



CHALMERS



Etiskt hållbar AI för nyanserade nyhetsflöden

En studie av maskininlärningsalgoritmers
påverkan på nyhetsspridning

Kandidatarbete inom Industriell ekonomi

LOUISE ALBINSSON
ELIN BERTHAG
LUKAS ILOSON

KLARA JAKOBSSON
VIKTOR JOHANSSON
KERSTIN WADMAN

**INSTITUTIONEN FÖR TENIKENS EKONOMI OCH ORGANISATION
AVDELNINGEN FÖR TEKNIK, VETENSKAP OCH SAMHÄLLE**

CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2021
www.chalmers.se
Kandidatarbete TEKX04-21-21

Etiskt hållbar AI för nyanserade nyhetsflöden

En studie av maskininlärningsalgoritmers påverkan på
nyhetsspridning

Ethically sustainable AI for balanced news feeds

A study of the impact of machine learning algorithms on news dissemination

LOUISE ALBINSSON
ELIN BERTHAG
LUKAS ILOSON

KLARA JAKOBSSON
VIKTOR JOHANSSON
KERSTIN WADMAN

Etiskt hållbar AI för nyanserade nyhetsflöden
En studie av maskininlärningsalgoritmers påverkan på nyhetsspridning

LOUISE ALBINSSON	KLARA JAKOBSSON
ELIN BERTHAG	VIKTOR JOHANSSON
LUKAS ILOSON	KERSTIN WADMAN

© LOUISE ALBINSSON, 2021	© KLARA JAKOBSSON, 2021
© ELIN BERTHAG, 2021	© VIKTOR JOHANSSON, 2021
© LUKAS ILOSON, 2021	© KERSTIN WADMAN, 2021

Kandidatarbete TEKX04-21-21
Teknikens ekonomi och organisation
Chalmers tekniska högskola
412 96 Göteborg
Sverige
Telefon + 46 (0)31-772 1000

Göteborg, Sverige 2021
Gothenburg, Sweden 2021

Ethically sustainable AI for balanced news feeds

A study of the impact of machine learning algorithms on news dissemination

LOUISE ALBINSSON

KLARA JAKOBSSON

ELIN BERTHAG

VIKTOR JOHANSSON

LUKAS ILOSON

KERSTIN WADMAN

Department of Technology Management and Economics

Chalmers University of Technology

SUMMARY

Today, most of people's news consumption consists of absorbing news and information through digital media. News feeds on platforms that use machine learning algorithms to filter the reader's content are being used as news sources even though they often do not show the whole picture to the reader, but only what is in the reader's interest. Filtering causes many problems as the reader is only met with information that supports its worldview and opinion. This creates polarization and contributes to an increased spread of misinformation in society.

The study aims to map the problems that arise due to digital news feeds and its underlying algorithms. Furthermore, the study also aims to understand how the same technology can be used in a more ethically sustainable way.

A qualitative study consisting of an in-depth literature study, qualitative interviews and a case study of Improvethenews.org, a website that uses an alternative approach to present news using machine learning algorithms, was conducted. Furthermore, an analysis was made where the different parts of the study were linked together to answer the study's questions.

The study concludes that the problems that arise due to digital news feeds can be divided into human and technical. The human problems include factors that are difficult to eliminate and hence greater emphasis was placed on the technical problems. A noticeable technical problem is that people do not understand how the technology behind the algorithms works. This contributes to reduced awareness of how individualized the news feeds are and increases the risk of polarization in society. An additional problem is who should be held responsible for providing readers with balanced information. It is not possible to hold the reader responsible and thus the developers of the algorithms and the news channels are responsible for following ethical guidelines and providing readers with sufficient information for them to understand how the algorithms affect their news consumption. The authors has compiled guidelines for ethical frameworks for machine learning algorithms, which can be found under section 7.2.9.

The conditions for machine learning to be used in an ethically defensible manner in news consumption are good if the technology companies are willing to cooperate to develop systems that meet the ethical guidelines and requirements addressed in the framework, see 7.2.9. Through increased transparency and responsibility of the companies behind the algorithms, a step is taken towards a better informed society.

Keywords: News filtering, machine learning algorithms, ethical frameworks, balanced news consumption.

Note: The report is written in Swedish.

SAMMANFATTNING

Idag utgörs största delen av människors nyhetskonsumtion av att ta till sig nyheter och information via digitala medier. Nyhetsflöden på plattformar som använder sig av maskininlärningsalgoritmer för att filtrera läsarens flöde är källor för information trots att de inte visar hela bilden för läsaren, utan endast det som är av läsarens intresse för att hålla kvar dennes uppmärksamhet. Filtreringen medför många problem då läsaren endast möts av information som stödjer dennes världsbild och åsikt. Detta skapar polarisering och bidrar till en ökad spridning av misinformation i samhället.

Studien syftar till att kartlägga de problem som uppstår på grund av digitala nyhetsflöden och dess bakomliggande algoritmer. Vidare syftar studien även till att förstå hur samma teknik kan användas på ett mer etisk hållbart sätt.

En kvalitativ studie bestående av litteraturfördjupning, kvalitativa intervjuer samt en fallstudie av Improvethenews.org, en webbsida som använder ett alternativt angreppssätt för att presentera nyheter med hjälp av maskininlärningsalgoritmer, genomfördes. Vidare gjordes en analys där de olika delarna av studien kopplades samman för att tillsammans besvara studiens frågeställningar.

Studiens slutsats är att problemen som uppstår på grund av digitala nyhetsflöden går att dela upp i mänskliga samt tekniska. De mänskliga problemen innefattar faktorer som är svåra att eliminera och därav lades större vikt vid de tekniska problemen. Ett märkbart tekniskt problem är att människor inte förstår hur tekniken bakom algoritmerna fungerar. Detta bidrar till minskad medvetenhet i hur individanpassat flödet är och ökar risken för polarisering i samhället. Ett ytterligare problem är gällande vem som bär ansvar för att förse läsarna med nyanserad information. Det går inte att hålla läsaren ansvarig och därmed bär utvecklarna av algoritmerna samt nyhetskanalerna ansvar för att följa etiska riktlinjer och förse läsarna med tillräcklig information för att de ska förstå hur algoritmerna påverkar deras nyhetskonsumtion. Författargruppen har sammanställt riktlinjer för etiska ramverk för maskininlärningsalgoritmer som finns under avsnitt 7.2.9.

Förutsättningarna för att maskininläring ska kunna användas på ett etiskt försvarbart sätt vid nyhetskonsumtion är goda om teknikföretagen är villiga att samarbeta för att ta fram system som uppfyller de etiska riktlinjer och krav som tas upp i ramverket, se 7.2.9. Genom ökad transparens och ansvarstagande hos företagen bakom algoritmerna tas ett steg mot ett mer nyanserat samhälle.

Nyckelord: Nyhetsfiltrering, maskininlärningsalgoritmer, etiska ramverk, nyanserad nyhetskonsumtion.

Förord

Denna rapport är en kandidatuppsats vid Chalmers tekniska högskola vårterminen 2021. Vi vill tacka vår handledare Alicja Ostrowska för alla tips och råd på vägen. Även ett tack till Karl de Fine Licht, Gordana Crnkovic och Dorna Behdadi för att ni delade med er av era tankar och idéer under intervjuerna.

Ordlista

Artificiell intelligens (AI): En maskins förmåga att efterlikna mänskliga tankeprocesser såsom inläring, beslutsfattande och kreativitet.

Bot: Ett datorprogram som automatiskt genomför operationer och uppgifter.

Bias: En persons sätt att värdera och reagera på omvärlden, det vill säga ett synsätt.

Clickwrap-avtal: Ett avtal som godkänns genom att man klickar på en knapp på en webbsida. Avtalet behöver inte läsas igenom innan det godkänns.

Desinformation: Felaktig information med skadlig avsikt. Medveten manipulering eller avsiktligt skapade konspirationsidéer är exempel.

Faktakoll: Undersökning om hur sann den fakta som visas för användaren egentligen är. Faktakollsverktyg utvecklas ofta med hjälp av maskininlärningsalgoritmer.

Fake news: Felaktig eller vilseledande information presenterad som nyheter. Fake news är ett samlingsbegrepp för alla nyheter som består av mis-, mal- eller desinformation.

Intelligent system: Tekniskt komplexa system som kan ta in data och analysera den samt kommunicera med andra system runt om dem. Dessa system är en form av artificiell intelligens.

Kognitiv bias: Ett av de psykologiska mönster som kan göra att en människas subjektiva bild av omvärlden avviker från verkligheten.

Kognitiv dissonans: Upplevd psykologisk stress som inträffar när en person agerar på ett sätt som strider mot en eller flera av deras motstridiga värderingar, idéer eller övertygelser. När denna motstridighet blottläggs kämpar personen för att nå konsensus, antingen genom att lägga till idéer eller genom att undvika ny information som kan skapa denna kognitiva dissonans.

Konfirmeringsbias: Innebär en typ av selektivt tänkande där en individ tenderar att observera det som bekräftar dennes redan befintliga åsikter. Det innebär även att individen ignorerar, bortser från eller undervärderar relevansen hos de åsikterna som säger emot de egna.

Maskininläring (ML): Ett område inom artificiell intelligens där datorer tränas att utifrån indata analysera och lösa uppgifter eller problem baserat på mönster i datan.

Maskininlärningsalgoritm: En algoritm som tränats till att fatta beslut och utföra uppgifter baserat på indata samt en målfunktion som representerar resultatet algoritmen ska uppnå.

Malinformation: Information som sprids med skadlig avsikt, exempelvis läckor av personlig information.

Misinformation: Oavsiktligt felaktig eller misledande information, skapas av misstag.

Nyhetsaggregator: En nyhetsplattform som samlar information från flera olika källor och presenterar dessa artiklar i ett gemensamt flöde för läsaren.

Nyhetsflöde: Ett digitalt flöde av information på någon medieplattform. Här ingår alla plattformar som används som informationskällor, oavsett om skaparna av plattformen har detta som avsikt eller ej.

Open source: Programmeringskod som finns upplagd på en internetbaserad webbplats så att andra utvecklare kan ta del av koden.

Ramverk: Generella riktlinjer som ska följas för att uppnå ett önskat resultat.

Robusthet: Inom datorvetenskap innebär robusthet den förmåga hos ett IT-system att fungera tillfredsställande. Vid fel ska programmet inte krascha eller producera felaktig utdata. Robusta program ska kunna förebygga fel och kontrollera dem.

Selektiv exponering: Översättning av engelskans "selective exposure". Fenomenet att individer har en tendens att konsumera medier som stämmer överens med deras tro och verklighet, och genom det undvika medier som ifrågasätter eller utmanar dess åsikter. Den psykologiska orsaken till selektiv exponering är konfirmeringsbias.

Tillgänglighetsbias: När människor värderar nya intryck och information baserat på den direkta kopplingen som görs vid första anblick. Innebär att händelser kan förvridas baserat på hur nyss de inträffade, om de var dramatiska eller fick stor uppmärksamhet medialt.

Innehåll

1	Introduktion	1
1.1	Bakgrund	1
1.2	Syfte	3
1.3	Problemformulering	3
1.3.1	Problem vid presentation av nyheter	3
1.3.2	Problem vid konsumtion av nyheter	4
1.3.3	Socialt ansvar hos nyhetsflöden	5
1.3.4	Frågeställningar	5
1.4	Avgränsningar	5
1.5	Definition av nyanserat nyhetsflöde	6
1.5.1	Övriga faktorer för ett nyanserat nyhetsflöde	7
1.6	Definition av etik	8
1.6.1	Konsekventialism	8
1.6.2	Deontologi	9
1.6.3	Dygdetik	9
2	Hållbar utveckling	11
3	Metod	13
3.1	Kvalitativa metoder	13
3.2	Fördjupning i teori och tidigare studier	14
3.2.1	Val av litteratur	14
3.2.2	Tillvägagångssätt	15
3.3	Intervjuer	16
3.3.1	Val av intervjuobjekt	16
3.3.2	Tillvägagångssätt	17
3.3.3	Etiskt ansvar gentemot intervjuobjekten	19
3.4	Fallstudie	19
3.4.1	Val av fallstudieobjekt	19
3.4.2	Tillvägagångssätt	20
3.5	Analysmetod	21
3.6	Kvalitetssäkring	21
4	Teori och tidigare studier	22
4.1	Tekniska aspekter	22
4.1.1	Maskininlärningsalgoritm	22
4.1.2	Algoritmer i nyhetsflöden	22
4.1.3	The Black Box Problem	23
4.1.4	Designmodellen ART	24
4.1.5	Angripa filterbubblor	26
4.2	Presentation av nyheter	26
4.2.1	Internets påverkan på källkritik	26
4.2.2	Plattformarnas makt över läsarna	27
4.2.3	Balans och faktakoll	28
4.3	Konsumtion av nyheter	29
4.3.1	Polarisering	29
4.3.2	Informationsbegränsning	30

4.3.3	Selektiv exponering	30
4.3.4	Konfirmeringsbias	30
4.3.5	Tillgänglighetsbias	31
4.3.6	News finds me	31
4.4	Etiska ramverk i verkligheten	31
4.4.1	EU kommissionens riktlinjer	32
4.4.2	Standarder från IEEE	33
5	Resultat från intervjuer	36
5.1	Karl de Fine Licht	36
5.2	Gordana Crnkovic	36
5.3	Dorna Behdadi	37
6	Resultat från fallstudie av Improve The News	39
6.1	Vad är Improve The News?	39
6.2	Hur fungerar Improve The News?	40
6.2.1	Reglage	41
6.2.2	FAQ	42
6.2.3	Integritetspolicy	43
6.3	Granskning utifrån teori	44
6.3.1	Standarder och ART	44
6.3.2	Sortering och filtrering	46
6.3.3	Konsumtion och medvetenhet hos läsaren	47
6.3.4	Balans av artiklar och källkritik	49
6.4	Koppling till studien	51
7	Analys	52
7.1	Problem och hinder vid nyhetskonsumtion	52
7.1.1	Bidragande faktorer till filterbubblor och ekokammare	52
7.1.2	Algoritmernas konsekvenser	54
7.2	Riktlinjer för etiskt ramverk	55
7.2.1	Meddelandeansvar	55
7.2.2	Ansvar	56
7.2.3	Transparens	57
7.2.4	Syfte och målgrupp med ett etiskt ramverk för nyhetsfiltrering	58
7.2.5	Metoder för att uppfylla kraven	59
7.2.6	Förslag på lösningar	59
7.2.7	Filterknapp	59
7.2.8	Certifiering	60
7.2.9	Riktlinjer för ett etiskt ramverk	61
7.3	Etiskt hållbar maskininlärning	63
7.3.1	Fallet Improve The News	63
7.3.2	Kommersiella plattformar	64
8	Slutsats	66
9	Diskussion	67
10	Förslag på framtida undersökningar	69

Referenser

71

Bilaga A

B1

1 Introduktion

Att ta till sig nyheter och information via internet har blivit allt vanligare. Vilken webbplats eller applikation läsaren använder sig av varierar. Det finns renodlade nyhetsaggregatorer men många människor har sociala medieplattformar som sin primära källa till nyheter. På grund av det överflöd av information som finns tillgängligt så används maskininlärning på webbplatser som individanpassar nyhetsflödet för varje läsare. Det innebär att vi inte längre tar till oss samma nyheter och information som alla andra människor, utan olika beroende på vad och vilka vi interagerar med eller utsätts för. En diskussion gällande vilka effekter det har på samhället och individen har vuxit fram och skapat ett ökat behov av forskning inom området ”nyheter, algoritmer och etik”.

Enligt ”Eu guidelines on ethics in artificial intelligence: Context and implementation” kommer standardisering spela en stor roll för framväxten av AI-marknaden (Madiaga, 2017). I bland annat EU-kommissionens etiska riktlinjer talas det om hur olika situationer inom AI leder till olika utmaningar vilket innebär att en sektorbaserad strategi kan vara aktuell att undersöka.

1.1 Bakgrund

Den 20 januari 2017 tillträdde Donald Trump presidentposten i USA. I mars samma år rapporterade The New York Times om företaget Cambridge Analytica och dess involvering i presidentvalet (Rosenberg et al., 2018). Cambridge Analytica var ett politiskt konsultbolag med säte i Storbritannien, dotterbolag till Strategic Communication Laboratories. The New York Times beskrev hur företaget hade tagit emot en investering på 15 miljoner dollar, i uppdrag att identifiera amerikanska väljare för att påverka deras beteende.

Genom att marknadsföra och kommunicera information till användare på Facebook hoppades Cambridge Analytica kunna influera väljare (Isaak & Hanna, 2018). Särskilt utsatta var användare som hade profilerats som möjliga att påverka inför valet. Sådana väljare kan vara personer som ännu inte bestämt hur de ska rösta. En ansats för att påverka människor kan vara att presentera verifierade nyheter tillsammans med desinformation i ett flöde på en plattform såsom Facebook, vilket gör det svårt för en person att avgöra vad som är en tillförlitlig källa och inte.

Internet har förändrat modellen för det sätt som människor i nutid får information och konsumerar nyheter. Internet-baserad media där deltagare rapporterar, delar och distribuerar nyheter har vunnit legitimitet i takt med att den tryckta journalistikens legitimitet har ifrågasatts (Pentina & Tarafdar, 2014). Konsumenter på exempelvis sociala plattformar bygger sitt eget nyhetsflöde genom att välja vilka de följer, vilka nyheter som är värda att dela och erbjuds möjligheten att filtrera bort de nyheter som känns irrelevanta eller inte överensstämmer med sin egen verklighetsuppfattning. Denna filtrering bidrar till att objektivitet och multiperspektiv kan gå förlorat. Resultatet av att välja sitt eget nyhetsflöde kan således bli att man hamnar i en så kallad *filterbubbla*.

Filterbubbla är ett uttryck som syftar på att ens informationsflöde kan vara överbelastat med information, ge en suboptimal grund för normativa lagar och bilda en partisk världsbild

genom att man matas med overifierad information (Pentina & Tarafdar, 2014). Filterbubblan blir alltså ett resultat av att man, medvetet och omedvetet, sällar information genom att välja källor som verifierar ens åsikter och omge sig av människor med liknande värderingar på internet. Konsekvensen blir att man låter bli att exponeras för utmanande nyheter och åsikter – man skapar ett eget socialt filter.

Det som möjliggör den här typen av individuell filtrering är *artificiell intelligens*, förkortat *AI*. Begreppet kan vara svårt att definiera men ett sätt att betrakta det är att relatera till människans intelligens genom att citera Marvin Minsky, som var forskare inom kognitionsvetenskap och artificiell intelligens samt en av grundarna av avdelning för AI på Massachusetts Institute of Technology (MIT): "the science of making machines do things that would require intelligence if done by men" (Kaplan & Haenlein, 2019). Citatet betyder ungefär: att få en maskin att uppföra sig på ett sätt som anses intelligent, på samma sätt som en människa hade ansetts vara av samma uppförande. Hur intelligent ett system är kan beskrivas som ett systems förmåga att tolka externa data rätt, lära sig av den, och använda dessa kunskaper till att uppnå ett särskilt mål eller utföra ett givet uppdrag. För plattformar som finansieras via reklamannonser så ligger det i ägaren av plattformens intresse att bibehålla användarnas uppmärksamhet för att öka värdet på annonsplatserna. Målet för det intelligenta systemet är då att upprätthålla användarens intresse och uppmärksamhet. Ett exempel på det är att läsa av hur en person interagerar med en viss typ av innehåll och sedan ge ytterligare förslag för liknande innehåll som användaren kan lägga tid på att konsumera.

Det var med hjälp av AI som Cambridge Analytica använde Facebooks data för att sprida information (Rosenberg et al., 2018). De samlade in data, identifierade potentiella väljare med algoritmer som tolkade informationen och angrep dessa personer genom att sprida information till kampanjens fördel. En kombination av att befinna sig i en filterbubbla, med ett individanpassat nyhetsflöde, och bristfällig källkritik innebär en risk att bilda sig en snedvriden verklighetsuppfattning. Det kan i sin tur bidra till polarisering i samhället.

En motreaktion på hur människor konsumerar nyheter via sociala medier är nyhetsplattformen *Improve The News* (<https://www.improvethe news.org/>). Improve The News är grundat av Max Tegmark som är professor i fysik och forskare på MIT. Tillsammans med en grupp studenter har han skapat webbplatsen i syfte att presentera nyheter med hjälp av maskininlärningsalgoritmer på ett nytänkande sätt (Webb, 2018). Webbplatsen är en nyhetsaggregator som är helt automatiserad där en läsare kan välja från vilket perspektiv nyheter ska presenteras. På webbplatsen finns reglage som visar från vilken politisk åskådning läsaren kan se nyheterna, hur djupgående nyheterna är eller om texten är provokativ istället för nyanserad. Detta är ett pågående projekt och tanken är att koden till webbplatsen kommer vara open source, vilket är öppen källkod som är tillgänglig för allmänheten. Tegmark hoppas kunna bidra till ökad medvetenheten gällande vinkling och filtrering hos nyhetskonsumenter och på så sätt motverka polarisering (deGrasse Tyson, 2021). En webbplats som Improve The News möjliggör inte samma typ av personligt riktad information som kan infinna sig i ett nyhetsflöde på sociala medier. Personligt riktad information och nyhetsinnehåll är vad Cambridge Analytica drog nytta av i Trump-kampanjen.

1.2 Syfte

Rapporten syftar till att kartlägga de problem som uppstår på grund av dagens nyhetsflöden och dess bakomliggande algoritmer. Författarna upplever att det finns en okunskap i samhället gällande algoritmers påverkan på individers förmåga att ta till sig nyanserad information via nyhetsflöden. Genom att sammanställa de olika problem som idag kan identifieras, och undersöka vilken anknytning det har till tekniken bakom nyhetsflödena, så kan rapporten skapa en tydligare bild av vilken problematik algoritmerna orsakar. Vidare syftar rapporten till att skapa förståelse för hur samma teknik kan användas på ett etiskt hållbart sätt. Genom att angripa problemområdet ur ett kritiskt granskande perspektiv kan rapporten förhoppningsvis även föra fram diskussionen och vara ett bidrag till en ökad granskning av algoritmer ur ett etiskt perspektiv.

1.3 Problemformulering

Gällande nyheter och informationsspridning finns det flera faktorer som påverkar läsarens förmåga att ta till sig information. I bakgrundsavsnittet 1.1 framgår det att sättet som nyheter presenteras påverkar hur källkritisk läsaren är. Något som också bör belysas är de mänskliga faktorer som försvårar processen att nyansera läsaren, vilket kan förklaras med teorier och begrepp inom psykologin. Vem som bör bära ansvaret för att ge läsaren en nyanserad bild, läsaren själv eller nyhetsförmedlaren, är också en viktig fråga att besvara. I följande avsnitt diskuteras de huvudsakliga problemen gällande presentation och konsumtion av nyheter, vilket även leder fram till studiens frågeställningar.

1.3.1 Problem vid presentation av nyheter

Kvaliteten och balansen av olika synvinklar i de nyheter som presenteras är två av de faktorer som ligger till grund för invånarnas politiska engagemang och åsikt, samt är en del av att forma bilden av samhället (Helberger, 2019). Den information som konsumeras påverkar också beteenden på andra områden. Exempelvis gäller detta frågor kring vaccination och klimatet. Om samhällsinvånare exempelvis lyssnar till korrekt information om förebyggande sjuk- och hälsovård och handlar därefter hindrar det ökade kostnader för sjukvård längre fram i tiden (Hed & Helte, 2020). Ett samhälle som lyssnar på verifierad information har alltså direkta politiska, ekonomiska och hälsomässiga vinster, och det motsatta gäller för sämre informationsspridning (Hed & Helte, 2020).

På de nyhetsplattformar som finns idag görs filtrering och kategorisering av nyhetsartiklar möjligt genom AI. Denna typ av kategorisering har utnyttjats av exempelvis Facebook och Youtube för att maximera intäkter genom att användare stannar på sidan ifråga (Frank, 2021). Sociala plattformar som dessa används som tidigare nämnt också för nyhetsspridning. Detta skapar problem när de sociala plattformarna har ekonomiska incitament som inte är i linje med samhällets vinster av välfungerande informationsspridning. Det har i debatten kring algoritmers användande lyfts att algoritmerna i sig inte behöver vara positiva eller negativa för samhället, utan att det snarare är var och hur de används som avgör om en specifik algoritm är etiskt hållbar (O’Neil, 2016a). Algoritmers upphovsmakares etiska ansvar bör alltså inte bedömas generellt utan utifrån varje unikt fall med dess kontext, det vill säga i den här studien fallet av nyhetsspridning och filtrering. För företagen bakom de sociala plattformarna innebär det en ny situation att behöva ta ansvar för att ens webbplats används på ett annat sätt än

det mest intuitiva, men utan detta ansvar kommer informationsflödena fortsätta att försvagas.

Ett av problemen som bör beaktas vid presentationen av nyheter är läsarens källkritiska förmåga att bedöma artikelns legitimitet på bästa sätt, exempelvis genom huruvida källor ifråga ska presenteras eller inte (Dias et al., 2020). Resultaten vid dessa typer av studier har inte alltid varit det intuitivt mest trovärdiga, som att vissa typer av presentation av källor inte påverkat läsarna nämnvärt (Dias et al., 2020). Därför är det viktigt att förstå vilka faktorer i presentationen som verkligen hjälper läsaren att bli nyanserad.

1.3.2 Problem vid konsumtion av nyheter

Den här studien har som utgångspunkt att människor inte har full potential att observera nyheter och information helt och hållet objektivt. Rapporten kartlägger övriga problem som hindrar människor från att idag konsumera nyheter på ett nyanserat sätt. Studien utgår från att problemet inte är att nyheter är vinklade, utan hur individer konsumerar de vinklingar som skildras i deras individanpassade nyhetsflöden. Det diskuteras inte huruvida åsiktsneutrala nyheter hade varit en bra lösning då författarna hävdar att åsikter har ett värde och således inte bör motverkas.

Ett psykologiskt begrepp som kan bidra till sämre nyansering inom nyhetskonsumtion är *konfirmeringsbias*. Konfirmeringsbias innebär att favorisera information som bekräftar en individs redan befintliga åsikter. Detta påverkar hur individer tar till sig, tolkar och återkallar information (Cherry, 2020). Ett besläktat fenomen är *group thinking* som innebär att en grupp människor fattar irrationella beslut som motiveras av sammanhang och lusten att anpassa sig (Psychology today, 2020). Group thinking kan främja beslut som inte stöds av kritiskt granskad information och kan innebära stora konsekvenser för samhället. Inom politiken finns det bland annat risk för att ledande grupper prioriterar ett specifikt mål och på vägen dit ignorerar etik och moral (Psychology today, 2020).

Två andra vanligt förekommande begrepp i diskussionen om nyhetsflöden på sociala medier är *ekokammare* och *filterbubblor*. Ekokammare är en slags atmosfär där en person som befinner sig i denna stöter på information som förstärker redan existerande, intagen, information. Den nya informationen bildar ett slags ”eko” av redan intagen liknande information, vilket kan innebära att motpartens ståndpunkter eller synvinklar inte synliggörs. Personen i fråga konsumerar alltså information som reflekterar och förstärker dess egna tankar (Thi Nguyen, 2019). Skillnaden mellan de två begreppen är att filterbubblor skapas av algoritmer och ekokammare av interaktion med andra människor (Sumpter, 2018).

Konfirmeringsbias, group thinking, filterbubblor och ekokammare kan alla appliceras på diskussionen kring huruvida Cambridge Analytica var inblandad i ”Do-So”-kampanjen i Trinidad och Tobago. Kampanjen innebar att influera unga, myndiga personer, att inte gå och rösta genom att skapa ett sammanhang för ungdomarna (Steuart, 2019). Detta demonstrerar ett exempel på hur stora konsekvenser dessa fenomen kan skapa. Att genom artificiell intelligens ha möjlighet att påverka demokratiska val kan ge konsekvenser för den moderna demokratin.

Ytterligare problem som bör belysas vid nyhetskonsumtion kan analyseras utifrån studier som

genomförts för att förstå om individer är bättre på att bedöma nyheters riktighet intuitivt eller genom att ge sig själv tid att vara eftertänksam. Dessa studier har haft varierande resultat, men där den traditionella synen att vara eftertänksam ses som konsensus för vad som är bäst i dagsläget (Bago et al., 2020).

1.3.3 Socialt ansvar hos nyhetsflöden

I och med att en stor del av samhällets invånare dagligen använder digitala plattformar som styrs av algoritmer uppkommer en diskussion om hur ansvarsfördelningen mellan plattform och läsare förhåller sig när det kommer till nyansering. De policys som finns idag skiljer sig inte mellan traditionell media och social media (White & Boatwright, 2020). Dessa policys behandlar aspekter såsom transparens, öppenhet, budskapens noggrannhet, sanningsenlighet samt respekt för meddelaren men tar inte hänsyn till hur sociala medier fungerar och vilka funktioner som ligger utanför det som behandlas av dessa policys.

White och Boatwright (2020) tar upp problemet med så kallade *clickwrap*-avtal, där användaren godkänner sekretessavtal utan att öppna upp dem genom att klicka i en "I Agree"-ruta. I en studie av Facebooks sekretessavtal framgick det att mer än hälften av Facebooks användare inte förstår språket i avtalet (Zeadally & Winkler, 2016). Det saknas studier som undersöker det sociala ansvar som sociala plattformar och dess nätverk har gentemot sina användare. Är det exempelvis skillnad på det etiska ansvar som en försäkringsbyrå har gentemot sina kunder jämfört med Facebook och deras användare? Båda använder data över sina kunder för att förutspå deras beteende men i olika syften.

Digitala plattformars disruptiva effekt på mediabranschen kräver en uppdatering av regelverken för att täcka in aspekter såsom integritetsrättigheter och att motverka filterbubblor, samt andra konsekvenser som social media har på samhället (White & Boatwright, 2020).

1.3.4 Frågeställningar

Utifrån problemanalysen formulerades följande tre frågeställningar:

- Vilka problem eller hinder finns idag för att ta till sig nyheter och samhällsinformation via nyhetsflöden på ett nyanserat sätt?
- Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?
- Hur kan maskininlärning användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?

1.4 Avgränsningar

Rapporten granskar områdets problem främst ur ett samhällsvetenskapligt perspektiv. De identifierade problemen berör huvudsakligen algoritmer vid nyhetsfiltrering och deras effekt på samhälle och individ, och analyseras således ur detta perspektiv. För att skapa förståelse för ämnet och de bakomliggande algoritmerna, studeras tekniken på tillräcklig detaljnivå för att möjliggöra analys av dess effekter på nyhetsflöden.

Arbetet behandlade flertalet plattformar genom att studera artiklar och rapporter som granskar olika appliceringar av algoritmer på nyhetsflöden. Studier av olika plattformar med individanpassade nyhetsflöden studerades också för att få en bred och djup förståelse av de olika problem som kan uppstå och varför de uppstår. Fallstudien syftade däremot endast till att utvärdera plattformen Improve The News och kopplingen till de problem som identifierades. Därmed avgränsades detta arbete till att inte djupgående analysera fler plattformar än Improve The News.

Fallstudien av Improve The News, förkortat ITN, utgick ifrån att maskininlärningsalgoritmerna fungerar som beskrivet på webbplatsen och därmed genomfördes ingen testning av algoritmerna. Utifrån studiens syfte att skapa förståelse för hur maskininläring kan användas på ett etiskt hållbart sätt så var det främst ITN:s koncept, och inte kvalitén på ITN:s algoritmer, som undersöktes. Granskningen utgick från problemen som togs upp i litteraturen och därmed är det inte tekniken bakom maskininläringen, utan konsekvenserna utav den, som är av intresse.

När rapportens andra frågeställning ”Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?” undersöktes var inte målet att formulera ett fullständigt ramverk då det hade varit för tidskrävande utifrån studiens tidsram. Rapporten syftar istället till att ta fram krav och riktlinjer som kan bli en start för en fortsatt utformning av ett framtida ramverk.

