Copilot med personvernbriller på

Sluttrapport fra sandkasseprosjektet med NTNU

Temaer: Generativ KI, orden i eget hus og vurdering av personvernkonsekvenser (DPIA)

November 2024

Datatilsynet

Innhold

November 2024

Denne pdf-en tilsvarer den første versjonen av rapporten, slik den ble publisert på Datatilsynets sider november 2024. Teknologien og jussen er stadig i utvikling, så det kan være behov for å justere eller presisere rapportene med tiden. Dersom denne pdf-en skiller seg fra det som står på Datatilsynets nettsider, kan du ta utgangspunkt i at det er nettsidens tekst som er gjeldende råd.

Sammendrag

Generativ kunstig intelligens (KI) er ikke lenger bare et artig verktøy på si, men er nå i ferd med å bli integrert i de digitale løsningene vi allerede bruker. Microsoft lanserte sin Copilot for Office-pakken i november 2023, med potensial for å forenkle arbeidshverdagen betraktelig. Noen har tatt det i bruk, helt eller delvis. Mange sitter på gjerdet. For hva skjer egentlig når du slår på Microsoft 365 Copilot?

Datatilsynet og NTNU har sett på hvilke personvernkrav som gjelder og hvilke vurderinger NTNU bør gjøre før Microsofts KI-assistent tas i bruk. Parallelt gjennomførte NTNU et pilotprosjekt for å undersøke om de er klar for å innføre M365 Copilot, samt foreslå et rammeverk for forvaltning, drift, vedlikehold og utvikling. NTNU har offentliggjort deres egen funnrapport. Den gir overordnet kunnskap om hvordan M365 Copilot det fungerer, og gir god innsikt for andre som vurderer å slå på copiloten.

Les NTNUs egen funnrapport

Datatilsynet stiller seg bak NTNUs rapport, men vi anbefaler å gå mer spesifikt til verks i vurderingen av personvernkonsekvenser (DPIA). Hver virksomhet må gjøre sine egne personvernkonsekvensvurderinger, basert på hvilke data de har og hvilke oppgaver de vil bruke M365 Copilot til.

M365 Copilot er en aktiv komponent som gjenfinner og gjenskaper informasjon på måter man tidligere ikke har vært vant til. Det er en utfordring at denne nye teknologiens evne til å formulere godt språk - også på norsk - kan få den til å fremstå menneskelig, som om den kan vurdere og foreta logiske resonnementer.

Det er også viktig å understreke at dette er nybrottsarbeid. Så vidt vi vet har ingen andre tilsynsmyndigheter sett på bruk av M365 Copilot opp mot personvernregelverket. Denne rapporten bør ses på som et første steg i å forstå og vurdere hvorvidt slike verktøy kan tas i bruk på en (forsiktig og stegvis) måte som er i samsvar med personvernregelverket.

Hovedpunkter

1. **M365 Copilot forutsetter at virksomhetsdata allerede ligger i Microsofts skyløsning.** M365 Copilot sitter på toppen av Microsofts M365 skyløsning. Før innføring av M365 Copilot, er det en forutsetning at dere har gjort alle nødvendige sikkerhets- og personvernvurderinger knyttet til selve M365-plattformen. Dere må også ha nødvendige ressurser og kompetanse for å forvalte tjenesteleverandører og skyløsningen på en ansvarlig måte over tid, særlig på grunn av hyppige endringer fra leverandørsiden. Ansvaret for dataene som brukes i copiloten ligger hos virksomhetene som tar verktøyet i bruk.

   DFØ har laget en veileder for offentlige virksomheter om anskaffelse av skytjenester, som kan være til hjelp. Se også punkt 4 i NTNUs funnrapport for nærmere informasjon og anbefalinger fra NTNU om forvaltning av systemet.

2. **Sørg for orden i eget hus.** Copiloten vil ha tilgang til alle de samme opplysningene som brukeren av verktøyet har. Det betyr at utfordringer og svakheter i «den digitale grunnmuren», som dårlig tilgangsstyring og kontroll over personopplysninger, vil bli synlige og kraftig forsterket av M365 Copilot. Det er viktig å understreke at Microsoft som tjenesteleverandør også forutsetter i svært høy grad at «alt er i orden» med forvaltning av den underliggende M365-plattformen, for at copiloten skal kunne brukes på en ansvarlig og lovlig måte. God orden i eget hus må altså være på plass først, og ethvert innføringsprosjekt vil trolig kreve grundige (re)vurderinger av egen informasjonsforvaltning. Dette krever innsats og ressurser, men er et kritisk og nødvendig steg ved innføring av ny teknologi.

   Digitaliseringsdirektoraret har utarbeidet en veileder om informasjonsforvaltning, og på Datatilsynets nettside finner du informasjon om hva en protokoll over behandlingsaktiviteter må inneholde.

   NTNU har selv kommet frem til at de ikke er klar til å innføre M365 Copilot i hele virksomheten ennå, blant annet fordi de vurderer at de ikke har god nok orden i eget hus.

3. **Identifiser og begrens hva copiloten skal brukes til.** Vurder hvilke oppgaver og tilhørende behandling av personopplysninger M365 Copilot skal, og ikke skal, brukes til. Noen oppgaver er dårlig egnet for bruk av generativ KI, for eksempel når det er viktig at svar er riktige og brukeren ikke har kompetanse eller tid til å kontrollere det som genereres. Videre vil bruk av M365 Copilot innenfor f.eks. HR og personalledelse innebære

ekstra høy risiko for personvernet. Dette fordi håndtering av tilgang til personopplysninger er krevende å forvalte og kontrollere, eller fordi konsekvensene for enkeltpersoner kan bli svært alvorlige. Oppgaver som involverer særlige kategorier (sensitive) personopplysninger bør også vurderes nøye eller unngås i forbindelse med bruk av M365 Copilot.

Kartlegg og beskriv behandlingsoperasjonene som skjer hvis M365 Copilot brukes til et bestemt formål, altså fra en instruks gis til copiloten til svaret kommer ut. Behandlingsprotokollen er et hensiktsmessig sted å starte, der man går gjennom og vurderer hver behandling for hvert formål. Det gir et godt utgangspunkt for å vurdere hvilke oppgaver dere vil og kan bruke copiloten til.

Dersom M365 Copilot har tilgang til informasjon med (sensitive) personopplysninger, må informasjonen være klassifisert, identifisert og merket, minimum på dokumentnivå. Vi understreker at Microsoft erkjenner at dette er nødvendig for at M365 Copilot kan benyttes på en ansvarlig måte.

4. **Vurder det rettslige grunnlaget.** Når oppgaver og tilhørende behandling av personopplysninger vurderes som «M365 Copilot-kandidater», må behandlingsgrunnlaget sjekkes. For behandlinger dere allerede utfører, må dere vurdere om bruken av M365 Copilot fører til endringer i behandlingen, for eksempel hvilke eller hvem sine personopplysninger som behandles. Hvis det er endringer, må dere vurdere om det eksisterende behandlingsgrunnlaget fortsatt kan brukes, herunder om behandlingen fortsatt er «nødvendig». Hvis svaret er nei, kan M365 Copilot ikke brukes til denne behandlingen.

   Behandlinger for nye formål krever at dere indentifiserer et passende behandlingsgrunnlag. Der det er snakk om å gjenbruke personopplysninger til nye formål, som ofte vil være tilfelle, må dere vurdere om den nye behandlingen er forenlig med det opprinnelige formålet.

5. **Vurder personvernkonsekvenser.** Det vil som hovedregel være et krav om å vurdere personvernkonsekvenser (DPIA) ved bruk av generativ KI som behandler personopplysninger. Dette fordi loven fremhever «bruk av ny teknologi» som en særlig viktig faktor, og forståelsen av risiko knyttet til generativ KI er ennå umoden. En DPIA må gjøres per behandling eller sett med lignende behandlinger. Oppgaver som i seg selv ikke krever behandling av personopplysninger, vil med M365 Copilot likevel kunne komme til å gjøre det, fordi copiloten bruker av all informasjon brukeren har tilgang til og kan dermed koble det sammen med personopplysninger.

   DPIA-prosessen skal identifisere tekniske og organisatoriske tiltak som kan redusere risikoen til et akseptabelt nivå, og disse må være på plass før eventuell bruk av M365 Copilot. Testing kan være et tiltak for å minimere risiko. Hvis risikoen er for høy også etter at tiltak er prøvd, bør dere trolig ikke bruke M365 Copilot til den aktuelle behandlingen likevel. Alternativt, kontakt Datatilsynet for en forhåndsdrøftelse.

6. **Vil bruken være i strid med e-postforskriften?** M365 Copilot logger alle samhandlinger. Historikken lagres på brukerens personlige område, og er i NTNUs tilfelle tilgjengelig for M365-administratorer. Samlet sett anser vi det som sannsynlig at samhandlingsloggen vil kunne rammes av forbudet mot overvåking av arbeidstakers bruk av elektronisk utstyr. Vi forstår det imidlertid slik at hovedformålet med samhandlingsloggen er å sikre at kvaliteten på tjenesten er slik den skal være. Dette formålet kan falle inn under unntaket for å administrere virksomhetens datanettverk. Hvorvidt det andre unntaket, å «avdekke eller oppklare sikkerhetsbrudd i nettverket», kan være aktuelt, må vurderes konkret opp mot hva som er formålet med samhandlingsloggen.

7. **Bruk av språkmodeller krever kompetanse og bevissthet.** Språkmodeller gir en ny brukeropplevelse for svært mange, med både muligheter og begrensninger som er uavklarte. Det kan være krevende å forstå hvilken informasjon som inngår i underlaget for formuleringer og hva som ikke er med. Det krever kompetanse for å formulere instrukser eller ledetekst (på engelsk, prompts) som gir relevante og gode svar. Det er virksomhetens ansvar at brukere av løsningen har tilstrekkelig kunnskap, bevissthet og opplæring i bruken av M365 Copilot. Denne kompetansen sikrer ikke bare god kvalitet i det som genereres, men også at løsningen brukes på en måte som ivaretar personvernet.

8. **Vurder alternative løsninger**. M365 Copilot kan brukes til svært mye. Det er derfor et omfattende arbeid å sikre at systemet brukes på en ansvarlig og lovlig måte. Noen av copilotens egenskaper kan utfordre formålsbegrensningsprinsippet og dataminimeringsprinsippet. Tiltak som i teorien kan redusere risiko og konsekvenser, kan i praksis være svært krevende å innføre. Derfor er det viktig å vurdere om andre KI-løsninger med lavere personvernrisiko kan oppfylle de konkrete behovene. Dette kan være løsninger som transkriberer lydopptak eller spesialtilpassede dialogroboter og støtteverktøy tilpasset spesielle formål og avgrenset til nøye utvalgte og kvalitetssikrede interne informasjonskilder.

9. **Gjør innføringen i små og kontrollerte steg.** Det er mulig for norske virksomheter å ta i bruk M365 Copilot, men ikke for alle og ikke for alt. Vår tydelige anbefaling er at slike løsninger blir innført kontrollert, i små steg, med utvalgte roller og for egnede behandlinger i virksomheten. Det må også lages strukturerte opplegg for etterkontroll og oppfølging av kvalitet på det løsningen produserer, både gjennom organisatoriske og tekniske tiltak.

Veien videre

NTNU har gjort en imponerende, samfunnsnyttig og omfattende jobb med å skaffe seg kunnskap og bevissthet rundt bruken av språkmodeller generelt og integrerte KI-løsninger som M365 Copilot spesielt. De har valgt å ikke innføre M365 Copilot for hele virksomheten, men heller ta i bruk verktøyet kontrollert i små steg med utvalgte roller først.

M365 Copilot er fortsatt tidlig i utviklingsløpet og mangler kontroll på et granulært nivå, som muligheten til å gjøre lokale og fleksible tilpasninger (f.eks. å slå av tilgang til brukeres e-postkasser, spesifikke slettepolicyer). Ubegrenset tilgang til brukers e-postkasse regnes antagelig av Microsoft som en viktig og sentral funksjon, samtidig som dette kanskje er et av elementene som skaper mest usikkerhet for mange virksomheter.

Datatilsynet forventer at utfordringer som kunder, virksomheter, myndigheter og samfunnet generelt identifiserer i produktet, tas på alvor av produktleverandør. Samtidig setter også løsningen klare krav til virksomheter som ønsker å hente gevinster ved bruk av verktøyet. Forutsetningen om en svært velfungerende informasjonsforvaltning og orden i eget hus kan gjøre veien for å lykkes med slike løsninger krevende, men åpenbart med en positiv oppside langt utover det å ta i bruk en bestemt løsning.

NB!

En vurdering av skytjenester generelt, overføringer av personopplysninger til tredjeland og Microsofts rolle(r) etter personvernforordningen, har vært utenfor omfanget av prosjektet. Vi vil likevel nevne at datatilsynet for EU-organene, EDPS, nylig fattet et vedtak som blant annet handler om Microsofts rolle i forbindelse med levering av en skytjeneste til flere EU-organer. Dette vedtaket er anket av både Microsoft og EU-kommisjonen. Utfallet av saken kan få betydning for hvordan bruk av skyløsninger i fremtiden må innrettes for å være i tråd med personvernforordningen.

Om prosjektet

NTNU er et internasjonalt universitet med hovedsete i Trondheim og campuser i Gjøvik og Ålesund. Universitetet har en teknisk-naturvitenskaplig hovedprofil, en rekke profesjonsutdanninger og en faglig bredde som også inkluderer humaniora, samfunnsvitenskap, økonomi, medisin, helsevitenskap, utdanningsvitenskap, arkitektur, entreprenørskap og kunstfag. NTNU har 9000 ansatte og 43 000 studenter.

