

Policy for utvikling og bruk av KI i Skatteetaten

Versjon 1.1
20. november 2023

Dokumentansvarlig: Harald Johannessen

| Versjon | Status | Merknad |
|---------|----------------------------------|--|
| 0.1 | Dokument opprettet | |
| 0.7 | Utkast til orientering i ELM | |
| 0.8 | Utkast til utvidet arbeidsgruppe | |
| 0.9 | Høringsutkast, intern høring SKE | |
| 0.95 | Drøftingssak Fellesmøtet | |
| 1.0 | Beslutningssak ELM | |
| 1.1 | Endelig versjon | Mindre språklige justeringer i tråd med tilbakemeldinger fra ELM |

Innhold

| | | |
|------|--|----|
| 1. | Innledning..... | 3 |
| 1.1. | Om dokumentet | 3 |
| 1.2. | Policyens virkeområde og avgrensninger..... | 3 |
| 1.3. | Regelverk | 4 |
| 1.4. | Sentrale begreper | 5 |
| 2. | Prinsipper for utvikling og bruk av KI i Skatteetaten..... | 6 |
| 2.1. | Ansvarlighet..... | 6 |
| 2.2. | Åpenhet | 6 |
| 2.3. | Rettferdighet | 8 |
| 2.4. | Teknisk robusthet, sikkerhet og beredskap | 8 |
| 2.5. | Datastyring og personvern | 9 |
| 3. | Anvendelse av prinsipper | 9 |
| 3.1. | Bruk av KI i oppgaveløsningen..... | 9 |
| 3.2. | Fra egenutviklet KI til bruk av standardverktøy | 12 |
| 3.3. | Utvikling og utforskning av KI-bruk | 12 |
| 4. | Forvaltning og revisjon | 13 |

1. Innledning

1.1. Om dokumentet

Dette dokumentet beskriver Skatteetatens policy for utvikling og bruk av kunstig intelligens (KI). Hva som omfattes av begrepet kunstig intelligens for denne policyen er nærmere omtalt i punkt 1.4 under.

Skatteetaten har som mål å utnytte det fulle potensialet som ligger i digitalisering. Økt bruk av data gir oss innsikt til å treffe bedre beslutninger i saksbehandlingen og i øvrige oppgaver. I Skatteetatens strategi er det ambisjoner om å forvalte informasjon slik at den bidrar til forenkling og digitalisering. Videre skal etaten bruke data, analyse og maskinlæring i stor skala i oppgaveløsningen¹.

KI er et område innenfor informasjonsteknologien som utvikler seg raskt. Offentlige myndigheters utvikling og bruk av KI innebærer muligheter for å ta ut gevinster, men det er også en viss risiko forbundet med å ta i bruk denne type teknologi. Kunstig intelligens kan gi store økonomiske og velferdsmessige fordeler for samfunnet, men kan samtidig øke risikoen for diskriminering og myndighetsmisbruk. Det er derfor avgjørende for Skatteetaten å sikre at vi anvender kunstig intelligens på en måte som er lovlig, rettferdig og ansvarlig.

Denne policyen definerer fem grunnleggende prinsipper for Skatteetatens utvikling og bruk av KI. Prinsippene skal være retningsgivende i alle situasjoner der etaten tar i bruk KI, og har til hensikt å gi overordnede føringer for det som på sikt skal bli et helhetlig rammeverk for kunstig intelligens i Skatteetaten. De fem prinsippene er:

- Ansvarlighet
- Åpenhet
- Rettferdighet
- Teknisk robusthet, sikkerhet og beredskap
- Datastyring og personvern

1.2. Policyens virkeområde og avgrensninger

Policy for utvikling og bruk av KI i Skatteetaten tar utgangspunkt i etatens overordnede styrende dokumenter, og gjelder innenfor rammene av øvrig regulatorisk rammeverk.

Policyen står på egne ben, men skal samtidig anvendes i tråd med andre sentrale styringsdokumenter og prinsipper i etaten. Her inngår f.eks.:

- *Styringssystem for sikkerhet og beredskap.*
- *Rammeverk for styring og forvaltning av data.*
- *Rammeverk for helhetlig risikostyring*
- *Juridisk instruks*
- *Retningslinje for å sikre etterlevelse av personvernregelverket*

¹ <https://www.skatteetaten.no/strategi>

Ovennevnte utgjør fundamentet for denne policyen. Policy for utvikling og bruk av KI omhandler først og fremst det som er særlig viktig å ta hensyn til ved bruk av KI.

1.3. Regelverk

Foreløpig har vi ingen gjeldende regelsett som eksplisitt regulerer utvikling og bruk av KI. Dette vil imidlertid endre seg på sikt, og det gjennomføres for tiden et arbeid med regelverksutvikling både nasjonalt og internasjonalt. For eksempel forventer EU å ha et regelsett (AI Act) klart i løpet av 2023 med forventet ikrafttredelse tidligst innen utgangen av 2025. Se nærmere omtale under.

