

Aarhus Universitet – ARTS – Informationsvidenskab, Digitalt liv

Etiske Retningslinjer for Dansk Sprogteknologi

Ethical guidelines for Danish language technology

Studerende: Toke Jøns Mulvad

Vejleder: Anders Albrechtslund

Studienummer: 201908926

Anslag: 158.342

Afleveringsfrist: 2024-06-03

Opgaven må offentliggøres: Ja

Abstract

This paper examines the Danish and European ethical regulation of artificial intelligence (AI) to achieve a better understanding of the regulation of language-based AI in Denmark and what influence ethical regulations have on Danish language models. The subject is approached with the research question: *What ethical implications do the regulations have on the development of Danish language technology based on artificial intelligence in Denmark?* To answer this primary question three subsequent questions are put forward. (1) *What is the ethical framework for development of language models in Denmark?* (2) *Which regulations does sprogteknologi.dk fulfil with regards to the metadata on language resources and which implications does the material have for the development of Danish language models?* And (3) *What is the role of the regulations of language technology based on artificial intelligence in the development of language models in Denmark?*

The theoretical perspective is based on actor-network theory (ANT), which is a sociological approach envisioning every phenomenon as a network relation and interactions between human and non-human actors. The paper mainly consists of three parts. A qualitative analysis of five regulatory documents consisting of two national regulations from Denmark, two regulations from EU and one advisory report from the European ‘High-level expert group on artificial intelligence’. A quantitative analysis of the Danish government’s initiative, ‘sprogteknologi.dk’, based on the available metadata. And a discussion of which implications the analysed ethical regulation of artificial intelligence has on development of language models in Denmark.

Finally, the paper concludes that regulation of AI in Denmark consists of seven ethical guidelines: Autonomy, Development, Responsibility, Transparency, Robustness and Justice. Sprogteknologi.dk complies with all the guidelines and contains a diverse collection of language resources and does not only support the development of monolingual language models but multilingual language models. The regulation of language models is a complex subject matter that should be approached based on a nuanced and informed foundation, due to the many actants in the network that language models exist in and consist of.

Forord

Først og fremmest vil jeg gerne takke min vejleder, Anders Albrechtslund, for den kyndige og motiverende vejledning under hele specialeforløbet.

Jeg skylder også at sige tak til Carl Frederik Bach Kirchmeier og Casper Blum Frohn fra sprogteknologi.dk, for at ville høre om projektet og give mig et bedre indblik i arbejdet og organisationen bag sprogteknologi.dk.

Endeligt må jeg takke min familie, venner og bekendte for at støtte mig i løbet af projektet. Uden dem havde specialet skulle undvære givende pointer, som er opstået på baggrund af inspirerende samtaler og diskussioner. Ydermere har jeg nydt fordel af nogle yderst hjælpsomme individer, der har været villige til at bruge deres tid og kræfter på at forskønne specialet.

Indhold

Abstract.....	1
Forord	2
1. Indledning	5
2. Forskningsoversigt	7
2.1. Etiske retningslinjer for KI	7
2.2. Undersøgelser af sprogmodeller	9
3. Metodeafsnit	10
3.1. Litteraturindsamling	10
3.2. Kvalitativ metode.....	11
3.3. Kvantitativ metode og dataindsamling	12
4. Teoriafsnit.....	14
4.1. Aktør-Netværks Teori	14
4.2. Etik	16
4.3. Moralsk status.....	17
4.4. KI Etik.....	19
5. Afgrænsning	20
5.1. Kunstig Intelligens, sprogmodeller og data.....	20
5.2. Relevante myndigheder	24
6. Analyse	25
6.1. Etiske retningslinjer for KI i Danmark.....	25
6.1.1. Lovgivning, rådgivning og rammesætning	26
6.2. Én etisk ramme.....	30
6.2.1. Autonomi	31
6.2.2. Udvikling	33
6.2.3. Ansvarlighed	34
6.2.4. Gennemsigtighed	37
6.2.5. Robusthed.....	40
6.2.6. Retfærdighed.....	42
6.2.7. Holisme.....	44
6.3. Sprogteknologi.dk.....	46
6.3.1. Baggrunden for sprogteknologi.dk	47
6.3.2. Ressourcernes metadata	49
6.3.3. Organisationerne bag ressourcerne	51

6.3.4.	Metadata og indhold	53
7.	Diskussion.....	58
7.1.	Sprogmodeller er ikke ”bare” teknologi.....	58
7.2.	Etiske reguleringers rolle for sprogmodeller i Danmark	63
7.2.1.	En kilde til objektivitet	63
7.2.2.	Er der overhovedet behov for regulering af KI	65
8.	Konklusion	70
	Litteraturliste	73

Etiske retningslinjer for dansk sprogteknologi

1. Indledning

Kunstig intelligens (Fra det engelske 'artificial intelligence', herefter kaldt 'KI') og sprogteknologi er siden OpenAIs offentlige lancering af ChatGPT 3.5 i 2019 blevet en del af den almene borgers digitale værktøjskasse. Som brugen af denne teknologi er blevet allemandseje, er det ligeledes blevet interessant at følge forekomsten af nye etiske udfordringer. Imens den teknologiske udvikling af KI fortsætter med at blive inkorporeret i alle mulige aspekter af vores hverdag, såsom i private og offentlige erhverv, på sociale medier og i nyhederne, er myndighederne i Danmark og EU gået i gang med en større regulatorisk indsats af området. Udover en generel markedsregulering af kunstig intelligens som produkt, har myndighederne også et ønske om, at man i EU og Danmark skal udvikle etisk KI. I tilgift har de danske myndigheder også fremlagt et ønske om at fremme udviklingen af dansksproget KI baseret i Danmark. Denne opgave har til formål at give et indblik i disse regulatoriske omstændigheder for udvikling af sprogteknologi baseret på KI i Danmark ved at besvare problemformuleringen:

Hvilke etiske implikationer har rammerne for udviklingen af dansk sprogteknologi baseret på kunstig intelligens i Danmark?

- a) Hvilke etiske retningslinjer gælder for udviklingen af sprogmodeller i Danmark?*
 - b) Hvilke retningslinjer opfylder sprogteknologi.dk med henblik på sprogressourcerne og deres tilhørende metadata, og hvad har materialet af implikationer for danske sprogmodeller?*
 - c) Hvilken rolle har de etiske reguleringer af sprogbasert kunstig intelligens for udviklingen af sprogmodeller i Danmark?*
-

Overordnet set er projektet udarbejdet på baggrund af en informationsvidenskabelig tilgang. Dertil undersøges KI ud fra en kontekstualistisk teknologiforståelse, hvormed teknologien anses som et fænomen, der udgøres af eksterne og interne processer, som består af både

materielle og sociokulturelle aspekter. Dette leder frem til, at KI bliver undersøgt ud fra en europæisk og dansk kontekst, hvorfor fundene bør blive vurderet på baggrund af denne faglige og geopolitiske afgrænsning.

Sproget gennem specialet vil være dansk i sådan en udstrækning, at der så vidt muligt vil blive anvendt danske udtryk og begreber, som ellers fremgår på engelsk i litteraturen og det empiriske materiale. Årsagen hertil er, at der er inden for det sprogteknologiske felt i Danmark er en bevægelse for at fremme brugen af dansk i udviklingen og forskningen af sprogteknologi (Kirchmeier et al., 2019, s. 44–48). Derfor anvendes det danske udtryk 'kunstig intelligens' og forkortelsen 'KI' fremfor det mere alment kendte engelske udtryk 'artificial intelligence' og den dertil hørende forkortelse 'AI'. Undervejs vil de engelske udtryk, som ikke har en officiel dansk modpart, blive formuleret i dansk oversættelse med det originale engelske udtryk præsenteret i parentes første gang udtrykkes benyttes.

For at opnå indsigt i reguleringen af KI bliver der anvendt en kvalitativ læsning af reguleringerne, hvori der findes frem til nøgleord, også kaldt koder, som danner rammen for analysen af teksterne. Til at undersøge sproressourcerne på sprogteknologi.dk udføres der en kvantitativ undersøgelse af deres metadata, hvilket gør det muligt at konkretisere fundene i visuelle fremstillinger af den indsamlede data.

I overensstemmelse med den kontekstualistiske teknologiforståelse af KI bygger projektets teoretiske perspektiv på aktør-netværks teori (herefter kaldt 'ANT'). Det betyder, at fænomener anses som netværk af aktanter, hvilket vil blive forklaret yderligere i teoriafsnittet, hvori specialets brug af 'KI etik' som teoretisk perspektiv på teknologi og etik også vil blive fremlagt.

Den inddragede empiri i projektets analyse består af fem reguleringer af KI i Danmark og EU. De danske reguleringer er Finans- og Erhvervsministeriets, "National Strategi for Kunstig Intelligens" fra 2019 og "Danmarks digitaliseringsstrategi – Ansvar for den digitale udvikling" af Digitaliserings- og Ligestillingsministeriet fra 2023. De europæiske reguleringer er loven "Europa-Parlamentets beslutning af 20. oktober 2020 med henstilling til Kommissionen om en ramme for etiske aspekter af kunstig intelligens, robotteknologi og relaterede teknologier (2020/2012(INL))", lovforslaget "Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence Act)

and amending certain Union legislative acts” og rapporten ”Ethiske Retningslinjer for Pålidelig Kunstig Intelligens”. Dertil er der indsamlet data fra sprogteknologi.dk til at undersøge initiativet med henblik på det andet underspørgsmål i problemformuleringen.

Projektet vil forløbe som følger. Først kontekstualiseres specialet i forskningskonteksten, som den skriver sig ind i, ved at præsentere toneangivende eller relevant forskning vedrørende KI etik. Derefter præsenteres metoderne, som er blevet anvendt gennem projektforløbet, med henblik på at synliggøre den fagligt begrundede tilgang til emnet. Så udfoldes det teoretiske grundlag, som danner rammen for hele projektets faglige ramme. Derefter gives en udførlig afgrænsning af de efterfølgende analyser i forhold til en fællesforståelse af de overordnede aspekter af sprogteknologi og indskrænkningen af det empiriske grundlag. Siden følger analysen af de inddragede reguleringer, hvor der laves en bred analyse af dokumenternes hensigt, efterfulgt af en udlægning af den etiske ramme på tværs af teksterne. Herfra går projektet over i den kvantitative undersøgelse af sprogteknologi.dk, som først kontekstualiseres i forhold til projektet, inden det analyseres. Endeligt diskuteres reguleringen af KI i Danmark.

2. Forskningsoversigt

De efterfølgende afsnit kontekstualiserer det nærværende projekt i det større akademiske landskab af forskning inden for KI etik.

2.1. Ethiske retningslinjer for KI

Changwu Huang, Zeqi Zhang, Bifei Mao og Xin Yao udgav i 2023 artiklen ”An Overview of Artificial Intelligence Ethics” i det digitale tidsskrift ”IEEE Transactions on Artificial Intelligence”. Undersøgelsen har til formål at give en oversigt over feltet for KI etik, med henblik på at afdække hvilke etiske problematikker og risici, der er relevante i forhold til KI. Dertil gives der også et indblik i de forskellige metodiske tilgange og evalueringsmetoder af KI etik, som er toneangivende inden for feltet. For at undersøge dette er der foretaget et multidisciplinært litteraturreview på en bred vifte af dokumenter og artikler udgivet siden 2010, omhandlende kunstig intelligens og etik. Materialet er fundet på baggrund af en

metodisk fremsøgning af litteratur ved hjælp af én søgestreng på tværs af flere forskellige databaser. Ud fra det litterære review er der udarbejdet en udførlig gennemgang af de mest repræsentative etiske emner og risici, som behandles og omtales i den fremsøgte litteratur (Huang et al., 2023, s. 801). Ydermere laves der en kvantitativ og kvalitativ undersøgelse af hvilke lande og organisationer, der har udgivet etiske retningslinjer for KI. Der bliver i alt identificeret elleve etiske principper for KI, hvoraf fem principper fremhæves som de vigtigste for reguleringen af KI: Gennemsigtighed (Transparency), Retfærdighed (Fairness and justice), Ansvarlighed (Responsibility), Ikke-skade (Nonmaleficence) og Retten til privatliv (Privacy) (Huang et al., 2023, s. 808). Endeligt konkluderes det, at KI etik ikke blot er et vigtigt felt inden for forskningssamfundet, men at det også er meget relevant og anvendeligt ude i samfundet (Huang et al., 2023, s. 816).

Senere i 2023 udkom bogen “The Ethics of Artificial Intelligence, Principles, Challenges and Opportunities”, hvori Luciano Floridi har bidraget med kapitlet ”A Unified Framework of Ethical Principles for AI”. Gennem kapitlet fremlægger Floridi argumenter for en ensretning eller sammenfatning af de etiske retningslinjer for KI. Floridi mener, at mængden og variationen af etiske retningslinjer for KI risikerer at resultere i et marked af principper, hvormed etisk regulering af KI vil miste sin effektivitet og validitet (Floridi, 2023, s. 58). Undersøgelsen af etiske retningslinjer baseres på dokumenter fra elleve organisationer, som hver består af en række etiske principper og retningslinjer for KI. Udvælgelsen af empirien er baseret på, hvorvidt organisationerne og dokumenterne opfylder fire kriterier, som bedømmer hvor stor betydning, de har for reguleringen af KI: (1) De skal være aktuelle og opdateret, (2) have indflydelse, (3) være relevante og (4) have autoritet i forhold til reguleringen af kunstig intelligens (Floridi, 2023, s. 59–60). De etiske reguleringer holdes op imod hinanden i en komparativ undersøgelse, hvor de i alt 47 identificerede principper koges ned til fem sammenfattende etiske principper for KI: Godgørenhed (Beneficence), Ikke-skade (Nonmaleficence), Autonomi (Autonomy), Retfærdighed (Justice) og Forklarlighed (Explicability) (Floridi, 2023, s. 59–60). Floridi henviser til, at de første fire etiske principper har rod i de fire kerneprincipper fra bioetikken, og at det sidste princip, forklarlighed, har kritisk betydning for det tekniske aspekt af KI (Floridi, 2023, s. 60 & 63–64).

Altså anvendes artiklen af Huang (et al.) og teksten af Floridi som en rammesætning af det metodiske og teoretiske udgangspunkt for indsamling af materiale og behandlingen af etiske retningslinjer i projektet.

2.2. Undersøgelser af sprogmodeller

Jorge Luis Morton Gutiérrez undersøger i artiklen “On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI” anvendeligheden af ANT i undersøgelsen af KI og en netværksbaseret forståelse af algoritmer (2023). Jorge viser dette gennem en undersøgelse af OpenAI’s samtaleprogram, ChatGPT 3.5, og analyserer det som et netværk af aktanter (Gutiérrez, 2023, s. 3–4). Først introducerer Jorge ANT og algoritmer i relation til KI og argumenterer for, hvorfor ANT er givende for forståelsen af sprogbasert KI som et netværk af handlinger mellem menneskelige og ikke-menneskelige aktører. Dermed understøtter ANT forståelsen for etisk og ansvarlig anvendelse af KI ved at fremstille teknologien som et produkt af socioteknologiske processer. KI teknologi udstilles derved som et komplekst system af relationer, hvor der konstant forekommer translationer mellem aktanterne. Dette understøtter også ideen om, at teknologi ikke er en neutral ting, men et artefakt som eksisterer i et netværk af konkurrerende processer (Gutiérrez, 2023, s. 11–12).

Risto Luukkonen (et al.) offentliggjorde i 2023 artiklen “FinGPT: Large generative models for a small language”. FinGPT er navnet på de finsksprogede sprogmodeller som Luukkonen (et al.) laver i forbindelse med deres arbejde med undersøgelsen af sprogmodeller i Finland. Det primære fokus for artiklen er at præsentere arbejdet bag finsksprogede sprogmodeller og en ny evalueringsmetode til at efterprøve finske sprogmodeller. I den forbindelse indgår der også overvejelser med henblik på, hvilke udfordringerne der er ved at udvikle sprogmodeller til et lille sprogområde (Luukkonen et al., 2023, s. 2710). Gennem artiklen forklares det fra hvilke databaser, der hentes sproressourcer til at træne de finsksprogede sprogmodeller. Derefter forklares det, hvordan sproressourcerne bliver behandlet, inden de anvendes som træningsdata (Luukkonen et al., 2023, s. 2711–2713). Gennem undersøgelsen udvikles der syv finske sprogmodeller, en flersproget sprogmodel og et nyt evalueringssæt til at teste finske sprogmodeller (Luukkonen et al., 2023, s. 2715 & 2718). På baggrund af deres evalueringssæt er det muligt at give indblik i sprogmodellernes uønskede kvaliteter og forskelsbehandlende

tendenser (Luukkonen et al., 2023, s. 2717–2718). Luukkonen (et al.) konkluderer, at deres sprogmodeller er bedre til at kommunikere på finsk end eksisterende finske sprogmodeller. Men deres evalueringssæt indikerer på samme tid, at der fortsat er begrænsninger og udfordringer i deres sprogmodeller. Dette forekommer på trods af, at de undervejs i udviklingen har forsøgt at modvirke tilstedeværelsen af uønskede kvaliteter og forskelsbehandling i sprogmodeller (Luukkonen et al., 2023, s. 2718).

Den første af de to artikler i dette afsnit iscenesætter projektets anvendelse af ANT som teoretisk forståelse af KI som et socioteknisk netværk af aktanter. Artiklen om FinGPT viser, at der er gode muligheder for udvikling af sprogmodeller selv inden for et lille sprogområde, men at der også er vedvarende udfordringer.

3. Metodeafsnit

3.1. Litteraturindsamling

Fremsøgningen af det litterære materiale til specialet har bestået af en systematisk søgningsproces med henblik på en hermeneutisk udvikling af søgestreng (Booth et al., 2016, s. 65). Litteratursøgningen til opgaven er hovedsageligt foregået på kb.dk, eftersom universitetsbiblioteket giver fri adgang til både enkeltstående materialer og andre databaser.

Forløbet begyndte med øvelser under det specialeforberedende forløb i efterårssemesteret 2023. På kurset blev vi, de studerende, under en kreativ øvelse bedt om at lave en liste med interessebegreber til vores specialeemne. På baggrund af listen blev der udviklet en oprindelig søgestreng: ("KI" OR "AI" OR (kunstig intell*) OR (artificial intell*)) AND (etik* OR ethic*). Ved en søgning på kb.dk med den nævnte søgestreng fås 673.304 resultater, hvilket er en meget stor mængde materiale, der inkluderer både mere og mindre relevante materialer i forhold til problemformuleringen. Årsagen til de mange og irrelevante resultater er mangel på afgrænsende interessebegreber (Flick, 2014, s. 71). En senere søgestreng bestod af flere og mere indskrænkende begreber, "constitutional" AND "responsible" AND "controllable" AND ("AI" OR (artificia* AND intelligen*)) AND ("DK" OR danis* OR denmark*) og gav 18 resultater på kb.dk. Det fremsøgte materiale blev derefter gennemgået ved at læse titler og abstracts på

de 18 resultater, hvilket ledte videre til en nærlæsning af det mest relevante materiale. På baggrund af det nærlæste materiale blev der efterfølgende udviklet endnu en søgestreng til fremsøgning af materiale til projektet. Projektets litteratur er indsamlet ved gentagne gange at gennemgå den beskrevne fremsøgning af materialer; dette inkluderer dog ikke det empiriske materiale, som vil blive forklaret under afgrænsningsafsnittet.

3.2. Kvalitativ metode

Den kvalitative metode er baseret på Sarah J. Tracys pragmatiske iterative tilgang til kvalitativ analyse af data i bogen “Qualitative Research Metodes: Collecting evidence, Crafting Analysis, Communicating Impact” (2013, s.183-202). Tracys analysemodel benyttes i dette projekt, fordi den lægger op til at finde mønstre på tværs af tekster, og den er derfor yderst velegnet til at formulere en sammenfatning af reguleringer.

Analyseforløbet involverer flere gennemlæsninger af empirien, hvor teksterne i hver omgang (*cycle*) læses med henblik på at udvikle koder på baggrund af teksterne. Koder er defineret som ord og passager, der opsummerer, er toneangivende eller essentielle for en given teksts fænomener eller udsagn med henblik på forskningskonteksten (Tracy, 2013, s. 189).

Kodningsprocessen er en kontinuerlig komparativ praksis, hvor koderne fra de individuelle tekster influerer på og influeres af andre teksters koder under analyseforløbet (Tracy, 2013, s. 190). Grundet det komparative analysearbejde er kodningsprocessen reflekterende og iterativ, eftersom koderne formes på tværs af teksterne.

Første omgang af kodning (*first-level codes*) foregår i dette projekt efter den første gennemlæsning af teksterne. I den første kodningsomgang anvendes koderne som et deskriptivt værktøj til at undersøge de umiddelbare udsagn og fænomener i teksterne (Tracy, 2013, s. 189). I dette projekt gøres det med henblik på at danne en ramme for de etiske retningslinjer for KI i Danmark, jf. problemformuleringen. På baggrund af koderne fra den første kodningsomgang dannes en kodebog, som er et analyseværktøj, der samler koderne fra teksterne, så de nemmere kan defineres og analyseres på tværs af de individuelle tekster (Tracy, 2013, s. 191). Herefter påbegyndes en anden kodningsomgang (*secondary-cycle coding*), som har til hensigt at identificere mønstre i teksterne ud fra de indsamlede koder. Efterhånden som forskellige mønstre og forbindelser identificeres under den anden

kodningsomgang, bliver koderne inddelt efter deres sammenhænge mellem koderne og i hierarkiske grupperinger, bestående af over- og underkategorier af andre koder (Tracy, 2013, s. 194–195). På baggrund af kodningsomgangene, kodebogen og systematiseringen af koderne analyseres det bearbejdede materiale med henblik på problemstillingen.

3.3. Kvantitativ metode og dataindsamling

Projektets kvantitative tilgang baseres på FAIR principperne for den digitale undersøgelse af sprogteknologi.dk. Det betyder, at datahåndteringen bygger på en grundlæggende gennemsigtighed og tilgængelighed gennem hele projektforløbet for at sikre, at undersøgelsens metoder og resultater er genskabelige (Barker et al., 2022, s. 1). FAIR-principperne omhandler både data og metadata, som bør være mulige at *finde* (*Findable*) selv efter projektet er afsluttet, få *adgang* (*Accessible*) til med hensyn til datarettigheder, at *integrere* (*Interoperability*) med andet software og hardware, og endeligt skal det også være muligt at *repetere* (*Reusable*) undersøgelsen (Wilkinson et al., 2016, s. 4). I praksis betyder det, at den anvendte data, datahåndtering og behandling af data i projektet enten bliver fremlagt direkte i teksten eller som et vedhæftet bilag.

De kvantitative undersøgelser er foretaget på en computer med Windows 11 som operativsystem. Det betyder, at nogle formuleringer i datamanipuleringen kan variere, hvis de udføres på en computer med et andet operativsystem.