Hovedformålet med prosjektet var å undersøke om og hvordan en stor offentlig organisasjon som NTNU kan ta i bruk M365 Copilot. Det er viktig å merke seg at Microsoft bruker begrepet «Copilot» på ulike måter, og at flere tjenester opererer under dette navnet. Dette prosjektet handler spesifikt om M365 Copilot, der KI blir integrert i eksisterende Microsoft 365-tjenester.

NTNU tok tak i flere problemstillinger knyttet til bruk av KI-verktøy i offentlig sektor. En sentral utfordring var om M365 Copilot kan brukes uten at personopplysninger behandles i konflikt med personvernforordningen. Et annet spørsmål var om folk vil akseptere at opplysningene deres kan brukes i andre sammenhenger enn de opprinnelig ble samlet inn for. I tillegg er det flere etiske og organisatoriske utfordringer knyttet til bruken av generative KI-verktøy generelt. NTNU skulle også undersøke risikoer knyttet til feilaktige beslutninger, som følge av for eksempel diskriminering og såkalt «hallusinasjon».

NTNU ønsket i tillegg å utvikle en verktøykasse med retningslinjer, rammeverk og vurderinger av personvernkonsekvenser, som kan brukes av andre offentlige og private virksomheter. Målet var å gjøre det enklere å vurdere om og eventuelt hvordan generative KI-verktøy som M365 Copilot kan implementeres i offentlig sektor på en ansvarlig måte. NTNU skulle også se på hvordan leverandører kan påvirkes til å tenke på innebygd personvern tidlig i utviklingsprosessen, og hindre at personvernspørsmål først diskuteres mot slutten av en anskaffelsesprosess.

Mål for sandkasseprosjektet

Omfanget av NTNUs prosjekt var bredt og dekket flere temaer enn bare personvern. Det var derfor viktig for Datatilsynet å snevre inn omfanget av det vi skulle se på og bistå med. Hovedmålet har vært å utforske og klargjøre **hva personvernregelverket krever** for at NTNU og andre offentlige organisasjoner kan bruke verktøy slik som M365 Copilot på en ansvarlig og lovlig måte.

For å gjøre dette, har det vært nødvendig å se på:

1. Hva M365 Copilot faktisk er og hvordan det fungerer, i tillegg til generative språkmodeller generelt.
2. Hvordan M365 Copilot kan forstås i lys av personvernregelverket på et overordnet nivå.
3. Hvilke forutsetninger som må være på plass, herunder «orden i eget hus».
4. Hvorvidt en eller flere vurderinger av personvernkonsekvenser kreves, og hva som er særlig relevant å vurdere i lys av M365 Copilot.
5. Anvendelse av den norske e-postforskriften.

Behandling av særlige kategorier personopplysninger, skytjenester generelt, overføringer av personopplysninger til tredjeland og Microsofts rolle etter personvernforordningen har vært utenfor prosjektets omfang.

Forholdet til NTNUs egen funnrapport

NTNU publiserte sin funnrapport 17. juni 2024, for å dele erfaringene med M365 Copilot med andre organisasjoner. Rapporten presenterer åtte hovedfunn som omhandler ikke bare personvern, men også etiske, juridiske, tekniske og organisatoriske problemstillinger.

Funnrapporten kan støtte og inspirere både offentlige og private virksomheter i deres planlegging og vurdering av generative KI-verktøy, samt bidra til utvikling av risikoreduserende tiltak. Vi fremhever særlig:

- NTNUs verktøykasse, som gir informasjon om hva generativ KI er og hvordan man kan bruke det på en smart, sikker og trygg måte.
- NTNUs KI-reise, med forslag til en KI-strategi, vurderinger per tjeneste og tips om anskaffelser (NTNUs funnrapport, sider 29–36).

- NTNUs forslag til retningslinjer for generativ KI.

Vi anbefaler at NTNUs funnrapport leses i tillegg til denne sluttrapporten, som er ment å utfylle NTNU-rapporten på utvalgte områder.

NTNUs verktøykasse inneholder en vurdering av personvernkonsekvenser (DPIA). NTNU valgte å gjøre en «overordnet» vurdering for M365 Copilot i driftsfasen, hvor de så på teknologien som helhet. NTNU har ikke vurdert konkrete behandlinger av personopplysninger i lys av spesifikke formål. Det betyr at arbeidet ikke oppfyller innholdskravene til en DPIA, jf. personvernforordningen artikkel 35. Arbeidet bidrar likevel med informasjon om NTNUs erfaring med M365 Copilot og generelt hvordan verktøyet fungerer, som kan være nyttig dersom en skal utarbeide en DPIA.

Avgrensning

M365 Copilot er et verktøy med mangfoldige funksjoner som kan brukes til en rekke forskjellige oppgaver. Gitt det brede spekteret av funksjoner, kan det også brukes til å behandle personopplysninger som en del av sine operasjoner. Det er ikke mulig å fastslå på generelt grunnlag at verktøyet kan brukes i tråd med personvernregelverket. Selv om enkelte oppgaver som M365 Copilot utfører i utgangspunktet ikke krever behandling av personopplysninger, vil slik behandling nærmest alltid forekomme på grunn av verktøyets iboende egenskaper.

I denne sluttrapporten har vi valgt å fokusere på noen grunnleggende temaer. Først forklarer vi hva M365 Copilot er og hvordan det fungerer. Deretter gir vi en generell beskrivelse av hvordan vi forstår M365 Copilot i lys av personvernregelverket, sammen med en gjennomgang av viktige begreper og forhold man bør være oppmerksom på. Vi ser også på grunnleggende forutsetninger for bruk av M365 Copilot, inkludert behovet for «orden i eget hus». Disse vurderingene er relevante også for andre KI-verktøy. Til slutt belyser vi viktigheten av personvernkonsekvensvurderinger (DPIA) og hva man bør ta hensyn til i forbindelse med M365 Copilot.

NTNU testet M365 Copilot med tre brukstilfeller: «utredningsstart», «referatfunksjon» og «saksbehandling på e-post».[1] Disse brukstilfellene ble valgt fordi de kan være relevante også for andre offentlige organisasjoner. I denne sluttrapporten har vi tatt utgangspunkt i NTNUs brukstilfeller, men vi har gjort brukstilfelle C litt mer spesifikt for å kunne ha et klart og tydelig formål. Dette eksempelet gjenspeiler ikke nødvendigvis hvordan NTNU faktisk jobber, men brukes for illustrasjonens skyld. Det er viktig å merke seg at hvert brukstilfelle kan innebære flere typer behandling av personopplysninger.

Brukstilfelle A: En utreder bruker M365 Copilot for å samle inn informasjon (datainnsamling) før utredningen kan starte. Utreder får tilgang til informasjon fra nett, tidligere dokumenter hen har skrevet selv eller dokumenter hen har tilgang til (men skrevet av andre). Utreder kan ved hjelp av M365 Copilot få ut en oversikt over relevant datamateriale for å gjøre nødvendige avveininger i tråd med instruksjonene for utredningen, få hjelp til selve skrivearbeidet (kladding) og renskriving/språkforbedring.[2]

Brukstilfelle B: En ansatt er ansvarlig for at man i et internt møte mellom to eller flere parter skal bli enige om noe. De kaller inn til et digitalt Teams-møte, eller et fysisk møte der Teams lytter aktivt til møtet. Møtet tas opp og transkriberes. M365 Copilot bruker transkripsjonen, informasjon i kalenderinvitasjonen og «nærliggende dokumenter» for å lage en oppsummering av møtet.[3]

Brukstilfelle C: En ansatt skal vurdere hvorvidt en søknad om opptak til masterprogram sendt inn per e-post er komplett (dvs. inneholder all informasjon som trengs), og svare på søknaden enten med en bekreftelse om at søknaden er komplett eller en anmodning om mer informasjon.

[1] NTNUs funnrapport, s. 43-51.

[2] Ibid., s. 43.

[3] Ibid., s. 46.

Hva M365 Copilot er og hvordan det fungerer

Microsoft 365 Copilot skiller seg fra samtalerobotene man kanskje kjenner fra før ved at språkmodellteknologien er integrert i de Microsoft 365-applikasjonene mange allerede bruker hver dag, som Word, Excel, PowerPoint, Outlook og Teams. Den representer dermed en viktig trend som omtales som «integrert KI», dvs. KI-løsninger og funksjonalitet som dukker opp i produkter man allerede bruker. Noen ganger er dette en funksjonalitet man kan velge å ta i bruk (M365 Copilot krever egen tilleggslisens). Andre ganger får man teknologien «på kjøpet» enten man ønsker det eller ei. Et eksempel på sistnevnte er tidligere «Bing chat» som nå også kalles «Copilot», men som er en egen samtalerobotløsning i likhet med f.eks. ChatGPT. Denne typen løsninger kan svare på generelle spørsmål og har ikke uten videre tilgang til brukerens og virksomhetens egen informasjon.

M365 Copilot fungerer derimot som en personlig assistent og kombinerer funksjonaliteten i de store språkmodellene med virksomhetens interne data og informasjonsstrukturer, gjort tilgjengelig via kunnskapsgrafer og indekserte semantiske søk (RAG – se vedlegg 1). Verktøyet kan også settes opp til å hente ekstern informasjon (fra internett) for å «berike» interne søk. Det er verdt å bemerke at en utfordring knyttet til dette, kan være at spørsmål som opprinnelig oppfattes som interne eller til og med konfidensielle, plutselig blir eksponert eksternt som søk i åpne søkemotorer. M365 Copilot er et produkt som krever at virksomheten har svært god kontroll med løsningens muligheter og begrensninger i praksis.

Ordet «Copilot» brukes for øvrig av Microsoft i mange sammenhenger og refererer ikke til et spesifikt produkt. Det er snarere et støtteverktøy implementert på forskjellige måter i forskjellige deler av Microsoft 365-plattformen. Løftet fra Microsoft er at M365 Copilot kan hjelpe deg med alt fra å generere innhold og analysere data til å forbedre kommunikasjon og samarbeid i organisasjonen – mot at man gir den tilgang til all informasjon en selv har tilgang til.

Kunnskapsgrafer (Microsoft Graph) og semantiske søk utgjør en dataplattform som kobler sammen data og tjenester på tvers av Microsoft 365. Den samler informasjon fra e-poster, kalendere, dokumenter, møter, chat-samtaler og mer. Ved å utnytte Microsoft Graph kan Copilot få tilgang til og forstå konteksten av informasjonen en bruker har tilgang til. I tillegg bygger løsningen over tid en erfaringsbasert profil av brukeren, noe som skal gi mer relevant og personlig assistanse.

Fordi M365 Copilot er tett integrert i Microsoft 365-applikasjonene, er løsningen tilpasset grensesnittet og arbeidsoppgavene i de forskjellige verktøyene, og den kan dermed assistere direkte i applikasjonen uten å måtte flytte informasjon mellom ulike applikasjoner eller grensesnitt.

Ved å automatisk hente inn informasjon fra møter, e-poster, dokumenter og samtaler (innenfor gitte tilgangsrettigheter) i kombinasjon med brukerprofilering over tid, har dermed løsningen potensialet for å gi mer skreddersydd bistand. Denne ganske omfattende tilgangen til enkeltbrukers og virksomhetens informasjon skal gjøre det mulig å automatisere komplekse oppgaver:

- **Outlook**: Oppsummere lange e-posttråder. Generere forslag til e-postsvar basert på kontekst, tone og tidligere kommunikasjon. Identifisere og foreslå kalenderavtaler eller oppgaver basert på innholdet i e-postene dine.
- **Word**: Utarbeide utkast til dokumenter basert på instruksjoner eller punkter. Forbedring av tekst, formuleringer, grammatikk og stil, og tilpasning av tone til målgruppe. Oppsummering av dokumenter og rapporter.
- **Excel**: Tolke datasett, identifisere trender og mønstre, og presentere oppsett på enkle måter. Lage komplekse formler ved å beskrive hva du ønsker å oppnå i naturlig språk. Foreslå passende diagrammer og grafer for å representere data visuelt.
- **PowerPoint**: Generere hele presentasjoner basert på et dokument eller en ide, inkludert forslag til tekst, bilder og design. Gi anbefalinger for layout, farger og grafiske elementer. Konvertere et dokument eller en rapport til en presentasjon ved å trekke ut de viktigste punktene.
- **Teams**: Oppsummere diskusjoner i sanntid, skrive referat og identifisere handlingspunkter. Generere en liste over neste steg og tildele oppgaver til teammedlemmer.
- **SharePoint**: Hjelpe med å lage og redigere innhold, inkludert tekstforslag og strukturering. Semantiske søk for å finne relevant innhold basert på betydning, ikke bare nøkkelord. Analysere innhold på tvers av SharePoint for å identifisere kunnskapsgap eller overlappende informasjon.

Det er med andre ord ikke lite som loves, og det hele muliggjøres av to ting: kraften i språkmodeller og mer eller mindre fri tilgang til all informasjon den enkelte bruker i virksomheten allerede har tilgang til.

Hvordan kan M365 Copilot forstås i lys av personvernregelverket?

Vi begynner med å trekke frem noen nøkkelkonsepter og begreper fra personvernlovgivningen som vi mener det er viktig å huske på når man vurderer eventuell innføring og bruk av M365 Copilot i virksomheten. Vi håper dette bidrar til at misforståelser som kan lede til følgefeil unngås fra starten.