Inntil ny spesiallovgivning kommer på plass vil de generelle reglene i forvaltningsregelverket og personvernregelverket, samt grunnleggende menneskerettigheter, utgjøre det sentrale regulatoriske rammeverket for Skatteetatens policy for utvikling og bruk av KI. Dette regelverket stiller for eksempel krav til hjemmelsgrunnlag, åpenhet, begrunnelse, kontradiksjon, klagerett, informasjon og innsyn, taushetsplikt, diskrimineringsforbud, rett til privatliv og forbud mot myndighetsmisbruk. Videre er Skatteetaten underlagt reglene i offentleglova om retten til offentlig innsyn og reglene om arkivplikt i arkivregelverket. Dette er regler som også kommer til anvendelse ved bruk av KI. I tillegg finnes andre relevante regelsett, som eks. regler for opphavsrett og andre immaterielle rettigheter som Skatteetaten må følge.

Policyen er utformet innenfor rammene som følger av gjeldende rett, men er ikke ment som en uttømmende kodifisering. Det betyr at utvikling og bruk av KI på vanlig måte må vurderes opp mot gjeldende regelverk, selv om prinsippene i policyen følges.

Særlig om EUs forordning for kunstig intelligens (AI Act)

Den 21. april 2021 la Europakommisjonen frem forslag til Artificial Intelligence Act (AI Act).

Forordningen skal gi en harmonisert regulering av KI til bruk i EU/EØS. Forslaget skal både tilrettelegge for teknologisk utvikling og innovasjon innenfor KI-området, samtidig som grunnleggende rettigheter og sikkerhet skal ivaretas.

Den foreslåtte KI-forordningen tar en risikobasert tilnærming når det gjelder ulike KI-systemer, hvor systemer med uakseptabel risiko er forbudt og de med høy risiko er strengt regulert. KI-systemer som utgjør en begrenset risiko, vil måtte oppfylle enkelte åpenhetskrav. For systemer som ikke faller inn under noen av de andre kategoriene (minimal risiko), oppstiller KI-forordningen ingen krav.

Det er lite som tyder på at Skatteetaten vil ta i bruk KI som vil omfattes av områdene for uakseptabel risiko. Når endelig forordningstekst foreligger må det imidlertid foretas en nærmere vurdering av hvorvidt Skatteetaten benytter, eller planlegger å ta i bruk, KI på områder som klassifiseres som høy risiko og som dermed pålegges mer omfattende krav.

Forslaget behandles nå i EU-systemet, hvor det pågår såkalte trilog-forhandlinger mellom Europaparlamentet, Rådet og Kommisjonen. Det forventes en beslutning i løpet av 2023, hvorefter forordningen tidligst vil kunne tre i kraft i løpet av 2025 i EU, og mulig enda senere i EØS.

1.4. Sentrale begreper

Her omtales noen sentrale begreper som ligger til grunn for tankegangen bak og forståelsen av Skatteetatens policy for utvikling og bruk av KI.

Kunstig intelligens

Det foreligger ikke en allmenngyldig, universell definisjon av *KI* og alternative definisjoner endrer seg gjerne i takt med hva som er teknologisk mulig. Følgelig kan et forsøk på å definere KI heller ikke sette et skarpt skille for avgrensningen av policyen. Det viktigste er de særegne egenskapene ved KI og hvordan vi skal forholde oss til disse i Skatteetaten. Definisjon av KI som anvendes her legger seg tett inntil definisjonen i EU AI Act og bør oppdateres når hensiktsmessig.²

«Kunstig intelligenssystem" (KI-system) betyr et maskinbasert system som er designet for å operere med varierende nivåer av autonomi, og som kan generere utdata som for eksempel prediksjoner, anbefalinger eller beslutninger, med eksplisitte eller implisitte mål, og som påvirker fysiske eller virtuelle miljøer.»

Maskinlæring

KI som fagområde favner bredere enn *maskinlæring* (ML), men all maskinlæring faller inn under området KI. ML er en vitenskapelig disiplin opptatt av design og utvikling av algoritmer som gjør datamaskiner i stand å lære fra og utvikle atferd basert på empiriske data. I denne konteksten er algoritmer avgrenset til maskinlæringsalgoritmer som er spesifikt utviklet for å lære fra data og trekke ut mønstre eller innsikt uten å være programmert med eksplisitte regler. Det finnes i denne kontekst flere typer algoritmer, herunder også nevrale nettverk/dyplæring. Treningen av algoritmene medfører ofte deler som vanskelig lar seg forklare. Ettersom algoritmene ofte er en sentral komponent av beslutningsstøttesystemer, er maskinlæring også et særlig viktig område innen KI for Skatteetatens policy.

