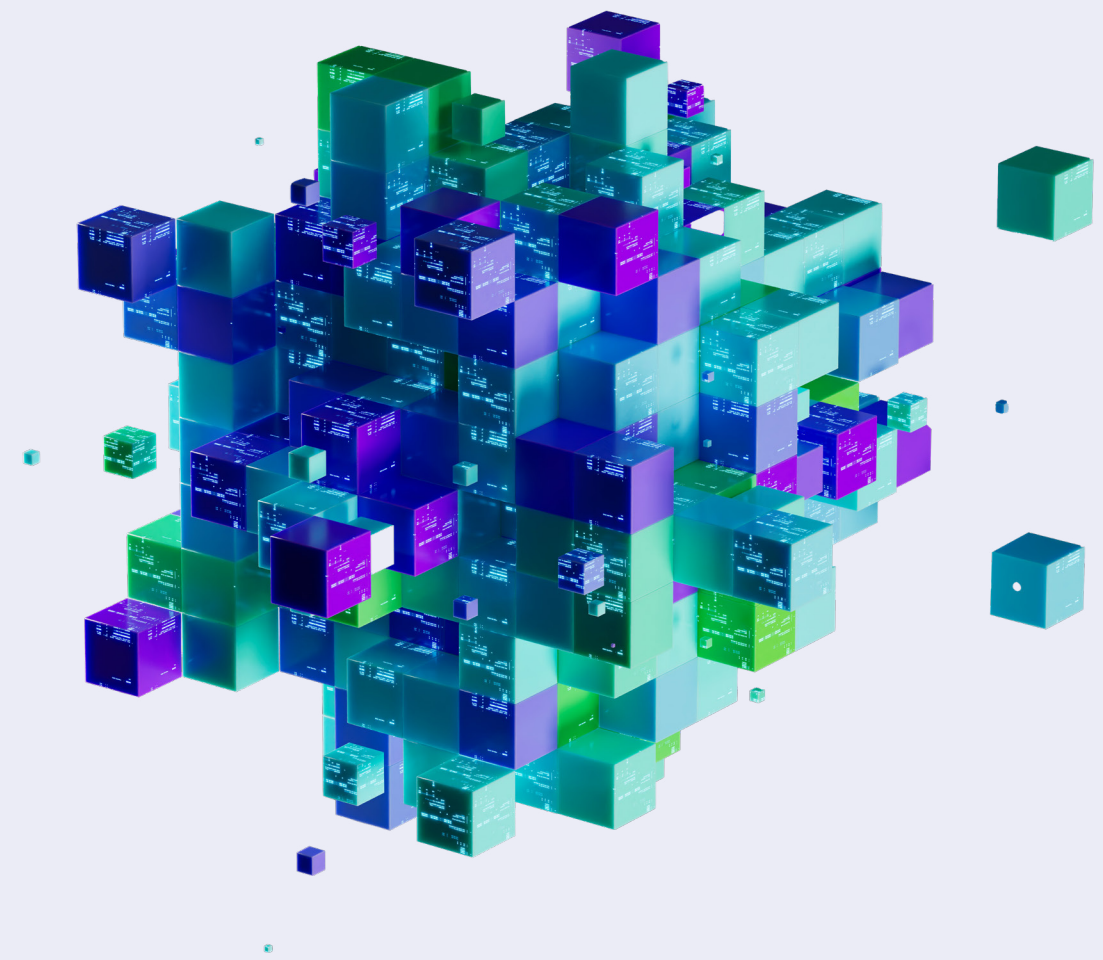


Hvidbog

Ansvarlig brug af AI-assistenter i den offentlige og private sektor



		Erhvervsstyrelsen
		Kombit
		ATP/Udbetaling Danmark
		Digitaliseringsstyrelsen
		Spar Nord
		PFA Pension
		TopDanmark
		Jyske Bank
		Netcompany
		Dubex
		Microsoft
		Trifork
		Kammeradvokaten
Afsnit 01	Introduktion & Formål	04
1.1	Målgruppe	06
1.2	Begreber	08
1.3	Afgrænsning	11
Afsnit 02	Hvad er en AI-assistent?	12
Afsnit 03	Fremgangsmådens Ni Punkter	14
3.1	Punkt 1: Definer AI-assistentens brugsscenarie	18
3.2	Punkt 2: Understøt AI-assistenten med en fleksibel teknisk platform	22
3.3	Punkt 3: Vurder nødvendige data og databehandling	26
3.3.1	Databehandling	28
3.4	Punkt 4: Forhold dig til de juridiske rammer	32
3.4.1	AIA	33
3.4.2	GDPR	38
3.5	Punkt 5: Afgræns AI-assistentens evner og ansvar	42
3.6	Punkt 6: Opbyg struktureret kvalitetssikring	44
3.7	Punkt 7: Mål og opbevar relevant data om AI-assistentens brug	48
3.8	Punkt 8: Planlæg organisatorisk implementering og uddannelse	52
3.9	Punkt 9: Etabler opfølgnings— og supportstrukturer	58
Afsnit 04	Afslutning	62
Afsnit 05	Referencer	64
Afsnit 06	Bilag	66
Afsnit 07	Addenda	74

Introduktion & Formål

C1



Vi befinder os i en tid, hvor teknologiske fremskridt i stadig stigende grad udfordrer vores måde at tænke, arbejde og leve på – og med kunstig intelligens som en af de helt store transformerende drivkræfter står vi ikke blot overfor nye udfordringer, men med et hidtil uset potentiale for at skabe bedre og mere effektive løsninger på tværs af sektorer.

I Danmark har vi en enestående mulighed for at udnytte teknologiens potentiale til gavn for hele samfundet, men der er samtidig tale om en nødvendighed, hvis vi vil sikre vores konkurrenceevne og styrke vores velfærd. Verden står i en brydningstid med stigende militære og økonomiske spændinger. Her kommer evnen til at kunne udvikle og anvende teknologier – ikke mindst kunstig intelligens – til at spille en stadig mere afgørende rolle for Europa.

Hvis vi skal kunne udvikle, implementere og anvende teknologien på den bedste, mest innovative og ansvarlige måde, kræver det et fælles kompas for, hvordan vi navigerer i den nye virkelighed – i juridisk, teknisk og organisatorisk forstand. Det er en vej, vi historisk set bedst har fundet i fællesskab.

Den fundamentale, digitale infrastruktur i Danmark er bygget i samarbejde mellem offentlige myndigheder og private leverandører. Derfor har centrale aktører fra DI Digital, anført af Netcompany, samt Kammeradvokaten, Microsoft, Dubex og Trifork, sammen med digitalt førerende offentlige organisationer (Digitaliseringsstyrelsen, Erhvervsstyrelsen, KOMBIT og UdbetalingDanmark), og den finansielle sektor (repræsenteret gennem ATP, Jyske Bank, PFA, Spar Nord og Topdanmark) indgået et unikt samarbejde om

udviklingen af en hvidbog, der har til formål at beskrive, hvordan AI-assisterter kan udvikles, implementeres og anvendes såvel innovativt som ansvarligt i den offentlige og private sektor i Danmark.

Hvidbogen tilbyder en konkret tilgang til anvendelse af AI-assisterter inden for eksisterende lovgivning som GDPR og EU AI Act (AIA). Den berører centrale udfordringer som risiko for misinformation i form af hallucinationer og bias, hvilket kræver nøje kvalitets sikring og robust risikostyring. Ligeledes indeholder den addenda med eksempler på tidligere implementering af AI-assisterter fra de deltagende organisationer, der kan fungere som inspiration og grundlag for fremtidige løsninger. Eksempler, der tjener som indblik i, hvordan teknologien succesfuldt anvendes i praksis, så den skaber mest mulig værdi for borgere og virksomheder.

Hvidbogen markerer en væsentlig milepæl for brugen af AI i Danmark og fremhæver, hvordan samarbejde mellem landets førende digitale aktører kan føre til udvikling af skalerbare, pålidelige og lovmedholdelige anvendelse af AI. Den er et værktøj, der kan hjælpe organisationer med at stille de rette spørgsmål, træffe informerede valg og bygge videre på fælles erfaringer.



Målgruppe

1.1.



Denne hvidbog er udviklet til brug i både den offentlige og den private sektor. Den er rettet mod såvel topledelsen, der sikrer den strategiske forankring og ressourceallokering, som projektteamet, der står for implementeringen af AI-assistenter – det værende alt fra projektledere, forretningsrepræsentanter, AI-udviklingsteamet, IT-arkitekter og juridiske rådgivere.

Begreber

1.2.

Logning

Automatiseret registrering af væsentlige systemhændelser, der giver indsigt i en AI-assistents ydeevne, sikkerhed og brugsmønstre. Logs anvendes til vedligeholdelse og problemløsning samt til dokumentation af brugen af AI-assistenten og dens databehandling.

On-premise

Lokale IT-tjenester drevet af en organisation eller leveret af en direkte leverandør til organisationen. Mindre sprogmodeller kræver mindre infrastruktur og kan derfor hostes on-premise.

Platform

En teknisk løsning, der muliggør udvikling og implementering af AI-assistenter.

AI-assistent

Et system eller en del af et system, der anvender en sprogmodel til at understøtte og automatisere en forretningsproces (se afsnit 2).

Hallucinationer

Sprogmodellers generering af faktisk forkert information, der kan virke overbevisende for brugeren [1].

Prompt Engineering

Proces omhandlende design af instruktioner, der effektivt guider sprogmodeller til at levere relevante og informative resultater af høj kvalitet [1].

Domæne

En organisations forretningsområde.

Sprogmodel

En klasse af generative modeller, der er trænet på store dataset. De generer typisk tekst baseret på en forespørgsel [1].

Bruger

En person, der interagerer med en AI-assistent. Brugeren kan være intern eller ekstern for organisationen. Interne brugere refereres til som 'medarbejdere', mens eksterne brugere benævnes som 'borgere'. I hvidbogen inkluderer 'borger' også kunder i private organisationer.

Cloud

Internetbaserede IT-tjenester leveret af tredjeparter. Kraftige sprogmodeller kræver ressourcetung infrastruktur og tilgås typisk gennem cloud.



1.3.

Afgrænsning

Hvidbogen fokuserer på AI-assistenter baseret på generativ AI, primært sprogmodeller, og dækker ikke brugen af anden type AI, så som prædiktive modeller. Den forudsætter, at organisationen allerede har identificeret AI-assistenter, der skaber forretningsværdi og kan understøttes af sprogmodeller.



Hvidbogen behandler relevante forpligtelser i GDPR samt AIA, da disse lovgivningsmæssige rammer er relevante for alle organisationer, der tager en AI-assistent i brug. Der er ikke tale om en udtømmende liste over alle juridiske forpligtelser, f.eks. er AI-assistenter, som AIA klassificerer som havende uacceptabel risiko, ikke omfattet af hvidbogen.

Forskellige organisationer har deres egne etiske retningslinjer og standarder. Det er uden for hvidbogens scope at definere en samlet etisk retningslinje.



Hvad er en AI-assistent?

AI-assistenter er systemer eller dele af systemer, der anvender sprogmodeller til at udføre dele af en forretningsproces med en vis grad af autonomi. De genererer svar baseret på de oplysninger og forespørgsler, de modtager, hvilket kan fremme interaktion med andre systemer, optimere beslutningsprocesser og skabe mere fleksible, data-drevne løsninger.

02

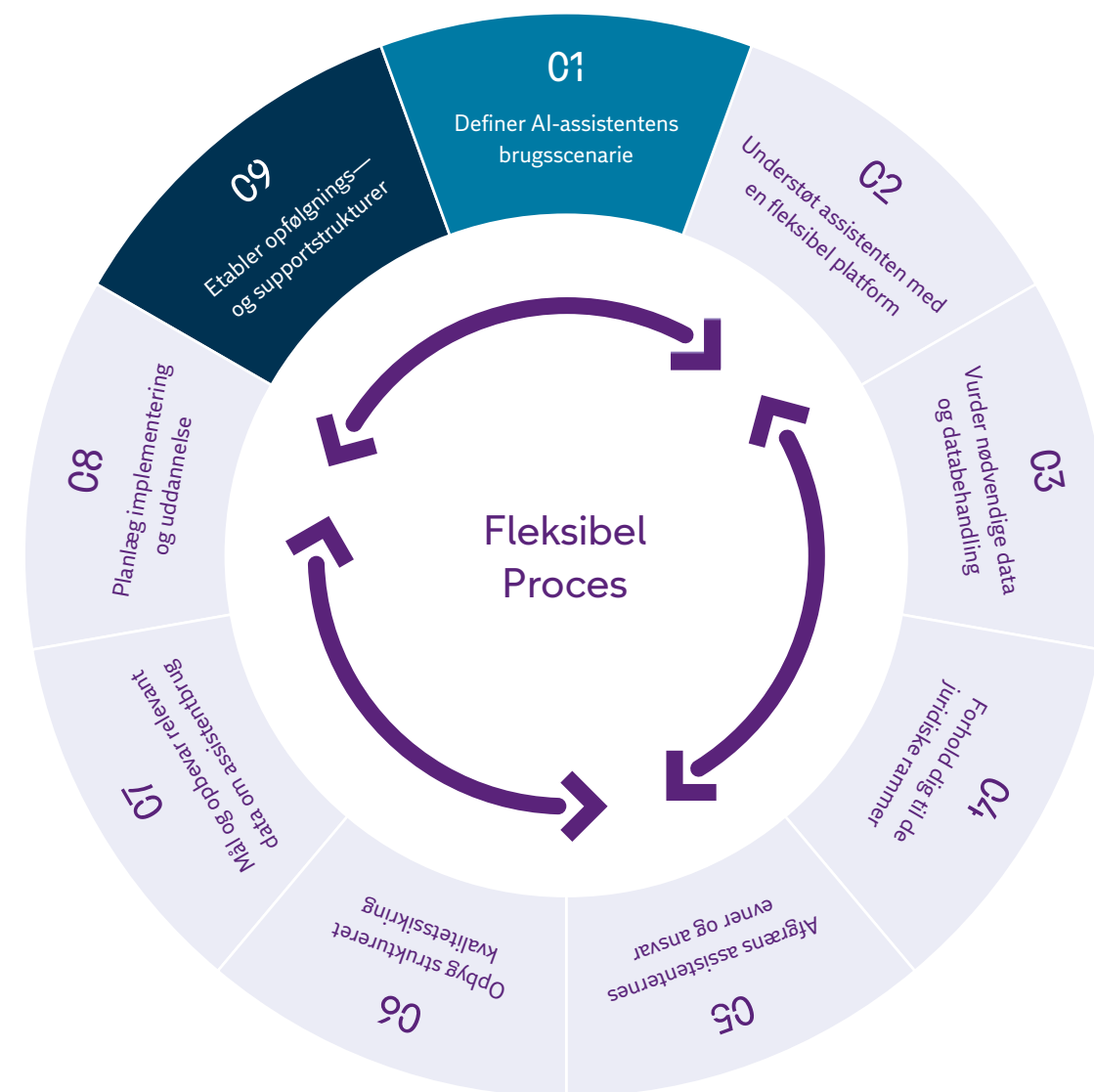
En AI-assistent kan således håndtere opgaver, der normalt kræver menneskelig vurdering, med anvendelse af værktøjer og data fra eksisterende forretningssystemer. Her kan være tale om udformning af mødereferater, opdatering af informationer baseret på e-mailkorrespondancer, støtte til juridisk rådgivning eller chatbots, der tilbyder personaliseret kundeservice. De implementeres således typisk med det formål at udføre specifikke opgaver mere ensartet, effektivt og med højere kvalitet.

AI-assistenterne kan udvikles med enkeltstående opgaver for øje eller indgå i komplekse arbejdsgange med flere assistenter, så længe det sker inden for gældende lovgivning og i henhold til særlige krav om transparens. Nedenfor opstilles en række principper, som AI-assistenten skal leve op til uanset forretningsdomænet. Principperne udgør et informativt grundlag, som organisationer kan anvende til at definere krav til deres AI-assistenter.

Princip	AI-assistenten præsenterer informationsgrundlaget for sine forespørgsler svar på en forståelig måde.
Rationale	Når brugeren kan se, hvilken information modellen bygger sine svar på, bliver det lettere at vurdere kvaliteten af resultaterne og identificere eventuelle hallucinationer. En typisk tilgang er, at modellen beskriver sit informationsgrundlag eller giver kildehenvisninger.
Princip	AI-assistenten benytter kun data, der er vurderet som relevante for en konkret forespørgsel.
Rationale	Ved anvendelse af relevant data sikres, at AI-assistentens svar forbliver fokuserede og præcise. Dette reducerer risikoen for at introducere irrelevante eller vildledende informationer, hvilket forbedrer kvaliteten og pålideligheden af svarene. Desuden optimerer det effektiviteten ved at begrænse den mængde data, der skal behandles.
Princip	AI-assistenten udfører selvstændigt nødvendige opgaver og informerer brugeren, hvis en opgave ligger uden for dens ansvar eller kræver assistance.
Rationale	Ved at operere selvstændigt maksimeres AI-assistentens nytteværdi og tillid. Når brugeren er opmærksom på begrænsninger eller manglende information, minimeres misforståelser, og når informationen hentes fra både forespørgsel og vidensbase undgås potentielt fejlagtige svar.
Princip	AI-assistenten forstår konteksten af en forespørgsel samt tidligere, relevante beskeder fra samme samtale.
Rationale	Ved at integrere tidligere interaktioner kan AI-assistenten give mere nuancerede og relevante svar. Det reducerer behovet for gentagelser og gør kommunikationen mere naturlig og effektiv.

Fremgangs- mådens Ni Punkter

C3



Figur 3.1: De ni punkter for AI-implementering

Dette afsnit præsenterer en systematisk og iterativ proces, der guider organisationer i udvikling, implementering og vedligeholdelse af sikre, effektive og ansvarlige AI-assistenter. Processen består af i alt ni punkter, der er vigtige aktiviteter i arbejdet med sikker og ansvarlig AI. De indbefatter alt fra indledende planlægning til løbende drift og optimering (Figur 3.1). Selvom processen minder om den, man typisk ser ved implementering af IT-systemer, har denne proces et særligt fokus på de aspekter, der er afgørende af væsentlig betydning for AI-assistenter.

Punkterne relevans vil nødvendigvis afhænge af den enkelte organisation og de respektive sektors behov og regulering. Det er dog

essentielt, at organisationer overvejer og dokumenterer årsagerne til eventuelle fravalg af specifikke punkter.

Punkterne er organiseret i en logisk rækkefølge, men bør betragtes som indbyrdes forbundne og gensidigt påvirkende. Tilgangen er iterativ, da organisationer ofte vil skulle gentage og revurdere, efterhånden som forståelsen bliver større, nye udfordringer opstår, eller anvendelsesområdet for AI-assistenten udvides. Dette kan f.eks. være ændringer i de juridiske rammer, ligesom ny data kan medføre yderligere overvejelser om lovgivning og sensitivitet. Således må man som organisation løbende holde sig opdateret på juridiske forhold for at sikre lovmedholdelighed.

