

TARTU ÜLIKOOL
ÕIGUSTEADUSKOND
Eraõiguse osakond

Narine Urban

**GENERATIIVSE TEHISINTELLEKTI MUDELITE TREENIMINE
PIRAATALLIKATELT KOGUTUD ANDMETEL: AUTORIÕIGUSE RIKKUMINE
VÕI ÕIGUSTATUD KASUTUS EUROOPA LIIDU JA AMEERIKA ÜHENDRIIKIDE
ÕIGUSES?**

Magistritöö

Juhendaja:
Prof. Aleksei Kelli

Tallinn
2026

SISUKORD

SISSEJUHATUS.....	3
1. AI TREENIMISE ÕIGUSLIKUD RAAMISTIKUD: EL JA USA.....	7
1.1 EL-i autoriõiguse raamistik AI treenimise kontekstis.....	7
1.1.1 DSM-direktiivi artiklid 3 ja 4: TDM-erandite struktuur.....	7
1.1.2 ELTK kohtupraktika TDM kontekstis: ACI Adam ja implikatsioonid.....	8
1.1.3 Tehisintellekti käsitleva määruse artikkel 53: ÜTEM-kohustused ja autoriõiguse vastavus.....	11
1.1.4 Eesti transponeerimise eripära: varasem TDM-erand vs DSM.....	13
1.2 USA autoriõiguse raamistik: <i>fair use</i> 'i neljaetapiline test.....	15
1.2.1 17 U.S.C. § 107 ja neljaetapiline test AI treenimise kontekstis.....	15
1.2.2 <i>Google v. Oracle ja Warhol Foundation</i> : transformatiivsuse pretsedendid.....	16
1.2.3 USA Autoriõiguse Ameti 2025. aastaaruanne: ametlik seisukoht.....	17
1.3 EL-i ja USA lähenemiste võrdlev kokkuvõte.....	18
2. PIRAATALLIKATE KÜSIMUS EL-I AUTORIÕIGUSES.....	20
2.1. Seadusliku juurdepääsu nõue piraatallikate kontekstis.....	20
2.2. <i>GEMA v. OpenAI</i> : memoriseerimine reproduktsioonina.....	23
2.3. <i>Kneschke v. LAION</i> : TDM-erandite kohaldamine ja <i>opt-out</i> mehhanism.....	25
2.4. Piraatluse mõju TDM-erandi kehtivusele: õigusdogmaatiline analüüs.....	28
2.5. Argumendid õigustatud kasutuse kasuks ja nende kriitika.....	31
3. PIRAATLUSE KÜSIMUS USA ÕIGUSES: <i>BARTZ, KADREY JA THOMSON REUTERS</i>.....	35
3.1. <i>Fair use</i> neljaetapiline test piraatandmete kontekstis.....	35
3.2. <i>Bartz v. Anthropic</i> : lahutamise doktriin.....	37
3.3. <i>Kadrey v. Meta Platforms</i> : ühtne hindamine ja turu kahjustamise teooria.....	39
3.4. Lahutamise doktriini ja ühtse hindamise lähenemise kriitiline võrdlus: kumb on doktrinaalselt põhjendatum?.....	43
3.5 <i>Thomson Reuters v. ROSS Intelligence</i> : Warholi standard ja konkureeriv kasutus.....	44
3.6. EL ja USA lähenemiste normatiivne võrdlus.....	48
4. SAASTUNUD MUDELI PROBLEEM JA <i>DE LEGE FERENDA</i> ETTEPANEKUD	49
4.1. Saastunud mudeli probleem: algoritmiline eraldamine.....	49
4.2. Praeguse regulatiivse raamistiku puudused.....	53
4.3. <i>De lege ferenda</i> ettepanekud.....	55
KOKKUVÕTE.....	62
SUMMARY.....	65
KASUTATUD ALLIKATE LOETELU.....	70
KASUTATUD KIRJANDUS.....	70
KASUTATUD ÕIGUSAKTID.....	74
KASUTATUD KOHTUPRAKTIKA.....	76

SISSEJUHATUS

Generatiivsed tehisintellekti mudelid – nagu OpenAI GPT-4, Anthropic’u Claude ja Meta LLaMA – kujutavad endast ühte määravamat tehnoloogilist arengut 21. sajandil. Nende mudelite treenimisprotsess hõlmab kahte autoriõiguslikult olulist toimingut: andmete kogumine (kopeerimine) ja mudeli eeltreenimine, mille käigus tekib memoriseerimise nähtus – mudeli võime teatavaid treeningandmete tekstijätke sõna-sõnalt taasesitada.

Autoriõigus on ajalooliselt reguleerinud teoste kasutamist inimestele suunatud väljundites – esmakordselt seisab õigus silmitsi olukorraga, kus teoseid kasutatakse masina sisendina, millest inimesele nähtavat väljundit ei teki. Just see struktuurne uudsus on tõstatanud autoriõiguslikud küsimused, millele kehtiv rahvusvaheline õigusraamistik ei ole seni suutnud ühtset vastust anda.

Treenimisandmete kogumise käigus ei ole AI-ettevõtted alati piirdunud seaduslikult ligipääsetava sisuga. Kohtuvaidlustes esitatud faktilised väited ja osaliselt tuvastatud asjaolud viitavad sellele, et treeningandmestikesse on kaasatud autoriõigusega kaitstud teoseid, mis pärinevad ebaseaduslikest veebikeskkondadest nagu LibGen (Library Genesis), Books3 ja Pirate Library Mirror.¹

Käesolevas töös kasutatakse terminit "piraatallikas" kirjeldava mõistena, tähistamaks veebikeskkondi ja andmekogumeid, mille kaudu tehakse autoriõigusega kaitstud teoseid kättesaadavaks ilma õiguste omajate loata. Õiguslikus tähenduses käsitatakse selliseid allikaid ebaseaduslikena Euroopa Liidu Kohtu praktika põhjal, eelkõige kohtuotsuses ACI Adam (C-435/12) kujunenud lähenemise kohaselt, mille järgi autoriõiguse erandid ei hõlma kopeerimist, mis põhineb ebaseaduslikult kättesaadavaks tehtud teostel.² Sellest tulenevalt kasutatakse käesolevas töös mõistet "piraatallikas" faktilise kategooriana, mille õiguslik kvalifikatsioon vastab "ebaseadusliku allika" mõistele Euroopa Liidu autoriõiguse doktriinis.

Selle asjaolu õiguslik kvalifitseerimine on käesoleva magistritöö keskseks uurimisobjektiks. Kui treeningandmestik sisaldab autoriõigusega kaitstud materjale, mis on omandatud ebaseaduslikest allikatest, tekib küsimus, kas selline treenimistegevus kujutab endast autoriõiguse rikkumist või võib see kuuluda autoriõiguse erandite ja piirangute kohaldamisalasse Euroopa Liidu õiguses. Käesolev töö lähtub seisukohast, et Euroopa Liidu

¹ Gao, L. jt. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. arXiv:2101.00027v1 [cs.CL], 2021, lk 1, 6-7; *Bartz v. Anthropic PBC*, nr 3:24-cv-05417-WHA (N.D. Cal., 23. juuni 2025), lk 4-7; *Kadrey v. Meta Platforms, Inc.*, nr 3:23-cv-03417-VC (N.D. Cal., 25. juuni 2025), lk 3-5.

² EKo 10.04.2014, C-435/12, *ACI Adam BV jt versus Stichting de Thuisakopie ja Stichting Onderhandelinge Thuisakopie vergoeding*, ECLI:EU:C:2014:254.

kehtiva õiguse kõige põhjendatuma tõlgenduse kohaselt ei ole piraatallikatelt pärinevate andmete kasutamine TDM-eranditega kaetud. Kuna Euroopa Liidu Kohus ei ole seda küsimust veel otseselt lahendanud, käsitatakse seda järeldust doktrinaalselt tugeva, kuid lõplikku kohtulikku kinnitust ootava seisukohana. Võrdlevas plaanis analüüsitakse, milline on sama tegevuse õiguslik kvalifikatsioon Ameerika Ühendriikide õiguses.

Käesolevas töös eristatakse analüüsi selguse huvides kolme tasandit. Esiteks andmete omandamist ebaseaduslikust allikast. Teiseks nende andmete kasutamist TDM-i või mudeli treenimise protsessis. Kolmandaks olukorda, kus mudel suudab treeningandmete osi hiljem sõna-sõnalt või sellele väga lähedaselt taasesitada. Töö põhiseisukoht on, et Euroopa Liidu õiguses välistab juba esimene tasand üldjuhul TDM-erandile tuginemise. Memoriseerimine ei asenda seda argumenti, vaid kujutab endast sellele lisanduvat iseseisvat autoriõiguslikku probleemi.³

Käesoleva magistritöö keskseks uurimisküsimuseks on, kas ja millistel tingimustel kujutab piraatallikatelt (st ebaseaduslikest allikatest) kogutud andmetel põhinev generatiivse tehisintellekti mudelite treenimine endast autoriõiguse rikkumist Euroopa Liidu õiguse kohaselt ning milline on selle tegevuse õiguslik kvalifikatsioon võrdluses Ameerika Ühendriikide õigusega.

Teema õiguslik ebaselgus on 2026. aasta seisuga erakordselt aktuaalne kolme paralleelse arengu tõttu. Esiteks on USA föderaalkohtupraktikas 2025. aastal kujunenud esimesed madalama astme kohtulahendid ja pooleliolevad kohtuvaidlused, mis puudutavad otseselt piraatlikelt allikatelt pärit treenimisandmeid.

Teiseks on Euroopa Liidus kujunenud 2025. aastal esimene kohtupraktika AI treenimise autoriõigusliku lubatavuse küsimuses, mis annab varasema akadeemilise kirjanduse teoreetilisele analüüsile sisulise kohtuliku kinnituse. Kolmandaks on EL-i õigusliku raamistiku tasandil toimunud oluline institutsionaalne areng: Euroopa Parlamendi ja nõukogu tehisintellektiseaduse üldotstarbeliste tehisintellekti mudelite kohustused jõustusid 2. augustil 2025, seades treeningandmete läbipaistvusele ja autoriõiguste järgimisele esmakordselt siduvad avalik-õiguslikud nõuded.⁴

³ *Bartz v. Anthropic PBC*, nr 3:24-cv-05417-WHA (N.D. Cal., 23. juuni 2025), kohtunik Alsup, lk 12-14. Kohus leidis, et piraatlikesse raamatukogusse allalaadimine on eraldiseisev autoriõiguse rikkumine, mida hilisem transformatiivne kasutamine ei õigusta. ; *Kadrey v. Meta Platforms, Inc.*, nr 3:23-cv-03417-VC (N.D. Cal., 25. juuni 2025), kohtunik Chhabria, lk 8-12. Kohus keeldus treenimist allalaadimisest lahutamast ning andis *Meta-le fair use'i* alusel kokkuvõtva otsuse, hoolimata sellest, et Meta oli kasutanud BitTorrenti teel allalaaditud raamatuid.

⁴ Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024, milles käsitletakse tehisintellektiga seotud ühtlustatud eeskirju (tehisintellekti käsitlev määrus). – ELT L, 12.07.2024. Jõustus 1. augustil 2024; V

Käesoleva magistritöö eesmärk on vastata küsimusele, kas generatiivse tehisintellekti mudelite treenimine piraatallikatelt kogutud autoriõigusega kaitstud andmetel kujutab endast autoriõiguse rikkumist Euroopa Liidu õiguse kohaselt ning selgitada võrdleva analüüsi kaudu, millised on selle küsimuse lahendamise alternatiivsed käsitlused Ameerika Ühendriikide õiguses.

Käesolev töö lähtub Euroopa Liidu õigusest kui peamisest normatiivsest raamistikust. Ameerika Ühendriikide õigust kasutatakse eeskätt võrdleva analüüsi vahendina, mille eesmärk on kontrasti kaudu esile tuua Euroopa Liidu autoriõiguse süsteemi struktuurseid eripärasid.

Käesoleva magistritöö hüpotees on, et Euroopa Liidu kehtiv autoriõiguslik raamistik, eelkõige "seadusliku juurdepääsu" nõude kaudu, välistab piraatallikatelt pärinevate treenimisandmete kasutamise TDM-erandite kohaldamisalast normatiivse ülesehituse poolest selgemalt ja ettearvatavamalt kui Ameerika Ühendriikide *fair use*'i doktriin, kuid samal ajal ei paku kumbki süsteem piisavalt tõhusaid õiguskaitsevahendeid juba treenitud mudelite suhtes, mistõttu on vajalik täiendav normatiivne sekkumine.

Püstitatud uurimisküsimustele vastamiseks kasutatakse käesolevas töös õigusdogmaatilist meetodit, mis hõlmab kehtivate õigusaktide tõlgendamist grammatilise, süstemaatilise, ajaloolise ja teleoloogilise meetodi abil.

Töö põhiliseks võrdlusmeetodiks on horisontaalne ja funktsionaalne võrdlev analüüs Euroopa Liidu ja Ameerika Ühendriikide õiguse vahel. Võrdleva meetodi eesmärk ei ole käsitleda mõlemat süsteemi normatiivselt võrdsena ega kasutada USA õigust Euroopa Liidu õiguse tõlgendamise otsese allikana, vaid kasutada seda kontrastse regulatiivse mudelina, mille abil esile tuua Euroopa Liidu autoriõiguse süsteemi struktuurseid eripärasid.

Ameerika Ühendriikide õiguse kaasamine on põhjendatud kolmel tasandil: (i) EL ja USA esindavad põhimõtteliselt erinevaid autoriõiguse erandite süsteeme (piiratud erandid vs avatud norm), (ii) USA kohtupraktikas on kujunenud esimesed AI treenimist puudutavad lahendid, ning (iii) tehisintellekti arendajad tegutsevad globaalsel turul, mistõttu USA õiguse arengud omavad ka kaudset mõju Euroopa turule.

Kohtupraktika analüüs moodustab töö primaarse allikmaterjali. Kohtulahendeid käsitletakse doktrinaalse analüüsi objektina, mitte empiirilise andmestikuna kitsas metodoloogilises tähenduses.

peatükk (ÜTEM-kohustused) kehtib alates 2. augustist 2025 vastavalt art 113 lg 1 p b-le (*edaspidi*: tehisintellekti käsitlev määrus).

Töö analüüsi struktureerimiseks käsitletakse nelja alauurimisküsimust:

- (1) Kas DSM-direktiivi⁵ "seadusliku juurdepääsu" nõue välistab piraatallikatelt pärinevate andmete kasutamise TDM-erandite kohaldamisalas?
- (2) Kuidas kohaldub USA *fair use*'i doktriin piraatandmetel põhinevale AI treenimisele?
- (3) Millised õiguskaitsevahendid on kättesaadavad juba treenitud mudelite suhtes?
- (4) Millised *de lege ferenda* muudatused on vajalikud tuvastatud lünkade kõrvaldamiseks?

Kohtupraktika analüüs moodustab töö peamise primaarse allikmaterjali. Töös kasutatavad allikad on eranditult akadeemilise tasemega: eelretsenseeritud teadusartiklid, institutsionaalsed raportid ning autoriõiguse valdkonna juhtivate teadlaste monograafiad. Blogipostitustele ja mitteteaduslikele veebiallikatele töö analüütilises osas ei toetuta.

Generatiivse tehisintellekti treenimise autoriõiguslikke küsimusi on teaduskirjanduses käsitletud nii Euroopa kui Ameerika perspektiivist. Quintais on analüüsinud DSM-direktiivi tekstikaeve-erandeid ja nende piire,⁶ Senftleben pakub välja väljundipõhise tasustamismehhanismi kui alternatiivi kehtivale regulatsioonile,⁷ ning Sag ja Lemley koos Casey'ga on argumenteerinud, et masinõpe kujutab endast mitteväljenduslikku kasutust USA autoriõiguse tähenduses.⁸ Küsimus, kas DSM-direktiivi seadusliku juurdepääsu nõue välistab kategooriliselt piraatlikelt allikatelt pärit treeningandmete kasutamise ning kuidas see erineb USA *fair use*'i doktriinist, on aga seni teaduskirjanduses süstemaatilise käsitluseta.

Püstitatud eesmärgi saavutamiseks on töö jagatud neljaks peatükiks⁹. Esimeses peatükis analüüsitakse EL-i ja USA autoriõiguse raamistikku. Teine peatükk käsitleb piraatandmete küsimust EL-i õiguses, kolmas peatükk USA õiguses ning neljas peatükk keskendub *de lege ferenda* ettepanekutele.

Märksõnad: andmekaeve, autoriõigus, digitaalpiraatus, generatiivne tehisintellekt, intellektuaalomand

⁵ Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790, 17. aprill 2019, autoriõiguse ja sellega kaasnevate õiguste kohta digitaalsel ühisel turul ning millega muudetakse direktiivid 96/9/EÜ ja 2001/29/EÜ. – ELT L 130, 17.05.2019, lk 92–125 (*edaspidi*: DSM-direktiiv).

⁶ J. P. Quintais. Generative AI, copyright and the AI Act. – Computer Law & Security Review 2025/56, art. 106107.

⁷ M. R. F. Senftleben. Generative AI and Author Remuneration. – IIC 2023/54, lk 1535-1560.

⁸ M. Sag. Fairness and Fair Use in Generative AI. – Fordham Law Review 2024/92, lk 1887-1921. ; M. A. Lemley, B. Casey. Fair Learning. – Texas Law Review 2021/99, lk 743-782.

⁹ Töö pealkirjas kasutatud väljendit "õigustatud kasutus" ei kasutata Euroopa Liidu õiguse kontekstis Ameerika Ühendriikide *fair use*'i doktriini tehnilise vastena. Euroopa Liidu õiguses tähistab see käesolevas töös üksnes küsimust, kas vaadeldav kasutus võib kuuluda autoriõiguse erandite ja piirangute kohaldamisalasse, eelkõige DSM-direktiivi teksti- ja andmekaeve erandite alla.

1. AI TREENIMISE ÕIGUSLIKUD RAAMISTIKUD: EL JA USA

Generatiivsete AI mudelite treenimisprotsessis tuleb autoriõiguslikult eristada vähemalt kahte põhitoomingut: andmete kogumist internetist ning nende kasutamist mudeli treenimisel. Neist eraldi tuleb käsitleda ka memoriseerimise nähtust, st olukorda, kus mudel suudab hiljem treeningandmete osi sõna-sõnalt või sellele väga lähedaselt taasesitada. Käesolev töö käsitleb kõiki kolme küsimust piraatallikatelt pärinevate andmete kontekstis. Analüüsi loogika on järgmine: esmalt hinnatakse, kas andmete päritolu välistab üldse TDM-eranditele tuginemise; seejärel käsitletakse, kas ja millises ulatuses tekitab treenimise tulemusena ilmnev memoriseerimine täiendava iseseisva reproduktsiooniprobleemi.

1.1 EL-i autoriõiguse raamistik AI treenimise kontekstis

1.1.1 DSM-direktiivi artiklid 3 ja 4: TDM-erandite struktuur

Euroopa Liidu autoriõiguse ajakohastamise keskseks instrumendiks on Euroopa Parlamendi ja nõukogu 17. aprilli 2019. aasta direktiiv (EL) 2019/790, mis käsitleb autoriõigust ja autoriõigusega kaasnevaid õigusi digitaalsel ühisel turul (edaspidi: DSM-direktiiv).¹⁰ Direktiivi II peatükk kehtestab teksti- ja andmekaeve (edaspidi: TDM) erandid, mille struktuur on generatiivse AI treenimise autoriõigusliku analüüsi jaoks määrava tähtsusega. DSM-direktiivi artikkel 2 lõige 2 defineerib TDM-i kui "automatiseeritud analüüsimeetodit, millega analüüsitakse digitaalkujul tekste ja andmeid, et saada teavet muu hulgas mustrite, suundumuste ja korrelatsioonide kohta".¹¹ Eesti õiguses on see definitsioon üle võetud AutÕS § 17¹ lõikes 1.¹²

DSM-direktiiv loob kahetasandilise TDM-erandite süsteemi. Artikkel 3 sätestab kitsama teadusuuringute erandi: teadus- ja kultuuripärandi asutused võivad reprodutseerida ja ekstraheerida teoseid, millele neil on "seaduslik juurdepääs" (ingl k *lawful access*), TDM-i eesmärgil ilma autori nõusolekuta.¹³ Artikkel 4 sätestab üldise TDM-erandi, mis kehtib igale isikule – sealhulgas ärilistele üksustele ja AI-ettevõtetele – ning lubab reprodutseerida ja ekstraheerida "õiguspäraselt ligipääsetavaid" teoseid mis tahes eesmärgil, kaasa arvatud

¹⁰ Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790, 17. aprill 2019, autoriõiguse ja sellega kaasnevate õiguste kohta digitaalsel ühisel turul ning millega muudetakse direktiivid 96/9/EÜ ja 2001/29/EÜ. - ELT L 130, 17.05.2019, lk 92-125 (edaspidi: DSM-direktiiv).

¹¹ DSM-direktiiv, art 2 lg 2.

¹² Autoriõiguse seadus (AutÕS). - RT I 1992, 49, 615; RT I, 12.07.2025, 9 (jõust. 01.10.2025), § 17¹ lg 1.

¹³ DSM-direktiiv, art 3 lg 1; Margoni, T.; Kretschmer, M. A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. - GRUR International, kd 71, nr 8, 2022, lk 685, 688.

kaubanduslikuks AI treenimiseks.¹⁴ Artikli 4 oluline piirang on õiguste omajate *opt-out*-mehhanism lõike 3 kohaselt: autorid võivad oma teoste kasutamise TDM-ks sõnaselgelt välistada asjakohasel viisil, sealhulgas masinloetavate vahenditega veebis üldkättesaadavaks tehtud sisu puhul.¹⁵ Artiklil 3 on lisaks tugev lepinguline kaitse: lepingu tingimused, mis piiravad teadusuuringute TDM-erandit, on tühised.¹⁶

Mõlemad artiklid sisaldavad ühise eeltingimusega nõuet, et teosel peab olema "seaduslik juurdepääs" või "õiguspäraselt ligipääsetavus". DSM-direktiiv jätab selle mõiste operatiivnormi tasandil defineerimata; direktiivi põhjenduspunkt 14 loetleb avatud juurdepääsu poliitika, lepingulisel kokkuleppel põhineva juurdepääsu ning muu seaduslikul teel saadud juurdepääsu, sealhulgas veebis vabalt kättesaadavaks tehtud sisu. Mõiste defineerimata jätmine operatiivnormi tasandil on akadeemilises kirjanduses tekitanud märkimisväärse tõlgendusvaidluse.¹⁷

1.1.2 ELTK kohtupraktika TDM kontekstis: ACI Adam ja implikatsioonid

Euroopa Liidu Kohtul puudub seni otsene lahend teksti- ja andmekaeve erandite kohta DSM-direktiivi alusel.¹⁸ Sellest hoolimata on ELTK varasem kohtupraktika, eelkõige kohtuasi *ACI Adam*, loonud doktrinaalse aluse, millele tuginedes on võimalik vastata käesoleva magistr töö kesksele küsimusele: kas piraatlikelt allikatelt pärit andmete kasutamine AI treenimiseks mahub DSM-direktiivi artiklite 3 ja 4 kohaldamisalasse.

Kohtuasjas *ACI Adam BV jt versus Stichting de ThuisKopie (C-435/12, ECLI:EU:C:2014:254, 10. aprill 2014)* küsis Madalmaade kõrgeim kohus ELTK-lt, kas infoühiskonna direktiivi artikli 5 lõike 2 punkt b, mis reguleerib kodu kopeerimise erandit, hõlmab ka ebaseaduslikest allikatest tehtud koopiaid.¹⁹ ELTK vastas ühemõtteliselt: "infoühiskonna direktiivi artikli 5 lõike 2 punkti b tuleb tõlgendada nii, et sellega ei ole hõlmatud juhtum, kus isiklikuks tarbeks tehtud koopia aluseks on ebaseaduslik allikas."²⁰

¹⁴ DSM-direktiiv, art 4 lg 1. AutÕS § 19² lg 1 kasutab sõnastust "õiguspäraselt juurdepääsetav teos", mis erineb terminoloogiliselt direktiivi sõnastusest. Vt Synodinou, T.-E.; Vrakas, G. Lawful Access as a Gatekeeper for TDM in the EU. - Verfassungsblog, 17. november 2025. DOI: 10.17176/20251117-142050-0.

¹⁵ DSM-direktiiv, art 4 lg 3; AutÕS § 19² lg 2. - RT I 1992, 49, 615; RT I, 12.07.2025, 9 (jõust. 01.10.2025).

¹⁶ DSM-direktiiv, art 7 lg 1; AutÕS § 19¹ lg 4. Rosati, E. Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790. Oxford: Oxford University Press 2021, lk 89-91.

¹⁷ DSM-direktiiv, põhjenduspunkt 14. Leistner, M.; Antoine, L. TDM and AI Training in the European Union - From 'LAION' to Possible Ways Ahead? – GRUR International, kd 74, nr 11, 2025, lk 1027-1044 (SSRN preprint: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5178237, 20.03.2026).

¹⁸ Käesolev töö on kirjutatud 2026. aasta kevadel. Kõik EL-i poolsed järeldused on esitatud oodatava ELTK lahendi (C-250/25) valguses tingimuslikult. Kohtu-jurist Szpunari ettepanek on oodata 3. septembriks 2026 ning lõplik lahend 2027. aastaks.

¹⁹ ELKo 10.04.2014, C-435/12, *ACI Adam BV jt versus Stichting de ThuisKopie ja Stichting Onderhandeligen ThuisKopie vergoeding*, ECLI:EU:C:2014:254, punkt 41 ja resolutiivosa punkt 1.

²⁰ *Ibid.*, punkt 41.

Kohus jõudis selle järelduseni kolme argumentatsiooni taseme kaudu: esiteks, põhjenduspunkt 22 keelab võltsitud või piraatkoopiade ebaseadusliku levitamise toetamise; teiseks, autoriõiguse kolmeastmeline test välistab sellise tõlgenduse, mis kahjustab õiguste omajate hüvesid; kolmandaks, hüvitiste arvutamise süsteem, mis ei erista seaduslikku ja ebaseaduslikku allikat, ei taga "õiglast tasakaalu" õiguste omajate ja kasutajate vahel.²¹ Kohus kinnitas seda põhimõtet hiljem kohtuasjas *Stichting Brein versus Filmspeler* (C-527/15), laiendades ebaseadusliku allika doktriini ajutiste reproduktsioonide erandi (InfoSoc art 5 lg 1) lubatavuse küsimusele. Kohtuasi *Stichting Brein vs Ziggo* (C-610/15) käsitleb eraldiseisvat küsimust - torrentiplatvormi pidaja otsest vastutust üldsusele edastamise eest (InfoSoc art 3 lg 1) - ega ole käesolevas töös käsitletava ebaseadusliku allika doktriini seisukohast otseselt asjakohane.²²

ACI Adam doktriini (ebaseadusliku allika doktriin) kohaldamine TDM-kontekstis ei tulene otseselt kohtupraktikast - ELTK ei ole TDM-erandite küsimuses seni sõna võtnud. Tegemist on doktrinaalse analoogiaga, mille poolt on argumenteerinud osa õiguskirjanduses, eelkõige Synodinou ja Vrakas (2025), kes järeldavad, et DSM-direktiivi artiklite 3 ja 4 seadusliku juurdepääsu nõue toimib samal põhimõttel nagu ACI Adam-i ebaseadusliku allika doktriin (piraatileheküljelt vabalt kättesaadav teos ei ole seaduslikult ligipääsetav direktiivi mõttes, sõltumata faktilisest kättesaadavusest). Selle analoogia autoriteetse kohtupraktilise kinnituse puudumine on oluline piirang. EL-i poolsed järeldused on selles küsimuses tingimusliku iseloomuga seni, kuni ELTK lahend asjas C-250/25 pole saabunud.²³ ACI Adam analoogia õigusteoreetilised piirid väärivad tähelepanu. Vastuargumendina on märgitud, et DSM-direktiiv (2019) võeti vastu viis aastat pärast ACI Adam'i (2014) ning põhjenduspunkt 14, mis loetleb seadusliku juurdepääsu näiteid, ei nimeta piraatlust *expressis verbis* – millest võiks järeldada seadusandja teadlikku valikut jätta küsimus lahtiseks. Selle argumendiga ei saa siiski nõustuda kahel põhjusel. Esiteks on TDM-erandite teleoloogiline tõlgendus suunatud seaduslike kasutuste hõlbustamisele, mitte ebaseadusliku algkäitumise seadustamisele; teiseks välistab kolmeastmeline test (Berni konventsiooni art 9(2); InfoSoc-direktiiv art 5(5)) tõlgenduse, mis kahjustaks õiguste omajate põhjendamatult. Seda

²¹ *Ibid.*, punktid 36-37. Vt ka punktid 39-40 (kolmeastmeline test); punktid 54-57 (õiglane tasakaal). Vt ka Euroopa Parlamendi ja nõukogu direktiiv 2001/29/EÜ, 22. mai 2001, infoühiskonna autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ütlustamise kohta. - EÜT L 167, 22.06.2001, lk 10-19, põhjenduspunkt 22.

²² ELKo 26.04.2017, C-527/15, *Stichting Brein versus Jack Frederik Willems*, ECLI:EU:C:2017:300, punkt 71; ELKo 14.06.2017, C-610/15, *Stichting Brein versus Ziggo BV ja XS4ALL Internet BV*, ECLI:EU:C:2017:456, punkt 47.

