LUNDS UNIVERSITET
Ekonomihögskolan

Institutionen för informatik

Mikroetik: Ett steg i rätt riktning för etik inom datadriven AI

En undersökning om etik och mikroetik hos svenska verksamheter som applicerar datadriven AI

Kandidatuppsats 15 hp, kurs SYSK16 i Informationssystem

Författare: Cecilia Minder
Leo Rasmusson

Handledare: Blerim Emruli

Rättande lärare: Umberto Fiaccadori
Odd Steen

Mikroetik: Ett steg i rätt riktning för etik inom datadriven AI: En undersökning om etik och mikroetik hos svenska verksamheter som applicerar datadriven AI

ENGELSK TITEL: Microethics: A step in the right direction for ethics in data-driven AI: A survey about ethics and microethics in Swedish organizations that apply datadriven AI

FÖRFATTARE: Cecilia Minder, Leo Rasmusson

UTGIVARE: Institutionen för informatik, Ekonomihögskolan, Lunds universitet

EXAMINATOR: Osama Mansour, PhD

FRAMLAGD: Maj, 2022

DOKUMENTTYP: Kandidatuppsats

ANTAL SIDOR: 85

NYCKELORD: Etik, AI, Mikroetik, AI-etik, Riktlinjer, Datadriven AI

SAMMANFATTNING:
Artificiell intelligens kan leda till disruptiva innovationer som verksamheter måste ta ställning till för att fortsätta vara konkurrenskraftiga. Med profilerade fall som Cambridge Analytica har etik i förhållande till datadriven AI fått en ny betydelse. Riktlinjer har framlagts för att vägleda organisationer till hur de ska förhålla sig till etik och AI. Det har framkommit att dessa dock inte följs i praktiken vilket har lett till uppkomsten av mikroetik. Ett sätt att genom tekniska instruktioner praktiskt sett implementera etik hos professionella verksamheter. Denna studie har för avsikt att undersöka hur pass medvetna organisationer, verksamma i Sverige, är om etik och mikroetik och hur det ska implementeras. Ett teoretiskt ramverk har sammanställts med litteratur kopplat till mikroetik och etablerade etiska riktlinjer i förhållande till AI. För att besvara forskningsfrågan i denna kvalitativa studie har tre intervjuer genomförts. En diskussion påbörjades sedan där resultatet tillsammans med teorin diskuteras i följande huvudområden: definitioner, etablerade riktlinjer, respons till etablerade riktlinjer och implementering av mikroetik. Den slutsats som kan dras är att det finns en medvetenhet om etik, men en låg medvetenhet om etiska riktlinjer och implementeringen av det, samt nästintill ingen medvetenhet om mikroetik.

Innehåll

Figurer

Tabeller

1 Introduktion

1.1. Bakgrund

Artificiell intelligens (AI) som begrepp kan härledas ända tillbaka till mitten av 1950-talet då det myntades av John Mccarthy, professor i datavetenskap på Stanford universitet (IEEE Computer Society, n.d). Uppgiften att definiera AI är minst sagt svår men kan övergripande förklaras som ett intelligent system som "hanterar information för att göra något målmedvetet" (Dignum, 2019, p. 9). Begreppet och tillika akronymen AI har sedan utvecklats och kan idag ses som ett paraplybegrepp till en mängd olika disruptiva innovationer, såsom Maskininlärning, Djupinlärning och Processautomation (Watson, 2018). Efter att McCarthy myntade begreppet 1955 stod den praktiska användningen av AI mer eller mindre stilla. Det var inte förrän informationsålderns ankomst vid tidigt 2000-tal och processorkraften i datorerna utvecklades som man verkligen kunde använda AI i praktiken (TEDx Talks, 2016). Varför denna tidpunkt kom att bli en vändpunkt för den praktiska användningen, var för att datorer nu snabbare kunde tillgodogöra sig data än vad människan kunde göra (TEDx Talks, 2016). Detta är i grund och botten vad AI handlar om, att samla in data, tolka den och sedan använda den för att kunna ta bättre och mer grundade beslut (European Commission, 2019).

Effekten av att en dator kan utföra många uppgifter bättre än en människa, har under informationsåldern växt exponentiellt. Medgrundaren till Intel, Gordon Moore, lade fram en hypotes om att utvecklingen av tillverkningen av datorchip, borde göra det möjligt att fördubbla antalet transistorer på ett chip vartannat år (Stair et al., 2018). Hypotesen blev känd som Moores lag och har sedan 40 år tillbaka blivit ett mål tillverkare uppnått, vilket har lett till att produktiviteten och prestandan kunnat förbättras och datorer har nu således blivit mer kraftfulla och en del av vårt dagliga liv (Stair et al., 2018). Ett historiskt känt exempel på att en dator kan utföra en uppgift bättre än en människa kan lyftas från 1997. Dåvarande världsmästaren i schack, Garri Kasparov, besegrades i en schackmatch av den uppgraderade superdatorn Deep Blue (History, 2021).

Eftersom processorkraften blivit så pass stark i datorer kan de nuförtiden hantera stora mängder data. Neil Assure och Kayvaun Rowshankish (2022) förutspår i ett blogginlägg från McKinsey att vid 2025 kommer nästan varje medarbetare att använda data i sitt dagliga arbete för att optimera sitt arbetssätt. Avsikten för detta datadrivna arbetssätt är att med hjälp av relevant och korrekt data kunna ta bättre beslut, vilket leder till att nya insikter och möjligheter skapas som således leder till mervärde för en verksamhet (Datadrivet, n.d).

Ett mer nutida exempel på hur långt utvecklingen har kommit och vad datadriven AI, i detta fall maskininlärning, kan göra nuförtiden blev tydligt efter den uppmärksammade Cambridge Analytica (CA) skandalen. Där drygt 87 miljoner Facebookanvändares personliga data användes av dataföretaget (Isaak & Hanna, 2018). CA använde sedan datadrivna AI-lösningar, för att rikta in sig på ambivalenta väljare (Miller, 2019). CA:s algoritmer användes för att hitta mönster i väljarnas data och därefter genom politisk marknadsföring manipulera väljaren i en viss riktning (Miller, 2019). Detta hade sedan stor påverkan på utfallet för

presidentvalet i USA 2016 (Isaak & Hanna, 2018). Skandalen gav upphov till mycket kritik och många kände att det gått så pass långt att vi inte längre hade någon kontroll över vår personliga data. Allmänheten fokuserade på att deras integritet och privatliv inkräktades av CA, medan en annan respons kom från professionella utövare och forskare. Denna var att se till förhållandet mellan AI och etik (Miller, 2019).

Forskare har således valt att lyfta och applicera etik i förhållandet till utvecklingen av AI. Rollen för etik är att ta tillbaka kontrollen över utvecklingen och att se till att faktorer såsom mänskliga, rättsliga och sociala tas i åtanke i arbetet med AI. Luciano Floridi (ed. 2021) menar på att detta är något som måste ske på rätt sätt och inte dagens "innovate first, fix later"-tankesätt, som kan vara mycket kostsamt och skapa motstånd till AI från samhället (ed. Floridi, 2021, p.43). Som en respons på detta lägger Floridi (ed. 2021, p.44) fram ett oxymoron, *Festina Lente*, det vill säga "skynda långsamt" och menar på att detta är något som måste appliceras som tankesätt när det kommer till innovationsarbete inom AI. Då kan man komma ifrån det förstnämnda destruktiva tankesättet (ed. Floridi, 2021).

I kölvattnet av kritiken har flertalet forskare därför lagt fram ett antal olika riktlinjer att förhålla sig till. Ett problem som uppstått är att dessa riktlinjer visat sig vara för abstrakta vilket innebär att de datavetare (från engelskans *computer scientist)* som utvecklar AI inte kan i praktiken, omvandla dem till något fysiskt applicerbart (Hagendorff, 2020). Eftersom riktlinjer är problematiska att implementera har det lett till att etiken i dagsläget huvudsakligen fungerar som syfte för marknadsföring (Hagendorff, 2020). Det skapar i praktiken en falsk trygghet för att lugna de som är oroliga att AI kan komma att bli den dystopiska framtidsskildring som man endast ser i Sci-fi filmer. Arbetet med etik inom AI har kommit att bli det nya miljöarbetet, ett yttre tryck från samhället som tvingar företaget att anpassa sig på ett visst sätt.

En lösning på problemen ovan har bland annat blivit att försöka minska klyftorna mellan forskare och professionella utövare inom datavetenskap. Detta kan göras genom att ändra forskarnas metodik som traditionellt sett haft en hög abstraktionsnivå och som ska vara generellt applicerbar, till en metodik som är mer specifik och praktiskt applicerbar (Hagendorff, 2020). Den mer påtagliga metodiken använder sig Europeiska kommissionens expertgrupp, High-Level Expert Group, av då de parallellt med de icke-tekniska riktlinjerna tagit fram tekniska och mer tillämpliga riktlinjer (European Commission, 2019). Hagendorff (2020) menar på att det även kommer bli nödvändigt att omvandla etik till så kallad "mikroetik" för att kunna nå ut till datavetare. Mikroetik kan sammanfattningsvis förklaras som att få forskare att använda en annan typ av metodik för att implementera etiska riktlinjer i professionella verksamheter. Det vill säga att göra tekniska instruktioner för hur implementationen av riktlinjerna ska utföras (Hagendorff, 2020). Detta kan till exempel göras genom att transformera etik till en mer branschspecifik dataetik eller maskinetik. På så sätt kan man specialisera implementationen av etik till specifikt område och minska abstraktionsnivån. Detta leder till att mänskliga, rättsliga och sociala riktlinjer når ut till de professionella områdena (Hagendorff, 2020).

1.2 Problemformulering

Artificiell intelligens kan leda till disruptiva innovationer som verksamheter måste ta ställning till för att kunna fortsätta vara konkurrenskraftiga. AI är fördelaktigt att tillämpa för att behandla data, den data som nuförtiden är alltmer betydelsefull och svårhanterlig. Efter att Sveriges riksdag (2018) införde Dataskyddsförordningen (GDPR) i Sverige fick dessutom verksamheter, som hanterar data i form av personuppgifter, många rättsliga krav på deras hantering av dem (Integritetsskyddsmyndigheten, 2021a). Vilket vidare problematiserar behandlingen av data i praktiken. Tanken med GDPR är att det ska skapa ett skydd för enskildas uppgifter så att de inte missbrukas i fall såsom tidigare nämnt med CA. Däremot genereras det utöver personuppgifter otroligt stora mängder data i den sociotekniska interaktion som sker mellan människa och dator genom exempelvis internet. Denna data ligger sedan som underlag till olika AI-lösningars algoritmer. Problemet är att data speglar och framhäver samhällets redan existerande fördomar, såsom social orättvisa, ojämlikhet och diskriminering (Leavy et al., 2020). Dessa fördomar eller bias i AI-sammanhang, är ofta kopplade till kön eller ras. En risk med att använda data med mycket bias är att de beslut som tas av AI-lösningarna kan få grava konsekvenser i form av rasism eller diskriminering (Ntoutsi et al., 2020).

För att bemöta problemen med databias och framhäva etik har flera forskare lagt fram olika etiska riktlinjer, exempelvis den ideella organisationen Association of Computing (ACM), som tagit fram ett antal uppförandekoder för professionella utövare inom mjukvaruutveckling. De har som mål att influera det praktiska tillvägagångssättet hos utövare för att de ska agera ansvarsfullt i förhållande till samhället (Association of Computing Machinery, 2018a). Andrew McNamara, Justin Smith och Emerson Murphy-Hill (2018) gjorde en omfattande undersökning över i vilken utsträckning ACM:s uppförandekoder används i praktiken, både genom att passivt analysera om professionella utövare använde riktlinjerna och genom att explicit be en kontrollgrupp att använda sig av riktlinjerna. Resultatet av McNamara et als (2018) undersökning visade på att dessa uppförandekoder inte hade någon som helst observerad effekt på det etiska beslutstagandet i praktiken. Som utgångspunkt i att dessa uppförandekoder inte hade någon påverkan ställer han då frågan om vilka tekniker alternativt riktlinjer som kan tänkas förbättra det etiska beslutstagandet vid mjukvaruutveckling i AI-sammanhang. Ytterligare ett problem har observerats av Hagendorff (2020), som menar på att abstraktionsnivån för dessa riktlinjer är så pass hög att det blir komplext att fysiskt applicera dem vid arbete med AI. Han menar på att riktlinjerna blir i praktiken istället en form av marknadsföring för att visa på att verksamheten tar sitt etiska ansvar. Louise Bezuidenhout och Emanuele Ratti (2021) menar ytterligare på att detta makroetiska förhållningssätt, som har en hög abstraktionsnivå, är svårapplicerbart hos datavetare i deras dagliga aktiviteter. Bezuidenhout och Ratti (2021) påpekar ytterligare att problemet grundar sig i hur dataetik lärs ut.

Den kritik som riktats mot AI är således inte helt obefogad. Relationen mellan AI och etik är, med andra ord, till viss del problematisk. Det har framlagts att det etiska perspektivet är väsentligt för att återta kontrollen och lindra bias men om de lösningar som lagts inte bär frukt kvarstår problemen. Dessutom blir etik i relation till AI som mest påtaglig i datadrivna AI-lösningar, då det som tidigare nämnts förekommer mycket bias kopplat till demografiska aspekter såsom kön och ras i datadrivna AI-lösningar. Nyare forskning lyfter begreppet mikroetik och menar på att detta kan ses som en lösning till problemet. Detta innebär att det därmed inte finns mycket forskning om innebörden av det och särskilt inte om mikroetik efterlevs i praktiken samt i synnerhet inte på den svenska marknaden. Med avstamp i de

problem som lyfts lämpar det sig således till att ställa följande forskningsfråga.

1.3 Forskningsfråga

I vilken utsträckning är verksamheter som arbetar med datadriven AI på den svenska marknaden medvetna om etik och mikroetik, samt vet de hur det kan implementeras?

1.4 Syfte

Syftet med studien är att få en klarare bild över hur organisationer på den svenska marknaden som arbetar med datadriven AI, eller är i tidigt skede med det, ser på relationen mellan AI och etik. Studien har en utgångspunkt i etablerade riktlinjer anpassade för AI, för att undersöka om dessa tillämpas av professionella utövare. Studien syftar slutligen till att undersöka om innebörden av mikroetik och hur det ska implementeras, är något som hittas i praktiken. Resultatet av detta klargör vilken medvetenhet som finns hos valda organisationer. Studien bidrar således med att precisera den nuvarande medvetenheten om etik i praktiken, vilket gör att ett ställningstagande kan göras kring hur fortsatt forskning om etik och framförallt mikroetik bör göras.

1.5 Avgränsningar

Då AI kan betraktas som ett paraplybegrepp med många utvecklingsområden är det väsentligt med avgränsningar. Därför kommer undersökningen inte fokusera på andra inriktningar av AI förutom de som ligger inom området för datadriven AI. Ytterligare görs en avgränsning inom fältet etik där fokus huvudsakligen kommer ligga på mikroetik. Ingen hänsyn kommer tas, utöver normativ etik, till de övriga inriktningarna inom etik. Denna studie kommer inte använda källor som är äldre än från det senaste decenniet. Detta då vi anser att IT-branschens utveckling och framför allt utvecklingen av AI sker i ständig takt, det blir således mest relevant att se till den senaste forskningen som finns. Slutligen kommer en avgränsning göras till verksamheter på den svenska marknaden. Med detta menas verksamheter som har kontor i Sverige och bedriver sin verksamhet på den svenska marknaden. Således kommer undersökningen undersöka verksamheter som arbetar på den svenska marknaden med datadrivna AI-lösningar.

2 Litteraturgenomgång

Detta kapitel redogör för den del av forskningen och litteraturen om AI och etik som lämpar sig bäst för att besvara den forskningsfråga som ställts. Inledningsvis kommer de två begreppen Artificiell Intelligens, etik och databias att beskrivas, där vi även redogör för vad som menas med datadriven AI då detta är den inriktning av AI som vi kommer att undersöka. Fortsättningsvis lyfts etablerade riktlinjer och dataskyddsförordningen, sedan beskrivs den respons som har kommit från dessa. Implementeringen av begreppet Mikroetik förklaras och kapitlet avslutas med att det teoretiska ramverket presenteras.

2.1 Definitioner

2.1.1 Artificiell Intelligens

Den definition av AI som studien utgår från beskrivs av Europeiska kommissionens High Level Expert Group (HLEG) i rapporten *Ethics Guidelines for Trustworthy AI* som publicerades 2019. Definitionen beskrivs som sådan:

"Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions." (European Commission, 2019)

2.1.2 Datadriven AI

All sorts AI innefattar någon sorts data, det som särskiljer datadriven AI från AI är just fokus på datakvalité och framför allt datamängd. Vad som menas med datadriven AI är att nuförtiden drivs AI-lösningar av stora mängder data i algoritmer (Ntoutsi et al., 2020). Med algoritmer menas modeller baserade på data från samhället, data som samlats in med datautvinning (Žliobaitė & Custers, 2016). Dessa modeller är förutsägande och kan således förutspå vilka beslut som bör tas med hjälp av beslutsunderlaget i insamlad data (Žliobaitė & Custers, 2016). Modeller som dessa bestämmer exempelvis vilka annonser och nyheter som visas för användare på internet samt vem som blir utvald till att genomgå en extra säkerhetskontroll på flygplatsen (Žliobaitė & Custers, 2016).

Ett av de vanligaste användningsområdena inom datadriven AI är maskininlärning. Inom datavetenskap betraktas maskininlärning som en del av AI där huvudfokuset ligger på att använda data och algoritmer (IBM, 2020). Algoritmer tränas till att göra klassifikationer eller förutsägelser genom att använda statistiska metoder. Dessa metoder kan ge viktiga insikter som senare kan driva beslutsfattandet (IBM, 2020). Maskininlärning handlar huvudsakligen

om att en algoritm observar data, bygger därefter en modell baserat på denna data och använder modellen både som en hypotes om algoritmens omgivning men även som problemlösande mjukvara (Russel & Norvig, 2022). Algoritmen kommer att utföra sina handlingar eller förutsägelser utifrån de mönster och likheter algoritmen känner igen från tidigare erfarenheter. På samma sätt som människor lär sig av sina misstag kommer algoritmen och dess precision samt utförande gradvis att förbättras. Detta innebär att utvecklare inte behöver skriva kod som täcker alla möjliga utfall (Dignum, 2019).

2.1.3 Etik

"Ethics is knowing the difference between what you have a right to do and what is right to do." - Potter Stewart (Dignum, 2019, p.35)

Med bakgrund i citatet ovan menar Virginia Dignum (2019) på att de olika systemen med AI-lösningar som utvecklas kommer att ta beslut, stora som små, som påverkar våra liv. Det kommer därför vara nödvändigt att AI tar hänsyn till samhällets värderingar och moraliska samt etiska överväganden. Dignum (2019) lägger fram att dagens AI-maskiner saknar förmågan att ta beslut som innehar en etisk natur. Hon argumenterar även för att innan en design av maskiner som tar det etiska beslutsfattandet i åtanke är det viktigt att skapa sig en förståelse för de teorier som beskriver hur etik används för att ta beslut. Detta beskrivs i följande stycke.

Nuförtiden delar dagens filosofer in etik i tre olika kategorier *meta-etik, applicerad etik* samt *normativ etik* (Dignum, 2019). Virginia Dignum påpekar i hennes bok *Responsible Artificial Intelligence: How to develop and use in a Responsible way* från 2019 att det är framför allt den sistnämnda kategorin av etik, *normativ etik*, som är mest användbar i förhållande till artificiella system. I *normativ etik* finns det därefter flera olika inriktningar, *konsekventialism, deontologi* och *dygdetik.* Det finns, förutom dessa tre, fler inriktningar men dessa tre är av särskild betydelse för att kunna förstå och applicera etik i förhållande till artificiella system (Dignum, 2019).

1. Konsekventialism är i stora drag att om man står inför ett val ska man välja det alternativ som ger det bästa möjliga utfallet oberoende på hur man kommer dit (Stanford Encyclopedia of Philosophy, 2016a). Med andra ord är konsekventialism en inriktning som förespråkar "ändamålet helgar medlen".

2. Den andra inriktningen vid namn deontologi är mer eller mindre motsatsinriktningen till konsekventialism. Deontologi handlar om att följa plikter och regler och lutar sig mot principen att hur bra ett visst utfall än är, finns det vissa val som är moraliskt sett förbjudna. Om man ska sätta dessa två inriktningar mot varandra skulle en person som följer den deontologiska inriktningen aldrig kunna mörda en människa för att rädda sju, medan en person som följer den konsekventialistiska däremot skulle kunna göra det (Stanford Encyclopedia of Philosophy, 2016b).

3. Den tredje inriktningen vid namn dygdetik är det förhållningssätt som betonar dygderna eller den moraliska karaktären. Det som urskiljer dygdetik från de övriga två är dygdens centralitet inom teorin. En dygd kan förklaras som en exemplarisk egenskap av ens karaktär. Dygdetik lägger vikt vid betydelsen av att utveckla goda

karaktärsdrag som till exempel välvilja, mod och ärlighet. Det innebär att en person som lever efter dygdetiken automatiskt utför goda eller rätt handlingar utifrån deras goda karaktärsdrag (Stanford Encyclopedia of Philosophy, 2016c).

2.1.4 Databias

Inom maskininlärning är bias inget nytt fenomen och syftar traditionellt sett till antaganden gjorda av specifika modeller (Ntoutsi et al., 2019). Bias är välstuderat i många olika discipliner men den definition som lämpar sig bäst för denna studie är den definition som beskriver bias som sådan:

> *"the inclination or prejudice of a decision made by an AI system which is for or against one person or group, especially in a way considered to be unfair" (*Ntoutsi et al., 2019 *p.3)*

Dagens AI-lösningar är som tidigare nämnt högst beroende av data. Data är sociotekniskt sett genererad då den via olika system skapas genom interaktionen mellan människa och dator. Detta gör att den data som skapas och därefter används i systemen grundas på de fördomar som människor och samhället innehar. Dessa fördomar eller bias kan till och med bli förstärkta i komplexa system såsom på internet (Ntoutsi et al., 2019). Detta resulterar i att algoritmerna som körs i AI-lösningarna avbildar och i vissa fall även förstärker redan existerande orättvisor och diskrimineringar. Detta kulminerar i att vissa grupper i samhället kan bli ofördelaktigt behandlade medan andra blir fördelaktigt behandlade. Rasism och sexism är vanliga exempel på bias i AI-system (Ntoutsi et al., 2019).

2.2 Etablerade riktlinjer

2.2.1 Europeiska kommissionens riktlinjer

Att en mer teknisk och specificerad metodik är nödvändig är något som Europeiska kommissionens expertgrupp, High-Level Expert Group (HLEG) börjat inse är nödvändig. Denna expertgrupp består av 52 experter från näringslivet, akademin och offentlig sektor. Gruppen har specifikt blivit handplockad för att främja utvecklingen av AI och för att understryka precis vad som gör AI tillförlitlig (European Commission, 2019). De tekniska instruktionerna för att förverkliga riktlinjerna nedan finns beskrivna i avsnitt 2.4.1. Riktlinjerna nedan är framtagna av gruppen och är det som ska förverkligas i avsnittet nämnt ovan. Riktlinjerna lutar sig på tre huvudpelare som ska uppfyllas i hela systemets livscykel och är det som gör systemet trovärdigt (European Commission, 2019). Dessa är följande:

1. *Lagligt* - Systemet ska vara lagligt och följa de lagar, riktlinjer och regler som finns.
2. *Etiskt* - Systemet ska vara etiskt och visa efterlevnad till de etiska principer och värderingar som finns.
3. *Robust* - Systemet ska vara robust och ha goda intentioner både från ett tekniskt och socialt perspektiv.

I en perfekt värld ska dessa komponenter arbeta i harmoni men i praktiken går de inte alltid ihop som de ska, exempelvis om omfattningen och innehållet i ett system skulle gå emot etiska normer (European Commission, 2019).

HLEG (2019) tar därefter upp ett antal fundamentala etiska principer som borde följas för att uppnå trovärdig AI-utveckling. Den första är *Respekt för Mänskligt Handlande* som kretsar kring att människor som interagerar med AI-system måste kunna, under hela livscykeln, ha möjlighet att vara en del i den demokratiska processen. Dessutom får inte ett AI-system lura eller manipulera människor, utan det är väsentligt att de är designade för att komplettera och bemyndiga människor. Den andra principen vid namn: *Förebyggande av Skada* syftar till att AI-system inte får göra eller förvärra skada mot människor. Detta täcker även in behovet av att skydda människors värdighet samt deras fysiska och psykiska integritet. Den tredje principen heter *Principen om Rättvisa* som i stora drag handlar om att utvecklandet, spridningen och användandet av ett AI-system måste vara rättvist. Principen syftar till att säkerställa att distributionen av både fördelar och kostnader är rättvisa, samt framför allt att individer och folkgrupper inte blir ofördelaktigt behandlande eller diskriminerade. Slutligen har den fjärde och sista principen följande namn: *Principen om Förklaring.* Denna princip kretsar kring öppenhet och transparens för att användaren av ett system ska lita på det. HLEG (2019) understryker i denna punkt vikten av att processerna ska vara tydliga med vad det finns för möjligheter och syfte med dem samt att de ska gå att förklara, till både de som är direkt och indirekt involverade i processen (European Commission, 2019).

Dessa fyra principer transformeras sedan till en icke begränsad lista av konkreta krav som tillgodoser perspektiv från systemet, individen och samhället. Den ser ut som följande:

1. *Mänskligt handlande och tillsyn*
    - Inkluderat fundamentala mänskliga rättigheter, mänskligt handlande och mänsklig tillsyn.
2. *Tekniskt robusthet och säkerhet*
    - Inkluderat motståndskraft till cyberattacker och säkerhet, en alternativ plan, träffsäkerhet, tillförlitlighet samt reproducerbarhet.
3. *Privatliv och datastyrning*
    - Inkluderat att respektera privatliv, kvalitet och tillgång till data samt integriteten av data.
4. *Transparens*
    - Inkluderat spårbarhet, att kunna förklara och kommunikation.
5. *Mångfald, icke-diskriminering samt rättvisa*
    - Inkluderat att undvika orättvis partiskhet, tillgång till universal design och intressentdeltagande.
6. *Miljö och samhällets välmående*
    - Inkluderat hållbarhet och miljömedvetenhet, social påverkan samt demokrati.
7. *Ansvar*
    - Inkluderat minimering av misstag, rapportering av negativ påverkan och avvägningar (European Commission, 2019).

2.2.2 Association for Computing Machinery

Association for Computing Machinery (ACM) är den största internationella associationen av professionella datavetare. De har publicerat en uppförandekod, Code Of Conduct, som är en frivillig kod att följa (Dignum, 2019). ACM (2018b) beskriver koden som sådan:

"a collection of principles and guidelines designed to help computing professionals make ethically responsible decisions in professional practice. It translates broad ethical principles

into concrete statements about professional conduct" (Association of Computing Machinery, 2018b).