Datamanipulering og visualisering er udført i softwareprogrammet R (Version 4.2.1) gennem brugerfladen RStudio (2023.06.1). R og RStudio er frit tilgængelige og åbne projekter (open source), som kan hentes ned og bruges på de fleste computere, da de både er understøttet af Windows, MacOS og Linux (The R Foundation, 2024a). R er et programmeringssprog og -miljø, der er udviklet med henblik på statistisk og grafisk computerdrevne undersøgelser. RStudio er et integreret udviklingsmiljø (*Integrated Development Environment (IDE)*), hvilket primært har betydning for at gøre programmeringsmiljøet mere brugervenligt gennem den visuelle brugerflade (Posit Software, 2024). Brugerfladen består af en række funktioner i R med hensyn til at opstille forskellige vinduer med information om f.eks. filer, data og andre hjælpefunktioner. R indeholder som programmeringssprog en række grundlæggende funktioner og metoder til manipulering og visualisering. Men i kraft af at teknologien er et

åbent projekt, hvilket betyder, at alle har adgang til koden, som projektet er opbygget af, findes der mange videreudviklinger i form af brugerudviklede funktionspakker (*packages*). Disse funktionspakker bliver distribueret, opbevaret og vedligeholdt ved hjælp af CRAN, som er et netværk af computere og protokoller (The R Foundation, 2024b). En nævneværdig funktionspakke, som benyttes til projektets undersøgelse, er pakken 'tidyverse', som indeholder en lang række funktioner med henblik på at gøre data klar til undersøgelse (*data cleaning*) og datamanipulering. Tidyverse er udviklet på baggrund af principperne om ordentlige data (*tidy data*). Disse principper er defineret i artiklen "Tidy data", af Hadley Wickham, som en standardisering af struktur og semantik vedrørende datasæt. Det betyder konkret, at data opstilles i skemaer, hvor hver variabel bør være en ny kolonne, hver observation bør være en ny række, og hver kolonne bør bestå af en type information (Wickham, 2014, s. 2–5).

Til indhentning af den data, der behandles i R, er sprogressourcernes metadata på "sprogteknologi.dk/group/corpora" blevet gennemgået manuelt i perioden 2024-04-14 til 2024-04-20. Derved er initiativets formidling af sprogressourcer blevet undersøgt på baggrund af de præsenterede metadata på hjemmesiden. Den indsamlede data fra sprogteknologi.dk består af: Ressourcens navn; Organisationen bag ressourcen; Typen af ressource; Sproget; Hvilke(n) filtype som ressourcen kan hentes i; Hvilken periode ressourcen stammer fra; Nøgleord om hvad ressourcen består af; Emneord om hvad ressourcen kan anvendes til; URL-link til ressourcens hjemmeside; Ressourcens licens; Dokumentation til at vejlede brugeren i anvendelse af ressourcen. Alle de overstående informationer for hver af ressourcerne på sprogteknologi.dk er blevet indskrevet i et excel-dokument i overensstemmelse med FAIR principperne for god datapraksis (Wickham, 2014, s. 4–5). Endeligt er datasættet blevet gemt som en csv-fil (komma-separerede værdier) i R-projektets arbejdsmappe.

4. Teoriafsnit

4.1. Aktør-Netværks Teori

Aktør-netværks teori (ANT) er en sociologisk tilgang til at undersøge fænomener som sociale netværk. Grundlaget for ANT er udtænkt af Michel Callon og Bruno Latour i teksten “Unscrewing the big Leviathan” (Callon & Latour, 1981). Teorien har rod i Thomas Hobbes bog fra 1651, “Leviathan”, om samfundets struktur og reguleringer, som er baseret på sociale kontrakter mellem mennesker (Hobbes, 2018, s. 24–26). Hobbes monster (Leviathan) er en analogi for samfundet, som er en konstruktion, der kun eksisterer i kraft af de aktører, som det udspringer af. Altså er samfundet et konstrueret uhyre, som består af en ansamling af aktører (Callon & Latour, 1981, s. 278). Aktører kan inddeles i makro- og mikroaktører. Eksempler på makroaktører er institutioner, organisationer, samfundsgrupper og stater, og mikroaktører kunne være individer, familier og mindre sociale grupper. Det er dog vigtigt at påpege, at Hobbes, Callon, og Latour ikke mener, at man a priori kan opdele aktører i de ovennævnte kategorier. Derfor må aktører som udgangspunkt anses som isomorfiske, hvilket betyder, at de er af samme form og struktur uanset størrelse. Inden man kan begynde at lave antagelser om en aktør må man derfor undersøge interaktionerne mellem aktørerne, som konstruerer dem (Callon & Latour, 1981, s. 279). Som konsekvens af dette må selv den individuelle aktør anses som et netværk. Derfor er det også mere givende for forståelsen af et netværk at fokusere på aktørernes handlinger og de processer, de indgår i. For selvom man ikke kan sige noget isoleret om den enkelte aktør, så kan man stadig iagttage og undersøge deres handlinger (Callon & Latour, 1981, s. 285–286). Disse processer kan være mere eller mindre betydende for netværket, de indgår i, hvilket kan defineres ud fra styrken af en proces. Styrke kan vurderes ud fra, i hvilken grad en proces kan lave indgreb (intervene), afbryde (interrupt), fortolke (interpret) eller have interesser i (interest) andre processer. Ydermere kan styrke vurderes på baggrund af, i hvilken grad processen kan skabe et asymmetrisk forhold mellem andre processer i netværket (Callon & Latour, 1981, s. 292). Denne magtkamp mellem processerne sker vel og mærke på samme tid og bliver derfor hele tiden udøvet og forandret af andre processer (Callon & Latour, 1981, s. 297).

Altså anses sociale netværk som konstruerede uhyrer bestående af isomorfe aktører. Disse aktører består i sig selv af konkurrerende processer, som er i konstant udarbejdelse af asymmetriske forhold. De asymmetriske forhold er udtrykt i regler for processer i et netværk og for, hvordan netværket i sidste ende udøver sit virke. Hvis disse regler kan blive så faste, at der ikke længere er tale om en konkurrence, men en alment gældende regel i netværket, kaldes dette en sort kasse (blackbox) (Callon & Latour, 1981, s. 285). En sort kasse er udtrykt i form af underliggende handlemønstre eller usagte normer og regler.

Med hensyn til anvendelsen af ANT i denne opgave er det også relevant at omtale anvendelsen af aktør som deltagende entiteter i sociale netværk. Latour argumenterer i en senere artikel for anvendelsen af begrebet 'aktant' (actant) frem for de menneskecentrerede betegnelser som 'aktør' og 'agent'. Årsagen hertil er, at 'aktant' som udtryk er mere elastisk end aktør, altså at det er mere anvendeligt til at beskrive alle entiteter, der indgår i ethvert netværk. Aktant omfavner både menneskelige og ikke-menneskelige, og en enkelt aktant kan have mange forskellige aktantielle roller (Latour, 1994, s. 33). Ved at benytte aktant bliver ikke-menneskelige aktører også anerkendt og aktivt inddraget som medvirkende i samfundet.

Latour beskriver gennem artiklen "On technical mediation – Philosophy, Sociology, Genealogy", hvordan den binære opdeling mellem menneskelige og ikke-menneske aktører ikke er retvisende for, hvordan samfundet udvikler sig i praksis. Derfor kan samfundet ikke blot betegnes som et socialt netværk, men et socioteknisk netværk af aktanter (Latour, 1994, s. 64). Denne ligestilling af menneskelige og ikke-menneskelige aktører er senere blevet diskuteret af Roberto Redaelli i artiklen "Different approaches to the moral status of AI: a comparative analysis of paradigmatic trends in Science and Technology Studies" (Redaelli, 2023, s. 5). Redaelli fremlægger argumentet om, at nedbrydningen mellem af barrieren mellem menneskelige og ikke-menneskelige aktører risikere at undergrave menneskets ansvarlighed over for sig selv og andre mennesker, uafhængigt af teknologi (Redaelli, 2023, s. 3–4). Herimod kan det dog argumenteres, at Latour grundlæggende anser, at beskrivelsen af mennesket som et socialt væsen er misvisende, da mennesket ud fra et filosofisk, sociologisk og genealogisk perspektiv er et socioteknisk væsen (Latour, 1994, s. 64).

ANT danner altså det teoretiske grundlag for projektets brug af aktant frem for aktører og antagelse om, at sociale fænomener er sociotekniske netværk bestående af aktanter.

4.2. Etik

Etik er et filosofisk felt med formålet om at undersøge og kortlægge 'det gode' og 'det onde'. Inden for feltet eksisterer der flere forskellige typer af etik: Metaetik, Normativ etik og Anvendt etik (Tzafestas S G, 2018, s. 7). I denne opgave anvendes den normative etik som den primære teoretiske position til at studere normer og regelsæt for handlinger (Tzafestas S G, 2018, s. 9). Dette fokus stemmer overens med projektets fokus på retningslinjer, som omhandler organisatoriske og institutionelle forsøg på at præge og skabe fremtidige etiske normer for korrekt handling. Dermed er den normative undersøgelse af etik ikke en analyse af urokkelige regler, men en analyse af kontinuerligt foranderlige sociotekniske fænomener, der formes af og i den kontekst, de eksisterer i (Van Den Hoven, 2010, s. 65).

Endvidere kan normativ etik vurderes ud fra tre grene inden for denne type: Nytteetik, pligtetik og dydsetik (Bartneck et al., 2021, s. 19–20). Inden for nytteetik er moralsk og etisk korrekthed vurderet på baggrund af konsekvenserne af en handling. Nytteetikken stammer fra omkring starten af 1800-tallet og har sine rødder i den britiske filosof Jeremy Bentham (Sverdlik, 2023a, s. 1–2). Inden for nytteetikken defineres godhed som oplevelsen af lykke og ondskab som oplevelsen af lidelse (Sverdlik, 2023b, s. 22). Hvis udfaldet af en handling eller ikke-handling øger den totale mængde af lykke, vil den altså blive anset som en etisk god handling. Det grundlæggende i nytteetikken er derfor, at en handling er god, hvis konsekvenserne af den resulterer i en større mængde lykke end lidelse. Både pligtetikken og dydsetikken står i kontrast til nytteetikken, eftersom de ikke vurderer en handling ud fra dens konsekvens, men dens ophav. Inden for pligtetik vurderes hensigten med handlingen, og i dydsetikken vurderes personen bag handlingen. Emmanuel Kants pligtetik bygger på antagelsen om, at menneskets besidder en indre moralitet, som stiller krav til, at man bør handle i overensstemmelse med moral. Moral er et felt inden for etik, og omhandler praktiseringen af etik korrekte normer og handlinger (Tzafestas S G, 2018, s. 9; Van Den Hoven, 2010, s. 64). Forholdet mellem den indre moralitet og at handle i overensstemmelse med den kalder Kant for en rationel handling. Dette skaber grundlaget for Kants beskrivelse af mennesker som 'rationelle agenter', der i

kraft af deres rationalitet besidder en iboende værdi, som gør dem til mål i sig selv. Denne sondring kaldes 'det praktiske imperativ', hvori rationelle agenter altid skal anses som mål og aldrig som et middel (Immanuel, 1999, s. 88–91). Dermed er pligtetikken i direkte modsætning til nytteetikken, eftersom man ifølge nytteetikken sagtens kan gøre mennesket til et middel, hvis det øger den totale mængde af lykke i samfundet. Dydsetikken stammer fra Platon og Aristoteles, og definerer det etisk korrekte ud fra, hvorvidt en handling eller ikke-handling kommer fra en god person. Altså er det gode ikke afhængigt af resultatet af en handling eller hensigten med den enkelte handling, men vurderes på baggrund af individets etiske karakter (Tzafestas S G, 2018, s. 12–13).

4.3. Moralsk status

Etik og moral skabes på baggrund af de eksisterende værdier i en given kontekst; denne definition af værdier kaldes teoretisk pluralisme, hvilket betyder, at én handling både kan være etisk korrekt og forkert, alt efter standpunktet det vurderes ud fra. Det er dog vigtigt at adskille pluralisme fra moralsk relativisme og skepticisme, hvori etik og moral efterlades uden nogen egentlig betydning. For at modarbejde dette bør en handling vurderes ud fra dens kontekst, hvorfra det etiske standpunkt må tages (Van Den Hoven, 2010, s. 65–67). Et eksempel på dette er, at korrekt etik i forhold til emner som race, seksualitet og køn kan variere, alt efter hvilken historisk tid og kulturelt standpunkt der spørges fra. Det er ikke hensigten med etik at fastsætte universelle sandheder for, hvad der er og ikke er korrekt adfærd, men at systematisere normerne og regelsættene, som er gældende i den aktuelle undersøgelseskontekst (Van Den Hoven, 2010, s. 67 & 75).

På baggrund af ovenstående definition af etik kunne det konkluderes, at etik kun er gældende for menneskelige aktører, men det er ikke tilfældet. Både ældre og nyere etiske teorier præsenterer eksempler på, hvordan ikke-mennesker også er relevante aktanter inden for etikens verden. Inden for pligtetikken fremgår det i forholdet mellem rationelle og ikke-rationelle agenter, der tildeles to forskellige moralske statusser. Selvom ikke-rationelle agenter ikke har en indre moralitet, er de ikke helt uden moralsk og etisk effekt, deres væren og behandling kan nemlig have indflydelse på den rationelle agent. Hvis en rationel agent har til hensigt at behandle ikke-rationelle agenter, såsom et dyr, dårligt, vil det skade den

rationelle agents indre moralitet (Jecker, 2010, s. 169–170). Rationelle agenter er derfor moralske agenter, som kan udføre etisk korrekte handlinger, og ikke-rationelle agenter, såsom dyr, kategoriseres som moralske patienter, der kan behandles korrekt uden selv at have noget eksplicit moralsk handlerum. En nyere tilgang til moralsk status er Mark Coeckelberghs 'kritiske tilgang', hvori moralsk status ikke er givet på forhånd, som i pligtetikken, men er noget som opstår i vores interaktion med verden (Coeckelbergh, 2018, s. 149). Med den kritiske tilgang forstås moralsk status som produktet af menneskelige aktørers tilskrivning af moralsk status til ikke-menneskelige entiteter gennem sociale, kulturelle og sproglige interaktioner (Coeckelbergh, 2018, s. 151–153). Det er derfor nødvendigt at undersøge netværket af interaktioner mellem menneskelige og ikke-menneskelige aktanter for at kunne definere, hvilken moralsk status aktanter tilskrives i en kontekst.

Ved at anvende den kritiske tilgang til moralsk status bliver tilskrivningen af moralsk agent, såsom rationel agent, og patient, såsom ikke-rationel agent, mindre kategorisk. En moralsk agent er defineret på baggrund af dens evne til at agere og udføre bevidste handlinger, også kaldt agent-orienteret (Gunkel, 2014, s. 115). Men eftersom moralsk status er en konsekvens af netværket af interaktioner, som aktanten indgår i, er moralsk status ikke noget man har, men noget man får i kraft af netværket. Til trods for at statussen som moralsk agent er agent-orienteret, bliver denne position, ligesom den patient-orienterede 'moralske patient', givet sin moralske status (Gunkel, 2014, s. 118). Statussen af moralsk agent anses derfor som værende et spektrum mellem etisk handlende aktanter og etisk ikke-handlende aktanter frem for en rigid kategorisering. Dermed bliver det muligt at få et nuanceret indblik i netværket og aktanternes roller i det: Etisk påvirkende agenter, etisk implicite agenter, etisk eksplicite agenter og fuldbyrdede etiske agenter. Den 'etisk påvirkende agent' er en aktant, som har etisk betydning i det netværk, som aktanten indgår i, uanset om den er agent eller patient (Wiegel, 2010, s. 203–204). Hertil inddrages det klassiske eksempel fra Langdon Winner om artefaktens politiske karakter, hvor broerne i Long Island, New York, fik en etisk rolle i adskillelsen af befolkningen baseret på sociale klasser og race. Altså havde en ikke-handlende infrastruktur en uetisk indflydelse, der kom til udtryk ved at forhindre den frie bevægelighed med busser, som primært blev anvendt af sorte- og fattige borgere (Winner, 1980, s. 123–124). Den næste aktant på spektret er den 'etisk implicite agent', som er en

aktant, der er designet til ikke at have evnen til at udføre uetiske handlinger. Dette kunne være fodgængerovergange, hvis eneste formål og funktion er at gøre bilister og fodgængere opmærksomme på vejkrydsninger for at undgå trafikulykker. Herefter kommer den 'etisk eksplicitte agent', som har etiske aspekter inkorporeret i sin funktion. Derved er det muligt for agenten at udføre handlinger i uforudsete situationer og med uforudsete beslutninger, hvilket betyder at denne agent, i modsætning til de ovennævnte agenter, har en grad af autonomi. Den sidste agent er den 'fuldbyrdede etiske agent', der har evnen til at udtrykke et menneskeligt niveau af moralsk bevågenhed, fornuft og autonomi (Wiegel, 2010, s. 203–204).

Det kan altså konkluderes, at moralsk status og moralsk agens ikke kun er et menneskeligt fænomen men også relevant for ikke-menneskelige aktører. Som konsekvens heraf er det relevant at undersøge KI som en etisk aktant i samfundet med hensyn til de processer i netværket, som teknologien indgår i.

4.4. KI Etik

KI etik er et nyere interdisciplinært felt inden for etikteori med det formål at adressere etiske fænomener vedrørende KI (Huang et al., 2023, s. 799). Undersøgelsesobjektet i KI etik er altså etik med henblik på én kategori af teknologi. Eftersom KI etik er et relativt nyt og fragmenteret forskningsfelt, findes der mange forskellige udlægninger af feltets begreber og metoder. Ikke desto mindre gives der herunder et indblik i feltets baggrund og anvendelse.

Der er en længere tradition for den kontekstualistiske teknologiforståelse, hvori teknologi anses som et mere socialt fænomen end den materialistiske teknologiforståelse, hvor teknologisk udvikling anses som et resultat af materielle faktorer. Her kan der henvises til Heideggers filosofiske udlægning af teknologi i bogen "The question concerning technology, and other essays", fra 1977. Heideggers teknologiforståelse trækker tråde hele vejen tilbage fra det gamle Grækenland og op til den moderne teknologifilosofi (Verbeek, 2005, s. 51–54). En vigtig pointe er her, at den moderne teknologi frembringer nye spørgsmål og problemstillinger, men den humanistiske bearbejdning af den fysiske verden og de praktiske videnskaber har altid været et aktuelt forskningsfelt (Verbeek, 2005, s. 54). Den sociotekniske karakter af teknologi er også klarlagt af Langdon Winner i artiklen fra 1980 "Do artefacts have politics?", hvori teknologiens sociale funktioner analyseres. For det første kan teknologi være

et middel til at skabe og vedligeholde magtstrukturer, og for det andet kan teknologi være indirekte eller direkte afhængig af en magtstruktur (Winner, 1980, s. 134). Dermed er det relevant at anse den etiske karakter af teknologi som konsekvens af de sociale strukturer, teknologi udspringer af (Hagendorff, 2020, s. 112).

Computeretik er ligesom KI etik et interdisciplinært felt inden for etikteori, men med det bredere fokus på computerteknologi i form af 'informations- og kommunikationsteknologi' (herefter kaldt IKT) (Van Den Hoven, 2010, s. 59). Inden for computeretik er der specielt fokus på gennemsigtighed, tilgængelighed, ansvarlighed, design og robusthed som ledende principper for IKT (Van Den Hoven, 2010, s. 59–61). Disse ledende principper danner grundlaget for en række ikke-tekniske retningslinjer, såsom retfærdighed og sikkerhed, hvilket kan anvendes i undersøgelsen af IKT (Van Den Hoven, 2010, s. 76).

KI etik bygger på ideerne fra den kontekstualistiske teknologiforståelse og tidligere interdisciplinære felter inden for etikteori vedrørende teknologi. Altså får man ved anvendelsen af KI etik et mere nuanceret billede af KI end den rene teknologiske anskuelse af maskiner, funktionalitet og data. Med det KI etiske perspektiv bliver teknologien et helhedsspørgsmål, som omfatter både dens sociale og teknologiske ophav, kontekst og virke.

5. Afgrænsning

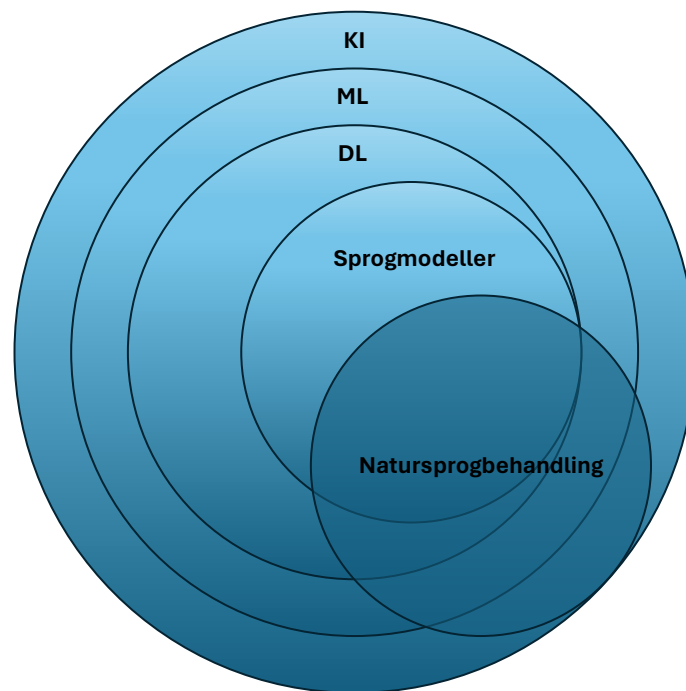
Herunder præciseres det, hvilken teknologi og hvilke myndigheder, der behandles gennem resten af projektet. Først fokuseres der på at definere og konkretisere kunstig intelligens, både som koncept og teknologi. Tillige behandles de relevante delelementer af teknologien, hvor det er nødvendigt at have et fælles udgangspunkt for at kunne gennemføre en fokuseret analyse og diskussion af emnet. Derefter gives der en begrundelse for det empiriske grundlag for analysen af etiske retningslinjer. Endvidere fremhæves andre relevante tekster i forhold til analysen, som ellers er blevet udeladt grundet projektets fokus og omfang.

5.1. Kunstig Intelligens, sprogmodeller og data

Den følgende afgrænsning af KI fokuserer hovedsageligt på sprogbaseret KI i form af 'sprogmodeller' (Kirchmeier et al., 2019, s. 15). I praksis er sprogmodeller den teknologi, der

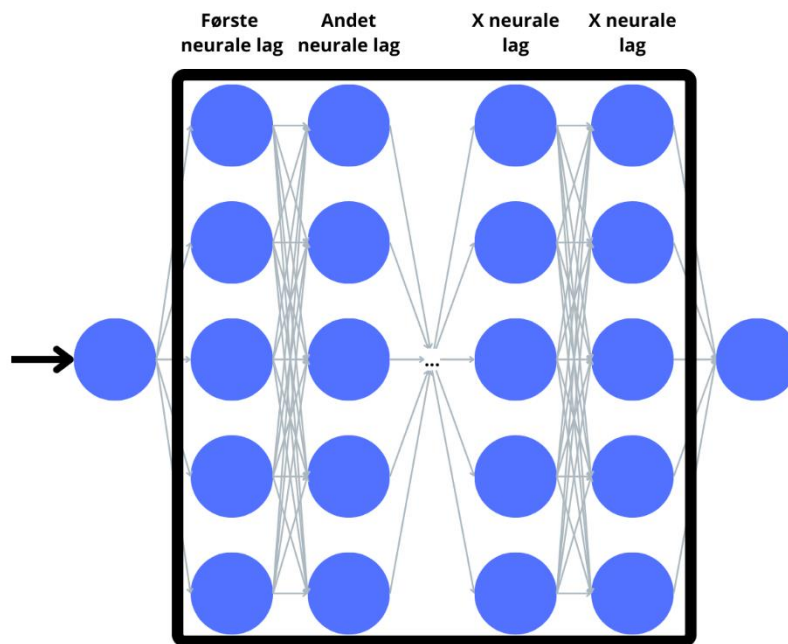
ligger bag samtaleprogrammer som ChatGPT, skabt af OpenAI, og Googles BARD. Disse former for KI kaldes også generativ KI.

KI er et paraplybegreb for teknologier, som gennem komplekse algoritmer og processer kan udføre handlinger, der til forveksling kunne tolkes som menneskelig intelligens (Gutiérrez, 2023, s. 2). Forholdet mellem sprogmodeller og KI er visualiseret i figuren herunder, som også er udgangspunktet for den videre forklaring af KI (Figur 1).