Generativ KI

Betegnelsen generativ KI har noen spesifikke kjennetegn ved at systemet kan generere tekst, bilder og andre media basert på semantiske instruksjoner («prompter»). Disse er som regel basert på dype nevrale nettverk. I sin tidlige form ble disse typisk benyttet til spesifikke oppgaver som klassifisering av eksisterende datasett. I dagens omtale har de en bredere funksjon primært knyttet til store språkmodeller (LLM) som blir trent med generelle tekstproduserende egenskaper³. Disse kan videre spisses mot avgrensede datasett/områder. Ettersom språk er en generell egenskap blir dette også tidvis knyttet til, men er på ingen måte det samme som, generell kunstig intelligens⁴. Dersom klassisk KI skal

² I arbeidet med EU AI Act er det foreløpig lagt frem forslag til tre ulike legaldefinisjoner. Skatteetaten har foreløpig lagt til grunn definisjonen som er foreslått av Parlamentet ettersom denne etter vårt syn fremhever de særegne egenskapene ved KI på best måte. Den engelske teksten som Parlamentet la frem lyder slik:

“Artificial intelligence system’ (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments;”

³ Disse modellene omtales gjerne som «foundation models» / grunnmodeller.

⁴ Generell kunstig intelligens refererer til en tilstand der et maskinsystem kan utføre enhver fysisk og kognitiv handling på samme nivå som et menneske. Dette er omdiskutert tema og er ikke et element som Skatteetaten tar stilling til eller omfatter i policyen.

defineres som motsats til generativ KI, vil denne fortelle/agere på grunnlag av eksisterende data, mens generativ KI vil skape data som ikke eksisterer. For Skatteetaten og denne policyen er det sentralt at vi har en bevisst tilnærming til nytte, begrensninger, og risikoer knyttet til bruk av de genererte dataene.

2. Prinsipper for utvikling og bruk av KI i Skatteetaten

2.1. Ansvarlighet

Skatteetaten skal ta i bruk KI på en tillitvekkende og ansvarlig måte. Vi skal alltid se hen til vårt samfunnsoppdrag, vår rolle i forvaltningen og den maktposisjon vi har overfor borgerne. All utvikling og bruk av KI-baserte løsninger skal skje i tråd gjeldende regelverk, god forvaltningsskikk, etiske retningslinjer, og prinsippene i denne policyen.

Skatteetatens KI-systemer må operere innenfor det regelverket som er gjeldende til enhver tid. Dette gjelder særlig, men er ikke begrenset til; internasjonalt regelverk med relevans for Norge, sektorspesifikke lover med særregulering for KI, rammeregulering for bruk og behandling av data og personopplysninger, lover som regulerer forvaltningen, offentlighet og arkiv, lover som ivaretar åndsverk og immaterielle rettigheter, samt konvensjoner og lover som ivaretar grunnleggende menneskerettigheter.

Det skal sikres nødvendig grad av menneskelig tilsyn og kontroll med KI-systemene, jf. pkt. 2.3 nedenfor. Det må etableres klare ansvarsområder og roller for alle involverte parter i utviklingen, implementeringen og bruken av KI-systemet. Dette inkluderer identifisering av ansvarlige personer, organisatoriske enheter eller team for overvåking av systemets ytelse, håndtering av risikoer og sikkerhetsbrudd, samt forvaltning av data og personvern.

Skatteetaten må påse at det finnes et styringssystem som skal sikre nødvendig kvalitet, risikostyring og internkontroll ved utvikling og bruk av KI. Styringssystemet skal dokumentere at Skatteetaten etterlever kravene i regelverket og interne retningslinjer. Styringssystemet innarbeides som en del av Skatteetatens øvrige prosesser for kvalitet og styring.

Det skal gjennomføres regelmessige evalueringer og revisjoner av KI-systemer for å vurdere lovligheten, effektiviteten, etikken og overholdelsen av interne retningslinjer. Det skal tas hensyn til nye funn, forskning eller lovendringer som kan påvirke systemets ansvarlighet og gjøres nødvendige justeringer og forbedringer.

2.2. Åpenhet

Åpenhet i offentlig sektor bidrar blant annet til å sikre kontroll med forvaltningen, det skaper notoritet for beslutningsprosesser og bidrar til å bygge tillit i befolkningen. For Skatteetaten er det derfor viktig å kunne forklare og dokumentere hvordan beslutninger tas, også der hele eller deler av beslutningsgrunnlaget er basert på bruk av KI.

All utvikling og bruk av KI-systemer skal skje på en måte som ivaretar åpenhet og gjennomsiktighet gjennom hele prosessen.

Mellom utvikler og bruker av et KI-system⁵, er det sentralt å sikre at det foreligger et nødvendig nivå av sporbarhet og forklarbarhet til KI-systemet. Dersom Skatteetaten skal kunne ta i bruk KI på en ansvarlig måte, så må det være mulig for oss å forstå hvordan KI-systemet er bygget opp og hvordan det genererer resultater, beslutninger og anbefalinger. Det er utvikleren av et KI-system som må sikre at sporbarhet og forklarbarhet er ivaretatt, og brukeren må påse at kravene er innfridd før systemet settes i produksjon.

Sporbarhet for data og KI-systemets aktiviteter, vil bidra til å forstå prosessen som leder til KI-systemets anbefalinger. Sporbarhet er også viktig for å oppdage feil, hendelser eller endringer i KI-systemet gjennom systemets livsløp. Funksjoner for automatisert logging, overvåking og oppfølging av endringer i datagrunnlaget og andre hendelser i KI-systemet, kan bidra til å sikre god sporbarhet.