²³ Synodinou; Vrakas (viide 9). Vt ka Rosati, E. Is Text and Data Mining Synonymous with AI Training? - *Journal of Intellectual Property Law & Practice*, kd 19, nr 12, 2024, lk 851, 857.

käsitlust kinnitavad 2025. aasta saksa kohtulahendid. LG München I leidis kohtuasjas *GEMA v. OpenAI* (Az. 42 O 14139/24, 11.11.2025; esimese astme otsus, apellatsioon menetluses OLG Münchenis), et autoriõigusega kaitstud sisu memoriseerimine AI mudeli parameetrites kujutab endast iseseisvat reprodutseerimist, mida TDM-erand ei kata, kuna teosed säilivad mudelis reprodutseeritaval kujul.²⁴ OLG Hamburg kinnitas kohtuasjas *Kneschke v. LAION (5 U 104/24, 10.12.2025)*, et DSM-direktiivi artikli 4 üldine TDM-erand kohaldub LAION-i CLIP-mudeli poolt teostatavale pildi-tekstikirjelduse vastavuse kontrollile – andmestiku kvaliteedikontrolli tegevusele. Kohus jättis sõnaselgelt lahtiseks küsimuse, kas TDM-erand kohaldub generatiivse AI mudeli treenimisele kui sellisele. ACI Adam'i põhimõttest ja Kneschke lahendist koosmõjus võib järeldada, et *opt-out*-mehhanismil on praktiline tähendus eeskätt siis, kui andmete algne omandamine toimus seaduslikust allikast. Kui andmekogumine toimus juba algselt ebaseaduslikust allikast, nihkub vaidluse kese *opt-out* küsimuselt seadusliku juurdepääsu puudumisele. Seda järeldust ei saa siiski omistada otseselt Kneschke lahendi resolutsioonile, sest nimetatud asjas käsitleti seaduslikult avaldatud pilte, mitte piraatallikate kasutamist generatiivse AI treenimisel.²⁵

Käesolevaks hetkeks on ELTK-le esitatud esimene otseselt generatiivse AI treenimist ja autoriõigust käsitlev eelotsusetaotlus kohtuasjas *Like Company Kft versus Google Ireland Ltd (C-250/25)*. Ungari meediaettevõtte väidab, et Google'i tehisintellekti assistent Gemini reprodutseerib ja kokkuvõtab tema autoriõigusega kaitstud ajakirjandustekste loata. Asi on suunatud ELTK suurkojale (15 kohtunikku), suuline menetlus toimus 10. märtsil 2026, ning kohtujürist Szpunari ettepanek on oodata 3. septembriks 2026, lõplik lahend on oodatav 2027. aastal. Kuni ELTK lahendini puudub autoriteetne EL-i kohtupraktika DSM-direktiivi TDM-erandite kohaldamise kohta AI treenimisele.²⁶

ELTK lahend *C-250/25* asjas võib võtta kolme suunda, millest igaüks mõjutab käesoleva töö argumenti erinevalt. Esimese stsenaariumi kohaselt – mis oleks käesoleva töö hüpoteesi jaoks kinnitav – leiab ELTK, et DSM-direktiivi art 4 TDM-erand ei kata piraatlikest allikatest pärit treeningandmeid, tugevdades seega seadusliku juurdepääsu nõude

²⁴ LG München I, 11.11.2025, 42 O 14139/24 (*GEMA v. OpenAI*, (LG München I, esimese astme otsus; apellatsioon menetluses)). Vt ka Leistner; Antoine (viide 12), lk 1035-1039.

²⁵ OLG Hamburg, 10.12.2025, 5 U 104/24 (*Kneschke v. LAION e.V.*). Vt ka Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024, tehisintellekti käsitlevate üldist raamistikku kehtestavate eeskirjade kohta (tehisintellekti käsitlev määrus). - ELT L, 12.07.2024, art 53 lg 1 p c.

²⁶ ELTK eelotsusetaotlus, *Like Company Kft versus Google Ireland Ltd, C-250/25*, esitatud 3. aprill 2025. - ELT C, 2025 (avaldatav). Suuline menetlus 10. märtsil 2026. Euroopa Autoriõiguse Ühing. Comment of the European Copyright Society on the Request for Preliminary Ruling in Case C-250/25 (Like Company). - JIPITEC, kd 16, 2025, lk 121 jj (SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5166831 , 20.03.2026).

kategoorilist iseloomu. Teise stsenaariumi kohaselt leiab kohus, et TDM-erand kohaldub ka ebaseaduslikest allikatest pärit andmete kasutamisele, mis nõuaks töö EL-i analüüsi põhjalikku ümberhindamist. Kolmanda stsenaariumi kohaselt tunnistab ELTK küsimused vastuvõetamatuks või jätab neile vastamata, millega *status quo* jääb – autoriteetne ELTK lahend puudub ning magistritöö põhineb doktrinaalsel analüüsil ilma kohtu kinnitusest. Kuni ELTK lahendini tugineb käesoleva töö EL-i argument ELTK varasemale kohtupraktikale (ACI Adam) ning 2025. aasta Saksa kohtulahendite analoogilisele tõlgendamisele. Töö EL-i poolsed põhijäreldused on tingimusliku iseloomuga, kuna kehtivat õigust ei esitata siin lõplikult väljakujunenud kujul. Töö kaitseb seisukohta, et olemasoleva normiteksti, ELTK varasema praktika ja 2025. aasta Saksa kohtupraktika kõige veenvama koosmõjulise tõlgenduse kohaselt ei laiene TDM-erandid piraatallikatelt pärinevatele treeningandmetele. See seisukoht ootab lõplikku autoriteetset kinnitust või korrektuuri ELTK tulevase lahendi näol asjas C-250/25.

1.1.3 Tehisintellekti käsitleva määruse artikkel 53: ÜTEM-kohustused ja autoriõiguse vastavus

Euroopa Parlamendi ja nõukogu 13. juuni 2024. aasta määrus (EL) 2024/1689, milles käsitletakse tehisintellektiga seotud ühtlustatud eeskirju (*edaspidi*: tehisintellekti käsitlev määrus), jõustus 1. augustil 2024 ning selle V peatükk, mis käsitleb ÜTEM-i, hakkas kohalduma 2. augustist 2025.²⁷ Tehisintellekti käsitlev määrus on esimene regulatiivne raamistik maailmas, mis muudab autoriõiguse järgimise ÜTEM-ide pakkujatele avalik-õiguslikuks kohustuseks. ÜTEM-i mõiste alla kuuluvad suurel hulgal andmeid läbi töötanud, märkimisväärset üldistusvõimet näitavad mudelid nagu OpenAI GPT-seeria, Anthropic’u Claude ja Meta LLaMA; süsteemse riskiga ÜTEM-ideks klassifitseeritakse mudelid, mille kumulatiivne treeningu mahu mõõtühik ületab 10²⁵ FLOPs-i künnise.²⁸

Artikkel 53 lõige 1 sätestab kõigile ÜTEM-i pakkujatele neli põhikohustust. Käesoleva töö seisukohalt on suurima tähtsusega kolmas kohustus (punkt c): pakkuja peab "kehtestama poliitika liidu autoriõigust ja autoriõigusega kaasnevaid õigusi käsitlevate õigusaktide järgimiseks, eelkõige tuvastama ning järgima, sealhulgas tiptasemel tehnoloogiate abil, direktiivi (EL) 2019/790 artikli 4 lõike 3 kohaselt väljendatud õiguste reservatsioone.“ Tehisintellekti käsitleva määruse põhjenduspunkt 105 kinnitab *expressis verbis*, et ÜTEM-ide

²⁷ Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024, milles käsitletakse tehisintellektiga seotud ühtlustatud eeskirju (tehisintellekti käsitlev määrus). - ELT L, 12.07.2024. Jõustus 1. augustil 2024, V peatükk kehtib alates 2. augustist 2025 vastavalt art 113 lg 1 p b-le.

²⁸ Tehisintellekti käsitlev määrus, art 3 lg 63; art 51 lg 1-2.

treenimiseks kasutatav TDM mahub DSM-direktiivi kohaldamisalasse. *Opt-out*-i eiramine on nüüd üheaegselt nii autoriõiguse rikkumine kui ka tehisintellekti käsitleva määruse rikkumine, mis toob kaasa trahvid kuni 15 miljonit eurot või 3% ülemaailmsest aastakäibest.²⁹ Neljas kohustus (punkt d) nõuab koostamist ja avalikult kättesaadavaks teha piisavalt üksikasjalik kokkuvõte ÜTEM-i treenimiseks kasutatud sisu kohta vastavalt AI büroo kehtestatud mallile; treeningandmete kokkuvõtte mall avaldati 24. juulil 2025.³⁰

Siinkohal on oluline art 53 lõike 2 erisus: avatud lähtekoodiga mudelite pakkujatele ei kohaldu punktides a ja b sätestatud kohustused, kuid autoriõiguse vastavuse poliitika (punkt c) ja treeningandmete kokkuvõtte (punkt d) kohustused kehtivad kõigile, sh avatud mudelite pakkujatele.³¹ ÜTEM-ide heade tavade koodeksi (lõplik versioon 10. juulil 2025) autoriõiguse peatükk täpsustab, et veebikraapimise käigus on kohustuslik piraatveebilehekülgede välistamine ning robots.txt-protokollide järgimine.³² Tehisintellekti käsitleva määruse põhjenduspunkt 106 kehtestab eksterritoriaalsuse põhimõtte: kohustused kehtivad kõigile ÜTEM-i pakkujatele, kes viivad mudeli EL-i turule, olenemata jurisdiktsioonist, kus treenimise aluseks olevad autoriõigust puudutavad toimingud toimuvad.³³ Eesti tasandil on Tarbijakaitse ja Tehnilise Järelevalve Amet (TTJA) nimetatud üheks pädevaks asutuseks, kuid V peatüki otsene järelevalvepädevus kuulub ainuisikuliselt AI büroole Euroopa Komisjoni struktuurüksusena.³⁴ Tehisintellekti käsitleva määruse artikkel 111 lõige 3 sätestab üleminekusätte juba turul olevate ÜTEM-i suhtes: nende pakkujad, kelle mudelid olid turule viidud enne V peatüki kohaldumiskuupäeva (2. augustit 2025), peavad võtma vajalikud meetmed artiklites 53-55 sätestatud kohustuste täitmiseks hiljemalt 2. augustiks 2026.³⁵ Käesoleva töö seisukohalt on sellest normist tulenev järeldus oluline – enne

²⁹ *Ibid.*, art 53 lg 1 p a-d; art 55 lg 1 (süsteemse riskiga ÜTEM-ide lisakõhustused); art 53 lg 1 p c; põhjenduspunkt 105. Quintais, J. P. Generative AI, Copyright and the AI Act. - Computer Law & Security Review, kd 56, 2025, artikkel 106107, lk 1-2 (SSRN: 4912701).

³⁰ *Ibid.*, art 53 lg 1 p d; põhjenduspunkt 107. Euroopa Komisjon. Template for General-Purpose AI Model Providers to Summarise Their Training Content. Brüssel: Euroopa Komisjon, 24. juuli 2025. – <https://digital-strategy.ec.europa.eu/en/faqs/template-general-purpose-ai-model-providers-summarise-their-training-content> (20.03.2026).

³¹ *Ibid.*, art 53 lg 2. Vt Peukert, A. Copyright in the Artificial Intelligence Act - A Primer. - GRUR International, kd 73, nr 6, 2024, lk 497, 503 (SSRN: 4771976).

³² Euroopa Komisjon. Üldotstarbeliste tehisintellekti mudelite heade tavade koodeksi lõplik versioon, 10. juuli 2025; Komisjon tunnistas koodeksi "piisavaks" 1. augustil 2025. – <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai> (20.03.2026), meede 1.2.

³³ Tehisintellekti käsitlev määrus, põhjenduspunkt 106. Quintais (viide 25), lk 7-9.

³⁴ *Ibid.*, art 77; art 88 lg 1-2. Tarbijakaitse ja Tehnilise Järelevalve Amet (TTJA) on nimetatud üheks pädevaks asutuseks tehisintellekti käsitleva määruse art 77 tähenduses, kuid tehisintellekti käsitleva määruse artikkel 88 lõike 1 kohaselt kuulub üldotstarbeliste tehisintellektimudelite pakkujate üle järelevalve ainuisikuliselt AI büroole Euroopa Komisjoni struktuurüksusena; TTJA pädevus piirdub liikmesriigi tasandi kohustuste täitmise järelevalvega (artikkel 77).

³⁵ Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024, tehisintellekti käsitlevate ühtlustatud eeskirjade kohta (tehisintellekti käsitlev määrus). - ELT L, 12.07.2024, artikkel 111 lõige 3: "Üldotstarbeliste

2025. aasta augustit treenitud mudelid ei ole tehisintellekti käsitleva määruse V peatüki kohaldamisalast automaatselt välja arvatud. Kui sellised mudelid on 2. augustiks 2026 endiselt EL-i turul ega vasta artikli 53 lõike 1 punktidest c ja d tulenevatele kohustustele, st pakkujal puudub autoriõiguse vastavuse poliitika või treeningandmete kokkuvõte, kujutab see endast tehisintellekti käsitleva määruse rikkumist sõltumata asjaolust, et mudeli treenimine toimus ajaliselt enne V peatüki jõustumist. Autoriõiguslik vastutus on seejuures eraldiseisev küsimus. Tehisintellekti käsitleva määruse art 111 lg 3 üleminekusäte reguleerib üksnes tehisintellekti käsitleva määruse kohustuste täitmise tähtaega, mitte piraatlikult allikalt treenimisest tuleneva tsiviilõigusliku kahju hüvitamise nõude aegumist. Viimase suhtes kohalduvad AutÕS §-dest 19¹ ja 19² ning InfoSoc-direktiivi artiklist 2 tulenevad sätted, mille kohaldamine ei ole seotud tehisintellekti käsitleva määruse V peatüki ajalise kehtivusega.

1.1.4 Eesti transponeerimise eripära: varasem TDM-erand vs DSM

Eestil oli ajavahemikul 2017-2022 laiem TDM-erand (AutÕS § 19 lõike 1 punkt 3¹), mis ei sisaldanud seadusliku juurdepääsu nõuet: kõigil isikutel oli lubatud teoste TDM-töötlemine mittekaubandusliku eesmärgiga, ilma et erand oleks eeldanud seaduslikku juurdepääsu lähteteosele.³⁶

DSM-direktiivi transponeerimisel Riigikogus 2021. aasta detsembris (RT I, 28.12.2021, 1, jõust. 07.01.2022) lisandus seadusliku juurdepääsu nõue mõlemasse uude paragrahvi: AutÕS § 19¹ lõige 1 nõuab seaduslikku juurdepääsu teadusuuringute TDM-erandi kasutamiseks, AutÕS § 19² lõige 1 nõuab, et teos oleks "õiguspäraselt juurdepääsetav" üldise TDM-erandi kohaldamiseks.³⁷

See ajalooline muutus on käesoleva töö jaoks normatiivselt oluline. Seadusandja teadlik valik piiravama sõnastuse kasuks kinnitab, et seadusliku juurdepääsu nõue ei tulene pelgalt direktiivi automaatsest ülevõtmisest, vaid esindab aktiivset poliitilist otsust ning peegeldab ELTK kohtuasjas *ACI Adam (C-435/12)* kujunenud üldpõhimõtet, mille järgi ükski autoriõiguse erand ei tohi toimida ebaseadusliku algkäitumise seadustajana.

tehisintellektimudelite pakkujad, kes on viinud oma mudelid turule enne käesoleva määruse artiklis 53 sätestatud kohustuste kohaldumiskuupäeva, tagavad vastavuse nimetatud kohustuste ja muude V peatüki sätetega hiljemalt 2. augustiks 2026."

³⁶ AutÕS § 19 lõike 1 punkt 3¹. - RT I, 07.07.2016, 1 (kehtetu alates 07.01.2022). Erand kehtis kõigile isikutele mittekaubandusliku TDM-i eesmärkidel ning selles puudus "seadusliku juurdepääsu" nõue. Vt Triaille, J.-P. jt. A European Solution for Text and Data Mining in the Research Sector. - Stockholm Intellectual Property Law Review, kd 4, nr 2, 2021, lk 9, 18.

³⁷ AutÕS § 19¹ lg 1 ja lg 4; § 19² lg 1 ja lg 2. - RT I 1992, 49, 615; RT I, 28.12.2021, 1 (jõust. 07.01.2022); RT I, 12.07.2025, 9 (jõust. 01.10.2025). DSM-direktiivi transponeerimistähtaeg oli 7. juuni 2021; Eesti hilines ligikaudu seitse kuud (eelno 368 SE, vastuvõetud 08.12.2021). Vt Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790, art 3 lg 1 ja art 4 lg 1.

Eesti-spetsiifiline avatud õiguslik küsimus puudutab mudeleid, mis treeniti ajavahemikul 2017-2022 siis kehtinud laiemal TDM-erandi alusel, kuid mis on jätkuvalt turul ja reprodutseerivad memoriseeritud kaitstud sisu. Küsimusele on võimalik anda kaks vastandlikku normatiivset vastust. Esimese positsiooni järgi – *tempus regit actum* põhimõttest lähtuvalt – kohaldub treenimistoimingule see õigus, mis kehtis toimingu tegemise ajal, mistõttu 2017-2022 perioodil treenitud mudelite suhtes ei saa seadusliku juurdepääsu nõuet tagasiulatuvalt rakendada. Teise positsiooni järgi kujutab piraatlikult treenitud mudeli jätkuv turul olemine – eelkõige iga memoriseeritud sisu reprodutseerimine kasutajapäringutele vastamisel – iseseisvat jätkuvat rikkumistoimingut (ingl k *continuing infringement*) InfoSoc-direktiivi artikli 2 tähenduses, mille suhtes kehtib praegune õigus ning mille aegumistähtaeg ei hakka kulgema seni, kuni rikkumine kestab. Mõlema positsiooni normatiivne alus on kaitstav, kuid küsimus on Eesti kohtupraktikas praegu lahendamata. See Eesti-spetsiifiline erijuht ei sea kahtluse alla töö põhihüpoteesi Euroopa Liidu õiguse kohta. See illustreerib pigem üleminekuolukordades tekkivaid probleeme, kus siseriiklik regulatsioon on ajutiselt olnud direktiiviga ühtlustatud raamistikust laiem. EL-i tasandil jääb töö põhiseisukoht muutumatuks: kehtiva õiguse kõige põhjendatuma tõlgenduse kohaselt ei ole piraatallikatelt pärinevate andmete kasutamine TDM-eranditega kaetud. Eesti eriküsimus puudutab üksnes varasemate mudelite suhtes kohaldatavat õigust ja jätkuva rikkumise võimalikkust.

Praktiline järeldus mõlemale lahendusele on ühine: Eesti autoriõiguse omajal puuduvad tõhusad täitemehhanismid piraatandmetel treenitud AI-mudeli vastu nõude esitamiseks, kuna ükski suurematest AI-ettevõtetest ei ole asutatud Eestis ning rahvusvaheline kohtualluvus tekitab olulisi praktilisi takistusi. Lisaks ei ole Eesti autoriõiguse seaduse §-d 19¹ ja 19² sõnaselgelt reguleerinud olukorda, kus TDM-i teostaja kasutab saastunud andmestikku, mis pärineb kolmandalt osapoolelt. Need lüngad toetavad töö *de lege ferenda* ettepanekuid nii EL-i tasandi direktiivimuudatuste kui Eesti AutÕS täpsustuse osas (ptk 4.3).³⁸

³⁸ Jätkuva rikkumise (*continuing infringement*) ja *tempus regit actum* põhimõtte kohta autoriõiguse kontekstis vt Tsiviilseadustiku üldosa seadus (TsÜS) §-d 146-150 (RT I 2002, 35, 216; RT I, 22.03.2021, 8); Leistner, M. EU Copyright Law. - IIC, kd 48, 2017, lk 381-403. *De lege ferenda* viide töö ptk 4.3 ettepanekutele.

1.2 USA autoriõiguse raamistik: *fair use*'i neljaetapiline test

1.2.1 17 U.S.C. § 107 ja neljaetapiline test AI treenimise kontekstis

Ameerika Ühendriikide autoriõiguse süsteem erineb Euroopa Liidu omast fundamentaalselt nii struktuurilt kui ka filosoofialt. Erinevalt DSM-direktiivi artiklites 3 ja 4 sätestatud kategooriaalsetest TDM-eranditest ei tunne USA autoriõiguse seadus (*Copyright Act of 1976*) ühtegi TDM-erandit ega generatiivset AI treenimist hõlmavat erisätet. Kõik AI treenimist puudutavad autoriõiguslikud vaidlused lahendatakse üldise *fair use*'i doktriini kaudu, mis kujutab endast positiivset kaitsealust (ingl k *affirmative defense*) – vastustajal (AI-ettevõttel) lasub kohustus tõendada, et tema tegevus vastab doktriini tingimustele.³⁹ *Fair use*'i doktriini aluseks on 17 U.S.C. § 107, mis sätestab neljaetapilise kaalumistesti: (1) kasutuse eesmärk ja iseloom, sh kas tegemist on kaubandusliku või mittetulundusliku kasutusega; (2) autoriõigusega kaitstud teose olemus; (3) kasutatud osa maht ja olulisus tervikteoses; ning (4) mõju autoriõigusega kaitstud teose potentsiaalsele turule või väärtusele.⁴⁰

Neljas faktor – turu mõju – on Ülemkohtu põhjal "vaieldamatult *fair use*'i kõige olulisem element" ning see faktor analüüsib kolme turu kahjuliiki: otsene turuasendus (väljundid asendavad originaalteoseid), turu lahjendamine (AI-genereeritud teosed küllastavad turgu) ning kaotatud litsentsimisvõimalused (kus toimivad AI treeningandmete litsentsimisturud).⁴¹ Kaks 2025. aasta California ringkonnakohtut jõudsid piraatlikelt allikatelt pärit andmete küsimuses täpselt vastupidistele järeldustele. Kohtuasjas *Bartz v. Anthropic* eristas kohtunik Alsup piraatluse treenimisest: 482 460 teose allalaadimine LibGenist ja PiLiMist moodustas iseseisva ja mitte-transformatiivse kasutuse, mida hilisem treenimine tagantjärele päästa ei saa. Kohtuasjas *Kadrey v. Meta* läks otsus aga Meta kasuks. Kohtunik Chhabria keeldus allalaadimist ja treenimist lahutamast ning hindas kogu protsessi ühe kõrgelt transformatiivse tervikuna. Meta sai *fair use* alusel täieliku võidu, hoolimata sellest, et LibGeni andmete kasutamine oli tuvastatud fakt. Hagejad kaotasid neljanda faktori osas, kuna toetusid substituutsioonil põhinevale turu kahjule, mida empiirilistelt tõendada ei suudetud - selle asemel, et kasutada kohtu poolt vihjeliselt avatud turu lahjendamise (ingl k

³⁹ 17 U.S.C. § 107.

⁴⁰ Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578-579 (1994) (Souter, J.); USA Autoriõiguse Amet (US Copyright Office). Copyright and Artificial Intelligence, Part III: Generative AI Training. Washington, D.C.: US Copyright Office, 9. mai 2025, lk 37-42 (edaspidi: USCO 2025). - <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> (20.03.2026).

⁴¹ USCO 2025 (viide 37), lk 54. 17 U.S.C. § 106; Harper & Row Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 566 (1985) (O'Connor, J.); USCO 2025 (viide 37), lk 61-73.

market dilution) argumenti. Chhabria lisas otsusele mittesiduva märkuse (*obiter dictum*), et turu lahjendamise teooria alusel oleks vastus enamikul juhtudel tõenäoliselt jah – kuid kuna hagejad seda teooriat ei kasutanud, ei moodusta see markus pretsedenti ega resolutsiooni.⁴²

1.2.2 *Google v. Oracle ja Warhol Foundation: transformatiivsuse pretsedendid*

Fair use'i esimese faktori tõlgendamine on viimase kolme aastakümne jooksul läbi teinud märkimisväärse arengu kahe fundamentaalselt erineva Ülemkohtu pretsedendi kaudu. *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021) (Breyer, J., 6:2) laiendas transformatiivset kasutusulatust – Google'i Java API deklaratiivse koodi (ligikaudu 11 500 rida) kopeerimine Androidi mobiilse platvormi tarbeks oli *fair use*, kuna see ühendas kopeeritud koodi "*Android Smart Phone Platformiks, mis täitis teistsuguses turusegmendis (nutitelefonid) põhimõtteliselt erinevat eesmärki*"⁴³. AI treenimise kontekstis on Oracle argumendiks, et teoste kopeerimine statistiliste mustrite ekstraheerimiseks on fundamentaalselt erineva eesmärgiga – analoogselt sellele, kuidas Java API kood kopeeriti mobiilse platvormi loomiseks.

Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508 (2023) (Sotomayor, J., 7:2) kitsendas oluliselt transformatiivsuse argumendi kasutusulatust.⁴⁴ Kohus leidis, et transformatiivsuse hindamisel tuleb vaadata konkreetset kasutust, mitte üldist kunstilist eesmärki: Goldsmithi foto ja Andy Warhol Foundationi "Orange Prince" litsentseerimispraktika jagas "sisuliselt sama eesmärki" – mõlemad teenisid kaubandusliku ajakirjandusliku illustratsioonina Printsi kuvandi loomiseks. Kohtu võtmesõnastus lk 533: "kui originaalteos ja teisene kasutus jagavad sama eesmärki ning teisene kasutus on kaubanduslik, kaalub esimene *fair use*'i faktor tõenäoliselt *fair use*'i vastu."⁴⁵ *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence, Inc.* (D. Del., nr 1:20-cv-00613-SB, 11. veebruar 2025) rakendas Warholi standardit otseselt AI treenimisele, leides, et Westlawi pealkirjaparagrahvide (ingl.k *headnotes*) kasutamine juriidilise AI otsinguteenuse treenimiseks ei olnud *fair use*. Pealkirjaparagrahvid on advokaatide koostatud originaalsed

⁴² *Bartz v. Anthropic PBC*, nr 3:24-cv-05417-WHA (N.D. Cal., 23. juuni 2025), kohtunik Alsup, lk 12-14; *Kadrey v. Meta Platforms, Inc.*, nr 3:23-cv-03417-VC (N.D. Cal., 25. juuni 2025), kohtunik Chhabria, lk 8-12; *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence, Inc.*, nr 1:20-cv-00613-SB (D. Del., 11. veebruar 2025), kohtunik Bibas, lk 18-22.

⁴³ *Google LLC v. Oracle America, Inc.*, 593 U.S. 1, 28-29 (2021) (Breyer, J., 6:2).

⁴⁴ *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528-533 (2023) (Sotomayor, J., 7:2).

⁴⁵ *Ibid.*, lk 533, 539-540. Kohtunik Kagan, eriarvamus, lk 17.

kokkuvõtted kohtulahendite õiguslikest printsiipidest – eraldiseisvad autoriõigusega kaitstud teosed, mis on eristuvad kohtuotsuse põhitekstist.⁴⁶

Google v. Oracle pretsedendi ülekandmisel generatiivse AI treenimisele loovate teoste puhul tuleb arvestada olulist eristust: Oracle käsitles API deklaratsioonide näol funktsionaalset koodi, mille puhul Ülemkohus rõhutas just funktsionaalset iseloomu *fair use* kaalumisel. Loovate teoste (ilukirjandus, muusika, kujutav kunst) puhul on *fair use*’i teine faktor – kaitstud teose olemus – oluliselt tugevam vastukaaluks, mistõttu Oracle’i pretsedent kohaldub generatiivse AI treenimisele kirjandusteostega üksnes piiratud analoogia kaudu, mitte otseselt.

1.2.3 USA Autoriõiguse Ameti 2025. aastaaruanne: ametlik seisukoht

USA Autoriõiguse Amet (*US Copyright Office*) avaldas 9. mail 2025 mahuka aruande *Copyright and Artificial Intelligence: Part III – Generative AI Training* (edaspidi: 2025. aastaaruanne), mis kujutab endast seni kõige põhjalikumat institutsionaalset analüüsi *fair use*’i ja AI treenimise suhete kohta.⁴⁷ Aruanne võtab spektripõhise seisukoha: *fair use*’i kohaldatavus AI treenimisele sõltub konkreetsetest asjaoludest ning ei võimalda üldreeglit ei ühes ega teises suunas. Spektri positiivses otsas on mitteväljenduslikud, uurimusliku või teadusliku eesmärgiga mudelid, mis ei võimalda teoste osade reprodutseerimist väljundites; spektri negatiivses otsas on ekspressiivsete teoste kopeerimine piraatlikelt allikatelt, et genereerida piiranguta sisu, mis konkureerib turul, kus litsentseerimiseks on mõistlikud võimalused.⁴⁸

Käesoleva magistritöö kontekstis on eriti oluline aruande seisukoht piraatlikelt allikatelt pärit andmete kohta. Lk 51: "Agentuuri hinnangul peaks piraatlikelt allikatelt või ebaseaduslikult ligipääsetud teostega koosnevate andmetike teadlik kasutamine kaaluma *fair use*’i vastu, olemata samas määrav."⁴⁹ Sõna "peaks" (*should*) – mitte "välistab" (*precludes*) – rõhutab, et tegemist on kaaluteguriga, mida saab kompenseerida teiste *fair use*’i kaalutlustega. See on magistritöö hüpoteesi kontrollimiseks kõige olulisem struktuurne erinevus EL-i ja USA raamistiku vahel – EL-i "seadusliku juurdepääsu" nõue toimib

⁴⁶ *Thomson Reuters* (viide 40), lk 18-22. Kolmanda ringkonnakohtu apellatsioonikohtu oodatav lahend (nr 25-2153) peaks andma esimese apellatsioonikohtulikku orientatsiooni, kuid ei ole oodatav enne 2026. aasta lõppu.

⁴⁷ USCO 2025 (viide 37), lk 1-3. Üle 10 000 kirjaliku kommentaari, avalikud kuulamised kirjanduse, visuaalkunstide, audiovisuaalse materjali ja muusika valdkonnas ning mitmeid sidusrühmade ümarlaudu.

⁴⁸ USCO 2025 (viide 37), lk 74 (spektripõhine üldjäreldus); lk 104-106 (seadusandlus soovitusel).

⁴⁹ USCO 2025 (viide 37), lk 51-52. Aruande sõnastus lk 51: "piraatlikest allikatest või ebaseaduslikult ligipääsetud teostega koosnevate andmetike teadlik kasutamine peaks kaaluma *fair use*’i vastu, olemata samas määrav."

lävenditingimusena, mille puudumisel TDM-erand ei kohaldu; USA õiguses on andmete ebaseaduslik päritolu aga üksnes kaaluv tegur. Aruannet ei käsitleta kohtutes siduva allikana, kuid selle erakordselt põhjalik analüüs annab talle arvestatava veenmisjõu Skidmore'i deferentsi tähenduses.⁵⁰

1.3 EL-i ja USA lähenemiste võrdlev kokkuvõte

Eelnevas kahes alapeatükis käsitletud põhjal on ilmne, et Euroopa Liidu ja Ameerika Ühendriikide autoriõiguse raamistikud erinevad generatiivse AI treenimise küsimuses nii struktuuriliselt, normatiivselt kui ka jõustamismehhanismi poolest. Kõige fundamentaalsem struktuuriline erinevus seisneb selles, et EL-i süsteem käsitleb TDM-i kui lubatud tegevust (ingl k *permitted act*) – seadusega ettenähtud erandit, mille tingimused on täidetavad *ex ante*, sõltumata kohtumenetlusest –, samas kui USA *fair use* toimib positiivse kaitsealusena (ingl k *affirmative defense*), mille puhul seaduslikkus selgub alles hagimenetluse käigus *ex post*.⁵¹

Seadusliku juurdepääsu nõue eristab kahe süsteemi lähenemist piraatlikelt allikatelt pärinevate andmete küsimuses kõige selgemini. EL-i raamistikus on "seaduslik juurdepääs" (DSM-direktiivi art 3 lg 1) ja "õiguspäraselt ligipääsetav" (art 4 lg 1) TDM-erandi eeltingimused: puudumise korral langeb kogu erandi kohaldatavus ära sõltumata kasutuse eesmärgist, mahust või turukahjust. ELTK kohtuasi *ACI Adam (C-435/12)* kinnitas, et see põhimõte kehtib EL-i autoriõiguse erandite süsteemis üldiselt: ebaseaduslikust allikast reprodutseerimine on kaetud ka Eesti õiguses AutÕS §§ 19¹ ja 19² kaudu. USA autoriõiguse seaduse § 107 ei sisalda ühtegi "seadusliku juurdepääsu" nõuet ega muid sõnaselgeid viiteid andmete päritolule; piraatlus on üksnes kaaluv tegur, mida saab kompenseerida teiste *fair use*'i kaalutlustega.

