinnalt vastata, et autori varalistest õigustest kasutatakse AutÕS § 13 lg 1 p-s 1 sätestatud reprodutseerimisõigust mudeli eeltreenimisel ning peenhäälestamisel mudeli algse arendaja või kolmandast isikust arendaja poolt. Samuti võidakse kasutada reprodutseerimisõigust mudeli hilisemal kasutamisel kasutaja poolt. AutÕS § 13 lg 1 p-s 5 sätestatud töötlemisõiguse kasutamine võib aset leida, kui mudeli kasutaja käskluse ja sisendandmete põhjal loob mudel väljundandmeid. Mudeli arendamine ja hilisem kasutamine ei pruugi rikkuda AutÕS § 13 lg 1 p-s 9¹ sätestatud üldsusele kättesaadavaks tegemise õigust. Küll aga võib see toimuda pärast mudeli treenimist juhul, kui andmekogu sisaldab hüperlinke, mis viivad teosteni, mis on tehtud kättesaadavaks ilma autori nõusolekuta ning andmekogu tehakse kättesaadavaks. Samuti, kui mudeli arendaja teeb üldsusele kättesaadavaks andmekogu, mis sisaldab kaitstud teoseid, sõltumata sellest, et teos oli varasemalt selle õiguste omaja nõusolekul juba mujal avaldatud. Mudeli treenimisel, peenhäälestamisel ja kasutamisel rikutakse autorile kuuluvatest isiklikest õigustest õigust autorsusele (AutÕS § 12 lg 1 p 1).

Käesolevas peatükis ilmnes ka, et üheks probleemkohaks mudelite treenimisel on treeningandmete läbipaistvus, mistõttu võib tõusetuda küsimus ka autorile kuuluvate õiguste teostamise võimalikkusest. Kuna kehtivast õigusest ei tulene mudeli arendajatele kohustust

¹¹⁴ Matulionyte, R. (viide 102), lk 13-14. Kõnealuse artikli autor on toonud näitena näotuvastusmudeli Clearview AI treenimise, mille andmekogu koosnes fotodest, kus oli alles jäetud kindel osa näost ning ülejäänud fotost eemaldatud. Vt ka Max Planck Institute for Innovation and Competition (viide 99), lk 12.

¹¹⁵ Vt selle kohta käesoleva töö peatükk 1.3.2.

teha kättesaadavaks teavet selle kohta, missuguseid andmeid, sh autoriõigusega kaitstud teoseid, on mudeli arendamiseks kasutatud, siis ei pruugi praegune õigusraamistik võimaldada autoritel oma ainuõigusi piisavalt tõhusalt tagada. Samuti võib olla õiguste omajal raskendatud autoriõiguse rikkumist tõendada kohtumenetluses. TsMS § 230 lg-st 1 tulenevalt on üldjuhul autoril autoriõiguse tõendamise kohustus.¹¹⁶ Kuigi kohtupraktika järgi võib teatud juhtudel kehtida erisus, mille järgi pööratakse tõendamiskoormis ümber ning mille korral peab hageja asemel tõendama kostja, et viimane ei ole autoriõigusega kaitstud teost kasutanud, siis on ebaselge, kas tõendamiskoormise ümber pööramine võiks olla võimalik ka mudelite arendamisel.¹¹⁷ Näiteks võib see olla küsitav olukorras, kus mudeli väljundandmetena loodud teos ei sarnane oluliselt õiguste omaja loodud teosega.

Kuigi mitmed riigid, nende hulgas näiteks ka võrdlusriikideks olevad Ameerika Ühendriigid ja Ühendkuningriik on pakkunud mittesiduvaid regulatiivseid suuniseid ja põhimõtteid mudelite arendamiseks, siis nende mittesiduva iseloomu tõttu ei loo need otseselt mudelite arendajatele täiendavaid kohustusi.¹¹⁸ Just treenimiseks kasutatavad andmekogud võimaldavad niivõrd võimekaid mudeleid treenida, mistõttu on treeningandmete kasutus tihedalt seotud ka mudelite arendajate äriliste huvidega.¹¹⁹ Seetõttu ei ole mittesiduvad regulatiivsed suunised ning põhimõtted piisavad selleks, et mudelite arendajad tagaksid läbipaistvuse mudelite arendamisel. Läbipaistvuse puudumine tähendab aga, et olukorras, kus mudeli arendaja kasutab treeningandmetena autoriõigusega kaitstud teoseid, võib õiguste omajal olla keeruline tõendada kohtus ainuõiguste rikkumist ning kasutada õiguskaitsevahendeid. Eeltoodud põhjustel ei pruugi kehtiv õigus käesoleva töö autori hinnangul tagada õiguste omajatele õiguste teostamiseks piisavat kaitset.

¹¹⁶ Tsviilkohtumenetluse seadustik. – RT I, 22.03.2024, 7.

¹¹⁷ Tõendamiskoormise ümber pööramiseks peab õiguste omaja näitama esmalt ära asjaolud, millest ilmneb, et tema teost on hilisema teose loomisel kasutatud. Lisaks peab olema piisavalt põhjendatud, et kaitstud teose kasutamine, näiteks mudeli arendamisel, on tõenäoline. RK praktikas on peetud õigustatuks ümber pööratud tõendamiskoormise kasutamist olukorras, kus hageja on kostjale loonud varasemalt teose ning näidanud ära olulised sarnasused oma loodud teose ning kostja kasutatud teose vahel. Vt RKTko 2-16-8751/184, p 24.2. Vt ka selles otsuses viidatud kohtuasjad RKTko 2-16-9016/34, p 11.2 ja RKTko 3-2-1-173-12, p 17.

¹¹⁸ Nt Ühendkuningriigis House of Lords. Large language models and generative AI. 1st Report of Session 2023-24. 02.02.2024. Kättesaadav: <https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/54.pdf> (20.03.2024); Ameerika Ühendriikides nt Select Committee on Artificial Intelligence of the National Science and Technology Council. The National Artificial Intelligence Research and Development Strategic Plan: 2023 Update. May 2023, lk 22. Kättesaadav: <https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf> (15.03.2024).

¹¹⁹ Andmekogude tähtsuse kohta vt käesoleva töö peatükk 1.1.

2. TREENING- JA SISENDANDMETE KASUTAMINE KEHTIVA NING KAVANDATAVA ÕIGUSE KOHASELT

2.1. Treening- ja sisendandmete kasutamine vaba kasutuse erandite alusel

2.1.1. Treening- ja sisendandmete ajutine või juhuslik reprodutseerimine

Olukorras, kus autoriõigusega kaitstud teost kasutatakse mudeli treening- või sisendandmetena ilma teose autori nõusolekuta, on autori nõusoleku kõrval võimalik ühe õigusliku alusena tugineda seaduses sätestatud vaba kasutuse erandile, mis välistavad autorile kuuluvate varaliste õiguste rikkumise.¹²⁰ Käesolevas peatükis selgitab autor esmalt välja, kas eelnevas peatükis käsitletud ainuõiguste kasutamise kaasaegne rikkumise võib välistada mõni AutÕS-s olev vaba kasutamise erand.¹²¹ Seeläbi jõuab autor vastuseni, kas ja millistel tingimustel on autoriõigusega kaitstud teose kasutamine generatiivse tehisintellekti treening- ja sisendandmetena õiguspärane. Lisaks kehtiva õiguse analüüsile analüüsib autor ka seda, kuidas mõjutab kavandatav tehisintellekti määrus kaitstud teose treening- ja sisendandmete kasutamise õiguspärasust. Selle tulemusena saab teha järelduse selle kohta, kas kehtiv õigus ja kavandatav tehisintellekti määrus tagavad tasakaalu nii autori huvide kui ka generatiivse tehisintellekt innovatsiooni vahel.

Generatiivse tehisintellekti arendamise seisukohast tuleb esmalt hinnata infoühiskonna direktiivi artikkel 5 lg-st 1 tulenevat vaba kasutuse erandit teose ajutiseks reprodutseerimiseks, mis on üle võetud Eesti õigusesse AutÕS §-ga 18¹.¹²² Säte paikneb AutÕS IV peatüki teises jaos ning selle järgi ei ole vaja maksta autori õiguste omajale õiguste teostamise eest tasu. AutÕS §-s 18¹ sätestatu esmaseks kumulatiivseks eelduseks on see, et teose reprodutseerimine oleks ajutine või juhuslik. Samuti peab see toimuma tehnilise protsessi lahutamatu ja olulise osana. Lisaks peab sellise tegevuse eesmärgiks olema teose edastamise vahendamine

¹²⁰ Vt nt AutÕS § 17.

¹²¹ Generatiivse tehisintellekti olemusest tulenevalt on analüüsist välja jäetud muuhulgas näiteks tsiteerimise erand (AutÕS § 19 lg 1 p 1), kuivõrd selle erandi kohaldumise eeldused, mh autori ja allika ära näitamine, motiveeritud maht ning teose mõtte moonutamine, ei ole täidetud ja seetõttu pikemat analüüsi konkreetne erand ei vaja.

¹²² Ühendkuningriigi autoriõiguste regulatsioonis on sarnaselt reguleeritud võimalus reprodutseerida ajutiselt või juhuslikult teoseid. Sisult vastab Ühendkuningriigis olev erand Euroopa Liidus oleva infoühiskonna direktiivi artiklis 5 sätestatud erandile, mistõttu sarnaneb see ka AutÕS §-ga 18¹. Vt CDPA § 28A.

kolmandate isikute vahel selliselt, et teost oleks võimalik seaduspäraselt kasutada.¹²³ Viimaks ei tohi sellisel toimingul olla iseseisvat majanduslikku eesmärki.¹²⁴

AutÕS § 18¹ lg 1 esmane eeldus viitab sellele, et reprodutseerimise käigus tekkinud koopiad tuleks koheselt hävitada, st reprodutseerimine peab olema ajutine või juhuslik. Generatiivse tehisintellekti arendamiseks kasutatud teoseid tohib seega kasutada üksnes lühikesel ajavahemikul ja teoseid ei tohiks säilitada pikkadeks aastateks selleks, et neid saaks ka edaspidi kasutada mudeli arendamiseks. Generatiivse tehisintellekti arendamisel on oluline vaadata tervikuna treenimise etappe ning seda, kus andmed paiknevad. Nimelt võivad treeningandmed olla talletatud kõvakettale (või ka pilvsalvestusruumi), muutmälusse või protsessoritesse (GPU) enne seda, kui need on mõnel kujul valminud mudeli kaaludes.¹²⁵ Lisaks muutmälus olevatele ja sinna salvestatud andmetele tuleb vaadelda ka mudeli arendamise järgmist sammu. Muutmälust hävitatakse treeningandmed pärast seda, kui need on jõudnud protsessoriteni ning protsessoritest omakorda pärast seda, kui need jõuavad mudeli kaaludeni.¹²⁶ Eelnevalt kirjeldatust nähtuvalt sisaldavad mudeli kaalud üksnes juhiseid autoriõigusega kaitstud teose taasloomiseks, mistõttu ei sisalda mudeli kaalud autoriõigusega kaitstud teoseid, mis vastavad autoriõiguslikult kaitse saava teoste eeldustele.¹²⁷ See tähendab, et kuigi mudeli treenimiseks ja peenhäälestamiseks võidakse kasutada treeningandmeid, mis on käsitatavad autoriõigusega kaitstud teostena, siis pärast mudeli valmimist enam mudel treeningandmeid ei sisalda. Kuivõrd mudel, sealhulgas mudeli kaalud, sisaldab üksnes õpetust selle kohta, kuidas teost saab uuesti luua, siis on treeningandmete liikumise etappe eraldiseisvalt hinnates võimalik jõuda järeldusele, et kaitstud teoste reprodutseerimine võib toimuda teatud mudeli treenimise etappides juhuslikult või ajutiselt.¹²⁸ Ka õiguskirjanduses on leitud, et mudeli arendamisel kasutatakse treeningandmeid üksnes ajutiselt, kuna pärast mudeli valmimist treeningandmed kustutatakse.¹²⁹

¹²³ Selle osas vt nt EKo C-527/17, *Filmspeler*, EU:C:2017:300.

¹²⁴ Vt ka C-5/08, *Infopaq International*, p 54, kus EK on selgitanud täpsemalt infoühiskonna artiklis 5 olevat erandit artikkel 2 kontekstis. Sisult vastasid lahendis välja toodud eeldused infoühiskonna direktiivi artiklis 5 loetletud eeldustele.

¹²⁵ Vt ka Joonis 2.

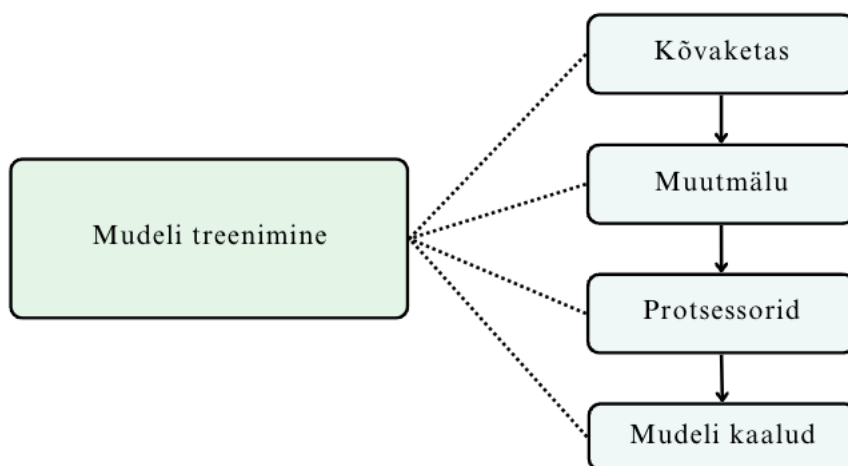
¹²⁶ Sandiumenge, I. Copyright Implications of the Use of Generative AI. Universitat Pompeu Fabra. SSRN 2023, lk 12-13. Kättesaadav: <http://dx.doi.org/10.2139/ssrn.4531912> (04.01.2024); Ørstavik, I. B. (viide 70), lk 17-18.

¹²⁷ Vt käesoleva töö peatükk 1.3.

¹²⁸ Geiger, C., Frosio, G., Bulayenko, O. jt. Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? *International Review of Intellectual Property and Competition Law* 2018, lk 821-822.

¹²⁹ Sandiumenge, I. (viide 126), lk 13-14; Ørstavik, I. B. (viide 70), lk 18.

Samas on õiguskirjanduses argumenteeritud, et kaitstud teoste reprodutseerimine ei toimu juhuslikult või ajutiselt, kuivõrd teoseid säilitatakse reeglina pikema ajaperioodi vältel, st ajaperiood, mil treeningandmed jõuavad kõvakettalt mudeli kaaludeni, on pikk.¹³⁰ Õiguskirjanduses on ka märgitud, et reprodutseerimine toimub mudeli tehnilise protsessi lahutamatu ja olulise osana ning treeningandmed kustutatakse pärast mudeli valmimist, mistõttu on reprodutseerimine ajutine.¹³¹ Autor nõustub, et ühest küljest on võimalik väita, et mudelite treenimisel on tegemist ajutise treeningandmete reprodutseerimisega, kuna pärast mudeli valmimist treeningandmed kustutatakse ning säilitatakse mudeli kaaludes üksnes juhisteana. Selle puhul tuleb siiski arvestada, et treeningandmeid reprodutseeritakse küll ajutiselt, kui vaadata igat treenimise etappi eraldiseisvalt, kuid neid etappe tervikuna vaadates ei ole võimalik tõenäoliselt järeldada, et tegemist oleks ajutise või juhusliku reprodutseerimisega. Seda esiteks põhjusel, et treeningandmete kõvakettalt mudeli kaaludeni jõudmine võib toimuda pikema perioodi vältel ning teiseks ei pruugi mudeli arendajad treeningandmeid pärast mudeli valmimist ära kustutada nende väärtuslikkuse tõttu.¹³²



Joonis 2. Treeningandmete liikumine mudelis.

EK on varasemas praktikas hinnanud infoühiskonna direktiivi artikkel 5 lg-st 1 tulenevat vaba kasutuse erandit, eelkõige seda, kas tegemist on ajutiste koopiatega olukorras, kus internetikasutaja sirvib veebilehekülgi.¹³³ EK seisukoha järgi on sellises olukorras tegemist ajutiste koopiatega, kuna arvutiekraanil olevad koopiad kustutati kohe, kui internetikasutaja

¹³⁰ Lee, K. jt (viide 83), lk 61-63.

¹³¹ Margoni, T. (viide 3), lk 17-19. Vt ka Sandiumenge, I. (viide 126), lk 15-16.

¹³² Selle kohta vt ka käesoleva töö peatükk 1.1.

¹³³ EKo C-360/13, *Meltwater*, ECLI:EU:C:2014:1194, p-d 47-50.

lahkus leheküljelt ning arvuti vahemällu salvestatud koopiad kustusid samuti. Täiendavalt on EK samas otsuses rõhutanud, et ajutise koopiaga on tegemist üksnes juhul, kui see on vajalik tehnoloogilise protsessi nõuetekohaseks toimimiseks ning tingimusel, et koopiad kustutatakse automaatselt, kui tehnoloogiline funktsioon on täidetud.¹³⁴ Nimetatud kohtuasjas ei nõustunud EK menetlusosalise väitega, et koopiad ei ole ajutised, kuna nende kustutamine sõltub inimese sekkumisest, st inimene peab ise veebileheküljelt lahkuma. Lõppkokkuvõttes leidis EK, et arvutiekraanil olevad koopiad on juhuslikud.¹³⁵ Sarnase seisukoha võib autori arvates võtta ka mudelite treenimisel. Mudeli treeningandmeteks olevad kaitstud teosed kustuvad kas automaatselt või kustutatakse mudeli arendaja poolt pärast igat mudeli treenimise etappi. Mudel kasutab väljundandmete loomisel üksnes mudeli kaaludes olevaid juhiseid, mis aga ei sisalda autoriõigusega kaitstud teoseid. See tähendab, et viidatud kohtuasjas esitatud tingimus, mille järgi koopiad tuleb kustutada automaatselt pärast tehnoloogilise funktsiooni täitumist, on täidetud ka mudeli treenimisel.