ACM Code of Ethics and Professional Conduct är tänkt att inspirera och guida professionella datavetare i deras etiska uppförande. Den grundar sig på tankesättet att allmänhetens bästa är det som alltid kommer vara av högsta prioritet. Koden riktar sig även till blivande och nuvarande utövare, studenter, instruktörer samt övriga parter som på ett effektfullt sätt använder datorteknik (ACM 2018b). Man kan även vända sig till koden i reaktivt syfte, när till exempel olika kränkningar inträffat. Uppförandekoden är skapad utifrån fyra principer som är formulerade som ansvarsförklaringar och varje princip består av ett antal riktlinjer. Dessa principer är följande: *Allmänna etiska principer, Professionellt ansvar, Professionella ledarskapsprinciper* och *Överensstämmelse med koden.* Förklaringar till varje riktlinje tillhandahålls för att hjälpa datavetaren att förstå principen och hur den ska appliceras. ACM påpekar dock att deras uppförandekod är tänkt att utgöra en utgångspunkt för etiskt beslutsfattande och inte är en exakt formel för att lösa etiska problem (ACM 2018b). Riktlinjerna kopplade till varje princip är översatta och sammanfattade. Dessa kan hittas i Appendix A.

2.2.3 Dataskyddsförordningen

Dataskyddsförordningen, även kallad GDPR (General Data Protection Regulation), är inte i samma bemärkelse riktlinjer, likt de som presenterats ovan, utan är den lagstiftning från Europaparlamentet 2016 som alla verksamheter som hanterar personuppgifter inom europeiska unionen måste följa (Integritetsskyddsmyndigheten, 2021b). Detta innefattar att följa de grundläggande principerna, som beskrivs nedan, samt att behandlingen av personuppgifter har en rättslig grund och sist men inte minst informerar de, vars personuppgifter verksamheter innehar, hur deras personuppgifter hanteras. Integritetsskyddsmyndigheten beskriver syftet bakom GDPR som följande:

*"Ett av syftena med dataskyddsförordningen (GDPR) är att skydda enskildas grundläggande rättigheter och friheter, särskilt deras rätt till skydd av personuppgifter. Rätten till privatliv uttrycks i den Europeiska konventionen om skydd för de mänskliga rättigheterna och de grundläggande friheterna (EKMR). I EKMR ges en rätt till respekt för privat- och familjeliv, hem och korrespondens. Konventionen har införts som lag i Sverige."(*Integritetsskyddsmyndigheten, 2021c).

Det centrala som de verksamheter som hanterar personuppgifter måste ta hänsyn till är en rad olika grundläggande principer. Dessa principer förklaras sammanfattande i form av en parafraserad punktlista.

1. All behandling av personuppgifter måste uppfylla tre element, den måste vara laglig, korrekt och öppen.
2. Personuppgifter ska endast samlas för ändamål som är uttryckligt angivna och berättigade.
3. Personuppgifter måste vara relevanta och korrekt för det ändamål i tanke. De ska inte heller vara för omfattande.
4. Hantering av personuppgifter ska präglas av uppdatering samt vara riktiga.
5. När ändamålet för personuppgifterna uppfyllts måste de tas bort.

6. Säkerhetsåtgärder måste tas för att skydda personuppgifterna på ett lämpligt sätt.
7. Uppvisande av att Dataskyddsförordningen med dess grundläggande principer följs ska kunna göras. Dessutom måste man kunna visa på hur de följs (Integritetsskyddsmyndigheten, 2021d).

2.3 Respons till etablerade riktlinjer

2.3.1 Mikroetik

Med avstamp i hur pass abstrakta och generella riktlinjer inom AI-etik historiskt sett varit, har frågor uppkommit om hur riktlinjerna ska kunna effektiviseras för att de ska kunna få en praktisk användning. Hagendorff (2020) problematiserar detta genom att peka på att dessa riktlinjer oftast använder benämningen "AI" istället för att använda sig av en mer specifik terminologi, då AI är ett paraplybegrepp för åtskilliga teknologier. Han understryker vidare att inte en enda framstående etisk riktlinje i hans efterforskning går in i teknisk detalj när det kommer till implementation, vilket tydliggör hur pass stor klyftan mellan datavetare gentemot etiker är.

Hagendorff (2020) argumenterar för att den till synes självklara framtiden för etik ligger i att göra tekniska instruktioner till de etiska riktlinjerna, exempelvis hur en datavetare ska kunna implementera transparens i ett visst system. Etiker inom AI måste åtminstone kunna greppa tekniska detaljer i deras ramverk, vilket innebär bland annat att kunna förstå hur data genereras, processas och används (Hagendorff, 2020). För att kunna göra detta på ett värdefullt sätt, föreslår han att etiken bör omvandlas till mikroetik. Mikroetik handlar om att transformera etik till något specifikt applicerbart inom ett professionellt område såsom maskinetik eller dataetik. Ett specifikt exempel på denna transformation är att inom maskininlärning använda sig av ett antal standardiserade dataset med tydliga innehållsförteckningar. Dataset förklaras som följande av Sveriges riksdag (n.d) "*Dataset är samlingar av strukturerad data som kan laddas ned och bearbetas vidare"*. Hagendorff (2020) beskriver att en datavetare kan sedan jämföra de olika dataseten för att avgöra vad intentionen bakom var, vilken data det innehåller samt hur den var insamlad, för att slutligen kunna avgöra vilket set som passar bäst för just dem. Genom denna metodik kan datavetaren ta ett mer informerat och rättvist beslut varför den valt ett visst dataset vilket leder till att arbetssättet blir mer transparent och således mer etiskt (Hagendorff, 2020).

I sin artikel *What does it mean to embed ethics in data science? An integrative approach based on microethics and virtues* från 2021 skriven av Louise Bezuidenhout & Emanuele Ratti föreslår författarna hur man kan "bädda in" etik för att bemöta de etiska problem som uppstår i det dagliga arbetet. Genom att använda modellen för mikroetik, som i sig grundar sig i ett ramverk av dygdetik, vill Bezuidenhout och Ratti visa hur man kan lära ut dagligt ansvar i digitala aktiviteter. När det kommer till dataetik har forskare de senaste åren ifrågasatt om den metodik som råder, som grundar sig på generella principer och fallstudier på högre nivå, är effektiv (Bezuidenhout & Ratti, 2021). Principer såsom de tidigare presenterade riktlinjerna från ACM (Association of Computing Machinery, 2018a) och genom Europeiska kommissionens initiativ HLEG (2019). Bezuidenhout & Ratti (2021) menar på att en del av kritiken som uppstått grundar sig i att det i en datavetenskapskontext blir svårt att applicera och lära ut denna metodik. De menar på att på principer och regler inte är samma sak, utan att regler är precisa och ordentliga. Principer har inte samma tyngd utan kan således vägas mot

varandra i beslutstagande. De understryker att trots denna skillnad mellan principer och regler så förväntar sig människor att principer är som regler. Författarna fortsätter med att forskning om hur detta kan teoretiseras och användas saknas. Således lägger författarna fram fullfjädrad karaktärisering av "inbäddnings"-etik och hur detta kan appliceras, vilket presenteras nedan i avsnitt 2.4.2. De lyfter den framväxande modellen av mikroetik och presenterar ett sätt att lära ut dagligt ansvar i samband till digitala aktiviteter (Bezuidenhout & Ratti, 2021).

2.3.2 ACM:s Uppförandekod i praktiken

För att uppmärksamma och ändra professionella utövares metodik till att bli mer etiska och ansvarstagande har, som tidigare nämnts, flertalet riktlinjer kopplat till etik framkommit. En av de mest etablerade är ACM Code of Ethics som nyligen uppdaterades år 2018. Dessa riktlinjer har som mål att kunna användas av datavetaren för att både proaktivt och reaktivt ta bättre beslut och förhindra diskriminering och orättvis behandling. McNamara et al. (2018) gjorde en omfattande etisk beteendestudie med 63 mjukvaruutvecklarstudenter och 105 mjukvaruutvecklare. Detta genom att analysera hur många av deltagarna i studien som i sitt vardagliga arbete använde sig av riktlinjerna och genom att explicit be deltagarna i studien att använda sig av riktlinjerna. McNamara et al. (2018) hittade inga statistiska bevis på att ACM:s riktlinjer hade någon som helst påverkan på vare sig studenterna eller de professionella utövares etiska besluttagande i praktiken.

2.3.3 Dygdetik

Hagendorff (2020) menar på att det utöver vikten av att applicera mikroetik och tekniska riktlinjer, även är väsentligt att försöka besvara frågan om hur de etiska riktlinjerna inom AI kan förbättras. Han lyfter vidare även att för att kunna förbättra riktlinjerna vid tillämpning och fullbordande är det nödvändigt att se till de olika etiska teorierna som finns etablerade.

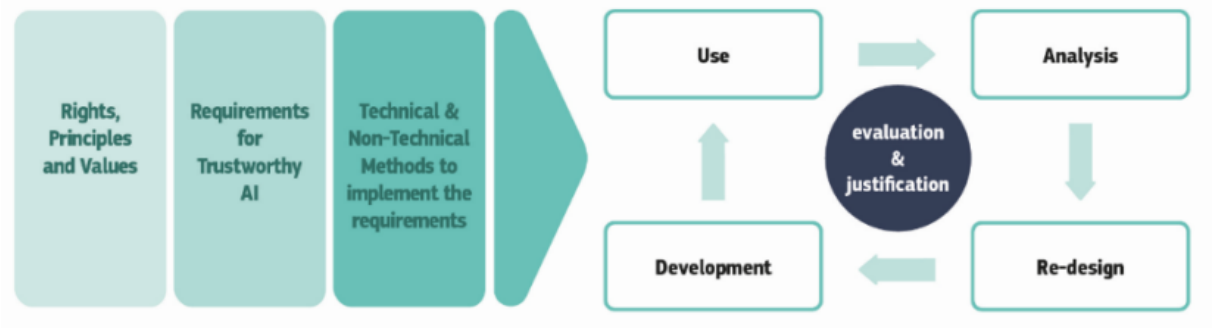
Det Hagendorff (2020) föreslår är att det deontologiska förhållningssättet bör utökas och kombineras med det dygdetiska, vars mål är att kultivera drag av moralisk karaktär. Karaktärsdrag såsom ärlighet, mod, rättvisa, empati, omsorg och artighet. Dessa drag eller dygder ska öka sannolikheten för ett etisk beslutsfattande i organisationer. Genom att kombinera dessa två förhållningssätt kan den traditionella AI-etiken, som i stora drag handlar om att uppfylla krav och regler för att följa normer kring AI, täckas av det deontologiska förhållningssättet medan det dygdetiska förhållningssättet används för att stärka värderingar, utveckla personligheter och ändra attityder. Etik kommer på så sätt utvecklas från att vara en kravlista som ska bockas av till att bli ett projekt som låter personligheter utvecklas, förändrar attityder och stärker ansvar vilket leder till att handlingar som anses vara oetiska fasas ut undermedvetet (Hagendorff, 2020).

Hagendorff (2020) understryker att om de tidigare nämnda förändringarna ska kunna äga rum är det väsentligt att IT-branschen accepterar det dygdetiska förhållningssättet. Dessutom bör flertalet institutionella förändringar göras, exempelvis att drastiskt förändra läroplaner inom tekniska utbildningar på universitet runt om i världen, för att främja det etiska perspektivet (ed. Floridi, 2021).

2.4 Implementering av Mikroetik

2.4.1 Förverkligande av AI

Europeiska kommissionens expertgrupp har i sin rapport från 2019, förutom de olika riktlinjerna som tidigare nämnts, anammat en mer implementerbar metodik för professionella utövare då de tar upp olika implementationsmetoder. Implementationsmetoderna kategoriseras som antingen tekniska eller icke tekniska. Dessa ska enligt HLEG (2019) agera som komplement till varandra eller som alternativ till varandra då olika kravbilder kräver olika metoder. HLEG (2019) påpekar även att metoderna kräver olika mognadsnivåer för att fungera samt att en del av metoderna finns på marknaden redan idag, medans andra behöver studeras vidare. Nedan kommer endast de tekniska metoderna tas upp och författarnas metodik kan ses i figur 1.



Figur 1: Förverkligande av trovärdigt AI genom hela livscykeln adopterad från Europeiska kommissionen (2019)

Den första metoden vid namn *Arkitektur för Trovärdigt AI* syftar till att omvandla de krav och riktlinjer som tidigare nämnts till procedurer och begränsningar vilket ska sedan förankras i systemets AI-arkitektur. Detta ska genomföras med hjälp av en *white list* med regler, beteenden eller situationer som systemet alltid ska följa samt en kompletterande *black list* som ger restriktioner på vilka beteenden och situationer som systemet aldrig får finna sig i. Övervakningen för att se att AI-systemet efterlever de båda listorna bör ske med ett separat system eller en separat process (European Commission, 2019).

Den andra metoden med namnet *Etik och Regler vid design (x-by-design)* handlar om att redan vid införandet av ett system, ska företaget inse innebörden och konsekvenserna av ett AI-system från början. De normer som deras AI-system bör följa ska identifieras innan systemet tas i bruk för att undvika negativa konsekvenser. Detta kretsar kring “by design” konceptet som redan är etablerat inom IT i andra sektorer exempelvis *privacy by design* samt *security by design* (European Commission, 2019).

Den tredje metoden som heter *Förklaringsmetoder* syftar till behovet att ett AI-system bara kan vara trovärdigt om man förstår exakt varför systemet betedde sig på ett visst sätt eller tolkade något på ett annat sätt. HLEG (2019) påpekar att det finns ett helt forskningsområde, *Explainable AI (XAI)* som försöker avhjälpa detta problem för att bättre kunna förstå systemets underliggande mekanismer och hitta lösningar för att öka förståelsen. Detta är viktigt då små förändringar i data-input i ett system ofta drastiskt kan ändra utfallet och hur datan tolkades av systemet (European Commission, 2019).

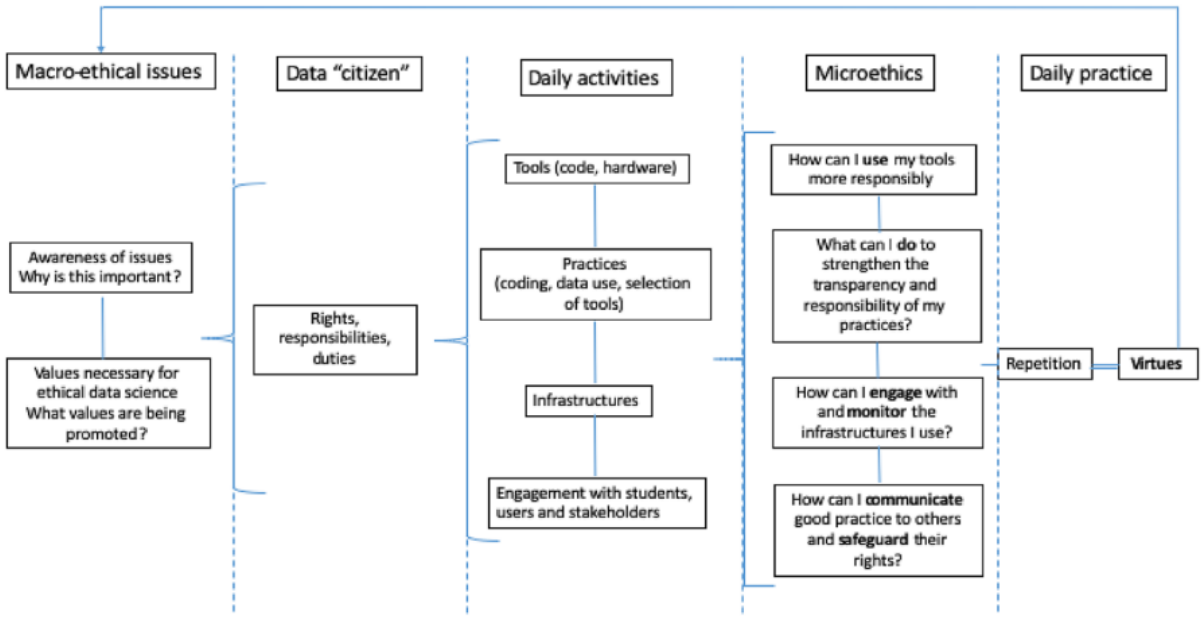
Den fjärde tekniska metoden som kallas *Testning och Validering* handlar om att understryka vikten av ordentlig testning av AI. Då AI-system är självlärande och högst kontextberoende räcker inte traditionell testning av systemet. Detta eftersom eventuella fel i koncept och representationer i många fall endast manifesterar sig när systemet ska processa surrealistisk data. HLEG (2019) argumenterar med det sagt för att modellen och systemet bör vara noggrant övervakat både när programmet tränar och distribueras. Det är därför väsentligt att testandet av systemet börjar så tidigt som möjligt i utvecklingsprocessen för att säkerställa att systemet beter sig som det är designat under hela livscykeln. HLEG trycker även på punkten att testningsprocessen bör göras av en så mångfaldig grupp som möjligt och att testningen ska inkludera alla delar av ett AI-system. Detta inkluderar komponenter såsom data, olika datamodeller, olika miljöer samt beteendet av systemet i sin helhet. Dessutom bör ett separat team med uppgiften att försöka ta sönder systemet för att hitta svagheter sjösättas, samt sist men inte minst bör output från systemet även jämföras med de tidigare definierade riktlinjerna, för att se att resultaten är i linje med dem (European Commission, 2019).

Den femte och sista metoden vid namn *Kvalitetsindikatorer* handlar om att införa kvalitetsindikatorer till AI-system som ska ge en grundläggande förståelse hos användarna att systemet blivit utvecklat med säkerhet och etikkrav i åtanke. Dessa indikatorer skulle kunna utvärdera träningen av algoritmerna och även indikera hur pass väl systemet fungerar utifrån traditionella mjukvaru-nyckeltal exempelvis funktionalitet, användbarhet och säkerhet (European Commission, 2019).

2.4.2 Inbäddningsetik

I artikeln från 2021 skriven av Bezuidenhout och Ratti tidigare presenterad lägger författarna fram sin karaktärisering av inbäddningsetik. Denna metodik riktar sig främst till etiker som lär ut etik till datavetare men är även applicerbart inom forskning och utveckling för datavetenskap. Metodiken tar form som en kombination av mikroetik och dygdetik. Författarna konkretiserar i sin metodik att etiska medvetna individer kan skapas genom att använda sig av mikrouppgifter som ett sätt främja etikarbetet. Bezuidenhout och Ratti (2021) menar på att genom att bädda in etisk reflektion i utövares dagliga uppgifter sker inbäddningen på ett diskret sätt. Tanken är att genom att etiken blir en del av den repetitiva vardagen innebär det att dygdetiken uppmärksammas och normaliseras på arbetsplatsen. Ett exempel på detta lyfts vid kodutveckling. Den kontextuella omständligheten blir här då vem som äger och har patent på koden som programmeras. Ett etiskt övervägande i denna aktivitet är att öppna för diskussion om vem som borde äga koden. Detta kan sedan problematiseras vidare genom att försöka få datavetaren att se till den etiska helheten och då fråga sig varför denna kod inte är open source kod, alltså publikt tillgänglig. Tanken med detta är att främja den mikroetiska metodiken för att främja kultiveringen av moraliska dygder (Bezuidenhout & Ratti, 2021).

Författarna har nedan gjort en modell, se figur 2, för att visualisera hela livscykeln från makroetiska problem till mikroetik på daglig nivå och preciserar hur den kan användas vid utbildning. Bezuidenhout och Ratti (2021) betonar att modellens huvudfokus ligger på att skapa förståelse för att etisk tänkande är något som sker integrerat i det dagliga, praktiska arbetet och inte är en fristående aktivitet som sker sporadiskt.



Figur 2: Faser av att lära ut mikroetik adopterad från Bezuidenhout & Ratti (2021)

2.4.3 Implementationsstrategier för ett framtida välmående samhälle

Artikeln vid namn *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, skriven av författarna Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Burkhard Schafer och Burkhard Schafer från 2018, presenterar olika etiska principer som grund för de rekommendationer de senare presenterar. Artikeln lägger fram resultatet av AI4People, ett Atomium - European Institute for Science, Media and Democracy initiativ som är tänkt att skapa grunden till ett bra AI-samhälle (Good AI Society). Atomium är en internationell ideell organisation baserat i Bryssel. Artikeln presenterar möjligheter samt risker, detta i form av 5 etiska principer och 20 rekommendationer som är tänkta till att bedöma, utveckla, agera som incitament och verka för bättre AI (Floridi et al., 2018).

Dessa rekommendationer ska ses som dynamiska då de inte handlar om att verksamheter ska göra en enskild investering eller införa en policy, utan något som ska kontinuerligt jobbas med för att fungera. De mest väsentliga punkterna för professionella verksamheter är följande:

- *Bedöma:* Bedöm vilka funktioner och beslutstagande mekanismer som inte ska bli delegerade till ett AI-system. Denna bedömning ska ligga i linje med sociala värderingar och samhällets åsikter. Bedömningen bör ta hänsyn till existerande lagstiftning, samt stödja den pågående dialogen om AI:s påverkan på samhället (Floridi et al., 2018).
- *Utveckla:* Utveckla funktioner för övervakning för att identifiera missgynnande utfall. Detta kan exempelvis göras i samarbete med försäkringsbranschen. Ett annat utvecklingsområde som är relevant för organisationer är att utveckla nyckeltal kopplade till pålitligheten för verksamhetens AI-system. Nyckeltalen ska sedan kunna användas för användarstyrd benchmarking. Ett av nyckeltalen föreslås kunna vara ett

"trust comparison index" som skulle kunna öka samhällets förståelse för systemet och framför allt skapa konkurrens mellan olika verksamheter när det gäller utvecklingen av en säker och samhällsförmånlig AI-lösning (Floridi et al., 2018).

- *Incitament:* Rekommendationerna kretsar kring att försöka få europeiska eller nationella institutioner att ge finansiella incitament till verksamheter för att driva viljan till att utveckla mer trovärdig, transparent samt etisk AI (Floridi et al., 2018).
- *Stödja:* Stödja utvecklandet av uppförandekoder som har ett etiskt fokus för datavetare och systemutvecklare. Dessutom bör utvecklingen av certifiering inom etisk AI stödjas, genom att skapa olika pålitliga certifikat kommer kunder inse det värde etisk AI har och kommer därefter kräva det från leverantörer av AI-system (Floridi et al., 2018).

2.5 Teoretisk sammanfattning

Tabell 1: Teoretisk sammanfattning

Huvudområde	**Beskrivning**	**Litteratur**
2.1 Definitioner	I 2.1 beskrivs de definitioner AI, datadriven AI och Etik som vi återkommande berör i undersökningen.	European Commission (2019), IBM (2020), Russell & Norvig (2022), Dignum (2019), Stanford Encyclopedia of Philosophy (2016) (Ntoutsi et al., 2020), (Žliobaitė & Custers, 2016)
2.2 Etablerade riktlinjer	I 2.2 lyfts etablerade riktlinjer från Europeiska kommissionen och Association of Computing Machinery samt Dataskyddsförordningen beskrivs.	European Commission (2019), Dignum (2019), ACM (2018), (Integritetsskyddsmyndigheten, 2021)
2.3 Respons till etablerade riktlinjer	I 2.3 redogörs för den respons som uppkommit från de etablerade riktlinjerna. Begreppet mikroetik förklaras, de etablerade riktlinjerna problematiseras och ett tillvägagångssätt för att lösa problemen presenteras.	Hagendorff (2020), Sveriges riksdag (n.d), Bezuidenhout & Ratti (2021), McNamara et al. (2018), Floridi (ed. 2021)
2.4 Implementering av Mikroetik	I 2.4 presenteras praktiska metoder för implementering av mikroetik i professionella verksamheter.	European Commission (2019), Bezuidenhout & Ratti (2021), Floridi et al. (2018),

3 Metod

I detta kapitel presenteras tillvägagångssättet för hur studien gått till. Inledningsvis presenteras vilket metodval som gjorts och hur datainsamlingen och urvalen gjordes. Fortsättningsvis beskrivs genomförandet av intervjuer och etiska överväganden förklaras. Kapitlet avslutas med en diskussion kring reliabilitet och validitet.

3.1 Metodval

Det metodval som vi valt att grunda vår undersökning på kommer vara av kvalitativ art. Med en kvalitativ metod menas att samla in data i form av ord (Jacobsen, 2002). Bakgrunden till valet av kvalitativ metod berodde delvis på att de är mer öppna för ny, överraskande information (Jacobsen, 2002). En kvalitativ metod är mer flexibel och fungerar väl för deduktiva ansatser. Det som skiljer induktiva och deduktiva ansatser från varandra är strategin för undersökningen (Jacobsen, 2002). Den ansats som valdes för undersökningen var deduktiv, då det först skapades en uppfattning om ämnet genom litteratur och teori (Jacobsen, 2002). Ett teoretiskt ramverk skapades utefter den insamlade litteraturen och därefter samlades empiri in i form av intervjuer. Detta för att se om ramverket hade stöd i praktiken. En induktiv studie däremot har en strategi som börjar från andra hållet, det vill säga där empiri blir till teori (Jacobsen, 2002). Den kvalitativa metoden kommer bestå av tre intervjuer med verksamheter som har kontor i Malmö. Valet av en kvalitativ studie gjordes då till skillnad från en kvantitativ, som grundar sig på numerisk data, lämpar sig en kvalitativ bättre när det sker en insamling och analys av data, där tyngden ligger på ord istället för siffror (Bryman, 2018). Då dessutom inga av resultaten från den empiriska studien är kvantitativa lämpar sig dessutom kvalitativa intervjuer bättre. En analys av empirin från undersökningen kommer därefter jämföras med den tidigare framlagda forskningen inom området som presenterats i kapitel 2.

3.2 Datainsamling

3.2.1 Sökning av litteratur och teori

De teorier och källor som tagits fram har hittats uteslutande genom Google Scholar och LUBsearch. Dessa har hittats genom följande sökord:

- AI
- Artificial intelligence
- AI Ethics
- Data driven AI
- Datadriven ethics
- Datadriven

- Microethics
- Mikroetik
- Ethics
- Virtue Ethics, Deontology Ethics, Consequentialism Ethics
- AI Ethics Implementation
- AI Ethics Guidelines
- Dataskyddsförordningen
- GDPR
- Databias

3.2.2 Urvalskriterier till organisation

Då området för vår undersökning var datadriven AI, ansåg vi det av största relevans att rikta oss mot företag som sysslade med utveckling, tillämpning och applicering av någon form av datadriven AI. Ytterligare ett kriterium var att organisationen skulle vara verksam i Sverige. Det gjordes även ett bekvämlighetsurval, där vi valde vi att rikta oss mot verksamheter i södra Sverige, detta av logistiska skäl.

