Politiets bruk av kunstig intelligens i arbeidet mot nettovergrep mot barn

Litteraturstudie

BACHELOROPPGAVE (OPPGAVE03)

Politihøgskolen

2022

Kand.nr : 70

Antall ord: 7677

Sammendrag

PrevBOT er en «crime prevention robot» som enda ikke er utviklet. Denne roboten vil baseres på kunstig intelligens (KI) i sitt arbeid mot nettovergrep mot barn. Ved hjelp av «authorship attribution», maskinlæringsalgoritmer og robotikk kan PrevBOT hovedsakelig gjøre to oppgaver: identifisere og kategorisere mulige overgripere eller forum på internett hvor det er høy risiko for at barn vil eller kan bli utsatt for nettovergrep. Beregningene til PrevBOT vil ikke regnes som bevis, men vil kunne bidra til både forebygging og etterforskning, alt ettersom hva politiet ønsker. Nettovergrep er en stor og utbredt kriminalitetsform, det ikke finnes en klar og effektiv løsning på. En stor fordel ved at PrevBOT nytter seg av KI er muligheten til å effektivisere jobben mot nettovergrep, ved at prosessen med å søke etter overgripere og problematiske steder blir automatisert. På den andre siden trenger PrevBOT virkelige samtaler mellom barn og overgripere for å skulle trenes opp og fungere best mulig. Dette gir utfordringer med tanke på personvernet til både barnet og overgriperen. I tillegg vil PrevBOT, som all annen KI, bli påvirket av personen som programmerer systemet. Dette kan føre til systematiske skjevheter, fordi ingen systemer er mer objektive enn menneskene som lager dem. PrevBOT er kun ett forslag til hvordan man kan treffe denne kriminalitetsutfordringen.

Innholdsfortegnelse

1 Innledning

Temaet for bacheloroppgaven er kunstig intelligens (KI) og hvordan roboten PrevBOT vil baseres på KI i arbeidet mot nettovergrep mot barn. PrevBOT er et nytt konsept, som enda ikke er utviklet. Jeg ønsker derfor å se på hvilke fordeler og ulemper som følger ved at PrevBOT baseres på KI.

1.1 Begrunnelse for valg av tema

Samfunnet er i stadig utvikling og spesielt de siste 20 årene har det skjedd store fremskritt på teknologifronten. Alt fra algoritmer som gir skreddersydd reklame til brukeren, selvkjørende biler og roboter. Politiet har en rekke store oppgaver som å forebygge kriminalitet og avdekke og stanse kriminell virksomhet (Politiloven, 1995, §2). Ny teknologi gjør at kriminalitetsformene også utvikler seg og blir mer komplekse. Dette gir politiet nye utfordringer, samtidig kan også politiet bruke denne teknologiske utviklingen til sin fordel.

I takt med min interesse for både politi og teknologi ønsket jeg å se på muligheten for å kombinere disse. Mange spår at KI vil få en voksende plass i samfunnet i fremtiden. Den globale teknologibedriften PwC (u.å.) skriver på sine nettsider at KI åpner en helt ny verden av muligheter. I den forbindelse kom jeg over en robot som er tenkt basert på KI for å jobbe mot nettovergrep mot barn. I følge ECPAT[1] har økning av sosiale medier, meldinger og live-stream apper de siste årene, ført til en dramatisk økning av rapporteringer av nettovergrep mot barn (Sunde & Sunde, 2021, s. 2). Internettrelaterte overgrep mot barn, er på listen over prioriterte sakstyper i 2022 (Riksadvokaten, 2022, s. 7). I tillegg slår FNs barnekonvensjon fast at alle barn har rett til at ingen skader eller misbruker dem (Barne- og familiedepartementet, 2003, s. 16, 25 & 27). Jeg ønsker derfor å se nærmere på hvordan PrevBOT kan møte denne utfordringen ved hjelp av KI.

1.2 Problemstilling

På bakgrunn av min interesse for både politi og teknologi, ønsket jeg å ha fokus på KI i min bacheloroppgave. Derfor kom jeg frem til følgende problemstilling: *Hvilke fordeler og ulemper finnes ved at roboten PrevBOT vil baseres på kunstig intelligens, i arbeidet mot nettovergrep mot barn?*

[1] ECPAT er et globalt nettverk av organisasjoner som jobber for å få slutt på seksuell utnyttelse av barn (ECPAT, u.å.).

1.3 Avgrensning

Det valgte temaet er stort og krever en nøye avgrensning. For å kunne svare ut problemstillingen min, kreves det at en del teori bli redegjort for, før selve diskusjonen. Jeg vil ikke drøfte andre forebyggende tiltak rettet mot nettovergrep mot barn annet enn PrevBOT. Allikevel kommer jeg til å nevne kort noen andre systemer som er sentrale for utviklingen av PrevBOT. Tyngden i oppgaven vil ligge på det tekniske bak PrevBOT, samt fordeler og ulemper ved at PrevBOT er basert på KI. Jeg vil i hovedsak fokusere på Norge og utviklingen av PrevBOT med norsk som språk. Dette fordi språklig uttrykksmåte ikke er direkte overførbar gjennom oversettelser. Allikevel vil prinsippene og teknologien bak roboten være den samme, uavhengig av hvilket land som utvikler den.

Jeg kommer ikke til å ta for meg hjemmelsgrunnlaget for at en slik robot kan operere på internett. Jeg kommer ikke til å ta for meg hvordan PrevBOT må vedlikeholdes, oppdateres og kontinuerlig driftes. Videre kommer jeg kun til å ta for meg teknologien slik den er tiltenkt på dette punkt. Jeg kommer ikke til å utdype hva slags ekstern kunnskap om teknologi som kreves for å utvikle roboten. Jeg kommer ikke til å ta for meg hva forebyggende politiarbeid eller etterforskning er, samt hvor PrevBOT favnes under det. I tillegg vil jeg ikke ta for meg personer som begår forhold som omhandler overgrepsmateriale, selv om mange av disse også begår nettovergrep.

1.4 Oppgavens oppbygning

For å belyse min problemstilling vil jeg definere sentrale begreper, samt utdype og forklare sentrale komponenter i oppgaven. Dette ved hjelp av faglitteratur og forskning. Deretter vil jeg se på hva som gjør at vi kan si at PrevBOT vil baseres på KI, samt fordeler og ulemper bruken av KI medfører. Avslutningsvis vil jeg oppsummere oppgaven.

1.5 Begrepsavklaring

Jeg vil i denne begrepsavklaringen redegjøre for begreper som brukes i oppgaven. Noen begreper vil beskrives ved hjelp av fotnoter i løpende teksten, slik at de som ønsker mer informasjon finner det der det trengs. Jeg vil bruke noen engelske ord og uttrykk, fordi det ikke finnes gode norske oversettelser. Redegjørelsen for blant annet PrevBOT og KI vil komme i teoridelen, da disse krever mer informasjon for å skaffe den forståelsen som kreves for å forstå oppgaven.

Internett betegnes som et nettverk av datamaskiner og nettverkskomponenter som er koblet sammen (Bjerknes et al., 2018, s. 265). Elektroniske kommunikasjonstjenester, herunder sosiale medier og nettforum, er hjelpemidler ved formidlingen av ytringer til en adressat som enten er menneske eller datamaskin (Sunde, 2016, s. 13-14). Her kan personer kommentere eller diskutere bestemte emner med en eller flere personer samtidig (Cambridge Dictionary, u.å.-a).

Ved seksuelle overgrep mot barn på internett, også kalt nettovergrep, misbruker overgriperen internett og mobiltjenester for å utnytte barn seksuelt (Politiet, u.å.). Politiet betegner nettovergrep som internettrelatert seksuell utnyttelse av barn ved straffbar handling via internett (Kripos, 2019, s. 13). Jeg vil her bruke ordet «barn» om personer under 16 år, som også er den seksuelle lavalder i Norge.

Ordene «problematic» eller «problematisk», som nevnes senere i oppgaven, vil knytte seg til steder der det er en viss risiko for at barn vil bli, eller forsøkt utsatt for nettovergrep. I tillegg vil ordene bli brukt om personer hvor det er beregnet en viss risiko for at disse vil begå, eller forsøke å begå, nettovergrep mot barn.

2 Metode

2.1 Valg av metode

I følge Aubert er metode er en fremgangsmåte eller et middel for å løse problemer og komme frem til ny kunnskap på (Hellevik, 2002, s. 12). Problemstillingen avgjør valg av metode. Metoden som velges er den som vil belyse problemstillingen best mulig (Dalland, 2017, s. 51). Jeg har valgt litteraturstudie. Dette innebærer å bruke etablerte teorier, argumenter, informasjon og relevant forskning for å belyse min problemstilling.