1.5 Definition av nyanserat nyhetsflöde

Vad som anses vara nyanserat tolkas olika mellan människor. För att tydliggöra hur begreppet tolkades och användes i detta arbete förs här en diskussion om vad nyanserat innebär. Diskussionen utgick från delar utav en av Skolverkets moduler kring definitionen av nyanserat som sedan applicerades på nyheters syfte att informera och utbilda samhället. Författarna menar att citatet behandlar två delar som reflekterar författarnas tankar kring nyanserat nyhetsflöde i stort. Skolverkets dokument användes då skolverkets arbete är inriktat på utbildning och därmed olika typer av information som ska vara enkel att förstå, samt att det anses vara en väl granskad förvaltningsmyndighet som bör tillhandahålla genomarbetade och allmänt vedertagna definitioner.

”Nyanserat betyder att man kan göra ett urval av exempel som representerar olika synvinklar eller fakta från olika källor och att man kan tolka eller värdera dem” (Svärd, 2016, s.2)

Skolverkets definition riktar sig mot elevers förmåga att utifrån en tilldelad text kunna genomföra och presentera något ur olika synvinklar, från olika källor och värdera samt tolka dem. Den första delen av citatet *”...göra ett urval av exempel som representerar olika synvinklar eller fakta från olika källor...”* är något maskininlärningsalgoritmer kan tänkas kunna genomföra bättre än människor och då blir nästa uppgift att presentera detta anpassat för läsarna, vilket motsvarar nyhetsaggregatorers arbetsprocess. Vad gäller den andra delen av citatet *”...att man kan tolka eller värdera dem”* är detta något nyhetsmottagaren ska ges största möjlighet till, precis som enligt skolverkets definition. Ett nyanserat flöde ska alltså vara ett urval av olika synvinklar som läsaren sedan själv kan värdera, tolka och bilda sig en uppfattning kring för att få en heltäckande bild av verkligheten, med samhället och dess utveckling.

1.5.1 Övriga faktorer för ett nyanserat nyhetsflöde

Utöver den initiala definitionen av ett nyanserat nyhetsflöde följer i detta avsnitt ett ytterligare antal faktorer som togs i beaktande då nyhetsflöden bedömdes och de tre frågeställningarna för arbetet besvarades. Ett förtydligande om hur ordet nyanserat tolkats i detta arbete kan förhoppningsvis bidra till förståelse och underlätta läsningen av fortsättningen av rapporten. Kategorisering av artiklar med hjälp av maskininlärningsalgoritmer beskrivs som en introduktion till fenomenets konsekvenser för nyhetsflöden. Övriga faktorer är variabler som är en del av alla nyhetsflöden. Samtliga stycken är en följd av diskussioner inom gruppen kring vad begreppet innefattar.

Kategorisering av artiklar är att dela in artiklar utefter ämne och synvinkel, alltså göra information från flera håll inom samma kategori tillgänglig för läsaren utan att denne själv behöver söka upp dem. Eftersom det finns en stor mängd information tillgänglig är någon form av kategorisering praktiskt för att hjälpa läsaren att visas ett nyhetsflöde som totalt sett kan göra denne nyanserad. Med detta menas att läsaren som vill ta del av hela bilden gällande något får hjälp med att välja artiklar som passar för att få olika perspektiv på en variabel såsom höger-vänster politiskt och på så sätt kunna lägga mer tid på att läsa och förstå samband, samt att risken för en inbillad bild av att vara nyanserad minskar. Det är detta som kräver maskininlärningsalgoritmer för att lösa över tid om stora mänskliga resurser inte ska behöva tas i beaktande, vilket i sin tur kan ses som ett ekonomiskt ohållbart alternativ.

Personlig ekonomisk situation kan vara ett hinder för att få ett nyanserat nyhetsflöde. Detta eftersom prenumerationskostnader för flera tidningar är nödvändigt för att läsaren ska bli nyanserad om inte kostnadsfria nyhetsaggregatorer finns tillgängliga. Det är alltså en viktig del i att nyansera samhällsinvånarna av att varje enskild läsare har tillgång till olika källor oavsett ekonomisk situation. Samtidigt krävs finansiering av den journalistik som läses, vilket är anledningen till att prenumerationskostnaderna finns. Finansieringen kan även lösas med hjälp av annonsering i tidsskrifterna, vilket är en lösning som även den innefattar en del problem, såsom påverkan och konsekvenser av riktad reklam. Att staten finansierar journalistik som syftar till att nyansera dess invånare är ett positivt bidrag, men det är viktigt att det finns journalistik som är fristående från staten för att förhindra spridning av ensidig propaganda och för att även statliga organ ska granskas.

Tidsmässigt kan ett flöde med fördel gå snabbt att ta del av i viss bemärkelse. Med detta menas att om flödet samlar en stor mängd källor ska dessa vara enkla att byta mellan och därmed skapa en fördel gentemot fysiska bibliotek på så sätt att informationen kan nås från en dator oberoende av fysisk plats. Då menas att en fördel med att vara oberoende av plats är att förflyttning inte tar någon tid, och att kunna byta mellan källor snabbt ger mer tid till läsning.

Utöver detta kan plattformen göra det möjligt att se hur rapporter och analyser om ett ämne förändrats över tid, alltså göra det möjligt för någon form av arkiv att etableras. Ju längre tid som en fråga diskuterats, desto enklare är det också att bli nyanserad gällande exempelvis olika grupper och individers syn på frågan då fler hunnit analysera och publicera sina ståndpunkter. Dessutom ger tiden bättre möjlighet för ambitiösa och heltäckande analyser, såsom genom exempelvis långsam journalistik vilket innebär att journalister följer en person eller en rörelse

på daglig basis under en längre tid och sedan beskriver detta i en efterföljande text.

Representation handlar om att alla de grupper som berörs av problemet ifråga ska ha möjlighet och utrymme för att uttala sig. Både om problemet i sig och deras situation i det. Detta för att ytterligare nyansera den information som finns. Även om all fakta i en artikel stämmer behöver inte det betyda att alla relevanta faktorer för alla inblandade tagits hänsyn till i resonemangen. Dessutom är sällan en grupp en så homogen enhet som den beskrivs, utan generaliseras felaktigt till det (Blanding, 2018). Att arbeta för representation är alltså en del i att öka nyanseringen. Över tid bör som sagt representationen inom ett ämne öka eftersom det innebär att fler fått möjlighet att uttala sig.

1.6 Definition av etik

För att ta reda på om maskininlärning kan användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation, så behöver först begreppet ”etiskt försvarbart” redas ut. Inledningsvis förklaras begreppet etik och exemplifiering ges av vilka etiska aspekter som ligger bakom beslutsfattande. Med hjälp av några kända etiska filosofier kan själva begreppet etik förklaras. Med dessa filosofier kan etik appliceras på maskininlärning och dess väg till etiskt hållbart beslutsfattande. De flesta människor har en känsla av vad de anser är etiskt. Dessa känslor grundar sig ofta i vilken social struktur man är uppvuxen i. Etik handlar om vad som anses vara rätt och fel, bra och dåligt samt vad man upplever är rättvisa, dygder och ansvar. Oftast finns en känsla av vad som är hjälpande kontra stjälpande. På grund av att människor har olika värderingar så kan gränsen mellan rätt och fel vara flytande individer emellan.

Normativ etik är ett av de ämnesområden som finns inom etik. Virginia Dignum behandlar normativ etik i sin bok *Responsible Artificial Intelligence* (2019). Detta område handlar om att förstå hur saker och ting ska, eller bör, vara genom att kartlägga hur vi som människor värderar saker och skiljer på rätt och fel. Inom normativ etik finns ett antal skolor, eller filosofier, som förklarar bakomliggande faktorer till olika etiska resonemang. Särskilt tre sådana skolor (*konsekventialism*, *deontologi*, *dygdetik*) är relevanta att presentera och applicera inom AI. Dessa tre som behandlas nedan, är baserade på Virginia Dignums förklaringar (Dignum, 2019).

1.6.1 Konsekventialism

Konsekventialism kallas även teologisk etik och är förankrad i utgången av en handling. Dignum skriver att med det menas att det mest etiska beslut som kan fattas är det beslut som genererar det bästa resultatet i slutändan. Det betyder att tillvägagångssättet för att uppnå något kan variera, om dessa tillvägagångssätt genererar samma resultat, och fortfarande ha samma grad av etisk uppfyllnad. Här frågar man sig alltså vad som räknas som goda konsekvenser, vem som drar nytta av dessa konsekvenser och vem som ska avgöra vad som är goda konsekvenser. I Dignums bok tas det även upp att en av de mest kända varianterna inom konsekventialism är utilitarism. Utilitarismen handlar om att agera utifrån den största nyttan som gynnar flest människor, där nytta definieras som människors välbefinnande.

1.6.2 Deontologi

Deontologi är ett etiskt synsätt som avgör hur god en handling är baserat på normativa lagar som råder, skriver Dignum. Med detta menas att så länge man förhåller sig till det som anses vara gott, i motsats till ont, och rätt, i motsats till fel, så har man handlat enligt ett etiskt perspektiv. Det är alltså själva handlingen som spelar roll, och inte lika mycket konsekvenserna av den, till skillnad från vad som förklaras inom konsekventialismen. Detta är en form av top-down beteende, alltså ett beteende som sker på grund av lagar och regler.

1.6.3 Dygdetik

Dygdetiken handlar om hur vi som människor mår i vårt handlande. En dygd kan beskrivas som goda vanor som gör inverkan på våra känslor. Viktigt för dygdetik är att människor personlighetsmässigt utvecklar goda vanor och välvilja. Dignum förklarar att beslutsfattande inom dygdetik grundar sig i vad människor anser är en dygd att handla därefter. Detta utan att ta lika mycket hänsyn till konsekvenserna som man gör inom konsekventialismen. Ett exempel kan vara att man känner att det är en dygd att alltid tala sanning, och handlar därefter, trots att sanningen kan sår en annan människa.

Utifrån de ovan beskrivna normativa etikskolor som Virginia Dignum inkluderar i sin bok *Responsible Artificial Intelligence* (2019) så kan en konkretisering göras relaterat till AI som gäller för detta arbete.

Etik inom AI bör handla om att försöka ta hänsyn till olika filosofier. Lagar ska följas enligt det deontologiska synsättet. Här blir frågan om rätt och fel enligt vad som är tillåtet enligt rådande lagar central för implementeringen av AI. Det handlar om hantering av personlig data, att människors, företags eller organisationers integritet ska hållas intakt. Ett exempel är att avtal som godkänns ska presenteras på ett tydligt sätt så att den intressent avtalet gäller har fått möjlighet att ta del av alla kriterier och krav. Att implementera AI etiskt enligt det deontologiska synsättet handlar även om att ta hänsyn till diversitet bland människor. Diskriminering ska undvikas och likabehandling bör efterföljas.

Det bör också vara centralt vid implementering av AI att försöka uppnå det bästa resultatet för så många människor som möjligt enligt det konsekventialistiska synsättet. Detta ska försöka uppnås utan att någon annan far illa. Samtidigt som det deontologiska synsättet, att efterfölja lagar, efterföljs bör också konsekvenserna av att implementera AI beaktas. Även om något är tillåtet enligt lag kan negativa konsekvenser uppstå. Det kan till exempel vara att en viss grupp privilegierade människor får tillgång till information bakom betalväggar, som annars begränsar åtkomst av innehåll för icke-betalande besökare, eller genom system som kräver vissa ekonomiska förutsättningar. Detta kan till exempel vara en internetbaserad undersökning som endast ett fåtal människor deltar i och som människor utan tillgång till internet inte kan nå, men som ligger till grund för beslut som påverkar en hel befolkning.

Dygdetiken kan vara svår att applicera då den kan vara väldigt individuell, men tack vare att dygder ofta speglas i lagar och normer så blir även den filosofin relevant i resonemanget kring artificiell intelligens och maskininlärning kopplat till reglering. Mest applicerbar blir dygdetik när utvecklares, designers och företags intentioner utreds i en kedja av implementering av AI.

Om ett resultat av en implementation av ett intelligent system anses vara negativt för en viss population så kan en utvecklarens intention enligt dygdetiken ändå ha ansetts vara etisk försvarbar. Utvecklaren kan ha handlat enligt vad denne anser vara en dygd som grundat sig i goda värderingar, men konsekvenserna är kanske illa förutspådda. Enligt detta sistnämnda exempel är det därför viktigt att lyckas kombinera olika etiska synsätt och alltid ha ett helhetsperspektiv på sitt intelligenta system, från designstadiet till implementation.

2 Hållbar utveckling

En viktig del i arbetet är att konkretisera vad som är hållbart och etiskt i sammanhanget ”nyheter och informationsspridning”. Under detta avsnitt fokuseras det endast på FN:s globala hållbarhetsmål och hur studien förhåller sig till dessa.

Gällande hållbar utveckling behandlar studien fem av de 17 mål som FN formulerat (Svenska FN-förbundet, 2021). Fokus ligger på underrubriker hos mål nummer 3, 10, 16 samt 17. För mål 9 är den övergripande formuleringen passande, men ingen underrubrik är direkt relaterad till filtreringsalgoritmer i nyhetsflöden. Detta utvecklas längre ner i texten.

- **3.3** Senast 2030 utrota epidemierna av aids, tuberkulos, malaria och försummade tropiska sjukdomar samt bekämpa hepatit, vattenburna sjukdomar och andra smittsamma sjukdomar.
- **3.4** Till 2030 genom förebyggande insatser och behandling minska det antal människor som dör i förtid av icke smittsamma sjukdomar med en tredjedel samt främja psykisk hälsa och välbefinnande.
- **9** Bygga motståndskraftig infrastruktur, verka för en inkluderande och hållbar industrialisering samt främja innovation.
- **10.3** Säkerställa lika möjligheter och minska förekomsten av ojämlika utfall, bland annat genom att avskaffa diskriminerande lagstiftning, politik och praxis och främja lagstiftning, politik och åtgärder av lämpligt slag i detta hänseende.
- **16.10** Säkerställa allmän tillgång till information och skydda grundläggande friheter, i enlighet med nationell lagstiftning och internationella avtal.
- **17.14** Föra en mer samstämmig politik för hållbar utveckling.

FN:s tredje mål (3.3 och 3.4) berör hälsa och välbefinnande. Kopplat till hantering och konsekvenser av algoritmer i nyhetsflöden kan problemet angripas genom att säkerställa att samhällsnyttig information når ut till samtliga medborgare och motverka spridningen av felaktig fakta. Det kan exempelvis vara att få korrekt information från myndigheter vid pandemier. Psykologiskt negativa effekter har kunnat kopplas till en icke-nyanserad bild av samhället på individnivå som kan ge upphov till oro (Ghersetti, 2018). Denna icke-nyanserade bild av samhället är något implementering av algoritmer på olika plattformar bidragit till (Spohr, 2017).

Gällande det nionde hållbarhetsmålet kan det undersökas ifall det i framtiden bör upprättas ett delmål för hur AI ska användas på ett hållbart sätt, både inom och mellan länder. Riktlinjer på hög politisk nivå bör eventuellt tillsättas för hur utvecklare av AI och liknande verktyg kan använda sin teknik och vad den får användas till. Begreppet infrastruktur kan inkludera informationsvägar, vilket således inkluderar användningen av plattformar som tillhandahåller nyhetsflöden i detta begrepp. Därmed blir mål 9 relevant att utveckla genom att addera ytterligare delmål som går att applicera på detta område mer specifikt.

Ett liknande mål är nummer 16 med delmål 16.10 som handlar om att säkerställa allmän tillgång till information och skydda grundläggande friheter, i enlighet med nationell lagstiftning och internationella avtal. Detta delmål skulle i teorin kunna uppnås, med dagens formulering, utan att ta hänsyn till maskininlärnings påverkan på journalistik och informationsspridning generellt. Eftersom ny teknik skapat sätt att producera och presentera nyheter genom fokus på användarnas interaktion med nyhetssidor, till skillnad från exempelvis traditionella papperstidningar, berörs detta område av delmål 16.10.

Målen med nummer 10 (10.3) och 17 (17.14) relateras till att partiska algoritmer eventuellt bör hanteras på högre politisk nivå för att motverka ojämlikhet, mer specifikt vad gäller nyhetsflöden. Algoritmer har i vissa exempel präglats av biaser liknande människans. Denna problematik har bland annat visat sig uppstå inom sjukvården och vid traditionell rekrytering, där självlärande algoritmer har baserat beslut på historiska data innehållande diskrimineringsmönster (Torstensson & Blomberg Cedergren, 2019; Petersson & Äng, 2020). Författarna menar att liknande bias kan förekomma vad gäller nyheter där läsaren exponeras för nyheter som är i linje med dennes uppfattning och åsikter. Konsekvenserna kan förstärkas då utvecklarna inte förstår sig på systemen tillräckligt väl.

Konsekvenser på grund av att inte förstå sig på ett intelligent system kan vara av typen att en viss grupp missgynnas, men också att hela samhället tar skada, som vid polarisering. Vidare bygger politik och demokrati bland annat på dialog och förståelse för olika åsikter, något som försvåras då nyhetskonsumentens åsikter och uppfattningar kan bekräftas och förstärkas då de utsätts för nyhetsfilter. För att uppnå detta krävs allmän tillgång till nyanserad information och att inte samhällsnyttiga nyheter filtreras bort för vissa individer eller grupper.

Vad gäller de etiska aspekterna finns det förhoppningar om att detta arbete bidrar till att skapa förståelse och ökad diskussion kring hur det går att använda denna teknik på ett etiskt hållbart sätt. Ett förslag på hur detta kan uppnås presenteras även mot slutet av rapporten. Författarnas uppfattning är även vad gäller de etiska aspekterna att arbetet snarare kommer hjälpa än att stjälpa. Detta eftersom arbetet, om något, kommer leda till en positiv utveckling vad gäller etiska frågor kopplat till användningen av maskininläring för individanpassade nyhetsflöden. Etisk hållbarhet är ett centralt tema för hela arbetet och det är utifrån detta som potentiella lösningar på identifierade problem senare i rapporten också föreslås.

3 Metod

Arbetets syfte är att kartlägga de problem som uppstår till följd av algoritmers påverkan på nyhetsflöden. Vidare ska rapporten skapa ökad förståelse för hur de etiska aspekterna som problemen berör kan förbättras genom att använda tekniken på ett mer etiskt hållbart sätt. För att uppnå syftet, och även svara på frågeställningarna, krävs en metod som ger tillräckligt med information och relevant underlag för att kunna göra en djupgående analys och dra trovärdiga slutsatser.

3.1 Kvalitativa metoder

Studien inleddes med en kartläggning av området ”nyheter, algoritmer och etik” för att få en förståelse för problematiken inom området och hur studien skulle avgränsas. Utifrån problemformuleringen gjordes valet att genomföra en *kvalitativ* studie bestående av fördjupning inom teori och tidigare studier, kvalitativa intervjuer samt en fallstudie av Improve The News. Holme och Solvang (1997) förklarar att en kvalitativ studie går på djupet inom ett noga utvalt segment och forskaren eftersträvar att beskriva och förstå samt fokuserar på det säregna inom det valda segmentet. De frågeställningar som togs fram avgränsade området till digitala nyhetsflöden och författarna ville få en djup förståelse kring problematiken inom det valda segmentet. Maskininlärningsalgoritmer används inom en mängd olika områden och genom denna studie önskas att förstå hur de används i just nyhetsflöden och vilka problem det innebär på samhällsnivå.

För att genomföra en kvalitativ studie är det viktigt att förstå sammanhanget och helheten som undersökningsobjektet förhåller sig till (Holme & Solvang, 1997). Därför inleddes studien med en kartläggning som sedan lade grunden för den teoretiska studien som genomfördes därefter. Utifrån teorin applicerades sedan de identifierade problemen i fallstudien för att undersöka hur Improve The News förhåller sig till problematiken bakom maskininlärningsalgoritmer. Studien använde således en *deduktiv* metodansats vilket innebär att forskaren först tar till sig teori och sedan applicerar denna kunskap på verkligheten och undersöker hur väl den överensstämmer med teorin (Åkerlund, 2017).

Eriksson och Wiedersheim-Paul (2008) rekommenderar att man för en kvalitativ studie bör använda sig av berättande intervjuer och fallstudier, vilket författarna också valde att använda sig av. För att få en nyanserad bild av ämnet så har *triangulering* användas vid både studie av litteratur och intervjuer. Triangulering innebär att använda sig av olika källor och perspektiv för att angripa ett område eller en fråga (Gunnarsson, 2020). Under studiens gång anpassas trianguleringen till de olika delarna där man under litteraturfördjupningen studerade ämnet ur olika perspektiv, i detta fall både tekniskt och samhällsvetenskapligt. Under intervjuerna användes triangulering genom att intervjua personer med olika relation till det studerade området.

Syftet med intervjuerna var att ge en ytterligare dimension kring problematiken där kunniga inom området kunde bidra med deras erfarenhet samt rekommendera fortsatta ämnen och riktningar för studien. Fallstudien kopplade därefter teorin till verkligheten genom att studera ett fall där utvecklarna valt ett annorlunda angreppssätt för att presentera individanpassade nyhetsflöden med hjälp av maskininlärningsalgoritmer, genom att ge

kontrollen över anpassningen till användaren.

I nedanstående tabell 1 redovisas valet av metod för att angripa respektive frågeställning.

Tabell 1: Frågeställning & metod

Frågeställning	Metod
Vilka problem eller hinder finns idag för att ta till sig nyheter och samhällsinformation via nyhetsflöden på ett nyanserat sätt?	Teori som behandlar problematik rörande nyhetsalgoritmer ur ett etiskt och samhällsvetenskapligt perspektiv.
Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?	Teori som behandlar olika metoder att motverka de problem som identifierats gällande maskininlärningsalgoritmer, exempelvis befintliga ramverk. Efterföljs av intervju med kunniga inom området "artificiell intelligens och etik". Avslutningsvis sammankoppling med fallstudie av Improve The News för jämförelse och verklighetsförankring.
Hur kan maskininläring användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?	Samtlig teori, intervjuer med kunniga inom området "artificiell intelligens och etik" och fallstudien av Improve The News används som underlag.

3.2 Fördjupning i teori och tidigare studier

Litteraturfördjupningen var studiens mest omfattande process då den syftade till att bidra med underlag för att angripa samtliga frågeställningar. I detta avsnitt ges en ingående genomgång av hur litteraturen som studerades valdes ut, samt hur fördjupningens arbete planerades och genomfördes.

3.2.1 Val av litteratur

För att hitta relevanta och trovärdiga källor användes Chalmers databaser, Google Scholar, ScienceDirect samt Scopus där det publiceras forskningsstudier och avhandlingar med innehåll av genomgående hög kvalitet. Litteraturfördjupningen behandlade ett omfattande ämne och delades därför upp utefter olika områden enligt det tillvägagångssätt som beskrivs i avsnitt 3.2.2. Detta innebar att författarna sökte efter en stor bredd av artiklar och studier för att hitta relevant information men genom att använda sig av kvalitetssäkrade sökkällor så har litteraturen hållit en genomgående god kvalitet.

Litteraturen som söktes efter har främst varit engelskspråkig och skrivits av författare av olika nationalitet. I inledningen av litteraturfördjupningen söktes det även med svenska sökord men utbudet upplevdes vara avsevärt mer begränsat, vilket gjorde att sökningarna därefter enbart bestod av engelska sökord. Vilka som har använts redovisas i tabell 2 nedan utefter respektive område. I jämförelse med övriga områden används få sökord för "Etik", detta på grund av

att annan litteratur än det som återfunnits i sökkällor använts för detta avsnitt. Vad gäller datasamlingen av etiska ramverk i verkligheten samt dataskyddsförordningen användes Google som sökmotor, då det ansågs vara snabbaste ruten till dokumenten.

Tabell 2: Litteraturfördjupingens sökord

Område	Sökord
Nyanserat nyhetsflöde	Nuanced definition, nuanced news, nyanserat definition, balanced news, nyanserad journalistik, balanced journalism
Etik	Filter bubbles and news, machine learning and news
Tekniska aspekter	Recommender systems, news recommendation, filter bubbles, algorithms ethics, news recommender systems, selective exposure, news feed algorithms, filter bubbles and news, machine learning and news
Presentation av nyheter	Balance in journalism, digital journalism, digital journalistik, fact-check, fact-check journalism, ethics journalism
Konsumtion av nyheter	Selective exposure, filter bubble, confirmation bias, confirmation bias, news consumption, fake news and AI, AI and ethics, filter bubble and news, echo chamber, consumption of news and AI, problems with news consumption AI, news finds me, biased artificial intelligence, polarization news and AI, problems social media AI, selective exposure and AI
Etiska ramverk i verkligheten	Ethical framework ai eu, ethics guidelines for trustworthy ai eu, eu guidelines on ethics in artificial intelligence, ieee, ethical framework ieee, gdpr autonoma beslut, skäl 71 gdpr, dataskyddsförordningen

I samband med ett handledningstillfälle tipsade Alicja Ostrowska om Virginia Dignum, professor i datavetenskap vid Umeå Universitet. Dignums bok "Responsible AI" behandlar många områden som tas upp i denna studie och har använts under stora delar av litteraturfördjupningen. Även andra böcker har använts då författarna ansåg att de artiklar och studier som återfanns i sökkällor inte täckte det tekniska området helt och hållet. Via Google sökmotor fann författarna böckerna "The Ethical Algorithm" av Michael Kearns och Aaron Roth samt "Weapons of Math Destruction" av Cathy O'Neil. Böckerna användes också under andra delar av litteraturfördjupningen då de även tar upp problemet med presentation av nyheter.

3.2.2 Tillvägagångssätt

Som visades i tabell 1 så är teorin en genomgående del av arbetet där materialet från litteraturen bidrog till att angripa samtliga frågeställningar. Litteraturfördjupningen lade därmed grunden för intervjuerna samt fallstudien och därför var den det första som författarna började med. För att kunna behandla alla de områden som studien syftade till att beröra så delades teorin upp i tydliga områden enligt följande:

- Nyanserat nyhetsflöde

- Etik
- Tekniska aspekter
- Presentation av nyheter
- Konsumtion av nyheter
- Etiska ramverk i verkligheten

Uppdelningen av områden initierades av den inledande litteraturfördjupningen som genomfördes vid studiens uppstart. Utifrån underlaget författarna fick fram enades de om ett antal områden som ansågs representera litteraturen väl och fick därmed utgöra ramen för den större litteraturfördjupningen.

Inledningsvis fokuserade författarna på tekniska aspekter, presentation av nyheter samt konsumtion av nyheter. Där delade författarna upp sig i par och genomsökte litteraturen för att täcka respektive område. Inledningsvis sökte författarna med stor bredd för att samla på sig mycket material. Litteraturfördjupningen skulle kunna klassas som kvantitativ inledningsvis men i och med att författarna sedan sammanställde och smalnade av den studerade litteraturen så blev resultatet kvalitativt och inriktat på de utvalda områdena.

Efter att litteraturfördjupningen pågått några veckor insåg författarna att det inte är självklart vad de menar med ”etiskt” eller ”nyanserat” då begreppen är av filosofisk karaktär och kan tolkas subjektivt av läsaren. Därmed inleddes även en studie av dessa begrepp samt egen reflektion av vad som menas med begreppen i denna studie. Detta kan hittas under avsnitt 1.5 och 1.6 i introduktionen för att läsaren tidigt ska få en förståelse för vad begreppen betyder.

Den sista delen av litteraturfördjupningen var undersökningen av etiska ramverk i verkligheten. Då denna del skulle inledas var studien redan halvvägs igenom och därmed hade planering av intervjuer samt fallstudien påbörjats. Författarna valde då att de tre personerna som hade ansvar för intervjuerna också skulle studera etiska ramverk medan resterande tre arbetade vidare med fallstudien. Denna uppdelning lämpade sig bra då intervjuerna inte var lika tidskrävande som fallstudien och därmed hade författarna tid med den sista delen av litteraturfördjupningen.

3.3 Intervjuer

Intervjuerna genomfördes delvis parallellt med litteraturfördjupningen för att komplettera den information författarna hittade och ge nya insikter och inriktningar för studien. I följande avsnitt beskrivs hur författarna valde ut sina intervjuobjekt samt hur intervjuerna gick till. Det har även tagits hänsyn till intervjuobjektens integritet, vilket redogörs i avsnitt 3.3.3.

3.3.1 Val av intervjuobjekt

I denna studie genomfördes kvalitativa, semi-strukturerade intervjuer med forskare inom arbetets område för att bidra med ytterligare perspektiv och synvinklar som inte nödvändigtvis presenterades i den studerade litteraturen. Ett ytterligare motiv var att få en helhetsbild och eventuellt värdefull information utöver det som direkt frågades efter eller tidigare söktes efter i litteraturen. På så sätt kunde intervjuerna också leda till ytterligare litterära studier inom

områden som inte behandlades men som då visade sig vara relevanta för arbetet.

Genom Chalmers AI Ethics committee, Chalmers etiska kommitté för AI (CHAIR), fick författarna kontakt med Karl de Fine Licht som svarade att han gärna ställde upp på intervju. För vidare val av framtida intervjuobjekt använde sig författarna av ett så kallat snöbollsurval, vilket innebär att respektive intervjuobjekt frågades efter att rekommendera någon som skulle vara passande för vidare intervju. På så sätt förenklades sökandet efter nya personer som kunde bidra med synvinklar som saknades då personer som är insatta inom ett område ofta känner till andra som är verksamma inom samma eller närliggande fält.

Karl de Fine Licht har en doktorexamen inom filosofi och är lärare vid Chalmers inom etik och teknik. Författarna ansåg att de Fine Licht skulle kunna bidra med en intressant synvinkel och intervjun genomfördes därmed tidigt under studien för att kunna få tips om inriktningar och områden som författarna borde undersöka under litteraturfördjupningen. Under intervjun frågade författarna om de Fine Licht kunde rekommendera någon person för vidare intervju varvid han rekommenderade Gordana Crnkovic och Dorna Behdadi. Författarna kontaktade respektive och båda ställde upp på intervju. Gordana Crnkovic är professor inom interaktionsdesign vid Chalmers samt medlem i Chalmers AI Ethics committee, Chalmers etiska kommitté för AI. Dorna Behdadi är doktorand inom filosofi och filologi vid Göteborgs Universitet. Behdadi har inom sina studier fokuserat på *normativ etik*, vilket är ett område inom moralfilosofi och etik. De Fine Licht ansåg att Crnkovic och Behdadi skulle komplettera varandra på ett bra sätt inom ämnet etik och att båda hade kompetens för att kunna tillföra nya synvinklar till studien.

Inom författargruppen diskuterades det även om det ansågs aktuellt med intervju av personer från näringslivet med insikt inom maskininlärning, nyhetsflöden och etik. Anledningen till att genomföra en sådan intervju skulle vara att få förståelse för hur trenden ser ut inom de företag som är verksamma inom området och vilka tendenser till förändringar som man kan se. Författarna kom fram till att ett försök att kontakta någon ansvarig på nyhetsplattformen Omni (<https://www.omni.se/>) skulle göras. Omni är en svensk nyhetsaggregator vars slogan är "Alla nyheter. Alla perspektiv.". Webbplatsens redaktörer sammanfattar nyheter för att sedan även länka till olika ursprungskällor inom ämnet. Dessvärre var försöket att kontakta Omni inte framgångsrikt och därmed blev en eventuell intervju aldrig av. Därmed valde författarna att utelämma webbplatsen ur studien, då författarna ansåg att endast en fallstudie var aktuellt i mån om studiens tidsram.

3.3.2 Tillvägagångssätt

Enligt E. Bohlin (föreläsning, 21 januari 2021) kännetecknas en kvalitativ intervju av att vara mindre strukturerad och fokus riktas mot den intervjuade snarare än på forskarens frågor. Utifrån detta är intervjuaren fri att styra i vilken riktning intervjun ska fortlöpa, vilket tydligt beskriver dess flexibla karaktär. Intervjumetoden genererar därmed ofta ett fylligt underlag med stora influenser från intervjuobjektet själv. I och med att författarna ansåg att intervjuerna främst var menade att bidra med nya synvinklar och rekommendationer så lämpade sig kvalitativa intervjuer väl för studien.