Planen består af følgende 9 punkter

C1

Definér AI-assistentens brugsscenarie

C2

Understøt assistenten med en fleksibel platform

C3

Vurder nødvendige data og databehandling

C4

Forhold dig til de juridiske rammer

C5

Afgræns assistenternes evner og ansvar

C6

Opbyg struktureret kvalitetssikring

C7

Mål og opbevar relevant data om assistentbrug

C8

Planlæg implementering og uddannelse

C9

Etabler opfølgings— og supportstrukturer

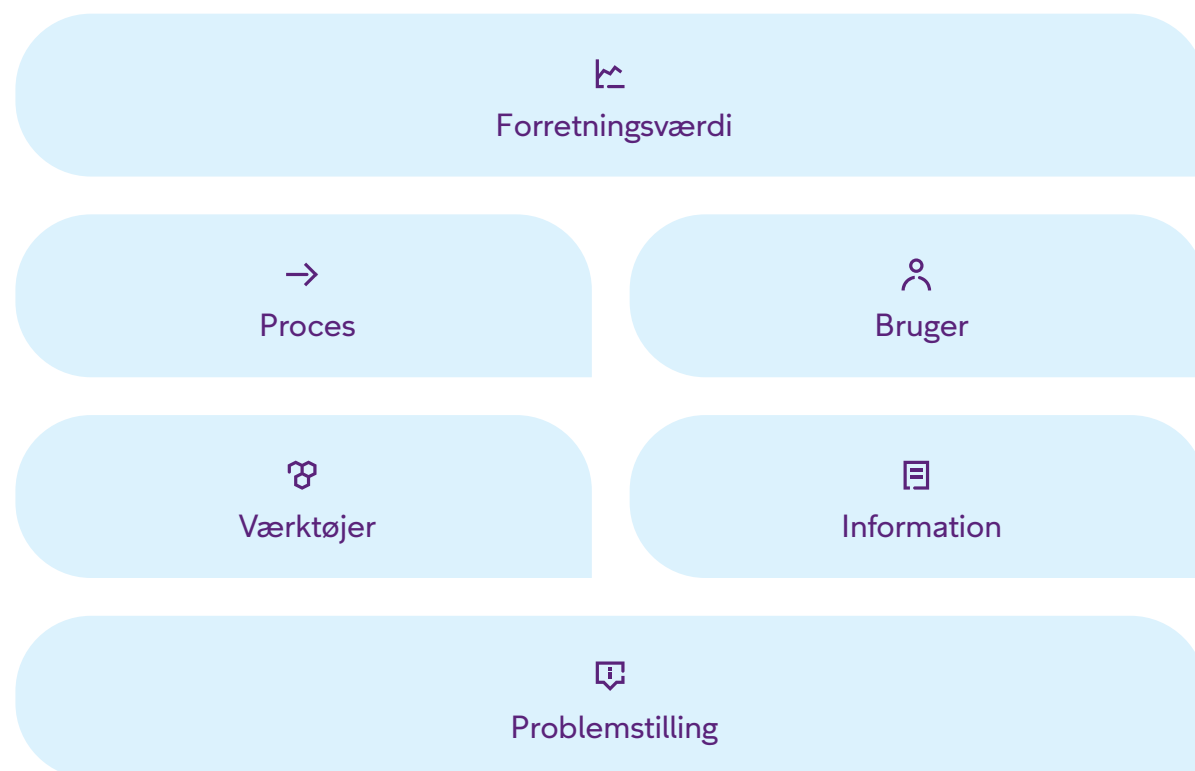
De enkelte punkter bidrager til forskellige faser i AI-assistentens livscyklus, fra planlægning og design til implementering og drift. Ved at følge disse punkter til implementering og vedligeholdelse iterativt, kan organisationer adoptere en holistisk tilgang, der tager højde for både tekniske, organisatoriske, juridiske og forretningsmæssige aspekter. Denne fleksible tilgang muliggør løbende læring, tilpasning og optimering i takt med udviklingen af teknologi, organisation, lovgivning og omverden. I de kommende afsnit vil hvert punkt blive uddybet.



3.1

Definer AI-assistentens brugsscenarie

Værdifulde brugsscenarier for AI-assisterter kan omfatte hele eller dele af arbejdsprocesser, der kræver samarbejde mellem flere personer i organisationen og som ikke kan løses med enkle funktioner. Selvom AI-assisterter kan udføre standardiserede opgaver, giver den generative natur af sprogmodeller dem mulighed for at håndtere mere dynamiske processer, hvor der kræves beslutningstagning undervejs. Det er vigtigt at overveje, hvordan organisationens nuværende teknologi kan være forældet. Sprogmodeller tilbyder innovative løsninger på opgaver, som eksisterende softwareløsninger ikke dækker, samtidig med at potentialet i AI-assisterterne udnyttes. →



Figur 3.2: De fire perspektiver, et brugsscenarie beskrives ud fra mellem problemstilling og forretningsværdi.

Implementeringen af en AI-assistent udvider organisationens digitale kapabiliteter. For at sikre en vellykket integration skal brugsscenariet transformere AI-assistentens vision til en konkret beskrivelse, der danner grundlag for videre aktiviteter. Dette brugsscenarie skal være forankret i den specifikke problemstilling, som AI-assistenten skal løse, med fokus på den forretningsværdi, den skal skabe. Når brugsscenariet udvikles, bør den overvejes ud fra fire perspektiver: proces, brugere, information og værktøjer.



Definer AI-assistentens brugsscenarie

Disse punkter danner fundamentet for den videre implementeringsproces. Den sikrer en klar retning og forståelse for, hvordan AI-assistenten integreres og skaber værdi i en konkret kontekst, ligesom den adresserer potentielle udfordringer og risici proaktivt.

Brugere

Det er væsentligt at overveje, hvem brugerne er, hvordan de tilgår og interagerer med AI-assistenten, samt hvordan resultaterne påvirker deres arbejde. Interaktion med AI-assistenten kan variere alt efter, om brugerne er interne medarbejdere, som kan uddannes i anvendelse, eller borgere, som blot vejledes heri, mens implementeringen kan hindres af modstand mod nye værktøjer. Afsnit 3.1 arbejder med uddannelse af brugere.

Problemstilling

Et brugsscenarie bør have et klart formål, som organisationen ønsker at adressere, såsom forbedring af kvalitet, optimering af services eller effektivisering af processer. Problematikken behøver ikke nødvendigvis at være en udfordring, den kan også være en mulighed for f.eks. at opgradere en ældre IT-løsning eller forbedre eksisterende processer.

Proces

Den arbejdsproces, som AI-assistenten skal støtte, kan være simpel, som at finde information i et dokument, eller kompleks, som at udarbejde analysegrundlaget for en vurdering. Den nuværende menneskelige arbejdsproces kortlægges, hvorefter den AI-understøttede proces beskrives.

Information

AI-assistenten skal have relevant viden tilgængeligt for at kunne udføre sine opgaver korrekt. Det kan omfatte data fra andre systemer, vejledninger osv. Med korrekt information bliver modellen bedre rustet til at operere i sit domæne, hvilket mindsker risikoen for hallucinationer [2]. Videre arbejde med data samt databehandling gennemgås i afsnit 3.3.

Værktøjer

De funktioner, AI-assistenten skal anvende for at udføre sine opgaver, kan omfatte integration med andre systemer til beregninger eller dataændringer. AI-assistenten, der udfører vurderinger, kan således f.eks. have brug for adgang til en model til kvantitative analyser, da sprogmodeller ikke er velegnede til dette.

Forretningsværdi

Definerede KPI'er for AI-assistentens formål er afgørende for at demonstrere dens værdi, f.eks. svartider, fejlrate, produktivetsforøgelse og kundetilfredshed. Ved at etablere en baseline for performance før implementeringen kan dens effekt bedre vurderes. Selvom det kan være udfordrende at opstille præcise mål for nye teknologier, bør subjektive faktorer som oplevet effektivitet og tilfredshed også måles. Desuden skal eventuelle omkostninger til drift og support tages i betragtning.

3.2

Understøt AI-assistenten med en fleksibel teknisk platform

For at fremme sikkerhed og effektivitet bør AI-assistenter centraliseres i en AI-platform med en robust sikkerhedsmodel. Dette forhindrer uautoriserede løsninger og muliggør genbrug samt integration med andre systemer, når flere brugsscenarier skal dækkes.

En organisation bør strategisk overveje, hvordan den ønsker at arbejde med AI-assistenter. Som udgangspunkt er der tre tilgange:



Egenudvikling

Giver fuld kontrol og tilpasning, men kræver betydelige ressourcer og teknisk ekspertise.



Leverandør-samarbejde

Balancerer kontrol og ekstern teknisk ekspertise.



Outsourcing

Hurtig implementering, men mindre kontrol over løsningen.

En organisation kan sagtens arbejde under forskellige strategier for brug af AI-assistenter. Ikke-forretningskritiske AI-assistenter kan således købes som standardprodukter, mens kerneforretningsprocesser kan kræve egenudvikling, og når AI-assistenten tænkes ind i organisationen, bliver den en del af organisationens samlede systemlandskab.

En platform tilbyder flere fordele som lavere omkostninger, kortere time-to-market og øget innovation gennem genbrug af komponenter, og valget mellem

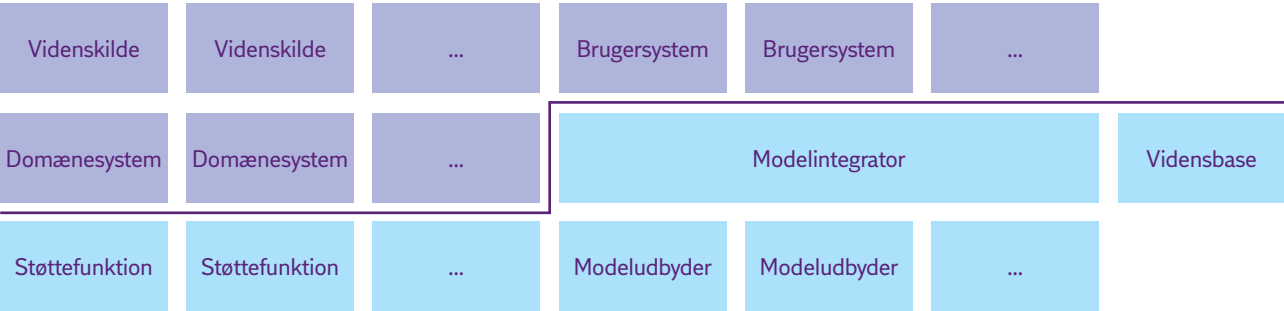
en standardplatform eller en skræddersyet løsning afhænger af organisationens behov. En standardplatform muliggør hurtigere implementering, lavere initialomkostninger og support, mens en skræddersyet platform tilbyder fuld kontrol og tilpasning til specifikke behov, men med længere udviklingstid og højere omkostninger.



Ved at tage organisationens langsigtede strategi og mål i betragtning kan simple AI-assistenter udvikles på en standardplatform, mens domænespecifikke AI-assistenter kan drage fordel af en tilpasset platform. Den valgte tilgang har indflydelse på ens ansvar ift. AIA, afsnit 3.4.1.

En platform bør være i stand til at integrere nye sprogmodeller, så organisationen kan skifte model i takt med den teknologiske udvikling. Figur 3.3 viser komponenterne i et systemlandskab, hvor AI-plattformen er løst koblet med modelintegratoren, modeludbydere, vidensbasen samt potentielle støttefunktioner.

■ Forretningssystemer ■ AI-plattform



Figur 3.3: Komponenterne i et systemlandskab centreret om modelintegratoren, der udstiller AI-assistenter.

Ved at tænke sine systemer ind i dette landskab muliggøres udskiftning af sprogmodeller i takt med teknologi- og prisudviklingen samt genbrug af AI-assistenter på tværs af f.eks. mobilapps, hjemmesider og desktopprogrammer. Nedenfor gennemgås systemlandskabets komponenter.

Modeludbyder

Modeludbydere er platforme, der gør sprogmodeller tilgængelige for AI-assistenter gennem systemgrænseflader. Disse udbydere kan være cloud-baserede fra en ekstern leverandør eller on-premise løsninger leveret af organisationen selv. Ved valg af modeludbyder skal organisationen overveje, om denne kan håndtere antallet af forventede forespørgsler, modellen vil modtage, samt sikre, at modeludbyderen håndterer data korrekt og overholder GDPR. Ved skift af model bør den nye modeludbyder kvalitetssikres som den tidligere udbyder, se afsnit 3.6. Modeludbyderen er nødvendigvis ikke udbyder ift. AI-assistenten ud fra AIA, se afsnit 3.4.1.

Modelintegrator

Modelintegratoren orkestrer interaktionen mellem forskellige komponenter såsom forespørgsler fra brugersystemer, valg af den passende model fra en modeludbyder og anvendelse af data fra vidensbasen. AI-assistenterne udstilles af modelintegratoren som services og er ikke direkte koblet til modeludbyderne. Dette sikrer, at AI-assistenterne er uafhængige af de enkelte modeludbydere og samtidig har adgang til relevante forretningssystemer.

Vidensbase

Videnskilder udstiller domænespecifik information til AI-assistenten, som den bruger til at udføre sine opgaver. Informationen kan være både struktureret og ustruktureret, ligesom visse informationer kræver forberedelse såsom indlejring. Ved integration af videnskilder skal det sikres, at data er nøjagtig, opdateret og relevant (se afsnit 3.3) samt at data er tilgængelig og i et format, der kan behandles af modelintegratoren.

Domænesystem

Domænesystemer er IT-systemer i en organisation, der understøtter forretningsprocesser. En AI-assistent benytter et domænesystem til at udføre handlinger, så som at oprette, ændre eller fjerne informationer. Handlingen, en AI-assistent udfører i systemet, skal være udstillet gennem en systemgrænseflade. AI-assistenten må ikke udføre en handling på vegne af brugeren, som brugeren ikke selv har rettigheder til. Hvis AI-assistenten for eksempel forsøger at slette information, som brugeren kun har læserettigheder til, skal handlingen afvises.

Brugersystem

Brugersystemer er de programmer, hvorigennem brugere interagerer med AI-assistenter. Disse kan være chatbaserede applikationer eller eksisterende systemer i organisationen. En bruger må ikke have adgang til en AI-assistent, der udfører opgaver, brugeren ikke selv har adgang til at udføre. Brugersystemet kan præsentere AI-assistenten som en chat-interaktion eller benytte AI-assistenten i baggrunden til at håndtere en proces automatisk. For at underbygge en god organisatorisk implementering er det vigtigt, at interaktionen med AI-assistenten er brugervenlig. Det bør ligeledes overvejes, hvordan kritiske opgaver fortsat kan udføres, hvis AI-assistenten er nede. Hvis AI-assistenten er beslutningsfremmende eller varetager behandling af individer, skal der tages højde for menneskeligt tilsyn, se afsnit 3.4.1. I disse tilfælde vil et brugersystem kunne benyttes.

Videnskilde

En vidensbase er et centraliseret lager til opbevaring og håndtering af information, som sprogmodellen bruger til at besvare forespørgsler. Vidensbaser er typisk implementeret gennem en database. Vidensbasen indeholder typisk data fra Videnskilder, som skal forberedes, før det kan bruges. Når en vidensbase vælges, er det vigtigt at overveje, hvordan dataene hænger sammen, og hvordan modelintegratoren hurtigt og effektivt kan finde den nødvendige information.

¹ Indlejring (embedding): En metode, hvor tekst omdannes til vektorer, der repræsenterer tekstens sammenhæng og kan benyttes til at lave senere opslag ved brug af en forespørgsels indlejring.

Støttesystem

Støttesystemer omfatter en række forskellige funktioner og teknologier, som sikrer, at AI-assistenter fungerer optimalt og i overensstemmelse med organisationens krav og standarder. De benyttes typisk til at bygge, træne og fintune modeller.

Et system kan sagtens have flere roller i systemlandskabet. For eksempel kan et sagsbehandlingssystem både være en videnskilde, hvorfra en AI-assistent henter informationer om en konkret sag, et domænesystem, hvor AI-assistenten opdaterer informationer på sagen samt selve brugersystemet, hvor sagsbehandleren interagerer med AI-assistenten. Når et system har flere roller, er det vigtigt at overveje de forskellige krav, dette kan medføre.



3.3

Vurder nødvendige data og databehandling

Visse opgaver kræver, at AI-assistenten har opgavespecifik viden som vejledninger, sagsinformationer eller lign. Data, der varierer i format, følsomhed og struktur. For at sikre tillid og undgå datalækage skal datahåndteringen være struktureret, sikker og dokumenteret. Data skal opdateres løbende for at sikre korrekt og relevant baggrundsinformation til AI-assistenten, ligesom processer for styring og forvaltning er påkrævet af AIA for at sikre datakvalitet (se afsnit 3.4.1).

Overordnet skelnes der mellem to former for data:
Struktureret og ustruktureret data



Struktureret data er ofte nemmere at håndtere systematisk end ustruktureret data, og data kan variere i graden af struktur. For eksempel kan en samling af dokumenter have ensartede overskrifter, som skaber en vis struktur, mens struktureret data kan indeholde felter med billeder eller formateret tekst. Når data skal forberedes, forslås følgende dimensioner som relevante at overveje, for at sikre kvalitet:

Korrekthed

Hvor nøjagtig og pålidelig informationen er ift. til domænet, det skal bruges i. Det kræver ofte en manuel gennemgang at sikre korrekthed af ustruktureret data, da den omliggende kontekst og fortolkning er vigtig. Struktureret data er lettere at validere på grund af det foruddefinerede format og kan ofte automatisk kontrolleres for fejl og inkonsistens.

Aktualitet

Hvor relevant og tidssvarende informationen er. Ustruktureret data kan indeholde tidssensitive oplysninger, som kræver regelmæssig opdatering, som for eksempel dokumentation for et system under udvikling. Struktureret data kan nemmere integreres med automatiske opdateringssystemer. For begge datatyper er det vigtigt at vurdere, hvor ofte data opdateres eller udgår.

Kontekst

Hvor tydelig og forklarende informationen er uafhængigt af dens kontekst. Ustruktureret data er typisk rig på kontekst, mens struktureret data ofte kræver supplerende dokumentation eller metadata for korrekt fortolkning.

Fortrolighed

Hvor sensitiv information er, f.eks. personoplysninger eller forretningskritisk strategisk data. Det kan være udfordrende automatisk at identificere fortrolige informationer i ustruktureret data, medmindre den fortro-

lige information er kendt og simpel. Struktureret data med fortrolige felter er ofte simplere at forberede.

Tilgængelighed

Hvor let og simpelt data tilgås ift. sikkerhed og opetid. For AI-assistenter kan datas tilgængelighed være en udfordring, da information ofte er spredt over forskellige systemer, med forskellige adgangsstyring samt forskellig forventning til, hvornår systemet kan tilgås.

Bias

Hvor repræsentative og retfærdige data er. Ustruktureret data kan have skjulte bias, som kan være svære at opdage uden kontekst. Struktureret data kan også være have bias, især hvis dataindsamlingen er begrænset. Bias kan ofte måles og korrigeres i data, men det kræver opmærksomhed på indsamlingsmetoder og repræsentativitet.

I henhold til ovenstående fokus på datakvalitet bør der under AI-assistentens dataarbejde være fokus på datasikkerhed samt eksisterende praksis som CIA -triaden².

² CIA-triaden består af tre nøgleelementer inden for informationssikkerhed: fortrolighed, integritet og tilgængelighed (eng. Confidentiality, Integrity, Availability). Fortrolighed: Data må kun være tilgængelig for autoriserede personer og beskyttet mod uautoriseret adgang. Integritet: Data skal være nøjagtig og fuldstændig og beskyttet mod uautoriseret ændring. Tilgængelighed: Data skal være tilgængelig og anvendelig, når autoriserede brugere har behov for det [9].

3.3.1

Databehandling

Når en AI-assistent forbereder og sender data til et værktøj eller underliggende sprogmodel, er det vigtigt at implementere robuste sikkerhedsforanstaltninger samt at overholde databeskyttelsesregler som GDPR, se afsnit 3.5. Hvad angår databehandling skelnes der mellem offentligt data og personoplysninger.



Offentligt data



Information, der er frit tilgængelig for offentligheden, som offentlige registre, åbne datasæt fra regeringer og virksomhedsregistre eller -information Offentligt data kræver ikke særlig beskyttelse eller samtykke for at blive behandlet

Personoplysninger



I persondatalovgivningen er personoplysninger kategoriseret efter følsomhed. Her skelnes mellem almindelige personoplysninger som navn, adresse, telefonnummer, og følsomme personoplysninger, der karakteriseres ved oplysninger som race, religion, helbred og seksuel orientering. GDPR regulerer yderligere børns data og straffetretlige oplysninger, mens personnummer (CPR-nummer) er særskilt reguleret i den danske databeskyttelseslov [3].

Databehandling

En organisation kan vælge at klassificere visse data, såsom konkurrencekritiske procesbeskrivelser, som fortrolige, selvom de ikke indeholder personoplysninger. I sådanne tilfælde kan et databrud få konkurrencemæssig betydning, hvilket gør det vigtigt for organisationer at have systematiske beskyttelsesforanstaltninger for disse data. Under arbejdet med data, er det vigtigt at tænke i principperne omkring databeskyttelse gennem design og databeskyttelse gennem standardindstillinger, se [4].

Når man overvejer, om en model skal være cloud-baseret eller on-premise, er det vigtigt at tænke på, hvordan dataene behandles. Offentlige data kan deles via en cloud-baseret model, men hvis persondata sendes til en cloud-løsning, skal databehandlingen hos udbyderen være korrekt og dokumenteret. On-premise sprogmodeller reducerer afhængigheden af tredjeparter og gør det muligt for virksomheder at implementere tilpassede sikkerhedspolitikker. Disse sprogmodeller er typisk mindre, hvilket begrænser kompleksiteten af de opgaver, sprogmodellerne kan udføre. Der ligger derfor et trade-off mellem fortrolighed og kvalitet, når der vælges mellem en cloud- eller on-premise model.