Bakgrunnen for valg av litteraturstudie er at jeg kom over en spesielt interessant forskningsartikkel som omhandlet temaet KI og nettovergrep. Som supplement til dette fant jeg annen forskning rundt KI og nettovergrep, som var ny og oppdatert. Dette er viktig for et tema som utvikler seg raskt. For å kunne presentere og svare ut problemstillingen min på best mulig måte, falt dermed valget mitt på litteraturstudie.

2.2 Forforståelse

Med forforståelse menes en slags forkunnskap, som gjør at man ikke oppfatter virkeligheten kun gjennom sansene (Thurén et al., 2009, s. 66). Thurén skriver videre at forforståelsen vil prege vår måte å se verdenen på, mer enn vi tenker. Det betyr at forforståelsen vil påvirke hvordan jeg svarer ut problemstillingen min, fordi jeg tolker kunnskapen jeg finner i lys av min forkunnskap. Det er derfor viktig å ha et bevisst forhold til egen forforståelse.

Jeg har lenge hatt en interesse for KI, som ble vekket gjennom muligheten til å skape noe som kunne lære på egenhånd. Helt siden jeg kom over KI for første gang har jeg lest både bøker og hørt utallige Podcaster om KI. Fra jeg begynte på Politihøgskolen, har jeg vært fascinert av hvordan politiet kan ta i bruk KI for å løse sitt samfunnsoppdrag. I løpet av min politikarriere har jeg ikke dannet meg noen erfaringer over utfordringer som blir løst ved hjelp av KI i politiet i dag. Mitt første møte med dette var forskningsartikkelen til Sunde og Sunde som omhandlet PrevBOT. Denne artikkelen fant jeg helt i begynnelsen av arbeidet med bacheloroppgaven. Utover dette har jeg ingen knagger å henge KI i politiet på.

2.3 Litteratursøk

Jeg har brukt søkemotorer som Google, Google Scholar, Oria, Nasjonalbiblioteket og Connected Papers. Her har jeg blant annet brukt søkeordene «Kunstig intelligens», «Artificial Intellegence», «PrevBOT» og «AI in the police». Videre har jeg også søkt på «nettovergrep» og «seksuelle overgrep mot barn på internett». Gjennom dette har jeg funnet forskrifter, rapporter og statistikk. I tillegg har jeg lent meg på Inger Marie Sunde og Nina Sundes artikkel om PrevBOT. Denne artikkelen har igjen ført meg videre til andre aktuelle forskningsartikler og offentlige dokumenter utgitt av blant annet Kripos og Politiet. I samme artikkel fant jeg også en masteroppgave fra NTNU om matematikken og teknologien bak «authorship analysis».

2.4 Kildekritikk

Kildekritikk handler om å være kritisk til det kildematerialet en velger å bruke i oppgaven (Dalland, 2017, s. 72). Dalland skriver videre at det handler om hvilke kriterier man benytter seg av under utvelgelsen av kildene. Jeg har hatt fokus på å finne ny og oppdatert forskning da KI er et fagfelt som er i kontinuerlig utvikling. Jeg har tatt i bruk kunnskap fra forskningsstudier, blant annet fra Nederland og Norge. Videre har jeg brukt Kripos sine rapporter om nettovergrep. I tillegg har jeg lent meg på offentlige dokumenter om KI og nettovergrep utgitt av Regjeringen.

Jeg har sett på disse kildene uavhengige av hverandre, for å finne ut av om funnene deres samsvarer. Kildene jeg har valgt er pålitelige, da de er skrevet og utgitt av blant annet Politiet, Kripos og norsk institutt for forskning om oppvekst, velferd og aldring (NOVA). Jeg har også brukt masteroppgaven «Automated detection of perpetrators in grooming conversations in Norwegian» fra NTNU. Jeg har brukt internasjonal litteratur om teknologi.

Hovedkilden min er forskningsartikkelen fra Inger Marie Sunde og Nina Sunde som forklarer konseptet PrevBOT. Dette er en kvalitativ forskningsartikkel, da den går i dybden på ett gitt konsept. PrevBOT blir kun beskrevet i denne artikkelen. Dette er grunnen til at jeg kun har brukt denne kilden for å forklare PrevBOT. Inger Marie Sunde er professor i rettsvitenskap og leder forskergruppen «Politiet i et digitalisert samfunn». Nina Sunde er politiutdannet og har en mastergrad innen informasjonssikkerhet og cyberkriminalitet fra NTNU. Mye av drøftelsen i oppgaven vil basere seg på allerede tolket og behandlet informasjon som er blitt publisert i bøker, som dermed anses å være sekundærinformasjon (Johannessen et al., 2016, s. 387).

3 Teori

3.1 Nettovergrep

Under vil jeg gi forståelse for fenomenet nettovergrep. I tillegg vil jeg utdype hvem overgriperen er, samt atferden til overgriperen. Dette er informasjon PrevBOT vil nytte seg av for å kunne gjøre sine oppgaver, noe jeg kommer tilbake til senere i oppgaven.

3.1.1 Fenomenforståelse

Nettovergrep og seksuell utnyttelse av barn på internett finnes over hele verden og er et samfunnsproblem på tvers av landegrenser (ECPAT, u.å.). Dette var en utfordring også før internett, men internett har åpnet nye muligheter fordi barn er på internett uten foreldre og foresatte (Sunde & Sunde, 2021, s. 1). Mennesker opplever mer intimitet og deler mer informasjon på nett, enn ansikt til ansikt. Hvis en person har til hensikt å utnytte denne åpenheten, kan tilbøyeligheten til åpenhet via nettkommunikasjon gjøre oss mer sårbare for å bli rammet av den andre (Aanerød & Mossige, 2018, s.15). Forskning viser at internett som miljø reduserer og fjerner hemninger mennesker ellers har i den virkelige verden (Aiken, 2017, s.13.14). Dette fordi internett har lavere oppdagelsesrisiko og gir høyere anonymitet. I tillegg bidrar internett til psykologisk distansering fordi overgriperen ikke står ansikt til ansikt med

offeret. Videre normaliseres overgrepene på forumene fordi her møter overgriperne likesinnede (Sunde, 2019, s. 185).

Det er mest vanlig å initiere kontakt mellom fornærmet og overgriper på chatterom, sekundært på gaming-plattformer og sosiale medier med chattefunksjoner. På disse plattformene finner vi barn helt ned i 8-årsalderen (Sunde & Sunde, 2021, s. 1-2). Samlet sett ser vi at teknologien har medført en kraftig økning i barns byrde som ofre for nettovergrep (Sunde, 2019, s. 178).

3.1.2 Hvem er overgriperen?

Majoriteten av de seksuelle overgrepene blir begått av menn (Kripos, 2019, s. 35). Andelen kvinner er estimert til rundt 15-20% av tilfellene. Kripos skriver at den største gruppen seksualovergripere mot barn på internett er barn eller ungdom selv. Over 50% av overgriperne i de alvorligste sedelighetssakene i Norge i perioden 2012-2016 var mellom 15 og 24 år (Aanerød & Mossige, 2018, s.23). De som har begått overgrep mot barn på internett er yngre, har mer utdannelse og deler i mindre grad bosted med barn, sammenlignet med de som har begått fysiske overgrep (Kripos, 2019, s. 37). Halvparten av voksne seksuelle lovbrytere begår sitt første overgrep allerede i barne- eller ungdomsårene, og antall unge som anmeldes for overgrep øker (Justis- og beredskapsdepartementet, 2021, s. 22).

Personer som blir prioritert for videre etterforskning er ofte i livssituasjoner som gir dem tilgang til barn enten som far, trener eller andre tillitsverv. Det finnes færre opplysninger om mennesker som begår seksuelle overgrep mot barn på internett, enn i den fysiske verden (Aanerød & Mossige, 2018, s.84). Dette medfører utfordringer for forståelsen av fenomenet. Noen personer flytter overgrepene fra fysiske arenaer til internett, andre kommer inn på internett uten tidligere bakgrunn (Justis- og beredskapsdepartementet, 2021, s. 22). Kripos (2019, s. 34) påpeker derimot at det blir stadig tydeligere at de som utnytter eller forgriper seg på barn over internett utgjør en heterogen gruppe, med stor variasjon i alder, etnisitet, yrke og sosial status.