Teine oluline struktuuriline erinevus puudutab õiguste omajate ennetavat kaitset. DSM-direktiivi artikkel 4 lõige 3 näeb ette sõnaselge ja tehnoloogiliselt vahendatava *opt-out*-õiguse; tehisintellekti käsitleva määruse artikkel 53 lõike 1 punkt c muudab selle *opt-out*-i järgimise ÜTEM-i pakkujatele avalik-õiguslikuks kohustuseks sanktsioonimäära kuni 15 miljonit eurot. USA autoriõiguse seaduses puudub *opt-out*'ile vastav mehhanism, st õiguste omajad ei saa proaktiivselt takistada AI treenimist, vaid peavad pärast toimunud rikkumist alustama kohtumenetlust. EL-i regulatsioon omab lisaks eksterritoriaalset ulatust –

⁵⁰ Skidmore v. Swift & Co., 323 U.S. 134, 140 (1944). Kohtud rakendavad Autoriõiguse Ameti seisukohtadele Skidmore'i, mitte Chevron'i deferentsi. Perlmutter v. Blanche et al., D.D.C., nr 1:25-cv-01659, DC ringkonnakohus otsustas 10. septembril 2025 Perlmutteri kasuks.

⁵¹ DSM-direktiiv, art 3 lg 1; art 4 lg 1; ELKo C-435/12 (viide 13), punkt 41. Võrdlevalt: 17 U.S.C. § 107 (kaaluv asjaolu, mitte eeltingimus).

tehisintellekti käsitleva määruse põhjenduspunkt 106 kohaldab kohustusi kõigile EL-i turule sisenevate mudelite pakkujatele, olenemata treenimise toimumiskohast.⁵²

Eeltoodust lähtudes loob EL-i süsteem piraatallikatelt pärinevatel andmetel põhinevale AI treenimisele vähemalt doktrinaalsel tasandil tugevama ja selgema õigusliku tõkke kui USA *fair use*'i doktriin, mille kohaldamine sõltub kohtuasja konkreetsetest asjaoludest ja hilisemast kohtulikust kaalumisest. Samal ajal tuleb arvestada, et Euroopa Liidu tasandil puudub selles küsimuses veel otsene ELTK lahend. See struktuurne erinevus on käesoleva magistritöö keskseks võrdleva analüüsi lähtekohaks.⁵³ Samaaegselt tuleb arvestada, et EL-i süsteemis on autoriteetne kohtupraktika DSM-direktiivi TDM-erandite kohaldamise kohta AI treenimisele seni puudu – oodatav ELTK lahend kohtuasjas C-250/25 (*Like Company v. Google*) peaks andma määrava selguse küsimusele, kas seadusliku juurdepääsu nõue kujutab endast tõepoolest kategoorilist välistust piraatlikelt allikatelt pärit andmetele tugineva AI treenimise suhtes.⁵⁴

Kokkuvõtvalt ilmneb, et Euroopa Liidu ja Ameerika Ühendriikide autoriõiguse süsteemid lähenevad tehisintellekti treenimise küsimusele erinevatest normatiivsetest eeldustest. Kui EL-i süsteem tugineb eelnevalt määratletud eranditele ja selgetele eeltingimustele (nt seadusliku juurdepääsu nõue), siis USA süsteem võimaldab paindlikku juhtumipõhist hindamist *fair use*'i doktriini kaudu. See erinevus kujundab kogu edasise analüüsi telje.

⁵² Tehisintellekti käsitlev määrus, põhjenduspunkt 106 (eksterritoriaalsus põhimõte). Quintais, J. P. Copyright, the AI Act and Extraterritoriality. The Lisbon Council 2025 (SSRN: 5316132).

⁵³ Senftleben, M. Generative AI and Author Remuneration. – IIC – International Review of Intellectual Property and Competition Law, kd 54, 2023, lk 1535, 1541–1544.

⁵⁴ ELTK eelotsusetaotlus C-250/25 (viide 21). Kohtujürist Szpunari ettepanek oodatav 3. septembriks 2026; lõplahend tõenäoliselt 2026. aasta lõpus või 2027. aastal.

2. PIRAATALLIKATE KÜSIMUS EL-I AUTORIÕIGUSES

Käesolev peatükk uurib, kas ja millistel tingimustel kujutab generatiivse tehisintellekti mudeli treenimine piraatlikest allikatest – nagu Library Genesis (LibGen), Books3 ja Pirate Library Mirror (PiLiMi) – kogutud andmete abil endast autoriõiguse rikkumist Euroopa Liidu õiguse kohaselt. Analüüsi lähtekohaks on DSM-direktiivi⁵⁵ TDM-erandite struktuur ja nende piirangud, mida täiendab 2025. aasta Saksamaa kohtupraktika.

2.1. Seadusliku juurdepääsu nõue piraatallikate kontekstis

DSM-direktiivi TDM-erandite kõige olulisemaks piiravaks tingimuseks on nõue, et TDM-toiminguid tohib teostada üksnes teostele, millele on seaduslik juurdepääs. Artikkel 3 lõige 1 näeb ette, et uurimisasutuste ja kultuuripärandi asutuste TDM-erand kohaldub "teostele ja muudele kaitstud objektidele, millele neil on seaduslik juurdepääs, teadustegevuse eesmärgil".⁵⁶ Artikkel 4 lõige 1 laiendab TDM-erandi kõigile isikutele mis tahes eesmärgil, kuid sätestab samuti nõude, et kasutatavad teosed peavad olema "seaduslikult kättesaadavad".⁵⁷ Mõlema sätte kontekstis on fundamentaalne küsimus, mida tähendab "seaduslik juurdepääs" või "seaduslikult kättesaadav" piraatlikest allikatest pärit andmete puhul.

Direktiivi põhjendus 14 selgitab "seadusliku juurdepääsu" mõistet, osutades, et see hõlmab juurdepääsu avatud juurdepääsu alusel, lepingulist juurdepääsu (nt tellimuse alusel) ning muud seaduslikku alust.⁵⁸ Seega eeldab "seaduslik juurdepääs" kahte elementi: esiteks, et juurdepääs teosele on faktiliselt võimalik; teiseks, et see juurdepääs toimub seaduslikul alusel – kas õiguste omaja loal, vastava erandi alusel või muus seaduses ettenähtud korras. Piraatlikud andmekogumid, mis sisaldavad autoriõigusega kaitstud teoseid, mille üleslaadimine on toimunud ilma õiguste omajate loata, ei vasta kummalegi nõudele. Kuigi need on faktiliselt kättesaadavad, puudub nende kasutamiseks igasugune seaduslik alus.

Euroopa Kohus on analoogset olukorda analüüsinud asjas *ACI Adam BV jt vs Stichting de ThuisKopie* (C-435/12, edaspidi: *ACI Adam*).⁵⁹ Ehkki nimetatud asi käsitles

⁵⁵ DSM-direktiiv: Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790, 17. aprill 2019, autoriõiguse ja sellega kaasnevate õiguste kohta digitaalsel ühtsel turul ning millega muudetakse direktiive 96/9/EÜ ja 2001/29/EÜ - ELT L 130, 17.05.2019, lk 92-119.

⁵⁶ DSM-direktiivi artikkel 3 lõige 1.

⁵⁷ DSM-direktiivi artikkel 4 lõige 1.

⁵⁸ DSM-direktiivi põhjendus 14: "Seaduslik juurdepääs tähendab juurdepääsu avatud juurdepääsupõhimõtte alusel, lepingulist juurdepääsu, näiteks tellimuslepingu alusel, või muud seaduslikku juurdepääsu, nagu juurdepääs piiratud valikuga avalikele andmebaasidele."

⁵⁹ Euroopa Kohtu otsus 10.04.2014, C-435/12, *ACI Adam BV jt vs Stichting de ThuisKopie ja Stichting Onderhandeligen ThuisKopie vergoeding*. – ECLI:EU:C:2014:254. (ACI Adam)

InfoSoc-direktiivi artikli 5 lõike 2 punktis b sätestatud isikliku koopia erandit, on kohtu arengusuunad analoogiliselt kohaldatavad ka TDM-erandite kontekstis. Kohus leidis, et isikliku koopia erand ei laiene koopiatele, mis on tehtud ebaseaduslikest allikatest.⁶⁰ Kohtu põhjendus tugines direktiivi eesmärgile ja süstemaatilisele tõlgendusele: erand on ette nähtud olukordadeks, kus isik on saanud seaduslikult juurdepääsu teosele ning soovib teha sellest isiklikuks otstarbeks koopia; erand ei ole mõeldud selleks, et seadustada tagantjärele juurdepääsu ebaseaduslikult levitatud teostele.⁶¹ Analoogiliselt kehtib sama loogika TDM-erandite suhtes: nende eesmärk ei ole seadustada piraatlikest allikatest andmete kogumist, vaid hõlbustada seaduslikult kättesaadavate teoste analüüsi.

Euroopa Kohtu seisukohta *ACI Adam* asjas on õiguskirjanduses laiendatult tõlgendatud kui üldpõhimõtet, mille kohaselt ükski autoriõiguse erand ei tohi toimida ebaseadusliku algkäitumise seadustajana.⁶² See põhimõte kujutab endast õigussüsteemi loomulikku tagajärge: erandi eesmärk on lubada teatud kasutusi, mis muidu oleksid keelatud, kuid seda eeldusel, et algele kontaktile teosega on olemas seaduslik alus. Geigeri, Frosio ja Bulayenko on nimetatud põhimõtet TDM-direktiivi analüüsimisel rõhutanud, märkides, et "seadusliku kättesaadavuse" nõue kehtestab selge piiri legaalse ja illegaalse andmekasutuse vahele.⁶³

Piraatlikelt allikatelt kogutud andmete kasutamise kontekstis on olulised ka kolmeastmelise testi nõuded (Berni konventsiooni artikkel 9(2); InfoSoc-direktiivi artikkel 5(5)).⁶⁴ Testi kolmas samm nõuab, et erand ei "kahjustaks põhjendamatult õiguste omaja seaduslikke huve". Kui TDM-erandi alla haarata ka piraatlikest allikatest pärit andmed, tekiks otsene kõrvalmõju: AI-arendajad saaksid kasutada teoseid tasuta, loobudes turust, mille eest tuleks õiguspärase kasutuse korral tasuda, mis on tegelikult piraatlik koopia.⁶⁵ Selline tõlgendus kahjustaks autorite ja kirjastajate majanduslikke huve ning lõhestaks AI andmekogumite turu, kus õigustepärane ligipääs võib tähendada litsentsimaksu maksmist. Seega ei läbi piraatandmete kasutamine TDM-erandina ka kolmeastmelise testi kolmandat sammu.

Eelneva analüüsi põhjal on põhjendatud seisukoht, et Euroopa Liidu kehtiva õiguse kõige veenvama tõlgenduse kohaselt jäävad piraatallikatelt pärinevad treeningandmed

⁶⁰ ACI Adam, punkt 39.

⁶¹ *Ibid.*, punkt 41.

⁶² *Ibid.*, punkt 56.

⁶³ Geiger, C.; Frosio, G.; Bulayenko, O. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – The Good, the Bad and the Ugly. – IIC, 49, nr 7, 2018, lk 814-844, siin lk 822.

⁶⁴ Berni konventsioon kirjandus- ja kunstiteoste kaitseks (Berni konventsioon), Pariisi akt, 24.07.1971, artikkel 9(2); InfoSoc-direktiivi artikkel 5 lõige 5.

⁶⁵ Geiger; Frosio; Bulayenko, TDM Exception, lk 823.

DSM-direktiivi TDM-erandite kohaldamisalast välja. "Seadusliku juurdepääsu" nõue toimib siin lävenditingimusena, mille puudumisel ei ole erandile võimalik tugineda sõltumata kasutuse eesmärgist, mahust või väidetavast ühiskondlikust kasust. Grammatiline, teleoloogiline ja süstemaatiline tõlgendus, ACI Adam'i analoogia ning kolmeastmelise testi loogika jõuavad kõik samale järeldusele. Selle seisukoha autoriteetne kinnitamine Euroopa Liidu Kohtu poolt on siiski veel ootel.⁶⁶

ACI Adam analoogia piirid ja selle ületamine. Vastuargumendina on võimalik märkida, et DSM-direktiiv võeti vastu 2019. aastal – viis aastat pärast *ACI Adam*'i otsust (2014). Kui EL-i seadusandja oleks soovinud piraatlikke allikaid sõnaselgelt TDM-eranditest välistada, oleks tal olnud võimalus seda teha. Direktiivi põhjenduspunkt 14 loetleb "seadusliku juurdepääsu" näitelisi juhtumeid – avatud juurdepääsuga artiklid, andmebaasitellimused, isiklikud koopiad –, kuid ei nimeta piraatlust expressis verbis. Sellest võiks järeldada, et seadusandja teadliku valiku kaudu jättis piraatluse küsimuse kohtupraktikale, st ei soovinud seda normatiivselt lahendada.⁶⁷

See vastuargument ei ole siiski veenev kolmel põhjusel. Esiteks on DSM-direktiivi teleoloogiline tõlgendus selgesuunaline: direktiivi eesmärk on tasakaalustada TDM-kasutuste võimaldamist autoriõiguse kaitsega, mitte seadustada ebaseaduslikke algolukordi. Põhjenduspunkt 18 sätestab, et erand ei tohi "kahjustada õiguste omajate huve ega seada ohtu teoste tavapärase kasutamist" – tingimus, mida piraatlikelt allikatelt kogumine rikub olemuslikult.⁶⁸ Seega ei jätnud seadusandja küsimuse lahendamata, vaid määratles erandile need eesmärgipärased piirid, millest piraatlikelt allikatelt treenimist katv kasutus paratamatult välja jääb.

Teiseks kinnitab kolmeastmeline test (Berni konventsioon art 9(2); InfoSoc-direktiiv art 5(5)) sama järeldust süstemaatilise tõlgenduse kaudu: erand ei tohi "kahjustada põhjendamatult õiguste omajate seaduslikke huve". Piraatandmete lubamine TDM-erandi alla murendaks litsentsiturgu, kuna AI-ettevõtetal kaoks majanduslik stiimul hankida andmeid seaduslikult.

Kolmandaks toetavad 2025. aasta Saksa kohtulahendid seisukohta, et ACI Adam'i loogika ülekandmine TDM-konteksti on Euroopa kohtupraktikas vähemalt kujunev suund.

⁶⁶ Quintais, J.P. The New Copyright in the Digital Single Market Directive: A Critical Commentary. – European Intellectual Property Review, 42, nr 1, 2020, lk 28-41, siin lk 36.

⁶⁷ Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790 teksti- ja andmekaeve kohta (DSM-direktiiv), põhjenduspunkt 14. Võrdle: *ACI Adam BV jt vs Stichting de Thuis kopie*, C-435/12, ELTK 10.04.2014, punkt 41: erand on ette nähtud seadusliku allika kasutajale, mitte piraatluse seadustamiseks.

⁶⁸ DSM-direktiiv, põhjenduspunkt 18. Vt ka InfoSoc-direktiivi art 5 lg 5 (Berni kolmeastmeline test): erand ei tohi põhjendamatult kahjustada õiguste omajate seaduslikke huve.

GEMA v. OpenAI käsitleb memoriseerimist ja reprodutseerimist AI kontekstis, *Kneschke v. LAION* aga täpsustab TDM-erandi ja *opt-out* mehhanismi toimimist seaduslikult avaldatud sisu puhul.⁶⁹ Need lahendid ei anna veel ammendavat ega lõplikku vastust piraatallikatelt treenimise küsimusele, kuid toetavad seisukohta, et seadusliku juurdepääsu nõudel on TDM-erandite kohaldamisel keskne tähendus.

2.2. *GEMA v. OpenAI*: memoriseerimine reproduktsioonina

Müncheni I Ringkonnakohus (*Landgericht München I*) tegi 11. novembril 2025 esimese astme otsuse asjas *GEMA vs OpenAI Ireland Limited jt* (Az. 42 O 14139/24, edaspidi: *GEMA vs OpenAI*).⁷⁰ Tegemist on ühe esimese Euroopa kohtulahendiga, mis käsitleb generatiivse tehisintellekti mudeli treenimise autoriõiguslikku kvalifikatsiooni. Otsus ei ole jõustunud, OpenAI on esitanud apellatsioonkaebuse OLG Münchenile, apellatsioonimenetlus on pooleli. Käesolev töö kasutab GEMA otsust toetava, mitte kandva argumendina. Töö põhiseisukoht ei sõltu selle kohtuasja lõpptulemusest, vaid tugineb iseseisvalt DSM-direktiivi tekstile, ACI Adam'i doktriinile ja nende süstemaatilisele tõlgendusele. GEMA otsus on oluline eeskätt seetõttu, et see illustreerib, kuidas sama probleem on hakanud ilmnema ka AI-spetsiifilises kohtupraktikas.

GEMA – Saksamaa heliteoste autorite kollektiivse esindamise organisatsioon – esitas hagi OpenAI Ireland Limited vastu, väites, et “GPT-4 ja GPT-4o reprodutseerivad lihtsate kasutajapäringute kaudu üheksa tuntud Saksa laulu sõnu sõna-sõnalt, ilma asjaomaste autorite litsentsilepinguta”.⁷¹

Kohus pidi lahendama kolm põhiküsimust: (i) kas treenimise käigus toimub reprodutseerimine InfoSoc-direktiivi artikli 2 tähenduses (UrhG § 16); (ii) InfoSoc-direktiivi art 5 lg 1 / ajutiste reproduktsioonide erandi analüüs; ja (iii) kas TDM-erand (UrhG § 44b) vabastab OpenAI vastutusest. Nende küsimuste sisuline analüüs järgneb vahetult allpool.

Kohtu analüüs algas küsimusest, kas suurte keelimudelite treenimine kujutab üleüldse endast reprodutseerimist. Toetudes InfoSoc-direktiivi artikli 2 laiale reproduktsioonimõistele, leidis kohus, et treenimise käigus toimub kaitstud teoste osaline reprodutseerimine mudelisse

⁶⁹ LG München I, 11.11.2025, 42 O 14139/24 (*GEMA v. OpenAI Ireland Limited, esimese astme otsus; apellatsioon menetluses*); OLG Hamburg, 10.12.2025, 5 U 104/24 (*Kneschke v. LAION e.V.*). Mõlemad Saksa kohtud kohaldavad ACI Adam doktriini analoogiat TDM-erandite kontekstis.

⁷⁰ *GEMA v. OpenAI Ireland Limited, Landgericht München I, 11.11.2025, Az. 42 O 14139/24, esimese astme otsus; apellatsioon menetluses; (GEMA vs OpenAI).*

⁷¹ LG München I, 11.11.2025, 42 O 14139/24 (*GEMA v. OpenAI*), lk 4-8. Vaidlusalused laulud: 'Helene Fischer "Atemlos durch die Nacht"', Herbert Grönemeier "Männer", Reinhard Mey "Über den Wolken" ja teised. Otsus ei ole käesoleva töö koostamise ajal jõustunud (OpenAI on esitanud apellatsioonikaebuse).

– parameetrite kujul, mis võimaldavad hiljem teoste sisu taasesitada.⁷² Kohus järgis seejuures laia autoriõiguslikku tõlgendust, mille kohaselt reproduktsioon ei piirdu teoste identse kopieerimisega, vaid hõlmab igasugust teose ülekandmist uuele andmekandjale viisil, mis võimaldab originaalse sisu tuvastamist.⁷³ Kohus märkis, et GPT-4o suutis lihtsa päringuga ("Kuidas on laul *“pealkiri”* sõnastatud?") reprodutseerida kõnealuste laulude refrääne peaaegu sõna-sõnalt, mis kinnitab, et kaitstud sisu on mudelisse talletatud.

Eriti oluline on kohtu käsitus nn memoriseerimise fenomeni kohta. Neuraalvõrkude kontekstis tähistab memoriseerimine olukorda, kus mudel on treenimise käigus omandanud teatavate tekstide täpse sisu selliselt, et suudab neid hiljem peaaegu veatult reprodutseerida. Carlini jt 2023. aasta uuringus tõendati, et suuremahulised keelimudelid suudavad sõna-sõnalt reprodutseerida kuni 1% treeningandmestikust – kuid see arv ei ole universaalne. Memoriseerimise ulatus sõltub mudeli parameetrite arvust, sellest mitu korda konkreetne tekst treeningkorpuses esineb, ning kasutatavast päringumeetodist.⁷⁴ Laulusõnad on eriti kõrge memoriseerimise riskiga just seetõttu, et need ilmuvad treeningandmetes korduvalt: korraga laulusõnade veebisaitidel, ajakirjandustekstides ja sotsiaalmeedia postitustes. GEMA asjas demonstreeritud reprodutseerimistulemused peegeldavad just seda kordumismehhanismi. Kohus leidis, et see memoriseerimine kujutab endast püsivat reproduktsiooni, mis erineb fundamentaalselt ajutistest reproduktsioonidest.⁷⁵

Küsimuse üle, kas kohaldatav on UrhG § 44a (ajutised reproduktsioonid) erand, leidis kohus, et see erand ei kohaldu käesoleval juhul.⁷⁶ UrhG § 44a, mis põhineb InfoSoc-direktiivi artikli 5 lõikel 1, kaitseb üksnes reproduktsioone, mis on "lühiajalised või juhusliku iseloomuga" ning mille ainus eesmärk on võimaldada tehnoloogilist protsessi seaduslikuks kasutuseks. Käesoleval juhul ei olnud treenimine vahepealse reproduktsiooniga seotud protsess, vaid treenimise eesmärk oli just nimelt luua mudel, mis säilitab teoste sisu. Reprodutseerimine ei olnud seega lühiajaline ega juhuslik, vaid eesmärgipärane.⁷⁷

TDM-erandi (UrhG § 44b) kohaldamise analüüsis rõhutas kohus, et nimetatud erand on mõeldud eelkõige teadusliku analüüsi hõlbustamiseks, mitte aga teoste sisu imendamiseks mudelisse sellisel kujul, et see on hiljem taasesitatav.⁷⁸ UrhG § 44b lõige 1 nõuab, et

⁷² *GEMA vs OpenAI*, punkt 28.

⁷³ *GEMA vs OpenAI*, punkt 34.

⁷⁴ Carlini, N. jt. Quantifying Memorization Across Neural Language Models. - 11th International Conference on Learning Representations (ICLR) 2023, lk 1-24, siin lk 8.

⁷⁵ *GEMA vs OpenAI*, punkt 45.

⁷⁶ UrhG § 44a (Vorübergehende Vervielfältigungshandlungen - ajutised reproduktsioonid) ja § 44b (Text und Data Mining).

⁷⁷ *GEMA vs OpenAI*, punkt 52.

⁷⁸ *GEMA vs OpenAI*, punkt 55.

reprodutseerimine toimuks "teadustegevuse eesmärgil" ja et nõuetele vastav TDM-tegevus "ei ületa seda, mis on protsessi jaoks hädavajalik". Kohtu hinnangul ületab memoriseerimine seda piiri: kui mudelisse on talletatud laulusõnade kaitstud sisu, mis võimaldab nende hilisemat täpset reprodutseerimist, on selle talletamise eesmärk midagi enamat kui lihtsalt teksti analüüs.⁷⁹ Lisaks rõhutas kohus, et TDM-toimingutele kohaldatakse nõuet, et reprodutseeritud koopiad kustutatakse pärast andmekaevet; käesoleval juhul on reprodutseerimine aga jäänud mudelisse alaliselt.⁸⁰

Kohus rahuldab GEMA nõuded suuremas osas ja kohustas OpenAI'd: (i) lõpetama kaitstud laulusõnade reprodutseerimise keelimudelites (*Large Language Models*) ning nende avalikuks tegemise vestlusroboti väljundites; (ii) avalikustama teabe treenimisel kasutatud andmete kohta (*Auskunftsanspruch*); ja (iii) hüvitama GEMA-le tekitatud kahju (*Schadenersatz*).⁸¹ Kohus lükas tagasi OpenAI argumendi, et tema tegevus on kaitstud TDM-erandi alusel. Kohus analüüsis TDM-erandi (§ 44b UrhG) kohaldamatust ning leidis, et TDM-erand ei kohaldu, kuna memoriseeritud laulusõnade talletamine mudelisse on üle selle, mis on TDM-protsessiks hädavajalik, ning kaitstud sisu on mudelis alaliselt taasesitatavana säilinud.

GEMA vs OpenAI otsuse tähtsus ei piirdu Saksamaa jurisdiktsiooniga, ehkki esimese astme iseloom nõuab sellele tingimuslikku toetust. Otsus kehtestab kaks põhimõtet, mis on tuletatavad EL-i ühtsest autoriõigusest. Esiteks, AI treenimine kujutab endast reprodutseerimist InfoSoc-direktiivi artikli 2 tähenduses, kui treenimise tulemusena on kaitstud teoseid võimalik mudelis tuvastada; teiseks, TDM-erand ei laiene memoriseerimisele mudelisse lõplikul kujul.⁸² Mõlemad põhimõtted põhinevad EL-i ühtsel autoriõigusel – need kehtivad sõltumata GEMA otsusest ning on relevantssed ka Eesti ja teiste liikmesriikide kohtute jaoks, ehkki OLG Müncheni apellatsioonimenetluse tulemus võib neid täpsustada.

2.3. *Kneschke v. LAION*: TDM-erandite kohaldamine ja *opt-out* mehhanism

Hamburgi Apellatsioonikohtu (*Oberlandesgericht Hamburg*, edaspidi: OLG Hamburg) 10. detsembri 2025. aasta otsus asjas *Kneschke vs LAION e.V.* (*Az. 5 U 104/24*, edaspidi: *Kneschke vs LAION*)⁸³ täiendab *GEMA vs OpenAI* otsust olulisest küljest. Kui Müncheni kohus keskendus treenimise autoriõiguslikule kvalifikatsioonile, käsitles Hamburgi

⁷⁹ *GEMA vs OpenAI*, punkt 63.

⁸⁰ *GEMA vs OpenAI*, punkt 67.

⁸¹ *GEMA vs OpenAI*, punkt 72 (*Anspruch auf Auskunft ja Schadenersatz*).

⁸² Quintais, DSM Commentary, lk 38.

⁸³ *Kneschke v. LAION e.V.*, Oberlandesgericht Hamburg, 10.12.2025, *Az. 5 U 104/24*. (*Kneschke vs LAION*).

apellatsioonikohus peamiselt küsimust, millistel tingimustel saab TDM-erandi kohaldamisest ette nähtud *opt-out* mehhanism olla õiguslikult siduv ning milline on "seadusliku kättesaadavuse" nõude sisu avaliku veebi kontekstis.

Fotograaf Robert Kneschke esitas hagi LAION e.V. vastu, väites et tema fotode kaasamine 5,85 miljardi pildi-teksti paari sisaldavasse avaliku andmekogumisse LAION-5B – URL-viidete ja CLIP-pisipiltide kujul – kujutab autoriõiguse rikkumist.⁸⁴

Kohus (esimene aste: LG Hamburg, 310 O 227/23, 27.09.2024; apellatsioon: OLG Hamburg, 5 U 104/24, 10.12.2025) pidi lahendama kolm põhiküsimust: (i) kas URL-viidete ja pisipiltide kogumine on reprodutseerimine UrhG § 16 mõttes; (ii) kas Kneschke fotod olid "seaduslikult kättesaadavad" DSM-direktiivi art 4 lg 1 tähenduses; ja (iii) kas robots.txt-fail kujutab art 4 lg 3 kohast kehtivat masinloetavat *opt-out* deklaratsiooni. Nende küsimuste sisuline analüüs järgneb allpool.

OLG Hamburg kinnitas esimese astme otsust põhiosas, kuid täpsustas oluliselt *opt-out* mehhanismi hindamise kriteeriume ning "seadusliku kättesaadavuse" mõiste sisu.⁸⁵ Apellatsioonikohus leidis, et Kneschke fotod olid oma avaldamise ajal "seaduslikult kättesaadavad", kuna need olid fotograafi enda poolt avaldatud avalikul veebisaidil ilma juurdepääsupiiranguteta. Kohus rõhutas, et "seaduslik kättesaadavus" ei nõua õiguste omaja sõnaselget luba TDM-toiminguteks, vaid tähendab seda, et juurdepääs teosele on toimunud sellises kontekstis, kus see on objektiivselt lubatav.⁸⁶ Selles osas erineb avalikult veebis avaldatud sisu fundamentaalselt piraatlikest andmekogumistest – kui õiguste omaja on ise avalikustanud oma teose veebis, on faktilise juurdepääsu alusel saavutatav ka "seaduslik" juurdepääs TDM-direktiivi tähenduses.

Asja keerukamaks pöördepunktiks osutus *opt-out* mehhanismi analüüs. DSM-direktiivi artikkel 4 lõige 3 sätestab, et artikli 4 lõikes 1 sätestatud TDM-erand ei kohaldu juhul, kui õiguste omajad on "masinloetaval viisil" selgelt väljendanud, et nad keelavad oma teoste kasutamise TDM-i eesmärkidel.⁸⁷ Kneschke argumenteeris, et tema veebisaidil kasutatud robots.txt-fail sisaldas keelu kraapimisrobotite suhtes (Disallow-direktiiv), mis peaks kujutama endast sellist masinloetavat *opt-out* deklaratsiooni.⁸⁸ OLG Hamburg ei pidanud seda

⁸⁴ LG Hamburg, 27.09.2024, 310 O 227/23 (*Kneschke v. LAION e.V.*, esimese astme otsus); OLG Hamburg, 10.12.2025, 5 U 104/24 (apellatsioonikohtu otsus). Vt Leistner, M.; Antoine, L. TDM and AI Training in the European Union – From 'LAION' to Possible Ways Ahead? – GRUR International, kd 74, nr 11, 2025, lk 1027-1044.