Lisaks on õiguskirjanduses vaidlusi tekitanud küsimus, kas mudeli treenimisel tehtavatel koopiatel on iseseisev majanduslik tähtsus AutÕS §-s 18¹ tähenduses.¹³⁶ Selle eelduse täitumist võib toetada eelkõige see, et teatud juhtudel ei pruugi koopiad iseseisvat majanduslikku tähtsust omada. Generatiivsele tehisintellektile on omane see, et seda tüüpi mudelit treenitakse suurte andmekogude peal. Kui vaadata andmekogudes sisalduvaid koopiaid iseseisvalt, siis selliselt iga üksik koopia majanduslikku tähtsust ei oma. Õiguskirjanduses leviva seisukoha järgi on majanduslik tähtsus valmis treenitud mudelil.¹³⁷ Sellest järelduvalt omavad tekste, pilte ja muud sisu loovate mudelite puhul tähtsust miljardid teostest moodustunud toonenid ning üksikul koopial iseseisvat majanduslikku tähtsust ei ole. Seega on majandusliku tähtsuse eeldus täidetud, kui mudel on veel treenimise faasis ega ole valminud.

Eeltoodust johtuvalt on võimalik mudeli treeningandmete kasutamisel tugineda AutÕS §-s 18¹ sätestatud erandile üldjuhul siis, kui mudelit treenitakse, mistõttu ei ole võimalik erandile tugineda pärast mudeli kasutusse laskmist, st sisendandmete kasutamisel. Samuti tuleb AutÕS §-s 18¹ sätestatud erandile tuginemisel arvestada, et see ei pruugi kohalduda kõigile mudeli

¹³⁴ Samas, p-d 29-32. Vt ka EKm C-302/10, *Infopaq International A/S v. Danske Dagblades Forening*, ECLI:EU:C:2012:16, p-d 30 ja 31.

¹³⁵ C-360/13, *Meltwater*, p-d 49-50.

¹³⁶ „Iseseisev majanduslik tähtsus“ kui ühte eeldust on käsitletud mudeli arendamise perspektiivist Isaac Sandiumenge. Vt Sandiumenge I. (viide 126), lk 16-18. Lisaks vt Guadamuz. A. (viide 18), lk 7-9.

¹³⁷ Guadamuz, A. (viide 18), lk 10; Ørstavik, I. B. (viide 70), lk 17-18. Inger B. Ørstavik on seisukohal, et mudeli treenimisel koopiade tegemine ei oma iseseisvat majanduslikku tähtsust, kuid valmis mudeli kolmandatele isikutele ärilisel eesmärgil pakkumine omab, mistõttu ei lange see infoühiskonna direktiivi art 5 lg 1 erandi alla.

treenimise etappidele, mistõttu ei välista AutÕS §-s 18¹ sätestatud erand vastutust autoriõigusliku rikkumise eest mudeli treenimisel tervikuna.

2.1.2. Treeningandmete reprodutseerimine teadustöös teksti- ja andmekaeve eesmärkidel

Generatiivse tehisintellekti treening- ja sisendandmete õiguspärase kasutamise hindamisel on olulised reprodutseerimisõigust puudutavad DSM-direktiivist tulenevad erandid. Tegemist on AutÕS §-ga 19¹ ja §-ga 19² üle võetud teksti- ja andmekaeve eesmärkidel teose kasutamist lubavate vaba kasutuse eranditega. Sarnaselt eelnevalt käsitletud AutÕS §-le 18¹ sisalduvad nii AutÕS § 19¹ kui ka § 19² AutÕS IV peatüki teises jaos. Ka nende erandite puhul ei hüvitata autori varaliste õiguste teostamist õiguste omajale.¹³⁸

Üheks levinud mudeli treenimise meetodiks on teksti- ja andmekaeve¹³⁹ (edaspidi ka TDM) tehnoloogia kasutamine, mille käigus muuhulgas kasutab mudeli arendaja veebilehekülgedel oleva sisu talletamiseks automatiseeritud lahendusi.¹⁴⁰ Õiguskirjanduses viidatakse DSM-direktiivist tulenevatele teksti- ja andmekaeve eranditele mudeli treenimise õiguspärasuse hindamisel.¹⁴¹ DSM-direktiivis sisalduva definitsiooni kohaselt on teksti- ja andmekaeve „automatiseeritud analüüsimeetod, millega analüüsitakse digivormingus tekste ja andmeid, et saada teavet muuhulgas mustrite, suundumuste ja korrelatsioonide kohta“. Samasugune definitsioon on üle võetud ka AutÕS § 17¹ lg-s 1. AutÕS § 19¹ lg 1 järgi on teose reprodutseerimine autori nõusolekuta ja autoritasu maksmiseta lubatud teksti- ja andmekaeve eesmärkidel. AutÕS §-le 19¹ sarnane teksti- ja andmekaevet võimaldav säte on olemas ka Ühendkuningriigi autoriõiguste regulatsioonis.¹⁴² Ameerika Ühendriikides ei ole sätestatud

¹³⁸ Autoriõiguse seaduse muutmise seadus (autoriõiguse direktiivide ülevõtmine) 368 SE, lk 28. Kättesaadav: <https://www.riigikogu.ee/tegevus/eelnoud/eelnou/d3d07943-9d1c-4ebe-94a4-8ae1ebdf7a68> (12.12.2023). Vt ka DSM-direktiiv, artikkel 3 ja 4 ning osaliselt ka DSM-direktiivi art 7 lg 1, kuivõrd vastav säte on Eesti õigusesse üle võetud AutÕS § 19¹ lg 4 näol.

¹³⁹ Ingliseelse termini *text and data mining* asemel kasutatakse sageli ka terminit *content mining*. Täpsemalt selle kohta vt Cocoru, D. Boehm, M. An analytical review of text and data mining practices and approaches in Europe. Policy recommendations in view of the upcoming copyright legislative proposal. Openforum, Europe 01.05.2016, lk 5. Kättesaadav: <https://openforumeurope.org/wp-content/uploads/2016/05/TDM-Paper-Diana-Cocoru-and-Mirko-Boehm-1.pdf> (02.02.2024); Rosati, E. (viide 71), lk 26. Käesoleva töö autori arvates on termini *content mining* kasutamine mõnes mõttes isegi täpsem, kuna teksti- ja andmekaeve teostamisel kogutakse laiemas tähenduses sisu (sh erinevaid kaitstud teoseid), mis hõlmab ka erinevaid tekste ja andmeid.

¹⁴⁰ Thompson, A. D. What's in my AI? A Comprehensive Analysis of Datasets used to Train GPT. 2022, lk 6. Kättesaadav: <https://lifearchitect.ai/whats-in-my-ai/> (02.02.2024).

¹⁴¹ Vt nt Andres Guadamuz (viide 18), lk 19-21.

¹⁴² Vt CDPA § 29A.

konkreetseid vaba kasutuse erandeid ning autoriõiguslikku kaitset omavat teost on seaduspäraselt võimalik kasutada, kui täidetud on nn õiglase kasutamise doktriini eeldused.¹⁴³

DSM-direktiivi kommenteerivas väljaandes on jagatud TDM protsess kolme-etapiliseks.¹⁴⁴ Esmaseks sammuks võib pidada juurdepääsu sisule. Teise sammuna on käsitletud sisu eraldamist ja kopeerimist algallikast, mis vastab oma olemuselt andmekraapele, ning kolmandana teksti ja andmete kaevandamist. Nende etappide juurde kuulub ka nn tokeniseerimine.¹⁴⁵ Tokeniseeritud andmetele saab rakendada erinevaid andmeanalüüsi meetodeid, millega tuvastada mustreid.¹⁴⁶ Nii TDM-i kui ka masinõppe käigus kasutatakse andmete analüüsi mustrite, suundumuste ja korrelatsioonide kohta teabe saamiseks. Selleks, et automatiseeritud analüüsimeetod vastaks TDM definitsioonile, peab olema võimalik selle tehnoloogia abil saada teavet näiteks mustrite, suundumuste ja korrelatsioonide kohta. Samas viitab TDM definitsioonis sisalduv sõna „muuhulgas“ sellele, et tegevus vastab TDM definitsioonile ka olukordades, kus tegemist ei ole üksnes mustrite, suundumuste ja korrelatsioonidega. TDM definitsiooni alla võib kuuluda näiteks ka tegevus, kus analüüsitakse fakte või andmeid, millele autoriõiguslik kaitse ei laiene.¹⁴⁷ Kui tõlgendada TDM definitsiooni kitsalt, siis võiksid selle tegevuse alla kuuluda üksnes analüütilised meetodid (sh meetodid, mida kasutatakse üksnes analüütilise tulemi jaoks) ning sellest võiksid välja jääda toimingud, mis hõlmavad endast uue sisu tootmist. Tegemist on aga väga kitsa tõlgendusega ning õiguskirjanduses on siiski leitud, et TDM hõlmab reeglina endas ka teisi tegevusi. Õiguskirjanduse kohaselt võib definitsioon hõlmata ka generatiivse tehisintellekti väljundini jõudmisega seotud etappe.¹⁴⁸ Seega tulenevalt TDM definitsiooni sõnastusest, võiks sellega olla hõlmatud nii mudeli treenimine kui ka peenhäälestamine.

Teksti- ja andmekaeve teostamine on võimalik kõigi eelnevalt kirjeldatud autoriõigusega kaitstud teoste suhtes, millega võidakse mudelit treenida, kuivõrd AutÕS § 19¹ lg-ga 1 on hõlmatud autoriõigusega kaitstud teosed ning ka kaasnevate õiguste objektid, näiteks

¹⁴³ USC, 17, § 107 (b). Õiglase kasutamise (*fair use*) doktriini kaudu hindab kohus, kas teose kasutamine teatud eesmärgiks on õiglane. Õiglase kasutuse hindamisel lähtub Ameerika Ühendriikides kohus järgnevatest eeldustest: i. kasutuse eesmärk ja iseloom; ii. kaitstud teose iseloom; iii. kasutatud teoseosa maht ja olulisus; iv. kasutuse mõju teose potentsiaalsele turule või väärtusele. Vt ka Gillotte, J. L. Copyright Infringement in AI-Generated Artworks. UC Davis Law Review, Vol. 53 No. 5, 2020, lk 32-38.

¹⁴⁴ Rosati, E. (viide 71), lk 34-36. Vt ka Geiger, C., Frosio, G., Bulayenko, O. jt (viide 128), lk 816-818.

¹⁴⁵ Tokeniseerimise kohta vt lähemalt käesoleva töö alapeatükk 1.1.

¹⁴⁶ Rosati, E. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Technical Aspects. European Parliament: 2018, lk 38.

¹⁴⁷ Rosati, E. (viide 71), lk 32-33.

¹⁴⁸ Rosati, E. (viide 71), lk 37. Generatiivse tehisintellekti toimimise kohta vt peatükk 1.1.

andmebaasid.¹⁴⁹ AutÕS § 19¹ lg-s 1 sätestatud erandi õigustatud subjektiks on teadus- ja kultuuripärandiasutused. Teadusasutuseks loetakse AutÕS § 17¹ lg 2 kohaselt teadus- ja arendustegevuse korralduse seaduse (TAKS) §-s 3 nimetatud asutusi.¹⁵⁰ Selle sätte kohaselt võib teadusasutuseks olla juriidiline isik või asutus, mille põhitegevus on muuhulgas alusuuringute, rakendusuuringute või arendustegevuse läbiviimine või põhitegevusega kaasneva tegevuseks on teadmiste levitamine õpetamise, publitseerimise või tehnoloogiasirde kaudu.¹⁵¹ AutÕS § 17¹ lg 2 loetleb teadusasutustena veel omakorda ülikooli ja selle raamatukogu, teadusinstituudi või muu asutuse, mille peamine eesmärk on teha teadusuuringuid või tegeleda õppetööga. Seejuures peab teadusasutus tegelema mittetulunduslikul alusel või investeerima kogu kasumi uuesti oma teadusuuringutesse. Kolmanda alternatiivina on teadusasutusel võimalus ka täita oma avalikes huvides olevaid ülesandeid selliselt, et teadusuuringute tulemused ei ole soodustingimustel kättesaadavad asutuse üle otsustavat mõju omavale ettevõtjale. AutÕS § 17¹ lg 3 järgi on kultuuripärandi asutusteks muuhulgas avalik raamatukogu, muuseum või arhiiv. AutÕS § 17¹ lg 2 loetelu on otse üle võetud DSM-direktiivi artiklis 2 olevate mõistete „teadusorganisatsioon“ ja „kultuuripärandiasutus“ järgi.

AutÕS § 19¹ lg-s 1 sätestatud erand võimaldab ka teadus- ja kultuuripärandiasutuste koostööd erasektoriga. Õiguskirjanduses on leitud, et infoühiskonna direktiivi artiklis 3 olev TDM erand võib julgustada üldiselt erasektori ettevõtjaid, sh teadusasutusest väljaspool olevaid tehisintellekti arendajaid, tegema koostööd teadusasutustega.¹⁵² Sarnase seisukoha võib võtta ka artikli 3 järgi üle võetud AutÕS §-s 19¹ oleva erandi osas. Kuigi AutÕS § 19¹ lg 1 järgi võivad TDM erandile tugineda üksnes teadus- ja kultuuripärandiasutused, siis AutÕS § 17¹ lg 2 sisalduva teadusasutuse definitsiooni kohaselt ei pruugi olla välistatud koostöö ka erasektoriga.¹⁵³ Nimelt võivad teadusuuringute tulemused olla kättesaadavad ka ettevõtjale, kes ei oma otsustavat mõju teadusasutuse üle. Ka DSM-direktiivi kommenteerivas väljaandes

¹⁴⁹ Vt selle kohta käesoleva töö peatükk 1.2.

¹⁵⁰ Teadus- ja arendustegevuse korralduse seadus. – RT I, 19.03.2019, 92.

¹⁵¹ TAKS, § 3 lg 1 p-d 1 ja 2. Infoühiskonna direktiivi põhjenduspunkti 12 järgi peaksid teadusasutuse definitsiooni alla kuuluma ka teadusuuringutega tegelevad haiglad. Kuigi TAKS ega AutÕS otseselt haiglatele ei viita, võib ka haigla olla käsitatav teadusasutusena, kuna vastab AutÕS-s sätestatud teadusasutuse eeldustele.

¹⁵² Bonadio, E., Dinev, P., McDonagh, L. Can Artificial Intelligence Infringe Copyright? Some Reflections. Research Handbook on Intellectual Property and Artificial Intelligence 2022, lk 10. Vt sissejuhatus viide 18 andmepesu kohta.

¹⁵³ AutÕS § 17¹ lg 2 sätestab, et „/.../ teadusuuringute tulemused ei ole soodustingimustel kättesaadavad ettevõtjale, kellel on otsustav mõju sellise asutuse üle“. Selle sõnastuse järgi on autori arvates teadus- ja kultuuripärandiasutusel võimalik jagada teadusuuringute tulemusi ja andmekogusid ka selliste ettevõtjatega, kellel ei ole asutuse üle otsustavat mõju.

ei ole välistatud olukorrad, kus ettevõtjad saavad siiski ligi teadusuuringute tulemustele ja andmekogudele, mis põhinevad erinevatel autoriõigusega kaitstud teostel, kui neil ei ole otsustavat mõju teadusasutuse üle.¹⁵⁴ Selline tegevus läheb autori hinnangul TDM erandi alla, kui ettevõtjad arendavad mudelit teadustöö raames.

Erinev on olukord, kus erasektoris tegutsevad äriühingud tuginevad erandile põhjusel, et nad teevad koostööd teadus- või kultuuripärandiasutusega, kuid tegelikkuses ei toimu nendevaheline koostöö AutÕS § 19¹ lg-s 1 välja toodud eesmärgil. Näiteks ei pruugi olla autori arvates võimalik välistada, et erasektor kasutab ka oma mudeli arendamiseks edaspidi teadus- ja kultuuripärandiasutustelt saadud andmekogu. Samuti on õiguskirjanduses välja toodud olukord, kus teadus- ja kultuuripärandiasutused treenivad mudelit AutÕS § 19¹ lg-s 1 sätestatud erandile tuginedes, kuid kolmandad osapooled võivad kasutada mudelit hiljem ka oma äritegevuseks.¹⁵⁵ Õiguskirjanduses on läbivalt leitud, et sellised olukorrad võivad tekitada niinimetatud „andmepesu“ juhtumeid, kuivõrd seaduses ei ole selgesõnaliselt välistatud hilisemat teadus- ja kultuuripärandiasutuse loodud andmekogu või nende asutuste loodud mudeli ärilisel otstarbel kasutamist selliste ettevõtete poolt, kellel ei ole otsustavat mõju asutuse üle. Autori hinnangul ei hõlma AutÕS § 19¹ lg-s 1 sätestatud erand andmekogu kasutamist hiljem ärilisel otstarbel, kuivõrd andmekogu on loodud teadusuuringute raames.¹⁵⁶ Lisaks ei ole õiguskirjanduses võetud seisukohta, kas teadus- ja kultuuripärandiasutuse arendatud mudelit võib hiljem ärilisel otstarbel kasutada. Teadus- ja kultuuripärandiasutuse arendatud mudeli hilisem ärilistel eesmärkidel kasutamine võib olla võimalik juhul, kui algne mudeli arendamine oli kooskõlas AutÕS § 19¹ lg-s 1 sätestatuga, kuivõrd valmis mudel ei sisalda enam autoriõigusega kaitstud teoseid ning sellisel juhul ei ole välistatud valmis mudeli hilisem kasutamine ärilisel eesmärgil.¹⁵⁷

Võrdlusriigis, Ühendkuningriigis on niinimetatud akadeemilise andmepesu tõttu räägitud TDM erandite täiendamisest, mis võimaldaks lisaks reprodutseerimisõigusele ka teiste autori

¹⁵⁴ Rosati, E. (viide 71), lk 26; Rosati, E. Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity. *Asia Pacific Law Review* 2020, lk 16-17. Vt ka käesoleva töö sissejuhataste viide 18.

