Europa-Kommissionens

EKSPERTGRUPPE PÅ HØJT NIVEAU OM KUNSTIG INTELLIGENS

UDKAST

ETISKE RETNINGSLINJER FOR PÅLIDELIG KUNSTIG INTELLIGENS

RESUMÉ

Arbejdsdokument til høring af interessenter

Bruxelles, den 18. december 2018

UDKAST TIL ETISKE RETNINGSLINJER

FOR PÅLIDELIG KUNSTIG INTELLIGENS

Ekspertgruppen på Højt Niveau om Kunstig Intelligens
Udkast til etiske retningslinjer for pålidelig kunstig intelligens

Europa-Kommissionen
Generaldirektoratet for Kommunikation

Kontaktperson Nathalie Smuha - Ekspertgruppens koordinator
E-mail CNECT-HLG-AI@ec.europa.eu

Europa-Kommissionen
1049 Bruxelles

Dette dokument blev offentliggjort den 18. december 2018 på engelsk.

Dette arbejdsdokument blev udarbejdet af Ekspertgruppen på Højt Niveau om Kunstig Intelligens, uden at dette berører medlemmernes individuelle holdning til specifikke punkter, og uden at dette berører den endelige udgave af dokumentet. Dette dokument vil fortsat blive videreudviklet, og en endelig udgave af dette dokument vil blive fremlagt i marts 2019 efter høringen af interessenterne via den europæiske alliance vedrørende kunstig intelligens (European AI Alliance).

 Ekspertgruppen på Højt Niveau om Kunstig Intelligens er eneansvarlig for indholdet af dette arbejdsdokument. Selv om medarbejdere i Kommissionens tjenestegrene har medvirket til at udarbejde retningslinjerne, afspejler de synspunkter, der gives udtryk for i dette dokument, ekspertgruppens holdning og kan under ingen omstændigheder tages som udtryk for Europa-Kommissionens officielle holdning. Dette er et udkast til ekspertgruppens første sæt konkrete resultater. Den endelige udgave heraf vil blive forelagt Kommissionen i marts 2019. Den endelige udgave af ekspertgruppens andet sæt konkrete resultater — dens anbefalinger om politikker og investering vedrørende kunstig intelligens ("AI Policy and Investment Recommendations") — vil blive fremlagt i midten af 2019.

Yderligere oplysninger om Ekspertgruppen på Højt Niveau om Kunstig Intelligens kan findes online (https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence).

RESUMÉ

Dette arbejdsdokument er et udkast til de etiske retningslinjer for kunstig intelligens, som er udarbejdet af Kommissionens ekspertgruppe på højt niveau om kunstig intelligens, og som forventes at foreligge i sin endelige udgive i marts 2019.

Kunstig intelligens (KI) er en af vor tids mest transformerende kræfter og vil uden tvivl ændre strukturen for vores samfund. Denne teknologi udgør en stor mulighed for at øge velstanden og væksten, og det skal Europa bestræbe sig på at opnå. I løbet af de sidste ti år blev der gjort store fremskridt som følge af de store mængder tilgængelige digitale data, kraftige computerarkitekturer og fremskridt inden for KI-teknikker såsom maskinindlæring. Udviklingen af kunstig intelligens i selvkørende biler, sundhedspleje, hjemme-/tjenesterobotter, uddannelse og cybersikkerhed forbedrer hver dag vores livskvalitet. Derudover er kunstig intelligens nøglen til at løse mange af verdens store udfordringer, f.eks. global sundhed og trivsel, klimaforandringer, pålidelige juridiske og demokratiske systemer og andre hindringer for at nå FN's mål for bæredygtig udvikling.

Kunstig intelligens har potentialet til at skabe enorme fordele for borgerne og samfundet, men det medfører også visse risici, som bør håndteres korrekt. Da fordelene ved kunstig intelligens generelt overstiger risiciene, må vi sørge for at følge den vej, der **maksimerer fordelene og samtidig minimerer risiciene**. For at sikre, at vi bliver på rette spor, er der behov for en **menneskecentreret tilgang til kunstig intelligens**, som tvinger os til at huske på, at udvikling og anvendelse af kunstig intelligens ikke bør ses som et mål i sig selv, men som et middel til at øge menneskers trivsel. **Pålidelig kunstig intelligens er vores ledestjerne**, idet mennesker kun vil være i stand til trygt og fuldt ud at høste fordelene ved kunstig intelligens, hvis de kan have tillid til teknologien.

