



Euroopa Õigusakadeemia töötuba: EL diskrimineerimise vastase õiguse rakendamine

DISKRIMINEERIMINE JA TEHISINTELLEKT: TEHNILISED JA ÕIGUSLIKUD ASJAOLUD

Prof. Dr. Philipp Hacker, LL.M. (Yale)
Digiühiskonna õiguse ja eetika õppetool
European New School of Digital Studies
European University Viadrina Frankfurt (Oder)



This training session is funded under the 'Rights, Equality and Citizenship Programme 2014-2020' of the European Commission.

1

Sisukord

- I osa: Diskrimineerimise allikad tehisintellektis
- II osa: Algoritmiline diskrimineerimine EL õiguses
- III osa: Algoritmiline õigus
- IV osa: Algoritmilise õigluse seaduslikud piirangud



2

I osa:
Diskrimineerimise allikad tehisintellektis



3

Ametipositsioonide valiku algoritmide

**The
Guardian**

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

Allikas: <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>



4

Tehisintellekt meditsiini valdkonnas

RESEARCH ARTICLE

ECONOMICS

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}

Allikas: Obermeyer et al., 366 Science 447 (2019)



5

Tarkade linnade dünaamika

Sisenemine avalikesse
hoonetesse



6

Tarkade linnade dünaamika

→ läbi näotuvastustehnoloogia (FRT)



ens

7

Näotuvastustehnoloogia probleemid

Proceedings of Machine Learning Research 81:1–15, 2018

Conference on Fairness, Accountability, and Transparency

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

JOYAB@MIT.EDU

Timnit Gebru
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

TIMNIT.GEBRU@MICROSOFT.COM

Editors: Sorelle A. Friedler and Christo Wilson



ens

8

Mõisted

Tehisintellekt (TI)

Definitsioonid:

Russel/Norvig,
2016, 2

Thinking Humanly

"The exciting new effort to make computers think ... *machines with minds*, in the full and literal sense." (Haugeland, 1985)

"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)

Acting Humanly

"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)

"The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)

Thinking Rationally

"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985)

"The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)

Acting Rationally

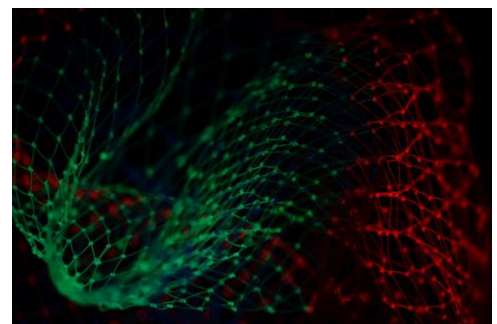
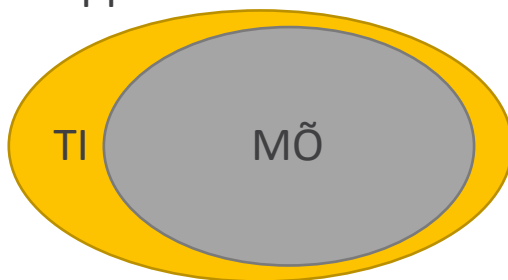
"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998)

"AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

9

Concepts

Masinõpe (MÕ): "Õppimine näidete kaudu"



Definition: A computer program is said to **learn** from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

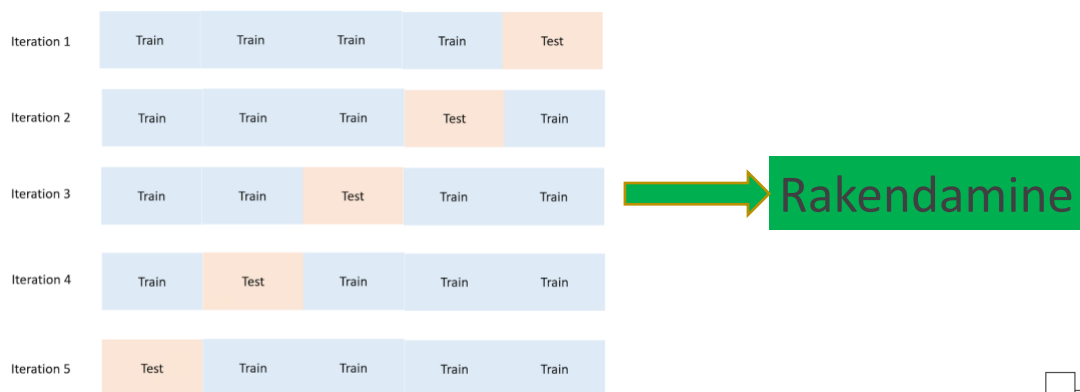
Mitchell, 1997, 2

ens

10

Juhendatud õpe

- Andmekogum: täielikult märgistatud (väljund iga sisendi jaoks)
- Treenimisandmed - testandmed



Cross-validation illustrated. The number of splits and iterations can be varied.