Forklarbarhet handler om å kunne forstå og redegjøre for hvordan KI-systemet gir sine anbefalinger eller er kommet til et gitt resultat basert på inndata. Forklarbarhet i KI-systemet kan in mange tilfeller oppnås ved å se hen til anerkjente og utprøvde forklarbarhetsmetoder⁶ innenfor KI-teknologi.

Overfor mottakere⁷ vil det være spesielt viktig å legge til rette for at vedkommende kan forstå at de kommuniserer med, eller at en beslutning er generert av, teknologi basert på KI med de muligheter og begrensninger det kan innebære. Spesielt der en KI-generert vurdering eller beslutning vil kunne påvirke mottakeren negativt, må vedkommende settes i stand til å få informasjon om KI-systemet som er brukt, herunder hvilke avgjørende faktorer, data og logikk som ligger til grunn. Dette er nødvendig for at mottaker skal kunne ivareta sine egne rettigheter og interesser, eller for å kunne utfordre en avgjørelse. Informasjon om KI-systemet må gis på en meningsfull og forståelig måte overfor mottakeren. Om slik informasjon skal gis proaktivt, reaktivt eller som en kombinasjon må vurderes ut fra en samlet vurdering der viktige elementer vil være summen av informasjonstiltak i sin helhet, hvilket virksomhetsområde KI benyttes på, og ev. konsekvenser bruken kan gi for mottakeren.

Dersom Skatteetaten tar i bruk KI direkte i sin kommunikasjon med mottakere, f.eks. i form av veiledningstjenester eller fastsettelse av vedtak, så vil også åpenhetskrav i forvaltningsretten og personvernregelverket komme til anvendelse. Fra forvaltningsretten kan dette være rett til kontradiksjon, begrunnelse, partsinnsyn og overprøving. Med mindre unntak⁸ gjør seg gjeldende, så har privatpersoner også rett til innsyn og informasjon om Skatteetatens behandling av personopplysninger. Dette vil omfatte informasjon om hvorvidt en beslutning er basert på helautomatisert behandling og profilering, samt logikken bak avgjørelsen.

⁵ I mange tilfeller vil Skatteetatens rolle være bruker av et KI-system, der dette er utviklet av en tredjepart. I noen tilfeller vil imidlertid Skatteetaten kunne opptre som både utvikler og bruker av et KI-system.

⁶ Eksempler på forklarbarhetsmetoder som Digdir viser til i sin [veileder](#) er LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), Heat maps, Saliency maps, Konseptdeteksjon og Kontrafaktiske forklaringer

⁷ Mottaker er en fysisk eller juridisk person som mottar en anbefaling eller beslutning generert ved hjelp av KI. Mottaker kan være både eksterne og interne personer avhengig av kontekst og anvendelsesområde.

⁸ Unntak finnes i personopplysningsloven §§ 16 og 17, personvernforordningen art. 13 nr. 4, art. 14 nr. 5, og art. 15 nr. 4.

2.3. Rettferdighet

KI-systemer skal utvikles og brukes på en rettferdig måte, som også respekterer grunnleggende menneskerettigheter og understøtter Skatteetatens etiske retningslinjer. Skatteetatens bruk av KI skal baseres på verdier som verdighet, inkludering, likhet, mangfold, menneskers autonomi og rett til privatliv. Tilsvarende skal vi avstå fra å bruke KI på en måte som er egnet til å diskriminere, manipulere eller på annen måte undertrykke enkeltpersoner eller sårbare grupper i samfunnet.

For at Skatteetaten skal kunne bruke KI på en rettferdig måte i tråd med verdiene ovenfor, er det viktig å vite hvilke typer risiko et KI-system vil kunne utgjøre mot grunnleggende rettigheter. Skatteetaten skal påse at mulige risikoer knyttet til slike rettigheter og verdier identifiseres og vurderes løpende både på utviklingsstadiet og under hele KI-systemets livsløp.

Skatteetaten skal påse at utvikling og bruk av KI skjer med nødvendig grad av menneskelig tilsyn og kontroll⁹, sett hen til formålet med systemet. Menneskelig tilsyn og kontroll kan bidra til å sikre at et KI-system ikke undergraver menneskers autonomi eller forårsaker andre uforutsette negative virkninger. Menneskelig tilsyn og kontroll er spesielt viktig der KI benyttes som del av oppgaveløsningen til Skatteetaten som kan få direkte konsekvenser for både fysiske og juridiske personers rettsstilling.

Ved trening, testing og validering av KI-systemer må det gjennomføres tiltak som skal sikre at systemet ikke gir urettferdige eller diskriminerende utfall som følge av skjevheter («bias») i algoritmen eller dataene. Det må implementeres tiltak for å motvirke, oppdage og rette slike skjevheter i KI-systemets levetid.