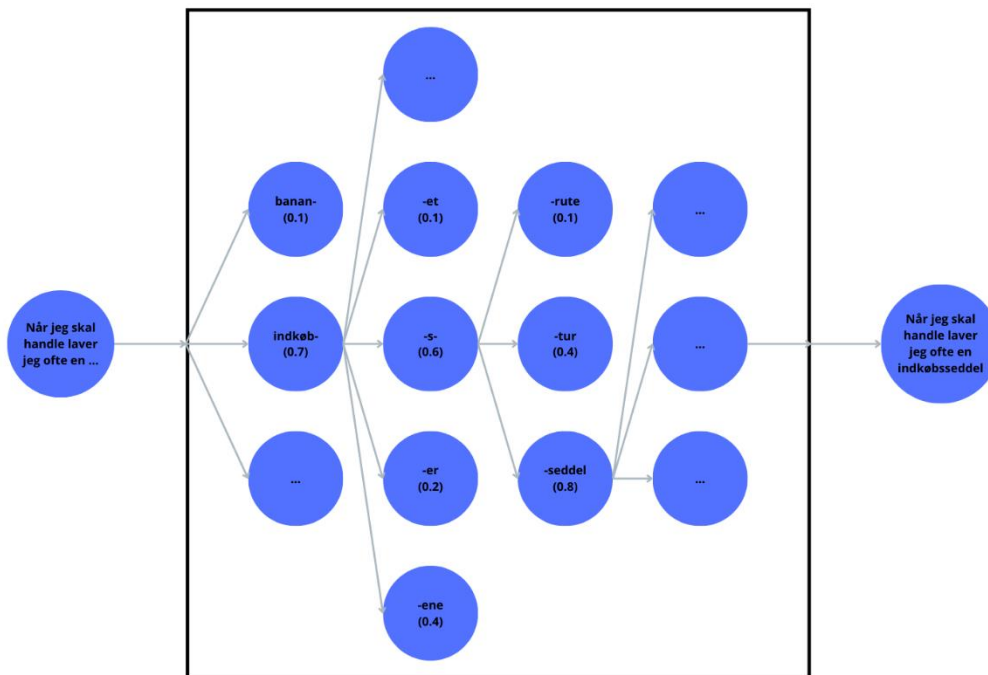
Figur 1.

Som det er vist i figur 1, består KI af en række underkategorier, hvoraf den første underkategori er 'Maskin Læring' (herefter kaldt ML). ML dækker over teknologier, som specialiseres i at identificere logiske sammenhænge i én type data. I forhold til sprogmodeller anvendes denne type teknologi f.eks. til at finde mønstre i, hvordan et sprog er opbygget. Ligesom KI er et paraplybegreb for ML, så dækker ML dækker over 'Dyb Læring (herefter DL). DL anvendes til systematisering af mere komplekse data end en enkelt type data. Teknologi baseret på DL har en karakteristisk struktur som et neuralt netværk bestående af lag af neuroner (se figur 2).



Figur 2. Eksempel på de neurale lag i et neuralt netværk

Anvendeligheden af DL er, at den ikke blot identificerer mønstre som ML, men at den kan lave forskellige produkter, alt efter hvilke data den er trænet på. For at kunne give et konkret eksempel på, hvordan DL fungerer, er det nødvendigt at gå videre til sprogmodeller (Kirchmeier et al., 2019, s. 16). Sprogmodeller består af statistiske modeller over det naturlige sprog, baseret på de mest hyppige og logiske sammenhænge mellem ord, ordsammensætninger og sætninger (Kirchmeier et al., 2019, s. 15). Den statistiske model foreligger i praksis som et neuralt netværk, hvis struktur kan illustreres ved figur 2, hvor hver cirkel inde i den sorte kasse symboliserer en statistisk sammenhæng. Der er givet et mere håndterbart eksempel i det hypotetiske eksempel i figur 3 herunder.



Figur 3. Eksempel på en simpel sprogmodel

Ligesom figur 2 består sprogmodellen her af en sort kasse, hvori der er X antal neurale lag. Alle neuronerne i hvert lag har en bestemt funktion i kraft af, at de indeholder en værdi, som vurderer i hvilken grad, de passer til resultatet af alle forrige lag. Altså starter vi med, at systemet gives en sætning, som skal afsluttes med det mest sandsynlige ord. Derfor giver denne hypotetiske sprogmodel svaret "indkøbsseddel", fremfor den mindre sandsynlige "indkøbsrute", da "-seddel" har en større værdi end "-rute" (Figur 3).

Værdierne, som hvert neuron indeholder, kaldes 'vektorer' og udregnes på baggrund af de statistiske sammenhænge mellem ord og ordsammensætninger (Kirchmeier et al., 2019, s. 15). For at udregne vektorerne træner man netværket på en kendt datamængde, der analyseres ved en metode kaldet 'natursprogsbehandling' (fra det engelske 'Natural Language Processing', herefter NLP) (Figur 1). Ved at analysere store mængder sprogbaseret data med NLP nedbrydes tekster, sætninger og ord til mindre elementer med henblik på at kortlægge de logiske sammenhænge mellem disse fragmenterede sprogelementer (Kirchmeier et al., 2019, s. 18). Kvaliteten af en sprogmodel er derfor direkte afhængig af ikke bare mængden men også

kvaliteten af den data, som den statistiske model baseres på (Kirchmeier et al., 2019, s. 16 & 18).

Altså er sprogmodeller en form for KI, som dækker over ML og DL, der begge er teknologier, der på hver sin måde er specialiseret i at systematisere og skabe sammenhænge mellem data. Sprogmodeller er en specifik form for DL, som er specialiseret i at afkode sprog og skabe logiske sammenhænge mellem sprogelementer. De logiske sammenhænge i en sprogmodel defineres ved at fragmentere det naturlige sprog med NLP. Dette betyder, at sprogmodeller er en refleksion af den datamængde den trænes på. Derfra kommer den direkte sammenhæng mellem sprogmodeller og mængden af kvalitetsdata, som er afgørende for, hvor succesfuld en sprogmodel ender med at være. Det er altså nødvendigt at være yderst opmærksom på datagrundlaget for udviklingen af sprogbasert KI.

Det er grundet dette kritiske forhold mellem data og KI, at projektet indeholder en analyse af sprogteknologi.dk, da det er et knudepunkt for sprogressourcer til udvikling af dansksprogede sprogmodeller.

5.2. Relevante myndigheder

Det er ikke muligt i én undersøgelse at analysere og diskutere alle de aktanter, som indgår i netværket for udviklingen af KI i Danmark. Derfor begrænses undersøgelsen til regulatoriske tiltag fra den danske regering og den europæiske union. Disse myndigheder har det til fælles, at de begge har direkte indflydelse på danske udviklere, udbydere og anvendere af KI. Andre myndigheder, organisationer og institutioner er også relevante og indflydelsesrige for rammerne, såsom 'The Institute of Electrical and Electronics Engineers IEEE', Dataetisk Råd og mange andre (Floridi, 2023, s. 65–66; IEEE, 2024; Nationalt Center for Etik, 2024). Grundet problemformuleringens fokus på aktuelle og gældende rammer for udvikling af KI i Danmark inddrages rådgivende tekster kun i et sådanne omfang, som de anvendes i reguleringerne, hvorfor de primære tekster i analysen af rammen for udvikling af KI i Danmark er baseret på dokumenter fra den danske regering og Den Europæiske Union (EU). De danske reguleringer består af Finans- og Erhvervsministeriets "National Strategi for Kunstig Intelligens" fra 2019 (herefter kaldt 'Nationalstrategien'), og "Danmarks digitaliseringsstrategi – Ansvar for den digitale udvikling" af Digitaliserings- og Ligestillingsministeriet fra 2023 (herefter kaldt

'Digitaliseringsstrategien'). Reguleringerne fra EU består af "Europa-Parlamentets beslutning af 20. oktober 2020 med henstilling til Kommissionen om en ramme for etiske aspekter af kunstig intelligens, robotteknologi og relaterede teknologier (2020/2012(INL))" (herefter kaldt '2020/2012(INL)'). Endvidere inddrages den højaktuelle europæiske retsakt om KI "Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence Act) and amending certain Union legislative acts" (herefter kaldt 'KI akten'), som stadig er under behandling i Parlamentet. Til trods for at tiltaget endnu ikke er vedtaget, er den relevant, fordi det er det første stykke KI lovgivning af sin slags, og fordi EU har en kontinuerligt voksende indflydelse på KI regulering i Danmark (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 26). I de to dokumenter fra EU henvises der til, at deres etiske grundlag er baseret på rapporten "Etiske Retningslinjer for Pålidelig Kunstig Intelligens" af 'Ekspertgruppe på højt plan om kunstig intelligens' (herefter kaldt 'HLEG'). Eftersom denne rådgivende tekst bliver direkte refereret i lovteksterne som grundlaget for den etiske rammesætning af KI, anses den som yderst relevant til besvarelsen af problemformuleringen (P9_TA(2020)0275, 2020, s. 2; 2021/0106(COD), 2024, s. 24).

Der er siden det projektets analytiske arbejde udkommet en ny version af KI akten, altså er der risiko for at nogle uddrag og pointer i forhold til KI akten ikke længere er gældende. Dette underkender dog ikke projektets undersøgelsesobjekt og formål med henblik på etisk regulering af KI i Danmark.

6. Analyse

6.1. Etiske retningslinjer for KI i Danmark

Først undersøges, hvad de inddragede reguleringer ønsker at opnå i forhold til KI. Derefter opstilles en harmoniseret retningslinje for etisk KI i Danmark på baggrund af en komparativ gennemgang af overensstemmende principper for at opnå en forståelse for de etiske retningslinjer for udviklingen af sprogmodeller i Danmark.

6.1.1. Lovgivning, rådgivning og rammesætning

Ved offentliggørelsen af Digitaliseringsstrategien skrev Digitaliserings- og Ligestillingsministeriet i et nyhedsopslag at “Kunstig intelligens er et af de områder, som regeringen ønsker at sætte retning for.” (Digitaliserings- og Ligestillingsministeriet, 2023b). I Digitaliseringsstrategiens forord beskrives retningen som et ønske om at skabe “klare rammer for udviklingen af fremtidens digitalisering og vækst.” (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 5). En måde, hvorpå Digitaliseringsstrategien sætter ramme og retning for udvikling af KI, er ved at lægge op til, at man skal ”sikre, at anvendelse, udvikling og udbredelse af kunstig intelligens sker på en etisk ansvarlig måde med mennesket og samfundet i centrum.” (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 11). Men Digitaliserings- og Ligestillingsministeriets giver kun en generel ramme for KI udvikling, og efterlader hvad etik, ansvarlighed og ”mennesket og samfundet i centrum” betyder for KI i det uvisse. Afsnittet ”Et stærkt, etisk og ansvarligt digitalt fundament” referer primært til etik som et spørgsmål om data, sikkerhed og ansvarlighed (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 26). Derudover er den nyere danske regulering af KI stærkt præget af et materialistisk syn, hvor teknologien er et værktøj, som ”Danmark skal udnytte [...] på en ansvarlig måde” (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 11) og dermed sikre, at udviklingen foregår i den retning, som er ønsket af KI, nemlig centreret om menneske og samfund. I stedet for at give en liste af kriterier og mål for udviklingen af KI danner digitaliseringsstrategien altså en konceptuel ramme for KI, bestående af 25 retningsgivende initiativer (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 1). Det er ét af disse initiativer, sprogteknologi.dk, som vil blive analyseret senere i opgaven. De 25 initiativer er inddelt i seks temaer, som skal sætte den langsigtede retning for den digitale udvikling i Danmark. Dermed består strategien af at lægge op til videre udvikling ved at fremhæve og promovere ønskværdige KI-relaterede tiltag.

Den tidligere danske Nationalstrategi er der anvendt en lignende tilgang ved at benytte sig af eksisterende initiativer. Men i stedet for at opstille temaer for KI indsatsen gives fire konkrete indsatsområder i form af sundhed, energi og forsyning, landbrug og transport. I modsætning til Digitaliseringsstrategien opstilles der i Nationalstrategien mere eksplicitte retningslinjer for, hvordan disse indsatser bør udføres med henblik på at opstille og definere seks etiske

principper for KI. Principperne fremlægges i forbindelse med den første af fire ”sigtelinjer” for tilgangen til udvikling af KI i Danmark (Finansministeriet & Erhvervsministeriet, 2019, s. 8–10). Den første af sigtelinjerne omhandler netop, hvad et etisk grundlag for KI bør være i Danmark: ”Danmark skal have et fælles etisk grundlag for kunstig intelligens med mennesket i centrum” (Finansministeriet & Erhvervsministeriet, 2019, s. 8). Hvor Digitaliseringsstrategien omtaler KI etik som noget, der skal sikres samtidig med udviklingen, bliver det nævnt i Nationalstrategien fra 2019 som et primært krav som både ”(...) virksomheder og offentlige myndigheder skal (egen formatering) have et stærkt fokus på(...)” (Finansministeriet & Erhvervsministeriet, 2019, s. 8). At etisk KI er det primære fokus giver genklang i de efterfølgende sigtelinjer som alle enten indeholder det etiske som et grundlag for teknologien eller som et mål i sig selv: ”Regeringen har som mål, at [...] Danmark bliver blandt de bedste lande i verden til at udnytte data og nye forretningsmodeller baseret på ansvarlig udvikling og anvendelse af kunstig intelligens” (Finansministeriet & Erhvervsministeriet, 2019, s. 9). Nationalstrategien har derfor et stærkt fokus på KI etik som en ramme og et produkt i sig selv, hvorimod Digitaliseringsstrategien fremsætter det som en faktor, som regulerer udviklingen af KI. Ydermere præsenterer Nationalstrategien konkrete retningslinjer for, hvilke faktorer der bør efterleves for at opnå etisk KI. Gældende for begge de nationale reguleringer af KI er dog, at de benytter sig af konkrete initiativer til at eksemplificere, hvilken type projekter de ønsker skal skabe den videre udvikling af KI.

I det endelige udkast til den europæiske regulering KI akten er det også muligt at finde eksempler på konkrete initiativer. Men disse eksempler anvendes som rammesætning af reguleringen og ikke udvikling af fremtidig KI (2021/0106(COD), 2024, s. 12–92). Det overordnede formål med KI akten er givet under artikel 1 som ”(...) to improve the functioning of the internal market and promoting the uptake of human centric and trustworthy artificial intelligence, while ensuring a high level of protection (...)” (2021/0106(COD), 2024, s. 93). KI akten indeholder altså to underformål, (1) at forberede det indre marked på KI ved at harmonisere lovgivningen, og (2) at promovere menneskecentreret og pålidelig KI. Den anden del af KI akten har hensigt på den etiske udvikling af KI og er baseret på at fremme risikohåndtering og europæiske værdier (2021/0106(COD), 2024, s. 11–12). Risikohåndtering er da også et centralt emne for den etiske ramme i Digitaliseringsstrategien, hvor både

ansvarlighed og etik relateres direkte med at modarbejde potentielle udfordringer og trusler forbundet med KI (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 16 & 22).

Hvad de ”europæiske værdier” er, som skal fremmes og danne det etiske grundlag for fremtidens KI, kan variere og forandres, alt efter hvilke aktører der sidder med ved forhandlingsbordet. Nationale aktører som Danmark vil forsøge at påvirke internationale reguleringer. Dette optræder i Nationalstrategien, som omtaler vigtigheden af at ”præge” udviklingen af KI baseret på ”vores fælles værdier” (Finansministeriet & Erhvervsministeriet, 2019, s. 5 & 7). Digitaliseringsstrategien omtaler dette under temaet ”Danmark i centrum af international digitalisering”, hvor det nævnes, at Danmark skal indgå i europæiske forordninger for at fremme ”afspejling af danske interesser.” (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 26). Sideløbende med at EU bliver præget af sine medlemsstater, forsøger de også at lave egen værdiudvikling. Det europæiske standpunkt i forhold til en etisk ramme for KI i EU, som fremlægges i KI akten, er delvist baseret på undersøgelser fra ’Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens’ og tidligere beslutninger for etiske principper vedrørende KI (2021/0106(COD), 2024, s. 14 & 24).

I undersøgelsen fra HLEG, ’Etiske retningslinjer for pålidelig kunstig intelligens’ fra 2019, præsenteres formålet som et ønske om ”At fremme pålidelig kunstig intelligens” og opstille ”en **ramme for opnåelse af pålidelig kunstig intelligens** (tekstens egen formatering).” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 2). I modsætning til de tre forrige reguleringer er undersøgelsen fra HLEG ikke et lovligt bindende dokument. ’Pålidelig KI’ defineres som opfyldelse af lovlige, etiske og robuste komponenter gennem alle teknologiens livscykler (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 2). Hertil understreges det, at undersøgelsens formål ikke er at have en direkte effekt på det lovlige aspekt af KI, men derimod at rådgive i at fremme og sikre etisk og robust KI. Operationaliseringen af rammen for etisk og robust KI, ifølge HLEG, fremlægges som en pilotversion af en eventuel evalueringsliste, der senere udkom som en selvstændig udgivelse i 2020 (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2020). Det er dog bemærkelsesværdigt, at det er undersøgelsen med pilotversionen, som omtales i KI akten, og ikke den endelige udgave af evalueringsskemaet. Derfor kan det antages, at senere anvendelse af undersøgelsen ikke er baseret på evalueringsskemaet, men de mere abstrakte

overvejelser, som bliver fremlagt undervejs i undersøgelsen, hvilket er i modsætning til, at undersøgelsens eksplicitte formål er at opsætte en ramme for pålidelig KI. Dog er det lykkedes HLEG at fremme pålidelig KI internt i EU, hvilket kan påvises ved, at den refereres i senere lovgivning (P9_TA(2020)0275, 2020, s. 2; 2021/0106(COD), 2024, s. 24).

Det leder os videre til de lovlige komponenter af KI, som i undersøgelsen overlades til eksisterende love og bestemmelser som i 2020/2012(INL). Formålet med denne regulering af KI er flerdelt og indeholder aspekter af alle de ovennævnte regulatoriske initiativer. Første mål er ”at opbygge tillid på alle niveauer hos de involverede interessenter og i samfundet til kunstig intelligens (...)” (P9_TA(2020)0275, 2020, s. 32). Dette omtales også i de andre reguleringer, der inddrager tillid som et aspekt af KI etik men ikke et mål i sig selv; i Nationalstrategien omtales tillid som et succeskriterium ”For regeringen er det derfor centralt, at danskerne fortsat kan være trygge og have tillid (...)” (Finansministeriet & Erhvervsministeriet, 2019, s. 7). Dernæst ønsker Europa-Parlamentet, at 2020/2012(INL) skal støtte udviklingen, udbredelsen og brugen af KI (P9_TA(2020)0275, 2020, s. 32). Disse ønsker om udvikling og udbredelse af KI er i overensstemmelse med det primære formål for Digitaliseringsstrategien. Det sidste formål med 2020/2012(INL) er ”at kræve gennemsigtighed og bedre informationsstrømme blandt borgere og organisationer, der udvikler, udbreder eller anvender kunstig intelligens (...)” (P9_TA(2020)0275, 2020, s. 32). Gennemsigtighed og data går også igen i de andre reguleringer; i Digitaliseringsstrategien og KI akten bemærkes vigtigheden af muligheden for adgang til data af høj kvalitet og teknologiforståelse (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 14 & 26; 2021/0106(COD), 2024, s. 46–47). Beslutningen 2020/2012(INL) favner vidt og fremhæver flere forskellige formål i modsætning til de andre reguleringer, som hver har mere fokuserede intentioner. Det er dog vigtigt at påpege, at det er muligt at identificere de individuelle formål på tværs af alle tiltagene som implicitte eller delelementer af deres overordnede formål.

Når teksterne inddeles i de regulatoriske niveauer ‘national’ og ‘international’, kan der opstilles et modsætningsforhold, hvor det nationale indeholder et stort fokus på iværksættende tiltag, og de internationale præsenterer langt mere regulatoriske tiltag. Dette defineres i rapporten fra HLEG som ”positive og negative forpligtelser”. Altså kan man skelne mellem positive reguleringer, der fremmer og opfordrer til bestemte handlinger, som man bør

eller skal gøre, og negative reguleringer, der hovedsageligt fokuserer på at forbyde og forhindre handlinger, man ikke må gøre (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 6). Opdelingen mellem disse to former for regulatoriske forpligtelser kan relateres til Emmanuel Kants pligtetik. De positive forpligtelser kan betegnes som en agent-orienteret pligtetisk forventning til moralske agenter. Derimod henvender de negative forpligtelser sig til den moralske agents pligt over for andre aktanter som moralske patienter, i form af en patient-orienteret pligtetisk forventning til den moralske agents handlen i overensstemmelse med en eksternt defineret moralitet. Derfor bliver reguleringerne og deres fremsatte principper et spørgsmål om, hvorvidt de henvender sig til regulering af KI på baggrund af en agent-orienteret eller patient-orienteret pligtetik.

På tværs af reguleringerne er der dog en bred enighed om vigtigheden af udvikling og udbredelse af etisk forsvarlig KI. Trods variationer i, hvorvidt der fokuseres på konkrete initiativer eller principielle værdier, kan reguleringerne tilsammen danne én ramme for etisk KI.

6.2. Én etisk ramme

Som sagt præsenterer reguleringerne forskellige versioner af etiske principper og krav til KI. Enten opstilles rammen som en liste af konkrete principper, som derefter udfoldes, forklares eller eksemplificeres, såsom Nationalstrategien, 2020/2012(INL) og rapporten fra HLEG (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 12–15; P9_TA(2020)0275, 2020, s. 33; Finansministeriet & Erhvervsministeriet, 2019, s. 28–29). Ellers indgår de etiske principper og krav som implicitte dele eller argumenter i større regulatoriske tiltag, som i Digitaliseringsstrategien og KI akten (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 10–13; 2021/0106(COD), 2024, s. 93–94). Herunder gives et overblik over de etiske principper, som fremgår i de ovennævnte tekster, og fremhæver hvilke principper, der optræder på tværs af reguleringerne (Tabel 1).

Etiske principper	National- strategien, 2019	HLEG, 2019	Europa- Parlamentet, 2020	Digitaliserings- strategien, 2023	KI akten, 2024
Autonomi					
Robusthed					
Gennemsigtighed					
Retfærdighed					
Udvikling					
Ansvarlighed					

Tabel 1. Visualisering af hvilke etiske principper for KI der optræder i og går på tværs af de individuelle reguleringer.

Principperne er fundet på baggrund af kodningen af teksterne af flere omgange. I første kodningsomgang har ord og sætninger i teksterne formet individuelle koder, og i efterfølgende kodningsomgange er koderne blevet anvendt på tværs af tekster, som er blevet samlet i en kodebog (Bilag A, s. 1–9). For eksempler på, hvordan denne proces ser ud i praksis, henvises der til Bilag B, som består af et uddrag fra den første kodningsomgang fra rapporten af HLEG og et uddrag af anden kodningsomgang fra Digitaliseringsstrategien (Bilag B, s. 10–11). Dette resulterede i dannelsen af hierarkiske koder, som blandt andet har muliggjort en konkretisering af de etiske principper.

I de følgende afsnit bliver principperne gennemgået med henblik på at definere og konkretisere deres betydning og indvirkning på reguleringen af KI i Danmark. Det er dog ikke alle principperne, som vil blive gennemgået individuelt, eftersom nogle af principperne kan beskrives som underkategorier af andre principper. Principperne, som ikke får en individuel behandling, vil derfor indgå i gennemgangen af det eller de principper, som de hører under.