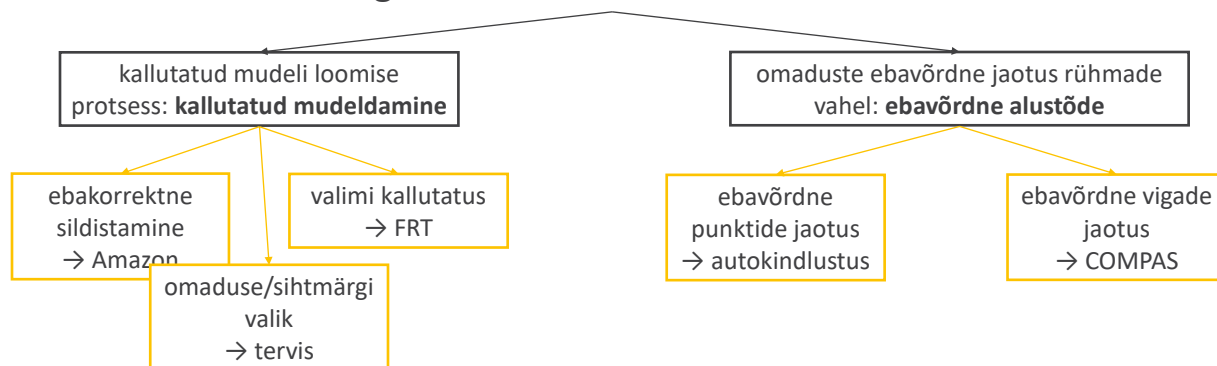
Maini/Sabri, 2017, 46



11

Algoritmilise diskrimineerimise allikad

Algoritmiline diskrimineerimine



12

II osa: Algoritmiline diskrimineerimine EL õiguses



13

Algoritmilise diskrimineerimise õigusraamistik

Diskrimineerimisvastane õigus (Hacker, 2018; vrd. ka Zuiderveen Borgesius, 2020; Wachter jt., 2020 & 2021)

- Hõlmab algoritmilist diskrimineerimist
- Kaudne diskrimineerimine: näiliselt neutraalne praktika seab kaitstud rühma konkreetselt ebasoodsasse olukorda
- Põhjendus: seaduslik eesmärk ja diskrimineeriva praktika proportsionaalsus
 - Kallutatud mudeldamine → (-)
 - Ebavõrdne alustõde → (+) kui alustõde läheneb tegelikkusele
- Probleem diskrimineerimise allikate eristamisel
- Jõustamisprobleemid: erineva kohtlemise tõestamine:
 - puudub ligipääs andmetele ja mudelile (Euroopa Kohus, C-415/10, *Meister*)



14

Algoritmilise diskrimineerimise õigusraamistik

Andmekaitseseadus (Hacker, 2018)

- Andmekaitse põhimõtted: *põhjendamata* diskrimineerimine kui ebaaus andmetöötlamine, art. 5(1)(a) GDPR (Artikkel 29 tööühm, 2018)
- Art. 22(3) GDPR: automaatne otsustamine: kallutatuse vähendamine õiguste ja vabaduste kaitsmiseks vajalike meetmetena
- Avalik jõustamine
 - trahvid, art. 83 GDPR
 - algoritmilised auditid, art. 58(1)(b) GDPR



15

Olemasolev kohtupraktika



16



17

Emaduse kaasus - faktid

R (The Motherhood Plan) vs. Her Majesty's Treasury [2021]
 EWHC 309 (Admin) (17 veebruar 2021)

- 2020: automaatne otsustussüsteem Ühendkuningriigis määramaks füüsilisest isikust ettevõtjale antava pandeemiatoetuse taset

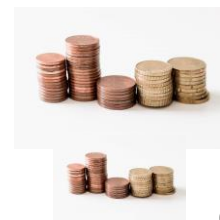
- Ühtegi inimest ahelas pole

- 80% eelmise 3 aasta keskmisest kasumist
- Probleem: samuti kui osaliselt emapuhkusel

Mees

Naine

Emadus



ens

18

Emaduse kaasus - diskrimineerimine?

Art. 14 ECHR + ÜK võrdõiguslikkuse seadus (avaliku sektori võrdsuskohustus)

Inglismaa ja Walesi kõrgema astme kohus (EWHC):
ei ole ei otsest ega kaudset diskrimineerimist:

■ “ebasoodne olukord [...] tuleneb isiku sissetulekute puudumisest või vähenemisest minevikus” (para. 62)

→ “ebasoodast olukorda **ei põhjustanud** [automaatne otsustussüsteem]” (para. 67)



19

Emaduse kaasus – diskrimineerimine?