Intervjuprocessen påbörjades några veckor efter att litteraturfördjupningen satts igång,

för att författarna först ville få grepp om vilka områden de eventuellt ville diskutera med intervjuobjekten. Som tidigare nämnt genomförde tre författare intervjuprocessen och resterande fortsatte därmed på litteraturfördjupningen. Processen inleddes med intervju av Karl de Fine Licht för att några veckor senare intervjuar Gordana Crnkovic och Dorna Behdadi runt samma tidpunkt. Inför intervjuerna formulerades potentiella frågor för respektive intervjuobjekt som redovisas i tabellen nedan.

Tabell 4: Förberedda frågor inför intervju

Intervjuobjekt	Förberedda frågor
Karl de Fine Licht	• Vad är din definition av etik? • Vad tycker du är viktigast att ha med i ett etiskt ramverk? • Vilka problem tycker du är störst inom området etiskt hållbar AI gällande filtreringsalgoritmer i nyhetsflöden? • Tror du att det går att använda ML på ett etiskt hållbart sätt inom området filtreringsalgoritmer som används vid sortering av nyhetsflöden? • Har du tips på ramverk för algoritmer och etik? • Har du någon person du kan rekommendera för vidare intervjuer?
Gordana Crnkovic	• Vad anser du vara det största problemet inom nyhetskonsument/presentation? • Anser du att det finns någonting som nyhetskanalerna kan göra för att minimera/ta bort problemen som finns idag? Och, vad kan de göra då? • Hos vem anser du att problemen ligger? Konsumenterna (de som läser nyheter) eller producenterna/intelligenta algoritmer (kanalerna)? Hos vem ligger ansvaret för att så många som möjligt ska ta till sig nyanserad information? • Tycker du att man ska använda olika typer av normativ etik beroende på vilken typ av AI det rör sig om? Isåfall, vilken typ av normativ etik ska användas inom vårt område?
Dorna Behdadi	• Vad anser du vara de största problemen inom nyhetsfiltrering? • Vem tycker du bär huvudansvaret för att människor ska ta till sig nyanserad och tillförlitlig information? • Vem bär ansvaret för konsekvenserna av smarta algoritmer som filtrerar olika typer av nyhetsflöden? • Vad tycker du om IEEE och EU kommissionens riktlinjer för etik inom AI? Vilka styrkor och vilka brister ser du hos dem? • Vi har tänkt att sammanställa något slags förslag till ett etiskt ramverk inom detta område. Vad ser du för potential hos ett etiskt ramverk? Vilka problem skulle det kunna lösa och vilka är svårare att nå/bota? • Vad anser du vara de viktigaste att ha med i ett etiskt ramverk för nyhetsfiltrering? (Krav, principer) • Har du några idéer på hur ett sådant ramverk skulle se ut?

I och med att det genomfördes kvalitativa intervjuer var frågorna inte menade att styra intervjun och författarna var beredda på att intervjun kunde ta olika vändningar beroende på intervjuobjektets reflektioner. Intervjuerna genomfördes över Zoom och spelades in samt antecknades av en av författarna medan de andra två ställde frågor och var mer delaktiga i intervjun. Efter intervjun sparades materialet för att senare redovisas som resultat från intervjuer i kapitel 5, Intervjuer, samt transkriberas under bilaga A. Där redovisas endast de delar som författarna ansåg bidrog till studiens syfte och frågeställningar eller hjälpte författarna vidare i studien på något annat sätt. De utvalda delarna transkriberades i efterhand för att läsare av rapporten ska ha tillgång till materialet.

3.3.3 Etiskt ansvar gentemot intervjuobjekten

Innan intervjuerna inleddes frågade författarna om den intervjuade godkände inspelning av materialet med hänsyn till deras integritet samt GDPR (Europaparlamentets och rådets förordning (EU) 2016/679). Intervjuobjekten har även fått läsa igenom det material som presenteras i kapitel 5, Intervjuer, för att godkänna att det författarna återgett stämmer med vad de menade. Intervjuobjekten har också haft möjlighet att vara anonyma i studien men samtliga godkände användning av deras namn.

3.4 Fallstudie

Genom fallstudien av Improve The News kunde författarna koppla litteraturen till verkligheten och undersöka huruvida nyhetsaggregatorn hanterar de problem som togs upp i litteraturen på ett bra sätt. I följande kapitel framgår det varför författargruppen gjorde en fallstudie av just Improve The News samt hur granskningen av webbplatsen gick till.

3.4.1 Val av fallstudieobjekt

Som visades i tabell 1 var syftet med fallstudien att bidra med underlag för att kunna besvara frågeställningarna ”Hur kan maskininlärning användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?” och ”Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?”. Valet att granska just Improve The News var ett förslag från en av författarna som är insatt i Max Tegmarks arbete och på så sätt hade kännedom om denna webbplats som grundats av Tegmark. Efter att resterande delar av författargruppen satt sig in i webbplatsens funktion enades samtliga om att Improve The News passade som fallstudieobjekt baserat på studiens fokusområde: digitala nyhetsflöden.

Den huvudsakliga orsaken till att en fallstudie genomfördes i detta arbete är att en fallstudie kan öka möjligheten till konkretisering. Webbplatsens sätt att använda maskininlärningsalgoritmer i ett nyhetsflöde skiljer sig avsevärt från de kommersiella plattformarna såsom Facebook, Twitter och Google News som författargruppen hade personlig erfarenhet av. Författarna ansåg att läsarens möjlighet att själv filtrera sitt flöde och välja perspektiv var ett intressant angreppssätt. Frågor gällande om det ökar transparensen och hur det påverkar läsaren dök upp, och gruppen enades om att en undersökning av webbplatsen kunde bidra till att koppla studien till verkligheten. Det fanns även intresse i att granska en webbplats som vid första anblick ger intrycket av att vara ett steg i rätt riktning mot mer etiskt hållbara maskininlärningsalgoritmer, något som författargruppen ville undersöka vidare. Därmed bidrog valet att undersöka Improve The News till formuleringen av den tredje frågeställningen, ”Kan maskininlärning användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?”.

Granskningen av Improve The News planerades också att användas vid konkretiseringen av ett etiskt ramverk för nyhetsflöden som använder maskininlärningsalgoritmer. Genom att koppla samman den information som återfanns i litteraturen om etiska ramverk med hur webbplatsen fungerar kunde ramverket få en mer anpassad struktur för just digitala nyhetsflöden. Författarna ville genom detta tillvägagångssätt skapa ett situationsanpassat ramverk för deras specifika område och därmed var en fallstudie av Improve The News även en

resurs för att besvara den tidigare nämnda frågeställningen ”Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?”.

3.4.2 Tillvägagångssätt

För att kunna göra en grundlig undersökning av Improve The News ansåg författarna att det var mest lämpligt att genomföra fallstudien efter litteraturfördjupningen och intervjuerna. Författarna ansåg att de i början av studien inte hade tillräckligt med kunskap och teoretiskt underlag för att kunna granska webbplatsen ur ett kritiskt perspektiv och därmed planerades fallstudien in i den senare delen av studien. Som nämns i avsnitt 3.1, kvalitativa metoder, innebär det att studien använde en deduktiv metodansats och därmed ville författarna undersöka de problem som litteraturen samt intervjuerna tog upp, för att sedan applicera dem på webbplatsen och granska hur väl den hanterar dem.

Efter genomförd litteraturfördjupning och intervjuer inledde tre författare fallstudien. Författargruppen valde att endast låta tre medlemmar fokusera på fallstudien för att övriga delar av studien skulle kunna fortgå parallellt. Gruppen bestämde vid inledning av fallstudien att den skulle bestå av två huvudsakliga delar, en beskrivning av webbplatsen Improve The News samt en granskning utifrån teorins underlag. Under ett tidigt skede av fallstudien hade en av författarna redan inlett arbetet med att beskriva webbplatsen. Genom att beskriva webbplatsen i detalj får läsaren av rapporten en tydlig bild av Improve The News utan att behöva besöka webbplatsen, och författarna fick själva en djup förståelse för sidans struktur och design. Resterande två författare kunde då läsa in sig på webbplatsen genom att både undersöka den själva samt läsa beskrivningen.

I inledningen av studien då det beslutades att en fallstudie av Improve The News skulle genomföras så hade författarna ambitioner om att även intervjua någon ansvarig, förslagsvis Max Tegmark. Anledningen var att få ytterligare underlag och kunna ställa frågor som dyker upp under fallstudiens gång. Dessvärre lyckades författarna inte få kontakt med någon ansvarig eller delaktig i projektet och beslutade därmed att stryka den delen av fallstudien. Däremot upptäckte författarna en publicerad intervju på plattformen YouTube, där Max Tegmark intervjuas av Niel deGrasse Tyson i podcasten StarTalk avsnitt ”Improving the News with Max Tegmark” (deGrasse Tyson, 2021). Intervjun kom därmed till användning både i beskrivning av webbplatsen samt i granskningen, då tekniska aspekter samt syftet och filosofin bakom webbplatsen förklaras under intervjun.

Resterande del av fallstudien bestod av att granska Improve The News utifrån de problem som återfanns i teorin. Utifrån litteraturen kunde författarna dela upp granskningen i fyra separata områden:

- Standarder och ART
- Sortering och filtrering
- Konsumtion och medvetenhet hos läsaren
- Balans av artiklar och källkritik

Författarna fördelade sedan arbetet sinsemellan genom att dela upp områdena baserat vilket respektive person hade mest kunskap inom, och därmed underlätta analysen och kopplingen

till teorin. Att dela upp granskningen baserat på teorin visade sig vara ett framgångsrikt tillvägagångssätt. Författarna upplevde att de kunde dra tydliga kopplingar mellan teori och empiri samt att analysens trovärdighet ökade då resonemangen kunde styrkas med hänvisningar till teorin.

3.5 Analysmetod

Enligt Eriksson och Wiedersheim-Paul (2008) genomförs en kvalitativ analys genom att komprimera informationen för att ge en sammanfattad översiktsbild. Baserat på informationen genomförs sedan en *mönsteranalys* där strukturer och relationer mellan olika perspektiv synliggörs. Efter att materialet från teorin, intervjuerna samt fallstudien sammanställdes i respektive kapitel så inledde författarna en sammankopplande analys. För att tydliggöra kopplingen till frågeställningarna så delades analysen upp i tre avsnitt, ett för respektive frågeställning. Uppdelningen förenklade arbetet med mönsteranalysen genom att tydligt kunna strukturera upp hur olika delar av studien angriper frågeställningarna och därmed gjordes det tydligt för läsaren vad de olika delarna av studien tillförde.

Författarna valde under analysen att dela upp arbetet så att en författare hade huvudansvar för att analysera materialet för att kunna svara på frågeställningen "Vilka problem eller hinder finns idag för att ta till sig nyheter och samhällsinformation via nyhetsflöden på ett nyanserat sätt?" då den endast berörs av teorins underlag. Frågeställningarna "Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?" och "Hur kan maskininlärning användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?" baseras på underlaget från både teorin, intervjuer samt fallstudien. Därför valde författarna att låta två författare som tillsammans varit delaktiga i samtliga delar genomföra analysen tillsammans. Resterande tre författare kunde därmed fortsätta arbetet med övriga delar av studien som kvarstod, såsom justering av metodkapitlet, justering baserat på underlag från tidigare opponeringstillfälle, sammanställning av intervjumaterial samt generell komplettering.

3.6 Kvalitetssäkring

För att kunna validera att studien håller god kvalitet och genomförs ur ett nyanserat och kritiskt granskande perspektiv så finns det ett antal aspekter som författare av en kvalitativ studie bör förhålla sig till. Enligt E. Bohlin (föreläsning, 21 januari 2021) bör författarna vara perspektivmedvetna och diskutera sin förförståelse för ämnet. De bör även vara noggranna med att använda rätt analysmetod och trovärdiga källor som stödjer resultaten. Genom hela studien bör det finnas en tydlig tankegång och resonemang som sedan styrker de slutsatser som författarna kommer fram till. Författarna har tagit hänsyn till dessa aspekter och under studiens gång diskuterat partiskhet, val av källor och analysmetod. Författarna har inte några kända konkurrerande ekonomiska intressen eller personliga relationer som kan ha påverkat studiens resultat. Däremot består författargruppen av sex stycken ingenjörstudenter vid Chalmers tekniska högskola, vilket kan ha påverkat tolkningen av resultaten och att ett visst urval av det studerade materialet gjordes.

4 Teori och tidigare studier

Inledningsvis presenteras tekniken bakom algoritmerna och den problematik som drabbar både utvecklarna och läsarna som algoritmerna appliceras på. Därefter följer två avsnitt som behandlar presentation samt konsumtion av nyheter och slutligen presenteras de ramverk som författarna behandlat och utgår från i deras analys av ramverk.

4.1 Tekniska aspekter

Ett antal tekniska problem och hinder identifierades för när människor ska ta till sig nyheter och samhällsinformation på ett nyanserat sätt. De tekniska problemen är av algoritmisk karaktär men det finns även problematik i design och organisation. Inledningsvis presenteras en bakgrund till varför maskininlärningsalgoritmer används i nyhetsflöden och de två i dagsläget mest använda teknikerna. Därefter presenteras problematik som identifierades i litteratur angående maskininlärningsalgoritmer.

4.1.1 Maskininlärningsalgoritm

I grunden är en algoritm en beskrivning av ett tillvägagångssätt att utföra en viss uppgift (Kearns & Roth, 2020). Ett klassiskt exempel är sorteringsalgoritmer, vars uppgift är att sortera indata utefter specifikationer designade av exempelvis ingenjörer. En sådan grundläggande uppgift kräver inte att algoritmen analyserar indatan utan det finns en tydlig instruktion för hur datan ska hanteras och resultatet är en sorterad lista. De algoritmer som istället genererar en modell utifrån data kallas maskininlärningsalgoritmer. I ett sådant fall är utdatan inte en manipulation av indatan, utan en algoritm i sig som sedan kan appliceras inom ett specifikt område. Algoritmens uppgift är att utifrån indata och en specificerad målfunktion, erhålla en modell som genererar det resultat som målfunktionen specificerar. Ett exempel är indata över studenter som läst på universitet och en målfunktion med de studenter som lyckades ta sin examen. Algoritmens uppgift är att härleda en modell som förutspår sannolikheten att en student som ansöker kommer ta sin examen.

4.1.2 Algoritmer i nyhetsflöden

I dagens samhälle kan befolkningen, det vill säga läsarna, ta till sig nyheter från en uppsjö av olika medier. Vid sidan av de analoga och digitala tidningarna har även nyhetsaggregationsplattformar i form utav till exempel Google News och Omni vuxit fram. Detta har skapat ett behov av att kunna avgränsa och sälla bland de nyheter läsaren vill konsumera (Karimi et al., 2018). Ett problem vid namn *information overload* har uppkommit där läsaren översköljs av nyheter vilket gör det svårt att avgöra vilka som är aktuella och relevanta för dem. Med hjälp av maskininlärning så kan algoritmer förutse läsarnas beteenden baserat på vad och vilka de interagerar med på webbplatsen. Med så kallade *rekommendationsalgoritmer* kan företagen bakom webbplatserna skapa ett individanpassat flöde av nyheter för varje läsare baserat på personlig data. I följande avsnitt kommer olika versioner av rekommendationsalgoritmer som används i nyhetsflöden presenteras.

Vid utformning av rekommendationsalgoritmer för nyheter ställs det höga krav på aktualitet då en nyhet kan ersättas av "senaste nytt" flera gånger om dagen (Karimi et al., 2018). Nyheters läsvärde upphör vid olika tillfällen beroende på vilket ämne som behandlas, och antalet artiklar som väljs bort av läsaren är generellt högt. Idag används olika tekniker för att tackla dessa

problem samt ge läsaren ett relevant flöde. Karimi et al. undersöker 112 artiklar som behandlar rekommendationsalgoritmer för nyhetsflöden och kommer fram till att framförallt *collaborative filtering* samt *content-based filtering*, ofta i kombination, som rekommenderas av litteraturen.

Collaborative filtering utgår från konceptet ”mun till mun” där åsikter från människor i ens omgivning ses ha stor påverkan på ens beslutstaganden (Machado Lunardi et al., 2020). Denna metod använder sig av en matris där läsaren kopplas ihop med vilket innehåll de föredrar baserat på vad de interagerar med. Genom att sedan titta på andra läsare med liknande resultat kan algoritmen ge en prognos för vilket innehåll läsaren skulle gilla. I sociala nätverk används ofta algoritmen för att skapa ett nyhetsflöde då man kan utgå från läsarens nätverk och utifrån vilka hen interagerar med visa det innehåll som de gillar. Något som pekas ut som en stor fördel med collaborative filtering är att algoritmen inte behöver någon information om innehållet i artiklarna utan endast utgår ifrån interaktionen med dem. En nackdel är däremot att en läsare som är ny på webbplatsen inte har någon tidigare interaktion med innehållet. Det skapar problem för algoritmen att koppla ihop den med andra läsare för att kunna rekommendera nytt innehåll, detta benämns i litteratur som *cold-start*-problemet.

Content-based filtering utgår ifrån att en läsare har ett enformigt intresse i vad hen väljer att konsumera för innehåll (Felfernig et al., 2018). Om läsaren visar intresse för ett visst ämne utgår algoritmen ifrån att den inte kommer ändra vare sig sin åsikt eller minska sitt intresse för ämnet i den närmsta framtiden. Baserat på vad läsaren interagerat med så kommer algoritmen därmed rekommendera liknande innehåll i framtiden. Metoden använder sig av nyckelord eller kategorier för att markera upp de olika nyheterna. Genom att gruppera nyheter länkade till samma nyckelord eller inom samma kategori så kan algoritmen rekommendera relevanta nyheter för läsaren.

En anledning till att dessa två tekniker ofta används i kombination är att content-based filtering kan lösa cold-start-problemet (Machado Lunardi et al., 2020). I och med att algoritmen inte behöver avancerad data om tidigare interaktioner med innehåll utan kan använda enstaka nyckelord eller kategorier för att guida vidare läsaren så uppkommer inte samma problem med vilket innehåll en ny läsare eventuellt kan vara intresserad av.

4.1.3 The Black Box Problem

En inlärningsalgoritm är ofta baserad på godtyckliga eller statistiska metoder för att analysera, jämföra eller dra slutsatser utifrån data. Det kräver ofta mycket data och över tid lär sig systemet att utvärdera mönster och kan applicera sin kunskap på liknande situationer. Denna inlärningsalgoritm kan visualiseras som en låda, box, på vilken det finns skruvar som justerar vad som händer inne i lådan. Inne i lådan finns alla justerbara parametrar, vilka vi inte kan se. Därför kallas transparensproblematiken ”the black box-problem” (Dignum, 2019).

En tydlig skillnad mellan maskininlärningsalgoritmer och traditionell programmering är att traditionell programmering är beroende av en expert på området som underhåller mjukvaran. Mjukvaran kan på egen hand inte identifiera vad som krävs om omständigheterna förändras och okänd indata uppkommer. Däremot kan en inlärningsalgoritm försöka hantera främmande data och identifiera när den lyckas hantera denna data och finna korrelation.

Vad som är tydligt med traditionell programmering är att systemets framgång går att identifiera. I koden framgår kriterierna av hur systemet tar sina beslut och på vilka grunder, även om det är en komplicerad kod. Den är så kallad transparent. Motsatsen riskerar att gälla när en inlärningsalgoritm ska tolkas. För att applicera maskininläring skriver en utvecklare en matematisk struktur som karaktäriserar en mängd valmöjligheter med justerbara parametrar, en algoritm. Dessa parametrar kan vara otroligt många, och justeras en stor mängd parametrar inne i en algoritm så är det svårt att hålla reda på vilka parametrar som justerades till vilken fördel (Dignum, 2019).

På grund av "the black box problem" förlorar utvecklarna en stor del av kontrollen. När det inte längre är en människa som konstruerat en modell och det finns en kunskapslucka i hur algoritmen fungerar så uppstår ett problem i att önskade etiska och moraliska sidoeffekter kan uppstå (Kearns & Roth, 2020). För att en algoritm ska ta hänsyn till vissa etiska aspekter behöver de specificeras som restriktioner eller ingå i målfunktionen. Här uppstår ett problem då etiska begrepp såsom integritet och rättvisa är alltför diffusa. En jurists eller filosofers definition är inte tillräckligt konkret för att kunna förklaras för en maskin.

Ett problem som kan uppkomma om man inför restriktioner som begränsar en algoritm i sitt beslutstagande är att resultatet hamnar längre ifrån den optimala målfunktionen (Kearns & Roth, 2020). Det är inte självklart att föredra en algoritm som följer vissa restriktioner, framför en annan som ger en person ett optimalt nyhetsflöde, exempelvis.

4.1.4 Designmodellen ART

Problematik i designprocessen av maskininläring och artificiell intelligens skriver Virginia Dignum om i sin bok Responsible Artificial Intelligence (2019). En designmodell kallad ART som riktar sig till utvecklare av AI-system presenteras i boken som ett steg i att lösa delar av problematik kring AI. Modellen ART är till för att skapa trovärdighet och tillförlitlighet i socio-tekniska miljöer där AI är involverat. Dignum menar att applicera ART på designen innebär ett socialt synsätt tillsammans med ett tekniskt synsätt på hela systemet som ska utvecklas. Alla intressenter, från företag och organisationer till användare och målgrupp behöver beaktas. ART står för Accountability, Responsibility och Transparency vilket direktöversatt till svenska blir ansvarighet, ansvar samt transparens.

Om accountability, ansvarighet, skriver Dignum att det handlar inte om ansvar för konsekvenser utan snarare om *meddelandeansvar*, vilket kommer användas som ord i rapporten. Man vill att systemet ska kunna rättfärdiga och förklara handlingar i efterhand, både till användare och andra aktörer. För att garantera att meddelandeansvar tagits så måste beslut som fattats gå att härleda från beslutsalgoritmerna. För att användare ska våga lita på systemet är det också viktigt att de kan lita på att designprocessen har varit säker, så att valmöjligheter, val och avgränsningar angående systemets mål och antaganden går att rapportera.

När man utvecklar förklaringsmekanismer i ett system är det viktigt att beakta vem det är som läser. Att ha en invecklad kod och förvänta sig att gemene man ska förstå den blir problematiskt. Dignum exemplifierar att förklaringar bör vara *kontrastiva* såsom "varför gjorde du såhär istället för det där?", *selektiva* det vill säga välja relevanta faktorer och presentera dem och förklaringar bör även vara *sociala*. Med sociala förklaringar menas att förklaringarna

anpassas efter vem det förklaras för och dennes bakomliggande kunskaper.

Responsibility, ansvar, i ART refererar till rollen människorna har till AI enligt Dignum. Det är vanligt att fråga sig om AI enskilt kan hållas ansvarigt för sina handlingar. Även om en inlärningsalgorithm kan anses ha någon form av intelligens så har den inget medvetande. Algoritmen är byggd av en anledning och med en given målfunktion, ett mål som är upprättat av en människa. I algoritmer som skrivs, med dess tillhörande system, så behöver samhällsliga och moraliska värderingar integreras, tillsammans med aktuell lag, i alla steg av utveckling och implementering. Detta innefattar allt från att samla data, designa, konstruera, implementera och evaluera systemet, (Dignum, 2019).

Dignum exemplifierar två skillnader i vem som bär ansvaret i användningen av AI. Antingen kan systemet bete sig precis som det är tänkt, och då ligger ansvaret hos användaren. Om systemet däremot inte beter sig som användaren förväntar sig kan felet och ansvaret istället ligga hos utvecklaren av systemet. Även om en inlärningsalgorithm genererar ett resultat som beror på inläring så kan inte ett potentiellt fel åläggas systemet i sig, utan det är en konsekvens som uppstått i designen av algoritmen. Här ligger en del av problematiken i ansvar, var och när övergår ansvaret hos designern till användaren? När tar problemlösningen i algoritmen en oväntad vändning och hade i så fall utvecklaren eller designern kunnat förutse detta?

Transparency, transparens, inom AI handlar om möjligheten att beskriva och förstå mekanismerna bakom vilka val systemet gör, inläringen och anpassningen till omkringliggande miljö och vilken data systemet använder. Tillförlitlighet till systemet ligger i hur transparent det är (Dignum, 2019).

Vidare beskriver Dignum att göra kod och algoritmer öppna och tillgängliga för inspektion genererar tillförlitlighet till systemet. Problematiken här ligger i att företagsstrategier och patent kan påverkas negativt. En annan aspekt i transparens är att en invecklad algorithm inte blir särskilt talande för gemene man. En enskild kodrad är nödvändigtvis inte alltför komplex, men på grund av det så kallade black box-problemet så är hela systemet i sig, och dess tillvägagångssätt, svårbegripligt. Transparens i den form av att göra algoritmerna synliga för alla intressenter är inte synonymt med generell förståelse av systemet.

Black box-problemet är svårt att komma ifrån i intelligenta system. Transparens är därför viktigt när det kommer till vad som sker vid ritbordet. Vilka värderingar, lagar, moraliska aspekter, etisk åskådning och vilken data som används är desto viktigare att presentera, snarare än att varje intressent har möjlighet att läsa källkoden, enligt Dignum. Här ligger då istället problematiken i hur detta ska förklaras. Ska utvecklarna för varje beslut som tas, varje algorithm som implementeras, redogöra för bakomliggande teori om etik och moral?

Transparens är högst subjektivt beroende på vem det är som ska förstå komplexiteten bakom ett resultat. Subjektiviteten gör det svårt att formulera ett övergripande teoretiskt ramverk då det beror på målgruppen för varje unik algorithm. Att identifiera sin målgrupp och undersöka deras förståelse, och låta det sätta gränsen för hur pass komplexa algoritmer som bör användas,

är ett sätt att angripa problemet (Kearns & Roth, 2020).

4.1.5 Angripa filterbubblor

På grund av att nyhetsflöden sorteras baserat på vad och vilka vi interagerar med så hamnar läsarna i olika filterbubblor (Kearns & Roth, 2020). Detta har visat sig bidra till polarisering inom politiken, gjort läsarna generellt sämre informerade och framförallt mindre toleranta till andra åsikter än deras egen. Kearns och Roth diskuterar huruvida det går att angripa dessa filterbubblor och minska deras effekt. En intuitiv lösning är att även rekommendera nyheter som ligger utanför läsarens filterbubbla. På samma sätt som algoritmen vet vilket innehåll som passar läsaren, så vet den också motsatsen. Att erbjuda läsarna innehåll som motsätter sig deras åsikt skulle dock antagligen främst leda till frustration. Däremot menar Kearns och Roth att istället erbjuda läsarna en möjlighet att justera sitt nyhetsflöde, från strikt sorterat till en bredare syn på vad som kan vara intressant, skulle kunna hjälpa läsarna att förflytta sig en bit utanför sin trygghetszon. De menar även att det kan öka transparensen om det tydliggörs vilken åsiktsgrupp som innehåll riktar sig mot.

4.2 Presentation av nyheter

De problem och hinder som finns för individer att ta del av nyanserade nyhetsflöden orsakas till viss del av hur nyheterna presenteras. Detta avsnitt beskriver nyhetsflödena på olika plattformar, hur skiftet från enbart tryckta papperstidningar till spridningen av nyheter och hur information på internet har påverkat flödet av nyheter och journalistiken i stort. Problem med fake news och specifikt misinformation behandlas för att slutligen presentera balans- och faktakollshemsidor som är två försök till att läsare enklare ska kunna bekräfta eller dementera fakta och påståendens relevans.

4.2.1 Internets påverkan på källkritik

Nyheter kan anses ha två dimensioner. En kvalitet, som bedöms efter faktorer såsom tillförlitlighet, val av källor och analytisk förmåga, och en vinkel. Historiskt har det visat sig svårt att samtidigt uppnå en bra kvalitet och ta upp flera vinklar i presentationen av nyheter. De traditionella papperstidningarna anses av många hålla högre kvalitet än exempelvis de korta nyhetsnotiser som framförallt skrivs på internet och som kan spridas till många trots bristande bakomliggande journalistiskt arbete (Regeringen, 2015). Papperstidningar håller generellt en högre kvalitet, men i princip alla har en politisk inriktning och få personer läser mer än en tidning av tids-, ekonomiska- eller intresseskäl (Regeringen, 2015). Att tala om en tidning som neutral säger inte heller något om att den räcker som källa eller tar upp alla aspekter (Media Bias Ratings, 2020).

Internet har delvis öppnat upp för mindre aktörer att uttrycka sig, vilket är positivt på så sätt att det ökar bredd och ger en möjlighet till större horisonter för en individ och att fler källor används. Det ökar samtidigt risken för att felaktigheter sprids eller att icke-kvalitativa nyheter får oproportionellt genomslag, vilket varit en av motiveringarna för många studier liknande vår (Benham, 2019). Det ställer högre krav på individens källkritiska förmåga då till exempel mer extrema åsikter har större chans att spridas än tidigare (Skolverket, 2021). En ytterligare fördel med internet är att det gjort det billigare att publicera material och även att ta in fler källor. Samtidigt finns kvalitativ journalistik tillgänglig, dock på ett sådant sätt att det som

sagt kan vara svårt att skilja mellan hög och låg kvalité och indelningar som vänster-höger och ytlig-detaljerad på ett nyanserat sätt.

Internet tillåter folk att hämta information om politik, debattera med andra och uttrycka sina åsikter till eliten. Sociala medieplattformar skulle kunna bekämpa spridningen av misinformation genom att prioritera mer mångfald och mindre polariserat innehåll. (Nikolov et al., 2021)

4.2.2 Plattformarnas makt över läsarna

I USA har två tredjedelar av alla vuxna en Facebookprofil och ungefär hälften använder bland annat Facebook för att konsumera nyheter (O'Neil, 2016b). Facebooks nyhetsflöde styrs av algoritmer med tillgång till tydliga nätverk av kontakter. Det skapar ett komplext flöde med ett blandat innehåll av nyheter, åsikter och reklam där prioriteringsordningen bestäms av algoritmernas bedömning huruvida innehållet bör spridas och till vilka. En vital skillnad mellan sättet som Facebooks algoritmer sätter ihop ett nyhetsflöde och hur en tidningsredaktion väljer ut sitt innehåll är att när tidningen väl är tryckt så läser alla samma innehåll. I och med det uppkommer diskussioner gällande vilka nyheter tidningen valt att ta upp och hur de porträtteras. Ett nyhetsflöde är unikt för varje läsare och därmed skapas inte samma debatt och källkritik, vilket ger algoritmerna stor möjlighet att påverka användarnas verklighetsbild. Utifrån empiriska studier som Facebooks ingenjörer utfört så har de lyckats ändra användares politiska engagemang samt deras emotionella tillstånd genom att göra förändringar i deras nyhetsflöden (Kramer et al., 2014) . Den potentiella skadan är stor och göms bakom kod som användarna inte förstår.