3.2.3 Urvalskriterier till intervjupersoner

Utgångspunkten för det urval vi gjort grundar sig i ett så kallat målstyrt urval. Innebörden av detta är att inte välja intervjudeltagare på slumpmässig basis, utan att det sker ett strategiskt val av deltagare som är av betydelse för forskningsområdet (Bryman, 2018). Val av urvalsmetoder landade i en kombination av *kriteriestyrt urval* samt *stratifierat målstyrt urval* (Bryman, 2018). Det kriteriestyrda urvalet användes eftersom vi endast valde intervjupersoner utifrån det område med tydliga avgränsningar och krav som beskrivits ovan. Stratifierat målstyrt urval är ett urval där exempelvis individer eller typiska fall väljs utifrån vilken subkategori de tillhör (Bryman, 2018). I vårt fall valde vi att rikta in oss på organisationer som sysslade med datadriven AI och således kontaktades personer för intervjun som arbetade med just detta. Dock var intervjupersonen från Malmö Stad, Jim Samuelsson, i uppstartsfasen med sitt arbete med AI och hade inte ännu börjat använda AI på deras data. Med tanke på att Jim hade en bakgrund av att arbeta med AI och data samt att det var intressant att se ett annat perspektiv, valde vi att genomföra intervjun med Jim ändå.

Tabell 2: Valda intervjupersoner

Företag	Intervjuperson	Roll
Capgemini	Informant 1	Konsult (inom maskininlärning, datavetenskap och datateknik). Med en bakgrund inom Natural Language Processing (NLP)
Malmö Stad	Jim Samuelsson	Data scientist på digitaliseringsenheten inom Arbetsmarknad och Socialförvaltningen

Nexer	Magnus Perman	Data science Tech Lead & Kompetensområdesansvarig för Data Science inom Region Syd

3.3 Intervjuer

3.3.1 Intervjuguide

I kapitel 2 har tidigare forskning presenterats, där området för forskningen varit AI och etik. Således har de intervjufrågor som framställts baserats på den litteratur som presenterats. I tabellen nedan listas de intervjufrågor vi har ställt och deras koppling till litteraturen specificeras.

Intervjuguiden har modifierats till viss del utifrån de olika intervjupersonerna, men strukturen på intervjuerna och de olika huvudområdena har varit detsamma. Detta för att personifiera frågorna då vissa frågor lämpade sig mer till en viss intervjuperson än till en annan. Dessutom uppkom vissa följdfrågor spontant under diskussionerna vilket ledde till att intervjuguiden kan vara missvisande i jämförelse med vissa av transkriptionerna. Se appendix B-D för kompletta transkriberingar.

Tabell 3: Intervjuguide

Huvudområde	**Beskrivning**	**Frågor**	**Litteratur**
Introduktion	- Intro, roll, ansvarsområde	*- Vad är din roll på …?* *- Hur länge har du arbetat inom området?* *- Vad har du för tidigare bakgrund?* *- Vad har du för arbetsuppgifter och ansvarsområden?* *- Hanterar ni personlig data i ert arbete?*	---
Definitioner	- Artificiell intelligens - Maskin-inlärning - Etik	*- Hur skulle du definiera AI och maskininlärning?* *- På vilket sätt använder ni AI på er data?* *- Vad är din definition av etik?* *- Hur ser du på förhållandet mellan AI och etik?* *- Hur skulle du påstå att etik påverkar ert arbete?* *- Påverkar yttre faktorer bl.a. politik och folkopinion ert arbete?*	European Commission (2019), IBM (2020), Russell & Norvig (2022), Dignum (2019), Stanford Encyclopedia of Philosophy (2016), (Ntoutsi et al., 2020), (Žliobaitė & Custers, 2016)

Etablerade riktlinjer	- Riktlinjer - Praktiskt användning - Etikarbete	*- Vilka etablerade riktlinjer inom AI-etik är du personligen medveten om?* *- Följer ni inom organisationen några etablerade riktlinjer?* *- Skulle du påstå att det i ditt arbete är nödvändigt att följa etiska riktlinjer?* *- Om ja, upplever du att det finns en hög abstraktionsnivå till dessa riktlinjer. Dvs, svåra att applicera i praktiken.* *- Upplever du att de används i praktiken på er arbetsplats?* *Har … en uppförandekod eller policy i förhållande till etik?* *- I och med att ni är ett konsultbolag, är det vanligt att era kunder har policys kopplade till etik som ni således måste följa i ert konsultarbete?* *- Känner du till någon på din arbetsplats eller andra arbetsplatser som arbetar med maskininlärningsetik eller AI-etik?*	European Commission (2019), Dignum (2019), ACM (2018)
Respons till etablerade riktlinjer	- Mikroetik - Inbäddad etik - Utbildning - Dataset - Marknads-föring	*- Är etik något som dagligen tas hänsyn till i ert arbete med data (framförallt personlig data)?* *- Är mikroetik något du har hört talas om?* *- Vet du om det finns tekniska instruktioner inom er organisation eller inom fältet AI för hur man ska implementera etisk AI?* *- Vad skulle du säga om det fanns standardiserade dataset eller förutbildade (pre-trained) modeller med tydliga innehållsförteckningar som skulle kunna användas till olika implementationsområden? (Datasets som är transparenta med tydliga innehållsförteckningar och där intentionen bakom datan och hur insamlingen av datan gick till finns beskrivet).* *- Tror du detta skulle ge värde för er i jämförelse med de dataset ni vanligtvis använder?* *- Har du fått någon utbildning inom etik på arbetsplatsen?* *- Var AI-etik en del av din universitetsutbildning?* *- Mycket forskning pekar på att företag som använder AI, endast använder etik i marknadsföringssyfte. Hur ser du på detta?* *- Vem tycker du etikansvaret ska falla på inom en organisation?*	Hagendorff (2020), Sveriges riksdag (n.d), Bezuidenhout & Ratti (2021), McNamara et al. (2018), Floridi (ed. 2021)

Implementering av Mikroetik	- Stöd - White/black list - Nyckeltal - Utmaningar	*- Får ni stöd i form av pengar, utbildning eller marknadsföring, för ert innovationsarbete eller känner till kunder som har fått det? (Ex Vinnova eller liknande institutioner)* *- Forskning visar på att det kan vara gynnsamt att ha en white/black list för AI-algoritmer, alltså regler som systemet alltid ska följa och beteende systemet aldrig får finna sig i. Hur ser du på detta?* *- Kanske är det något ni redan har?* *- Ett förslag för att få en organisation att bli mer etisk är att "bädda" in etisk reflektion i en utvecklares dagliga arbete. Detta skulle innebära att etiken normaliseras på arbetsplatsen och öppnar upp för diskussion i frågor som berör Open Source-kod eller datasets. Hur ser du på detta, tror du att det praktiskt skulle fungera på er arbetsplats?* *- Det finns forskning som visar på att skapa certifieringar i form av nyckeltal i hur pass pålitligt ett AI-system är, exempelvis ett sorts "Trust comparison index". Detta skulle sedan användas för användarstyrd benchmarking. Detta är ett förslag på hur man skulle kunna skapa konkurrens mellan olika verksamheter när det gäller utvecklingen av samhällsförmånliga AI-lösningar. Tror du en sådan lösning skulle kunna vara genomförbar i praktiken?* *- Inom testning och validering av en AI-lösning, är ett förslag att sjösätta ett separat team med uppgiften att försöka hitta svagheter och mer eller mindre ta sönder systemet, hur ser du på detta förslag?* *- Vad är de största utmaningarna med ert AI-arbete?*	European Commission (2019), Bezuidenhout & Ratti (2021), Floridi et al. (2018),

3.3.2 Genomförande av intervju

Alla tre intervjuer valdes att genomföras digitalt via Microsoft Teams. Samtliga intervjupersoner gavs möjligheten till att antingen ha intervjun på plats men alla valde att ha intervjun digitalt då detta alternativ lämpade sig bäst för samtliga intervjupersoner. Alla intervjuer var mellan 30 och 50 minuter långa. Två av intervjuerna hölls på svenska och en genomfördes på engelska. Som Oates (2006) understryker är det lämpligt att låta intervjupersonerna få ta del av intervjufrågorna i förväg. Detta för att ge utrymme för reflektion från deras sida och för att inge en känsla av trovärdighet (Oates, 2006). Således

valde vi att erbjuda alla intervjupersoner att ta del av frågorna i förväg om de så önskade, vilket alla intervjupersoner ville.

Inledningsvis påbörjades intervjun med att vi introducerade oss själva och kortfattat beskrev ämnet samt omfattningen av vår undersökning. Därefter tillfrågades intervjupersonen om de ville vara anonyma eller inte samt om vi fick deras godkännande till att spela in intervjun. Vi erbjöd intervjupersonen att få ta del av den transkriberade intervjun om de önskade detta. När momenten ovan var genomförda påbörjades inspelningen och intervjun utefter intervjuguiden.

3.4 Dataanalys

3.4.1 Transkribering

Vi valde att spela in våra intervjuer då det var av största vikt att vara uppmärksam vid intervjun och inte bli distraherad av avbrott, vilket skulle kunna ske om man tog anteckningar under intervjuns gång. Vi ville båda även vara närvarande och följa upp på saker som sagts under intervjun (Bryman, 2018). Då inga anteckningar togs lämpade det sig därför vid ett senare tillfälle att transkribera intervjuerna, detta för att få ut så mycket meningsfull information som möjligt. Att ha all information skriftligt underlättar arbetet med den efterföljande analysen av datan.

3.4.2 Analysmetod

När en intervju var genomförd påbörjades arbetet omedelbart med transkriberingen. Detta då den information som getts i intervjun fortfarande var ny och värdefulla insikter således inte skulle gå förlorade. När underlaget var gjort, det vill säga transkriberingen, påbörjades arbetet med analysen av den. Här lyfter Jacobsen (2002) att analysprocessen utgörs av tre element, *beskrivning, kategorisering & kombination*. Den första brukar gå under namnet som tjocka beskrivningar, vilket handlar om att få en så djup och detaljerad beskrivning av datan som möjligt. *Beskrivningen* ska genomsyras av variationer och en nyanserad analys (Jacobsen, 2002). Följande element är *kategorisering* som handlar om att kategorisera, gallra och schematisera data. Detta för att få en tydlig överblick (Jacobsen, 2002). När dessa två element är genomförda ska dessa i det sista steget *kombineras* för att kunna tolkas genomgående (Jacobsen, 2002). I detta element blir det möjligt att få fram förhållanden som tidigare varit dolda (Jacobsen, 2002). Då vi gjort en kvalitativ undersökning kommer dessa tre element att ske parallellt med varandra (Jacobsen, 2002). Genom att applicera detta förhållningssätt blev analysen av empirin mer organiserad och överskådlig.

3.4.3 Kodning av intervjuer

Efter transkriberingarna strukturerades de olika intervjuerna genom att ge varje fråga och svar varsin rad som dessutom blev numrerad. Varje intervjuperson blev döpt i transkriberingen efter sina initialer, exempelvis Jim Samuelsson blev JS. En av intervjupersonerna ville vara anonym och hen fick således namnet Informant 1 med initialen INF1. Genom att strukturera på detta sätt kunde uttalanden från transkriberingen med enkelhet kopplas till empirin och vice versa. Denna referering fungerar på det sättet att, efter ett uttalande lagts till i empirin, skrivs initial och radnummer inom parentes i slutet av meningen, exempelvis (JS, rad 46). Vi

har även färgkodat intervjuerna utefter huvudområden definierade i det teoretiska ramverket. Detta innebär att varje gång en intervjuperson nämner något kopplat till ett av områdena stryks detta över med vald färg. Färgvalet utefter huvudområdet beskrivs i tabellen nedan.

Tabell 4: Färgkodning

Huvudområde	**Färg**
Definitioner	Blå
Etablerade riktlinjer	Grön
Respons till etablerade riktlinjer	Röd
Implementering av mikroetik	Lila

3.5 Etiska överväganden

Vid en intervju är det viktigt att vara medveten om de etiska principer som finns. Några av dessa är fyra olika krav, det vill säga, *informationskravet*, *samtyckeskravet*, *konfidentialitetskravet* och *nyttjandekravet* (Bryman, 2018). *Informationskravet* handlar om att man som forskare ska berätta om undersökningens syfte för berörda intervjupersoner. Här ska det bli tydligt för intervjupersonen hur upplägget för intervjun ser ut, att deltagandet är frivilligt och att det inte finns några konsekvenser om man inte längre vill delta. *Samtyckeskravet* innebär att en intervjuperson har själv rätt att bestämma över i vilken mån den ska medverka. *Konfidentialitetskravet* innebär med att all data och information om de deltagande i studien, ska respekteras och behandlas med stor konfidentialitet. Personuppgifter ska dessutom förvaras så att ingen utomstående kommer åt dem. *Nyttjandekravet* innebär att den data som samlas in, endast ska användas i forskningssyfte (Bryman, 2018). För att minska risken för att information som intervjuperson delgett kopplas till egen person, har vi som tidigare nämnts erbjudit alla intervjudeltagare möjligheten till att vara anonyma. Vi har dessutom klarlagt syftet för undersökningen och i vilken utsträckning intervjun kommer vara. Detta för att minimera risken för att intervjupersonerna vägrar delta i undersökningen (Jacobsen, 2002).

Varje deltagande intervjuperson fick även möjligheten att ta del av sin transkriberade intervju. Om det mot förmodan skulle ha uppstått några misstolkningar i vår transkribering eller något annat som inte stämde, skulle varje intervjuperson få möjlighet att notera om detta så att det således skulle kunna korrigeras.

3.6 Studiens reliabilitet & validitet

Jacobsen (2002) lyfter att den empiri som samlas in i undersökningen ska uppfylla två krav. Dessa två krav berör validitet och reliabilitet, det vill säga att empirin har krav på sig att vara giltig och relevant samt tillförlitlig och trovärdig (Jacobsen, 2002). Bryman (2018) understryker att validitet är ett av de viktigaste forskningskriterierna att uppfylla och förklarar det som bedömningen om de slutsatser som tagits har en kausal koppling till undersökningen.

Inom validitet har vi både fokuserat på extern validitet och intern validitet, men framför allt på det sistnämnda.

Angående extern validitet pekar Bryman (2018) på att det finns en hel del problem som kretsar kring kvalitativa studier i förhållande till möjligheten att uppnå hög extern validitet. Detta eftersom extern validitet handlar om att täcka ett så stort urval som möjligt för att kunna generalisera urvalsgruppens åsikter (Bryman, 2018). Detta är något som är lättare att uppnå i kvantitativa studier (Jacobsen, 2002). Eftersom vår studie endast täcker tre företag från området blir studiens externa validitet inte särskilt hög.

Intern validitet handlar till skillnad från extern validitet om hur pass väl slutsatser gjorda hör ihop med varandra, om x lett till att y skett eller om det är någon annan variabel som i själva verket lett till y (Bryman, 2018). I vår studies fall pekar flera variabler på de slutsatser vi tagit. Dock har vi inte tidigare forskat inom området för AI och etik. Vi har inte heller någon praktisk arbetslivserfarenhet inom ämnet vilket kan påverka hur pass väl den interna validiteten är. Detta kan dessutom leda till att studien har lägre pålitlighet, vilket leder oss in på det andra kravet, reliabilitet.

Reliabilitet handlar i stora drag om hur pass pålitliga resultatet från undersökningen är och om de skulle vara identiska om studien gjorts om på nytt, detta är så kallad *extern reliabilitet* (Bryman, 2018). Begreppet rör också frågan om studien påverkats av tillfälligheter eller ej (Bryman 2018). *Intern reliabilitet*, till skillnad från extern, handlar om personerna som utför undersökningen har kommit överens om en gemensam tolkning (Bryman, 2018). Som tidigare nämnts har vi båda inte någon erfarenhet inom ämnet, vilket leder till att den *externa reliabiliteten* blir relativt låg. Däremot blir den *interna reliabiliteten* hög, eftersom vi inte har några tidigare förutfattade meningar inom området och gemensamt har hittat allt material tillsammans. Vi har således skapat en gemensam uppfattning därefter.

4 Empiri

I detta kapitel presenteras ett sammanfattande resultat av de tre intervjuer som gjorts. I metodkapitlet presenterades en tabell över de informanter som intervjuades, på vilket företag de jobbar och vilken roll de har. Strukturen för detta kapitel följer samma struktur som i kapitel 2 litteraturgenomgång.

4.1 Definitioner

Informant 1 lyfte att hens utbildningsbakgrund är inom matematik och fysik och att hen aldrig haft en lärare som gett hen en definition av AI. Utifrån egen erfarenhet gav Informant 1 sin definition som sådan: *"any program that learns from data in order to make predictions or analysis"* (INF1, rad 6). Jim Samuelssons definition av AI är: "*metoder om mjukvara som hjälper oss att bli klokare, som kan lära sig själv i någon mening"* (JS, rad 10). Magnus Perman gör en distinktion mellan AI och ML genom att ge följande exempel: *"Man ser en självkörande bil så är maskininlärningen och hela den biten att förutse vad som kommer hända om inte bilen svänger om 10 meter medan AI är mer att sätta en algoritm ovanpå det som utger själva handlingarna och tolkar maskininlärningen och sätter handlingar ovanpå det för att få någon sorts autonomi alltså självlärande, att den tar över en människas jobb i det fallet"* (MP, rad 10). Magnus understryker att i det arbete som de gör med AI är det centrala att ta fram insikter och hitta nya vinklar. Han menar att eftersom AI är ett "flashigt begrepp" är det ofta så att det finns förväntningar att insikterna sedan ska sättas in i till exempel ett autonomt marknadsföringssystem (MP, rad 10 & 12). Magnus understryker dock att det inte i praktiken alltid fungerar så utan det är här som människan kommer in i bilden igen och agerar på de insikter som AI skapat, enligt Magnus erfarenheter (MP, rad 12). Både Informant 1 och Jim Samuelsson underströk dock att AI är något som är svårt att definiera och att det troligtvis finns bättre definitioner att hitta. Jim menade även på att det är något som går att utveckla något otroligt (INF1, rad 6; JS, rad 10).

Hur etik definieras svarade Informant 1 med beskrivningen *"Doing one's best to treat sentient beings as they would like to be treated"* (INF1, rad 8). Informant 1 påpekar även att mycket av hens personliga ängslan kring etik kommer från de nyheter som rapporteras om i media (INF1, rad 16). Jim Samuelsson beskrev etik som följande *"... det är någon sorts rekommendation att man beter sig så att man kan se sig själv i ögonen …"* (JS, rad 14). Magnus Perman beskrev att det är viktigt att anta ett mottagarperspektiv i förhållande till etik. Han menar på att när man skapar något är det viktigt att tänka sig in i hur mottagaren skulle uppfatta det som skapats. Det är viktigt att man hela tiden är medveten om att till exempel kunder har olika tankesätt som måste tas hänsyn till och att ett etiskt perspektiv därför blir viktigt att inta (MP, rad 36).

När det gäller förhållandet mellan AI och etik är intervjupersonerna eniga om att det är viktigt att värna om. Jim och Magnus är överens om att förhållandet mellan AI och etik är mest relevant och centralt när det gäller känsliga personuppgifter och Magnus påpekar att detta framför allt förekommer i branscher där AI tar livsavgörande beslut (JS, rad 16; MP, rad 16). Informant 1 tycker att det är av yttersta vikt att förhållandet fortsätter existera och menar på att med tiden kommer AI-etik bara bli viktigare när AI-lösningar tar fler och fler avgörande beslut (INF1, rad 10). Informant 1 påpekar även att etikfrågor är högst knutna till vilka

resurser som finns inom en organisation, då hen menar att detta är något som huvudsakligen större organisationer tar hänsyn till i hens erfarenhet (INF1, rad 12). Jim pekar dock på att förhållandet är konfliktfyllt och att man bör följa försiktighetsprinciper i sitt AI-arbete medan Magnus understryker vikten av transparens och beskriver att detta skapas genom att en medvetenhet hos kunden skapas om hur systemet fungerar samt vilken data som sparas (JS, rad 16; MP, rad 14).

4.2 Etablerade riktlinjer

Rent generellt kände ingen av intervjupersonerna till några etiska riktlinjer kopplade till AI (INF1, rad 20; JS, rad 22; MP, rad 22). Däremot lyfte alla tre att de var medvetna om riktlinjer, uppförandekoder och lagstiftning som kunde kopplas till etik men inte explicit till AI. I Informant 1:s fall beskrev hen hur arbetsplatsen hen arbetar på nu har en intern uppförandekod och att det finns en anonym supportlinje anställda kan rikta sig till med frågor som berör etik (INF1, rad 22). Jim Samuelsson har varit delaktig i ett projekt som haft riktlinjer men dessa kan han inte erinra sig om vid intervjutillfället. Däremot påpekar han om att han är säker på att det finns riktlinjer lokalt i Sverige och på nationell nivå (JS, rad 22). Han poängterar även att han följer flera olika riktlinjer och lagar kring datahantering med känsliga uppgifter. Till exempel att alla personuppgifter de har hand om måste vara krypterade och inte får skickas till andra myndigheter. De riktlinjer Jim följer är inte specifikt kopplade till etik men Jim menar på att lagstiftning ofta har sitt ursprung i etiska överväganden (JS, rad 30). Magnus lyfter GDPR men gör däremot understrykningen om att de inte specifikt är kopplade till etik utan handlar om lagar kring transparens och vad kunderna har för rättigheter (MP, rad 22).

Alla tre menar att det är viktigt att följa riktlinjer (INF1, rad 24; JS, rad 28; MP, rad 26). Informant 1 understryker att ur ett helt kapitalistiskt synsätt är riktlinjerna endast relevanta för att inte skada organisationens rykte (INF1, rad 24). Ytterligare något Informant 1 tar upp är att företag inom IT-branschen är väldigt resultatdrivna och att det är en underförstådd överenskommelse att allt annat är sekundärt, inklusive etik (INF1, rad 44). Jim och Magnus är överens om att riktlinjerna är högst centrala att följa när det gäller känsliga persondata och Magnus påpekar att det är viktigt att försöka sätta sig i mottagarens sits samt tänka en extra gång vad som skulle kunna hända om persondata hamnar i fel händer (JS, rad 18; MP, rad 26). Både Jim och Magnus känner till personer inom branschen som arbetar med AI-etik och etiska riktlinjer men inte någon de känner personligen (JS, rad 43; MP, rad 38). Informant 1 understryker att den enda hen känner till som jobbat uteslutande med AI-etik var en kvinna som arbetade på Google (INF1, rad 26).

4.3 Respons till etablerade riktlinjer

Varken Informant 1, Jim eller Magnus kände till mikroetik (INF 1, rad 31; JS, rad 47; MP, rad 42). Dock på tal om mikroetik tog Magnus upp *Explainable AI*, något som han jobbat en hel del med. Han beskriver det som att man ska kunna förklara varför modeller och algoritmer agerar så som de gör och att man ska kunna följa de olika stegen som tagits. Man ska kunna förstå varför de beslut som gjorts har tagits. Detta är något som enligt Magnus utvecklats och

implementerats på senare tid för att förhindra att AI blir *black-boxat*, som det har en tendens att bli. *Black-boxat* innebär att mycket av den information som kommer in i AI-algoritmen försvinner och det är inte alltid enkelt att se varför ett visst beslut har tagits av algoritmen. Magnus menar att detta både blir användbart för utvecklarna samt för mottagarna av systemet (MP, rad 45 & 47).

Vidare kände intervjupersonerna inte heller till några tekniska instruktioner inom branschen för hur man ska implementera etisk AI (INF1, rad 34; JS, rad 51; MP, 55 rad). Däremot lyfte Jim Samuelsson ett globalt projekt med UNICEF han var delaktig i vid namn *Barn och AI*. Inom detta projekt hade Lunds kommun reviderat deras dåvarande riktlinjer till att bli mer fokuserade på data, AI och etik i förhållande till projektet. I dessa riktlinjer menade Jim att det även fanns tekniska instruktioner för hur de ska implementeras och hur Lunds kommun skulle hantera data i samband med AI (JS, rad 53). Informant 1 hade dock reflekterat över tekniska instruktioner och minns en konferens hen deltog i där någon utvecklat en datamodell där man kunde specificera vilka begränsningar modellen skulle ha innan den togs i bruk. Informant 1 påpekade vidare att man skulle kunna använda denna metodik på stora förintränade (pre-trained) datamodeller där det finns mycket bias. Man skulle på så sätt kunna ta bort mycket av den existerande biasen (INF1, rad 34).

De riktlinjer som intervjupersonerna kommit i kontakt med hade nästintill alltid en hög abstraktionsnivå, något som de menade på kan bli problematiskt vid implementering (JS, rad 56; MP, rad 30). De riktlinjer som Jim tagit del av kretsar kring data och personuppgifter. Han menar att de är väldigt abstrakta och svåra att applicera i praktiken då till exempel begreppet personuppgift är ett vagt begrepp. Han påpekar även att samma problematik gällande abstraktionsnivån med riktlinjer finns vid lagar och regler kring data. I Jims fall gäller samma regler för personuppgifter även om datan är krypterad och han finner det problematiskt att lagen inte särskiljer på dessa. Jim ställer frågan om krypterad data bör falla inom ramen för känslig uppgift som det i nuläget gör. Jim menar på att om man inte ser namn, adress eller personnummer borde han kunna hantera data på ett friare sätt. Lagstiftningen som den ser ut idag kan vara hindrande i deras arbete då den data de har tillgång till är begränsad till deras egen förvaltning vilket innebär att de inte kan ta del av andra förvaltningars data. De får inte heller inte skicka data mellan de olika myndigheterna vilket vidare problematiserar området säger Jim. Genom att kombinera data från olika myndigheter såsom Jims avdelning och skolförvaltningen skulle man enligt honom tidigare kunna fånga upp barn som far illa. Till exempel om någon signal från ett barn i skolan skulle kunna kombineras med orosanmälan från socialförvaltningen, hade det gett stora fördelar. Enligt Jim lever dessa nu i två separata världar. Jim har till och med varit i kontakt med intergritetsskyddsmyndigheten i Stockholm för att se om det finns klara direktiv kring personuppgifter och hantering av krypterad data, det vill säga vad du får och inte får göra. Den respons som Jim fick från integritetsskyddsmyndigheten var att varje fall var unikt och att därefter måste behandlas så. Problematiken är trots att data och datorer funnits länge, är det ett nytt fenomen och arbetsområde att knyta ihop data från flera olika instanser menar Jim (JS, rad 30 & 32). Magnus som arbetat mycket med GDPR pekar på att det inte finns en hög abstraktionsnivå till just lagstiftningen, men menar på att han förstår problematiken kring abstraktion i hans arbete genom följande: *“man försöker ju alltid få ner helt konkret vad man får göra med data som en analytiker”* (MP, rad 30).