Nøkkelkonsepter- og begreper

Personopplysning	Dette er definert i personvernforordningen artikkel 4 nr. 1 som «enhver opplysning om en identifisert eller identifiserbar fysisk person». Selv om en opplysning om noen er feil, f.eks. når en språkmodell har generert noe feil om en person, er det en (uriktig) personopplysning. Det samme gjelder forutsigelser og antakelser om en person.
Den registrerte	En identifisert eller identifiserbar fysisk person, jf. personvernforordningen artikkel 4 nr. 1. Med andre ord er det enkeltpersonen som opplysningene kan knyttes til.
Sletting og riktighet	Det må treffes ethvert rimelig tiltak for å sikre at personopplysninger som er uriktige med hensyn til formålene de behandles for, uten opphold slettes eller rettes i henhold til riktighetsprinsippet, jf. personvernforordningen artikkel 5 nr. 1 bokstav d. Det betyr at brukere må være tilstrekkelig opplært, og NTNU må ha rutiner for å minske risikoen for at feil personopplysninger skapes av M365 Copilot. Hvis det likevel skjer, må de slettes eller rettes straks.
Behandling	I personvernforordningen artikkel 4 nr. 2 blir behandling definert som «enhver operasjon eller rekke av operasjoner som gjøres med personopplysninger, enten automatisert eller ikke, f.eks. (...)». Lovgiveren har bevisst gitt begrepet «behandling» en vid rekkevidde. Dette fremgår både av utrykket «enhver operasjon» og av den ikke-uttømmende karakteren i definisjonen, tydeliggjort ved bruken av «f.eks.».[4] Hver behandling må ha et rettslig grunnlag, jf. personvernforordningen artikkel 6. For å fastsette riktig grunnlag må man først klargjøre formålet med behandlingen og hvilke personopplysninger som skal behandles. Det er også nødvendig å kartlegge hvilke konkrete behandlingsoperasjoner som vil finne sted, før det rettslige grunnlaget kan vurderes og velges. Dersom særlige kategorier personopplysninger skal behandles, må det i tillegg identifiseres et gyldig unntak fra forbudet i artikkel 9. «M365 Copilot i driftsfase» eller «innføring av M365 Copilot» er ikke en konkret behandling.[5]
Formålet med behandlingen og formålsbegrensnings-prinsippet	«Formål» er selve hjørnesteinen i personvernforordningen. Formålet er **årsaken** til at en behandling skjer, og det er formålet som setter grensene for hvilke personopplysninger som skal behandles og hvordan. Personvernforordningen sier at personopplysninger må samles inn for spesifikke, uttrykkelig angitte og berettigede formål og ikke viderebehandles på en måte som er uforenlig med disse formålene, jf. artikkel 5 nr. 1 bokstav b. Særlig viktig i kontekst av bruken av M365 Copilot er vilkårene «spesifikke» og «uttrykkelig angitte». Formålet må fastsettes senest på tidspunktet for innsamlingen av personopplysningene,[6]

[4] Dom av 24. februar 2022 [C5], *Valsts ieņēmumu dienests*, C-175/20, EU:C:2022:124, avsnitt 35.

[5] NTNUs funnrapport, s. 102.

[6] Dom av 24. februar 2022 [C5], *Valsts ieņēmumu dienests*, C-175/20, EU:C:2022:124, avsnitt 64.

	med mindre et nytt formål er forenlig med det opprinnelige formålet i henhold til personvernforordningen artikkel 6 nr. 4. Det er nødvendig å se til formålet for å kunne overholde blant annet dataminimeringsprinsippet, hvor personopplysninger må være adekvate, relevante og begrenset til det som er strengt nødvendig **for å oppnå formålet**.
Dataminimerings-prinsippet	Dataminimeringsprinsippet går ut på at personopplysninger må være adekvate, relevante og begrenset til det som er **nødvendig for formålene** de behandles for. Det er derfor nødvendig å se hen til formålet (som skal ha vært utformet i tråd med formålsbegrensningsprinsippet) når en vurderer hvilke og hvem sine personopplysninger som er adekvate, relevante og nødvendige **for å oppnå formålet**.
Mottaker	Begrepet er definert i personvernforordningen artikkel 4 nr. 9 som «en fysisk eller juridisk person, offentlig myndighet, institusjon eller ethvert annet organ som personopplysninger utleveres til, enten det dreier seg om en tredjepart eller ikke». M365 Copilot kan lett personifiseres eller legemliggjøres, fordi det selv utgir seg for å være en fysisk person gjennom måten det svarer på. Dette kan føre til en uriktig tenkemåte når verktøyet vurderes fra et personvernperspektiv og kan føre til følgefeil i senere vurderinger. For eksempel er ikke M365 Copilot en «mottaker» i personvernforordningens forstand.
Behandlingsprotokoll	Hver behandlingsansvarlig har plikt til å føre protokoll over sine behandlingsaktiviteter, hvor blant annet formålet med behandlingen oppgis, samt hvilke og hvem sine personopplysninger som behandles for å oppnå formålet, jf. personvernforordningen artikkel 30. Informasjonen i behandlingsprotokollen er langt på vei sammenfallende med det man må informere de registrerte om. Enkelte nye behandlinger vil uunngåelig oppstå ved selve innføringen av M365 Copilot, for eksempel loggføring av interaksjoner (samhandlingsloggen). Nye behandlinger må oppføres i behandlingsprotokollen, og de registrerte må bli informerert i henhold til personvernforordningen artikkel 13.

Kartlegg og beskriv behandlingen

Det er gjennomgående slik at bestemmelsene i personvernforordningen knytter seg til en «behandling» av personopplysninger, som definert i personvernforordningen artikkel 4 nr. 2 (se over). M365 Copilot er ikke i seg selv en behandling, men et verktøy eller sett med funksjoner – altså et «middel» – som kan benyttes for å behandle personopplysninger på mange forskjellige måter og for ulike formål. Det på forhånd definerte formålet, og hva som er nødvendig for å oppnå dette formålet, setter imidlertid grenser for hvilke personopplysninger som kan behandles og på hvilken måte. Det første man bør gjøre er derfor å kartlegge og beskrive behandlingen som vil skje hvis M365 Copilot tas i bruk i forbindelse med et bestemt formål, altså fra når en instruks legges inn i M365 Copilot til svaret kommer ut. Ofte vil man ønske å bruke M365 Copilot i forbindelse med behandlinger som allerede gjøres, og ved å kartlegge hva som er nytt ved bruk av M365 Copilot vil man være i stand til å sammenligne den «gamle» behandlingen og den «nye» behandlingen, og da identifisere de nye behandlingsoperasjonene («rekke[n] av operasjoner») som kan eller vil oppstå.

En systematisk beskrivelse av den «nye» behandlingen gir flere fordeler.

- **Valg av rettslig grunnlag:** Det blir mulig å fastslå hvilket rettslig grunnlag som passer best for behandlingen.
- **Vurdering av nødvendighet og proporsjonalitet:** En sammenligning mellom gammel og ny behandling bidrar til å vurdere behovet for og forholdsmessigheten av den nye behandlingen.
- **Risikobegrensing:** Det blir mulig å identifisere hvilke tekniske eller organisatoriske tiltak som bør innføres for å begrense risiko, f.eks. ved å ha spesifikke retningslinjer eller prosedyrer for effektiv instruksjonsutforming, ved å endre hva slags tilgang en bestemt brukerrolle skal ha eller ved å slå av eller på tilgjengelige innstillinger i M365 Copilot.

NTNU så på tre utvalgte brukstilfeller i sandkasseprosjektet (NTNUs funnrapport, s. 43-51). Imidlertid valgte de bevisst ikke å se på disse i DPIA-en sin. De ser i stedet på teknologiproduktet på et overordnet nivå. NTNU bør konkretisere og beskrive de nye behandlingsoperasjonene som vil skje ved bruk av M365 Copilot for hver behandling.

Behandlingsprotokollen kan være et hensiktsmessig sted å starte. Det finnes mer informasjon om systematiske beskrivelser av behandlinger i Datatilsynets sjekkliste for DPIA-er.

Ofte er det flere behandlingsoperasjoner som gjøres med personopplysninger for et spesifikt formål. Når en saksbehandler behandler en søknad om opptak til et studieprogram og må besvare en henvendelse, kan hen først lete etter relevant informasjon i databaser hen har tilgang til. Dette kan inkludere tidligere korrespondanse med andre søkere om lignende henvendelser, tidligere vedtak eller interne retningslinjer. Slike søk innebærer behandling av personopplysninger, hvor resultatene kan inneholde både relevant og irrelevant informasjon og personopplysninger. Det er saksbehandleren som bestemmer hva som er relevant og hva hen vil ta med videre. En av nyvinningene med M365 Copilot er at slike behandlingsoperasjoner gjøres automatisk, og at innholdet oppsummeres og blir gjort tilgjengelig på en annen måte enn tidligere. Det er ikke gitt at dette medfører en *ny* behandlingsoperasjon, men det må vurderes. Hvis for eksempel flere personopplysninger behandles eller personopplysninger sammenstilles på en annen måte når M365 Copilot brukes i forbindelse med oppgaven, er det viktig å kartlegge og beskrive disse nye behandlingsoperasjonene.

Eksempel: I **brukstilfelle A** innebærer ikke oppgaven nødvendigvis behandling av personopplysninger uten bruk av M365 Copilot.

Ved å ta i bruk M365 Copilot vil opplysninger om brukeren og eventuelt andre kunne behandles fordi M365 Copilot vil lete etter og bruke opplysninger den finner i «nærliggende dokumenter» (e-poster, chatter, kalenderinvitasjoner osv.) for å berike instrukser og skape output som er mer relevant for brukeren. Omfanget av personopplysninger som kan behandles vil være avhengig av den enkelte brukerens tilgangsstyring, og kan også påvirkes av innstillingene til M365 Copilot (f.eks. ved å slå av «Graph-grounded chat») eller ved bruk av «effektiv instruksjonsutforming» (det vil si hvordan instruksen er utformet). Formålet med behandlingen kan beskrives som å hjelpe brukeren med å skrive en utredning mer effektivt. Dette kan anses som en helt ny behandling, for et formål som ikke eksisterte før M365 Copilot ble tatt i bruk i forbindelse med oppgaven.

Eksempel: I **brukstilfelle B** innebærer oppgaven behandling av personopplysninger uten bruk av M365 Copilot, og skal være beskrevet i behandlingsprotokollen allerede. Dette kan være beskrevet som følgende: Formålet med å føre møtereferat fra interne møter er å dokumentere interne beslutninger som blir tatt i virksomheten. Møtedeltakernes navn, rolle og (en oppsummering av) det som ble sagt i møtet nedtegnes skriftlig og lagres et sted hvor de som har saklig behov for det har tilgang.

Ved bruk av M365 Copilot vil nye behandlingsoperasjoner oppstå i form av opptak og transkripsjon av møtet som vil innebære behandling av flere personopplysninger enn før, slik som stemme, stemmeleie, uttrykksform, kjønn (antakelse fra stemmeleie), samt (person)opplysninger som M365 Copilot finner når den leter etter «nærliggende dokumenter». I tillegg kan de valgfrie innstillingene påvirke hvilke andre behandlingsoperasjoner som skjer (f.eks. oversikt over hvem snakker, når og hvor lenge).

Eksempel: I **brukstilfelle C** innebærer oppgaven behandling av personopplysninger også uten bruk av M365 Copilot, som skal være beskrevet i behandlingsprotokollen allerede, hvor formålet er å behandle søknader til opptak i et studieprogram. Saksbehandleren skal vurdere om all nødvendig informasjon er med i søknaden og besvare henvendelsen, som et ledd i saksbehandlingen. Det brukes personopplysninger som er mottatt i e-posten som er relevante for å vurdere om all nødvendig informasjon er mottatt.

Ved bruk av M365 Copilot kan nye behandlingsoperasjoner oppstå, men dette må vurderes opp mot hvordan søknader behandles i dag, herunder eventuelle eksisterende behandlingsaktiviteter knyttet til for eksempel søk. Det er viktig å undersøke om omfanget av personopplysninger som behandles er utvidet. Opplysninger om søkeren, brukeren og

eventuelt andre kan behandles fordi M365 Copilot vil lete etter og bruke opplysninger den finner i «nærliggende dokumenter» (e-poster, chatter, kalenderinvitasjoner osv.) for å berike instrukser og skape output som er mer relevant for brukeren, f.eks. et skreddersydd svar rettet mot søkeren men som ligner svar gitt til tidligere søkere. I tillegg kan de valgfrie innstillingene påvirke hvilke andre behandlingsoperasjoner som skjer (f.eks. ved å slå av «Graph-grounded chat»), eller ved bruk av «effektiv instruksjonsutforming» (det vil si hvordan instruksen er utformet).

Det er også viktig å vurdere nye behandlingsoperasjoner som vil oppstå uavhengig av hvilken oppgave M365 Copilot brukes i forbindelse med, f.eks. samhandlingsloggen og muligens profilering av brukeren, som kan skje for andre, nye formål.

Vurder det rettslige grunnlaget

Ifølge personvernforordningen artikkel 6, er behandling av personopplysninger kun lovlig dersom ett av vilkårene i nr. 1 bokstav a til f er oppfylt. NTNU har i DPIA-en sin skrevet at det er vanskelig å definere ett eller flere klare og tydelige formål for bruk av M365 Copilot. NTNU konkluderer med at behandlingsgrunnlaget for «M365 Copilot i driftsfase» er berettiget interesse, jf. personvernforordningen artikkel 6 nr. 1 bokstav f. Det er viktig for oss å påpeke at dette ikke harmoniserer med personvernforordningen, fordi «M365 Copilot i driftsfase» ikke er en «behandling». M365 Copilot kan brukes som et middel til å utføre mange forskjellige behandlingsoperasjoner til forskjellige formål med ulike rettslige grunnlag.

