

BÆREKRAFTIG MASKINLÆRING – ETISKE PERSPEKTIVER

Aksel Braanen Sterri

SINTEF seminar om bærekraftig maskinlæring
21. september 2023

OSLOMET

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Embedded
EthiCS @Harvard

 **SENTER FOR
LANGSIKTIG POLITIKK**

MIN BAKGRUNN

- PhD i filosofi fra UiO
- Postdoktor på OsloMet
- Fagsjef i et nystartet Senter for langsiktig politikk
- Jobber med etikken rundt kunstig intelligens
 - Embedded EthiCS og Honors på UiO
 - Ulike former for risiko knyttet til kunstig intelligens

TEKNIKK OG ETIKK



Teknikk



Etikk

STRUKTUR

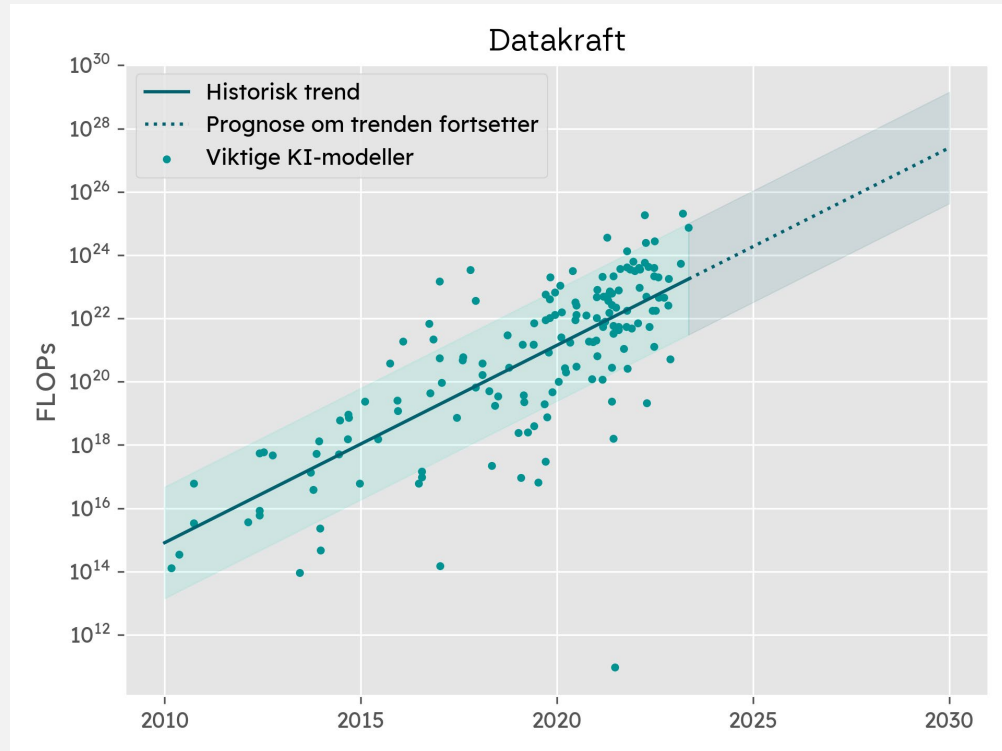
1. Problembeskrivelse, situasjonen og utvikling
2. En test av ulike etiske prinsipper og løsninger
 - a) Klimanøytralitets-løsningen
 - b) SDG-løsningen
 - c) NTP-løsningen
 - d) Karbonskatt-løsningen
3. Konklusjon

RELEVANTE BÆREKRAFTS- HENSYN

1. **Tredjepartsskade**
Klimagassutslipp
2. **Beslag på knappe ressurser** (vann, knappe råmaterialer)
3. **Utslppsreduksjon**
Maskinlæring kan brukes til å redusere utslipp, effektivisere ressursbruk og ressursallokering
4. **Trygg og SDG-forenlig bruk**



SITUASJONEN OG FRAMTIDEN



- I dag bruker vi rundt 0.2% av verdens kraftforbruk på kunstig intelligens.
- Om den eksplorative veksten i datakraft fortsetter vil KI bruke tilsvarende **17 % - 267 % av dagens globale kraftforbruk i 2030.**



MÅLENE FOR FORSKERE, TEKNOLOGER, BEDRIFTER OG SAMFUNN



Bruke ML til å
redusere utslipp



Redusere utslippene
ved ML



Gitt skadene, bruk ML
til noe nyttig.