¹⁵⁵ Lisaks vt Guadamuz, A., lk 9 (viide 18).

¹⁵⁶ Selle teema kohta vt lisaks Baio, A. AI Data Laundering: How Academic and Nonprofit Researches Shield Tech Companies from Accountability. – *Waxy Blog*, 30.09.2022. Kätesaadav: <https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/> (03.04.2024); Yew, R.-J. Break It 'Til You Make It An Exploration of the Ramifications of Copyright Liability Under a Pre-training Paradigm of AI Development. *Proceedings of the Symposium on Computer Science and Law*, 3/2024, lk 67-68.

¹⁵⁷ Mudeli kaalude kohta vt lähemalt käesoleva töö peatükk 1.3, kus on hinnatud, kas valmis mudel võib sisaldada autoriõigusega kaitstud teoseid.

ainuõiguste piiramist.¹⁵⁸ Ühendkuningriigi autoriõiguste regulatsioon sätestab sarnaselt Eestile erandi teadustöö raames toimuvaks teksti- ja andmekaeveks ning võimaluse ajutiseks teose reprodutseerimiseks.¹⁵⁹ Ühendkuningriigi valitsuse 2023. aasta ettepanekust ilmneb, et 2022. aasta alguses sooviti muuta TDM erandit viisil, et see oleks lubatud mistahes eesmärgil (sh tehisintellekti arendamiseks ja äritegevuseks), seejuures andmata kirjeldatud *opt-out* võimalust õiguste omajatele.¹⁶⁰ Autoriõiguste omajate huve esindavate isikute survega loobuti sellest ideest.¹⁶¹ Käesoleva töö loomise hetkel ei ole autorile teadaolevalt võetud Ühendkuningriigis kindlaid meetmeid teksti- ja andmekaeve erandi muutmise osas.

Lisaks piirangule, et tegemist oleks teadus- ja kultuuripärandiasutusega, peab AutÕS § 19¹ lg 1 kohaselt olema teadus- ja kultuuripärandiasutusel teosele ka seaduslik juurdepääs. Kuigi AutÕS ei määratle, mida võib pidada seaduslikuks juurdepääsuks, on autoriõiguse seaduse muutmise eelnõu seletuskirjas, millega võeti üle DSM-direktiivi konkreetsed artiklid, viidatud DSM-direktiivi põhjenduspunktidele.¹⁶² DSM-direktiivi põhjenduspunkti järgi on seaduslik juurdepääs „sisule avatud juurdepääs poliitika või õiguste omajate ja teadusorganisatsioonide või kultuuripärandiasutuste vaheliste lepinguliste kokkulepete alusel, nagu abonemendid, või muudel seaduslikel viisidel“.¹⁶³ Samuti kuulub seadusliku juurdepääsu alla sisu, mis on veebis tasuta kättesaadav. Seega võib AutÕS 19¹ lg 1 sisalduvat seaduslikku juurdepääsu tõlgendada laialt ning selle alla võivad kuuluda esiteks olukorrad, kus juurdepääs sisule, mis muidu oleks piiratud, on antud õiguste omaja nõusolekul, näiteks avatud juurdepääsu poliitika või litsentsilepingu alusel. Samuti kuuluvad seadusliku juurdepääsu alla olukorrad, kus juurdepääs ei ole piiratud, näiteks juhul, kui teos on veebis tasuta kättesaadav.¹⁶⁴ See tähendab, et mudeli arendamisel on eeldus täidetud juhul, kui treeningandmetena kasutatakse andmekogusid, mis on omakorda koostatud selliselt, et andmekogus olevate andmete kasutamiseks on olemas õiguslik alus, näiteks mõni vaba kasutuse erand.¹⁶⁵

¹⁵⁸ Intellectual Property Office. The Government's Code of Practice on Copyright and Artificial Intelligence. 29.06.2023. Kättesaadav: <https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai> (06.03.2024).

¹⁵⁹ CDPA, §-d 28A, 29, 29

¹⁶⁰ The Government's Code of Practice on Copyright and Artificial Intelligence (viide 157).

¹⁶¹ Montagnon, R., Cho, S. UK Withdraws Plans For Broader Text And Data Mining (TDM) Copyright And Database Right Exception. – Mondaq, 07.03.2023. Kättesaadav: <https://www.mondaq.com/uk/copyright/1290378/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception> (06.03.2024).

¹⁶² AutÕS-i seletuskiri, lk 25.

¹⁶³ DSM-direktiivi põhjenduspunkt 14; AutÕS-i seletuskiri, lk 26.

¹⁶⁴ Samas. Vt ka Rosati, E. (viide 71), lk 28-30.

¹⁶⁵ Selle alla võivad kuuluda ka avatud juurdepääsuga (*open access*) andmed. Vt nt Kamocki, P, lk 123; DSM-direktiivi põhjenduspunkt 10; Rosati, E. (viide 71), lk 45 ja 76.

Täiendavalt on mudeli arendamise seisukohast põhjendatud hinnata, kui kaua on andmeid võimalik AutÕS § 19¹ lg-s 1 sätestatud erandile tuginedes säilitada. TDM erandile tuginedes on kaitstud teostest tehtud koopiaid võimalik säilitada seni kuni need on näiteks teadustöö tegemiseks vajalikud. AutÕS § 19¹ lg-s 1 sätestatud erand ei kohaldu juhul, kui ettevõtjal on teadusasutuse üle otsustav mõju. Otsustav mõju võib esineda olukorras, kus ettevõtja kuulub näiteks juhatusse ja on seega teadusasutuse struktuuri osa.¹⁶⁶ Sellest lähtudes on AutÕS §-s 19¹ sätestatud erand kohalduv juhul, kui tegemist on AutÕS §-s 17¹ ja TAKS §-s 3 välja toodud teadus- ja kultuuripärandiasutustega, kes tegelevad teksti- ja andmekaevega teadustöö eesmärgil. Muud juriidilised isikud, sealhulgas mudeli arendajad, kes ei kuulu teadus- ja kultuuripärandiasutuste määratluse alla, saavad küll teatud juhtudel ülikoolidega koostööd teha, kuid seda üksnes olukorras, kus neil puudub otsustav mõju teadusasutuse üle.

Eelnevast analüüsist järeldub, et mudeli arendamisel on võimalik tugineda reprodutseerimisõiguse teostamisel AutÕS § 19¹ lg-s 1 sätestatud erandile siis, kui mudeli arendamine on samuti otseselt seotud teadustöö eesmärgiga. Kõnealune erand võimaldab ka erasektori ja teaduasutuste vahelise koostöö. Samas, kui mudeli arendaja kasutab teadus- või kultuuripärandiasutuse AutÕS § 19¹ lg-s 1 sätestatud erandile tuginedes loodud andmekogu ärielistel eesmärkidel oma loodava mudeli arendamiseks, siis selline kasutus ei ole sätte eesmärgiga hõlmatud. Praegune õiguskord jätab siiski võimaluse, et teadustöö eesmärgil arendatud mudelit on võimalik kasutada hiljem ka ärielistel eesmärkidel mõne kolmanda osapoole poolt, kui mudel on arendatud algselt kooskõlas AutÕS § 19¹ lg-s 1 sätestatuga. Eeltoodu on aktuaalne ka mujal Euroopa Liidu liikmesriikides, mille TDM erandid põhinevad samuti DSM-direktiivil. Eeltoodud väite ilmestamiseks toob autor näitena DSM-direktiivi artiklil 3 põhinevad TDM erandid Prantsuse ning Soome kehtivast õigusest.¹⁶⁷ Prantsusmaa intellektuaalse omandi seaduses on DSM-direktiivi artiklist 3 tulenev TDM sõnastatud sarnaselt AutÕS § 19¹ lg-s 1 sätestatud erandile.¹⁶⁸ Samuti ei ole ärielist otstarbel valmis mudeli kasutus välistatud Soome autoriõiguse seaduses.¹⁶⁹ Eeltoodut arvestades võiks olla põhjendatud täpsustada kehtivas õiguskorras olevat teadustöö teksti- ja andmekaeve erandit

¹⁶⁶ Rosati, E. (viide 71), lk 45; Kumar, R jt. Training AI and Copyright Infringement: Where does the law stand? Indian Journal of Integrated Research in Law Volume II Issue I 2022, lk 11.

¹⁶⁷ Vt Prantsusmaa intellektuaalse omandi seadus (*Code de la propriété intellectuelle*), artikkel L122-5-3(II). Kättesaadav: https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006069414 (19.03.2024) ja Soome autoriõiguse seadus (*Tekijänoikeuslaki*), § 13b. Kättesaadav: <https://www.finlex.fi/fi/laki/ajantasa/1961/19610404> (19.03.2024).

¹⁶⁸ Prantsusmaa intellektuaalse omandi seadus, artikkel L122-5-3(II).

¹⁶⁹ Soome autoriõiguse seadus, 13b §.

selliselt, et oleks välistatud mudeli hilisem ärilisel otstarbel kasutamine juhul, kui käesolevast tööst ilmneb, et õiguste omajate huvid võiksid õigustada täiendavate piirangute kehtestamist. Vastasel juhul võib AutÕS § 19¹ lg-s 1 sätestatud erandi täiendamine eelnevalt soovitatud kujul pärssida liigselt innovatsiooni ja mudelite arendamist.

2.1.3. Treeningandmete reprodutseerimine väljaspool teadustööd teksti- ja andmekaeve eesmärkidel

Lisaks DSM-direktiivi artiklis 3 sätestatud erandile on treening- ja sisendandmete kasutamise juures asjakohane analüüsida ka artiklis 4 sätestatud AutÕS §-ga 19² üle võetud erandit. AutÕS § 19² võimaldab autori nõusolekuta ja autoritasu maksmiseta reprodutseerida teost teksti- ja andmekaeve eesmärkidel. Erinevalt AutÕS § 19¹ lg-s 1 sätestatust ei ole vajalik, et teksti- ja andmekaeve toimuks teadustöö eesmärgil.¹⁷⁰ Lisaks peab seejuures tegemist olema õiguspäraselt juurdepääsetava teosega. Võrreldes AutÕS § 19¹ lg-ga 1, ei ole AutÕS § 19² lg-s 1 õiguste subjektid piiratud ning sellele erandile saavad tugineda ka erasektoris tegutsevad juriidilised isikud.

AutÕS § 19² sõnastusest „ilma et see mõjutaks“ saab järeldada, et konkreetne säte ei ole vastuolus AutÕS §-s 19¹ sätestatuga, mistõttu saab nendele eranditele tugineda ka samaaegselt. Samamoodi sätestab DSM-direktiivi artikkel 4 lg 4, et artiklis 4 sätestatud erand ei piira artiklis 3 sätestatud erandi kohaldamist. Seega, kui AutÕS §-le 19¹ on võimalik tugineda teadusuuringutes ja AutÕS §-le 19² on võimalik tugineda väljaspool teadusuuringuid ning teadusuuringute käigus muutub näiteks uuringute läbiviija, siis saab autori arvates hiljem tugineda AutÕS §-le 19². Ühe näitena saab tuua olukorra, kus ettevõtja ei oma algselt otsustavat mõju teadusasutuse üle, kuid kui see muutub teadustöö käigus, võib ettevõtja hiljem tugineda AutÕS § 19¹ asemel AutÕS §-le 19².

AutÕS § 19² kohaselt eeldab sellele erandile tuginemine õiguspärasest juurdepääsust teosele. Sarnaselt seaduslikule juurdepääsule ei ole ka õiguspärase juurdepääsu kohta regulatsioonil konkreetset määratlust. Siiski on õiguskirjanduses välja toodud, et õiguspärase juurdepääsu alla võivad sarnaselt seaduslikule juurdepääsule kuuluda juhtumid, kus on saadud õiguste omajalt juurdepääs kaitstud teosele või tegemist on teosega, mis on tasuta veebis

¹⁷⁰ Ühendkuningriigis on teksti- ja andmekaeve lubatud üksnes mitteäriilistel eesmärkidel. Vt Bently, lk 260 (viide 20) ja CDPA § 29A.

kättesaadav.¹⁷¹ Nendest esimese all kuuluvad juhtumid, kus mudeli arendaja kasutab teost autori nõusolekul, näiteks varasemalt sõlmitud litsentsilepingu kaudu.

AutÕS § 19² lg 3 järgi võib reproduktsioone säilitada nii kaua, kui see on teksti- ja andmekaeve jaoks vajalik. Tegemist on erinevusega infoühiskonna direktiivi artikli 5 lg-s 1 sätestatud ja AutÕS §-ga 18¹ üle võetud ajutise reprodutseerimise vaba kasutuse piirangust. Selline lai sõnastus võib kahtlemata hõlmata õiguskirjanduse kohaselt ka püsivat kopeerimist erinevalt AutÕS §-s 18¹ sätestatud erandist.¹⁷² Mudeli treenimine ja peenhäälestamine koosnevad sellistest protsessidest, mille käigus analüüsitakse hulgaliselt tekste ja andmeid, mis võimaldab omakorda saada teavet erinevate mustrite, suundumuste ja korrelatsioonide kohta. Nende pinnalt on valmis mudel hiljem võimeline looma kasutajate käsklustel põhinevaid väljundandmeid. Mudeli treenimisel ja peenhäälestamisel liiguvad andmed lõpuks mudeli kaaludeni, mille järel andmed kustutatakse. Kuna eelnevalt on põhistatud, et teksti- ja andmekaeve hõlmab nii mudeli treenimist kui ka peenhäälestamist, siis järelikult säilitatakse treeningandmeid kui reproduktsioone nii kaua, kui see on teksti- ja andmekaeveks vajalik, mistõttu langeb see ka AutÕS § 19² lg-s 3 sätestatu alla.

Kaitstud teoste õiguste omajad saavad ka selgesõnaliselt välistada AutÕS §-s 19² sätestatud erandile tuginemise võimalikkuse ehk kasutada niinimetatud *opt-out* võimalust.¹⁷³ Küsitav on, et kuna AutÕS § 19² sisaldub AutÕS IV peatüki teises jaos, kus on võimalik erandit kasutada ilma õiguste omaja nõusoleku ja tasu maksmiseta, siis kas AutÕS §-s 19² lg-s 2 sisalduva *opt-out* võimaluse tõttu võiks konkreetne erand olla eraldiseisvas jaos. Samas on AutÕS §-s 19² lg-s 2 oleva *opt-out* võimaluse näol tegemist nn vaikimisi antava nõusolekuga, mille puhul nõusolekut eeldatakse juhul, kui õiguste omaja ei ole vastupidist väljendanud. DSM-direktiivist ega selle põhjenduspunktidest ei nähtu, kas õiguste omaja välistuse tegemine peaks toimuma *ex ante* või *ex post*.¹⁷⁴ Analüüsisides AutÕS § 19² lg-s 2 sätestatu olemust, võib olla autori hinnangul õiguste omajal võimalik teoste kasutamine teksti- ja andmekaeveks välistada nii *ex ante* kui ka *ex post*. AutÕS §-s 19² lg-st 2 tulenevalt on lubatud välistuse tegemine ka masinloetavate vahenditega veebis üldkättesaadavaks tehtud sisu puhul. Õiguste omajal on võimalik seega näiteks veebilehe kasutamise tüüptingimustesse lisada välistus, mille järgi ei

¹⁷¹ Rosati, E. (viide 71), lk 76-77.

¹⁷² Franceschelli, G., Musolesi, M. Copyright in generative deep learning. Cambridge University Press 05/2022, lk 17.7-17.6.

¹⁷³ AutÕS § 19² lg 2 ja DSM-direktiiv, artikkel 4 lg 3. Vt ka Ørstavik, I. B. (viide 70), lk 15.