När det gäller standardiserade dataset inom en viss bransch med tydliga innehållsförteckningar och tydliga intentioner höll samtliga intervjupersoner med om att detta hade varit användbart (INF1, rad 36; JS, rad 87; MP, rad 59). Informant 1 påpekar att det finns många databanker med dataset ute på nätet som är open-source men att hen aldrig sett

någon databank som specificerar intentionen bakom datan och hur insamlingen av datan gick till (INF1, rad 36). Jim understryker problematiken kring att göra dataseten generella nog för en bred användning samtidigt som de ska vara specifika nog för en uppgift. Däremot menar Jim på att man skulle kunna tänka sig ett scenario där till exempelvis Malmö Stad varje gång de samlar ihop data för att göra ett dataset för inträning måste vara beskrivna på olika sätt. Datan skulle i så fall vara beskriven bland annat för hur den är insamlad, datamängd, intention bakom, vad den får användas till samt att det är klart och tydligt var den kommer ifrån. Jim understryker att man tappar mycket av den information som kommer in i en AI- algoritm, då AI ofta har en tendens att vara *black-boxat*. Då kan det vara väldigt bra att ha tydliga beskrivningar på den ursprungliga datan och dessutom en kontaktperson till den som samlat och kategoriserat datan (JS, rad 60). Magnus lyfter kring standardiserade dataset att yttre situationer, till exempel Covid-19 pandemin, kan komma att påverka köpmönster och att man då måste vara snabb på att revidera alternativt skapa nya dataset. Ytterligare något som Magnus lyfter är att det inte alltid är så att företag vill dela med sig av sin kunddata till konkurrerande företag och att det då kan bli problematiskt med standardiserade dataset inom just hans bransch, alltså detaljhandel. Dock menar Magnus på att standardiserade dataset potentiellt skulle kunna vara väldigt användbart inom andra branscher som inte har lika organisationsspecifik data (MP, rad 59 & 63).

På påståendet om att många företag endast använder etik i marknadsföringssyfte var ingen av intervjupersonerna särskilt förvånade över att så var fallet utan det var något även de till viss del upplevt. Informant 1 menar på att han sett etik användas i marknadsföringssyfte inom företaget han just nu befinner sig på, dock har han inte jobbat där länge nog för att se om det är något som levs upp till i praktiken. Informant 1 påpekar dock på att etik finns på tapeten inom företaget och diskuteras en hel del (INF1, rad 42). Jim menar på att han känner igen att AI-etik ofta används i marknadsföringssyfte. Jim tror att etisk AI inte praktiskt sett kommer appliceras av systemutvecklare i AI-lösningar förrän det får konsumentmakt, det vill säga att kunder faktiskt efterfrågar etisk AI (JS, rad 75). När det gäller vem som ska ha etikansvar inom en verksamhet tycker Magnus att det ska falla på de som sitter och utvecklar koden, det vill säga de som implementerar tekniken. Han menar på att de som jobbar med utvecklingen ska vara medvetna om etiken, att de använder data på rätt sätt och att de har stenkoll på vad man får göra med den (MP, rad 85). Jim håller med Magnus och säger på att det hade varit konstigt att ha en chef över etik eller någon sorts etikförvaltning utan det bästa är att låta ansvaret ligga på personnivå, så att det kan diskuteras kollegor emellan (JS, rad 77 & 79).

Ingen av informanterna hade fått någon etikutbildning inom deras universitetsutbildningar (INF1, rad 38; JS, rad 66; MP, rad 65). Dock hade Informant 1 fått rätt omfattande etikutbildning på sin nuvarande arbetsplats (INF1, rad 38). Jim som varken fått etikutbildning på universitetet eller på sin arbetsplats upplever att någon utbildning inte skulle behövas då de har en daglig levande diskussion mellan kollegor, angående vad de får använda och vad de kan eller inte kan göra (JS, rad 62 & 64). Jim känner att en försiktighetsprincip kopplat till persondata och etik är något som genomsyrar hans avdelning (JS, rad 16). Magnus har precis som Jim varken fått etikutbildning på universitetet eller hans arbetsplats men menar på att han nu är till viss del självlärd (MP, rad 65 & 72). Magnus sitter dessutom med i ledningsgruppen för en data-scientistutbildning på en yrkeshögskola som finns i Helsingborg och Malmö. Där har det nu på senare tid blivit aktuellt att lägga in kurser kopplat till etik och moral i kursplanen för just data-scientistutbildningen (MP, rad 65). Något som Jim tar upp kopplat till just utbildning är vikten av att försöka undvika bias. Jim menar på att det hade varit användbart om alla som arbetar med data och AI hade fått någon sorts utbildning just kring databias. Jim understryker att genom att inte koda in fördomar blir en algoritm bättre, inte bara ur en etisk synpunkt, utan även en teknisk vilket kanske kan tilltala en utvecklare mer

(JS, rad 73). Informant 1 understryker även vikten av att vara medveten om vilken sorts bias din modell har innan den praktiskt sett tas i bruk (INF1, rad 30). Magnus påpekar även kring just databias att det är otroligt viktigt att träna upp sina modeller på ett så stort dataset som möjligt som i största möjliga utsträckning representerar en hel population (MP, rad 22).

4.4 Implementering av mikroetik

Både Jim och Magnus känner till projekt som har fått finansiellt stöd av Vinnova men har själva har inte varit involverade i dessa. Jim menar dock på att stödet inte direkt varit kopplat till deras AI-arbete utan till Malmö stads digitaliserings- och dataarbete (JS, rad 81). Magnus har ett svagt minne av att de har haft ett innovationsprojekt som fått finansiellt stöd men att det mer handlade om grön IT (MP, rad 89).

På frågan om vad de anser om en white/black list var det något som Jim inte kunde uttala sig om då de officiellt inte påbörjat sitt AI-arbete än, utan menar på att det är något som man kan förhålla sig till när man har något konkret att arbeta med (JS, rad 85). Informant 1 har inte hört talas om att de inom företaget har någon white/black list men tror att det är något som skulle kunna finnas på kundsidan på framtida projekt för att kunna kontrollera en slutprodukt (INF1, rad 52). Magnus tycker att det låter som en bra idé. Han utvecklar att han personligen inte är delaktig i den nivån av utveckling där en aspekt av begränsning är nödvändig. Däremot lyfter han att en människa mer eller mindre alltid är delaktig i mitten och att inget således blir 100 % självkörande (MP, rad 93). Något som han tar upp vid denna fråga är *OpenAI* där det finns så kallade avancerade transformationsmodeller där man själv kan styra modellen till att ha ett visst beteende, till exempel vara snäll eller sarkastisk. Han fortsätter med att sarkasm inte är något man vill ha just i chattbottar i en kundsupport. Då menar han på att det kan vara en bra idé att en white lista som avgränsar modellen till att enbart ha vissa beteenden (MP, rad 95).

När det gäller förslaget att bädda in etisk reflektion i det dagliga arbetet var informanterna positiva till detta (INF1, rad 54; JS, rad 87; MP, rad 97). Informant 1 påpekar att det är av yttersta vikt att upprätthålla en form av medvetenhet till etiken och menar på att daglig diskussion om etisk reflektion är särskilt viktig i situationer där AI tar avgörande beslut (INF1, rad 10 & 54). Jim ser även detta som en bra idé men understryker att det är något som kommer behöva träning och mest sannolikt ske på teamnivå. Detta främst eftersom innebörden av etisk reflektion är något som man behöver diskutera. Han menar att man då behöver tala med varandra och till och med granska varandras arbete. Jim understryker att de redan i dagsläget på arbetsplatsen till viss del bäddar in etisk reflektion, med tanke på hur pass känslig den data som de behandlar är, är det omöjligt att inte ha en daglig etisk reflektion. Dock tycker Jim att reglerna kring känslig persondata är mer hindrande än bra då de inte riktigt kan göra vad de vill, men diskussionen kring det etiska perspektivet kopplat till data är ständig (JS, rad 87). Magnus tycker att det verkar som en vettig idé att bädda in etik i en systemutvecklares dagliga arbete. Han lyfter att det är på utvecklarnas nivå som vissa av besluten kommer att tas huruvida modeller ska “tweakas” och ställas in och hur man hanterar data. Han pekar ytterligare på att det är viktigt att etik och moral är något som genomsyrar hela organisationen. Ytterligare något som Magnus lyfter är att han tror att man som utvecklare inte är medveten om den makt man besitter om man utvecklat en modell. Han menar på att det då hade varit bra att man som utvecklare gått en kurs i etik så att man blir medveten om vad det kan få för konsekvenser om man använder data på fel sätt (MP, rad 97 & 99).

Både Jim och Magnus är positiva till tanken på att sjösätta ett team för testning och validering (JS, rad 109 & 111; MP, rad 101). Jim menar dock på att man bör se till vilket område som systemet finner sig i, till exempel om systemet används inom medicinsk utrustning på en intensivvårdsavdelning. Då är det ett system som tar livsavgörande beslut och pekar således på att det kan vara väldigt aktuellt med ett sådant team (JS, rad 111). Magnus understryker att det är ett väldigt rimligt förslag och som borde appliceras på AI-lösningar för att hitta svagheter och kryphål. Detta på samma sätt som inom cybersäkerhet. Inom detta berättar Magnus att det finns så kallade "pen-tester" för att försöka penetrera lösningen. Magnus är övertygad om att det bara är bra att stressa AI-modeller och försöka se om det finns ett sätt att lura dem. Slutligen säger Magnus att det finns inget att förlora på att sjösätta ett sådant team (MP, rad 101).

När Jim fick frågan om hans tankar om ett "trust comparison index" säger Jim att det är en god tanke men inte så lätt att genomföra. Hans initiala tanke var på vem som egentligen ska sätta denna stämpel. Jim understryker på att det troligtvis borde vara en stämpel på extern eller nationell nivå men nämner även att faktorer som kostnad kan påverka utfallet av stämpeln. Jim nämner att något sådant hade kunnat göras av integritetskyddsmyndigheten eller någon annan myndighet på hög nivå. Han påpekar även att det hade varit behjälpligt i Malmö stads arbete att ha en myndighet som helt enkelt utvecklat riktlinjer med krav som måste uppfyllas. Detta bör riktas speciellt för de som arbetar med AI inom den offentliga förvaltningen men menar på att ett "trust comparison index" kan vara att dra det lite långt (JS, rad 93–107).

5 Diskussion

I detta kapitel sammanvävs teorin och litteraturen tillsammans med den empiri som presenterats. Syftet med detta kapitel är att med ett analytiskt förhållningssätt diskutera resultatet med utgångspunkt i den litteratur som lagts fram.

5.1 Definitioner

I alla tre intervjupersonernas personliga definition av AI finns det element som överensstämmer med den presenterade definition som HLEG (2019) lagt fram. Informant 1 lyfter att det är ett program som lär sig av data för att kunna göra förutsägelser och analyser. Hens definition är intressant nog den enda som berör data och ligger således i linje med HLEG:s beskrivning som bland annat lyfter just datainsamlingen och tolkningen av datan. Just datainsamling eller datautvinning, som är extra relevant i datadriven AI, är det som ligger till grund för de förutsägelser som AI-lösningen ska göra, samt det som lösningen ska lära sig av (Žliobaitė & Custers, 2016). Aspekten av självlärande är något som både Jim och Magnus nämner i sina definitioner och detta element beskrivs även i HLEG (2019). Både Informant 1 och Jim underströk dock att AI är något som är svårt att definiera och de menar på att det säkerligen finns bättre definitioner att hitta. Jim påpekade även att det går att utveckla definitionen av AI oändligt. Att AI är något svårdefinierat stämmer överens och ligger i linje med det som Dignum (2019) påpekar.

Det blev tydligt att mötet mellan människor hamnar i centrum vid frågan om definitionen av etik. Informant 1 menar att det handlar om att göra sitt allra bästa för att behandla andra människor som de själv vill bli behandlade. Jims definition handlar om att man ska behandla andra människor så att man själv kan se sig i ögonen. Magnus lyfter just mottagarperspektivet och menar på att det är viktigt att förstå hur mottagaren uppfattar det som skapats. Med utgångspunkt i detta blir det tydligt att den tredje etikinriktningen som lyfts i litteraturen, dygdetik, är den som överensstämmer mest och ligger i linje med hur alla tre intervjupersoner ser på etik. Dygdetik framhäver vikten av att utveckla goda karaktärsdrag och att personer som lever efter denna typ av etik automatiskt utför goda eller rätt handlingar (Stanford Encyclopedia of Philosophy, 2016c). Något som både Jim och Magnus var överens om var att etikansvar inom en organisation ska ligga på medarbetarnivå, vilket går att uppnå genom att applicera dygdetik genom diskussion mellan medarbetare. Detta för att få individer att få upp ögonen för vad som anses vara goda och rätta handlingar att utföra, så att handlingarna så småningom sker undermedvetet och automatiskt.

Något som intervjupersonerna lyfte kring förhållandet mellan AI och etik, är att ju känsligare uppgifter och ju mer livsavgörande beslut som AI-systemet ska ta, desto större är behovet av att uppmärksamma förhållandet. Dessutom bör relationen fortsätta att existera då AI-lösningarna med tiden kommer ta större plats i våra liv. Detta är något som ligger i linje med det Dignum (2019) förutspår, alltså att AI-lösningarna som utvecklas kommer ta beslut, stora som små, som kommer påverka människan. Dignum (2019) påpekar även att dagens AI-system saknar förmågan att ta beslut av etisk natur.

Som går att utröna från diskussionen ovan är både AI och etik svårdefinierbart. AI är dessutom, som tidigare nämnts, ett paraplybegrepp som täcker en rad olika områden och har

många olika användningsområden (Watson, 2018). Användningen av AI ser också olika ut i olika branscher vilket vidare problematiserar möjligheten att definiera ämnet. Det är med det sagt inte svårt att se varför det finns många olika definitioner av AI. De olika definitionerna av AI understryker även svårigheten att precisera regler och riktlinjer kring AI och därefter förhållandet till etik. Vilket leder till nästa punkt, etablerade riktlinjer.

5.2 Etablerade riktlinjer

De etablerade riktlinjerna undersökta i studien kommer från ACM och Europakommissionen. Ingen av dessa kom på tal under intervjuerna och intervjupersonerna kände inte heller till några andra etiska riktlinjer explicit kopplade till AI. Varför ingen hade hört talas om ACM:s riktlinjer kan vara för att de är frivilliga att följa (Dignum, 2019). Bristen på medvetenhet om just dessa kan vara för att de är tänkta att inspirera och guida (ACM, 2018b) och att de således inte har lika stor tyngd som lagstiftning. Exempelvis var GDPR något som alla intervjupersoner kände till och nämnde under sina intervjuer. Denna medvetenhet om lagstiftningen kan förklaras genom att dra en parallell till det som Bezuidenhout och Ratti (2021) lyfter. De menar på att regler är precisa och ordentliga och inte kan vägas mot varandra. Man måste alltså följa dem och en regel kan inte förbises till skillnad från riktlinjer. Medvetenheten kan även förklaras med att avvikelser från att följa dataskyddsförordningen leder till konsekvenser såsom sanktioner (Integritetsskyddsmyndigheten, 2021a). Jim menar ytterligare på att hans dagliga arbete påverkas mycket av den lagstiftning som finns, ibland till den grad att det blir hindrande. Eftersom lagstiftningen som redan finns på plats hindrar det praktiska arbetet, menar Jim att det går att förstå varför det kan finnas ett motstånd eller okunskap på professionell nivå till icke obligatoriska riktlinjer som inte har några rättsliga konsekvenser. Jim menar dock på att mycket av den lagstiftning som finns på plats ofta har sitt ursprung i etiska överväganden, vilket kan vara en orsak till varför etablerade etiska riktlinjer hindras från att ta fäste, då det etiska perspektivet redan till synes täcks av lagstiftningen. Alla tre intervjupersoner var dock eniga om att etiska riktlinjer är viktiga att följa. Informant 1 pekar på att ur ett helt kapitalistiskt synsätt är riktlinjers enda funktion att inte skada en organisations rykte mot den breda massan.

Med utgångspunkt i att utforska medvetenheten om etik inom branschen i förhållande till riktlinjer, tillfrågades intervjupersonerna om de kände någon som arbetade explicit med AI-etik. Informant 1 sa att den enda personen han kände till som jobbat explicit med AI-etik är en kvinna som har jobbat på Google. Detta blir intressant i relation till det som Informant 1 tidigare sagt om att det enligt hen krävs stora resurser för att ta hänsyn till etikfrågor. Hen menar på att i hens personliga erfarenhet är det bara större organisationer som satsat mer på etik medans det i mindre företag, bland annat start-ups, är något som helt förbisetts eller som inte har resurserna att ta hänsyn till det. Detta kan förklaras med att eftersom Google är ett av världens största dataföretag blir det nödvändigt att perspektivet av etik tas hänsyn till och att det läggs resurser på det. Google är ett exempel på den sortens företag som idag sysslar med AI-etik men det är nödvändigt att även mindre organisationer som applicerar och hanterar AI-lösningar ska kunna ta hänsyn till AI-etik. Detta då Informant 1 pekar på att AI-etik med tiden kommer bli mycket viktigare då AI kommer ta fler och fler livsavgörande beslut. För att kunna applicera AI-etik krävs riktlinjer och lagstiftning som ska kunna vara praktiskt implementerbar, alltså med andra ord efterfölja *mikroetik*. Detta diskuteras i nästa stycke.

5.3 Respons till etablerade riktlinjer

Det var inte särskilt förvånande att ingen av de intervjuade kände till mikroetik. Det är ett relativt nytt begrepp som i nuläget mest diskuteras bland forskare. Alla tre intervjupersoner menade på att de riktlinjer de kommit i kontakt med har hög abstraktionsnivå, det vill säga att de är svåra att applicera i praktiken. Då mikroetik handlar om att omvandla riktlinjer med hög abstraktionsnivå till mer tekniska instruktioner är detta något som absolut skulle kunna tänkas vara applicerbart. Jim lyfter att han möter dataskyddsförordningen dagligen i sitt arbete, speciellt eftersom den data de hanterar är känslig. Något som han dock pekar på är att även denna lagstiftning har en hög abstraktionsnivå. Jim lyfter ett tydligt exempel på detta då han understryker att i dataskyddsförordningen faller även krypterad persondata under samma kategori som känsliga uppgifter. Dessutom har Jim behövt ringa till Integritetsskyddsmyndigheten i Stockholm för att försöka klargöra lagstiftningens principer. Detta skiljer sig från vad Bezuidenhout och Ratti (2021) lyfter då de menar på att regler och lagar är precisa och ordentliga, vilket kan tolkas som enkla att följa i praktiken. Magnus är av samma åsikt som Bezuidenhout och Ratti (2021) och pekar på att i hans erfarenhet har dataskyddsförordning inte en hög abstraktionsnivå. En förklaring till varför det finns en diskrepans mellan åsikterna hos intervjupersonerna kan vara att de hanterar olika typer av data. Magnus hanterar egentligen inte känsliga uppgifter i sitt arbete utan han nämner att det känsligaste de behandlar är hemadresser. Jim däremot hanterar personuppgifter för personer i utsatta situationer, som måste krypteras för att ens kunna användas. Således kan man tolka det som att ju mer känslig data man hanterar desto högre abstraktionsnivå blir det på principerna. I vår erfarenhet och utifrån att ha läst lagstiftningen finns det en hel del likheter när det kommer till abstraktionsnivå mellan riktlinjerna och dataskyddsförordningens principer. Denna abstraktionsnivå skulle således kunna minskas genom att applicera mikroetik. Till exempel att mikroetik, i form av tekniska instruktioner, skulle kunna användas för att förklara hur dataskyddsförordningen ska appliceras i praktiken.

Förutom mikroetik fanns det även en annan respons på de etiska riktlinjerna. En av de punkter Hagendorff (2020) lyfte var att etiska riktlinjer i praktiken används just i marknadsföringssyfte och att dessa riktlinjer inte efterlevs i praktiken. När denna fråga ställdes till intervjupersonerna så upplevdes påståendet som negativt från deras sida, detta överensstämde även med vår ursprungliga tolkning. Magnus påpekade att om man inte lever som man lär i förhållande till marknadsföring kommer företaget slutligen ta skada när det kommer fram i ljuset. Dock finns det en annan sida av detta mynt, den positiva sidan av marknadsföringen. Informant 1 uttryckte att mycket av hens personliga oro kommer från nyheter och media samt att det är anledningen till att hen tänker på etikfrågor överhuvudtaget. Däremot är det också det som driver hens etiska medvetenhet framåt. Jim framlägger en hypotes kring just marknadsföring om att det finns mer makt i den än vad man kan tro. Om nyheter och media skulle börja uppmärksamma etik i förhållande till AI skulle detta således bidra till att kunder får upp ögonen för etik. De skulle då börja efterfråga etisk AI och på så sätt skulle de få en konsumentmakt och pressa leverantörer att ta ställning till de riktlinjer kring etisk AI som finns. En koppling kan här göras till i vilken utsträckning intervjupersonerna känner till etablerade riktlinjer. Då kännedomen av riktlinjerna inte är särskilt hög, kan det förklaras bland annat med att om etik endast pratas om i marknadsföringssyfte så är det inget som praktiskt diskuteras i en organisations dagliga arbete. Skenet om att organisationen är etisk uppfattas således endast från externa parter, medan någon aktiv diskussion inom organisationen inte sker. Följaktligen blir den interna medvetenheten om etik och i sin tur etiska riktlinjer låg. En annan anledning till varför organisationer kanske inte är så etiskt medvetna lyftes av Informant 1. Då hen menar på att

inom IT-branschen finns det ett så stort fokus på resultat att allt annat blir sekundärt, inklusive etik. Detta då etik inte är lika påtagligt och mätbart som andra resultat.

Tidigare lyftes mikroetik och för att kunna applicera detta och för att få medarbetare att bli mottagliga för det, krävs det etisk kompetens och ansvar. Detta är något som ACM:s riktlinjer (punkt 2, se Appendix A) tar upp och de lägger framför allt stor vikt på professionellt ansvar. Punkt 2.2 bland riktlinjerna pekar på att professionell kompetens bland medarbetare är det som ska driva etikarbetet framåt. Dock på tal om ACM:s riktlinjer var dessa ingenting som intervjupersonerna kände till eller hade hört talas om. Detta ligger i linje med det som McNamaras et al. (2018) undersökning kom fram till, det vill säga att dessa inte är allmänt kända eller används i praktiken överhuvudtaget. Dock gick det att utröna en del av innebörden av punkt 2.2 i intervjupersonernas resonemang. Både Jim och Magnus var överens om att etikansvar bör ligga på personlig nivå, det vill säga hos varje medarbetare. För att uppnå hög kompetens inom etik krävs det både intern och extern utbildning. Även om både Jim och Magnus var överens om att etikansvar bör ligga på medarbetarnivå, är det intressant att ingen av dem hade fått någon intern etikutbildning och någon etik fanns inte heller i deras akademiska utbildningar. Magnus är dock medveten om att det krävs förändring i akademisk utbildning i förhållande till etik. Han sitter med i ledningsgruppen för en data-scientistutbildning och där har det på senare tid blivit tydligt att etik och moral måste inkorporeras i kursplanen. Detta är något som överensstämmer med Floridis (ed. 2021) tankar om att institutionella förändringar måste göras genom att drastiskt ändra akademiska läroplaner till att främja det etiska perspektivet på universiteten världen över. Utöver akademisk utbildning finns det sätt att lära ut etik på en arbetsplats och ett sätt går att finna via Bezuidenhout och Ratti (2021) som lägger fram en modell som kan appliceras vid utbildning. Den syftar till att bädda in mikroetik i form mikrouppgifter i dagligt arbete för att öka etisk kompetens. Något som är värt att nämna är att Informant 1 har fått en rätt omfattande etikutbildning på sin arbetsplats, så det kan skilja sig arbetsplatser mellan. Däremot var Informant 1:s etikutbildning inte alls kopplad till det arbete med data och AI som hen utför. Det verkar med det sagt som att även Informant 1:s arbetsplats tror på att en generell kompetens inom etik ska genomsyra det arbete som hen gör inom AI. Något som kom på tal kring just etikutbildning i de intervjuer vi genomförde var databias, med andra ord de fördomar från människor som genom det sociotekniska utbytet med en dator nu finns i data. Jim påpekade att databias är ett av de största problemen kopplat till AI-arbete och understryker att alla som arbetar med data och AI borde få någon utbildning direkt kopplat till databias. Ntoutsi et al. (2019) understryker att i komplexa AI-system kan fördomarna till och med bli förstärkta. Det är med det sagt därför av yttersta vikt, som både Informant 1 och Magnus påpekar, att vara medveten om de bias ens modell har och att träna den med ett så stort dataset som möjligt och så gott det går representera en hel population.

Ett sätt att minska databias och samtidigt applicera mikroetik är att använda sig av standardiserade dataset med tydliga intentioner och innehållsförteckningar enligt Hagendorff (2020). Det fanns en positivitet från alla intervjupersoner till användningen av just detta. Informant 1 beskrev att hen kände till dataset i databanker men att hen aldrig tagit del av en databank där en specifikation över intention och insamling funnits med. Jim påpekar problematiken kring att göra generella dataset specifika nog för en bred användning. Han kan däremot se ett scenario där Malmö stad måste följa de krav som finns i förhållande till data. Det vill säga att dataseten kommer med tydliga beskrivningar över hur den är insamlad, vilken datamängd den innehåller, vad det är för intention bakom, att det är tydligt var datan kommer ifrån samt sist men inte minst en kontaktperson till den som skapade datasetet. Magnus lyfte även att yttre situationer såsom Covid-19 snabbt kan ändra köpmönster och det kan därför vara svårt med standardiserade dataset då de i sådana situationer snabbt måste revideras,

vilket inte alltid är särskilt lätt. Både Magnus och Jim lyfter också, vilket vi även lagt märke till i vår forskning, att AI har en tendens att bli *black-boxat.* Genom att använda sig av standardiserade dataset kan en professionell utövare ta ett mer informerat och rättvist beslut om vilket dataset den ska använda. Dessutom kan de beslut som tagits av algoritmen lättare kopplas tillbaka till den ursprungliga datan. Detta gör att den professionella utövarens arbetssätt blir mer transparent, något som både Hagendorff (2020 och Magnus underströk är det absolut viktigaste inom AI-etik. Krav på transparens är även något som HLEG (2019) listar som punkt 4 i deras lista av konkreta krav. Magnus tar även upp att ett vanligt sätt att tackla *black-boxad AI* är att applicera *Explainable AI (XAI).* Magnus beskriver att detta används för att förklara varför ett system agerat på ett visst sätt samt förstå vilka steg som tagits i processen som ledde till slutresultatet. Detta är något som HLEG (2019) tar upp som en teknisk implementationsmetod för att uppnå högre trovärdighet. Fler tekniska implementationsmetoder kommer i nästa punkt.

5.4 Implementering av mikroetik

Den metodik som Bezuidenhout och Ratti (2021) lägger fram om att bädda in etisk reflektion i en professionell utövares dagliga arbete är något som alla intervjupersoner var positivt inställda till. Bezuidenhout och Ratti (2021) menar att genom att använda mikroetik kommer man bland annat komma ifrån den höga abstraktionsnivån och ineffektivitet som råder hos generella riktlinjer och principer. De föreslår att detta ska implementeras med hjälp av mikrouppgifter som ska bli en del av utövarens dagliga repetitiva arbete för att framhäva dygdetik, som syftar till att framhäva goda karaktärsdrag. Jim och Informant 1 har tidigare lyft den höga abstraktionsnivån och således kan metodiken för mikroetik vara ett sätt att bemöta detta problem. Jim betonade att de dagligen har en levande diskussion kring etik och Informant 1 underströk att det är väldigt viktigt att det sker en daglig diskussion om etik. Detta gäller speciellt i situationer där AI-lösningen tar beslut som är livsavgörande, exempelvis i AI-system som ska styra medicinsk utrustning. Dessa två exempel visar på att en metodik för mikroetik skulle kunna appliceras relativt friktionsfritt.

