

Kommunal- og moderniseringsdepartementet

Postboks 8112 Dep
0032 OSLO

Offl. § 15.2

Deres ref:

Vår ref: 2021/54582-2

Arkivkode: 053

Dato: 23.06.2021

Innspill til norsk posisjon til EU-kommisjonens forslag til regulering av kunstig intelligens

Bufdir har gjennomgått EU-kommisjonens forslag til regulering av kunstig intelligens (harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts). Direktoratet ønsker å komme med noen generelle merknader til direktivet, som kan danne grunnlag for en norsk posisjon.

Om Bufdirs rolle som fagdirektorat for likestilling

Bufdir ønsker å gi et innspill i denne saken for å ivareta Bufdirs rolle som fagdirektorat for likestilling og universell utforming. Bufdir har et sektorovergripende ansvar for likestilling. Det vil si at direktoratet bistår, og er pådriver overfor andre sektorer med kunnskap om likestilling. Vi analyserer utviklingen og vi gjennomfører og samordner handlingsplaner. Vi bidrar også til at likestillingsfeltet er godt koordinert, i vårt samarbeid med andre myndigheter.

Direktoratet skal jobbe for likestilling og ikke-diskriminering knyttet til kjønn, graviditet, permisjon ved fødsel eller adopsjon, omsorgsoppgaver, etnisitet, religion, livssyn, funksjonsnedsettelse, seksuell orientering, kjønnsidentitet, kjønnsuttrykk og alder.

Om Bufdirs arbeid med diskriminering og kunstig intelligens

Både EU-kommisjonens utkast til regulering av kunstig intelligens og Europarådet "Feasibility study on a legal framework on AI design, development and application based on Council of Europe's standards" fremhever at innføring og bruk av kunstig intelligens hos private aktører og statlige myndigheter medfører risiko for diskriminering. Bufdirs gjennomgang av kunnskap om diskriminering og kunstig intelligens viser at bruk av kunstig intelligens kan medføre ulike former for forskjellsbehandling og diskriminering.

Bruk av KI som beslutningsstøtte / forutsigelsesprodukter¹ for offentlig myndighetsutøvelse har i enkelte tilfeller ført til diskriminering². Amerikanske sykehus brukte KI for å forutsi hvilke pasienter som hadde

¹ Terminologi hentet fra Overvåkningskapitalismens tidsalder, Shoshana Zuboff

² <https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf>

behov for ekstra medisinsk oppfølging. Forskning på et utvalg av over 200 millioner pasienter viste at denne bruken av KI systematisk favoriserte «hvite» pasienter over «fargede».³ Denne formen for diskriminering omtales ofte som «harms of allocation».

Diskriminering kan også oppstå der en kunstig intelligens er trent på og dermed tilpasset mangfoldet i befolkningen. Et kjent eksempel er stemmegjenkjenningsprogrammer som forstår menn bedre enn kvinner.⁴ Dette omtales ofte som «harms of representation».

I vurderingen av forbud mot bruk av kunstig intelligens fremhever EU-kommisjonen at barn og personer med funksjonsnedsettelse er særlig sårbare for kunstig intelligens som bruker manipulerende teknikker.⁵ I sin «feasibility study on AI»⁶ forklarer Europarådet at slik manipulering kan være i strid med retten til privatliv og psykologisk og moralsk integritet (EMK art. 8). For å sikre likestilt beskyttelse av menneskerettighetene er det vesentlig at utsatte grupper beskyttes mot manipulerende og skadelig bruk av kunstig intelligens.

Diskriminering ved bruk av kunstig intelligens i forutsigelsesprodukter kan skje ved at personer som søker om offentlige ytelser eller tjenester, må utlevere sensitive og private opplysninger i større grad enn befolkningen ellers, og i større grad enn det som er rimelig. Omfattende innhenting av data kan være nødvendig for at kunstig intelligens skal gi riktige forutsigelser. Å invitere det offentlige inn i sin private sfære kan imidlertid oppleves som invaderende.