¹⁷⁴ Õiguskirjanduses on käsitletud seda kui ühte võimalikku probleemi, mille osas puudub õiguselgus mudelite treenimisel. Vt nt Sandiumenge, I. (viide 126), lk 19-20.

ole mudelite treenimiseks andmekraabe lubatud. Sellisel juhul ei oleks andmete kraapimine lubatud ka sisu osas, mis ei ole kaitstud autoriõigustega.¹⁷⁵ Seega oleks mudeli arendamisel vajalik igakordselt kindlaks teha, kas kasutatavate teoste osas on õiguste omaja välistanud sõnaselgelt teksti- ja andmekraave võimaluse.¹⁷⁶ AutÕS §-s 19² lg-st 2 tuleneva välistuse tegemine on võimalik ka erinevate tehniliste vahendite abil, s.o veebilehe metaandmetesse või lähtekoodi näiteks robots.txt või ai.txt faili lisamisega.¹⁷⁷ Just robots.txt faili on varasemalt kaalutud ühtse tehnilise masinloetava vahendina, mille abil oleks võimalik kindlaks teha, kas veebilehel, kus teostatakse andmekraabet, on AutÕS § 19² lg-st 2 tulenevalt õigusvastane selline teksti- ja andmekraave, mille puhul on välistatud AutÕS §-s 19² sätestatud erandile tuginemine.¹⁷⁸ Kuivõrd mudelite andmekogud on suuremahulised, võimaldab robots.txt faili kasutamine analüüsida, kas veebilehtedel, kus andmekraabet treeningandmete kogumiseks teostatakse, on andmete kraapimine teksti- ja andmekraave eesmärkidel lubatud. Õiguskirjanduse kohaselt tagab faili robots.txt kasutamine seeläbi õiglase tasakaalu õiguste omajate ning mudeli arendajate huvide vahel.¹⁷⁹ 2024. aasta märtsikuu seisuga kasutab vaid 10% maailma populaarseimatest lehtedest võimalust läbi robots.txt faili teksti- ja andmekraape välistamist.¹⁸⁰

Õiguskirjanduses on täiendavalt kritiseeritud ühtse standardi puudumist, mida õiguste omajad saaksid kasutada ning mille kaudu õiguste kasutajad tuvastada, kas teose kasutus teksti- ja andmekraaveks on välistatud.¹⁸¹ Autori arvates on eelnev kriitika põhjendatud arvestades seda, et teksti- ja andmekraave on saanud mudelite arendamise oluliseks osaks ning ilma ühtsete

¹⁷⁵ Novelli, C. jt. Generative AI in EU Law: Liability, Privacy, Intellectual Property, Cybersecurity, 2024, lk 18-19.

¹⁷⁶ Selle kohta vt täpsemalt peatükk 2.3.

¹⁷⁷ Ducato, R., Strowel, A. Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to „Machine Legibility“. International Review of Intellectual Property and Competition Law 2018, lk 22-24. Robots.txt ja ai.txt failid annavad veebilehe kasutajale teabe selle kohta, kas sellel olevat sisu (sh pilte, tekste jms) saab kasutada mudelite treenimiseks. Need on kaks tüüpilist näidet olemasolevatest standarditest, kuid käesoleva töö kirjutamise hetkel on *opt-out* võimalust pakkuvaid standardeid veel rohkem. Vt ka Keller, P., Warso, Z. Defining best practices for opting out of ML training. – Open Future, 29.09.2023, lk 8. Kättesaadav: https://openfuture.eu/wp-content/uploads/2023/09/Best_practices_for_optout_ML_training.pdf (19.02.2024).

¹⁷⁸ Senftleben, M. AI Act and Author Remuneration – A Model for Other Regions? Institute for Information Law 2024, lk 10. Vt ka Hugenholtz, P. B. The New Copyright Directive Text and Data Mining Articles 3 and 4. – 24.07.2019, Kluwer Copyright Blog. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/> (20.02.2024).

¹⁷⁹ Dermawan, A. Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese „nonenjoyment“ purposes? The Journal of World Intellectual Property 2023, lk 58-59.

¹⁸⁰ Langley, H. Top websites block Google from training AI models on their data. – Yahoo News, 15.03.2024. Kättesaadav: <https://www.yahoo.com/tech/top-websites-block-google-training-194847790.html> (23.04.2024).

¹⁸¹ Vt nt Sandiumenge, I. (viide 126), lk 19-20.

tehniliste standardite võib olla ühest küljest tagada mudelite arendamise õiguspärasust ning teisalt õiguste omajate huvide kaitset.

Treeningandmete kogumisel, s.o andmete kraapimisel võib aga probleeme tekitada asjaolu, et puudub teave selle kohta, missuguseid autoriõigusega kaitstud teoseid on arendajad oma mudelites kasutanud. Kuigi on olemas mitmeid tehnilisi lahendusi, mis võimaldavad õiguste omajatel AutÕS § 19² lg-st 1 tulenevalt erandile tuginemist välistada, siis on ebaselge, kas ja kuidas mudelite arendajad seda järgivad, kuna reeglina ei avalda arendajad mudeli treenimiseks kasutatud treeningandmeid.¹⁸² See on omakorda lisaks eelpool kirjeldatud läbipaistvuse probleemidele üheks täiendavaks keerukuseks treeningandmete õiguspärasel kasutamisel, kuivõrd muudab pea võimatuks mudeli kasutajal ja õiguste omajal hinnata, kas treeningandmete seas leiduvad sellised kaitstud teosed, mis on saadud kooskõlas AutÕS § 19² lg-s 2 sätestatuga.

Nii AutÕS §-s 19² kui ka AutÕS §-s 19¹ sätestatud erandid kohalduvad üksnes reprodutseerimisõiguse rikkumise korral. Seetõttu ei ole õigusliku alusena võimalik tugineda nendele eranditele, kui mudeli arendamisel on rikutud mõnda muud autorile kuuluvat õigust, näiteks õigust üldsusele kättesaadavaks tegemisele.¹⁸³ Autori hinnangul ei oleks AutÕS-i muutmise põhjendatud, kuna eelnevast analüüsist nähtuvalt on üldsusele kättesaadavaks tegemise õiguse rikkumisega tegemist üksnes juhul, kui pärast mudeli treenimist avaldatakse mudeli treenimiseks kasutatav kaitstud teostest koosnev andmekogu.¹⁸⁴ Seega puudutab konkreetse õiguse rikkumine treeningandmete kasutamist pärast mudeli treenimist ega ole seotud mudeli arendamise ega valmimisega. See jätab ära ka vajaduse kehtestada vaba kasutuse erand, mis hõlmaks lisaks reprodutseerimisele ka üldsusele kättesaadavaks tegemise õigust. Nimetatud üldsusele kättesaadavaks tegemise õiguse rikkumise ületab see, kui mudeli arendaja ei avalikusta treeningandmete kasutamiseks mõeldud andmekogu. Kuigi vahel andmekogu koos mudeli arendamiseks kasutatud treeningandmetega siiski avaldatakse, ei ole see praktikas liiga sagedane.

Eelnevast nähtuvalt ei ole erinevalt Ühendkuningriigist, teises võrdlusriigis, Ameerika Ühendriikides teksti- ja andmekaeve erandit. Selle asemel on võimalik õiguskirjanduse

¹⁸² Täpsemalt selle kohta vt ka käesoleva töö peatükk 2.3., kus on analüüsitud tehisintellekti määrusega generatiivse tehisintellekti arendajatele pandavaid kohustusi seoses autoriõiguste kasutamiseiga.

¹⁸³ Vt AutÕS § 13 lg 1 p 9¹.

¹⁸⁴ Vt selle kohta peatükk 1.1.

kohaselt rakendada mudelite treenimiseks kasutatava teksti- ja andmekaeve puhul üldist õiglase kasutuse doktriini.¹⁸⁵ Õiglase kasutuse doktriini esimese eelduse, kasutuse eesmärgi ja iseloomu osas on New Yorgi osariigi kohus hinnanud, kas tegemist võib olla edasiarendava kasutusega¹⁸⁶ raamatu osade avaldamisel.¹⁸⁷ Selle kohtuasja otsuses leidis kohus, et õiglase kasutuse doktriin on täidetud, kuna kaitstud teoseid kasutati edasiarendaval eesmärgil. Sarnasele seisukohale võiks jõuda ka mudelite arendamisel, sh teksti- ja andmekaeve teostamisel, kuna kaitstud teoseid kasutatakse üksnes mudeli treeningandmetena, mis täidab eraldiseisvat eesmärki. Kuigi õiglase kasutuse doktriinile tuginemine mudelite arendamisel ei ole praktikas veel selge, siis viitavad varasemad kohtuasjad sellele, et mudelite arendamise juhud langevad tõenäoliselt õiglase kasutuse doktriini alla.¹⁸⁸ Seega võib varasemast kohtupraktikast tulenevalt järeldada, et treeningandmete kasutamine on võrdlusriigis Ameerika Ühendriikides tõenäoliselt õiguspärane. Sellest sõltumata on õiglase kasutuse doktriini kasutus mudelite treenimise seisukohast õiguskirjanduses laialdaselt kritiseeritud. Õiglusteadlaste arvates ei taga õiglase kasutuse doktriin mudelite arendajatele kindlust, missugustel juhtudel erandile tuginemine on võimalik. Seetõttu on ka Ameerika Ühendriikides õiguskirjanduses arutletud selle üle, kas on vajalik luua eraldi erand, mis sarnaneks Euroopa teksti- ja andmekaeve erandiga.¹⁸⁹

Eeltoodust tulenevalt annab seega AutÕS § 19² lg-s 1 sätestatud teksti- ja andmekaeve erand ühest küljest mudelite arendajatele piisavalt võimalusi mudelite arendamiseks, kuid teisalt võimaldab eelkõige lg-s 2 sisalduv niinimetatud *opt-out* võimalus õiguste omajatel oma õigusi tõhusamalt kaitsta. Kokkuvõtvalt on võimalik AutÕS § 19² erandile tugineda muuhulgas erasektoris tegutseval juriidilisel isikul, kuivõrd õigustatud subjektide ring ei ole piiratud nii mudeli treenimisel kui ka peenhäälestamisel.

¹⁸⁵ USC, 17, § 107 (b).

¹⁸⁶ Edasiarendav kasutus kui *transformative use*. Vt ka Lee, K. jt (viide 83), lk 95-96; Quang, J. Does training AI violate copyright? Berkeley Technology Law Journal 1/2023, lk 1416-1417.

¹⁸⁷ Ameerika Ühendriikide kohtuasjas Authors Guild v. Google, Inc. leidis kohus, et tegemist on edasiarendava kasutusega. Selles kohtuasjas esitas Authors Guild Google vastu hagi, kuna Google võimaldas potentsiaalsetel ostjatel tutvuda raamatu osaga. Vt Authors Guild v. Google, Inc., 804 F.3d 202 (2nd Cir. 2015), p 207. Edasiarendava kasutuse kohta vt ka Authors Guild v. HathiTrust, 755 F.3d 87 (2nd Cir. 2014), p D19.

¹⁸⁸ Vt Ameerika Ühendriigi kohtuasjadest lisaks eelnevalt viidatud Authors Guild, Inc. v. Google, Inc. asjale ka Perfect 10, Inc. v. Amazon.com, Inc. Sarnaselt eelnevalt viidatud kaasusele on ka viimasena märgitud kohtuasi seotud otsingumootoriga. Mõlemates asjades on asunud seisukohale, et õiglase kasutuse doktriinile tuginemine on võimalik, kui lõpptootel/-teenusel on funktsionaalne eesmärk ja teenus/toode pakub sotsiaalset kasu. Mudelite funktsionaalne eesmärk võiks autori arvates väljenduda selles, et mudel suudab vastata kasutajate küsimustele ja/või täita kasutajate antavaid ülesandeid. Sotsiaalne kasu aga selles, et mudel võimaldab kiire ligipääsu baasteadmistele (sarnaselt Google Books teenusele). Vt lisaks Quang, J. (viide 185), lk 1419-1420.

¹⁸⁹ Quang, J. (viide 185), lk 1425-1433.

2.1.4. Sisendandmete reprodutseerimine isikliku kasutamise eesmärkidel

AutÕS § 18 sätestab võimaluse teost reprodutseerida ja tõlkida isikliku kasutamise eesmärkidel.¹⁹⁰ Erinevalt eelnevalt analüüsitud TDM eranditest, on AutÕS § 18 sätestatud AutÕS IV peatüki 1. jaos.¹⁹¹ Sellest sõltumata on võimalik erandi tingimuste täitumisel reprodutseerida ja tõlkida teost ilma õiguste omaja nõusolekuta ning autoritasu maksmiseta, kuivõrd säte näeb ette, et teos peab olema õiguspäraselt avaldatud.¹⁹² Võrreldes Eesti õigusega, oli Ühendkuningriigis varasemalt sätestatud isikliku kasutamise eesmärkidel reprodutseerimist võimaldav erand CDPA §-s 28B, kuid erand ei võimaldanud õiguste omajatel õiglast hüvitist saada, mistõttu tunnistas kohus selle kehtetuks.¹⁹³

AutÕS §-s 18 sätestatud erandile ei ole võimalik tugineda mudeli arendajal või ärielistel eesmärkidel tegutsevatel juriidilistel isikutel. Isikliku kasutamise eesmärkidel reprodutseerimist võimaldava erandi kohaldamise vajadus võib tekkida eelkõige juhul, kui valmis mudelit on arendatud teostega, mis ei ole piisavad kasutaja soovitud väljundi loomiseks. Selle lahendamiseks peab kasutaja lisama valmis mudelisse sisendandmete näol kaitstud teose, et mudel suudaks luua sobiva väljundi. Sellisel juhul võib mudeli kasutajal olla võimalik tugineda AutÕS §-s 18 sätestatud erandile, mis võimaldab teose vaba reprodutseerimise isikliku kasutamise eesmärkidel ning mistõttu on tegemist erandiga, millele võivad tugineda valmis mudeli kasutajad. AutÕS § 18 lg 1 järgi on teose vaba reprodutseerimine lubatud isikliku kasutamise eesmärkidel, kui tegemist on õiguspäraselt avaldatud teosega ning eeldusel, et tegevus ei taotle ärilisi eesmärgi. Isikliku kasutamise eesmärgi AutÕS ei täpsusta, seega sellest tulenevalt võib teost eelduslikult reprodutseerida nii õppetööks kui ka enda huviks.¹⁹⁴

¹⁹⁰ AutÕS §-ga 18 on üle võetud infoühiskonna direktiivist tulenev erand teose vabaks reprodutseerimiseks (AutÕS § 13 lg 1 p 1) ja tõlkimiseks (AutÕS § 13 lg 1 p 4) isikliku kasutamise eesmärkidel. Kuigi vastav erand hõlmab ka teose tõlkimist, siis käesoleva töö raames selle analüüs asjakohane ei ole, mistõttu käsitleb autor töös edaspidi üksnes teose reprodutseerimist isikliku kasutamise eesmärkidel.

¹⁹¹ AutÕS IV peatüki „Autori varaliste õiguste teostamise piiramine (teose vaba kasutamine)“ esimene jagu on „Põhisätted“. Alles teise jao pealkiri „Teose kasutamine autori nõusolekuta ja tasu maksmiseta“ viitab sellele, et erandeid on võimalik kasutada ilma autori nõusolekuta ja tasu maksmiseta.

¹⁹² Vt infoühiskonna direktiivi artikkel 5 lg 2 p b ja põhjenduspunkt 35. Need näevad ette, et õiguste omajad peaksid saama ka „õiglase hüvitise“. AutÕS §-s 18 on õiglase hüvitise hõlmatud sellega, et teos peab olema õiguspäraselt avaldatud, st õiguste omajad on saanud tasu mõnes muus vormis.

¹⁹³ Kõnealune säte pärines ajast, mil Ühendkuningriik kuulus veel EL-i, kuid kohus tunnistas sätte vahetult enne EL-ist välja astumist kehtetuks. Kohtuasja kohaselt oli CDPA § 28B vastuolus infoühiskonna direktiiviga, kuna säte ei võimaldanud õiguste omajatel õiglast hüvitist saada. Vt selle kohta lähemalt R (British Academy of Songwriters, Composers and Authors and others) v Secretary of State for Business, Innovation and Skills 2015 EWHC 2041, p 3 ja Bently, L. jt, lk 257 (viide 20).

¹⁹⁴ Isikliku kasutamise eesmärkidel reprodutseerimise erandi kohta vt ka Pisuke, H. Autoriõiguse alused. Tallinn: Ellington Trükiagentuur 2006, lk 80-81.

Samuti ei tulene AutÕS §-st 18 mahupiiranguid reprodutseerimisele, kuid samas saavad üksnes füüsilised isikud erandile tugineda. Lõige 2 loetleb teosed, mille suhtes AutÕS § 18 lg-s 1 sätestatud erand ei kehti.

Hinnates, kas tegemist on õiguspäraselt avaldatud teosega, võib autori hinnangul lähtuda sarnastest eeldustest nagu ka seadusliku juurdepääsu (AutÕS § 19¹ lg 1) või õiguspäraselt juurdepääsetava (AutÕS § 19² lg 1) teose puhul. Seega tuleks eelkõige tuvastada, kas kasutajal on autori nõusolek teose kasutamiseks või teos oli eelnevalt avaldatud õiguspäraselt.¹⁹⁵ Sarnaselt võivad viimase alternatiivi alla kuuluda ka näiteks sellised olukorrad, kus teos on vabalt veebis kättesaadav.¹⁹⁶ Kui mudeli kasutaja on varasemalt kaitstud teose endale soetanud või muul viisil õiguspärase ligipääsu teosele saanud, siis on täidetud ka AutÕS § 18 lg 1 esimene eeldus. Teise eelduse kohaselt ei tohi kasutaja oma tegevusega taotleda ka ärilisi eesmärke. Eeltoodut võib illustreerida hüpoteetiline olukord, kus kasutaja soovib luua raamatupoest soetatud värskest proosateosest enda jaoks kokkuvõtte. Sellisel juhul on täidetud mõlemad eeldused – raamat kui kaitstud teos on õiguspäraselt avaldatud ning kasutaja tegevus ei taotle ärilisi eesmärke. Eelnevast johtuvalt on mudeli kasutajal võimalik sisendandmetena reprodutseerida kaitstud teoseid autoriõigusi rikkumata eeldusel, et mudeli kasutajal oli autori nõusolek teose kasutamiseks või teos oli avaldatud õiguspäraselt. Lisaks ei tohi kasutaja sisendandmeid ärielistel eesmärkidel reprodutseerida.