⁸⁵ *Kneschke vs LAION*, OLG Hamburg, punkt 31.

⁸⁶ *Ibid.*, punkt 38.

⁸⁷ DSM-direktiivi artikkel 4 lõige 3.

⁸⁸ *Kneschke vs LAION*, OLG Hamburg, punkt 43.

argumendita kergekäeliseks, kuid leidis siiski, et robots.txt ei vasta DSM-direktiivi artikli 4 lõike 3 nõudele.

Kohus analüüsis, milline standard kehtib "masinloetava" *opt-out* deklaratsiooni suhtes, ning jõudis kaalutletud järeldusele, mis omab olulist tähendust tulevasele DSM-erandi tõlgendamisele.⁸⁹ Kohtu hinnangul ei ole robots.txt piisav *opt-out* instrument, kuna: (i) robots.txt on primaarsetel eesmärkidel mõeldud otsingumootorite indekseerimise piiramiseks, mitte autoriõigusliku *opt-out* deklaratsioonina; (ii) LAION tegi andmekogumi 2021. aastal, mil ei olnud veel välja kujunenud selgeid standardeid TDM-spetsiifilise *opt-out* deklaratsiooni osas; ja (iii) *opt-out* deklaratsioon peab olema selgelt suunatud TDM-tegevuse keelustamisele, mitte lihtsalt veebikraapimisele üldiselt.⁹⁰ Kohus märkis, et alates 2022.-2025. aastast välja töötatud spetsiifilised tehnilised standardid – nagu IPTC foto-metadata standard, tdm-reservation HTTP-päis ning Schema.org TDM-spetsiifilised omadused – kujutavad endast asjakohaseid masinloetavaid *opt-out* instrumente, mida tulevikus rakendades saaks fotograaf oma teosed TDM-kasutusest välistada.

Kneschke vs LAION asja tulem – et LAION tegevus oli TDM-erandi poolt kaitstud – tundub esmapilgul vastuolus piraatandmete kasutamise autoriõigusliku kvalifikatsiooniga. Siiski on kahe olukorra vahel fundamentaalne erinevus: LAION-5B andmekogumi aluseks olnud pildid olid avaldatud avalikul veebil nende õiguste omajate poolt, ning juurdepääs neile oli seega "seaduslik", kuigi *opt-out* deklaratsiooni polnud esitatud. Piraatliku andmekogumi puhul – nagu Books3, mis sisaldab autoriõigusega kaitstud raamatuid, mis on üles laaditud ilma kirjastajate loata – puudub igasugune seaduslik juurdepääs.⁹¹ See eristus on käesoleva töö põhiargumendi seisukohast kriitiline: DSM-direktiivi TDM-erandid kaitsevad "seadusliku kättesaadavuse" korral TDM-toiminguid, kuid ei laiene piraatlikest allikatest pärit andmetele.

Kneschke v. LAION lahend on kahepoolse sõnumiga: kuigi kohus kinnitas *opt-out* deklaratsiooni masinloetavuse nõuet, jäi LAION *opt-out*-deklaratsiooni puudulikkuse tõttu siiski TDM-erandi kaitseulatusse ja võitis kohtuasja. Seega seisneb hoiatus *opt-out* nõude tehnilises täpsuses – masinloetavus, robots.txt standard – mitte TDM-tegevuse keelustamises kui sellises. Positiivne sõnum AI-arendajatele on järgmine: seadusliku *opt-out* mehhanismi range järgimine – sh masinloetav robots.txt või muu *opt-out* deklaratsioon – tagab TDM-erandi kaitseulatusse DSM art 4 alusel. Alates 2022.-2025. aastast välja töötatud spetsiifilised tehnilised standardid (IPTC foto-metadata standard, tdm-reservation HTTP-päis,

⁸⁹ *Ibid.*, punkt 48.

⁹⁰ *Kneschke vs LAION*, OLG Hamburg, punkt 51.

⁹¹ Senftleben, M.; Buijtelar, L. Robot Creativity: An Incentive-Based Neighbouring Rights Approach. – European Intellectual Property Review, 42, nr 12, 2020, lk 797-811, siin lk 804.

Schema.org TDM-spetsiifilised omadused) kujutavad selleks sobivaid instrumente. LAION-i tegevus oli TDM-erandi alusel lubatav just seetõttu, et pildid olid seaduslikult avaldatud – piraatlikust allikast pärit andmete kasutamine aga ei oleks TDM-erandi kaitseulatusse jäänud. Isegi kui täna on teatud andmekogumid faktiliselt kättesaadavad ning *opt-out* standardid ebaselged, peab AI-arendaja tuleviku standardite arengut silmas pidama. *Opt-out* mehhanismi eiramine hiljem kehtestatud selge standardite korral ei pruugi olla enam kaitstud TDM-erandi poolt, mis tähendab, et andmekogumite pidev uuendamine ja *opt-out* nõuete järgimine on EL-i õiguses oluline riskimaandamise meede.

2.4. Piraatluse mõju TDM-erandi kehtivusele: õigusdogmaatiline analüüs

Alapeatükkides 2.1-2.3 esitatud analüüsi põhjal on võimalik sõnastada selge ja põhjendatud positsioon EL-i autoriõiguse seisukohast: piraatlikest allikatest pärit andmete kasutamine generatiivse AI treenimiseks ei kuulu DSM-direktiivi artiklite 3 ega 4 kohaldamisalasse. See järeldus tugineb mitmele üksteist toetavale argumendile.

Esimene argument on tekstuaalne – mõlemad TDM-erandid – artikkel 3 lõige 1 ja artikkel 4 lõige 1 – nõuavad selgesõnaliselt, et TDM-toiminguid teostatakse teostel, millele on "seaduslik juurdepääs"⁹² või mis on "seaduslikult kättesaadavad".⁹³ Piraatlikud andmekogumid ei vasta sellele nõudele mingil juhul: teosed on piraatlikesse andmebaasidesse laaditud ilma õiguste omajate loata, mistõttu puudub juurdepääsule igasugune seaduslik alus. See on selge lävendiküsimus – ilma seadusliku juurdepääsuta ei saa TDM-erand kohalduda.

Teine argument tugineb kohtupraktikale. *ACI Adam* doktriini kohaselt ei laiene autoriõiguserandid koopiatele, mis on tehtud ebaseaduslikest allikatest, kuna see tooks kaasa rikkumise *de facto* seadustamise.⁹⁴ Ehkki *ACI Adam* käsitles isikliku koopia erandit, on selle põhimõtte laiendamine TDM-eranditele süstemaatiliselt põhjendatav: mõlemad erandid eeldavad, et kasutaja kontakt teosega on õiguspärane, ning mõlemad on ette nähtud seaduslikult kättesaadavate teoste piiratud kasutuse hõlbustamiseks, mitte ebaseadusliku levitamise seadustamiseks.⁹⁵

Kolmas argument tuleneb direktiivi eesmärgist. DSM-direktiivi TDM-erandid on vastu võetud eesmärgiga edendada teadusuuringuid, innovatsiooni ja kultuuripärandi digiteerimist

⁹² DSM-direktiivi artikkel 3 lõige 1.

⁹³ *Ibid.*, artikkel 4 lõige 1.

⁹⁴ *ACI Adam*, punkt 41.

⁹⁵ Synodinou, T.-E.; Vrakas, G. Lawful Access as a Gatekeeper for TDM in the EU. - Verfassungsblog, 17.11.2025. DOI: 10.17176/20251117-142050-0.

viisil, mis on kooskõlas õiguste omajate huvidega.⁹⁶ Direktiivi põhjendus 18 rõhutab, et erand ei tohi kahjustada õiguste omajate huve ega seada ohtu teoste tavapärase kasutamist. Piraatlikest allikatest andmete kogumise lubamine TDM-erandi alusel oleks selgelt vastuolus selle eesmärgiga: see annaks AI-arendajatele stimulatsiooni hankida andmeid piraatlikest allikatest selle asemel, et maksta litsentsitasusid, mis tähendaks otsest kahju autorite ja kirjastajate majanduslikele huvidele.⁹⁷

Eriliselt keeruline on küsimus olukordadest, kus AI-arendaja ei olnud teadlik, et tema poolt kasutatud andmekogum sisaldas piraatlikust allikast pärit teoseid. Selliseid olukordi on kirjeldatud näiteks asjas, kus andmekogumi koostas kolmas isik ning AI-arendaja omandas selle laialt levinud uurimisandmestikuna. EL-i autoriõiguse üldpõhimõtete järgi on autoriõiguse rikkumine objektiivne – teadlikkus rikkumisest mõjutab üksnes kahju ulatust, mitte rikkumise tuvastamist.⁹⁸ Selline lähenemine on kooskõlas InfoSoc-direktiivi ja DSM-direktiivi loogikaga, kuid tekitab praktilist ebakindlust AI-arendajatele, kes soovivad tegutseda heas usus.

Eraldi küsimus on nn "saastunud mudeli" autoriõiguslik staatus – olukord, kus AI-mudel on treenitud andmekogumil, mis sisaldab osaliselt piraatlikust allikast pärit teoseid. Kas kogu mudel on "saastunud" (mõistet käsitletakse täpsemalt ptk 4.1-s), või piisab vastutuse vältimiseks sellest, et eemaldada piraatlikud osad andmekogumist? *GEMA vs OpenAI* otsus viitab sellele, et kohus hindab eelkõige, kas mudel on võimeline konkreetseid teoseid memoriseeritult reprodutseerima – kui jah, siis kujutab see endast autoriõiguse rikkumist.⁹⁹ "Saastunud mudeli" küsimus on aga keerukam kui lihtsalt treeningandmete puhastamine: kas ja kuidas saab mudelit täiendavalt treenida selliselt, et varem memoriseeritud sisu kustutatakse, on avatud nii tehniline kui ka õiguslik küsimus, mida käsitletakse käesoleva töö 4. peatükis.

Olulise perspektiivi lisab Euroopa Kohtule esitatud eelotsusetaotlus asjas *Like Company SRL vs Google Ireland Ltd jt (C-250/25)*,¹⁰⁰ kus muuhulgas palutakse Euroopa Kohtul selgitada "seadusliku kättesaadavuse" täpset tähendust DSM-direktiivi artikli 4 kontekstis. Euroopa Kohtu seisukoht selles küsimuses on oodatavalt oluline täpsustus praegusele õigusliku ebakindluse olukorrale. Ehkki eelotsusetaotluse sisu ja kohtu võimalikku seisukohta ei ole käesoleva töö koostamise ajal teada, annab otsuse ootamine

⁹⁶ DSM-direktiivi põhjendus 18.

⁹⁷ Husovec; Quintais, *How to License*, lk 517.

⁹⁸ Elkin-Koren, N.; Hachohen, G. *Cognitive Copyright*. - Yale Law Journal, 133, nr 1, 2023, lk 1-73, siin lk 42.

⁹⁹ LG München I, 11.11.2025, 42 O 14139/24 (*GEMA v. OpenAI GmbH*).

¹⁰⁰ *Like Company SRL v. Google Ireland Ltd jt, Euroopa Kohtu eelotsusetaotlus (2025), C-250/25 (menetluses)*.

lisaargumendi ettepanekuks, et EL-i seadusandja peaks selgitama "seadusliku juurdepääsu" mõistet ning piraatlikele allikatele viitamist DSM-direktiivi muudatuste kaudu – selleks esitab käesoleva töö autor ettepanekud 4. peatükis.¹⁰¹

Peatüki kokkuvõttena võib öelda, et EL-i autoriõigusraamistik ei paku AI-arendajatele kaitset piraatlikelt allikatelt kogutud andmete kasutamise korral. "Seadusliku juurdepääsu" nõude grammatiline, teleoloogiline ja süstemaatiline tõlgendus koos ACI Adam doktriiniga moodustavad käesoleva töö normatiivse aluse. Seda tõlgendust kinnitab jõustunud *Kneschke vs LAION* (OLG Hamburg, apellatsioonikohtu lõplik otsus) seadusliku kättesaadavuse standardite osas, ning lisatuge pakub – apellatsioonimenetluse pooleli olekut arvestades tingimusliku kaaluga – *GEMA vs OpenAI* esimese astme otsus (LG München I). Need allikad koos moodustavad ühtse terviku, mille kohaselt: (i) piraatlikest allikatest pärit andmete kasutamine on InfoSoc-direktiivi artikli 2 kohane autoriõigusliku reprodutseerimisõiguse rikkumine; (ii) DSM-direktiivi TDM-erandid ei kohaldu puuduva "seadusliku juurdepääsu" tõttu; ja (iii) memoriseerimine kujutab endast eriti selget rikkumist, kuna mudelisse talletatud kaitstud sisu on reprodutseeritav. Need järeldused loovad aluse piraatandmetel treenitud mudelite suhtes autoriõiguslike nõuete esitamiseks ning on oluliseks lähtepunktiks *de lege ferenda* arutelu jaoks, mida käsitletakse käesoleva töö 4. peatükis.¹⁰²

Hea usu kaitse ja objektiivne vastutus EL autoriõiguses. Iseseisev, kuid praktikas oluline küsimus on, millist rolli mängib AI-ettevõtte teadlikkus piraatlikust allikast tema vastutuse kindlaksmääramisel. EL-i autoriõiguse üldpõhimõtete kohaselt on autoriõigusliku rikkumise tuvastamine objektiivne: InfoSoc-direktiivi artiklist 2 tulenev reprodutseerimisõigus laieneb sõltumata sellest, kas rikkuja teadis oma tegevuse ebaseaduslikkusest.¹⁰³ Teadlikkuse puudumine ei välista vastutust – see mõjutab üksnes kahju ulatust.

IPRED-direktiivi artikkel 13 lõige 2 sätestab erisuse: kui rikkuja ei teadnud ega pidanud mõistlikult teadma, et ta tegeles rikkumisega, võib kohus kohaldada immateriaalset hüvitist (nt eeldatavat litsentsitasu) karistusliku hüvitise asemel.¹⁰⁴ See tähendab praktikas, et AI-ettevõtte, kes kasutas kolmanda osapoole koostatud andmekogumit teadmata selle piraatlikust päritolust, ei väldi vastutust, kuid võib saada oluliselt madalama kahjuhüvitise kui tahtliku rikkumise korral. See eristus on käesoleva töö *de lege ferenda* argumentatsiooni

¹⁰¹ Elkin-Koren; Hacoen, Cognitive Copyright, lk 55.

¹⁰² Quintais, J.P. jt. Licensing Copyright for AI Training: Considerations and Policy Options. SSRN Working Paper 2024, lk 8.

¹⁰³ InfoSoc-direktiiv 2001/29/EU, art 2: reprodutseerimisõiguse rikkumine on objektiivne. Vt Leistner, M. EU Copyright Law. IIC, kd 48, 2017, lk 381-403.

¹⁰⁴ Euroopa Parlamendi ja nõukogu direktiiv 2004/48/EU intellektuaalomandi õiguste jõustamise kohta (IPRED), art 13 lg 2: kui rikkuja ei teadnud ega pidanud mõistlikult teadma, et ta tegeles rikkumisega, kohalduv immateriaalne hüvitis.

jaoks oluline kahel põhjusel. Esiteks näitab see, et due diligence kohustuse kehtestamine (ptk 4.3) ei ole pelgalt haldusnõue, vaid omab otsest mõju vastutuse ulatusele: tõendatud due diligence täitmine on kõige tugevam argument "mõistliku mitteteadmise" kasuks IPRED art 13 lg 2 tähenduses. AI-ettevõtte, kes on süstemaatiliselt kontrollinud treeningandmete päritolu, on oluliselt paremas positsioonis kahjuhüvitise vähendamise nõudmisel.¹⁰⁵ Teiseks kinnitab see objektiivse vastutuse loogika, et n-ö süütu rikkuja seisund ei mõjuta vastutuse tuvastamist ennast – piraatandmetel treenimine jääb rikkumiseks sõltumata tahtlusest. Seetõttu ei ole teadlikkuse puudumine EL-i õiguses iseseisvaks kaitsevahendiks, erinevalt mõnedest USA õiguse tõlgendustest.

Eriliselt keeruline on küsimus osaliselt saastunud andmekogumite õigusliku staatuse kohta. Reaalsuses ei koosne paljud AI-treeningandmestikud täielikult piraatlikest allikatest – sageli on suurem osa seaduslik, kuid teatud osakaal (nt 5%) võib sisaldada ebaseaduslikku materjali. Kas sellisel juhul langeb TDM-erandi kaitse täielikult ära kogu andmestiku suhtes? EL-i autoriõigus ei tunne selgesõnaliselt *de minimis* erandit: InfoSoc-direktiivi reprodutseerimisõigus on põhimõtteliselt absoluutne, mida kinnitab kohtuasi *Infopaq International A/S vs Danske Dagblades Forening* (C-5/08), kus ELTK leidis, et isegi 11-sõnaline tekstilõik võib olla kaitstud. Seega ei ole väike osakaal piraatlike teoseid andmestikus iseenesest relevantne TDM-erandi kohaldatavuse hindamisel – iga konkreetne ebaseaduslikust allikast pärit teos kujutab eraldiseisvat rikkumist. Küsimus, kas kogu mudel on "saastunud" ja kas vastutus on proportsionaalne piraatlikest allikatest pärit teoste osakaaluga, jääb EL-i kohtupraktikas avatuks ning sõltub eeldatavasti tulevaste ELTK lahendite suunistest.

2.5. Argumendid õigustatud kasutuse kasuks ja nende kriitika

Käesoleva töö pealkiri küsib, kas piraatlikelt allikatelt kogutud andmetel treenimine kujutab endast autoriõiguse rikkumist või õigustatud kasutust. Analüütiline ausus nõuab selle dihhotoomia teise poole tõsist käsitlemist – mitte üksnes formaalselt, vaid sisuliselt, arvestades argumente, millel on intellektuaalne kaal. Järgnev analüüs esitab neli argumenti, mida võiks kasutada õigustatud kasutuse kasuks EL-i ja USA õiguse raames ning toob igaühe suhtes esile peamised vastuargumendid.

Esimene argument: transformatiivsuse fundamentaalne loomus.

¹⁰⁵ Tehisintellekti käsitleva määruse art 53 lg 1 p c kohustab ÜTEM-pakkujaid kehtestama autoriõiguse vastavuse poliitika. *Due diligence* kohustuse normatiivse aluse kohta vt kaesoleva töö ptk 4.3.

Väide seisneb selles, et AI treenimine erineb olemuslikult teose reprodutseerimisest inimkasutaja jaoks: mudel ei loe teksti nagu inimene, vaid töötleb statistilisi mustreid – sõnade tõenäosusjaotusi, süntaksit ja semantilisi seoseid. Lõpptulemus ei sisalda ühtegi lauset originaaltekstist ning mudelisse ei ole kodeeritud ühtki konkreetset teost identifitseeritaval kujul. Selle argumendi järgi on AI treenimine sarnane kunstnikule, kes vaatab tuhat maali ning õpib stiili ja tehnikat – "stiili" ei saa autoriõigusega kaitsta, seega ei ole ka AI mudelisse "talletatud" stiil kaitstud.¹⁰⁶ Argument on teoreetiliselt huvitav, kuna mõtestab AI treenimist fundamentaalselt erinevana traditsioonilistest reprodutseerimistoimingutest.

Vastuargument: empiiriline reaalsus ei toeta "ainult statistilised mustrid" positsiooni. Carlini jt 2023. aasta uuring kinnitas, et suurtel keelemudelitel esineb memoriseerimise nähtus – kuni 1% treenimisandmetest on otseselt reprodutseeritav vastusena lihtsatele päringutele.¹⁰⁷ *GEMA v. OpenAI* esimese astme otsus (LG München I, apellatsioon menettluses) rakendas just seda empiirilist leidu kohtus: GPT-4o reprodutseeris konkreetseid laulusõnu sõna-sõnalt, mis tõendas, et "ainult statistilised mustrid" väide ei vasta tegelikkusele kõigi mudelite puhul. Seega ei kehti siin kunstniku-analoogia: kunstnik ei suuda reprodutseerida meenutusest sõna-sõnalt tervet maali – kuid suurmudel suudab reprodutseerida terve peatüki. Transformatiivsuse argument on seega teoreetiliselt huvitav, kuid praktiliselt ümberlükatav memoriseerimise tõendusmaterjaliga.

Teine argument: ühiskondlik kasu kui õigustav tegur.

Autoriõiguse fundamentaalne eesmärk ei ole ainult autorite kaitse – see on tasakaal autorite õiguste ja ühiskondliku kasu vahel. Berni konventsioon, InfoSoc-direktiiv ja USA Copyright Act tunnustavad kõik seda tasakaalu. Generatiivsed AI mudelid pakuvad väidetavalt suurt ühiskondlikku kasu: teadusuuringute kiirendamine, hariduslik juurdepääs, ligipääsetavuse parandamine ning innovatsioon tervishoius ja muudes valdkondades. USA Autoriõiguse Amet märgib oma 2025. aasta aruandes otseselt, et "*fair use* analüüs peab arvestama AI ühiskondliku kasuga".¹⁰⁸ Argumendi kaalukus seisneb selles, et kui autoriõigus blokeeriks AI arengu täielikult, võib küsida, kas selline tulemus vastab autoriõiguse kujundamise algele eesmärgile.

Vastuargumendiks on see, et ühiskondlik kasu ei saa õigustada konkreetset autoriõiguse rikkumist juhul, kui seaduslik alternatiiv on kättesaadav. Litsentsimisturud on aktiivselt

¹⁰⁶ Transformatiivsuse argumendi kohta vt Fisher, W.W. Promises to Keep. Stanford UP, 2004; Lemley, M.A.; Casey, B. Fair Learning. Texas Law Review, kd 99, 2021, lk 743-775.

¹⁰⁷ Carlini, N. jt. Quantifying Memorization Across Neural Language Models. ICLR, 2023 (arXiv:2202.07646): kuni 1% treenimisandmetest on otseselt reprodutseeritav. *GEMA v. OpenAI* otsus kinnitas seda empiirilist leidu kohtupraktikas.

¹⁰⁸ USA Autoriõiguse Amet. Copyright and Artificial Intelligence. Washington D.C., 2025, lk 48-51.

arenenud: Getty Images, Associated Press, Axel Springer ja The New York Times on kõik sõlminud AI treenimise lepinguid.¹⁰⁹ Kui seaduslik tee on avatud, ei ole ühiskondliku kasu argument piisav, et õigustada ebaseadusliku tee kasutamist. Lisaks ei kohaldu ühiskondliku kasu kaalutus EL-i TDM-erandite kontekstis: need erandid on sõnastatud konkreetsete lubade, mitte üldise kaalumisklauslina – seega ei saa EL-i kohus kaaluda abstraktset ühiskondlikku kasu TDM-erandi lubatavuse hindamisel samal viisil nagu USA *fair use* test seda teeb.

Kolmas argument: piraatluse moraalne ambivalentsus teadusliku ligipääsu kontekstis.

LibGen ja Sci-Hub on teadusringkondades laialdaselt kasutatavad alternatiivid kõrgete paywallide taga olevale teaduskirjandusele – eriti arengumaades ja piiratud eelarvega ülikoolides. Teadusartiklite taga olevad paywallid on iseenesest akadeemilises maailmas vastuolulised ning mõnede autorite arvates ebaõiglased. Selle argumendi järgi ei ole piraatlikust allikast treeningandmete kasutamine moraalselt samaväärne kaubanduslike raamatute varastamisega, kui algne materjal oli niigi ebaseaduslikult piiritletud. See küsimus on relevantne eriti DSM art 3 teaduserandi kontekstis.

Vastuargumendiks on – õigusanalüüs ei saa sõltuda moraalsetest hinnangutest konkreetsete piraatplatvormide suhtes. Oluline on vahet teha kahe eri tüüpi piraatlikul allikal: teaduskirjanduse piraatplatvormid (nt Sci-Hub) ja kaubanduslike teoste piraatplatvormid.¹¹⁰ Books3 ja PiLiMi andmekogumid, mida kasutati *Bartz v. Anthropic* asjades tuvastatud viisil, ei sisalda peamiselt teadusartikleid – need sisaldavad peamiselt kaubanduslikke romaane, mille puhul teadusliku ligipääsu argument ei kehti. Lisaks, AI-ettevõtted kasutavad neid andmeid kaubanduslikku kasumit taotlevateks eesmärkideks, mitte teaduseesmärkidel, mistõttu DSM art 3 teaduserand (mis nõuab mitte-ärilist eesmärki) ei kohaldu.

Neljas argument: *opt-out* süsteemi struktuuriline ebaõiglus.

DSM-direktiivi art 4 *opt-out* mehhanism eeldab, et autorid aktiivselt keelavad oma teoste kasutamise masinloetava märgistuse kaudu. Enamik autoreid ei tea sellest mehhanismist, ei suuda seda rakendada ja ei oma ressursse selle jälgimiseks. *De facto* kaitsevad end suurkirjastajad, kellel on õigusosakonnad ja tehniline võimekus. Väikeautorid jäävad praktiliselt kaitseta sõltumata sellest, kas andmekogum on seaduslik või piraatlik. Argument on, et piraatlikest allikatest treenimine ei kahjusta tegelikult väikeautoreid erinevalt

¹⁰⁹ Seadusliku litsentsimisaktiivuse kohta vt Getty Images v. Stability AI, 1:23-cv-00135 (D. Del., 2025); AP-OpenAI koostooleping, 2023; Axel Springer-OpenAI kokkulepe, 2024; The New York Times Co. v. Microsoft Corp. jt, nr 23-cv-11195 (S.D.N.Y., 2025).

¹¹⁰ Books3 ja PiLiMi andmekogumite koosseisu kohta vt Gao, L. jt. The Pile: An 800GB Dataset. arXiv:2101.00027v1 [cs.CL], 2021, lk 6-7. Bartz v. Anthropic, nr 3:24-cv-05417-WHA (N.D. Cal., 23. juuni 2025), lk 4-7.

seaduslikust treenimisest, kuna mõlemal juhul puudub autori nõusolek ja tasu.¹¹¹ See on süsteemse ebaõigluse argument – küsib, kas kehtiv *opt-out* mudel on üldse autoriõiguse eesmärgiga kooskõlas.

Vastuargumendiks on see, *opt-out* süsteemi ebatäiuslikkus on regulatiivne probleem, mis vajab lahendust – täpselt mida käesoleva töö *de lege ferenda* ettepanekud pakuvad. Süsteemi struktuurne nõrkus ei õigusta süsteemi ignoreerimist. Kui kehtiv süsteem on ebaõiglane, on lahendus süsteemi parandamine, mitte selle eiramist legitimeeriv kohtulahend. Lisaks on tehisintellekti käsitleva määruse art 53 lg 1 p c muutmisel *opt-out* jälgimine nüüdseks ÜTEM-pakkujate avalik-õiguslikuks kohustuseks, mis parandab süsteemi toimimist mõnevõrra.

Nende nelja argumendi analüüs näitab, et "õigustatud kasutuse" positsioon ei ole tühine – selles on intellektuaalne sisu ja see väärib tõsist käsitlemist. Kuid ühelgi neist argumentidest ei ole piisavalt jõudu, et ületada piraatlikelt allikatelt treenimise põhitakistust EL-i õiguses – seadusliku juurdepääsu nõuet, mis toimib lävenditingimuseks ning mille puudumisel ei saa TDM-erandile tugineda sõltumata treenimise eesmärgist, ühiskondlikust kasust või *opt-out* süsteemi toimivusest. USA õiguses on oluline katsetaks neljas faktor – turu kahjustamine –, mida käsitletakse käesoleva töö 3. peatükis. Seetõttu on töö pealkiri "rikkumine või õigustatud kasutus?" mõlemapoolse argumendi kohane – kuid EL-i kehtiva õiguse tekstuaalse, teleoloogilise ja süstemaatilise tõlgenduse alusel toetab analüüs kehtiva õiguse kõige põhjendatuma tõlgenduse kohaselt rikkumise positsiooni, mille lõplikku kinnitust oodatakse ELTK lahendist asjas C-250/25.

¹¹¹ DSM-direktiiv, art 4 lg 3. *Opt-out* süsteemi struktuurse nõrkuse kohta vt Montagnani, M.L. *Opt-Out in EU Copyright Law*. GRUR International, kd 73, 2024, lk 211-228.

3. PIRAATLUSE KÜSIMUS USA ÕIGUSES: *BARTZ, KADREY JA THOMSON REUTERS*

Käesolev peatükk ei käsitle Ameerika Ühendriikide õigust iseseisva normatiivse süsteemina, vaid võrdleva analüüsi komponendina. Selle eesmärk on selgitada, millisel viisil erineb *fair use*'i doktriin Euroopa Liidu kategooriaalsest erandisüsteemist ning milliseid alternatiivseid lahendusi pakub USA kohtupraktika piraatandmete kasutamisele tehisintellekti treenimisel.

Käesolev peatükk käsitleb piraatlikest allikatest pärit treenimisandmete autoriõiguslikku reguleerimist USA õiguse perspektiivist. USA õigus läheneb piraatlikest allikatest pärit treenimisandmete küsimusele fundamentaalselt erinevalt EL-i raamistikust: kuna USA autoriõiguse seaduses puudub üldine TDM-erand, lahendatakse kõik treenimisega seotud vaidlused *fair use*'i doktriini kaudu. 2025. aasta kohtupraktika on selles küsimuses andnud vastuolulisi tulemusi, mis muudab doktrinaalse analüüsi eriti vajalikuks.

3.1. *Fair use* neljaetapiline test piraatandmete kontekstis

USA autoriõigusliku analüüsi lähtepunktiks AI treenimise kontekstis on *fair use* doktriin vastavalt 17 U.S.C. §-le 107.¹¹² Erinevalt Euroopa Liidu DSM-direktiivi selgesõnaliselt sätestatud TDM-eranditest on *fair use* paindlik faktispetsiifiline doktriin, mis nõuab kõigi nelja faktori holistilist kaalumist.¹¹³ Kõrgeima Kohtu praktika raamistikus tähendab see, et ühtegi faktorit ei saa analüüsida isoleeritult ning et *fair use* võimalus sõltub konkreetsetest asjaoludest. Käesoleva töö seisukohast on põhiline küsimus, kas ja mil viisil mõjutab treeningandmete ebaseaduslik päritolu – ehk nende pärinemine piraatlikest allikatest – *fair use* hindamist.