AI-assistenten skal kun have adgang til data, som er relevante i forhold til opgaven, og som brugeren af AI-assistenten har adgang til (se afsnit 3.5). Der er flere metoder, der kan anvendes for at minimere risikoen under databehandling, når modellen udbydes over cloud.

C1

Retrieval-augmented generation (RAG)

En metode, hvor relevant information på forhånd forberedes og ved forespørgsel hentes fra en vidensbase. Informationen indsættes som kontekst i forespørgslen. Ved brug af RAG generer sprogmodellen mere præcise og kontekstbaserede svar. I modsætning til traditionel træning af sprogmodellen, hvor data er integreret i modellen, eksisterer data selvstændigt fra modellen, når RAG-metoden benyttes. Data kan der-for nemmere opdateres uden at gentræning af modellen. RAG sænker samtidig risikoen for hallucinationer, da information bliver en del af modellens kontekst [2].

C3

AI-baseret anonymisering

Brug af on-premise model til at identificere følsomme data, som derefter maskeres. Her kan både anvendes sprogmodeller og specialiserede maskeringsmodeller. Der er en risiko for, at nogle personoplysninger ikke identificeres korrekt, hvis information står i en struktur, modellen ikke har set før, og derfor deles utilsigtet med cloud-udbyderen.

C2

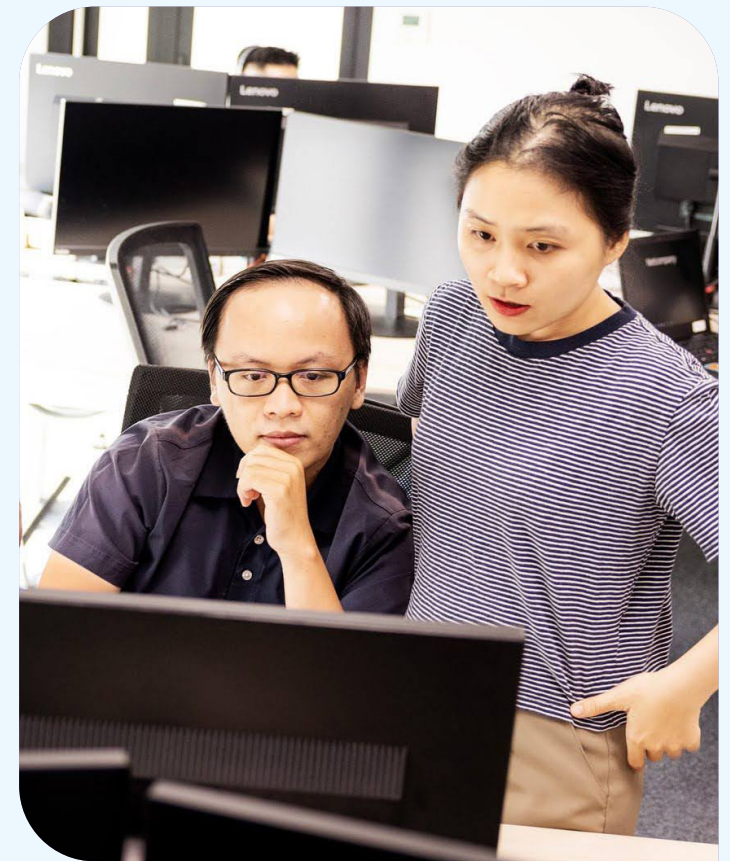
Systematisk datamaskering

Kendte, følsomme data i en forespørgsel erstattes med midlertidige værdier, før de behandles. Når behandlingen er afsluttet, genindsættes de oprindelige data i svaret. Data, der ikke er kendt på forhånd, kan ikke systematisk identificeres. Ved effektiv datamaskering, hvor informationer ikke kan spores tilbage til personer, undgår man behandling af personoplysninger.

C4

Dataminimering

En metode, hvor der sendes information relateret til en person eller dennes situation, uden at personoplysninger videreformidles. Dette kan være aldersgruppe, lovgivning relateret til personens situation eller vilkår for en ydelse. AI-assistenten modtager relevant information, men ingen individuelle personoplysninger. Dette muliggør, at en AI-assistent kan understøtte sensitive processer såsom sagshåndtering, uden at personoplysninger bliver delt med en cloud-baseret model. Dataminimering medfører, at kontekst fjernes fra en forespørgsel, hvilket kan gøre svar fra modellen mere abstrakte.



Tillid gennem AI og datasikkerhed

En organisation har pligt til at dokumentere tilgangen til databehandling. Ved at implementere robuste sikkerhedsforanstaltninger, overholde databeskyttelsesregler, håndtere risici samt implementere styrings- og forvaltningsprocesser omkring AI-assistentens data, skabes grundlaget for en tillidsfuld integration i organisationens processer og sikrer en værdifuld understøttelse uden kompromittering af datasikkerheden.

Forhold dig til de juridiske rammer



På baggrund af punkt 1, se afsnit 3.1, og den foretagne forretningsmæssige samt tekniske afklaring, skal en organisation først identificere de relevante juridiske krav, AI-assistenten skal overholde, før denne tages i brug. Som nævnt i punkt 1 skal dette skridt udføres løbende i AI-assistentens livscyklus, især ved ændringer i anvendelsen.

Der er særligt to regelsæt, som altid skal overvejes og tages højde for – databeskyttelsesforordningen (GDPR) [5] og forordningen om kunstig intelligens (AIA) [6] , som behandles i dette afsnit. Med henblik på konkretisering og operationalisering af relevante krav fra GDPR og AIA er der udarbejdet en tjekliste, der er vedlagt som Bilag A.

Afhængig af AI-assistentens konkrete brug

skal anden relevant lovgivning dog også tages i overvejelse. Her kan være tale om sundhedslovgivning, forbrugerbeskyttelseslovgivning, regler om tavshedspligt, finansiell lovgivning, immaterialret, forvaltningsregler, udbudsregler, NIS2 mv. Dette vil ligeledes omfatte eventuel lovgivning, som AI-assistenten skal besvare spørgsmål om, hvilket tillige skal identificeres og håndteres som en del af kortlægningen af de juridiske rammer for assistenten.

Tjeklisten i Bilag A kan både anvendes i tilfælde af egenudvikling, og AI-assistenten købes som standardløsning. Desuden kan tjeklisten anvendes, hvis man ønsker at sikre sig, at en assistent følger god praksis for ansvarlig udvikling og anvendelse af AI, uanset om assistenten er omfattet af AIA eller ej.

AIA

3.4.1

AIA gælder for ”AI-systemer” i henhold til forordningen. Det er forudsat, at AI-assistenten betragtes som et AI-system i AIA’s forstand, og det er desuden antaget, at AI-assistenten falder inden for AIA’s anvendelsesområde. Overordnet set er der på den baggrund to forhold, som skal afklares i relation til AIA:

³ AI-forordningens artikel 3, nr. 1): ” »AI-system«: et maskinbaseret system, som er udformet med henblik på at fungere med en varierende grad af autonomi, og som efter idriftsættelsen kan udvise en tilpasningsevne, og som til eksplicitte eller implicitte mål af det input, det modtager, udleder, hvordan det kan generere output såsom forudsigelser, indhold, anbefalinger eller beslutninger, som kan påvirke fysiske eller virtuelle miljøer” [6].

C1

Afdæk AI-assistentens risikoniveau, jf. punkt 3.4.1.1

C2

Afdæk egen rolle under AIA, jf. punkt 3.4.1.2

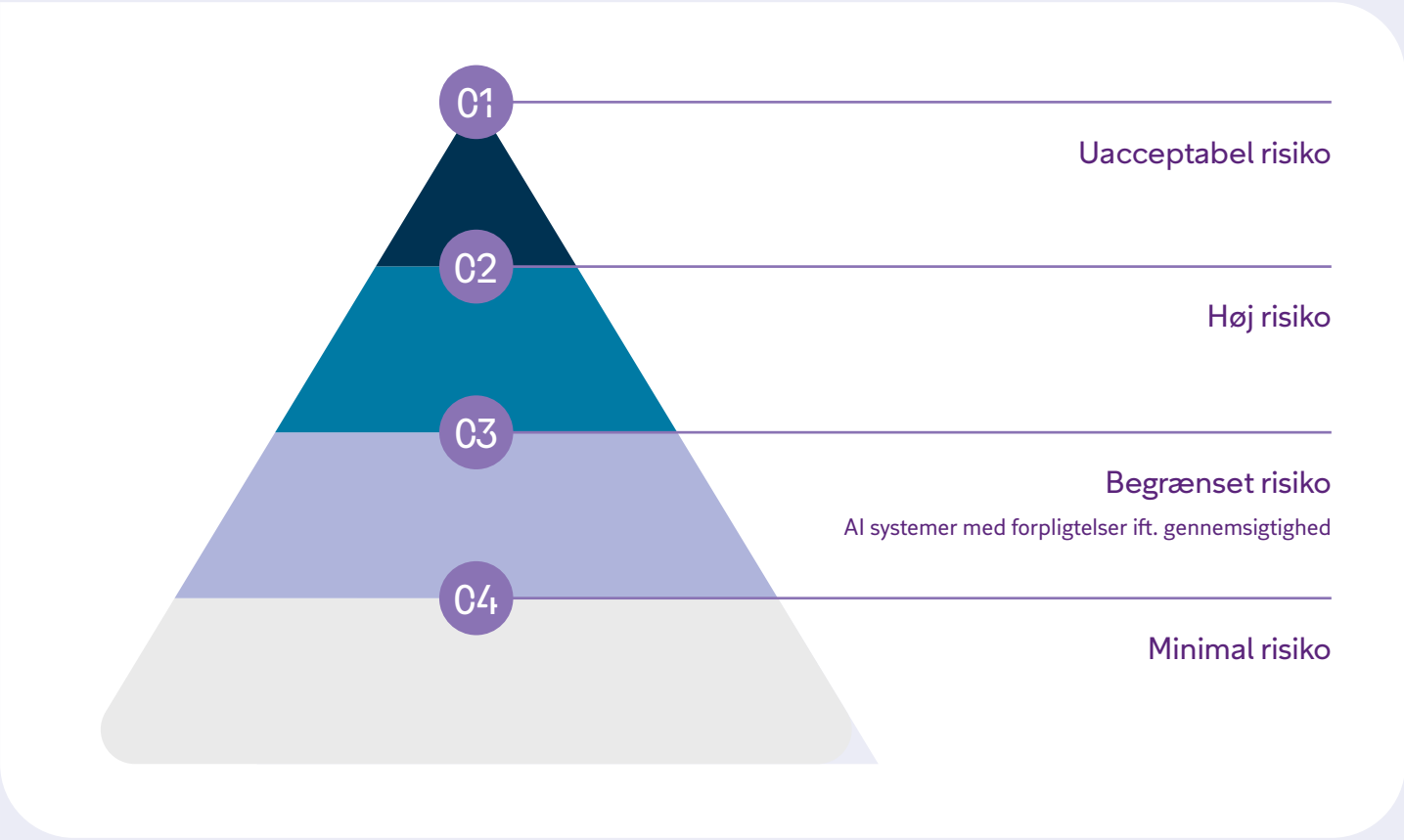
Disse forhold uddybes nedenfor. Der henvises samtidig til Bilag A, punkt 2, i forhold til kravene til hhv. udbydere og idriftsættere af højrisiko- AI-assisterter.

Afdæk AI-assistentens risikoniveau

3.4.1.1

AIA opererer med en risikobaseret tilgang, og der gælder derfor under AIA forskellige forpligtelser afhængig af risikoen ved AI-systemet og organisationens rolle.

Den risikobaserede tilgang i AIA er illustreret i Figur 3.4.



Niveau 01

Forbudte AI-systemer

Flere typer af AI-systemer er forbudt i henhold til AIA, da de anses for at udgøre en betydelig trussel mod menneskers sikkerhed og borgeres rettigheder eller værdighed. Eksempler inkluderer sociale scoringssystemer og stemmeassistent-teknologier, der fremmer farlig adfærd.

Det skal derfor undersøges og afklares, hvorvidt AI-assistenten anvendes på en måde, der er forbudt efter AIA, som hvis assistenten anvender manipulerende eller vildledende teknikker til at overtale en bruger til at træffe en bestemt beslutning til skade for brugeren.



Niveau 02

Højrisiko

Et AI-system betragtes som højrisiko, hvis følgende betingelser er opfyldt:

1. AI-systemet tilsigtes anvendt som en sikkerhedskomponent i et produkt, eller er i sig selv et produkt, der er omfattet af den EU-harmoniseringslovgivning, der er anført i bilag I til AIA.
2. Sikkerhedskomponenten eller produktet skal underkastes en overensstemmelsesvurdering foretaget af en tredjepart med henblik på omsætning eller ibrugtagning af produktet i henhold til den EU-harmoniseringslovgivning, der er anført i bilag I til AIA.

Herudover betragtes de AI-systemer, der er omtalt i Bilag III til AIA, også som højrisiko. En AI-assistent kan således falde ind under højrisiko-kategorien, hvis den f.eks. anvendes i:

- kritisk infrastruktur (f.eks. transport), der kan bringe borgernes liv og sundhed i fare.
- uddannelse eller erhvervsuddannelse, der kan være afgørende for en persons adgang til uddannelse og erhvervsforløb (f.eks. bedømmelse af eksamener).
- beskæftigelse, forvaltning af arbejdstagere og adgang til selvstændig erhvervsvirksomhed (f.eks. CV-sorteringssoftware til ansættelsesprocedurer).
- væsentlige private og offentlige tjenester (f.eks. kreditvurdering, der nægter borgerne mulighed for at opnå et lån).

- retshåndhævelse, der kan gribe ind i menneskers grundlæggende rettigheder (f.eks. evaluering af dokumentationens pålidelighed)

Hvis en AI-assistent anses for at være et højrisiko-AI-system, er denne underlagt strenge forpligtelser, før den kan bringes i omsætning. Kravene er som følger:

- Passende risikovurderings- og risikobegrænsningssystemer.
- Høj kvalitet af de datasæt, der leverer data til systemet, for at minimere risici og diskriminerende resultater.
- Logning af aktivitet for at sikre sporbarhed af resultater.
- Detaljeret dokumentation af alle de nødvendige oplysninger om systemet og dets formål, så myndighederne kan vurdere, om det er i overensstemmelse med kravene.
- Klare og tilstrækkelige oplysninger til idriftsætteren.
- Passende foranstaltninger til menneskeligt tilsyn for at minimere risikoen.
- Høj grad af robusthed, sikkerhed og nøjagtighed.

Der henvises til tjeklisten i Bilag A, punkt 1.2.1 og 1.2.2, for nærmere angående krav relateret til højrisiko-assistenter.

Niveau 03

Begrænset risiko

Begrænset risiko henviser til de risici, der er forbundet med manglende gennemsigtighed i anvendelsen af kunstig intelligens. AIA indeholder specifikke gennemsigtighedsforpligtelser for at sikre, at mennesker informeres, når det er nødvendigt, og dermed fremmer tilliden. Når der f.eks. anvendes AI-systemer såsom chatbots, bør brugere gøres opmærksomme på, at de interagerer med en maskine, så de kan træffe en informeret beslutning om at fortsætte eller træde et skridt tilbage. Udbydere skal også sikre, at AI-genereret indhold kan identificeres. Desuden skal AI-genereret tekst, der offentliggøres med det formål at informere offentligheden om spørgsmål af offentlig interesse, mærkes som kunstigt genereret. Dette gælder også for lyd- og videoindhold, der udgør deep fakes⁴.

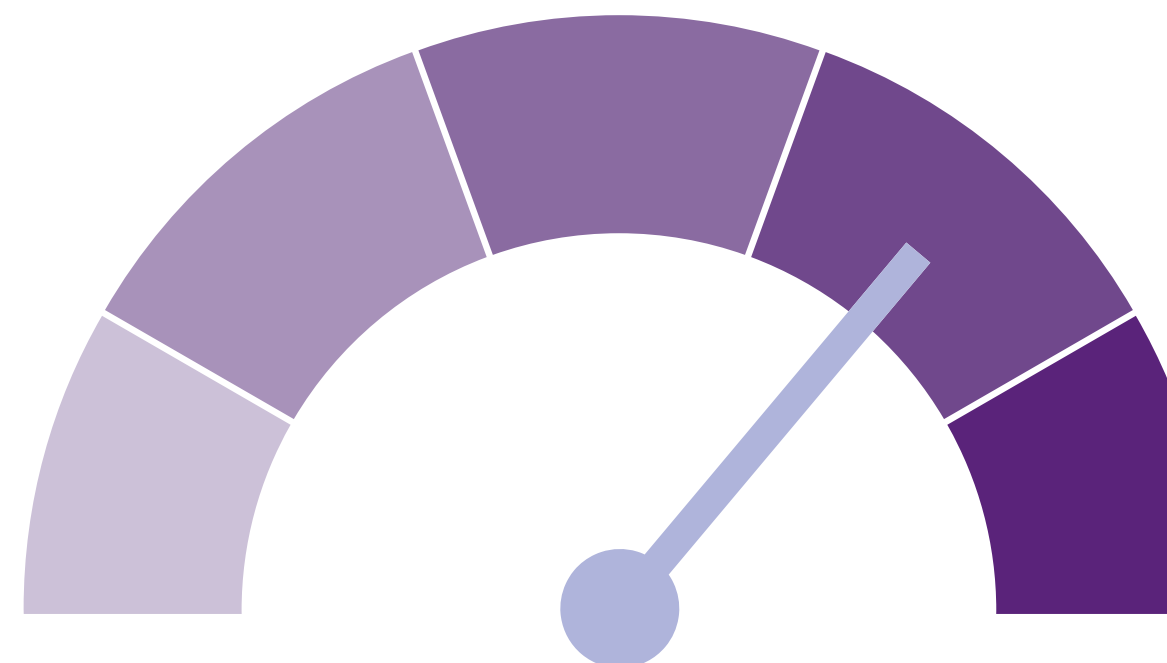
Assistenter, der falder ind under kategorien begrænset risiko, skal derfor overholde AIA's artikel 50.

⁴ AI-generede billeder, lyd eller video, der angiver en eksisterende person, genstand, sted eller anden form for entitet eller begivenhed, der er falske, men fremstår som autentiske samt sandfærdige [1].

Niveau 04

Minimal eller ingen risiko

AIA giver mulighed for fri anvendelse af kunstig intelligens med minimal risiko. Dette omfatter applikationer som AI-aktiverede videospil eller spamfiltre. Langt størstedelen af de AI-systemer, der i øjeblikket anvendes i EU, falder ind under denne kategori.



3.4.1.2

Afdæk din rolle under AIA

Omfanget og typen af dine forpligtelser under AIA afhænger af din rolle. Der er flere roller under forordningen, men typisk vil spørgsmålet gå på, hvorvidt din organisation i relation til AI-assistenten er en udbyder eller idriftsætter⁵:

⁵ Definitioner angivet i dette afsnit fremgår af AIA, artikel 3 [6].



Udbyder

En fysisk eller juridisk person, en offentlig myndighed, et agentur eller et andet organ, der udvikler eller får udviklet et AI-system eller en AI-model til almen brug og bringer dem i omsætning eller ibrugtager AI-systemet under eget navn eller varemærke, enten mod betaling eller gratis.



Idriftsætter

En fysisk eller juridisk person, en offentlig myndighed, et agentur eller et andet organ, der anvender et AI-system under sin myndighed, medmindre AI-systemet anvendes som led i en personlig ikkeerhvervsmæssig aktivitet.



Langt de fleste af forpligtelserne efter AIA påhviler udbydere af højrisiko AI-systemer (Bilag A, punkt 1.2.1). Du skal være opmærksom på, at en række forpligtelser påhviler andre led i AI-systemernes værdikæde, f.eks. udbydere af:

01

AI-model til almen brug

En AI-model, herunder en model der er trænet med store mængder data ved hjælp af selvovervågning i stor skala, udviser betydelig generalitet og har evnen til at udføre en lang række forskellige opgaver, uanset hvordan modellen bringes i omsætning. Desuden kan den integreres i forskellige downstream-systemer eller -applikationer, med undtagelse af AI-modeller, der anvendes til forsknings-, udviklings- og prototypeaktiviteter, inden de tages i brug.