2.2. Treening- ja sisendandmete kasutamise kooskõla kolmeastmelise testiga

Eelnevalt käsitletud vaba kasutuse erandid võimaldavad kasutada autorile kuuluvaid varalisi õigusi tasu maksmiseta ning autori nõusolekuta. Eraldiseisvalt on seaduses ettenähtud ka juhud, millega kaasneb autorile hüvitise maksmine.¹⁹⁷ Vaba kasutuse erandite kaheks jaotumine on oluline, kuivõrd ühel juhul – autorile hüvitise maksmiseta – peab kasutus erandile tuginemiseks läbima kolmeastmelise testi.¹⁹⁸ AutÕS §-s 17 sätestatud kolmeastmelise testi läbimiseks on esmalt vaja kindlaks teha, kas seaduses ettenähtud vaba kasutuse erandi

¹⁹⁵ Vt võrdluseks Ühendkuningriigi CDPA § 28B (4), mille järgi teos peab olema alaliselt seaduslikult omandatud kasutaja poolt, st teosele ei tohi olla üksnes ajutist juurdepääsu.

¹⁹⁶ Rosati, E., lk 45 ja 76 (viide 71).

¹⁹⁷ Vt C-457/11 *VG Wort/Kyocera*, ECLI:EU:C:2013:246, p-d 33-35; DSM-direktiivi põhjenduspunkt 6.

¹⁹⁸ AutÕS § 17. Kolmeastmelise testi eelduste täidetust peab kontrollima kohus õigust kohaldades. Vt nt RKTko 2-16-17491/52, p-d 21.1. ja 21.4.

tingimused on täidetud.¹⁹⁹ Eelnevalt käsitletust ilmneb, et treening- ja sisendandmete kasutamisel kohalduvad vaba kasutuse erandid järgmistel juhtudel:

- i) AutÕS § 18¹ lg-s 1 sätestatud erand osaliselt treeningandmete kasutamisele, st erand ei kohaldu tervikuna treeningandmete kasutamisele mudeli arendamisel;
- ii) AutÕS § 19¹ lg-s 1 sätestatud erand tingimusel, et mudeli arendamine on otseselt seotud teadustöö eesmärgiga ning treeningandmeid ei kasutata ka hiljem muul eesmärgil. Kõnealune erand kehtib üksnes treeningandmete kasutamisele;
- iii) AutÕS § 19² lg-s 1 sätestatud erand treeningandmete kasutamisele;
- iv) AutÕS § 18 lg-s 1 sätestatud erand sisendandmete kasutamisele.

Seega on eeltoodud juhtudel täidetud ka kolmeastmelise testi esmane eeldus. Kolmeastmelise testi teise ning kolmanda eelduse täituvust tuleb hinna igakordselt sõltuvalt asjaoludest.²⁰⁰ Esmase eelduse kohaselt peab hindama, ega treeningandmete ajutine reprodutseerimine (AutÕS § 18¹), teksti- ja andmekaeveks kasutamine (AutÕS § 19¹ ja AutÕS § 19²) ning sisendandmete isiklikul eesmärgil kasutamine (AutÕS § 18) ole vastuolus teose tavapärase kasutamisega. Tavapärase kasutuse puhul tuleb õiguskirjanduse kohaselt hinnata muuhulgas seda, kas erandile tuginev teose kasutus võib kattuda teose tavapärase kasutusega.²⁰¹ Näitena on tavapärase kasutuse kohta õiguskirjanduses toodud olukord, kus raamatu monograafia fotokopeeritakse.²⁰² Kuivõrd raamat on mõeldud tavapäraselt lugemiseks kirjastaja avaldatud formaadis, ei saa taolist fotokopeerimist käsitada ka tavapärase kasutusena.²⁰³ Lisaks eeltoodule, peab koleastmelise testi puhul ka hindama, ega kasutus kahjusta põhjendamatult autori seaduslikke huve.²⁰⁴ Kolmanda eelduse puhul on kohtupraktika sisustanud, kuidas tuleb hinnata autori seaduslike huvide kahjustumist. Näiteks on EK praktikas leitud, et teoste kasutamine kommertseesmärkidel võib õiguste omajate huve põhjendamatult kahjustada.²⁰⁵

¹⁹⁹ Täpsemalt kolmeastmelise testi kohta vt nt Senftleben, M. R. F. *Copyright, Limitations and the Three-step test. An Analysis of the Three-Step Test in International and EC Copyright Law*. Kluwer Law International, 2004. Vt ka C-265/16, *VCAST*, p-d 53-54 kolmeastmelise testi eelduste täidetavuse kohta.

²⁰⁰ Pisuke, H. (viide 193), lk 80.

²⁰¹ Vt nt Wymeersch P. *EU Copyright Exceptions and Limitations and the Three-Step Test: One Step Forward, Two Steps Back*. GRUR International, 7/2023 lk 10-13.

²⁰² Pisuke, H. (viide 193), lk 82.

²⁰³ Samas.

²⁰⁴ Vt ka *Football Association Premier League Ltd v QC Leisure* (C-403/08) EU:C:2011:631, p 164 tasakaalu leidmise kohta erandile tugineja (kasutaja ja kasutuse) ning autori (õiguste omaja) huvide vahel. Mõjuvaks asjaoluks selle eelduse hindamisel on muuhulgas see, kas õiguste omaja võib jääda ilma oma tulust. Vt Wymeersch, P., lk 15-17 (viide 200).

²⁰⁵ C-162/10 *Phonographic Performance (Ireland)*, ECLI:EU:C:2012:141, p 74-76. Sarnasele seisukohale on jõudnud ka Riigikohus. Vt RKTko 2-16-17491, p 21.1 ja selles p-s olevad viited EK praktikale.

Õiguskirjanduses on veel selgitatud, et õiguste omajate huve kahjustava kasutusega on tegemist eelkõige siis, kui teost kasutatakse viisil, mis konkureerib teose autoriga.²⁰⁶

Eeltoodut arvestades hindab autor esmalt kolmeastmelise testi teise ja kolmanda eelduse täidetust TDM erandite ja ajutist reprodutseerimist võimaldava erandi osas. Õiguskirjanduses on eristatud tavapärasest teose oma algkujul reprodutseerimist, näiteks koopiamasinaga raamatutest koopiategemist ning mudelite treenimise kontekstis vaid teksti- ja andmekaeve tulemusel mustrite leidmise eesmärkidel reprodutseerimist.²⁰⁷ Teksti- ja andmekaeve käigus analüüsitakse erinevaid teoseid, et tunda ära mustreid ja korrelatsioone. Juhul, kui õiguste omaja ei ole loonud teost spetsiifiliselt teksti- ja andmekaeveks, siis ei ole vastuolu näiteks kirjandus- või kunstiteose tavapärase kasutamisega. Kuna teose algne eesmärk on pakkuda meelelahutuslikku või kunstilist väärtust, mitte võimalust statistilisteks analüüsideks, siis ei ole ka TDM erand vastuolus teose tavapärase kasutamisega. Kolmanda eelduse osas tuleb hinnata ega erandile tuginev kasutus ei kahjusta põhjendamatult autori seaduslike huve. Jätkates eelnevalt välja toodud näitega, kui kirjandus- või kunstiteose autorid loovad oma teoseid eesmärgiga pakkuda kunstilist või meelelahutuslikku väärtust, siis on ka nende seaduslikud huvid seotud sarnaste kasutusviisidega. Kui aga teoseid kasutatakse teisel eesmärgil, näiteks teksti- ja andmekaeveks, siis eelduslikult ei jää teose õiguste omaja ilma võimalikust tulust, kuna teose loomisel ei olnud sellise tuluga arvestatud.²⁰⁸ Teose kasutamine teksti- ja andmekaeveks ei kahjusta teose eesmärgipärasest kasutamist ega kahanda selle väärtust.

Ka AutÕS §-s 18¹ sätestatud erandi puhul on täidetud kolmeastmelise testi teine eeldus, kuna algne õiguste omaja ei ole tõenäoliselt loonud teost spetsiifiliselt selleks, et teost kasutataks ajutiselt mõne tehnilise protsessi lahutamatu ja olulise osana.²⁰⁹ Samuti ei pruugi AutÕS §-le 18¹ tuginev kasutus põhjendamatult kahjustada autori seaduslike huve, kuivõrd teose kasutamine tehnilise protsessi tarbeks, sh mudeli treenimiseks, omab teistsugust eesmärki, mis ei konkureeri teose õiguste omaja kasutusega. Seega on täidetud ka kolmeastmelise testi kolmas eeldus.

²⁰⁶ Gervais, D. J. Exploring the Interfaces Between Big Data and Intellectual Property Law. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law (JIPITEC)*, 10/2019, lk 32-33.

²⁰⁷ Sandiumenge, I. (viide 126), lk 23-24. Viidatud artikli autor on selgitanud, et tavapärase kasutuse hindamisel tuleb eristada, kas teose reprodutseerimine toimub selleks, et teosest eraldada mustreid ja omadusi või toimub samal eesmärgil, milleks autor teose on loonud.

²⁰⁸ Samas, lk 25.

²⁰⁹ Samas, lk 26; Wymeersch, P. (viide 200), lk 8-9.

Vaba kasutuse eranditele tuginemise võimalikkuse hindamisel tuleb analüüsida ka kasutatava teose mahtu. Varasemat kohtupraktikat hinnates jõeldub, et üksnes motiveeritud mahu treeningandmete kasutamine on õiguspärane.²¹⁰ Eelkõige võib õiguste omajate seaduslikele huvidele tekkida kahju, kui treeningandmetena on kasutatud teoseid tervikuna ning mudelid suudavad taastada teoseid sisuliselt algsel kujul – sellisel juhul konkureerivad teksti- ja andmekaeve eesmärkidel kogutud teosed õiguste omajate kasutusega ning nende kasutamine võib kahjustada õiguste omajate seaduslike huve.²¹¹ Kuivõrd kirjeldatud teose taasloomise probleemi puhul on tegemist praktikas harva esineva mudeli veaga, siis võib asuda seisukohale, et enamikel juhtudel on AutÕS § 18¹ lg 1, § 19¹ lg 1 ning § 19² lg 1 puhul kolmeastmelise testi teine ja kolmas eeldus täidetud.²¹²

AutÕS § 18 lg-s 1 sätestatud erandi puhul, millele on võimalik tugineda sisendandmete kasutamisel, on samuti kolmeastmelise testi teine ja kolmas eeldus täidetud. Tegemist ei ole kasutusega, mis oleks vastuolus teose tavapärase kasutusega ning kahjustaks autori seaduslike huve. Isiklikul otstarbel teose sisendandmetena kasutamisel ei saa õiguste omaja jääda enam ilma võimalikust tulust, kuna kasutaja on sisendandmeteks oleva teose juba soetanud. Seetõttu ka ei konkureeri isiklikul eesmärgil sisendandmete reprodutseerimine õiguste omaja kasutusega, st ei kahjusta põhjendamatult autori seaduslike huve. Samuti ei ole AutÕS §-s 18 sätestatud erandil motiveeritud mahu nõuet, mistõttu on võimalik kaitstud teoseid isiklikul otstarbel sisendandmetena kasutada täielikul kujul.

Kokkuvõtlikult on eelnevalt kirjeldatud tingimuste täitumisel võimalik treeningandmete kasutamisel tugineda AutÕS § 18¹ lg-s 1, § 19¹ lg-s 1 ja § 19² lg-s 1 ning sisendandmete kasutamisel AutÕS § 18 lg-s 1 sätestatud vaba kasutuse eranditele. Kehtiv õigus võimaldab üksnes teatud juhtudel kaitstud teoseid treening- ja sisendandmetena õiguspäraselt kasutada.²¹³ Näiteks ei hõlma AutÕS § 18¹ lg-s 1 sätestatud erand kõiki mudeli treenimise etappe. Seejuures on ärielistel eesmärkidel võimalik mudeleid arendada üksnes kooskõlas AutÕS § 19² lg-s 1 sätestatuga ning tingimusel, et õiguste omaja ei ole keelanud AutÕS § 19² lg-st 2 tulenevalt

²¹⁰ RKTko 2-16-17491/52, p 21.5. Kuigi RK hindas konkreetses kohtuasjas AutÕS § 75 lg-s 2 sätestatud eelduste täidetust, on autori hinnangul motiveeritud mahu nõue teatud juhtudel oluline ka teksti- ja andmekaeve puhul (eelkõige seoses mudeli võimekusega taastada algsed treeningandmeteks kasutatud teoseid).

²¹¹ Vt käesoleva töö viide 79 mudeli meeldejätmise võime ja teose taasloomise kohta. Vt ka New York Times v. OpenAI, Inc., No. 1:23-cv-11195, p 80.

²¹² Seda, et TDM erandid läbivad kolmeastmelise testi on märgitud ka õiguskirjanduses. Vt nt Guadamuz, A. (viide 18), lk 10-11.

²¹³ Autoriõigusega kaitstud teoste treening- ja sisendandmetena kasutamise õiguspärasest võivad mõjutada ka andmekaitseõiguslikud probleemid, kuid neid autor käesoleva töö raames ei käsitle.

oma teoste teksti- ja andmekaeveks kasutamist. Välistatud ei ole ka AutÕS § 19¹ lg 1 kohaselt arendatud mudeli hilisem ärilisel otstarbel kasutamine. Arvestades aga autoriõiguste potentsiaalse rikkumise tuvastamise keerukust, võib õiguste omajatel olla sisuliselt võimatu oma ainuõigusi tõhusalt teostada, mistõttu ei taga kehtiv õigus õiguste omajatele piisavat kaitset. Autor käsitleb järgmises peatükis seda, missugune mõju on kavandataval tehisintellekti määrusel kehtiva õiguse pinnalt üles kerkinud probleemkohtadele.

2.3. Tehisintellekti määrusest tulenevad nõuded treeningandmete kasutamisele

2021. aastal esitas Euroopa Komisjon Euroopa Parlamendile ettepaneku tehisintellekti määruse vastu võtmiseks.²¹⁴ Kuigi 2021. aastal avaldatud määruse eelnõu ei sisaldanud mudeli arendajate jaoks autoriõigusega seotud kohustusi, siis 2022. aastal hoogustunud generatiivse tehisintellekti tõus muutis aktuaalseks ka käesolevas magistritöös käsitletava võtmeküsimuse – kas ja missuguses ulatuses on autoriõigusega kaitstud teoste kasutamine generatiivse tehisintellekti treenimisel õiguspärane. Selle tõttu lisati määruse eelnõusse ka autoriõigusi puudutav artikkel 53, mis küll ei sätesta õiguste omajatele täiendavaid õigusi ega vaba kasutuse erandeid, kuid näeb ette lisakohustused mudelite arendajatele.²¹⁵ Järgnevalt analüüsib autor, kas tehisintellekti määruses sisalduvad mudeli arendajatele loodud kohustused aitavad tagada treeningandmete mudelis kasutamise õiguspärasust ning seda, kas määrus lahendab eelnevates peatükkides esile kerkinud probleemid.

Mudelite õiguspärase treenimisega seotud probleemid võib jagada üldpildis kahte kategooriasse. Esiteks probleemid, mis on seotud Eesti õigusesse AutÕS § 19² lg-ga 2 üle võetud DSM-direktiivi artikkel 4 lg-st 3 tuleneva *opt-out* võimalusega ning teiseks probleemid, mis on seotud õiguste omajate ainuõiguste tõhusa teostamisega.²¹⁶ Eelnevalt käsitletust nähtuvalt, kumab mõlemast probleemist läbi vajadus tagada parem läbipaistvus mudelite arendamisel. Määruse artikli 53 lg 1 p c järgi peavad mudeli arendajad kindlaks tegema kõigi

²¹⁴ Vt käesoleva töö viide 25. Määrusest tulenevad nõuded kehtivad mudeli arendajatele sõltumata sellest, kus mudelit arendatakse ning nõuded puudutavad treeningandmete kasutamist. Selle kohta vt ka käesoleva töö peatükk 1.1.

²¹⁵ Vt võrdluseks 2021. aastal Euroopa Komisjoni tehtud ettepaneku p 2.1 (viide 9) ning tehisintellekti määruse eelnõu põhjenduspunkt 5 (viide 25).