3.1.3 Atferden til overgriperen

Forskning indikerer at overgriperen normalt skjuler eller maskerer sin identitet på internett (Sunde & Sunde, 2021, s. 4). De tre mest brukte måtene å skjule sin identitet på var å late som man var yngre enn man egentlig var, late som man var barn eller tenåring, eller ved bruk av falskt profilbilde. I tillegg var det vanlig å lyve om hvilket kjønn man var. Strategien med å

lyve om alder og kjønn er brukt i flere norske saker om nettovergrep (Sunde, 2019, s. 178). Overgripere kommer i kontakt med mindreårige barn basert på at barn og unge tror de snakker med jevnaldrende barn og unge på nett (NTNU, u.å.). Ved hjelp av høy teknisk kompetanse, klarer overgriperen å skjule identiteten sin.

Overgriperne går vanligvis etter ukjente personer, både barn og voksne, og tar hovedsakelig kontakt med barn eller tenåringer som publiserer seksuelt innhold i en eller annen form. Dette kan eksempelvis være gjennom profilbilder, brukernavn eller meldinger (Sunde & Sunde, 2021). Overgriperne ønsker å være anonyme og er svært sikkerhetsbevisste. I tillegg til å bruke fiktiv identitet, kan de veksle mellom flere ulike identiteter (Sunde, 2019, s. 181-182). En måte overgriperne kontrollerer risikoen på, er å isolere barnet ved for eksempel og lytte kommunikasjonen til andre mer private plattformer, samt holde kommunikasjonen skjult fra foreldre og foresatte. Sunde skriver videre at risikostyring vil skje tidligere på internett enn i den virkelige verden.

Ordinære samtaler, da uten seksuell karakter, spiller en stor rolle i starten av kontakten hvor overgriperen da fokuserer på å danne et vennskap med barnet (Sunde & Sunde, 2021, s. 4). I tillegg bruker overgriperne ofte overtalelsesteknikker når de tar kontakt med barnet, enten ved utpressing eller ved å love kjærlighet og omtanke. Videre nevner Sunde & Sunde at det er vanlig at overgriperen sender overgrepsmateriale til barnet for å normalisere overgrep både på internett og i den fysiske verden.

Overgriperne blir delt inn i to typer: de som har til hensikt å møte barnet og begå fysiske overgrep[2], såkalt kontaktdrevet, og de som har til hensikt å utnytte barnet seksuelt over internett, såkalt fantasidrevet (Sunde & Sunde, 2021, s. 3). European Online Grooming Project har delt gjerningspersonene inn i tre hovedtyper som ansees bedre enn skillet mellom kontaktdrevet og fantasidrevet. Disse tre er intimitetssøkende, tilpasningsdyktige og hyperseksualiserte (Kripos, 2019, s. 38). Den intimitetssøkende forandret ikke identiteten sin, men ønsket å bli likt for den han eller hun var. Den tilpasningsdyktige tilpasset sin identitet og modus til den de kom i kontakt med. Videre mente han eller hun at barna var modne og i stand til å stoppe overgrepet dersom de ønsket det. Den hyperseksualiserte sin kontakt med de unge var svært seksualisert

[2]Prosessen hvor en voksen oppretter kontakt med et barn i den hensikt å møte barnet og begå et seksuelt overgrep, er også kjent som grooming (Eirik Teigestad, 2017).

og eskalerte fort. Han eller hun ønsket umiddelbar seksuell tilfredsstillelse og hadde ikke et mål om å skape en relasjon med fornærmede (Sunde & Sunde, 2021, s. 11).

3.2 Kunstig intelligens

KI kan være vanskelig å forstå. Jeg har under forsøkt å gi en forklaring på hva KI er, samt en beskrivelse på hva som regnes som KI og ikke.

3.2.1 Definisjon

Hva som regnes som KI kan være utfordrende å forholde seg til, fordi det ikke finnes en presis definisjon på hva KI er (Elements of AI, u.å.). Definisjonen og forståelse om begrepet endrer seg i takt med teknologiutviklingen og endringer skjer kontinuerlig. Helsingfors Universitetet skriver videre gjennom sitt nettkurs Elements of AI, at det å ramse opp typiske egenskaper for KI, er for noen en mer hensiktsmessig måte å definere KI på. Slike egenskaper er blant annet autonomi og adaptivitet. Autonomi er evnen til å utføre oppgaver i komplekse omgivelser uten kontinuerlig hjelp fra mennesker. Adaptivitet er evnen til å forbedre prestasjonen ved å lære av erfaringer. Allikevel vil disse egenskapene, så snart de anses selvsagte, bli utdatert og ikke lenger være en del av definisjonen (JOU 2020:5, s. 11).

Det finnes tre hovedgrunner til at KI er vanskelig å definere: 1) det er i kontinuerlig utvikling, 2) kulturen skaper forventinger som gir et uriktig bilde av hvordan teknologien utvikler seg og 3) KI-systemer har problemer med å gjøre oppgaver mennesker anser enkle, mens det er relativt lett å lage et KI-system som kan løse oppgaver som ansees komplekse og vanskelige for mennesker (JOU 2020:5, s. 11).

For å gjøre KI enklere å forstå vil jeg i denne oppgaven bruke EUs ekspertgruppes definisjon på KI. Det er også denne definisjonen som er gjengitt i Nasjonal strategi for kunstig intelligens:

> «Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål. Enkelte KI-systemer kan også tilpasse seg gjennom å analysere og ta hensyn til hvordan tidligere handlinger har påvirket omgivelsene». (Kommunal- og moderniseringsdepartementet, 2020, s. 9).

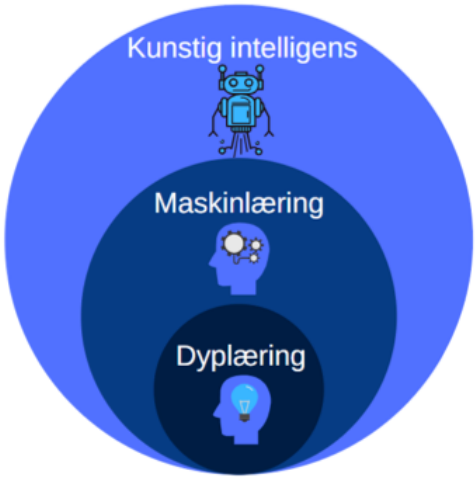
Systemer som tolker data[3] eller får data og tar avgjørelser basert på denne dataen favnes her av definisjonen. Mange KI-systemer tar avgjørelser basert på læring, og data er en viktig del av avgjørelses- og læringsprosessen. Avgjørelsene som blir tatt av KI-systemer, baserer seg ofte på sannsynlighet. Innenfor datateknologi representerer sannsynlighet sjansen for at en avgjørelse er riktig. Sannsynligheten kan være beregnet på bakgrunn av datasettet systemet har lært av. Datasystemene kan ha visse mønster som kan kategoriseres og brukes i sannsynlighetsberegning i senere tid (JOU 2020:5, s. 11-12).

3.2.2 Hva er kunstig intelligens?

Det skilles mellom sterk og svak KI. KI som ligner på menneskelig intelligens, blir omtalt som kunstig generell intelligens. Kunstig generell intelligens er sterk KI. Svak KI regnes som systemer som gir spesifikke løsninger utviklet med tanke på én oppgave. For eksempel bildebehandling eller mønstergjenkjenning for bestemte formål (Kommunal- og moderniseringsdepartementet, 2020, s. 9-10). Eksempler på praktiske anvendelser av KI i dag er blant annet robotikk som kan brukes til å utvikle autonome fartøy som biler, skip og droner. Videre kan KI brukes til å gjenkjenne mønstre eller avvik for eksempel til å avsløre bank- og forsikringssvindel. I tillegg kan man bruke KI til å identifisere objekter i bilder, såkalt «computer vision». Her favnes blant annet ansiktsgjenkjenning (Kommunal- og moderniseringsdepartementet, 2020, s. 11).

Det finnes ulike nivåer innenfor KI; kunstig intelligens, maskinlæring og dyplæring. Maskinlæring dreier seg om å lære opp et system, basert på data (JOU 2020:5, s. 11-12). Dyplæring, også kalt nevrale nettverk, er inspirert av hjernen. Disse er bygget opp av flere «lag» av prosesseringsenheter, kalt nevroner, som er knyttet sammen via synapser (Tidemann, 2021). Dyplæring er de mest komplekse systemene innen KI.

[3] Når data tillegges mening eller betydning blir det informasjon. Data er i seg selv lite nyttig, men ved kunnskap om hva den representerer og hvordan den skal tolkes blir den nyttig (Nätt, 2021).