02

AI-system til almen brug

Et AI-system, som er baseret på en AI-model til almen brug, og som har kapacitet til at opfylde en række forskellige formål, både til direkte anvendelse og til integration i andre AI-systemer.

03

Downstream

En udbyder af et AI-system, herunder et AI-system til almen brug, som integrerer en AI-model, uanset om AI-modellen leveres af udbyderen selv og integreres vertikalt eller leveres af en anden enhed på grundlag af kontraktforhold.”

Når du ”finetuner⁶” en sprogmodel, der understøtter en AI-assistent, vil det således være relevant at overveje, hvorvidt du er en udbyder af et AI-system til almen brug, eller om du er en udbyder af et høj risiko AI-system. Du skal derfor kende forpligtelserne i AI systemernes værdikæde som beskrevet i AIA artikel 25 og vurdere under hvilken led i værdikæden en assistent falder ind under.

⁶ Videre træning af en model ovenpå et konkret dataset med formålet at specialisere modellen.

3.4.2

GDPR

GDPR finder anvendelse, hvis der behandles personoplysninger ved udvikling og/eller anvendelse af AI-assistenten. Hvis man som organisation anvender en AI-assistent, som man enten selv har udviklet eller købt som en standardydelse (SaaS-løsning), vil man som oftest være dataansvarlig for behandlingen af personoplysningerne i AI-assistenten. Du skal derfor være opmærksom på den dataansvarliges forpligtelser i GDPR.

Nedenfor gennemgås de typiske krav og temaer, der er i spil, når man som dataansvarlig behandler personoplysninger ved udvikling/anvendelse af en assistent.

Få et overblik over behandlingen af personoplysninger

Du skal starte med at få et overblik over databehandlingen, herunder hvilke personoplysninger der behandles om hvilke personer til hvilke formål i hele løsningens livscyklus, herunder både ved udvikling, test og drift/gentræning af løsningen. Det gælder både inputdata, herunder eventuelle personoplysninger i prompts, samt outputdata. Vær opmærksom på, om brugen af assistenten indebærer, at der i dens outputdata er generet nye personoplysninger; det kan f.eks. være, at assistenten gennem profilering udleder følsomme personoplysninger om brugeren eller øvrige personer.

Håndter databeskyttelsesreglerne i hele løsningens livscyklus

Det er et grundlæggende krav, at den dataansvarlige skal kunne dokumentere overholdelse af databeskyttelsesreglerne. Det følger af kravet om databeskyttelse gennem design og gennem standardindstillinger i GDPR-artikel 25, at den dataansvarlige skal indtænke databeskyttelsesreglerne og håndtere risici igennem hele AI-assistentens livscyklus, dvs. allerede fra designfasen og frem til overvågning af løsningen i drift. Dette gælder uanset,

om der er tale om standardløsninger ("off the shelf") eller løsninger, der er udviklet specielt til en bestemt opgave hos den dataansvarlige.

Ved behandling af personoplysninger ved udvikling og drift af AI-assistenten skal de grundlæggende principper om bl.a. lovmæssighed, rimelighed og gennemsigtighed, formålsbegrænsning, dataminimering og datakvalitet i GDPR-artikel 5 overholdes. Du skal også sørge for, at der er behandlingsgrundlag (hjemmel) for behandlingen af personoplysninger. Virksomheder, der er dataansvarlige, vil ofte skulle overveje brug af samtykke fra de registrerede, kontrakt eller brug af reglen om legitim interesse ved behandling af ikke-følsomme personoplysninger efter GDPR-artikel 6. Offentlige myndigheder vil ofte skulle finde hjemmel-sgrundlaget i GDPR-artikel 6, stk. 1, litra e, om behandling af personoplysninger som led i offentlig myndighedsudøvelse eller en opgave i samfundets interesse samt særlov-givningen på det område, som assistenten skal anvendes på. Jo mere indgribende behandlingen er for de registrerede – dvs. jo mere behandlingen har indvirkning på borgerens økonomiske, uddannelsesmæssige, sociale, sundhedsmæssige eller lignende forhold – jo større krav er der til hjemmel-sgrundlagets klarhed. Hvis der behandles følsomme personoplysninger, skal der være hjemmel i GDPR-artikel 9, stk. 2, jf. GDPR-artikel 6.

Underret de registrerede om behandlingen af personoplysninger og håndter de registreredes rettigheder

Som udgangspunkt skal de registrerede underrettes om behandlingen af personoplysninger, som AI-assistenten indebærer, jf. GDPR-artikel 13-14. Dette kan ske i en privatlivspolitik eller på anden måde. Bemærk, at udvikling og drift af AI-assistenten skal anses for to separate formål, og de registrerede skal således underrettes om begge formål.



De registrerede skal også informeres om tilstedeværelse af profilering og konsekvenserne heraf. Og hvis der med AI-assistenten foretages automatiske, individuelle afgørelser efter GDPR-artikel 22, skal de registrerede underrettes herom og som minimum modtage meningsfulde oplysninger om logikken heri samt betydningen og de forventede konsekvenser af en sådan behandling for de registrerede.

Du skal også kunne håndtere de registreredes rettigheder, herunder bl.a. give indsigt i behandlingen af personoplysninger og kunne efterleve berigtigelsesretten, indsigelsesretten, retten til sletning m.v.

Vær særligt opmærksom på, om du bruger AI-assistenten til at lave automatiske individuelle afgørelser

Der gælder som udgangspunkt et forbud mod anvendelse af automatiske, individuelle afgørelser, jf. GDPR-artikel 22. Sådanne afgørelser kan dog efter bestemmelsen ske lovligt, hvis undtagelserne til forbuddet er opfyldt, herunder hvis der indhentes udtrykkeligt samtykke, behandlingen er nødvendig for indgåelse eller opfyldelse af en kontrakt mellem den registrerede og den dataansvarlige eller hvis behandlingen er hjemlet i EU-ret eller national ret. Der gælder dog i visse situationer ret for de registrerede til at kunne få menneskelig indgriben i afgørelsen m.v.



GDPR



Få styr på sikkerheden og håndter risici effektivt

Behandlingen af personoplysninger ved udvikling og anvendelse af AI-assistenten skal leve op til de grundlæggende krav om behandlingssikkerhed. Der skal derfor fastsættes passende organisatoriske og tekniske sikkerhedsforanstaltninger i hele løsningens livscyklus, og dette skal ske med udgangspunkt i en konkret risikovurdering vedrørende de registreredes rettigheder og frihedsrettigheder, jf. GDPR-artikel 32. Ved udarbejdelsen af risikovurderingen skal der bl.a. tages udgangspunkt i formålet med brugen af AI-assistenten og dennes begrænsninger samt de involverede personoplysningers karakter. Det vil f.eks. være mere risikofyldt at anvende AI-assistenten som beslutningsstøtte til at træffe afgørelser eller foretage profilering af borgere, kunder eller medarbejdere end at give f.eks. medarbejdere mulighed for at bruge AI-assistenten til at få svar på spørgsmål interne regler på intranettet eller anvende AI-assistenten til at producere indhold til markedsføringsmateriale.

Hvis behandlingen af personoplysninger sandsynligvis medfører en høj risiko for de registre-

redes rettigheder og frihedsrettigheder, skal du forud for behandlingen af personoplysninger udarbejde en konsekvensanalyse vedrørende databeskyttelse (DPIA) efter GDPR-artikel 35. En sådan DPIA er navnlig relevant ved behandling af personoplysninger ved brug af ny teknologi, og det vil normalt være tilfældet ved udvikling og brug af AI-løsninger. Det gælder navnlig, hvis der sker behandling af følsomme personoplysninger i stort omfang, behandling af personoplysninger om sårbare registrerede eller brug af profilering eller automatiske, individuelle afgørelser.

Følgende er blandt de typiske risici ved udvikling og anvendelse af AI-assistenter, som man skal håndtere:

- Faktuelt forkerte svar og hallucinationer, dvs. hvor AI-assistenten kommer med forkerte eller opdigtede svar (output).
- Usaglig forskelsbehandling (bias), dvs. hvor AI-assistenten medfører ulovlig forskelsbehandling på baggrund af et eller flere sensitive kriterier som f.eks. køn eller etnicitet.

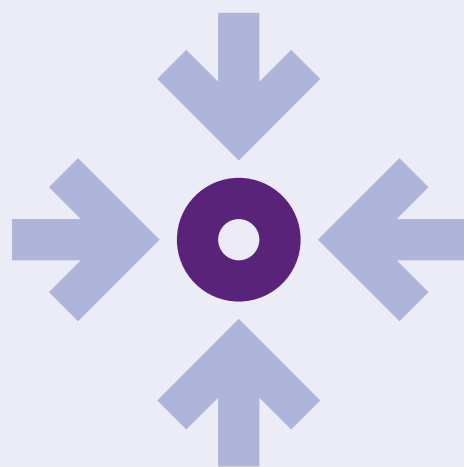
- Automatiseringspartiskhed (automation bias), dvs. hvor de mennesker, der anvender AI-assistenten, ukritisk følger eller stoler på AI-assistentens output eller lægger det til grund for en afgørelse, fordi de stoler på løsningen.
- Utilstrækkelig behandlingssikkerhed, dvs. hvor uvedkommende får adgang til personoplysninger eller manipulerer med AI-assistenten, så den virker ukorrekt.

Det er derfor også vigtigt, at der udarbejdes og gives de nødvendige instrukser og oplysninger til medarbejdere om, hvordan de skal anvende AI-assistenten korrekt og sikkert, så de kender til AI-assistentens formål og begrænsninger, lovlig brug og involverede risici. Det kan også være relevant at foretage kvalitetskontroller af AI-assistentens output m.v., ligesom det kan overvejes at stille krav om, at medarbejdere tydeligt skal markere, når et output er genereret af AI-assistenten.

Overførsler til tredjelande og deling af personoplysninger med leverandøren

Assistenter, der leveres som SaaS-løsninger i et cloudmiljø, kan indebære, at der sker overførsler af personoplysninger til usikre tredjelande, herunder til underleverandører. Du skal derfor kortlægge, om der sker sådanne overførsler af personoplysninger til usikre tredjelande, og i så fald vurdere, om dette lovligt kan ske inden for databeskyttelsesreglerne. I praksis betyder dette, at der skal udarbejdes en såkaldt Transfer Impact Assessment (TIA), der kortlægger overførslerne samt vurderer lovligheden heraf.

Det vil også være relevant at undersøge, om du ved brugen af løsningen videregiver personoplysninger til leverandøren af løsningen til brug for leverandørens egne formål, f.eks. træning af løsningen, produktudvikling, markedsføring etc. I så fald skal du vurdere, om der er hjemmel til denne videregivelse, samt om leverandørens brug af personoplysningerne til egne formål er forenelig med det oprindelige formål med behandlingen af personoplysningerne, jf. GDPR-artikel 5, stk. 1, litra b [5].



3.5

Afgræns Al-assistentens evner og ansvar

Al-assisterter skal afgrænses til konkrete, veldefinerede opgaver, ansvarsområder og værktøjer, så risikoen for fejl og uhensigtsmæssig brug minimeres. Dette er især vigtigt, når flere specialiserede Al-assisterter skal samarbejde, og opgaver automatisk fordeles mellem dem. Værktøjer kan udvide Al-assistentens muligheder fra blot sproglig transformation til mere komplekse opgaver. Mens de underliggende sprogmodeller håndterer naturligt sprog, kan en Al-assistent for eksempel udføre kvantitative analyser, hvis den kombineres med de rette værktøjer eller modeller.

Veldefinerede opgaver fremmer også systematisk monitorering og forbedring af Al-assistenternes ydelse og pålidelighed. Modne modeller og Al-platforme kan allerede have værktøjer til at understøtte afgrænsning, og organisationen bør undersøge, hvordan de kan bygge ovenpå eksisterende funktionalitet.

Der findes mange metoder til at afgrænse Al-assistenters evner og ansvar. Her er nogle af de mest udbredte.

Prompt engineering

Præcise instruktioner, der målretter Al-assistentens arbejde inden for dens ansvars-område, sikrer effektiv løsning af komplekse opgaver. Prompt engineering kan dog ikke alene afgrænse Al-assistenten, da denne kan ignorere en sådan afgrænsning med angreb som prompt injection. Prompt engineering tilbyder forskellige værktøjer, der håndteres sammen med Al-plattformen, hvor følgende er gængse metoder:

Agent

En dynamisk form for prompt chaining, hvor Al-assistenten kan benytte sig af mere specialiserede Al-assisterter, funktioner eller andre værktøjer, kaldet kompetencer. Al-assistenten analyserer forespørgslen og identificerer så, hvilken samling af kompetencer der skal løse opgaven. Falder en opgave ikke ind under en af disse kompetencer, må Al-assistenten sige fra.

Struktureret prompting

Al-assistenten skal returnere et svar i et bestemt struktureret, foruddefineret format. Dette giver mere ensartede, maskinlæsbare resultater, der efterfølgende kan valideres og benyttes i andre sammenhænge. F.eks. function calling, hvor Al-assistentens strukturerede svar sendes videre til en funktion eller som en forespørgsel til et eksternt system.

Single— og few-shot prompting

Al-assistenten får et til flere eksempler på det ønskede svar til at løse opgaven. Dette giver modellen kontekst at arbejde ud fra, hvilket kan forbedre præcisionen for visse opgavetyper [8].

Prompt chaining

Nedbryder komplekse opgaver i simple trin, som løses et ad gangen. Da løsningen af hvert trin bygger videre på resultaterne af de foregående, op-bygger Al-assistenten en dybere forståelse af opgaven. Dette ansporer Al-assistenten til et mere gennemtænkt, nuanceret og præcist resultat.

Klassificering af forespørgsler

Brugeres forespørgsler bliver klassificeret ind i en liste af godkendte kategorier, Al-assistenten må behandle eller forbudte kategorier, den ikke må behandle. Klassificeringen vil typisk blive udført med en vis sikkerhed, hvorfor visse forespørgsler kan risikere stadig at komme igennem til modellen. Klassificering er typisk hurtig at udføre og kræver mindre og mere specialiserede modeller end sprogmodeller.

Sikkerhedsmodeller

Al-assisterter skal betragtes som en systembruger i organisationens systemlandskab og må derfor kun få adgang til data og systemer på vegne af autoriserede brugere. Adgangen skal begrænses til relevante data. Når Al-assisterter arbejder i borgerspecifikke kontekster, skal deres funktioner og ansvar også baseres på den adgang, brugerne har som brugere.



En tydelig definition af opgaver, afgrænsning af ansvarsområder og adgang til tilgængelig baggrundsinformation fremmer ikke kun præcision, men øger også tilliden til Al-assistenten.

Opbyg struktureret kvalitetssikring

Kvalitetssikring er essentiel både under udvikling og implementering af sikre og ansvarlige AI-assistenter. Sprogmodellers svar er ikke deterministisk. Det betyder, at deres svar kan variere, selvom det samme spørgsmål gentages, hvilket gør kvalitetssikringsprocessen særligt udfordrende.

Effektiv kvalitetssikring kræver derfor en struktureret tilgang, der omfatter både en udviklingsfase, en pilotfase og løbende monitorering. Ud over traditionelle testmetoder, som funktionelle tests og ydelsestests, er der behov for metoder designet specielt til at håndtere AI-assistenternes adfærdsmønstre, så de ikke kun er teknisk velfungerende, men også leverer værdi til organisationen på en ansvarlig måde. Som med andre systemer skal almindelige tests, som funktionelle test eller ydelsestests, stadig vedligeholdes, særligt i forbindelse med AI-assistentens funktionelle værktøjer.

Dette punkt fokuserer på en række metoder specifikke for AI-assistenter, som organisationer bør overveje, når de kvalitetssikrer AI-assistenter

C1

Forretningsdrevet Design

En AI-assistent bør designes i tæt samarbejde mellem forretningseksperter og udviklere. Forretningseksperterne opstiller kriterier for, hvordan AI-assistenten bør svare i forskellige scenarier. Udviklerne omsætter kriterierne til et teknisk design, der bør inkludere prompts, adgang til relevante videnskilder og integrationer med forretningssystemer. Slutbrugere kan inddrages tidligt i processen ved at teste forskellige kombinationer af scenarier, prompts og baggrundsinformation på modeludbydere-nes hjemmesider.

Hvis udviklerne alene designer AI-assistenten uden at indsamle viden fra forretningen, risikerer organisationen, at AI-assistenten kun fungerer i en begrænset virkelighed baseret på udvikler-nes forståelse. Dette kan føre til, at AI-assistenten ikke kan håndtere domænespecifikke udfordringer, som udviklerne ikke kender til, og i værste fald at AI-assistenten foretager forkerte beslutninger eller giver mangelfulde svar.

C2

Red Teaming

En tilgang, hvor eksperter med teknisk og forretningsmæssig indsigt i AI-assistenten forebyggende identificerer risici for skadelige resultater ved brug af AI-assistenten. Dette kan f.eks. være, at AI-assistenten giver løfter, som organisationens ikke kan overholde eller deler data, der ikke skal deles. Der er to tilgange til Red Teaming:

Neutral

Eksperten tester den funktionalitet, som AI-assistenten er udviklet til, ud fra forventet brug. Risici, der identificeres gennem den neutrale tilgang, er kritiske, da de kan forekomme hyppigere sammenlignet med risici identificeret gennem den fjendtlige tilgang.

Fjendtlig

Eksperten tester aktivt AI-assistenten for at identificere risici. Alle kneb gælder, og eksperten skal identificere både sandsynlige og usandsynlige risici.

C3

Pilotforløb

Efter en AI-assistent er udviklet, men før den brede udrulning i organisationen sker, køres en pilot med en gruppe forretnings-eksperter, der afprøver AI-assistenten i deres reelle arbejde. Eksperterne giver jævnligt udviklerne feedback på AI-assistenten, så de kan forbedre kvaliteten af dens resultater. Sprogmodeller er ikke-deterministiske, og specielt for AI-assistenter, der tilgås gennem en chat, er der stor risiko for, at designet ikke forudser alle udfald. Gennem pilotforløbet opbygges ligeledes en kernegruppe af superbrugere, der kan vejlede og supportere deres kolleger i brug af AI-assistenten. De kan hjælpe organisationen med den videre udrulning ved at supportere deres kollegaer. Pilotforløb kan køres med eksterne brugere som kunder, hvor en inviteret målgruppe tester og giver feedback på AI-assistenten. Da målgruppen typisk ikke kan give løbende aktiv feedback, vil disse forløb mere agere som dataindsamling.

05

Kontrol

For at støtte den løbende kvalitetsstyring skal organisationen overvåge AI-assistentens resultater, når den er i produktion. Kontrol kan udføres gennem forskellige metoder:

Manuel kontrol

En udvikler eller forretningsekspert gennemgår AI-assistentens svar manuelt for at identificere forkerte eller tvetydige svar. Manuel gennemgang kan anvendes i alle faser af en AI-assistents implementering, men bør især bruges, når AI-assistenten rulles ud, for at opdage uforudsete risici. Manuel kontrol drager fordel af ekspertdrevet indsigt og menneskelig intuition. Dog begrænser den store arbejdsbyrde og afhængigheden af eksperter effektiviteten af manuel kontrol. Over tid bør manuel kontrol automatiseres for at lette kvalitetssikringen, når AI-assistenten ændres.

Regelbaseret kontrol

Kontrol af AI-assistentens resultater på baggrund af definerbare regler. Hvis det forventede resultat for eksempel inkluderer en postadresse, kan regelbaseret kontrol anvendes til at sikre, at feltet er inkluderet og korrekt. En regelbaseret tilgang er transparent og giver mulighed for at bruge en procentsats til at måle kvaliteten. Kontrollen kræver løbende vedligeholdelse eller tilføjelser, når AI-assistentens funktionelitet udvides, og de kan ikke identificere skjulte eller abstrakte usikkerheder.

Model-baseret validering

AI-assistentens resultater valideres, inden de præsenteres for brugerne, typisk ved hjælp af en sprogmodel eller en anden AI-model. Ydeevnen for AI-assistenten falder, da brugeren må vente på, at resultaterne bliver valideret. Til gengæld reduceres risikoen for fejl, promptmanipulation eller hallucination, og modellen kan identificere ukendte og abstrakte usikkerheder. For visse kritiske forretningsprocesser er præcision vigtigere end ventetid eller omkostninger, hvilket gør automatisk validering til en foretrukken løsning. Modelbaseret validering kan også anvendes uden brugerinvolvering som en del af testningen, hvor en række spørgsmål og svar automatisk behandles af modeller for at sikre, at løsningen opretholder kvaliteten, inden en ny version lanceres.