²¹⁶ Sarnasest õiguspärase kasutamise probleemide jaotusest on lähtunud ka õiguslalases kirjanduses ning artiklites. Vt Senftleben, M. (viide 178), lk 10; Keller, P. A first look at the copyright relevant parts in the final AI Act compromise. – Kluwer Copyright Blog, 11.12.2023. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/> (25.03.2024).

mudelis kasutatavate treeningandmete puhul, mis on kaitstud autoriõigustega, kas õiguste omanik on keelanud ära teose kasutamise mudeli treenimise tarbeks toimuvaks teksti- ja andmekaeveks tulenevalt AutÕS § 19² lg-st 2.²¹⁷ Sätestatud kohustus aitab kaasa sellele, et kaitstud teoseid ei kasutata treeningandmetena, kui õiguste omanik on selle AutÕS § 19² lg 2 kohaselt välistanud, mistõttu aitab määrusega sätestatud kohustus kaasa ka sellele, et treeningandmete kasutamine mudelite treenimisel oleks õiguspärane. Õiguskirjanduses on märgitud, et määruse artikkel 53 lg 1 p-st c tulenev kohustus seab üldiselt mudeli arendajatele üsna suure koormuse, kuivõrd võimekamate mudelite kasutamiseks on vaja hulgaliselt andmeid ning mudelite andmekogude koostamisel võidakse reprodutseerida mitu miljardit teost.²¹⁸ See tähendab, et ka potentsiaalseid AutÕS § 19² lg-st 2 tulenevaid õiguste omanike tehtud välistusi võib olla võimatu tuvastada.²¹⁹ Isegi kui mudeli arendajatel on olemas ülevaade mudeli treenimiseks kasutatavatest andmekogudest, ei pruugi olla neil ülevaadet kõikidest teostest, mida treeningandmetena kasutatakse.²²⁰ Kuigi eelnevast alapeatükist nähtuvalt on loodud mitmed tehnilised vahendid, näiteks robots.txt ja ai.txt, mis aitavad artikkel 53 lg 1 p-st c tuleneva kohustuse täitmisele kaasa, siis võib kohustuse täitmine osutada mudeli arendajate jaoks koormavaks.

Täiendavalt kehtestatakse artikkel 53 lg 1 p-ga d mudeli arendajatele kohustus koostada ja avalikustada piisavalt üksikasjalik kokkuvõte sellest, kuidas mudeleid treenitakse, s.o selle kohta, missuguseid treeningandmeid mudelite treenimisel kasutatakse.²²¹ Kohustuse sisu ei ole määruses selgemalt sisustatud, mistõttu jääb ebaselgeks, mida kujutab endast „piisavalt üksikasjalik kokkuvõte“. Määruse eelnõus on täiendavalt selgitatud, et artikkel 53 lg 1 p-st d tulenev kokkuvõte peab olema tehniliselt üksikasjaliku asemel laiaulatuslik, et see hõlbustaks muuhulgas autoriõiguste omanikel kasutada oma õigusi. Näitena on toodud, et kokkuvõte peab sisaldama loetelu andmekogudest, mida kasutati mudeli treenimiseks. Punkti d järgi peab kokkuvõte vastama tehisintellekti ameti²²² esitatud näidisele.²²³ Sellest nähtuvalt on

²¹⁷ Eraldiseisev küsimus viidatud probleemi osas on, kas AutÕS § 19² lg-s 2 sätestatud välistuse tegemine peab toimuma *ex ante* või *ex post*. Vt selle kohta alapeatükk 2.3.

²¹⁸ Lemley M. A., Casey, B. Fair Learning. Texas Law Review Volume 99 Issue 4 2021, lk 743-744.

²¹⁹ Täiendavalt on õiguskirjanduses arutletud selle üle, et määruse artikkel 53 lg 1 punktis c oleva kohustuse täitmist raskendab lisaks asjaolu, et mudeli arendajad peavad läbi viima ulatusliku õigusliku auditi selleks, et tuvastada kõikide treeningandmete seast need, mis vastavad AutÕS §-s 4 sätestatud tingimustele ehk autoriõigusega kaitstud teoste tunnustele. See võib tähendada aga miljardite teoste läbitöötamist. Vt Hacker, P. jt. Regulating ChatGPT and other Large Language Generative AI Models. ArXiv, 2023, lk 1115.

²²⁰ Sellele probleemile on viidanud ka Senfleben, M. (viide 178), lk 12.

²²¹ Tehisintellekti määruse eelnõu artikkel 53 lg 1 p d ja põhjenduspunkt 104.

²²² Tehisintellekti ameti (*AI Office*) ja selle pädevuste kohta vt lähemalt Euroopa Komisjon. Euroopa tehisintellekti amet. Kättesaadav: <https://digital-strategy.ec.europa.eu/en/policies/ai-office> (28.02.2024).

²²³ Tehisintellekti määruse eelnõu põhjenduspunktid 96, 101 ja 107.

tehisintellekti ametil pädevus otsustada, missugusel kujul määruse artikkel 53 lg 1 p-s d sätestatud kohustust täidetakse. Eeltoodust tulenevalt oleks põhjendatud, kui ameti koostataja ja esitatav näidis võimaldaks ühest küljest õiguste omajatel oma õigusi tõhusamalt teostada, kuid teisalt võimaldaks mudeli arendajatel ka seda kohustust mõistlikkuse piires täita. Autori hinnangul oleks võimalik seeläbi tagada, et mudelite arendajale artikkel 53 lg 1 p-s d sätestatud kohustus ei takistaks mudelite arengut ning oleks täidetav selliselt, et ei koormaks liigselt arendajat. Määruse artikkel 53 lg 1 p-st d tulenev kohustus aitab kaasa sellele, et õiguste omajatel on võimalik aru saada, kas nende teoseid on kasutatud mõne konkreetse mudeli treenimisel või peenhäälestamisel.²²⁴ Ühtlasi võimaldab p-s d sätestatud kohustus tagada õiguste omajate kaitse juhul, kui nad on teose kasutamise AutÕS § 19² lg 2 järgi välistanud, kuid mudeli arendajad kasutavad endiselt teoseid ning rikuvad seega AutÕS § 13 lg 1 p-st 1 tulenevat reprodutseerimisõigust, mida esimesest peatükist nähtuvalt teostatakse mudeli treeningandmete kasutamisel. Samuti on õiguskirjanduses märgitud, et määruse artikkel 53 p-s d sätestatud kohustus aitab õiguste omajatel tuvastada potentsiaalseid õiguste rikkujaid, kuna õiguste omajad saavad ülevaate, kus nende teoseid on kasutatud.²²⁵

Analüüsidest võrdlusriike, on ka Ühendkuningriik võtnud Euroopa Liiduga sarnase suuna treeningandmete kasutamise õiguspärasuse probleemide lahendamiseks. Nimelt on ka Ühendkuningriigis hetkel loomisel tehisintellekti käsitlev õigusakt, mis peaks aitama tagada muuhulgas mudelite treenimise läbipaistvuse ning seeläbi nende õiguspärase treenimise.²²⁶ Võrreldes Euroopa Liidu tehisintellekti määrusega, on Ühendkuningriigi tehisintellekti käsitleva õigusakti eelnõu aga veelgi konkreetsem, kui võrd sisaldab kohustust esitada mudeli arendajatel kõik mudeli treenimiseks kasutatavad treeningandmed.²²⁷ See sarnaneb mõneti varasemale Euroopa Liidu määruse eelnõule, kus sisaldus samuti kohustus praeguse andmekogusid sisalduva kokkuvõtte asemel esitada loetelu kõikidest kasutatavatest treeningandmetest. Teises võrdlusriigis, Ameerika Ühendriikides on avaldatud 2023. aastal Ameerika Ühendriikide presidendi korraldus, mille üheks eesmärgiks on edendada

²²⁴ Vt ka Communia Policy Paper. Using Copyrighted Works for Teaching the Machine. Communia Association, 26.04.2023. Kättesaadav: <https://communia-association.org/wp-content/uploads/2023/04/Policy-Paper-15-on-using-copyrighted-works-for-teaching-the-machine.pdf> (02.03.2024).

²²⁵ Senftleben, M. (viide 178), lk 12; Hacker, P. jt (viide 219), lk 1117.

²²⁶ Ühendkuningriigis on parlamendis käesoleva töö kirjutamise hetkel tehisintellekti käsitlev õigusakt. Vt Artificial Intelligence (Regulation) Bill. HL Bill 11. 22.11.2023. Kättesaadav: <https://bills.parliament.uk/publications/53068/documents/4030> (02.03.2024).

²²⁷ Samas, § 5(1)(a)(i).

innovatsiooni ning selgitada tehisintellekti ja intellektuaalse omandiga seotud küsimusi.²²⁸ Eelnevalt käsitletu järgi on Ameerika Ühendriikides TDM erandi asemel õiglase kasutuse doktriin, mis on oma olemuselt TDM eranditest laiaulatuslikum. Õiguskirjanduses põhistatu järgi võib eeltoodu koosmõjus määrusega kehtestataivate nõuetega tähendada Ameerika Ühendriikide tehnoloogiaettevõtete jaoks paremaid võimalusi mudelite arendamiseks võrreldes Euroopa Liidu mudeli arendajatega.²²⁹

Kokkuvõtlikult on võimalik eelnevalt käsitletu põhjal järeldada, et tehisintellekti määruises sisalduvad mudeli arendajatele sätestatud kohustused aitavad kaasa treeningandmete õiguspärasele kasutamisele mudeli treenimisel ja peenhäälestamisel. Määruse artikkel 53 p-d c ja d aitavad senisest paremini tagada autori ainuõiguste kaitse ning õiguste tõhusa teostamise ning selle abil saavad mudelite arendajad kindluse, et nende loodud mudeleid on arendatud õiguspäraselt ja on seega kooskõlas autoriõigusega. Võrreldes aga määrusega õiguste omajatele pakutavat kaitset ning äärmiselt laialdasi mudeli arendajatele sätestatud kohustusi, võib esineda oht, et rangete nõuete tulemusel lahkuvad mudelite arendajad Euroopa Liidu turult ega tee ka mudeleid siinsetele tarbijatele kättesaadavaks, kuivõrd näiteks Ameerika Ühendriikides on mudelite õiguspärasest treenimist lubavad erandid leebemad kui Euroopa Liidus. Seega võib ebatasakaal õiguste omajatele pakutava kaitse ja mudeli arendajatele seataivate kohustuste vahel soodustada sellise olukorra tekkimist, mis osutub kahjulikuks mõlemale osapoolle – arendajad lahkuvad Euroopa Liidu turult ning autorid jäävad sellega ilma ka Euroopa Liidus kehtivast kaitsest. Teisalt lahendab kavandatav määrus eelnevalt kirjeldatud mudeli arendamisega seotud läbipaistvuse probleemid, eelkõige osas, mis puudutab autorite ainuõiguste teostamise tõhusust ning aitab tagada AutÕS § 19² lg-st 2 sätestatud teksti- ja andmekaeve eesmärgi täitumise. Seetõttu tagab kavandatav määrus koosmõjus kehtiva õigusega osaliselt siiski tasakaalu õiguste omajate ja mudeli arendajate huvide vahel, kuid võib õhnestada innovatsiooni võrdlemisi laialdaste ja raskesti täidetavate kohustuste tõttu, mis seatakse mudeli arendajatele.

²²⁸ The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 30.10.2023, p 5.2 c)(iii). Kättesaadav: <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> (04.03.2024). Näiteks peab korralduse järgi andma autoriõiguse ameti direktor (*Director of the United States Copyright Office*) presidendile soovitusi autoriõiguste ja tehisintellektiga seotud võimalike rakendusmeetmete võtmise kohta 270 päeva pärast korralduse kuupäevast (s.o alates 30.10.2023) või 180 päeva pärast Ameerika Ühendriikide Kongressi Raamatukogu (*United States Copyright Office of the Library of Congress*) uuringu avalikustamist. Töö autorile teadaolevalt ei ole uuringut käesoleva töö kirjutamise hetkel veel avalikustatud.

²²⁹ Senftleben, M., lk 14-15 (viide 178).

KOKKUVÕTE

Generatiivse tehisintellekti vallas on praegu vaieldamatult toimumas murrangulised muutused. Sarnaselt igale uuele tehnoloogiale, on ka generatiivse tehisintellekti avalikkuse ette jõudmisega kaasnenud mitmed õiguslikud probleemid ning kohtuvaidlused, mille tulemused on praegu veel lahtised. Kuna generatiivse tehisintellekti arendamine ja kasutamine on suuresti võimalik tänu suurtele andmekogudele, mis sisaldavad ka autoriõigusega kaitstud teoseid, on peamised vaidlused üles kerkinud just seoses treening- ja sisendandmete kasutamisega. Autoriõigusega kaitstud teoseid kasutatakse tihti mudeli treeningandmetena selle treenimisel ja peenhäälestamisel, eelkõige mudeli algse arendaja või ka hilisema kolmandast isikust arendaja poolt. Samuti võidakse autoriõigusega kaitstud teoseid kasutada sisendandmetena mudeli hilisemal rakendamisel mudeli kasutaja poolt.

Käesoleva töö üheks eesmärgiks oli välja selgitada, kas ja millistel tingimustel on autoriõigusega kaitstud teose kasutamine generatiivse tehisintellekti treening- ja sisendandmetena õiguspärane. Töö autor püstitas eeltoodud arvestades hüpoteesi, mille kohaselt autoriõigusega kaitstud teose kasutamisega generatiivse tehisintellekti treening- ja sisendandmetena kaasneb autori ainuõiguste kasutamine, kuid olemasolevad vaba kasutuse erandid võivad teatud juhtudel välistada autoriõiguse rikkumise. Töö esimese eesmärgi saavutamiseks ning hüpoteesi ümber lükkamiseks või kinnitamiseks analüüsis autor kahte uurimisküsimust.

Esimese uurimisküsimuse raames hinnati esimeses peatükis seda, millistel juhtudel ning missuguses ulatuses kasutatakse teose generatiivse tehisintellekti treening- ja sisendandmetena kasutamisel mõnda autorile kuuluvat ainuõigust. Selle tulemusena leiti, et autorile kuuluvatest varalistest õigustest võib mudeli treening- ja sisendandmete kasutamisel olla tegemist eelkõige reprodutseerimisõiguse kasutamisega, kuid sisendandmete kasutamisel võib toimuda ka töötlemisõiguse kasutamine. Isiklikest õigustest võidakse analüüsist nähtuvalt treening- ja sisendandmete kasutamisel tõenäoliselt kasutada õigust autorsusele. Lisaks analüüsiti töös varalistest õigustest ka üldsusele kättesaadavaks tegemise õigust ning isiklikest õigustest au ja väärkuse kaitse õigust, kuid nende puhul mudeli treenimine, peenhäälestamine ega kasutamise etapis ei kasutata autori ainuõigusi, mistõttu ei kaasne sellega ka autoriõiguste rikkumist. Küll aga võidakse üldsusele kättesaadavaks tegemise õigust kasutada erandlikult siis, kui mudeli arendaja teeb treeningandmed kättesaadavaks pärast mudeli treenimist. Treeningandmete

kasutamisel võib õiguste rikkujaks olla kas algse mudeli arendaja või hilisem mudeli peenhäälestaja ning sisendandmete kasutamisel võib rikkujaks olla mudeli kasutaja.

Täiendavalt nähtus esimesest peatükist, et senises õigusraamistikus esineb mitmeid probleemkohti, mis puudutavad mudelite arendamist ning õiguste omajate huvide kaitse tagamist. Kõige problemaatilisemaks on kehtivas õiguses see, et mudelite arendajad ei ole kohustatud avaldama informatsiooni selle kohta, missuguseid autoriõigusega kaitstud teoseid on mudelite treenimisel kasutatud. See on viinud olukorrani, kus ühest küljest on teoste autoritele seadusega tagatud autoriõiguslik kaitse, kuid autoritel on oma ainuõigusi keeruline teostada. Seda näitlikustavad ka mitmed Ameerika Ühendriikides menetluses olevad kohtuasjad, kus õiguste omajatel on keeruline tõendada autoriõiguste rikkumisele viitavaid asjaolusid.

Teises peatükis hinnati esmalt AutÕS-is sätestatud vaba kasutuse erandeid, mille põhjal sai vastata teisele uurimisküsimusele, kas ja missugustel tingimustel välistab autoriõiguse rikkumise mõni seaduses kehtestatud vaba kasutuse erand. Autor võrdles AutÕS regulatsiooni ka Ameerika Ühendriikides ning Ühendkuningriigis oleva õigusraamistikuga, millest omakorda nähtus, et siinsetel mudeli arendajatel ja kasutajatel võivad olla Ühendkuningriigiga võrreldes paremad võimalused treening- ja sisendandmete õiguspäraseks kasutamiseks. Kuigi selle osas puudub veel selge kohtupraktika, tundub Ameerika Ühendriikides olevat siinsest veelgi leebem õiguskord, mis võimaldab eelkõige treeningandmete õiguspärast kasutamist. Eestis kehtiva õiguse kohaselt on kaitstud teoste kasutamisel mudelite treenimiseks võimalik üksnes osaliselt tugineda AutÕS §-s 18¹ sätestatud erandile. Nimelt ei välista AutÕS §-s 18¹ sätestatud erand autoriõiguste rikkumist mudeli treenimisel tervikuna, kuivõrd mudeli treenimine on pikaajalisem ning ka mudelis sisalduvaid treeningandmeid ei talletata mudelis üksnes ajutiselt. Samuti ei räägi AutÕS §-s 18¹ sätestatud erandile tuginemise kasuks asjaolu, et praktikas mudeli arendajad pärast treenimise lõppu treeningandmeid tavaliselt ära ei kustuta, sest ka andmekogud ise on juba väärtuslikud.

Mudelite arendamisel on tavapärane ka teksti- ja andmekaeve kasutamine. Kehtivas õiguses on võimalik teostada teksti- ja andmekaevet teatud tingimuste täitumisel nii teadustöö eesmärgil (AutÕS § 19¹) kui ka väljaspool teadustööd (AutÕS § 19²). Kuigi kehtivas õiguskorras ei ole selgesõnaliselt välistatud teadus- ja kultuuripärandiasutuse loodud andmekogu või nende asutuste loodud mudeli hilisemat ärilisel otstarbel kasutamist selliste ettevõtete poolt, kellel ei ole otsustavat mõju asutuse üle, siis AutÕS § 19¹ lg-s 1 sätestatud erand ei võimalda

andmekogu hilisemat ärilisel otstarbel kasutamist. Teadustöö eesmärgil loodud mudeli hilisema ärilise kasutamisega kaasneb samuti mitmeid probleemkohti, kuivõrd praegune AutÕS § 19¹ sõnastus võimaldab iseenesest kasutada mudelite hiljem ka ärilistel eesmärkidel, mis loob ka töös viidatud andmepesu juhtumite jaoks sooduspinna. Samas ei pruugi töö autori arvates AutÕS § 19¹ sõnastuse täpsustamine, millega välistatakse hilisem mudeli ärilistel eesmärkidel kasutamine, enam tagada tasakaalu õiguste omajate ning mudeli arendajate huvide vahel, mistõttu ei oleks õiguskorra muudatus ilmselt põhjendatud.

