

Ansvarlig udvikling og anvendelse af AI

A. INDLEDENDE OPLYSNINGER	
Aktivetsområde	Indsatsområdet Digital sikkerhed, tillid og dataetik
Institut	Alexandra Instituttet
Titel <i>Dækker indholdet af aktiviteterne</i>	Ansvarlig udvikling og anvendelse af AI
Nummerering <i>Af beskrivelsen</i>	5
Version	1
Periode <i>Forventet start og slut</i>	1/1 2021-31/12 2021
Kontaktperson	Kristian Krämer

B. ÆNDRINGER
Angiv her, hvis en planlagt aktivitet er ændret i forhold til den forudgående version af beskrivelsen.

C. BESKRIVELSE	
1. Mål <i>Hvorfor? Hvad er målet for aktiviteterne? Hvordan bidrager de til det overordnede mål for indsatsområdet?</i>	Målet med aktiviteterne er at skabe grundlag for en kvalificering af teknologiernes potenti-ale for at skabe reel <i>impact</i> gennem udvikling af metoder og værktøjer og teknologiske komponenter, der styrker tilliden til AI-løsninger. At sikre tillid til AI-løsninger handler blandt andet om, at brugerne forstår, hvad løsningerne gør og kan, og hvordan de bruger dem i praksis. Målet er at styrke teknologiforståelsen og udvikle nye metoder og inddragselsesprocesser, der bidrager til, at brugere og borgere oplever, at de kan have tillid til AI – en aktivitet, planen også vil indeholde. En af de største udfordringer i forhold til at gennemføre AI-projekter i virksomheder er at komme fra <i>proof of concept</i> til produktion. Mange virksomheder efterspørger redskaber, retningslinjer, standarder og benchmarks for at kunne sætte moderne løsninger i drift. Da data kan være følsomme, er det altafgørende med modeller med fokus på ansvarlighed. Det er vigtig med nem adgang til uvildig rådgivning inden for tillid og digital sikkerhed, hvilket er en central GTS-rolle. Aktiviteterne bidrager til de overordnede mål for indsatsen ved at skabe grundlaget for udvikling af teknologiske services målrettet danske virksomheders behov for at arbejde ansvarligt med datadrevne modeller, der kan skabe forretningsmæssig indsigt og værdi. Aktiviteten bidrager samtidig med fagligt input i form af oplæg og artikler til den tværgående videnspredning. Disse mål og indikatorer understøtter direkte indsatsområdets mål om digital sikkerhed, tillid og dataetik.
2. Indhold <i>Hvad skal der ske? Hvilke(n) konkret(e) aktiviteter udføres?</i>	Kunstig intelligens til beslutningsstøtte og automatiske beslutninger medfører nye udfordringer for gennemsigtighed, forklaringer og ansvarlig anvendelse, såvel teknologiske som organisatoriske. Diskursen om <i>ansvarlig AI</i> er dog særdeles luftig, og skal nødvendigvis oversættes til værktøjer, der kan anvendes i praksis, hvis vi skal have held med at implementere principperne i danske virksomheder. Vi vil således i denne aktivitetsplan beskrive, hvordan vi, ud fra et praksisperspektiv, vil arbejde med udvikling og udbredelse af