Üsna **ebaveenev** (vrd. Allen/Masters, 2021a)

■ **Ebasoodne olukord:** jah

■ Mineviku sissetulek langes otseselt **emaduse tõttu**

■ Andmed faktiliselt korrektsed, kuid valesti võetud valim (vrd. **kallutatud valim**)

■ Vajalik naiste ebasoodne olukord

■ Probleem automaatse otsustussüsteemi **sisene**, mitte väline

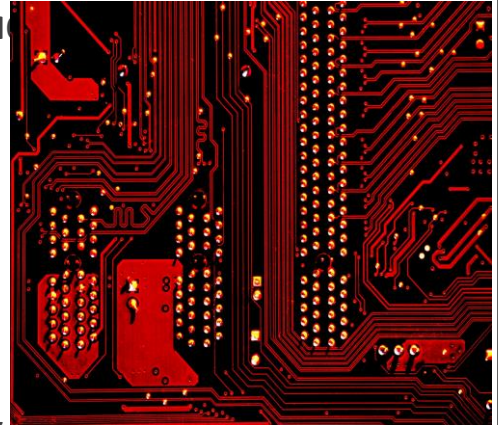


20

Emaduse kaasus – põhjendus

EWHC: automaatse otsustus-süsteemi kasutamine **põhjendatud** (mitte ilmselgelt ilma mõistliku aluseta)

- Kiirem
- Odavam
- Lihtsam
- Esineb vähem pettusi kui inimesi kasutava süsteemi puhul (para. 77-85)



ens 

21

Emaduse kaasus - põhjendus

Veenev? **Kindlasti mitte**

- automaatne otsustussüsteem on alati kiirem, odavam, „lihtsam“ ja otsekohesem
- Ei hõlma raskeid juhtumeid
- Ei saa olla õige meede

Pigem:

- **Kallutatud mudeldamise** juhtum (kaasa arvatud emapuhkus)

→ Üldiselt **mitte põhjendatud**

ens 

22

Emaduse kaasus - põhjendus

Emadusplaanis:

- Kiirus ja maksumus: automaatse otsustussüsteemi sobivus & vajadus (+)
- Kuid puudub **õiglane tasakaal** (vrd. Campbell, 2021, p. 1217; Hacker, 2018)
 - Andmestik puudulik/kallutatud
 - TI operaator: **mõistlikud pingutused saamaks tasakaalustatumaid andmeid**
 - Siin: lihtne lahendus:
 - **Isikuline enesetuvastamine**: emaduspuhkus
 - **Inimese poolne ülevaatus** ja varasemate kasumite käsitsi arvutamine
 - Mõistlikkus: rakendatakse abikõlblikkuse, mitte maksetaseme kriteeriumi
- **Põhjendamata diskrimineerimine**



23



24

Deliveroo Itaalia kaasus - faktid

- “Aus”: teenindusaegade jaotamise süsteem, mis eelistab kõrgete punktisummadega inimesi

Arvutamine: varasemate tulemuste põhjal

- Broneeritud, kuid **vastamata teenindusaegade** arv
- Osalemine **tippnõudluse** kellaajal, reede õhtul kl 20-22

→ Mõlemaid tingimusi raske saavutada pere ja haigete laste kõrvalt



ens

25

Deliveroo Itaalia kaasus - faktid

Bologna töökohus, juhtum 2949/2019:

- Kaudne **diskrimineerimine** (puue, ametiühingu tegevus= streik; naised)

■ Põhjendamata:

- Ei tunnista raskeid juhtumeid = head põhjused puudumise kohta
 - Lapse haigus, puue, streik



ens

26

Lahendamata probleemid

Jõustamine

- Teadmine, et *prima facie* diskrimineerimine eksisteerib ja selle tõendamine



- Ex-ante* ennetus *ex-post* vastutuse asemel



ens

27

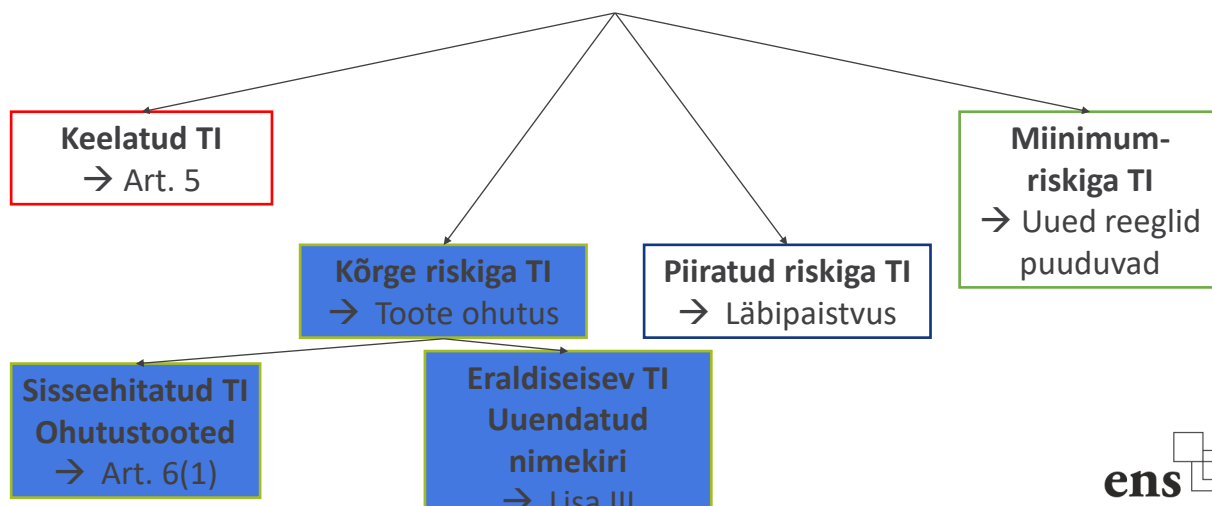
Kas uus Euroopa Liidu tehisintellekti seadusandlus muudab midagi?