6.2.1. Autonomi

Generelt bliver autonomi omtalt i forbindelse med, at KI skal være menneskecentreret med hensyn til menneskelig selvbestemmelse, kontrol og frihed (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 12). Både de europæiske og danske retningslinjer nævner menneskelig autonomi som et af de primære principper, hvilket også kan ses i tabel 1, hvor det fremgår i alle reguleringerne. Autonomi beskrives som en grundlæggende rettighed

inden for EU-regi, og i de danske reguleringer fremhæves den demokratiske værdi i at sikre borgerens rettigheder (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 5; 2021/0106(COD), 2024, s. 93). Nationalstrategien anvender autonomi under et andet navn, "Selvbestemmelse", men har menneskelige og demokratiske rettigheder nævnt under et separat etisk princip, "Værdighed" (Finansministeriet & Erhvervsministeriet, 2019, s. 28). De europæiske retningslinjer forholder sig primært til autonomi som et grundlæggende spørgsmål om at forsvare lovsikrede rettigheder (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 12). Ved at referere til 'Den Europæiske Unions charter om grundlæggende rettigheder' baseres lovsikringen på normative vurderinger af menneskelig ret og værdi. Heri beskrives den menneskelige ret og værd som et produkt af Unionens principper om demokrati og ret (Den Europæiske Unions charter om grundlæggende rettigheder, 2012, s. 395). Den europæiske appel til demokratiske principper og værdier bliver også omtalt i Digitaliseringsstrategien, som gentagne gange henviser til sikringen af "demokratiet" og "demokratiske værdier" som en central faktor for udviklingen af KI (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 5 & 8 & 16 & 22 & 26).

Autonomi kan derfor anses som et etisk princip, der fremhæver mennesket som et mål i sig selv frem for et middel, enten på baggrund af at overholde europæisk lovgivning eller demokratiske principper. Eftersom ethvert individ har en grundlæggende ret til at handle og have selvbestemmelse i deres ageren, bør udvikling, udbredelse og anvendelse af KI altid udføres med henblik på at sikre den menneskelige autonomi. Derfor er autonomi som etisk princip baseret på en patient-orienteret pligtetik, der retfærdiggøres på baggrund af, hvad KI gør ved menneskets autonomi, og ikke hvad mennesket gør med KI. Denne rammesætning er dog ikke entydig, da autonomi, i de danske reguleringer, også forbindes med menneskets handlingspotentiale i forhold til KI. Ifølge de danske reguleringer skal det enkelte individ kunne forholde sig til KI selvstændigt, for at der er tale om reel autonomi (Finansministeriet & Erhvervsministeriet, 2019, s. 28). Derfor fremlægges autonomi også som et agent-orienteret princip i forlængelse af pligtetikken krav til den moralske agents evne og pligt til selvbestemmelse. Hvis befolkningen ikke besidder evnerne til "at anvende, forstå og forholde" sig kritisk til KI, så vil selv den mest autonomi-understøttende KI anses som en potentiel risiko (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 14).

Derfor fremlægges menneskets demokratiske ret til autonomi som en ramme for udvikling, udbredelse og anvendelse af KI på baggrund af en iscenesættelse af befolkningen som både etiske agenter og patienter. Altså bør mennesket blive anset som et mål i sig selv uanset hvilken position det har i forhold til KI, som gør at mennesket ikke bør anvendes som et middel i produktionen af KI (Immanuel, 1999, s. 88–91).

6.2.2. Udvikling

I reguleringerne fremgår 'udvikling' som et etisk princip på forskellige grundlag. I de danske reguleringer er udvikling nævnt i sammenhæng med, at digitalisering, i form af KI, er motoren for den danske samfundsudvikling. Det etiske aspekt heraf er, at vedvarende samfundsudvikling er en nødvendighed for at sikre velfærdsmæssig og økonomisk vækst gennem øget konkurrenceevne og produktivitet (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 7 & 20). Derfor er udvikling af KI en etisk korrekt handling i kraft af, at den "skaber markant bedre levevilkår for mange mennesker." (Finansministeriet & Erhvervsministeriet, 2019, s. 5). For at sikre denne udvikling er det nødvendigt at udnytte KI med hensyn til de potentielle udfordringer og muligheder forbundet med teknologien (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 8). Hertil fremhæves små og mellemstore virksomheder (herefter kaldt 'SMV'er') som en vigtig komponent i den danske og europæiske økonomi (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 20; P9_TA(2020)0275, 2020, s. 5). For at SMV'er har mulighed for at innovere og udvikle, er der behov for gode omstændigheder for at skabe teknologiske løsninger, hvortil KI rummer et enormt vækstpotentiale (Finansministeriet & Erhvervsministeriet, 2019, s. 25). Hertil opfordrer de europæiske reguleringer til afventende juridiske handlinger overfor KI, og hvornår det kan tages i brug. Hvis der opstilles for mange juridiske reguleringer af KI, vil det hindre udviklingen af SMV'er. Derfor opfordrer EU nationale KI-aktører til at handle ud fra et proportionalitetsprincip, hvor lovgivning og juridiske virkemidler kun tages i brug, når de er nødvendige og proportionelle, for dermed at undgå overregulering (P9_TA(2020)0275, 2020, s. 8 & 31; 2021/0106(COD), 2024, s. 173). Vurderingerne, som skal ligge til grund for, hvornår det er proportionelt at styre udviklingen med reguleringer, er fremlagt i form af 'risikovurderinger', 'omkostning-udbytte analyser' og 'konsekvensanalyser' (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 21–22; P9_TA(2020)0275, 2020, s. 5 & 55).

Udvikling indgår altså i reguleringen af KI i form af eventuelle og aktuelle konsekvenser eller risici af dens udvikling, udbredelse og/eller anvendelse. Konsekvenserne af en digital løsning kan måles på mange forskellige parametre, hvilket også kommer til udtryk i reguleringerne, som fremhæver, at KI har et stort og varieret potentiale. Dermed er princippet om udvikling en etisk retningslinje formet af nytteetik, hvor en handling vurderes på baggrund af dens konsekvens eller udbytte. Altså hvis fortjenesterne opvejer omkostningerne, som kan være potentielle risici, aktuel skade eller brud på etiske principper, kan en handling betegnes som etisk korrekt. Fortjenester kan være både materielle, i form af ressourcer og monetære goder, eller immaterielle, i form af sociale og velfærdsmæssige forbedringer. En vigtig pointe hertil er dog, at KI akten forsøger at modarbejde en ren konsekvensanalytisk tilgang til KI ved at skabe en kategori for 'høj-risiko' KI, som medfører en mere kritisk og omfattende regulering (2021/0106(COD), 2024, s. 111–112 & 248–251).

I reguleringerne bliver udvikling som etisk retningslinje altså inddraget på et nytteetisk grundlag, hvor den potentielle totale produktion af lykke i form af velfærd og vækst, kan forsvare midlerne for at opnå et mere værdigt mål. Dette afspejler derfor en tilgang til teknologien som en balancegang mellem håbefuldst udvikling og varsom risikohåndtering (Howard, 2022, s. 97). Hertil er det dog også muligt at observere en potentiel modsætning mellem de etiske retningslinjer, eftersom den nytteetiske grundantagelse i forhold til udvikling om, hvad der er etisk korrekt, modsiger det pligtetiske rationale bag autonomi som etisk retningslinje. Hvis mennesket er et mål i sig selv, bør det nytteetiske regnestykke aldrig kunne gå op, eftersom maksimering af den totale lykke i samfundet inddrager mennesket som middel og ikke mål.

6.2.3. Ansvarlighed

"Kunstig intelligens skal anvendes ansvarligt" er et krav og etisk princip, som går igen i alle reguleringerne (Finansministeriet & Erhvervsministeriet, 2019, s. 7). Ligesom autonomi har ansvarlighed rod i juridiske mekanismer, hvorved den praktiske håndhævelse af det etiske princip sker gennem lovgivning (P9_TA(2020)0275, 2020, s. 7 & 11). Dette betyder, at det skal være muligt at stille et menneske til ansvar for konsekvenserne af KI, hvilket danner en forbindelse mellem ansvarlighed og lovlighed (Finansministeriet & Erhvervsministeriet, 2019,

s. 64). Det skal derfor være muligt at stille et menneske til ansvar for konsekvenserne og valgene i forbindelse med et KI system, uanset hvilken fase teknologien befinder sig i. Der eksisterer dog uenighed blandt reguleringerne, med hensyn til hvem der kan defineres som ansvarshavende. KI akten specificerer, at den ansvarlige både kan være et naturligt menneske, altså et individ, men også et lovligt individ, i form af organisationer, institutioner og virksomheder (2021/0106(COD), 2024, s. 52), hvorimod den tidligere udarbejdede lovgivning fra EU specificerer, at ”hele ansvaret for beslutningen om at udforme, udvikle, anvende og benytte AI-systemer skal påhvile menneskelige operatører” (P9_TA(2020)0275, 2020, s. 23). Dette understøttes af den danske Nationalstrategi, som bestemmer, at det skal være muligt at stille mennesker til ansvar for konsekvenserne af KI i alle led af udvikling og anvendelse (Finansministeriet & Erhvervsministeriet, 2019, s. 28). Hertil påpeger HLEG, at ansvarlighed ikke kun er noget, som foregår på et individuelt niveau. Selvom det forventes, at den enkelte KI-aktant handler ansvarligt, så er der også kollektiv ansvarlighed i forhold til KI. Eftersom KI består af tekniske og sociale aspekter, tilfalder ansvarligheden for teknologiens potentiale og konsekvenser også det sociotekniske netværk, som den indgår i. Derfor omhandler ansvarlighed ikke blot den enkeltes handlinger, men ansvarlighed under alle faserne i et KI systems levetid, som består af såvel, eksplicite og implicite KI-aktanter (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 5).

Det gøres også til en pointe i reguleringerne, at ansvarlighed ikke blot er en øvelse i at udpege gerningspersoner. Ansvarlighed omfatter også en større øvelse i at uddanne mennesker til at tage del i komplekse overvejelser, hvori de udrustes med en hensigtsmæssig forståelse for KI (2021/0106(COD), 2024, s. 55–56 & 129). Altså bør den menneskelige ansvarlighed i forhold til KI sikres ved, at beslutninger træffes på baggrund af modne og hensigtsmæssige tilgange til KI (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 22–23; P9_TA(2020)0275, 2020, s. 22). Eftersom ansvarlighed ikke blot skal være gældende under anvendelsen af KI, bør man bestræbe sig på at uddanne interessenter med hensyn til udvikling, udbredelse og anvendelse. Til trods for Digitaliseringsstrategiens store fokus på uddannelse og opkvalificering af KI-kompetencer, forbindes det ikke direkte med ansvarlighed, men behandles som et adskilt etisk aspekt af KI. Derimod bliver ansvarlighed udelukkende anvendt i forbindelse med juridiske mekanismer i form af demokratiske

rettigheder, som beskrevet under afsnittet om autonomi (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 8 & 14–16).

For at kunne garantere ansvarlighed skal det være muligt at forklare afgørelser foretaget af KI (P9_TA(2020)0275, 2020, s. 11). Endnu et aspekt af ansvarlighed er derfor afhængigt af ” (...) mulighed for revision, minimering og rapportering af negative virkninger, afvejninger og klageadgang.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 15). Dette kan opnås ved at tilgå KI med velbegrundede metodiske fremgangsmåder, på baggrund af den ovennævnte dannelse af relevante interessenter. Et teknologisk krav hertil er at sikre, at et KI system udstyres med en revisions- eller evalueringsmodel, som muliggør at finde frem til, hvem der står til ansvar for hvilke konsekvenser, en KI måtte producere (Finansministeriet & Erhvervsministeriet, 2019, s. 28). Ansvarlighed kan altså komme til udtryk allerede før og under udviklingen af KI gennem organisatoriske tiltag og udgangspunktet for, hvorfor KI udvikles (Finansministeriet & Erhvervsministeriet, 2019, s. 26). Med hensyn til ansvarlighed for lovlige mennesker kan ansvarlighed derfor også udtrykkes som styringsrammer for en organisation og omfatter, hvordan etiske dimensioner af KI sikres gennem velbegrundede ledelses-, tilsyns- og rådgivningsopgaver. Ansvarlighed kan altså påtages gennem tiltag på ledelsesniveau i organisationer ved at inddrage interne eller eksterne personer og udvalg med særlig indblik i etiske spørgsmål vedrørende KI (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 25; Finansministeriet & Erhvervsministeriet, 2019, s. 31).

Med hensyn til ansvarlighed som et etisk princip bliver de etiske retningslinjer og efterlevelsen af dem beskrevet som en ansvarlig handling i sig selv (Finansministeriet & Erhvervsministeriet, 2019, s. 8). Altså kan ansvarlighed både beskrives som en etisk retningslinje i sin egen ret, jf. de overstående afsnit. Men ansvarlighed kan også anses som en supplerende retningslinje til alle de andre etiske principper, da det sætter krav til gennemførelsen af ”en rationel og metodisk” anerkendelse af relevante interessenter og værdier samt potentielle konflikter og trusler, som KI kan føre til (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 21–22).

6.2.4. Gennemsigtighed

Gennemsigtighed som retningslinje er stærkt betonet af værdien tillid og består af både tekniske og ikke-tekniske krav til alle KI-aktører samt mange aspekter, som er tæt beslægtede med principperne om ansvarlighed og udvikling (P9_TA(2020)0275, 2020, s. 31–32). I

Nationalstrategien fremlægges en kontinuerlig høj tillid til digitalisering og ny teknologi, såsom KI, som et hovedmål, der skal sikre, at alle i det danske samfund er trygge ved udviklingen (Finansministeriet & Erhvervsministeriet, 2019, s. 7 & 25). Hertil præsenteres en række tekniske og ikke-tekniske krav vedrørende gennemsigtighed som udgangspunkt for design af KI, der skal fremme gennemsigtighed under udviklingen af KI.

I HLEG fremlægges gennemsigtighed som forklarlighed, hvorved ”procedurer skal være gennemsigtige [...] og at beslutninger – så vidt muligt – kan forklares til alle, der er direkte eller indirekte berørt.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 13). Årsagen til, at beslutninger ikke nødvendigvis kan forklares i sin helhed, skyldes uigennemskuelige algoritmer (’Blackbox-algorithms’), som det derfor ikke er muligt at give en reel forklaring på som resultat af den høje kompleksitet, der gør dem nær uforklarlige (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 13). Dette kan dog nuanceres ved at betragte tilstedeværelsen af uigennemskuelige algoritmer i KI som et brud på det etiske princip om ansvarlighed. Uigennemskuelige algoritmer og de deraf følgende uforklarlige KI resultater er et spørgsmål om designtilgang. Hvis KI udvikles uden hensyn til forklarlighed, vil efterfølgende positive og negative konsekvenser risikere kun at være forklarlige med at ’den gør sådan, fordi den er designet til at gøre det’ (Kroll, 2018, s. 3). Derimod bør forklarlighed være et designprincip, som er til stede under alle dele af KI, hvorved uigennemskuelige algoritmer er en fejl i KI og ikke en konsekvens af KI (Goebel et al., 2018, s. 296–297; Kroll, 2018, s. 11). KI udviklet med forklarlighed som førende princip kaldes også ’Forklarlig KI’ (XAI), hvor der altid som minimum bør være en generel forståelse for de underliggende mekanismer og netværk. Men grundet strukturen af og opbygningen af neurale netværk, samt den konkrete logik bag de numeriske værdier i sprogmodellers vektorer, er XAI fortsat et kritisk forskningsområde inden for etisk KI (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 23). Som konsekvens af dette perspektiv på uigennemskuelige algoritmer er gennemsigtighed og forklarlighed af KI et resultat af, at de

teknisk ansvarlige KI-aktanter handler etisk korrekt under udviklingen af KI, hvilket blandt andet kan opnås ved brug af ikke-tekniske metoder (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 19). Når gennemsigtighed gøres til et metodisk krav til udviklere af KI, bliver gennemsigtighed et agent-orienteret etisk princip, hvor de tekniske valg bør tages af reflekterende og moralsk handlende agenter på baggrund af ikke-teknologiske metoder, såsom designprincipper, der i større udviklingsvirksomheder besluttet gennem organisatoriske beslutningsprocesser. I tillæg er det vigtigt at se ud over udviklingen af selve sprogmodellerne, med henblik på at gennemsigtighed foregår i løbet af alle teknologiens livsfaser, såsom i forbindelse med en brugervenlig brugerflade, anvendelig metadata og gode muligheder for aktindsigt (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 8; Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 42; P9_TA(2020)0275, 2020, s. 42).

Fokusset på designprocessen er tæt relateret til evalueringsaspektet af ansvarlighed fra forrige afsnit. I forbindelse med gennemsigtighed fremhæver HLEG ”sporbarhed, forklarlighed og kommunikation” som underkategorier, hvorigennem KI ”kan forstås og spores af mennesker” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 15). Ved at relatere gennemsigtighed og ansvarlighed bliver gennemsigtighed et etisk spørgsmål om rettigheder, blandt andet ”retten til god forvaltning og aktindsigt” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 11). Dermed kan gennemsigtighed også defineres som et patient-orienteret etisk princip, eftersom en EU-borger bør have ret til indsigt i KI, ikke fordi det er det etisk korrekte at gøre, men fordi det er en rettighed, som er givet til dem (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 12).

Retningslinjerne opfordrer dog ikke til en total gennemsigtighed, men fremhæver også undtagelserne til princippet som konsekvens af princippet om udvikling. I Nationalstrategien nuanceres gennemsigtigheden af KI: ”Forklarlighed er ikke ensbetydende med fuld transparens omkring algoritmer”, hvilket gøres for at tage hensyn til konkurrenceevnen i erhvervslivet (Finansministeriet & Erhvervsministeriet, 2019, s. 28). Dog holdes der fast i en generel åbenhed og gennemsigtighed ved brug af algoritmer hos offentlige myndigheder, eftersom de ikke arbejder på samme vilkår som det private erhverv. Det samme budskab bliver præsenteret i 2020/2012(INL) som en ”skelnen mellem gennemsigtighed af algoritmer

og gennemsigtighed i brugen af algoritmer” (P9_TA(2020)0275, 2020, s. 11). Hvilket betyder, at KI-aktører altid bør være åbne om, hvornår KI anvendes, men ikke at alle KI-systemer skal være frit tilgængelige og åbne projekter for offentligheden.

Ydermere refererer princippet om gennemsigtighed til rollen af data for KI. Grundet forholdet mellem et KI systems endelige funktionalitet og den data, som det trænes på, er dataetik et vigtigt aspekt i reguleringen af KI (P9_TA(2020)0275, 2020, s. 10). Herunder fremhæves det i de danske reguleringer, at tiltag i forhold til dansksproget data er et særligt fokuspunkt, eftersom der er ”Behov for flere data til kunstig intelligens på dansk” for at sikre gode digitale løsninger på dansk (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 16; Finansministeriet & Erhvervsministeriet, 2019, s. 16). Vigtigheden heraf er flerdelt, i Digitaliseringsstrategien er sikringen af dansk data først og fremmest et spørgsmål om ”effektiv anvendelse af data i høj kvalitet” på et grundlag af ”dataetik, sikkerhed og ansvarlighed” (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 22). Disse punkter går igen i Nationalstrategien, som ønsker at ”skabe bedre adgang til data af høj kvalitet” for at understøtte det videnskabelige miljø, offentlige myndigheder og erhvervslivet (Finansministeriet & Erhvervsministeriet, 2019, s. 9–10 & 17 & 21 & 33–35). I forhold til adgang til data er der ikke blot et fokus på danske data, men datadeling på tværs af landegrænser i EU for at udnytte det fulde potentiale af de enorme mængder data, som eksisterer inden for Unionens grænser (Finansministeriet & Erhvervsministeriet, 2019, s. 39–40).

Gennemsigtighed som etisk princip for KI i Danmark indeholder altså mange referencer til andre etiske principper, især med hensyn til ansvarlighed og udvikling. Det er dog relevant at forholde sig til gennemsigtighed som en separat retningslinje grundet dets metodiske betydning inden for KI. Det vil sige, at gennemsigtighed opstilles i reguleringerne som et designprincip, hvori de andre etiske principper er inkorporeret undervejs i forløbet for at garantere gennemsigtighed på et praktisk niveau. Udvikling som etisk retningslinje omhandler primært den socioøkonomiske indvirkning, som KI har og kan have på samfundet og den etiske betydning deraf. Ansvarlighed indgår primært i reguleringerne grundet den juridiske betydning for, hvem, hvor og hvordan ansvar pålægges. Så hvor udvikling og ansvarlighed er konkrete fænomener, er gennemsigtighed en metode, som danner rammerne for udvikling og ansvarlighed. Derfor er gennemsigtighed en agent-orienteret pligtetisk retningslinje med

henblik på, at KI-aktanter handler i overensstemmelse med den bedste praksis inden for udvikling af KI.

6.2.5. Robusthed

Robusthed præsenteres i KI akten som et organisatorisk og teknisk foretagende, der skal modarbejde udviklingen, udbredelsen og anvendelsen af KI-systemer, som producerer skadelige eller uønskede resultater, under hensyn til den kontekst og det miljø som den befinder sig i (2021/0106(COD), 2024, s. 50). Denne opdeling af robusthed i forhold til KI er klarlagt i rapporten fra HLEG, der præsenterer robusthed som et af de tre bærende komponenter for pålidelig KI: ”den skal være robust fra både et teknisk og et socialt perspektiv, da AI-systemer kan forårsage uforsætlig skade, selv med gode hensigter.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 2). Yderligere defineres robusthed som værende både et formål og et værktøj i arbejdet med KI. Dermed går nogle af de samme mønstre fra gennemsigtighed som etisk retningslinje igen med hensyn til robusthed som etisk retningslinje. Ligesom gennemsigtighed fremlægges i Nationalstrategien som et tillidsopbyggende etisk princip, er robusthed, ifølge HLEG, dét tillidsopbyggende etiske princip. Ydermere påpeges det tekniske og sociale perspektiv som en nødvendig afklaring af omfanget af robusthed: ” (...) fra et teknisk perspektiv (dvs. at systemet er teknisk robust, som det kræves i en bestemt sammenhæng, f.eks. anvendelsesområde eller livscyklusfasen) og fra et socialt perspektiv (under hensyntagen til den sammenhæng eller det miljø, hvori systemet opererer).” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 7).

Ved udlægning af ”princippet om forebyggelse af skade” angives den tekniske robusthed som et bærende element for at beskytte mennesker mod mental og fysiske skade (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 13). Altså præsenteres det tekniske perspektiv af robusthed som en agent-orienteret og pligtetisk praksis, som udføres ved, at der anvendes konkrete indgreb i teknologien for at modarbejde kendte og ukendte potentielle risici. Det sociale perspektiv af robusthed er ligeledes farvet af en agent-orienteret pligtetisk praksis, hvor hensigtsmæssige valg bliver taget med hensyn til den sammenhæng og det miljø, som KI indgår i. Men robusthed er ikke inddraget på et entydigt pligtetisk grundlag. Som følge af rollen, som robusthed spiller for pålidelig KI i HLEG, bliver robusthed også et

spørgsmål om udvikling: ”En pålidelig tilgang er nøglen til at sikre ’ansvarlig konkurrencedygtighed’ idet den kan tilvejebringe et fundament for, at alle, der berøres af AI-systemer, kan have tillid til, at deres design, udvikling, og anvendelse er lovlig, etisk og robust.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 5). Dermed kan robusthed som etisk retningslinje for KI også forsvares ud fra et nytteetisk standpunkt, hvormed robusthed skal sikres som et middel til at opnå målet i form af økonomisk og social vækst (se afsnit 6.2.2).