Diskriminering kan også oppstå ved at KI reproducerer stereotypier. Et eksempel er oversettelser fra Google translate som oversetter «han er sykepleier» til «hun er sykepleier» som følge av at historiske data om kjønn og yrke er matet inn i algoritmen.⁷

Bufdirs kommentarer til direktivet – innspill til norsk posisjon

EU-kommisjonen fremhever av forslaget til regulering skal sikre at bruk av kunstig intelligens i EU er i samsvar med EU charter on fundamental rights og art. 21 om ikke-diskriminering. Bufdir vil peke på enkelte forhold ved direktivet som direktoratet mener kan inngå i en norsk posisjon til EU-kommisjonens forslag til regulering av kunstig intelligens:

Art. 5 a)

Den foreslåtte art. 5 a) forbyr bruk av kunstig intelligens som

«subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;»

Bufdirs oppgave er blant annet å sikre et likestilt vern av grunnleggende rettigheter. Dette innebærer blant annet å sikre en likestilt beskyttelse mot vold og psykisk skade for barn og diskriminerte grupper. Norske myndigheter skal også sikre likestilt beskyttelse av økonomiske, sosiale og kulturelle rettigheter.

Den foreslåtte artikkelen forbyr kunstig intelligens som manipulerer personers atferd, og som kan påføre personen selv eller andre fysisk eller psykisk skade. Bufdir vil påpeke at kunstig intelligens oftere benyttes til å manipulere enkeltpersoner på en måte som kan forårsake økonomisk skade, snarere enn fysisk og psykisk skade.

³ <https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070>

⁴ <https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases>

⁵ <https://www.denverpost.com/2021/03/28/what-a-gambling-app-knows-about-you/>

⁶ <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>

⁷ <https://algorithmwatch.org/en/google-translate-gender-bias/>

Direktoratet vil foreslå at en norsk posisjon tar opp behovet for at forbudet mot manipulerende teknikker også gjelder for kunstige intelligens som forårsaker eller mest sannsynlig vil forårsake vesentlig økonomisk skade.

Art. 6

EU-kommisjonens forslag definerer enkelte bruk av kunstig intelligens som høyrisiko, jf. art. 6 andre ledd og annex III. Utkastet definerer blant annet bruk av kunstig intelligens i arbeid og utdanning som høyrisiko kunstig intelligens, jf. annex III punkt 3 og 4.

For arbeidslivet kategoriserer kommisjonens utkast bruk av kunstig intelligens både i rekruttering, i medarbeidervurdering og andre forhold i arbeidslivet som høyrisiko. Dette kan støtte opp under det særlig strenge forbudet mot diskriminering i arbeidslivet i diverse EU-direktiver og likestillings- og diskrimineringsloven § 6, jf § 9 andre ledd og § 29.

Kunstig intelligens brukes i flere sammenhenger i utdanning. Forslaget fra EU-kommisjonen definerer bruk av kunstig intelligens ved opptak til studier og utdanning som høyrisiko. Retten til utdanning er en sentral menneskerettighet, og skal gjennomføres på en likestilt måte for hele befolkningen. Bufdir vil påpeke at bruk av kunstig intelligens i den løpende vurderingen av elevene, for å tilrettelegge og tilpasse undervisningen, i læringsplattformer og i læremidler medfører stor risiko for diskriminering av særlig utsatte elevgrupper, samtidig med at slik teknologi kan gi store gevinster.

Bufdir mener derfor at Norge bør jobbe for at all bruk av kunstig intelligens i utdanning og undervisning bør defineres som høyrisiko.

Tittel IX, art 69

EU-kommisjonens utkast foreslår ikke regler for kunstig intelligens som faller utenfor det som forbudt, defineres som høyrisiko eller «visse typer» kunstig intelligens i art. 52. Kommisjonen foreslår imidlertid frivillige «codes of conduct» for kunstig intelligens som faller utenfor reguleringen.

Bufdirs gjennomgang av risikoer for diskriminering ved bruk av kunstig intelligens viser at det er vesentlig diskrimineringsrisiko for bruk som ligger utenfor den bruk som forslaget definerer som forbudt, høyrisiko eller visse typer kunstig intelligens. Bufdir mener norske myndigheter bør spille inn at det er behov for at direktivet i art. 69 synliggjør visse grunnleggende krav som gjelder for all kunstig intelligens som retter seg mot borgerne, også fra private aktører, blant annet forbudet mot diskriminering.

Informasjon og reklame på nett og sosiale plattformer er direkte rettet mot den enkelte borger. Det kan være spesielt vanskelig for den enkelte å identifisere forskjellsbehandling og diskriminering på digitale plattformer, siden man ikke har kunnskap om hva andre mottar av informasjon. Det kan derfor være behov for et generelt og risikobasert tilsyn med kunstig intelligens som brukes i samhandling med borgerne.

Med hilsen

Wenche Mobråten (e.f)
divisjonsdirektør

Maren Anna Lervik
fagdirektør

Dokumentet er godkjent elektronisk og har derfor ikke håndskrevet signatur.

Tenk miljø – velg digital postkasse fra e-Boks eller Digipost på www.norge.no