ens

28

Välja pakutud TI seaduse struktuur (AIA)

Neli riskitaset



Kõrge riskiga TI süsteemid

Hõlmavad (Lisad II, III AIA):

- näotuvastustehnoloogia
 - tööhõive
 - meditsiini valdkonna TI
 - krediitpunktide saamine
 - sotsiaaltoetused
 - kohtusüsteem
- Meie juhtumid

Ei hõlma:

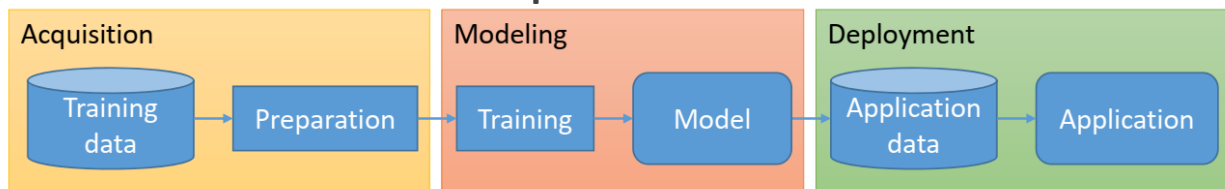
- E-kaubandus
 - Otsingumootorid
- Digitaalsete turgude seadus (DMA)

Diskrimineerimise keelamine kõrge riskiga TI süsteemides

Masinõppe liin

• Art. 11/13 AIA: lähbipaistvus

• Art. 14 AIA: inimene ahelas



• Art. 10 AIA: treenimisandmete režiim

- Korrektsus
- Esinduslikkus

• Art. 15 AIA: jõudlus

- Täpsus
- Kallutatud tagasiside leevendamine

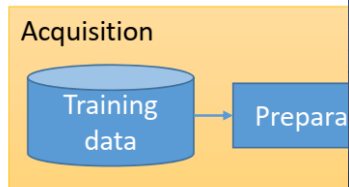
*Hacker, A Legal Framework
for AI Training Data (2021)*



31

Diskrimineerimise keelamine kõrge riskiga TI süsteemides

ML pipeline:



• Art. 10: training d regime

- Correctness
- Representativeness

In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation > The bot learned language from people on Twitter—but it also learned values

BY OSCAR SCHWARTZ | 25 NOV 2019 | 4 MIN READ

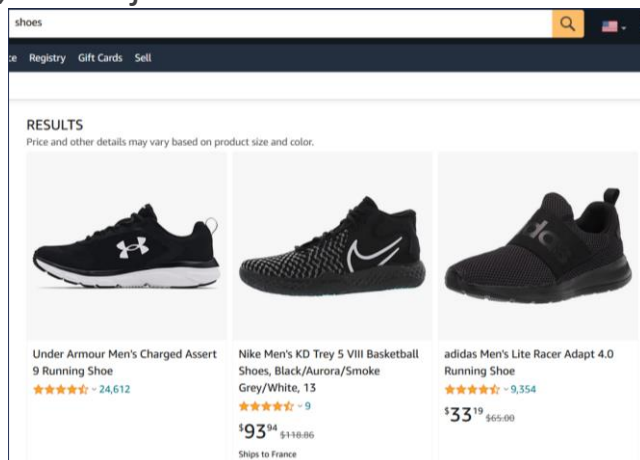


<https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation>

32

Digitaalsete turgude seadus (DMA)

- Teabekorraldajatele (Google, Amazon, Meta)
- Art. 6(5) DMA: läbipaistvad, ausad ja diskrimineerimist vältivad järjestused
- Põhjendus vajalik
- **ND** e-kaubanduses



33

Lahendamata probleemid

- Jõustamine
 - Teadmine, et *prima facie* diskrimineerimine eksisteerib ja selle tõendamine
- Ex-ante ennetus ex-post vastutuse asemel
- Ikka veel problemaatiline
 - Isikuline enesetuvastamine AIA-s
 - Piirang kõrge riskiga tehisintellektile ja teabekorraldajatele
- Parema AIA-s
 - Kuid sõltub operaatori heast tahtest ja tõhusatest heidutusmeetmetest



34

III osa: Algoritmiline õigus



35

Algoritmiline õigus

Õigluse definitsioonid: 2 peamist gruppi (Dwork, 2012;
Friedler jt., 2016; Pessach/Shmueli, 2020)

- Individuaalne õigus: sarnane sisend → sarnane väljund
 - Aristoteles, „Nikomachose eetika“, V raamat, § 3, 1131a10
 - Euroopa Kohus: võrdsus seaduse ees, art. 20 põhiõiguste harta
- Grupiõigus: st. iga rühma jaoks sama positiivne valikumäär (statistiline võrdsus)
 - Tulemus-võrdõiguslik kontseptsioon (Binns, 2018)
 - Kaudse diskrimineerimise võimatus
- Kompromiss vajalik: rohkem grupiõigust ↔ vähem individuaalset õigust