Figur 1: Illustrasjon av nivåer innen kunstig intelligens (JOU 2020:5, s. 12)

3.2.3 Maskinlæring

Teknologi som benytter seg av KI i dag, er som regel løsninger som baserer seg på maskinlæring. I maskinlæring blir reglene utledet fra de dataene systemet trenes på. Det vil si at reglene ikke er gitt av mennesker, men utledet av informasjonen systemet får ut av dataene. Ved utvikling av KI-systemer med maskinlæring, vil maskinlæringsalgoritmer bygge matematiske modeller basert på eksempeldata eller treningsdata, som deretter brukes til å ta beslutninger. Disse systemene lærer på 3 ulike måter; veiledet læring, ikke-veiledet læring eller forsterkende læring (Kommunal- og moderniseringsdepartementet, 2020, s. 11). Jeg vil ikke utdype disse i oppgaven.

3.2.4 Dyplæring

Dyplæring er en læreprosess innen maskinlæring, som nevnt over. Det går ut på å trene opp såkalte «dype kunstige nevrale nettverk». Prinsippet dreier seg om at datamaskiner skal tilegne seg kunnskap om noe den ikke vet eller kan fra før. Det er dyplæring som har stått for de største gjennombruddene de siste årene innen maskinlære, når det gjelder maskinell forståelse av bilder, tekst og sekvenser (Tidemann, 2021). Noen dyplæringsalgoritmer kan sammenlignes med en sort boks, der man ikke har innsyn i modellen som forklarer hvordan en gitt inndataverdi har fått sitt resultat. Dette er noe som bidrar til lite transparens og det kan være ønskelig å bruke en annen tilnærming enn dyplæring, nettopp fordi man ønsker å se hvordan en kommer frem til resultatene (Kommunal- og moderniseringsdepartementet, 2020, s. 12 & 58).

3.3 Introduksjon til PrevBOT

PrevBOT er en forkortelse for «crime prevention robot» og er per nå et konsept som enda ikke er utviklet. PrevBOT er tenkt som et verktøy politiet kan bruke for å forhindre nettovergrep mot barn ved å identifisere problematiske forum og personer på internett. Disse stedene er forholdsvis chatterom og forum, som ikke er trygge med tanke på nettovergrep. PrevBOT vil bli skapt etter programmet Sweetie 2.0 som kan observere åpne samtaler og samhandle automatisk i chatterom. I tillegg vil PrevBOT inneholde komponenter fra systemet AiBA som bruker maskinlære og «authorship analysis» til å forutsi alder og kjønn bak online alias som snakker seksuelt med barn. Hva dette er, vil jeg ta for meg senere i oppgaven. AiBA kan også finne ut om personen faktisk er den han eller hun utgir seg for å være (Sunde & Sunde, 2021, s. 6-7).

3.3.1 Sweetie 2.0

Sweetie 2.0 var et programvaresystem som ble brukt over hele verden, 24/7 i 10 uker, for å identifisere mulige overgripere. Formålet med Sweetie 2.0 var å forhindre nettovergrep mot barn. Sweetie 2.0 ble utviklet både som en chatbot og en 10 år gammel dataanimert virtuell filippinsk jente (Sunde & Sunde, 2021, s. 6). Bilder av henne ble lagt ut på chatterom og datingsider som gjorde at personer kunne kontakte henne. Da personer begynte å skrive til Sweetie 2.0 på en seksuell måte, observerte hun eller automatisk samhandlet i samtalene, uten menneskelig innblanding (Sunde, 2019, s. 179). Programmet kunne brukes i flere chatterom samtidig. Alle chattene ble lagret og brukt til å advare, spore opp eller etterforske gjerningspersoner. (Terre Des Hommes, u.å.).

3.3.2 AiBA

AiBA står for «author input behavioural analysis» og er et system som kan avsløre personer på nett som utgir seg for å være en person de ikke er (Sunde & Sunde, 2021, s. 6). Dette skjer basert på tastetrykkdynamikk[4] og stylometri[5] (NTNU, u.å.). AiBA bruker atferdsbiometri[6] og lærende algoritmer for å analysere chatsamtaler, for å avgjøre alder og kjønn på chatdeltagerne. Ved maskinlæring gjør systemet en kontinuerlig analyse av alle samtalene og risikovurderer de

[4] Tastetrykkdynamikk dreier seg om hvordan en person bruker tastaturet, rytme og slagkraft (Aakervik, 2020).
[5] Stylometri er analysen av litterære stiltrekk som kan kvantifiseres statistisk, slik som setningslengde, ordforrådsmangfold og frekvenser av ord og ordformer (Sunde & Sunde, 2021, s. 5).
[6] Atferdsbiometri handler om hvordan man gjør ting, for eksempel hvordan man bruker tastaturet på en PC når man skriver. Disse kjennetegnene kalles for biometriske kjennetegn og er unike for enkeltpersonen, samtidig som de er stabile eller permanente over tid. Ved å måle disse kjennetegnene kan de benyttes til å gjenkjenne en person, eller bekrefte en persons påståtte identitet (Aakervik, 2020).

utfra bestemte kriterier. Dersom AiBA detekterer en høy risiko for mulige seksuelle overgrep i chatten, vil barnet varsles om dette (Aakervik, 2020). Teknologien bak AiBA kombinerer atferdsbiometri og språkvitenskap. Målet er at teknologien skal implementeres i plattformer og applikasjoner der barn befinner seg, som MoviestarPlanet, Snapchat og Instagram (Furberg, 2019).

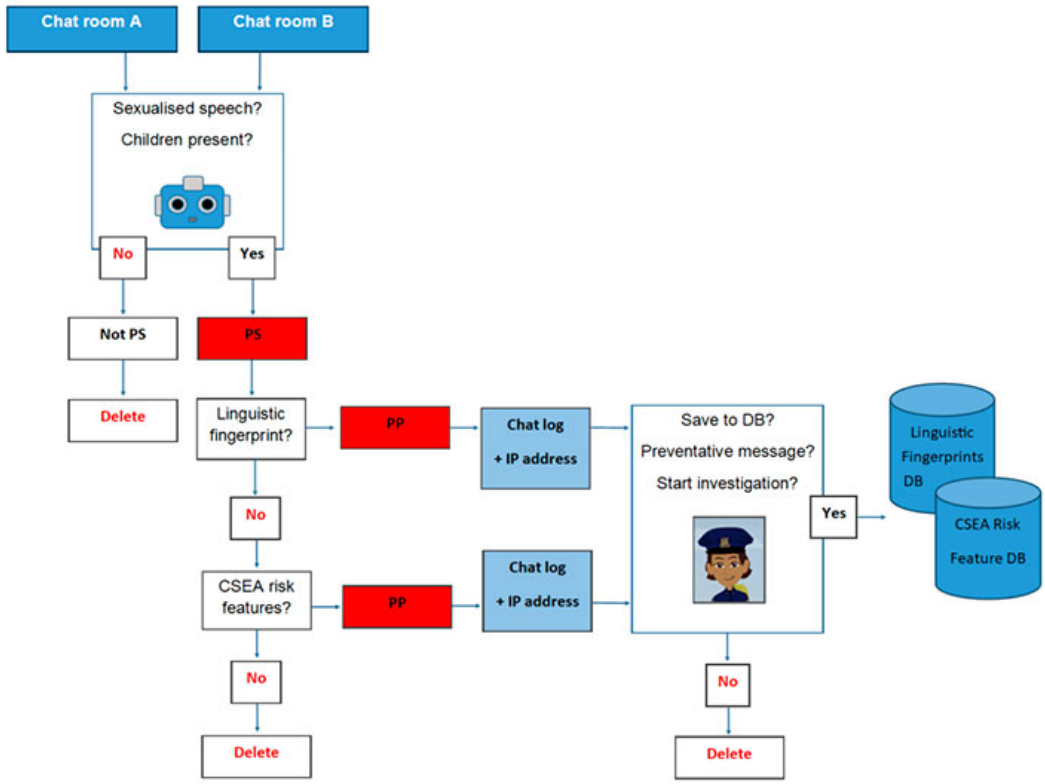
4 Resultat

4.1 Teknologien bak PrevBOT

PrevBOT vil predikere og beregne informasjon som er relevant til forebygging og etterforskning av nettovergrep mot barn. PrevBOT utfører hovedsakelig to handlinger: identifisering og kategorisering (Sunde & Sunde, 2021, s. 1). Dette dreier seg om å klassifisere og identifisere «problematic spaces» (PS) og «problematic persons» (PP) på internett. Gjennom maskinlæringsalgoritmen vil PrevBOT kunne predikere og kategorisere alder og kjønn på personer bak brukernavn som snakker seksuelt med barn, samt identifisere tidligere overgrepsdømte som har gjenopptatt ulovlig aktivitet på internett. Resultatet blir et sannsynlighetsutsagn, gitt med en viss grad av usikkerhet (s. 8). Dette gir et bilde på hvor stor risiko det er for at et barn vil bli utsatt nettovergrep av en gitt person eller via et gitt forum. PrevBOT bruker kunnskap om nettovergrep og nettovergriperen, for å kategoriseringen personene og forumene.