Makten har sedan internets introduktion, och ett allt större användande av medier via internet, skiftat från nyhetsredaktioner till publiken och till andra aktörer såsom programmerarna bakom algoritmerna. Något som beskrivs som: "... de allt mer grumliga gränserna för vad som är journalistik och journalister, introducerar nya frågor såsom vem som bestämmer vem som ska vara publiken? Journalistikens gränser innefattas numera även av publiken, algoritmerna och analysen av data". [Egen översättning] (Duffy & Peng, 2019). När digitaliseringen påbörjades genomfördes förändringar av journalistiken med syfte att bibehålla läsare. Denna anpassning av journalistiken till det digitala har inneburit en introduktion av algoritmer för att följa läsarnas intresse för att både leda de till artiklar som förutspås tilltala dem, men har också påverkat sättet nyheter produceras på i sig. Dessutom har nyhetsaggregatorer skapat nya former av marknader ekonomiskt för nyhetsbyråer. Dessa digitala förändringar ökar behovet att konstruera sätt för journalister att kunna fortsätta bedriva sitt arbete efter bästa sed och samtidigt få läsarna att ge störst tilltro åt de som gör det bäst, tekniskt, socialt och ekonomiskt. (Duffy & Peng, 2019)

Vad gäller förändringen i produktionen av nyheter har icke-traditionella medier som sagt kunnat öka sitt inflytande jämfört med andra genom sociala plattformar, såsom genom att producera emotionellt innehåll. En av de största empiriska studierna av fake news och misinformation hittills sett till antal studerade bitar av information visade att det som i studien kategoriseras som fake news både sprids djupare i fler led från ursprungskällan, samt bredare då fler upptäcker den på egen hand. De sprider det sedan vidare från sin egen sida som andrahandskälla, vilket är ett tecken på att aktörer som tidigare inte skulle kunna sprida sitt

budskap mer än till sin vänskapskrets nu har större potential att göra så (Vosoughi et al., 2018).

Avslutningsvis är en relevant aspekt vilka källor sidan ifråga samlar information ifrån. Alltså, vad som kvalificerar en tidning eller liknande att överhuvudtaget inkluderas på webbplatsen. Här råder delade meningar om vad som egentligen är journalistik, särskilt frågor kopplade till gränsen mellan aktivism och journalistik har blivit allt mer aktuella (Blanding, 2018; Elfving Persson, 2014).

4.2.3 Balans och faktakoll

Balans definieras i praktiken ofta som slutresultatet av journalistens arbete, nämligen artikeln i sig, medan författare till flertalet artiklar menar att det snarare borde vara en inställning hos journalisten eller en arbetsmetod att följa (Benham, 2019; Suarez, 2018; American Press Institute, u.d). Det finns två inställningar som förespråkas av de olika lägren. Den ena är den så kallade vetenskapliga, att återge precis det som sagts utan att ge någon kontext eller lägga in någon värdering. Den andra är att försöka värdesätta hur viktigt det kandidater säger är och låta det mest relevanta ta större plats och då ofta genom att placera uttalanden i en kontext exempelvis genom en text som: "... detta stöds av konsensus inom vetenskapen då 99% av forskarna på området ser det som den korrekta slutsatsen av problemet [KÄLLA]". Det tillvägagångssättet kan ses som en form av transparens precis som det förstnämnda. Samtidigt finns också två läger av läsare. Det ena där kontext förespråkas och ett annat där man inte vill läsa några som helst värderande ord vilket man menar öppnar upp för subjektivitet från journalistens sida (Benham, 2019). Ett sätt att bedriva den vetenskapliga sidan är att frångå det strikt återgivande med argument såsom att journalisten alltid i slutändan blir subjektiv i sitt val av vad som valts att återges (Vosoughi et al., 2018). På så sätt skapas en kontext. Det viktiga här blir då att kontextualisera på ett sätt som i sig stöds av tillförlitliga källor och resonemang, i enlighet med den vetenskapliga principen (Suarez, 2018).

De faktakoll-hemsidor med högt anseende idag som Slopes och Full Fact följer denna syn på balans och objektivitet. De utgår från vad som sagts i en förstahandskälla och anstränger sig för att tolka detta så vetenskapligt som möjligt, vilket betyder i sin kontext, objektivt och i relation till övrig relevant information på ämnet (Tompkins, 2020). De används i dagsläget av journalister och andra som vill bekräfta antaganden och påstådda fakta eller större teorier. Ett kommande användningsområde på bred front skulle kunna vara explicit i andra nyhetssidors artiklar. På så sätt kan läsaren direkt se om fakta stöds av dessa sidor och själv läsa vidare om denne tvivlar på det underliggande resonemanget för deras bedömning. Den objektivitet som ena lägret av journalister eftersträvar, finns att finna i vetenskapens vikt vid dess metod för att dra slutsatser (Suarez, 2018; American Press Institute, u.d). Genom att följa en artikels resonemang enligt denna metod med underliggande faktakoll som i sig kan åskådliggöra detta resonemangs underbyggnad, kan en väg dit skapas (Walter et al., 2019; Barrera et al., 2020). Faktakoll-hemsidors påverkansgrad har diskuterats, men majoriteten av arbetets undersökta källor är överens om att de som strävar efter att vara nyanserade hjälps av det på grund av sin källkritiska förmåga och då på så sätt att de ökar sin kunskap men inte ändrar beteenden till lika stor grad. Utöver det ställs makthavare till svars på ett annat sätt då åtminstone faktakoll-hemsidorna själva verifierar deras uttalanden, och genom att de inför högre krav som på sikt kan ge dem än större inflytande. Att faktakoll skulle få motsatt effekt har också studerats och risken anses öka med extrema åsikter och polariserade frågor, men anses liten generellt (Wood & Porter, 2019; Sippit, 2019).

Läsaren blir själv i slutändan alltid den som väljer vad den tror på, men introduktionen av faktakoll är ett sätt att uppmärksamma läsare som vill bli nyanserade på uppenbara fel och kan också fungera i lärande syfte för att se hur andra bekräftar eller dementerar fakta och resonemang, något som tros öka deras inverkan över tid då källkritik kommer mer på agendan och fler utbildats och uppmärksammats på att denna typ av hemsidor finns (Tompkins, 2020; Walter et al., 2019).

En del rapporter poängterar på denna not att ett informationsflöde som innehåller en struktur för att finna fel och påpeka dessa gör att det kan förbättras över tid. Det har gjorts analyser kopplade till att olika sidor politiskt i USA är bättre och sämre på detta, att skapa en konstruktiv feedback-loop (Faris et al., 2017).

4.3 Konsumtion av nyheter

Utöver tekniska aspekter och presentationen av nyheter så skapas problem och hinder för nyanserade nyhetsflöden av hur individerna som läser nyheterna fungerar. För att förstå vilka problem och hinder som finns, samt vilka konsekvenser de filtrerade nyhetsflödena ger, presenteras olika psykologiska och sociologiska fenomen i detta avsnitt, tillsammans med förklaringsmodeller till varför individer tar del av och tolkar nyheter på det sätt de gör. Detta för att skapa en bild av hur individer fungerar och veta vad olika verktyg som syftar till att lösa problematiken måste förhålla sig till.

4.3.1 Polarisering

Då fler individer använder sig av sociala medier som huvudsaklig källa till nyheter ökar dessa plattformars inflytande på vilka nyheter som konsumeras av vilka individer. När radikalt olika nyheter konsumeras av olika individer ökar polariseringen i samhället, till följd av de skilda verklighetsuppfattningarna som skapas (Spohr, 2017). I Spohrs artikel framhävs två huvudorsaker till den växande polariseringen. Första orsaken påstås vara algoritmer som placerar oss i ekokammare och filterbubblor som filtrerar de nyheter vi möter online. Under samma orsak presenteras även fake news som en kritisk faktor. Den andra huvudorsaken har diskuterats längre tid och har rötter i psykologi och beteendevetenskap. Denna orsak innebär att fenomen som bland annat konfirmeringsbias skapar polarisering både online och offline. (Spohr, 2017)

Den viktigaste konsekvensen av polarisering tycks vara bristande mångfald i nya synvinklar och åsikter. I ett tidigt stadie ansågs internet primärt öka expansionen för politisk mångfald eftersom det skulle överbrygga sociala och geografiska begränsningar. Dock var det optimistiskt då den stora översvämningen av information snarare ledde till att människor blev mer selektiva i sina val av nyhetskanaler. (Spohr, 2017)

Det pågår en diskussion kring vad som påverkar polarisering, informationsintag och att människor blir mer selektiva i informationen som de tar till sig. Enligt Spohr (2017) behövs bland annat ideologisk polarisering, algoritmer och *selektiv exponering*, vilket förklaras vidare i avsnitt 4.3.3, undersökas vidare för att framställa en policy kring ämnet. Sociala medier så som Facebook och Google måste ta mer ansvar över fake news och desinformation. Människor

behöver också bli medvetna om att nyhetskonsumention bör vara en aktiv process eftersom ideologiskt skiftande, kvalitetsrikt innehåll inte finner oss bara för att vi är online (Spohr, 2017).

4.3.2 Informationsbegränsning

Sociala relationer bidrar med bristfällig ideologisk mångfald oavsett om kontakten sker fysiskt eller via internet. Viss forskning tyder på att internet exponerar individer för bredare spann av synvinklar än klassisk interaktion med grannar, kollegor eller familjemedlemmar medan andra rapporter förespråkar det motsatta. Sammantaget finns det inte empiriska bevis för att kunna dra någon slutsats om huruvida sociala medier totalt sett begränsar eller vidgar bredden av konsumerad information. Det finns flera orsaker till denna konflikt inom forskningen. Ett problem är vaga begreppsdefinitioner som gör det svårt att jämföra olika forskningsresultat. Detta går inte heller att göra när olika plattformar har studerats på olika sätt och även att det har varit för smalt fokus på data från mycket aktiva användare, vilka är icke-representativa. Ett ytterligare problem är den dynamiska naturen där plattformarna oannonserat justerar sina filtreringsalgoritmer. Detta leder till begränsning av generalisering av empiriska resultat över tid, eftersom samma plattform fungerar annorlunda och således inte kan behandlas som samma. (Kitchens et al., 2020)

Konceptet informationsbegränsande miljöer definieras dels som sociala nätverks dragning till samma saker, dels som att filtreringsalgoritmer begränsar informationskällorna som individer väljer att konsumera, skyddar dem från åsiktsutmanande information och uppmuntrar dem att anta mer extrema synvinklar. Detta koncept fångar in huvudproblemen kring ekokammare och filterbubblor. (Kitchens et al., 2020)

4.3.3 Selektiv exponering

I artikeln "Fake news and ideological polarization: Filter bubbles and selective exposure on social media" diskuteras selektiv exponering som ett fenomen som innebär att individer tenderar att konsumera information som stämmer överens med deras tro och verklighet, och genom detta undvika information som ifrågasätter eller utmanar ens åsikter (Spohr, 2017). Det finns överensstämmande bevis på att detta fenomen har en påverkan på polarisering. Spohrs artikel hänvisar till en studie där mänskliga klick begränsade mängden tvärgående innehåll mer än vad algoritmerna gjorde (Bakshy et al., 2015). Ytterligare en undersökning visade på att även när människor blir visade en utmanande sida av sina åsikter ledde det ändå till att personen i fråga betedde sig selektivt. Detta kunde förklaras med att de antingen undvek den andra informationen eller undersökte den i syfte att hitta fel. Flera undersökningar har nått slutsatsen att människor primärt samlar information pragmatiskt och i enlighet med överlevnadsinstinkt, och sekundärt genom att vara kritiska och ha på "lögndetekorn".

4.3.4 Konfirmeringsbias

Konfirmeringsbias kan förklaras ur dels ett kognitivt, dels ett mer sociologiskt beteendeperspektiv. En numera vedertagen definition av konfirmeringsbias är att när någon väl har format en tro på något så är den relativt motståndskraftig mot motargument (Ross & Anderson, 1982). Festinger (1957) såg det som ett sätt att undvika kognitiv dissonans och hjälpa att behålla och bekräfta vår känsla av egen identitet (Williams et al., 2016).

Gruppsyck påverkar individer så att de ideologiska biasar som finns inom gruppen behålls och det arbetas aktivt för att inte tas in influenser som inte är enligt dessa. Kopplat till konfirmeringsbias kan dagens teknik för nyhetskonsumtion möjliggöra att ens sociala sfär dras åt samma håll. (Ling, 2020)

En ytterligare dimension av huruvida individer tar till sig nyanserade nyheter eller inte avgörs av om de anstränger sig för att ta del av olika perspektiv. På detta tema har undersökningar genomförts på personer som upplever oro. Oro tycks inte öka tiden för informationssökning men kan öka kvaliteten på sökningarna och därmed öka lärandet. (Valentino et al., 2009)

4.3.5 Tillgänglighetsbias

Ytterligare ett psykologiskt fenomen som påverkar polarisering är *tillgänglighetsbias*, vilket innebär att människor tenderar att tro att saker kommer hända igen eftersom de har hänt i det förflutna. Det blir en mental genväg där beslut fattas beroende på känslor, åsikter från nära och kära samt bilder som skapar enkla intryck i våra hjärnor. Genvägarna tas eftersom det minimerar ansträngningen som krävs av hjärnan för att genomföra en viss uppgift (The Decision Lab, 2021).

4.3.6 News finds me

En tredje observation kring människors beteende vid nyhetskonsumtion är "News finds me". Begreppet innebär att människor blir så matade med information och nyheter att de inte aktivt måste söka efter dem själva längre. De tror sig vara uppdaterade om viktiga nyheter genom att endast passivt läsa det som kommer till dem genom generellt internetanvändande och sociala medier. En följd av att människor förlitar sig på att de blir välinformerade utan att aktivt söka efter nyheter kan vara att de får en brist av ideologisk mångfald (Spohr, 2017).

4.4 Etiska ramverk i verkligheten

Genom att undersöka redan existerande etiska ramverk som brukas inom området AI finns möjlighet att reda ut vilka delar av dem som kan appliceras på det avsmalnade området maskininlärningsalgoritmer som filtrerar nyhetsflöden. Två omfattande etiska ramverk undersöktes och beskrivs i följande avsnitt. Senare i analysen presenteras vilka delar som kan användas som ett eget förslag på riktlinjer till framtida etiska ramverk för detta område.

EU:s riktlinjer har tagits fram av deltagare fördelade mellan företagsrepresentanter och forskare som beskrivs genom sin forskningstitel på respektive universitet. Forskarna har ofta en teknisk bakgrund men även flera samhällsvetenskapliga discipliner finns representerade. Alla namn på deltagarna finns skrivna ihop med ett företagsnamn eller universitet och stundtals även ihop med personens specialområde.

IEEE dokumentets författare är främst forskare inom datateknik, AI, filosofi/etik, policy av olika slag och information. Forskarna sträcker sig från professorer till universitetsstudenter, där majoriteten är på doktorandnivå och uppåt. Ämnesmässigt är de flesta forskarna tekniskt eller filosofiskt inriktade.

4.4.1 EU kommissionens riktlinjer

En expertgrupp bestående av branschaktiva och forskare inrättades 2018 av Europeiska kommissionen för att ta fram etiska riktlinjer för tillförlitlig AI. EU kommissionens ramverk syftar till att ge riktlinjer för AI-tillämpningar i allmänhet och därmed skapa en generell grund för att forma tillförlitlig AI (Europeiska kommissionen, 2018). Tanken bakom ramverket är att skapa en utgångspunkt för diskussioner kring tillförlitligt europeisk AI samt främja forskning och reflektion kring ett etiskt ramverk för AI på global nivå. Allt i detta avsnitt är hämtat från EU kommissionens etiska riktlinjer för tillförlitlig AI (Europeiska kommissionen, 2018).

Ramverket utgår från tre komponenter som presenteras som krav att inkluderas under ett AI-systems hela livscykel.

- Den bör vara laglig
- Den bör vara etisk
- Den bör vara robust

Laglig innebär att den följer de lagar och förordningar som finns. Etisk syftar till att garantera upprätthållning av de etiska principerna. Robust bör utvärderas ur både en teknisk och samhällelig synvinkel av den anledning att AI trots goda intuitioner kan orsaka skada. Robust programmering innebär att programmet ska kunna förebygga och hantera vissa typer av fel. Det är grundläggande att de tre komponenterna är balanserade och kompletterar varandra.

Utifrån de tre komponenterna är ramverket uppbyggt av tre kapitel som associeras till varandra. Uttryckligen tas dock inte första komponenten ”laglig” upp. Första kapitlet identifierar de etiska principerna som bör tas hänsyn till med utgångspunkt från de grundläggande rättigheterna. Dessa etiska principer behöver upprätthållas för att säkerställa etisk och robust AI. Andra kapitlet övergår de etiska principerna i sju krav som det anses att AI-system bör uppfylla. Förutom det skildras även tekniska samt icke-tekniska metoder som anses användbara vid genomförandet av de sju kraven. Dessa krav anses primärt vara lika viktiga, men då de tillämpas inom olika områden bör de anpassas efter sammanhanget. Tredje kapitlet är det mest konkreta och innehåller en bedömningsmall för att utvärdera tillförlitlig AI i praktiken. Det bör betonas att mallen är icke uttömmande och bör därför skraddarsys efter vilket system som tillämpas. Förutom dessa tre kapitel innehåller dokumentet ett sista kapitel som definierar några av de möjligheter och farhågor som AI-system kan bidra till.

Det grundläggande målet med ramverket är att uppnå tillförlitlig AI. Expertgruppen menar att mänskligheten endast kommer kunna tillvarata fördelarna med AI ifall de känner fullt förtroende för tekniken, samt människorna bakom tekniken. Tillförlitligheten är även en grundpelare för fortsatt utveckling, spridning och användning av AI. Expertgruppen menar på att samhällen och människor endast kommer uppnå denna tillit om det finns en tydlig och heltäckande ram för AI.

Expertgruppen erkänner även de positiva effekter som AI redan har och kommer ha i framtiden, men betonar betydelsen av att vara medveten om de risker som kan förekomma. Förutom detta understryks även betydelsen av att anpassa riktlinjer efter vilken situation AI-systemet befinner sig i och att detta ramverk endast utgör en grund för att främja tillförlitlig AI i

allmänhet. De anser att det bör undersökas huruvida en sektorbaserad strategi är behövlig, då utformningen av AI-system är så pass beroende av sammanhanget.

Expertgruppen anser även att AI-systemen bör vara människocentrerade, vilket innebär att de bör användas i människans tjänst med avsikt att förbättra människors välfärd och frihet. De menar på att AI kan förändra samhället kraftigt och att det bland annat är en lovande metod för att öka det mänskliga välbefindandet.

I dokumentet lyfts även att det är väsentligt att involvera alla berörda parter i AI-systemets hela livscykel, genom att bland annat främja utbildning för att skapa medvetenhet om tillförlitlig AI. Expertgruppen beskriver även att det finns forskning som visar på att det är avgörande att ha en ledning som är engagerad för att lyckas skapa en förändring och uppfylla de krav som lyfts i ramverket (Johnston et al., 2017). Även acceptansen för införandet av nya processer hos alla parter i organisationen lyfts som betydelsefullt. Utifrån detta rekommenderas därmed processer som skapar delaktighet.

Slutligen skriver expertgruppen att deras mål med ramverket innebär att skapa en kultur av tillförlitlig AI för Europa och genom det möjliggöra att ta del av fördelarna med AI. De fördelarna bör då hjälpa till att garantera respekt för mänsklighetens grundläggande värden: Grundläggande rättigheter, demokrati och rättsstatsprincipen. De understryker även att ramverket i takt med att teknik, samhälle och kunskap utvecklas kommer uppdateras för att säkerställa dess relevans i förhållande till tiden.

4.4.2 Standarder från IEEE

Institute of Electrical and Electronics Engineers (IEEE) är en organisation som består av över 396 000 medlemmar från över 160 olika länder som tillsammans inför tekniska standarder, publicerar vetenskapliga tidskrifter, håller i konferenser och andra professionella och lärande aktiviteter inom teknisk innovation för en bättre framtid (Institute of Electrical and Electronics Engineers, 2021). Bland många andra publiceringar inom teknisk design har de publicerat dokumentet "Ethically Aligned Design (EAD)" (2019). I EAD behandlas etiken kring Autonoma och Intelligent System (A/IS) med mål om att öka förståelsen för A/IS, belysa vikten av att inkludera etik i utvecklingen av A/IS samt att införa situationsanpassade lösningar i form av ramverk. EAD kommer att behandlas nedan. A/IS innefattar i dokumentet allt ifrån autonoma robotar, tekniker till algoritmer. Dokumentet är organiserat efter specifika problem rörande A/IS följt av rekommendationer som betraktas som eventuella lösningar till de specifika problemen. Lösningarna som tas upp grundar sig i åtta olika generella principer listade i EAD, där nummer fem, transparens, främst kommer behandlas i denna studie. Samtliga rekommendationer i EAD riktar sig mot så kallade "stakeholders", likt svenskans intressenter. EAD:s definition på en så kallad stakeholder vad gäller A/IS är någon som är inblandad i forskning, design, tillverkning eller information kring A/IS inom universitet, organisationer, stater eller företag som tillsammans möjliggör att A/IS blir verklighet för samhället (EAD, 2019).

Transparens innefattar följande tre delar enligt EAD; spårbarhet, förklarbarhet och tolkbarhet. Hos A/IS är transparens viktigt av olika anledningar för varje enskild intressentgrupp menar EAD. Vidare presenteras fem olika intressentgrupper och varför transparens anses vara viktigt

för var och en av dessa grupper. Tre av dessa fem intressentgrupper är *användare*, *skapare* och *allmänheten* som anses vara speciellt viktiga för denna rapport. De övriga två är *olycksutredare* samt *de som är inblandade i rättsliga processer* gällande A/IS. De två sistnämnda är inte direkt kopplade till problematiken kring nyhetsflöden och kommer därför inte behandlas under detta avsnitt.

För *användaren* av ett A/IS är det viktigt, enligt EAD, att systemet ska vara kapabelt till att förklara sitt resonemang på ett, för användaren, begripligt sätt utan några försvårande tekniska termer. Alltså att ge användarna en ökad förståelse av vad A/IS gör och varför det gör något. Detta behandlas mer djupgående i sektion 2, problem 2, där behovet av transparens från implementeringen till spridningen av A/IS belyses. EAD vill, ifall systemet inte kan förklara sina resonemang på ett tydligt och begripligt sätt för användaren, att användaren ska ha möjlighet att vända sig till en utvecklare eller ansvarig för systemet för en förtydligande förklaring. Vidare refererar denna del av EAD till General Data Protection Regulation (GDPR) som säger att för automatiserade beslut baserat på personliga data har individen, användaren, rätten till att överklaga ett automatiserat beslut baserat på personliga data samt rätten till en förklaring till det automatiserade beslutet. I Europaparlamentets och rådets förordning står följande: ”Denna form av uppgiftsbehandling bör under alla omständigheter omgärdas av lämpliga skyddsåtgärder, som bör inkludera specifik information till den registrerade och rätt till mänskligt ingripande, att framföra sina synpunkter, att erhålla en förklaring till det beslut som fattas efter sådan bedömning och att överklaga beslutet” ((EU) 2016/679).

Vad gäller *skaparna* av A/IS, som enligt EAD inkluderar både de som godkänner A/IS och de som certifierar dessa system, är det viktigt, enligt EAD, att systemen ska vara transparenta vad gäller dess processer och hur systemen hanterar data-input. Till sist vad gäller *allmänheten* anses transparens vara viktigt för att bygga upp ett förtroende för A/IS. Något som även kan ses som betydande i EU kommissionens riktlinjer, se 4.4.1.

Vad EAD har som rekommendationer för att bland annat lösa liknande problem är följande; införa nya standarder som möjliggör mätbar transparens hos A/IS. Dessa typer av standarder, menar EAD, kommer även underlätta en bedömning av systemets transparens redan under utvecklingsstadiet och på så sätt ge utvecklarna en bra chans att behandla och möjliggöra transparens vid ett tidigt stadie.

Något som även behandlas i EAD under ”A/IS for Sustainable Development” är i linje med vad denna rapport bland annat vill belysa, nämligen betydelsen och vikten av sanningsenlig information i ett demokratiskt samhälle. EAD menar att dagens informationsinfrastruktur saknar transparens som försvårar kunskapen att veta, och i vissa fall omöjligt att ta reda på, bland andra följande punkter:

- vem som skapade och finansierade spridningen av informationen
- andelen av informationen som spridits av bottar
- samt vem som finansierar dessa bottar.

Dessa tre punkter är tagna från en lista bestående av sex punkter, de övriga tre anses inte vara relevanta för denna del av uppsatsen.

De rekommendationer som EAD kommer med är bland annat att öka transparensen vad gäller A/IS som används vid profilering för att minska manipulation av användarna. Vidare vill EAD se A/IS tekniker som kan detektera lögnaktig information. Dessa A/IS ska i sin tur vara övervakade av demokratiska organ vilket, enligt EAD, minimerar risken för censurering. Man vill avslutningsvis förplikta företag att vara mer transparenta vad gäller bland annat deras algoritmer och finansiering (EAD, 2019).

5 Resultat från intervjuer

I detta kapitel sammanfattas och belyses de viktigaste och mest relevanta delarna från de tre olika intervjuerna som genomfördes. Samtliga delar från de olika respondenterna som tas upp i detta kapitel kan hittas under bilagor. Där hittas avsnitt med transkriberade delar från de behandlade delarna från samtliga intervjuer. Intervjuerna genomfördes i samma ordning som rubrikerna följer. Resultaten från intervjuerna blev av olika djup, detta då författarna under Behdadis och Crnkovics intervjuer hade mer kunskap och teori kring ämnet jämfört med de Fine Lichts intervju. Därför blev resultaten från Behdadis och Crnkovics intervjuer mer djupgående än från de Fine Licht.

5.1 Karl de Fine Licht

Snöbollsurval är något som tas upp i avsnitt 3.3.1, och var precis vad en del av resultatet blev från intervjun med de Fine Licht.

Vid ett tidigt stadie under intervjun, vid frågan om det var något speciellt gällande etik som borde tas i beaktning i arbetet, nämnde de Fine Licht normativ etik. Vidare tipsade de Fine Licht om att istället för att prata om det stora begreppet etik, som var tanken från början, skulle man istället fokusera på de normativa teorierna som finns och koppla dem mer direkt till problematiken kring AI. Något som behandlas mer ingående vid diskussionen kring definitionen av etik, under 1.6.

Vidare in i intervjun frågade författarna om de Fine Licht ansåg att det fanns något speciellt viktigt att ta med i ett eventuellt etiskt ramverk gällande nyhetfiltrering, och isåfall om han hade några allmänna tips om vad som skulle kunna vara bra att ha med. De Fine Licht tipsade då om att granska de stora, redan existerande, ramverken. Framförallt lyftes två ramverk upp, EU:s och IEEE:s. Båda dessa dokument behandlas under rubriken etiska ramverk i verkligheten 4.4. EU:s ramverk tyckte de Fine Licht var bra då det fanns mycket intressant diskussionsunderlag kopplat till denna. Vidare in i intervjun belyste de Fine Licht att de stora ramverken liknar varandra då det ofta är samma personer som sitter i de olika grupperingarna. Mot slutet av intervjun ställdes frågan ifall han hade några tips på personer som skulle kunna vara bra att kontakta för fortsatta intervjuer. Då kom namnen Gordana Crnkovic och Dorna Behdadi upp. Dessa två skulle, enligt de Fine Licht, komplettera varandra på ett bra sätt inom området etik. Här kom snöbollsurvalet att spela sin roll.

5.2 Gordana Crnkovic

Ur denna intervju lyftes många intressanta aspekter. Först och främst ansåg Gordana att ett stort problem inom nyhetskonsumering är att mångfalden försvinner om algoritmer gör ett urval, istället för att människor kommunicerar med varandra. Vidare förklarade hon att de stora företagen som äger de artificiella intelligenta systemen bestämmer hur det går till när algoritmerna filtrerar. Idag är de relativt självständiga i de besluten som fattas och överallt i hela världen diskuteras det att alla andra stakeholders också måste vara delaktiga i beslutsfattandet. Hon benämnde information som det centrala och att ifall endast de stora företagen kan ta bort, välja ut och sammanställa information så fås en helt viss bild som möjligen inte innefattar alla infallsvinklar.

Övriga medborgare vill också förstå och kunna påverka vad som sker bakom kulisserna. Hon ansåg således att det största problemet är att övriga människor inte är delaktiga i hur detta går till, speciellt eftersom det kan påverka demokratin. Vidare betonade hon att de stora tekniska företagen vill göra etiska handlingar men att det är en fråga om involvering från alla intressenter, vilket kan komma från en lång process av förhandling och diskussioner där de inblandade behöver förstå varandras intressen.

Företagens största hinder för att göra etiska handlingar är att de har vuxit organiskt och gått från små företag till enorma jättar, vilket har lett till att de ekonomiskt har blivit större än vissa stater. Att företag kan besitta så stor makt är nytt i historien och det har aldrig funnits företag som har haft så stor påverkan på demokratin förr. Gordana menar på att alla krafter i samhället, inkluderande de stora tekniska företagen, måste balanseras för att kunna skydda de demokratiska processerna. Gordana syftar dock på att de tekniska företagen är de som i första hand bär ansvar för att minimera spridningen av missinformation. Anledningen till detta ansåg hon vara att de besitter kunskapen om algoritmerna och kan hjälpa de andra involverade att förstå hur de fungerar.

Angående ansvar betonade även Gordana att slutanvändarna inte kan hållas ansvariga. Hon menade på att konsumenterna tar de enklaste vägarna för att ta till sig information, då de varken besitter tid eller energi för att konsumera för mycket nyheter. Det användarna kan göra är att balansera sig med olika källor från olika länder och med olika typer av information.

Slutligen underströk hon vikten av att AI blivit stort inom etik och att många numera diskuterar etiska problem rörande AI. Hon menar på att AI är en fantastisk teknologi och att många ser stora fördelar med den. Därefter nämnde hon att AI likt alla mäktiga teknologier kommer med stora risker och att de negativa konsekvenserna som kan komma ur AI måste förebyggas.

5.3 Dorna Behdadi

Ur denna intervju föddes många nya idéer som kommer att behandlas ytterligare vid diskussionen kring hur etiska ramverk ska etableras, under 7.2.6. I början av intervjun diskuterades frågan om A i ART, alltså accountability (ansvarfrågan). Att lägga ansvaret hos användaren ansågs vara en för enkel väg att gå, påstod respondenten. Det diskuterades vidare och ansågs även orättvist då denne skulle behöva sätta sig in i ett långt juridiskt dokument med försvårande tekniska termer, i värsta fall kod, och förväntas förstå innehållet och på så sätt ha en tillräcklig grund för att kunna ta sitt eget ansvar.

Detta ledde till att T, transparency (transparens), i ART nämndes för första gången i intervjun. Här diskuterades att transparens inte spelar någon roll ifall det är svårbegripligt. Ifall det råder transparens, i form av att kod och juridiska dokument är lättillgängliga, men samtidigt råder brist i den typ av transparens som är kopplad till begriplighet, spelar det inte någon stor roll om dokumenten är lättillgängliga, menar respondenten. Vidare in i intervjun menar respondenten att det blir likt ”tomma ord” som presenteras ifall användaren inte kan tyda det. Att något är transparent kan ha olika betydelse och olika vikt, vilket kommer att behandlas i rapportens analys.

Vidare in i intervjun besvarades frågan ifall respondenten trodde att det skulle räcka med ett välformulerat etiskt ramverk som man uppmanar de stora aktörerna att följa för att få dem att faktiskt ta till sig innehållet och följa ramverken. Detta trodde inte respondenten skulle räcka, men ansåg att det skulle vara ett bra första steg. Vidare diskuterades frågan hur det skulle vara möjligt att reglera företag till att följa etiska ramverk rent juridiskt. Även om respondenten själv uttryckte att det var lite långt ifrån dennes område så kom förslaget och idén om en slags tredjeparts certifiering upp, vilket analyseras vidare i avsnitt 7.2.8.

Idéen om en certifiering återkommer senare under intervjun och kopplas samman med transparens. Tankar om att, istället för att användaren möts av en lång försvårande text som kryptiskt förklarar att det pågår filtrering, kunna ha en liten informativ ruta tog då upp. Om det pågår filtrering så visas rutan i samband med att användaren ansluter till sidan, föreslår respondenten. Idéer om att en certifiering skulle kunna vara ett mått på uppnådd transparens uppkom från detta.

Mot slutet av intervjun diskuterades en eventuell knapp. Denna skulle ge användaren möjlighet att enkelt själv kunna välja att bli presenterad en helt ofiltrerad version av sidan. Knappen som förslag på lösning analyseras vidare i avsnitt 7.2.7.