	ansvarlig AI. Dette arbejde vil tage udgangspunkt i et helhedsperspektiv, som dækker både teknologi, forretning, processer og menneskelige kompetencer. En række af de aktiviteter vi vil arbejde med, inkluderer 1) Kortlægning af eksisterende arbejde med ansvarlig AI i danske virksomheder samt behovet for værktøjer og teknologier til understøttelse af samme. 2) Kortlægning af eksisterende/state-of-the-art-værktøjer og teknologier til understøttelse af ansvarlig AI, herunder huller i udbuddet, når dette holdes sammen med ovenstående afdækning af behov. Mulighederne vil blive demonstreret gennem én til flere virksomheds-cases i en tværfaglig tilgang med øje for teknologiske muligheder, brug, organisation og kontekst. Vi starter ikke på bar bund: Afdækning samt udvikling af værktøjer vil blive udviklet på baggrund af viden og kompetencer, der er indsamlet og oparbejdet i bl.a. forrige RK-periode, hvor vi allerede har identificeret en lang række af de udfordringer, virksomheder står overfor, når de skal udvikle og implementere ansvarlig AI, herunder mål for indsatsen, komplicerede juridiske og etiske spørgsmål, inddragelse af datasubjekter og brugere af løsningerne¹. Dette skal dog gøres mere systematisk, da forrige RK primært beskæftigede sig hovedsageligt med Explainable AI (forklarlig kunstig intelligens), som kan være en delkomponent, når vi taler om Ansvarlig AI.
3. Aktører <i>Hvem udfører aktiviteterne? Hvilken afdeling af instituttet? Evt. hvilke eksterne parter er med (videninstitutioner, virksomheder, erhvervsorganisationer, myndigheder, klyngeorganisationer eller andre.)</i>	Alexandra Instituttets medarbejdere på tværs af to faglige labs vil udføre aktiviteterne. Det gælder <i>Artificial Intelligence and Data Analytics Lab</i> og <i>People, Technology and Business Lab</i>. Derudover vil centrale eksterne samarbejdspartnere være: • Dansk Standard • Københavns Universitet, DIKU og Juridisk Fakultet • Dansk Industri, DI Garagen • Aalborg Universitet – AI for the People • DIREC, dansk national forskningscenter for digitale teknologier
4. Sammenhæng med andre projekter <i>Indgår aktiviteten i andre eksternt finansierede projekter?</i>	Vi vil i projektet skabe synergi med projektet AI:DK hos Industriens Fond, der har til formål at understøtte SMV'er med at komme hurtigere i gang med at udnytte data og AI-værktøjer i deres digitale omstillingsproces. Her vil de to projekter kunne geare hinanden, da formålet med nærværende aktiviteter ikke vil have som formål at bygge hele ML-modeller. Denne del vil AI:DK kunne byde ind med og dermed skabe grundlag for cases, hvor Ansvarlig AI kan spille en rolle.
5. Følgegruppe <i>Har følgegruppen forholdt sig til aktiviteten? I så fald hvordan?</i>	Samarbejdet med følgegruppen i denne aktivitet sker gennem løbende involvering af en undergruppe af interessenter gennem aktivitetsbeskrivelsen "Demonstration af anvendelighed og værdiskabelse ift. indsatsområdet Digital sikkerhed, tillid og dataetik", hvor risikostyring omkring potentiel konkurrence også vil blive håndteret. Aktiviteten bidrager med faglig viden og forslag på teknologiske services, som følgegruppen orienteres om og efterfølgende forholder sig til.

¹ [Kunstig intelligens i praksis, oplevelser og erfaringer fra +20 virksomheder. Alexandra Institut, September 2020](#)

6. Formidling af resultater <i>Hvordan/hvor kan interesserede virksomheder m.fl. få viden om resultaterne af aktiviteterne?</i> <i>Anføres/tilføjes hvis det ikke allerede fremgår af beskrivelsen ovenfor, f.eks. ved links til konferencer, hjemmeside, publikationer etc.</i>	I forrige RK-periodes arbejde med Explainable AI begyndte vi at arbejde med overbygningen <i>Ansvarlig AI</i>. Som afkast fra dette arbejde udviklede vi bloggen <i>Ansvarlig AI</i> (https://medium.com/ansvarlig-ai), som dette projekt vil bygge videre på. Ligeledes påbegyndte vi en kortlægning af viden om især Explainable AI, men også begyndende ansvarlig AI, som vi deler på en GitHub-side (https://github.com/alexandrainst/responsible-ai), der er tilgængelig for alle interesserede. Formidling vil desuden ske i koordination med indsatsens overordnede videnspredning i aktiviteten "Demonstration af anvendelighed og værdiskabelse ift. indsatsområdet Digital sikkerhed, tillid og dataetik", hvor denne aktivitet generelt vil levere fagligt indhold. OPDATERING: Væsentlige aktiviteter og resultater opnået i 2021: • Deltagelse i konferencen Respons ability med oplægget: "Who needs an explanation? Service design as a tool for explaining AI" • Deltagelse i DIREC-arbejdsgruppe for "AI technologies for explainability" • Bidrag til udvikling af Dansk Standard for "Bias i AI-systemer og AI-assisteret beslutningstagning" • Opstart af state-of-the-industry analyse af danske virksomheders arbejde med ansvarlig og forklarlig kunstig intelligens
--	---