36

Lõhe ületamine: meie mudel

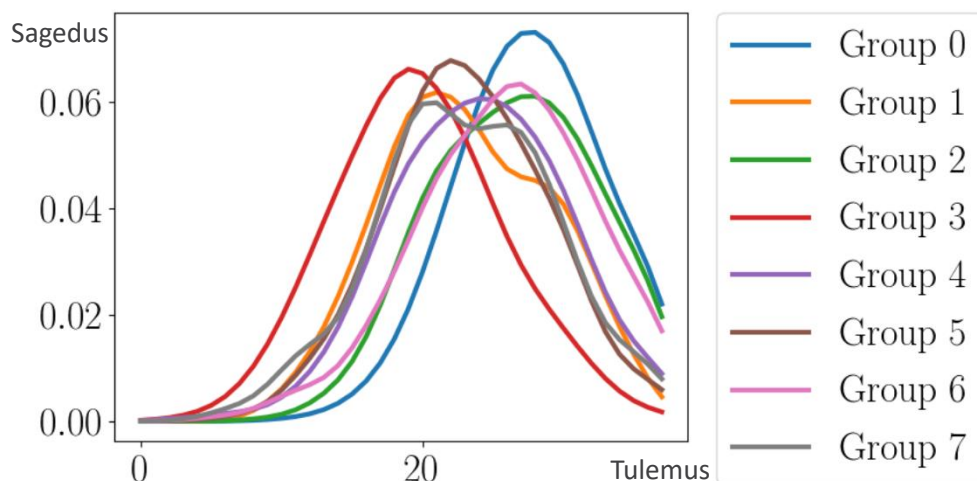
Zehlike/Hacker/Wiedemann, Koodi ja seaduse sobitamine, 34
Andmete kaevandamine ja teadmiste avastamine 2020, 163:

- Pidev laveerimine individuaalse ja grupiõiguse mõõdete vahel
- Parameeter $\theta \in [0;1]$: grupidistributsiooni umbkaudse hinnangu aste
- $\theta = 0 \rightarrow$ individuaalsed erinevused täielikult säilitatud (IF)
 - $\theta = 1 \rightarrow$ grupidistributsioonid täielikult kaardistatud peritsentril (GF)
- Minimaalne informatsiooni kadu otsustaja jaoks