6 Resultat från fallstudie av Improve The News

Syftet med fallstudien är att undersöka om webbplatsen *Improve The News* (<https://www.improvethenews.org/>) kan anses vara ett verktyg för att hjälpa människor att ta till sig nyheter på ett nyanserat sätt. Samtidigt ska det även undersökas om webbplatsen tar hänsyn till användarens integritet och om den implementerade AI:n är etiskt försvarbar.

6.1 Vad är Improve The News?

Som nämnt i bakgrunden är Improve The News, förkortat ITN, en webbplats för nyhetskonsumtion. Webbplatsen är ett icke-vinstdrivande projekt som grundats med utgångspunkt i frågan hur polarisering uppstår i samhället.

I februari 2021 medverkade grundaren Max Tegmark i en videointervju med podcasten StarTalk (deGrasse Tyson, 2021), i avsnittet "Improving the News with Max Tegmark". Där förklarar Tegmark hur de använt sig av maskininlärning för att klassificera de artiklar som webbplatsen publicerar. Tegmark berättar att de använt sig av en datasamling tillhandahållen av The New York Times där de låtit människor läsa igenom 1,4 miljoner artiklar och klassificera dem inom ungefär 600 olika kategorier. Genom att mata maskininlärningsalgoritmer med datasamlingen som indata, med mål att uppnå samma resultat av klassificering som människorna, har de tränat algoritmerna att kunna klassificera artiklar, vilket Tegmark klargör i intervjun med deGrasse Tyson. Inom vissa kategorier har även andra webbplatser använts för att klassificera artiklar, vilket tas upp i avsnitt 6.2.1.

I intervjun i podcasten StarTalk går de utöver den tekniska aspekten in på webbplatsens bakomliggande syfte och filosofi. Tegmark kritiserar storföretag som använder maskininlärning för att klassificera sina användare och kontrollerna vilket innehåll de ser. Tegmark säger även att företagen bär ansvar för vilka konsekvenser deras algoritmer har, precis som ett tobaksbolag bär ansvar för den skada som deras produkt gör på kundernas hälsa. Däremot beskyller Tegmark inte företagen för att avsiktligt bidra till polarisering utan menar att det är en sidoeffekt av en teknik vars avsikt är att hålla kvar läsarens uppmärksamhet för att tjäna pengar på reklamannonser. Ytterligare material från intervjun kommer användas i analysen av Improve The News och refereras i fortsättningen till som *StarTalk-intervjun*.

En del av filosofin kring ITN är att en medveten konsumtion av nyheter precis som en medveten konsumtion av mat är att föredra, med stöd i Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines (2020) av Bago et al. som länkas på webbplatsens FAQ. Arbetet menar att falska nyheter blir mindre betrodda då läsaren får tid att tänka på hur trovärdig den är jämfört med vid kortare tid. Det verkar logiskt men är inte självklart och det är en av flera faktorer som är kritiska för att webbplatsen ska nå sitt mål. Här går det dock att nämna att Tegmark uttalat att webbplatsen är en del i ett större led av försök att bekämpa polariseringen, och att de som blir inspirerade av sidan uppmuntras att förbättra den då källkoden är öppen.

6.2 Hur fungerar Improve The News?

Webbplatsen är en nyhetsaggregator som, istället för att innehålla artiklar, länkar till artiklar som läsaren själv får klicka sig vidare till och hänvisas därefter till respektive nyhetskanal som äger artikeln. På ITN finns reglage som går att positionera godtyckligt där användaren ges möjlighet att välja från vilket perspektiv nyheterna ska vara skrivna. Dessa reglage är indelade i sex kategorier.

- Politisk hållning: vänster - höger
- Hållning till etablissemang: kritisk - för
- Stil på text: provocerande - nyanserad
- Djup: ytligt - detaljerat
- Hållbarhetstid: kort - lång
- Aktualitet: gammal - ny

Parametrarna kan regleras oberoende av varandra, vilket möjliggör ett stort antal olika kombinationer av inställningar. Det är dock svårt att beräkna exakt hur många möjliga kombinationer av inställningar som är möjliga, då skjutreglagen inte har ett givet antal lägen, utan ställs in godtyckligt och kan stanna var som helst längs axeln. En bild på dessa reglage återfinns i figur 1.



Figur 1: Slidebars på improvethenews.org

Varje gång användaren ändrar någon parameter så laddas webbplatsen om och ett nytt flöde med rubriker till artiklar presenteras. Artiklarna presenteras utifrån olika kategorier, här översatta till svenska. Kategorierna är Världen, Politik, Hälsa, Brott och Rättvisa, Vetenskap och Teknologi, Sociala Frågor, Sport, Pengar, Underhållning, Miljö/Energi, Militär, Kultur,

Väder och Media. Varje gång användaren klickar på en länk så öppnas ett nytt fönster eller ny flik där läsaren hamnar på respektive nyhetsmedias artikel.

6.2.1 Reglage

Det finns en förklarande text på ITN om hur reglagen fungerar och tanken bakom dess existens. Texten om reglagen på webbplatsen är en introduktion som ger läsaren en inblick i hur kategoriseringen är uppbyggd och texten gör även en ansats till att motivera läsaren att våga testa att ändra inställningarna vid användning av ITN.

För det reglage som kallas Political Stance (Politisk Hållning) så värderas artiklarna i hur höger- kontra vänstervridna medierna är. Denna värdering är gjord med hjälp av en webbplats som heter AllSides (2021). Även denna webbplats är en nyhetsaggregator som presenterar nyheter genom att tydligt kategorisera utefter uttalad politisk hållning. AllSides har en flik som listar över 800 nyhetsmedier och dess ranking på en politisk skala som går från vänster till höger. Här kan man som användare av webbplatsen klicka i ifall man håller med eller inte om den politiska hållningen som AllSides presenterar. På Improve The News står det att denna lista är det som i huvudsak använts för att kategorisera nyheterna efter vänster- eller högerpolitik.

Establishment Stance (Hållning Till Etablissemangen) och dess reglage går från "kritisk" till "för". Med etablissemangen menas en liten del av populationen som innehar mycket makt. Artiklarna kategoriseras utefter vad respektive nyhetsmedia anses ha för relation till bestämmande organ, alltså hur nära de har till "makten". Har nyhetskällan en tradition av att acceptera eller utmana ståndpunkter från bestämmande organ såsom regering eller inflytelserika företag? För att ranka artiklar i denna kategori så framgår det att Improve The News har använt Wikipedia tillsammans med en webbplats som heter Swiss Policy Research (2021). Webbplatsen tillhandahåller en sida kallad The Media Navigator, som likt AllSides listar medieföretag baserat på politisk tillhörighet och deras relation till makt. Denna webbplats listar dock endast 72 stycken medieföretag.

När det kommer till vilken stil på texten som önskas så finns två kategorier, Writing Style (Stil På Text) och Depth (Djup). Stil på texten går från nyanserad till provocerande. Med nyanserad menas en respektfull och saklig ton i texten, medan den provocerande texten kan innehålla personangrepp, slag under bältet och häckande av meningsmotståndare. Reglaget som reglerar artiklarnas djup syftar till att ge läsaren möjlighet att välja hur djupgående texten ska vara, från "Breezy" till "Detailed". Med breezy menas att texten rör ett ämne på ytan, utan att göra några djupare analyser eller verifieringar av de påståenden som kan förekomma i texten. I motsats till detta finns även alternativet med mer detaljerade texter, där artiklar har mer pålitliga källor, djupare analyser och bekräftade uppgifter samt stundtals en mer akademisk stil på texten.

Det finns en möjlighet att välja hur tidlösa artiklarna ska vara. Shelf-life (Hållbarhetstid) och dess reglage går från kort till lång och kategoriserar artiklarna utefter hur länge de anses vara relevanta. Den korta tiden på "hyllan" syftar på aktuella ämnen som kan förändras med tiden, medan de som har lång tid på hyllan är av sådant slag att de kan lämna ett djupare avtryck på samhället och är mer analytiska. Reglaget för Recency (Aktualitet) listar artiklar utefter

hur nyligen de är publicerade.

Det är endast de reglage som representerar politisk tillhörighet och syn på etablissemangen som har någon sorts förklaring direkt på ITN angående hur kategoriseringen har gått till.

6.2.2 FAQ

På webbplatsen finns en flik som heter FAQ, där några frågor om Improve The News besvaras. Inledningsvis står det att ITN är en nyhetsaggregator som utvecklats av en grupp forskare på bland annat MIT, skapad för att ge läsaren kontroll över sin nyhetskonsumtion. Därtill så länkas det till en video på YouTube, (<https://www.youtube.com/watch?v=PRLF17Pb6vo>), där Tegmark förklarar vad webbplatsen gör och syftet med den.

I videon berättar Tegmark att ITN hjälper läsaren att själv bestämma sina åsikter genom att ta del av information från ett stort spann av perspektiv. Därefter förklarar han hur reglagen fungerar och varför läsaren bör testa att reglera dem åt olika håll. Han framhäver reglaget som filtrerar nyheter efter huruvida de är för eller kritiska mot etablissemangen, och menar att detta är särskilt intressant då de vanligaste och mest populära nyhetsmedierna sällan kritiserar regering och stora väl ansedda företag. Han menar att även om enskilda personer kritiserar mycket i media så är det sällan som ett helt styrande organ blir det. Han berättar även att dessa reglage kan få läsarna att se nyheter som inte vanligtvis visas bland de vanligaste medierna. I videon framgår ingen kritik om, eller påtalade brister hos, Improve The News. Däremot uppmanar Tegmark att användare återkopplar sin upplevelse av webbplatsen genom ett formulär för framtida förbättring och utveckling. Efter den inledande texten och videon så följer svar på ett antal frågor:

”Hur fungerar det?”

Denna fråga besvaras med att det kommer läggas ut en öppen källkod på GitHub, som är en webbplats där utvecklare kan ladda upp sina koder, så snart koden är accepterad för publikation.

”Kommer inte detta bidra till filterbubblor?”

Som svar på den frågan följer en diskussion om att det finns vetenskaplig litteratur om hur maskininlärningsalgoritmer på bland annat Facebook och Google har polariserat och delat samhället. De menar att samhället delas in i grupper där respektive grupp bara får idéer som de redan håller med om presenterade för sig. Därefter ställs en förtydligande fråga: om en nyhetskonsument får medveten kontroll över vilket perspektiv de vill läsa nyheter från, kommer inte det förstärka problematiken med att få sina egna idéer bekräftade?

Som svar på den förtydligande frågan hänvisas det till den studie Bago et al. gjort som heter Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines (2020). Resultatet i studien påvisar att människor blir mindre mottagliga för ”fake news” och partiskhet när de får enkel tillgång till information av stor räckvidd. I studien framgår det även att många människor är intresserade av att nås av åsikter som inte är linje med deras egen, om dessa är uttryckta på ett nyanserat och respektfullt sätt. Problemet är dock att människor upplever att de sällan träffas av information presenterat på det sättet.

”Hur lyder integritetspolicyn?”

Integritetspolicyn finns under en egen flik och förtydligas i nästkommande avsnitt.

”Hur kan man kontakta Improve The News med återkoppling?”

Eftersom webbplatsen är under uppbyggnad så uppmanas användarna till att återkoppla hur de upplever ITN, och detta kan man göra via ett formulär på webbplatsen.

6.2.3 Integritetspolicy

Under fliken *Integritetspolicy* förklaras det framförallt hur data från användarna sparas och vad den används till. Improve The News är gratis och tillgängligheten är att jämföra med public service, vilket gör att de inte har annonserplatser eller liknande för att tjäna pengar. Med detta som argument så förklaras det att personlig data därför aldrig delas vidare till tredje part, så länge inte lagen kräver det. Webbplatsen tillhandahålls på en molnplattform som kan komma att lagra IP-adresser av säkerhetsskäl. Dessa raderas när de inte längre behövs.

På webbplatsen används cookies (på svenska: kakor) för att spara inställningarna för reglagen. Kakor är textfiler som sparas för att förhöja en användares upplevelse av en webbplats, så som att slippa välja språk varje gång en person navigerar sig runt på olika sidor på en och samma webbplats. Den tid som dessa kakor sparas i användarens webbläsare beror på användaren själv, då det är möjligt att radera kakor när man vill.

Inställningar på reglage, besökta sidor, IP-adresser och tidsåtgång är sådan data som kan sparas. Anledningen till detta är för att analysera data och till exempel se hur intresset för ett särskilt ämne förändras beroende på hur väl det uppmärksammas på flera olika mediaplattformar.

Improve The News nyhetspresentation är i form av länkar som tar läsaren vidare till det nyhetsmedie som skrivit artikeln. Om läsaren vill ta del av en artikel så kommer den behöva besöka en webbplats utöver ITN. Med detta i åtanke så har ITN inte längre kontroll över hur läsarens data används, utan läsaren själv behöver beakta de villkor som råder på nyhetssidan som man klickar sig vidare till.

Feedback-formuläret och dess metadata sparas tillsvidare för att kunna ta del av de åsikter och förslag som kommer in via formuläret. Detsamma gäller för analytisk data, som kan vara värdefull vid framtida efterforskning.

Om en användare av webbplatsen har intresse av att se vilken data som ITN har sparat, så är det möjligt att få detta skickat till sig. Det finns även möjlighet att be ITN att radera personlig data. All data som ITN samlar på sig för en person sparas på en säker server. Om det skulle bli intrång i en server så har dock ITN ingen möjlighet att meddela sina användare om detta då de aldrig ber om, eller sparar, kontaktuppgifter till sina användare.

Till sist så klargörs det att Improve The News inte tar emot någon data från tredje part. De använder ingen som helst data för att profilera användare eller automatisera beslutsfattande.

De är inte en del av någon reglerad industri med upplysningskrav.

6.3 Granskning utifrån teori

Med utgångspunkt i den genomförda litteraturfördjupningen valde författarna ut fyra områden att granska Improve The News utifrån. Följande avsnitt tar upp respektive område och undersöker hur ITN förhåller sig till de problem som hittades i litteraturen.

6.3.1 Standarder och ART

Något som tagits upp återkommande i teorin är faktumet att transparens är subjektivt och beror på vem innehållet är menat att vara transparent gentemot. The Black Box Problem som togs upp under tekniska aspekter går att applicera på det problem som uppstår då maskininlärning används vid sortering av nyheter som i fallet ITN. Designmodellen ART presenterades i teorikapitlet som ett sätt att delvis lösa problematiken kring The Black Box Problem. Värdet i att applicera modellen ligger i att skapa tillförlitlighet och trovärdighet i socio-tekniska miljöer, vilket är relevant vid skapandet och användandet av en nyhetssida.

Transparens och meddelandeansvar finns med i EAD som tas upp under under avsnitt 4.4.2, Standarder från IEEE. I EAD framgår det att transparens är en viktig del i lösningen av problematiken kring förståelse av A/IS, där algoritmer är inkluderade. Enligt EAD ska systemet kunna förklara sina resonemang, i detta fall sorteringen av artiklar, på ett förståeligt sätt för användaren utan att blanda in komplicerade tekniska termer. Om detta inte är möjligt så ska användaren kunna vända sig till utvecklare eller ansvarig för systemet för att få en förtydligande förklaring.

Bokstaven A i ART, accountability, handlar om meddelandeansvaret gentemot olika intressenter. I denna fallstudie är läsaren den viktigaste intressenten, då syftet med webbplatsen främst är att hjälpa människor att bli mer nyanserade i sin nyhetskonsumention. En fråga att ställa sig vid användandet av webbplatsen, då en användare har ställt in skjutreglagen på önskat läge, är: meddelas användaren om varför den har blivit presenterad en särskild artikel?

För att förstå vilka artiklar som presenteras i samband med olika val av parametrar på ITN så måste läsaren först gå igenom hur reglagen fungerar. Det som är beskrivet om reglagen i denna rapport är den information som går att finna på webbplatsens presentation av dem. Politisk Hållning och Hållning Till Etablissemangen har kategoriserats med hjälp av AllSides, Swiss Policy Research och Wikipedia. Information gällande kategorisering utifrån vilket ämne en artikel avser, där de använt sig av en datasamling tillhandahållen av The New York Times, fås från StarTalk-intervjun som författarna själva letat upp.

En användare av ITN har alltså tillgång till viss grad av förklaring på webbplatsen och kan utifrån StarTalk-intervjun få en tydligare bild av hur maskininlärningen fungerar. Däremot är det ytterst svårt att få tag i någon ansvarig personligen, något som författarna själva märkt då de sökt personer och upphovsmän involverade i ITN för intervju. För djupare förståelse av hur ITN har implementerat maskininlärning presenteras det under FAQ att källkoden för webbplatsen är menad att publiceras på GitHub.

Att publicera en källkod går inte helt i linje med vad meddelandeansvar innebär, särskilt inte enligt EAD från IEEE där ett system ska förklara sina resonemang på ett förståeligt sätt. Här behöver man välja relevanta faktorer och presentera dem rätt utefter målgruppen som ska förstå resonemanget. Att få tillgång till koden faller enligt IEEE inte inom ramen för att öka förståelsen för användarna. Den större delen av samhället har inte den kunskap som krävs för att kunna förstå maskininlärningsalgoritmer bättre via tillgång till koden. Att publicera den är snarare till nytta och nöje för andra datavetenskapsingenjörer.

En fråga man bör ställa är hur djupt ingående informationen till användaren ska vara för att webbplatsen ska ha tagit sitt meddelandeansvar. Vilken nivå av förståelse för sorteringen behöver läsaren få för att webbplatsen ska anses vara etiskt hållbar inom den aspekten?

I designmodellen ART står R för responsibility, ansvar. Om Improve The News fungerar som det är tänkt så ligger ansvaret hos användaren och hur denne använder sig av reglagen i syfte att försöka bli nyanserad. Om systemet inte beter sig så som det är tänkt så ligger ansvaret hos utvecklaren, alltså om kategorisering och filtrering inte sker i enighet med det som utlovats. Förutom ansvaret för hur olika artiklar presenteras så finns även ett ansvar kring hur maskininläringen har gått till. Eftersom ITN har använt externa verktyg för att träna sina algoritmer så har ITN ett ansvar att studera de externa verktygen och säkerställa att de är i linje med aktuella lagar och standarder. Det framgår inte hur ITN ser på sitt ansvar i användandet av dessa externa verktyg, eller vem som bär ansvaret om algoritmerna inte fungerar som det är tänkt.

Vidare till ansvaret kring riktigheten i artiklarna så tar ITN inget av det ansvaret alls. Då ITN endast länkar vidare till andras artiklar har de ingen kontroll över innehållet. Här kan man då säga att ansvaret ligger både hos läsaren genom att vara källkritisk, och ansvarig utgivare av artikeln. ITN har aldrig påstått att de garanterar helt sanningsenlig information eller högkvalitativ journalistik - de hävdar bara att de hjälper folk att bli nyanserade. Därför är det okej att fördela ansvaret här ut mot varje användare att avgöra vad för typ av artikel denne faktiskt läser.

Den sista designaspekten, tillika bokstaven, i ART är transparency, transparens. För att uppleva transparens inom AI så behöver en intressent förstå olika designaspekter likväl som själva tillvägagångssättet för implementering av AI. Tillvägagångssättet för att öka transparensen genom att förstå designaspekter hade till exempel inkluderat att presentera varför just AllSides används för kategorisering av politik. Man hade även kunnat presentera hur längden av hållbarhetstiden avgörs, för att förtydliga vad "shelf-life" betyder.

Något som styrker ITN:s transparens runt sina maskininlärningsalgoritmer är att de är tydliga med vad som är indata och vad som är utdata. Läsaren förstår att algoritmerna matas med artiklar som sedan klassificeras och placeras i olika kategorier. Jämför man med en kommersiell plattform vars utdata har beräknats vara relevant innehåll för läsaren, är det mer otydligt vad som är indata och därmed vilka parametrar som spelar in i besluten. I den aspekten är ITN ett gott föredöme kring hur de hanterar tekniska termer på ett begripligt sätt och får läsaren att förstå helheten i ett komplext sorteringssystem.

I problemformuleringen under 1.3.3, Socialt ansvar hos nyhetsflöden, beskrivs problemet med dolda avtal, så kallade clickwrap-avtal, som riskerar att minska läsarens förståelse för vilken personlig data webbsidan använder sig av och i vilket syfte. I och med att ITN inte klassificerar sina läsare så finns inte något behov av att samla in profilerande data och därmed skrivs inget avtal mellan läsaren och webbsidan. Ur transparensynpunkt finns det alltså ingen dold datahantering eller automatiserat beslutsfattande på webbplatsen som berör läsarens integritet. För den data som sparas, inställning av reglage exempelvis, så framgår det tydligt på webbplatsen hur den behandlas. Detta är beskrivet med ord som gemene man kan förstå och det framgår även tydligt hur man går tillväga för att få personliga data raderad om så önskas.

I presentationsvideon för ITN samt i StarTalk-intervjun förklarar Tegmark hur algoritmerna fungerar. Denna förklaring, tillsammans med publikation av algoritmerna, genererar en transparens kring hur ITN fungerar när artiklar filtreras och kategoriseras. Enda sättet att förstå algoritmen utan kod eller utan beskrivning av den vore genom omfattande observationer av algoritmen. Bara då kan en användare förstå resultaten vid justering av reglagen, alltså vad som presenteras för läsaren vid olika parametrar. Detta styrker vikten av att vara transparent och förklara sina designaspekter och implementering av AI.

Transparensen gällande kod och design kan dock ha en nackdel i att patent blir inskränkta. I det fallet med ITN, som har som syfte att fungera i likhet med public service, så bör det inte finnas någon problematik i att vara öppen med koder och implementering. Här finns snarare en vilja att andra plattformar för nyhetskonsumention och information ska ta efter ITN:s koncept.

6.3.2 Sortering och filtrering

Transparens i relation till presentation av nyheter kan förklaras från olika perspektiv. Det kan vara transparens gällande vem som har skrivit artikeln, i vilket syfte den skrivits och vem som beställt artikeln. Med att beställa en artikel menas här att finansiera den, och alltså vore det fördelaktigt om det gick enkelt att ta reda på vilka intressenter som ligger bakom att en artikel skrivits, likt IEEE:s krav i avsnitt 4.4.2. Det kan också handla om varför en läsare fått just en särskild artikel presenterad för sig. Det sistnämnda perspektivet berör sortering och filtrering och därmed transparensen av maskininlärningsalgoritmer och beslut som tagits av mjukvaran.

Reglagen gör det inte tydligare för läsaren vem som har skrivit artikeln och i vilket syfte, än vad som redan framgår på respektive nyhetssida. Eftersom ITN endast publicerar länkar till andras material genererar ITN ingen ytterligare transparens gällande hur och varför en artikel har skrivits. Det står heller inget om hur artikeln finansierats.

Något som skiljer ITN från andra nyhetsaggregatorer och sociala medier är att algoritmerna inte används för att kartlägga användare och deras beteende, utan kartlägger artiklar. Genom att vara en icke-vinstdrivande organisation så finns det inget ekonomiskt intresse i att påverka tiden som läsarna spenderar på webbsidan, till skillnad mot en kommersiell plattform. En kommersiell plattform finansieras i regel av reklamannonser och arbetar därför aktivt för att hålla kvar läsarens uppmärksamhet. Det är då det uppstår en konflikt mellan plattformens intressen och läsarens intressen.

Transparensen på ITN gällande varför läsaren får ett visst utbud av artiklar presenterade för sig är förhållandevis bra, jämfört med kommersiella webbplatser. Varför det är mer tydligt på ITN är på grund av att läsaren vet att det är de önskade preferenserna, regelens inställningar, som gör att artiklar filtreras på ett visst sätt. På många kommersiella webbplatser framgår det ofta inte hur eller varför ett visst urval av artiklar gjorts.

Vissa vinklar och perspektiv kan falla bort ur ett läsarperspektiv på ITN om webbplatsen är läsarens enda källa till nyheter. Detta är på grund av begränsningen av tillgång till data för kategorisering. ITN har utgått från externa kategoriseringsverktyg och artiklarna som presenteras på ITN tillhör en medieplattform som då även finns på till exempel AllSides, där nyhetsmedier kategoriserats utefter politisk tillhörighet. Att ITN är beroende av ett externt verktyg gör att utbudet av olika nyhetsmedier begränsas.

ITN har inget tydligt mål om att vara transparenta, utan motiverar skapandet av webbplatsen med att vilja hjälpa människor att aktivt välja vilket eller vilka perspektiv de vill ha nyheter från. De uppmuntrar även människor till att läsa nyheter från olika källor. Vad som ändå gör att de upplevs som att de vill vara transparenta är deras tydliga förklaringar på hur de använder data från läsare och hur en läsare kan gå tillväga för att få sin data raderad. De gör även en ansats till att vara transparenta gällande sina algoritmer och sidans uppbyggnad, då de avser att publicera koden för allmänheten att beskåda. Problematiken kring hur förståelig koden är för gemene man kvarstår dock.

Frågan om ITN bidrar till ökad eller minskad transparens är svår att avgöra. Den frågan måste ställas i relation till en annan nyhetskälla. Jämfört med till exempel Facebook bidrar ITN till ökad transparens, eftersom Facebook kategoriserar användare till skillnad mot ITN som kategoriserar artiklar. Det som med största sannolikhet är positivt för en läsare som använder ITN är medvetenheten kring kategorisering och filtrering. Om en användare aktivt har använt sig av regelns så förväntar den sig ett visst perspektiv, och förhoppningsvis tar personen till sig informationen med ett kritiskt öga.

ITN bidrar till transparens genom att berätta hur sidan skapades, vilka som är intressenterna samt öppenheten kring koden. De säger inte att ett perspektiv är bättre eller sämre än ett annat, utan uppmuntrar till att konsumera båda delar. Detta kan bidra till att våga uttrycka sina åsikter, våga berätta varför en text är skriven så som den är och även bidra till att människor vågar läsa sådant som de inte är säkra på att de håller med om - just för att man blir uppmärksam på att läsa med ett kritiskt öga.

6.3.3 Konsumtion och medvetenhet hos läsaren

I rapportens teoridel avsnitt 4.3 beskrivs det hur algoritmer i nyhetsflöden bidrar till en ökad polarisering i samhället. Orsaken tycks vara att läsaren placeras i filterbubblor på grund av det filtrerade innehållet samt det psykologiska begreppet konfirmeringsbias, vilket innebär att favorisera information som bekräftar ens befintliga åsikt.

Kommersiella nyhetsflöden använder sig av maskininlärningsalgoritmer för att kunna

presentera det mest relevanta innehållet för varje läsare och på så sätt fånga så stor del av deras uppmärksamhet som möjligt. För att möjliggöra detta matas algoritmerna med personlig data för att klassificera och sedan delge innehåll som avspeglar läsarens intressen. Det är delvis på grund av detta som läsarna hamnar i filterbubblor. Det filtrerade nyhetsflödet håller kvar läsaren och företaget tjänar mer pengar på sina reklamannonser. I linje med konfirmeringsbiasens innebörd så kommer innehållet som presenteras för läsaren avspegla åsikter som hen stödjer, då sådant innehåll favoriseras av läsaren.

Här går det att se en stor skillnad mellan ITN och kommersiella nyhetsflöden. ITN använder maskininlärning i ett annat syfte än att binda läsarnas uppmärksamhet. Algoritmerna används för att underlätta att hitta nyheter ur olika synvinklar och vilket innehåll som syns i flödet styr läsaren över. Den ökade polariseringen som skapas av individanpassade flöden blir därmed inte lika påtaglig då läsaren snabbt och enkelt kan justera reglagen och hela tiden är medveten om att det finns alternativa nyheter att ta del av. Medvetenheten om att det finns andra synvinklar rörande samma nyhet och hur lätt det är att ta del av dem på ITN kan öka toleransen och förståelsen för att andra människor inte har samma åsikt.

ITN lämnar ansvaret till läsaren att bestämma hur nyanserad hen vill sträva efter att bli i sin nyhetskonsumtion. I teoriavsnittet 4.3 förklaras begreppet selektiv exponering vilket innebär att människor tenderar att konsumera information som stämmer överens med deras tro och verklighet. Tillsammans med konfirmeringsbias skapar det stora hinder för att människor ska få en nyanserad bild av ett ämne. Kommer läsarna att läsa andra nyheter än de som stödjer deras åsikt? Psykologiska faktorer såsom dessa är ett tydligt hinder på vägen mot ett mer nyanserat samhälle.

I StarTalk-intervjun får Tegmark frågan om han tror att läsarna kommer välja att läsa artiklar från kategorier som motsätter deras egen åsikt. Tegmark svarar här att det inte går att förutse hur människor kommer navigera runt på sidan, men att han motsätter sig "Big Brother"-synsättet där storföretagen tar alla beslut åt användaren. Tegmark vill att läsaren själv ska få välja sin nyhetsmix på samma sätt som man väljer sin egen diet. Företag som använder sig av algoritmer för att ge läsaren det innehåll som de anser att hen kommer få störst nöje utav lämnar inget utrymme för läsaren att själv söka nyheter.

Problemet med att webbsidorna väljer ut vilket innehåll läsaren ser har tidigare tagits upp i rapporten under 4.3 och benämnts "News finds me". Begreppet innebär att läsaren möts av nyheter utan att behöva söka upp dem själv och kan leda till minskad ideologisk mångfald då läsaren tror att hen får tillgång till alla aktuella nyheter. Precis som Tegmark säger i StarTalk-intervjun så bidrar inte ITN till detta problem då webbsidan uppmuntrar läsaren till att själv söka sig runt bland kategorierna och välja vilken del av spektrat som hen vill läsa utifrån. Tegmark jämför maskininlärning med en kniv, det finns risk för skada men också stor nytta. Genom att sortera och presentera nyheter ur olika vinklar för läsaren och sedan låta läsaren välja vad hen vill konsumera används maskininläringen som ett redskap för kategorisering, och inte för att styra och kontrollera människors uppmärksamhet.

Att använda sig av filtrering av nyheter på internet kommer aldrig kunna undvikas då det finns för stort urval av källor. De filterbubblor som skapas motarbetas därför rimligen genom att

öka transparensen för läsaren gentemot vilket innehåll hen inte får se. Genom att ge läsaren inblick i vad som presenteras för människor med andra åsikter så kan man öka förståelsen för varandra och minska polariseringen. ITN underlättar för läsaren genom att samla nyheterna och möjliggöra filtrering, samtidigt som läsaren själv väljer vad som ska visas. Ansvaret för nyansering delas därmed mellan ITN och läsaren, där webbsidan ger bra förutsättningar och läsaren själv får fatta aktiva beslut i sin nyhetskonsumention och blir därmed mer medveten om vad hen läser för vinklingar.

6.3.4 Balans av artiklar och källkritik

I avsnitt 4.2, Presentation av nyheter, beskrevs två synsätt på begreppet balans. Det ena som har en positiv syn på att skapa en kontext i sina artiklar genom att jämföra och beskriva sammanhang mer i detalj när så krävs, och en annan som vill undvika detta kontextualiserande i största möjliga mån. ITN skapar kontext genom sina rubriker som exempelvis "Energy". Det skapas också en kontext genom reglagen som visar var på olika skalor inom ämnet artiklarna hamnar, och på så sätt jämför den med artiklar på andra ställen på samma skalor.

ITN är däremot inte kontextualiserande på samma sätt som en papperstidning eller tidsskrifter på nätet som placerar artiklar i en viss ordning efter redaktionens bild av läsarnas intresse. Just detta kontextualiserande togs inte upp under begreppet balans, men har diskuterats i förståelsen och begreppsliggörandet av digital journalistik. Där beskrivs hur makten av att välja vad läsarna ska läsa delvis lämnat den traditionella redaktionen och istället hamnat hos algoritmer, nyhetsaggregatorer och publiken. Gällande ITN är det tydligt att både algoritmerna och publiken, det vill säga läsarna, ges mer makt och ansvar för flödet och den information som slutligen når läsaren. Detta kommer påverka hur väl ITN når sitt mål att nyansera läsarna på ett bättre sätt än andra. Ett exempel är hur makten läsaren har över vilka ämnen och på vilka ställen på skalan den väljer att läsa artiklar, kommer påverka hur väl densamme kommer nås av hela det spektrum som finns tillgängligt på sidan och som läsaren uppmuntras att ta del av. Dessutom får läsaren genom denna ökade makt ett större utrymme att misslyckas med att värdera information på ett proportionerligt sätt då detta jobb traditionellt gjorts av grupper av ämnesspecifika och yrkesskickliga journalister.