2.4. Teknisk robusthet, sikkerhet og beredskap

KI-systemer må oppfylle et egnet nivå for nøyaktighet, robusthet og informasjonssikkerhet sett hen til tiltenkt bruk/formål gjennom systemets levetid. Dette er tiltak for å sikre pålitelighet til KI-systemet, som igjen er nødvendig for at vi skal kunne stole på at den fungerer i tråd med intensjonen.

Skatteetaten skal identifisere trusler, samt gjennomføre konsekvens- og sårbarhetsanalyser for å minimere risikoen for at uønskede eller utilsiktede hendelser påvirker sikkerheten og/eller robustheten til KI-systemet. Herunder skal KI-systemet sikres mot misbruk, endring eller manipulering. Videre skal det implementeres egnede tiltak for å sikre konfidensialitet, integritet og tilgjengelighet av dataene som brukes av KI-systemet. Dette kan omfatte kryptering, tilgangskontroll, logging og sikker datalagring.

Det skal utføres grundig trening, testing og validering av KI-systemet for å sikre at det oppfyller robusthets- og sikkerhetskravene. Dette inkluderer testing av systemet mot angrep, feil- og hendeshåndtering, samt avdekking av mulige svakheter i KI-systemet. Det bør også etableres en kontinuerlig overvåkingsprosess for å identifisere kvalitetssvikt, uønsket atferd eller potensielle sikkerhetsbrudd i KI-systemet.

⁹ Menneskelig tilsyn og kontroll («Human oversight») kan oppnås gjennom styringsmekanismer som baseres på f.eks. human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC). Se også AI Act art. 14 som også omtaler krav til Human Oversight for høy-risiko KI-systemer.

Ansatte i Skatteetaten som tar del i utvikling og/eller bruk av KI-baserte løsninger må få tilstrekkelig opplæring for sikker bruk av systemet. Dette kan inkludere opplæring i sikkerhet, personvern, forvaltning av data og rapportering av mistenkelige og uønskede hendelser (avvik).

2.5. Datastyring og personvern

God styring og forvaltning av data er sentralt ved utvikling og bruk av KI-systemer. Dette er viktig både for at KI-systemet skal kunne fungere effektivt og pålitelig, men også for at Skatteetaten skal bevare kontrollen over egne data. Videre skal Skatteetaten sikre at taushetsplikt, retten til privatliv og personvern prioriteres når vi jobber med KI.

Data, eller datasett, som skal benyttes til trening, validering og testing av KI, skal underlegges datastyrimmekanismer som står i forhold til den tiltenkte bruken.

Skatteetaten skal ha et bevisst forhold til hvilke data som er nødvendige å bruke for de ulike stegene ved utvikling og bruk av et KI-system. Vurderingene skal dokumenteres. Ved trening, validering og testing av KI-systemet, skal Skatteetaten påse at dataene som benyttes opprettholder tilstrekkelig nivå av datakvalitet og integritet. Dataene som benyttes må være så fullstendige, relevante og oppdaterte som mulig sett hen til formålene med bruken. I tillegg bør dataene inneholde egnede statistiske egenskaper.

Skatteetaten skal ikke ta i bruk KI på en måte som utgjør et uforholdsmessig inngrep i de private integritet eller negativt påvirker den personlige autonomi. KI-systemer som forutsetter behandling av personopplysninger enten under utvikling eller i produksjon, må følge kravene i personopplysningsloven og personvernforordningen.¹⁰ Av særlig viktighet for Skatteetatens utvikling og bruk av KI fremheves kravet til rettslig behandlingsgrunnlag, prinsippene om formålsbegrensning og dataminimering, krav til personvernkonsekvensvurdering, innebygget personvern og forbudet mot helautomatiserte avgjørelser.

Skatteetaten skal minimere mengden personopplysninger og konfidensielle data som behandles i KI-systemet, ved å vurdere alternative metoder og teknikker (f.eks. gjennom bruk av syntetiske og/eller anonymiserte data) som kan oppnå ønsket formål eller funksjonalitet. Slike metoder må likevel avveies mot eventuelle negative konsekvenser for kvalitet, robusthet og pålitelighet i KI-systemet.

Opplysninger som er gradert etter beskyttelsesinstruksen eller sikkerhetsloven skal som hovedregel ikke benyttes ved utvikling eller bruk av KI. Dette skyldes det ekstraordinære skjermingsbehovet som alltid gjør seg gjeldende for slike opplysninger. Det må iverksettes tiltak for å motvirke utilsiktet avsløring av slike graderte opplysninger som kan oppstå ved kobling av ulike datasett som inneholder geolokalisering opplysninger.

