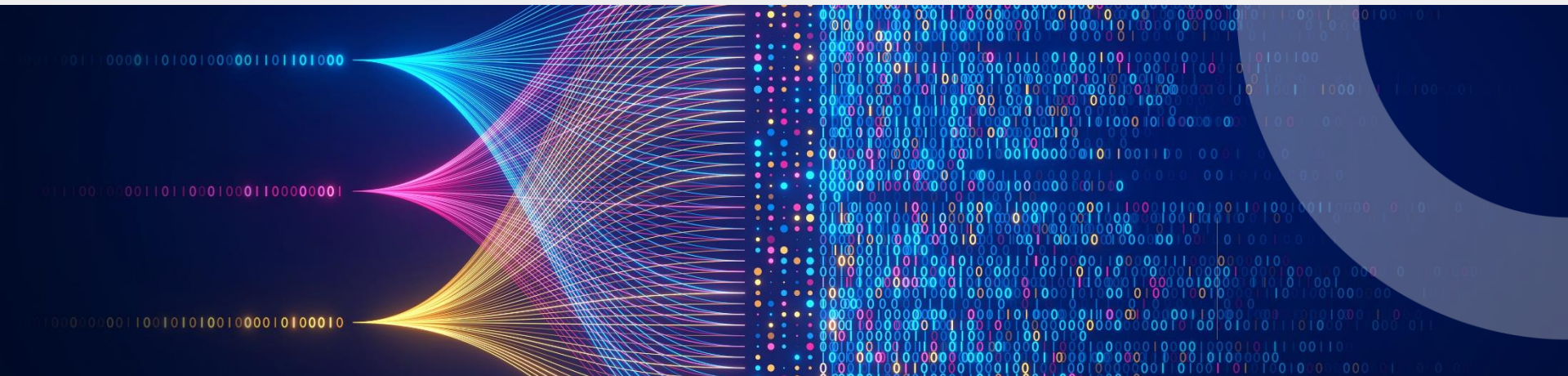


Språkteknologi i Norge 2023

Sammenstilling av funn fra to ulike spørreundersøkelser om bruk og utvikling av språkteknologi





Innhold

Del 1: Språkteknologi i digitalisering og kunstig intelligens

- Datagrunnlag og metode
- Oppsummering av hovedfunn
- Tidligere kartlegging av bruk og behov for språkteknologiske løsninger
- Dagens og fremtidens bruk av språkteknologiske løsninger
- Egenskaper og funksjonaliteter ved språkteknologien, samt håndtering av språkdata

Del 2: Ressursbehov til utvikling av språkteknologi

- Datagrunnlag og metode
- Oppsummering av hovedfunn
- Om virksomhetene og virksomhetenes aktiviteter
- Bruk av grunnlagsressurser og språkdata
- Behov for grunnlagsressurser
- Offentlige anbud

Del 1: Språkteknologi i digitalisering og kunstig intelligens

Denne delen sammenstiller funn fra en undersøkelse om bruk av språkteknologi i digitalisering og kunstig intelligens i offentlige virksomheter.



Datagrunnlag og metode

På oppdrag for Språkrådet har Oslo Economics gjennomført en spørreundersøkelse blant offentlige virksomheter i Norge. Formålet med undersøkelsen har vært å kartlegge bruk av og behov for språkteknologiske i offentlige virksomheter, samt virksomhetenes håndtering av språkdata som grunnlagsressurs for språkteknologi.

Språkteknologi handler om prosessering av språk og omfatter alle verktøy som involverer automatisert behandling av språk, også allmenne verktøy som stavekontroller og skriveprogrammer. Den gjennomførte undersøkelsen er utformet av Språkrådet. Oslo Economics har gitt innspill til utforming og oppbygging av undersøkelsen.

Listen over offentlige virksomheter ble utarbeidet ved hjelp av data fra Oslo Economics bedriftsdatabase og enhetsregisteret i Brønnøysundregistrene. Fra enhetsregisteret hentet vi informasjon om alle virksomheter i Norge med tilhørende organisasjonsform, næringskode og sektorkode. Private virksomheter ble filtrert ut fra utvalget (se Vedlegg 1 gj. Fra Oslo Economics' bedriftsdatabase hentet vi andre virksomhetsopplysninger som epostadresser. Virksomheter hvor vi ikke hadde informasjon om epost-adresser ble filtrert ut av utvalget. Et lite utvalg av virksomhetene, som ble definert som særlig relevante for undersøkelsen, ble kontaktet i forkant av utsendelsen. Formålet var å innhente kontaktinformasjon til aktuelle respondenter.

Undersøkelsen ble sendt til 2 767 offentlige virksomheter. Undersøkelsen ble sendt ut i starten av desember 2023, og det ble sendt ut en påminnelse i uke 50. Undersøkelsen ble avsluttet ved årsskiftet og fikk totalt 188 svar.