Fokusset på den tekniske robusthed fremgår også i de andre reguleringer. I 2020/2012(INL) fremgår robusthed hovedsageligt som et spørgsmål vedrørende datastyring med hensyn til personoplysninger, ”der skal træffes vigtige og stærke foranstaltninger for at forhindre sikkerhedsbrud, dataudslip, dataforgiftning, cyberangreb og misbrug af personoplysninger” (P9_TA(2020)0275, 2020, s. 11–12). Denne rammesætning af robusthed som sikker håndtering af data med hensyn til at modarbejde potentiel risiko er et primært fokus for alle de regulerende tekster. Nationalstrategien og Digitaliseringsstrategien fremhæver begge de digitale risici, som er forbundet med den øgede digitalisering. Altså har tilstedeværelsen af risikoskabende elementer, såsom ureguleret eller mangelfuld regulering af KI, skabt behovet for en etisk ramme, hvori robusthed indgår som et grundlæggende etisk princip for etiske retningslinjer (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 20–21; Finansministeriet & Erhvervsministeriet, 2019, s. 17 & 26).

Robusthed i både de nationale og internationale reguleringer omfavner modarbejdelsen af skadelig eller uønsket KI aktivitet, samt sikkerhedsforanstaltninger vedrørende data som indgår i udviklingen og anvendelsen af KI (Finansministeriet & Erhvervsministeriet, 2019, s. 26; 2021/0106(COD), 2024, s. 126). Der opstår dog en differens mellem tilgangene i reguleringerne til, hvordan disse foranstaltninger skal realiseres. De internationale reguleringer indeholder primært restriktive tiltag, som sætter krav til KI-aktørerne, med særligt henblik på de menneskelige KI-aktanter som udvikler, udbreder og anvender høj-risiko KI, hvorfor de også refererer og opfordrer til en øget forskningsindsats og investering i udvikling af robust KI. Derimod tager de nationale reguleringer en mere aktiv rolle, der i Nationalstrategien fremsættes som, at ”Regeringen vil sikre, at data om fx borgernes personoplysninger fortsat håndteres forsvarligt, og at man kan stole på den kunstige intelligens, dens algoritme og

resultater.” (Finansministeriet & Erhvervsministeriet, 2019, s. 27). Altså opfordrer de internationale reguleringer til, at robusthed opnås på baggrund af individuelle KI-aktanters handlinger i deres interaktion med KI, hvor de nationale reguleringer pålægger ikke-statslige KI-aktanter en mere passiv rolle med hensyn til robusthed (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 22; 2021/0106(COD), 2024, s. 125). Det er dog vigtigt, at alle reguleringerne også indeholder positive forpligtigelser og tiltag, som har til formål at fremme udviklingen og anvendelsen af KI. De danske reguleringer refererer kontinuerligt til de offentlige initiativer, som skal understøtte den ønskede samfundsudvikling (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 15 & 17 & 19 & 21 & 23 & 25 & 27; Finansministeriet & Erhvervsministeriet, 2019, s. 27–31). De europæiske myndigheder bestyrer da også flere forskellige initiativer, men disse indgår blot ikke direkte i de regulerende tekster (ELRC, 2024; EPIC, 2018).

6.2.6. Retfærdighed

På tværs af alle reguleringerne bliver objektivitet og upartiskhed præsenteret som det ultimative formål med princippet om retfærdighed (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 11; P9_TA(2020)0275, 2020, s. 40; 2021/0106(COD), 2024, s. 82; Finansministeriet & Erhvervsministeriet, 2019, s. 7). For at opnå retfærdighed er der nogle grundlæggende problematikker, såsom de tekniske aspekter af KI.

Grundelementerne i sprogbasert KI er nemlig baseret på den algoritmiske træning på data, som forsøger at danne mønstre, i form af vektorer, som kan anvendes til at generere sprogmodeller (Kirchmeier et al., 2019, s. 16–19). Dermed indebærer retfærdighed som etisk princip for KI en grundlæggende problemstilling på alle niveauer af teknologien fra de sociale aspekter af anvendelsen af KI til de mere tekniske aspekter, som diskrimination gennem data (Finansministeriet & Erhvervsministeriet, 2019, s. 25).

I forhold til social retfærdighed fremlægger HLEG ”Ligestilling, ikkediskrimination og solidaritet” som grundlæggende rettigheder, der former fundamentet for retfærdighed som en etisk retningslinje for KI (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 11). HLEG fremhæver da også i den forbindelse, at hvis KI ikke udvikles på baggrund af princippet om retfærdighed, vil der være risiko for at udstøde sårbare personer og grupper, da

der vil risikere at blive genereret negative forudindtagelser (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 11 & 13). Grundet de potentielle farer ved denne problemstilling bliver retfærdighed fremsat som et af de fire grundlæggende etiske principper, som HLEG opfordrer til, at videre KI udvikling og regulering udarbejdes på baggrund af (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 12). Mere præcist defineres retfærdighedsprincippet som et krav til KI med ordene ”Diversitet, ikkediskrimination og retfærdighed (...) Herunder undgåelse af urimelig skævhed, tilgængelighed og universelt design samt inddragelse af interessenter” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 15). Dette stemmer overens med Nationalstrategiens bud på et konkret værktøj for at sikre retfærdighed i KI i form af ”demografisk og faglig diversitet” i forbindelse med KI arbejde (Finansministeriet & Erhvervsministeriet, 2019, s. 28). HLEG rapporten henviser da også til, at retfærdighed er den etiske baggrund for et universelt menneskecenteret designprincip, som styrer KI arbejde i hele dets livscyklus (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 20 & 25–26).

For at kunne opnå diversitet under udarbejdelsen af KI er der blandt andet behov for ligestilling i erhvervet med hensyn til kvindelige it-specialister. Digitaliseringsstrategien fremhæver, at situationen i Danmark er, at det kun er hver fjerde it-specialist i Danmark, som identificerer sig som kvinde. Ydermere ligger Danmark på sjettepladsen i EU, når det kommer til andelen af kvindelige it-specialister (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 10). Der er altså et stort behov for diversitet blandt udviklere af KI, og der er stadig lang vej igen, især når landet med den sjette bedste fordeling i forhold til køn har så skæv en kønsfordeling. Denne problematik er da også baggrunden for Digitaliseringsstrategiens fokus på uddannelse af it-specialiser og generel teknologiforståelse i Danmark (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 14). Ud over dette punkt forholder digitaliseringsstrategien sig ikke eksplicit til retfærdighed som et etisk princip, men behandler det som en underkategori af autonomi og ansvarlighed. Retfærdighed forbindes også med de etiske principper autonomi og ansvarlighed med hensyn til værdighed og rettigheder i 2020/2012(INL), som dog også behandler retfærdighed som et selvstændigt etisk princip for KI (P9_TA(2020)0275, 2020, s. 7 & 12). Ud fra dette perspektiv er retfærdighed en pligtetisk

bestræbelse, som både indeholder aspekter af en agent-orienteret og patient-orienteret pligtetik. Det agent-orienterede perspektiv tager udgangspunkt i, at retfærdighed kan være en aktivitet, som bliver fremmet ved at moralske agenter stræber efter at forbedre demografisk og faglig diversitet. F.eks. fremlægger Nationalstrategien retfærdighed som en aktiv handling i form af "Lighed og retfærdighed" inden for KI, som indebærer modarbejdelse af uønskede fordomme og bestræbelser på at undgå "kategorisering, som diskriminerer på baggrund af fx etnicitet, seksualitet og køn." (Finansministeriet & Erhvervsministeriet, 2019, s. 28). Defineres retfærdighed derimod som en grundlæggende rettighed, vil det ikke være et spørgsmål som primært omhandler handlingerne, som udføres af KI-aktanter på baggrund af deres indre moralitet, men som et universelt krav, der er givet til dem, som oplever direkte eller indirekte indflydelse fra KI.

6.2.7. Holisme

Dette sidste punkt bør ikke ses som en etisk retningslinje for KI men har til hensigt at præcisere og klargøre, hvordan reguleringerne mener, de etiske principper bør anvendes i praksis. Det er blevet omtalt i løbet af de forrige afsnit, hvilke forventninger der er til hvad, hvornår og hvem principperne er gældende for, men det vil blive sammenfattet i dette afsnit.

Under behandlingen af hver af de etiske principper er det på den ene eller anden vis blevet beskrevet, hvordan reguleringen er gældende under "udvikling, udbredelse og anvendelse" af KI, hvilket er en formulering som går igen i alle reguleringerne (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 11; Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 2; P9_TA(2020)0275, 2020, s. 3; 2021/0106(COD), 2024, s. 11–12; Finansministeriet & Erhvervsministeriet, 2019, s. 28–29). I KI akten og 2020/2012(INL) er disse faser af KI mere konkret beskrevet som 'KI værdikæden', som omfatter sikring af nøjagtighed, robusthed og sikkerhed under alle livscykler i KI (P9_TA(2020)0275, 2020, s. 19; 2021/0106(COD), 2024, s. 66). Det er ydermere gældende for aktanter indblandet i udvikling, import, udbredelse, produktion, autorisering, evaluering og under påvirkning af KI i og uden for EU, der har til hensigt at udvikle, udbrede eller anvende en pågældende KI inden for EU's grænser (2021/0106(COD), 2024, s. 94). Disse aktanter bliver refereret til med begreberne 'KI-aktører' eller 'KI-interessenter', som interagerer med eller påvirkes af KI på baggrund af direkte

eller indirekte konsekvenser af alle processer og funktioner under hele teknologiens livscyklus (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 6).

Nationalstrategiens og digitaliseringsstrategiens reguleringer af KI bliver sat i direkte relation til fire primære grupper af interessenter i form af forskere, virksomheder, det offentlige og den enkelte borger (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 5; Finansministeriet & Erhvervsministeriet, 2019, s. 8–10). Alle disse aktanter er hver især i besiddelse af forskellige etiske positioner og praktiske roller med hensyn til KI. Der henvises endvidere til den etiske ansvarlighed med hensyn til, at ”det bredere samfund, andre følede væsener og miljøet anses for interessenter i hele AI-systemets livscyklus.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 20). Dette omfatter også bæredygtighed og økologi, i den udstrækning det er relevant for den pågældende KI, samt hensyn til fremtidige generationer i forhold til ressourceanvendelse og energiforbrug (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 21). Dette er også et perspektiv, som behandles i digitaliseringsstrategien, som følge af fokuset på mulighederne med KI i form af håndtering af klimaudfordringerne og effektivisering af den grønne omstilling. I forhold til anvendeligheden af KI fremhæves initiativer vedrørende ”grøn databehandling- og opbevaring” (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 5 & 8 & 18).

Det enorme omfang af KI-aktanter gør, at HLEG opfordrer til, at digitale løsninger og KI undergår en kontinuerlig evaluering og retfærdiggørelse gennem hele teknologiens liv for at kunne sikre, at de etiske principper overholdes (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 22 & 26–28). Der er dog undtagelser til denne systematiske gennemgang af KI under alle faser og facetter af dets livscyklus, blandt andet kan der være domæne-specifikke KI, hvortil der må tages kontekstspecifikke hensyn (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 6). Domæne-specifik KI omfatter blandt andet sprogmodeller med en meget snæver anvendelseskontekst, som eksempelvis indeholder meget præcise regler for, hvad den kan og skal (Patel et al., 2023, s. 2).

”Indsatsen for at opnå pålidelig kunstig intelligens omfatter derfor ikke kun selve AI-systemets pålidelighed, men kræver en holistisk og systemisk tilgang, der omfatter pålideligheden af alle aktører og processer, der er en del af systemets sociotekniske sammenhæng i hele dets livscyklus.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 5).

Eftersom KI eksisterer i foranderlige sociotekniske miljøer, som ændrer sig konstant i takt med konteksten, de er i, vil der altid være undtagelser eller mangler i praksis, når det kommer til evalueringer eller efterlevelse af de etiske principper (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 12). Dette gør, at man i løbet af langt de fleste KI systemers levetid vil opleve spændinger mellem principperne, hvor nogle principper må brydes, for at andre kan overholdes, for eksempel ved at begrænse brugerens autonomi for at sikre robusthed og retfærdighed, eller ved at gå på kompromis med gennemsigtighed for at understøtte virksomheders konkurrenceevne i udviklingen af KI (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 14; Finansministeriet & Erhvervsministeriet, 2019, s. 28).

Det kan derfor konkluderes, at de etiske retningslinjer for KI i Danmark forsøger at tage højde for mange KI-aktanter med direkte eller indirekte forbindelse til KI. Uanset om man er udviklende virksomhed eller individuel forbruger af KI, er det relevant at forholde sig til, hvordan myndighedernes retningslinjer og reguleringer opstiller positive og negative forpligtigelser til individet og samfundet som helhed. Dertil medregnes de komplekse forhold ved at regulere sociale fænomener med hensyn til at måtte bryde med nogle principper for at kunne overholde andre. Men på baggrund af den foreløbige analyse kan det konkluderes, at de etiske retningslinjer for KI i Danmark består af en holistisk ramme for, hvordan KI-aktanter bør efterleve principper vedrørende autonomi, udvikling, ansvarlighed, gennemsigtighed, robusthed og retfærdighed.

6.3. Sprogteknologi.dk

De efterfølgende afsnit er en analyse af metadata for tekst- og lydressourcer, som er tilgængelige på sprogteknologi.dk. Metadata er data om data, altså består metadata af oplysninger om data (Kirchmeier et al., 2019, s. 16). Undersøgelsen består af en kvantitativ analyse af ressourcernes metadata med henblik på at opnå indblik i, i hvilken grad initiativet sprogteknologi.dk opfylder de ovenstående retningslinjer for KI. Endvidere inddrages de specifikke retningslinjer, som initiativet selv refererer til, med hensyn til deres formål om at fremme udviklingen af dansk sprogteknologi ved hjælp af tilgængeliggørelse af store mængder data og metadata af høj kvalitet (Digitaliseringsstyrelsen, 2024e, 2024f). Årsagen til

fokusset på metadata er begrundet i undersøgelsen af initiativets anvendelighed som en brugervenlig udviklingsressource (Kirchmeier et al., 2019, s. 27). Dette er et velegnet kriterium til vurdering af anvendeligheden for KI-aktanter, i kraft af hvor lettilgængelige sprogressourcerne er på sprogteknologi.dk, uden at der er behov for at nærstudere de individuelle sprogressourcer. Vigtigheden af metadata fremgår da også på sprogteknologi.dk, hvor indholdet på platformen beskrives som en samling af metadata om danske sprogressourcer. Altså fungerer sprogteknologi.dk som mellemmand for sprogressourcer ved at forbinde udbydere og forbrugere af sprogressourcer, hvor sprogteknologi.dk gennem metadata skaber kontakt mellem KI-aktanter (Digitaliseringsstyrelsen, 2024d, 2024e).

Analysens datagrundlag og hele R projektet ligger tilgængeligt i henholdsvis bilag C og bilag D med supplerende forklaringer og begrundelser for de kvantitative indgreb i manipuleringen og visualiseringen af data fra sprogteknologi.dk (Bilag C, s. 12; Bilag D, s. 13–27).

6.3.1. Baggrunden for sprogteknologi.dk

I Nationalstrategien begrundes behovet for en fælles dansk sprogressource med, at man i Danmark må følge med den stigende udvikling af digitale løsninger (Finansministeriet & Erhvervsministeriet, 2019, s. 37). Problematikken fremlægges som et spørgsmål om det danske sprogs anvendelse til udvikling og formidling af KI i fremtiden på et nationalt og internationalt niveau:

”I takt med, at der udvikles stadig mere avancerede digitale løsninger på engelsk, som vi kan interagere med, er der en risiko for, at det i stigende grad kommer til at foregå på engelsk frem for dansk. Skal borgere og danske virksomheder fortsat have adgang til og mulighed for at udvikle gode løsninger på dansk, kræver det, at dansk står som et stærkt sprog i den internationale konkurrence.” (Finansministeriet & Erhvervsministeriet, 2019, s. 36).

I praksis er sprogteknologi.dk en database over ressourcer, som kan understøtte udviklingen af dansk sprogteknologi. Det overordnede mål med sprogteknologi.dk beskrives på hjemmesiden som at ”mindske barriererne og styrke udviklingen af sprogteknologiske løsninger til understøttelse af kunstig intelligens på dansk af høj kvalitet.” (Digitaliseringsstyrelsen, 2024a). Dette formål stemmer overens med målsætningerne i Nationalstrategien og Digitaliseringsstrategien samt Digitaliseringspagten fra 2019, som blev

udarbejdet i et samarbejde mellem Regeringen, Kommunernes Landsforening og Regionerne (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 16; Finansministeriet & Erhvervsministeriet, 2019, s. 37; Regeringen et al., 2019, s. 4).

Ud over de etiske retningslinjer, som blev klarlagt i forrige analyse, kan sprogteknologi også vurderes på baggrund af dens efterlevelse af selverklærede standarder og principper, målrettet databaser. Der refereres til den europæiske standard "Data Katalog Vokabular Applikations Profil" (DCAT-AP) "Metadata om sprogteknologiske ressourcer er i overensstemmelse med DCAT-AP-specifikationen." (Digitaliseringsstyrelsen, 2024f). DCAT-AP-specifikationerne er en standardisering af, hvilke termer og koder der anvendes til at opbygge databaser for at kunne mindske samarbejdsvanskeligheder på tværs af sprogområder (Nuffelen, 2024). Standardiseringen bygger blandt andet på "Open Data Goldbook for Data Managers and Data Holders – Practical guidebook for organisations wanting to publish Open Data" (herefter kaldt 'Åben data rapporten'), som er en rådgivende rapport fra Den Europæiske Dataportal (Carrara et al., 2019). I Åben data rapporten fremlægges fem principper for, hvordan offentlige myndigheder bør handle i forhold til håndtering og deling af offentligt tilgængelige data; det er dog vigtigt at påpege at retningslinjerne for datastyring bygger på det Europæiske direktiv om åben data og genanvendelse af informationer fra den offentlige sektor, som er vedtaget ved lov (Europa-Parlamentets og Rådets direktiv (EU) 2019/1024 af 20. juni 2019 om åbne data og videreanvendelse af den offentlige sektors informationer (omarbejdning), 2019). Første princip 'åbent data som udgangspunkt' betyder, at al data fra den offentlige sektor som udgangspunkt bør være tilgængeligt for offentligheden. Andet princip vedrører kvaliteten og kvantiteten af data, som bør blive offentliggjort i sin helhed samt være understøttet af forståelig og fyldestgørende metadata. Tredje princip omhandler tilgængeligheden af den offentliggjorte data, som bør være brugbar for alle menneskelige og ikke-menneskelige aktanter. Det fjerde princip opfordrer offentlige myndigheder til at offentliggøre den tilegnede tekniske ekspertise i takt med den offentlige sektors anvendelse af digitale løsninger. Sidste princip er rettet mod, at lande med en særlig økonomisk styrke går forrest med at investere i udvikling af dataforståelse og KI (Carrara et al., 2019, s. 10 & 13). Ud over at dette giver endnu et sæt principper, som sprogteknologi.dk kan ses i lyset af, viser det også, at initiativet er

udarbejdet med hensyn til projektets virke i det større samfund ved at have en områdespecifik tilgang til udviklingen af projektet. Anvendelsen af DCAT-AP kan betragtes som en efterlevelse af princippet om ansvarlighed og gennemsigtighed, eftersom de tekniske aspekter af sprogteknologi.dk er lavet ud fra en fagligt funderet og genanvendelig tilgang. Dertil opfylder sprogteknologi.dk også princippet om udvikling som følge af, at platformen er baseret på en international standard, hvormed den kan indgå i internationalt samarbejde om udarbejdelse af sprogteknologi. Derfor bliver ressourcerne på sprogteknologi lettere at tilgå for flere KI-aktanter.

Det anerkendes dog også, at sprogteknologi.dk er et større projekt, som forsøger at understøtte den generelle digitalisering i Danmark ved at medvirke i licensarbejde og ved at tilbyde et datagrundlag ”i konstant udvikling.” (Digitaliseringsstyrelsen, 2024a, 2024f). Derfor er den kontinuerlige udvikling og udbygning et vigtigt perspektiv i undersøgelsen af sprogteknologi.dk ”Den digitale platform vil løbende blive udbygget med nye metadata om relevante, eksisterende sproressourcer.” (Digitaliseringsstyrelsen, 2024a).

6.3.2. Ressourcernes metadata

I forbindelse med den konstante tilføjelse af data og metadata til sprogteknologi.dk er det givende at se på omstændighederne for det kontinuerlige arbejde. På undersiden ”Optagelse i kataloget” ligger hovedfokusset på, at organisationerne som bidrager til sprogteknologi.dk. bestræber sig på at supplere deres sproressourcer med fyldestgørende og opdateret metadata (Digitaliseringsstyrelsen, 2024d). Mere konkret refereres der under betingelserne for at blive en bidragende organisation til etiske, juridiske og teoretiske retningslinjer for, hvordan sproressourcer bør eksistere på sprogteknologi.dk (Digitaliseringsstyrelsen, 2024c). Med hensyn til de etiske retningslinjer omtales ”Etiske anvendelse af sproressourcer”, som afspejler alle de etiske retningslinjer fra forrige analyse af de etiske retningslinjer for KI i Danmark. ”Udviklingen af sprogteknologiske løsninger rummer store muligheder for innovation.” stemmer overens med retningslinjen for udvikling (Digitaliseringsstyrelsen, 2024c). ”Både dataudstillere og dataanvendere opfordres til aktivt at forsøge at undgå bias (...)” efterlever retningslinjerne for retfærdighed og ansvarlighed samt tager højde for den holistiske ramme for udvikling af KI ved at henvende sig til flere forskellige KI-aktanter

(Digitaliseringsstyrelsen, 2024c). ”Brugere af sprogteknologi.dk forventes til en hver tid at anvende data på en etisk forsvarlig måde, som på ingen måde kan formodes at vildlede eller volde skade.” opfordrer til, at retningslinjerne om autonomi og robusthed efterleves (Digitaliseringsstyrelsen, 2024c). Endeligt relateres de etiske retningslinjer til gennemsigtighed og ansvarlighed ved at opfordre KI-aktanter til at være tydelige ” (...) når brugere interagerer med en maskine og ikke et levende menneske.” (Digitaliseringsstyrelsen, 2024c).