06

Struktureret dokumentation

Det er kritisk, at AI-assistenten og processen for kvalitets-sikring er formaliseret og veldokumenteret. Følgende typer dokumenter bør indgå i dokumentationen:

Funktionelt design

Beskriver AI-assistentens ansvar, rolle og formål på designniveau med udgangspunkt i AI-assistentens brugsscenarie.

Brugerguide

Beskriver hvordan AI-assistenten benyttes samt begrænsningerne for denne.

Juridisk dokumentation

Beskriver databehandlingen samt de nødvendige risikovurderinger, der skal udføres på baggrund af GDPR og AIA. Se afsnit 3.4 ift. de juridiske rammer.

Teknisk design

Beskriver AI-assistentens tekniske opbygning, dernæst prompts, interne funktioner, samt hvordan den integrerer med andre systemer.

Sikkerhedsdokumentation

Beskriver sikkerhedsforanstaltninger, herunder adgangskontrol og procedurer for håndtering af sikkerhedshændelser. Dokumentationen bør også omfatte en vurdering af potentielle sårbarheder og tiltag til at afbøde disse.

Test-design

Dokumenterer, at AI-assistenten agerer hensigtsmæssig i forhold til krav og design. Da sprogmodeller ikke er deterministiske, skal tests fokusere på, hvad det forventede resultat forventes at indeholde.

Operationel dokumentation

Beskriver procedurer for daglig drift som overvågning og rutiner for opdateringer og vedligeholdelse af AI-assistenten.

Den underliggende sprogmodel, AI-assistenten benytter, skal ligeledes være dokumenteret. Dokumentationen bør vise, hvordan der sikres, at modellen arbejder med information gjort tilgængeligt i dens forespørgsel, og hvordan bias er håndteret. Benyttes en eksisterende model, bør udviklerne af denne have den nødvendige dokumentation.

Mål og opbevar relevant data om AI-assistentens brug

For at sikre gennemsigtighed og overholdelse af krav er det nødvendigt at etablere en detaljeret dataaudit og -lineage. Målet er en opsætning, som skal kunne logge alle interaktioner og handlinger udført af AI-assisterter og give fuld gennemsigtighed i forhold til dataflow og AI-assistentens beslutningstagning.

Data audit

Processen for logning og overvågning af alle relevante dataaktiviteter. Dette gør det muligt at dokumentere, hvordan data er blevet anvendt, hvem der har haft adgang, og hvilke data beslutningerne er baseret på.

Data lineage

Kortlægning af datas bevægelse fra oprindelse til anvendelse. Sporingen giver indsigt i, hvilke data der er blevet brugt, og hvordan de er blevet transformeret og behandlet undervejs. Når RAG benyttes, kan det f.eks. være relevant at logge, hvad RAG-processen introducerede på en given forespørgsel.

Når det kommer til data audit og logning af AI-assistenters aktiviteter, er det vigtigt at skelne mellem to typer af logning:

Generel logning

Anvendes når AI-assistenten ikke har adgang til personoplysninger. Denne type logning fokuserer på anonyme, overordnede interaktioner.

Kontekstspecifik logning

Anvendes når AI-assistenten har adgang til personoplysninger. Kontekstspecifik logning er mere detaljeret og knytter aktiviteter til specifikke borgere.



I en generel kontekst, hvor AI-assistenten ikke har adgang til personoplysninger, behandles eventuelle borgere anonymt. Dette inkluderer typisk logning af forespørgsler, AI-assistentens svar og konteksten bag disse svar, uden at knytte informationen til en bestemt borger. Hvis AI-assistenten derimod arbejder med en borgers personoplysninger, anvendes kontekstspecifik logning. I dette tilfælde skal logningen inkludere information om den specifikke borger, hvis oplysninger behandles.

Denne skelnen mellem generel og kontekstspecifik logning er ikke kun vigtig for sporbarheden, men også for at kunne påvise over for tilsynsmyndigheder og andre interessenter, at AI-assistenten opererer i overensstemmelse med relevant lovgivning.

Det er nødvendigt at have en centraliseret og sikker løsning på plads i platformen for at kunne sikre korrekt logning og håndtering af data. Løsningen bør have strenge adgangskontroller, så kun autoriseret personale kan tilgå og administrere logfilerne. Afhængigt af hvad AI-assistenten skal bruges til, kan det desuden være nødvendigt at logge data på et meget detaljeret niveau. For mange AI-assisterter betyder det, at man skal kunne lave rapporter, der viser, hvordan de bruges helt ned til hver enkelt samtale. Dette gør det muligt at følge og analysere, hvordan data bevæger sig, og hvordan beslutninger bliver taget.

Opbevaring og håndtering af de store mængder data kræver planlægning. Da mængden af loggede data kan være betydelig, bør der være klare regler, med udgangspunkt i GDPR og AIA, for hvor længe data skal opbevares. Ifølge AIA skal data opbevares i minimum seks måneder og derefter i en passende periode afhængigt af behov. Det er også vigtigt at sikre, at dataene opbevares på en måde, der beskytter dem mod uautoriseret adgang og samtidig gør det muligt at genfinde og analysere dem ved behov.

For at sikre den rette sikkerhed, skal der være dedikerede ressourcer til håndtering af problemer. Dette team skal kunne reagere hurtigt og effektivt på eventuelle sikkerhedsbrud eller dataproblemer, der måtte opstå. Incident-håndtering er essentiel for at kunne opretholde tilliden til AI-assisterterne både internt i organisationen og eksternt over for kunder og tilsynsmyndigheder.



Mål og opbevar relevant data om AI-assistentens brug

For at etablere en detaljeret dataaudit-opsætning anbefales relevante informationer, der dækker specifikke datapunkter.

Disse er præsenteret i Tabel 3.1.



Datapunkt	Beskrivelse	Begrundelse
AI-assistent	Den specifikke AI-assistent, der er tilknyttet de øvrige datapunkterelementer.	Det er nødvendigt for sporbarhed at identificere, hvilken AI-assistent der har behandlet en forespørgsel.
Bruger	De brugere, der har interageret med AI-assistenten. Brugere er menneskelige brugere, men også andre systemer eller AI-assisterter.	For at sikre sporbarhed og overholde krav til dataejerskab og adgangskontrol.
Session	Et overordnet dataelementpunkt, der binder andre elementer punkter sammen og registrerer sessionens start og slut.	Sessioner muliggør en helhedsorienteret forståelse af interaktioner og hændelser.
Forespørgsel	Den specifikke forespørgsel sendt til AI-assistenten. Følsomme oplysninger maskeres før logning.	Nødvendigt for at kunne spore, analysere og validere AI-assistentens adfærd og sikre overholdelse af GDPR.
Tilladelser	AI-assistentens tilladelser til at fuldføre relevante opgaver.	For at sikre at AI-assistenten er begrænset til kun at udføre handlinger, der ikke er relevante for den.
Data og kilde	Den data, som AI-assistenten har anvendt til at generere sit resultat.	Dokumentation af datakilder er vigtig for at forstå og verificere beslutningsgrundlaget.
Resultat	Det genererede output, som AI-assistenten producerer.	Da output fra sprogmodeller ikke er deterministisk, logges det for at muliggøre efterfølgende analyse.
Klassificering	Hvordan forespørgslen er klassificeret af AI-assistenten samt en forklaring af det genererede resultat. En forespørgsel kan gennemgå flere klassificeringer.	Klassificering giver indsigt i beslutningsprocessen og sikrer forståelighed i resultatet samt muliggør opfølgning ved fejl-klassificering.
Sprog	Hvilket sprog beskeden er skrevet på.	Sproget påvirker forståelsen og tolkningen af forespørgsler, hvilket er relevant for resultatets kvalitet.
Kontekst	Den kontekst eller prompt engineering, som AI-assistenten har været udsat for under produktionen af resultatet.	Dokumentation af kontekst er vigtig for at forstå og reproducere AI-assistentens adfærd.
Kvalitetssikring	Procedurer for kvalitetssikring af data og resultater. Hvis AI-assistenten udfører kvalitetssikring selv, logges dette.	Kvalitetssikring er afgørende for at sikre nøjagtighed og pålidelighed i AI-assistentens output.
Datoer og tidspunkter	Dato og tidspunkt for alle øvrige dataelementer.	Tidspunktet for hver handling er kritisk for korrekt audit og analyse.

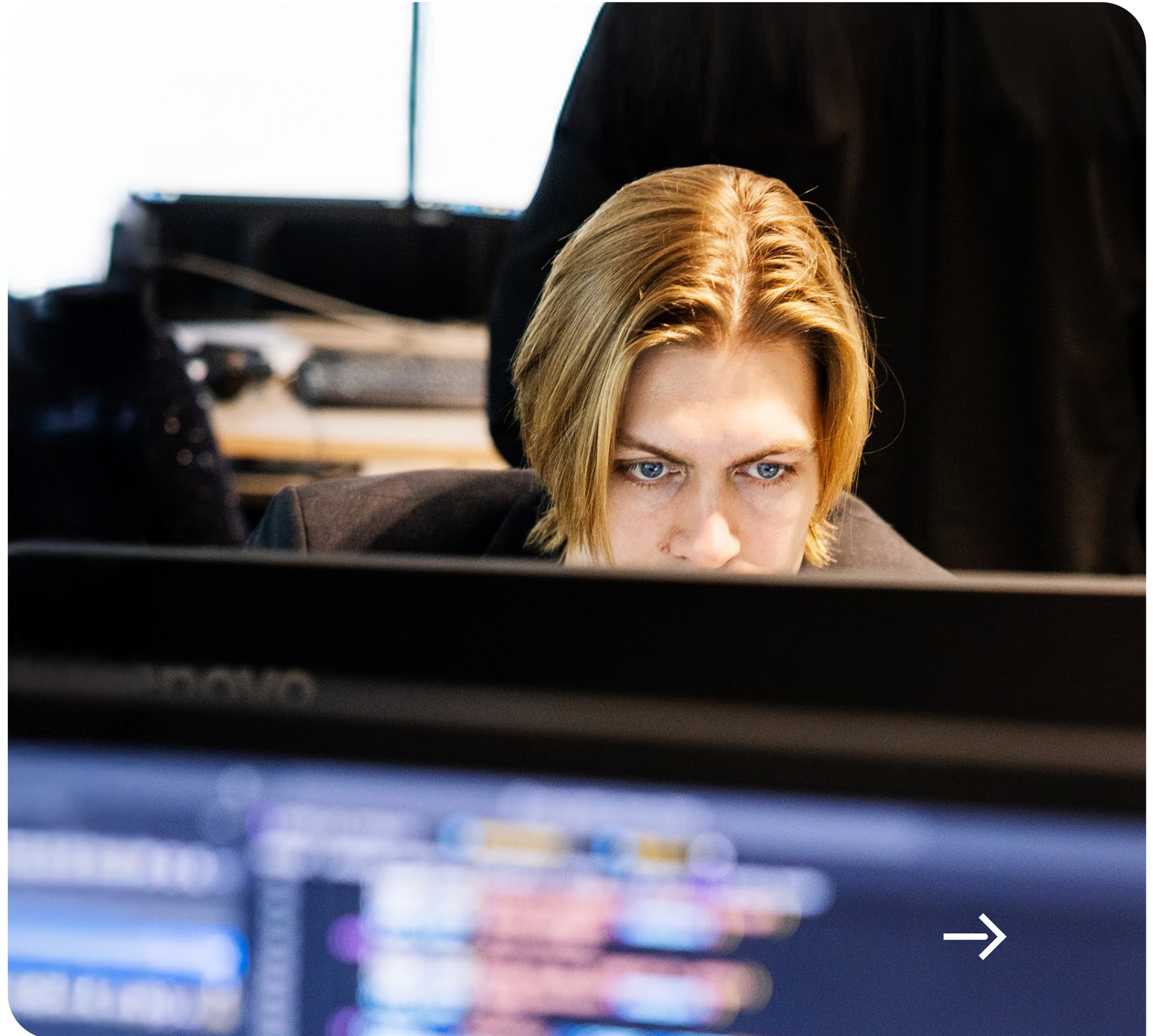
Tabel 3.1: Datapunkter der generelt er relevante at logge for AI-assisterter.

3.8

Planlæg organisatorisk implementering og uddannelse



Organisationer, der implementerer AI-assistenter, skal undervise og forberede deres medarbejdere på de forandringer, AI-assisterne medfører. Ifølge AIA er det en pligt at sikre, at medarbejderne bl.a. er informeret om, at de interagerer med AI-assistenter, samt at de modtager den nødvendige træning i at anvende AI-assisterne, med der er taget højde for deres tekniske kompetencer. Hvis organisationen og medarbejderne ikke forberedes tilstrækkeligt, kan der opstå modstand for forandringen.



Planlæg organisatorisk implementering og uddannelse

Dette afsnit præsenterer fire fokusområder for denne organisatoriske implementering (Figur 3.5). Afsnittet inkluderer ikke en udtømmende liste af organisatoriske implementeringsaktiviteter, da den nødvendige indsats afhænger af forskellige parametre, som det omfang hvori organisationen påvirkes, de konkrete målgrupper samt organisations forandringsparathed og AI-modenhed.

**01**

Forstå påvirkningen og lav en strategi

**03**

Uddan og byg kompetencer

**02**

Etabler en kernefortælling

**04**

Skab tillid og forankring i organisationen

Figur 3.5: Organisatoriske fokusområder ved implementering

C1

Forstå påvirkningen og lav en strategi

Det er nødvendigt at forstå, hvordan organisationen påvirkes af at en AI-assistent introduceres. Dette kan opnås ved at undersøge de forretningsgange og processer, der bliver påvirket samt at analysere hvilke medarbejdergrupper, der vil blive berørt. Implementeringen af AI-assistenten påvirker både de medarbejdere, som skal anvende den, IT-medarbejderne, som skal vedligeholde den, og de eventuelle borgere, der skal interagere med AI-assistenten.

Ved at udføre proces-, påvirknings- og målgruppeanalyser kan man afdække, hvordan organisationen skal ændre sin måde at arbejde på, og hvordan processer og arbejdsgange vil ændre

sig. Dette hjælper med at identificere behovet for nye arbejds-metoder, viden og kompetencer. Når omfanget af transformati-onen er klar, skal organisationen etablere en implementerings-strategi og -plan. Denne plan skal sikre en glidende overgang, så organisationen er klar til at tage AI-assistenten i brug ved lancering. Den skal tage højde for indsatsområder som uddan-nelse, træning og kommunikation, som alle er nødvendige for at klæde medarbejderne på til at anvende AI-assistenten.

C2

Etabler en kernefortælling

Det næste fokusområde er at etablere en klar, struktureret og gennemsigtig fortælling om implementeringen af AI-assistenten. En klar fortælling er med til at formidle essentielle kommunikationsbudskaber som;

Hvorfor

Skab en samlet forståelse af, hvorfor man ønsker at implementere AI-assistenten, og hvilken værdi den skal tilføre organisationen. Det er f.eks. effektivisering af arbejds-gange og eliminering af manuelt arbejde.

Hvornår og hvordan

Informer om processen, som organisationen skal igennem – f.eks. hvornår og hvor-dan medarbejderne vil få træning i at bruge AI-assistenten.

Hvad betyder det for medarbejderen

Skab overblik over hvad ibrugtagen af AI-assistenten betyder for medarbejderen, og hvad medarbejderen kan forvente.

Budskaberne kan bygge på brugsscenarioet defineret i afsnit 3.1. Fortællingen anvendes til at sikre, at de rigtige budskaber om AI-assistenten får taletid. Dette skaber opmærksomhed, buy-in og support i organisationen. Hvis fortællingen ikke præsenteres korrekt, risikerer man, at der opstår bekymring og forvirring om forandringen, hvilket kan lede til modstand. F.eks. ses det, at de medarbejdere, der skal anvende AI-assistenten, er bekymrede for, at AI-assistenten vil erstatte dem, og derved bliver medarbejderne modstandere af forandringen. I stedet bør medarbejderne se AI-assistenten som en virtuel kollega, der udfører det kedelige arbejde, så der for medarbejderen frigives tid til andre opgaver.

C3

Uddan og byg kompetencer

Uddannelse i at anvende AI-assistenten er centralt for at lykkes med implementeringen, og etableringen af træning sikrer, at medarbejderne opbygger kompetencer til at bruge AI-assistenten og forstå deres nye arbejdsgange. Dette gør sig gældende både for medarbejderne, der skal anvende AI-assistenten samt medarbejderne i IT, der skal vedligeholde AI-assistenten.

Som en del af træningen er det vigtigt at medarbejderne introduceres til, hvordan AI fungerer. Dette skaber indsigt i, hvordan AI-assisterter kan bibringe værdi i organisationen. Dernæst kan superbrugere med fordel anvendes. Superbrugere er en gruppe af særligt udvalgte medarbejdere, der oplæres i at uddanne resten af organisationen og være fortalere for forandringen. Et netværk af superbruger på tværs af organisationen bidrager til en solid forankring af og fællesskab omkring AI-assistenten i organisationen og er med til at øge ibrugtagningen af assistenten.

For at opbygge kompetencer i at bruge AI-assistenten anbefales praktisk træning med konkrete øvelser. Medarbejderne vil herved få erfaring i at anvende AI-assistenten og vide, hvordan

de agerer. Træningsmetoden afhænger dog af AI-assistentens kompleksitet, da nogle opgaver vil kræve mere interaktion med AI-assistenten end andre. Medarbejderen skal ligeledes have en overordnet forståelse for den information, som AI-assistenten trækker på således, at de ved, hvad der ligger til grund for de svar, den giver. Hvis AI-assistentens resultater benyttes til videreudvikling, kan det være relevant at undervise i datakvalitet, samt hvordan ny data påvirker AI-assistentens resultater. En bedre forståelse for AI-assistentens underliggende data bidrager til forklarlighed og gennemsigtighed

C4t

Skab tillid og forankring

Det sidste fokusområde handler om at skabe tillid hos brugerne af AI-assistenten. Dette kan være interne brugere som medarbejdere i organisationen eller eksterne brugere som organisationens kunder.

Ved implementering er det vigtigt, at medarbejderen føler sig tryk i både at anvende AI-assistenten samt i de resultater, som AI-assistenten giver. Ellers motiveres medarbejderen ikke til at benytte det nye værktøj og vil i stedet fortsætte med at udføre sin opgave som vanligt. Derfor skal kernefortællingen og uddannelsen fokusere på, hvor data kommer fra, og hvordan den behandles. Ligeledes skal medarbejderen lære at kvalitets-sikre AI-assistents resultater. Medarbejderen skal f.eks. lære at afkode, hvornår AI-assistenten hallucinerer. Her kan f.eks. indsættes kontrolspørgsmål, hvorpå medarbejderen kender svaret. Herved vil medarbejderen opnå en større forståelse for,

hvordan AI-assistenten fungerer, så denne kan føle sig mere tryk i brugen af AI-assistenten.

For at skabe tillid hos kunder kan det være en fordel at forklare kunderne, at assistenterne er med til at forbedre kundeservicen vha. hjælpetekst, hvor AI-assistenten anvendes. Dernæst skal AI-assistenten give en personlig betjening af kunderne, eks. ved at præcisere personlige detaljer i det omfang det er lovligt. På den måde er det bl.a. muligt at øge trygheden for kunderne i at anvende AI-assistenten.



3.9

Etabler opfølgnings— og support-strukturer

Den digitale transformation ender ikke ved go-live. Organisationen skal fortsætte med at sikre, at de nye arbejdsgange følges samt fortsat at opbygge tillid til AI-assistenten. Der skal derfor etableres strukturer, som sikrer effektiv vedligeholdelse og fortsat lovmedholdelighed. Målet med opfølgnings- og supportstrukturerne er at lette forandringen i organisationen, forbedre AI-assistenten og minimere risikoen ved at identificere problemstillinger, så snart de opstår.

Følgende værktøjer kan tages i brug



01 Ekspertes

Forretnings- og tekniske eksperter, der vejleder og besvarer spørgsmål for brugere. Forretningseksperter benyttes til side-mandsoplæring, mens tekniske eksperter vejle-der i AI-assi-stentens opbygning. Den første gruppe af eksperter kan dannes ud fra et pilotforløb, som beskrevet i afsnit 3.5.