37

LSAT tulemused (rahvus): kirjeldavad andmed

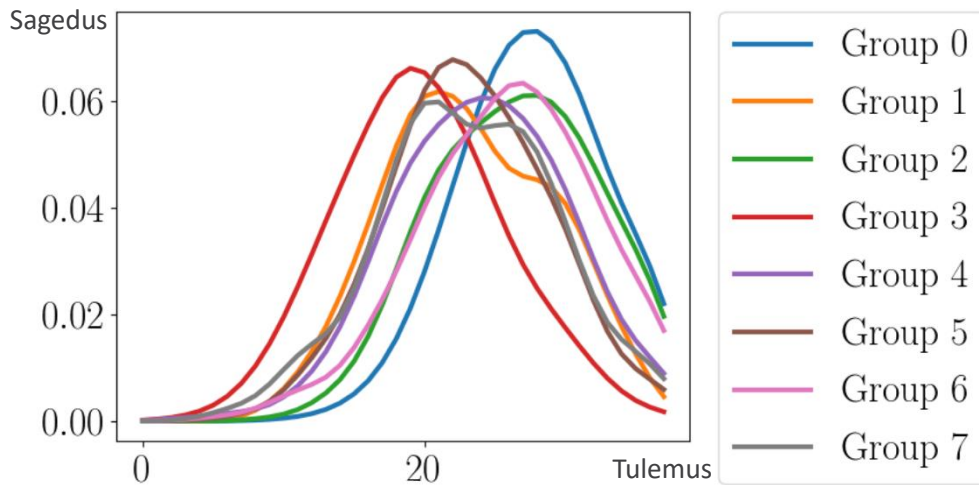


Allikas: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163



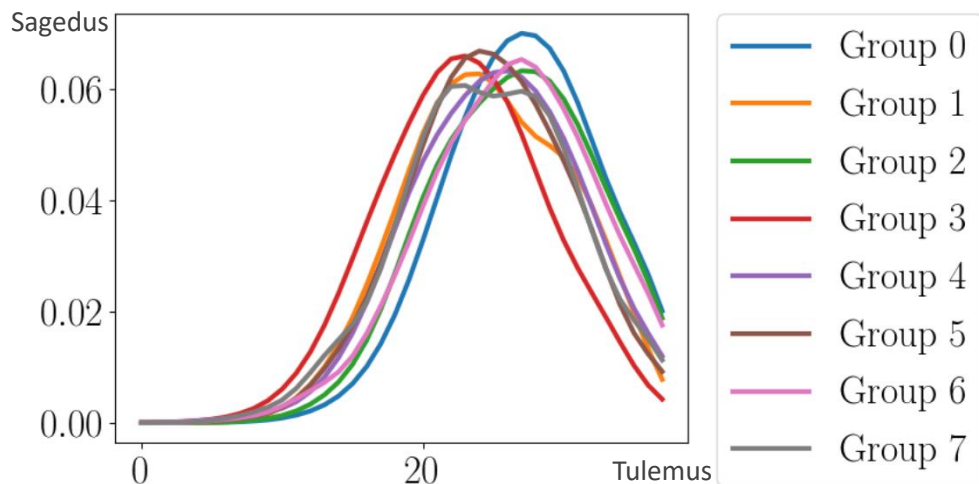
38

LSAT tulemused (rahvus) $\theta = 0$



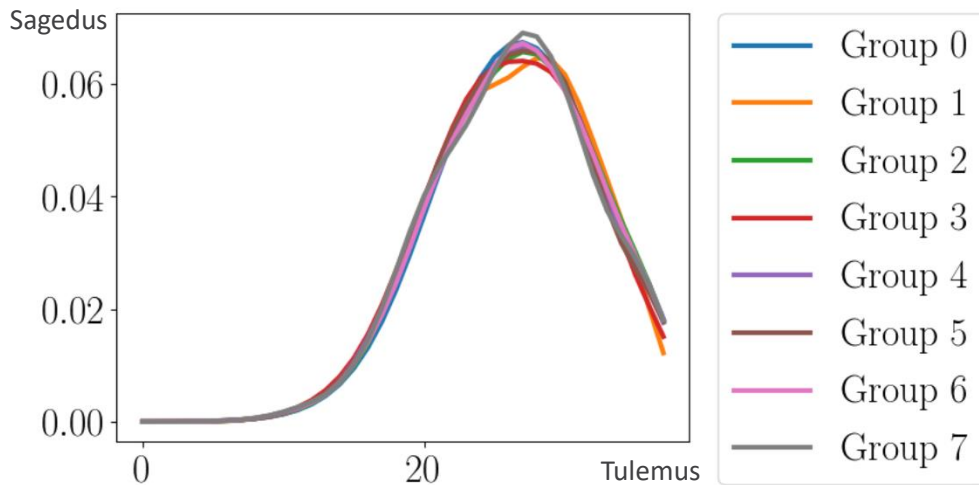
Allikas: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

LSAT tulemused (rahvus) $\theta = 0.5$



Allikas: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

LSAT tulemused (rahvus) $\theta = 1$



Allikas: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

Praegune laiendus

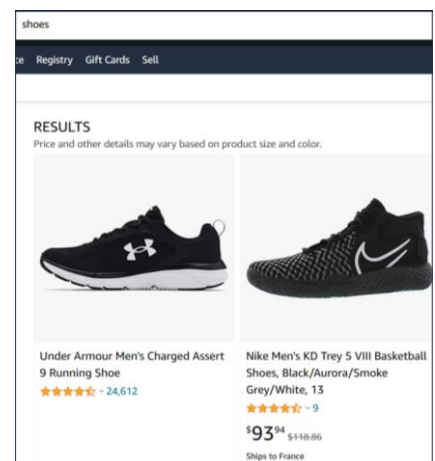
Koos Zalandoga (Saksa Amazon):

■ Kompromiss 3 õigluse mõõtme osas:

- Kalibreerimine
- Negatiivse klassi tasakaalustamine (valepositiivsed)
- Positiivse klassi tasakaalustamine (valenegatiivsed)

■ COMPAS

■ E-kaubanduse tooted (Art. 6(5) DMA: õiglane paremusjärjestus, vt. Hacker, KI ja DMA [= TI ja DMA], GRUR 2022, 1278)