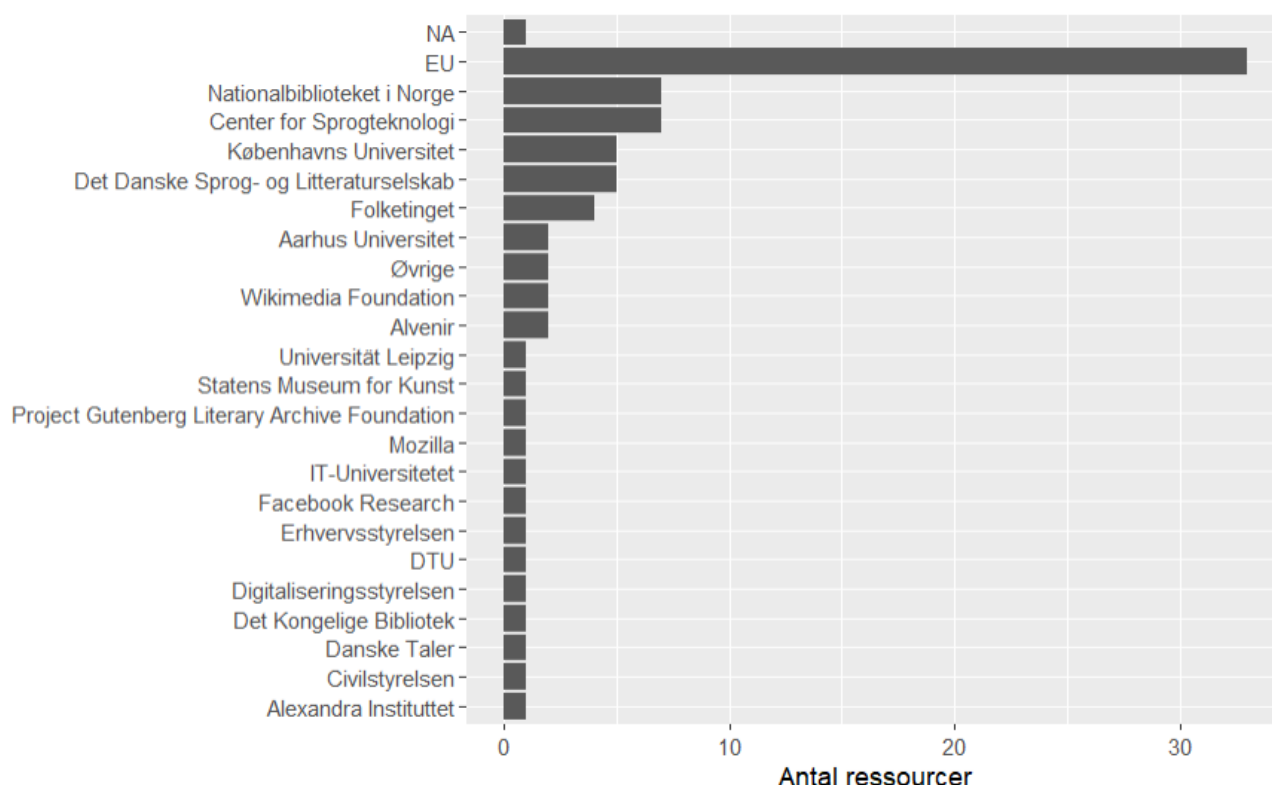
De juridiske retningslinjer er udarbejdet med henblik på betingelserne for anvendelse af de sprogressourcer, som gøres tilgængelige på sprogteknologi.dk. Hertil opfordres organisationerne til at informere brugerne af sprogteknologi.dk om, hvilken licens deres sprogressource hører under. I overensstemmelse med det overordnede formål med sprogteknologi.dk opfordres organisationerne til at gøre deres data så tilgængelig som muligt ved at underlægge deres data ”en så åben licens som muligt (se fx Creative Commons).” (Digitaliseringsstyrelsen, 2024c). Dermed forsøger sprogteknologi.dk gennem de generelle betingelser for hjemmesiden at understøtte den åbne deling af data og udvikling af KI i Danmark ved at spore bidragende organisationer i retning af mindre restriktive licenser.

Den teoretiske vinkel for håndtering af metadata bliver givet i form af FAIR-principperne, som sprogteknologi.dk forsøger at efterleve ved både selv at opfylde principperne og pålægge bidragende organisationer samme værdier (Digitaliseringsstyrelsen, 2024c). FAIR-principperne beskrives på sprogteknologi.dk ved at data skal gøres brugbare som følge af, at de er ”findbare”, ”tilgængelige”, ”systemuafhængige” og ”anvendelige – så både mennesker og computere kan læse dem” (Digitaliseringsstyrelsen, 2024c). For at kunne leve op til disse kriterier for data og metadata er der behov for en metodisk og konsekvent tilgang til oprettelsen af sprogressourcerne og deres tilhørende metadata. Men til trods for at sprogteknologi.dk opstiller forskellige retningslinjer for oprettelsen af sprogressourcer på deres side, forekommer der alligevel kritiske mangler på tværs af sprogressourcernes metadata. Manglerne inkluderer 43 tilfælde af manglende dokumentation, ni tilfælde af manglende licens og endda tre manglende destinationssider på sprogressourcerne, samt manglende oplysning af organisation i et tilfælde (Bilag D, s. 16). Ydermere kan der også forekomme fejl i den oplyste metadata, såsom ”Referater fra Folketinget”, der indeholder en

destinationsside, men leder til en blank side (Bilag C, linje 9). Denne kritik af sprogressourcerne på sprogteknologi.dk bør ikke ses som en underkendelse af initiativets anvendelighed, men en kritik rettet mod den arbejdsmodel, som projektet er baseret på. Her refereres der til, at den oplyste metadata på sprogteknologi.dk er udarbejdet centralt af sprogteknologi.dk, og man anmoder den bidragende organisation om at validere den fremfundne metadata, inden den offentliggøres. Efterfølgende er det dog et fælles ansvar mellem sprogteknologi.dk og den bidragende organisation at opdatere den offentliggjorte metadata. Dette betyder, at vedligeholdelsesarbejdet vedrørende metadata bliver decentraliseret, hvilket medfører, at metadata bliver mindre konsekvent og systematisk på tværs af sprogressourcerne (Digitaliseringsstyrelsen, 2024d). En vigtig nuancering af denne kritik er, at der kun blev identificeret to kritiske tilfælde af manglende metadata ud af 83, hvor det ikke var muligt at få adgang til sprogressourcen gennem sprogteknologi.dk (Bilag C, linje 9 & linje 75).

6.3.3. Organisationerne bag ressourcerne

Ved at undersøge hvilke organisationer, der står bag ressourcerne på sprogteknologi.dk, er det muligt at opnå indsigt i, hvor mangfoldige ressourcerne er. Dette parameter kan dermed indgå i vurderingen af, i hvor høj grad sprogteknologi.dk lever op til den etiske regulering af retfærdighed gennem diversitet. Grunden til, at retfærdighed er relevant for sprogressourcerne, er, at den endelige model er et produkt af det data, som den er blevet trænet på. Derfor vil sprogbasert KI trænet på et datasæt kendetegnet ved mangfoldige data give et bedre grundlag for ikke-diskriminerende KI. De forskellige organisationer bag sprogressourcerne på sprogteknologi.dk kan ses på graf 1 herunder (Bilag D, s. 17-18):



Graf 1. Et horisontalt søjlediagram over, hvilke organisationer, har lagt ressourcer ud på sprogteknologi.dk, samt hvor mange individuelle tekst- og/eller talebaserede datasæt hver organisation har gjort tilgængelige.

Som det fremgår på grafen, er den største bidrager til sprogteknologi.dk den internationale myndighed, Den Europæiske Union, som har fået tilgængeliggjort 33 sprogressourcer. Dette kan sammenlignes med i alt syv sprogressourcer fra danske myndigheder som Regeringen, Digitaliseringsstyrelsen, mv. De resterende bidragende organisationer består af institutioner for uddannelse og kultur, samt syv private virksomheder og aktører, dog med langt færre antal sprogressourcer per bidrager. Den største andel af sprogressourcerne kommer altså fra forskellige organisationer, som hver især kan bidrage med den type data og det perspektiv, som de er i besiddelse af, hvilket understøtter diversiteten på sprogteknologi.dk. Dog stammer en markant andel af sprogressourcerne fra én bidrager, EU, hvilket har den modsatvirkende effekt og øger risikoen for at mønstre og tendenser i de europæiske sprogressourcer kan blive toneangivende.

En vigtig pointe hertil er som pointeret i forrige afsnit, at der forekommer mangler og uklarheder i metadata på sprogteknologi.dk. To af sprogressourcerne står under

organisationen 'Øvrige', som ikke giver et tydeligt indblik i, hvem der faktisk står bag den offentliggjorte data. Det er dog muligt at opnå indblik i, hvem organisationerne er, ved at undersøge destinationssiderne for sprogressourcerne. Her bliver det tydeligt, at 'Øvrige' anvendes som en organisationskategori for private individer, som har bidraget med sprogressourcer uden at være relateret til en organisation. Ydermere er én af sprogressourcerne oplyst uden nogen organisation, hvilket resulterer i værdien 'NA', som betyder, at der ikke er nogen værdi. Igen kan det ved nærmere undersøgelse af sprogressourcen ses, at det er et bidrag fra Wikimedia Foundation. Men ud over navnet på sprogressourcen 'Dansk Wikiquote' er der ikke nogen metadata på sprogteknologi.dk, som blot leder til en tom side (Digitaliseringsstyrelsen, 2024b). Dermed fejler sprogteknologi.dk i enkelte tilfælde den anden retningslinje for data i Åben data rapporten med hensyn til kvaliteten af metadata, som bør være fyldestgørende for den tilgængelige data. Dette går også imod gennemsigtighedsprincippet, i forhold til at ukomplet data modarbejder tilgængeligheden og kvaliteten af en database. Det er dog kun tre ud af 83 sprogressourcer, som har disse problemer, og det er ikke ensbetydende med, at sprogteknologi.dk fejler i sit overordnede mål om at samle og tilgængeliggøre data til udvikling af dansk KI-teknologi. En af årsagerne til denne problematik kan blandt andet have rod i arbejdsmodellen, som blev nævnt i forrige afsnit, med hensyn til det kontinuerlige arbejde med metadata og opdatering af sprogressourcer (Digitaliseringsstyrelsen, 2024a).

Altså består sprogressourcerne på sprogteknologi.dk af data fra forskellige afsendere, der får mulighed for at blive inkluderet i fremtidig udvikling af dansk sprogteknologi. Dermed opfylder sprogteknologi.dk de etiske principper for retfærdighed og gennemsigtighed, til trods for enkelte mangelfulde eksempler.

6.3.4. Metadata og indhold

For at skabe et overblik på tværs af sprogressourcernes data analyseres tre parametre af den indsamlede metadata: 'navnene' på sprogressourcerne samt de 'tags' (herefter kaldt 'nøgleord') og 'emner', der er blevet anvendt til at beskrive sprogressourcernes indhold. Først fokuseres der på at give et indblik i navnene, og derefter behandles nøgleordene og til sidst emnerne.

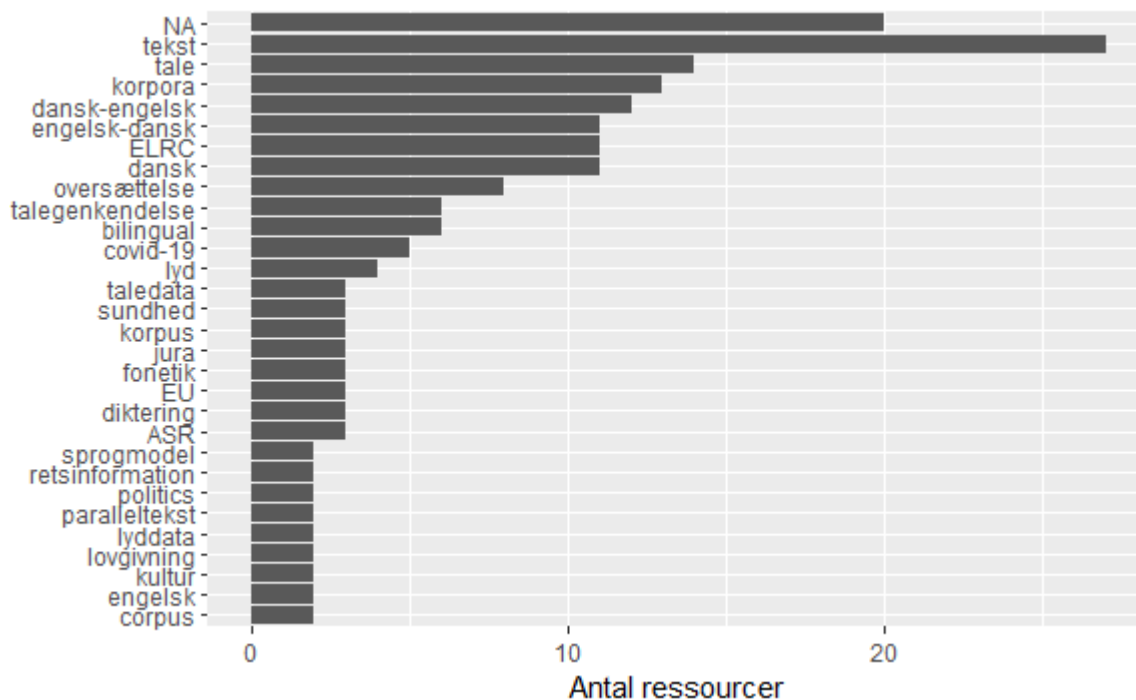


Figur 4. Ordsky af ord som forekommer mere end 3 gange på tværs af navnene på sprogressourcerne på sprogteknologi.dk. Ikke-betydningsbærende ord er filtreret fra, se Bilag D for yderligere forklaring.

Figur 4 er en ordsky over de mest hyppigt forekommende betydningsbærende ord på tværs af alle navnene på sprogressourcerne (Bilag D, s. 21-22). Dette er opnået ved at fragmentere navnene på sprogressourcerne og frasortere såkaldte stopord (Bilag D, s. 19-21). Ordskyen er derfor ikke retvisende i forhold til, hvilke ord der egentligt er mest hyppige i datasættet, eftersom fyldord såsom 'og', 'from' og 'of' ikke er inkluderet. Størrelsen på ordene i ordskyen indikerer hvilke af ordene forekommer hyppigst, det vil sige at ordet 'danish' optræder flest gange på tværs af navnene på sprogressourcerne. Tre

andre hyppigt forekommende ord som er interessante at fremhæve i relation til 'danish' er ordene: 'english', 'bilingual' og 'parallel'. Disse hyppigt forekommende ord refererer alle til, hvilke sprog sprogressourcerne indeholder og kan sammenlignes med metadata om, hvilke sprog sprogressourcerne indeholder. Optælles sprogene på tværs af sprogressourcernes metadata er dansk det klart hyppigst forekommende sprog, men dette er kun tilfældet, når sprogene tælles individuelt. Laves der derimod en optælling af alle de sprog, som indgår i sprogressourcerne, fremgår det, at 30 sprogressourcer består af både dansk og engelsk, hvorimod 23 sprogressourcer kun består af dansksproget data (Bilag D, s. 22-24). Den store tilstedeværelse af danske sprogressourcer er dog ikke nogen overraskelse, eftersom sprogteknologi.dk som sagt har til formål at fremme udviklingen af danske sprogmodeller. Årsagen til, at det fremgår på engelsk, skal findes i den store andel af sprogressourcer fra EU. Det er da også bidragene fra EU, som forårsager den hyppige forekomst af 'bilingual' og 'website', der henviser til, hvor sprogressourcerne stammer fra, hvilket er flersprogede europæiske og danske hjemmesider (Bilag C, linje 57 – 74). Endeligt er 'corpus' det engelske udtryk for korpus, som er en samling af sprogligt data, som refererer til hvilken type data, sprogressourcerne består af.

Som nævnt indeholder sprogressourcernes metadata også nøgleord, der kan give et umiddelbart indblik i, hvad en sprogressource består af. Ydermere medvirker disse nøgleord også i fremsøgningen af sprogressourcer på sprogteknologi.dk, hvilket underbygger vigtigheden af, at de er retvisende og relevante for den tilhørende data. Dette er blevet visualiseret herunder (Graf 2; Bilag D, s. 24-25).



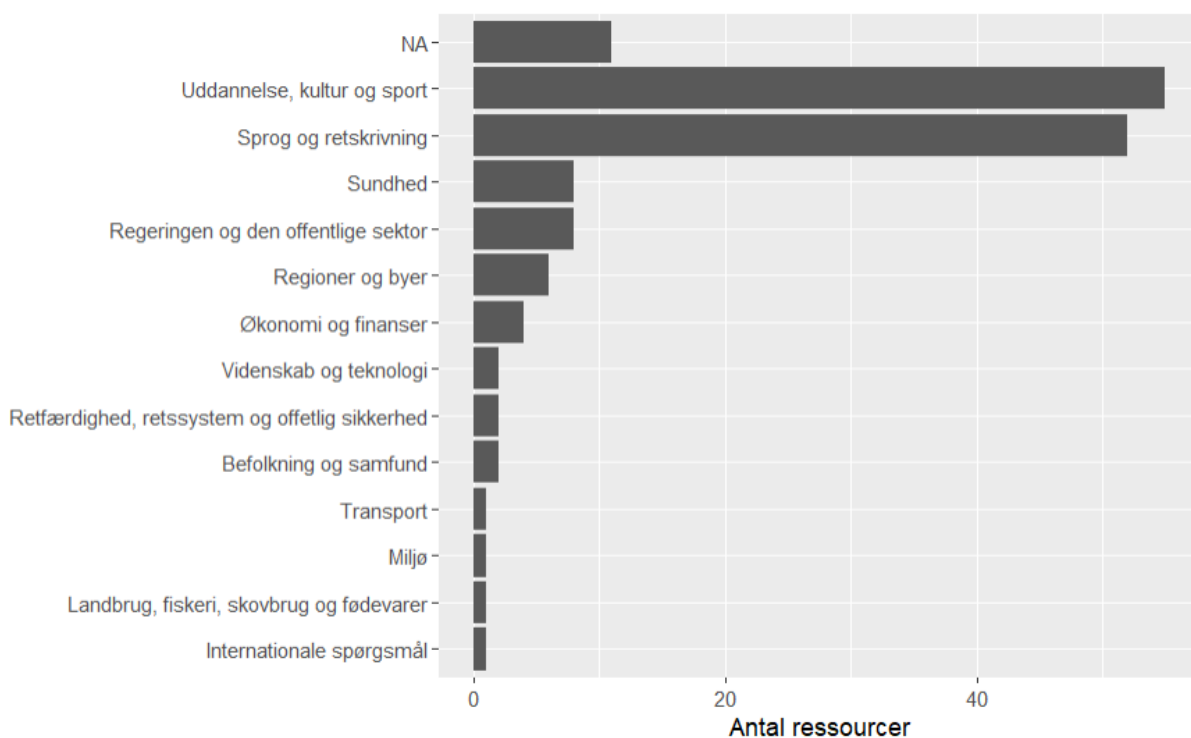
Graf 2. En graf over nøgleordene som de bidragende organisationer har angivet for at give brugere af sprogteknologi.dk indblik i sprogressourcens indhold.

Først og fremmest er der manglende metadata på en fjerdedel af sprogressourcerne, der ikke har fået tilføjet nogen nøgleord (Graf 2). Dernæst er det hyppigste nøgleord 'tekst', hvilket tyder på, at størstedelen af sprogressourcerne er tekstbaseret frem for at være fokuseret på lyddata, som dog også fremgår i grafen som det tredje mest hyppigt anvendte nøgleord. Herefter fremgår en række nøgleord, som refererer til danske sprogmodeller med henblik på anvendelse mellem fremmedsprog, mere specifikt engelsk. Nøgleordet 'ELRC' er en forkortelse for det europæiske projekt "European Language Ressource Coordination". ELRC omhandler styring af sprogressourcer af europæiske sprog på tværs af sprogområder med henblik på at understøtte et flersprogligt Europa (ELRC, 2024). Dette fremhæver konsekvensen af, at en større mængde data stammer fra EU, som ikke nødvendigvis er

målrettet udvikling af en enkeltsproget dansk sprogmodel, men derimod flersproget KI.

Fundene fra graf 2 stemmer derfor generelt overens med pointerne fra ordskyen over sprogressourcernes navne, den store andel af sprogressourcer fra EU og omstændighederne for metadata på sprogteknologi.dk i forhold til tendensen med manglende metadata.

Indtil videre har undersøgelsen af metadata givet indblik i sprogressourcernes navne og nøgleord, til sidst vil analysen fokusere på emnerne som sprogressourcerne omhandler. På baggrund af disse kategorier er det muligt at undersøge, hvilken overordnet retning sprogressourcerne på sprogteknologi.dk trækker udviklingen i, eftersom sprogmodeller er produkter af den data, som de er baseret på (Graf 3).



Graf 3. Visualisering af sprogressourcernes emner. Hver sprogressource på sprogteknologi.dk er blevet tildelt op til fire emner, som indikerer hvilket formål ressourcen opfylder i udviklingen af KI.

Især to emner er repræsentative for, hvilke sprogressourcer der er mest almindelige på sprogteknologi.dk: 'Uddannelse, kultur og sport' og 'Sprog og retskrivning' (Bilag D, s. 25-26). Til trods for at der ikke er en direkte forklaringsmodel for metadata, som er gjort tilgængelig for offentligheden, er det tydeligt, at emnerne refererer til anvendelsen af sprogressourcerne. Det første emne 'Uddannelse, kultur og sport' tilskrives data, som omhandler sociale aspekter.

Den hyppige forekomst af dette emne stemmer overens med målsætningerne om udvikling af dansksproget KI i nationalstrategien og digitaliseringsstrategien. De nationale strategier fremhæver vigtigheden af at udvikle og fremme dansk KI baseret på **”vores fælles værdier”** og **”afspejling af danske interesser”** (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 26; Finansministeriet & Erhvervsministeriet, 2019, s. 7). Dermed efterlever sprogteknologi.dk altså den politiske retningslinje for KI ved at have en stor grad af data, der indeholder og afspejler danske værdier. Den anden kategori 'Sprog og retskrivning' omhandler med al sandsynlighed data, som indeholder grammatisk og sprogligt materiale på dansk, med henblik på at systematisere afkodning og formulering af sprogdata. Dette er grundlæggende for at skabe velformulerede sprogmodeller, eftersom mere dansk data af høj kvalitet er afgørende for at kunne opnå gode færdigheder inden for det danske sprog (Kirchmeier et al., 2019, s. 18–19 & 27).

Altså består sprogressourcerne på sprogteknologi.dk i en relativt høj grad af data, som henvender sig til udviklingen af dansksprogede modeller på baggrund af flersproget data. Dette er blandt andet en konsekvens af de store bidrag fra EU med forbindelse til ELRC, som har til formål at fremme et flersproget EU (ELRC, 2024). Derfor lægger sprogressourcerne på sprogteknologi.dk ikke kun op til udviklingen af rene danske sprogmodeller, men indebærer også et stort potentiale for udvikling af sprogmodeller med henblik på dansk i forhold til engelsk og omvendt. På baggrund af den varierende manglende metadata er der desværre en risiko for, at visualiseringerne og konklusionerne på baggrund af metadata ikke er retvisende for de tilgængelige sprogressourcer. Disse usikkerheder har blandt andet rod i arbejdsbyrden i vedligeholdelsen af sprogressourcernes metadata, hvilket er en naturlig konsekvens af at have et samarbejde mellem flere aktanter. Ud over den mulige påvirkning af den nærværende undersøgelse af sprogteknologi.dk kan den mangelfulde metadata også have en potentiel indvirkning på anvendelse af sprogressourcerne på sprogteknologi.dk. Dette er begrundet i den kritiske betydning, som data har for udviklingen af sprogmodeller. Derfor kan konsekvenserne af manglende metadata være, at nogle data som ellers ville være relevante for udviklingen af en sprogmodel, bliver udeladt eller at noget ikke-egnet data bliver tilvalgt grundet uklarheder.

7. Diskussion

For at diskutere hvilken rolle etiske reguleringer af sprogbaseret KI har for udviklingen af sprogmodeller i Danmark, bliver emnet tilgået fra det teoretiske perspektiv af ANT. Det betyder, at alle menneskelige og ikke-menneskelige aktører i stedet defineres som aktanter, der konstituerer netværket i kraft af handlingerne mellem aktanterne.

Første afsnit går i dybden med at klargøre udlægningen af sprogmodeller som netværk, ved at gennemgå hvordan de forskellige aktanter relaterer sig til hinanden gennem deres handlinger med hensyn til udviklingen af sprogmodeller. Derefter vil reguleringerne af KI blive diskuteret i forhold til tekniske og teoretiske overvejelser vedrørende sprogmodeller som en aktant og et produkt af aktanter. Det efterfølgende diskussionsafsnit tager udgangspunkt i bogen "AI – mellem fornuft og følelse" af Jan Damsgaard, professor ved 'Institut for Digitalisering ved Copenhagen Business School' (CBS) og medlem af 'Akademiet for de Tekniske Videnskaber' (Damsgaard & Sayers, 2023). Damsgaards perspektiver på regulering af KI i Danmark anvendes som modstykke til de danske og europæiske myndigheders tilgang til udvikling og regulering af KI. På baggrund af de nævnte afsnit afsluttes diskussionen med at besvare det tredje underspørgsmål af problemformuleringen: Hvilken rolle har de etiske reguleringer af sprogbaseret kunstig intelligens for udviklingen af sprogmodeller i Danmark?