AutÕS §-s 19¹ sätestatust laialdasemat treeningandmete õiguspärasest kasutamist võimaldab AutÕS § 19², kuivõrd sellele tuginedes saab mudeleid arendada ka väljaspool teadustööd. Lisaks sellele, et täidetud peavad olema AutÕS §-st 19² tulenevad eeldused, on praktikas problemaatiline see, kas ja kuidas õiguste omaja saab kasutada AutÕS § 19² lg-s 2 sisalduvat *opt-out* võimalust, mida on võimalik teha ka masinloetavate vahenditega. Peamisi keerukusi tekitab ühtse tehnilise standardi puudumine AutÕS § 19² lg-s 2 sätestatu jaoks ning see, kas ja mis määral mudelite arendajad üleüldse järgivad AutÕS § 19² lg-s 2 olevat välistuse tegemise võimalust. Seda eriti juhul, kui on ebaselge, missuguseid andmeid mudeli treenimiseks kasutatakse.

Isegi kui eelnevalt nimetatud vaba kasutuse erandite eeldused on täidetud ning need läbivad kolmeastmelise testi, siis on nendele tuginedes võimalik välistada üksnes reprodutseerimisõiguse rikkumine. Treening- või sisendandmete kasutamisel võidakse rikkuda ka teisi autori ainuõigusi, nt õigust autorsusele ning töötlemisõigust. Samas selgus analüüsi tulemusena, et täiendavate vaba kasutuse erandite loomine ei pruugi olla siiski põhjendatud. Eelnevast tulenevalt leidis kinnitust ka esimene töös püstitatud hüpotees, mille järgi treening- ja sisendandmete kasutamisel teostatakse küll autori ainuõigusi, kuid rikkumise võivad teatud juhtudel välistada olemasolevad vaba kasutuse erandid.

Teiseks töö eesmärgiks oli hinnata, kas ja milliseid muudatusi oleks õiguskorda vaja, et tagada tasakaal autori huvide kaitse ja generatiivse tehisintellekti innovatsiooni vahel, sh kas kavandatavas tehisintellekti määruses olevad autoriõigusi käsitlevad sätted aitavad neid eesmärke saavutada. Püstitati ka teine hüpotees, mille kohaselt võimaldavad nii kehtiv AutÕS kui ka kavandatav tehisintellekti määrus tagada tasakaalu autori huvide kaitse ja innovatsiooni vahel. Eelnevast tulenevalt analüüsiti kolmanda uurimisküsimusena seda, missugune mõju on kavandataval tehisintellekti määrusel autoriõigusega kaitstud teose treeningandmetena kasutamisele.

Töös selgus, et tehisintellekti määruse abil lahendatakse mitmed läbipaistvusega seotud probleemid. Nii peavad mudeli arendajad esitama ülevaate sellest, missuguseid andmekogusid on mudeli arendamiseks kasutatud ning tegema kindlaks, missuguste teoste puhul on kasutatud AutÕS § 19² lg-st 2 tulenevat *opt-out* võimalust. Töö analüüsist ilmnevalt võib aga kavandataval tehisintellekti määrusel olla negatiivne mõju innovatsioonile seeläbi, et määrusega mudeli arendajatele seatavad kohustused on praktikas raskesti täidetavad. Arvestades eelkõige mudelites sisalduvate andmete mahtu ning seda, et võimekamad mudelid kasutavad oma andmekogudes rohkem kaitstud teoseid, kui ükski inimene suudab oma elu ajal läbi sünteesida, võib mudeli arendajal olla keeruline kõnealust kohustust täita.

Eeltoodu tõttu leidis üksnes osaliselt kinnitust teine töös püstitatud hüpotees õiguste omajate huvide ja generatiivse tehisintellekti arendamise tasakaalu tagamise kohta. Ühest küljest lahendab kavandatav määrus mudelite arendamisel ja kasutamisel tekkinud läbipaistvusega seotud probleemid, aidates õiguste omajatel tõhusamalt oma ainuõigusi teostada ning tagades, et mudeli arendajad järgivad autorite tehtud AutÕS § 19² lg-st 2 tulenevaid välistusi. Võttes aga arvesse määrusega sätestatavate kohustuste iseloomu ning seda, kuivõrd keeruline võib olla analüüsist nähtuvalt kehtestatavaid kohustusi täita, võib see tähendada mudelite arendamise seisukohalt innovatsiooni hääbumist.

Mudeli arendamist ning seeläbi ka innovatsiooni võib toetada see, kui luuakse ühtne tehniline standard AutÕS § 19² lg-s 2 sätestatu jaoks. Nimelt arvestades mudeli arendamiseks kasutatavate andmekogude mahtu, võib AutÕS § 19² lg-ga 2 seotud kohustuse täitmist lihtsustada ühtse tehnilise standardi loomine, mis võimaldaks õiguste omajate tehtud välistusi kergemini tuvastada. Mudeli arendajatele seatud kohustuste koormavust võib aidata leevendada muuhulgas see, missugusel kujul nende täitmist nõutakse. Eeltoodu osas on teinud autor ka töös läbivalt soovitusi ning ettepanekuid, mis võiksid aidata kaasa parema tasakaalu tagamisele nii õiguste omajate kui ka mudelite arendajate õiguste ja huvide vahel ning toetada innovatsiooni säilimist.

THE LEGALITY OF USING A COPYRIGHTED WORK AS TRAINING AND INPUT DATA IN GENERATIVE ARTIFICIAL INTELLIGENCE

Abstract

As a result of rapid development and innovation in artificial intelligence (hereinafter “AI”), there has been an increasing focus on AI-related copyright issues, including generative AI (hereinafter “GenAI” or “model”) and its development. The increasing accessibility of GenAI models, which can by now be used by almost everyone for tasks ranging from work to entertainment, has highlighted various concerns about potential copyright infringements. Regardless of the initiation of several lawsuits, mostly in the United States, it is still unclear whether the models may be trained and used in accordance with copyright law. Due to the new technical features, the capabilities of models have been enhanced enormously. These new features allow for the more effective processing of online data, including copyrighted works, which are later used as a training or input data in GenAI. GenAI models have been trained using billions of pieces of online content and if development of these models constitutes a copyright infringement, it could lead to substantial economic losses for the technology companies involved.

The purpose of this master's thesis is to use a comparative-analytical method to assess whether and under what conditions the use of copyrighted works as training and input data for GenAI is lawful. In addition, the thesis will analyse whether and what changes are needed in the legal framework to ensure a balance between protecting the interests of authors and fostering innovation in GenAI. As the AI Act (hereinafter “AIA”) will be soon adopted in the European Union, this work also examines whether the copyright-related provisions in the proposed AIA would help achieve these goals. As there is currently no relevant case law in the European Union yet, the author used the United States as a comparison country, including ongoing court cases there. The United Kingdom was also used as a comparison country, as it holds a leading position in academic discussions relevant to this thesis. Using comparison countries helps to achieve the work's objectives and allows for more comprehensive conclusions about the legitimate use of training and input data.

Arising from the described objectives, the author analyses three research questions. Firstly, under what circumstances and to what extent is the use of GenAI training and input data

considered a use of an author's exclusive rights. Secondly, whether and under what conditions an existing free use exception in the law might exclude the infringement of an author's exclusive rights. Thirdly, the impact of the proposed AIA on the use of copyrighted works as training and input data. The author of the master thesis hypothesised that the use of copyrighted works as training and input data for models leads to the use of the author's exclusive rights, but existing exceptions for free use may exclude copyright infringement in certain cases. Additionally, the author hypothesised that both the current Copyright Act (hereinafter “CA”) and the proposed AIA allow for a balance between the protection of the author's interests and innovation. In situations where the hypothesised outcomes are not fully or partially confirmed, it may be necessary to amend or update the legal framework to ensure the protection of the author's exclusive rights as well as the development of technology and innovation when using copyrighted works as training and input data for models.

In the first chapter, an assessment of the circumstances and extent to which the use of GenAI training and input data is considered as a use of an author's exclusive rights. Copyrighted works used as model training and input data may involve the use of personal and proprietary rights, which could constitute a copyright infringement under CA. Before addressing this question, it is crucial to understand how data used for training GenAI is obtained, processed, and stored, to conclude whether and in which stages of training or model use a potential use of authors exclusive right occurs. Model development and use can be broadly divided into three stages: i) model training (including data collection and compilation for training); ii) model fine-tuning, and iii) the use of the finished model. Assessing the technological nature resulted in the conclusion that training, and input data are used in different ways by different individuals at various stages of model training, which may also affect the use of an author's exclusive rights, i.e., copyright infringement, and the possibility of relying on exceptions to avoid such infringement and liability.

The CA stipulates several exceptions for free use of works (i.e., limitations on exercise of economic rights of authors) that could exclude copyright liability, especially the exceptions for text and data mining arising from Directive EU 2019/790 of the European Parliament and the Council (hereinafter “Directive EU 2019/790”), as well as exceptions arising from Directive 2001/29/EC of the European Parliament and the Council (hereinafter “Directive 2001/29/EC”). The exceptions from Directive 2001/29/EC are particularly for temporary or incidental

reproduction and for personal use. Similarly, the United Kingdom has established free use exceptions and in the United States, it is possible to use the doctrine of fair use instead.

The assessment of free use exceptions was preceded by an analysis of the use cases in which an author's exclusive rights are used during model training and subsequent use. Training and input data subject to copyright protection can be considered as various literary texts, visual works, databases, and longer expressions that meet specific cumulative criteria stipulated in the CA. Several author's rights may be violated during model training and use if copyrighted works are used as training and input data. In summary, it revealed that during the development (including fine-tuning of the model) and use of models, author's reproduction right is primarily used from the author's economic rights. The right of alteration of the work may also be used in certain cases when the model's user is adding the protected work as input data. In exceptional cases, the right of making the work available to the public may also be exercised, but only if the training data is made available to the public after the development of the models. From author's moral rights, the right to authorship may be violated during model training, fine-tuning, and use.

In addition, the analysis showed that one of the problematic areas in model training is the transparency of collection of training data, which may raise questions about the possibility of exercising an author's rights. Since current law does not require model developers to make available information about the data used for model development, including copyrighted works, the current legal framework may not allow authors to effectively ensure their exclusive rights, i.e., the right to authorship. Although several countries, including the comparison countries of the United States and the United Kingdom, have offered non-binding regulatory guidelines and principles for model development, which include guidelines for publishing data used for model development, it is questionable how many developers follow them. In a situation where a model developer uses copyrighted works as training data, it may be difficult for the rights holder to prove infringement in court proceedings.

In the second chapter, the author first assessed the exception for temporary and incidental reproduction arising from Directive 2001/29/EC of the European Parliament and the Council, which has been transposed into the CA through section 18¹. The analysis showed that in most cases, it is not possible for model developers to rely on CA section 18¹.

One common method for collecting model training data is using text and data mining technology, where the model developer uses automated solutions to store content from web pages. One of the text and data mining exceptions from Directive EU 2019/790 of the European Parliament and the Council is implemented into CA section 19¹ which allows the free use of work in scientific research for the purpose of text and data mining. During model development, it is possible to rely on the exception set out in CA § 19¹ (1) if the development of the model is also directly related to research purposes. This exception also enables cooperation between the private sector and research institutions. However, if a model developer uses a data collection created by a research or cultural heritage institution based on the exception stated in CA section 19¹ (1) for commercial purposes in developing their model, such use is not covered by the purpose of the provision. The current legal framework, however, leaves room for the data laundering meaning that private sector technology companies could exploit the exception set out in CA section 19¹ (1), and presumably use the developed model for commercial purposes.

The free use of work for the purpose of text and data mining for purposes other than scientific research is set out in CA section 19². Rights holders of protected works can also explicitly exclude the possibility of relying on the exception set out in CA section 19², known as the opt-out option. Pursuant to CA section 19² (2), it is also allowed to use the exclusion for using machine-reading tools on content made generally available online. The exclusion mentioned under CA section 19² (2) can be implemented using various technical means, e.g., adding the opt-out to a website's metadata or source code, such as through robots.txt or ai.txt file. However, the absence of a uniform technical standard has attracted several criticisms on this topic. The exceptions set out in both CA sections 19² and CA section 19¹ apply only in the case of reproduction rights infringement. Although CA section 19² is somewhat broader, it still sets clear limits on when it is possible to rely on the provision.

For input data, the most relevant exception arises from Directive 2001/29/EC of the European Parliament and the Council for personal use. A user of the GenAI can reproduce protected works as input data without infringing copyrights, provided that the model user had the author's consent to use the work, or the work was lawfully published. It is not yet entirely clear, however, whether the aforementioned exceptions pass the three-step test, which is necessary for reliance on the discussed free use exceptions. Therefore, the first hypothesis put forward is that the use of copyrighted works as training and input data for models leads to the infringement

of the author's exclusive rights, but existing free use exceptions may exclude liability for copyright infringement in certain cases.

The analysis of the second chapter also revealed how the proposed AI regulation affects the legitimacy of using protected work as training and input data and allowed for a conclusion about whether current law and the proposed AI regulation ensure a balance between the interests of the author and the innovation of GenAI. Problems related to the lawful training of models can be broadly divided into two categories. Firstly, problems related to the opt-out option arising from Article 4 (3) of the DSM Directive, implemented into Estonian law by CA section 19² (2), and secondly, problems related to the effective exercise of rights holders' exclusive rights. The obligations set forth in the AIA for model developers contribute to the lawful use of training data during model training and fine-tuning.

However, comparing with each other the protections offered to rights holders by the AIA and the extremely broad obligations set for model developers, there is a risk that the stringent requirements will result in developers leaving the European Union market and not making models available to consumers here, as, for instance, the exceptions facilitating model training in the United States are more lenient than in the European Union. In the author's view, an imbalance between the protection offered to rights holders and the obligations set for model developers may result in a situation that is detrimental to both parties – developers leave the European Union market, and thus authors are left without the protections which exist in the European Union.

Nevertheless, it may not be justified to change the existing legal framework with additional free use exceptions, as this could harm the legitimate interests of authors. The author's opinion is that the consideration should be given to create other uniform standards, especially those that allow authors to use their opt-out option effortlessly and exclude the use of works as training data for text and data mining purposes outside of research work. As a result, the second hypothesis proposed in the thesis regarding ensuring a balance between the protection of the author's interests and innovation of GenAI was only partially confirmed.

KASUTATUD LÜHENDID

AutÕS	Autoriõiguse seadus
CDPA	Copyright, Designs and Patents Act
EK	Euroopa Kohus
EMS	Eesti märksõnastik
EWHC	England and Wales High Court
OECD	Organisation for Economic Co-operation and Development
RKTK	Riigikohtu tsiviilkolleegium
TAKS	Teadus- ja arendustegevuse korralduse seadus
TsMS	Tsiviilkohtumenetluse seadustik
USC	United States Code
VÕS	Võlaõigusseadus
WIPO	World Intellectual Property Organization

KASUTATUD ALLIKAD

Kasutatud õigusaktid

1. Autoriõiguse seadus. – RT I, 29.06.2022, 16.
2. Berni kirjandus- ja kunstiteoste kaitse konventsioon. – RT II 1994, 16, 49.
3. Code de la propriété intellectuelle. Kättesaadav: https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006069414 (03.03.2024).
4. Copyright Act of 1976, 17. U.S.C. Kättesaadav: <https://www.copyright.gov/title17/> (13.12.2023).
5. Copyright, Designs and Patents Act 1988. Kättesaadav: <https://www.legislation.gov.uk/ukpga/1988/48/contents> (13.12.2023).
6. Ebaausa konkurentsi takistamise ja ärisaladuse kaitse seadus. – RT I, 07.12.2018, 2.
7. Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790, 17. aprill 2019, mis käsitleb autoriõigust ja autoriõigusega kaasnevaid õigusi digitaalsel ühtsel turul ning millega muudetakse direktiive 96/9/EÜ ja 2001/29/EÜ. – EÜT L 130/92, 17.05.2019.
8. Euroopa Parlamendi ja nõukogu direktiiv 2001/29/EÜ, 22. mai 2001, autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ühtlustamise kohta infoühiskonnas. – ELT L 167, 22.06.2001.
9. Euroopa Parlamendi ja nõukogu direktiiv 96/9/EÜ, 11. märts 1996, andmebaaside õiguskaitse kohta. – EÜT L 77, 27.03.1996.
10. Teadus- ja arendustegevuse korralduse seadus. – RT I, 19.03.2019, 92.
11. Tekijänoikeuslaki. Kättesaadav: <https://www.finlex.fi/fi/laki/ajantasa/1961/19610404> (03.03.2024).
12. Tsiviilkohtumenetluse seadustik. – RT I, 22.03.2024, 7.
13. Võlaõigusseadus. – RT I, 06.07.2023, 116.