För att stödja etiskt arbete understryker Floridi et al. (2018) att incitament kan ges till företag som vill utveckla mer trovärdig och transparent AI. Dessa incitament kan komma från nationella institutioner och vi valde att fråga om Vinnova som är Sveriges innovationsmyndighet. Bland intervjupersonerna var det ingen som hade fått stöd för deras AI-arbete direkt kopplat etik. Jim hade dock hört talas om tidigare projekt kopplade till data och digitalisering inom Malmö stad. Även Magnus hade hört talas om projekt som fått stöd men dessa var kopplade till grön IT. Det går att konstatera att finansiella incitament inom branschen har getts. Med det sagt går det inte att utesluta att ytterligare stöd kommer att ges i framtiden. Följaktligen hade det varit intressant att se om Vinnova gett stöd till andra organisationer för utveckling av etisk AI, då de tidigare projekt som lyfts i intervjuerna inte hade ett fokus på just etisk AI.

Metoden för testning och validering som HLEG (2019) presenterar om att sjösätta ett team för att försöka ta sönder ett AI-system, behövs eftersom författarna menar på att då AI-system är självlärande och beroende av kontext räcker inte traditionell testning till. Förslaget mottogs av Jim som en bra idé. Han menar även på att ett sådant förslag är väldigt kontextberoende, det vill säga vilket område systemet finns i. Skulle AI-systemet användas för något livsavgörande kan det absolut vara nödvändigt, men om så inte är fallet, anser Jim att det inte är lika väsentligt. Magnus håller med Jim i att det är bra att stressa AI-modeller för att hitta sätt att

lura dem. Han drar en parallell till cybersäkerhet där så kallade “pen-tester” ofta förekommer. I många fall kan det vara en kostnadsfråga. Om det finns resurser inom företaget för att tillsätta ett sådant team leder detta till att det är positivt både i aspekten av säkerhet samt även i förhållande till etik. HLEG (2019) påpekar även att teamet kan fungera som en slutlig kontroll, då de kan analysera slutprodukter från AI-lösningen för att se att den uppfyller tidigare bestämda riktlinjer. Ännu en implementationsmetod för att förverkliga mikrotetik kommer även den från HLEG (2019), i form av en *white list* respektive *black list.* Det vill säga listor som förklarar systemets beteende, situationer det får befinna sig i och vilka regler det alltid ska följa, respektive en lista med motsatser. Vid denna fråga leder Magnus in samtalet på *OpenAI* och berättar att det finns så kallade transformationsmodeller där man själv kan styra beteendet. Han exemplifierar detta genom att ta upp chattbottar där man kan styra vilket beteende chattbotten ska ha och här menar han på att till exempel beteende som sarkasm skulle kunna vara bra att svartlista.

Något som Informant 1 tog upp angående hur man ska kunna få bort databias är att försöka bestämma vilka begränsningar AI-lösningen ska ha innan den tas i bruk. Detta är något som ligger i linje med HLEG:s (2019) andra implementationsmetod *x-by-design.* Denna syftar till att normer som systemet ska följa ska identifieras innan systemet börjar användas för att undvika negativa konsekvenser, i detta fall databias. Även Jim lyfte idén om att ställa in systemet i förväg och begränsa det innan det används, då han pratade om hur Malmö stad i ett tänkbart scenario skulle kunna förändra sin metodik vid uppträning av datamodeller. I Jims fall handlade det då om att ha koll på var datan kom ifrån, vilka intentioner den hade och vem som samlat in den. Om datan är ordentligt genomtänkt från grunden, eller så kallat *by design,* kommer detta leda till att förekomsten av bias i systemet minskar. Trots att Jim och Informant 1 pratar om skilda områden, är det dock viktigt att lyfta den grundtanke som blir synlig hos de båda. Både Jim och Informant 1 lyfter idéen om en design som gör begränsningar eller anger vissa förinställningar innan designen tas i bruk. Detta överensstämmer med *by-design*-koncept som är vanligt inom IT-branschen (European Commission 2019). Denna tankegång hos de båda intervjuade visar således på att det kan finnas utrymme för en acceptans av ett tillvägagångssätt som kretsar kring ta etik i åtanke tidigt skede.

Som nämnts i tidigare punkt, tar Jim upp konsumentmakt och att företag som utvecklar AI inte kommer få upp ögonen för etisk AI förrän det finns konsumentmakt. Ytterligare ett sätt som skulle kunna driva på konsumentmakten är något som Floridi et al. (2018) tar upp i punkterna *Stödja* och *Utveckla.* Den första punkten syftar till att skapa certifieringar kopplat till etisk AI för att få kunder att inse värdet av det. När certifieringarna potentiellt skulle etableras skulle kunder sedan kunna kräva detta av leverantörer av AI-system. Den andra punkten, *utveckla,* avser att skapa nyckeltal för att öka samhällets förståelse för samhällsförmånlig AI. Nyckeltal kopplat till etik och samhällsförmånlig AI är något som HLEG (2019) också tar upp som ett sätt att implementera etik. Angående detta påpekar Jim dock att det kan vara problematiskt och svårt att genomföra, då dessa nyckeltal måste skapas av någon på myndighetsnivå. Vi håller med och tänker att Integritetsskyddsmyndigheten som i nuläget styr överseendet av dataskyddsförordningen i Sverige skulle vara en passande aktör, både i fallet av att skapa certifieringar och nyckeltal. Framför allt då de arbetar nära med frågorna om känslig datahantering. De bör dessutom samarbeta med någon organisation med hög kompetens inom AI, exempelvis Vinnova, för att genomföra arbetet med certifieringar och nyckeltal.

6 Slutsats

Detta kapitel ämnar sig till att lyfta slutsatser för att besvara forskningsfrågan. Den forskningsfråga studien grundar sig på är följande: I vilken utsträckning är verksamheter som arbetar med datadriven AI på den svenska marknaden medvetna om etik och mikroetik, samt vet de hur det kan implementeras?

6.1 Slutsatser

Det vi kan konkludera utifrån vår undersökning är att det finns en viss medvetenhet om etik hos de personer vi har intervjuat. Med viss medvetenhet menas att etik övergripande var något som alla intervjupersoner hade i åtanke och något som de uttryckte var viktigt i förhållande till AI. Det påpekades även att det framtida behovet av etik inom AI kommer få en större betydelse då fler livsavgörande beslut kommer tas av datadrivna AI-system. Apropå det framtida behovet av etik framgick det i studien att den medvetenhet, och i den mån de arbetar med AI-etik inom organisationen, kan vara kopplat till de resurser som de har att tillgå. Ju större organisation, desto mer utrymme att jobba med etik.

I undersökningen kopplades etik ihop med etiska riktlinjer, och dessa riktlinjer hade intervjupersonerna inte någon större medvetenhet om. De hade endast tagit del av några få riktlinjer, men ingen av de etablerade riktlinjer vi presenterat i denna studie. En anledning till att intervjupersonerna inte hade hört talas om riktlinjerna kan vara, som Dignum (2019) påpekar, att det inte finns något tvång att följa dem. De är helt enkelt till för att inspirera och guida arbetet med etik, något som i praktiken blir för abstrakt. Samtliga intervjupersoners åsikter låg i linje med Dignums (2019) då de underströk att de riktlinjer de kommit i kontakt med hade hög abstraktionsnivå. Något som framkom i studien är att AI är ett paraplybegrepp och att det är svårdefinierat, likaså etik. Detta understryker svårigheten i att precisera regler och riktlinjer kring förhållandet däremellan. Ytterligare något som skulle kunna förklara den låga medvetenheten, är att IT-branschen är väldigt resultatdriven, således blir allt annat sekundärt.

Till skillnad från den låga medvetenheten om riktlinjer nämnde samtliga intervjupersoner dataskyddsförordningen, vilket påvisar en högre medvetenhet om förordningen än riktlinjerna. Vid första anblick kan riktlinjerna och förordningen ses som vitt skilda men det påpekades att lagstiftning ofta har sitt ursprung i etiska överväganden. En slutsats som kan dras från detta är att etiska riktlinjer hindras från att ta fäste då de till synes redan täcks av tvingande lagstiftning. På så sätt kanske etiska riktlinjer bör avvecklas och i stället bör huvudfokus för etiker ligga i att koppla riktlinjerna till, exempelvis dataskyddsförordningen. Något som dessutom framgick i undersökningen är att i praktiken kan dataskyddsförordningen även ha samma abstraktionsnivå som etablerade riktlinjer. Detta kan variera beroende på hur pass känslig data en organisation behandlar. Med det sagt skulle mikroetik kunna gå att applicera även på dataskyddsförordningen ute hos verksamheter.

Det påvisades i undersökningen, apropå mikroetik, ingen medvetenhet om begreppet men delar av innebörden lyftes av intervjupersonerna. Det går att utröna i undersökningen att *black-boxed* AI är ett vanligt problem i branschen. Något som teoretiskt sett går att bemöta med mikroetik. I praktiken däremot, där begreppet mikroetik inte kändes till, fanns det andra responser till problemet, nämligen *Explainable AI* och *OpenAI*.

I studien framkom det att för att kunna implementera mikroetik krävs det etisk kompetens och ansvar. Ett ansvar som bör ligga på en personlig nivå, enligt de intervjuade. Om man har personal som är kompetent inom etik, kommer etikutbildning ske automatiskt genom daglig diskussion. Detta leder till att individer får upp ögonen för vad som anses vara goda och rätta handlingar att utföra, så att handlingarna så småningom sker undermedvetet. Grunden för kompetensutveckling inom etik ska helst läggas på universitetsnivå, det vill säga tas hänsyn till så tidigt som möjligt. Ansvaret ska sedan inkorporeras i det dagliga arbetet genom mikrouppgifter. Kompetensen på personlig nivå är också det som främst kan förebygga databias i AI-system, genom ett tankesätt av dygdetik.

En implementationsmetod som är högst relevant inom datadriven AI för att minska databias och applicera mikroetik, är att använda sig av standardiserade dataset som specificerar intention, insamlingsmetod, datamängd och har en tydlig innehållsförteckning. Detta var något som intervjupersonerna hade kommit i kontakt med, var relativt medvetna om och resonerade utförligt om. Likaså implementationsmetoderna såsom sjösättning av ett testningsteam samt listor med regler och restriktioner. Ytterligare i undersökningen påvisades det, att det i praktiken är förmånligt att redan i tidigt skede ta etik i åtanke. Detta exemplifieras genom ett förhållandesätt som redan existerar i branschen, nämligen *x-by-design* konceptet. Detta är ett koncept som är ett av de bästa sätten för att kunna arbeta etiskt och förebygga problem såsom databias.

Avslutningsvis kan det konstateras utifrån vår studie att konsumentmakten kan komma att styra hur utvecklingen av etisk AI ser ut. Denna konsumentmakt kan bli i högsta grad påverkad av marknadsföring, nyheter och media. Licensieringar och nyckeltal lyfts för att öka konsumentmakten och därmed också samhällets förståelse för samhällsförmånlig AI. Det framkom i undersökningen att det finns en viss problematik kring nyckeltalen och licensieringar då de måste implementeras av en objektiv part på hög nivå. Förslagsvis skulle intergritetssskyddsmyndigheten i samarbete med Vinnova kunna skapa och utveckla dessa två. Ytterligare bör nationella institutioner som vill främja utvecklingen av etisk AI ge finansiella incitament till organisationer för att stödja deras utveckling av etisk och transparent AI.

6.2 Vidare forskning

Det finns ett tydligt behov av att etik i förhållande till AI bör uppmärksammas och tas hänsyn till, då AI kommer ta fler avgörande beslut som kommer påverka människor i framtiden. Då mikroetik är ett nytt begrepp och behovet av etik kommer växa ser vi ett stort behov av ytterligare forskning om ämnet i framtiden. Då denna undersökning endast gjordes hos tre organisationer bör liknande studier med större omfång ge ytterligare insikter. Dock kan studien agera som underlag till vidare forskning inom samma område.

Då det blev tydligt i studien att det fanns en låg medvetenhet om riktlinjer kan det vara av intresse att vidare forska kring hur organisationer bör förhålla sig till dessa. Bör man jobba mot att en större medvetenhet skapas eller bör riktlinjerna integreras i lagstiftning?

Appendix A

ACM:s uppförandekod

1. *Allmänna etiska principer*

   1.1 *Bidra till samhället och till mänskligt välbefinnande, genom att erkänna att alla människor är intressenter i datoranvändning*

   - Datavetare bör använda sin kompetens för att gynna samhället, dess medlemmar och miljön kring dem. De bör överväga om resultatet av deras ansträngningar kommer vara respektfullt för alla intressenter.

   1.2 *Undvik skada*

   - Datavetare bör följa best practices för att undvika skada, som i denna uppförandekod menas med negativa konsekvenser.

   1.3 *Var ärlig och trovärdig*

   - Datavetare ska vara transparenta, ärliga om sina kvalifikationer och skapa en rättvis bild av en organisations policys och procedurer.

   1.4 *Var rättvis och ta handling för att inte diskriminera*

   - Datavetare bör främja rättvist deltagande av alla människor

   1.5 *Respektera arbetet för att skapa ny idéer, innovationer, kreativt arbete och datorartefakter*

   - Datavetare ska kreditera skapare av nya idéer, innovationer, arbete och artefakter.

   1.6 *Respektera integritet*

   - Datavetare ska respektera integritet och inte kränka individer eller gruppers rättigheter

   1.7 *Hedra sekretessen*

   - Datavetare ska skydda konfidentialitet förutom i fall där det finns bevis på brott mot lag, organisationsbestämmelser eller koden

2. *Professionellt ansvar*

   2.1 *Sträva mot att uppnå hög kvalité i både processer som i produkter av det professionella arbetet*

- Datavetare ska insistera och stödja hög kvalité på arbetet, både från de själva och från kollegor.

2.2 *Upprätthålla höga krav på professionell kompetens, uppförande och etisk praxis*

- Satsning på professionell kompetens hos de anställda ska göras

2.3 *Känn till och respektera existerande regler som tillhör det professionella arbetet*

- Datavetare ska följa regler. I regler inkluderas lokala, regionala, nationella och internationella lagar och förordningar, samt alla policyer och förfaranden för de organisationer som datavetare tillhör.

2.4 *Acceptera och förse lämplig professionell granskning*

- Datavetare ska söka och använda sig av intressentgranskning, samt ge konstruktiva, kritiska recensioner av andras arbete.

2.5 *Ge omfattande och noggranna utvärderingar av datorsystem och deras påverkan, inkluderande analys av potentiella risker*

- Datavetare bör sträva efter att vara lyhörda, noggranna och objektiva när de utvärderar, rekommenderar och presenterar systembeskrivningar och alternativ

2.6 *Utför arbete endast inom kompetensområden*

- Datavetare är ansvariga för att utvärdera potentiella arbetsuppgifter

2.7 *Främja allmänhetens medvetenhet och förståelse för datoranvändning, relaterad teknik och deras konsekvenser*

- Datavetare ska dela teknisk kunskap med allmänheten, främja medvetenhet om datoranvändning och uppmuntra förståelse för datoranvändning om det lämpar sig för kontextens och ens förmågor.

2.8 *Ges tillgång till dator- och kommunikationsresurser endast när det är tillåtet eller när det är påtvingat av allmännyttan*

- Datavetare ska inte få tillgång till andra datorsystem, mjukvara eller data om det inte är en handling som är auktoriserad eller en övertygande övertygelse om att det är förenligt med allmännyttan

2.9 *Designa och implementera system som är robusta och användbart säkert*

- Datavetare ska se till att systemen fungerar som avsett, och vidtar lämpliga åtgärder för att säkra resurser mot oavsiktlig och avsiktlig missbruk, modifiering och denial of service

3. Professionella ledarskapsprinciper

3.1 *Se till att det allmännas bästa är det centrala i allt professionellt datorarbete*

- Datavetare ska alltid ha allmänhetens bästa i åtanke vid olika uppgifter kopplade till undersökning, kravanalys, design och implementation m.m.

3.2 *Artikulera, uppmuntra acceptans av, och evaluera uppfyllandet av socialt ansvar av medlemmar i organisationen eller gruppen*

- Datavetare ska uppfylla relevanta sociala ansvarsområden och motverka tendenser att göra något annat

3.3 *Hantera personal och resurser för att öka kvalitén av det professionella arbetet*

- Professionella ledare ska se till att de förbättrar arbetslivskvalitén.

3.4 *Artikulera, tillämpa, och ge stöd till policies och processer som reflekterar principen av koden*

- Professionella ledare ska utöva de organisatoriska policies som är förenliga med koden och effektivt kommunicera dessa till relevanta intressenter.

3.5 *Skapa möjligheter för medlemmar av organisationen och gruppen till att utvecklas som proffs (professionals)*

- Professionella ledare ska se till att deras anställda får möjligheten till att utöka sina kunskaper och färdigheter.

3.6 *Var varsam vid modifieringen eller avveckling av system*

- Professionella ledare ska agera varsamt vid exempelvis ändringar av gränssnitt och mjukvaruuppdateringar. Datavetare ska hjälpa och förklara för systemanvändare varför vissa ändringar är nödvändiga.

3.7 *Känn igen och ta hand om system som integreras i infrastrukturen av samhället*

- Professionella ledare ska agera pålitliga förvaltare av nya system som utvecklas och blir en viktig del av samhällets infrastruktur.

4. *Överensstämmelse med koden*

4.1 *Vidmakthålla, främja, och respektera principerna av koden*

- Datavetare ska följa uppförandekodens principer och bidra till att förbättra dem. Datavetare som upptäcker överträdelser av koden bör handla så att de etiska problem de upptäcker.

4.2 *Behandla överträdelser av koden som oförenliga med medlemskap i ACM*

- Varje ACM-medlem bör uppmuntra och stödja efterlevnad av alla datavetare oavsett ACM-medlemskap.

Appendix B

Transkriptionsprotokoll Capgemini
Medverkande personer:
Cecilia Minder (CM)
Leo Rasmusson (LR)
Informant 1 (INF1)
Datum och tid: 2022-04-26 15:00-15:45

#	Person	Fråga/Svar
1.	CM	Okay! Can you describe your role at Capgemini?
2.	INF1	I was recently hired in the beginning of this month, so officially I am a consultant but the little work I have done so far on the few projects I have been introduced to so far has had to do with Data science, machine learning, and data engineering. In my role before this, which I had for one and a half years, I did a lot of data engineering work and data science specifically with natural language processing and stuff.
3.	CM	I think that's a nice segment into our next question since you mentioned NLP could you explain a little bit more about NLP?
4.	INF1	Yes of course, at my previous workplace a media monitoring company so it would go through, alot of it was based on a web crawler that would scan articles throughout the internet and also print data from magazines and newspapers and things like that. All to analyze how a particular company is being spoken about wherever on the internet and even on TV in a few cases. So there was a lot of text data that came with that basically. One of our primary goals was to analyze if the data was positive or negative or neutral for this particular entity, this person or the company. In 2018 there were some big advances in natural language processing with neural networks and stuff so alot of our work was based on some really big deep learning models for text based on architectures and design that were less than four years old.
5.	CM	So to your ability and knowledge, how would you define artificial intelligence and machine learning?
6.	INF1	Okay so my educational background is in math and physics so i have never sat in front of a professor and had them define any of these things for me. I don't know, any program that learns from data in order to make predictions or analysis. I mean that's not a great answer but I think it's hard to have a line. People would say that

		linear regression and stuff, these things that in the realm of my mind are more statistics and I have learned about them in that context. You could also consider them as machine learning applications. So I guess it is a little bit fuzzy in a way.
7.	CM	Moving on to the next question, what is your definition of ethics?
8.	INF1	I guess it would also be some kind of personal definition, doing one's best to treat sentient beings as they would like to be treated.
9.	CM	That is really nice, so how do you view or what are your thoughts of the relationship between AI and ethics?
10.	INF1	I think it's really important and only gonna get more important as we use AI for more and more things and make more and more decisions that have a greater impact on people's lives. Of course it's really important to have those discussions and understand how it can impact people's lives I guess.
11.	CM	Okay, thanks, so in what way would say that ethics affect your work?
12.	INF1	I guess for a lot of these sorts of workplace questions I kinda have to split them into two since it's very different here at Capgemini compared to the startup where I was working before. So at that startup there were a total of three people on the tech team. So extremely small and ethics were almost never discussed. so if it was gonna affect my work it would be a completely internal understanding of if this is right or if this is wrong and maybe thankfully i have never come across a situation in my work here or there where a module that a trained or work that i did would directly impact someone else's life where i could measure and see. Like in social media or other places where it's directly dealing with personal data or models that make really important decisions for people. Maybe if that was the case ethics would be a bigger discussion but it was always seen as something that we get to think about in the future once we are out of this startup phase but here at capgemini, i have been here for little over three weeks and i have had to do so many different trainings and things about awareness about ethics and doing what's right and capgemini's seven core values and none of them have specifically pertained to AI but i feel like it is part of the conversation or at least its clear that it is important and expected that ethical considerations should take place for everything that might be important. because that other place didn't have the resources or the size, it was not talked about in the same way.
13.	CM	What external factors such as regulations, political reasons and social aspects affected your previous work with AI and your current with ML?

14.	INF1	So I'm only aware of a few, I have never worked with anyone's personal data so something like GDPR has never affected me. I'm aware of plenty of cases here at Capgemini where it has but it never really did at my previous place of work. I'm aware of that at least. In the US there is HIPAA and I mean that's about the extent of my knowledge, and in regards to the social aspects , what exactly do you mean by that?
15.	LR	Probably like the media and society, more political and the political external pressure.
16.	INF1	Yeah yeah I'm definitely aware of that. I think that's the primary drive when it comes to my concern about things when I think about these questions and maybe the reason why I think about these questions and why I think about these things at all. It is the fact that's in the news and in the media, it is on my radar for those reasons and something that I definitely think about even though especially where I work before I did not really feel that from my environment.
17.	CM	Meaning your work environment?
18.	INF1	Yeah, my work environment.
19.	CM	Okay, thank you, so what established guidelines within AI Ethics are you aware of?
20.	INF1	None, and i made sure to not look them up either
21.	CM	That is totally fine, and then do you within your organization follow any established ethical guidelines?
22.	INF1	Established ethical guidelines, i don't know, definitely at my other workplace where i worked before no, but here i mean there is an internally established code of conduct and ethics and people i know where i can go to with ethical questions and support lines and anonymous support lines to go to with ethical questions but i don't know if its some kind of externally established guideline.
23.	CM	Okay, thanks, would you argue that it is necessary to follow these guidelines from a business perspective and I mean even if you said that you are not aware of any established guidelines but I guess you could say if you were, would you argue that it is necessary to follow these guidelines then?
24.	INF1	Yes, I would say so, from a coldhearted capitalism perspective i think, which i would think a lot of the people i worked with in the startup would have that perspective i think even they would see it as necessary in so far that it doesn't affect reputation. The business

		logic line of it and i would say that at capgemini that has been much more strongly stated. I feel that it is necessary but personally internally I would agree wholeheartedly for personal reasons.
25.	CM	Yes, thanks, do you know anyone in your workplace or other workplaces that work with or worked with AI ethics?
26.	INF1	So I'm aware of people here who work with ethics and compliance and stuff but not particularly with AI or ML. I would say that they only person that i have ever been aware of working explicitly with AI ethics would be the woman who used to work at google at ethics and AI, there was a semi high profile case where i don't know it was last year or something or there was someone who were working at google in that field and she was fired.
27.	LR	I guess that's how she got famous.
28.	INF1	So I guess officially no I don't know anyone who works or worked with AI ethics.
29.	LR	Okay, should we move on? In your opinion, is ethics something that needs to be considered and taken into account on a daily basis, in the work process of developing and applicating machine learning?
30.	INF1	In my opinion, yeah yes absolutely, especially like there are companies out there that use AI -systems to develop for example like car insurance quotas and levels, I know there is a company in the US that does that. And in a situation like that, where you're deciding how much someone pays, how much of a risk someone is, it is incredibly important to figure out what kind of biases your model has, yeah for sure.
31.	LR	Thanks, is microethics something you've heard of?
32.	INF1	No.
33.	LR	In your experience, do you know of any technical instructions in the industry or in the company that explains how to implement ethical AI?
34.	INF1	Technical instructions, no. Essentially just principles, but definitely not technical instructions. I've thought about it in the past, just vaguely, I remember being at a kind of AI conference or actually it was just virtually, but there was some kind of model in kind of related to the field I was working in with NLP. Someone was developing a kind of model where you could specify constraints for the model upfront, sort of ground truth information and I was thinking if you could modify that somehow to remove some of the biases that do exist in those big pre-trained models regarding race and gender. If you could input rules that say, to do that, but that has only been a thought that I've had. So no, I haven't.