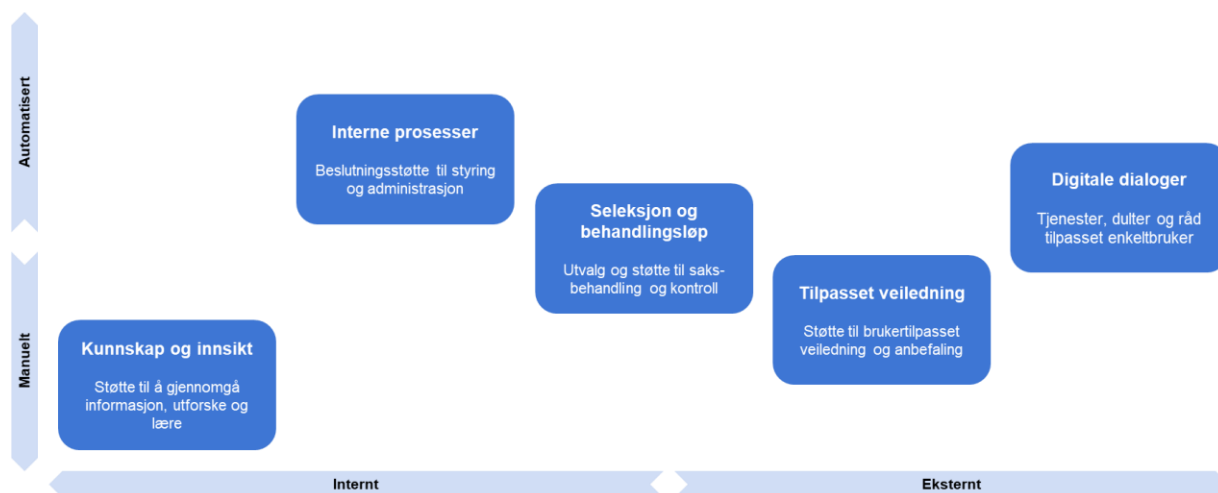
3. Anvendelse av prinsipper

3.1. Bruk av KI i oppgaveløsningen

Gjeldende og fremtidig anvendelse av KI i Skatteetaten kan grupperes på ulikt vis. For anvendelse av prinsippene i policyen, er det særlig relevant å se på hvor direkte bruken av KI treffer mottaker, samt i

¹⁰ Se også [Intranett - Retningslinje for å sikre etterlevelse av personvernregelverket \(sits.no\)](https://sits.no)

hvilken grad oppgaven vil kunne være gjenstand for automatisering. Forenklet sett er det da fem ulike oppgaveområder som peker seg ut, se Figur 1.



Figur 1: Områder i etaten der KI/maskinlæring blir benyttet/kan benyttes, etter grad av automatisering og eksternt orientering.

Skatteetatens erfaring med bruk av KI springer særlig ut fra to av disse oppgaveområdene. For det første *Seleksjon og behandlingsløp*, der automatiserte utvalg er grunnlag for ulik håndtering. Etter manuell filtrering og kartlegging muliggjør det mer målrettet kontroll og saksbehandling. Tilnærmingen har vært benyttet bredt innen fastsetting av skatt, avgift, innkreving og andre områder. For det andre *Digitale dialoger*, der skattepliktig dultes og mottar råd direkte i våre tjenester, tilpasset hver enkelt. Sistnevnte område er mer automatisert, i den forstand at dulter eksponeres uten mellomliggende saksbehandling.

Tilpasset veiledning kan betraktes som både mykere og mer manuelt støttet enn de digitale dialogene. I veiledning har etaten først og fremst erfaring med bruk av statistiske modeller, samtidig som KI-systemer, inkludert løsninger med generativ KI, prøves ut. Innen rent *Interne prosesser* er det potensial til å benytte mer automatisert KI – både til å støtte IT-utvikling og kontorstøtte, og som beslutningsstøtte inn mot styring. Oppgaver innen *Kunnskap og innsikt* er også interne, men innebærer trolig mer manuell bruk av KI til analyse og informasjonsutforskning, kontrastert mot potensialet for automatisering som ligger i administrative støtteprosesser.

Anvendelse av prinsippene på ulike oppgaveområder

Avsnittene nedenfor er eksempler på hvordan prinsippene kan tolkes på ulike områder, der oppgaven løses med hjelp fra KI. Det skal utarbeides ytterligere styringsdokumenter som utdyper hvordan Skatteetaten tolker prinsippene og implementerer dem i prosesser for utvikling og bruk av KI.

Ansvarlighet: Alle oppgaveområder krever kompetanse på hva innebygget KI vil kunne bety, og forståelse av hvor ansvaret til enhver tid ligger. Det inkluderer tydelig plassering av ansvar for oppfølging av risiko, lovlighet, kvalitet og ytelse på områder der KI anvendes. Der Skatteetaten anvender KI i eksterne dialoger, er det viktig å vurdere hvordan etaten møter mottakere samlet sett. Ulike løsninger og modeller kan forholde seg til prinsippene hver for seg, men vi bør også vurdere den samlede innsatsen mot enkelte skattytere.

Åpenhet: Der KI treffer mottaker direkte, slik som i digitale dialoger, er det særlig viktig å kunne forklare årsakene bak ulike utslag. Det betyr at bakenforliggende modeller må kunne forstås og forklares. I veiledning må det tilsvarende være tydelig hvilke kilder som ligger til grunn. Overfor skattepliktige må avgjørelser som fattes ved bruk av KI kunne begrunnes, og det skal kunne redegjøres for hvilke egenskaper som er utslagsgivende. For interne prosesser er det viktig at etatens medarbeidere (saksbehandlere, kontrollører mv.) forstår at KI anvendes og på hvilken måte. Det gjelder også når kunnskapsgrunnlag produseres helt eller delvis ved bruk av KI, og det gjelder når det benyttes verktøy/løsninger med innebygget KI fra andre leverandører.