Kasutatud kirjandus

14. Alzubaidi, L. jt. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data 08/2021.
15. Arvisto, Ä. Tehisintellekti loomingu autorsus. Tartu 2022.
16. Autoriõiguse seaduse muutmise seadus (autoriõiguse direktiivide ülevõtmine) 368 SE, lk 28. Kättesaadav: <https://www.riigikogu.ee/tegevus/eelnoud/eelnou/d3d07943-9d1c-4ebe-94a4-8ae1ebdf7a68> (12.12.2023).
17. Bently, L. jt. Intellectual Property Law (6th edition). Oxford University Press 2022.

18. Bommasani, R. jt. On the Opportunities and Risks of Foundation Models. ArXiv 2021.
19. Bonadio, E., Dinev, P., McDonagh, L. Can Artificial Intelligence Infringe Copyright? Some Reflections. Research Handbook on Intellectual Property and Artificial Intelligence 2022.
20. Cheng-Davies, T. Honour in UK Copyright Law is is Not „A Trim Reckoning“ – Its Impact on the Integrity Right and the Destruction of Works of Art. Oxford Journal of Legal Studies, Volume 36 Issue 2 2016.
21. Dermawan, A. Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese „nonenjoyment“ purposes? The Journal of World Intellectual Property 2023.
22. Dingliwal jt. Prompt Tuning GPT-2 language model for parameter-efficient domain adaptation of ASR systems. InterSpeech. ArXiv 2021.
23. Ducato, R., Strowel, A. Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to „Machine Legibility“. International Review of Intellectual Property and Competition Law 2018.
24. Euroopa Komisjon. Ettepanek: Euroopa Parlamendi ja Nõukogu määrus, millega nähakse ette tehisintellekti käsitlevad ühtlustatud õigusnormid (tehisintellekti käsitlev õigusakt) ja muudetakse teatavaid liidu õigusakte. 21.04.2021. Kättesaadav: <https://eur-lex.europa.eu/legal-content/EN-ET/TXT/?from=EN&uri=CELEX%3A52021PC0206> (06.09.2023).
25. Euroopa Parlamendi 13. märtsi 2024. aasta otsus tehisintellekti määruse kohta. Kättesaadav: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (18.03.2024).
26. Feuerriegel, S. jt. Generative AI. Business & Information Systems Engineering, 2024.
27. Franceschelli, G., Musolesi, M. Copyright in generative deep learning. Cambridge University Press 5/2022.
28. Geiger, C., Frosio, G., Bulayenko, O. jt. Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? International Review of Intellectual Property and Competition Law 2018.
29. Gervais, D. J. AI Derivatives: The Application to the Derivative Work Right to Literary and Artistic Productions of AI Machines. Seton Hall Law Review Vol. 53, 2022.
30. Gervais, D. J. Exploring the Interfaces Between Big Data and Intellectual Property Law. Journal of Intellectual Property, Information Technology and Electronic Commerce Law (JIPITEC) 10/2019.

31. Gillotte, J. L. Copyright Infringement in AI-Generated Artworks. *UC Davis Law Review*, Vol. 53 No. 5 2020.
32. Guadamuz, A. A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs. *GRUR International* 2/2024.
33. Hacker, P. jt. Regulating ChatGPT and other Large Language Generative AI Models. *ArXiv* 2023.
34. Harshvardhan, G. M. jt. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 11/2020.
35. Justin K. The Right of Communication to the Public in EU Copyright Law. Bloomsbury Publishing 2020.
36. Kelli, A jt. Vestlusrobotid ja autoriõigus. *Juridica* 5/2020.
37. Kelli, A. jt. The Extent of Legal Control over Language Data: the Case of Language Technologies. *Proceedings of CLARIN Annual Conference 2019: CLARIN 73 Annual Conference, Leipzig, Germany, 30 September – 2 October 2019*. Ed. K. Simov and M. Eskevich. CLARIN. Kättesaadav: <http://hdl.handle.net/10138/327736> (09.12.2023).
38. Krotov, V., Johnson, L., Silva, L. Tutorial: Legality and Ethics of Web Scraping. *Communications of the Association for Information Systems* 2020.
39. Kumar, R. jt. Training AI and Copyright Infringement: Where does the law stand? *Indian Journal of Integrated Research in Law*, Volume II Issue I 2022.
40. Lee, K. jt. Talkin' Bout AI Generation: Copyright and the Generative-AI Supply Chain (July 27, 2023). Forthcoming, *Journal of the Copyright Society* 2024.
41. Lucchi, N. ChatGPT: A Case Study on Copyright Challenges for Generative AI Systems. *European Journal of Risk Regulation* 2023.
42. M. A., Casey, B. Fair Learning. *Texas Law Review* Volume 99, Issue 4, 2021,
43. Margoni, T. Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI? *CREATE Working Paper* 12/2018.
44. Margoni, T., Kretschmer, M. A deeper look into EU Text and Data Mining Exceptions: Harmonisation, Data Ownership and the Future of Technology. *GRUR International* 8/2022.
45. Matulionyte, R. Can AI Infringe Moral Rights of Authors and Should We Do Anything About it: An Australian Perspective. *Law, Innovation and Technology* 2023.
46. Matulionyte, R. Reconciling Trade Secrets and AI Explainability: Face Recognition Technologies as a Case Study. *European Intellectual Property Review* 2022.

47. Max Planck Institute for Innovation and Competition. Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate. Max Planck Institute for Innovation and Competition Research Paper No. 21-10 2021.
48. Novelli, C. jt. Generative AI in EU Law: Liability, Privacy, Intellectual Property, Cybersecurity 2024.
49. Ørstavik, I. B. Access to data for training algorithms in machine learning: copyright law and “right-stacking”. Edward Elgar Publishing 2022.
50. Pandey, M. jt. The transformational role of GPU computing and deep learning in drug discovery. Nature Machine Intelligence 2022.
51. Pisuke, H. Autoriõiguse alused. Tallinn: Ellington Trükiagentuur 2006.
52. Quang, J. Does training AI violate copyright? Berkeley Technology Law Journal 1/2023.
53. Rabin, M. R. I. jt. Memorization and Generalization in Neural Code Intelligence Models. Information and Software Technology, Elsevier 2021.
54. Radiya-Dixit E., Wang, X. How fine can fine-tuning be? Learning efficient language models. International Conference on Artificial Intelligence and Statistics. Italy: PMLR Volume 108 2020.
55. Ray, P. P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, Volume 3 2023.
56. Rosati, E. Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790. Oxford University Press 2021.
57. Rosati, E. Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity. Asia Pacific Law Review 2020.
58. Rosati, E. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Technical Aspects. European Parliament: 2018.
59. Rosentau, M. Intellektuaalse omandi õigused infotehnoloogias. Autori isiklikud õigused. Juridica 9/2007.
60. Rosentau, M. Intellektuaalse omandi õigused infotehnoloogias. Autori varalised õigused. Juridica 10/2010.
61. Sandiumenge, I. Copyright Implications of the Use of Generative AI. Universitat Pompeu Fabra. SSRN 2023. Kättesaadav: <http://dx.doi.org/10.2139/ssrn.4531912> (04.01.2024).

62. Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science 2021.
63. Senftleben, M. R. F. Copyright, Limitations and the Three-step test. An Analysis of the Three-Step Test in International and EC Copyright Law. Kluwer Law International, 2004.
64. Sõrmus, S. Digiühiskonna direktiivis sätestatud teksti- ja andmekaeve erandi autoriõiguslikud väljakutsed. Tallinn 2020.
65. The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 30.10.2023. Kättesaadav: <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> (04.03.2024).
66. UK Parliament. Artificial Intelligence (Regulation) Bill. HL Bill 11. 22.11.2023. Kättesaadav: <https://bills.parliament.uk/publications/53068/documents/4030> (02.03.2024).
67. Uszkoreit, J. Transformer: A Novel Neural Network Architecture for Language Understanding. 31st Conference on Neural Information Processing Systems NIPS 2017. Kättesaadav: <https://arxiv.org/pdf/1706.03762> (11.01.2024).
68. Vahlberg, J.-A. Tehisintellekti loomingu autoriõiguslik kaitse. Tartu 2017.
69. Varul, P. jt. Võlaõigusseadus II. Komm vlj. Juura 2019.
70. Vaswani, A. jt. Attention is All You Need. 31st Conference on Neural Information Processing Systems. NIPS 2017.
71. Vilbaste, K. Keeletehnoloogia ja Bürokratt. Õiguskeel 2/2023.
72. Wischmeyer, T., Rademacher, T. Regulating Artificial Intelligence. Springer 2020.
73. Wymeersch P. EU Copyright Exceptions and Limitations and the Three-Step Test: One Step Forward, Two Steps Back. GRUR International 7/2023.
74. Yew, R.-J. Break It 'Til You Make It An Exploration of the Ramifications of Copyright Liability Under a Pre-training Paradigm of AI Development. Proceedings of the Symposium on Computer Science and Law 3/2024.

Kasutatud kohtupraktika

75. RKTKo 2-16-8751.
76. RKTKo 2-16-17491/52.
77. RKTKo 2-14-5664.
78. RKTKo 2-16-8751/184.

79. RKTko 3-2-1-67-07.
80. RKTko 2-16-9016/34.
81. RKTko 3-2-1-173-12.
82. EKo C-476/17, *Pelham jt.*, ECLI:EU:C:2019:624.
83. EKo C-762/19, *CV-Online Latvia SIA v. Melons SIA*, ECLI:EU:C:2021:434.
84. EKo C-5/08, *Infopaq International A/S v. Danske Dagblades Forening*, ECLI:EU:C:2009:465.
85. Ekm C-302/10, *Infopaq International A/S v. Danske Dagblades Forening*, ECLI:EU:C:2012:16.
86. EKo C-30/14, *Ryanair Ltd v. PR Aviation BV*, ECLI:EU:C:2015:10.
87. EKo C-403/08, *Football Association Premier League Ltd v QC Leisure*, EU:C:2011:631.
88. EKo C-310/17, *Levola Hengelo*, ECLI:EU:C:2018:899.
89. EKo C-160/15, *GS Media BV v Sanoma Media Netherlands BV jt.* ECLI:EU:C:2016:644.
90. EKo C-360/13, *Meltwater*, ECLI:EU:C:2014:1195.
91. EKo C-162/10 *Phonographic Performance (Ireland)*, ECLI:EU:C:2012:141.
92. EKo C-457/11 *VG Wort/Kyocera*, ECLI:EU:C:2013:246.
93. EKo C-527/17, *Filmspeler*, EU:C:2017:300.
94. EKo C-466/12, *Nils Svensson jt v Retriever Sverige AB* ECLI:EU:C:2014:76.
95. EKo C-597/19, *Mircom*, EU:C:2020:1063.
96. EKo C-263/18, *Tom Kabinet*, EU:C:2019:111.
97. EKo C-610/15, *Stichting Brein*, EU:C:2006:764.
98. EKo C-135/10, *SCF*, EU:C:2012:140.
99. EKo C-265/16, *VCAST v. RTI SpA*, EU:C:2017:913.
100. EKo C-392/19, *VG Bild Kunst*, EU:C:2020:696.
101. EKo C-145/10, *Painer*, EU:C:2011:798.
102. *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202 (2nd Circ. 2015).
103. *Authors Guild v. HathiTrust*, 755 F.3d 87 (2nd Cir. 2014).
104. *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007).
105. *Paul Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223 (menetlus pooleli).
106. *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 1:24-cv-01515 (menetlus pooleli).
107. *Sarah Silverman, Christopher Golden & Richard Kadrey v. OpenAI, Inc.*, No. 3:23-cv-03416 (menetlus pooleli).
108. *New York Times v. OpenAI, Inc.*, No. 1:23-cv-11195 (menetlus pooleli).
109. *FAPL v. British Communications*, 2017 EWHC 480.

110. R (British Academy of Songwriters, Composers and Authors and others) v Secretary of State for Business, Innovation and Skills, 2015 EWHC 2041.

Muud allikad

111. AlphaSense. The Future of Generative AI: An Analysis of the Leaders, Opportunities and Threats. 01/2023. Kättesaadav: <https://www.alpha-sense.com/resource/report/the-future-of-generative-ai-an-analysis-of-the-leaders-opportunities-and-threats/> (17.01.2024).
112. Anguiano, D. Hollywood writers strike ends studio deal. – The Guardian, 27.09.2023. Kättesaadav: <https://www.theguardian.com/culture/2023/sep/26/hollywood-writers-strike-ends-studio-deal> (08.12.2023).
113. Baio, A. AI Data Laundering: How Academic and Nonprofit Researches Shield Tech Companies from Accountability. – Waxy Blog, 30.09.2022. Kättesaadav: <https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/> (03.04.2024).
114. Cocoru, D. Boehm, M. An analytical review of text and data mining practices and approaches in Europe. Policy recommendations in view of the upcoming copyright legislative proposal. Openforum, Europe 01.05.2016. Kättesaadav: <https://openforumeurope.org/wp-content/uploads/2016/05/TDM-Paper-Diana-Cocoru-and-Mirko-Boehm-1.pdf> (02.02.2024).
115. Commission of the European Communities. Putting knowledge into practice: A broad-based innovation strategy for the EU. 13.9.2006, lk 6. Kättesaadav: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52006DC0502> (07.01.2024).
116. Communia Policy Paper. Using Copyrighted Works for Teaching the Machine. Communia Association, 26.04.2023. Kättesaadav: <https://communia-association.org/wp-content/uploads/2023/04/Policy-Paper-15-on-using-copyrighted-works-for-teaching-the-machine.pdf> (02.03.2024).
117. Eesti Statistikaamet. Tehisintellekti tehnoloogiate kasutamine ettevõtetes on tõusutrendis. 15.09.2023. Kättesaadav: <https://www.stat.ee/et/uudised/tehisintellekti-tehnoloogiate-kasutamine-ettevotetes-tousutrendis> (17.01.2024).
118. Emanuilov, I. Margoni, T. Memorisation in generative models in EU copyright law: an interdisciplinary view. – Kluwer Copyright Blog, 26.03.2024. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2024/03/26/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/> (04.04.2024).

119. Euroopa Komisjon. Euroopa tehisintellekti amet. Kättesaadav: <https://digital-strategy.ec.europa.eu/en/policies/ai-office> (28.02.2024).
120. Google DeepMindi avaldatud raport. Google DeepMind. Gemini: A Family of Highly Capable Multimodal Models. 02.04.2024, lk 3. Kättesaadav: <https://arxiv.org/pdf/2312.11805.pdf> (12.04.2024).
121. Guadamuz, A. Photographer sues LAION for copyright infringement. 05.05.2023. Kättesaadav: <https://www.technollama.co.uk/photographer-sues-laion-for-copyright-infringement> (07.01.2024).
122. House of Lords. Large language models and generative AI. 1st Report of Session 2023-24. 02.02.2024. Kättesaadav: <https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/54.pdf> (20.03.2024).
123. Hugenholtz, P. B. The New Copyright Directive Text and Data Mining Articles 3 and 4. – 24.07.2019, Kluwer Copyright Blog. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/> (20.02.2024).
124. Intellectual Property Office. The Government's Code of Practice on Copyright and Artificial Intelligence. 29.06.2023. Kättesaadav: <https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai> (06.03.2024).
125. Keller, P. A first look at the copyright relevant parts in the final AI Act compromise. – Kluwer Copyright Blog, 11.12.2023. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/> (25.03.2024).
126. Keller, P., Warso, Z. Defining best practices for opting out of ML training. – Open Future, 29.09.2023. Kättesaadav: https://openfuture.eu/wp-content/uploads/2023/09/Best-_practices_for_optout_ML_training.pdf (19.02.2024).
127. Langley, H. Top websites block Google from training AI models on their data. – Yahoo News, 15.03.2024. Kättesaadav: <https://www.yahoo.com/tech/top-websites-block-google-training-194847790.html> (23.04.2024).
128. Majandus- ja Kommunikatsiooniministeerium jt. Eesti tehisintellekti tegevuskava 2024–2026. 2024.
129. Matulionyte, R. The (forgotten) moral rights of the age of AI. – Kluwer Copyright Blog, 07.02.2022. Kättesaadav: <https://copyrightblog.kluweriplaw.com/2022/02/07/the-forgotten-moral-rights-in-the-age-of-ai/> (30.03.2024).

130. Montagnon, R., Cho, S. UK Withdraws Plans For Broader Text And Data Mining (TDM) Copyright And Database Right Exception. – Mondaq, 07.03.2023. Kättesaadav: <https://www.mondaq.com/uk/copyright/1290378/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception> (06.03.2024).
131. OECD. AI Principles. 2019. Kättesaadav: <https://oecd.ai/en/ai-principles> (02.02.2024).
132. Reisner, Alex. The Flaw That Could Ruin Generative AI. – The Atlantic, 11.01.2024. Kättesaadav: <https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/> (08.03.2024).
133. Select Committee on Artificial Intelligence of the National Science and Technology Council. The National Artificial Intelligence Research and Development Strategic Plan: 2023 Update. May 2023. Kättesaadav: <https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf> (15.03.2024).
134. Thompson, A. D. What's in my AI? A Comprehensive Analysis of Datasets used to Train GPT. 2022, lk 6. Kättesaadav: <https://life architect.ai/whats-in-my-ai/> (02.02.2024).
135. United States Copyright Office. Artificial Intelligence and Copyright, 30.08.2023. Kättesaadav: <https://www.copyright.gov/ai/docs/Federal-Register-Documents-Artificial-Intelligence-and-Copyright-NOI.pdf> (28.03.2024).
136. WIPO. Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence. 21.05.2020. Kättesaadav: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf (09.09.2023).