35.	LR	Alright, thanks. So what would you say if there were standardized datasets with clear table of contents implemented, datasets that are transparent in how the data was collected and the purpose behind the data was clearly described. Do you think that is something that could be usable in your field?
36.	INF1	Yeah, for sure. I mean there are banks of datasets out there, but they, it's not required that you know all that information that you just mentioned. Where it comes from, are people aware that this exists and might have data that, I don't know, might have words that they wrote, a tweet that they wrote, is somewhere in this database, yeah that kind of information, I don't think is available.
37.	LR	Okay, alright. Have you had any education in ethics in your current or in your previous workplace or in your technical education when you went to college?
38.	INF1	Yeah current yes. Previous, no. It never came up, at least as a topic of education.
39.	LR	Alright. A lot of research that we found suggests that professional organizations and companies use ethics mostly as a marketing tool to address societal concerns and are not really used in practice. What is your opinion on this statement?
40.	INF1	Um though, I would say that my previous place of work didn't really mention it or it wasn't important to them neither…
41.	LR	So not even from a marketing perspective?
42.	INF1	Not really, it wasn't, I never saw them use those words or speak about it to investors or anything like that, maybe in the future if they grow they will, but not right now. Here I would say that so far I've definitely seen that it is out there as a marketing tactic, and I'm sort of yet to see how ingrained that is in what I'm actually doing here, i would say they had lots of internal education which I guess is a good first start, personal awareness, people I guess are aware about those things at the smallest level, but I don't know if I'm going to end up on a project someday and see that on a certain level once you reach a certain scale those rules disappear or if people actually respect them. That's something I kind of have my ear to the ground for, but and I hope that it's something people respect but I'm yet to know for sure.
43.	LR	Alright, okay, thanks. Let's see. To your knowledge, do you feel that clients value ethics when they inquire for machine learning or AI solutions?
44.	INF1	To my knowledge, no. I guess it would depend on the client, like I'm sure if Amnesty International wanted something done they would

		have requirements when it comes to that, but I think in general, no. I think they're very much interested in results and anything else kind of at least tacitly secondary.
45.	CM	Would you, I mean I know you haven't been long at Capgemini, but I guess you could use your previous experience as well, do you think the subject of ethics is something that will, like people will start talking about it more. Will it become more of an apparent thing or even more apparent to what it is now?
46.	INF1	I, I think so. Out of necessity, yeah, I doubt, I think it's already become better, atleast at a place like Capgemini, I think, I can see the ways that events of the past like with Facebook, and data privacy breaches becoming more and more of a thing, and something that affects ultimately companies bottom lines, that's what I think makes this happen, makes companies pay more attention and I think that will continue to move in that direction. I see no reason why it would stop.
47.	LR	Do you know if Capgemini or your previous employer had any KPI, so Key performance indicators, when it comes to strictly around data bias, justice, equality? Something like that.
48.	INF1	My previous place, no, not at all. And here I actually don't have any personal KPIs yet, I started when they like resetted them and analyzed them for the next upcoming year so I don't have them, but to my knowledge, no. As far as I have heard from other people, at least on an individual level, I don't know if on the level of our department or something, if that's something that they pay attention to but it hasn't reached my ears at least.
49.	LR	Do you know if Capgemini has a white or black list for the algorithms that you work with? So rules that the system always needs to follow or the situations or behaviors that the system never can have. I'm not sure you are aware of the white and black lists, or if that is commonly known in the branch, I'm not sure?
50.	INF1	No, I haven't heard of anything like that. I would be surprised if they did have something like that. If I could ask a question, like what kind of algorithms would be white or black listed?
51.	LR	No that is more like the, it is more the rules that would be, rules would be whitelisted, rules that the system always needs to follow would be considered white rules and the black rules would be like behaviors or situations that the systems, that should never happen.
52.	INF1	Okay, in that case I think, I don't think Capgemini would have any rules like that, being a consultant company, I think it might come from the client side, I could imagine a case where a client would

		expect and demand something like that from a result that. But I haven't heard of anything like that from speaking with any of my coworkers, I haven't seen it as a requirement.
53.	LR	Alright okay, thanks. So let's see. We have seen a lot of research that's point to embedding ethical reflection in a developers or a system scientist's daily work, would mean ethics would be normalized in the workplace, so for example daily discussions about for example when it comes to open source code "who owns the code" "who should own the code". What is your opinion on this, is that something that could practically work in your workplace?
54.	INF1	Yeah, I think so. I think it is important to maintain that awareness, especially for certain situations where AI is making important decisions, yeah yes I would agree.
55.	LR	Alright, I think that was all of our questions actually. I think we can pause the recording.

Appendix C

Transkriptionsprotokoll Malmö stad

Medverkande personer:
Cecilia Minder (CM)
Leo Rasmusson (LR)
Jim Samuelsson (JS)
Datum och tid: 2022-05-02 14:00-15:00

#	Person	Fråga/svar
1.	CM	Då Jim Får du gärna berätta om din roll på Malmö stad?
2.	JS	Ja så jag är anställd som så kallad data scientist inom Malmö stad på arbetsmarknad och socialförvaltningen och jobbar i en enhet som heter digitaliseringenheten där vi är ungefär 15 kollegor som jobbar med olika typer av digitalisering som sker i arbetsmarknad och socialförvaltningen. Som data scientist är jag ansvarig för kan man väl säga analys av den data som susar runt i våra system och ett huvudsystem kan vi säga. Där allting ute i verksamheten när det gäller ekonomiskt bistånd eller arbetsmarknads-kopplade åtgärder, narkotikaproblem, allt det sociala. Det är en bred palett och det registreras då i ett huvudsakligt system och då uppkommer det frågor om det kan extraheras någonting från huvud-datan som kan vara användbart vid beslutsprocesser samt förbättringsprocesser. Det är det och det ska väl också sägas att jag jobbar mycket med andra kollegor. Jag sitter inte alls isolerad, jag sitter med folk som kan de olika domänerna.
3.	CM	Tack, och hur länge har du arbetat inom området.
4.	JS	Ja alltså med data har jag arbetat i 30 år.
5.	CM	och vad har du för bakgrund som tidigare?
6.	JS	Alltså rent akademiskt är jag doktor inom teoretisk fysik och sedan så har jag under många år arbetat inom läkemedels och biotech industrin i köpenhamn, där jag jobbade i många år innan jag kom hit. så att det väl liksom varit data i olika former och utvecklat algoritmer och metoder för att kunna hantera data, beräkning. Bara liksom förvara data på ett bra sätt och så vidare. Olika aspekter av Data.
7.	CM	Egentligen vad har du för arbetsuppgifter och ansvarsområden samt är det så att ni hanterar personlig data?
8.	JS	Ja jag tror nog att mina arbetsuppgifter beskrev jag kort innan så att säga, personlig data hanterar vi och den data är ju väldigt känslig för de

		handlar om människor som befinner sig i väldigt utsatta situationer i nöd och sånt. Det finns mycket lagstiftning och sånt om detta. Den data jag primärt använder är den data jag får ut från systemets backend. Handläggare och socialsekreterare och så sitter i frontend och knappar in och registrerar. Den data som analyseras på min sida är det vi tar upp på baksidan och där finns det personuppgifter och personnummer men det är krypterat så jag ser aldrig personnummer, jag ser heller aldrig några namn. Jag ser aldrig heller några bostadsadresser eller något sånt. Det enda explicita namn jag ser är namn på handläggarna. Det råder mycket tankar här om vad vi får och inte får göra här. Ibland vill man kanske göra saker som man inte kan på grund av att lagstiftning är som den är. Såatte men det var ett långt svar på en kort fråga, ja vi hanterar personlig data , ja det gör vi
9.	CM	Ja men det uppskattar vi, tack så mycket! Då tänker jag gå vidare egentligen. Så gott du kan, hur skulle du definiera AI och maskininlärning?
10.	JS	Det är en svår definition. Metoder om mjukvara som hjälper oss att bli klokare som kan lära sig själv i någon mening. Det finns säkert någon jättefin definition på wikipedia. Det är väl det de handlar om metoder som hjälper oss att bli klokare och också att det finns någon sorts lärande alltså något självlärande i det. Det kan man ju brodera ut så mycket som helst ju. Men jag kommer säkert komma på något bra efter intervjun.
11.	CM	Du får du dra iväg ett kompletterande mail! Ja på vilket sätt tänker ni då i framtiden att ni kommer använda AI på er data?
12.	JS	Jag tror vi kommer använda den, just nu så de här första två åren här när jag jobbar har det varit väldigt mycket konkreta frågor, Hur många finns det av den sorten? Hur många arbetslösa har vi som samtidigt går på ekonomiskt bistånd? Hur många barn och unga far illa i stadsområdet öster? Alltså riktigt konkreta frågor som i någon mening varit bra här inledningsvis. De flesta människor har väl sett eller hört ai men vad är det egentligen. Det är ju rätt abstrakt så vi har ju egentligen börjat i andra änden och besvarat konkreta frågor som folk har haft och så har vi sett att vi kan faktiskt med hjälp av vår data få ut svar som kan hjälpa oss samtidigt som vi jobbar med det har jag försökt strukturera datan på ett sådant vis att när vi väl finner ett problem som vi tycker kunna vara lämpligt så har det liksom ett format som vi kan kasta AI eller ML metoder på det så att säga. Det vet vi ju. Ska det vara lämpligt ska det vara rätt problem och datan ska vara av god kvalite men det ska också vara ett visst format för att det ska göra lösningar på datan. Vi strävar liksom i den riktningen och nu börjar vi nog komma till en punkt när man skulle kanske kunna tänka sig AI eller maskinlärande algoritm. Då gäller det att hitta mönster som man inte kan se själv, ytterligare kanske en sak med föregående fråga är att liksom att kunna se mönster som liksom det mänskliga ögat inte själv kan upptäcka.

13.	CM	Och en annan definitionsfråga, vad är din definition av Etik?
14.	JS	Ja, den är ju, det finns säkert också någon fin wikipedia definition av det. Ja men man skulle nästan kunna säga att man det är någon sorts rekommendation att man beter sig så att man kan se sig själv i ögonen. kanske , någonting sånt. Vara snäll skulle Bamse sagt.
15.	CM	Det är väl jättebra och om vi får binda ihop de här två hur ser du på förhållandet mellan AI och Etik?
16.	JS	Det är väl det vi sa tidigare. vi handskas ju med känsliga uppgifter och jag tror att när man väl kastar in känsliga uppgifter i ett AI-system så vet ni ju själv att AI boxar kan vara rätt svåra. Den som gjort AI programmet vet naturligtvis vad som hänt där inne men mottagaren av resultatet av AI kanske inte kan se vad som hänt här inne eller vad som beslut har tagits, liksom varför det som kommit ut på andra sidan är som det är. Så att det är ju, det kan ju vara ett konfliktfyllt förhållande och där får man ta någon försiktighetsprincip och det är nog liksom det korta versionen på det tror jag.
17.	CM	Tack. Hur skulle du påstå att etik påverkar ert arbete idag men framförallt främst kommande arbete?
18.	JS	Det påverkar mycket om man med etik menar den lagstiftning som följer efter något sorts etiskt övervägande. Vi är ju omringade av väldigt mycket lagstiftning så att, jag ska inte säga var gång men ganska ofta om vi ska ta in någon ny data får vi tala med folk som känner till lagstiftning och vad vi kan göra. Det är ju mer lagstiftning men den har ju ofta sitt ursprung i etiska överväganden så att det är något som förekommer hela tiden så att säga. Jag kommer till exempel om jag har någon data som gäller och som jag tagit fram angående personuppgifter kan jag inte bara skicka iväg den i ett mail utan att ta bort personuppgifterna eller kryptera det på något sätt. Det förekommer ständigt.
19.	CM	Och då egentligen, det kanske du har besvarat nu men jag lyfter den ändå, det är om det finns ytterligare något att påpeka. Nu när ni tänker att ni ska börja använda AI, finns det andra yttre faktorer, bland annat politik eller folk opinion och hur det kommer påverkar arbetet mer än vad det gör idag?
20.	JS	Det vet jag nog inte, jag tror att det kommer bli så att om vi vill göra något som ligger i gränsområdet så kommer vi nog tala med våra Jurister också får dem uttala sig om, och de uttalar sig ju naturligtvis när det gäller lagstiftning, kanske inget direkt kopplat till den dagliga folk opinionen. Jag hoppas iallafall inte att det kommer påverka.
21.	CM	Bra, vilka etablerade riktlinjer inom AI-etik är du personligen medveten

		om?
22.	JS	Där måste jag erkänna att jag är svag på det området. Jag var faktiskt med med några av mina kollegor på ett projekt förra året som handlade om förenta nationerna UNICEF som är UNICEFS barnavdelning och de har startat ett jättestort projekt som heter Barnen och AI. Det är lite att AI kommer på stark frammarsch och då kommer det ju naturligtvis också påverka barnen och då har de lagt upp ett stort program för detta och det finns rätt många riktlinjer kring detta, jag har de inte i huvudet och då var vi med lite på det och tog fram några saker och jag är helt säker på att det finns många lokala i Sverige också och nationella men inga jag känner till generellt.
23.	LR	Bara en följdfråga där, var det där projektet specifikt inom Malmö Stad eller ?
24.	JS	Nä det var faktiskt helt globalt och så var det tre kommuner i skåne som var med på ett hörn, Malmö, Lund och Helsingborg.
25.	CM	Okej, Följer ni inom Malmö Stad några etablerade riktlinjer? Jag har några underfrågor till den också men
26.	JS	Jaja, Det var väl som jag sa innan, vi har när det gäller AI inte kommit så långt i AI så vi tänker inte så mycket AI men när det gäller data har vi väldigt mycket regler. Vi har ju folk som är heltidsanställda för detta. Så svar på den frågan brett är Ja, vi följer riktlinjer.
27.	CM	Okej, Och skulle du påstår i ditt arbete och i synnerhet som kommunal organisation att det är nödvändigt att följa etablerade riktlinjer?
28.	JS	Ja det är det såklart, speciellt med den typen av problem där det är svaga människor som är utsatta på olika sätt, då är det extra viktigt.
29.	CM	Även om du kanske inte riktigt nu hade någon av dessa riktlinjerna i åtanke, Jag tänker att om du vill minnas när du stött på dessa riktlinjer om det finns en hög abstraktionsnivå på dessa?
30.	JS	Ja det finns det faktiskt, skulle jag vilja säga. och ni skickade ju frågorna och jag skrev ut dem och ni frågar ju här sedan dvs svåra att applicera i praktiken och det kan man faktiskt säga ibland då personuppgifter. Så fort det finns några uppgifter knutet till person är det personuppgift men det är ett rätt vagt begrepp. Till exempel jag som jobbar med krypterad data. Men hur känsligt är det då egentligen?, när man inte ser något namn eller adress eller sådär. Där är det faktiskt väldigt svårt skulle jag vilja säga att vända sig till någon och nu eftersom vi jobbar med krypterad data kan vi egentligen inte då agera lite friare. En sak till exempel som ni kanske inte känner till är att personuppgifter får hanteras inom en myndighet men det får inte per automatik skickas över mellan myndigheter. Till exempel i Malmö stad där jag jobbar på

		Arbetsmarknad och socialförvaltningen är en myndighet och skolförvaltningen är en annan myndighet så man kan inte bara ta och skicka inom staden, personuppgifter bara sådär. Detta kan vara en hindrade sak om vi till exempel är intresserade av ungar som far illa så vi in en del på socialförvaltningen men så kan vi också vara intresserade hur det till exempel går för den här ungen i skolan och det hade vi velat knyta ihop den här datan. Detta är något vi generellt sätt inte kan göra nu.
31.	LR	Detta gäller även när det är krypterad data?
32.	JS	Ja så vitt vi vet, Jag ska inte säga att vi försöker pusha linjen lite men vi försöker utmana lite vad det finns för möjligheter. Vi kommer säkert dit tillslut och kommer kunna göra det men man kan liksom inte bara gå till rad 28 och så står svaret där utan och jag har faktiskt varit i kontakt med Stockholm så att säga och pratat med integritetsskyddsmyndigheten och dem kan inte heller säga, detta får du göra och detta får du inte göra utan det handlar alltid om fall till fall. Det beror nog på att det här är rätt nytt ändå även om vi har haft data och datorer länge så just det här med att kunna knyta ihop data och sånt. I och med det här med AI och Maskininlärning har det öppnats upp väldigt många möjligheter där de stora möjligheterna ligger i att just kunna knyta ihop data från väldigt många olika håll. Då kan AI se saker som vi inte kan se visuellt till exempel. Det ska helst inte som vi som personer eller lagstiftning riktigt lika långt komna som tekniken är. Vi kan göra det men vi får inte göra det.
33.	CM	Vill du påstå att om man hade kunnat knyta ihop de olika områden så att man får mer data eller en annan typ av data, hade det lett till att beslutstagandet hade sett annorlunda ut?
34.	JS	Det är jag helt säker på. Och inte bara jag utan mina kollegor med.
35.	CM	Även i den positiva riktningen?
36.	JS	Då hade man fått en bättre förståelse till exempel då hade man kanske tidigare kunnat fånga upp barn som far illa. Man får liksom någon signal i skolan och har vi sett något på orosanmälan och nu lever dessa två i separata världar så att säga. Det skulle ge stora fördelar helt säkert.
37.	CM	Väldigt intressant! Du har nog egentligen pratat lite om det här tidigare just om Malmö stad har en uppförandekod eller policy i förhållande till etik?
38.	JS	Ja, Vi har massa policies och koder men jag kan inte säga att det är just den, den finns på den hemsidan. Men just i förhållande till etik får jag nog säga att jag inte vet då jag inte sett själva dokumentet.
39.	CM	Ja, då ska min kollega ta över…
40.	LR	Det var en fråga till eller? Eller ja vi kan köra på den.

41.	CM	Ja juste!
42.	LR	Känner du någon på din arbetsplats eller andra arbetsplatser som då arbetar enbart med maskininlärningsetik eller AI-etik?
43.	JS	Inte på vår arbetsplats i nu..sen känner jag andra människor som gör det så att säga, till daglig dags.
44.	CM	Jag tänkte bara fråga om det var liksom inom kommunen då eller om det var..?
45.	JS	Inte vad jag vet inom kommunen, det kanske är någon, det är ju många som är anställda och det finns liksom 8 *ohörbart* på olika ställen och sen är det ju lite en definition vad man menar med AI och maskininlärande. Men jag känner ingen inom kommunen som liksom jobbar med det till daglig dags så. Men på andra arbetsplatser känner jag folk som gör det.
46.	LR	Okej. Ska vi se, jag tycker redan han svarat på den. Men då kör vi direkt in på… Är mikroetik något du har hört talas om personligen?
47.	JS	Nej.
48.	LR	Okej.
49.	JS	Jag läste frågan här när jag fick den, i slutet av förra veckan tänkte ja, nej jag ska inte gå ut på nätet och läsa, jag ska vara så okunnig som jag är. Jag vet inte vad det är.
50.	LR	Det är helt rätt. Okej, ska vi se, känner du att det finns några tekniska instruktioner inom er organisation eller inom fältet AI för hur man ska implementera då etisk AI?
51.	JS	Vi har inte haft den diskussionen riktigt nu hos oss skulle jag vilja påstå. Så jag får nog svara nej. Det betyder inte att det inte finns, det kan finnas någonstans i något dokument som jag inte sett men jag känner inte till det.
52.	LR	Okej, och det är inget du har hört talas om inom branschen eller så där inom vad ska man säga, ja men kanske inom data eller inom AI?
53.	JS	Kanske inte direkt kopplat till, jag har nog mer hört kopplat mer till det här generella, som liksom hur hanterar vi data och när det är känsliga uppgifter sånt där. Vi har ju liksom, jo faktiskt jag kan faktiskt säga att, i samband med det här projektet, UNICEF projektet, som vi jobbade med förra året så valde liksom, vi var tämligen fria att göra vad vi ville med projektet bara det på nåt vis var kopplat till data/AI. Och vi gjorde, vi analyserade data hos oss, men det var liksom ingen AI sak. Var Lunds kommun gjorde det va, de i någon mening, de gjorde inget alls rent tekniskt, utan de har tydligen om jag förstod det rätt redan nån sorts riktlinjer kopplat till data, AI och etik och då inom ramen för det projektet så reviderade de dem riktlinjerna så att säga. Och då tror jag

		också det, liksom ingick att det fanns liksom instruktioner, från de här riktlinjerna kommer det säkert tekniska instruktioner för hur man får hantera data och sånt i samband med AI. Så att, jag känner till att det är på det viset, men jag kan inte detaljerna.
54.	LR	Tack så mycket.
55.	CM	Jag tänkte bara, är det kryptering som är liksom, vad säger man, den främsta tekniska aspekten av när man just hanterar personlig data, om det går att svara på?
56.	JS	Jag önskar faktiskt att jag kunde svara på den frågan, jag vet inte riktigt om det, om det vore på det viset, så hade det varit bra, för då hade vi liksom kunnat göra rätt mycket genom att kryptera data, men jag tror det också finns andra aspekter som återknyter då till de frågorna tidigare, det är lite svårt att få liksom grepp om vad det är som är, liksom det som avgör huruvida man får använda det som man vill eller inte. Det står här en hög abstraktionsnivå, man skulle faktiskt kunna vara så fräck och säga kanske, att det finns en viss luddighetsnivå i det skulle jag vilja säga. Det är inte riktigt klart.
57.	LR	Okej, yes, nu är det en lång fråga här men, Vad skulle du säga då om det fanns standardiserade datasat eller då förutbildade (pre-trained) modeller med då tydliga innehållsförteckningar som skulle kunna användas till olika implementationsområden och då framförallt hur den här datan då blivit insamlad och att man då kan se intentionen bakom datan och hur själva insamlingen gick till. Tror du det skulle vara användbart?
58.	JS	Ja alltså, jag tänkte när jag läste det, betyder det, bara så att jag förstår det, det betyder att det skulle, vart skulle de datamängderna komma ifrån så att säga?
59.	LR	Alltså antingen om de skulle vara open-source eller om de bara skulle vara tillgänglig inom er organisation men mer bara så att det är tydligt hur de har samlats in, för vi vet i alla fall att det finns mycket open-source dataset online där det är väldigt luddigt om man skulle säga så, var datan har kommit från, helt enkelt.
60.	JS	Nej men det tror jag, är värdefullt. Sen gäller det såklart när man när man väl är i den aktuella situationen så, så är det inte säkert att de dataseten som då håller den här höga kvalitén med liksom, beskrivningar, att det är det dataset man kan använda just i den uppgift man vill lösa. Men man skulle kunna tänka sig att man kanske ser att om vi ska använda AI maskininlärande i Malmö Stad, så var gång vi gör det så måste vi, göra de här sakerna som vi beskriver, det ska vara beskrivet hur de är insamlade, ja liksom annoterade på lite olika sätt, så att säga, datamängden, så det är klart och tydligt var den kommer ifrån, kanske vad den får användas till. Så att, det kanske är, det kanske nästan är något, min, för när datan väl hamnar, liksom i AI-boxen, så och sen

		kommer det ut något på andra sidan, då har man så att säga tappat rätt mycket av vad som kom in. Men AI är ju liksom, hur många dimensioner som helst, projiceringar av vektorer och hit och dit, och då kan det vara väldigt bra att kanske ha, men den ursprungliga datan som analyserade, det var denna, bestod av denna, och insamlingen, alltså den och den, och kanske även någon kontaktperson, det tycker jag låter hur vettigt som helst.
61.	LR	Okej, tack. Ska vi se, och då, har du fått någon utbildning inom etik på just arbetsplatsen?
62.	JS	Nej det har jag inte fått.
63.	CM	Är det någonting du känner skulle vara, kanske, vad säger man, gynnsamt?
64.	JS	Ja, jag vet inte. Jag tycker vi ändå har en, för i och med att vi ändå har en rätt levande diskussion i det dagliga arbetet, vad vi kan använda, vad får vi göra, så känns det inte som, även om det ibland kan vara svårt, att svara på direkta frågor, så liksom är det ändå något som genomsyrar, vi får vara försiktiga med datan, vi får tänka på vad vi gör, så jag känner inte att det är en brist direkt att vi inte har haft någon direkt utbildning.
65.	LR	En lite liknande fråga, men var AI-etiken en del av din universitetsutbildning?
66.	JS	Nej.
67.	LR	Det var det inte, okej.
68.	CM	Skulle jag kunna komma med en sidofråga, tänker du att, kanske svårt, men din personliga åsikt då, tycker du det är något som borde, kanske läggas till på olika utbildningar inom ja…
69.	JS	Ja, jag är nog så naiv så jag hade trott att om man går en liten längre AI eller maskininlärande utbildning på universitetet idag så kommer det med en delkurs men det kanske det inte gör? Jag vet inte, det var så länge sedan jag var vid universitet.
70.	LR	Vad jag vet på LTH, i alla fall, så tror jag inte att det finns någon explicit etikutbildning där även om de hanterar mycket maskininlärning och AI-algoritmer i deras utbildning. Det är bara min tjej som jag vet därifrån.
71.	JS	Ja för det finns ju mycket prat om det här att när man utvecklar och det handlar inte bara om AI utan när man utvecklar algoritmer rent allmänhet att, så det är nog ingen överdrift att säga att de flesta som skrivit kod för maskininlärning är liksom vita män.
72.	LR	Okej.
73.	JS	Och, man skulle kunna tänka sig att, sådana personer designar sina program och algoritmer så att det smyger sig in någon form av det som

		kallas bias, ja fördomar och sånt kanske. Så det kanske inte vore fel att ha gått igenom några poäng AI-etik innan man sätter sig ner, eller ja, medan man gör den här typen av jobb. Så det skulle säkert vara en god idé. Om inte annat ur det just vetenskapliga, kanske inte så mycket etikens synpunkt, för jag kan tänka mig att om man kommer in med ett, man ska ju liksom inte koda in fördomar och då blir det en dålig algoritm i sig, så även rent liksom tekniskt så kanske det är en fördel att man tänker på det nästa gång innan man börjar liksom. Det var bara högst en personlig reflektion.
74.	LR	Ja men det är helt rätt. Ska vi se här. Vi har hittat mycket forskning, som pekar på att organisationer och företag som använder AI i praktiken använder det mer eller mindre etik då i marknadsföringssyfte, hur ser du på det här påståendet?
75.	JS	Ja det är väl ingen som är förvånad. Konstigt vore annat. Nej, det är ju inte bra men. Sen anser man kanske att det, ja men om, om det finns någon konsumentmakt så kanske man kan se till att det blir verkstad av det hela men det vet jag inte. Men AI är ett rätt svårt område, så det är liksom inte bara, det är ju inte som att köpa tandkräm direkt, utan det är rätt abstrakt ju. Så nej jag är inte förvånad att det är på det viset, ja precis det är marknadsföring.
76.	LR	Vem tycker du etikansvaret ska falla på inom en organisation?
77.	JS	Nej det får nog alla ta ansvar för.
78.	LR	Alltså personligen då?
79.	JS	Ja, det tro jag. Chef över etik låter lite…Etikförvatlningen kanske man kan ha, nej. Men jag tror nog att det får vara på den personliga nivån, så får man tala med sina kollegor om det.
80.	LR	Yes. Får ni, eller i det arbete du har gjort med data då, än så länge, och för ert kommande arbete med AI, får ni stöd på nåt sätt i form av pengar, utbildning eller marknadsföring?
81.	JS	Ja kopplat, mikroetik som sagt känner jag ju inte till, men alltså, jag har inte jobbat med externt stöd men det finns ju andra som har gjort det här ju. Och jag vet inte om, om man till exempel, vi har kollegor som har jobbat till exempel med Vinnova, kanske inte direkt AI, men låt oss kalla det för data/digitalisering. Och jag vet inte om det har varit några liksom etiska aspekter av det. Men att det har kommit lite pengar utifrån, så är det.
82.	LR	Det har det ja, okej.
83.	JS	Men om det har liksom varit nån, “om ni får de här pengarna måste ni också liksom göra något inom etikområdet”, det vet jag inte.