#1 KLIMANØYTRALITET

- Kan vi lete etter etiske prinsipper som kan veilede arbeidet?
- **Klimanøytralitetsprinsippet:** Maskinlæring er bare etisk forsvarlig dersom den er netto klimanøytral.
 - Noen applikasjoner kan være negative, men totalt må bruken av maskinlæring enten være netto nøytral eller redusere de totale utslippene.
 - Hvorfor?
 - Global oppvarming enorm utfordring.
 - Utslippsnøytralitet anerkjent mål.
 - Implisitt i en del kritikk.

PROBLEMER MED KLIMANØYTRALITET



Individuerinsproblemet

At noe bruk av KI reduserer utslipp rettferdiggjør ikke utslipp fra samfunnsunyttig bruk av KI.



Målmangfoldsproblemet

Hvis KI kan ”kurere kreft” og det koster utslipp er vel det verdt det?



Alternativproblemet

Er utslippene større enn det plausible alternativet?

#2 SDG- LØSNINGEN



- Bruk av maskinlæring er bare etisk forsvarlig dersom den bidrar til å fremme bærekraftsmålene.
- Løser:
 - Individueringsproblemet
 - Målmangfoldproblemet
 - Alternativproblemet

ANDRE PROBLEMER MED SDG-LØSNINGEN

- **Tilstrekkelighetsproblemet:** Å fremme mange bærekraftsmål i større grad er bedre enn å fremme ett bare litt. Men prinsippet skiller ikke mellom dem.
- **Prioriteringsproblemet:** Ingen systematisk måte å vekte de ulike målene mot hverandre ved konflikt: Hvor mye klimagassutslipp skal vi tåle for en gitt mengde fattigdomsreduksjon?
- **Incentivproblemet:** Hvordan sikre at bedrifter i et kapitalistisk system prioriterer bærekraft når det kommer i konflikt med bunnlinjen?
- **Byrdefordelingsproblemet:** Hvordan sikrer vi en rettferdig fordeling av byrdene for å oppnå bærekraftsmålene? Hvis det er frivillig, så blir flere byrder lagt på selskaper som lever opp til sine etiske forpliktelser.

#3 NTP-LØSNINGEN

- NTP-prinsippet etter Nasjonal Transportplan.
 - Gitt at maskinlæring har høye utslipp må vi prioritere de viktigste prosjektene (de med størst samfunnsnytte) og stoppe de prosjektene med lav nytte.
- Har en annen fordel: Det gir oss muligheten til å forhindre prosjekter med betydelig risiko for alvorlige negative samfunnskONSEKVENSER.



NTP LØSER NOEN PROBLEM

- **Tilstrekkelighetsproblemet:** Løser tilstrekkelighetsproblemet siden de prosjektene med *mest* samfunnsnytte blir prioritert.
- **Prioriteringsproblemet:** En overordnet vurdering av samfunnsnytte gjør det mulig å prioritere mellom ulike mål.
- **Incentivproblemet:** Bedrifter vil ha et insentiv til å prioritere samfunnsnyttige prosjekter.
- **Byrdefordelingsproblemet:** De selskapene som bidrar med mest samfunnsnytte blir belønnet. Det kan sies å være rettferdig.

(NYE) PROBLEMER MED NTP-LØSNINGEN

- **Verdiproblemet:**
 - Hvordan definerer vi samfunnsnytte?
- **Informasjonsproblemet:**
 - Hvordan greier byråkrater å vurdere samfunnsnyten til maskinlærings-applikasjoner?
- **Byråkratiproblemet:**
 - Tid og rapporteringskrav er en skatt på teknologisk utvikling.
 - Forsinker utviklingen, noe som kan ha en høy alternativkostnad.
- **Regulatory capture:**
 - Bedrifter vil bruke ressurser på å påvirke den politisk-byråkratiske behandlingen.