Det er viktig å vite hvilket rettslig grunnlag man kan bruke for hver planlagt behandling helst før DPIA-stadiet, og i hvert fall før M365 Copilot tas i bruk. Hvis særlige kategorier personopplysninger skal behandles, må slik behandling hjemles i et av unntakene i artikkel 9. Når de samme personopplysningene behandles for forskjellige formål, må behandlingen for hvert formål forankres i et eget rettslig grunnlag.[7]

Dersom det er snakk om en eksisterende behandling, må det gjøres en fornyet vurdering av vilkårene i det opprinnelige rettslige grunnlaget, basert på beskrivelsen av nye behandlingsoperasjoner.

Alle alternativene i personvernforordningen artikkel 6 nr. 1 bokstav b til f inneholder et vilkår om at «behandlingen [av personopplysninger] er **nødvendig**» (vår utheving). Nødvendighetsvilkåret vil være oppfylt dersom *formålet* med behandlingen ikke med *rimelighet* kan oppnås *like effektivt* med andre midler som er *mindre inngripende* i de registrertes rettigheter og friheter.[8]

Nødvendighetsvilkåret skal fortolkes innskrenkende, gitt at det tillater behandling av personopplysninger uten at den registrerte har gitt sitt samtykke.[9] Nødvendighet må for øvrig vurderes i sammenheng med dataminimeringsprinsippet som er nedfelt i artikkel 5 nr. 1 bokstav c, som krever at personopplysningene skal være «adekvate, relevante og begrenset til det som er nødvendig for formålene de behandles for».[10] Dataminimeringsprinsippet gir utrykk for forholdsmessighetsprinsippet.[11] Forholdsmessighet krever blant annet at fordelene ved å begrense en rettighet ikke oppveies av ulempene ved å utøve denne rettigheten.

Vi drøfter betydningen av nødvendighetsvilkåret i kontekst av personvernforordningen artikkel 6 nr. 1 bokstav e og f i det følgende.

---

[7] EDPB retningslinjer 1/2024 (ikke endelig vedtatt) om behandling av personopplysninger basert på personvernforordningen artikkel 6 nr. 1 bokstav f, s. 6, tilgjengelig via https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-12024-processing-personal-data-based_en.

[8] Dom av 4. juli 2023 [GC], *Bundeskartellamt*, C-252/21, EU:C:2023:537, avsnitt 108 og dom av 22. juni 2021 [GC], *Latvijas Republikas Saeima*, C-439/19, EU:C:2021:504, avsnitt 110 og 113.

[9] Dom av 4. oktober 2024 [C5], *Koninklijke Nederlandse Lawn Tennisbond*, C-621/22, EU:C:2024:857, avsnitt 31.

[10] Dom av 4. juli 2023 [GC], *Bundeskartellamt*, C-252/21, EU:C:2023:537, avsnitt 109.

[11] Dom av 22. juni 2021 [GC], *Latvijas Republikas Saeima*, C-439/19, EU:C:2021:504, avsnitt 98.

I enkelte norske rettskilder er det tatt til orde for at effektiv saksbehandling i offentlig forvaltning kan anses som en «viktig allmenn interesse» i personvernforordningen artikkel 9 nr. 2 bokstav g.[12] Etter vårt syn kan effektivitet i noen tilfeller tolkes inn i formål som baserer seg på personvernforordningen artikkel 6 nr. 1 bokstav e. Dette kan være relevant for NTNU som en offentlig institusjon.

NTNUs vurdering av nødvendighetsvilkåret under artikkel 6 nr. 1 bokstav e bør blant annet ta hensyn til følgende:

- Er M365 Copilot egnet til å oppfylle NTNUs formål på en bedre måte?
- Hvor mye bedre oppnår NTNU formålet med behandlingen om man tar i bruk M365 Copilot?
- Er det andre måter NTNU med rimelighet kan oppnå formålet like godt på?
- Hvor mye mer inngripende er de nye behandlingsoperasjonene for de registrertes personvernrelaterte rettigheter og friheter?
- Er det tiltak NTNU kan gjennomføre som gjør behandling med Copilot mindre inngripende?

Hvis det ikke er mulig å vurdere om nødvendighetsvilkåret oppfylles på dette stadiet, kan det vurderes i DPIA-stadiet hvor det skal vurderes «om behandlingsaktivitetene er nødvendige og står i et rimelig forhold til formålene», «en vurdering av risikoene for de registrertes rettigheter og friheter» og hvilke tiltak som kan håndtere risikoene og sikre vern av personopplysninger, jf. personvernforordningen artikkel 35 nr. 7.

Hvis nødvendighetsvilkåret ikke kan oppfylles, selv etter at tiltak identifisert i DPIA-en gjennomføres, kan M365 Copilot ikke tas i bruk på behandlingen.

For behandlinger som NTNU gjør i dag for et formål de har bestemt selv og som baserer seg på forfølgelse av en berettiget interesse jf. personvernforordningen artikkel 6 nr. 1 bokstav f, kan NTNU vurdere å ta inn effektivitet som en berettiget interesse de vil forfølge. EU-domstolen har sagt at å gjøre en tjeneste mer effektiv ikke kan utelukkes som en berettiget interesse.[13] Dette vil imidlertid ofte innebære en justering av formålet for behandlingen og utvide hvilke behandlingsoperasjoner som er nødvendig for å oppnå de berettigede interessene. Å kunne gjøre dette forutsetter at

- behandlingen ikke utføres som ledd i utførelsen av en offentlig myndighets oppgave, jf. personvernforordningen artikkel 6 nr. 1 annet ledd
- det nye formålet er forenlig med det opprinnelige formålet hvis, som ofte vil være tilfelle, personopplysninger som skal behandles ble innsamlet for et annet formål, jf. personvernforordningen artikkel 6 nr. 4
- NTNU gjennomfører en ny og oppdatert interesseavveining[14] som kommer ut i NTNUs favør
- NTNU overholder alle de andre pliktene i personvernforordningen[15]

Dersom et av vilkårene over ikke kan oppfylles, selv etter at tiltak identifisert i DPIA-en gjennomføres, kan M365 Copilot ikke tas i bruk på behandlingen.

For behandlinger for et nytt formål, må NTNU kunne identifisere et rettslig grunnlag for behandlingen på vanlig måte.

For at NTNU kan bruke samtykke som rettslig grunnlag for en behandling, må det være blant annet frivillig, spesifikt, informert og utvetydig. I lys av M365 Copilot betyr dette blant annet at NTNU må kunne forklare på en klar og tydelig måte, som gir forutsigbarhet til den registrerte, hvordan personopplysninger skal behandles når M365 Copilot tas i bruk. Dette kan være vanskelig, særlig hvis den registrerte har lite kunnskap om generativ KI og måten M365 Copilot fungerer på. I tillegg må man også se på styrkeforholdet mellom NTNU og den enkelte. For eksempel vil normalt ikke offentlige myndigheter eller arbeidsgivere kunne bruke samtykke som behandlingsgrunnlag siden den enkelte er i et avhengighetsforhold. Det betyr ikke at bruk av samtykke som behandlingsgrunnlag i lys av Microsoft 365 Copilot er helt utelukket, men det må vurderes konkret per brukstilfelle og tilhørende behandling hvorvidt samtykkevilkårene kan

---

[12] Prop. 135 L 2019-2020 pkt. 5.3.3.

[13] Dom av 4. juli 2023 [GC], *Bundeskartellamt*, C-252/21, EU:C:2023:537, avsnitt 122.

[14] Det europeiske Personvernrådets retningslinjer 1/2024 om behandling av personopplysninger basert på personvernforordningen artikkel 6 nr. 1 bokstav f er på offentlig høring og gir veiledning om hvordan å gjøre en interesseavveining. De er tilgjengelig via https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-12024-processing-personal-data-based_en.

[15] Dom av 4. oktober 2024 [C5], *Koninklijke Nederlandse Lawn Tennisbond*, C-621/22, EU:C:2024:857, avsnitt 50.

oppfylles. Også i arbeidstaker-/arbeidsgiverforhold kan det være tilfeller der arbeidsgiveren kan påvise at samtykke er faktisk frivillig, hvor det ikke vil være noen ulemper for den ansatte hvis de ikke samtykker til behandlingen.[16] Det er også viktig å huske på at en registrert kun kan samtykke til behandling av sine egne personopplysninger, mens bruk av Copilot ofte kan innebære at flere personers opplysninger behandles, selv om input eller output bare gjelder én person.

Derfor må det vurderes nøye om samtykke er et egnet rettslig grunnlag i lys av den enkelte behandlingen man står overfor, og om vilkårene for samtykke kan oppfylles.

[16] Retningslinjer 05/2020 om samtykke etter personvernforordningen, avsnitt 22.

Orden i eget hus

Orden i eget hus **er en grunnleggende forutsetning for å etterleve personopplysningsloven og personvernforordningen. Dette blir særlig viktig når M365 Copilot skal tas i bruk, fordi verktøyet fungerer som en «akselerator» og kan bringe opp til overflaten all informasjon den har tilgang til i løpet av sekunder.**

M365 Copilot kan anses som en «klone» av brukeren. M365 Copilot har de samme tilgangene og rettighetene som brukeren. Det vil si at alle dokumenter, e-poster, chatter og annet brukeren har tilgang til, er tilgjengelig for M365 Copilot. Selv om brukeren ikke vil få tilgang til ny informasjon med M365 Copilot, gjør verktøyet det mulig å raskt hente fram informasjon som tidligere har vært vanskelig tilgjengelig. Det kan være informasjon brukeren ikke skulle hatt tilgang til, og sannsynligvis ikke visste at hen hadde tilgang til. Dette øker risikoen for utilsiktet eller uautorisert bruk av data. Derfor må tilgangsstyring være nøye knyttet til brukerens rolle og behov i virksomheten.

Digitaliseringsdirektoratet har utarbeidet en veileder for orden i eget hus, inkludert en modenhetsmodell som hjelper offentlige virksomheter med å kartlegge og forbedre informasjonsforvaltning, med fokus på oversikt over egne datasett. Dette er en ressurs vi anbefaler alle offentlige virksomheter å gjøre seg kjent med.

NTNU må først ha god oversikt og kontroll over

1. avtaleverket og innstillingene for selve skytjenesten som M365 Copilot sitter på toppen av
2. informasjonsforvaltning generelt, herunder klassifisering, kategorisering og tilgangsstyring
3. behandling av personopplysninger, herunder en oppdatert og uttømmende behandlingsprotokoll

Dette er en utfordrende oppgave, både for store virksomheter med mye data og mange forskjellige systemer, slik som NTNU, men også for mindre virksomheter som kanskje ikke har den nødvendige kompetanse som kreves.

Informasjonsforvaltning

God informasjonsforvaltning bidrar til å nå flere mål:

- **Kvalitet på informasjonen:** Sikrer at informasjonen er nøyaktig, oppdatert og pålitelig.
- **Sikkerhet:** Beskytter sensitiv informasjon mot uautorisert tilgang og sikkerhetsbrudd.
- **Etterlevelse:** Hjelper organisasjonen med å overholde juridiske krav som personvernforordningen og offentlighetsloven.
- **Effektivitet:** Forbedrer arbeidsflyt og beslutningstaking ved å gjøre informasjon lett tilgjengelig og forståelig.
- **Redusert risiko:** Minimerer risikoen for tap av data, juridiske sanksjoner og omdømmeskade.

God informasjonsforvaltning forutsetter etablering av grunnleggende retningslinjer for hvordan informasjon skal håndteres innad i organisasjon. Informasjonskartlegging for å identifisere hvilken informasjon som finnes, hvor den lagres, og hvem som har tilgang, er et viktig steg og legger også grunnlaget for tilgangskontroll. Relevant opplæring av ansatte er et viktig ansvar for virksomheten.

For å lykkes med dette kreves robuste rutiner som løpende klassifiserer informasjon basert på sensitivitet og juridiske krav, inkludert personvernforordningen. Tilgang til de ulike kategoriene informasjon må begrenses ut fra saklig behov. Livssyklusadministrasjon fra opprettelse til arkivering eller sletting er også en del av denne prosessen.

Moderne informasjonsforvaltning krever automatisering og bruk av verktøy som sikrer effektivitet, enkelhet og konsistens i prosessene. Disse må også kunne håndtere behovet for å regelmessig evaluere og oppdatere praksiser for å tilpasse seg endrede behov og reguleringer, samt understøtte behovet for å gjennomføre interne og eksterne revisjoner for å sikre overholdelse av lover og interne retningslinjer.

Krav til moderne informasjonsforvaltning kan også utløse behov for organisatoriske endringer, hvis rollene for å understøtte prosessene ikke allerede er etablert med klare mandater.

Som det understrekes i funn fra NTNU-rapporten, vil et verktøy som M365 Copilot kunne påvirke organisasjonen og bør først og fremst ansees å være et organisasjonsendringsprosjekt og et informasjonsforvaltningsprosjekt heller enn et IT-prosjekt.

Virksomheten bør vurdere om det er prosesser som bør tilpasses eller endres for å integrere M365 Copilot på en effektiv måte, like mye som å forsøke å tilpasse produktet til organisasjonens eksisterende prosesser. Denne type verktøy kan konfigureres til en viss grad for å møte spesifikke behov, men det er viktig å forstå verktøyets begrensninger og styrker. Da trenger man bevissthet og kunnskap om hvilke tiltak som best understøtter behovet for å hente ut gevinster, samt å sikre at man etterlever kravene til ansvarlighet og lovlighet.