Rettferdighet: Skatteetatens bruk av KI skal baseres på rettferdighet som beskrevet i 2.3. I mange tilfeller oppstår rettferdighetsproblemer i KI-systemer som følge av skjevheter som allerede finnes i samfunnet implisitt eller eksplisitt. Disse skjevhetene kan f.eks. være diskriminering basert på hudfarge, kjønn eller fødeland, og kan være reflektert i historiske datasett som brukes til å utvikle KI-systemer. Der KI brukes i digitale dialoger eller på andre eksternt orienterte områder, vil systematisk skjevhet kunne være særlig kritisk i kraft av å eksponeres direkte ut til mottakere. Det fordrer særskilt overvåkning og oppfølging. For delautomatiserte prosesser innen seleksjon og valg av behandlingsløp, eller andre interne prosesser der det fattes beslutninger, kan slike iboende skjevheter også bidra til å forsterke eventuelle skjevheter i KI-systemet. Menneskelig tilsyn og kontroll med KI spiller derfor en viktig rolle for å fange opp skjevhet både i eksterne og interne prosesser. Ekstra kontroll er også tilrådet der generativ KI benyttes.

Teknisk robusthet, sikkerhet og beredskap: For alle del- eller helautomatiserte prosesser der KI anvendes er det nødvendig med regelmessig oppfølging av uventede eller uønskede utslag. Alle løsninger der KI anvendes bør være underlagt revisjonsprosesser som sikrer at integritet og konfidensialitet ivaretas. For det eksternt orienterte må det være tydelig beredskap på plass som sikrer kontinuitet dersom en KI-løsning kompromitteres eller gjøres utilgjengelig. Der etaten benytter standardløsninger med KI, er det viktig med tekniske og organisatoriske tiltak som sikrer at konfidensielle data ikke eksponeres for uvedkommende.

Datastyring og personvern: Kontinuerlig oppfølging av datakvalitet er viktig for alle oppgaveområder. For interne løsninger bør personvern bygges inn overfor medarbeidere, for eksterne løsninger skal det bygges inn overfor mottaker. Det skal videre være tydelig for mottaker hvorvidt en KI-støttet dialog brukes for å veilede en skattepliktig, eller om det fattes en automatisert avgjørelse med de rettigheter og plikter dette innebærer.

Hvordan henger dette sammen med risiko?

KI-teknologier og deres modenhet må vurderes opp mot hva som er akseptabel (rest)risiko på ulike bruksområder i etaten. Etaten har et eget risikostyringsrammeverk som skal legges til grunn, samtidig som det i slike vurderinger kan være aktuelt å se hen til risikoinngangen i EU sin *AI Act*. Denne kategoriserer risiko som:

- Uakseptabel risiko
- Høy risiko
- Begrenset risiko
- Minimal/ingen risiko.

Alle prinsippene i denne policyen er relevante uavhengig av oppgaveområdet KI anvendes på. Samtidig vil risikonivået kunne påvirke den konkrete anvendelsen av prinsippene. Graden av automatisering og graden av direkte eksponering er omtalt ovenfor, og er viktige elementer for vurderingen av konkret risiko. Samtidig kan også andre faktorer være relevante i risikovurdering, slik som at:

- bruken av KI er standardisert og utprøvd over tid
- bruken av KI kan påvirke skattepliktige/brukere fysisk eller økonomisk
- bruken av KI kan berøre sensitive forhold eller rettigheter hos skattepliktige/brukere
- bruken av KI medfører beslutninger eller støtter opp under slike
- bruken av KI berører et stort eller mindre omfang av skattepliktige/brukere.

3.2. Fra egenutviklet KI til bruk av standardverktøy

KI-systemer i Skatteetaten – i stor grad modeller – har i hovedsak vært utviklet av spesialister med kompetanse på analyse og modellutvikling som fag, og god kunnskap om de dataene som benyttes i trening og produksjonsbruk av en modell. Videre har dette ofte vært gjennomført som en del av prosjekter der det gjøres juridiske vurderinger i forhold til gjeldende regelverk.