7.1. Sprogmodeller er ikke "bare" teknologi

Sprogmodeller består ikke kun af koder, algoritmer og data, men er et produkt af interaktioner mellem menneskelige og ikke-menneskelige aktanter. Derfor bør sprogmodeller anses som sociotekniske netværk af aktanter, der hver især indgår i netværket for udviklingen af sprogmodeller ved at opfylde deres forskellige roller og funktioner i forhold til de andre aktanter i netværket. Herunder fremgår en visualisering af de mest fremtrædende aktanter i netværket for udvikling af danske sprogmodeller.



Figur 5. Visualisering af netværket af KI-aktanterne involveret i udviklingen af Danske Sprogmodeller.

En vigtig pointe i forhold til netværket er, at det umiddelbart ser forholdsvis lille og simpelt ud. Dette kan begrundes i to sideløbende forklaringer: (1) projektets fokus i kraft af problemformuleringen og afgrænsningen af emnet, samt (2) Latours definition af aktanter som isomorfiske. For det første ønsker projektet ikke at opnå indsigt i, hvordan sprogmodeller udvikles, men betragter de etiske aspekter af reguleringen for udviklingen af KI. I undersøgelsen af emnet er det relevant at have forståelse for både de materialistiske og sociale aspekter af teknologien for at kunne have en velinformeret og nuanceret diskussion. For det andet er aktanterne i netværket isomorfiske, hvilket betyder, at hver aktant i sig selv består af et netværk. Figur 5 indeholder dermed de mest relevante og repræsentative aktanter fra empirien for at kunne give indblik i og undersøge netværket.

Reguleringen af udviklingen af sprogmodeller i Danmark bliver udarbejdet af de danske myndigheder og EU. Reguleringer kunne blive anset som handlinger fra de danske myndigheder og EU, men bliver i denne undersøgelse en aktant i sin egen ret, grundet henvisningen til reguleringer som selvstændige aktanter (Damsgaard & Sayers, 2023, s. 10; 2021/0106(COD), 2024, s. 14). Derfor er reguleringerne aktanter, som bliver skabt af de politiske myndigheder. Det er dog ikke kun myndighederne, som laver indgreb i udarbejdelsen af reguleringerne, men også ekspertgrupper såsom HLEG, Dataetisk Råd og det bredere

forskningssamfund (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 22; P9_TA(2020)0275, 2020, s. 7). Hertil er det interessant at referere til forskellene og lighederne mellem de videnskabeligt definerede etiske retningslinjer for KI af Huang (et al.) og Floridi, og de identificerede etiske retningslinjer for KI i Danmark (Floridi, 2023, s. 61–64; Huang et al., 2023, s. 808–809). Ved en nærmere læsning af, hvordan de etiske retningslinjer defineres af de tre aktanter, viser det sig, at variationerne mellem retningslinjerne primært skyldes sproglige forskelle. Udvikling som etisk retningslinje bliver fremlagt i Floridis undersøgelse som 'godgørenhed', men begge omhandler prioriteringen af velstand (Floridi, 2023, s. 61). Forskellen mellem Floridi og de danske og europæiske reguleringer opstår i myndighedernes inddragelse af økonomisk vækst, hvorimod Floridi henviser til bæredygtighed og humanisme (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 5; Floridi, 2023, s. 61). Denne difference kan blandt andet findes i aktanternes forhold til andre i netværket. De danske myndigheder og EU har en mere direkte relation til erhvervslivet i form af organisationer og virksomheder, som gør, at der er en interesse i at have et kapitalistisk frem for et humanistisk forhold til udvikling. Dertil må der også tages hensyn til, at danske og europæiske myndigheder er politiske institutioner, som begge er under pres fra offentlighedens interesser, som har betydning for hvilke tiltag, der fremhæves og understøttes i reguleringerne.

Tillige udspringer der også aktanter fra myndighederne og reguleringerne i form af offentlige initiativer, såsom sprogteknologi.dk. Som aktant fortolker sprogteknologi.dk sine egne målsætninger og formål på baggrund af interesserne i de danske myndigheder og reguleringernes indgreb i initiativets eksistensgrundlag (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 16; Digitaliseringsstyrelsen, 2024a; Finansministeriet & Erhvervsministeriet, 2019, s. 36–37). Reguleringerne og initiativerne er direkte relateret til virksomheder og organisationer, som har til hensigt at udvikle eller arbejde med sprogmodeller. Eksempelvis er virksomheder og organisationer sammen med EU de primære bidragende aktanter af sprogressourcerne på sprogteknologi.dk. Dermed har virksomheder og organisationer mulighed for at have indflydelse på datagrundlaget for udvikling af danske sprogmodeller ved at medvirke i initiativer som sprogteknologi.dk. Altså kan virksomheder og organisationer fremme deres egne interesser ved at indgå i deling af deres data, til trods for at myndighederne, reguleringerne og sprogteknologi.dk har deres individuelle interesser.

Hvor de forudgående aktanter består af større og abstrakte forsamlinger af aktanter, repræsenterer den følgende gennemgang umiddelbart mere håndgribelige aktanter. Til den praktiske udvikling af sprogteknologi er det nødvendigt at inddrage udviklerne, aktanterne, som aktivt sidder og udarbejder de algoritmer, der i sidste ende resulterer i sprogmodellerne. Udover at udviklere er underlagt begrænsningerne i deres arbejde gennem reguleringer, så kan de også være en del af virksomheder og organisationer, som pålægger udviklerne endnu et lag af eventuelle interesser i, begrænsninger af, indgreb i eller fortolkning af produkter, de udvikler. Alt efter om udviklerne agerer selvstændigt eller som en del af en større enhed, kan de have flere direkte forbindelser til andre aktanter i netværket. Udviklere udarbejder som sagt sit arbejde ved at programmere og kode, og dette resulterer i algoritmer som bliver formet af udvikleren og den data, som teknologien trænes på. Endelig resulterer interaktionerne mellem udviklere, data og algoritmer i sprogmodeller. Altså har relationerne mellem udviklere, data og algoritmer stor betydning for udformningen af sprogmodeller, men det er, som netværket viser, ikke en isoleret proces. Udvikleren kan enten have sine egne interesser eller være en del af en virksomhed eller organisation med interesser i at udvikle en sprogmodel. Eftersom sprogmodeller kræver enorme mængder af data for at blive udviklet, kræver det, at udvikleren eller deres virksomhed eller organisation enten selv har den nødvendige data eller kan få adgang til den nødvendige data. Denne data er nødt til at være af en høj kvalitet for at kunne udvikle gode og anvendelige sprogmodeller. Hvilken data, der er nok af og er af høj nok kvalitet til udvikling af en sprogmodel, er et spørgsmål, som virksomheder og organisationer, EU, danske myndigheder, reguleringer, sprogteknologi.dk og udviklere har mulighed for at have indflydelse og fortolke på, alt efter hvilke interesser de hver især har. Dertil er udvælgelsen af data af enorm betydning for sprogteknologi, og denne del af netværket risikerer nemt at blive til en sort kasse, som en konsekvens af de mange processer der går forud for udviklingen af sprogmodeller. Hvis ikke der sikres en vis grad af gennemsigtighed med hensyn til hvilken data, der er anvendt, vil det være umuligt for udefrakommende at opnå indsigt i, hvilke potentielle data der ligger til grund for teknologien.

Endeligt er det også relevant at inddrage aktanter, som udbreder og bruger sprogmodeller, eftersom sprogmodellerne ikke bare bliver udviklet for at kunne eksistere uden noget formål, og fordi den sproglige data har et konkret ophav. Alle de menneskelige aktanter i netværket

har en eller anden interesse i et aspekt af udviklingen eller modarbejdelsen af sprogmodeller. De danske myndigheder og EU kan have en interesse i at lave indgreb i udviklingen af sprogmodeller på baggrund af den offentlige debat blandt direkte og indirekte brugere af KI. Endvidere har udbredelsen af sprogmodeller også indvirkning på hvilke interesser, der kan være i udviklingen. Hvis en sprogmodel med henblik på kommercielt udbytte ikke kan udbredes, fejler den på forhånd, eftersom den ikke kan komme ud til brugerne. Ud over den direkte relevans af brugere og udbredelse af sprogmodeller bliver de da også omtalt som hovedaktanter i reguleringerne, hvor ”udviklere, udbredere og brugere” er (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 11; Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 2; P9_TA(2020)0275, 2020, s. 6; 2021/0106(COD), 2024, s. 11–12; Finansministeriet & Erhvervsministeriet, 2019, s. 28–29). Dertil er det også nødvendigt at påtale den enorme mængde af sproglige data, som dagligt bliver produceret af alle de menneskelige aktanter i netværket. Denne data udgør en potentiel kritisk indgangsvinkel til udvikling af en dansk sprogmodel på naturligt sprog, som reflekterer det brede danske samfund, da en af de store udfordringer for udviklingen af dansksproget KI er de mange sociale og kulturelle variationer, der er i det danske sprog, til trods for at dansk repræsenterer en lille sproggruppe (Kirchmeier et al., 2019, s. 44–45).

Udviklingen af sprogmodeller er altså et netværk af flere forskelligartede og mere eller mindre abstrakte aktanter, som hver især har indflydelse på hinandens processer. Både menneskelige og ikke-menneskelige aktanter har indflydelse på hvilke processer, der fremmes, begrænses og bliver lavet indgreb i. Det er altså nødvendigt at medregne såvel tekniske som sociale aspekter i netværket og udviklingsprocessen af danske sprogmodeller, da de statistiske modeller forsøger at kortlægge det naturlige sprog, de baseres på. Derfor er det kritisk i diskussionen af KI-teknologi at være meget bevidst om at medtænke gennemsigtighed i udviklingen af sprogmodeller. Hvis ikke aktanterne i netværket bestræber sig på at efterleve de etiske principper, risikerer man en udvikling af KI, hvori der ikke tages højde for nogle af de positive forpligtelser, som reguleringer omfatter. Reguleringerne af udviklingen bestræber sig på, at netværket ikke bliver en sort kasse af uigennemskuelige normer og regler, gennemtvunget af aktanterne med den bedste evne til at skabe asymmetriske processer. Ved at opstille tydelige og ensartede retningslinjer sikres en udtalt

ramme for udviklingen af KI baseret på gennemsigthed, ansvarlighed, autonomi, udvikling, robusthed og retfærdighed.

7.2. Ethiske reguleringers rolle for sprogmodeller i Danmark

Med hensyn til at besvare, hvilken rolle de etiske retningslinjer har for udviklingen af sprogmodeller, er det givende at se på hvilke specifikke ønsker, der er til teknologien. Efterfølgende diskuteres det, hvorvidt ønskerne reelt set er teoretisk holdbare. Til sidst diskuteres den overordnede betydning af de etiske retningslinjer for udviklingen af sprogmodeller.

7.2.1. En kilde til objektivitet

På nuværende tidspunkt burde det stå klart, at KI er alt andet end entydigt og simpelt. Udviklingen af sprogbasert KI i form af sprogmodeller består af mange forskelligartede processer, som udføres af aktanter, der hver især har egne interesser og har indflydelse på andres processer under udarbejdelsen af enhver sprogmodel. Det er derfor bemærkelsesværdigt, at en række af reguleringerne vælger at have en målsætning om objektivitet med hensyn til KI og retfærdighed.

I nationalstrategien fremhæves det, at KI er systemer baseret på algoritmer, som har til formål at "identificere den mest hensigtsmæssige løsning" (Finansministeriet & Erhvervsministeriet, 2019, s. 6). Derfra videreføres denne logik til, at sprogmodeller skal "være objektive, saglige og uafhængig af personlige forhold" (Finansministeriet & Erhvervsministeriet, 2019, s. 7). Dette bliver mere konkret nævnt i forbindelse med, at KI skal være fri for fordomme og sikre ligebehandling ved at være objektiv. En lignende sammenhæng mellem retfærdighed og KI bliver lavet i 2020/2012(INL), som i forbindelse med ligebehandlingsprincippet nævner, at det kun er objektive og ikkediskriminerende KI, der er tilladt i Unionen (P9_TA(2020)0275, 2020, s. 54). Objektivitet bliver i rapporten fra HLEG også forbundet med ligestilling, ikkediskrimination og solidaritet, men her påpeges det også at "der skelnes mellem forskellige situationer baseret på objektive begrundelser." (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 11). Hvilket betyder, at objektivitet, ifølge HLEG, er noget, som KI kan blive udviklet med, og som skal gøre det muligt, at teknologien besidder et moralsk kompas.

Altså skal KI være en etisk eksplicit agent, som skal have etiske aspekter inkorporeret i dets funktionalitet, hvormed universelle etisk korrekte handlinger kan udføres og vurderes (Wiegel, 2010, s. 203–204). Men som det er blevet fremlagt, er sprogmodeller ikke bare et stykke teknologi, som kan blive programmeret til at være objektivt. For det første består sprogmodeller af aktanter, som danner et netværk af processer, som leder til den konkrete sprogmodel, der er formet efter deres forskelligrettede interesser. For det andet er teknologi, i et kontekstualistisk perspektiv, et socioteknologiske produkt, der aldrig er neutralt, især ikke med en sådan kompleksitet og handlingspotentiale som sprogmodeller besidder. Derimod er et stykke teknologi altid et produkt af dets historiske kontekst samt de materielle, sociale, kulturelle og politiske miljøer, som det udtænkes og skabes i (Winner, 1980, s. 130). For det tredje er der eksempler fra Danmark, hvor man har ønsket at benytte sig af KI som et middel til at opnå objektive vurderinger. I 2017 blev Gladsaxe-modellen implementeret til sagsbehandlingen af udsatte børn i Gladsaxe kommune med henblik på at opspore mistvivl på baggrund af en vurdering, der var fri for menneskelige forudindtagelser (Andersen, 2020). Men efter omkring et år blev modellen skrottet på grund af tvivlsom brug af data og KI. Senere blev Gladsaxe-modellen endvidere anklaget for, at et sådant projekt indebærer en høj risiko for forskelsbehandling (Andersen, 2023). Gladsaxe-modellen var et forsøg på at udvikle et stykke objektiv teknologi, men problemet med teknologi er, at det er udviklet i en socioteknisk kontekst, der er under konstant indflydelse af de involverede aktanter. Som det er blevet påpeget flere gange undervejs i dette og forrige afsnit, er antagelsen om at teknologi er objektiv en fejlslutning, som glemmer at medregne, at teknologien et socioteknisk fænomen.

Med hensyn til objektivitet af KI er det givende igen at rette blikket tilbage på rapporten af HLEG, hvor der henvises til, at objektivitet kan og bør understøttes af ”de teams, der designer, udvikler, tester og vedligeholder, udbreder og/eller indkøber disse systemer, overvejer diversiteten af brugere og af samfundet generelt.” (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 25). Ved at repræsentere et bredere udsnit af samfundet vil det altså være muligt at fremme objektivitet i kraft af, at flere perspektiver præsenteres i de forskellige processer under udviklingen, udbredelsen og anvendelsen af KI. KI er altså ikke objektivt på baggrund af den rene konsekvens af, at mennesker har forskellige

perspektiver, som vil have indvirkning på teknologien. Men dette kan ifølge HLEG modvirkes ved at inddrage mennesker fra forskellige sociale, kulturelle og økonomiske baggrunde.

Det dog nødvendigt at henvise til, at sprogmodeller er et netværk af flere aktanter end blot de menneskelige og organisatoriske aktanter, for der ligger endnu en komplikation for ønsket om objektivitet i forhold til, at sprogmodeller også bygger på data. Som beskrevet består data af information, og den information har også et ophav, som ikke er uden intention eller fordomme. Det er altså også relevant at henvise til, at data i sig selv ikke er objektivt. Denne kendsgerning betyder dog ikke, at reguleringerne har fejlet på forhånd, eller at udviklingen af brugbar KI er nyttesløst. Derimod det et spørgsmål om, at 'objektivitet' er et misvisende begreb at anvende i denne sammenhæng. For til trods for at reguleringerne ønsker, at KI skal være objektivt i forbindelse med retfærdighed, så ønsker de også, at KI skal udvikles, udbredes og anvendes i overensstemmelse med europæiske og danske værdier (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 26; Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 4; Finansministeriet & Erhvervsministeriet, 2019, s. 7; 2021/0106(COD), 2024, s. 11–12). Altså ønsker reguleringerne ikke, at KI og sprogmodeller skal være objektive, men at danske og europæiske sprogmodeller bør have velovervejede fordomme, som afspejler de værdier, som ønskes tilgodeset.

7.2.2. Er der overhovedet behov for regulering af KI

KI med rod i etiske retningslinjer fremstilles på tværs af reguleringerne som en potentiel konkurrenceevne, hvilket fremgår i kraft af udvikling som etiske retningslinjer, hvor vækst som konsekvens af KI anses som etisk rigtige handlinger. I KI akten fremgår konkurrenceevnen i forhold til etisk KI i form af, at alle KI-aktanter, der er involveret i standardiseringsprocessen af etisk KI i EU, også skal understøtte investering i og innovation af KI med henblik på at vækste det indre marked (2021/0106(COD), 2024, s. 153). Altså er udviklingen af KI direkte forbundet med etik som konkurrenceelement i forhold til teknologiens potentielle rolle som vækstskaber i Europa, som EU må sikre gennem regulering af KI. Ifølge Damsgaard er dette dog den forkerte tilgang til KI i Europa og Danmark, og KI akten omtales som "(...) verdens første og mest restriktive lov på AI-området" (Damsgaard & Sayers, 2023, s. 9). Hertil mener Damsgaard, at KI akten vil have en negativ indvirkning på udviklingen af KI og dermed

konkurrenceevnen af etisk KI. Altså vil den europæiske regulering af KI modarbejde den ønskede udvikling som konsekvens af lovgivning, bøder og regulering af KI. Damsgaard mener, at denne regulatoriske tilgang til udvikling af etisk KI er kontraintuitiv i forhold til målet, da EU langt hellere burde fokusere på at udvikle frem for at regulere (Damsgaard & Sayers, 2023, s. 9).

I modsætning til Damsgaard mener den europæiske regulering, at der er et behov for international regulering af KI på EU-niveau, eftersom ” (...) en sådan lovgivningsmæssig ramme er af afgørende betydning for at undgå en opsplittning af det indre marked som følge af forskellig national lovgivning, og at den vil bidrage til at fremme hårdt tiltrængte investeringer, udvikle datainfrastruktur og støtte forskning.” (P9_TA(2020)0275, 2020, s. 4). Det samme rationale går igen i digitaliseringsstrategien og nationalstrategien, hvor en koordineret indsats på internationalt og europæisk niveau menes at ville have en større og mere effektiv indvirkning på KI-feltet, end hvis Danmark forsøgte at ændre KI-landskabet på egen hånd (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 26–27; Finansministeriet & Erhvervsministeriet, 2019, s. 19).

Dog mener Damsgaard, at KI etik i langt større grad bør være op til den enkelte virksomhed at afgøre, ved at vurdere hvorvidt det pågældende KI-projekt er etisk holdbart, værdiskabende og i overensstemmelse med deres individuelle strategier (Damsgaard & Sayers, 2023, s. 137).

Damsgaard taler ikke for en anarkistisk og ureguleret udvikling af KI men mener, at KI-virksomheder agerer ud fra et menneskecentreret standpunkt, såvel som et markedsorienteret standpunkt (Damsgaard & Sayers, 2023, s. 136). Konkret mener Damsgaard, at etik bør være et ansvarsområde for den enkelte virksomheds bestyrelse, som bør håndtere risikostyring, etik og datasikkerhed (Damsgaard & Sayers, 2023, s. 140–141). Ud fra denne logik ville reguleringer altså være mindre centraliseret og potentielt blive defineret på et mere borgernært niveau af samfundet: ”Det er vigtigt, at vi ikke som samfund gør AI til et spørgsmål om jura eller teknokratisk detaljerytteri. Vi skal definere et formål. Måske skal vi også definere nogle AI-udviklinger, vi absolut ikke vil tillade.” (Damsgaard & Sayers, 2023, s. 146–147). Men Damsgaard modsætter sig ideen om overordnede regler for KI som i de danske og europæiske reguleringer, hvortil han tilføjer ”Vi kan ikke lave regler, der vil passe til alle situationer.” (Damsgaard & Sayers, 2023, s. 147). Dermed går Damsgaard imod de danske

myndigheders ønske om, at det danske erhvervsliv som udgangspunkt udelukkende skal sigte efter at gøre etisk KI til et konkurrenceparameter, eftersom det bør være op til den enkelte virksomhed at tage stilling til emnet (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 20; Finansministeriet & Erhvervsministeriet, 2019, s. 9). Derimod mener Damsgaard, at "Vi bør (...) udråbe Danmark til AI-laboratorium, hvor den enkelte borger, civilsamfundets organisationer, det offentlige og virksomhederne i fællesskab afprøver, hvad ChatGPT og andre former for AI kan bruges til, og hvor de kræver tilpasning til vores værdier og normer." (Damsgaard & Sayers, 2023, s. 149). Altså mener Damsgaard, at etik er et emne, som bør afgøres på baggrund af den konkrete kontekst af bestyrelser i de pågældende virksomheder, frem for at blive behandlet af myndigheder gennem bredt gældende reguleringer.

Årsagen til, at Damsgaard har dette syn på etik som et mere individnært emne, skal findes i hans syn på mennesker som dyds- og/eller pligtetiske væsner: "Mennesker vil generelt sig selv og hinanden det godt (...)" (Damsgaard & Sayers, 2023, s. 145). Ud fra en pligtetisk læsning betyder dette, at mennesker generelt ønsker at handle i overensstemmelse med deres indre moralitet, og dermed handle etisk korrekt for ikke at skade ens egen morale status. Senere udbygger Damsgaard sin etiske forståelse af mennesket ved at påstå, at "Vi har som mennesker en etisk fordring om at gøre det gode, ikke blot at undgå det onde." (Damsgaard & Sayers, 2023, s. 148). Hvilket henviser til at mennesker som udgangspunkt handler på et dydsetisk grundlag, eftersom mennesket er født med en indre godhed af natur, som gør at vores handlinger er etisk korrekte på baggrund af den rene intention med vores handlinger. Det er på grund af denne menneskeforståelse, at Damsgaard har en meget positiv indstilling i forhold til virksomheders evne til at selvregulere. Fordi menneskene i virksomhedernes bestyrelser vil andre mennesker det godt, så bør vi sætte dem fri til at handle i overensstemmelse med deres naturlige etiske higen og indre moralitet.

En af Damsgaards pointer i forhold til regulering af KI er, at der ikke kan laves absolutte regler for teknologi, eftersom den praksis, som teknologi anvendes i, er kompleks (Damsgaard & Sayers, 2023, s. 147). Dette stemmer overens med projektets portrættering af KI som et netværk af aktanter, hvilket er årsagen bag kompleksiteten af teknologien. Men Damsgaards iagttagelse af reguleringer som udelukkende restriktive er forsimplet i forhold til de nærværende reguleringer. Især KI aktens negative forpligtelser omhandler primært

udviklingen, udbredelsen og anvendelsen af generel KI og høj-risiko KI (2021/0106(COD), 2024, s. 58–59). Generel KI (også kaldt 'general purpose AI') er en form for KI, som overgår andre KI-modellers handlingspotentiale ved at udvise et højere og mere varieret handlingspotentiale (2021/0106(COD), 2024, s. 104). Høj-risiko KI er defineret på baggrund af de mulige risici, som er forbundet med den kontekst eller data, som teknologien anvendes i forbindelse med. KI akten giver en liste over områder, hvori anvendelsen af KI ville blive kategoriseret som høj-risiko, og omhandler primært KI, der risikerer at have negative konsekvenser, hvis der brydes med en eller flere af de juridiske eller etiske retningslinjer (2021/0106(COD), 2024, s. 248–251). Dertil opfordres der også i KI akten til, at myndigheder anvender sig af et proportionalitetsprincip, som netop har til formål at modvirke overregulering af KI, der ikke kan kategoriseres som høj-risiko (2021/0106(COD), 2024, s. 80). Dette er en af årsagerne til, at det nærværende projekt har identificeret udvikling som en etisk retningslinje i sig selv.