Vidare på detta spår kommer här beskrivas hur väl de faktorer som tas upp under avsnittet nyanserat finns tillgängliga på ITN. Den första och mest tydliga faktorn är det som benämnts som vinklar, vilket främst innebär att konsumera nyheter från olika håll på olika skalor, exempelvis både vänster och högerpolitiskt innehåll. Denna skala är beroende av tre saker. Först och främst bristen på journalister i USA, där webbplatsen hämtar största delen av sin information ifrån, som har blivit allt tydligare de senaste åren (Doctor, 2015). Denna brist som i sin tur skapar en obalans i ITN:s flöde kan ITN själva inte göra något åt på kort sikt och därför inte kritiseras för. Den andra faktorn är de kriterier ITN fastställt för vilka källor de tar med från den totala tillgängliga mängden, vilket som sagt är otydligt precis som den totala mängden källor som då uppfyller dessa och därför finns representerade på sidan. Utöver Tegmarks uttalanden i StarTalk-intervjun att den största begränsningen är det engelska språket samt att källan respekterar demokrati och inte bidrar till hets mot folkgrupp, vilket inte konkretiseras fullständigt, finns inte mycket att gå på gällande denna punkt. För det tredje hur ofta och på vilket sätt dessa källor visas på sidan i proportion till varandra. Med antagandet att ITN:s algoritmer fungerar som förväntat kan det sägas att sidan balanseras utefter urvalet och att den tredje faktorn är uppfylld.

Den andra delen av nyansering är tidsfaktorn, alltså att flera sidor går snabbt att ta del av. Det är något som är mycket väl uppfyllt av ITN, med reservation för att vissa artiklar kräver en ekonomisk insats för att läsas. Med tanke på detta är det en svaghet för att bli nyanserad att behöva betala för innehåll, men samtidigt kräver kvalitativ journalistik finansiellt stöd för att kunna bedrivas. Kopplat till det ekonomiska perspektivet är frågan hur sidor som ITN påverkar journalistikens affärsmodell fortsatt öppen och ämne för framtida studier.

Representation i form av olika samhällsgrupper och olika personer inom samhällsgrupper är svårt att bedöma i brist på transparens gällande vilka källor som ITN hämtar sina artiklar från totalt sett och hur ofta dessa visas proportionellt. Källornas antal, ungefär 100, talar för att representationen potentiellt är bättre än i en traditionell tidning. I brist på exakta namn på källorna och hur ofta de visas är det fortsatt svårt att säga exakt hur väl denna potential faktiskt uppfylls i praktiken. Dessutom är även denna punkt precis som tidigare faktorer beroende av den totala mängden journalister i USA, i detta fall särskilt journalistkårens diversitet. Även algoritmernas urval påverkar representationen genom att välja vilka källor som visas.

Gällande skillnaden mellan analys och fakta som tas upp vid ett par tillfällen, varav en gång i avsnitt 4.2, finns ingen integration av faktakoll i ITN:s artiklar. ITN kan inte själva anses ansvariga för detta då läsaren skickas till webbplatsen för artikelns upphovsmakare då användaren klickar på artikeln. Dock kan det heller inte ses som en omöjlighet för ITN att integrera detta ändå via en teknisk lösning som kopplas till dessa sidor likt tilläggsapplikationer. Detta får ses som en svaghet hos sidan då teorin påvisar en förbättring i att bedöma artiklars sanningshalt hos läsaren då faktakoll används explicit i artiklar och hävdanden.

En tidsmässig nyansering inom en nyhets område möjliggörs genom någon form av arkiv som beskrivits tidigare. ITN har gällande detta både ett reglage för om en artikel anses relevant över tid, och ett för att gå bakåt i tiden ca 4 veckor maximalt inom en kategori. Med denna funktion uppmärksammas läsaren på denna faktor och kan enkelt gå bakåt i tiden för att se hur det som skrivits om en händelse förändrats över tid. Vid specifika ändamål såsom att gå bakåt i arkiv kan också lämpligen en annan källa än ITN användas, vilket gör att denna svaghet hos ITN inte behöver ha lika allvarliga konsekvenser som andra.

Sammantaget skapar dessa faktorer inom det nyanserade förutsättningen för hur väl läsaren kan göra ett urval av olika exempel och fakta från olika synvinklar för att sedan tolka eller värdera dem.

I ett flöde baserat på spektrum, som ITN, kommer en debatt mellan åsiktsgrupper skapas. Detta då två motsatta sidor på ett spektrum med stor sannolikhet kommer arbeta för att påvisa eventuella svagheter i den motsatta sidans resonemang. Enligt vissa är det botemedlet mot polariseringen i USA, som av samma grupp anses skapat av bristen på sådana i vissa politiska sammanhang där journalistiken är för självbeträktande till den sida den sympatiserar med, som beskrivet i avsnitt 5.2. Dessutom kan en nyanserad konsumtion av nyheter från flertalet etablerade källor möjligen motverka att emotionellt innehåll sprids djupare än de nyheter som

verkligen är mest relevanta för samhället att fokusera på. I och med att ITN endast återger artiklar från etablerade tidningar och nyhetswebbsidor så minskar mängden emotionellt innehåll jämfört med om läsaren hämtar nyheter från sociala medier såsom Facebook och Twitter. Det kan minska risken för polarisering då läsaren får bilda sin egen uppfattning om nyheterna och därmed vara mer källkritisk i sin läsning.

6.4 Koppling till studien

ITN är som tidigare nämnt ett forskningsprojekt och inte en etablerad kommersiell nyhetsplattform. Syftet med webbsidan är att testa ett nytt koncept vid nyhetspresentation och Max Tegmark som är ansvarig för sidan hoppas kunna bidra till en medvetenhet kring filtrering och därmed minska polariseringen som sker i USA. ITN är ett steg i riktningen mot att ge tillbaka makten till läsarna över vad för nyheter de konsumerar via internet.

Syftet med granskningen av ITN är att undersöka om webbsidan också är ett steg i rätt riktning mot att använda maskininlärning på ett etiskt hållbart sätt vid nyhetskonsumtion. Fallstudien är på så sätt en del av arbetet mot att kunna besvara frågeställningen ”Kan maskininlärning användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?” och kommer därmed återkomma i nästkommande analyskapitel. Genom att koppla samman fallstudien med analys av etiska ramverk i verkligheten, 4.4, så kan en ytterligare dimension av huruvida ITN uppfyller etiska krav analyseras.

7 Analys

I detta avsnitt analyseras arbetets resultat utifrån teori och tidigare studier för att reda ut problematiken och försöka besvara frågeställningarna. Skillnader och likheter mellan olika resultat redovisas, förslag på lösningar föreslås och spekulationer om framtiden presenteras avslutningsvis.

7.1 Problem och hinder vid nyhetskonsumtion

I problemformuleringen i 1.3.1 förklarades några mänskliga faktorer som påverkar möjligheten att ta till sig nyheter och samhällsinformation på ett nyanserat sätt. Dessa var konfirmeringsbias, ekokammare och filterbubblor. En av studiens frågeställningar blev därmed ”Vilka problem eller hinder finns idag för att ta till sig nyheter och samhällsinformation via nyhetsflöden på ett nyanserat sätt?” för att kunna gå djupare in på dessa mänskliga, och även tekniska, problem som uppstår då människor ska konsumera nyheter via ett nyhetsflöde som använder maskininlärningsalgoritmer. Följande avsnitt sammanställer och analyserar det underlag som återfanns i teorin runt problematiken vid nyhetskonsumtion.

7.1.1 Bidragande faktorer till filterbubblor och ekokammare

Filterbubblor är inriktade mot algoritmers påverkan på människan, och är ett problem som uppstår om människan inte är uppmärksam på sitt informationsflöde på till exempel sina sociala medier. Mänskliga faktorer är det som i första hand beaktas för att förstå hur maskininläring genererar hinder för människor att bli nyanserade.

I avsnitt 4.3 som behandlar konsumtion av nyheter förklarades det att det är svårt att säga om internet exponerar individer för bredare spann av synvinklar än klassisk interaktion med grannar, kollegor eller andra fysiska möten. Vissa rapporter förespråkar detta, medan andra dementerar tesen. Vad som dock står klart är att utbudet av information på internet kräver att informationen behöver sällas, då det annars blir ett överflöd av information för den enskilde individen. Ett koncept som kallas informationsbegränsande miljöer presenteras under samma avsnitt. Det kan vara sociala nätverks dragning till samma innehåll, eller filtreringsalgoritmer, som begränsar informationskällorna som individer väljer att konsumera. Denna brist på information från olika perspektiv kan bidra till polasiering i samhället eftersom människor behöver vara selektiva i sina val, medvetet som omedvetet, när de väljer sina nyhetskanaler.

Återkommande i hela avsnittet är att bristen på olika synvinklar är det som bidrar till filterbubblor och ekokammare. Det nämns att det fanns en förhoppning om att internet skulle öka expansionen för politisk mångfald eftersom internet tillåter överbryggning av sociala och geografiska begränsningar. Denna expansion begränsas av filtreringsalgoritmerna som sällar materialet då det är för stor mängd information för att en individ ska kunna hantera den på egen hand.

I avsnitt 4.3.3 diskuterades selektiv exponering. Här framkommer det att människans vilja att bli nyanserad spelar stor roll i hur nyhetskonsumtionen ser ut. Trots att det finns tillgång till många olika perspektiv på nyheter så är läsaren inte alltid villig att ta del av andra perspektiv. Antingen undviker personer artikeln helt, eller läste den i syfte att hitta fel. Det framgår att

varför människor beter sig på detta sätt är för att man samlar på sig information pragmatiskt i enlighet med sin överlevnadsinstinkt. Att samla information och samtidigt aktivt vara kritiskt faller sig inte naturligt på samma sätt som att vara ute efter att samla information som man tror sig ha nytta av. Att söka information och hela tiden vara beredd att dementera det som står kräver mycket tid och uppmärksamhet från läsaren.

För att konsumera nyheter från flera perspektiv utan att bli mentalt utmattad av att hela tiden ha "lögndetektorn" på så behövs ett enkelt sätt att bli medveten om de perspektiv som presenteras. En person som befinner sig på sociala medier gör kanske inte det för att bli allmänbildad eller uppdatera sig på vad sker i världen. En person som befinner sig på sociala medier för att hålla sig uppdaterad på vad ens vänner eller förebilder gör, men ändå träffas av nyheter, är inte mentalt inställd på att ta ställning till vad som presenteras.

"News finds me" är ett begrepp som tas upp i 4.2. "News finds me" kallas det fenomen som sker när en person slentrianmässigt läser nyheter samtidigt som denne befinner sig på sina sociala medier och därmed endast konsumerar de nyheter som dyker upp i flödet. Detta passiva sätt att konsumera nyheter på kan bli direkt skadligt när en person tror att denne håller sig uppdaterad på vad som sker i världen, men egentligen bara får veta saker som algoritmer bestämt att personen ska få presenterat för sig.

I avsnitt 4.2 förklaras det varför nyheter presenteras med hjälp av algoritmer. När digitaliseringen möjliggjorde nyhetskonsumention via internet så har det medfört att maskininlärningen har introducerats för att fånga upp läsare, bibehålla dem och på så sätt generera pengar genom exempelvis marknadsföring eller prenumeration. Problematik uppstår när rekommendationsalgoritmer gör val så att ett individanpassat flöde medför att vissa perspektiv sällas bort. För läsaren kan detta upplevas som tillfredställande och intressant, men i syfte att bli nyanserad är individanpassade nyhetsflöden negativa. Nyanserad nyhetskonsumention via sociala medier kräver en aktiv läsare som söker sig till olika perspektiv och är källkritisk i sin läsning vilket "News finds me" inte möjliggör. Konflikten mellan läsarens intresse i att få en aktuell och korrekt bild av händelser som skett i världen och företagets intresse i att hålla kvar läsarens uppmärksamhet skapar en ogynnsam miljö för nyanserad nyhetskonsumention. Återigen blir fenomenet "News finds me" aktuellt att ha i åtanke vid presentation av nyheter i ett nyhetsflöde på sociala medier där även åsikter och missinformation kan spridas och lätt blandas ihop.

Förutom att nyheter filtreras av olika anledningar så har internetbaserad nyhetsmedia medfört ytterligare negativa konsekvenser. Olikt traditionella papperstidningar så har internet möjliggjort en mer slit och släng-attityd till artiklar. För att generera pengar, vara intressant och relevant så måste en nyhetssida uppdatera sina medier ofta. Kvaliteten som väl genomarbetade artiklar ofta har kan då gå förlorad, vilket också tas upp i avsnitt 4.2. Då artiklar kan presenteras som korta notiser får heller inte läsaren alltid hela historien korrekt framställd, då slagkraftiga rubriker prioriteras, snarare än att snabbt ge ett representativt helhetsintryck. Detta kan medföra att viktig information går förlorad, en person får fel uppfattning om en situation eller ett ämne, och polarisering kan uppstå på grund av bristfällig information.

Att internet har öppnat upp för mindre aktörer är både positivt och negativt. Fler personer har möjlighet att göra sin röst hörd, vilket är viktigt för ett jämställt samhälle, men det ökar också risken för att felaktig fakta sprids. Denna paradox finns det ingen enkel lösning på, men försök görs med faktakollhemsidor, där artiklars riktighet ska ha vetenskapligt stöd eller vara strängt baserade på förstahandskällor.

Tekniska problem och hinder som identifierades i avsnitt 4.1 hade olika infallsvinklar. Det är rent algoritmiska problem, men även designmässiga. Problemen som uppkommer gör det på grund av att andra problem har funnit sina lösningar. Med detta menas att när ett överflöd av information behöver sällas så har man använt sig av till exempel maskininläring. Den ur människans perspektiv omedvetna sällningen som görs med ML bidrar till filterbubblor och på lång sikt polarisering i samhället.

7.1.2 Algoritmernas konsekvenser

De rekommendationsalgoritmer som presenterades i denna rapport är så kallade collaborative filtering och content-based. Collaborative filtering avgör vad en läsare ska få presenterat för sig genom att utvärdera vad läsarens kontaktnät interagerar med för slags information eller nyheter. Därav kan collaborative filtering bidra till att en person hamnar i en filterbubbla i och med att personens kontaktnät avgör vilka synvinklar som presenteras. De personer som en användare interagerar mest med, och därmed har störst inverkan på dennes algoritmer, stödjer rimligtvis samma världsbild som användaren själv och vidgar därmed inte perspektiven hos användaren. Content-based filtering utgår ifrån att läsarens intresse är enformigt och läsaren rekommenderas innehåll som går i linje med vad hen tidigare konsumerat. Även denna filtreringsmetod kan placera läsaren i en filterbubbla då innehållet blir av samma typ och därmed innebär att läsaren inte ser helheten och förstår ämnens relation till varandra. Att endast konsumera innehåll av en typ, exempelvis endast ekonominyheter, kan skapa en skev uppfattning om vad som pågår i världen och därmed blir människor endast insatta inom det som intresserar dem vilket minskar den allmänna allmänbildningen.

En designaspekt som bidrar till att människor omedvetet hamnar i filterbubblor är att meddelandeansvaret i designmodellen ART inte efterföljs. Ett meddelandeansvar som efterföljs skulle kunna innebära att en person som konsumerar nyheter får i direkt koppling till presentation av en nyhet även information om varför en särskild artikel är presenterad. När detta inte finns så uppstår ett hinder i att bli nyanserad - en person får svårt att hålla koll på sitt informationsflöde och stannar därmed i sin filterbubbla. För att meddelandeansvar ska efterföljas så behöver någon ta ett ansvar. I designmodellen ART tas ansvarsdelen upp som en aspekt att beakta vid framtagning av AI och maskininlärningsalgoritmer. Relaterat till filterbubblor blir frågan vem som ska ta ansvar för att uppmärksamma nyhetskonsumenter om filtreringen av innehållet i ett specifikt flöde. Problemet i dagsläget ser ut att vara att ingen verkar ta detta ansvar.

I linje med meddelandeansvar och även transparens, som också återfinns i ART, är "the black box-problem" ett begrepp som väl täcker in algoritmernas huvudsakliga brist. Att varesig kunna förklara varför ett visst innehåll presenteras för läsaren eller kunna förklara hur den filtreringen tekniskt gått till tyder på en stor brist inom båda områdena. I och med att läsaren i regel inte har kunskap om maskininlärningsalgoritmer så krävs det stora ansträngningar från företagens

sida om de vill kunna leva upp till vad begreppen meddelandeansvar och transparens innebär. Transparensens subjektivitet i frågan om nyhetsflöden kan lättast angripas genom att låta de minst kunniga i målgruppen sätta standarden. Nyhetsflöden berör alla delar av samhället både åldersmässigt och utbildningsmässigt vilket innebär att förklaringen av de intelligenta systemen bör anpassas efter de med lägst förståelse. Tekniken bakom maskininlärningsalgoritmer är banbrytande när det kommer till presentation av innehåll vilket därmed också kräver ett banbrytande förhållningssätt till läsaren av innehållet. Plattformarna som använder sig av denna teknik befinner sig på en disruptiv marknad där de etiska aspekterna förs fram mer och mer i ljuset. Det företag som lyckas skapa transparens och ta ansvar för teknikens konsekvenser kan därmed även göra stora vinster i framtiden då vi rör oss mot en mer reglerad marknad för AI.

7.2 Riktlinjer för etiskt ramverk

För att reda ut frågeställningen ”Hur skulle ett etiskt ramverk för maskininlärningsalgoritmer vid nyhetsfiltrering kunna formuleras?”, undersöktes två befintliga ramverk för det bredare området ”AI”. Fallstudien av Improve The News samt erkända problem som maskininlärningsalgoritmer medför idag användes också för att besvara frågeställningen.

Analysen som kopplar till den här frågeställningen har brutits ner i olika delar. Först analyseras hur de tre delarna av designmodellen ART kan appliceras på ett etiskt ramverk. Därefter utreds vilket syfte och målgrupp som är relevant för ett etiskt ramverk inom nyhetsfiltrering. Vidare presenteras metoder för att uppfylla kraven och därefter förslag på lösningar. Till sist presenteras en sammanfattad lista över föreslagna riktlinjer som kan lämpa sig i ett etiskt ramverk för dagens maskininlärningsalgoritmer vid nyhetsfiltrering.

De nedanstående riktlinjerna är inte ett fullständigt ramverk utan enbart riktlinjer som författarna anser måste förekomma i ett framtida ramverk, tillsammans med fler riktlinjer, krav och hjälpmedel för att enklare kunna efterfölja ramverket.

7.2.1 Meddelandeansvar

Meddelandeansvar avser underrättande om teknikens beslut och inte konsekvenser som tekniken orsakar. Detta beskrevs under designmodellen ART i, 4.1.4. Spårbarhet och förklarbarhet är två aspekter som presenterades i EU-kommissionens riktlinjer 4.4.1, samt i IEEE:s dokument EAD i 4.4.2. Förklarbarhet är något som även visat sig vara nödvändigt enligt lag, vilket även togs upp i EAD med hänvisning till GDPR. Utvecklarna är alltså skyldiga att meddela användaren om att ett visst automatiserat beslut tas baserat på användarens personliga data om användaren önskar detta. Förklaringen i sig ska vara i linje med det som tas upp under 4.1.4. Alltså ska den vara kontrastiv, selektiv och social. Detta betyder att förklaringen som presenteras ska kunna besvara frågan ”varför gjorde du såhär och inte såhär?” genom att hänvisa till relevanta faktorer som påverkat beslutet. Allt detta ska kunna presenteras på ett sådant sätt att användaren förstår oavsett dennes bakomliggande kunskapsnivå, en fortsättning på detta tas upp under 7.2.3. Att kunna förklara något på ett begripligt sätt är väsentligt för att skapa ett förtroende hos allmänheten för intelligenta system, något som ansågs vara viktigt både enligt EU kommissionens riktlinjer och EAD. Alla dessa delar måste finnas med i ett etiskt ramverk för att göra nyhetsfiltrering hållbart.

7.2.2 Ansvar

För att besvara frågan hur ett etiskt ramverk skulle kunna formuleras behöver det redas ut vem som bär ansvar för vad. Gällande ansvaret för maskininlärningsalgoritmer och dess konsekvenser blir frågan om ansvaret ligger hos antingen företaget som driver plattformen med maskininlärningsalgoritmer för filtrering och därmed valt vilken målfunktion algoritmen har, eller hos användarna som väljer att använda plattformen som nyhetskälla. Ett argument för att användarna ska hållas ansvariga är att det är de själva som valt att använda sig av plattformen och konsumera dess information som nyheter på ett sätt som inte alltid är i linje med plattformens uttalade syfte.

Efter intervjuerna med Gordana Crnkovic och Dorna Behdadi, tillsammans med problematiken gällande om det är utvecklaren eller användaren som bär ansvaret som presenterades i 4.1.4, framkom det att företaget som driver plattformen bör hållas ansvarig för konsekvenserna som dess algoritmer orsakar. Detta gäller även om de används på ett annat sätt än de avsett. Om de vill avsäga sig detta ansvar bör de tydligt uppmärksamma användarna om detta genom att i detta fall upplysa användaren genom att tydligt skriva exempelvis: "Denna sida bör inte användas som en källa till nyheter eller för att ge en verklighetsbaserad bild av omvärlden. Eftersom den länkade informationen och inläggen på sidan som visas för just dig har valts ut baserat på personlig data ges därmed en vinkling som inte representerar samhällets åsikter och trender. Sidan är inte skapad för att presentera nyheter och innehållet är inte begränsat till att endast hämtas från tillförlitliga källor av journalistisk kvalitet". Detta tillsammans med en länk till var användaren kan ta reda på mer om möjliga konsekvenser, likt hur läkemedelsföretag tydligt måste informera om hur deras medicin ska användas och vilka biverkningar den kan medföra skulle göra avsägandet av ansvar fullständigt.

Företagen som driver de stora plattformarna är de som har kunskapen om hur tekniken de använder fungerar. Kunskapsnivån är för låg hos användarna för att lägga ansvaret på dem eftersom de inte vet vad det är för produkt de använder, om de använder den på fel sätt och inte informerats om dess eventuella konsekvenser. Gällande var gränsen går för när ansvaret skiftar från företag till användare går att hantera på samma sätt som för läkemedel. Om användaren blivit tydligt informerad om hur den ska användas och har försetts med tydlig information om eventuella konsekvenser av att använda den på fel sätt, men ändå väljer att använda den på ett sätt som varnats för eller är uppenbart felaktigt så ligger ansvaret hos användaren.

Följer däremot användaren instruktionerna och undviker det som varnats för, ligger ansvaret för att leverera det som utlovats utan oförväntade bieffekter hos företaget. Alltså är det företaget som är ansvarig för att förse användarna med tillräcklig information för att de ska kunna ta ett medvetet beslut om hur de vill använda plattformen i fråga. Med den makt de besitter är det också rimligt att ta ansvar över sina handlingar. Stora sociala medieplattformar som inte utger sig för att vara nyhetsplattformar men som i viss utsträckning används som detta och inte varnar för att de inte ska användas på detta sätt, skulle kunna göra stor skillnad genom att börja ta etiskt ansvar. Stater har ansvar över att säkerställa dess invånares säkerhet, men eftersom stora sociala plattformar kan ses som större än vissa stater, ur inflytelse- och ekonomisk synpunkt, är det en god idé att dessa bör samarbeta för att upprätta regler som motverkar effekterna av skadliga filtreringsalgoritmer.

En ytterligare anledning till att den enskilda nyhetskonsumenten inte bör hållas ansvarsskyldig är med bakgrund i de mänskliga faktorerna som behandlats i avsnitt 4.3, om konsumtion av nyheter, som förklarar varför det är svårt för personer att ta till sig nyheter på ett nyanserat sätt. I intervjun med Gordana Crnkovic som presenterades i avsnittet 5.2, betonade hon att det är ett systemproblem och att ansvaret inte enbart ligger på att media ska lösa det. Hon menar att det är teknikföretagen som i första hand bör stå som ansvariga eftersom det är de som sitter på kunskap om hur algoritmerna fungerar. Hon understryker även att det ligger i människans natur att vara lat och välja den enklaste vägen att ta till sig information. Därför vore det irrationellt att sätta ansvaret på individen. Ansvarsskyldigheten innebär att det behövs införas mekanismer som säkerställer ansvarstagande för AI-system och de resultat som de utger.

7.2.3 Transparens

Transparens bör vara utgångspunkten för arbetet mot etiskt hållbara algoritmer i nyhetsflöden. Ett etiskt ramverk bör innefatta krav på transparens för att kunna skapa förtroende hos allmänheten för maskininlärningsalgoritmer som filtrerar nyhetsflöden. Ett krav på transparens från EU kommissionens riktlinjer som presenterades i 4.4.1 består av de tre underrubrikerna spårbarhet, förklarbarhet och kommunikation. Det innebär förmågan att förklara de tekniska processerna i AI-system samt de medföljande mänskliga besluten. Spårbarhet kan underlätta granskning av systemen samt förklara hur de fungerar. Det bör dokumenteras vilka processer som leder fram till att ett AI-system fattar vissa beslut så man kan undvika att misstag sker.

Förklarbarhet innebär att det ska vara möjligt för människor att förstå hur systemen fungerar. Detta anses grundläggande inom algoritmer och nyhetsflöden då AI-system inom den kategorin har en stor inverkan på människors liv och således bör det finnas möjlighet att få en förklaring på systemets beslutsprocess. Här innebär kommunikation att människor behöver veta att de interagerar med AI. Förslagsvis bör nyhetskanalerna presentera för läsaren på ett tydligt sätt att de använder sig av algoritmer för att filtrera flödet.

Spårbarhet och förklarbarhet innebär även att människan ska förstå besluten som ett AI-system fattar och kunna spåra det. Kravet är viktigt inom nyhetsfiltrering för att få människor att förstå varför och hur deras flöde blir filtrerat. Genom att öka medvetenheten bör människor således få ökad förståelse för att de kan utsättas för polarisering och utifrån vetskapen också reducera risken för det. I 4.4.1 beskrevs också förklarbarhet som centralt för att bygga samt behålla användarnas förtroende för AI-system, vilket ytterligare motiverar att inkludera förklarbarhet i ramverket för nyhetsfiltrering. För att åstadkomma detta, menar de att processerna behöver vara transparenta. Detta är något som även EAD presenterade som en viktig del för ett ökat förtroende för A/IS tekniker, såsom automatiserad nyhetsfiltrering.

Införandet av någon form av mätbar transparens som granskas av ett oberoende organ, likt det som presenterades från EAD i avsnittet 4.4.2 är ett förslag på hur ökad transparens kan uppnås.

7.2.4 Syfte och målgrupp med ett etiskt ramverk för nyhetsfiltrering

Ett grundläggande syfte med ett etiskt ramverk för nyhetsfiltrering baserad på AI är att främja denna teknik och ta vara på dess potential att förbättra dagens nyhetskonsumtion, samt göra den mer robust. Olika situationer inom AI leder till olika utmaningar och som presenterades i 4.4.1 anser EU-kommissionen att standardsiering spelar stor roll för framväxten av marknaden för AI. En del av denna standardisering uppfylls genom att upprätta en ram inom vilken tekniken måste passa in. Denna ram kan fastställas i form av ett ramverk.

Ett ytterligare syfte med ett etiskt ramverk är att underlätta för utvecklare som vill utveckla etiskt hållbart redan under ett tidigt utvecklingsstadium. Ett ramverk kan fungera som ett bollplank som utvecklare kan spela mot och på så sätt kunna användas som en slags checklista för att säkerställa etisk hållbarhet redan från start.

Ett ramverk skulle även kunna förebygga eventuella skador som kan uppstå relaterade till nyheter, exempelvis en missvisande världsbild eller polarisering som presenterades i 4.3, genom att öka medvetenhet om nyhetsfiltrering med dess konsekvenser och hur dessa kan motverkas.

I avsnitt 4.4 behandlades redan existerande etiska ramverk för det bredare området AI, där två av dessa undersöktes. Problemet med att applicera dessa på området filtrering av nyhetsflöden är att de är övergripande och behandlar inte AI specifikt i denna kontext. Det krävs etiska ramverk som är anpassade till mer specifika tillämpningsområden om de ska kunna appliceras mer direkt och utnyttjas på bästa sätt.

Till skillnad från utvecklarna som påverkas direkt av ett ramverk, påverkas även andra intressenter indirekt. Dessa presenterades utifrån IEEE i avsnitt 4.4.2. Här var fokus framförallt på användare av dessa nyhetsflöden. Målet med ett ramverk är att ge användarna mer transparenta nyhetsflöden på plattformar som går att lita på. Detta kommer i sin tur eventuellt kunna stärka den positiva bilden av AI som hjälpmedel vid nyhetsfiltrering och på så sätt bygga upp ett stärkt förtroende hos allmänheten. Detta har belysts tidigare i rapporten och är något som även EAD behandlade, se 4.4.2. Författarnas förhoppning är att ökad medvetenhet hos allmänheten kring förekomsten av filtreringalgoritmer blir en följd av ett sådant ramverk. En kombination av att införa ramverk och införa en tredje oberoende part som granskar och utfärdar eventuella certifikat till de som följer ramverket kan alltså både öka medvetenheten och förtroendet hos allmänheten kring AI-filtrering.

Samtliga intressenter kan gynnas av att följa ett välformulerat etiskt ramverk gällande nyhetsfiltrering. Företagen gynnas då de får hjälp att utveckla ett etiskt hållbart system eller teknik som i längden ökar dess konkurrenskraft. Användarna gynnas då de får ökat förtroende för tekniken och kommer kunna använda denna på ett hälsosammare sätt. För att uppfylla kraven ett etiskt ramverk ställer kommer det krävas arbete från företagen och att vissa processer förändras. Det finns en risk att företag inte upplever att det är värt för dem att arbeta mot det när alternativet att inte göra det finns.

7.2.5 Metoder för att uppfylla kraven

För att uppfylla kraven från ett etiskt ramverk kan det användas både tekniska och icke-tekniska metoder, vilket presenterades i 4.4.1, om EU kommissionens riktlinjer och i 4.1.4, om designmodellen ART.

Det bör vara obligatoriskt att företagen utför testning och validering av deras AI-system. Det bör ske så tidigt som möjligt för att säkerställa att systemet beter sig som det är tänkt. Det bör även ske kontinuerligt under hela livscykeln för att garantera att systemet fortsätter bete sig som det är avsett. Detta kopplar tillbaka till meddelandeansvar, ansvar och transparens.

En icke-teknisk metod är att sprida kunskap till alla inblandade parter om hur AI-systemen fungerar. Författarna anser att det är extra viktigt att sprida kunskap om systemens eventuella konsekvenser. Det bör finnas grundläggande kunskaper inom AI i hela samhället. Det borde krävas av nyhetsflödena att de informerar om att de använder algoritmer och på vilket sätt de gör det.

7.2.6 Förslag på lösningar

För att förhindra att etiska ramverk bara utvecklas utan att efterföljas gäller det att skapa incitament för att följa dem. Att införa lagar tvingar till anpassning, men med trögrörliga rättsystem kan det vara svårt att uppdatera lagarna i takt med teknikens utveckling. Strikt lagstiftning riskerar även att verka hämmande för utveckling av ny teknik när inte företagen tillåts experimentera eller tjäna pengar på samma saker som förut. Detta drabbar inte bara företagen och deras bidrag till statskassan, utan även samhällets potentiella vinster från tillgång till de tjänster som tekniken möjliggör. Givetvis är lagstiftning nödvändigt för att skydda samhället från många faror och i många fall är en strikt och hämmande lagstiftning att föredra över en vag som ger stora spelutrymmen. Den juridiska frågan lämnas med det sagt här för att ge plats åt hur det skulle vara möjligt att skapa incitament för företagen att ta ansvar på ett etiskt försvarbart sätt.