Behandlingsprotokoll

Alle virksomheter som behandler personopplysninger skal føre protokoll over behandlingsaktivitetene sine, jf. artikkel 30. Behandlingsprotokollen skal vise formålene med hver enkelt behandling, en beskrivelse av hvem sine og hvilke personopplysninger som behandles, mottakere av personopplysninger (hvis aktuelt) og hvorvidt personopplysninger overføres til land utenfor EØS. I tillegg, hvis mulig, må tidsfristene for sletting vises, samt en generell beskrivelse av de tekniske og organisatoriske sikkerhetstiltakene nevnt i artikkel 32 nr. 1.

Før man vurderer om M365 Copilot kan innføres, og for å kunne vurdere hvorvidt og hvordan det kan skje, er det helt nødvendig å ha en uttømmende og oppdatert behandlingsprotokoll på plass, slik vi allerede har nevnt ovenfor. Da kan virksomheten som første steg vurdere, per behandling, hvorvidt M365 Copilot er egnet til å tas i bruk i forbindelse med den og, hvis ja, hvordan.

Ved å ta i bruk M365 Copilot, vil flere nye behandlinger uunngåelig oppstå. Dette inkluderer som nevnt lagringen av hver brukers logg over samhandlinger med M365 Copilot (samhandlingsloggen). Dette er en ny behandling, og det må vurderes særlig hva formålet med den er, når loggen skal slettes og hvem i virksomheten som skal ha tilgang til den. Vi drøfter også anvendelse av den norske e-postforskriften lenger ned. Administratorer har tilgang til brukerens samhandlingslogg og har mulighet til å søke i den ved bruk av eDiscovery.[17] Andre nye behandlinger som oppstår kan være at M365 Copilot genererer opplysninger når den svarer på instrukser, og disse kan være personopplysninger. Det er også uklart for NTNU hvorvidt det skjer en profilering av brukeren. NTNU har valgt å regne det som høyst sannsynlig at profilering skjer.[18] Hvilke nye behandlingsoperasjoner bruk av M365 Copilot medfører, vil variere noe ut fra hvilke innstillinger som skrus av eller på, og disse nye behandlingene må også inn i behandlingsprotokollen.

Tilgangsstyring

I det følgende skriver vi om tilgangsstyring som sikkerhetstiltak etter personvernforordningen artikkel 32.

Det overordnede kravet etter personvernforordningen artikkel 32 nr. 1 er at den behandlingsansvarlige gjennomfører «egnede tekniske og organisatoriske tiltak for å oppnå et sikkerhetsnivå som er egnet med hensyn til risikoen» ved behandlingen av personopplysninger. Hensikten er at sikkerhetstiltakene skal stå i et rimelig forhold til den konkrete risikoen ved behandlingen.

Personvernforordningen stiller ikke spesifikke krav til det nærmere innholdet i sikkerhetstiltakene. For offentlige myndigheter gjelder likevel en plikt til å bruke etablerte standarder ved anskaffelse, utvikling, oppsett, drift og bruk av IT-løsninger, jf. forskrift 5. april 2013 nr. 959 om IT-standarder i offentlig forvaltning § 14. Det finnes en rekke slike standarder for personopplysningssikkerhet, som gjennomgående stiller krav om at tiltak som tilgangsstyring, logging og loggkontroll er på plass, se for eksempel ISO/IEC 27002:2022 kapittel 5 og 8. Det er på det rene at tilgangsstyring er et nødvendig element i tiltakene som er påkrevd etter artikkel 32.

---

[17] eDiscovery, eller «electronic discovery» er prosessen med å identifisere, samle inn, og analysere elektronisk lagret informasjon som kan brukes som bevis i juridiske saker eller interne undersøkelser. Microsoft Purview eDiscovery er en løsning innen Microsoft Purview som hjelper organisasjoner med å håndtere eDiscovery-prosessen.

[18] NTNUs funnrapport, s. 119.

Hva menes egentlig med tilgang?

Vi har merket oss at begrepet «tilgang» i praksis anvendes på litt ulike måter når folk snakker om M365 Copilot. Vi vil derfor forklare noe nærmere hva vi mener med tilgangsstyring som sikkerhetstiltak.

NTNU understreker i sin funnrapport viktigheten av «aktiv stillingtaken til hvilke data M365 Copilot skal ha tilgang til».[19] For å hindre at verktøyet *brukes* på feil data, vil tiltak for å styre sluttbrukernes *tilgang* til data være hjelpsomt. Tilgangsstyring kan understøtte arbeidet med å sikre at personopplysninger bare brukes innenfor rammene av det en har rettslig grunnlag for etter personvernforordningen artikkel 6 og 9. Men det er ikke dette som ligger i kjernen av tilgangsstyring som *sikkerhetstiltak*. Som sikkerhetstiltak er målet med tilgangsstyring først og fremst å sørge for at personopplysninger har et egnet konfidensialitetsvern, jf. personvernforordningen artikkel 5 nr. 1 bokstav f og artikkel 32 nr. 1 bokstav b.

Eksempel fra UiO

Betydningen av dette kan illustreres med et eksempel fra virkeligheten. Khrono omtalte 27. juni 2024 en avvikshendelse hos UiO, hvor jobbsøkeres CV og ansettelseskomiteens vurderinger lå åpent tilgjengelig for alle ansatte ved universitetet.[20] Som formildende omstendighet ble det trukket fram at opplysningene var vanskelig tilgjengelig. I avviksmeldingen som ble sendt til Datatilsynet, gjengitt i Khronos artikkel, skrev UiO:

> *«For å finne [opplysningene] må ansatte enten lete aktivt, eller komme over det ved en tilfeldighet. Dette reduserer sannsynligheten for at opplysningene faktisk har blitt eksponert for uvedkommende, og risikoen for at hendelsen kan medføre skade for de berørte, men UiO kan ikke utelukke at det har skjedd en utilsiktet eksponering av personopplysninger til uvedkommende.»*

Slike argumenter har vanligvis en viss gyldighet når det kommer til vurderingen om hvorvidt det er sannsynlig at et sikkerhetsbrudd har påvirket den registrerte. Det vil imidlertid stille seg annerledes for virksomheter som bruker M365 Copilot. M365 Copilot henter enkelt fram opplysninger fra obskure kilder, som brukeren selv kanskje ikke visste at hen hadde tilgang til. Dette øker sannsynligheten for at personopplysninger eksponeres ulovlig.

Datatilsynets erfaring er at denne typen avvik – hvor personopplysninger lagres på steder som gjør dem tilgjengelige for uvedkommende – er svært vanlig. Tilgangsstyring og klassifisering av informasjon bør derfor være et høyt prioritert sikkerhetstiltak dersom virksomheten vurderer å innføre M365 Copilot.

---

[19] NTNUs funnrapport side 2.

[20] https://www.khrono.no/alle-uio-ansatte-hadde-tilgang-til-informasjon-om-jobbsokere/885123 (sist besøkt 04.09.2024).

Vurdering av personvernkonsekvenser

Plikten til å gjøre en konkret vurdering av personvernkonsekvenser følger av personvernforordningen artikkel 35. I den engelske versjonen omtales dette som *Data Protection Impact Assessment* – derav forkortelsen DPIA. DPIA-er skal minst inneholde de fire elementene som fremgår av artikkel 35 nr. 7 bokstav a til d:

a. En systematisk beskrivelse av de planlagte behandlingsaktivitetene og formålene med behandlingen, herunder, dersom det er relevant, den berettigede interessen som forfølges av den behandlingsansvarlige.
b. En vurdering av om behandlingsaktivitetene er nødvendige og står i et rimelig forhold til formålene.
c. En vurdering av risikoene for de registrertes rettigheter og friheter.
d. De planlagte tiltakene for å håndtere risikoene, herunder garantier, sikkerhetstiltak og mekanismer for å sikre vern av personopplysninger og for å påvise at denne forordning overholdes, idet det tas hensyn til de registrertes og andre berørte personers rettigheter og berettigede interesser.

Plikten til å gjøre en DPIA inntreffer dersom det er sannsynlig at «en type behandling, særlig ved bruk av ny teknologi (…) vil medføre en høy risiko for fysiske personers rettigheter og friheter», jf. artikkel 35 nr. 1. Det er også verdt å merke seg at ordlyden «fysiske personers rettigheter og friheter» må forstås som en henvisning til EUs charter om grunnleggende rettigheter, som langt på vei samsvarer med EMK. Det er altså ikke bare konsekvensene for personvernet som skal analyseres, men også konsekvensene for rettigheter som ytrings- og informasjonsfrihet og ikke-diskriminering.[21] Charteret er ikke del av EØS-avtalen, men får indirekte virkning gjennom personvernforordningen.

For sammenhengens del kan vi også nevne at offentlige virksomheter som tar i bruk det som i henhold til KI-forordningen klassifiseres som «høyrisiko»-KI-systemer[22], i de fleste tilfeller får en lignende konsekvensanalyse-plikt gjennom KI-forordningen artikkel 27 som etter personvernforordningen artikkel 35.

Ny teknologi, nye konsekvenser?

Om det kreves en DPIA ved bruk av M365 Copilot, altså hvorvidt behandlingen av personopplysninger medfører en høy risiko for fysiske personers rettigheter og friheter, avhenger av flere faktorer. Det er relevant å se på de spesifikke oppgavene verktøyet skal utføre og for hvilke formål, sammenhengen verktøyet brukes i samt arten og omfanget av behandlingen av personopplysninger. Det som kan være vanskelig, særlig med bruk av ny teknologi, slik som generativ KI, er at man ikke er kjent med hvordan teknologien eller produktet fungerer, og det gjør det krevende å fastslå hvilke potensielle risikoer det innebærer, for ikke å nevne sannsynligheten for slike risikoer.

Vi vurderer det slik at en DPIA som hovedregel vil kreves ved bruk av generative KI-verktøy slik som M365 Copilot i forbindelse med behandling av personopplysninger, ettersom «bruk av ny teknologi» er fremhevet som en særlig viktig faktor, og forståelsen av risiko knyttet til generativ KI ennå er umoden. Å gjennomføre en eller flere DPIA-er, uansett om det er påkrevd eller ikke, vil hjelpe NTNU å vurdere spesifikke risikoer og sannsynligheten for slike risikoer i en gitt kontekst. Det vil også bidra til å belyse det NTNU ikke vet om enten produktet, teknologien eller forutsetningene for teknologien sett i sammenheng med en bestemt behandling (f.eks. for å kunne vurdere om en har god nok «orden i eget hus»). Det vil også bidra til å påvise etterlevelse av ansvarsprinsippet etter personvernforordningen artikkel 5 nr. 2.

Flere DPIA-er?

Det kan høres tungt ut å måtte gjøre flere DPIA-er, særlig hvis virksomheten har flere hundre forskjellige behandlingsaktiviteter, men hvis ikke man gjør det slik, vil man sannsynligvis ikke klare å etterleve prinsippene om formålsbegrensning, dataminimering og lovlighet. Dette er årsaken til at det sjeldent vil være ansvarlig eller lovlig å slå på M365 Copilot for hele virksomheten og for alle rollene i virksomheten. Vi vurderer det slik at en stegvis tilnærming til innføring av M365 Copilot er mest hensiktsmessig, hvor innføring først vurderes for et begrenset område, f.eks. en rolle og de tilhørende behandlingene som denne rollen utfører.

---

[21] Den europeiske unions charter om grunnleggende rettigheter artikkel 11 og 21.

[22] Vi har ikke vurdert om M365 Copilot vil klassifiseres som et «høyrisiko»-KI-system.

Det følger imidlertid av artikkel 35 nr. 1 siste setning at én DPIA kan omfatte flere lignende behandlingsaktiviteter som innebærer tilsvarende høye risikoer. Det er derfor mulig å vurdere flere behandlinger som M365 Copilot vurderes brukt til i én og samme DPIA, så lenger de «ligner», og her vil både formålet, omfanget av personopplysningene og det som gjøres med dem være relevante. Når det kommer til omfanget av personopplysningene som kan behandles, er det viktig å se på hvilken rolle som skal utføre oppgaven og hva deres tilgang er. En saksbehandler vil for eksempel ikke ha de samme tilgangene som en fra HR eller en fra ledelsen.

Dessuten vil informasjon eller vurderinger i én DPIA ha overføringsverdi til en annen DPIA.

Vurder før konsekvensene kan komme

Det er viktig at vurderingene gjøres *før* behandlingsoperasjonene starter. Dersom M365 Copilot skal anvendes i forbindelse med en eksisterende behandling, må vurderingen gjøres før verktøyet brukes. Men det stopper ikke der. Som NTNU poengterer i sin funnrapport, er M365 Copilot tidlig i utviklingsløpet og krevende å forvalte på grunn av hyppige endringer som påvirker risikobildet.[23] DPIA-er må derfor gjennomføres som en kontinuerlig prosess, jf. også artikkel 35 nr. 11. En av NTNUs klare anbefalinger er å utarbeide en exit-strategi i tilfellet det skjer endringer som gjør at bruken f.eks. må anses ulovlig.[24] Ytterligere veiledning om DPIA-er er å finne på Datatilsynets hjemmesider[25] og i publikasjoner fra henholdsvis Artikkel 29-gruppen og EDPB.[26]

I det videre går vi gjennom noen utvalgte temaer fra en DPIA som vi har sett særlig på i sandkasseprosjektet. Det er likevel viktig at NTNU også vurderer de andre temaene som påkreves når de utfører DPIA-er på konkrete behandlinger.