4.1.1 «Problematic spaces»

Et PS er et sted på internett der det er en viss risiko for at barn som er tilstede, vil bli utsatt eller forsøkt utsatt for nettovergrep (Sunde & Sunde, 2021, s. 7). Disse stedene er som nevnt ulike forum. Spørsmål PrevBOT ønsker å søke svar på for å avgjøre dette, er 1) om det er seksualisert språk i forumet og 2) om det er barn og voksne til stede i forumet (s. 8). Dersom PrevBOT finner ut at det både er barn til stede og at det foreligger seksualisert språk, vil PrevBOT flagge forumet som et PS og politiet bør følge nærmere med. Se figur under.



Figur 2: Illustrasjon av PrevBOTs beslutningsprosess (Sunde & Sunde, 2021, s. 7).

4.1.2 «Problematic persons»

Dersom PrevBOT har lokalisert et PS, vil roboten jobbe videre for å mulig lokalisere en PP. Det er tenkt at PrevBOT skal være koblet til en database med chatsamtaler fra tidligere dømte nettovergripere (Sunde & Sunde, 2021, s. 9). PrevBOT vil gjøre en sanntidsanalyse av chatten og kalkulere et såkalt «linguistic fingerprint»[7]. Det er dette språklige fingeravtrykket som forsøkes matchet mot databasen. På denne måten kan PrevBOT finne ut om en overgriper har gjenopptatt ulovlig aktivitet. Dersom PrevBOT ikke får treff mot databasen, fortsetter letingen ved hjelp av andre momenter, såkalte «CSEA risk features»[8]. Dersom PrevBOT får treff, enten ved «linguistic fingerprint» eller «CSEA risk features», blir personen flagget som PP og politiet bør følge mer med på denne personen (s. 8). Se figur over.

[7] Direkte oversatt betyr «linguistic fingerprint» språklig fingeravtrykk. Konseptet er basert på antakelsen om at folk bruker språket forskjellig, og at forskjellen kan observeres med høy grad av sikkerhet, i likhet med fysiske fingeravtrykk. «Linguistic fingerprint» kalkuleres dermed på bakgrunn av de språklige dataene og funksjonene som gjør skrivestilen til en person unik (Sunde & Sunde, 2021, s. 6).

[8] «CSEA risk features» er ulike egenskaper som knyttes til økt risiko for at en person ønsker å begå nettovergrep. Disse er alder, kjønn og seksualisert språk, som sammen brukes for å identifisere og kategorisere en person (Sunde & Sunde, 2021, s. 6).

4.1.3 Etter identifisering av PS og PP

Dersom PrevBOT kategoriserer en chat eller en person som problematisk, vil deltagerne i chatten bli varslet om risikooppdagelsen til politiet (Sunde & Sunde, 2021, s. 9). I tillegg kan politiet velge om de vil sende ut en forebyggende melding som informerer den flaggede personen om at aktiviteten han eller hun gjør er ulovlig og at vedkommende har snakket med politiet. Personen kan også bli varslet om at politiet har lagret samtalen og IP-adressen[9] til personen, på grunn av problematisk oppførsel rundt barn på internett (s.8). Politiet kan også motivere personen til å søke profesjonell hjelp. Dersom en flagget person er uenig i risikooppdagelsen til PrevBOT, er det mulig for personen å gi en tilbakemelding til politiet. Sammen med meldingen politiet sender til den flaggede personen, vil det være en lenke som forteller vedkommende hvordan han eller hun kan klage på hendelsen (s.9).

Å identifisere en PP og en PS kan skje både ved passiv observasjon av forumet eller ved automatisk samhandling i åpne samtaler (Sunde & Sunde, 2021, s. 7). I tillegg kan PrevBOT observere og samhandle i flere chatter samtidig. Politiet bestemmer hvilke chatterom PrevBOT skal inn i, samt monitorerer informasjonen som kommer frem av samtalen. Både kategorisering og identifisering forutsetter en viss mengde tekst, som er vanskelig å forutsi (s. 9). Derfor må det påregnes at PrevBOT starter en-til-en samtale for å kunne få tilstrekkelig med informasjon til å gjøre fullstendige beregninger.

4.2 Hvordan baseres PrevBOT på KI?

PrevBOT nytter seg av ulike teknologikomponenter og systemer som regnes å være kunstig intelligente, noen av disse er allerede nevnt. Jeg har valgt å dele disse i tre undergrupper: «authorship analysis», maskinlæringsalgoritme og robotikk. Undergruppene er ikke selvstendige, og må ses i sammenheng.

4.2.1 Authorship analysis

«Authorship analysis» er PrevBOTs maskinlæringskomponent. «Authorship analysis» handler om å utlede egenskapene til forfatteren av en tekst, utfra kjennetegnene som kommer frem av selve teksten (Juola, 2007, s. 119). På den måten kan man si noe om forfatteren, uten at dette er låst til forfatterens identitet. Det finnes to ulike underkategorier av dette som er svært sentrale

[9] Alle enheter som kommuniserer over internett, har en IP-adresse som gjør det mulig å motta og sende datapakker over internett. IP-adressen er unik og sikrer at informasjon ikke blir sendt feil (Kripos, 2019, s. 13).

for utviklingen av PrevBOT: «authorship attribution» og «author profiling» (Sunde & Sunde, 2021, s. 5). Dette gir flere ulike tilnærminger PrevBOT kan bruke for å identifisere og kategorisere personer og forum.

«Authorship attribution» handler om å kunne identifisere den faktiske forfatteren av en tekst basert på skrivestilen til personen (Juola, 2007, s. 120). Dette gjennomføres ved at «linguistic fingerprints» forsøkes matchet mot databasen, som nevnt over. «Author profiling» handler om å klassifisere forfattere i ulike grupper, for eksempel alder, kjønn, morsmål, personlighetstrekk eller følelser. Dette gjøres ved bruk av stylometri og tastetrykkdynamikk (Sunde & Sunde, 2021, s. 5). «Author profiling» kan gi forskere informasjon på om du er mann eller kvinne, ung eller gammel, med opptil 80% sikkerhet. Dersom man i tillegg legger til hvilke ord de bruker, kan dette avgjøres med opptil 90% sikkerhet (Aakervik, 2020).

4.2.2 Maskinlæringsalgoritmen

En algoritme er en fremgangsmåte eller en oppskrift som beskriver steg for steg hva som skal gjøres for å løse et problem eller oppnå et bestemt resultat (Moe, 2019). Maskinlæringsalgoritmer utvikler kunnskap ved å tilegne seg unike egenskaper basert på treningsdataene. Det vil si at algoritmene lærer underveis mens de utfører oppgavene de er utviklet for å løse (Sunde & Sunde, 2021, s. 10). Maskinlæringsalgoritmen i PrevBOT vil bygge matematiske modeller basert på treningsdata, som deretter brukes til å ta beslutninger og utforme risikoutsagn i form av en gitt sannsynlighet. Ved hjelp av maskinlæringsalgoritmen vil «authorship analysis», inkludert sammenligning av «linguistic fingerprints», skje uten menneskelig innblanding (s.11). Det samme gjelder prosessen med å lete etter «CSEA risk features», dersom det ikke er treff på «linguistic fingerprints» mot databasen. Etter hvert som PrevBOT får mer erfaring og følgelig lærer mer, er tanken at PrevBOT vil kunne identifisere PS og PP med høyere nøyaktighet.

4.2.3 Robotikk

PrevBOT er per nå kun et konsept som er tenkt å være en «crime prevention robot». Cambridge Dictionary (u.å.-b) definerer roboter som maskiner som blir kontrollert av en datamaskin, som utfører jobber automatisk. Dette skjer ved at roboten sanser og interagerer med omgivelsene (Søraa, 2022). De tar inn data, gjør vurderinger basert på disse og gjennom maskinlæringsalgoritmer utfører de handlinger. Roboter lærer å utføre sine oppgaver fra

mennesker gjennom maskinlæring. Videre skriver Søraa at roboter består av programvare, såkalt software, og maskinvare som gir roboten en kropp å bevege seg i, såkalt hardware. Det er usikkert hvordan PrevBOT er tenkt utviklet i fremtiden, men Søraa skriver videre at i dagligtalen brukes ordet robot også om programmer som ikke er kroppsliggjort, som for eksempel chatboter. De mer avanserte robotene har innebygd KI, slik at de selv lærer hvordan de best kan utføre oppgavene de er satt til å gjøre.