IV osa: Algoritmilise õigluse seaduslikud piirangud



43

Algoritmilise õigluse seaduslikkus

Õiguslikud piirangud

Tundlike grupiandmete
töötlemine

Art. 9(1) GDPR
→ ebaseaduslik?
Nüüd: Art. 10(5),
tehisintellekti seadus

Algoritmiline positiivne
diskrimineerimine

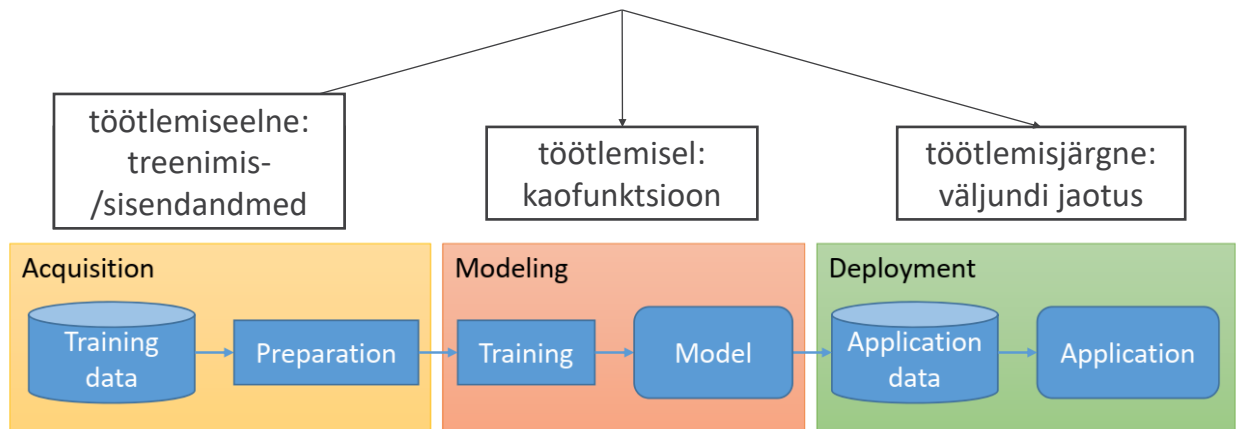
Positiivse tegevuse
doktriin
→ piirangud



44

Algoritmilise õigluse protseduurid

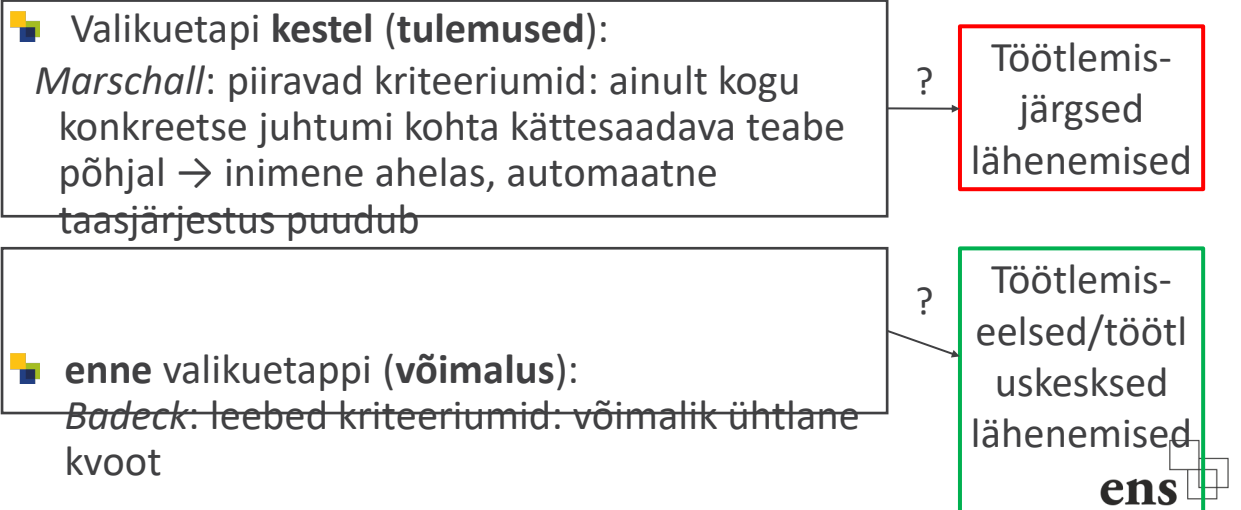
Algoritmilise õigluse tüübid



45

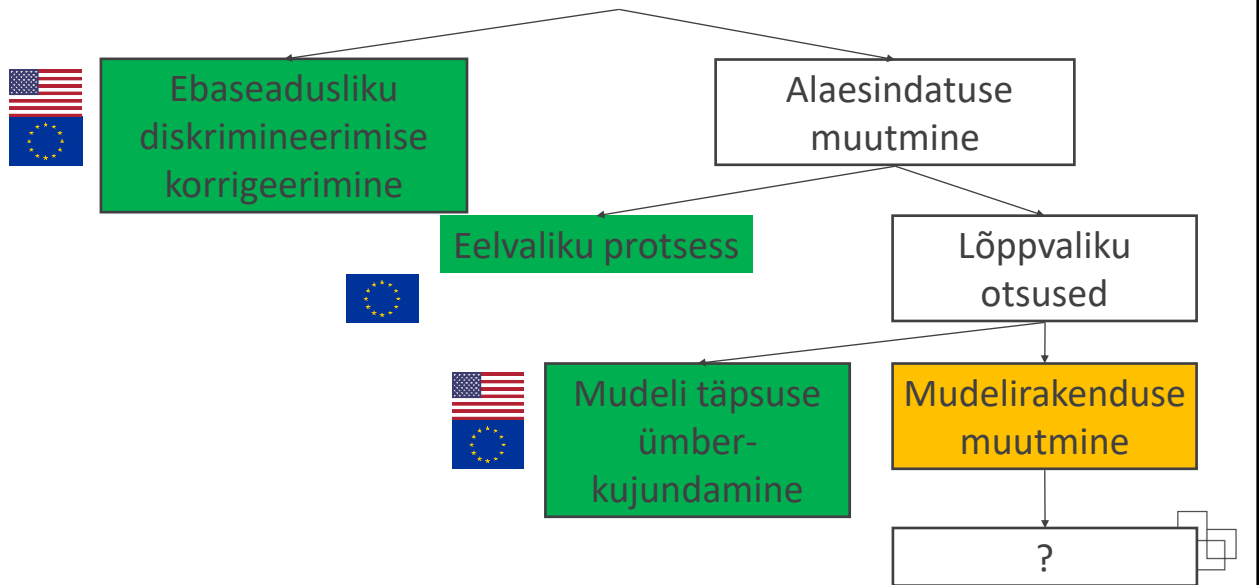
Algoritmiline positiivne diskrimineerimine