En systematisk beskrivelse av behandlingen

Det er viktig å merke seg at en DPIA etter artikkel 35 krever at man konkretiserer de planlagte behandlingsaktivitetene og formålet, herunder den berettigete interessen som skal forfølges dersom det er relevant. Dette sammenfaller med «kartlegg og beskriv behandlingen» som vi omtaler over og inkluderer alle behandlingsoperasjonene som inngår i den. Det er likevel slik at mye av informasjonen som NTNU har fremskaffet i utarbeidelsen av deres overordnede vurdering, vil ha overføringsverdi og vil lette arbeidet med de konkrete DPIA-ene betydelig.

Nødvendighet og proporsjonalitet av behandlingen

NTNU har sagt at det ikke vil være gjennomførbart for dem å kunne begrunne nødvendighet og relevans for formålet for hver enkelt av variablene i de datasettene som finnes i en brukers tilgang til Microsoft 365-plattformen, uten en mer grundig gjennomgang og systematisk oppfølging.[27] Nødvendighet og relevans bør imidlertid kunne vurderes hvis de(n) aktuelle behandlingen(e) er beskrevet systematisk først i lys av et konkret brukstilfelle.

Formålsbegrensning

Formålsbegrensningsprinsippet sier at personopplysninger skal samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles på en måte som er uforenlig med disse formålene. Når man skal vurdere behandlingens formål, er det viktig å huske på at M365 Copilot er et verktøy eller funksjon – et middel – til å oppnå formålet med behandlingen. Bruken av M365 Copilot er ikke et formål i seg selv.

---

[23] NTNUs funn 4 og 5, se NTNUs funnrapport s. 12-15.

[24] NTNUs funnrapport s. 12 og 31.

[25] Se https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/vurdering-av-personvernkonsekvenser/ (sist besøkt 16. juli 2024).

[26] Artikkel 29-gruppen: *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is «likely to result in a high risk» for the purposes of Regulation 2016/679* (WP 248), tilgjengelig via https://ec.europa.eu/newsroom/article29/items/611236/en.

EDPB: *Recommendation 01/2019 on the draft list of the European Data Protection Supervisor regarding the processing operations subject to the requirement of a data protection impact assessment*, tilgjengelig via https://www.edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendation-012019-draft-list-european-data_en.

[27] NTNUs funnrapport, s. 104.

«Å hjelpe brukeren med å gjennomføre sine oppgaver» er for vagt og generelt, men formålsbegrensningsprinsippet kan oppfylles hvis det spesifiseres og konkretiseres mer, f.eks. ved å spesifisere nærmere hva slags oppgave det er snakk om og hvorfor den utføres.

Det er som nevnt en særskilt utfordring ved M365 Copilot at formålet, og hvilke personopplysninger som brukes for å oppnå formålet, i praksis defineres (kontrolleres) av den enkelte brukeren i hver enkelt instruks. Bruk av M365 Copilot gjør at personopplysninger kan behandles i en annen kontekst – og for et annet formål – enn opprinnelig tiltenkt. Det skjer fordi M365 Copilot bruker all informasjon som er tilgjengelig for brukeren via Microsoft Graph. NTNU identifiserte funksjoner med M365 Copilot som gjør at personopplysninger samlet inn for ett formål kan bli viderebehandlet til nye eller andre formål.[28] Det er derfor viktig at NTNU setter klare rammer for sine brukere, for eksempel i form av retningslinjer, rutiner og opplæring, for i størst mulig grad å sikre blant annet formålsbegrensning ved bruk av M365 Copilot. Dette bør helst ses i kontekst av brukerens «rolle» i virksomheten, som også vil samsvare med denne rollens tilganger.

Det er en åpning for å viderebehandle personopplysninger for nye formål enn formålet, så lenge det nye formålet er forenlig med det opprinnelige. Personvernforordningen artikkel 6 nr. 4 oppstiller en ikke-uttømmende liste over hva som skal vektlegges i denne vurderingen. Per nå er det ingen retningslinjer eller rettsavgjørelser på hvordan denne bestemmelsen skal anvendes eller forstås, men en slik forenlighetsvurdering må gjøres per behandling og ikke for M365 Copilot som helhet.

Dette er åpenbart vanskelig å praktisere, men kan være lettere ved bruk av M365 Copilot for noen utvalgte roller som utfører et begrenset omfang av behandlinger. Det som er særlig vanskelig med dagens M365 Copilot, er at det ikke er mulig å slå av tilgang til en brukers e-postkasse. Det betyr at e-poster og personopplysningene de inneholder lett kan brukes til andre formål enn det som var opprinnelig tiltenkt. Datatilsynet har ikke fasitsvaret på hvordan den behandlingsansvarlige skal sørge for at formålet med behandling av personopplysninger i kontekst av M365 Copilot er forenlige med det opprinnelige formålet personopplysningene ble samlet inn for.

Et mulig tiltak kan være å lære opp brukere til å avgrense søkeområdet gjennom selve instruksen ved å bruke effektiv instruksjonsutforming. I denne sammenhengen vil det være en forutsetning at virksomheten har god «orden i eget hus» og retningslinjer. I tillegg fremgår det av NTNUs funnrapport[29] at M365 Copilot kan stilles inn til ikke å bruke opplysninger fra bestemte områder, f.eks. Teams chat. Vi synes at Microsoft bør utvikle innstillingene som gjør det mulig å stenge av opplysninger også fra e-poster, hvor det praktisk talt er umulig å ha kontroll på hva som er omfattet.

Dataminimering

Dataminimering er et ufravikelig krav etter personvernforordningen artikkel 5 nr. 1 bokstav c: «[personopplysninger skal] være adekvate, relevante og begrenset til det som er nødvendig for formålene de behandles for». Dataminimering må derfor vurderes med utgangspunktet i formålet med den aktuelle behandlingen, for å finne ut hvilke personopplysninger som trengs for å oppnå formålet. Som sagt over, har ikke NTNU identifisert en konkret behandling med et spesifikt formål ennå («M365 Copilot i driftsfase» er ikke et formål), og dette er det NTNU som må gjøre.

Dataminimeringsprinsippet er kanskje en av de vanskeligste pliktene å overholde ved bruk av M365 Copilot, fordi verktøyet er bygd sånn at det har tilgang til alt en bruker har tilgang til og derfor har mulighet til å behandle dataene det selv «vurderer» som relevant ut fra instruksen etter berikingsprosessen. Det er ikke mulig å vite akkurat hvordan M365 Copilot «velger» hva som er relevant ut ifra instruksen, på grunn av både black box-problematikken og at det uansett vil være proprietær informasjon.

Effektiv instruksjonsutforming kan være et tiltak som kan brukes til å minimere kildetilfang, men det er uklart hvorvidt det vil oppfylle dataminimeringsprinsippet fullt ut. Et annet potensielt tiltak er aktivering av Double Key Encryption (DKE) for å sperre av filer som ikke skal brukes av M365 Copilot.[30] Det beste hadde imidlertid vært mer granulære innstillinger i M365 Copilot hva gjelder kildetilfang, særlig angående tilgang til en brukers e-postkasse.

---

[28] NTNUs funnrapport, s. 68.

[29] NTNUs funnrapport, s. 18.

[30] NTNUs funnrapport, s. 117.

Riktighet

Riktighetsprinsippet i personvernforordningen innebærer at «det må treffes ethvert rimelig tiltak for å sikre at personopplysninger som er uriktige med hensyn til formålene de behandles for, uten opphold slettes eller rettes». Som nevnt over, kan svarene som genereres av M365 Copilot være feil. De kan samtidig fremstå overbevisende og riktig. Risikoen for uriktige svar vil aldri kunne elimineres. Det er derfor viktig at brukere møter KI-genererte svar med en viss skepsis.

Et KI-generert svar er basert på sannsynlighet og er avhengig av modellens treningsdata og vekting. NTNU skriver at det er sannsynlig at M365 Copilot vil kunne finne på ting som både er usant og feilaktig[31] (også kjent som «hallusinering»), og at det ligger i verktøyets natur at det kan gi uriktige opplysninger.[32] M365 Copilot kan gjøre feilslutninger også når den tilsynelatende har tilgang til god nok informasjon.[33] Sannsynligheten for feil øker om brukeren gir upresise instrukser til språkmodellen. Dette kan bli et enda større problem hvis brukeren skal kontrollere svaret, men ikke har nok tid eller informasjon til å sjekke om svaret inneholder feil eller ikke.

Dette kommer enda mer på spissen ved behandling av personopplysninger. I henhold til riktighetsprinsippet er det en lovpålagt plikt at personopplysninger skal være korrekte, og det må treffes ethvert rimelig tiltak for å sikre at personopplysninger som er uriktige med hensyn til formålene de behandles for uten opphold slettes eller rettes. Hvis M365 Copilot genererer feil personopplysninger om noen, kan det for det første være vanskelig for brukeren å kontrollere hvorvidt svaret inneholder feil, og for det andre kan det innebære en høy risiko for den registrertes rettigheter.

NTNU ønsker at brukere skal kunne ta aktivt stilling til informasjonen løsningen gir og være grunnleggende kritisk til informasjon som en språkmodell gir nettopp for å motvirke faren for at uriktige opplysninger blir oppfattet som riktige. Samtidig anerkjenner NTNU at det særlig bør vurderes om M365 Copilot skal benyttes i prosesser hvor det er naturlig med kontradiksjon. Det er derfor fornuftig å vurdere hvilke områder eller oppgaver som ikke egner seg for bruk av generative KI-verktøy. Dette kan for eksempel være noen oppgaver innenfor HR eller utøving av offentlig myndighet, som krever en høy grad av presisjon og riktighet og hvor konsekvensene av feil kan være alvorlige. Dette påpeker NTNU i sin funnrapport (funn 2).

I kontekst av generativ KI mener vi at dette betyr både at det bør være implementert tiltak som minsker risikoen for at feil personopplysninger genereres (f.eks. ved effektiv instruksjonsutforming eller regler om hva M365 Copilot ikke skal brukes i forbindelse med) og tiltak som retter eller sletter feil personopplysninger uten opphold (f.eks. effektiv etterkontroll av det som genereres). Hvis M365 Copilot skal brukes som beslutningsstøtte, må tiltak som regler, retningslinjer, opplæring og valg av brukere med riktig kompetanse vurderes og innføres. Det er viktig at den som bruker M365 Copilot er kritisk til svarene og har både tid og kompetanse til å oppdage og rette feil personopplysninger som kan forekomme i output.

Dette kan begrense hva M365 Copilot vil kunne brukes til. F.eks. vil det som hovedregel være uaktuelt å bruke M365 Copilot hvor personer eller personopplysninger er i kjernen av oppgaven som skal utføres. Vi er også enige med NTNU når de sier at terskelen bør settes veldig høyt for bruk av M365 Copilot innenfor utøvelse av offentlig myndighet, hvor riktighet står sentral.[34]

---

[31] NTNUs funnrapport, s. 71.

[32] NTNUs funnrapport, s. 105.

[33] NTNUs funnrapport, s. 9.

[34] Ibid.

De registrertes rettigheter og friheter

NTNU har identifisert en rekke risikoer knyttet til de registrertes rettigheter og friheter ved NTNUs bruk av M365 Copilot på et overordnet nivå. Disse beskrives i NTNUs funnrapport s. 106-117. Her vil vi særlig fremheve retten til informasjon, retten til innsyn, retten til å protestere og forbudet mot automatiserte avgjørelser.

Retten til informasjon

Hjørnesteinen i de registrertes rettigheter er rett til informasjon. Den behandlingsansvarlige må forklare, på en klar og enkel måte, hvordan personopplysninger behandles. Dette er en forutsetning for at de registrerte skal kunne utøve sine rettigheter. Personvernforordningen artikkel 12 til 14 samt åpenhetsprinsippet i artikkel 5 pålegger behandlingsansvarlige å gi registrerte informasjon om hvordan deres personopplysninger skal behandles, og disse pliktene oppfylles oftest ved å bruke både eksterne og interne personvernerklæringer.

Når ny teknologi innføres, er det viktig at påvirkningen teknologien kan ha på den enkelte gjenspeiles i aktuell informasjon og aktuelle personvernerklæringer og at det er klart og tydelig når, hvordan og i hvilken kontekst slik teknologi skal tas i bruk på den registrertes personopplysninger. For at det skal være forståelig for de registrerte, vil det ofte også være nødvendig å forklare på en kortfattet måte hvordan selve teknologien fungerer.

NTNU har identifisert at personvernerklæringen må oppdateres med informasjon om bruk av M365 Copilot, basert på en realitetsvurdering av det som faktisk skjer.[35] Dette bør inkludere informasjon om hvorvidt M365 Copilot brukes i forbindelse med behandlingen og for hvilket formål, samt hvilke nye behandlingsaktiviteter og behandlinger som oppstår ved bruk av M365 Copilot og hvilket rettslige grunnlag som brukes per behandling. For å sørge for åpenhet, bør det også forklares på en så enkel måte som mulig hvordan verktøyet fungerer.

Ved innføring av M365 Copilot, bør de registrerte få tilsendt informasjon direkte, hvor mulig, om hvilken behandling NTNU skal bruke M365 Copilot i forbindelse med og hva det betyr for den enkelte. For eksempel kan dette gjøres i en e-post som sendes direkte til den registrerte. Hvis det ikke er mulig å ta direkte kontakt med registrerte, bør informasjonen komme tydelig frem i kontaktpunktene de registrerte har med NTNU, f.eks. på NTNUs nettside.