4.3 Fordeler ved at PrevBOT baseres på KI

Den generelle digitaliseringen av samfunnet fører til at stadig mer av kriminalitet som rammer befolkningen er digital, både ved at den digitale hverdagen rammes direkte eller ved at kriminelle benytter digitale hjelpemidler for å begå tradisjonelle former for kriminalitet. Det kreves av politiet å møte rustet til den nye hverdagen (Barnholt et al., 2021, s. 25). PrevBOT er et godt forslag til et tiltak som møter denne digitaliseringen og utfordringene som følge av den.

Tanken ved å ta i bruk automatisering og moderne teknologi gjør at politiet mer effektivt og presist kan identifisere steder på internett hvor det er stor risiko for at barn blir utsatt for nettovergrep (Sunde & Sunde, 2021, s. 2). Internett er stort og blir stadig større for hver dag. Teknologibedriften SkillCore (u.å.) skriver på sine hjemmesider at internett har doblet seg hvert år siden 2012, og i 2019 eksisterte det angivelig 5,85 milliarder nettsider. Norsk politi har ikke kapasitet eller ressurser til å sitte manuelt på alle forum som eksisterer. Det å bruke en automatisert robot, som PrevBOT, gjør at man kan sette ut flere slike samtidig og arbeide mer effektivt mot problemet (Sunde, 2019, s. 179).

PrevBOT kan observere og samhandle i flere chatter og forum samtidig (Sunde & Sunde, 2021, s. 9). En robot som PrevBOT blir ikke påvirket av lavt blodsukker, en dårlig dag eller lite søvn og kan jobbe døgnet rundt (Datatilsynet, 2018, s. 12). PrevBOT vil fungere som en bruker og krever ikke installasjon eller inn-programmering av de som driver nettsiden (Sunde & Sunde, 2021, s. 9). I tillegg kan politiet bestemme hvilke forum og chatter PrevBOT skal overvåke og samhandle i. Denne prosessen kan også skje automatisk ved at PrevBOT gjør selvstendige vurderinger. Videre trenger ikke PrevBOT å snakke direkte med en person og utgi seg for å være et barn, for å identifisere problematiske steder og personer (s.3). PrevBOT kan også gjøre dette ved å kun observere samtalene. Dersom politiet gjør en vurdering på en person der de selv anser at disse er mulige PP og PS, vil de manuelt kunne flagge personer og forum (s.8).

Dette gjør at man kan kombinere menneskelig forståelse med PrevBOTs beregninger og prediksjoner, uten at det hindrer systemets forutsetninger til å gjøre en god jobb.

I forbindelse med sin masterstudie fikk Bendiksen (2019) tilgang til ekte samtaler mellom barn og nettovergripere. Dette prosjektet, som er en del av AiBA, ga 89% suksessrate ved å finne riktig alder og kjønn på personene i studien. De 10 ukene Sweetie 2.0 ble testet på internett, identifiserte forskere tusen potensielle overgripere fra 71 land (Sunde, 2019, s. 179). Denne statistikken viser effektiviteten ved automatisering og KI, svært godt.

Informasjonen PrevBOT beregner og predikerer, gir politiet handlingsrom og ulike muligheter for hva de skal gjøre med informasjonen de får (Sunde & Sunde, 2021, s. 9). Det nevnes noen eksempler som å varsle barnet om oppdagelsen, varsle overgriperen om at det han eller hun driver med er ulovlig, eller oppfordre overgriperen til å søke hjelp. Eventuelt kan politiet hoppe rett i etterforskningssporet. PrevBOT gir politiet mulighet til å forebygge nettovergrep på et tidlig stadium, som før ikke har vært mulig på samme måte (s.4). Erfaring, fra blant annet Sweetie 2.0, viser at det tar kort tid fra en mindreårig går inn i et chatterom, til han eller hun får seksuelt motiverte henvendelser (Sunde, 2019, s. 178). Desto tidligere politiet får forebygget nettovergrep, jo mer skånes barnet for usunne faktorer som kan hemme og skade deres utvikling (Aanerød & Mossige, 2018, s.13).

Ved at denne jobben blir automatisert, forhindres muligheten for at mennesker gjør uriktige vurderinger. Som nevnt vil overgriperne ta i bruk flere digitale hjelpemidler for å forholde seg anonyme, noe som gjør det vanskelig for ordinære mennesker å finne ut særlig mye om personen. PrevBOTs prediksjoner kan gi politiet informasjon som er svært vanskelig for mennesker å oppdage eller forstå (Sunde & Sunde, 2021, s. 2). Det å avgjøre om det er seksuelt språk i en chat eller på et forum kan fint detekteres av mennesker. Derimot kreves det gjerne datametoder for å forutsi alder og kjønn basert på «authorship analysis». Dersom politiet i tillegg skulle lest alle chattene selv, ville det tatt både kapasitet og ressurser politiet ikke har. Dette ville gått utover andre arbeidsoppgaver (s.6). Dette viser at KI er en god mulighet for å løse denne kriminalitetsutformingen med tanke på ressurser og effektivitet.

AiBA kan peke ut meldinger av seksuell karakter i etterkant av en samtale, som et etterforskningshjelpemiddel (Aakervik, 2020). Dette forenkler etterforskningen slik at en betjent slipper å gjennomgå og lese alle meldinger i en sak. Dette er noe PrevBOT også vil

kunne gjøre. Som vist i figur 2 på side 15, er det viktig å påpeke at all informasjon som ikke ansees relevant for politiet å vite vil bli slettet. Dette er en fortløpende vurdering og er en del av beslutningsprosessen til PrevBOT. På den måten opprettholdes integriteten, respekten og personvernet til menneskene på plattformen, som ikke gjøre noe straffbart (Datatilsynet, 2018, s. 15).

4.4 Ulemper ved at PrevBOT baseres på KI

For å trene opp maskinlæringsalgoritmen til PrevBOT trengs det treningsdata, eller såkalte datasett. Kvaliteten og mengden data vi har tilgjengelig når vi skal trene opp algoritmen, påvirker læringsmodellen. En modell utviklet med virkelige samtaler, vil være annerledes enn en modell utviklet på datasett som etterligner virkelige aktiviteter (Sunde & Sunde, 2021, s. 12). Bruken av chatlogger ved utviklingen av PrevBOT vil utfordre flere av de grunnleggende prinsippene om behandling av personopplysninger[10] (Datatilsynet, 2018, s. 14). Samtaler fra norske nettovergrepssaker vil inneholde personopplysninger. Disse samtalene vil derfor normalt sett ikke leveres ut, og per nå vil det si at PrevBOT ville fått for lite relevante treningsdata mot utfordringen den skal jobbe med. Dette vil påvirke PrevBOTs evne til å utføre sine tiltenkte oppgaver.

Selv om en kunne gått til andre land for å få relevant treningsdata fra faktiske samtaler mellom overgriper og fornærmet, vil språket bli en hindring. Språket som blir brukt i treningsdataene spiller en stor rolle dersom modellen skal brukes i en tekstbasert tilnærming (Sboev et al., 2016, s. 140). For at PrevBOT skal være så effektiv som mulig, bør modellen trenes på tekst på språket til nasjonen som planlegger å bruke den (Sunde & Sunde, 2021, s. 9). Det vil si at dersom PrevBOT skal operere i Norge, kreves det treningssett av samtaler som er skrevet på norsk. Dette fordi uttrykksmåte ikke er overførbart på tvers av språk, som snevrer inn mulighetene for å skaffe relevant og kvalitetsmessig treningsdata til utviklingen av PrevBOT (s.10). Dette er en direkte konsekvens av at PrevBOT nytter seg av KI. En annen slagside er at det må lages en egen PrevBOT for hvert land eller språk, som skal benytte seg av konseptet (s. 12). Selv om man ikke nødvendigvis må starte helt på nytt, ville det vært enklere å forholde seg til en universell robot på tvers av landegrenser og språk.

[10] Personopplysninger er alle opplysninger som kan knyttes til en enkeltperson. Opplysningene kan være direkte, ved for eksempel navn eller fødselsnummer, eller indirekte, ved at personen kan identifiseres på bakgrunn av en kombinasjon av ett eller flere elementer som er spesifikke for personens identitet (Datatilsynet, 2018, s. 14)

AiBA bruker samtaler mellom barn og unge til å trene opp algoritmen til å kunne forutsi alder og kjønn bedre (NTNU, u.å.). Personer som ønsker å delta i utviklingen av dette programmet, logger seg på til faste tidspunkter for å snakke med andre på nett. Deltagerne oppgir kjønn og alder før de begynner å snakke. PrevBOT vil kunne bruke en slik tilnærming, men dette gir ingen kvalitetsmessige gode treningsdata. Allikevel får PrevBOT øvd seg på å predikere alder og kjønn. Generelt krever datasystemer mye mer data enn mennesker, for å lære det samme. Dette er en begrensning ved maskinlære som kompenseres med ved å benytte betydelige datamengder (Datatilsynet, 2018, s. 8). Datatilsynet nevner videre at det hjelper lite med mye data, dersom det ikke representerer hele spekteret av hva modellen senere skal jobbe med.