Endvidere er det ikke retvisende udelukkende at referere til reguleringer af KI som en barriere for udviklingen, fordi reguleringen af KI også forsøger at håndtere barriererne for udviklingen af KI i Danmark og EU. De danske reguleringer fremhæver, at nogle af de største barrierer for udvikling af KI i Danmark skyldes en række mangler. Først og fremmest en mangel på overordnede etiske retningslinjer for KI, dernæst en mangel på dansk data som konsekvens af det lille sprogområde, efterfulgt af en mangel på it-specialister i det danske erhverv og endeligt manglende investeringer i de danske marked (Digitaliserings- og Ligestillingsministeriet, 2023a, s. 9; Finansministeriet & Erhvervsministeriet, 2019, s. 17).

Damsgaard henviser til, at man bør være opmærksom på KI fra USA og Kina, som er de mest dominerende nationer i forhold til vækst og udvikling på baggrund af KI. Hertil ønsker de danske myndigheder lige præcis at undgå amerikanske eller kinesiske tilstande, hvor der udvikles KI uden den nødvendige rod i etisk og ansvarlig KI, som man forsøger at opnå i Danmark og EU gennem en regulatorisk tilgang (Finansministeriet & Erhvervsministeriet, 2019, s. 8). Damsgaard mener dog, at dette er en forkert tilgang til KI, som kun vil resultere i, at EU vil ende "som en tech-ørken", mens resten af verden vil nyde godt af KI-baseret vækst og velstand (Damsgaard & Sayers, 2023, s. 9). Fremstillingen af EU og dermed også Danmark som en tech-ørken er dog på nuværende tidspunkt ikke en realitet, og der er da også

eksempler på private og offentlige KI initiativer i Danmark med henblik på at udvikle større danske sprogmodeller (Dansk Erhverv, 2024; Mønsted, 2024). Det ene eksempel er Det Kongelige Biblioteks forslag om at basere en sprogmodel på de omkring ni millioner gigabyte data, som deres digitale database består af. En af barriererne for, at udviklingen endnu ikke er gået i gang, er blandt andet uklarhederne i den danske lovgivning og KI aktens betydning for dansk KI (Mønsted, 2024). Hvilket altså understøtter Damsgaards forrige pointe med hensyn til faren for at ” (...) vi ikke som samfund gør AI til et spørgsmål om jura eller teknokratisk detaljerytteri” og bør overveje erklæringen af Danmark som et KI-laboratorium (Damsgaard & Sayers, 2023, s. 146 & 149).

Årsagen bag Damsgaards grundlæggende kritik af reguleringer er, at han mener, der er en bred tendens i samfundet, hvor især medier giver et meget negativt og dystopisk billede af KI, som er baseret på urealistiske og følelsesbaserede vurderinger af teknologien. I forhold til reguleringer og lovgivning af KI omtales dette som en konsekvens af manglende forståelse for teknologien og en utryghed, som gør, at ” (...) hård regulering og stram lovgivning er den eneste rigtige måde (...) at tøjle en teknologiudvikling (...)” (Damsgaard & Sayers, 2023, s. 8). Derfor har hans bog til formål at give ”et konstruktivt og holistisk syn på teknologi og AI” (Damsgaard & Sayers, 2023, s. 8). Altså skal bogen fungere som et modstykke til, hvad han definerer som en irrationel reaktion på teknologi, der har skabt ”et stort behov for at inddrage faglig fornuft i diskussionerne om AI, så det ikke udelukkende er følelserne, der styrer.” (Damsgaard & Sayers, 2023, s. 11). Dette er en unuanceret iscenesættelse af reguleringer, som Damsgaard her direkte forbinder med følelsesbaseret reaktioner på udvikling af KI. Men det er misvisende derfra at fremstille reguleringer som KI akten som udelukkende negativt forpligtende (Damsgaard & Sayers, 2023, s. 10). I stedet burde Damsgaard nemlig forsøge at give et holistisk perspektiv ved også at henvise til reguleringernes positive forpligtigelser, der understøtter og fremmer udviklingen af KI. Eftersom KI akten er objektet for hans skarpe kritik af EU's regulatoriske håndtering af KI, er det vigtigt at påpege, at KI akten blandt andet er baseret på etiske retningslinjer fra rapporten fra HLEG, som både består af medlemmer fra forskningssamfundet og erhvervslivet (Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens, 2019, s. 43).

Endeligt er det relevant at fremhæve Damsgaards pointer med hensyn til, at der kan være en unødvendig tendens til at fokusere på negative fremstillinger af KI. Det er derfor givende, at der fremgår perspektiver som Damsgaards, der minder os om, at man i EU og Danmark må være mindst lige så opmærksom på mulighederne som på udfordringerne ved KI. Hvis vi ikke ønsker at blive overhalet indenom af udefrakommende KI systemer, er det meget muligt, at man i nogle tilfælde må være mere risikovillig end i andre tilfælde. Dog må Damsgaards kritik af reguleringer læses med nogle forbehold med hensyn til hans meget positive tiltro til teknologi og mennesker. Dette er konkluderet på baggrund af Damsgaards menneskeforståelse og portrættering af regulering som primært negativ forpligtende og følelsesbaseret. Damsgaard har dog relevante pointer i sin kritik af faren ved at have for stor en tillid til, at reguleringer er svaret på alle KI problematikker. Ydermere er det givende at inddrage Damsgaards tilgang til en menneskenær tilgang til etisk KI og modarbejdelse af bureaukrati i den videre diskussion af KI udvikling. Men hans kritik bliver også mindre rammende i kraft af projekts fund i forhold til reguleringernes opfordringer til proportionalitet, samt udvikling, ansvarlighed og holisme som etiske retningslinjer for KI udvikling i Danmark, hvilket nuancerer Damsgaards påstande om KI akten og reguleringer.

8. Konklusion

For at kunne konkludere på specialets problemformulering er det nødvendigt at gennemgå de forskellige underspørgsmål.

De etiske retningslinjer for udvikling af KI i Danmark består af en række etiske retningslinjer, som danner en holistisk ramme for dansk sprogteknologi: Autonomi, Udvikling, Ansvarlighed, Gennemsigtighed, Robusthed og Retfærdighed. Disse etiske retningslinjer forholder sig primært til de menneskelige aktanter, hvor udvikling, udbredelse og anvendelse af KI ses som en praksis, der kan reguleres gennem etiske og designmæssige retningslinjer. Altså adresseres de menneskelige aktanter som handlende individer, der i kraft af deres moralske status bør handle etisk korrekt med henblik på at efterleve et europæisk og dansk sæt af værdier. Dog fremgår reguleringerne af KI ikke blot i form af en række etiske forpligtigelser,

men også i form af understøttende initiativer med henblik på nytteetiske vurderinger af KI over for pligtetiske hensyn til menneskelige aktanter som moralske agenter og patienter.

Sprogteknologi.dk er et af de initiativer, som er opstået på baggrund af danske myndigheders regulering og understøttelse af sprogteknologi. Til trods for at sprogteknologi.dk ikke som udgangspunkt er udviklet med henblik på etisk KI, fremgår efterlevelse af alle de etiske retningslinjer under de generelle betingelser for sprogteknologi.dk. Af de selverklærede retningslinjer forekommer der enkelte brud på baggrund af de få observerede mangler i sprogressourcernes metadata, som gør, at den ikke er fyldestgørende. Sprogressourcerne på sprogteknologi.dk kommer fra en bred vifte af organisationer, dog med en stor andel fra én organisation, EU, hvilket afspejles i alle undersøgelserne af metadata. Dog anses sprogteknologi.dk som et vellykket initiativ i det primære formål om at fremme udviklingen af sprogteknologi ved at understøtte tilgængeligheden af sprogressourcer. Ud fra undersøgelsen af materialet konkluderes det, at sprogressourcerne generelt trækker i retning af flersproget sprogteknologi, baseret på dansk- og engelsksproget data.

Endeligt viser undersøgelsen af de danske og europæiske reguleringer af KI, at sprogmodeller ikke blot er et teknisk fænomen men eksisterer som én aktant i et netværk af interagerende aktanter. Dette er en af årsagerne til, at ønsket om objektivitet i sprogmodeller er mere kompliceret end at programmere teknologien til at være objektiv. Grundet de mange sociotekniske aspekter af teknologien betyder det, at sprogmodeller aldrig vil være sandfærdige i alle tilfælde. Derfor opfordres det til, at man bør øge diversiteten under alle livsfaserne af en sprogmodels levetid. Tillige er det misvisende at benytte sig af ordet 'objektivitet', eftersom der mere specifikt er tale om, at reguleringerne ønsker at teknologien afspejler europæiske og danske værdier og normer i overensstemmelse med love og etiske retningslinjer. Det er ligeledes vigtigt at nuancere anvendeligheden af generelle reguleringer af etik. Det er nødvendigt at være bevidst om, hvordan reguleringer både kan hæmme og fremme udviklingen af KI. Reguleres der unødvendigt, vil reguleringerne fremstå som negative forpligtelser, som kun har til formål at begrænse. Derimod indeholder reguleringerne også understøttende og positive forpligtelser, der kan medvirke i en effektiv og indflydelsesrig udvikling af etisk KI. Men for at dette kan realiseres, er der behov for en nuanceret og fagligt funderet tilgang til KI fra alle aktanter gennem hele netværket. Dermed har reguleringerne af

sprogteknologi i Danmark vidtrækkende implikationer for udviklingen, som både initierer etisk understøttende tiltag og sætter grundlæggende etiske krav til alle KI-aktanter.

Litteraturliste

- Andersen, T. (2020, januar 29). *Gladsaxe-modellen spørger: Nyt AI-projekt skal forudsige mistrivsel hos børn* | Version2. version2.dk. <https://www.version2.dk/artikel/gladsaxe-modellen-spoeger-nyt-ai-projekt-skal-forudsige-mistrivsel-hos-boern>
- Andersen, T. (2023, december 18). *Danske forskere går galt i byen med underdrivelser om Gladsaxe-model og social profilering*. Radar. <https://radar.dk/holdning/danske-forskere-gaar-galt-i-byen-med-underdrivelser-om-gladsaxe-model-og-social-profilering>
- Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022). Introducing the FAIR Principles for research software. *Scientific Data*, 9(1), 622. <https://doi.org/10.1038/s41597-022-01710-x>
- Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). What Is Ethics? I C. Bartneck, C. Lütge, A. Wagner, & S. Welsh (Red.), *An Introduction to Ethics in Robotics and AI* (s. 17–26). Springer International Publishing. https://doi.org/10.1007/978-3-030-51110-4_3
- Booth, W. C., Colomb, G. G., Williams, J. M., Bizup, J., & Fitzgerald, W. T. (2016). *The Craft of Research* (4. udg.). University of Chicago Press. <http://ebookcentral.proquest.com/lib/asb/detail.action?docID=4785166>
- Callon, M., & Latour, B. (1981). Unscrewing the big Leviathan: How actors macro-structure reality and how sociologists help them to do so. I K. Knorr-Cetina & A. V. Cicourel (Red.), *Advances in social theory and methodology: Toward an integration of micro- and macro-sociologies*. Routledge & Kegan Paul. <http://www.bruno-latour.fr/sites/default/files/09-LEVIATHAN-GB.pdf>
- Carrara, W., Enzerink, S., Oudkerk, F., Radu, C., & Steenbergen, E. van. (2019). *Open Data Goldbook for Data Managers and Data Holders—Practical guidebook for organisations wanting to publish Open Data*. Den Europæiske Dataportal. https://data.europa.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

- Coeckelbergh, M. (2018). Why Care About Robots? Empathy, Moral Standing, and the Language of Suffering. *Kairos. Journal of Philosophy & Science*, 20(1), 141–158.
<https://doi.org/10.2478/kjps-2018-0007>
- Damsgaard, J., & Sayers, S. (2023). *AI - mellem fornuft og følelse* (1. udgave.). Djøf.
- Dansk Erhverv. (2024, maj 6). *Nyt ambitiøst initiativ bringer tunge spillere fra det offentlige og private sammen for at udvikle danske sprogmodeller*. Dansk Erhverv.
<https://www.danskerhverv.dk/presse-og-nyheder/nyheder/2024/maj/nyt-ambitiost-initiativ-bringer-tunge-spillere-fra-det-offentlige-og-private-sammen-for-at-udvikle-danske-sprogmodeller/>
- Den Europæiske Unions charter om grundlæggende rettigheder, C 326/391 Den Europæiske Unions Tidende (2012). https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=uriserv%3AOJ.C_.2012.326.01.0391.01.DAN&toc=OJ%3AC%3A2012%3A326%3ATOC
- Digitaliserings- og Ligestillingsministeriet. (2023a). *Danmarks digitaliseringsstrategi – Ansvar for den digitale udvikling*. Regeringen.
<https://digmin.dk/digitalisering/nyheder/nyhedsarkiv/2023/nov/ny-digitaliseringsstrategi-saetter-retning-for-den-digitale-udvikling-af-danmark>
- Digitaliserings- og Ligestillingsministeriet. (2023b, november 16). *Ny digitaliseringsstrategi sætter retning for den digitale udvikling af Danmark*.
<https://digmin.dk/digitalisering/nyheder/nyhedsarkiv/2023/nov/ny-digitaliseringsstrategi-saetter-retning-for-den-digitale-udvikling-af-danmark>
- Digitaliseringsstyrelsen. (2024a). *Baggrund* [Sprogteknologi.dk]. Baggrund. Set 2024-05-02.
<https://sprogteknologi.dk/pages/baggrund>
- Digitaliseringsstyrelsen. (2024b). *Fejl 404* [Sprogteknologi.dk]. Fejl 404. Set 2024-05-23.
<https://sprogteknologi.dk/dataset/dansk-wikiquote>
- Digitaliseringsstyrelsen. (2024c). *Generelle betingelser* [Sprogteknologi.dk]. Generelle betingelser. Set 2024-05-23. <https://sprogteknologi.dk/pages/generellebetingelser>
- Digitaliseringsstyrelsen. (2024d). *Optagelse i kataloget—Sprogteknologi.dk*. Set 2024-05-23. [sprogteknologi.dk. https://sprogteknologi.dk/pages/optagelse-i-kataloget](https://sprogteknologi.dk/pages/optagelse-i-kataloget)

- Digitaliseringsstyrelsen. (2024e). *Sprogressourcer* [Sprogteknologi.dk]. Sprogressourcer. Set 2024-05-23. <https://sprogteknologi.dk/pages/omsprogressourcer>
- Digitaliseringsstyrelsen. (2024f, april 29). *Applikation og datamodel* [Sprogteknologi.dk]. Applikation og datamodel. Set 2024-05-23. <https://sprogteknologi.dk/pages/applikation-og-datamodel>
- Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens. (2019). *Ethiske retningslinjer for pålidelig kunstig intelligens | Shaping Europe's digital future*. Europa-Kommissionen. <https://digital-strategy.ec.europa.eu/da/library/ethics-guidelines-trustworthy-ai>
- Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens. (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment | Shaping Europe's digital future*. Europa-Kommissionen. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342
- ELRC. (2024). *ELRC - European Commission*. Language-Data-Space.Ec.Europa.Eu. Set 2024-05-02 https://language-data-space.ec.europa.eu/related-initiatives/elrc_en
- EPIC. (2018). DIGITALEUROPE [Digitaleurope.org]. *Recommendations on Artificial Intelligence Policy*. Set 2024-05-23. <https://www.digitaleurope.org/resources/digitaleurope-recommendations-on-ai-policy/>
- Europa-Parlamentets beslutning af 20. oktober 2020 med henstillinger til Kommissionen om en ramme for etiske aspekter af kunstig intelligens, robotteknologi og relaterede teknologier (2020/2012(INL)), C 404/63 Den Europæiske Unions Tidende (2020). <https://eur-lex.europa.eu/legal-content/DA/TXT/PDF/?uri=CELEX:52020IP0275>
- Europa-Parlamentets og Rådets direktiv (EU) 2019/1024 af 20. juni 2019 om åbne data og videreanvendelse af den offentlige sektors informationer (omarbejdning), L 172/56 Den Europæiske Unions Tidende (2019). <http://data.europa.eu/eli/dir/2019/1024/oj/dan>
- Finansministeriet & Erhvervsministeriet. (2019). *National strategi for kunstig intelligens*. Regeringen. <https://fm.dk/udgivelser/2019/marts/national-strategi-for-kunstig-intelligens/>
- Flick, U. (2014). *An introduction to qualitative research* (5. udg.). Sage.

- Floridi, L. (2023). A Unified Framework of Ethical Principles for AI. I L. Floridi (Red.), *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press. <https://doi.org/10.1093/oso/9780198883098.003.0004>
- Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., & Holzinger, A. (2018). Explainable AI: The New 42? I A. Holzinger, P. Kieseberg, A. M. Tjoa, & E. Weippl (Red.), *Machine Learning and Knowledge Extraction* (s. 295–303). Springer International Publishing. https://doi.org/10.1007/978-3-319-99740-7_21
- Gunkel, D. J. (2014). A Vindication of the Rights of Machines. *Philosophy & Technology*, 27(1), 113–132. <https://doi.org/10.1007/s13347-013-0121-z>
- Gutiérrez, J. L. M. (2023). On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00314-4>
- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hobbes, T. (2018). *Leviathan*. Lerner Publishing Group.
<http://ebookcentral.proquest.com/lib/asb/detail.action?docID=5443203>
- Howard, D. (2022). Whence and W(h)ither Technology Ethics. I S. Vallor (Red.), *The Oxford Handbook of Philosophy of Technology* (s. 0). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780190851187.013.6>
- Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799–819.
<https://doi.org/10.1109/TAI.2022.3194503>
- IEEE. (2024). *IEEE at a Glance* [ieee.org]. About. Set 2024-05-23.
<https://www.ieee.org/about/at-a-glance.html>
- Immanuel, K. (1999). *Grundlæggelse af sædernes metafysik* (T. Bøgeskov, Overs.; 2. udgave.). Hans Reitzel.
- Jecker, N. S. (2010). Animal subjects research Part I: Do animals have rights? I G. Van Norman, S. Jackson, S. Rosenbaum, & S. Palmer (Red.), *Clinical Ethics in Anesthesiology: A Case-Based Textbook* (s. 168–173). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511841361.030>

- Kirchmeier, S., Henrichsen, P. J., Diderichsen, P., & Hansen, N. B. (2019). *Dansk Sprogteknologi i Verdensklasse*. Dansk Sprognævn. <https://dsn.dk/wp-content/uploads/2021/01/sprogteknologi-i-verdensklasse.pdf>
- Kroll, J. A. (2018). The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180084. <https://doi.org/10.1098/rsta.2018.0084>
- Latour, B. (1994). On Technical Mediation: Philosophy, Sociology, and Genealogy. *Common Knowledge*. <http://www.bruno-latour.fr/sites/default/files/54-TECHNIQUES-GB.pdf>
- Luukkonen, R., Komulainen, V., Luoma, J., Eskelinen, A., Kanerva, J., Kupari, H.-M., Ginter, F., Laippala, V., Muennighoff, N., Piktus, A., Wang, T., Tazi, N., Scao, T. L., Wolf, T., Suominen, O., Sairanen, S., Merioksa, M., Heinonen, J., Vahtola, A., ... Pyysalo, S. (2023). *FinGPT: Large Generative Models for a Small Language* (arXiv:2311.05640). arXiv. <http://arxiv.org/abs/2311.05640>
- Mønsted, S. (2024, april 30). *Dansk sprogmodel er en mulighed – skal vi gøre det?* DM Digi. <https://dm.dk/digi/artikler/ai/dansk-sprogmodel-er-en-mulighed-skal-vi-goere-det/>
- Nationalt Center for Etik. (2024). *Dataetisk Råd*. Dataetisk Råd - En del af Nationalt Center for Etik. Set 2024-04-02. <https://dataetiskraad.dk/>
- Nuffelen, B. V. (2024). *DCAT-AP 3.0*. SEMIC. <https://semiceu.github.io/DCAT-AP/releases/3.0.0/>
- Patel, S., Kane, H., & Patel, R. (2023). *Building Domain-Specific LLMs Faithful To The Islamic Worldview: Mirage or Technical Possibility?* (arXiv:2312.06652). arXiv. <http://arxiv.org/abs/2312.06652>
- Posit Software. (2024). *RStudio Desktop* [Posit.co]. Posit. Set 2024-05-23. <https://www.posit.co/download/rstudio-desktop/>
- Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 8115/21, EUR-lex (2024). <https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf>

- Redaelli, R. (2023). Different approaches to the moral status of AI: A comparative analysis of paradigmatic trends in Science and Technology Studies. *Discover Artificial Intelligence*, 3(1), 25. <https://doi.org/10.1007/s44163-023-00076-2>
- Regeringen, KL, & Danske Regioner. (2019). *Digitaliseringspagt—En ny retning for det fællesoffentlige samarbejde*. Digitaliseringsstyrelsen.
<https://digst.dk/media/19919/digitaliseringspagt-en-ny-retning-for-det-faellesoffentlige-samarbejde.pdf>
- Sverdlik, S. (2023a). Bentham's Chapter I. I S. Sverdlik (Red.), *Bentham's An Introduction to the Principles of Morals and Legislation: A Guide*. Oxford University Press.
<https://doi.org/10.1093/oso/9780190089900.003.0002>
- Sverdlik, S. (2023b). Bentham's Chapter II and the Formal Argument. I S. Sverdlik (Red.), *Bentham's An Introduction to the Principles of Morals and Legislation: A Guide*. Oxford University Press. <https://doi.org/10.1093/oso/9780190089900.003.0003>
- The R Foundation. (2024a). *R: What is R?* [R-project.org]. Set 2024-05-23. <https://www.r-project.org/about.html>
- The R Foundation. (2024b). *The Comprehensive R Archive Network* [Cran.r-project.org]. Set 2024-05-23. <https://cran.r-project.org/>
- Tracy, S. J. (2013). Data analysis basics: A pragmatic iterative approach. I *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Wiley.
- Tzafestas S G. (2018). *Information, communication, and automation ethics in the knowledge society age*. Nova Science Publishers.
- Van Den Hoven, J. (2010). The use of normative theories in computer ethics. I L. Floridi (Red.), *The Cambridge Handbook of Information and Computer Ethics* (1. udg., s. 59–76). Cambridge University Press. <https://doi.org/10.1017/CBO9780511845239.005>
- Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Pennsylvania State University Press.
<http://ebookcentral.proquest.com/lib/asb/detail.action?docID=6224410>
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59, 1–23.
<https://doi.org/10.18637/jss.v059.i10>

Wiegel, V. (2010). The ethics of IT-artefacts. I L. Floridi (Red.), *The Cambridge Handbook of Information and Computer Ethics* (1. udg., s. 201–218). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511845239.013>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A.,
Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J.,
Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T.,
Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data
management and stewardship. *Scientific Data*, 3(1), 160018.
<https://doi.org/10.1038/sdata.2016.18>

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus (Cambridge, Mass.)*, 109(1), 121–136.