17 U.S.C. § 107 sätestab, et *fair use* hindamisel arvestatakse järgmisi faktoreid: (1) kasutuse eesmärk ja iseloom, sealhulgas see, kas kasutus on kommertsiaalse iseloomuga või mittekasumliku hariduslikku eesmärki teeniv; (2) kaitstud teose iseloom; (3) kasutatud osa hulk ja olulisus terviku suhtes; ja (4) kasutuse mõju kaitstud teose potentsiaalsele turule või väärtusele.¹¹⁴ Need faktorid moodustavad mitteammenduva loetelu, mida ei saa mehaaniliselt kohaldada – kohus peab hindama kõiki faktoreid koosmõjus ning arvestama kasu kõiki asjaolusid.¹¹⁵ Praktikas on Kõrgeim Kohus rõhutanud, et esimene faktor (transformatiivsus) ja

¹¹² 17 U.S.C. § 107 (2018) - *fair use* erandi säte.

¹¹³ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

¹¹⁴ 17 U.S.C. § 107.

¹¹⁵ *Campbell*, lk 577-578.

neljas faktor (turu mõju) on kaalukaimad analüütilised kategooriad, kuigi kõik neli on formaalselt võrdses seisus.¹¹⁶

Esimese faktori juures on keskse transformatiivsuse kontseptsioon, mille Ameerika Kõrgeim Kohus kehtestas 1994. aasta otsuses *Campbell v. Acuff-Rose Music, Inc.*¹¹⁷ Kohus leidis, et *fair use* analüüsis on otsustav küsimus, kas kasutus "lisab midagi uut, muutes esmaseid materjale või lisades neile uue väljenduse, tähenduse või sõnumi".¹¹⁸ Seda transformatiivsuse nõuet täpsustasid edaspidi *Authors Guild, Inc. vs Google* (raamatute digitaliseerimine ja tsitaatide näitamine on transformatiivne)¹¹⁹ ning *Google LLC vs Oracle America, Inc.*, kus Kõrgeim Kohus jõudis järeldusele, et Java API-de kasutamine Androidi süsteemis oli transformatiivne, luues uue platvormi uutel eesmärkidel.¹²⁰ AI treenimise kontekstis on peamine argument, et mudelite treenimine on transformatiivne akt: mudel ei reprodutseeri teoseid kasutajale edastamiseks, vaid kasutab neid statistiliste mustrite väljaselgitamiseks – otstarbeks, milleks teoseid algselt ette nähtud ei olnud.¹²¹

Transformatiivsuse argumenti oluliselt kitsendas aga Kõrgeima Kohtu 2023. aasta märgiline otsus asjas *Andy Warhol Foundation for the Visual Arts, Inc. vs Goldsmith*.¹²² Asjas käsitleti Warholi portree-seeriat, mis põhines fotograaf Lynn Goldsmithi pildil laulja Prince'ist, ning küsimust, kas selle müümine ajakirjale kommertsliku albumikaane kujunduseks kujutab endast *fair use*'i. Kohus leidis 7:2 häältega, et ei kujuta – mõlemad teosed (foto ja kunstniku töö) täitsid samal turusegmendil samasugust kaubanduslikku eesmärki.¹²³ Kohus rõhutas, et esimese faktori hindamisel tuleb keskenduda konkreetse kasutuse eesmärgile, mitte abstraktsele transformatiivsusele: küsimus on, "kas ja mil määral on kõnealusel kasutusel erinev eesmärk või iseloom" originaalteos vaates.¹²⁴ Warholi lahend on oluline hoiatus AI-arendajatele, mis räägib, et transformatiivsus ei paku absoluutset kaitset kommertsliku kasutuse korral, mis konkureerib originaalteoste turul.

USCO avaldas 2025. aastal kõikehõlmava aruande AI ja autoriõiguse kohaldamise kohta,¹²⁵ käsitledes ka küsimust, kas *fair use* doktriin kaitseb AI treenimist piraatlikelt

¹¹⁶ Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 566 (1985): "The fourth factor is undoubtedly the single most important element of *fair use*."

¹¹⁷ *Campbell*, lk 576-579.

¹¹⁸ *Campbell*, lk 578-579.

¹¹⁹ *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202, 214 (2d Cir. 2015).

¹²⁰ *Google LLC v. Oracle America, Inc.*, 593 U.S. 1, 30 (2021).

¹²¹ Lemley, M.A.; Casey, B. Fair Learning. – Texas Law Review, 99, nr 4, 2021, lk 743-784, siin lk 758.

¹²² *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

¹²³ Warhol, lk 532.

¹²⁴ Warhol, lk 545-546.

¹²⁵ USCO. Copyright and Artificial Intelligence. Part 2: Generative AI Training. Washington: U.S. Copyright Office 2025.

soetatud andmetel. USCO võttis seisukoha, et piraatlikelt allikatest pärit andmete teadlik kasutamine peaks kaaluma *fair use*'i vastu, kuigi ei ole iseenesest määrav – see on kaaluv tegur, mida teised *fair use* faktorid (eelkõige transformatiivsus) põhimõtteliselt kompenseerida saavad.¹²⁶ Aruandes märgiti, et "kui treenimine toimub teostega, mis on tehtud kättesaadavaks ilma õiguste omajate nõusolekuta, tekitab selline kasutus selge autoriõigusliku kahju, mis neljandas faktoris peab saama otsustavat kaalu".¹²⁷ USCO soovitas Kongressil kaaluda eraldi seadusandlikku selgitust AI treenimise *fair use* staatuse osas, et vältida õiguslikku ebakindlust.

Neljas faktor – kasutuse mõju turule ja originaalse teose väärtusele – on Kõrgeima Kohtu poolt nimetatud "kõige olulisemaks" *fair use* faktoriks.¹²⁸ AI treenimise kontekstis kätkeb see faktor kahte eraldiseisvat küsimust: esiteks, kas treenimine asendab originaalteoste ostmist (otsene turumõju); teiseks, kas treenimine kahjustab olemasolevat või tekkivat litsentsimisturgu (kaudne turumõju).¹²⁹ Piraatlikest allikatest pärit andmete kasutamine on otseselt kahjulik litsentsimisturule. Nimelt, suured kirjastajad ja autorite organisatsioonid on hakanud looma AI-treenimise litsentsimise programme, mis näitab, et see turg on realiseerumisele üle minemas; piraatandmete kasutamine nullib selle turu, kuna AI-arendajad saavad teoste statistilist väärtust ilma nõuetekohast litsentsitasu maksmata.¹³⁰ Kokkuvõtlikult – piraatandmetel treenimine ei läbi *fair use* testi neljanda faktori kontekstis, kuna see kahjustab otseselt tekkivat turgu, samas kui muud faktorid (transformatiivsus) ei ole iseenesest piisavad selle kahjuliku mõju kompenseerimiseks.¹³¹

3.2. *Bartz v. Anthropic*: lahutamise doktriin

Autorite hagide seerias Anthropic PBC vastu Põhja-California föderaalkohtus (N.D. Cal.) kujunes 2025. aastal eriti märgiliseks asi *Bartz vs Anthropic PBC* (nr 3:24-cv-05417-WHA (Alsup), edaspidi: *Bartz vs Anthropic*).¹³² Nimetatud asi koondas rea autorite nõudeid seoses Claude'i keelimudelite treenimiseks piraatlikest allikatest – sealhulgas Library Genesisist (LibGen) ja Pirate Library Mirror'ist (PiLiMi) – kogutud

¹²⁶ USCO 2025, lk 52.

¹²⁷ USCO 2025 (viide 133), lk 54-55.

¹²⁸ *Harper & Row*, lk 566.

¹²⁹ 17 U.S.C. § 107 - neljas faktor: "the effect of the use upon the potential market for or value of the copyrighted work".

¹³⁰ Henderson, P. jt. Foundation Models and Fair Use. – *Journal of Machine Learning Research*, 24, nr 1, 2023, lk 1-50, siin lk 35-36.

¹³¹ Sag, M. Copyright Safety for Generative AI. – *Houston Law Review*, 61, nr 2, 2024, lk 295-350, siin lk 320.

¹³² *Bartz v. Anthropic PBC*, No. 3:24-cv-05417-WHA (Alsup). (*Bartz vs Anthropic*)

andmete kasutamisega ning sai kaks olulise arengut: esmalt tegi kohus *summary judgment*¹³³ otsuse, mis kehtestas lahutamise doktriini (ingl *separation doctrine*); seejärel lõpetati asi 2025. aastal poolte kokkuleppega.

Kohtuasjas *Bartz v. Anthropic PBC* (nr 3:24-cv-05417-WHA, N.D. Cal.) väitsid hagejad – autorite grupp – et Anthropic laadis Claude'i keelimudelite treenimiseks alla 482 460 teost piraatlikest andmebaasidest, sh Library Genesisist (LibGen) ja Pirate Library Mirror'ist (PiLiMi), esitades nõuded nii allalaadimise kui ka treenimistoimingu osas.¹³⁴ Kohtunik Alsup langetas 23. juunil 2025 *summary judgment* otsuse, kehtestades lahutamise doktriini; asi lõpetati poolte vahelise kokkuleppega 2025. aasta augustis. Doktriini sisuline analüüs järgneb allpool.

Oma analüüsis eristas kohtunik Alsup kaks eraldiseisvat autoriõiguslikku sündmust. Esimene on allalaadimine ise – see on iseseisev reprodutseerimisakt 17 U.S.C. § 106 lg 1 tähenduses.¹³⁵ Teine on treenimistoiming – eraldiseisev kasutusakt, mille *fair use* hinnang võib erineda.¹³⁶ Kohus leidis, et need kaks sündmust ei ole autoriõiguslikult kõrvuti asuvad, vaid selgelt eraldiseisvad. Oluline reservatsioon on siin see, et kuna asi lõpetati poolte kokkuleppega enne apellatsioonimenetlust, ei kujune see analüüs siduvaks pretsedendiks. Tegemist on käesoleva töö normatiivse eelistusega, mitte kehtiva USA õigusega.

Esimese sündmuse (allalaadimise) osas leidis kohus, et see ei läbi *fair use* testi, sest piraatlikelt allikatelt allalaadimise eesmärk on teose saamine ilma litsentsita, mis on otseses vastuolus autoriõiguse eesmärkidega.¹³⁷ Kohus rakendas Warholi standardit – küsimus on, kas alla laadimisel on erinev eesmärk või iseloom originaalteosel. Allalaadimine kommertsliku mudeli treenimiseks ei lisa originaalteosele midagi uut – see on lihtsalt ebaseaduslik koopeerimisakt, mille eesmärk on säästa litsentsitasudelt. Neljas faktor – turu mõju – on samuti selgelt kahjulik, nimelt kätkeb piraatlikelt allikatelt allalaadimine teose omandamist ilma maksmata, mis vähendab autorite tulu otseselt.¹³⁸

Teise sündmuse – seaduslikult soetatud koopiade kasutamine treenimiseks – osas jõudis kohus aga erinevale järeldusele.¹³⁹ Kohus leidis, et seaduslikult soetatud koopiade (nt raamatu

¹³³ *Summary judgment* tähendab, et kohus langetab otsuse juba tõendite esitamise järel, ilma täieliku kohtuistungita, kui faktiline vaidlus sisulistes küsimustes puudub – st üks pool võidab "paberite pealt". Vt *Federal Rules of Civil Procedure, Rule 56(a)* (USA föderaalkohtute tsiviilmenetluse reeglite kogumik).

¹³⁴ *Bartz v. Anthropic PBC*, nr 3:24-cv-05417-WHA (N.D. Cal., 23. juuni 2025), kohtunik Alsup, lk 4-7. Hagejate grupp koondas rea autorite nõudeid; kokkuleppe üksikasjad ei ole täies mahus avalikustatud.

¹³⁵ 17 U.S.C. § 106 lg 1 - reprodutseerimise ainuõigus.

¹³⁶ *Bartz vs Anthropic*, otsus, lk 12-14.

¹³⁷ *Ibid.*, lk 18-19.

¹³⁸ *Ibid.*, lk 19-20.

¹³⁹ *Ibid.*, lk 22-23.

poest ostetud, legaalselt omandatud andmebaasi) kasutamine treenimiseks võib teatud tingimustel olla *fair use* kaitsega kaetud. Transformatiivsuse analüüsis leidis kohus, et mudeli treenimine on transformatiivne kasutus. Nimelt leidis ta, et originaalse teose funktsioon on inimestele lugemiseks mõeldud kirjanduslik teos; mudeli treenimisel kasutatav funktsioon on statistiliste mustrite väljaselgitamine keeles ja mõistetesüsteemides.¹⁴⁰ Neljas faktor – seaduslikult soetatud koopiate puhul – ei kahjusta litsentsimisturgu samal määral, kuna ostmine on juba toimunud ja litsentsitasu on makstud. Kohus ei eitanud, et kõige selle analüüsi lõplik lahendus jääb õhju, kuid kehtestas selge põhimõtte lahutamise doktriini näol.

Asi lõpetati poolte vahel saavutatud kokkuleppega 2025. aasta augustis.¹⁴¹ Kokkuleppe üksikasju ei ole täies mahus avalikustatud, kuid teateid avaldanud allikate kohaselt nägid kokkuleppe tingimused ette olulist hüvitiste maksmist autoritele ning rangema protokolliga kehtestamist tulevaste treeningandmete soetamise osas. Ehkki Alsup'i *summary judgment* otsus on sisuliselt põhjalik, jäi see mittesiduvaks pretsedendiks, kuna asi lõpetati poolte kokkuleppega enne apellatsioonimenetlust. Doktriini analüütiline väärtus ei muutu — kuid seda ei saa nimetada kehtivaks referentspunktiks USA kohtupraktikas.

Alsup'i analüüs on autoriõiguslikult loogiline. Kui rikkumine toimub juba allalaadimise hetkel, ei saa hilisem transformatiivne treenimine seda saneerida. See lähenemine on kooskõlas Warholi standardiga, mis suunab küsimuse konkreetsele toimingule, mitte abstraktsele protsessile. Siiski on oluline pidada silmas, et see on käesoleva töö normatiivne eelistus – kehtiv USA kohtupraktika liigub teises suunas. Ainuke lõplik otsus (Kadrey, Chhabria J.) läks AI-arendaja kasuks. Lahutamise doktriini praktilised tagajärjed sõltuvad apellatsioonikohtu kinnitusest. Kui apellatsioonikohus selle kinnitab, tähendaks see praktikas, et AI-arendajad peavad treeningandmete päritolu aktiivselt kontrollima – teadmatuse ei ole piisav vabandus. See nõue ühtiks EL-i TDM-raamistiku loogikaga, kus opt-out eiramine muudab kasutuse ebaseaduslikuks sõltumata treenimise eesmärgist.

3.3. *Kadrey v. Meta Platforms*: ühtne hindamine ja turu kahjustamise teooria

Kohtuasjas *Kadrey v. Meta Platforms, Inc.* (nr 3:23-cv-03417-VC, N.D. Cal.) esitasid kirjanikud Richard Kadrey, Sarah Silverman ja Christopher Golden hagi Meta vastu, väites et

¹⁴⁰ *Ibid.*, lk 23-24.

¹⁴¹ *Bartz vs Anthropic*, kokkuleppe teadaanne (august 2025). Kokkuleppe üksikasju ei ole täies mahus avalikustatud.

Meta Llama keelimudelite treenimiseks kasutati BitTorrenti teel allalaaditud raamatuid, sh LibGeni andmestikku.¹⁴²

Kohtunik Chhabria lükkas 2023. aasta novembris esialgsed kaebused tagasi, kuna need ei vastanud *plausibility* standardi nõuetele¹⁴³ – st hagejad ei esitanud piisavalt konkreetseid faktilisi väiteid, mis muudaksid nõude usutavaks, vaid toetusid üldistele järeldustele. Hagejatele anti luba esitada täiendav kaebus, milles tuli üksikasjalikumalt kirjeldada väidetavat turukahju. Ümberesitatud kaebusele järgnes 25. juunil 2025 *summary judgment* (kokkuvõttev otsus eel-arutamisel; *Fed. R. Civ. P. 56*) Metale. Kohus leidis Meta kasutuse kõrgelt transformatiivseks ja leidis, et hagejad ei esitanud piisavalt empiirilist tõendusmaterjali turu kahjustuse kohta vaatamata laiendatud võimalusele. Otsus ei ole Metale absoluutne lubakiri – Chhabria hoiatas sõnaselgelt, et litsentsimisturu lahjendamise (*market dilution*) argumentidega oleks võinud tulemus olla teistsugune. Kohtu analüüsi sisuline käsitlemine järgneb allpool.

Ümberesitatud kaebusel (ja sellele järgneval lõppotsuses) märkis kohus märgatavalt detailsemat analüüsi, käsitledes kõiki nelja faktorit eraldiseisvalt.¹⁴⁴ Esimese faktori – transformatiivsuse – analüüsis tunnistas kohus, et treenimine on teataval määral transformatiivne, kuna see loob uut funktsionaalsust (keelimudelit), mis erineb originaalsete teoste funktsioonist (kirjandusliku naudingu pakkumine).¹⁴⁵ Kohus leidis aga, et transformatiivsuse argument ei ole piiritu kaitse vahend: jälgides Warholi standardit, märkis kohus, et esimese faktori analüüsis peab hindama kõnealuse kasutuse konkreetset eesmärki – mitte abstraktset transformatiivsust – ning seda, kas see eesmärk konkureerib originaalse teose eesmärgiga. Kui Meta Llama suudab genereerida kõrgetasemelist ilukirjandust ja seega mõnel määral asendab originaalteoste lugemise tarbimist, kahaneb transformatiivsuse argument.¹⁴⁶

Kolmas faktor – kasutatud osa kogus ja olulisus – leidis käsitlust lühidalt: kohus leidis, et kõigi koopiaste kasutamine treenimiseks ei ole iseenesest *fair use*'i välistav, sest statistilise mudeli treenimine eeldab kõikehõlmavaid andmeid.¹⁴⁷ See on kooskõlas autoriõiguse

¹⁴² *Kadrey v. Meta Platforms, Inc.*, nr 3:23-cv-03417-VC (N.D. Cal., 25. juuni 2025), kohtunik Chhabria, lk 3-5 ja 8-12. Esialgsed kaebused lükati tagasi 2023. aasta novembris puuduliku tõendatuse tõttu; ümberesitatud kaebus viis 2025. aasta *summary judgment*-ile.

¹⁴³ *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 (2007); *Ashcroft v. Iqbal*, 556 U.S. 662 (2009). Standard nõuab, et kaebus sisaldaks "*enough facts to state a claim to relief that is plausible on its face*" ehk pelgalt teoreetiline võimalikkus ei ole piisav.

¹⁴⁴ *Kadrey vs Meta*, kohtu lõppotsus (N.D. Cal. juuni 2025).

¹⁴⁵ *Kadrey vs Meta*, lõppotsus, lk 18.

¹⁴⁶ *Kadrey vs Meta*, lõppotsus, lk 22-23.

¹⁴⁷ *Ibid.*, lk 28.

üldpõhimõttega, et kolmas faktor kätkeb eelkõige küsimust, kas võetud osa on "süda" originaalteostest – mis stiililiste teoste treenimise korral ei ole eriti relevantne, kuna mudel ei reprodutseeri konkreetset stiili, vaid õpib statistilisi mustreid.¹⁴⁸

Neljanda faktori – turu kahjustamise – analüüs kujunes asja juriidiliselt keerukaimaks küsimuseks.¹⁴⁹ Hagejad väitsid, et Meta Llama'l põhinevad rakendused suudavad genereerida ilukirjanduslikku sisu, mis konkureerib hagejate teoste omadega turul – see tähendaks hagejate tulevaste müügitulude ja litsentsimisvõimaluste kahjustamist. Kohus mõõns, et see argument on teoreetiliselt kaalukas, kuid leidis, et hagejad ei esitanud piisavat empiirilist tõendusmaterjali selle kohta, kuidas konkreetset Meta Llama'l põhinevad rakendused kahjustavad nende konkreetseid teoseid konkreetset turusegmendil.¹⁵⁰ Kohus rõhutas, et spekulatiivne tulevane turukahjustamine ei ole piisav – hagejad peavad näitama konkreetseid turu-andmeid, millest nähtub, et AI-genereeritud sisu on hakanud asendama nende teoste ostu või kahjustama nende litsentsimisvõimalusi.¹⁵¹

Kadrey vs Meta turu kahjustamise teooria käsitlemine on tähelepandav kõrvutuses USCO 2025. aasta aruandega, milles leiti, et piraatlikest allikatest pärit andmete kasutamine kahjustab otseselt tekkivat AI-litsentsimisturgu.¹⁵² USCO positsioon on seega jõulisem kui kohtu nõuded hagejatele turu kahjustamise tõendamisel: kui USCO hindab kahju turul lähtuvalt normatiivsest, siis kohus nõuab empiirilist tõendust konkreetsest turukahjustamisest. Sag on märkinud, et see eristus on oluline, sest kohtute kõrge tõendamisstandard turu kahjustamise osas võib de facto piirata autoriõiguse kaitset tõhusalt, kuna empiiriline tõendamine AI mõjude kohta turul on metodoloogiliselt keerulises seisus.¹⁵³

Kadrey vs Meta lõppotsus (25.06.2025) andis Metale täieliku võidu *fair use*'i alusel: kohtunik Chhabria leidis kogu treenimisprotsessi – sh LibGeni piraatandmete kasutamise – "kõrgelt transformatiivseks", kuna keelimudelite loomine erineb originaalteoste lugemiseesmärgist põhimõtteliselt. Erinevalt kohtunik Alsup'ist (ptk 3.2) keeldus Chhabria allalaadimist ja treenimist eraldamast, hinnates neid ühe protsessina. Neljas faktor (turu kahjustamine) toetas samuti Metat, kuid kohtu arutlus oli sisuliselt keerukam.

Kohus käsitles *obiter dictum* korras turu lahjendumise (*market dilution*) teooriat. Selle kohaselt kahjustab AI-süsteemide poolt suurel mahul konkureeriva sisu genereerimine teoste

¹⁴⁸ *Harper & Row*, lk 564-565; Lemley; Casey (viide 129), lk 762.

¹⁴⁹ *Kadrey vs Meta*, lõppotsus, lk 33.

¹⁵⁰ *Ibid.*, lk 33-35.

¹⁵¹ Sag, Copyright Safety, lk 322-325.

¹⁵² USCO 2025, lk 55.

¹⁵³ Sag, Copyright Safety, lk 326-328.

turgu. Chhabria lisas, et "*enamikul juhtudel on tõenäoliselt vastus jah*", et selline treenimine kujutaks sellele teooriale tuginedes rikkumist. Siinkohal tuleb teha oluline reservatsioon. See kommentaar on *obiter dictum*, ehk kohtu mittesiduv märkus, mis ei moodusta otsuse resolutsiooni ega ole pretsedentne. Chhabria andis hagejatele sisuliselt vihje, millisel teorial oleks olnud tugevam alus, kuid ta ei teinud selle teooria alusel iseseisvat hinnangut. Hagejad kaotasid, sest nad ei arendanud turu lahjendumise argumenti välja. Nad tuginesid hoopis substituutsioonist tulenevale turu kahjule, mida nad ei suutnud empiirilisel tõendada. Kohus lisas aga sõnaselge hoiatuse: "*This decision does not stand for the proposition that Meta's use is lawful. It stands only for the proposition that these plaintiffs made the wrong arguments and failed to develop a record in support of the right one.*" (eesti k.: "*See otsus ei kujuta endast seisukohta, et Meta tegevus on seaduspärane. See seisab üksnes selle eest, et käesolevad hagejad esitasid valesid argumente ja ei kujundanud välja tõendusbaasi õigetele argumentidel.*"; autori tõlge) – rõhutades, et otsus ei anna AI-ettevõtetele üldist seaduspärasest tegutsemisluba piraatandmetel treenimiseks. See reservatsioon on kaheti oluline: kohus andis Metale faktispetsiifilise võidu, mitte üldise doktriinilise kinnituse – ning viitas samas selgelt, millisel teorial (turul lahjendumise argument) oleks hagejatel olnud õiguslik alus.¹⁵⁴

Kadrey v. Meta summary judgment Metale *fair use*'i alusel (vaatamata LibGeni kasutamisele) on käesoleva töö hüpoteesi jaoks otsene väljakutse, mida tuleb ausalt tunnistada: 2025. aasta seisuga on ainus lõplik USA ringkonnakohtu otsus piraatandmetel treenimise *fair use* küsimuses läinud AI-arendaja (Meta) kasuks. See tähendab, et USA õiguse tegelik olukord (*de facto*) on ühtse hindamise lähenemise tõttu praegu AI-ettevõtetele soodsam kui töö argumentatsioon esmakordsel lugemisel võib viidata. EL-i süsteem on normatiivsel tasandil selgem, kuna seadusliku juurdepääsu nõue toimib eeltingimusena, mille puudumisel erandi kohaldamine langeb ära, kuid USA praktiline olukord on hetkel erinev. Käesoleva töö hüpoteesis on seega täpsem öelda, – EL-i poolne osa leidis kinnitust; USA-poolne osa on doktrinaalses murrangupunktis, kus kehtiv pretsedent (*Kadrey*) toetab AI-arendajaid, kuid see pretsedent on teoreetiliselt vaieldav (lahutamise doktriin) ning ebastabiilne kuni apellatsioonikohtude lahendini.

¹⁵⁴ Warhol, lk 532-533.

3.4. Lahutamise doktriini ja ühtse hindamise lähenemise kriitiline võrdlus: kumb on doktrinaalselt põhjendatum?

Kohtuasjade *Bartz v. Anthropic* ja *Kadrey v. Meta* analüüs paljastab USA kohtupraktikas fundamentaalse metoodilise vastuolu küsimuses, kas piraatlikelt allikatelt andmete allalaadimine ja nende kasutamine treenimiseks tuleb hinnata eraldiseisvate autoriõiguslike sündmustena või ühe tervikliku protsessina. Kohtunik Alsup'i lahutamise doktriin ja kohtunik Chhabria ühtsekindla hindamise lähenemine jõudsid sisuliselt identsetel faktilistel alustel vastupidistele tulemustele, mis nõuab siinkohal kriitilist hinnangut küsimusele, kumb neist on doktrinaalselt tugevamini põhjendatav.

Lahutamise doktriini kasuks räägib esmalt selle kooskõla Kõrgeima Kohtu Warholi (598 U.S. 508, 2023) standardiga, mille kohaselt *fair use*'i esimese faktori hindamisel peab küsimus olema suunatud konkreetsele kasutusele, mitte abstraktssele protsessile tervikuna. Kui "kasutus" Warholi tähenduses on defineeritud konkreetse toiminguna, siis piraatlikult allikalt allalaadimine kujutab endast üht toimingut ja treenimistoimingud teist – ning nende eraldiseisev hindamine on Warholi testi loomulik rakendus. Ühtsekindla hindamise lähenemine, mille kohaselt allalaadimine on transformatiivse treenimisprotsessi lahutamatu osa, riskib transformatiivsuse mõiste lahjendamisega ulatuses, mis murendaks Warholi otsuse piirava eesmärgi: takistada abstraktssele "uuele kontekstile" tuginemist kaubandusliku kasutuse õigustamiseks.

Teiseks on lahutamise doktriin kooskõlas autoriõiguse ennetava kaitsefunktsiooniga. Kui hilisem transformatiivne kasutamine saab tagasiulatuvalt heastada varasema ebaseadusliku allalaadimise, tekib moraalne hasard: AI-ettevõtetal puudub stiimul andmete päritolu kontrollida, kuna transformatiivne treenimine on alati kättesaadav kaitsena. Seda riski on tunnistanud ka USCO 2025. aasta aruanne, leides, et treeningandmete päritolul on iseseisvat tähendust *fair use*'i hindamisel sõltumata sellele järgnevate protsesside transformatiivsusest.

Ühtsekindla hindamise lähenemise kasuks räägib seevastu praktiline argument – kui digitaalse allalaadimise iga akt tuleb hinnata eraldiseisvalt, tekivad AI-arendajatele ulatuslikud ja raskesti prognoositavad vastutusriskid kõigi tehniliste vaheetappide eest, mis kaasnevad andmete kogumisel. Chhabria lähenemise kohaselt piisab sellest, et treenimisprotsess tervikuna on transformatiivne ning et turukahju on ebapiisavalt tõendatud, mitte ei pea iga allalaadimisakt eraldiseisvalt läbima *fair use*'i analüüsi. See on

administratiivselt lihtsam ja ettevõtjasõbralikum, kuid ohverdab õiguskaitseks vajaliku ennetavuse.

Kaaludes mõlema lähenemise argumente, on käesoleva töö hinnangul lahutamise doktriin doktrinaalselt veenvam, arvestades selle kooskõla Warholi pretsedendi, ennetava kaitsefunktsiooni ning USCO 2025. aastaaruande seisukohaga. Samas on oluline arvestada, et *Kadrey v. Meta* (Chhabria J., 25. juuni 2025) on 2025. aasta seisuga ainus lõplik USA ringkonnakohtu otsus AI treenimise *fair use* küsimuses – ning see otsus kohaldas ühtsekindlat hindamist ja läks AI-arendaja kasuks. Lahutamise doktriin (Bartz, Alsup J.) on analüütiliselt atraktiivne ja töö *de lege ferenda* argumentatsiooni toetav, kuid ei kujuta hetkel kehtivat operatiivset USA õigust. Seega ei saa magistritöö väita, et lahutamise doktriin "tõenäoliselt kinnitatakse" apellatsioonikohtus — tegemist on analüütilise eelistusega, mitte empiirilise prognoosiga. Selle asemel on täpsem öelda: USA raamistik on doktrinaalses murrangupunktis, mis on iseenesest peamine argument EL-i struktuurse lähenemise üleoleku kasuks – süsteem, kus kaks ringkonnakohtu jõuavad identsetes olukordades vastupidistele tulemustele, ei paku piisavat õiguskindlust.

Samas väärrib eraldi tähelepanu, et *Kadrey v. Meta Platforms* lõppotsus (nr 3:23-cv-03417-VC, N.D. Cal., 25. juuni 2025) ei kujuta endast üldist piraatandmetel treenimise seaduspärasuse kinnitust. Kohtunik Chhabria andis küll Metale *fair use*'i alusel summary judgment, leides treenimisprotsessi „kõrgelt transformatiivseks”, kuid lisas sõnaselge reservatsiooni, et otsus seisab üksnes selle eest, et hagejad esitasid valesid argumente ega kujundanud välja tõendusbaasi õigetel argumentidel, viidates, et turu lahjendumise alusel esitatud hagil oleks olnud suurem edu tõenäosus. Käesoleva töö hinnangul on see tulemus Warholi standardi, moraalse hasardi ja USCO 2025. aasta aruande vastuseisu tõttu doktrinaalselt vaieldav, kuid *Kadrey v. Meta* kinnitab, et USA õigus on hetkel doktrinaalses murrangupunktis — ning see ebakindlus annab iseenesest lisaargumendi EL-i struktuurse lähenemise üleoleku ja käesoleva töö *de lege ferenda* ettepanekute kasuks.