02 Feedback

Funktion, hvor borgere kan indrapportere fejl. En god feed-backmekanisme hjælper med at identificere specifikke områ-der, hvor AI-assistenten ikke svarer korrekt eller et område, den burde kunne udføre, men ikke kan. Dette gør det muligt at justere modellen baseret på konkrete præcise eksempler. En simpel feedbackmekanisme kan være muligheden for at mar-kere, om svaret er godt eller dårligt. Hvad der for brugeren kan fremstå som et dårligt svar er imidlertid ikke nødvendigvis et forkert svar, hvorfor det kan være relevant at indsamle uddyb-ning eller forslag til korrekt svar.

03 Overvågning

Overvågning af logs og feedback så observerede udfordringer håndteres, inden de når at påvirke borgerne i større omfang. Løbende overvågning af en AI-assistent bør forgå automatisk. Overvågning kan baseres på datapunkterne introduceret i afsnit 3.2. Der er to primære tilgange til overvågning: Regelba-seret og intelligent.

Regelbaseret overvågning

Regler benyttes til at identificere adfærd i løsningen, der ikke går som forven-tet eller optimalt. F.eks. vil kvaliteten af et svar i RAG kunne måles i form af relevans og nøjagtighed, så længe distancen i en forespørgsel bliver beregnet, hvilket kan indikere en hallucination er forekommet. Regelbaseret overvåg-ning bygger kun på kendte problemer.

Intelligent overvågning

En AI-assistent implementeret med det formål at overvåge loggen og eskalere, hvis AI-assistenten finder noget uhensigts-mæssigt. AI-assistenten kan vurdere svar baseret på input og genkende forsøg på sikkerhedsbrud såsom prompt-injection-angreb⁷ og derfra opskalere. Hvis der anvendes en cloud-udby-der, skal det sikres, at et databrud i den overvågede AI-assistent ikke og-så medfører et databrud i den overvågende AI-assisten-ten. De to systemer bør være adskilt, så et sikkerhedsbrud i det ene system ikke påvirker det andet.

⁷ En teknik, hvor AI-assistent manipule-res ved at få indsat uventet tekst i dens input med formålet om at ændre assi-stentens adfærd eller få den til at gene-rere uønskede eller forkerte svar [10].

04 Automatisk eskalering

Borgere der søger rådgivning kan opleve, at en AI-assistent ikke kan besvare deres spørgsmål. Hvor problemstillinger internt håndteres gennem feedback og support, kan borgere sidde i en frustrerende situation. Derfor bør AI-assistenten have indbyg-ge-de kontrolmekanismer, der kan registrere, hvis det samme spørgsmål bliver stillet flere gange eller ikke kan finde relevant information, så den kan hjælpe borgeren vi-dere i sin proces. Dette kan være ved at give et telefonnummer eller føre borgeren videre til en supportchat med en menneskelig supporter. Her-med sikres det, at bor-gere altid får den nødvendige hjælp, selv når AI-assistenten ikke kan levere et tilfredsstillende svar.

05 Opfølgning

Menneskelig opfølgning på feedback og overvågning med målet om at gøre AI-assistenten bedre sikrer en effektiv løbende optimering af AI-assistentens ydeevne. Ekspertes bør regel-mæssigt evaluere overvågningsdata og feedbackrapporter samt implementere de nødvendige justeringer. Da opfølgning kan være tidskrævende, bør de logs, der gennemgås, være filtrere-de på forhånd for at reducere arbejdsbyrden. Kombinationen af automatiske og manuelle processer sikrer, at AI-assistenten kon-stant forbedres og forbliver troværdig.

06 Vedligeholdelse af vidensgrundlag

For at sikre at AI-assistenten forbliver præcis og opdateret, bør der etableres en pro-ces for løbende vedligeholdelse af dette vidensgrundlag. Selvom en del opdatering vil kunne udføres automatisk, vil der være behov for at ajourføre dens kilder, f.eks. ved introduktion af ny lovgivning. Ekspertes, der vedli-geholder videnskilder, bør samar-bejde med både juridiske rådgivere og tekniske teams for at sikre, at kilder er ajour-førte og korrekte.



Disse strukturer – kombinationen af eksperter, feedbackmekanis-mer, automatisk overvågning, eskalering og menneskelig opfølg-ning – sikrer, at AI-assisterter løbende optimeres og opretholder de højeste standarder for korrekthed, compliance og pålidelighed i deres interaktioner.

Afslutning

C4t

Denne hvidbog har præsenteret en tilgang for, hvordan organisationer kan udvikle og implementere AI-assisterter på den bedste, mest innovative og ansvarlige måde i en tid, hvor Europa står overfor de største udfordringer i nyere tid.

Vi står med en unik mulighed for at udnytte teknologien til at skabe mere effektive, bæredygtige og robuste samfund, og selvom specifikke teknologier og implementeringsmetoder udvikler sig over tid, forbliver de grundlæggende overvejelser i hvidbogen relevante, da de bygger på teknologi-agnostiske gode praktiser.

Ved at anvende denne fremgangsmåde kan organisationer opbygge en proces for at anvende AI-assisterter, så de tager højde for lovkrav og risiko. Denne proces vil ikke kun understøtte den aktuelle AI-assistents implementering, men samtidig danne grundlag for fremtidige AI-projekter.

En central fordel ved at følge hvidbogen er at den er med til at øge viden og kompetencer internt i organisationen ift. at arbejde med AI. På den måde vil organisationen blive mere erfaren og bedre klædt på til at implementere AI-assisterter i organisationen i fremtiden.

Organisationer opfordres til at overveje en AI-plattform tidligt i deres AI-rejse bygget på gode principper om IT. Ved at vælge en robust platform samt implementere centrale overvågnings- og styringskomponenter fra starten, kan organisationer lægge grundlaget for mere effektiv skalering og integration af fremtidige AI-assisterter.

Mens AI-teknologien fortsætter med at udvikle sig, forbliver organisationers forpligtelse til ansvarlig, etisk og effektiv implementering uforandret. Ved at bygge på det fundament, der er præsenteret i hvidbogen, kan organisationer med tillid navigere i det spændende landskab af AI-assisterter og udnytte deres fulde potentiale til gavn for virksomheden og samfundet som helhed.

De tilknyttede addenda præsenterer virkelige cases, der giver konkrete eksempler på, hvordan principperne i hvidbogen kan anvendes i praksis. Casene bør udforskes for yderligere indsigt og inspiration.

Referencer

- [1] EU-U.S. Trade and Technology Council, EU-U.S. Terminology and Taxonomy for Artificial Intelligence - Second Edition. The Directorate General for Communications Networks, Content and Technology, 2024. Set: 26. september 2024. [Online]. Tilgængelig hos: <https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence-second-edition>
- [2] P. Béchard og O. M. Ayala, "Reducing hallucination in structured outputs via Retrieval-Augmented Generation", apr. 2024, [Online]. Tilgængelig hos: <http://arxiv.org/abs/2404.08189>
- [3] Datatilsynet, "Hvad er personoplysninger?" Set: 26. september 2024. [Online]. Tilgængelig hos: <https://www.datatilsynet.dk/hvad-siger-reglerne/grundlaeggende-begreber/hvad-er-personoplysninger>
- [4] Datatilsynet, Justitsministeriet, og Digitaliseringsstyrelsen, "Behandlingssikkerhed - Databeskyttelse gennem design og standardindstillinger", jun. 2018. [Online]. Tilgængelig hos: www.datatilsynet.dk,
- [5] Europa-Parlamentets og Rådets forordning (EU) 2016/679 af 27. april 2016 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger og om ophævelse af direktiv 95/46/EF (generel forordning om databeskyttelse) (EØS-relevant tekst). EUR-Lex, 2016.
- [6] Europa-Parlamentets og Rådets forordning (EU) 2024/1689 af 13. juni 2024 om harmoniserede regler for kunstig intelligens og om ændring af forordning (EF) nr. 300/2008, (EU) nr. 167/2013, (EU) nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 og (EU) 2019/2144 samt direktiv 2014/90/EU, (EU) 2016/797 og (EU) 2020/1828 (forordningen om kunstig intelligens) (EØS-relevant tekst). EUR-Lex, 2024. [Online]. Tilgængelig hos: <http://data.europa.eu/eli/reg/2024/1689/oj>
- [7] C. and T. Directorate-General for Communications Networks, "Retsakt om kunstig intelligens". Set: 26. september 2024. [Online]. Tilgængelig hos: <https://digital-strategy.ec.europa.eu/da/policies/regulatory-framework-ai>
- [8] T. B. Brown m.fl., "Language Models are Few-Shot Learners", 2020. [Online]. Tilgængelig hos: <https://commoncrawl.org/the-data/>
- [9] European Union Agency For Network and Information Security, "Guidelines for SMEs on the security of personal data processing", s. 10, 2016, doi: 10.2824/867415.
- [10] CERT-EU Team, "CERT-EU Security Guidance 23-002 Potential impact and risks of Generative AI in EUIBAs", maj 2023. Set: 26. september 2024. [Online]. Tilgængelig hos: <https://cert.europa.eu/publications/security-guidance/security-guidance-23-002---potential-impact-and-risks-of-generative-ai-in-euibas/>

05

Bilag

Có

Bilag A

Dette bilag indeholder en juridisk tjekliste, som har til formål at konkretisere og operationalisere hvidbogens punkt 4 (afsnit 3.4) nærmere. Tjeklisten omhandler således de juridiske forhold og de spørgsmål, der skal undersøges og afklares, når du arbejder med at afklare de juridiske rammer for AI-assistenten.

Tjeklisten er til inspiration og vejledning – og udgør ikke en facitliste. Der skal foretage en konkret juridisk vurderingen i relation til den pågældende AI-assistent, herunder med iagttagelse af andre juridiske krav, der gælder for din organisation.

Tjeklisten adresserer AIA og GDPR og er opbygget i henhold hertil.

Tjeklisten er i al væsentlighed opstillet som spørgsmål, og spørgsmålene baserer sig på bagvedliggende juridiske krav, der er udtrykt i AIA eller GDPR. Der er relevante steder lavet henvisninger til disse bagvedliggende juridiske krav.

Uanset om AI-assistenten er omfattet af forordningen eller ej, kan tjeklisten anvendes til sikring af, at AI-assistenten udvikles/ anvendes i overensstemmelse med principperne for ansvarlig AI, som i vid udstrækning bygger på de samme principper som forpligtelserne i AIA.

1. AIA

1.1. Risikoniveau og rolle

Du skal – som angivet i hvidbogens punkt 4 – indledningsvist afklare:

1. Assistentens risikoniveau
2. Din rolle under AIA

Tjeklisten er opbygget i henhold til disse afklaringer, og de relevante spørgsmål og forhold er således opdelt i forhold til risikoniveau og rolle, jf. nedenfor. Hvis AI-assistenten således vurderes at være højrisiko, og du har rollen som idriftsætter, skal du orientere dig i punktet herom (konkret afsnit 1.2.1). Det skal dog bemærkes, at du bør orientere dig i de andre krav, så du er bekendt hermed, herunder for at sikre, at en eventuel leverandør af AI-assistenten efterlever de relevante krav, f.eks. som udbyder.

1.2. Risikoniveau: AI-assistenten vurderes at være ”højrisiko”

1.2.1 Krav til udbydere – højrisiko

Som udbyder skal du overholde en række krav, hvis AI-assistenten vurderes at være et højrisiko AI-system, jf. AIA, Afdeling 2 ”Krav til højrisiko-AI-systemer”. Du skal således – som udbyder - vurdere og sikre overholdelsen af disse krav ved udvikling af anvendelse af AI-assistenten:

1.2.1.1 Overordnede forpligtelser (AIA, artikel 16)

Overordnet har du følgende forpligtelser som udbyder af AI-assistenten, der vurderes at være et højrisiko-AI-system:

- a) sørge for, at AI-assistenten overholder de krav, der er fastsat i afdeling 2 i AIA, se nærmere under punkt 1.2.1.2-1.2.1.7 nedenfor
- b) angive dit navn, registrerede firmanavn eller registrerede varemærke og kontaktadresse på AI-assistenten, eller, hvis dette ikke er muligt, på dets emballage eller i den medfølgende dokumentation, alt efter hvad der er relevant
- c) have indført et kvalitetsstyringssystem, der er i overensstemmelse med AIA, artikel 17
- d) opbevare dokumentation, der er omhandlet i AIA, artikel 18

e)	opbevare de logfiler, der automatisk genereres af AI-assistenten, når logfilerne er under deres kontrol, som omhandlet i AIA, artikel 19
f)	sørge for, at AI-assistenten underkastes den relevante overensstemmelsesvurderingsprocedure som omhandlet i AIA, artikel 43, inden AI-assistenten bringes i omsætning eller ibrugtages
g)	udarbejde en EU-overensstemmelseserklæring i overensstemmelse med AIA, artikel 47
h)	anbringe CE-mærkningen på AI-assistenten i overensstemmelse med AIA, artikel 48
i)	overholde de registreringsforpligtelser, der er omhandlet i AIA, artikel 49, stk. 1 – er AI-assistenten registreret i EU-databasen?
j)	foretage de nødvendige korrigerende tiltag og fremlægge oplysningerne som krævet i henhold til AIA, artikel 20
k)	efter begrundet anmodning fra en national kompetent myndighed påvises, at AI-assistenten er i overensstemmelse med de krav, der er fastsat i AIA, afdeling 2
l)	sikre, at AI-assistenten overholder tilgængelighedskravene i overensstemmelse med direktiv (EU) 2016/2102 og (EU) 2019/882.

1.2.1.2 Risikostyringssystem (AIA art. 9)

- Er der oprettet og gennemført et risikostyringssystem, som dokumenteres og vedligeholdes? (art. 9, nr. 1)
- Dækker det AI-assistentens livscyklus?
- Har du lavet kortlægning og analyse af kendte og forudsigelige risici forbundet med AI-assistenten, herunder
 - vurdering og evaluering af de kendte og forudsigelige risici, der kan opstå, når AI-assistenten anvendes
 - evaluering af andre risici, der kan opstå, på grundlag af analyse af data indsamlet fra AI-assistenten til overvågning efter omsætningen,
- Har du indført passende risikostyringsforanstaltninger?

1.2.1.3 Data og datastyring (AIA artikel 10)

- Har du etableret styrings- og forvaltningsprocesser for sikring af datakvalitet af trænings- validerings- og afprøvningsdatasæt. Følge databeskyttelse gennem design og databeskyttelse gennem standardindstillinger, se [4].
- Har du sikret, at trænings-, validerings- og afprøvningsdatasæt er underlag sådanne processer, navnlig:
 - de relevante valg med hensyn til udformning
 - dataindsamlingsprocesser og dataenes oprindelse og, hvis der er tale om personoplysninger, det oprindelige formål med dataindsamlingen
 - relevant dataforberedelsesbehandling såsom annotation, mærkning, rensning, opdatering, berigelse og aggregering
 - formuleringen af antagelser, navnlig med hensyn til de oplysninger, som dataene skal måle og repræsentere
- en vurdering af tilgængeligheden, mængden og egnetheden af de datasæt, der er nødvendige
- undersøgelse med hensyn til mulige bias, der sandsynligvis vil påvirke personers sundhed og sikkerhed, have negativ indvirkning på de grundlæggende rettigheder eller føre til forskelsbehandling, som er forbudt i henhold til EU-retten, navnlig hvis dataoutput påvirker input til fremtidige operationer
- passende foranstaltninger til at påvise, forebygge og afbøde de mulige bias, der er identificeret i henhold til litra f)

- kortlægning af relevante datamangler eller datautilstrækkeligheder, som forhindrer overholdelse af denne forordning, og hvordan de kan afhjælpes.

1.2.1.4 Teknisk dokumentation (AIA artikel 11)

- Har du udarbejdet al relevant teknisk dokumentation? Se kravene til dokumentation i AIA, Bilag IV.
- Har du sikret processer for løbende ajourføring af den tekniske dokumentation?
- Har du sikret, at dokumentationen kan opbevares i overensstemmelse med AIA artikel 18?

1.2.1.5 Gennemsigtighed, logning og formidling af oplysninger (AIA artikel 12-13)

- Har du indført foranstaltninger, der sikrer gennemsigtighed? Har du udarbejdet en brugsanvisning? Har du indført mekanismer til at informere Medarbejdere/Brugerne om årsagerne til og kriterierne for AI-assistentens resultater? Kommunikeres dette på en klar og forståelig måde til målgruppen? Har du etableret processer, der tager hensyn til brugernes feedback og anvender dette til at tilpasse AI-assistenten? Har du formidlet potentielle eller opfattede risici, f.eks. skævheder? Har du, afhængigt af anvendelsen, også overvejet kommunikation og gennemsigtighed. over for andre målgrupper, tredjemand eller offentligheden?
- Har du vurderet, hvorvidt AI-assistentens afgørelser og dermed resultater kan forstås?
- Har du indført mekanismer, der gør det lettere at revidere AI -assistentens for Medarbejdere/Brugere, f.eks. ved at sikre sporbarhed og registrering af AI-assistentens processer og resultater?
- Har du sikret, at hændelser (”logfiler”) registreres automatisk i AI-assistenten, når den er i drift? Har du anvendt anerkendte standarder eller fælles specifikationer for logning?
- Har du sikret, at det kan forklares, hvorfor AI-assistenten traf et bestemt valg, der førte til et bestemt resultat, på en måde, der er forståelig for alle Medarbejdere/Brugere, der måtte ønske en forklaring?
- Har du vurderet, i hvilken udstrækning AI-assistentens afgørelse påvirker Organisationens beslutningsprocesser?
- Har du fra starten designet AI-assistenten, så det kan fortolkes? Har du undersøgt og forsøgt at anvende den enkleste og mest fortolkningsvenlige model til den pågældende anvendelse? Har du vurderet, om du kan analysere dine oplærings- og testdata? Kan du ændre og ajourføre disse over tid? Har du vurderet, om du efter modeloplæringen og -udviklingen har mulighed for at undersøge fortolkningsevnen, eller om du skal vurdere modellens interne workflow?
- Har du præciseret, hvad formålet med AI-assistenten er, og hvem eller hvad der kan få gavn af produktet/tjenesten? Er anvendelsesscenarierne blevet specificeret og klart formidlet, også under hensyntagen til alternative kommunikationsformer for at sikre, at de er forståelige og hensigtsmæssige for den pågældende målgruppe? Har du, afhængigt af anvendelsen, taget hensyn til den menneskelige psykologi og potentielle begrænsninger, f.eks. risiko for forvirring, forudindtaget bekræftelse eller kognitiv træthed?

1.2.1.6 Menneskeligt tilsyn (artikel 14)

- Interagerer AI-assistenten med Medarbejderes og Brugeres beslutningstagning (f.eks. anbefalede aktioner eller beslutninger, der skal træffes, med angivelse af muligheder)? Er der i disse tilfælde risiko for, at AI-assistenten påvirker menneskers autonomi ved at gribe ind i Brugerens/medarbejderens beslutningsproces på en utilsigtet måde? Har du overvejet, om AI-assistenten bør meddele Brugerne/Medarbejderne, at en beslutning, indhold, rådgivning eller en konklusion er resultatet af en algoritmisk afgørelse?
- Hvis AI-assistenten er en ”chatbot” eller et samtalsystem, gøres de menneskelige Brugere da opmærksomme på, at de interagerer med en ikkemenneskelig agent?
- Hvis AI-assistenten er implementeret i en arbejdsproces, har du da overvejet opgavefordelingen mellem AI-assistenten og Medarbejderne for at sikre meningsfulde interaktioner og passende menneskelig kontrol? Har du sikret, at det menneskelige tilsyn udføres af tilstrækkeligt kompetente medarbejdere?
- Forbedrer eller styrker AI-assistenten menneskelige færdigheder? Har du truffet sikkerhedsforanstaltninger for at forhindre

	dre overdreven tillid til eller afhængighed af AI-assistenten i arbejdsprocesser?
5.	Har du overvejet, hvilket niveau af menneskelig kontrol der er hensigtsmæssigt for den specifikke AI-assistent og den specifikke anvendelse? Kan du beskrive niveauet af menneskelig kontrol eller inddragelse, hvis det er relevant? Hvem har den menneskelige kontrol, og hvornår og med hvilke værktøjer kan der ske menneskelig indgriben? Har du indført mekanismer og foranstaltninger, der sikrer sådan potentiel menneskelig kontrol, eller som sikrer, at beslutninger træffes på menneskers ansvar?
6.	Har du truffet foranstaltninger for at muliggøre revision og afhjælpe problemer vedrørende styring af AI-autonomi?
7.	Hvis der er tale om en selvstående eller autonomt AI-assistent eller anvendelsestilfælde, har du da indført mere specifikke kontrol- og tilsynsmekanismer?
8.	Hvilken type detektions- og responsmekanismer har du etableret for at vurdere, om noget kan gå galt?
9.	Har du etableret en ”stopknap” eller procedure, der kan afbryde en operation på en sikker måde, hvis det bliver nødvendigt? Afbryder denne procedure processen helt eller delvist, eller overlader den kontrollen til et menneske?