Följande två avsnitt belyser två alternativa lösningar för att öka medvetenheten hos läsarna samt uppmuntra till mer etiskt hållbara algoritmer.

7.2.7 Filterknapp

Det är i dagsläget upp till användaren av filtrerade nyhetssidor att ta reda på hur ens data påverkat vad som visas för en och utifrån detta eventuellt fundera kring åt vilket håll ens verklighetsbild är vinklad. Användaren kan med detta upplägg befinna sig i en filterbubbla utan att få någon varning eller information om detta och på så vis förbli omedveten om det.

Ett förslag på lösning är att införa ett enkelt digitalt verktyg i form av en knapp där användaren själv får klicka i om den vill visa individuellt anpassat innehåll eller om den vill ha en ofiltrerad standardvy som skulle kunna visas för vem som helst. Denna knapp skulle i bästa fall finnas på varje sida som på något sätt filtrerar urval för användaren baserat på dennes data. Knappen ska vara väl synlig med ett enkelt användarvänligt gränssnitt och se likadan ut på

alla sidor av denna typ så att användaren lättare känner igen den och vet hur den ska användas.

Utvecklingen av detta verktyg skulle kunna ske av en oberoende tredje part som uppmuntrar plattformarna att implementera den i deras system. Om det är tekniskt möjligt vore det också bra om användarna själva kunde ladda ner verktyget och använda det på vilket filtrerat nyhetsflöde som helst, ifall vissa plattformar inte väljer att erbjuda den själva.

Knappen skulle lika väl kunna utvecklas av de stora plattformarna som i analysen om ansvar under 7.2.2 pekades ut som ansvariga för många problem som behandlades i detta arbete. Där beskrevs det även att de har stor makt och med tanke på den kunskap och arbetskapacitet dessa företag besitter skulle ett verktyg som knappen vara relativt enkelt för dem att utveckla och implementera. Med det sagt är inte detta ett argument för att dessa företag bör ses som onda eller misstänkas för att ha orsakat problemen med mening. Detta är snarare ett försök att hjälpa till och uppmuntra till större etiskt ansvarstagande.

7.2.8 Certifiering

För att få företag som idag tjänar pengar på användaranpassade flöden att ta hänsyn till ett etiskt ramverk behöver det skapas ekonomiska incitament. Att införa en certifiering av produkter som lever upp till etiska krav är ett sätt att skapa sådana incitament. Eftersom konsumenter i viss utsträckning är beredda att betala mer för eller hellre använda en certifierad produkt innebär detta konkurrensfördelar för de som certifieras. På så sätt kan det bidra till företags fortsatta överlevnad och lönsamhet, trots anpassning till ett etiskt ramverk.

Certifieringen skulle även kunna bidra med transparens eftersom det blir tydligt för konsumenter när ett företag eller en tjänst uppfyller kraven för certifikatet eller inte. Här skulle det kunna finnas olika certifieringsnivåer beroende på vad det är för plattform och i vilken utsträckning de uppfyller vilka krav. När medvetenheten om denna typ av certifikat sprids, kommer den medvetna användaren tydligt se hur etiskt hållbar en viss plattform är utifrån dess certifiering eller avsaknad av den. En ytterligare idé är att det organet som utfärdar dessa certifikat också upprätthåller en svartlista på de företag eller specifika nyhetsflöden som inte tar något etiskt ansvar genom att ignorera arbetet mot ramverket helt, det vill säga varken varnar för felaktig användning av dess produkt eller arbetar mot att förbättra den enligt det etiska ramverket. De skulle kunna sammanställa en lista över alla sidor som de avråder från att använda för nyhetskonsumtion, där sidorna som står på denna kan välja att ta ansvar genom att uppmana sina användare att inte använda sidan på det sättet.

Certifieringen skulle behöva utföras av ett oberoende, icke-vinstdrivande organ med hög kunskapsnivå inom området. Oberoendet är viktigt så att inte kraven för certifiering utformas för att gynna några egenintressen. Icke-vinstdrivande så att målet inte blir att utfärda så många certifikat som möjligt, utan att säkerställa att de som certifieras faktiskt lever upp till de kraven som ställs. Detta organ skulle även vara lämpat för att ansvara för att utveckla och förse företag och användare med knappen som behandlades i 7.2.7. En ytterligare funktion detta organ skulle kunna ha är att vara granskande, likt det som föreslogs i 7.2.5. Vidare förslag är att utveckla och ansvara för ett faktakolls-verktyg i form av något slags AI-system som kan detektera vilseledande, lögnaktig information samt desinformation och flagga för detta. Detta

presenterades som ett förslag av EAD i 4.4.2 och kan precis som det uppmanades till övervakas av detta oberoende och demokratiska organ.

7.2.9 Riktlinjer för ett etiskt ramverk

Följande punkter är förslag på riktlinjer för ett etiskt ramverk för AI inom filtrering av nyhetsflöden, med företagen som äger nyhetsflödena och är ansvariga för utvecklingen av dess filtreringsalgoritmer som främsta målgrupp. Punkterna har delats in utefter delarna i designmodellen ART i hopp om ökad tydlighet. Samtliga rubriker, Meddelandeansvar (A), Ansvar (R) och Transparens (T) väger lika tungt, trots att de innehåller olika antal riktlinjer. Orsaken till varför det finns fler riktlinjer under vissa rubriker och vice versa är att det till exempel finns fler punkter inom transparens som kan uppnås än vad det finns punkter att uppnå inom meddelandeansvar. Med detta sagt väger en punkt under exempelvis rubriken A lika tungt som en punkt under rubriken T. Med "sidan" menas den vy där användaren tar del av ett filtrerat nyhetsflöde.

Meddelandeansvar (A):

- Sidan ska vara transparent utformad för allmänheten så att i princip alla kan förstå en beskrivning av hur den ska användas, hur den fungerar och hur algoritmerna filtrerar på olika tekniska nivåer, där kod är den högsta nivån och där den lägsta beskrivningen ska kunna förstås av ett läskunnigt barn. Koden behöver inte publiceras för allmänheten om det är känsligt ur patent- eller säkerhetssynpunkt, men beskrivningar bör alltid finnas tillgängliga.
- Beskriva vilken personlig data från användaren som används på vilket sätt och hur det påverkar vad som visas. Detta kan presenteras i en informationsruta som förklarar varför det som visas faktiskt visas för användaren.

Ansvar (R):

- Algoritmerna bör vara människocentrerade och utformas för att uppfylla det större syftet med nyheter och informationsspridning i samhället som är att ge folket intressant och korrekt information, och inte för att tjäna pengar. Samtidigt måste lösningarna vara ekonomiskt hållbara för att oberoende journalistik från alla rimliga vinklar ska kunna finansieras och finnas tillgänglig i praktiken, annars finns ingen chans till relevanta och heltäckande nyheter.
- Sidan bör uppmärksamma läsaren på att det går, och algoritmerna bör konstrueras för, att konsumera nyheter från flera synvinklar på ett effektivt sätt och därmed bli nyanserad enligt författarnas definition av begreppet i avsnitt 1.5.
- Sidans ägare är ansvarig för att algoritmerna som används fungerar som utlovat och för eventuella bieffekter som inte varnats för, samt att upplysa användaren om att sidan inte ska användas som nyhetsflöde om den inte är utformad för det men används så. Ansvaret förflyttas däremot till användaren om den väljer att använda sidan på ett sätt som det på sidan tydligt varnats för. Detta först genom att användaren blir informerad och utifrån det väljer att godkänna ansvarsförflyttningen.
- Om sidan saknar uttalad politisk ideologi eller eventuell annan agenda så bör den stå opartisk till de nyheter den publicerar och därmed i största möjliga mån undvika partiskhet.

- Sidans ägare bär det juridiska ansvaret för att dess algoritmer följer nuvarande lagar.
- Sidans ägare bär ansvar för att dess algoritmer är robusta enligt gällande standarder och algoritmerna bör med jämna mellanrum undersökas av ett oberoende externt organ.
- Sidan bör arbeta för att kontextualisera information för läsaren eller alternativt göra det enkelt att kontextualisera för läsaren själv, enligt de sätt som tagits upp tidigare i arbetet i 6.3.4.
- Sidans ägare bör arbeta för att följa upp forskning på ämnet nyanserad nyhetskonsumtion och utveckla metoder som alltmer tillgodoser detta behov.
- Hela samhället bör involveras genom att ta in förslag till förbättringar och försöka tillgodose alla parter intressen. Ett förslag till hur är att ha en förslagsruta där ansvarig på plattformen återkopplar med hur förslaget behandlats.
- Öka medvetenheten och utbilda för att öka kunskapsnivån hos användarna gällande hur plattformen fungerar och hur algoritmerna filtrerar tillsammans med dess eventuella konsekvenser och påverkan på användaren som individ och samhället i stort. Även förklara hur de kan motverka dessa konsekvenser.
- Som plattformsdrivare öppet och kontinuerligt reflektera över hur dess filtrering av nyheter påverkar individ och samhälle och hur de kan ta etiskt ansvar. Detta tillsammans med hur de ställer sig till stora etiska frågor, så som vilket syfte plattformen drivs med och på vilket sätt detta syfte säkerhetsställs.
- Plattformens algoritmer bör användas i människans tjänst med avsikt att förbättra människors välfärd och frihet. I detta fall genom att förse användare med nyheter som bidrar till ökad kunskap och en nyanserad bild av samhället och aktuella debatter.

Transparens (T):

- Beskriva vilken data algoritmerna som används tränats på och vad som är indata respektive utdata.
- Om det gjorts ett urval av från vilka källor algoritmerna hämtar nyheter från så ska en lista på dessa finnas öppet tillgänglig. Det ska även visas vilka kriterier som fanns för detta urval.
- Tydligt källhänvisa var enskilda nyheter hämtats. Om det är en plattform som inte är begränsad till att enbart hantera journalistik bör det tydligt framgå vilka inlägg som är publicerade på nyhetsplattformar med en ansvarig utgivare och vilka som saknar detta och kan räknas som personliga inlägg.
- Visa vem som utvecklat algoritmen och dennes intressen eller målet med algoritmen. Både om den är en produkt av någon annans arbete, företagets egna eller en blandning.
- Företaget/organet bakom nyhetssidan ska vara transparent vad gäller finansieringen bakom filtreringsalgoritmerna.
- Sidan ska vara öppen med dess bakomliggande filosofi och syfte.
- Det ska göras skillnad på fakta, analys och missinformation genom att implementera faktakoll genom algoritmer som detekterar fake-news när det är möjligt.

- Sidan bör vara transparent med hur användarens data hanteras och hur användaren kan påverka denna användning. Detta för att öka tillförlitligheten till sidan och dess upphovsmakare.
- Implementera ett stöd som fyller samma funktion likt knappen som föreslås i avsnitt 7.2.7, för att ge användaren möjlighet att ta del av information på ett icke-individualiserat sätt och för att uppmärksamma om att algoritmer styr ens flöde som således inte visas på samma sätt för alla individer i samhället.
- Sidan bör antingen ta etiskt ansvar genom att arbeta likt hur de skulle för att bli certifierade enligt avsnitt 7.2.8, eller så måste de ifrånsäga sig sitt ansvar genom att tydligt varna för att den inte ska användas för nyhetskonsumention.
- Det ska framgå hur väl resten av alla kriterier för transparens uppfyllts, vilket skulle kunna tydliggöras genom certifiering.

7.3 Etiskt hållbar maskininlärning

Arbetets första frågeställning tog upp problem kopplade till dagens nyhetsflöden. Den andra behandlade riktlinjer för att förbättra dem. I den tredje ”Hur kan maskininlärning användas på ett etiskt försvarbart sätt för att hjälpa människor att ta till sig nyheter och samhällsinformation på ett nyanserat sätt?” kommer det utvärderas vad som är det verkliga syftet med att överhuvudtaget använda sig av maskininlärningsalgoritmer i nyhetssyfte. Det utvärderas även om det finns möjlighet att uppnå det som eftersträvas i nyhetskonsumention med hjälp av dem, samt på vilket sätt i så fall.

7.3.1 Fallet Improve The News

En del av syftet med fallstudien var att undersöka huruvida ITN är ett steg i rätt riktning gällande etiskt hållbar användning av maskininlärningsalgoritmer. I granskningen av Improve The News utifrån teorin i avsnitt 6.3 förde författarna fram delade åsikter gällande hur väl ITN hanterar problematiken bakom maskininlärningsalgoritmer. Gällande transparens så anser författarna att tillgång till koden inte ökar läsarens förståelse men däremot finns det andra delar på webbsidan som gör det. Exempelvis finns det information om hur filtreringen går till. Även om den stundtals är bristfällig på webbsidan så finns det en publicerad intervju med Max Tegmark där det förklaras mer ingående. Webbsidan är tydlig med vilken in- och utdata som används och det sker ingen analys av läsarens beteende på webbsidan utöver att inställningen av reglagen sparas och därmed görs inga övertramp på läsarens integritet.

Framförallt ger ITN tillbaka en stor del av makten och beslutstagandet över sin nyhetskonsumention till läsaren. Läsaren själv får bestämma hur nyanserad hen vill vara i sin nyhetskonsumention och alla synvinklar finns lättillgängliga genom en enkel justering. Just denna medvetenhet om att andra synvinklar på samma ämne finns tillgängliga kan i sig öka läsarens källkritiska förmåga och minska blindheten gentemot andra åsikter som annars kan uppstå och skapa den polarisering som finns i samhället idag. Genom att inte placera läsarna i olika kategorier utan istället låta dem välja fritt mellan kategorier av innehåll uppnår ITN:s filtrering dess grundläggande mål, att presentera intressant innehåll för läsaren, utan att placera läsarna i filterbubblor.

Improve The News är som tidigare nämnt ett pågående forskningsprojekt och kan inte jämföras helt och hållet med en kommersiell plattform. Däremot anser författarna att webbsidan kan ses som ett steg i rätt riktning mot ökat meddelandeansvar och transparens, något som presenterats som två av de viktigaste aspekterna vid etiskt hållbar maskininläring.

7.3.2 Kommersiella plattformar

I tidigare exempel av certifiering har det visats att de på lång sikt kan ändra beteenden hos både konsumenter och följaktligen även producenter då efterfrågan förändras. För arbetets område, nyhetsflöden, gäller att de som efterföljer certifikaten i högst mån kommer ges större chans att konkurrera ut alternativ som inte gör det och att standarden totalt sett höjs. I hela landskapet av olika plattformar som används som nyhetsflöden används vissa utan att vara tänkta eller skapta för det. Syftet med nyheter är att informera samhällets befolkning om lokala och globala viktiga händelser. Att plattformar som inte har som huvudsyfte att informera och sprida nyheter används som nyhetsplattformar anses vara ett av de största problemen gällande nyhetsflöden idag. Med en certifiering bör de flöden som syftar till att tillgängliggöra nyheter på ett nyanserat sätt användas i allt större utsträckning. Det kan även bidra till att ocertifierade medier inte används som nyhetskällor och därmed minska risken för spridning av fake news och missinformation.

Det som särskiljer ett nyhetsflöde som använder maskininlärningsalgoritmer, jämfört med ett som inte gör det, är som sagt kategoriseringen av antingen individer eller innehåll. För det nyhetsflöde som är tänkt och skapat för att nyansera sina läsare är främst kategoriseringen av innehåll intressant. Kategoriseringen möjliggör att presentera nyheter för läsaren från flera vinklar på ett strukturerat sätt. Som tagits upp i rapporten är detta ett sätt att underlätta för läsaren i dagens informationsöverflöd. Det hjälper även läsaren att till exempel förstå skillnaden mellan två politiska synvinklar samt vilka nyhetsmedia och artiklar som representerar dessa. På så sätt görs ett nyanserat intag av nyheter möjligt. Ihop med de riktlinjer för ramverk som beskrivits i arbetet är denna fördel något som ökar chanserna att nyansera läsare och samhällsmedborgare på detta sätt jämfört med utan maskininlärningsalgoritmer.

I intervjun med Gordana Crnkovic benämner hon att AI är en fantastisk teknologi och att många ser fördelar med den. Därmed finns det förutsättningar att använda maskininläring på ett etiskt försvarbart sätt för att hjälpa människor ta till sig samhällsinformation nyanserat, men således krävs det att de negativa konsekvenserna som kan komma från AI förebyggs. Enligt Crnkovic var det största problemet att den enskilde samhällsinvånaren inte är medveten om hur teknologin bakom algoritmerna fungerar. Hon berättar att det är de stora tekniska företagen som besitter kunskapen och för att bearbeta konsekvenserna av missinformation måste alla intressenter involveras i teknologin och de beslut som påverkas av den. Utifrån det här kan det anses relevant att de stora tekniska företagen tar på sig ansvar för att informera och involvera övriga intressenter i hur tekniken fungerar.

Även Dorna Behdadi betonade i intervjun att det är osannolikt att lägga ansvar på den enskilde användaren vilket styrker det faktum att de stora bolagen bör vara ansvarsskyldiga. Det kan anses överflödigt att engagera samhällsinvånarna på teknisk detaljnivå men det anses väsentligt att de förstår vilka beslut som kommer från teknologin samt hur och varför den fattar de besluten. Behdadi poängterar också att transparens inte spelar någon roll ifall det

är svårbegripligt, vilket innebär att den information som framförs måste gemene man förstå för att kunna ta till sig den. Det är även de tekniska bolagen som utvecklar algoritmerna och blir de mer lyhörda till övriga samhällsinvårares åsikter kan således inblandade intressenter också påverka vad som sker bakom kulisserna. Därmed kan processen mot ett mer nyanserat samhälle ta fart då det krävs att alla involverade parter kan förstå teknologin samt vara med och påverka hur den ska fungera.

Framtiden för maskininlärningsalgoritmer som sorterar nyheter ser ljus ut om teknikföretagen är öppna för att samarbeta eller på annat sätt uppfyller de krav som ramverket tar upp. Det övriga samhället har de senaste åren ställt allt högre krav på de stora plattformarna och i grupper av samhället sukts det efter nyanserade sätt att konsumera nyheter. På lång sikt kommer nya typer av maskininläring utvecklas och det är svårt att säga vilka dessa kommer vara i dagsläget och vilka bieffekter de kommer medföra. Men med traditionella verktyg som certifiering och press på makthavare i samhället ihop med ny teknologi som maskininlärningsalgoritmer, kan dagens nyhetsflöden och individers nyhetskonsumtion förbättras avsevärt på färden mot ett alltmer nyanserat samhälle.

8 Slutsats

I undersökningen synliggjordes ett flertal problem och hinder för att ta till sig nyheter och samhällsinformation på ett nyanserat sätt. Problemen hanterades inom tre olika kategorier: tekniska, konsumtion och presentation. Tidigt i undersökningen ansågs att de minst kritiska problemen är de som faller under hur den enskilde individen konsumerar nyheter. Till största del berodde detta på att de innefattar mänskliga faktorer som är svåra att eliminera. Därmed riktades fokus på att mer djupgående reda ut de tekniska hinder och de problem som innefattar hur nyheter presenteras.

Med hjälp av rekommendationsalgoritmer filtreras nyhetskonsumenternas flöde vilket bidrar till att mångfald av information kan gå förlorad. Ett märkbart problem som tillkännagavs är att människor inte förstår hur algoritmerna fungerar. Detta bidrar i synnerhet till minskad medvetenhet i hur individanpassat flödet är och vilka konsekvenser det bidrar till.

Ett ytterligare avsevärt problem handlar om vem som bär ansvaret för att förse konsumenterna med nyanserad information. En slutsats är att det inte är möjligt att hålla den enskilde användaren som ansvarig. Det finns istället två primära aspekter inom ansvarsfrågan. Å ena sidan krävs det att utvecklarna av algoritmerna ansvarar för att algoritmerna beter sig som förväntat och följer de etiska riktlinjer som rekommenderas. Det krävs även att de ansvarar för att systemens beslutsprocesser går att förklara samt spåra och granska för att undvika eventuella fel. Å andra sidan behöver nyhetskanalerna förse användarna med tillräcklig information för att användarna ska förstå hur algoritmerna påverkar deras nyhetskonsumtion samt vilka konsekvenser det kan bidra till. Genom att öka den medvetenheten får användarna förutsättningar för att minimera de negativa konsekvenser som kan uppkomma när deras flöde blir filtrerat. Denna transparens medför även ett större förtroende för filteringsalgoritmer vilket blir en förutsättning för fortsatt användning av deras positiva egenskaper i framtiden.

Det som författargruppen ansett vara det mest väsentliga inom det behandlade området är meddelandeansvar, ansvar och transparens. Därför har gruppen valt att utforma riklinjerna i det etiska ramverket utifrån designmodellen ART. Idéer för att uppnå de tre delarna av ART har bland annat uppstått ur, diskussioner inom författargruppen kring riklinjer i redan existerande ramverk, det som uppfattades saknas hos ITN i fallstudien från den genomförda samt tankar som uppstått ur de genomförda intervjuerna. Riklinjerna från det framtagna ramverket riktar sig främst åt utvecklare, men ska även gå att läsas och undersökas av nyfikna användare. De etiska riklinjerna finns att hitta under avsnitt 7.2.9.

Slutligen har studien kommit fram till att det finns ett flertal förutsättningar för att använda maskininlärning etiskt försvarbart för att hjälpa människor att ta till sig nyheter och samhällsinformation nyanserat. Framförallt krävs det att de stora teknikföretagen är villiga att samarbeta för att ta fram system som uppfyller de etiska riktlinjer och krav som ramverket tar upp. Kombinationen mellan att sätta press på makthavare i samhället och framställa ny teknologi kopplade till nyheter kan ge goda förutsättningar för att dagens nyhetsflöden och individers nyhetskonsumtion förbättras avsevärt. Att stora delar av samhället fortsatt ställer höga krav på nyhetsplattformars sätt att presentera nyheter är ett stort steg i rätt riktning mot ett mer nyanserat samhälle.

9 Diskussion

Denna studie har bidragit till debatten och diskussionen kring användandet av maskininlärningsalgoritmer i nyhetsfiltrering och huruvida dessa algoritmer kan användas och utvecklas på ett etiskt hållbart sätt inom området. Problem, risker och omfattning kring området har belysts i studien men även eventuella lösningar och hopp om framtiden har framförts. Detta har gjorts på en nivå anpassat för icke-kunniga inom området för att, även hos den gruppen, väcka debatt och medvetenhet kring filtreringsalgoritmer i våra nyhetsflöden. Att väcka diskussion och debatt hos grupper som till vardags inte är med i utvecklingen av algoritmerna har ansetts vara viktigt då dessa grupper är direkt påverkade av resultatet av den utvecklingen. Därför har arbetet riktat sig även mot dessa grupper.

De valda intervjuerna skedde genom snöbollsurval. Detta kan ha påverkat urvalet, då första respondenten tipsade om de andra två. Gruppen hade flertalet kandidater utanför de tipsade kandidaterna, dock blev några intervjuer med dessa aldrig av. Då den första respondenten menade att de andra två respondenterna kompletterade varandra i deras olika områden uppstår risken för en icke objektiv åsikt. Trots detta genomfördes intervjuerna med tanken om att det skulle bidra till heterogena diskussioner i intervjuerna, vilket det tillslut ansågs ha gjorts.

Detta är i linje med diskussionen kring de två större granskade riktlinjerna, nämligen att det sitter i stort sätt samma grupp människor i de olika grupperingarna som tar fram dessa riktlinjer. Där yrkestitlar sträcker sig mellan forskare till företagsledare för EU kommissionens riktlinjer. Vidare vad gäller IEEE:s dokument tillhör den större delen forskare, på allt från student- till professornivå. Någon fråga om avsaknad kompetens kring ämnet inom grupperna förekommer därmed inte anser författarna. Dock förekommer en avsaknad av övriga samhällsinvånare, som inte är yrkesverksamma i utvecklingen av AI men som frekvent använder de system som riktlinjerna är avsedda för. Detta kan även ha påverkat resultatet i form av att det sker en homogen diskussion i de olika grupperingarna. Det här har författarna av studien tagit i beaktning och således inte enbart tagit fram de egna etiska riktlinjerna endast baserade på dessa riktlinjer. Många av de punkter som finns med i de framtagna riktlinjerna har tagits fram genom interna diskussioner inom författargruppen samt diskussioner kring fallstudien. Att ha tittat på mindre ramverk, och granskat ifall kommittéerna är tillräckligt heterogena, är något som hade kunnat göras annorlunda och på så sätt eventuellt påverkat det slutgiltiga resultatet.

Mycket av arbetets teori, analyser och slutsatser har grundat sig i designmodellen ART. En diskussion bör föras kring dess för- och nackdelar. Å ena sidan ingår endast frågor kring meddelandeansvar, ansvar och transparens i modellen och därmed finns risken att andra eventuellt väsentliga aspekter förbises. Å andra sidan har det under arbetets gång i underlag från intervjuer, fallstudie och litteratur visat sig vara just meddelandeansvar, ansvar och transparens som är de väsentliga aspekterna i utformningen av hållbar AI. Med det sagt så anses ART vara en heltäckande designmodell för ändamålet och därmed valdes den av författarna som grund till utformningen av etiskt hållbara maskininlärningsalgoritmer vid nyhetsfiltrering.

Författarna av studien har under studiens gång ständigt haft öppna sinnen kring problematiken.

Även om vissa förutfattade meningar existerade i de tidigare stadierna av studien anses inte dessa ha blockerat intaget av ny information och anses därmed inte heller ha påverkat resultatet av studien.

Å ena sidan har författarna till studien inte haft en djup kunskap inom den tekniska delen av området, mer specifikt hur koden för AI-baserade nyhetsflöden ser ut och fungerar. Detta kan möjligtvis ha bidragit till att alla riktlinjer som tagits fram ej är tekniskt görbara. Å andra sidan är majoriteten av riktlinjerna sådana att de inte kräver djupgående teknisk förståelse för att konstruera. Flera riktlinjer grundar sig snarare i etik och moral, välvilja och ärlighet, något som teknik inte kan blockera. Riktlinjerna av mer teknisk natur anses som sagt inte heller vara för komplexa för att kunna implementeras relativt enkelt av de som berörs av det.

10 Förslag på framtida undersökningar

Under arbetets gång har vissa frågor lämnats utan vidare undersökning eftersom de låg utanför detta arbetets omfång. Det har även uppkommit idéer till framtida undersökningar som tar vid där detta arbete slutar.

Ett första förslag på framtida undersökning är att behandla den ekonomiska frågan djupare genom att undersöka hur det går att göra en ekonomiskt hållbar nyhetstjänst vars syfte är att nyansera människor, oavsett ekonomisk bakgrund. Det vore även intressant att undersöka hur det går att gynna kvalitativ, samhällsnyttig journalistik i kommersiella nyhetsflöden. Kan en etiskt hållbar nyhetstjänst främja sådan journalistik och hur ska denna utformas i så fall?

Den ekonomiska frågan kan även angripas genom att undersöka de ekonomiska möjligheterna för en sida som ITN. Detta genom att undersöka vad det skulle kosta att driva den i stor skala och hur detta skulle påverka de tidsskrifter den hämtar nyheter från. Förblir tidsskrifternas inkomstflöden konstanta eller skapar denna förändring av marknaden behov att finansiera tidsskrifterna på nya sätt? Detta med tanke på att antalet journalister redan visat sig sjunka i flertalet länder och att lokala tidsskrifter har fått det allt svårare att gå runt.

En idé som diskuterats internt är hur en prenumerationstjänst likt hur Spotify erbjuder musik, skulle fungera för nyhetsartiklar. Istället för att visa ett flöde av artiklar som innehåller länkar till nyhetsmedierna som publicerat dem så skulle idén vara att prenumerationstjänsten betalar de olika källorna för varje gång deras artiklar visas. Användaren betalar bara en prenumerationsavgift för att få läsa allt och slipper processen att skickas vidare till flertalet olika nyhetsmedier. Här råder en oklarhet i hur det påverkar journalisterna, likt hur det finns ett missnöje från artisterna vars låtar är publicerade på Spotify.

Ett annat förslag på framtida undersökning är att behandla den juridiska frågan gällande etiska ramverk. Vilka delar av det som i detta arbete formulerats mot ett etiskt ramverk skulle göra störst nytta om de blev lag och när är det fördelaktigt att låta det vara frivilligt? Är certifieringsidéen en effektiv medelväg eller medför den oönskade konsekvenser?

På temat juridik kan det även undersökas vilka brister som finns i dagens lagstiftning för att nyansera läsare med etiskt hållbara nyhetsflöden. Detta genom att undersöka de lagar som finns idag och jämföra med de riktlinjer till etiska ramverk som presenterats här för att försöka reda ut om striktare lagar skulle kunna bidra mot detta syfte.

Att genomföra intervjuer med fler personer som arbetar inom detta arbetets område är ett annat förslag som skulle kunna ge andra resultat baserat på deras erfarenheter från ämnet. Ett ytterligare förslag är att involvera personer som inte arbetar inom området också, för att få en bättre spegling av samhällets syn på frågan. Dessa intervjuer skulle vara av en annan typ eftersom syftet med dem skiljer sig från de som genomfördes i detta arbete. En fristående undersökning skulle även kunna vara att undersöka hur hela samhället bäst kan involveras i frågorna om hur företag kan ta större etiskt ansvar och vilken information eller vilka verktyg de önskats få tillgängliga i vilka lägen. Vidare vad som upplevs vara de största problemen och

hur medvetna konsumenterna faktiskt är.

Ett till förslag är att undersöka hur villiga företag är att arbeta mot etiska ramverk. Är det något de önskar för att hjälpa dem att ta större etiskt ansvar eller upplever de att deras verksamhet och affärsmodell inte skulle vara lönsam om det etiska ramverket applicerades. Har de resurser nog för att arbeta fram en ny, etiskt hållbar strategi eller är detta ett problem som kräver hjälp att lösas utifrån?

Kopplat till fallstudien av ITN som genomfördes i detta arbete vore det intressant att undersöka vad användning av sidan har för påverkan. Detta genom att låta en testgrupp använda den över tid för att sedan genom exempelvis en enkät som undersöker vilken information de tagit del av undersöka hur väl innehållet är balanserat mellan klassiskt amerikansk vänster- och högerpolitik. Bristen av amerikanska journalister från högersidan har klargjorts, men att testa sidans påverkan på dess användare vore intressant för att se om denna brist syns hos användarna och om fler brister i hur sidan presenterar nyheter kan synliggöras.

Utifrån fallstudien av ITN uppstod tankar på en undersökning som behandlar om det som sker vid sidans reglage verkligen justerar nyhetsflödet så som de påstår. Även att undersöka vilka alternativa sätt reglagen kan utformas på och hur denna utformning skulle påverka nyhetsflödet, exempelvis att justera i flera dimensioner istället för binärt.

En ytterligare undersökning skulle vara att analysera kod från exempelvis ITN och ge förslag på hur den ska förstås bäst på olika nivåer. Utifrån detta skulle det kunna leda till en standardiserad mall på hur andra sidor kan förklara sina algoritmer i text för sina användare i syfte att öka sidans transparens.

Ett sista förslag är att experimentera med faktorer som att användaren aktivt ska försöka få en nyanserad bild av samhället med hjälp av ITN och jämföra med det densamme tar till sig när den spenderar tid och söker nyheter så som den brukar.

Referanser

Akkor Gül, Aysen Ertürk, Yıldız Dilek Elmer, Paul. (2020). *Digital Transformation In Media Society*. Istanbul University Press.
<https://cdn.istanbul.edu.tr/file/JTA6CLJ8T5/2B799F5AB7A14F628F90BFC458B38029>

Allsides. (2021). *Media Bias Ratings*. <https://www.allsides.com/media-bias/media-bias-ratings>

Bago, B., Rand, D. G., & Pennycook, G. (2019, July 22). *Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines*. doi: 10.1037/xge0000729

Barrera, O., Guriev, S., Henry, E., & Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182. doi: 10.1016/j.jpubeco.2019.104123

Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science Magazine* 348(6239): 1130–1132. <http://science.sciencemag.org/content/348/6239/1130>.