Der er **to kriterier** for, at en kunstig intelligens kan være pålidelig: 1) Den skal overholde de grundlæggende rettigheder, den gældende lovgivning og de grundlæggende principper og værdier, hvorved det sikres, at den har et "**etisk formål**". 2) Den skal være **teknisk robust** og pålidelig, da en mangel på teknologisk beherskelse kan forårsage uforsætlig skade, selv med gode hensigter.

Disse retningslinjer opstiller derfor en **ramme for pålidelig kunstig intelligens**:

- **Kapitel I** handler om at sikre etiske formål for kunstig intelligens ved at fastsætte de grundlæggende rettigheder, principper og værdier, som skal overholdes.
- Ud fra disse principper udstikkes der i **kapitel II** retningslinjer for virkeliggørelsen af pålidelig kunstig intelligens, hvor der tages højde for både det etiske formål og den tekniske robusthed. Dette gøres ved at opstille en liste over krav til pålidelig kunstig intelligens og give et overblik over de tekniske og ikketekniske metoder, der kan indgå i dens anvendelse.
- **Kapitel III operationaliserer** derefter kravene ved at opstille en konkret, men ikke udtømmende evalueringsliste for pålidelig kunstig intelligens. Denne liste tilpasses derefter til hvert enkelt anvendelsesformål.

I modsætning til andre dokumenter vedrørende etisk kunstig intelligens har retningslinjerne derfor ikke til formål at fremsætte endnu en liste over centrale værdier og principper for kunstig intelligens, men snarere at give vejledning om den konkrete gennemførelse og operationalisering heraf i KI-systemer. Denne vejledning gives på tre abstraktionsniveauer fra det mest abstrakte i kapitel I (grundlæggende rettigheder, principper og værdier) til det mest konkrete i kapitel III (evalueringsliste).

Retningslinjerne er rettet til alle **relevante interessenter, der udvikler, udbreder eller anvender kunstig intelligens**, herunder virksomheder, organisationer, forskere, offentlige tjenester, institutioner, enkeltpersoner og andre enheder. Den endelige udgave af disse retningslinjer vil indeholde en

mekanisme, der skal gøre det muligt for interessenter frivilligt at tilslutte sig dem.

Det er vigtigt at bemærke, at disse retningslinjer ikke er tænkt som erstatning for nogen form for politik eller regulering (dette behandles i ekspertgruppens andet sæt konkrete resultater: anbefalingerne om politikker og investering, der efter planen skal foreligge i maj 2019), og de har heller ikke til formål at hindre indførelsen af disse. Retningslinjerne bør desuden betragtes som et levende dokument, der med tiden skal ajourføres regelmæssigt for at sikre, at det forbliver relevant, efterhånden som teknologien og vores viden om den udvikler sig. Dette dokument bør derfor danne udgangspunkt for debatten om "**Pålidelig kunstig intelligens fremstillet i Europa**".

Selv om Europa kun kan udbrede sin etiske tilgang til kunstig intelligens, hvis vi er konkurrencedygtige på globalt plan, er en **etisk tilgang til kunstig intelligens vigtig for at opbygge en ansvarlig konkurrenceevne**, da dette vil skabe tillid hos brugerne og fremme bredere anvendelse af kunstig intelligens. Det er ikke meningen, at disse retningslinjer skal kvæle innovationen inden for kunstig intelligens i Europa — formålet er derimod at bruge etik som inspiration til at udvikle en unik slags kunstig intelligens, som har til formål at beskytte og gavne både enkeltpersoner og almenvellet. Dette vil gøre det muligt for Europa at indtage en førende stilling inden for avanceret, sikker og etisk kunstig intelligens. Kun ved at sikre pålidelighed vil de europæiske borgere kunne drage de fulde fordele af kunstig intelligens.

Uden for Europas grænser tager disse retningslinjer også sigte på at **fremme overvejelser og debat** om de etiske rammer for kunstig intelligens på **globalt plan**.

VEJLEDNING

Hvert kapitel i retningslinjerne giver vejledning i at opnå pålidelig kunstig intelligens og er rettet til alle relevante interessenter, der udvikler, udbreder eller anvender kunstig intelligens. En sammenfatning af hvert kapitel findes her:

Kapitel I: Central vejledning til sikring af etiske formål:

- Det skal sikres, at kunstig intelligens er **menneskecentreret**: Kunstig intelligens skal udvikles, udbredes og anvendes med et "**etisk formål**", som bygger på og afspejler de grundlæggende rettigheder, samfundsmæssige værdier og følgende etiske principper: *godgørenhed* (vær god), *ikke-skadevolden* (gør ikke skade), *menneskers autonomi*, *retfærdighed* og *forklarlighed*. Dette er afgørende for arbejdet med at opnå **pålidelig kunstig intelligens**.
- De mulige fremtidige virkninger af kunstig intelligens på mennesker og almenvellet skal vurderes på grundlag af de grundlæggende rettigheder, etiske principper og værdier. Der skal rettes **særlig opmærksomhed** mod situationer, der involverer mere **udsatte grupper** såsom børn, personer med handicap eller minoriteter, eller situationer med **asymmetriske magt- eller informationsforhold**, f.eks. mellem arbejdsgivere og arbejdstagere eller mellem virksomheder og forbrugere.
- Det skal anerkendes og bemærkes, at selv om kunstig intelligens medfører væsentlige fordele for enkeltpersoner og samfundet, kan der også være negative virkninger. Der kræves årvågenhed på områder af kritisk betydning.

Kapitel II: Central vejledning til realisering af pålidelig kunstig intelligens:

- **Kravene til pålidelig kunstig intelligens** skal indarbejdes allerede **fra den tidligste designfase**: ansvarlighed, datastyring, design for alle, styring af kunstig intelligens (kontrol udført af mennesker), ikke-diskrimination, overholdelse af menneskelig autonomi, overholdelse af privatlivets fred, robusthed, sikkerhed og gennemsigtighed.

- Der skal overvejes tekniske og ikketekniske metoder til at sikre opfyldelsen af disse krav i KI-systemet. Desuden skal disse krav holdes for øje ved udformningen af både det hold, der skal arbejde med systemet, selve systemet, testmiljøet og systemets potentielle anvendelsesmuligheder.
- **Interessenterne (kunder, ansatte osv.) skal underrettes** på en klar og proaktiv måde om KI-systemets kapacitet og begrænsninger, så de kan opstille realistiske forventninger. Det er i den forbindelse afgørende at sikre KI-systemets **sporbarhed**.
- Pålidelig kunstig intelligens skal være **en del af organisationens kultur**, og interessenterne skal underrettes om, hvordan princippet om pålidelig kunstig intelligens gennemføres i KI-systemernes udformning og anvendelse. Pålidelig kunstig intelligens kan også indgå i organisationers deontologiske chartre eller adfærdskodekser.
- **Interessenterne skal deltage og inddrages** i udformningen og udviklingen af KI-systemet. Derudover er det vigtigt at sikre **mangfoldighed** i forbindelse med oprettelsen af de hold, der udvikler, gennemfører og tester produktet.
- Der skal gøres en indsats for at **lette kontrol og revision** af KI-systemerne, især under kritiske forhold eller i kritiske situationer. Så vidt muligt skal systemet udformes, således at individuelle beslutninger kan spores til forskellige input: data, forudoplærte modeller osv. Desuden skal der defineres **forklaringsmetoder** for KI-systemet.
- Der skal sikres en særlig proces for **ansvarsstyring**.
- Der skal planlægges **undervisning og uddannelse**, således at ledere, udviklere, brugere og arbejdsgivere kender til og er uddannet i pålidelig kunstig intelligens.
- Der gøres opmærksom på, at der kan være grundlæggende spændinger mellem forskellige mål (gennemsigtighed kan åbne døren for misbrug, og konstatering og korrektion af skævheder kan indebære svækkelse af beskyttelsen af privatlivets fred). Disse afvejninger skal dokumenteres og kommunikeres.
- Forskning og innovation skal fremmes med henblik på at opfylde kravene til pålidelig kunstig intelligens.

<u>Kapitel III: Central vejledning til evaluering af pålidelig kunstig intelligens:</u>

- Der skal vedtages en **evalueringsliste** for pålidelig kunstig intelligens ved udviklingen, udbredelsen og anvendelsen af kunstig intelligens, og listen skal tilpasses systemets specifikke anvendelsesformål.
- Det bør understreges, at en evalueringsliste aldrig vil være udtømmende, og det at gøre en kunstig intelligens pålidelig ikke kun handler om at afkrydse punkter på en liste, men om en løbende proces med at fastlægge krav, evaluere løsninger og sikre bedre resultater i hele KI-systemets livscyklus.

Denne vejledning er en del af en vision med en menneskecentreret tilgang til kunstig intelligens, som vil gøre Europa i stand til at blive en globalt førende innovator inden for etisk, sikker og avanceret kunstig intelligens. Dens formål er at lette og muliggøre **"Pålidelig kunstig intelligens fremstillet i Europa"**, som skal fremme de europæiske borgeres velfærd.