1.2.1.7 Nøjagtighed, robusthed og cybersikkerhed (artikel 15)

- Har du sikret et passende niveau af nøjagtighed af AI-assistenten og angivet det og relevante parametrene i brugsanvisningen?
- Har du sikret, at AI-assistenten er modstandsdygtige over for fejl, svigt og uoverensstemmelser, der kan forekomme i systemet eller i det miljø, som systemet fungerer i, navnlig på grund af systemets interaktion med fysiske personer eller andre systemer.
- Har du sikret AI-assistentens robusthed ved hjælp af tekniske redundansløsninger, som kan omfatte backupplaner eller fejlsikre planer?
- Har du sikret, at AI-assistenten er modstandsdygtig over for uautoriserede tredjeparters forsøg på at ændre dets anvendelse eller ydeevne ved at udnytte AI-assistentens sårbarheder?
- Har du indført passende tekniske løsninger til afhjælpning af AI-specifikke sårbarheder f.eks. foranstaltninger til forebyggelse af og kontrol for angreb, der forsøger at manipulere træningsdatasættet (”dataforgiftning”), input, der har til formål at få modellen til at begå fejl (”fjendtlige eksempler”), eller modelfejl.

1.2.1.8 Kvalitetssikringssystem (artikel 17)

- Har du etableret et kvalitetsstyringssystem, der sikrer overensstemmelse med AIA. Herunder dokumenteres systemet på en systematisk og velordnet måde i form af skriftlige politikker, procedurer og anvisninger?
- Har du sikret, at dokumentationen som minimum omfatter følgende aspekter:
 - en strategi for overholdelse af lovgivningen, herunder overholdelse af overensstemmelsesvurderingsprocedurer og procedurer for forvaltning af ændringer af højrisiko-AI-systemet
 - teknikker, procedurer og systematiske tiltag, der skal anvendes til udformningen, kontrollen af udformningen og verifikationen af udformningen af højrisiko-AI-systemet
 - teknikker, procedurer og systematiske tiltag, der skal anvendes til udvikling, kvalitetskontrol og kvalitetssikring af højrisiko-AI-systemetundersøgelser-, afprøvnings- og valideringsprocedurer, der gennemføres før, under og efter udviklingen af højrisiko-AI-systemet, samt den hyppighed, hvormed de gennemføres
 - tekniske specifikationer, herunder standarder, der anvendes, og, hvis de relevante harmoniserede standarder ikke anvendes fuldt ud eller ikke omfatter alle de relevante krav, der er fastsat i afdeling 2 i AIA, de midler, der skal anvendes for at sikre, at højrisiko-AI-systemet overholder disse krav
 - systemer til og procedurer for dataforvaltning, herunder dataopsamling, dataindsamling, dataanalyse, datamærkning, datalagring, datafiltrering, datamining, dataaggregering, dataopbevaring og enhver anden handling vedrørende data,

	som udføres forud for og med henblik på omsætning eller ibrugtagning af højrisiko-AI-systemer
f)	det risikoforvaltningssystem, der er omhandlet i artikel 9 i AIA
g)	oprettelse, implementering og vedligeholdelse af et system til overvågning efter omsætningen i overensstemmelse med artikel 72 i AIA
h)	procedurer for indberetning af en alvorlig hændelse i overensstemmelse med artikel 73 i AIA
i)	håndtering af kommunikation med nationale kompetente myndigheder, andre relevante myndigheder, herunder dem, der giver eller understøtter adgang til data, bemyndigede organer, andre operatører, kunder eller andre interesserede parter
j)	systemer til og procedurer for registrering af al relevant dokumentation og alle relevante oplysninger
k)	ressourceforvaltning, herunder foranstaltninger vedrørende forsyningssikkerhed
l)	en ansvarlighedsramme, der fastlægger ledelsens og det øvrige personales ansvar.

1.2.2 Krav til idriftsættere – højrisiko (Artikel 26 og 27)

Som idriftsætter skal du overholde en række krav, hvis AI-assistenten vurderes at være et højrisiko AI-system. Du skal således – som idriftsætter - vurdere og sikre overholdelsen af nedenstående krav ved udvikling af anvendelse af AI-assistenten. Du bør dog – som bemærkes indledningsvist – tillige orientere dig i kravene, der gælder til udbydere.

- Har du sikret, at højrisiko-AI-systemet anvendes i overensstemmelse med brugsanvisningen?
- Har du sørget for menneskeligt tilsyn med AI-assistenten til en person med tilstrækkelige kompetencer, uddannelse og myndighed samt den nødvendige støtte?
- Har du sikre relevante og repræsentative inputdata i lyset af formålet med AI-assistenten?
- Har du sikret, at driften af AI-assistenten overvåges på grundlag af brugsanvisningen?
- Har du sikret at logfiler opbevares, hvis de er under idriftsætterens kontrol?
- Har du registreret AI-assistenten i en EU-database (artikel 49)?
- Har du udarbejdet en konsekvensanalyse vedrørende databeskyttelse (DPIA) på baggrund af oplysningerne i brugsanvisningen?
- Har du udarbejdet en Fundamental Rights impact Assessment (FRIA) i overensstemmelse med AIA artikel 27 i AIA? Herunder har du vurderet den bredere samfundsmæssige indvirkning af AI-assistentens anvendelse ud over den enkelte Bruger, f.eks. de potentielt indirekte berørte aktionærer? Bemærk, at Fundamental Rights impact Assessment kun relevant for ”idriftsættere, der er offentligretlige organer eller private enheder, der leverer offentlige tjenester, og idriftsættere af de højrisiko-AI-systemer, der er omhandlet i bilag III, punkt 5, litra b) og c)”, jf. art. 27. Dvs.
 - AI-systemer, der tilsigtes anvendt til at evaluere fysiske personers kreditværdighed eller fastslå deres kreditvurdering, med undtagelse af AI-systemer, der anvendes til at afsløre økonomisk svig
 - AI-systemer, der tilsigtes anvendt til at foretage risikovurdering og prisfastsættelse for så vidt angår fysiske personer i forbindelse med livs- og sygeforsikring

2. Overholdelse af GDPR

Du skal vurdere overholdelsen af følgende **generelle databeskyttelsesretlige krav ved behandling af personoplysninger ved udvikling og anvendelse af AI-assistenten. Bemærk, at listen over de nævnte databeskyttelsesregler og risici ikke er udtømmende og varierer fra projekt til projekt, men giver et solidt fundament for databeskyttelsesretlig compliance.**

2.1. Overblik over behandlingen af personoplysninger og din rolle

1. Har du et overblik over databehandlingen, herunder hvilke personoplysninger der behandles om hvilke personer til hvilke formål i hele løsningens livscyklus?
2. Har du afklaret din databeskyttelsesretlige rolle, herunder de databehandlinger som du er dataansvarlig for?

2.2. Håndtering af databeskyttelsesreglerne i hele løsningens livscyklus

1. Har du sørget for – på en dokumenterbar måde – at indtænke databeskyttelsesreglerne igennem hele AI-assistentens livscyklus, dvs. allerede fra designfasen og frem til overvågning af løsningen i drift?
2. Har du forholdt dig til, om du overholder de grundlæggende principper om bl.a. lovlighed, rimelighed og gennemsigtighed, formålsbegrænsning, dataminimering og datakvalitet i GDPR-artikel 5?
3. Har du identificeret et behandlingsgrundlag (hjemmel) for behandlingen af personoplysningerne ved udvikling/drift af AI-assistenten?

2.3. Oplysningspligten og de registreredes rettigheder

1. Har du sørget for at underrette de registrerede om behandlingen af personoplysninger ved både udvikling og drift af AI-assistenten?
2. Underretter du de registrerede herom, hvis der anvendes profilering, herunder konsekvenserne heraf?
3. Hvis der anvendes automatiske, individuelle afgørelser, bliver de registrerede da underrettet herom og modtager de som minimum meningsfulde oplysninger om logikken heri samt betydningen og de forventede konsekvenser af en sådan behandling for de registrerede?
4. Hvis der anvendes automatiske, individuelle afgørelser, er der i så fald identificeret et klart hjemmelsgrundlag herfor, og kan I håndtere de registreredes rettigheder, herunder retten til at få menneskelig indgriben, hvor relevant?
5. Er der procedurer for håndtering af de registreredes øvrige rettigheder, herunder indsigts-, berigtigelses- og indsigelsesretten samt retten til sletning?

2.4. Sikkerhed, konsekvensanalyse og effektiv håndtering af risici for de registrerede

1. Har du med udgangspunkt i en risikovurdering vedrørende behandlingssikkerhed identificeret risici vedrørende de registreredes rettigheder og frihedsrettigheder og implementeret passende organisatoriske og tekniske sikkerhedsforanstaltninger i hele løsningens livscyklus for at mitigere disse risici?
2. Har du taget stilling til, om der skal udarbejdes en konsekvensanalyse vedrørende databeskyttelse (DPIA) og i så fald gennemført analysen med inddragelse af en eventuel DPO?
3. Har du udarbejdet retningslinjer og givet instrukser til medarbejderne om, hvordan de skal anvende AI-assistenten korrekt og sikkert, så de kender til AI-assistentens formål og begrænsninger, lovlig brug og involverede risici?

2.5. Overførsler til tredjelande og deling af personoplysninger med

leverandøren

1. Har du identificeret samtlige eventuelle overførsler af personoplysninger til tredjelande og vurderet, om disse overførsler er lovlige ved at udarbejde en TIA herfor?
2. Hvis du anvender en SaaS-løsning, har du da gennemgået leverandørens kontraktvilkår og databehandleraftale med henblik på at undersøge, om du ved brugen af løsningen videregiver personoplysninger til leverandøren af løsningen til brug for leverandørens egne formål, f.eks. træning af løsningen, produktudvikling, markedsføring etc. Har du i så fald skal vurderet, om denne videregivelse af personoplysninger er lovlig?

Addenda

C7

Erhvervsstyrelsens branchekodevælger

Indledning

En branchekode er en kode der indikerer, hvilken branche en virksomhed arbejder inden for. Det er et krav, at virksomheder skal have en branchekode, men det er notorisk svært at vælge den rigtige. Ved hjælp af kunstig intelligens er der mulighed for, at virksomheder, fx i forbindelse med registrering eller ændring af virksomheden på Virk.dk, kan få hjælp til at vælge branchekode ved at indtaste fritekst om deres aktivitet, hvorefter branchekodevælgeren giver en række forslag med de bedste bud på relevante branchekoder.

Business case

Ved at bruge kunstig intelligens til at understøtte virksomheders valg af branchekode kan Erhvervsstyrelsen opnå et kvalitetsløft af data og forbedring af services for brugeren. De primære brugere af branchekodevælgeren er virksomheder. Assistenten på virk.dk fører brugeren hurtigt igennem trin, hvor de får hjælp til at tjekke deres virksomheds nuværende branchekode og registrere en evt. ny branchekode, hvis det er nødvendigt.

Det er vigtigt, at virksomheder registrerer den korrekte branchekode, da den er afgørende for en række ting, fx forsikringspræmie, affaldsafgift m.v. Desuden danner branchekoderne grundlag for de data, som Erhvervsstyrelsen udstiller om virksomhederne. Korrekte branchekoder øger datakvaliteten i CVR såvel som Danmarks Statistik, som bl.a. fører statistik om virksomhedernes omsætning, beskæftigelse, investeringer og serviceydelser.

Juridiske forhold

De dataprompt, som branchekodevælgeren behandler, består i virksomhedens egen frivilligt indtastede overordnede beskrivelse af, hvad virksomheden beskæftiger sig med. Der behandles

hverken personoplysninger eller fortrolige oplysninger.

Henset til at branchekodevælgeren ikke behandler oplysninger, der ikke kan henføres til en person, falder behandlingen uden for databeskyttelsesforordningens og -lovens anvendelsesområde.

Branchekodevælgerens output er ikke en afgørelse eller beslutning, men er alene et forslag til virksomheden om, hvilken eller hvilke branchekoder der kan være relevant for virksomheden at angive. Virksomheden har mulighed for selv at angive en specifik branchekode uden at benytte branchekodevælgeren. Ansvar for, hvilken branchekode, der vælges, vil således være virksomhedens eget uanset om branchekodevælgeren anvendes eller ej.

I lyset af at branchekodevælgeren alene er et frivilligt, understøttende værktøj, som hjælper brugeren med at vælge branchekode ved at komme med forslag hertil, og som ikke indeholder personoplysninger eller fortrolige oplysninger, er det ikke nødvendigt med en eksplicit lovhjemmel for at stille branchekodevælgeren til rådighed for brugeren.

Branchekodevælgerne vil umiddelbart være omfattet af AI-forordningens definition af et AI-system. Ligeledes vil styrelsen umiddelbart blive betegnet som udbyder i forordningens forstand.

Tekniske aspekter

Sprogmodellen bag branchekodevælgeren ligger på Erhvervsstyrelsens egen platform. Platformen er en skræddersyet avanceret datadrevet platform, som baserer sig på en sky-baseret machine learning-platform til drift af machine learning modeller. Platformen understøtter de mange forskellige nye teknologier, som anvendes i machine learning.

Skalering og indpakning af modellerne medvirker til en stabil drift mellem

de mange specialiserede teknologier, som indgår i kunstig intelligens. Platformen leverer desuden en række services, som sikrer, at modellernes udsagn og tilknyttede metrikker tilflyder de rette fagsystemer. En del af disse data gemmes af hensyn til krav om transparens og forklarbarheden af styrelsens praksis med beslutningsunderstøttelse.

Branchekodevælgeren er baseret på en open source-sprogmodel, XLM-RoBERTa, som udgør kernen af dens sprogforståelse. Sprogmodellen er fine-tunet til at oversætte virksomheders aktivitetsbeskrivelser til branchekoder. Den er trænet på data udleveret af det norske Brønnøysundregisteret, hvor det er sagsbehandlere der manuelt tildeler brancher til virksomheder, på baggrund af virksomhedens aktivitetsbeskrivelse. Dette er maskinoversat til dansk. Datagrundlaget herudover er følgende:

- Aktivitetsbeskrivelser fra regnskaber, indleveret til Erhvervsstyrelsen
- Formålstekster fra CVR-registeret
- Navne på selskaber
- Branchebeskrivelser fra Danmarks Statistik

Modellen er for nyligt opdateret til de nye branchekoder, der træder i kraft fra 2025.

Implementering i organisationen

Branchekodevælgeren har ikke krævet en digital transformation af interne arbejdsgange, hvorfor der ikke har været behov for at udvikle en implementeringsstrategi eller opbygge nye kompetencer hos ansatte.

Erhvervsstyrelsens brug af en AI assistent til kodeskrivning

Indledning

Erhvervsstyrelsen har en række applikationer og systemer i drift til at understøtte styrelsens opgaveløsning. For at forbedre deres funktionalitet og stabilitet skal frameworket Grails opdateres løbende. Dette gør man ved at omskrive koden, og denne proces kan automatiseres ved hjælp af generativ kunstig intelligens.

Business case

Der udgives ofte nye Grails versioner til Erhvervsstyrelsens applikationer, og det er en tidskrævende proces at opgradere en applikation til en ny version. Desuden er manuel kodeskrivning til framework og opgave, som ikke er motiverende for medarbejdere.

Ved at bruge generativ kunstig intelligens til at automatisere kodeskrivningen går Erhvervsstyrelsen fra at bruge 60-100 timer på en opdatering af Grails versioner til 7-31 timer, hvilket giver en besparelse i tid på 69-88 %.

Juridiske forhold

De data, som den pågældende AI-assistent behandler, består alene af kodedata, hvori der hverken behandles personoplysninger eller fortrolige oplysninger. Henset til, at der ved brug af AI-assistenten ikke behandles oplysninger, der kan henføres til en identificeret eller identificerbar person, hverken direkte eller indirekte, falder behandlingen uden for databeskyttelsesforordningens og -lovens anvendelsesområde.

AI-assistents output er ligeledes hverken en afgørelse eller beslutning, men er alene et forslag til kode. Den pågældende medarbejder, der anvender AI-assistenten, skal validere assistentens output og har mulighed for selv at justere eller rette det foreslåede output hvis nødvendigt. Ansvar for kodens korrekthed og validitet vil ligge hos den pågældende, der anvender AI-assistenten. Det er alene medarbejdere der har de fornødne forudsætninger, som anvender AI-assistenten.

Data der behandles, som led i brug af AI-assistenten, hverken lagres, bruges til læring el.lign. af AI-assistenten.

Tekniske aspekter

Platformen, som benyttes til at automatisere kodeskrivningen, består af en grundlæggende arkitektur med AI-understøttede forretningsprocesser, såsom automatiseret dataforberedelse. Platformen har en indbygget AI-governance, som understøtter styring og overvågning af AI-initiativerne fx overvågning af AI's præstation, revisionsspor for AI-beslutninger og kontrolmekanismer for at sikre, at Erhvervsstyrelsens brug af en AI assistent til kodeskrivning opfører sig som forventet. Desuden er platformen modeluafhængig, hvilket giver Erhvervsstyrelsen mulighed for at vælge en anden model alt efter, hvordan markedet udvikler sig.

Den store sprogmodel, som p.t. bruges til at automatisere kodeskrivning i Erhvervsstyrelsen, er GPT4 og bliver hostet i en af Erhvervsstyrelsens eksterne IT-leverandørers cloud. Selve kodebasen ligger lokalt hos Erhvervsstyrelsen.

IT-sikkerhed

Det kan være vanskeligt at bruge en standard udbyder af sprogmodeller (Chat-GPT, Microsoft Copilot, Bard m.m.) samtidig med, at man bevarer kontrollen over, hvilke forretningsdata, der deles med modeludbyderen.

Fokus for it-sikkerhed er ISO 27001 og de retningslinjer, som styrelsens direktion har vedtaget.

Grundig risiko- og konsekvensvurdering foretages for data for hvert projekt, der påtænkes effektiviseret ved de automatiserede kodeskrivninger.

Implementering i organisationen

Automatiseringen af kodeskrivning kræver ikke en transformation af interne arbejdsgange, hvorfor der ikke har været behov for at udvikle en implementeringsstrategi eller opbygge nye kompetencer hos ansatte.

Topdanmarks kundesvendte GenAI chatbot

Indledning

Topdanmark, en førende dansk forsikringsvirksomhed, har lanceret TopGPT, den første kundesvendte generative AI-assistent i den danske finanssektor. Formålet med TopGPT er at revolutionere kundeservice ved at tilbyde personlige interaktioner, der er skræddersyet ud fra kendskabet til hver kundes eksisterende forsikringer. Tidligere anvendte Topdanmark en regelbaseret chatbot, der manglede personalisering og krævede omfattende vedligeholdelse. Med TopGPT sikres effektiv, præcis og 24/7 kundeservice, hvilket har resulteret i ti gange højere kundetilfredshed end den tidligere regelbaserede chatbot, og danner grundlaget for yderligere forbedringer, gennem personaliserede svar baseret på kundens samtalehistorik og kendskabet til kundens forsikringsvilkår. Udfordringen i implementering af løsningen var i brugen af generativ AI i en stærkt reguleret branche, samtidig med at man overholder GDPR-krav og undgår bekostelige fejl.

Business Case

TopGPT er en kundespecifik AI-assistent, der har til formål at forbedre kundeservice gennem generativ AI. Målet er at sikre høj kvalitet og effektiv service lige ved hånden. TopGPT håndterer kundekontakt fra simple forespørgsler til avancerede forsikringsspørgsmål og leverer personlige svar baseret på individuelle kundedata. TopGPT evalueres via en række målbare KPI'er, heriblandt responstid, kundefeedback og -tilfredshed, omkostninger og intern auditering. TopGPT skal kunne interagere med eksterne brugere, såsom borgere og kunder.