Benham, J. (2020) Best Practices for Journalistic Balance: Gatekeeping, Imbalance and the Fake News Era, *Journalism Practice*, 14(7), 791-811. doi: 10.1080/17512786.2019.1658538

Blanding, M. (2018). Where Does Journalism End and Activism Begin?. *Nieman Reports*. <https://niemanreports.org/articles/where-does-journalism-end-and-activism-begin/>

Cherry, K. (2020). *How confirmation bias work*. Verywellmind.
<https://www.verywellmind.com/what-is-a-confirmation-bias-2795024>

Cihon, P. (2019). *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*.
https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_FHI – Technical – Report.pdf

deGrasse Tyson, N. (20210226). *StarkTalk: Improving the News with Max Tegmark*. [Video podcast]. Curved Light Productions. <https://www.youtube.com/watch?v=9M2CZlWq3zM>

Dias, N., Pennycook, G., & Rand, D. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Misinformation review*. <https://misinforeview.hks.harvard.edu/article/emphasizing-publishers-does-not-reduce-misinformation/>

Dignum, V. (2019). *Responsible Artificial Intelligence*. Springer.

Duffy, A., & Peng H.A. (2019) Digital Journalism: Defined, Refined, or Re-defined, *Digital Journalism*, 7(3), 378-385. doi: 10.1080/21670811.2019.1568899

Elfving Persson, B. (2014). *Var börjar och slutar journalistiken? : En kritisk diskursanalys av journalistikens gränser*. <http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-56465>

Eriksson, L.T., Wiedersheim-Paul, F. (2008). *Rapportboken*. Malmö: Liber AB.

Europeiska kommissionen. (2018). *Etiska riktlinjer för tillförlitlig AI*. doi: 10.2759/346720

Europaparlamentets och rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning). *Europeiska Unionens Officiella Tidning*, L119/1, 1-173. Hämtad från <https://eur-lex.europa.eu/legal-content/SV/TXT/?uri=celex%3A32016R0679>

Faris, R. M., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election. *Berkman Klein Center for Internet Society Research Paper*, 91. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33759251>

Felfernig, A., Boratto, L., Stettinger, M., Tkalčič, M. (2018). *Group Recommender Systems*. doi: 10.1007/978-3-319-75067-5

Frank, R. H. (Feb 11, 2021). The Economic Case for Regulating Social Media. *The New York Times*. <https://www.nytimes.com/2021/02/11/business/social-media-facebook-regulation.html>

Ghersetti, M. (2018). *Alarmerande nyheter: Innehåll i och effekter av nyheter om hot och larm* [Studierapport, Institutionen för journalistik, medier och kommunikation, Göteborgs universitet]. MSB. <https://rib.msb.se/filer/pdf/28792.pdf>

Gunnarsson, R. (2020) *Validitet och reliabilitet*. Hämtad 2020-02-11 från <https://infovoice.se/validitet-och-reliabilitet/>

Hed, I., & Helte, E. (2020). *Klimatet på agendan: En studie om Dagens Nyheter och The New York Times gestaltning av klimatkrisen under november 2020*. [Examensarbete, Institutionen för informatik och media]. DiVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1522708dswid=6322>.

Helberger, N. (2019). On the Democratic Role of News Recommenders. *Digital journalism*, Volym(7), 993-1012. doi: 10.1080/21670811.2019.1623700

Holme, I.M. & Solvang, B.K. (1997). *Forskningmetodik - Om kvalitativa och kvantitativa metoder*. Studentlitteratur

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H. & Aral, S. (2020). *The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify*. <http://ide.mit.edu/sites/default/files/publications/SSRN-id3555927.pdf>

Institute of Electrical and Electronics Engineers (IEEE). (2021). *Advancing Technology for Humanity* Hämtad 2021-04-09 från <https://www.ieee.org/about/index.html>

Institute of Electrical and Electronics Engineers (IEEE) (2019): *Ethically Aligned Design, (EAD), First edition, A Vision for Prioritizing Human Well-being with Autonomous and Intelligent systems*.

Hämtad från

https://engagestandards.ieee.org/rs/211-FYL-955/images/EAD1e.pdf?mkt_tok=eyJpIjoiWkRVME1UVm1OREE1TVRSbSIzLnQiOiIiY3RONFl6YXh0cWxSRUpLNE9taUtwQllpaXNkYktmd3FDM2lOQ1ZNXC9YUURKV3Z4b2dJc3d3ekNDREdTd24zMHNcL0xUTEFqeFFoYTN4NWNqQUZRclY0amMyTzhXeU9VXC9yNjhneWlleHFHV3lSMU1rRGxmeUJSTU9cL3dDeXZmN1AifQ%3D%3D

Isaak, J. & Hanna, M. (2018). *User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection..* IEEE Xplore. doi: 10.1109/MC.2018.3191268/

Johnston A, Lefort F, Tesvic J. (2017). Secrets of successful change implementation. *McKinsey & Co*.

<https://www.mckinsey.com/business-functions/operations/our-insights/secrets-of-successful-change-implementation>

Kaplan, A. & Haenlein, M. (2019). *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*. ScienceDirect. doi: 10.1016/j.bushor.2018.08.004

Karimi, M., Jannach, D., Jugovac, M. (2018). News recommender systems – Survey and roads ahead. *Information Processing Management*, 54(6), 1203-1227. doi: 10.1016/j.ipm.2018.04.008

Kearns, M., Roth, A. (2020) *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford: OUP USA.

Kitchens, B., Johnson, S. & Gray, P. (2020). UNDERSTANDING ECHO CHAMBERS AND FILTER BUBBLES: THE IMPACT OF SOCIAL MEDIA ON DIVERSIFICATION AND PARTISAN SHIFTS IN NEWS CONSUMPTION. *MIS Quarterly Vol. 44 No. 4*, pp. 1619-1649. doi: 10.25300/MISQ/2020/16371

- Kramer, A., Guillory, J., Hancock, J. (2014). Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Sciences of the United States of America* 111, no.24, 8788-8790. doi: 10.1073/pnas.1320040111.
- Ling, R. (2020). Confirmation Bias in the Era of Mobile News Consumption: The Social and Psychological Dimensions. *Digital Journalism*, 8:5, 596-604. doi: 10.1080/21670811.2020.1766987
- Machado Lunardi, G., Medeiros Machado, G., Maran, V., Palazzo M. de Oliveira, P. (2020). A metric for Filter Bubble measurement in recommender algorithms considering the news domain. *Applied Soft Computing*, 97(A), 106771. doi: 10.1016/j.asoc.2020.106771
- Madiega T. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. *European Union*.
[https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI\(2019\)640163](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2019)640163)
- Nikolov, D., Flammini, A., & Menczer, F. (2021). Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *Harvard Kennedy School (HKS) Misinformation Review*, 1 (7). doi: 10.37016/mr-2020-55
- O'Neil, C. (2016a, november). *Weapons of Math Destruction* [Video]. YouTube.
<https://www.youtube.com/watch?v=TQHs8SA1qpk>
- O'Neil, C. (2016b). *Weapons of Math Destruction*. London: Penguin books.
- Parmelee, J. H. & Roman, N. (2020). Insta-echoes: Selective exposure and selective avoidance on Instagram. *Telematics and Informatics*, volym(52). doi: 10.1016/j.tele.2020.101432
- Pentina, I., Tarafdar, M. (2014). From “information” to “knowing”: Exploring the role of social media in contemporary news consumption. *Computers in Human Behavior. ScienceDirect*. doi: 10.1016/j.chb.2014.02.045
- Peterson, L. Äng M. (2020). Etisk problematik vid implementering av maskininlärning i sjukvården. [Kandidatuppsats, Uppsala universitet]. DiVA.
<https://www.diva-portal.org/smash/get/diva2:1443400/FULLTEXT01.pdf>
- Psychology Today. (2020, 10 februari). *Groupthink*.
<https://www.psychologytoday.com/intl/basics/groupthink>

Regeringen. (2015). *Medieborgarna medierna, En digital värld av rättigheter, skyldigheter – möjligheter och ansvar* (SOU 2015:94). Wolters Kluwers.
<https://www.regeringen.se/4ab02f/contentassets/aca20d317683426797fc15c0056004f3/medieborgarna-medierna-en-digital-varld-av-rattigheter-skyldigheter-mojligheter-och-ansvar-sou-201594>.

Rosenberg, M., Nicholas Confessore, N., Cadwalladr, C. (2018). How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*.
<https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>

Skolverket. (2021). *Medieutvecklingen förändrar ungas nyhetskonsumtion*.
<https://www.skolverket.se/skolutveckling/inspiration-och-stod-i-arbetet/inspiration-och-reportage/medieutvecklingen-forandrar-ungas-nyhetskonsumtion>

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, Vol. 34(3) 150–160. doi: 10.1177/0266382117722446 journals.sagepub.com/home/bir.

Steuart, J. (2019, 6 augusti). *Netflix's 'The Great Hack' highlights Cambridge Analytica's role in Trinidad Tobago elections*. Global Voices adVox.
<https://advox.globalvoices.org/2019/08/06/netflixs-the-great-hack-highlights-cambridge-analyticas-role-in-trinidad-tobago-elections/>

Sumpter, David. (2018). *Outnumbered: From Facebook and Google to fake news and filter-bubbles - the algorithms that control our lives*. Bloomsbury Sigma.

Svenska FN-förbundet, 2021. *Globala målen för hållbar utveckling*.
<https://fn.se/globala-malen-for-hallbar-utveckling/>

Swiss Policy Research. (2021). *The Media Navigator*. <https://swprs.org/media-navigator/>

Swärd, S.P (2016, mars). *Modul: Analysera och kritiskt granska. Del 6. Värdeutvecklat, utförligt och nyanserat*. Skolverket, Lärportalen. Hämtad: 23 februari, 2021, från https://larportalen.skolverket.se/LarportalenAPI/api-v2/document/path/larportalen/material/inriktningar/5-las-skriv/Gymnasieskola/024_anlaysera-o-kritiskt-granska/del_06/Material/Flik/Del_06_MomentA/Artiklar/M24_gym_06A_01_valutvecklat_utforligt_nyanserat.docx

The Decision Lab. (2021). *Why do we tend to think that things that happened recently are more likely to happen again?*.
<https://thedecisionlab.com/biases/availability-heuristic/?gclid=CjwKCAiAyc2BBhAaEiwA44-wW8qzYte>

Thi Nguyen, C. (2019). *The problem about living inside echo chambers*. TheConversation.
<https://theconversation.com/the-problem-of-living-inside-echo-chambers-110486>

Tompkins, A. (2020). *Is fact-checking effective? A critical review of what works – and what doesn't*. DW Akademie.

<https://www.dw.com/en/is-fact-checking-effective-a-critical-review-of-what-works-and-what-doesnt/a-55248257>

Torstensson, K. Blomberg Cedergren, A. (2019). AI vid rekrytering. [Kandidatuppsats, Lunds universitet]. Lund University Libraries. <https://lup.lub.lu.se/luur/download?func=downloadFilerecordOId=8983179fileOId=8983192>

Vosoughi, S. Roy, D. Aral, S. (2018). *Science*, 359(6380), 1146-1151. doi: 10.1126/science.aap9559

Walter, N., Cohen, J. R., Holbert, L., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350-375. doi: 10.1080/10584609.2019.1668894

Webb, R. (2018). “If we do it wisely, AI can be the best thing ever for humanity”. New Scientist. doi: 10.1016/S0262-4079(20)31250-1

White, C. L., Boatwright, B. (2020). Social media ethics in the data economy: Issues of social responsibility for using Facebook for public relations. *Public Relations Review*, 46(5), 101980. doi: 10.1016/j.pubrev.2020.101980

Wood, T., & Porter, E. (2019). The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence. *Political Behavior* 41, 135–163. doi: 10.1007/s11109-018-9443-y

Zeadally, S., Winkler, S. (2016). Privacy policy analysis of popular web platforms. *IEEE Technology and Society Magazine*, 75-85.

Åkerlund, D. (2017). *Guide till Akademiskt skrivande*. (Upplaga 2). Karlstad: Karlstads Universitet <https://kau.diva-portal.org/smash/get/diva2:1138556/FULLTEXT01.pdf>

Bilagor

A: Transkribering av intervjuer

Intervju 1

07:16 Klara: Är det någonting speciellt du tänker att vi borde ta med när vi pratar om etik i arbetet så, det är väl ganska svårt sak att säga kanske...

de Fine Licht: Ah men istället för att kanske fokusera på vad etik är så skulle man till exempel kunna titta på normativa teorier då till exempel som finns och relatera de till dem idéer som finns, om dem här algoritmerna är bra eller dåliga och har orättvisa/rättvisa effekter och så vidare. Och då finns det en massa olika normativa teorier, har ni hört talas om utilitarism till exempel?

16:42 Klara: Men sen då kan vi ju komma in på det också för vi funderade lite också på vad du tycker är viktigast att ha med om man ska göra något slags etiskt ramverk eller försöka sammanställa liksom, ja men olika idéer som man ska ha med i det? Är det några grejer eller punkter som du tänker att, såhär, det här är viktigt att ha med isåfall?

de Fine Licht: Asso jag tänker ju, det är ju bra då att gå igenom såklart, det här är ju något som alla säger som ni själva har gjort och gör också, och det är att man går igenom de stora ramverken så och ser hur dom ser ut.

19:40 de Fine Licht: Oftast fokuserar man bara på en policy och tänker att, asso, vad som ska stå i policyn. Men hur ska policyn vara och vilket instru... vad behöver vi kring policyn för att det ska vara bra.

44:21 Klara: Har du några bra tips förresten på några ramverk som du tycker är bra liksom? Eller såhär också kanske lite lättare att förstå som inte bara säger någonting som är lite luddigt och diffust.

de Fine Licht: Ja men EU, ehm, jag skulle kolla på, har ni kollat på EU:s?

de Fine Licht: För där kan det vara rätt bra att börja. Det ni kommer se är att alla dom här stora ramverken som... Eu har ett då, dom har en sån grupp, ageleg, som håller på att arbeta med att utveckla det där. Ett skäl att det är bra att titta på det är att det finns en massa diskussion som är lätt att komma åt. Såhär det finns folk som är jättekritiska och andra är jättepositiva. Vissa tycker det kommer gå för långt och andra tycker att det går för kort osv. Så på det sättet tycker jag det är en bra startpunkt. Mycket bloggar, mycket asso såhär, det är väldigt mycket diskussion. Och sen så tycker jag också att dom, asso det ser ungefär ut så som många av dem här stora ramverken ser ut. Det är inte jättestor skillnad på det och på A tripple E (IEEE) är en annan sån stor då, från tekniska hållet. Dom har alltså, deras ramverk ser ungefär likadant ut. Och det är för att det är ofta samma personer som sitter i de här olika grupperingarna. Alltså fysiskt samma.

52:31 Klara: Vi tänkte bara kolla är det några fler personer som du tänker på?

de Fine Licht: Ah men jag tänker Gordana i den etiska kommittén, hon är bra på sånt här med etiska policys.

52:45 de Fine Licht: Sen kan ni alltid kolla med, eller bolla med, Dorna på. Ni kan söka på Dorna AI FLOV så hittar ni Dorna. Dorna är väldigt duktig på dem etiska frågorna under AI

och har, tänker med en helt annan grupp av personer. Så dem två kan komplettera varandra rätt väl.

Intervju 2

0.55 Klara: Vad anser du vara det största problemet inom nyhetskonsumtion eller presentation?

Gordana: En sak som man skulle kunna tänka om man bara fokuserar på nyheter är att mångfalden försvinner om en algoritm gör ett urval, filtrerar och sammanställer, istället för att vi pratar med andra människor så blir det en stor förlust av mångfald. Men den processen om vi tittar historiskt sett har pågått i väldigt många år.

2.37 Gordana: Vad som händer med algoritmer är att stora företag som äger dessa artificiella intelligenta system bestämmer över hur det ser ut och det är filtrering och det är klassificering och gruppering i trådar och rankning. Så de gör hela spektrat av val som är deras egna och de är som de ser ut idag ganska oberoende, ganska självständiga i de valen om besluten som de fattar över hur det här ska gå till. Och det som man säger överallt i hela världen är att man behöver ha mer insyn hur man gör och fatta gemensamma beslut och ta med i beslutsprocessen alla andra stakeholders. Inte bara stora tekniska företag som är väldigt få i hela världen, utan olika medborgare, medborgargrupper och alla demokratiska aktörer, eftersom det handlar om demokrati. För information är det centrala, man styr hela samhället med information och om man tar bort, väljer, sammanställer på olika sätt så får man en helt annan bild. Om man vill först vara medveten om hur det går till och på vilket sätt man skapar nyheter, på vilket sätt man placerar nyheter till oss användare, på vilket sätt man värderar saker osv. Så det finns en hel del saker som pågår bakom kulisserna när man automatiserar processen, som vi gärna som medborgare vill förstå och ha någonting att säga om. Och det är en process som har redan börjat, man har börjat att bli mer medveten genom incidenter som har hänt, som ni alla vet om bias när man tolkar bilder exempelvis. Så det var en rad incidenter som var att vi insåg att algoritmer inte är helt neutralt, logiskt, välstrukturerat och fattar rätt beslut eftersom det är matematik bakom. Man förstår att algoritmen är full av val, värderingar, den gör någonting som vi vill att den gör. Och den viljan och de värderingarna vill vi vara med om och medverka i när vi skapar algoritmer som vi alla påverkas av. Så det största problemet är att vi fortfarande inte är delaktiga i hur det här går till. Vi har inte insyn och det är en väldigt central demokratisk fråga. Så det är något som kommer både i USA där de flesta finns, men också i Europa och över hela världen. Det är en process som pågår som är mycket viktig för demokratin. Och det är det jag ser som det största problemet är att på vilket sätt man påverkar demokratin, från stora företagens sida som inte är en formell del av den demokratiska processen. De har uttalat önskan om att bete sig etiskt. Ni har både Google och Microsoft och nästan alla Facebook och andra har skrivit etiska policys. De har deklarerat att de vill vara etiska och det är bra att de vill det. Och använda deras goda vilja för att integrera deras sätt att jobba på med demokratiska principer och demokratiska beslutsprocesser. Det är något som återstår. Det är inte så att de är emot etiska principer eller värderingar eller någonting. De vill gärna göra bra jobb, men vad betyder bra jobb från alla våra olika synpunkter? Det är en fråga om involvering av stakeholders och en lång process av förhandling, diskussioner, för att vi allihopa ska förstå varandras intressen och på vilket sätt vi bygger demokratin med de här nya verktygen och det är verktyg och de är väldigt bra att ha. För vi har så mycket information, den produceras överallt, mer och mer och vi måste på något sätt ha översyn över vad som pågår och strukturera saker och få snabbt nyheter och så. Så det är en klar fördel att ha hjälp av teknologin, men på vilket sätt det händer det kan vi säga någonting om som medborgare och som medlemmar i olika demokratiska strukturer. Så det är det som jag inte ser som ett problem, utan som ett

jobb, som alla tillsammans måste göra.

8:43 Elin: Okej, men en följdfråga på detta då. Om Google och alla de stora tekniska bolagen du pratar om, om de vill bete sig etiskt, vad ser du som deras största hinder för att göra det då? Är det de du säger med att man inte riktigt förstår och kan hantera..

Gordana: Jag tror att de har vuxit organiskt och har blivit från små företag till så stora jättar, att de har blivit större än vissa stater, ekonomiskt. Och de har en sån enorm makt, så det är något som är helt nytt i historien. Det har aldrig funnits såna företag som var så centrala för demokratin och samtidigt hade så mycket makt själva. Så det är väldigt unikt i historien, så vad vi måste jobba mot är att alla krafter i demokratin tillsammans med stora företag, som givetvis har som största intresse: profit, de vill tjäna pengar och det är helt legitimt, de ska tjäna pengar det är därför de finns. Men de kan inte tjäna pengar och samtidigt strida emot värderingar och etik och så. Och de vill inte heller strida mot etik eller våra gemensamma värderingar. Det är bara en fråga om på vilket sätt vi integrerar dem i demokratiska processer och demokratiska strukturer. Eftersom just nu är de oberoende och jättejättemäktiga. Så det är den motsättningen mellan demokratin som vill representera alla krafter och allting i samhället, balansera maktförhållanden och så och sådana här jättar som har den makten. De har bara vuxit fram så.

10.43 Elin: Så du anser att vi måste samarbeta mer alla krafter tillsammans lixom?

Gordana: Ja ja, jag tror att för att kunna få en balans i samhället, i hela världen är det nödvändigt att få alla krafter tillsammans för inte en enda människa eller en enda struktur kan ändra på det här. Det är en jätteprocess.

11.08 Klara: Hur skulle man då kunna samla alla goda krafter ihop och lixom jag vet inte, sälla bort dolda agendor eller sådana saker. Är det att det kommer uppifrån eller finns det risker med det också?

Gordana: Det finns professionella organisationer som har redan att börja diskutera. Det finns grupper av filosofer, etiker, professionella inom detta. Vad vi kan göra var och en som individer det är en sak, och vad samhället kan göra, för detta är egentligen ingen ny sak på så sätt att vi löser olika motsättningar mellan olika intressen eller att vi försöker att få gemensamma intressen för en stat, men även globalt.

15.35 Klara: Hur stort är problemet egentligen? Tror du att för rent teoretiskt så borde det ju vara så att eftersom man kan sprida information mycket mer nu och på global nivå att folk kan prata med varandra och diskutera med högt uppsatta personer i olika forum istället för att bara få information genom att prata med sina grannar och kanske få en tidning hemskickad till sig. Skulle du säga att det har bidragit till fler positiva eller negativa konsekvenser?

Gordana: Det finns både positiva och negativa konsekvenser självklart. Och det är jättebra och positivt att man kan kommunicera med människor som annars är väldigt svåra att nå, så vi har idag seminarier och föreläsningar av folk över hela världen och vi kan samla en panel av deltagare från var som helst så det är fantastiskt bra. Det är också bra att det finns ett sätt för politiker att nå väljare.

24:50 Elin: Jag tror att jag har kikat lite på den länken någon gång tidigare, för jag känner igen det lite. Men anser du att det är någonting som nyhetskanalerna kan göra för att ta bort problemen med filterbubblor och misinformation och desinformation?

Gordana: På vilket sätt man kan kontrollera att informationen man får är korrekt eller att man

kan lita på källan eller så?

Elin: Asså mer vad nyhets.. asså just om man kollar på vad kanalerna som sprider informationen, om dem på något sätt kan lixom ah eller vad de kan göra för att ta sitt ansvar i att minimera problemen med att sprida misinformation?

Gordana: Som jag sa så tror jag att detta är ett systemproblem. Så det ligger inte bara på media att lösa det utan på teknikföretag i första hand eftersom dem är de enda som har kunskapen om algoritmer och de bygger upp algoritmer och de kan hjälpa oss att förstå hur algoritmerna funkar och de kan diskutera med de andra varför de väljer vissa saker och inte de andra. Så vi har satt att de är väldigt lyhörda när det är någon chattbot som sprider rasism så kommer dem att ändra det direkt och om det är någon algoritm som har bias när de tolkar bilder så blir det jättereaktion i hela världen, då ändrar dem det. Men vi ser inte allting och det finns massa algoritmer som gör olika saker. Inte bara aggregerar nyheter utan också sorterar problem inom rättsväsendet.

27:49 Elin: Jag tänkte säga det att vi inom gruppen eller vårt projekt, amen lite vems problem det är. Vi inser ju det att det inte bara är en sida som står för alla problem men just det här med är det algoritmerna, sen har vi också diskuterat mycket att vi som människor och läsare till nyheter drar oss ah men med group thinking, konfirmeringsbias, att vi drar oss till den informationen vi vill höra lixom och de som jag redan tycker är det som jag vill läsa för att inte bli lixom ah att få lixom mina åsikter vad säger man, att någonting som konkurrerar med vad jag redan tycker typ och lixom ah vart det största problemet ligger då, eller ah jag vet inte riktigt vad jag ska säga.

Gordana: Jag håller med om det, mmm och om du bara vill titta på oss som användare, som users, vad kan vi göra personligen? Vi kan välja fler källor, inte isolera oss bara i en liten bubbla utan försöka att balansera oss med flera olika källor, helst från hela olika länder och olika typer av information och så. Jag litar mest på vetenskaplig information.

31:30 Gordana: Källor måste kombineras det är kort sagt. Om man kan läsa originalet är det väldigt bra, och om man inte kan läsa det så försök att läsa flera källor och försök att diskutera med människor offline.

32:10 Klara: Hur mycket ansvar tycker du att konsumenterna har i det här?

Gordana: Jag tror att vi har samhälleligt ansvar för att utbilda och informera konsumenter. Så det kommer inte spontant. Och om vi har några källor som vi kan lita på åtminstone någorlunda. Vi kan lita på skolan, vi kan lita i viss mån på andra aktörer. Om de berättar för oss någonting då kan vi följa nästa steg. Konsumenter är spontana och de tar lättaste vägar och kortaste vägar, alla vi har inte tid att, vi har inte energi för att konsumera allting. Och alla nyheter idag kommer från hela världen, inte bara från Göteborg, utan från precis hela världen och man kan inte bara följa nyheter så det är väldigt lätt att man får serverat något som är totalt konstigt. Och man märker det, ibland när man läser flera olika källor så ser man det att aha den här försöker att placera något som inte stämmer.

34:39 Gordana: Jag tror också att vi inte ska ta hela ansvaret själva som slutanvändare. Vi är bara en stakeholder i hela gruppen och de som är stora spelare är stora teknikföretagen som har AI teknologin. Det som är bra är att AI har kommit stort inom etik och alla diskuterar etiska problem inom AI. Och är seriösa och vill verkligen, alla ser väldiga fördelar med AI, AI är en fantastisk, i princip fantastisk teknologi. Den kommer som alla jättemäktiga teknologier med stora risker, och så den kan påverka oss på väldigt många negativa sätt och vi måste förebygga

detta.

42:00 Klara: Och det är jättespännande verkligen, men vi tänkte också på det här gällande de olika normativa etiker, eller vad man ska säga. Ser du, eller tycker du att de ska såhär kategoriseras in beroende på vilken AI tillämpning som det rör sig om, och isåfall om det är någon normativ etik som ska användas i just det här området?

Ja, det är rätt som ni säger. Det är beroende på kontext, hur man använder etiska överväganden och vilka teorier som passar bäst. Så till exempel om man tittar från utilitarism synpunkt så är det typiskt när man har stora grupper som samhället och många människor samtidigt. Och det är väldigt mycket utilitarism etik man tittar på när man tittar på AI etik. Man tittar från högsta nivån och försöker hitta lösningar som skulle vara bra för alla.

Intervju 3

03:20 Dorna: Det är väldigt lätt att säga att det är användarens ansvar. Att ah men läs bara liksom avtalen eller liksom sätt dig in i det, men det är så mycket och det är så snabb utveckling hela tiden på många fronter så det blir väldigt svårt för användare ehm, tänker jag att sätta sig in i det här, så det kan ju bli lite orättvist att kräva en sån sak. Ehm, jag tänker att transparens är jätteviktigt ehm och i det liksom att transparens spelar kanske inte så stor roll om man gör det svårtillgängligt alltså om informationen kring hur dom här algoritmerna fungerar och vad det är man går med på till exempel. . . .

09:30 Dorna: Transparens är tror jag viktigt men transparens spelar inte så stor roll om det inte är tillgängligt liksom för dom flesta fattar inte kod eller dom flesta kan liksom inte sätta sig in i det rent tekniska utan det tillgängliggörs liksom. För annars blir det bara tomma ord på något sätt.

14:29 Klara: men tänker du att, eller tror du att det skulle liksom räcka med att ha ett liksom välformulerat etiskt ramverk som man bara uppmuntrar stora aktörer att följa?

Dorna: Ne det tror jag absolut inte. Sen kanske det måste vara ett första steg liksom, det kanske inte går att lagstifta liksom på en gång. Men jag tror inte det räcker med frivilliga ramverk. Jag tycker inte man ser att det räcker inom något annat fält heller. Ehm sen kan det vara jättefint liksom men risken är att det bara blir. . . Jag tittade ganska nyligen på ett arbete där dom tittade på olika typer av policys och sådär och det dem kritiserade var just att dem inte var juridiskt bindande. Att det är problematiskt när dom inte är det. För att det räcker ju inte så långt. Man får ju liksom tänka på vad som finns på andra sidan av det. Jag menar om jag driver ett företag till exempel, ehm då är ju inte mitt, jag har ju andra intressen som kommer väga tyngre ehm så jag tror, det kanske är ett jättebra första steg och liksom, ha det på en policynivå liksom där man kan fundera kring och diskutera det och prova det till och med på olika sätt. Kanske bolla det också med användarna. Ehm men jag personligen tror inte att det räcker.

16:58 Klara: Men på vilket sätt skulle man kunna reglera det? Eller det kanske är en för juridisk fråga. . .

Dorna: Det är nog lite för långt ifrån det jag sysslar med men en tanke är att göra någon slags såhär mellan. . . Ett mellansteg och det kanske ni redan har tänkt på, det är ju att man kan ha en slags märkning liksom, någon slags certifiering, en tredjeparts certifiering av något slag. För då är det ju frivilligt, men det är en tredje part, som till exempelvis krav då om man tänker kravmärkning. Som har tittat på det och går i god för att, jo men det här uppfyller dem här

och dem här principerna så då kan ju konsumenten i alla fall utgå ifrån det när man tycker och tänker och väljer liksom.

28:45 Dorna: Hur har ni tänkt kring eller finns det förslag om att liksom en del av liksom sidan eller appen eller så presenterar ofiltrerat? Ehm, en liten del eller liten del liksom eller inte vet jag det beror ju på vad det är för typ av sidan eller app eller vad vi pratar om men att, typ att, en del av första sidan eller någon knapp man kan trycka på som tar en ehm liksom till en ofiltrerad ehm..

30:20 Elin: Men det är ju en bra tanke att man bara skulle typ kunna trycka på en knapp och ba, ah hur ser mitt flöde ut när det inte filtrerar överhuvudtaget liksom. Ehm.

40:39 Dorna: Det kan ju så klart finnas en bortre gräns för på hur mycket filtrering som är försvarbar. Jag vet inte hur ni tänker där ehm. Det skulle ju också kunna vara någonting som skulle kunna ingå i någon slags märkning eller certifikat. Här sker filtrering men vi är transparenta med det vi har ett reglage som användaren kan använda när som helst som är liksom, osv som finns där. Som ni var inne på själva.

Elin: Ja det är ju jättebra verkligen, det är ju faktiskt något som kan hade kunna, typ om vi nu skulle ta fram ett etiskt ramverk för nyhetsfiltrering, ha med där att, om du filtrerar så måste du berätta att du gör det och visa på vilket sätt typ.

Dorna: Precis och det kanske inte räcker med den här jättelånga väggen av text i något slags användaravtal eller så, utan ehm att det kanske är någonting man blir påmind om varje gång man använder eller går in på den här sidan till exempel. Ehm och jag vet inte hur möjligt det är att ha en sådant reglage eller någonting liknande men det kan nog vara jättebra att ha en knapp eller ett reglage eller någonting där det står, här finns liksom, ehm, ofiltrerad eller vad man ska säga ehm, en ofiltrerad version av sidan.

Elin: Ja jag tänkte precis säga det att just det där med knappen är väldigt bra att man ska kunna trycka där, och ba så, nu är det ofiltrerat.

INSTITUTIONEN FÖR TEKNIKENS EKONOMI OCH ORGANISATION
AVDELNINGEN FÖR TEKNIK, VETENSKAP OCH SAMHÄLLE
CHALMERS TEKNISKA HÖGSKOLA

Göteborg, Sverige 2021
www.chalmers.se



CHALMERS