Euroopa Kohtu juhised:



46

Algoritmilise õiguse seaduslikkus algoritmilises keeles



47

Seaduslikud õiglust taastavate sekkumiste kriteeriumid

Konkreetses operatsioonimudeli muutmine

Näide: rakendusandmete algoritmiline muutmine (eeltöötlus) või taotleja skooride algoritmiline muutmine (järeltöötlus)

■ Individuaalne kahju → *Marschall*

■ Lubatavus: õiguste ja huvide tasakaal



48

Seaduslikud õiglust taastavate sekkumiste kriteeriumid

Mudeli muutmine valiku käigus alaesindatud rühma kasuks on õigustatud:

- Raske mõõta väärtusi gruppide vahel (nt. kallutatud alustõde)
→ Individuaalne õiglus ei toimi
- Grupipõhine **sisuline** võrdsus > väärtuspõhine **formaalne** individuaalne võrdsus
→ Põhioskused hilisemaks konkurentsiks väärtuste põhjal
nt. põhiharidus; ainult erandlikult: punktide kogumine tööle kandideerimisel
- Aga: üldiselt vajalik, et inimene vaataks üle
- “Sisemine võimaluste võrdsus”



Kokkuvõte

- Olemasolev kohtupraktika tehisintellekti ja diskrimineerimise alal
- Jõustamise pudelikael
- Mitmed algoritmilised õiglusindikaatorid arvutiteaduses
- Meie algoritmiline mudel:
 - Ületab lõhe individuaalse- ja grupiõigluse vahel
 - Samas minimeerib informatsiooni kadu
- Õigluse lõhe replikatsioon positiivse diskrimineerimise õiguses
 - Algoritmilise õigluse seaduslikud piirangud
 - Inimeste ja masinate meeskonnatöö stiimulid alaesindatuse leevendamiseks

Aitäh!

hacker@europa-uni.de



51

Philipp Hacker'i publikatsioonid

■ **Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law**, 55 Common Market Law Review 1143-1186 (2018), avalik ligipääs:

<https://ssrn.com/abstract=3164973>

■ **Matching Code and Law: Achieving Algorithmic Fairness with Optimal Transport**, 34 Data Mining and Knowledge Discovery 163-200 (2020) (Meike Zehlike, Philipp Hacker and Emil Wiedemann), avalik ligipääs:

<https://arxiv.org/abs/1712.07924>

■ **A Legal Framework for AI Training Data**, 13 Law, Innovation and Technology 257-301 (2021), avalik ligipääs:

<https://doi.org/10.1080/17579961.2021.1977219>

■ **KI und DMA – Zugang, Transparenz und Fairness für KI-Modelle in der digitalen Wirtschaft**, GRUR 2022, 1278-1285 (ingliskeeelne ettevalmistamisel)

52

Lisaallikad

- Allen, R. & Masters, D. (2021a). The pandemic, social benefits, and automated decision making (ADM): Just because it is quicker to use a machine, is it consistent with the principle of non-discrimination?, *AI Law Hub Blog* (19 April 2021), <https://ai-lawhub.com/2021/04/19/the-pandemic-social-benefits-and-automated-decision-making-adm-just-because-it-is-quicker-to-use-a-machine-is-it-consistent-with-the-principle-of-non-discrimination/>.
- Allen, R. & Masters, D. (2021b). An Italian lesson for Deliveroo: Computer programmes do not always think of everything!, *AI Law Hub Blog* (18 January 2021), <https://ai-lawhub.com/2021/01/18/an-italian-lesson-for-deliveroo-computer-programmes-do-not-always-think-of-everything/>.
- Article 29 Data Protection Working Party (2018). Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, WP251rev.01, <https://ec.europa.eu/newsroom/article29/redirection/document/49826>.

53

Lisaallikad

- Campbell, M. (2021). The austerity of lone motherhood: discrimination law and benefit reform. *Oxford Journal of Legal Studies*, 41(4), 1197-1226.
- Maini, V., & Sabri, S. (2017). Machine learning for humans. https://www.dropbox.com/s/e38nil1dnl7481q/machine_learning.pdf?dl=0.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Russell, S. J. (2010). *Artificial Intelligence: A Modern Approach*. 3rd. ed. Pearson Education.
- Wachter, S., Mittelstadt, B., & Russell, C. (2020). Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. *W. Va. L. Rev.*, 123, 735.

54

Lisaallikad

- Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41, 105567.
- Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, 24(10), 1572-1593.