De ulike forumene PrevBOT fremtidig skal operere i kan være forskjellige, noe som krever at PrevBOT får nødvendige endringer og justeringer for ulike plattformer. Dette gjør at det kreves domeneekspertise, altså kunnskap om hvordan ulike domener[11] fungerer (Sunde & Sunde, 2021, s. 10). På denne måten vil man forsikre seg om at de funksjonene som er relevante for domenet blir tatt i betraktning, så PrevBOT klarer å fungere optimalt. I 2020 passerte Norge 800 000 domenenavn (NORID, 2020). Dette vanskeliggjør politiets mulighet til å skaffe seg kunnskap og oversikt nok, for at PrevBOT skal være effektiv på de fleste domenene.

PrevBOT er bygd rundt det å sette overgriperne i bås utfra hvilken type overgriper de er, og her er det klare skillelinjer. Generelt håndterer datasystemer dårlig at en person flyter mellom ulike kategorier (Kommunal- og moderniseringsdepartementet, 2020, s. 10). Derfor kreves eksperthjelp i utviklingen av PrevBOT, slik at overgripere som opptrer flytende mellom de ulike kategoriene, ikke faller utenfor. Personene som vil utarbeide og utvikle PrevBOT, har mye makt over hvordan systemet vil fungere og utvikle seg. Algoritmer og modeller er imidlertidig ikke mer objektive enn menneskene som lager dem eller personopplysningene som benyttes til dette (Datatilsynet, 2018, s. 15). Dette kan føre til systematiske skjevheter i PrevBOTs programmering, som igjen vil påvirke resultatet. Datatilsynet skriver videre at dette gir mulighet for misbruk og kan ha et stort skadepotensial om det ikke gjøres rett. PrevBOT er nødt til å lære seg hvilke opplysninger den kan legge vekt på og ikke (s.16). Derfor kreves det en tilstrekkelig testingsperiode før roboten tas i bruk.

[11] Domene er et administrativt delområde i et nettverk. På internett brukes domene om en organisasjons unike navn på internett, for eksempel .no eller .com (Rossen, 2022).

Analysen PrevBOT kommer med er basert på sannsynlighet gitt med en viss risiko (Sunde & Sunde, 2021, s. 12). Det at PrevBOT nytter seg av sannsynlighet, vil gjøre det vanskelig å identifisere suksessraten til beregningene dens. Desto mer trening PrevBOT får, jo mer nøyaktig vil sannsynlighetsutsagnet bli. Derimot vil det aldri være en garanti for at PrevBOT sine beregninger stemmer. Dette fordi sannsynlighet representerer eller uttrykker usikkerhet (Terje Aven, 2021). Dette gjør at beregningene til PrevBOT ikke kan brukes som bevis i en straffesak, men heller som et etterforskningsskritt nettopp på bakgrunn av usikkerhetsgraden (Sunde & Sunde, 2021, s. 8).

5 Oppsummering

Jeg har i denne oppgaven tatt et dypdykk i hvordan roboten PrevBOT vil baseres på KI, i arbeidet mot nettovergrep mot barn. Vi har sett at PrevBOT hovedsakelig gjør to oppgaver: kategoriserer og identifiserer PP og PS. Hovedvekten i oppgaven har ligget på de tekniske komponentene som gjør at PrevBOT regnes som KI; herunder robotikk, «authorship analysis» og maskinlæringsalgoritme. Maskinlæringsalgoritmen gjør det mulig å automatisere bruken av «authorship analysis». PrevBOT utnytter det at mennesker bruker språket forskjellig, og at disse språklige uttrykksmåtene kan skilles fra hverandre med høy grad av sikkerhet. Ved hjelp av robotikk kan PrevBOT, uten menneskelig innblanding, gjøre selvstendige vurderinger av PS og PP.

Fordeler ved at PrevBOT vil baseres på KI er blant annet effektivitet og automatisering, som minimerer mulighetene for menneskelige feil. I tillegg kan PrevBOT også fungere som et etterforskningshjelpemiddel. Ulempene ved at PrevBOT vil baseres på KI er blant annet at det kreves treningsdata fra faktiske overgrepssaker i utviklingsfasen av PrevBOT. Dette kolliderer med de strenge personvernreglene, fordi disse samtalene inneholder personopplysninger om både overgripere og fornærmede. I tillegg kan det at mennesker utvikler systemer som PrevBOT føre til skjevhet i konseptet, fordi programmene ikke er mer objektive enn menneskene som lager dem. Det er viktig å være bevisst både fordelene og ulempene ved at PrevBOT vil baseres på KI.

6 Litteraturliste

Aakervik, A.-L. (2020). *Algoritmer kan hindre overgrep på nett*. Gemini. https://gemini.no/2020/12/algoritmer-kan-hindre-overgrep-pa-nett/

Aiken, M. (2017). *The cyber effect : a pioneering cyberpsychologist explains how human behaviour changes online*. John Murray.

Aanerød, L. M. T. & Mossige, S. (2018). *Nettovergrep mot barn i Norge 2015–2017* (NOVA Rapport 10/2018). Norsk institutt for forskning om oppvekst, velferd og aldring. https://oda.oslomet.no/oda-xmlui/bitstream/handle/20.500.12199/5127/Nettutg-NOVA-Rapport-Nettovergrep-10-2018.pdf?sequence=1&isAllowed=y

Barne- og familiedepartementet. (2003). *FNs konvensjon om barnets rettigheter*. Regjeringen. https://www.regjeringen.no/globalassets/upload/kilde/bfd/bro/2004/0004/ddd/pdfv/178931-fns_barnekonvensjon.pdf

Barnholt, K., Bjørn, K., Greve, M., Ousdal, S., Paulsen, J. E., Rjaanes, M., Strand, M. & Thorsberg, L. (2021). *Teknologiutviklingens betydning for politiet, PST og Den høyere påtalemyndighet* (21/02532). Forsvarets Forskningsinstitutt. https://publications.ffi.no/nb/item/asset/dspace:7255/21-02532.pdf

Bendiksen, J. (2019). *Automated detection of perpetrators in grooming conversations in Norwegian* [Masteroppgave, NTNU].

Bjerknes, O. T., Fahsing, I. A. & Bergum, U. (2018). *Etterforskning : prinsipper, metoder og praksis*. Fagbokforl.

Cambridge Dictionary. (u.å.-a). *Forum*. Hentet 31.mars 2022 fra https://dictionary.cambridge.org/dictionary/english/forum

Cambridge Dictionary. (u.å.-b). *Robot*. Hentet 18.april 2022 fra https://dictionary.cambridge.org/dictionary/english/robot

Dalland, O. (2017). *Metode og oppgaveskriving* (6. utg. utg.). Gyldendal akademisk.

Datatilsynet. (2018). *Kunstig intelligens og personvern* (Rapport). https://www.datatilsynet.no/globalassets/global/dokumenter-pdfer-skjema-ol/rettigheter-og-plikter/rapporter/rapport-om-ki-og-personvern.pdf

ECPAT. (u.å.). *Hvem er vi?* Hentet 22.april 2022 fra https://ecpatnorge.no/om-ecpat

Eirik Teigestad. (2017, 22.november 2017). *Hva er grooming?* Overgrep.no. Hentet 19.april 2022 fra https://www.overgrep.no/hva-er-grooming/

Elements of AI. (u.å.). *Hvordan skal vi definere kunstig intelligens (KI)?* https://course.elementsofai.com/no/1/1

Furberg, K. (2019). *NTNU-teknologi kan avsløre overgripere på nett*. Universitetsavisa. Hentet 17.mars 2020 fra https://www.universitetsavisa.no/nyheter/ntnu-teknologi-kan-avslore-overgripere-pa-nett/116629

Hellevik, O. (2002). *Forskningsmetode i sosiologi og statsvitenskap* (7. utg. utg.). Universitetsforl.

Johannessen, A., Christoffersen, L. & Tufte, P. A. (2016). *Introduksjon til samfunnsvitenskapelig metode* (5. utg. utg.). Abstrakt.