Juridiske forhold

TopGPT er designet med henblik på at sikre høj datasikkerhed og overholdelse af GDPR's krav om databeskyttelse gennem design og standardindstillinger. Maskering af persondata fører til anonymisering, hvilket eliminerer risikoen for data-brud og uautoriseret adgang. Der er desuden implementeret mekanismer til at filtrere og identificere ondsindet brug, hvilket yderligere understøtter kravene til sikkerhed.

Før idriftsættelsen af TopGPT blev der gennemført en omfattende risikovurdering af løsningen. Denne proces involverede flere enheder, herunder en juridisk afdeling, IT-sikkerhed og et tech-board, som alle skulle bidrage med input til og evaluere løsningen i forhold til Topdanmarks eksisterende systemlandskab, IT-strategi, juridiske krav og sikkerhedsstandarder.

TopGPT er vurderet til at falde under kategorien begrænset risiko i henhold til AIA. Der er sikret gennemsigthed ved at informere brugerne om, at de interagerer med en AI-assistent, og hvad de kan forvente AI-assistenten kan hjælpe med.

Tekniske aspekter

TopGPT benytter et sofistikeret netværk af AI-modeller og databehandling, som hver især er designet til at håndtere forskellige aspekter af løsningen. Den tekniske platform er bygget efter en mixed model og modulær tilgang, således, at sprogmodellerne kan skiftes ud efterhånden som markedet udvikler sig over tid eller behovene ændrer sig. Infrastrukturen kører via Azure og udnytter de nyeste tilgængelige OpenAI modeller i Azure OpenAI.

Samarbejdende AI-model netværk: Den indledende fase af løsningen involverer flere AI-modeller, der i et netværk behandler indkomne beskeder. Denne fase starter med at maskere enhver brugerfølsom data, hvilket er afgørende for at opretholde GDPR-overholdelse og håndteres fuldt ud af modeller, der er hostet lokalt for at sikre datasikkerhed og privatliv. Desuden identificeres og filtreres eventuelle skadelige beskeder og uønskede elementer såsom spørgsmål om konkurrenter.

Intention og kontekst analyse: Efter bearbejdningen vurderer et andet sæt AI-agenter kundens intention. Uanset om kunden spørger om policedetaljer eller stiller generelle spørgsmål, klassificerer disse agenter intentionen. Klassificeringen af kundens intention er afgørende, da det direkte påvirker den efterfølgende informationshentning og strategien for besvarelse.

Retrieval Augmented Generation (RAG): TopGPT trækker information fra en vidensbase, som består af data fra vilkårsdokumenter, hjemmesidetekst og forsikringsspecifikke data fra forretningssystemer via API-kald. Når kundens intention er klar, hentes relevant data, inklusive dem specifikt relateret til kundens egne forsikringsaftaler. En anden gruppe agenter gennemgår derefter den hentede data og sikrer, at kun den mest relevante information til forespørgslen sendes videre til det næste lag af agenter.

Dynamiske interaktionsstrategier: Afhængigt af forespørgslen kompleksitet anvendes forskellige svarstrategier dynamisk i forhold til kundens klassificerede intention og den hentede data. Nogle AI-agenter kan svare direkte baseret på den hentede data, mens andre kan stille afklarende spørgsmål for at forbedre deres forståelse.

TopGPT modtager informationer fra forretningssystemer, men har ikke rettigheder eller egenskaber til at ændre i data. Kunden kan interagere med TopGPT både gennem Topdanmarks hjemmeside og mobilapp. TopGPT svarer ud fra de generelle forsikringsvilkår og dækninger, der gælder for kunden, men den har ikke oplysninger om, hvem den specifikke kunde er.

Implementering af ekstern AI-assistent

TopGPT sikrer at brugerne oplyses grundigt om dens funktioner og begrænsninger ved starten af hver chat. Brugerne informeres om, at de interagerer med en AI-assistent, der er under oplæring, og at de ikke skal dele personlige oplysninger. Brugerne får også at vide, at AI-assistenten fungerer som en digital assistent med viden om kundens privatforsikringer, mens specifik forsikringsrådgivning kræver kontakt via telefon eller e-mail. Ved at etablere en klar og gennemsigtig kernefortælling om, hvorfor og hvordan AI-assistenten anvendes, skabes der en samlet forståelse og tillid hos brugerne. Ydermere, er TopGPT trænet til ikke at svare på en måde, så den agerer som forsikringsformidler, men nærmere oplysende, og assisterende.

For at sikre en vellykket implementering af TopGPT har Topdanmark fokuseret på flere centrale områder:

Ekspertes til rådighed: Forretnings- og tekniske eksperter vejleder brugerne og vedligeholder assistenten.

Feedbackmekanismer: Brugerne kan rapportere fejl og stemme på svar, hvilket hjælper med at justere og forbedre AI-assistenten over tid.

Automatiseret overvågning: Logs og feedback overvåges proaktivt for at håndtere udfordringer. Regelbaseret og intelligent overvågning anvendes for at identificere og eskalere problemer.

Indbyggede kontrolmekanismer: Sikrer at brugerne får hjælp, selv når assistenten ikke kan levere tilfredsstillende svar. Modellen henviser til Topdanmark kundeservice, når det ikke er muligt at svare på kundes spørgsmål.

Løbende opfølgning: Ekspertes evaluerer regelmæssigt data for at optimere ydeevnen. Vidensgrundlaget opdateres løbende for at sikre præcision og ajourføring.

Afslutning

I begyndelsen af marts 2024 blev TopGPT lanceret. Siden lanceringen har TopGPT håndteret over 85.000 samtaler per 1. november 2024 og modtager ros for sin nøjagtighed og relevans. TopGPT har en auditeret korrekthed på fagspecifikke spørgsmål som performer langt bedre end forventet og overstiger bekvemt korrekthedsgrænserne som vi har opsat. Topdanmark har rapporteret en tidobling i antallet af positive anmeldelser sammenlignet med deres tidligere regelbaseret chatbot løsning.

Udviklingssamarbejdet mellem Topdanmark og Netcompany på TopGPT har også givet en værdifuld viden og erfaring for udviklingen af Netcompanys avancerede AI-plattform, EASLEY AI.

Assistent til effektivisering og kvalitetssikring af projektdokumentation i Netcompany

Indledning

For at sikre en konsistent levering af IT-ydelser til tiden anvender Netcompany en ensartet, struktureret tilgang til projektledelse, systemudvikling, og drift. Netcompanys metode standardiserer blandt andet de dokumenter, der produceres gennem implementeringen af et system, samt hvad dokumenterne skal indeholde. Tilgangen sikrer høj kvalitet af systemer og leverance til tiden med lav risiko, da planen for dokumentation er forudbestemt samt fleksibilitet, da et projekt nemmere kan trække på Netcompanys talentpulje, fordi processen for at rulle kollegaer ombord bliver simple.

Som led i Netcompanys strategiske fokus på at automatisere vigtige arbejdsgange gennem AI, har Netcompany udviklet en AI-assistent, der arbejder med dokumentering inden for tre aktiviteter:

1. Udarbejdelse: Assistenten skriver udkast på baggrund af medarbejderens noter, guidelines for det konkrete dokument samt eksisterende projektdokumentation.
2. Kvalitetssikring: Assistenten foreslår forbedringer til et dokumentets struktur og indhold, både sprogligt i form af termer og stavfejl, men også stemme et dokument overens med anden dokumentation på projektet.
3. Informationssøgning: Assistenten svarer på spørgsmål om et projekt ud fra dens dokumentation.

Assistenten er integreret i Netcompanys system til projektstyring, Toolkit, og benyttes flittigt af Netcompanys 8.000+ medarbejdere.

Business Case

Netcompany afsætter den nødvendige tid til at skrive god dokumentation og oplære nye medarbejdere heri. Ved at implementere dokumentationsassistenten på tværs af organisationen kan Netcompany opnå en mere effektiv og konsistent dokumentationsproces. Dette sparer tid, ressourcer og ensarter dokumentation på tværs af organisationen, så medarbejdere opnår fælles forståelse for, hvad god dokumentation er. Samtidig forbedrer assistenten projektleverancer og kundetilfredshed, fordi medarbejdere kan bruge mere tid på produktet eller blive hurtigere færdig med leverancen. Det var en vigtig prioritet, at tidsforbruget for dokumentation kunne nedsættes uden at forringe kvaliteten af dokumenterne.

Assistenten er bygget på Netcompanys AI-plattform, EASLEY AI, hvor Netcompanys andre assistenter ligeledes bygges.

Juridiske forhold

Assistenten er udviklet og implementeret under hensyntagen til EU's AI Act (AIA) og databeskyttelsesforordningen (GDPR). Assistenten er vurderet til at falde under kategorien ”begrænset risiko” under AIA. Denne klassificering er baseret på assistentens anvendelse til intern dokumentation og det faktum, at Netcompanys dokumentation ikke indeholder personfølsomme data. Brugere skal aktivt aktivere EASLEY AI, således de ikke er i tvivl om at de aktiverer en AI-assistent, hvilket opfylder AIAs krav om gennemsigtighed for systemer i denne risikokategori.

Selvom assistenten ikke behandler personfølsomme data, er assistenten udviklet til at overholde GDPR. Dette omfatter principper som dataminimering og formålsbegrænsning, så assistenten kun benytter data, som er nødvendig i henhold til assistentens formål. For at sikre sporbarhed logger EASLEY AI alle forespørgsler til assistenten. Dens svar gemmes sammen med de data, som assistenten finder og anvender. Dette muliggør løbende kvalitetskontrol og forbedring af assistenten.

Tekniske aspekter

EASLEY AI-plattform består af softwarekomponenter til udvikling, drift og monitorering af AI assistenter. Netcompany besluttede at udvikle en platform selv, da assistenterne dækker unikke, forretningskritiske processer.

Platformen håndterer sikkerhed, forberedelse af data og kommunikation med sprogmodeller samt projektværktøjer. EASLEY AI har en brugergrænseflade, hvor medarbejdere kan skrive dokumentation i samarbejde med assistenten. Den har ligeledes en integration til Word, så assistenten kan kommentere og foreslå rettelser direkte i dokumenter.

Arkitektur: EASLEY AI adskiller assistenten fra dens vidensbase og model. Dette muliggør simpel udskiftning mellem de tre: Flere assistenter kan benytte den samme information, så længe det er relevant for disse. Vidensbasen kan nemt opdateres, når et projekts dokumentation ændres og modellen kan skiftes ud, hvis en bedre model bliver tilgængelig i markedet.

Vidensbase: Med arkitekturen kunne Netcompany udvikle én dokumentationsassistent, der arbejder inden for konteksten af et projekt. Dens viden om et enkelt projekt kommer af tilknytningen mellem brugeren af assistenten og projektet, således assistenten ikke blander projekter sammen. EASLEY

AI benytter relevant projektdokumentation fra dokumenthåndteringssystemet samt håndbøger om Netcompanys arbejdsmetode. Vidensbasen opdateres automatisk, når et dokument tilføjes, ændres, eller slettes, så assistenten altid er opdateret med nyeste information. Medarbejderen kan kun benytte assistenten til de projekter, denne har adgang til. Assistenten angiver altid hvilke dokumenter, den har brugt til at formulere sit svar, så medarbejderen kan opsøge yderligere information på egen hånd eller validere assistens svar.

Integration: Søgning og kvalitetssikring med assistenten er integreret i Netcompanys dokumentbibliotek. Medarbejdere kan søge efter dokumenter, indhold og svar fra en chatboks på bibliotekets brugergrænseflade. Medarbejdere kan også efterspørge et review af et dokument. Assistenten indsætter kommentarer direkte i Word-dokumentet, så brugeren arbejder i den kontekst, der er relevant for dem.

Implementering i organisationen

For at sikre en korrekt udrulning af assistenten i organisationen, gennemgik Netcompany tre trin:

Pilotforløb: Netcompany gennemførte to pilotforløb for at teste og finjustere assistenten inden den blev udrullet til hele virksomheden. Udover forløbende gav indsigt i uforudsete brugsmønstre, der understøttede bedre kvalitetssikring af modellen, opbyggede det også to grupper ambassadører, der kunne tale for assistenten og hjælpe kollegaer med at tage den i brug.

Udmelding: På Netcompanys intranet blev der opslået en nyhed om, at assistenten var tilgængelig. Desuden blev der udarbejdet en fast side, der kort præsenterede selve assistenten og en håndbog omkring at ibrugtage assistenten.

Opfølgning: Efterfølgende blev der aktivt opsat aktiviteter til at følge op med brugen af assistenten. Ud over mængden af forespørgsler mod assistenten, blev der også opsamlet kvalitativt input.

Den gradvise udrulning og fokus på brugerinddragelse bidrog til en positiv modtagelse og integration af assistenten i de daglige arbejdsprocesser.

Afslutning

Dokumentationsassistenten blev lanceret i juni 2024. I den første måned har 240 unikke brugere fordelt på 30 initierende projekter benyttet dokumentassistenten til enten at udarbejde dokumentation eller få information om projektets dokumentation. Assistenten har i denne periode kvalitets-sikret 150 dokumenter. Netcompany har for disse projekter effektiviseret procesarbejdet omkring udarbejdelse og godkendelse af dokumentation med 20%, udover gevinsterne i det daglige administrationsarbejde.

Regnskabsanalyse — automatiseret med GenAI

Indledning

Spar Nord – en af Danmarks største banker – etablerede tilbage i maj 2023 afdelingen Virksomhedsanalyse. Regnskabsanalyserne af bankens erhvervskunder har tidligere været en opgave, der har ligget decentralt hos de lokale banker, men opgaven er nu blevet delvist centraliseret, hvor større og risikofyldte engagementer analyseres i denne centrale afdeling.

Virksomhedsanalyse har til formål at registrere regnskabsdata og udarbejde regnskabsanalyser på bankens erhvervskunder, hvilket skal understøtte:

- En mere effektiv proces for bankens regnskabsregistrering og -analyse
- En forbedring af datakvaliteten for regnskabsdata
- At højne kvalitet i beslutningsgrundlag gennem ensartede og præcise analyser

I Virksomhedsanalyse er der i efteråret 2024 udviklet en AI-assistent, der automatiserer regnskabsanalyserne. Hensigten med assistenten har været at understøtte målsætningen med etableringen af Virksomhedsanalyse ved at effektivisere og muliggøre det centrale bidrag til regnskabsanalyser i markant større skala end ved traditionelle manuelle metoder.

Business Case

AI-assistenten til regnskabsanalyse har til formål at effektivisere, forbedre og ensarte regnskabsanalyser gennem anvendelsen af generativ AI (GenAI).

Løsningen er tilgængelig for medarbejdere i Virksomhedsanalyse og skal agere assistent, der leverer et udkast til regnskabsanalyser med udgangspunkt i regnskabsoplysninger fra bankens egne systemer og offentlige kilder.

Analytikerne i Virksomhedsanalyse arbejder dernæst videre med udkastet, som leveres i Word-format, til den respektive erhvervsrådgiver, der efterfølgende har mulighed for at berige regnskabsanalysen med oplysninger, der opnås gennem dialog med kunden.

Med formålet om at effektivisere regnskabsanalyser har fokus været at reducere tidsforbruget uden at forringe kvaliteten af analyserne samt at nedbringe risici for menneskelige fejl. Det er vurderingen, at AI-assistenten gennemsnitligt kan bidrage til minimum at halvere det samlede tidsforbrug ved udarbejdelse af regnskabsanalyser. Reduktion af tidsforbruget har

også medført hurtigere sagsbehandling og dermed bidraget til en forbedret kundeoplevelse.

Juridiske forhold

Forud for idriftsættelse af AI-assistenten er der sket en omfattende vurdering heraf, bl.a. vurdering af IT-sikkerhed, databeskyttelse, dataetik, datakvalitet og operationelle risici, herunder modelrisiko.

I arbejdet med assistenten har bankens AI-udvalg været involveret, og en lang række afdelinger har bidraget til de forskellige risikovurderinger.

Risici forbundet med anvendelsen

Der er identificeret en række risici forbundet med assistenten. En af de potentielt væsentligste risici er, at beslutningstagere ikke forholder sig tilstrækkeligt kritisk over for de AI-genererede analyser. Denne risiko er håndteret ved at opsætte tydelige rammer for brugernes stillingtagen til analyserne, herunder at sikre opmærksomhed på eventuelle fejl og hallucinationer til trods for en velformuleret og overbevisende analyse. Dertil foretages løbende uvildig validering af de AI-genererede output.

Der er desuden en operationel risiko relateret til brug og vedligeholdelse af assistenten, som er håndteret gennem interne change management procedurer samt beskyttelse af prompten mod ændringer.

Tekniske aspekter

AI-assistenten består af flere elementer, der hver især bidrager til et solidt og driftssikkert system. Datagrundlaget for analysen kommer fra bankens egne systemer og offentlige kilder (CVR).

Herudover er der bygget forskellige skabeloner afhængig af, om der er tale om f.eks. driftsselskaber, holdingselskaber eller ejendomsselskaber. Hver skabelon indeholder en række forskellige prompts, der fokuserer på de enkelte områder i analysen, bl.a. indtjeningssevne og kapitalforhold. Herudover er der defineret en række overordnede instrukser, der tager udgangspunkt i bankens kreditpolitik og instrukser for regnskabsanalyser.

Prompts styres centralt og kan ikke redigeres af brugerne af AI-assistenten. Dette for at sikre høj datakvalitet og validering

af AI-assistentens generelle performance. Brugerne skal alene indtaste et CVR-nummer i applikationen, hvorefter regnskabsanalysen automatisk genereres.

Der anvendes ikke GenAI til udledning af regnskabstal eller beregning af nøgletal. Dette sker med almindelige algoritmer. GenAI finder således primært anvendelse i den verbale formulering af regnskabsanalysen, herunder forklaring af udvikling i nøgletal og identifikation af opmærksomhedspunkter.

I forhold til anvendelsen af GenAI anvendes Azure OpenAI. Herudover bruges Microsoft som en integreret del af assistenten. F.eks. eksporteres den endelige analyse til et Word-dokument, der kan redigeres og beriges yderligere af en erhvervsrådgiver, inden den indgår som en del af en låneindstilling.

Implementering i organisationen

Som beskrevet er AI-assistenten implementeret i afdelingen Virksomhedsanalyse, og brugen er dermed begrænset til ansatte i denne afdeling. De AI-generede regnskabsanalyser vil derfor altid have ”human in the loop”, inden de deles som udkast med erhvervsrådgiverne i de lokale banker. For at sikre en korrekt udrulning af anvendelsen af AI-assistenten i organisationen, er implementeringen opdelt i fire trin:

Trin 1

Intern test i afdelingen

AI-assistenten blev testet internt i Virksomhedsanalyse på regnskabsanalyser, som udarbejdes centralt.

Trin 2

Pilotforløb

Fem større erhvervsafdelinger har deltaget i et pilotforløb, hvor erhvervsrådgivere har bidraget til en bredere test af AI-genererede regnskabsanalyser.

Trin 3

Bred implementering

Generel implementering i alle lokale banker påbegyndt i november.

Trin 4

Opfølgning/feedback

Afdelinger og analytikere kommer løbende med forbedringsforslag og generel feedback, som Virksomhedsanalyse indarbejder i AI-assistenten.

Der følges løbende op på brugen af AI-assistenten via statistik over antallet af generede udkast til decentrale analyser i de lokale banker. AI-assistenten vil løbende blive optimeret og videreudviklet. Perspektiverne er mange, bl.a. udvidelse med henblik på, at AI-assistenten kan bruges hos de enkelte erhvervsrådgivere – uden involvering af Virksomhedsanalyse. Dette vil dog stille nye og andre krav til AI-assistenten, lige som det kan være forbundet med andre risici.

Afslutning

Spar Nord har gennem lang tid ønsket at automatisere arbejdet med regnskabsanalyse og har gennem arbejdet med AI-assistenten opnået mange gode erfaringer.

Spar Nord har med AI-assistenten mere end halveret tidsforbruget for regnskabsanalyser ved visse virksomhedsformer og brancher. AI-assistenten skal løbende optimeres, således samme effektivisering opnås ved øvrige virksomhedsformer og brancher.

Herudover vurderes det, hvordan AI-assistenten kan understøtte arbejdet med regnskabsanalyser generelt i banken.