Informasjon til de registrerte må gjennomgås og vurderes kontinuerlig eller med jevne mellomrom i tråd med teknologiutviklingen og hvorvidt og hvordan M365 Copilot tas i bruk på nye områder i NTNU.

Det bør også komme tydelig frem hvilket innhold som er laget med hjelp fra generativ KI. Dette er særlig viktig når innholdet inkluderer personopplysninger, og det vil gjøre det lettere for de registrerte å utøve rettighetene sine og ha kontroll over personopplysninger sine.

Retten til innsyn

Output generert av M365 Copilot kan inneholde personopplysninger og vil da være gjenstand for retten til innsyn fra den det gjelder. NTNU har sagt at det må vurderes hvorvidt et innsynskrav kan besvares fullt ut da NTNU vil ha utfordringer med å identifisere alle steder personopplysninger kan bli behandlet i M365 Copilot.[36] Dette er også identifisert av NTNU som et problem ved bruk av Microsoft 365-plattformen uten bruk av M365 Copilot.

Som sagt over, bør innhold som genereres med hjelp av generativ KI merkes. Hvor generert innhold lagres vil være avhengig av det spesifikke brukstilfellet, men det bør samsvare med lagringsstedet der andre dokumenter for oppgaven/brukstilfellet er lagret før bruk av M365 Copilot. En ting som er særlig nytt med bruk av M365 Copilot, er lagring av samhandlingsloggen. Samhandlingsloggen vil lagres i tråd med NTNUs gjeldende «retention policy» (se delen om lagringsbegrensning over) og vil kunne søkes i av administratorer. Den registrerte vil likevel kun ha rett til innsyn i sine egne personopplysninger som eventuelt er lagret i samhandlingsloggen, og ikke hele samhandlingsloggen generelt. Retten til innsyn skal ikke ha negativ innvirkning på andres rettigheter og friheter, jf. personvernforordningen artikkel 15 nr. 4.

---

[35] NTNUs funnrapport, s. 106-7.

[36] NTNUs funnrapport, s. 107.

Retten til å protestere

Registrerte kan protestere mot NTNUs behandling av sine personopplysninger når behandlingsgrunnlaget er personvernforordningen artikkel 6 nr. 1 bokstav e eller f, jf. personvernforordningen artikkel 21 nr. 1. Hvis protesten innvilges, kan det være vanskelig for NTNU å etterleve protesten på grunn av M365 Copilots iboende egenskaper, hvor dens tilganger gjenspeiler brukerens tilganger. NTNU har løftet en mulig løsning for dette, som er å bruke Double Key Encryption (DKE) som kan aktiveres for filer som inneholder de aktuelle personopplysningene.[37]

Forbud mot automatiserte individuelle avgjørelser

Samtidig som en vurderer behandlingsgrunnlag, bør en se hen til forbudet mot automatiserte individuelle avgjørelser i forordningen artikkel 22, som setter noen grenser for hva M365 Copilot kan brukes til. Personvernforordningen artikkel 22 inneholder et forbud mot automatiserte individuelle avgjørelser som består av tre kumulative vilkår: (1) det må være en «avgjørelse», som er et begrep som skal tolkes vidt[38], (2) avgjørelsen må utelukkende baseres på automatisert behandling, og (3) den må ha rettsvirkning for eller på tilsvarende måte i betydelig grad påvirke vedkommende. Det en høy terskel for hva som omfattes av forbudet. Utgangspunktet er at det meste ikke rammes, men vurderinger kan rammes dersom beslutningstaker i realiteten lener seg blindt på M365 Copilots vurdering.

> **Eksempel**: Vurderingen av søknaden om tilrettelegging i brukstilfelle C vil ha «rettsvirkning» for eller «i betydelig grad påvirke» søkeren, og kan derfor ikke overlates til M365 Copilot alene. Et menneske kan imidlertid bruke M365 Copilot som en beslutningsstøtte, så lenge hen ikke lener seg blindt på M365 Copilots vurdering.

NTNU identifiserte funksjoner hvor en bruker kan spørre M365 Copilot om å vurdere en kollegas atferd og arbeidsprestasjon. Selv om denne handlingen i utgangspunktet ikke rammes av forbudet i personvernforordningen artikkel 22, vil det også innebære en behandling av personopplysninger som neppe vil ha et gyldig rettslig grunnlag.

Risikoreduserende tiltak

Dersom man kommer frem til at man har behandlingsgrunnlag og DPIA-en gir grønt lys for behandlingen, er testing et anbefalt risikominimerende tiltak. Testing må skje innenfor rammene av det/de identifiserte behandlingsgrunnlagene. Avhengig av hva som er formålet med testingen og hva som testes, kan personvernforordningen artikkel 32 også utgjøre et såkalt supplerende behandlingsgrunnlag.

NTNU har identifisert mange forskjellige risikoer for de registrertes rettigheter og friheter samt mulige risikoreduserende og skadebegrensende tiltak, som omtales i deres funnrapport, s. 121-126. Hele 41 tiltak er listet opp. Disse risikoene og tiltakene er identifisert basert på en overordnet gjennomgang av M365 Copilot. Likevel vil mange av de risikoreduserende tiltakene være relevante å vurdere når det gjøres en mer spesifikk vurdering av personvernkonsekvenser i lys av en konkret behandling eller et sett med lignende behandlinger.

En særskilt utfordring når det gjelder M365 Copilot, er at formålet med en behandling i praksis defineres (kontrolleres) av den enkelte bruker i hver enkelt instruks. Det er derfor viktig at den behandlingsansvarlige setter klare rammer for sine ansatte, blant annet i form av rutiner og opplæring, for i størst mulig grad å sikre at behandlingen utføres lovlig.

Det vil imidlertid ikke være realistisk for den behandlingsansvarlige å oppnå fullstendig kontroll. Derfor vil det være nødvendig med en etterfølgende kontroll av den faktiske bruken. En slik kontroll må i seg selv ha et rettslig grunnlag. Her er det verdt å merke seg at bestemmelsene i forordningen selv kan utgjøre et supplerende rettsgrunnlag etter artikkel 6 nr. 3. Jo større virksomheten er, jo vanskeligere vil det være å oppnå kontroll, og jo større risiko vil det være for uønskede hendelser. NTNU, med sine 70 000 brukere, omtaler det som «utopisk» å få brukerne til å følge rutiner.

---

[37] NTNUs funnrapport, s. 117.

[38] Dom av 7. desember 2023 [C5], *Schufa*, C-634/21, EU:C:2023:957, avsnitt 60.

Involvering av personvernombud

Når en virksomhet vurderer å bruke et nytt KI-verktøy slik som M365 Copilot, er det viktig at personvernombudet involveres tidlig. Personvernombudet bør anses som en nøkkelperson i både vurderingen, innføringen og etterkontrollen av KI-verktøyet. Personvernforordningen artikkel 39 sier noe om personvernombudets oppgaver, som blant annet inkluderer å gi råd om personvernforpliktelsene og vurderinger av personvernkonsekvenser samt kontrollere gjennomføringen av slike vurderinger. Personvernombudet skal utføre oppgavene sine på en uavhengig måte.

Personvernombudet må ha en forståelse av hele livssyklusen til KI-systemet som virksomheten vurderer å skaffe og hvordan det fungerer. Dette betyr at personvernombudet blant annet må få informasjon om når, hvorfor og hvordan et slikt system behandler personopplysninger, hvordan dataflyten fungerer (input og output) samt beslutningstakingsprosesser i modellen.

Forbudet mot overvåking i e-postforskriften

I Norge har vi en egen forskrift som gjelder for arbeidsgivers innsyn i ansattes e-postkasse og annet elektronisk lagret materiale (e-postforskriften).

Når en bruker samhandler med Copilot, lagres det data om disse samhandlingene som inneholder brukerens instrukser og M365 Copilots svar, blant annet referanser og henvisninger til kildemateriale (samhandlingsloggen).[39] Samhandlingsloggen lagres i en skjult mappe i brukerens «mailbox». Slike skjulte mapper er ikke utformet til å bli direkte tilgjengelig for brukere eller administratorer, men kan søkes opp av etterlevelsesadministratorer med «eDiscovery tools».[40]

Sletting av loggen

Hvorvidt og når samhandlingsloggen slettes permanent er avhengig av virksomhetens lagringspolicy. En bruker kan ha mulighet til å slette samhandlingsloggen selv som et alternativ i M365 Copilot-innstillingene, men det er uklart for NTNU om sletting av egen logg også medfører sletting av loggen som administrator kan se.[41] Microsoft sier selv at *«Messages visible in Copilot are not an accurate reflection of whether they are retained or permanently deleted for compliance requirements»*.[42] Samhandlingsloggen vil først flyttes til «the SubstrateHolds»-mappen. Først etter at lagringsperioden satt av virksomheten utløper, vil samhandlingsloggen slettes permanent.

E-postforskriften[43] omfatter både e-postkasser, personlige områder i virksomhetens datanettverk og annet elektronisk utstyr som en arbeidsgiver har stilt til arbeidstakerens disposisjon til bruk i arbeidet ved virksomheten. Forskriften gjelder tilsvarende for opplysninger som er slettet, dersom de finnes på sikkerhetskopier eller tilsvarende. Det er på det rene at samhandlingsloggen faller inn under forskriftens virkeområde.

E-postforskriften inneholder for det første vilkår for når arbeidsgiveren har rett til å gjøre enkeltstående innsyn i opplysninger som er lagret på ovennevnte område. For det andre inneholder forskriften i § 2, andre ledd et forbud mot overvåking av arbeidstakers bruk av elektronisk utstyr, med mindre formålet med overvåking er

a. å administrere virksomhetens datanettverk eller
b. å avdekke eller oppklare sikkerhetsbrudd i nettverket

Datatilsynets veiledning[44] omtaler at forbudet gjelder hvis : (1) tiltaket dreier seg om overvåking, (2) overvåkingen retter seg mot arbeidstakeres bruk av elektronisk utstyr og (3) arbeidsgiver har tilgang til opplysningene.

Er samhandlingsloggen overvåking?

Det er klart at samhandlingsloggen viser historikken over en arbeidstakers bruk av elektronisk utstyr. Spørsmålet blir derfor om samhandlingsloggen er å anse som overvåking, etter e-postforskriften. Samhandlingsloggen kan regnes som overvåking, selv om dette ikke har vært formålet fra arbeidsgivers side. Relevante momenter i vurderingen er blant annet hva slags opplysninger og mengden opplysninger som framgår av loggen, hvor lenge opplysningene lagres og hvor stor del av arbeidshverdagen som kan spores.

---

[39] Data, personvern og sikkerhet for Microsoft Copilot for Microsoft 365 | Microsoft Learn (åpnet på 26. august 2024)

[40] Learn about retention for Microsoft Copilot for Microsoft 365 | Microsoft Learn (åpnet 26. August 2024)

[41] NTNUs funnrapport, s. 79.

[42] Learn about retention for Microsoft Copilot for Microsoft 365 | Microsoft Learn (åpnet 26. August 2024)

[43] Forskrift om arbeidsgivers innsyn i e-postkasse og annet elektronisk lagret materiale av 2. juli 2018

[44] Overvåking av ansattes bruk av elektronisk utstyr | Datatilsynet

Samhandlingsloggen kan avsløre mye om noens atferd og arbeidshverdag. M365 Copilot er bygget inn i alle verktøyene de ansatte bruker til daglig, slik som Word, Excel, Power Point og Teams. Den omfattende kartleggingen som skjer når de ansatte bruker M365 Copilot taler for at samhandlingsloggen er å anse som overvåking.

Tilgang til loggen

Dersom bare arbeidstakeren selv hadde hatt tilgang til loggen, ville vi likevel vært utenfor forbudet. Som beskrevet ovenfor, gjelder forbudet hvis arbeidsgiver har tilgang til opplysningene.

NTNU identifiserte at arbeidsgiver har tilgang til samhandlingsloggen ved bruk av eDiscovery og Purview. Det finnes i tillegg en funksjon som fører til at en brukers instruks sendes til kontroll hvis bestemte kriterier oppfylles – med andre ord at en «alarm» utløses. Disse kriteriene kan endres av NTNU selv. Det er likevel uklart hvor mye av samhandlingsloggen som sendes til kontroll, til hvem og til hvilket formål.

Samlet sett anser vi det som sannsynlig at samhandlingsloggen vil kunne rammes av forbudet mot overvåking av arbeidstakers bruk av elektronisk utstyr i e-postforskriften § 2 andre ledd. For at samhandlingsloggen skal kunne opprettes lovlig, må formålene med samhandlingsloggen falle inn under ett av unntakene.

Unntak fra forbudet

Det ene unntaket kommer til anvendelse dersom formålet er å «administrere virksomhetens datanettverk». Dette forstår vi som alle praktiske og tekniske tiltak som er nødvendige for at systemene, nettverk, utstyr og programvare skal fungere.[45] Vi har forstått det slik at hovedformålet med samhandlingsloggen er å sikre at kvaliteten på tjenesten er slik den skal være. Ved å gå gjennom brukerens instrukser og M365 Copilots svar kan NTNU avdekke svakheter og forbedringspotensialer ved systemet. Mulige systematiske feilsvar kan bli fanget opp og rettet opp. Dette formålet kan falle inn under unntaket for å administrere virksomhetens datanettverk.