3.5 Thomson Reuters v. ROSS Intelligence: Warholi standard ja konkureeriv kasutus

USA senini kõige lõplikuma ja selgema kohtuotsuse AI treenimise *fair use* küsimuses tegi Delaware föderaalringkonnakohtus (D. Del.) 11. veebruaril 2025 asjas *Thomson Reuters Enter. Ctr. GmbH vs ROSS Intelligence Inc.* (nr 1:20-cv-00613-SB, edaspidi: *Thomson Reuters vs ROSS*).¹⁵⁵ Ehkki asi käsitleb konkreetset juriidilise teabebaasi Westlaw

¹⁵⁵ *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc.*, No.1:20-cv-00613-SB, otsus (D. Del. 11.02.2025). (*Thomson Reuters vs ROSS*)

pealkirjaparagrahvide kasutamist AI otsinguriista treenimiseks – mitte üldist generatiivset AI-d – pakub see otsus kõige põhjalikuma USA kohtupraktikast tuleneva raamistiku *fair use* analüüsiks AI treenimise kontekstis.

Thomson Reuters esitas 2020. aastal hagi ROSS Intelligence Inc. vastu, väites et ROSS kasutas Westlaw juriidilise andmebaasi pealkirjaparagrahve oma tehisintellektil põhineva juriidilise otsinguteenuse treenimiseks pärast seda, kui Thomson Reuters keeldus ROSS'ile litsentsi andmast.¹⁵⁶

Delaware föderaalringkonnakohus (D. Del., nr 1:20-cv-00613-SB, 11. veebruar 2025) leidis, et Westlaw pealkirjaparagrahvid on autoriõigusega kaitstud teosed ning ROSS'i kasutus ei ole *fair use*, rakendades Kõrgeima Kohtu Warholi standardit. *Fair use*'i sisuline analüüs järgneb allpool.

Oluline täpsustus käesoleva töö kontekstis (lisatav lõpuosas): ROSS ei olnud generatiivne AI klassikalises mõttes – tegemist oli otsingumootori tüüpi süsteemiga, mis ei genereerinud uut ekspressiivset sisu, vaid vahendas olemasolevaid juriidilisi andmeid kasutajapäringutele vastates. Seetõttu kohaldub Warholi test eriti otsekoheselt, kuna kasutuse eesmärk ühtis originaalmaterjali eesmärgiga täielikult. Generatiivse AI juhtumitele tuleb ROSS'i pretsedenti üle kanda ettevaatlikult, nagu käesolev töö allpool täpsustab.

Fair use analüüsis käsitas kohus kõiki nelja faktorit, rakendades selgelt Kõrgeima Kohtu Warholi 2023. aasta standardit.¹⁵⁷ Esimese faktori – kasutuse eesmärk ja iseloom – analüüsis leidis kohus, et ROSS'i kasutus ei olnud transformatiivne Warholi-järgse tõlgenduse kohaselt.¹⁵⁸ Kohus analüüsis ROSS'i konkreetset eesmärki: luua AI otsinguriist, mis konkureerib Westlaw'iga juriidilise teabeotsingu turul. Kuivõrd Westlaw pealkirjaparagrahve kasutatigi eelkõige selle tüüpi teabeotsingul, täitis ROSS'i kasutus täpselt sama kaubanduslikku eesmärki mis originaalteos. Warholi testi rakendades jõudis kohus järeldusele: kuna ROSS'i ja Westlaw eesmärgid on identsed – juriidiline teabeotsing –, puudub igasugune transformatiivse kasutuse element.¹⁵⁹

Teise faktori – teose iseloom – osas leidis kohus, et pealkirjaparagrahvid kätkevad loomingulisi elemente, mis pooldab Thomson Reuters'i positsiooni. Kolmas faktor –

¹⁵⁶ Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence, Inc., nr 1:20-cv-00613-SB (D. Del., 11. veebruar 2025), kohtunik Bibas, lk 3-8. Kuna Thomson Reuters keeldus ROSS'ile Westlaw litsentsi andmast, pöördus ROSS ettevõtte LegalEase poole, kes kogus Westlaw pealkirjaparagrahve. ROSS on vaidlustanud otsuse Kolmandas Apellatsiooniringkonnas (3d Cir., nr 25-8018, 17.06.2025).

¹⁵⁷ Warhol, lk 532 - "the first factor asks whether and to what extent the use in question has a further purpose or different character."

¹⁵⁸ *Thomson Reuters vs ROSS*, lk 21.

¹⁵⁹ *Ibid.*, lk 21-22.

kasutatud osa kogus – pooldas kõige rohkem ROSS'i, kuna ROSS kasutas vaid osa Westlaw pealkirjaparagrahvidest (mitte tervet andmebaasi).¹⁶⁰ Kuid neljas faktor – turu kahjustamine – osutus kaalukaimaks elemendiks, mis kinnitas Warholi kohaldamist kõige selgemalt.¹⁶¹ Kohus leidis, et ROSS'i kasutus kahjustas otseselt Thomson Reuters'i Westlaw'i litsentsimisturgu: ROSS kasutas pealkirjaparagrahve just selleks, et luua konkureeriv toode, mis muudab Westlaw tellimuste vajaduse osaliselt üleliigseks juristide jaoks, kes vastasel korral oleksid kasutanud Westlaw'i andmebaasi. See on klassikaline turu kahjustamise näide – kasutatavad teosed ei täiendanud originaalset toodet erinevatel turgudel, vaid asendavad seda samal turul.¹⁶²

ROSS teatas otsuse vaidlustamisest Kolmandas Apellatsiooniringkonnas (3d Cir., nr 25-8018, 17.06.2025).¹⁶³ See apellatsioonimenetlus (nr 25-8018) on eraldiseisev varasemast apellatsioonikaebusest (nr 25-2153), mis käsitleb ringkonnakohtu varasemat lahendit ning mille otsus on oodata 2026. aasta lõpuks. Apellatsioonikohtu otsus on oodatavalt üks olulisimaid AI autoriõiguse lahendeid lähiajal USA-s, sest see oleks esimene apellatsioonikohtu taseme otsus, mis käsitleb AI treenimise *fair use* küsimust. *Thomson Reuters vs ROSS* eristub *Bartz vs Anthropic* ja *Kadrey vs Meta* asjadest põhjalikult: ROSS'i kasutuse eesmärk oli selgelt ja otseselt konfigureeritud nii, et see konkureeris pealkirjaparagrahvide algse kasutuse eesmärgiga – see muutis Warholi testi kohaldamise loogiliselt obligatoorseks.

Thomson Reuters vs ROSS'i otsuse õigusdogmaatiline tähtsus seisneb selles, et see annab kohtutele kasutuskõlbliku raamistiku generatiivse AI juhtumiteks: kui AI-mudeli lõppeesmärk on luua toode, mis konkureerib originaalteoste või originaalteoste alusel loodud teenustega samal turusegmendil, ei saa *fair use* kaitset rakendada. See raamistik võib koos Warholi standardiga laieneda generatiivse AI üldisematele kasutusjuhtumitele: näiteks ChatGPT kasutamine konkureerib ajalehe vööi entsüklopeedia lugemisega samal turul – mis tähendab, et *fair use* argument jääb nõrgaks.¹⁶⁴

Kõnealuse pretsedendi ülekandmisel generatiivse AI kontekstile tuleb aga arvestada üht olulist eristust: *Thomson Reuters v. ROSS* ei käsitlenud generatiivset AI-d klassikalises mõttes. ROSS oli juriidiline otsinguriist, mis ei loonud uut sisu, vaid valis ja esitas juba olemasolevaid juriidilisi andmeid. Tema eesmärk oli otseselt ja täielikult identne

¹⁶⁰ *Ibid.*, lk 28.

¹⁶¹ *Ibid.*, lk 34-35.

¹⁶² *Ibid.*, lk 35-37.

¹⁶³ *Thomson Reuters vs ROSS*, interlocutory appeal, 3d Cir., No. 25-8018 (17.06.2025).

¹⁶⁴ Henderson jt, *Foundation Models*, lk 42; Sag, *Copyright Safety*, lk 346-348.

originaalteose (Westlaw pealkirjaparagrahvide) eesmärgiga – mõlemad andsid vastuseid juriidiliste otsinguküsimuste peale.¹⁶⁵ Generatiivsed keelemudelid, nagu GPT, LLaMA ja Claude, genereerivad originaalse tekstiväljundi, mis ei kattu struktuuriliselt originaalteoste otseesmärgiga: nende lõpptööde on uus tekst, mitte originaalteoste agregeeritud esitlemine. See ei tähenda, et *fair use* argument on automaatselt tugev – USCO 2025. aasta aruanne väidab vastupidist –, kuid tähendab, et ROSS-i pretsedenti tuleb üle kanda ettevaatlikult ja mitte mehaaniliselt. ROSS on väärtuslik raamistik, kuid puudub analoogiline struktuuriline identsus selliste generatiivse AI juhtumitega nagu *Bartz v. Anthropic* või *Kadrey v. Meta*.

Käesoleva peatüki analüüs näitab, et USA *fair use* doktriini kohaldamine piraatlikelt allikatelt kogutud treenimisandmete suhtes on doktrinaalselt ebaselge ja kohtupraktikas lahknevalt lahendatud. Lahutamise doktriin (*Bartz vs Anthropic*) sõnastab, et piraatlikelt allikatelt andmete alla laadimine on *per se* autoriõiguse rikkumine, mida *fair use* ei kata – see rikkumine ei ole saneeritav treenimise transformatiivsuse argumentiga. Kohtunik Alsup'i lahend on aga jäänud mittesiduvaks pretsedendiks, kuna asi lõpetati pooltevahelise kokkuleppega 2025. aasta augustis, enne apellatsioonikohtu kontrolli.

Seevastu kohtunik Chhabria kohtuasjas *Kadrey vs Meta* jõudis vastupidisele järeldusele: keeldudes allalaadimist treenimisest lahutamast, hindas ta kogu protsessi ühtsena "kõrgelt transformatiivseks" ning andis Metale *fair use*'i alusel kokkuvõtva otsuse, vaatamata sellele, et tuvastatud faktina oli kindlaks tehtud LibGeni piraatandmete kasutamine. *Kadrey* on 2025. aasta seisuga ainus lõplik ringkonnakohtu otsus AI treenimise *fair use* küsimuses – ja see otsus läks AI-arendaja kasuks. Samas lisas Chhabria otsusesse sõnaselge reservatsioon: "*This decision does not stand for the proposition that Meta's use is lawful. It stands only for the proposition that these plaintiffs made the wrong arguments and failed to develop a record in support of the right one.*" (eesti k.: "See otsus ei kujuta endast seisukohta, et Meta tegevus on seaduspärane. See seisab üksnes selle eest, et käesolevad hagejad esitasid valesid argumente ja ei kujundanud välja tõendusbaasi õigetel argumentidel."); autori tõlge) – mis rõhutab, et kohus andis faktispetsiifilise võidu, mitte üldise doktriinilise kinnituse piraatandmetel treenimise seaduspärasuse kohta, kuid viitab ka sellele, et neljandal faktoril põhinev turu lahjendumise argument oleks olnud teostatav.

Thomson Reuters vs ROSS on ajalooliselt esimene USA kohtuotsus, mis lükkas *fair use* argumenti AI treenimise kontekstis selgelt tagasi, rakendades Warholi standardit:

¹⁶⁵ *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc.*, nr 1:20-cv-00613-SB (D. Del., 11. veebruar 2025), kohtunik Bibas: ROSS oli juriidiline otsinguriist, mis ei genereerinud uut sisu.

konkureeriv turupositsioon elimineerib transformatiivsuse argumendi. *ROSS* eristub aga *Bartz'i ja Kadrey* asjadest struktuuriliselt – tegemist ei olnud piraatlikelt allikatelt pärit andmetega, vaid *Thomson Reutersi* otsese litsentsimiskeeldumisega, mistõttu *ROSS*-i pretsedendi ülekannet generatiivse AI piraatandmete konteksti tuleb teha ettevaatlikult.

Koos moodustavad need arengud pildi doktrinaalsest murrangupunktist, kus kaks ringkonnakohut on jõudnud identsetes olukordades vastupidistele tulemustele. USA kehtiv operatiivne seis on järgmine: piraatandmetel treenimine on *de facto* lubatud *Kadrey*-pretsedendi tõttu, kuid see olukord on ebakindel kuni Kolmanda Apellatsiooni ringkonna otsuseni asjas *Thomson Reuters v. ROSS* (nr 25-2153, oodatav 2026. aasta lõpus), mis saab olema esimene apellatsioonikohtu taseme otsus küsimuses. Seaduslikult soetatud andmetega treenimise küsimus jääb avatud analüüsipunktiks.

3.6. EL ja USA lähenemiste normatiivne võrdlus

EL-i ja USA analüüsi põhjal ilmneb, et kahe süsteemi peamine erinevus seisneb piraatandmete käsitlemise õiguslikus struktuuris. Euroopa Liidu õiguses on "seadusliku juurdepääsu" nõue eeltingimus, mille puudumisel langeb TDM-erandi kohaldatavus täielikult ära. Seevastu USA õiguses ei ole andmete päritolu iseseisev välistav kriteerium, vaid üks kaalutegur *fair use*'i analüüsis.

Sellest tulenevalt pakub EL-i süsteem suuremat õiguskindlust, kuid on samas jäigem, samas kui USA süsteem võimaldab paindlikumat, kuid ettearvamatumat hinnangut. Alapeatükis 1.3 teoreetiliselt tuvastatud struktuurne erinevus leidis käesolevates peatükkides läbi viidud kohtupraktika analüüsi käigus kinnitust.

4. SAASTUNUD MUDELI PROBLEEM JA *DE LEGE FERENDA* ETTEPANEKUD

Käesolev peatükk liigub *de lege lata* analüüsilt normatiivse hinnangu ja ettepanekute suunas. Eesmärk ei ole üksnes tuvastada kehtiva õiguse puudused, vaid pakkuda süsteemseid lahendusi piraatandmetel treenitud mudelitega seotud probleemide lahendamiseks.

Käesolev peatükk käsitleb küsimust, mis eelnevates peatükkides jäi taustaks: mis saab mudeliga, mis on juba treenitud piraatlikel andmetel? Kehtiv regulatiivne raamistik ei anna sellele küsimusele rahuldavat vastust kummagi uuritud süsteemi puhul. Analüüsi põhjal esitatakse *de lege ferenda* ettepanekud, mis on suunatud kolmele tasandile: preventsiiooni tugevdamisele, õigusliku ebakindluse kõrvaldamisele ning tõhusate heastamisvahendite loomisele.

4.1. Saastunud mudeli probleem: algoritmiline eraldamine

"Saastunud mudel" (ingl *contaminated model* või *tainted model*) tähistab käesolevas töös analüütilise tööriistana generatiivset tehisintellektimudelit, mis on treenitud andmekogumil, mis sisaldab osaliselt või täielikult piraatlikest allikatest pärit teoseid. Mõiste on käesoleva töö kontekstis kirjeldava ja normatiivse funktsiooniga: kirjeldav tähendus viitab empiirilisele olukorrale, kus treeningandmestik on koostatud ebaseaduslikest allikatest; normatiivne tähendus viitab sellele, et selline mudel on treenitud autoriõigust rikkudes ning selle olemasolu loob kehtiva autoriõiguse alusel nõudeõiguse. Käesolev töö ei pretendeeri sellega uue juriidilise kategooria loomisele, vaid kasutab mõistet analüütilise vahendina olemasolevate normide kohaldamisel AI-spetsiifilisele olukorrale.

Kui eelnevates peatükkides on tuvastatud, et selline treenimine on EL-i ja USA õiguse kohaselt autoriõiguse rikkumine, tekib järgnev loogiline küsimus: milline on sellest treeningprotsessist lähtuva mudeli õiguslik saatus? Kas mudel ise on täielikult saastunud ning milliseid heastamisvahendeid peaks õigusraamistik pakkuma?

Vastuse leidmiseks on abistavaks Ameerika Ühendriikide Föderaalkaubanduskomisjoni (FTC) välja töötatud algoritmiline eraldamine (ingl *algorithmic disgorgement*), mida kohaldati esimest korda asjades *FTC v. Everalbum, Inc. (2021)*¹⁶⁶ ja *FTC v. WW International, Inc. (2022)*.¹⁶⁷ FTC v. Everalbum asjas oli ettevõtte kogunud kasutajate pilte ilma nende nõusolekuta ning kasutanud neid näotuvastuse algoritmi treenimiseks. FTC

¹⁶⁶ FTC. In the Matter of Everalbum, Inc., FTC Docket No. C-4782, Consent Order (2021).

¹⁶⁷ FTC. In the Matter of WW International, Inc., FTC Docket No. C-4818, Consent Order (2022).

kohustusotsus nõudis, et Everalbum ei kustutaks mitte ainult kasutajate andmeid, vaid hävitaks kõik algoritmid, mis olid treenitud illegaalselt kogutud andmetel.¹⁶⁸ Sama loogika kehtis FTC v. WW International asjas, kus WW International'i kaalukaotuse laste programm Kurbo kogus alla 13-aastaste laste isikuandmeid ilma vanema kontrollitava nõusolekuta, rikkudes COPPA-t (*Children's Online Privacy Protection Act*) — nendel andmetel treenitud mudel tuli samuti hävitada.¹⁶⁹ Algoritmiline eraldamine kui kontseptsioon tähendab seega, et ebaseadusliku andmekasutuse "vili" – treenitud algoritm – tuleb hävitada koos treeningandmetega, mitte üksnes kustutada andmeid.

EL-i kontekstis pakub analoogia korras lähtepunkti GDPR artikkel 17 – kustutamisoigus (ingl *right to erasure*).¹⁷⁰ Ehkki GDPR käsitleb isikuandmeid, on kustutamisoiguse loogika laiem: kui andmete töötlemine on olnud ebaseaduslik, peab selle tulemused – sh töötlemise "vili" – olema kustutamisele kuuluvad.¹⁷¹ GDPR põhjendus 65 täpsustab, et kustutamisoigus kohaldub juhul, kui andmete säilitamine on vastuolus liidu õigusega.¹⁷² Autoriõiguse rikkumise kontekstis võib sellest järeldada, et mudeli parameetrid, mis kodeerivad piraatlikest allikatest pärit teoste sisu, on vaidlustatavad "rikkumise viljana", mis peaks olema kustutatav. Probleem on aga tehniline: erinevalt konkreetsetest isikuandmetest, mida saab andmebaasist sihtotstarbeliselt kustutada, on AI-mudeli parameetrid statistiline abstrakt, kus konkreetse teose sisu ei ole selgelt identifitseeritav.

Algoritmilise eraldamise kohustuse õiguslik teostatavus sõltub otseselt sellest, milliseid tehnilisi tingimusi saab kohus kontrollitavaks kohustuseks sõnastada. Selleks on vajalik eristada kolme kontseptuaalselt erinevat dimensiooni, mida õiguskirjanduses on sageli ühe probleemina käsitletud.

Esimene dimensioon on memoriseerimise tuvastamine: kas konkreetne autoriõigusega kaitstud teos on üldse mudelisse memoriseeritud selliselt, et see on reprodutseeritav. Carlini jt 2023. aasta uuringud näitavad, et memoriseerimine ei ole universaalne – üksnes umbes 1% treenimisandmetest on otseselt reprodutseeritav, kusjuures sagedamini esinenud tekstid on

¹⁶⁸ *FTC vs Everalbum*, Consent Order, punkt II.B: kohustus hävitada kõik algoritmid, mis on treenitud ebaseaduslikult kogutud andmetel.

¹⁶⁹ *FTC vs WW International*, Consent Order, punkt III.C. Vt ka Lathrop, E. Algorithmic Disgorgement: An FTC Enforcement Theory and Its Implications. - North Carolina Journal of Law & Technology, 24, nr 2, 2022, lk 45-78, siin lk 52.

¹⁷⁰ Euroopa Parlamendi ja nõukogu määrus (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus). - ELT L 119, 04.05.2016, lk 1-88 (GDPR).

¹⁷¹ GDPR artikkel 17 lõige 1 - kustutamisoigus (nn "õigus olla unustatud").

¹⁷² GDPR põhjendus 65: "Füüsilisele isikule tuleks anda õigus lasta oma isikuandmed kustutada, kui andmete säilitamine on vastuolus käesoleva määrusega või liidu või liikmesriigi õigusega."

memoriseerimisele märkimisväärselt altimad kui harvem esinenud.¹⁷³ *GEMA* v. *OpenAI* lahend rakendas just seda loogikat: *GEMA* demonstreeris, et GPT-4o reprodutseerib vaidlusesemeid laulusõnu vastusena lihtsatele päringutele. Seega on esimene dimensioon – tuvastamine – tehniliselt teostatav.

Teine dimensioon on sihtotstarbelise kustutamise ehk *machine unlearning*'i¹⁷⁴ teostatavus: kas tuvastatud memoriseeritud teos on eemaldatav mudelist ilma selle terviklikku ümberkoolitamist nõudmata. Siin on olukord märgatavalt keerulisem. Praeguses tehnikaarengus ei ole tõendatult usaldusväärsed meetodeid, mis võimaldaksid konkreetse autori kaitstud teose mõju täielikku eemaldamist mudeli parameetritest selliselt, et seda saaks kohtus kontrollitaval viisil verifitseerida.

Machine unlearning'u valdkond on pärast 2022. aastat kiiresti arenenud. Varasemad klassikalised käsitlused (Cao jt 2015) on tänaseks olulises osas täiendatud uuemate meetoditega – sealhulgas *gradient-based forgetting*'u ja model *editing*'u tehnikatega ning nn selektiivse kustutamise lähenemistega (Nguyen jt 2022; Eldan jt 2023). Nende meetodite abil on teoreetiliselt võimalik vähendada konkreetse andmeelemendi mõju mudeli parameetrites.

Siiski on valdkonna peamine probleem käesoleva töö jaoks tõenduslik, mitte ainult tehniline. Carlini jt (2023) memoriseerimise tuvastamise metodoloogia toimib ühes suunas: saab tuvastada, kas mudel reprodutseerib konkreetset teksti. Kuid see ei tõenda, et mõju on täielikult eemaldatud – ainult seda, et konkreetne päring ei anna otsest reproduktsiooni. Kohtus kontrollitaval viisil on "kustutamise" lõpetatuse tõendamine seega ikka veel lahendamata probleem. Seetõttu ei ole *machine unlearning*'u kasutamine praegu rakendatav kohtukohustusena – puudub verifitseerimisstandard, millele kohus saaks tugineda. Cao jt varasemates uuringutes ning hiljutistes *machine unlearning*-käsitlustes on näidatud, et küll on olemas meetodeid, mis "vähendavad" mõne andmeelemendi mõju mudelis, kuid need ei ole piisavalt täpsed konkreetse autori loominguhulgal.¹⁷⁵ Seega ei saa kohus praegu kohustada AI-ettevõtet "kustutama" konkreetset teost mudelis viisil, mida oleks võimalik usaldusväärselt kontrollida.

Kolmas dimensioon on ümberkoolitamine alternatiivina: kas kogu mudeli ümberkoolitamine ilma piraatlikke andmeid sisaldava andmekogumita on teostatav ja

¹⁷³ Carlini, N. jt. Quantifying Memorization Across Neural Language Models. - 11th International Conference on Learning Representations (ICLR) 2023, lk 1-24, siin lk 8.

¹⁷⁴ *Machine unlearning* (eesti k masin-unustamine) tähistab tehnilist protsessi, mille eesmärk on eemaldada konkreetse treeningandme mõju juba treenitud mudeli parameetritest ilma kogu mudelit uuesti treenimata.

¹⁷⁵ Cao, Y. jt. Towards Making Systems Forget with Machine Unlearning. - IEEE Symposium on Security and Privacy 2015, lk 463-480. Uuemas kirjanduses: Nguyen, T.T. jt. A Survey of Machine Unlearning. - arXiv:2209.02299 (2022).

proportsionaalne meede. Suurte keelemudelite ümberkoolitamine on ressursimahukas – GPT-suurusega mudeli ümberkoolitamine nõuab hinnanguliselt kümneid miljoneid eurosid arvutusressursse – kuid see on tehniliselt teostatav ning selle tulemus on verifitseeritav: uut mudelit saab testida memoriseerimise suhtes sama metoodikaga, mida kasutati rikkumise tuvastamiseks. FTC pretsedent (Everalbum, WW International) lähtus just sellest kolmandast dimensioonist: kogu algoritmi hävitamise kohustus on sobiv heastamismeede olukorras, kus osaline kustutamine ei ole usaldusväärselt teostatav.

Nendest kolmest dimensioonist tuleneb käesoleva töö jaoks oluline normatiivne järeldus: kõige tõhusam õiguslik heastamisinstrument ei ole mitte algoritmiline eraldamine *post factum*, vaid ennetav *due diligence* nõue enne treenimist. Kui mudel on juba treenitud ja turul, on kõik olemasolevad heastamismeetmed kas tehniliselt usaldusväärselt täitmatud, ebaproportsionaalselt kulukad või kohtus raskesti kontrollitavad. See järeldus on 4.3. alapeatüki tuumikettepanekute peamine normatiivne alus.

GEMA v. OpenAI esimese astme otsus (LG München I; apellatsioon OLG Münchenis pooleli) pakub saastunud mudeli probleemile ühe senistest konkreetsematest käsitlustest esimese astme kohtupraktikas, arvestades, et apellatsioonikohtu seisukoht on veel teadmata. Kohus kohustas OpenAI'd mitte ainult lõpetama laulusõnade kasutamist tulevaste toodete loomisel, vaid ka kustutama kaitstud sisu talletavad mudelikomponendid Saksamaa infrastruktuuris. Apellatsioonimenetluse tulemus võib selle kohustuse ulatust täpsustada.¹⁷⁶ Kohus mõõnis, et selle nõude territoriaalne ulatus piirdub Saksamaaga,¹⁷⁷ kuid õiguslik põhimõte on laiemalt rakendatav: käsk mudeli kaitstud sisu talletavad osad kustutada või mudel ümber koolitada on õiguslik ekvivalent algoritmilisele eraldamisele EL-i kontekstis. See lahend kinnitas, et ei piisa ainult tulevatest rikkumistest hoidumisest – õiguste omajatel on nõudeõigus ka olemasolevate rikkumistoodete eemaldamiseks.

Õiguslikust perspektiivist on saastunud mudeli küsimuse keerukus järgmine: kui mudel on treenitud andmekogumil, mis sisaldab osaliselt (nt 5%) piraatlikelt allikatelt pärit teoseid, kas siis kogu mudel on "saastunud", või piisab vastutuse välistamiseks sellest, et eemaldada piraatlikud andmed treeningandmestikust ja koolitada mudel uuesti? IPRED artikkel 10 näeb ette parandusmeetmete kohaldamise autoriõigust rikkuvate kaupade suhtes,¹⁷⁸ ning 17 U.S.C.

¹⁷⁶ LG München I, 11.11.2025, 42 O 14139/24 (*GEMA v. OpenAI*), punkt 72 (*Anspruch auf Auskunft ja Anspruch auf Unterlassung* – keelava ettekirjutuse nõue ja avalikustamise nõue).

¹⁷⁷ LG München I, 11.11.2025, 42 O 14139/24 (*GEMA v. OpenAI*), punkt 77 (territoriaalne keelav ettekirjutus – mudeli kasutamine Saksamaal).

¹⁷⁸ Euroopa Parlamendi ja nõukogu direktiiv 2004/48/EÜ, 29. aprill 2004, intellektuaalomandi õiguste jõustamise kohta. - ELT L 157, 30.04.2004, lk 45–86, artikkel 10 (parandusmeetmed) (IPRED). ; IPRED, artikkel 10 lõige 1 punkt b: "ebaseaduslikult valmistatud kaupade võõrandamine ärikäibest" - analoogia: "saastunud mudeli" turult kõrvaldamine.

§§ 502–504 kehtestab samaväärsed vahendid USA õiguses.¹⁷⁹ Küsimus on, kas mudel ise kujutab endast autoriõigust rikkuvat "kaupa". Kui mudel on võimeline reprodutseerima konkreetseid kaitstud teoseid (memoriseerimise nähtusel), pooldab GEMA v. OpenAI vastus jaatavat seisukohta: selline mudel on rikkuv kaup, mida tuleb parandada või hävitada.

Nendest kaalutlustest tuleneb normatiivne järeldus, mida senine analüüs ei ole piisavalt rõhutanud: praeguses tehnikaarengus on kõige teostatavam ja proportsionaalsem post factum heastamisvahend mitte *machine unlearning* (mis on kohtus raskesti kontrollitav ja tehniliselt ebausaldusväärselt täidetav) ega kohene hävitamine (mis on AI-arendajate jaoks ebaproportsionaalselt koormav), vaid kasutuskeeld (injunctive relief). IPRED-direktiivi artiklid 10 ja 11 ning 17 U.S.C. § 502 võimaldavad kohtul keelata saastunud mudeli kasutamise ja turustamise kuni seaduslik versioon on välja töötatud.¹⁸⁰ GEMA v. OpenAI esimese astme otsus (LG München I, 11.11.2025; apellatsioon menetluses) rakendas analoogset kohustust: kohus nõudis, et OpenAI lõpetaks kaitstud sisu reprodutseerivate komponentide kasutamise Saksamaa infrastruktuuris.

Kasutuskeeld on heastamisvahendina kolmel põhjusel eelistatav. Esiteks on see tehniliselt teostatav ja kohtus kontrollitav: mudeli kasutamise lõpetamine on selge, jõustatav kohustus, erinevalt *machine unlearning*'i tehnilistest verifikatsiooninõuetest. Teiseks on see proportsionaalne: mudeli kasutamise keeld kuni seadusliku ümberkoolitamise lõpetamiseni jätab AI-ettevõttele võimaluse probleemi lahendada ilma kogu investeringu hävitamiseta. Kolmandaks loob see tõhusa stiimuli ennetavaks due diligence'iks: AI-ettevõtte, kes teab, et piraatandmete avastamine tähendab kasutuskeeldu, on motiveeritud treeningandmeid eelnevalt kontrollima. See seos kasutuskeelu ja *due diligence*'i vahel on käesoleva töö 4.3. alapeatüki tuumikettepanekute struktuuriline alus.