84.	LR	Ja nej, okej. Ska vi se, ja, nu kommer några lite längre frågor i rad här, men vi har också hitta lite forskning som visar på att det kan vara gynnande att ändå ha en white/black list för AI-algoritmer vilket då helt enkelt är regler som systemet alltid ska följa och då beteende som systemet aldrig får finnas i. Tror du det är något ni hade kunnat applicera i ert kommande arbete?
85.	JS	Jag vet faktiskt inte. Alltså jag kan varken svara ja eller nej på det för att det får man nog nästan se konkret för att man ska kunna förhålla sig till just, jag får nog svara jag vet inte på den frågan.
86.	LR	Okej, ja. Då är det en lång fråga till här. Men ett förslag för att få en organisation att bli mer etisk då är att bädda in etisk reflektion i en utvecklares eller AI arbetares dagliga arbete och det skulle då innebära att etiken normaliseras på arbetsplatsen och man öppnar upp diskussionen i frågor som open-source kod eller datasets. Hur ser du på detta?
87.	JS	Ja men jag tror, jag tycker det är lite knutet till det vi pratade om innan, att man det inte är bra att lägga in bias, inte bra är ju direkt fel, men alltså att det finns en risk för det, så att, det är väl…ska bara se på frågan… Ja det tror jag nog skulle kunna göra, ehm, det krävs säkert lite träning, det är ju ingenting man sett liksom "nu ska jag vara etisk när jag skriver" utan det är nästan på nåt vis ha tänkt igenom vad det innebär, det skulle nästan, alltså nu tänker jag jättehögt men alltså, det kanske till och med krävs att man nästan är i team där man liksom får tala med varandra och till och med granska varandras kod, det är ju inte ovanligt när man utvecklar i teams tillsammans, så ja det skulle jag kunna tänka mig skulle fungera.
88.	CM	Jag tänkte bara om jag ska flika in, alltså från tidigare, nu när du har beskrivit och så, är det ändå en ganska, även om man inte kanske skulle säga att det, ni har en etisk diskussion dagligen alltså i kafferummet eller på fikan eller så, men ändå att den aspekten finns dagligen, alltså just även om man inte " nu har vi en etisk diskussion här"...
89.	JS	Ja men det skulle jag påstå, den är, den är i stort sett dagligen, just på grund av att den data vi har är så pass känslig så att säga, så att, sen är det ofta att man tycker att det är mer jobbigt än bra för att de här reglerna runt omkring hindrar oss lite från att arbeta kanske som vi vill och men det är hela tiden, närvarande.
90.	LR	Vi kan köra den här frågan också, den är också väldigt lång om du läser den där lite.
91.	JS	Ja, jag kan läsa den här, det är nummer fyra?
92.	LR	Exakt nummer fyra.
93.	JS	Ja, uhm. Ja men det kanske, trust comparison index, pålitligt, bara så att jag förstår, pålitligt här är det liksom i, hur bra det är liksom för att

		producera bra resultat eller ska det här ses som i ett etik sammanhang.
94.	CM	Ja det är nog mer det vi tänker.
95.	LR	Ja det är mer i etiksammanhang skulle jag säga. Att det är pålitligt därifrån. Det är ganska vagt.
96.	JS	Nej nej det är bara så att jag förstår frågan.
97.	CM	Det är väl mera att vi, man har lyft lite att antingen att man skapar en sorts certifiering för att liksom, som du säger att man kanske till och med vet lite vad den här boxen som vi nu liksom stoppar datan att man vet att det är, det som kommer ut kan vara etiskt bra resultat. Om man kan förklara det på.
98.	LR	Ja men det skulle man kunna säga att det mer, alltså mer att ett system får en stämpel av att det här är liksom, ja lite som vi snackade om dataset att det är gjort med rätt intentioner i tanke om vi säger det så.
99.	JS	Ja men, det låter som en god idé, frågan är bara vem som ska sätta den stämpeln egentligen.
100.	LR	Ja det är sant. Det får ju vara nån extern kanske…
101.	JS	Ja nationell, ja nån extern ja, ja alltså man kan ju, det låter ju bra, men man skulle också kunna tänka sig att det då skulle vara otympligt och kanske dyrt då också, jag vet inte. Men det...
102.	CM	Och det låter som aspekter som man inte, som man inte kommer kunna komma ifrån, det känns ju att pris kan styra väldigt mycket, liksom hur...
103.	JS	Ja, ja den är ju…
104.	CM	Att det blir liksom en ekonomisk fråga istället för en etisk fråga, kanske?
105.	JS	Ja precis, om det är någon som ska sätta den här stämpeln, nån ska ju göra det så att säga ja, krävs det, hur ska den, vem ska göra det och kan man göra det internt i Malmö Stad, det låter som en tilltalande idé men den är ju nog inte så himla lätt att genomföra.
106.	CM	Men, detta kan ju visa lite på den här abstraktionsnivån vi pratade om tidigare liksom, det låter bra på papper men ja…
107.	JS	Man hade ju kunnat tänka sig, och det var lite därför jag ringde då till integritetsskyddsmyndigheten faktiskt för bara några veckor sedan för att, det hade varit skönt att luta sig mot liksom nationell myndighet, för där liksom, där har de liksom hundratals människor som jobbar med sånt, men det är ju inte på det viset. Utan, men det hade ju och det är ju något som vi kommuner i allmänhet, i alla fall ha någon sorts riktlinjer i alla fall hjälper ju ofta, att kunna luta sig mot det, sen om det ska gå så långt att det är liksom ett trust comparison index, det vet jag inte men, man skulle kunna tänka sig att nån myndighet, ja men om ni ska använda AI

		inom den offentliga förvaltningen så måste de här kriterierna uppfyllas eller ni måste tänka på det här och så vidare, nånting, skulle man kunna tänka sig.
108.	LR	Absolut, okej. Yes, inom testning och validering av en AI lösning är ett förslag att sjösätta ett separat team med uppgiften att försöka hitta svagheter och mer eller mindre försöka ta sönder systemet. Hur ser du på det här?
109.	JS	Ja, ja det är ju en rolig tanke. Ehm, ja det är ju.
110.	LR	Det skulle ju då vara i en extern miljö, nån VM gissar jag i så fall.
111.	JS	Ja, som sagt, det måste ju vara någon som kommer utifrån, det är, det kanske också jag vet inte, alla program och alla, allting inom AI, maskininlärande är ju liksom inte livsavgörande så att säga. Så det kan ju bero lite på inom vilket område systemet ska användas liksom, är det något AI system som ska styra medicinsk utrustning på en intensivvårdsavdelning ja då kanske det hade varit vettigt att sjösätta ett sånt team, men är det bara liksom något som ska hjälpa någon att spara lite tid på en arbetsplats kanske det inte är lika aktuellt. Det kan vara lite från fall till fall.
112.	LR	Ja men det är en bra poäng.
113.	CM	Om jag får flika in nu nu när ni är i början av ert AI arbete, om det är många du pratar med eller så. Vad skulle du säga att folks generella bild av AI är, är det mer Oj ska ni verkligen ta er an det eller hur är det?
114.	JS	Vi har faktiskt inte haft den diskussionen så mycket för att vi har ju som jag sagt tidigare tagit fram väldigt konkreta problem och försökt röra oss i ett håll när vi till slut kommer vara redo för AI. Både jag och min chef tycker att vi börjar med det konkreta och skapar förtroende hos de människor som använder våra tjänster. Detta får man nog säga att vi har lyckats med och men vi inte, det finns liksom i andra kommuner i Sverige där man kommit med sin AI låda, man har alltså kommit med sitt verktyg och så försöker man hitta uppgifter för lådan. Detta har varit svårsålt för de flesta människor vet ju inte riktigt vad AI är. Det är något man ser som spännande och intressant men hur ska det passa in i min konkreta arbetsvardag. Där på de platserna skulle jag kunna tänka mig att ha en diskussion men vi har inte kommit dit ännu och jag tror faktiskt att när vi kommit dit har vi byggt upp ett sådant förtroende mellan oss att folk säger: Låt oss testa. Vi säger ju inte att det är lösningen på allt, man får ju liksom se vad som kommer ut på andra sidan i form av resultatet men vi har inte haft diskussionen men när den kommer tror jag den kommer bli bra. Detta eftersom vi har börjat med konkreta frågeställningar
115.	LR	Okej, tack. Också sista frågan här, vad tror de största utmaningarna kommer vara med ert kommande AI-arbete?

116.	JS	Jag tror faktiskt det blir rent konkret, inte kopplat till etik, jo kanske lite att vi måste få grönt ljus från Jurister osv vilket jag tror vi kommer få när vi väl har ett bra case, alltså att detta vill vi använda AI så tror jag vi kommer få det. Men rent konkret är frågan om vi kommer bli klokare av det som kommer ut på andra sidan av AI-boxen och dra slutsatser samt ta bättre beslut ifrån dem. För det vill ju till att data är av sådan natur att AI eller maskininlärning kan komma med någonting. Så det vill ju till att det finns någon form av data. Vi har en massa data och god sådan men det är inte säkert att den är lämpad för AI men det får vi helt enkelt testa. Det får vi testa. Det tror jag är den största utmaningen att verkligen att hitta områden och hitta datamängder som kan hjälpa oss. Det är bara trial and error, vi prövar med något och så får vi se.
117.	CM	Och det är data ni kan använda för att testa?
118.	JS	Ja vi har ju mycket data och som är intressant och då skulle man kunna hitta mönster i den här datan. Just nu tittar vi konkret på folk som är arbetslösa och så kommer det med i något som kallas arbetsmarknadsåtgärder. De här människorna är ofta de som också får ekonomiskt bistånd det som tidigare kallades för socialbidrag. Då är ju frågan liksom, de som kommer med i arbetsmarknadsåtgärder, blir de hjälpta? alltså får de ett jobb, blir de självförsörjande?, kommer de bort från ekonomiskt bistånd? Och då kan vi se om det är folk som kommer bort från ekonomiskt bistånd och blir självförsörjande och de som inte kommer bort från ekonomiskt bistånd trots de hamnar i en sådan här arbetsmarknadsåtgärd, och då ställer man sig frågan. Har de människor som lyckats någon speciell egenskap, har de någon arbetslivserfarenhet, är det något i deras utbildning? är det kön? Är det en viss typ av familjesammansättning? Vi har ju all den datan. Det förstår ni ju själv att detta är något vi hade kunnat kasta på lite AI och ML på. Finns det någonting där vi kan plocka ut? Just nu vet vi inte svaret. Nä det kanske inte är något som sticker ut som en AI algoritm kan hitta eller tvärtom om du har faktiskt har den utbildningen eller den arbetslivserfarenheten och du har varit hos den här delen av socialförvaltningen, de går vidare. Till exempel, det vet vi inte ännu. Men man skulle kunna tänka sig det, ja det var en liten parantes.
119.	LR	Det var jätteintressant, Det var nog allt vi hade här så vi kan pausa inspelningen. Alla frågor är tagna!

Appendix D

Transkriptionsprotokoll Malmö stad

Medverkande personer:
Cecilia Minder (CM)
Leo Rasmusson (LR)
Magnus Perman (MP)
Datum och tid: 2022-05-02 13:00-14:00

Ytterligare information:
Då Magnus Perman hade tystnadsplikt mot ett av företagen som de var konsulter hos har företagsnamnet blivit utbytt mot "Alfa" i transkriptionen nedan. Dessutom har namnet på företagets kundklubb, som blir nämnt en del, blivit utbytt mot "kundklubb".

#	Person	Fråga/Svar
1.	CM	Ja, då ska vi se, så att vi har det, vad har du för roll på Nexer?
2.	MP	Ja, min roll på Nexer är Data science tech lead och tomorrow pilot, vilket kanske är lite svårt att veta om man inte jobba på Nexer men det är typ kompetensområdesansvarig på Nexer då. Så jag jobbar mot olika kunder inom retail och media och industri just inom AI och data science. Sedan har jag internt inom Nexer då rollen av att vara lite ambassadör inom data science. Vilket innebär hjälpa till med rekryteringar och prata på lite olika konferenser och hålla i lite olika månadsmöten också vidare för gruppen av data scientist som jobbar på Nexer då i region Syd för min del
3.	CM	Jättebra, och hur länge har du arbetat inom området av data science?
4.	MP	Jo men det är intressant, jag började ju på nexer eller som Sigma som det hette 2003. Det är ju rätt längesedan nu men då började jag som Javautvecklare sedan rätt snabbt snubblade jag in på det spåret som på den tiden, ja det heter väl det fortfarande, men business intelligence 2006. Så jag var med och byggde upp Alfas globala data warehouse Alfas kundklubb när den rullades ut mellan 2006-2012 så jag var lösningsarkitekt för det liksom, Då var det on premise stora data warehouse lösningar i oracle också vidare. Alfas kundklubb fanns i 5 länder 2006 och så skulle man rulla ut det till 25 länder inom ett par år. Så jag var med och byggde upp den lösningen för att samla all kunddata runt Alfas kundklubb kunddata. Den vägen kom jag in på själva business intelligence eller analytics spåret. På den tiden var det mer, handlade mycket om olika KPI rapporter, business objects, olika verktyg inom det. Det handlade mer då om att titta i backspegeln om

		hur många kunder har vi, hur många kunder fick vi föregående månad per land också vidare. Ren historiskt data så det var ju den business intelligence. Sen när jag kom tillbaks i min andra vända till Alfa 2016 då ville man börja använda datan lite mer avancerat alltså mer i just data science för att kunna använda all den data man hade samlat in runt 200 miljoner kunder, alla kvitton, all webbhistorik också vidare och hur kan man använda allt det för att titta framåt och inte bara i backspegeln. Vad tror vi kommer hända med kunderna, vilka tror vi kommer lämna, vilka tror vi kommer handla för mycket nästa år. Man kan se mönster i historiken för att titta framåt. Då 2016 ville Alfa börja mer på det sättet. På den vägen kom jag in på data science och fick ta fram en data science lösning för Alfa. Fick kolla en del på customer science engagement. Alfa är ju enormt och det finns en mängd olika områden på Alfa men just det här handlar om liksom om marknadsföring och customer engagements, liksom Alfas kundklubb engagements så att säga, reklam och analys av kundbasen. Sedan har jobbat inom det sedan 2016.
5.	CM	Intressant, ja men då fick vi lite bakgrund eftersom du gick rätt långt tillbaka också. Men har du någon akademisk utbildning?
6.	MP	Ja jag är ju civilingenjör i datateknik från början från LTH. Så jag gick ut där 20.. ja nu känns det ju som en evighet sedan men 2002. Men sedan 2003 började jag på Nexer som javautvecklare från början men halkade rätt snabbt in på analytics spåret då. Men på den tiden fanns det inga machine learning utbildning eller det var inte en del av av det man kunde läsa inom civilingenjör men man läste statistik och man läste programmering och sedan har jag domänkunskapen inom retail nu och de tre tillsammans, brukar man säga att är man bra på dem tre så har man en bra grund för data science.
7.	CM	Ja bra, och egentligen, jag tycker ändå du har nuddat lite vad du har för ansvarsområde och de du jobbar med så jag ställer den underfrågan om du kommer i kontakt med och hanterar personlig data?
8.	MP	Jo men det gör jag ju, vi har ju hand om hela Alfa's kundklubb kunder och kunddata inom Alfa då vilket är enorma mängder, alltså många miljoner kunder ungefär och all profildata runt om kunderna då. Så ja det kommer jag i kontakt med och jag sitter ju som arkitekt för en data science lösning inom customer engagements som egentligen handlar om att man egentligen ger tillgång till all den här datan till Alfas olika marknadsavdelningar i de olika länderna då. Så till exempel tyskland då har ju sin egen organisation inom Alfa och där har man data scientist som sitter och jobbar på den tyska datan för att ta fram vilka kampanjer de ska köra och i USA har de en annan organisation. Men allt bygger på samma datamodell kan man säga så jag har hand om det systemet som på något sätt gör den datan tillgänglig för alla länderna till sina organisationer och globalt också, vissa funktioner har vi globalt som

		sitter och körs på alla länder på samma gång.
9.	CM	Okej, tack jättebra, då går vi vidare, ja hur skulle du definiera AI och maskininlärning?
10.	MP	Ja AI skulle jag säga är att om man försöker få något att snurra av sig själv om vi säger machine learning och data science handlar väl mer om att ta fram modellerna och algoritmerna som kan ta fram nya insikter så är AI att sätta någon typ av, vad ska man säga, handlingar ovanpå det. Man ser en självkörande bil så är maskininlärningen och hela den biten att förutse vad som kommer hända om inte bilen svänger om 10 meter medan AI är mer att sätta en algoritm ovanpå det som utger själva handlingarna och tolkar maskininlärningen och sätter handlingar ovanpå det för att får någon sorts autonom alltså självlärande, att den tar över en människa jobb i det fallet.
11.	CM	Ja men jättebra och på vilket sätt använder ni AI på er data?
12.	MP	Ja det är ju att försöka se mönster i stora mängder data och att kunna räkna på det, ja man har ju all historisk data sedan många år tillbaka, man kanske dock inte använder data från så många år tillbaks i tiden men om man ser mönster på det som har hänt tidigare just runt kundbeteende också vidare. Så kan på något sätt räkna ut var, med en viss sannolikhet vad kunderna kommer att vilja har för erbjudanden och vad som är intressant för dem i nästa steg och sen är det ju AI är ju ett flashigt begrepp men det handlar mer om att skapa insikter och det är inte alltid man sätter det här i en kontext där det liksom blir något självkörande marknadsföringssystem utan det kanske fortfarande kommer vara människor som tar liksom actions på de insikterna och skickar ut kampanjer men det beror lite på var man sätter gränsen mellan AI och machine learning. På något sätt att tolka mönster i stora mängder data och skapa och titta lite mer framåt i tiden.
13.	CM	Jättebra, Tack, Och om vi ska introducera etik, hur ser du då på förhållandet mellan AI och etik?
14.	MP	Jo men det är ju ett intressant område man läser och hör väldigt mycket om det och just i de fallen där jag jobbar så försöker man vara så transparent som möjligt mot kunderna. Vad har vi och vilken data kommer vi spara från kunderna och vad kommer vi göra med datan, Sedan har man rullat ut väldigt många lösningar där man liksom synliggör det för själva kunden. Om man kollar på en webbsida så har man ju det klassiska att du trycker på en sådan här cookie consent. vill man läsa texten om det så är det en lång utläggning på flera sidor som man kanske inte orkar läsa igenom men att man försöker istället för det då gör man det så tydligt som möjligt med olika sliders. Om vi får lov att spara din köphistorik kommer vi kunna ge dig personliga rekommendationer om vad tror kommer vara nästa intressanta produkt för dig också vidare. Så kan man då slå av och på dessa sliders pers

		funktion. Det är väl den typen av etik jag har varit ganska involverad i att jobba med då.
15.	CM	Och skulle du påstå att etik påverkar ditt arbete?
16.	MP	Aa men absolut men där jag sitter inom reklam och marknadsföring så är det ju inte några livsavgörande beslut för kunden men om man ska få rekommendation för en produkt eller någon annan men kan tänka mig att det finns andra områden inom bank och försäkring osv då där det kanske är lite mer känsliga typer av erbjudande då och vad man får erbjuda kunderna men absolut till viss del handlar ju etik om all den datan som sparas kring kunderna då och på något sätt göra det transparent och få kunderna medvetna om att de har ett val och kan också att välja att inte i det här fallet då med Alfa sparar deras data och vad det då får för konsekvenser då, etik handlar väl om att göra kunderna medvetna om vilken typ av data vi spara för dem och vad det är. Både fördelar och nackdelar med det för kunderna, såklart försöker vi i vårt fall lyfta fram fördelarna men att göra det så enkelt som möjligt att stänga av om de vill.
17.	CM	Jag hade en underfråga och jag tänker att vi tar den också. Det är om du i ditt arbete känner av yttre faktorer som är kopplade till etik som påverkar ditt arbete och det kan vara allt ifrån lite politik till folkopinion, men ja allt kring yttre faktorer?
18.	MP	Ja men det är väl yttre faktorer är väl hårddraget lagstiftning som kan se rätt olika ut från olika länder och där försöker man då oftast följa de som har strängast möjliga stiftning för att just med det här med cloudact som har varit mycket på tapeten nu med att om man sparar europeisk kunddata i Usa som i tex google analytics men att man följer det liksom och där kommer man i Usa med lösningar nu och man får lov att utbyta den datan så att alla länder är medvetna och kan godkänna så att sådana yttre faktorer är man såklart väldigt påverkad av och även om just om man kolla från ett kundperspektiv och liksom företag så räcker det att man har 10 eller 100 väldigt missnöjda kunder i ett land och i och med dagens teknik, facebook, och twitter och sådant kan anmäla dig till domstolar i varje land så kan det bli en stor grej även om det bara är ett fåtal kunder som är påverkade av till exempel om man har datakvalitetsproblem eller om de har råkat få se något som de inte borde få se eller om de var missnöjda med erbjudande också vidare.
19.	LR	Jag har bara en lite följdfråga där men du nämner att det var att vissa länder hade lite strängare lagar gällande det här, är det vanligt att ni har eller generellt sett ett arbetssätt där ni generellt sett utgår från det mest stränga landet och sedan applicerar eller generaliserar det på resten av världen?
20.	MP	Jo men exakt vi försöker ju hålla samma lösning för alla länder för annars blir det ohållbart att ha 29 olika lösningar på hur man sparar kunddata. Men absolut vi försöker förhålla oss på samma sätt och

		använder den strängaste möjliga tolkningen. Om man kollar på kina och vissa andra länder så tillåter inte de att man ens sparar kunddata utanför Kina så då får man göra en speciallösning. Då använder man ett annat ekosystem helt enkelt med lösningar som ligger i Kina.
21.	CM	Tack. Vilka etablerade riktlinjer inom AI-etik är du personligen medveten om?
22.	MP	Ja nu det är väl lite det vi snackat om, just GDPR men det är inte just etik iförsig utan mer lagar då men just transparens för kunderna gör det möjligt för dem på ett enkelt sätt att gå in och ur. Samt vilket typ av data man sparar runt om och även får dem att förstå vad det betyder om de inte går med på det. Att få dem att se både fördelarna med det hela men även göra det enkelt att hoppa ur. Sen är ju jag ju medveten om modeller har en viss typ av bias om du tränar upp en modell på en för snäv population så kan det ju bli lite fel för du ska inte liksom bara välja ut män över 40 eller 30 och sedan träna en modell på vad de har köpt för produkter. Det säger väl sig själv litegrann men ja men var liksom medveten om att om man tränar upp modeller att använda sig av så stort dataset som möjligt som representerar en hel population. Och ja som sagt Alfa i det fallet där jag jobbar sparar man ju inte känslig data för kunder. Det känsligaste man sparar är vilken adress man bor på ner på gatunivå och det är ändå inte tillgängligt för den stora massan av analytiker utan det är bara en liten, liten andel. Så vi sparar lite kring vilken typ av kön och ålder kunderna har och vart i landet man bor och dessutom köphistorik. Hur man har surfat runt på olika webbsidor och appar som företaget tillhandahåller.
23.	CM	Ja, tack, Du har ju nämnt GDPR, sedan kanske ni har andra tillvägagångssätt för att uppnå transparens. men följer ni liksom inom ja när du jobbar mot kund i ditt fall Alfa eller inom Nexer, några etablerade riktlinjer som du känner till?
24.	MP	När inga jag känner till på det sätter alltså riktlinjer runt just etik och moral. Det är mer dem lagliga kraven sen kanske det finns andra på dem företagen jag jobbar som sitter mer med de här etikreglerna också vidare. Nexer i sig följer kundernas och försöker vara rådgivande ute där vi sitter i uppdrag. Jag jobbar ju inte med data science liksom på Nexer som företag utan det är alltid ut mot Nexers kunder inom det här området.
25.	CM	Ja men skulle du påstå att är i ditt område nödvändigt att följa etiska riktlinjer?
26.	MP	Jo men det tycker jag att det är , att man är försiktig med kunddata om vem som har access till det, tänk alltid lite mer vad som kan hända om kunddata hamnar i fel händer och även man får försöka se det från en kunds perspektiv också. Jag kanske inte själv inte är så känslig vad företag sparar just på mig personligen men det finns många andra som

		tycker det kan vara obehagligt när företag börjar räkna ut vilken typ av livssituation kanske man befinner sig i baserat på vad man har handlar för olika produkter också vidare. Folk vill inte ha den typen av att analytikerdatasetföretag liksom räknar ut det då med hjälp av data. Så att det förstår jag då att det kan vara känsliga ämnen.
27.	CM	Ja, och du som ändå jobbar på då liksom en arkitektur nivå om man kan säga så, de riktlinjer som du har kanske då, eller så här upplever du att det finns en hög abstraktionsnivå till dessa riktlinjer? Det kan ju vara mer..
28.	LR	Du kan ju koppla det till GDPR då antar jag
29.	CM	Då som vi beskrivit då alltså lite svårare att applicera i praktiken och man kan tolka det lite som man vill, men till exempel då till GDPR.
30.	MP	Ja men precis. Man försöker ju alltid få ner helt konkret vad man får göra med data som en analytiker. Då går vi alltid igenom när jag jobbar med de som blir använda av det systemet. Man måste kolla dels vad man har för olika consent alltså vad man som kund gått med på, det måste man ha med i alla modeller och alla uträkningar, all marknadsföring som skickas ut så har man det som ett obligatoriskt krav. Så där finns ju en ren koppling till GDPR och vad kunden själv har sagt och vad som är lagligt att göra med datan. Sedan försöker vi alltid eller alltid där jag utvecklar försöker vi alltid ha spårbarhet. Så om kunden nu undrar varför den fick reklam för just den här produkten eller varför fick jag den här reklamen som verkar vara riktad på det här sättet så ska man alltid kunna gå tillbaks till modellen och så vilka olika faktorer har gjort att just den här modellen predikterat på just det här sättet för den här kunden. Man måste kunna gå ner för kund till kund för att kunna se vilka variabler det var som avgjorde och sedan kanske också kunna förklara för kunden då i värsta fall om det är någon som tar kontakt med Alfa i det här fallet och vill exakt veta vilken algoritm det var som valde just det här erbjudandet. Nu kanske inte det är så vanligt då för just Alfa då men man försöker de riktlinjer för spårbarhet och transparens som finns och lagliga krav.
31.	CM	Då kan man nästan liksom säga att du upplever att de används i praktiken?
32.	MP	Jo men absolut.
33.	CM	Ja nu blir det en fråga kopplad till Nexer. Men du kanske kan koppla den till Alfa också och det är om det finns en uppförandekod eller policy i förhållande till etik som ni följer?
34.	MP	Ja på Nexer skulle jag säga att där jobbar vi mot kunderna så vi lyder deras modeller eller koder. Men i Alfa fallet vet jag inte riktigt om det finns en uppförandepolicy runt etik men det är ett väldigt viktigt ämne

		just, etik, kunddata vad vi gör med den och vad vi får lov att göra med den och transparens också vidare. Det kanske inte just finns någon nedskrivet, jag har inte sett något nedskrivet iallafall men allting kokar ju ner till vad man får lov att göra med kunddatan och att det ska se bra ut för kunderna och att de ska vara medvetna om vad man gör med datan. Men jag faktiskt inte sett något nedskrivet på det sättet.
35.	CM	Du nämnde lite tidigare just när ni pratade om etik och ert förhållande till kunder och de ska gärna få avsäga sig det de liksom inte vill ta del av och jag tänkte att vi bara kunde understryka vad din personliga definition av etik är?
36.	MP	Ja men det är ju liksom så att man får tillgång till enormt stora mängder data och man får tänka sig in i den individuella kundens och om man ser sig själv som kund och får man då dessutom tänka att det finns kunder som har lite annat tankesätt och om man har kunder som tycker det är obehagligt också vidare och att man försöker se det från det håller och inte bara stirra sig blind på data och vad man får göra med hjälp av data. Men att man ska ju, jag kan tycka att det underlättar väldigt mycket liksom det vi gör med data och det kan skapa bra erbjudanden och skapa relevant kommunikation för kunder också men att man får tänka på att man ska se saker från kundens perspektiv också. Så det är väl min uppfattning av etik. Man får tänka sig in lite i mottagaren av det man skapar.
37.	CM	Jättebra, tack, då ska vi se, Känner du någon på din arbetsplats eller andra arbetsplatser som arbetar med maskininlärning-etik eller AI-etik?
38.	MP	Jo men det kan jag ju säga rent att det finns ju andra på i det här fallet Alfa som jobbar mycket på att ta fram riktlinjer kring etik och att följa de lagliga kraven men även att det finns. Sedan känner jag inte dem personligen men jag känner till att funktionen finns, man jobbar väldigt mycket med det här de senaste åren med just att utveckla appar och webbsidor som ska vara som jag sa innan transparenta och som ska göra det enkelt för kunden. Jag känner som sagt inte dem personlig men känner till att de finns många som jobbar inom det hos kund.
39.	LR	Okej, då går vi vidare. Då tar jag över lite här.
40.	MP	Ja, ja.
41.	LR	Då, jag tycker han svarat på den. Men ja, är mikroetik någonting du har hört talas om?
42.	MP	Nej, faktiskt inte. Det har jag faktiskt inte hört talas om. Så det är intressant.