JOU 2020:5. (2020). *Kunstig intelligens: Muligheter og risikoer i velferdsforvaltningen*. Arbeids- og velferdsdirektoratet ved Seksjon for informasjonsforvaltning. https://www.uio.no/studier/emner/jus/jus/JUS5502/JOU-jus5502/jou/2020-5-kunstig-intelligens.pdf

Juola, P. (2007). Future trends in authorship attrubution. I P. C. S. Shenoi (Red.), *Advances in digital forensics III* (s. 119-132). Springer. https://doi.org/https://doi.org/10.1007/978-0-387-73742-3_8

Justis- og beredskapsdepartementet. (2021). *Forebygging og bekjempelse av internettrelaterte overgrep mot barn*. Regjeringen. https://www.regjeringen.no/contentassets/2915ff68eb2849edb3218055be32d8cb/strategi-mot-internettrelaterte-overgrep-mot-barn_uu.pdf

Kommunal- og moderniseringsdepartementet. (2020). *Nasjonal strategi for kunstig intelligens*. Regjeringen. https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf

Kripos. (2019). *Seksuell utnyttelse av barn og unge på internett*. Politiet. https://www.politiet.no/globalassets/04-aktuelt-tall-og-fakta/seksuelle-overgrep-mot-barn/seksuell-utnyttelse-av-barn-over-internett.pdf

Moe, M. J. (2019). *Hva er en algoritme?* NDLA. Hentet 28.mars 2022 fra https://ndla.no/nb/subject:1:f7d7f164-fb40-4d21-9813-6a171603281d/topic:2:172361/topic:2:190388/resource:f24cda0a-4548-48e1-a543-2b92e969d92f

NORID. (2020, 10.desember 2020). *Nøkkeltall om domenenavn*. Hentet 22.april 2022 fra https://www.norid.no/no/om-domenenavn/nokkeltall/

NTNU. (u.å.). *Hjelp oss å stoppe overgripere på nett*. Hentet 17.mars 2022 fra https://www.ntnu.no/aiba

Nätt, T. H. (2021, 13.september 2021). *Data*. Store Norske Leksikon. Hentet 31.mars 2022 fra https://snl.no/data

Politiet. (u.å.). *Seksuell utnyttelse av barn på internett*. Hentet 15.mars 2022 fra https://www.politiet.no/tjenester/tips-politiet/seksuell-utnyttelse-av-barn-pa-internett/

Politiloven. (1995). *Lov om politiet* (LOV-1995-08-04-53). Lovdata. https://lovdata.no/dokument/NL/lov/1995-08-04-53

PwC. (u.å.). *Hva er kunstig intelligens?* https://www.pwc.no/no/teknologi-omstilling/digitalisering-pa-1-2-3/kunstig-intelligens.html

Riksadvokaten. (2022). *Mål og prioriteringer for straffesaksbehandlingen i 2022* (1/2022) [Rundskriv]. Den Høyere Påtalemyndighet. https://www.riksadvokaten.no/document/riksadvokatens-mal-og-prioriteringer-for-2022/

Rossen, E. (2022, 8.februar 2022). *Domene (IT)*. Store Norske Leksikon. Hentet 20.april 2022 fra https://snl.no/domene_-_IT

Sboev, A., Litvinova, T., Gudovskikh, D., Rybka, R. & Moloshnikov, I. (2016). Machine learning models of text categorization by author gender using topic-independent features. *Procedia Computer Science*, *101*, 135–142. https://doi.org/https://doi.org/10.1016/j.procs.2016.11.017

SkillCore. (u.å.). *Hvor stort er internettet?* https://skillcore.no/hvor-stort-er-internettet/

Sunde, I. M. (2016). *Datakriminalitet : en fremstilling av strafferettslige regler om datakriminalitet*. Fagbokforl.

Sunde, I. M. (2019). Sweetie, et politibarn eller en politistyrke på internett. I *Det digitale er et hurtigtog!* (s. s.177-205).

Sunde, N. & Sunde, I. M. (2021). Conceptualizing an AI-based Police Robot for Preventing Online Child Sexual Exploitation and Abuse: Part I – The Theoretical and Technical Foundations for PrevBOT. *Nordic Journal of Studies in Policing*, *8*(2), 1-21. https://doi.org/10.18261/issn.2703-7045-2021-02-01

Søraa, R. A. (2022, 22.mars 2022). *Robot*. Store Norske Leksikon. Hentet 31.mars 2022 fra https://snl.no/robot

Terje Aven. (2021). *Sannsynlighet*. Store Norske Leksikon. Hentet 29.mars 2022 fra https://snl.no/sannsynlighet

Terre Des Hommes. (u.å.). *Sweetie, our weapon against child webcam sex*. Hentet 17.mars 2022 fra https://www.terredeshommes.nl/en/programs/sweetie

Thurén, T., Gjerpe, K. & Gjestland, D. (2009). *Vitenskapsteori for nybegynnere* (2. utg. utg.). Gyldendal akademisk.

Tidemann, A. (2021, 5.juli). *Dyplæring*. Store Norske Leksikon. Hentet 23.mars 2022 fra https://snl.no/dyplæring

6.1 Selvvalgt litteratur

Aiken, M. (2017). *The cyber effect : a pioneering cyberpsychologist explains how human behaviour changes online*. John Murray. (3s)

Barnholt, K., Bjørn, K., Greve, M., Ousdal, S., Paulsen, J. E., Rjaanes, M., Strand, M. & Thorsberg, L. (2021). *Teknologiutviklingens betydning for politiet, PST og Den høyere påtalemyndighet* (21/02532). Forsvarets Forskningsinstitutt. https://publications.ffi.no/nb/item/asset/dspace:7255/21-02532.pdf (42s)

Bendiksen, J. (2019). *Automated detection of perpetrators in grooming conversations in Norwegian* [Masteroppgave, NTNU]. (25s)

Datatilsynet. (2018). *Kunstig intelligens og personvern* (Rapport). https://www.datatilsynet.no/globalassets/global/dokumenter-pdfer-skjema-ol/rettigheter-og-plikter/rapporter/rapport-om-ki-og-personvern.pdf (28s)

JOU 2020:5. (2020). *Kunstig intelligens: Muligheter og risikoer i velferdsforvaltningen*. Arbeids- og velferdsdirektoratet ved Seksjon for informasjonsforvaltning. https://www.uio.no/studier/emner/jus/jus/JUS5502/JOU-jus5502/jou/2020-5-kunstig-intelligens.pdf (13s)

Juola, P. (2007). Future trends in authorship attrubution. I P. C. S. Shenoi (Red.), *Advances in digital forensics III* (s. 119-132). Springer. https://doi.org/https://doi.org/10.1007/978-0-387-73742-3_8 (13s)

Justis- og beredskapsdepartementet. (2021). *Forebygging og bekjempelse av internettrelaterte overgrep mot barn*. Regjeringen. https://www.regjeringen.no/contentassets/2915ff68eb2849edb3218055be32d8cb/strategi-mot-internettrelaterte-overgrep-mot-barn_uu.pdf (80s)

Kommunal- og moderniseringsdepartementet. (2020). *Nasjonal strategi for kunstig intelligens*. Regjeringen. https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf (28s)

Kripos. (2019). *Seksuell utnyttelse av barn og unge på internett*. Politiet. https://www.politiet.no/globalassets/04-aktuelt-tall-og-fakta/seksuelle-overgrep-mot-barn/seksuell-utnyttelse-av-barn-over-internett.pdf (75s)

Sboev, A., Litvinova, T., Gudovskikh, D., Rybka, R. & Moloshnikov, I. (2016). Machine learning models of text categorization by author gender using topic-independent features. *Procedia Computer Science*, *101*, 135–142. https://doi.org/https://doi.org/10.1016/j.procs.2016.11.017 (7s)

Sunde, N. & Sunde, I. M. (2021). Conceptualizing an AI-based Police Robot for Preventing Online Child Sexual Exploitation and Abuse: Part I – The Theoretical and Technical Foundations for PrevBOT. *Nordic Journal of Studies in Policing*, *8*(2), 1-21. https://doi.org/10.18261/issn.2703-7045-2021-02-01 (15s)

Aanerød, L. M. T. & Mossige, S. (2018). *Nettovergrep mot barn i Norge 2015–2017* (NOVA Rapport 10/2018). Norsk institutt for forskning om oppvekst, velferd og aldring. https://oda.oslomet.no/oda-xmlui/bitstream/handle/20.500.12199/5127/Nettutg-NOVA-Rapport-Nettovergrep-10-2018.pdf?sequence=1&isAllowed=y (71s)

6.2 Figurliste