4.2. Praeguse regulatiivse raamistiku puudused

EL-i autoriõigusliku raamistiku peamine puudus käesoleva töö kontekstis on "seadusliku juurdepääsu" nõude vähene operatiivsus. Kuigi DSM-direktiivi artiklid 3 ja 4 välistavad piraatlikest allikatest pärit andmete kasutamise TDM-erandi alusel, on see välistus sõnastatud implitsiitselt mõiste tõlgenduse kaudu, mitte eksplitsiitse keeluna. See tähendab, et AI-ettevõtted saavad mõiste ebaselgusele tugineda oma kaitseks, ning kohtud peavad iga kord

¹⁷⁹ 17 U.S.C. §§ 502-504 (kodifitseeritud autoriõiguse kaitsevahendid USA õiguses: keelav ettekirjutus, hüvitis, kasumi väljamõistmine).

¹⁸⁰ IPRED-direktiiv 2004/48/EU, art 10 ja 11. 17 U.S.C. par 502 sätestab samaväärsed võimalused USA kohtutele. GEMA v. OpenAI otsus (11.11.2025) kohustas lõpetama kaitstud sisu reprodutseerivate komponentide kasutamise.

lähima *ACI Adam*'i analoogilise kohaldamise argumendi – protsessi, mis eeldab ELTK autoriteetset kinnitust enne, kui see saab kõigis liikmesriikides ühtlaselt rakendatavaks.

Tehisintellekti käsitleva määruse läbipaistvuskohustuse artikkel 53 lõige 1 punkt d kohustab ÜTEM-pakkujaid avaldama kokkuvõtte treeningandmetest,¹⁸¹ kuid Euroopa Komisjoni poolt 2025. aastal kehtestatud malli nõuded on piisavalt laiad, et vältida piraatlikele allikatele osutamist.¹⁸² Mall nõuab andmekogumite nimede ja mahtude avaldamist, kuid mitte spetsiifiliselt dokumenteerimist selle kohta, kas andmekogumid sisaldavad piraatlikest allikatest pärit materjali. Lisaks puudub kohustuslik audit või sõltumatu järelevalve, mis kontrolliks treeningandmete seaduslikku päritolu. Veelgi olulisem on see, et artikkel 53 lõige 1 punkt c kohustab ÜTEM-pakkujaid kehtestama autoriõiguse vastavuse poliitika, mis hõlmab DSM-direktiivi artikli 4 lõike 3 *opt-out* mehhanismide jälgimist,¹⁸³ kuid kohustus on protseduuriline, mitte sisuline: AI-büroole piisab poliitikadokumendi olemasolust, mitte selle täitmise tõendamisest.

Opt-out mehhanismi rakendamine on killustunud. DSM-direktiivi artikkel 4 lõige 3 nõuab masinloetavat *opt-out* deklaratsiooni,¹⁸⁴ kuid nagu Kneschke v. LAION otsus näitas, ei ole standardid selle nõude rakendamiseks kõigis kontekstides selged.¹⁸⁵ Schema.org metaandmete standard, IPTC foto-metadata standard ja tdm-reservation HTTP-päis kujutavad endast erinevaid lahendusi erinevate meedialiikide jaoks, kuid ühtne kohustuslik standardiseeritud *opt-out* protokoll puudub EL-i tasandil. Direktiivi põhjendus 18 märgib, et masinloetavus on nõutav, kuid ei täpsusta konkreetset tehnilist standardit.¹⁸⁶

USA autoriõigusliku raamistiku peamine puudus on *fair use* doktriini struktuuriline ettearvamatus AI treenimise kontekstis. *Thomson Reuters v. ROSS*, *Bartz v. Anthropic ja Kadrey v. Meta* otsused loovad kujuneva pildi, kuid kõik kolm on esimese astme kohtute lahendid ilma apellatsioonikohtu kinnitusega. USA Kongress ei ole kehtestanud AI treenimist reguleerivat spetsiifilist seadust, ehkki USCO on soovitanud selliseid meetmeid 2025. aasta

¹⁸¹ Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024 (tehisintellekti käsitlev määrus). - ELT L, 12.07.2024, artikkel 53 lõige 1 punkt d.

¹⁸² Euroopa Komisjon. Tehisintellekti üldotstarbeliste mudelite koolitusandmete avaliku kokkuvõtte mall, 24.07.2025 (avaldatud AI-ameti veebisaidil).

¹⁸³ Tehisintellekti käsitleva määruse artikkel 53 lõige 1 punkt c: "panna paika eeskirjad liidu autoriõiguse ja sellega kaasnevate õiguste järgimiseks."

¹⁸⁴ DSM-direktiivi artikkel 4 lõige 3: "õiguste omajad võivad ära jätta loa anda [...] masinloetaval viisil teatavaks tehtud reservatsiooni kaudu."

¹⁸⁵ OLG Hamburg, 10.12.2025, 5 U 104/24 (*Kneschke v. LAION e.V.*), punkt 51 (tulevased masinloetava *opt-out* standardid), (*opt-out* nõude standardiseerimise vajadus); DSM-direktiivi artikkel 4 lõige 3: "õiguste omajad võivad ära jätta loa anda [...] masinloetaval viisil teatavaks tehtud reservatsiooni kaudu."

¹⁸⁶ DSM-direktiiv (EL) 2019/790, põhjenduspunkt 18.

aruandes.¹⁸⁷ USCO soovib Kongressil kaaluda üldist deklaratiivset seadust, mis selgitaks, et treenimine seaduslikult soetatud teostel on *fair use*'ga kaitstud, kuid piraatlikelt allikatelt kogutud andmetel treenimine mitte. Selline selgitus vähendaks märgatavalt õiguslikku ebakindlust.¹⁸⁸

Rahvusvahelisel tasandil puudub koordineeritud lähenemine AI treenimise autoriõiguslikele küsimustele. WIPO arutelud AI ja intellektuaalomandi kohta on käimasolevad,¹⁸⁹ kuid WIPO ei ole andnud siduvaid soovitusi piraatlikelt allikatelt kogutud andmete kasutamise kohta. Kõigis arenenud jurisdiktsioonides tegutsevatele AI-arendajatele tähendab see vastavust erinevatele, potentsiaalselt vastuoluliste normatiivide komplektidele: EL-i TDM-erandid, USA *fair use*, Suurbritannia autoriõiguse erand ning teiste riikide sätted. See suurendab regulatiivset kulu ning loob võistlusmoonutusi ettevõtjate vahel, keda mõjutavad erinevad normid. Kuigi õiguslik killustumine on intellektuaalomandi valdkonnas tavaline, on AI treenimise kontekst piisavalt eripärane – globaalsed andmekogumid, piiriülene infrastruktuur, koheselt jälgitavad mõjud – et rahvusvaheline koordinatsioon oleks eriti vajalik.¹⁹⁰

4.3. De lege ferenda ettepanekud

Iga ettepanekut hinnatakse kolme kriteeriumi alusel: (i) õiguspoliitiline põhjendus, (ii) rakendatavus praktikas ning (iii) proportsionaalsus autoriõiguse kaitse ja innovatsiooni vahel.

Eelnevalt tuvastatud regulatiivsed lüngad ei lahene kohtupraktika arengu kaudu üksinda. Käesoleva alapeatüki ettepanekud jagunevad selgelt kahte tasandisse. Tuumikettepanekud on töö analüüsis otseselt tulenevad, lähitulevikus institutsionaalselt realistlikud ning läbitud normatiivse prooviga, mis vastab neljale küsimusele: mis probleem laheneb, miks kehtiv õigus seda ei lahenda, miks just see abinõu on proportsionaalne ning milline on selle praktiline teostatavus. Täiendavad ettepanekud on pikemaajalise iseloomuga ning neid esitatakse regulatiivse visiooni osana, mille teostatavus sõltub laiemate protsesside arengust. Ettepanekud järgivad õiguspoliitilist loogikat: esmalt tuleb kehtestada selged

¹⁸⁷ USA Autoriõiguse Amet (US Copyright Office). Copyright and Artificial Intelligence, Part III: Generative AI Training. Washington, D.C.: US Copyright Office, 9. mai 2025 (USCO 2025), lk 65–66 (soovitud Kongressile).

¹⁸⁸ USCO 2025 (viide 24), lk 68–70 (kõikehõlmav litsentsimine AI treenimiseks).

¹⁸⁹ WIPO. Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Summary (2019). Vt WIPO/IP/AI/GE/20/1 Rev. (2020). ; WIPO. Scoping Paper on Copyright and Artificial Intelligence (2021), lk 14–18 (AI treenimine ja autoriõigus).

¹⁹⁰ Sag, M. Copyright Safety for Generative AI. - Houston Law Review, 61, nr 2, 2024, lk 295–350, siin lk 346–348.

normid piraatandmete keelustamiseks (preventsioon), seejärel luua tõhusad heastamisvahendid rikkumise tuvastamise korral (*remediatio*) ning lõpuks tagada õiguste omajate õiguste reaalse jõustamise võimalused (*executio*).

Tuumikettepanekud

Esimene tuumikettepanek: tehisintellekti käsitleva määruse artikli 53 täiendamine due diligence'i kohustusega

Kehtiv regulatiivne lünk seisneb selles, et tehisintellekti käsitleva määruse artikkel 53 lõige 1 punkt c kohustab ÜTEM-pakkujaid kehtestama autoriõiguse vastavuse poliitika,¹⁹¹ kuid kohustus on protseduuriline, mitte sisuline. Tehisintellekti käsitleva määruse põhjendus 105 märgib küll, et ÜTEM-pakkujad peaksid järgima liidu autoriõiguse õigust,¹⁹² kuid see on soovituslik, mitte sanktsiooniga tagatud kohustus piraatlikest allikatest andmete kasutamise suhtes. *Bartz v. Anthropic* faktilised asjaolud näitasid, et ettevõtte kasutas teadlikult Pirate Library Mirror'it hoolimata siseprotseduuridest. Kehtiv õigus ei lahenda seda lünka, kuna poliitika olemasolu ei välista selle süstemaatilist eiramist.

Ettepanek on täiendada tehisintellekti käsitleva määruse artiklit 53 lõiget 1 *due diligence'i* kohustusega, mis nõuab ÜTEM-pakkujatelt treeningandmete päritolu kontrollimist proportsionaalses ulatuses enne andmekogumi kasutamist ning kontrollitulemuste dokumenteerimist viisil, mis on AI-büroole kättesaadav järelevalve käigus. Proportsionaalsus tähendab astmelisust: avaliku veebi kraapimise teel kogutud andmestikud nõuavad rangemate kontrollipõhimõtete rakendamist – eelkõige tuntud piraatplatvormide aktiivset välistamist – kui litsentsitud andmekogumid. Ettepanek on proportsionaalne kahel põhjusel. Esiteks ei nõuta iga teose algallika individuaalset kontrollimist, vaid süstemaatilise kontrollimehhanismi olemasolu ja dokumenteerimist, mis on AI-ettevõtete praeguse võimekuse piires. Teiseks on GDPR-is väljakujunenud vastutuse põhimõte – artikkel 5 lõige 2 – on põhimõtteliselt analoogiline, kuid analüüsis tuleb arvestada, et GDPR reguleerib isikuandmete töötlemist, mitte intellektuaalomandi vastavust. Analoogia on pedagoogiliselt kasulik, kuid ei ole normatiivne argument selles mõttes, et GDPR *accountability*-kohustus ei laiene automaatselt autoriõiguse konteksti. Ettepanek peaks selgelt möönma analoogia struktuurilist piiri ning tuginema iseseisvatele põhjendustele. Praktiliseks teostatavuseks on

¹⁹¹ Tehisintellekti käsitleva määruse artikkel 53 lõige 1 punkt c: "panna paika eeskirjad liidu autoriõiguse ja sellega kaasnevate õiguste järgimiseks."

¹⁹² Tehisintellekti käsitleva määruse põhjenduspunkt 105: "Kui üldotstarbelise tehisintellektimudeli pakkujad täidavad käesolevast määrusest tulenevaid kohustusi, peaksid nad eelkõige järgima liidu õigust autoriõiguse kohta."

see, et *Bartz v. Anthropic* kohtuasjas saavutatud kokkulepe näitas, et AI-ettevõtted on valmis *due diligence* protokolle kehtestama vabatahtlikult.¹⁹³

Tugevaim vastuargument innovatsioonikahju kohta on nõrk: *due diligence* kohustus ei piira treenimist seaduslikult soetatud andmetel, vaid loob ainult dokumenteerimiskohustuse, mille täitmiseks vajalik ressursikulu on marginaalne võrreldes suurte keelemudelite arendamise üldkuludega. Andmekogumite täielik manuaalne audit ei ole suures andmemahus teostatav – seda tuleb tunnistada. Käesolev ettepanek seda ei nõuagi. *Due diligence* kohustuse sisu peaks olema kolmeastmeline, näiteks: esiteks, tuntud piraatplatvormide automaatne välistamine kraapimisloendist. LibGen, Pirate Library Mirror, Books3 ja nende teadaolevad peegelsaidid on identifitseeritavad domeenid. Nende välistamine andmekogumist ei ole tööstuslikult keeruline – see on filter, mille rakendamine on AI-ettevõtete jaoks tehniline rutiin. Teiseks: statistiline valimkontroll. Andmestiku juhuslik alamhulk auditeeritakse ning tulemused dokumenteeritakse. Antud lähemine ei garanteeri täielikku puhtust, kuid tõendab, et ettevõtte on võtnud mõistlikke meetmeid. Kolmandaks võiks olla protsessidokumentatsioon, mis on AI-büroole kättesaadav järelevalve käigus. Sellise astmelise *due diligence* täitmine muudaks ettevõtte positsiooni IPRED art 13 lg 2 tähenduses oluliselt. Tõendatud mõistlik hoolsus viib tahtliku rikkuja ja heauskse rikkuja eristamiseni, mis mõjutab otseselt kahjuhüvitise suurust. See on nii normatiivne eesmärk kui praktiline stiimul.

Teine tuumikettepanek: DSM-direktiivi artiklite 3 ja 4 sõnaselge täiendamine

Kehtiv regulatiivne lünk seisneb selles, et DSM-direktiivi artiklite 3 ja 4 "seadusliku juurdepääsu" nõue välistab küll piraatlikelt allikatelt pärit andmete kasutamise TDM-erandite kaitseulatuses, kuid see välistus ei ole operatiivnormi tasandil sõnaselge. Direktiiv jätab mõiste "seaduslik juurdepääs" defineerimata, mistõttu kohtud peavad rakendama *ACI Adam*'i analoogiat – tõlgendusreservi, mis eeldab ELTK autoriteetset kinnitust (C-250/25) enne, kui see saab kõigis liikmesriikides ühtlaselt rakendatavaks. Kehtiv õigus ei lahenda seda lünka, kuna puuduv operatiivne definitsioon loob õigusliku ebakindluse olukorras, kus AI-ettevõtted viitavad mõiste ebaselgusele oma kaitseks.

Ettepanek on täiendada DSM-direktiivi artikleid 3 lõiget 1 ja 4 lõiget 1 klausliga järgmises sõnastuses: "Teos ei ole käesoleva artikli mõttes seaduslikult ligipääsetav, kui juurdepääs sellele on saavutatud autoriõigust rikkudes, sõltumata sellest, kas teos on faktiliselt kättesaadav." See täiendus on proportsionaalne kolmel põhjusel. Esiteks ei laienda

¹⁹³ *Bartz v. Anthropic PBC*, nr 3:24-cv-05417-WHA (N.D. Cal.), kokkuleppeteadaanne (august 2025) – *due diligence*-protokolli nõue andmekogumite päritolu kontrollimiseks.

see direktiivi kohaldamisala ega piira seaduslikku TDM-tegevust – see üksnes kodifitseerib tõlgenduse, mis tuleneb juba kehtivast *ACI Adam*’i doktriinist ja kolmeastmelisest testist.¹⁹⁴ Teiseks on sõnastus piisavalt täpne, et välistada moonutav tõlgendus, mille kohaselt ka marginaalne ebaseaduslik element andmekogumis muudaks kogu TDM-erandi kohaldamatuks. Kolmandaks tuleks vältida liiga kitsaste erandite kehtestamist, mis häiriks seaduslikku teadusuuringute tegevust.¹⁹⁵ EL-i õigusega koskõla on täielik, kuna täiendus tugineb ELTK väljakujunenud põhimõtetele. Praktiline teostatavus on realistlik: Euroopa Komisjoni 2026. aasta digitaalse autoriõiguse reformikava hõlmab AI-treenimise autoriõiguse selgitamist, ning selline sõnaselge klausel on minimaalseim ja poliitiliselt vähem vastuoluline sekkumine kui TDM-erandite sisuline piiramine. Tugevaim vastuargument on, et selline täiendus võib luua õigusliku ebakindluse hea usu puhul tegutsevate AI-ettevõtjate jaoks, kelle andmekogumid sisaldavad tahtmatult vähesel määral ebaseaduslikku materjali. See risk on maandatav sellega, et klausel peaks sõnaselgelt kohalduma üksnes olukorras, kus juurdepääs ebaseaduslikule allikale oli teadlik või kus allpool kirjeldatava *due diligence* kohustuse täitmine oleks rikkumise ilmsiks toonud.’

Kolmas tuumikettepanek: IPRED-direktiivi täiendamine kasutuskeelu mehhanismiga saastunud AI-mudelite suhtes.

Kehtiv regulatiivne lünk seisneb selles, et IPRED-direktiivi artikkel 10 näeb ette parandusmeetmete kohaldamise autoriõigust rikkuvate kaupade suhtes, kuid see raamistik ei ole kohandatud olukorrale, kus rikkuv "kaup" on immateriaalse loomuga mudel. *GEMA v. OpenAI* näitas, et kohtud on valmis andma korralduse lõpetada kaitstud sisu reprodutseerivate komponentide kasutamine, kuid selline kohustus nõuab selget õiguslikku alust.

Ettepanek on täiendada IPRED-direktiivi artiklit 10 sättega, mis kehtestab saastunud AI-mudelite suhtes kohaldatava astmelise heastamismehhanismi järgmiselt. Esimeses etapis peab õiguste omaja tõendama kahte asja. Esiteks, et treeningandmestik sisaldas piraatlikest allikatest pärit teoseid – faktiliselt tuvastataval kujul, näiteks andmekogumi koosseisu analüüsi teel. Teiseks, et mudel on võimeline vähemalt üht sellist teost memoriseeritult reprodutseerima – standardiseeritud memoriseerimise tuvastamise testi alusel Carlini jt (2023) meetoodika kohaselt. Mõlemad tingimused peavad olema täidetud korraga: pelgalt

¹⁹⁴ Geiger, C.; Frosio, G.; Bulayenko, O. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market. – GRUR International, kd 67, nr 8, 2018, lk 824-836, siin lk 830-832.

¹⁹⁵ DSM-direktiivi artikkel 7 lõige 1: liikmesriigid peavad tagama, et artiklite 3-6 eranditest kinnipidamist ei piira lepingulised sätted.

andmestiku koosseis ei ole piisav, kui mudel ei reprodutseeri kaitstud sisu; pelgalt reprodutseerimisvõime ei ole piisav, kui andmepäritolu ei ole tuvastatud. See kahekordne lävi väldib ebaproportsionaalset vastutust juhtudel, kus piraatlikku materjali on andmestikus marginaalses mahus ega avaldu mudeli väljundis. Käesolev töö ei paku *de minimis* künnist protsentides – see küsimus jääb ELTK kujundada. Kuid põhimõte on selge: lävenditingimuseks on reprodutseeritavuse faktiline tuvastamine, mitte andmestiku abstraktne koosseis.

Kui see on tõendatud, kohaldub teises etapis kohustus lõpetada saastunud mudeli kasutamine ja turustamine (kasutuskeeld – *injunctive relief*) mõistliku tähtaja jooksul, mille kohus määrab sõltuvalt rikkumise ulatusest. Kasutuskeeld kehtib kuni AI-ettevõtte tõendab, et on välja töötanud seadusliku ümberkoolitatud versiooni, mis läbib memoriseerimise tuvastamise testid. Kogu mudeli hävitamine jääb kolmandaks astmeks üksnes siis, kui ümberkoolitamine ei ole proportsionaalse aja jooksul teostatav ja rikkumine on ulatusliku ning tahtliku iseloomuga.

See lähenemisviis on koherentne käesoleva töö ptk 4.1 järelausega, et *machine unlearning* ei ole praeguse tehnika tasemel usaldusväärselt kohtus kontrollitav: kasutuskeeld ei eelda *machine unlearning*'i teostatavust, vaid asendab selle selge jõustatava meetmega.

Siin võib tekkida põhjendatud küsimus: miks just IPRED-direktiiv, mitte näiteks tehisintellekti käsitleva määruse (2024/1689) täiendamine intellektuaalomandi artiklitega või eraldi AI-autoriõiguse jõustamise instrumendi loomine? IPRED-direktiiv on üldine intellektuaalomandi jõustamise raamistik, kuid seda ei kasutata esimest korda sektorispetsiifilisel eesmärgil – analoogselt on ravimite võltsimise vastane direktiiv (2011/62/EL) lisanud sellesse valdkonnaspetsiifilise kihi. Põhiline praktiline eelis on see, et IPRED sisaldab juba toimivat heastamisvahendite raamistikku (artiklid 10–11), millele AI-spetsiifilise kasutuskeelu lisamine on kiirem ja vähem ressursimahukas kui täiesti uue instrumendi loomine nullist. Tehisintellekti käsitleva määruse täiendamine oleks struktuuriliselt elegantne, kuid nõuaks pikka seadusandlikku protsessi – IPRED täiendamine on sihtotstarbelisem lahendus, mis ei sõltu tehisintellekti käsitleva määruse regulaarsest läbivaatamise tsüklist.

Täiendavad ettepanekud

Järgnevad kolm ettepanekut on pikemaajalise iseloomuga ning neid esitatakse regulatiivse visiooni osana, mitte töö põhiargumendina.

EL-i tasandi kohustusliku kollektiivlitsentsimise süsteem AI treenimiseks lahendaks piraatlikelt allikatelt treenimise majandusliku ajendi: kui seaduslik alternatiiv on taskukohane ja kättesaadav, kaob piraatlike allikate kasutamise ratsionaalne majanduslik põhjendus.¹⁹⁶ DSM-direktiivi artiklites 12-13 sätestatud laiendatud kollektiivlitsentsimise mudel¹⁹⁷ ning InfoSoc-direktiivi artikkel 9¹⁹⁸ pakuvad valmis raamistiku, millele saaks tugineda. Quintais jt on süstemaatiliselt analüüsinud kollektiivlitsentsimise praktilisust AI treenimise kontekstis, tehes järelduse, et süsteem on teostatav tingimusel, et koostatakse piisavalt esinduslik õiguste omajate kollektiivne esindusorganisatsioon, kehtestatakse selge tasu jagamise mehhanism ning AI-arendajatele tagatakse õiguslik kindlus treeningandmete kasutamisel.¹⁹⁹

Eesti tasandil tuleks vähemalt kaks muudatust rakendada. Esiteks tuleks AutÕS §-le 19² lisada selgitav lõige, et TDM-erand ei kohaldu teostele, millele on juurde pääsetud autoriõigust rikkudes.²⁰⁰²⁰¹ See täpsustus annaks Eesti kohtutele selge aluse TDM-erandi tagasilükkamiseks piraatandmete kasutamise korral, ilma vajaduseta tugineda üksnes EL-i direktiivi analoogilisele tõlgendamisele. Teiseks tuleks kaaluda vastutava ministeeriumi koordineeritavat sektoritevahelise ümarlauaprotsessi käivitamist autorite, kirjastajate ja AI-arendajate osavõtul, eesmärgiga töötada välja Eestis kohaldatav AI-litsentsimise raamistik.²⁰²

WIPO tasandil ühtse masinloetava *opt-out* standardi kehtestamine on käesoleva töö kõige pikemaajalisem ettepanek, mida esitatakse teadliku reservatsiooniga: selle teostatavus eeldab rahvusvahelist lepinguloomet, mille ajahorisont ulatub tõenäoliselt kümnendi lõppu. Siduvaks instrumendiks jõudmine nõuab konsensust peamiste osapoolte vahel, mis ei ole lähiajal realistlik. Selle ettepaneku väärtus seisneb seetõttu suuna näitamises, mitte koheses rakendatavuses: globaalne AI treeningandmete turg vajab globaalset normatiivset koordinatsiooni, ning praegune jurisdiktsioonide killustumine on pikemas perspektiivis jätkusuutmatu nii autoriõiguse omajate kaitse kui ka AI-arendajate õigusliku kindluse

¹⁹⁶ Quintais, J.P. jt. Licensing Copyright for AI Training: Considerations and Policy Options. SSRN Working Paper 2024, lk 8-12.

¹⁹⁷ DSM-direktiivi artiklid 12–13 (laiendatud kollektiivlitsentsimine ja vabatahtlik kollektiivlitsentsimine).

¹⁹⁸ Euroopa Parlamendi ja nõukogu direktiiv 2001/29/EÜ (InfoSoc-direktiiv), artikkel 9 – kaabelülekanalilitsentsid koos kohustusliku kollektiivse esindamisega.

¹⁹⁹ Quintais, J.P. jt. Licensing Copyright for AI Training: Considerations and Policy Options. SSRN Working Paper 2024, lk 14-17.

²⁰⁰ Autoriõiguse seadus (AutÕS). – RT I, 12.07.2025, 9 (jõust. 01.10.2025), § 19² lõige 1 (teksti- ja andmekaeve üldine erand).

²⁰¹ AutÕS § 19² lõige 2: *opt-out* mehhanism.

²⁰² Kelli, A.; Välimäe, V.; Hoffmann, A. Legal Framework for Artificial Intelligence in Estonia. -Juridica International, 2022, nr 31, lk 3–15, siin lk 12. ; Kelli, A. jt. Eesti autoriõiguse reform. - Juridica, 2021, nr 4, lk 249262, siin lk 259.

tagamise seisukohalt.²⁰³ WIPO käimasolev protsess AI ja intellektuaalomandi küsimuste on sobiv foorum sellise standardi algatamiseks, kuid käesolev töö ei loe seda töö põhiargumendi, vaid pikaajalise poliitilise eesmärgi osana.

Kokkuvõtlikult näitavad käesolevas peatükis esitatud analüüs ja ettepanekud, et kolmest tuumikettepanekust on käesoleva töö hinnangul kõige kõrgema prioriteediga tehisintellekti käsitleva määruse *due diligence* kohustus, kuna see tegutseb seal, kus õiguslik sekkumine on kõige tõhusam: enne rikkumise tekkimist. DSM-direktiivi sõnaselge täiendus tagab pikemas perspektiivis õigusliku selguse ning IPRED-i astmeline mehhanism lahendab heastamise probleemi *post factum*. Need kolm ettepanekut moodustavad koherentse terviku, milles preventsiiooni, selguse ja heastamise instrumendid täiendavad üksteist. "saastunud mudeli" probleem on tehniliselt tõsisem väljakutse kui piraatluse tuvastamine ise: kui piraatlikest allikatest andmete kogumine on juba toimunud ja mudel on treenitud, ei ole lihtsat tagasivaatavat lahendust. Seetõttu on ennetava poliitika prioriteet – selged normid andmekogumite seadusliku päritolu kohta, tõhusad *opt-out* mehhanismid ja proportsionaalsed heastamisvahendid – kõige tähtsam samm AI treenimise autoriõiguslike küsimuste lahendamisel.²⁰⁴

²⁰³ Senftleben, M.; Buijelaar, L. Robot Creativity: An Incentive-Based Neighbouring Rights Approach. - European Intellectual Property Review, 42, nr 12, 2020, lk 797-811, siin lk 808-810 ; WIPO. Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Summary (2019). Vt WIPO/IP/AI/GE/20/1 Rev. (2020).;

WIPO. Scoping Paper on Copyright and Artificial Intelligence (2021), lk 14-18.

²⁰⁴ ELTK eelotsusetaotlus, Like Company Kft versus Google Ireland Ltd, C-250/25, esitatud 3. aprill 2025. Suuline menetlus toimus 10. märtsil 2026.

KOKKUVÕTE

Käesolev töö vastas püstitatud kesksele uurimisküsimusele, näidates, et piraatallikatelt pärinevate andmete kasutamine generatiivse tehisintellekti treenimisel kujutab Euroopa Liidu kehtiva õiguse tõlgenduse kohaselt endast autoriõiguse rikkumist, samas kui Ameerika Ühendriikide õiguses sõltub selle lubatavus juhtumipõhisest *fair use*'i analüüsist.

Magistritöö uuris küsimust, kas generatiivse tehisintellekti mudelite treenimine piraatlikelt allikatelt kogutud andmetel kujutab endast autoriõiguse rikkumist Euroopa Liidu ja Ameerika Ühendriikide õiguse alusel ning millised õiguslikud tagajärjed peaksid sellisest treenimistegevusest tulenema. Töö analüüs hõlmas DSM-direktiivi TDM-erandite süsteemi, Eesti autoriõiguse seaduse transponeerimise eripära, tehisintellekti käsitleva määruse artiklist 53 tulenevaid ÜTEM-ide kohustusi ning USA *fair use*'i doktriini neljaetapilist testi.

Töö hüpoteesi EL-i poolne osa leidis doktrinaalset tuge, kuid mitte lõplikku kohtupraktikalist kinnitust. Seadusliku juurdepääsu nõude grammatiline, teleoloogiline ja süstemaatiline tõlgendus koos *ACI Adam* analoogiaga moodustavad tugeva argumendi seisukoha kasuks, et piraatlikelt allikatelt kogutud treeningandmed jäävad TDM-erandite kohaldamisalast väljapoole. Lõplikuks kohtupraktikalist kinnituseks on vaja ELTK lahendit asjas C-250/25, mis on käesoleva töö esitamise hetkel ootel. USA õiguse osas on olukord keerulisem. Lahutamise doktriin (*Bartz v. Anthropic*, Alsup J.) toetab hüpoteesi teoreetiliselt, kuid ainus lõplik USA ringkonnakohtu otsus – *Kadrey v. Meta* (Chhabria J., 25. juuni 2025) – läks AI-arendaja kasuks, lubades piraatandmetel treenimist *fair use*'i katte all. Otsus ei kujuta endast siiski üldist doktriinilist kinnitust piraatandmetel treenimise seaduspärasusele – Chhabria lisas otsusesse sõnaselge reservatsiooni, mida käsitletakse täpsemalt allpool. USA õiguse tegelik olukord on seega hetkel AI-ettevõtetele soodsam kui töö esmakordne lugemine võiks jätta mulje. Hüpotees leidis doktrinaalset tuge (lahutamise doktriin), kuid mitte operatiivset kohtupraktikat USA-s.