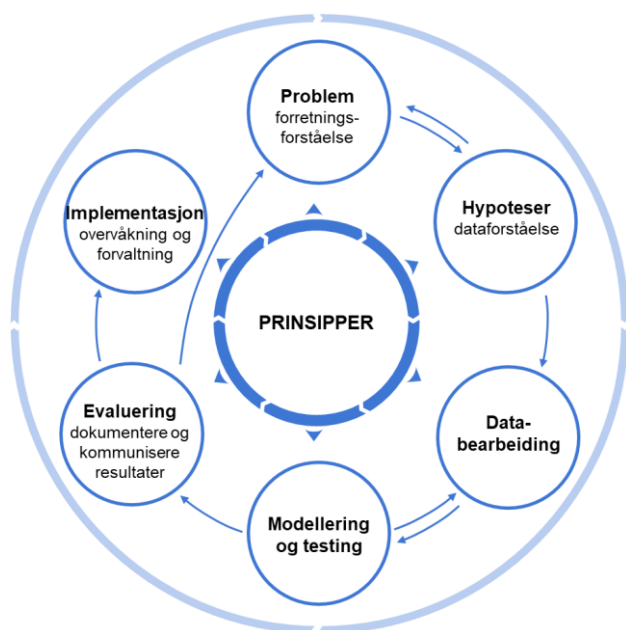
KI blir nå i større grad bygget inn i standardverktøy/produktivitetsverktøy; eksempler er Office 365, Microsofts Powerplattform, eller generative språkmodeller som Open AIs *ChatGPT*. I noen tilfeller vil det være veldig tydelig at det er et KI-system (for eksempel GPT) mens bruker får tilgang til en funksjon/knapp der det ikke er veldig tydelig at det ligger KI bak. Eksempler på anvendelsesområder som den enkelte ansatte *kan* velge å ta i bruk disse løsningene på, er støtte til produksjon av dokumenter og presentasjoner, oppsummering av dokumenter og sakskomplekser, programmering, generering av testdata og analyse av datakvalitet.

Hva som i praksis ligger bak disse standardløsningene, hvordan de er trent og også hvordan våre data blir benyttet, er ofte ikke tilgjengelig for brukerne av dem. Samtidig kan den økte tilgjengeligheten føre til en økt risiko for at prinsippene som er trukket opp her ikke følges. Det vil kunne være opp til den enkelte bruker, i det enkelte brukstilfelle, å vurdere om bruken er forsvarlig, i tråd med prinsippene og om f.eks. krav til taushetsplikt overholdes eller om noen data kan «lekke» til en tredjepart.

I hele spennet fra «kunnskap og innsikt» til «Digitale dialoger» må etaten derfor vurdere om det enkelte produkt, «knapp» eller funksjon *skal* tilbys til brukere, eventuelt hvilken veiledning og opplæring det er behov for overfor de som skal bruke den.

3.3. Utvikling og utforskning av KI-bruk

Prinsippene i dette dokumentet er ikke bare relevante i ulike forvaltningssituasjoner som omtalt over, men allerede når bruk av KI er i utforskning eller under utvikling. Data science og bruk av KI innebærer ofte en eksplorativ utviklingstilnærming, med iterasjoner på flere nivåer før vi er over i reell leveransemodus. Prosessen minner ofte mer om forskning og utvikling, enn klassisk smidig systemutvikling som gjerne har vekt på dekomponering og kontinuerlige leveranser. Prosessen fremgår i Figur 2 nedenfor, og er en tilpasning av de facto standard for data science-arbeid.



Figur 2: Prosess for data science og egenutvikling av maskinlæringsmodeller/KI-løsninger. Tilpasning av CRISP-DM (Cross-Industry Standard Process for Data Mining).

Den ytterste sirkelen indikerer behov for å gå hele løpet flere ganger på ulike nivå, og er et vesentlig poeng. Typisk vil det innledningsvis være behov for å utforske forretningsverdi gjennom pilotprosjekter og mer eksperimentelle løp som etter evaluering går en ny runde med ny problem- og forretningsforståelse. Rammen rundt bør være tydelig og avgrenset/lukket.

Senere iterasjoner av prosessen innebærer utviklingsløp mot tydeligere produksjonsorienterte leveranser. Innenfor utviklingsløpet er det likevel normalt å bevege seg frem og tilbake mellom faser, eksempelvis mellom forretningsforståelse og dataforståelse, samt mellom databearbeiding, modellering og testing (jf. toveispilene i Figur 2). Til tross for at prosessen ikke er rett frem, er det viktig å vektlegge alle prinsippene i samtlige faser ved et slikt produksjonsorientert løp. Det inkluderer utviklingsløpet, men også drift og forvaltning («ML-ops»).

Prinsippene i denne policyen skal operasjonaliseres for all utvikling og bruk av KI uavhengig av hvilken metodikk som legges til grunn. Det vil kunne innebære såkalte ikke-funksjonelle krav (IFK), men kan også påvirke det funksjonelle. Kravene må baseres på føringene fra denne policyen.

Team som jobber med utvikling og forvaltning av KI-løsninger skal etablere roller (faglig, juridisk, teknisk mv.) med ansvar for å sikre etterlevelsen på de respektive områdene.

4. Forvaltning og revisjon

Policy for utvikling og bruk av KI i Skatteetaten besluttet av ELM og forvaltes av SKD Juridisk avdeling. Mindre endringer og presiseringer av innholdet i denne policyen, kan besluttet av SKD Juridisk avdeling. Det må utarbeides konkrete rutiner og retningslinjer for å understøtte etterlevelse av prinsippene i denne policyen.