Det andre unntaket kan være aktuelt dersom formålet med samhandlingsloggen er å «avdekke eller oppklare sikkerhetsbrudd i nettverket». Med sikkerhetsbrudd forstår vi brudd på informasjonssikkerheten generelt. Det er vanlig å si at det handler om å sikre at informasjon ikke blir kjent for uvedkommende (konfidensialitet), ikke blir endret utilsiktet eller av uvedkommende (integritet) og er tilgjengelig ved behov (tilgjengelighet). Hvorvidt dette unntaket kommer til anvendelse må vurderes konkret opp mot hva som er formålet med samhandlingsloggen.

For begge unntakene må NTNU være oppmerksom på prinsippene om dataminering og formålsbegrensning.[46]

---

[45] Du kan lese mer om formålene i Datatilsyets veileder til Overvåking av ansattes bruk av elektronisk utstyr | Datatilsynet

[46] En nærmere beskrivelse og eksempler på dataminimering og formålsbegrensning for begge unntakene kan du lese om i Overvåking av ansattes bruk av elektronisk utstyr | Datatilsynet.

Avslutning

M365 Copilot har et betydelig potensial, men krever at man tar små og kontrollerte steg. Løsningen forutsetter fra Microsofts side at man har svært god kontroll på egen informasjonsforvaltning, en grad av kontroll som nok mange virksomheter ikke vil kunne påberope seg.

Det vil neppe være mulig å benytte M365 Copilot på en ansvarlig og lovlig måte uten å legge ned betydelig forarbeid. Dette inkluderer både å få orden i eget hus og å gjennomføre grundige vurderinger av personvernkonsekvenser for de planlagte bruksområdene. Teknologien stiller også høye krav til opplæring av ansatte, og til bevissthet og kunnskap blant virksomhetens brukere.

Den positive siden ved dette arbeidet er at et sterkt fokus på informasjonsforvaltning kan gi store gevinster som strekker seg langt utover selve bruken av verktøyet. Effektiv og velfungerende informasjonsforvaltning er selve grunnlaget for å lykkes med digitalisering, samfunnsnyttig datadeling og kostnadseffektiv etterlevelse av lover og regler, inkludert personvernforordningen. Å ta i bruk ny og avansert teknologi uten å forberede seg grundig, forstå muligheter og begrensninger, og sikre nødvendig kompetanse, vil ikke være ansvarlig.

Det er mulig for norske virksomheter å bruke M365 Copilot, men bruksområdene bør velges med omhu for å sikre etterlevelse av blant annet personvernkrav. Samtidig er det både riktig og viktig å prøve ut ny teknologi og skaffe få praktisk erfaring med mulighetene den gir. Forutsetningen er at de nødvendige vurderingene blir gjort i forkant, og at virksomheten har gode prosesser og etablerer tiltak for å redusere identifiserte risikoer.

Den samme språkmodellteknologien som ligger til grunn for M365 Copilot, kan også brukes på flere og mer spissede måter enn bare som et generelt kontorstøtteverktøy. Slike tilnærminger kan medføre lavere krav til organisatoriske endringer, raskere uthenting av gevinst og ikke minst bedre kontroll med kvalitet og etterlevelse av regelverk. Derfor er det viktig å vurdere om det finnes andre KI-løsninger som kan dekke virksomhetens konkrete behov, men som har mindre personvernrisiko. Det kan også være at slike mer fokuserte løsninger kan være et godt startsted for senere bruk av integrerte KI-løsninger som M365 Copilot. Ved først å sikre «orden i eget hus» og etablere støttemekanismer for etterlevelse av krav, kan virksomheter stå bedre rustet til å ta i bruk avanserte løsninger når produkter som M365 Copilot kanskje har fått litt tid til å modnes og tilpasses ytterligere.

Veien videre

NTNU har valgt å ikke innføre M365 Copilot for hele virksomheten, men heller ta i bruk verktøyet kontrollert i små steg med utvalgte roller først. I og med at M365 Copilot krever egne lisenser, må økte kostnader forsvares gjennom faktisk og realiserbar gevinst, og det er viktig at både direkte og indirekte kostnader tas med i helhetsvurderingen. M365 Copilot er fortsatt tidlig i utviklingsløpet og mangler kontroll på et granulært nivå, som muligheten til å gjøre lokale og fleksible tilpasninger (f.eks. å slå av tilgang til brukeres e-postkasser, spesifikke slettepolicyer osv.). Ubegrenset tilgang til brukers e-postkasse regnes antagelig av Microsoft som en viktig og sentral funksjon, samtidig som dette kanskje er et av elementene som mange virksomheter føler skaper mest usikkerhet.

Datatilsynet forventer at utfordringer som kunder, virksomheter, myndigheter og samfunnet generelt identifiserer i produktet, tas på alvor av produktleverandør. Samtidig setter også løsningen klare krav til virksomheter som ønsker å hente gevinster ved bruk av verktøyet. Forutsetningen om en svært velfungerende informasjonsforvaltning og orden i eget hus kan gjøre veien for å lykkes med slike løsninger krevende, men åpenbart med en positiv oppside langt utover det å ta i bruk en bestemt løsning.

NTNU har gjort en imponerende, samfunnsnyttig og omfattende jobb med å skaffe seg kunnskap og bevissthet rundt bruken av språkmodeller generelt og integrerte KI-løsninger som M365 Copilot spesielt. Ønsker NTNU å utvide bruken av M365 Copilot videre, er det imidlertid viktig at nødvendige vurdering av personvernkonsekvenser gjøres på konkrete behandlinger i lys av gitte brukstilfeller.

Vedlegg

Språkmodeller

Språkmodeller (Large Language Models – LLM), slik som GPT (Generative Pre-trained Transformer), er maskinlæringsmodeller som er trent på svært store mengder tekstdata. Disse modellene bearbeider og genererer tekst ved å bruke sammenhenger i treningsdataene til å forutsi neste ord i en tekst eller når den svarer på et spørsmål. De brukes i en rekke applikasjoner, som samtaleroboter, tekstgenerering og språkanalyse.

Språkmodellene er bygget på nevrale nettverk. De lagrer ikke språk og ord som tekst, men som numeriske representasjoner kalt vektorer, som effektivt beskriver svært komplekse sammenhenger mellom språkelementer. En språkmodell «forstår» ikke språk i menneskelig forstand, men modellerer sammenhengen mellom språkelementer basert på hvordan språk brukes av mennesker.

Fordi språkmodeller er så velartikulerte, er det lett å oppfatte språkmodeller som kunnskapsmodeller, men det er de ikke. De er modeller for språket i seg selv og hvordan det brukes i praksis. Dagens språkmodeller har med andre ord ikke innebygd kunnskap om fagområder som juss, kjemi, fysikk, filosofi og matematikk. Samtidig gjenspeiler språk i sin natur informasjon om verden rundt oss. Språkmodeller trenes på store mengder tekst som inneholder (tilfeldige) opplysninger ulike temaer, og på denne måten blir forskjellige typer informasjon ofte avspeilet i språket som modellen er trent opp på. Teknologien er i rivende utvikling, og språkmodeller som kombineres med kunnskapsmodeller er både under utprøving og vil i økende grad også bli tilgjengelig for allmenn bruk, med potensial for mer presise og pålitelige svar.

En viktig utfordring med språkmodeller er fenomenet «hallusinering». Dette betyr at modellen genererer tekst som språklig sett er korrekt, men som inneholder feil eller oppdiktet informasjon. Språkmodeller fungerer som avanserte, statistiske modeller uten en innebygd forståelse av fakta. De har en viss tilfeldighet innebygget (for å kunne gi varierte og/eller alternative formuleringer av svar), men mangler mekanismer for å vurdere innholdets sannhet, noe som kan føre til at generert tekst framstår troverdig, men faktisk er feil.

I tillegg til hallusinering kan feil oppstå på grunn av misoppfatninger i treningsdataene. Hvis en utbredt feil eller misoppfatning er til stede i datagrunnlaget, vil modellen kunne gjenta eller forsterke denne feilen. For eksempel, dersom det i treningsmaterialet finnes mange kilder som feilaktig påstår noe, vil modellen sannsynligvis gjenspeile dette som om det var korrekt. Dette kan være problematisk når modeller brukes i situasjoner der det kreves høy grad av presisjon eller faglig korrekthet.

I praksis vil allikevel mange responser være gode og relevante, gitt at det store flertallet av tekstene de er trent på inneholder den riktigste informasjonen, og fordi de språklige sammenhengene i mange tilfeller også inneholder de relevante faktaene.

De fleste store leverandørenes språkmodeller er primært trent på engelskspråklige data, noe som gjør at ofte generere bedre responser på engelsk. Det pågår arbeid for å tilpasse språkmodeller til nasjonale språk. Men vil de også klare å tilpasse dem til nasjonale kulturer? Dette er et viktig å være oppmerksom på ved bruk av språkmodeller, fordi de også reflekter kulturell kontekst. En overvekt av engelske og amerikanske tekstkilder i treningsmaterialet vil bety at tekst som genereres også er påvirket av og reflekterer engelsk og amerikansk kultur.

De største og mest dominerende språkmodellene, som OpenAIs modeller, er laget, trent og driftet av store, private, amerikanske selskaper. Microsoft benytter OpenAI-modeller for å levere språkmodelltjenester på sin Azure-plattform, med justeringer og tilpasninger til egne produkter som f.eks M365 Copilot.

I motsetning til «klassiske» KI-/maskinlæringssystemer som i hovedsak er modellert og trent for spesifikke formål, har språkmodeller den egenskapen at de kan brukes til mange og uspesifiserte oppgaver. Derfor omtales de også som generelle grunnmodeller (foundation models). Dette gjør dem svært anvendelige, men samtidig utfordrende med tanke på å sikre korrekthet, relevans og ansvarlig bruk.

Tilpasning av språkmodeller

M365 Copilot benytter flere teknikker for å tilpasse produktet, inkludert kunnskapsgrafer (Knowledge Graphs) og Retrieval-Augmented Generation (RAG). Hensikten med RAG er å styre kvaliteten på tekstbaserte svar ved å hente

utvalgt og oppdatert informasjon fra interne informasjonskilder før svaret genereres. Denne nye tilleggsinformasjonen vektoriseres og indekseres i samme numeriske format som den opprinnelige grunnmodellen.

RAG består av tre hovedkomponenter:

1. **Retrieval (gjenfinning):** Modellen søker etter informasjon i en database eller eksterne kilder. Dette likner på hvordan tradisjonelle søkemotorer fungerer, men RAG benytter såkalte semantiske søkemetoder for å finne relevant informasjon basert på språklig kontekst, i stedet for kun å basere seg på søkeord.
2. **Augmented (forsterkning):** Den innhentede informasjonen brukes til å berike språkmodellens svar. Dette gjør at svarene blir mer presise og faktuelle, sammenlignet med grunnmodellene som primært baserer seg på forhåndstrente data.
3. **Generation (generering):** Etter at relevant informasjon er hentet, genereres responser som inneholder tilleggsinformasjon fra steg 1 og 2 ved hjelp av selve grunnmodellen.

RAGs semantiske søk baserer seg på den språklige konteksten i et spørsmål (instruks/ledetekst). Dermed kan systemet finne informasjon som er relevant selv om de eksakte ordene ikke samsvarer. Dette gjør søkene mer fleksible og kan finne flere relevante sammenhenger for brukeren.

RAG har flere egenskaper:

- **Oppdatert kunnskap:** I motsetning til språkmodeller som bare kan trekke på egne statiske treningsdata, kan RAG hente virksomhetsintern informasjon og informasjon fra nye kilder man selv kontrollerer, noe som gir mer nøyaktige svar.
- **Fleksibilitet:** Systemet kan oppfatte komplekse intensjoner og sammenhenger i spørsmål, selv om ikke alle søkeord er tilstede.
- **Nøyaktighet:** Ved å spesifisere og styre hvilken informasjon som skal brukes som underlag for svar, reduserer RAG risikoen for feil eller "hallusinasjoner" sammenliknet med svar produsert av grunnmodellen alene.

Bruksområder for RAG

Hensikten med RAG er å i større grad kunne styre hvilken informasjon som skal inngå i svar fra språkmodeller. Den muliggjør også noe tilpassing til spesifikke fagdomener for å øke sannsynligheten for at informasjonen som presenteres er relevant og korrekt. Dette forutsetter imidlertid svært god kontroll på hvilken informasjon som inngår i RAG-modellen. Generelle språkmodell-løsninger som M365 Copilot vil i større grad kunne forankre svar til virksomhetens egen informasjon, men er fremdeles avhengig av kvaliteten på denne informasjonen. Uklassifisert, gammel, utdatert eller feil informasjon i interne kilder vil påvirke kvaliteten negativt.

Selv om RAG kan forbedre språkmodellers evne til å gi svar forankret i virksomhetens egen informasjon, er det også utfordringer knyttet til implementering, inkludert overvåking av kvalitet på svar. I tillegg er det tekniske utfordringer knyttet til skalering og ytelse når disse modellene brukes i stor skala, bl.a. kan mange tilleggsoperasjoner medføre lenger svartid.

Språkmodeller som GPT representerer en viktig teknologisk innovasjon innen tekstgenerering, men har sine begrensninger når det gjelder oppdatert og faktabasert informasjon. RAG-systemet kan imidlertid, hvis det er implementert riktig, øke kvaliteten på svarene ved å bruke virksomhetsintern informasjon.

Datatilsynets regulatoriske sandkasse for ansvarlig kunstig intelligens

Besøksadresse:
Trelastgata 3, Oslo

Postadresse:
Postboks 458 Sentrum
0105 Oslo

sandkasse@datatilsynet.no
Telefon: +47 22 39 69 00

datatilsynet.no/sandkasse
personvernbloggen.no
twitter.com/datatilsynet