KARBONSKATT-LØSNINGEN

- Vi ønsker å gi prioritet til samfunnsnyttige lavutslippsprosjekter og at bedrifter skal forsøke å gjøre sine prosjekter mer nyttige og lavere utslipp.
- En løsning: En skatt på utslipp (karbonskatt) og ellers la markedet avgjøre hvilke prosjekter som er nyttige.

KARBONSKATT GIR SVAR PÅ

- **Verdiproblemet:**
 - Samfunnsnytte defineres som betalingsvillighet, hvor fremtidige generasjoners estimerte betalingsvilje er inkludert.
- **Prioriteringsproblemet**
 - Betalingsvillighet er en felles måleenhet for alle mål. De prosjektene det er størst etterspørsel etter vinner. Det løser prioriteringsproblemet.
- **Incentivproblemet**
 - Skatt på utslipp sikrer at bedrifter i et kapitalistisk system prioriterer utslippsreduksjon fordi det lønner seg på bunnlinjen.
 - En skatt skaper også etterspørsel etter utslippsreducerende tiltak.

KARBONSKATT GIR SVAR PÅ FLERE PROBLEMER

- **Byrdefordelingsproblemet**
 - Forurensere betaler.
 - Den som skaper samfunnsnyttige produkter vinner markedskonkurransen.
- **Byråkratiproblemet:**
 - Byråkratenes jobb blir å finne ut riktig nivå på karbonskatten, ikke å vurdere hver enkelt prosjekts samfunnsnytte. Og det er enklere.
- **Informasjonsproblemet:**
 - Informasjonsproblemet løses av prismekanismen.

KARBONSKATTENS PROBLEMER

- **Måleproblemet:** Hvordan måle utslipp som følge av maskinlæring?
 - Skatten bør legges på selskap og selskaper kan få frihet til å flytte kostanden til kundene slik de mener er best.
 - Selskaper betaler for datakraft via datasenter som måler energibruk. Dette kan brukes til å gi anslag på utslipp per relevant enhet (her kan ML hjelpe)
- **Rapporteringsproblemet:** Hvordan sikre troverdig rapportering av utslipp?
 - Myndighetene kan kreve at datasentre rapporterer til dem og dette kan gå noenlunde automatisk.

BETALINGSVILLIGHET ER IKKE SAMFUNNSNYTTE

- **Verdiproblemet:** Betalingsvillighet er en imperfekt proxy for samfunnsnytte.
 - Selv med karbonskatt er det en rekke tredjepartsvirkninger som ikke inkluderes. Tredjeparts betalingsvillighet er derfor ikke tilstrekkelig dekket.
 - Rike får flere «stemmesedler» enn fattige, men deres velferd bør telle like mye.
 - Dyrs velferd teller ikke
 - Naturen teller ikke
- **Tidshorisontsproblemet:** Bedrifter har en mye høyere diskonteringsrate enn samfunnet. Karbonskatt er utilstrekkelig for å sikre at langsiktige initiativ blir igangsatt.

LØSNINGER

- Supplementere karbonskatt ved nye finansieringsmetoder
 - Myndighetene bør finansiere prosjekter med høy samfunnsnytte.
 - Grønne subsidier.
 - Innovasjonspriser
 - Advanced Market Commitments (som for vaksiner)

KONKLUSJON

- Karbonskatt bør være basis.
- Suppleres av grønne subsidier
- Subsidier av utvikling av KI-modeller som kan brukes til å redusere utslipp
- Det er uavhengige grunner til å kontrollere de kraftigste (og minst grønne) KI-modellene. Her kunne NTP-modellen kanskje vært nyttig.
- Massiv utjevning, og i fravær av det, statlig finansiering for å sikre at ML utvikles til gode for alle.
- Utvikle instrumenter for å inkludere dyrs velferd og naturens verdi i samfunnsnytte-kalkylen