EL-i õiguse vaatenurgast viib kehtiva õiguse kõige veenvam tõlgendus järelduseni, et piraatallikatelt kogutud treenimisandmete kasutamine ei ole DSM-direktiivi TDM-eranditega kaetud. DSM-direktiivi artiklid 3 ja 4 saavad mõlemad TDM-toimingule eeltingimuseks seadusliku juurdepääsu olemasolu – nõude, mille grammatiline, teleoloogiline ja süstemaatiline tõlgendus jõuavad ühisele järeldusele, et piraatlikust allikast pärit teos ei ole direktiivi mõttes "seaduslikult ligipääsetav", hoolimata sellest, et see on faktiliselt veebis

kättesaadav. ACI Adam'i (C-435/12) doktriin kinnitab seda tõlgendust: autoriõiguse erandid ei ole loodud ebaseadusliku algkäitumise tagantjärele seadustamiseks. Seda suunda toetavad ka 2025. aasta Saksa kohtulahendid AI-spetsiifilises kontekstis. Müncheni I Ringkonnakohtu esimese astme otsus asjas *GEMA v. OpenAI* (apellatsioon menetluses) käsitleb memoriseerimist iseseisva reprodutseerimisprobleemina, samas kui Hamburgi Apellatsioonikohtu jõustunud otsus asjas *Kneschke v. LAION* täpsustab seadusliku *opt-out* mehhanismi standardit. Ükski neljast käsitletud argumendist – transformatiivsus, ühiskondlik kasu, piraatluse moraalne ambivalentsus ega *opt-out*-mehhanismi kriitika – ei ole piisav, et ületada seadusliku juurdepääsu nõudest tulenevat põhitakistust kehtiva õiguse kõige põhjendatuma tõlgenduse kohaselt. Selle seisukoha otsene autoriteetne kinnitamine ELTK poolt on ootel asjas C-250/25.

USA kohtupraktika pilt on märkimisväärselt keerukam ja doktrinaalselt lahtisem. Kõige olulisem on tunnistada, et 2025. aasta seisuga on ainus lõplik ringkonnakohtu otsus AI treenimise *fair use* küsimuses antud *Kadrey v. Meta* asjas – ja see otsus läks AI-arendaja kasuks, hoolimata tuvastatud piraatandmete kasutamisest. Siinkohal tuleb lisada täpne reservatsioon. Chhabria kommentaar "*enamikul juhtudel on tõenäoliselt vastus jah*" on *obiter dictum*, ehk ei ole otsuse resolutsioon ega siduv pretsedent. Sellele toetuda hüpoteesi põhjendamisel oleks tõenduslikult nõrk positsioon. Täpsem on öelda, et kohus viitas mittesiduvalt, millisel teoorial oleks olnud parem alus. Hagejad ei kasutanud seda teooriat ja seetõttu kaotasid. Sellele lisandub Chhabria sõnaselge reservatsioon: otsus "seisab üksnes selle eest, et käesolevad hagejad esitasid valesid argumente ja ei kujundanud välja tõendusbaasi õigetest argumentidest" (autori tõlge) – rõhutades, et faktispetsiifiline võit ei tähenda üldist kinnitust AI treenimise seaduspärasusele. Kohtunik Alsup'i lahutamise doktriini sõnastanud otsus *Bartz v. Anthropic* asjas jäi mittesiduvaks, kuna asi lõppes kokkuleppega. Käesolev töö eelistab lahutamise doktriini, kuna see sobib paremini Warholi standardiga. Warhol suunab hindamise konkreetsele toimingule – ja allalaadimine on iseseisev toiming, mitte transformatiivse protsessi osa. Kuid see on normatiivne eelistus, mitte kirjeldus kehtivast USA õigusest. Kehtiv USA õigus – nagu seda esindab ainus lõplik otsus (Kadrey, 2025) – liigub vastupidises suunas. Kuid see analüütiline eelistus ei muuda kehtivat operatiivset seisut: USA õiguses on piraatandmetel treenimine *de facto* lubatud kuni Kolmanda Apellatsiooniringkonna oodatava lahendini (*Thomson Reuters v. ROSS*, nr 25-2153). Just see ebakindlus on iseenesest argument EL-i struktuuri selgema lähenemise kasuks.

Töö hüpoteesi teine pool, et kumbki süsteem ei paku piisavalt tõhusaid õiguskaitsevahendeid juba treenitud mudelite suhtes, leidis samuti kinnitust. EL-i ega USA õigus ei paku tõhusat *post factum* vahendit olukorras, kus mudel on piraatandmetel treenitud ja turul. Praegune tehnikaareng ei võimalda *machine unlearning*'ut usaldusväärselt teostada: konkreetse teose memoriseeritud mõju eemaldamine mudeli parameetritest ilma kogu mudeli ümberkoolitamiseta on kohtus kontrollitaval viisil lahendamata tehniline probleem. USA õigus pakub algoritmilist eraldamist FTC pretsedendi kaudu, kuid selle teostatavus AI-mudelite suhtes on vähendatud just verifitseerimisstandardite puudumise tõttu. Sellest tulenevalt on praegu kõige teostatavam *post factum* vahend kasutuskeeld – kohustus lõpetada saastunud mudeli turustamine kuni tõendatult seaduslik versioon on turule toodud. Just see tõdemus annab käesoleva töö kolmele tuumikettepanekule nende praktilise aluse. Kui *post factum* heastamine on piiratud, jääb ainsaks realselt toimivaks meetmeks preventatsioon, mis hõlmab *due diligence* kohustust, selget normatiivset välistust ja tõhusat õiguslikku vastutust.

Kolm tuumikettepanekut on esitatud prioriteedi järjekorras. Kõige kiireloomulisem on tehisintellekti käsitleva määruse artikli 53 täiendamine *due diligence*'i kohustusega, mis nõuaks treeningandmete päritolu proportsionaalset kontrolli enne andmekogumi kasutamist – meede, mis toimib preventiivsel tasandil ja mille rakendamine on AI-büroole delegeeritud rakendusaktide kaudu kiiremini saavutatav kui seadusandlik muudatus. Sellele järgneb DSM-direktiivi artiklite 3 ja 4 sõnaselge täiendamine piraatlikke allikaid välistavate klauslitega, mis kõrvaldaks praeguse normatiivse ebakindluse ilma ELTK tõlgendust ootamata. Kolmandaks on vajalik IPRED-direktiivi täiendamine AI-spetsiifilise kasutuskeelu mehhanismiga, mis muudaks "saastunud mudeli" turult kõrvaldamise operatiivselt teostatavaks.

Käesoleva magistritöö analüüs näitab, et kahe õigussüsteemi struktuuriline erinevus – EL-i kategooriline "seadusliku juurdepääsu" nõue *versus* USA kaalumistest – ei ole pelgalt tehniline normidevaheline erinevus, vaid peegeldab fundamentaalselt erinevat õigusfilosoofilist lähenemist küsimusele, kelle kanda peaks jääma autoriõiguse kaitse koorem AI arengu kontekstis. USA õiguse lõplik vastus sõltub Kolmanda Apellatsiooniringkonnakohtu otsusest Thomson Reutersi asjas (nr 25-2153, oodatav 2026). Kui ringkonnakohus kinnitab, et *fair use* ei kata AI treenimist, toetab see töö hüpoteesi ka USA-s. Kui ringkonnakohus tühistab lahendi, jäävad *de lege ferenda* ettepanekud ainult veelgi vajalikumaks. Mõlemal juhul jääb EL-i normatiivne seisukoht samaks: kehtiva õiguse kõige veenvama tõlgenduse kohaselt ei ole piraatlikest allikatest pärit treenimisandmete kasutamine TDM-eranditega kaetud.

SUMMARY

Training Generative Artificial Intelligence Models on Data Collected from Pirated Sources: Copyright Infringement or Justified Use under European Union and United States Law?

This master's thesis examines whether the training of generative artificial intelligence (AI) models on data collected from pirated sources constitutes copyright infringement or justified use under the applicable legal frameworks. The research covers two primary legal systems – the European Union's copyright framework (including its transposition into Estonian law) and the United States fair use doctrine – and addresses both the current state of the law and forward-looking proposals for reform.

The first chapter establishes the conceptual foundations of the study. Generative AI models, including large language models (LLMs) such as GPT-4, Claude, Gemini and LLaMA, require training on datasets of extraordinary scale. The composition of these datasets has become increasingly controversial, as AI developers have turned not only to publicly licensed or owned data but also to content scraped from the open web, including from shadow libraries and pirate repositories such as LibGen, Books3 and Sci-Hub. The chapter defines “piracy” in the AI training context not merely as end-user downloading but as systematic unauthorised reproduction for training purposes and introduces the concept of the “contaminated dataset” – a training corpus that contains materials derived from unlawful sources. The scale of such contamination fundamentally transforms the traditional understanding of copyright infringement.

The first chapter also critically examines the strongest counterarguments advanced in favour of treating AI training as justified use – as required by the disjunctive framing of the thesis title. Four principal arguments are considered. First, the transformative nature argument holds that AI models process statistical patterns rather than storing text verbatim, analogous to a painter studying technique from existing works. Second, the societal benefit argument invokes the balance between authors' rights and the public interest recognised in the Berne Convention, the InfoSoc Directive and the US Copyright Act. Third, the moral ambivalence argument questions whether training on pirate repositories is morally equivalent to commercial piracy, particularly where the original access barriers are themselves contested. Fourth, the opt-out system critique argues that the DSM Article 4 mechanism *de*

facto protects only major publishers, leaving individual authors exposed regardless of dataset legality. Each argument is assessed and rebutted: the memorisation phenomenon documented in Carlini et al. (2023) and confirmed in *GEMA v. OpenAI* undermines the statistical-patterns claim; the availability of lawful licensing markets (Getty, AP, Axel Springer, The New York Times) defeats the societal-benefit exception; the commercial character of Books3 and PiLiMi data refutes the academic-access analogy; and the opt-out system's weaknesses are better addressed through the *de lege ferenda* proposals than by legitimising non-compliance. The analysis demonstrates that these arguments, while not without intellectual substance, do not overcome the lawful access requirement as interpreted in this thesis, according to which pirated sources fall outside the scope of the EU TDM exceptions.

The second chapter analyses the European Union legal framework, with particular focus on the TDM exceptions in Articles 3 and 4 of the DSM Directive (EU) 2019/790 and their transposition into Estonian law through sections 19¹ and 19² of the Estonian Copyright Act (AutÕS). The analysis centres on the “lawfully accessible” requirement that both provisions impose as a prerequisite for invoking the TDM exception. By applying the principle established in *ACI Adam BV v Stichting de ThuisKopie (C-435/12)*, which holds that unlawful sources cannot benefit from reproduction exceptions, it is concluded that AI developers who train models on datasets containing unlawfully reproduced works cannot rely on the DSM TDM exceptions. The German Landgericht München I in *GEMA v. OpenAI* (Az. 42 O 14139/24, 11 November 2025; first-instance decision, under appeal before OLG München) reasoned at first instance that memorised works reproduced by a model constitute an independent act of reproduction – a principle grounded in the EU's common copyright framework and therefore analytically applicable regardless of the appeal outcome and the Hanseatisches Oberlandesgericht Hamburg in *Kneschke v. LAION* (Az. 5 U 104/24, 10 December 2025) clarified the standard for a valid machine-readable opt-out under DSM Article 4(3), holding that robots.txt does not constitute a sufficient opt-out declaration – a conclusion which, by analogy with *ACI Adam*, implies that the opt-out mechanism cannot retroactively cure initial data collection from an unlawful source. The chapter further analyses the doctrine of objective liability under EU copyright law: while infringement is established without proof of knowledge, IPRED Article 13(2) allows courts to award non-punitive damages where the infringer neither knew nor had reasonable grounds to know of the infringement – making documented due diligence the primary instrument for reducing

damages exposure and the primary normative basis for the due diligence proposal in Chapter 4.3.

The third chapter examines the United States copyright framework, where the analysis revolves around the four-factor fair use test under 17 U.S.C. § 107. The Supreme Court's 2023 decision in *Andy Warhol Foundation v. Goldsmith* (598 U.S. 508) substantially narrowed the transformativeness argument: commercial-purpose use cannot rely on abstract transformativeness unless it serves a genuinely different purpose from the original. Applying this standard in *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence, Inc.* (D. Del., 11 February 2025), the Delaware district court found that ROSS's training purpose was functionally identical to the original use of Westlaw headnotes – a significant precedent, though one that must be applied cautiously to generative AI, which produces novel output rather than directly substituting for the original. The separation doctrine articulated by Judge Alsup in *Bartz v. Anthropic* (No. 3:24-cv-05417-WHA, N.D. Cal.; case settled August 2025, ruling non-binding) holds that the unlawful downloading of pirated materials is an independent act of infringement, not curable by the transformativeness of subsequent training. This thesis finds this position more consistent with the Warhol standard and with copyright's preventive function, though the operative 2025 precedent (Kadrey/Chhabria) moved in the opposite direction. However, the operative 2025 precedent points in the opposite direction: Judge Chhabria granted Meta summary judgment in *Kadrey v. Meta Platforms* (N.D. Cal., 25 June 2025) on fair use grounds despite confirmed use of LibGen data, finding the training "highly transformative." Importantly, Chhabria introduced a "market dilution" theory in the fourth-factor analysis – holding that LLMs generating competing works at scale could harm the market for originals – and stated that "in most cases the answer will likely be yes" that AI training on copyrighted works is infringement on this basis. The plaintiffs nonetheless lost because they failed to develop the market dilution argument, relying instead on substitutive market harm they could not prove. Chhabria also included a pointed reservation: "*This decision does not stand for the proposition that Meta's use is lawful. It stands only for the proposition that these plaintiffs made the wrong arguments and failed to develop a record in support of the right one.*" This is the only final district court ruling on point as of 2025. The US copyright framework on AI training therefore remains genuinely unsettled pending the Third Circuit's decision in *Thomson Reuters v. ROSS* (No. 25-2153, expected late 2026). That doctrinal uncertainty is itself the strongest argument for

the EU’s categorical lawful-access approach – a system that produces divergent outcomes in materially identical cases fails to provide adequate legal certainty.

The fourth chapter addresses the contaminated model problem and advances *de lege ferenda* proposals. Chapter 4.1 analyses the contaminated model – a generative AI model trained on a dataset containing pirated material – through three distinct dimensions: memorisation detection (technically feasible, as demonstrated by Carlini et al. 2023 and illustrated in the first-instance decision *GEMA v. OpenAI*, currently under appeal); targeted erasure through machine unlearning (currently unproven at scale, as existing gradient-based and model-editing techniques cannot reliably verify that a specific author’s work has been fully removed); and full retraining as the more effective but resource-intensive alternative. The chapter concludes that the most practically enforceable *post factum* remedy is not machine unlearning or immediate model destruction, but an injunction: prohibiting the use and commercialisation of the contaminated model until a lawfully retrained version has been developed and verified. This injunctive relief mechanism is available under IPRED Article 10 and 17 U.S.C. § 502 and was applied in substance by the Munich court in *GEMA v. OpenAI* (first-instance decision, under appeal). Chapter 4.2 identifies six structural gaps in the current regulatory framework: the DSM Directive does not explicitly address contaminated datasets; the AI Act’s transparency obligations under Article 53 do not resolve copyright liability; opt-out mechanisms are globally fragmented and lack binding force; US fair use uncertainty persists pending appellate review; there is no international coordination mechanism; and existing enforcement tools are insufficient to deal with remedies at model level.

On the basis of this analysis, the thesis advances six *de lege ferenda* proposals, presented in order of institutional feasibility. First and most urgently, the AI Act (Article 53) should be supplemented with a due diligence obligation requiring GPAI providers to document the lawful origin of training data, with a proportionate audit mechanism accessible to the AI Office – achievable through implementing acts without full legislative procedure. Second, Articles 3 and 4 of the DSM Directive should be amended to include an explicit clause stating that works are not “lawfully accessible” where access was obtained through copyright infringement, thereby codifying the ACI Adam analogy and eliminating reliance on interpretive derivation. Third, the IPRED Directive (2004/48/EU) should be amended to introduce a specific injunctive relief mechanism for contaminated AI models, requiring suspension of use and commercialisation until a compliant retrained version is available –

replacing any reference to algorithmic disgorgement as the primary remedy, since machine unlearning is not currently verifiable in judicial proceedings. Three further proposals are longer-term in character: establishing an EU-level mandatory collective licensing system for AI training; supplementing the Estonian Copyright Act with an explicit clause excluding datasets that contain unlawful copies; and establishing a standardised international opt-out mechanism under WIPO auspices. Of the six proposals, the first three are institutionally realistic in the near term, as elaborated in Chapter 4.3.

Under EU law, the doctrinal analysis advanced in this thesis leads to the conclusion that training generative AI models on pirated data falls outside the scope of the DSM TDM exceptions and therefore constitutes copyright infringement, although authoritative confirmation from the CJEU remains pending in case C-250/25. Under US law, the doctrinal analysis points in the same direction, even if the only final district court ruling to date (*Kadrey v. Meta*, Chhabria J., 2025) reached the opposite result. The two systems differ in how they get there. EU law is structured more rigidly around the threshold requirement of lawful access: works are either lawfully accessible or they are not, and this thesis argues that pirated sources do not satisfy that requirement. US law is less rigid – the unlawful origin of training data is a significant factor against fair use, but not an automatic disqualifier. That flexibility is what allowed Meta to prevail in *Kadrey*. Still, the separation doctrine articulated in *Bartz v. Anthropic* is doctrinally more coherent, it treats the act of downloading from a pirate source as a standalone infringement, consistent with the Supreme Court's Warhol standard and with copyright's function as a preventive safeguard.

Effective enforcement remains substantially hampered by the difficulty of identifying contaminated models and applying appropriate remedies at model level and the most feasible *post factum* remedy is injunctive relief rather than machine unlearning. Urgent legislative action is therefore required, both at EU level and internationally, to adapt the copyright framework to the pace of generative AI development and to ensure that the rights of creators are meaningfully protected in the age of large-scale machine learning.

Keywords: *generative artificial intelligence, copyright, text and data mining, piracy, fair use, DSM Directive, data training, AI Act, machine unlearning, injunctive relief, de lege ferenda.*

KASUTATUD ALLIKATE LOETELU

KASUTATUD KIRJANDUS

1. Brown, T. jt. Language Models are Few-Shot Learners. – Advances in Neural Information Processing Systems, kd 33, 2020.
2. Cao, Y. jt. Towards Making Systems Forget with Machine Unlearning. – IEEE Symposium on Security and Privacy 2015.
3. Carlini, N. jt. Extracting Training Data from Large Language Models. – Proceedings of the 30th USENIX Security Symposium, 2021.
4. Carlini, N. jt. Quantifying Memorization Across Neural Language Models. – Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
5. Elkin-Koren, N.; Hacohen, G. Cognitive Copyright. – Yale Law Journal, 133, nr 1, 2023.
6. Euroopa Autoriõiguse Ühing. Comment of the European Copyright Society on the Request for Preliminary Ruling in Case C-250/25 (Like Company). – JIPITEC, kd 16, 2025. – https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5166831 (20.03.2026).
7. Euroopa Komisjon. Template for General-Purpose AI Model Providers to Summarise Their Training Content. Brüssel: Euroopa Komisjon, 24.07.2025. – <https://digital-strategy.ec.europa.eu/en/faqs/template-general-purpose-ai-model-providers-summarise-their-training-content> (20.03.2026).
8. Euroopa Komisjon. Üldotstarbeliste tehisintellekti mudelite heade tavade koodeksi lõplik versioon, 10.07.2025. – <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai> (20.03.2026).
9. Gao, L. jt. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. arXiv:2101.00027v1 [cs.CL], 2021.
10. Geiger, C.; Frosio, G.; Bulayenko, O. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market. – IIC – International Review of Intellectual Property and Competition Law, kd 49, nr 7, 2018.

11. Henderson, P. jt. Foundation Models and Fair Use. – Journal of Machine Learning Research, 24, nr 1, 2023.
12. Husovec, M.; Quintais, J. P. How to License Article 3 TDM Research Exception? A Baseline Standard and Way Forward. – GRUR International, kd 70, 2021.
13. Ilin, I.; Kelli, A. Natural Language, Legal Hurdles: Navigating the Complexities in Natural Language Processing Development and Application. – Journal of the University of Latvia. Law, kd 17, 2024.
14. Kelli, A. jt. The Impact of Copyright and Personal Data Laws on the Creation and Use of Models for Language Technologies. – Selected Papers from the CLARIN Annual Conference 2019. Linköping: Linköping University Electronic Press, 2020.
15. Kelli, A. jt. Building a Chatbot: Challenges under Copyright and Data Protection Law. – Ebers, M. jt (toim), Contracting and Contract Law in the Age of Artificial Intelligence. Oxford: Hart Publishing, 2022.
16. Kelli, A. jt. Eesti autoriõiguse reform. – Juridica, 2021, nr 4.
17. Kelli, A. jt. Keeleandmete õigusliku režiimi mõju nende abil loodud keelemudelitele. – Eesti Rakenduslingvistika Ühingu aastaraamat, kd 16, 2020. DOI: 10.5128/ERYa16.04.
18. Kelli, A.; Välimäe, V.; Hoffmann, A. Legal Framework for Artificial Intelligence in Estonia. – Juridica International, 2022, nr 31.
19. Lathrop, E. Algorithmic Disgorgement: An FTC Enforcement Theory and Its Implications. – North Carolina Journal of Law & Technology, 24, nr 2, 2022.
20. Lee, J.-A.; Chang, R.-C. Out from the Shadows: Developing Effective Copyright Laws for AI Training Datasets and Shadow Libraries. – Journal of Intellectual Property Law & Practice, kd 21, nr 2, 2026.
21. Lee, K.; Cooper, A. F.; Grimmelmann, J. Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain. – Journal of the Copyright Society of the USA, kd 70, nr 3, 2024.
22. Leistner, M. EU Copyright Law: From the Package to the Directive Proposal. – IIC – International Review of Intellectual Property and Competition Law, kd 48, 2017.

23. Leistner, M.; Antoine, L. TDM and AI Training in the European Union – From ‘LAION’ to Possible Ways Ahead? – GRUR International, kd 74, nr 11, 2025.
24. Lemley, M. A.; Casey, B. Fair Learning. – Texas Law Review, 99, nr 4, 2021.
25. Lucchi, N. Generative AI and Copyright: Training, Creation, Regulation. Euroopa Parlament, JURI komitee uuring IUST_STU(2025)774095. Brüssel: Euroopa Parlament, 2025.
26. Margoni, T.; Kretschmer, M. A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership and the Future of Technology. – GRUR International, kd 71, nr 8, 2022..
27. Nguyen, T. T. jt. A Survey of Machine Unlearning. arXiv:2209.02299, 2022.
28. Peukert, A. Copyright in the Artificial Intelligence Act – A Primer. – GRUR International, kd 73, nr 6, 2024.
29. Quintais, J. P. Copyright, the AI Act and Extraterritoriality. The Lisbon Council, 2025 (SSRN: 5316132).
30. Quintais, J. P. Generative AI, Copyright and the AI Act. – Computer Law & Security Review, kd 56, 2025, artikkel 106107.
31. Quintais, J. P. The New Copyright in the Digital Single Market Directive: A Critical Commentary. – European Intellectual Property Review, 42, nr 1, 2020.
32. Quintais, J. P. jt. Licensing Copyright for AI Training: Considerations and Policy Options. SSRN Working Paper, 2024.
33. Rosati, E. Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790. Oxford: Oxford University Press, 2021.
34. Rosati, E. Is Text and Data Mining Synonymous with AI Training? – Journal of Intellectual Property Law & Practice, kd 19, nr 12, 2024.
35. Sag, M. Copyright Safety for Generative AI. – Houston Law Review, 61, nr 2, 2024.
36. Senffleben, M. Generative AI and Author Remuneration. – IIC – International Review of Intellectual Property and Competition Law, kd 54, 2023.
37. Senffleben, M.; Buijelaar, L. Robot Creativity: An Incentive-Based Neighbouring Rights Approach. – European Intellectual Property Review, 42, nr 12, 2020.

38. Sootak, J.; Siimets-Gross, H. (toim). Õigusteaduslik uurimistöö: kirjutamise ja vormistamise juhend. 4. tr. Tallinn: Juura, 2020.
39. Synodinou, T.-E.; Vrakas, G. Lawful Access as a Gatekeeper for TDM in the EU. – Verfassungsblog, 17.11.2025. DOI: 10.17176/20251117-142050-0.
40. Touvron, H. jt. LLaMA: Open and Efficient Foundation Models. arXiv:2302.13971v1 [cs.CL], 27. veebruar 2023.
41. Triaille, J.-P. jt. A European Solution for Text and Data Mining in the Research Sector. – Stockholm Intellectual Property Law Review, kd 4, nr 2, 2021.
42. USA Autoriõiguse Amet (US Copyright Office). Copyright and Artificial Intelligence, Part III: Generative AI Training. Washington, D.C.: US Copyright Office, 9. mai 2025. – <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> (20.03.2026).
43. WIPO. Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Summary (2019–). WIPO/IP/AI/GE/20/1 Rev. Genève: WIPO, 2020.
44. WIPO. Scoping Paper on Copyright and Artificial Intelligence. Genève: WIPO, 2021.

KASUTATUD ÕIGUSAKTID

Eesti õigusaktid

- 45. Autoriõiguse seadus (AutÕS). – RT I 1992, 49, 615; RT I, 12.07.2025, 9 (jõust. 01.10.2025).
- 46. Autoriõiguse seaduse muutmise seadus (autoriõiguse direktiivide ülevõtmine). – RT I, 28.12.2021, 1 (jõust. 07.01.2022).
- 47. Tsiviilseadustiku üldosa seadus. – RT I 2002, 35, 216; RT I, 22.03.2021, 8.

Euroopa Liidu õigusaktid

- 48. Euroopa Parlamendi ja nõukogu direktiiv 2001/29/EÜ, 22. mai 2001, infoühiskonna autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ühtlustamise kohta. – EÜT L 167, 22.06.2001, lk 10–19.
- 49. Euroopa Parlamendi ja nõukogu direktiiv 2004/48/EÜ, 29. aprill 2004, intellektuaalomandi õiguste jõustamise kohta (IPRED). – ELT L 157, 30.04.2004, lk 45–86.
- 50. Euroopa Parlamendi ja nõukogu direktiiv (EL) 2019/790, 17. aprill 2019, autoriõiguse ja sellega kaasnevate õiguste kohta digitaalsel ühtsel turul ning millega muudetakse direktiivid 96/9/EÜ ja 2001/29/EÜ (DSM-direktiiv). – ELT L 130, 17.05.2019, lk 92–125.
- 51. Euroopa Parlamendi ja nõukogu määrus (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus). – ELT L 119, 04.05.2016, lk 1–88.
- 52. Euroopa Parlamendi ja nõukogu määrus (EL) 2024/1689, 13. juuni 2024, tehisintellekti käsitlevate ühtlustatud eeskirjade kohta (tehisintellekti käsitlev määrus). – ELT L, 12.07.2024.

Välisriikide ja rahvusvahelise õiguse aktid

- 53. Berni konventsioon kirjandus- ja kunstiteoste kaitseks (Pariisi akt, 24. juuli 1971). Eesti ratifitseeris 26. oktoobril 1994. – RT II 1994, 16/17, 49.

54. Saksamaa autoriõiguse seadus (*Urheberrechtsgesetz – UrhG*), 09.09.1965 (muudetud kujul).
55. 17 U.S.C. § 107 (USA *fair use* erandi säte).
56. 17 U.S.C. §§ 502–504 (USA autoriõiguse kaitsevahendid: keelav ettekirjutus, hüvitis, kasumi väljamõistmine).

KASUTATUD KOHTUPRAKTIKA

Euroopa Liidu Kohtu lahendid

- 57. EKo 10.04.2014, C-435/12, *ACI Adam BV jt versus Stichting de ThuisKopie ja Stichting Onderhandeligen ThuisKopie vergoeding*, ECLI:EU:C:2014:254.
- 58. EKo 26.04.2017, C-527/15, *Stichting Brein versus Jack Frederik Wullems*, ECLI:EU:C:2017:300.
- 59. EKo 14.06.2017, C-610/15, *Stichting Brein versus Ziggo BV ja XS4ALL Internet BV*, ECLI:EU:C:2017:456.
- 60. ELK eelotsusetaotlus, C-250/25, *Like Company Kft versus Google Ireland Ltd jt*, esitatud 3. aprill 2025 (menetluses).

Saksa kohtute lahendid

- 61. OLG Hamburg (apellatsioonikohtu otsus), 10.12.2025, Az. 5 U 104/24, *Kneschke versus LAION e.V.*
- 62. LG Hamburg, 27.09.2024, Az. 310 O 227/23, *Kneschke versus LAION e.V.* (esimese astme otsus).
- 63. LG München I (esimese astme otsus; apellatsioon menetluses) 11.11.2025, Az. 42 O 14139/24, *GEMA versus OpenAI GmbH*.

USA kohtute lahendid

- 64. *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).
- 65. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).
- 66. *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021).
- 67. *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985).
- 68. *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015).
- 69. *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence, Inc.*, No. 1:20-cv-00613-SB (D. Del., 11. veebruar 2025).
- 70. *Bartz v. Anthropic PBC*, No. 3:24-cv-05417-WHA (N.D. Cal. 2025) (kokkulepe august 2025).

71. *FTC v. Everalbum, Inc.*, FTC Docket No. C-4782, Consent Order (2021).
72. *FTC v. WW International, Inc.*, FTC Docket No. C-4818, Consent Order (2022).
73. *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC (N.D. Cal., 25. juuni 2025).
74. *Perlmutter v. Blanche et al.*, No. 1:25-cv-01659 (D.D.C., 10. september 2025).
75. *Skidmore v. Swift & Co.*, 323 U.S. 134 (1944).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Narine Urban,
(*autori nimi*)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

GENERATIIVSE TEHISINTELLEKTI MUDELITE TREENIMINE
PIRAATALLIKATELT KOGUTUD ANDMETEL: AUTORIÕIGUSE
RIKKUMINE VÕI ÕIGUSTATUD KASUTUS EUROOPA LIIDU JA
AMEERIKA ÜHENDRIIKIDE ÕIGUSES?

(*lõputöö pealkiri*)

mille Aleksei Kelli
juhendaja(d) on

(*juhendaja nimi*)

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni;

3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile;
4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Narine Urban

29.04.2026