43.	LR	Vi kan dra en snabb på den. Ja men mikroetik är i alla fall en ny metodik som har kommit, som uppkommit nu de senaste två åren ungefär, som handlar helt enkelt om att man ska göra om då etiska policys och riktlinjer till att bli mer branschspecifika och mer tekniskt implementerbara och försöka liksom, jag vet inte koka ner det här till någonting som skulle kunna gå att implementera på en professionell arbetsplats. Helt enkelt, det är nog den snabba förklaringen.
44.	CM	Ja det är verkligen, det är i namnet, mikroetik, så att ja…
45.	MP	Ja, koka ner det på branschnivå och så vidare. Något som jag jobbat ganska mycket eller och har kommit i kontakt med är något som kallas för explainable AI.
46.	LR	Ja.
47.	MP	Och jag vet inte, det har ni säkert också hört talas om, men just att kunna förklara varför en modell gör på ett visst sätt och ja, och kunna följa liksom stegen, nu kanske enkelt i rena linjära modeller där liksom där man har ett linjärt samband mellan olika variabler och så vidare, då kanske klarar att ja men, på grund av de här olika variablerna hade det, men om du kollar på en deep learning modell liksom som ett neuralt nätverk med 50 olika steg och så kan det va, men att man även där liksom kan ha nån typ av att man kan koka ner det till nån typ av explainable AI.
48.	LR	Ja men verkligen.
49.	MP	Som ni säkert har hört talas om rätt mycket. Ja så att, det är ett intressant område.
50.	LR	Amen det är väl ett bra sätt att så att man undviker att AI blir en så himla black-boxat som det…
51.	MP	Precis
52.	LR	Intressant!
53.	MP	Ja och det kan både vara för de som är mottagare liksom men även de som utvecklar algoritmerna, så att de förstår sig på vad de gör liksom.

54.	LR	Ja men exakt. Yes, ska vi se, så då går vi in lite mer på det här med mikroetik då, men vet du om det finns några tekniska instruktioner då inom er organisation eller på er arbetsplats eller mer generellt inom fältet AI, för hur man då ska implementera till exempel etisk AI då.
55.	MP	Eh, instruktioner, inte på, alltså det är väl mer att, att det finns en funktion som jobbar med det som jag känner till, och de är ganska synliga och just på Alfa fallet så har man hela, att man har tatt etik liksom hela vägen upp till till alltså ledningsgruppen genom att man har en CDO som sitter med i Alfas ledningsgrupp som jobbar mycket med digitalt och hon, hon nu har hon slutat men hon drev väldigt mycket just det här med transparens och ja, och hela den biten. Men sen att jag inte känner till exakt tekniska instruktioner det, nej, det är väl mer att,ja, nej jag har faktiskt inte sett dem. Att, att i och med att man har implementerat det här liksom top-down och det har kommit hela vägen ner liksom till alla som jobbar med data i princip, så ja alla, har alla det i sina tankar liksom och hur de kan jobba, just med etik. Och så vidare. Men sen har jag inte sett några mer nedskrivna exakta instruktioner.
56.	LR	Alright, tycker du vi ska ta den eller?
57.	CM	Ja.
58.	LR	Då är det en liten lång fråga här, men, vad skulle du då säga om det till exempel fanns standardiserade dataset, som är då pretrained, eller med pretrained modeller, som har helt enkelt tydliga innehållsförteckningar då framförallt kopplat till hur datan är insamlad, och vad det finns för intention bakom datan, just kretsat kring transparens då, är det något som du känner till finns i branschen eller nånting som du tror skulle vara användbart?
59.	MP	Ja alltså, jag kan tänka mig att det kan vara användbart just, just att man har liksom färdiga dataset som man har, och det kan även vara användbart kanske inom, om vi nu kollar på Alfa, att man tittar om på samma grej om och om igen utan att man har liksom ett bra dataset som vi använder för att träna upp en viss algoritm och då kanske, kanske nån annan del av Alfa kan kan använda det datasetet för att träna upp om de vill göra någonting annat liksom, eh, så att definierade dataset är ju bra sen kan det ju hända väldigt mycket i världen om man kollar liksom med Covid-19 och så vidare, när alla köpmönster helt enkelt ändras väldigt snabbt alltså från en vecka till en annan så handlar alla online. Då får man ju liksom vara snabb på att och publicera nya såna här dataset också, i så fall. Men absolut, jag kan se liksom nyttan i det, att man liksom har bra datasets som man kan använda till att träna upp,

		men sen är det lite branschspecifikt att just i vissa branscher så kollar man på lite olika variabler liksom om man kollar en retail hade man ju behövt de variablerna som är viktiga där och där kanske man inte vill dela med sig av sin kunddata mellan liksom konkurrerande företag…
60	LR	Det är sant
61.	MP	Men, men andra andra för andra branscher så liksom kanske det är lite enklare att dela data liksom.
62.	LR	Så inom retail skulle det då bli mer företagsspecifikt egentligen.
63.	MP	Ja, jag tror det. Liksom, i och med att man kollar ju på just om man handlar produkter på Alfa och vi vill se en viss typ av produkter så blir det ganska branschspecifikt liksom då om man tittar på kundbeteende och så vidare, jämfört med om man kollar folk som köper bilar, kanske, vad det nu kan vara, så ja, det blir ganska branschspecifikt tror jag just inom, där jag sitter i alla fall. Jag kan ju se nyttan av att ha liksom, färdigcheckade och kollade datasets för andra typ av appliceringar.
64.	LR	Okej. Ska vi se, nu var det ju ett tag sen du, vad ska man säga, gick klart din universitetsutbildning men minns du om AI-etik just eller ML etik var en del av den utbildningen?
65.	MP	Nej det var det ju inte faktiskt, man kunde ju inte plugga det på den tiden liksom, utan då var det statistik var ett ämne liksom och sen var programmering ett annat ämne och det fanns inget ämne som kopplade ihop de två till data science eller machine learning, så nej, så på den tiden fanns det ju inte, men jag sitter med i ledningsgruppen för en annan utbildning, just för data scientist, alltså EC-utbildningen, den här yrkes, och där vet jag att det är uppe på tapeten att ta med etik och moral liksom inom AI och så vidare att man har ett liksom, att man har det i kursplanen, och det kan ju plugga…
66.	LR	Förlåt, var det yrkesutbildningen i Malmö eller är den?
67.	MP	Njae Malmö Helsingborg, EC. Så jag vet att de inte har de som en egen kurs, alltså inte som ett eget ämne men att man bäddar in det i de andra kurserna.

68.	CM	Ja det är ju jätteintressant för vi har ju inte haft, nu har ju inte vi, nu tillhör ju vi Ekonomihögskolan så att vi har ju inte eh, vi har haft en del ja men mer programmering men inte alls etik…
69.	LR	Nej inte alls.
70.	MP	Nej, men man ser lite alltså att tekniken har kommit före ja vad ska man säga etiken men även i alla fall före lagarna och så vidare, att man försöker komma ikapp i många fall med lag, alltså lagar och så vidare runt i och just etikbiten har väl också kommit lite efter, ja alltså efter själva tekniken när man ser att då att det kanske har blivit fel i vissa fall, eh, och där man ja helt enkelt AI:n har inte gjort det som, har inte blivit bra i vissa fall, då har man så att säga vissa konstiga konsekvenser av det hela och då försöker man komma ikapp liksom med, både etiken och lagstiftning så att säga.
71.	LR	Ja men verkligen. Ska vi se, men inom Alfa till exempel har ni fått någon utbildning där om kring etik eller, hur ska man säga, etik till AI, eller så där?
72.	MP	Nej, nej faktiskt inte. Inte vad jag, utan det har väl mer varit när de har varit stora drives liksom runt GDPR för några år sedan, ja nu är det rätt många år sen också, men när den kom ut liksom själva eh att man att det skulle föras ut med höga böter för alla som inte följde det och så vidare och mycket fokus på det på Alfa så att man liksom följde det väldigt bra, men just inom AI-etik, nej inga utbildningar som jag gått, jag är självlärd.
73.	LR	Yes, eh, vi har hittat mycket forskning i vår research helt enkelt som pekar på att det är många företag som använder just AI, eller etik kretsat kopplat till AI endast i marknadsföringssyfte, vad säger du om det här påståendet?
74.	MP	Förlåt, om man använder etik…
75.	LR	Etik, eller man använder AI-etik bara för att eller bara i marknadsföringssyfte, så ut mot publik ja, vad ska man säga, ja ut mot samhället.
76.	MP	Ah okej, så att man säger utåt “att vi är väldigt liksom…”

77.	CM	Tar ett etiskt ansvar.
78.	LR	Exakt.
79.	MP	Men sen bakom kulisserna så gör man lite annorlunda
80.	LR	Ja eller hur, exakt.
81.	MP	Ah så om jag har sett det? Eller vad jag säger om det?
82.	LR	Ja vad du säger om det?
83.	MP	Ja men jag tycker att det inte är snyggt att göra på det sättet att liksom att hålla en fasad utåt och en annan, att göra på ett annat sätt för jag tror tillslut kommer det ändå ut på ett eller annat sätt om man inte lever som man lär så att säga. Även folk som jobbar där eller att kunder liksom lyckas lista ut att det liksom, att det inte funkar på det sättet som de säger, så att och i dagens samhälle liksom så är det så lätt att sprida ett missnöje liksom eller att få publicitet runt nånting så kan ju företag hamna, de kan ta stor skada liksom, både legalt och liksom i kundmissnöje om sånt kommer ut så det har man allt att förlora på. Men det tycker jag är knasigt liksom och det är därigenom, där jag sitter på Alfa har man försökt få ut hela vägen fram till kunden liksom i front-end systemen så har du de här slidsen eller vad man ska säga som då gör att du kan, ja du kan acceptera eller inte på olika sätt och det är såklart följer hela vägen ner till datan. Så att ja.
84.	LR	Tack. Vem tycker du etikansvaret ska falla på inom en organisation?
85.	MP	Nej men jag tycker liksom det ska, vad ska man säga, det ska vara de som gör skickar ut, implementerar själva tekniken måste vara medvetna och så egentligen, det ska ju hela vägen ner längs organisationen ner till utvecklarna och modellerna så annars blir det ju lite som du sa att det kanske blir att man man har nån typ av organisation som jobbar mycket med att ta fram flashiga eller vad man ska säga liksom ja, påstående och så vidare, men ja, det måste ju vara de som jobbar med själva utvecklingen som är medvetna om etiken också annars blir det svårt att genomföra det.
86.	CM	Tänker du det att de, alltså just de, amen ska det vara typ att komma, ska det vara riktlinjer som de ska följa liksom från ett annat håll eller

		hur tänker du då att alltså hur ska de bli medvetna om den, det etikansvaret?
87.	MP	Ja men det, man för det första måste man liksom bestämma lite vad får vi lov att göra med datan, att vad hur ska vi, nu är jag mycket inom retail och kund liksom, men hur ska vi visa mot kunderna, hur ska vi låta kunderna själv får bestämma osv att det finns implementerat liksom och då måste det följa hela vägen ner till de som utvecklar modellerna att de verkligen använder sig av datan på rätt sätt och i det fallet där jag sitter, så har vi liksom stenkoll på vad man får lov att göra med datan och vilken typ av data man får lov att använda i vissa fall och så vidare på grund av att det finns vissa, vad ska man säga variabler i datan som säger till liksom vad, vad har den här kunden gett medgivande till och vad har den här kunde inte, ja, så att man ser det mer liksom på ja, mer på varje individ i princip. Så att det är väl mer i varje själva, går mer på själva datamodellen så måste ju ja, det måste ju finnas spårbarhet av det som är bestämt hela vägen ner till datan, så man ser vad man får lov att göra på ett väldigt enkelt sätt.
88.	LR	Okej, ska vi se, får ni på Nexer eller när du är ute hos kunder, har du varit med om att man får stöd i form pengar, utbildning, marknadsföring för innovationsarbete av tex Vinnova eller liknande institutioner?
89.	MP	Eh, ja det är en bra fråga, jag vet att, jag har inte själv varit inblandad i det men jag vet att, ja nu var det rätt länge sen, när vi hette Sigma så hade vi nåt Vinnova projekt, men jag kommer faktiskt inte ihåg vad det handlade om, tror det var något med grön IT eller, ja men det får jag lov att vara osagt, men jag har inte själv kommit i kontakt med innovation, alltså att man tar del av innovationspengar på det sättet, för att utveckla etik.
90.	LR	Okej, yes, vi har då hittat lite forskning som visar på att det kan vara gynnsamt i när man utvecklar AI algoritmer att ha en white/black list då, som är helt enkelt regler som systemet alltid ska följa eller beteenden systemet aldrig får finnas i. Hur ser du på detta?
91.	MP	Hur menar ni med, vilket typ av system då? Eller generellt…
92.	LR	Ah, alltså en AI algoritm helt enkelt att den är utvecklad på så sätt att den har, eller att man har listor helt enkelt som systemet, eller med regler som systemet alltid ska följa och beteenden de aldrig får finnas i.
93.	MP	Ah okej, ja men det är väl, det låter vettigt, jag själv är inte liksom så involverad i den typen av utveckling där man där liksom där man skulle

		vilja liksom begränsa, för det finns alltid eller i många fall, jag ska säga att i 99% har man en människa i mitten som gör kampanjerna och som är medveten så det blir inget självkörande på det sättet liksom men ja du menar om att det ska finnas vissa grejjer av AI:n inte får göra att skicka ut något som…
94	LR	Ja men exakt det är nog, tanken är nog lite mer åt autonomt system.
95.	MP	Ja att den skulle helt ta, ja men jag vet open AI där man kan få tillgång till olika sådana här transformationsmodeller som är rätt avancerade så kan man själv styra modellen till att att om den ska vara sarkastisk eller liksom snäll eller så vidare, och där hade man inte kanske velat att ha den i just chattbotar och så vidare att som börjar svara med taskiga svar i kundsupport och så vidare. Det låter som en vettig idé att ha nån typ av sån white label och att man liksom jobbar på det sättet att man alltid vill avgränsa en modell till att den inte får ta vissa typer av beteenden liksom, ja.
96.	LR	Ska vi se, ett förslag på att för att få en organisation är att blir mer etisk är att bädda in då etisk reflektion i en utvecklares dagliga arbete och det är helt enkelt att försöka normalisera etik på arbetsplatsen och öppna upp för diskussion som berör typ open-source code eller dataset då, hur ser du på det här?
97.	MP	Jo men jag tror det är vettigt, lite som jag sa där innan att verkligen att man i just etik och moral och hela den biten att det kommer ner till alla som jobbar liksom med i princip, de utvecklarna för det är i princip där vissa beslut kommer tas liksom för hur modellerna ska tweakas och ställas in och vilka data man får jobba med och hur man kan jobba med den och så vidare så absolut. Det tror jag hade varit vettigt att det finns hela vägen, genom en organisation.
98.	LR	Ja för upplever du att det finns en så pass mycket, vad ska man säga, makt eller om man säger så, just hos utvecklarna när det gäller, vad ska man säga, etikansvaret då kanske?
99.	MP	Njae, men jag tror faktiskt att där är makt liksom hos de som sitter och utvecklar de att man kanske inte själv, jag tror det är i många fall man kanske inte är medveten om den makt man sitter på om man har utvecklat en modell som liksom ska göra en viss grej, då kan det ju va väldigt vettig att ha gått igenom någon typ av etik, ja etikutbildning, vad det kan få för konsekvenser om man använder data på fel sätt liksom, ja. Så att där jag sitter så tycker jag att, där känns det som att alla har liksom stenkoll, eftersom det har varit så mycket snack om det liksom att. Men ja det är liksom inrutat i organisationen på nått sätt,

		tror jag. Men på många andra företag tror jag att det hade varit, ja rent generellt tror jag det hade varit bra liksom om man kunde säga att man har den typen av utbildning för utvecklarna också, absolut.
100.	LR	Inom testning och validering av en AI-lösning finns det förslag, framförallt kopplat lite till mer vad ska man säga mer livsavgörande AI-lösningar att sjösätta ett team helt enkelt som jobbar enbart med att försöka ta sönder eller hitta hål i lösningen. Vad tror du om nåt sånt här?
101.	MP	Ja men det tror jag är viktigt, för oftast finns det liksom i just om man kollar på Cybersecurity och Website och så vidare så har man oftast en penetrations, alltså pen-tester, folk som sitter och försöker jobba med att hitta svagheter i lösningen, hitta kryphål, sätt att komma in, på samma sätt så tycker jag att man ska jobba med AI modeller försöka stressa dem modellerna och försöka se om det finns ett sätt att lura de och det vinner väl alla på om man kan få till den typen av testning, så absolut. Det tror jag bara är bra.
102.	LR	Yes, och så sista frågan. Vad tycker du är de största utmaningarna med ert AI-arbete?
103.	MP	Ja men det är en intressant fråga. Ja men just att på nåt sätt skapa relevanta och bra lösningar för kunder, kunderna,och där jag sitter som sagt så blir det inte jätte etik och moral det handlar mer om att göra det synligt för kunder om vad man gör och vad man har deras data för och vad det kan ge för positiva effekter och så vidare, men nu hade jag glömt frågan…
104.	LR	Ja det är mer generellt vad de största utmaningarna med ert AI-arbete är.
105.	MP	Ja nej men det är, det händer väldigt mycket inom branschen och det går snabbt framåt liksom och att kunna använda senaste tekniken men att se lite kritiskt på den också liksom, att i alla fall kanske att man inte behöver använda de allra flashigaste och nyaste algoritmerna, att man kan använda enkla modeller också som kanske går bra, och ger samma resultat. Så det är väl det, ja att hålla sig, vad ska man säga i spetskompetens men ändå vara lite kritisk till allt nytt som kommer ut, kanske i alla fall att man inte behöver använda det allra senaste liksom. Sen är etik och moral liksom ja, hänga med och se vad modellerna och det här black-box tänkandet lite hur man kan komma ifrån det och hur man på nått sätt kan förklara vad modellerna gör för gemene man och att man gör det liksom, ja, göra det enkelt greppbart för alla som vill

		veta vad modellerna gör, inte bara för utvecklarna, så att, ja det kan vara lite tröskel att ta sig över, på nåt sätt förklara det för vem som helst.
106.	LR	Alright, ja men det var nog allt vi hade här.
107.	CM	Ja det var det, det var alla frågor.
108.	LR	Så vi kan pausa vår inspelning här.

Referenser

Association of Computing Machinery. (2018a). ACM Code of Ethics and Professional Conduct [pdf], Available at: https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-and-professional-conduct.pdf [Accessed 4 April 2022]

Association of Computing Machinery. (2018b). ACM Updates Code of Ethics, Available at: https://www.acm.org/articles/bulletins/2018/july/new-code-of-ethics-released [Accessed 5 April 2022]

Assure, N. & Rowshankish, K. (2022) The data-driven enterprise of 2025, web blog post, Available at: https://www.mckinsey.com/business-functions/quantumblack/our-insights/the-data-driven-enterprise-of-2025 [Accessed 2 May 2022]

Bezuidenhout, L. & Ratti, E. (2021). What Does It Mean to Embed Ethics in Data Science? An Integrative Approach Based on Microethics and Virtues, AI & SOCIETY, vol. 36, no. 3, pp 939-953, Available online: https://link.springer.com/content/pdf/10.1007/s00146-020-01112-w.pdf [Accessed 12 April 2022]

Bryman, A. (2018). Samhällsvetenskapliga Metoder. Stockholm: Liber

Datadrivet. (n.d) Alla pratar om datadrivet, Available online: https://www.datadrivet.ai/ [Accessed 2 May 2022]

Dignum, V. (2019). Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way [e-book] New York: Spring International Publishing AG, Available through: LUSEM University Library website https://www.lusem.lu.se/library [Accessed 13 April 2022]

European Commission. (2019). High-Level Expert Group on Artificial Intelligence [pdf], Available at: https://42.cx/wp-content/uploads/2020/04/AI-Definition-EU.pdf [Accessed 5 May 2022]

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U. & Rossi, F. (2018). Ai4people—an Ethical Framework for a Good Ai Society: Opportunities, Risks, Principles, and Recommendations, Minds and Machines, vol. 28, no. 4, pp 689-707, Available online: https://link.springer.com/content/pdf/10.1007/s11023-018-9482-5.pdf [Accessed 15 April 2022]

Floridi, L. (ed.). (2021). Ethics, Governance, and Policies in Artificial Intelligence, [e-book] Cham, SWITZERLAND: Springer International Publishing AG, Available online: LUSEM University Library website https://lubcat.lub.lu.se/cgi-bin/koha/opac-detail.pl?biblionumber=6954916 [Accessed 22 April 2022]

Hagendorff, T. (2020). The Ethics of Ai Ethics: An Evaluation of Guidelines, Minds and Machines, vol. 30, no. 1, pp 99-120, Available online: https://link.springer.com/content/pdf/10.1007/s11023-020-09517-8.pdf [Accessed 22 April 2022]

History. (2021). Deep Blue defeats Garry Kasparov in chess match, Available online: https://www.history.com/this-day-in-history/deep-blue-defeats-garry-kasparov-in-chess-match [Accessed 4 April 2022]

IBM. (2020). What is Machine Learning?, Available online: https://www.ibm.com/cloud/learn/machine-learning [Accessed 12 April 2022]

IEEE Computer Society. (n.d). John McCarthy Biography, Available at: https://www.computer.org/profiles/john-mccarthy [Accessed 5 April 2022]
Integritetsskyddsmyndigheten. (2021a). Vad kan tillsynen leda till?, Available online: https://www.imy.se/om-oss/vart-uppdrag/sa-arbetar-vi-med-tillsyn/vad-kan-tillsynen-leda-till/?fbclid=IwAR3ZrQOmIwfdF3ifTt2mafmAaZMHJYgKwDIQGRVPTe40EzMvPzcdhoe8f9Q [Accessed 9 May 2022]
Integritetsskyddsmyndigheten. (2021b). Det här gäller enligt dataskydds-förordningen, Available Online: https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/ [Accessed 9 May 2022]
Integritetsskyddsmyndigheten. (2021c). Syfte och tillämpningsområde, Available online: https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/introduktion-till-gdpr/syfte-och-tillampningar/ [Accessed 9 May 2022]
Integritetsskyddsmyndigheten. (2021d). Grundläggande principer, Available online: https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/grundlaggande-principer/ [Accessed 9 May 2022]
Isaak, J. & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection, Computer, vol. 51, no. 8, pp 56-59, Available online: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8436400 [Accessed 25 April 2022]
Jacobsen, D. I. (2002). Vad, Hur Och Varför: Om Metodval I Företagsekonomi Och Andra Samhällsvetenskapliga Ämnen, Lund: Studentlitteratur AB
Leavy, S., O'Sullivan, B. & Siapera, E. (2020). Data, Power and Bias in Artificial Intelligence, arXiv preprint arXiv:2008.07341. Available Online: https://arxiv.org/pdf/2008.07341.pdf [Accessed 2 May 2022]
McNamara, A., Smith, J. & Murphy-Hill, E. (2018). Does Acm's Code of Ethics Change Ethical Decision Making in Software Development? Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering, 2018. 729-733, Available online: https://dl.acm.org/doi/pdf/10.1145/3236024.3264833 [Accessed 8 April 2022]
Miller, S. (2019). Machine Learning, Ethics and Law, Australasian Journal of Information Systems, vol. 23, no. Available online: https://journal.acs.org.au/index.php/ajis/article/view/1893/851 [Accessed 22 April 2022]
Norvig, P., & Russel, S. (2022) Artificial intelligence: A Modern Approach, London: Pearson Education Limited
Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S. & Krasanakis, E. (2020). Bias in Data-Driven Artificial Intelligence Systems—an Introductory Survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 3, pp e1356. Available online: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1356 [Accessed 2 May 2022]
Oates, J. B. (2006) Researching Information Systems and Computing, London: SAGE Publications Ltd
Stanford Enclyclopedia of Philosophy. (2016a) Consequentialism, Available online: https://plato.stanford.edu/entries/consequentialism/ [Accessed 13 April 2022]
Stanford Enclyclopedia of Philosophy. (2016b) Deontological Ethics, Available online: https://plato.stanford.edu/entries/ethics-deontological/ [Accessed 13 April 2022]
Stanford Enclyclopedia of Philosophy. (2016c) Virtue Ethics, Available online: https://plato.stanford.edu/entries/ethics-virtue// [Accessed 13 April 2022]

The Quest for the Master Algorithm. (2016). YouTube video, added by TedxTalks [Online], Available at: https://www.youtube.com/watch?v=qIZ5PXLVZfo [Accessed 1 April 2022]

Sveriges riksdag. (2018). GDPR - införande och stöd till företag och företagare, Available online: https://www.riksdagen.se/sv/dokument-lagar/dokument/skriftlig-fraga/gdpr---inforande-och-stod-till-foretag-och_H5111155 [Accessed 9 May 2022]

Sveriges riksdag. (n.d). Om dataset, Available online: https://data.riksdagen.se/dokumentation/om-dataset/ [Accessed 27 April 2022]

Stair, R., Reynolds, G., & Chesney, T. (2018). Principles of Information Systems, Andover: Cengage Learning.

Watson, M. (2018) Why Business Leaders Should Think of AI as an Umbrella Term, web blog post, Available online: https://medium.com/opex-analytics/why-business-leaders-should-think-of-ai-as-an-umbrella-term-dba8badc55e4 [Accessed 6 April 2022]

Žliobaitė, I. & Custers, B. (2016). Using Sensitive Personal Data May Be Necessary for Avoiding Discrimination in Data-Driven Decision Models, Artificial Intelligence and Law, vol. 24, no. 2, pp 183-201. Available online: https://link.springer.com/content/pdf/10.1007/s10506-016-9182-5.pdf [Accessed 4 May 2022]