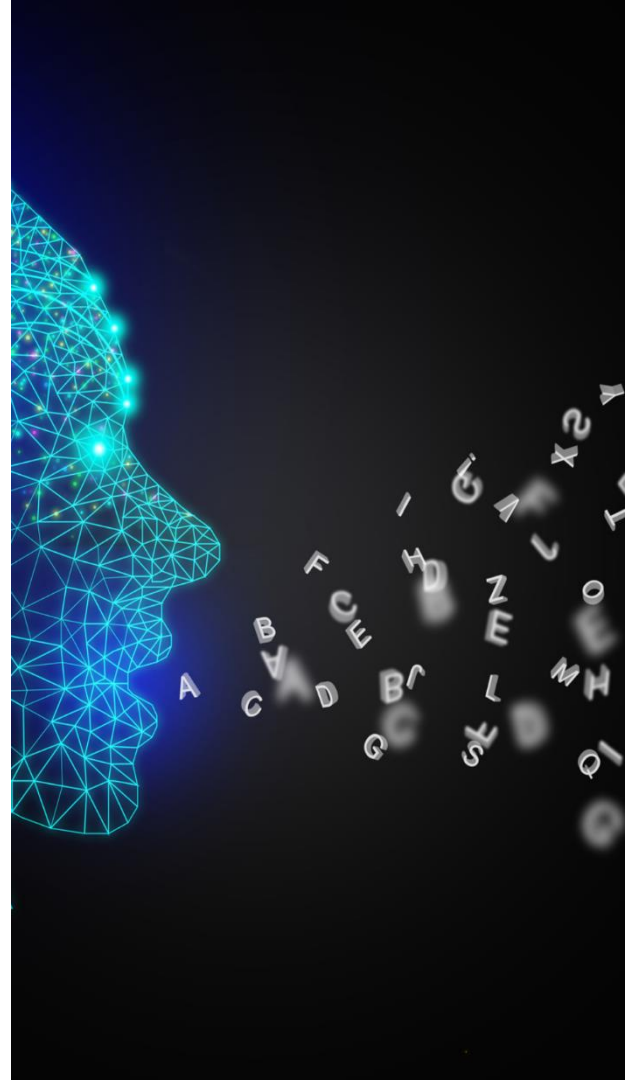
Undersøkelsen var anonym og det ble ikke innhentet noen bakgrunnsinformasjon om respondentene. Av denne årsaken har det ikke vært mulig å sammenligne kjennetegn ved respondentene med den totale populasjonen. Det kan derfor være mulige skjevheter i informasjonsgrunnet. Dette, sammen med en relativt lav svarprosent, gjør at resultatene må tolkes med varsomhet.

Tidligere kartlegging av bruk og behov for språkteknologiske løsninger

Undersøkelsen som nå har blitt gjennomført blant offentlige virksomheter bygger videre på en tidligere undersøkelse fra 2019/2020, gjennomført av Ipsos på oppdrag for Språkrådet og Direktoratet for forvaltning og IKT (Difi). Undersøkelsen ble gjennomført blant ledere med ansvar for utviklings-/IKT-arbeid i statlige virksomheter og et utvalg av kommunale og fylkeskommunale virksomheter.

Undersøkelsen fant at 9 av 10 offentlige virksomheter benyttet språkteknologiske løsninger, hvorav én av tre benyttet dette i stor eller svært stor grad. Undersøkelsen fant også at et flertall av virksomheten opplevde at språkteknologiske løsninger bidro til effektivisering av arbeidet og bedrer tjenestene til brukerne/innbyggerne. Videre avdekket undersøkelsen at om lag 70 prosent av virksomhetene som brukte språkteknologiske løsninger hadde et behov for nye eller flere løsninger.

Kun 1 av 10 av virksomhetene kjente til ressursene for utvikling av språkteknologiske løsninger i Språkbanken, mens 1 av 5 virksomheter oppga at de utvekslet informasjon om bruk av språkteknologiske løsninger med Språkrådet, norske utdanningsinstitusjoner med kompetanse i språkteknologi og andre virksomheter.





Oppsummering av hovedfunn

Av de 188 virksomhetene som har besvart undersøkelsen oppgir 78 prosent at de bruker språkteknologiske løsninger. Til tross for at disse resultatene ikke er direkte sammenlignbare med resultatene fra den tidligere undersøkelsen, kan de indikere at bruken av språkteknologi har blitt mer utbredt. 69 prosent av virksomhetene oppgir at de ønsker eller planlegger å ta i bruk en eller flere nye løsninger i fremtiden.

De fleste virksomheter (86 prosent) benytter språkteknologi på norsk, men kun 37 prosent tar hensyn til støtte på både bokmål og nynorsk når de går til anskaffelse av ny teknologi.

Blant virksomhetene som oppgir at de ønsker/planlegger å ta i bruk nye løsninger, finner vi at dette vanligst er i form av automatisk referatskriving, arkivering, transkribering og tekstsammendrag.

Av de 146 virksomhetene som bruker språkteknologiske løsninger, oppgir 50 prosent at de kjenner til og følger opp lovpålagte språklige krav ved anskaffelse og bruk av digitale verktøy i det offentlige.

25 prosent av virksomhetene er kjent med at Språkrådet kan tilby rådgivning knyttet til språkteknologi, men kun 5 prosent har benyttet seg av denne tjenesten.

Kun 10 prosent av virksomhetene oppgir at de er kjent med muligheten for å levere fra seg språkdata til Språkbanken, mens kun 2 prosent oppgir at de også har gjort dette.

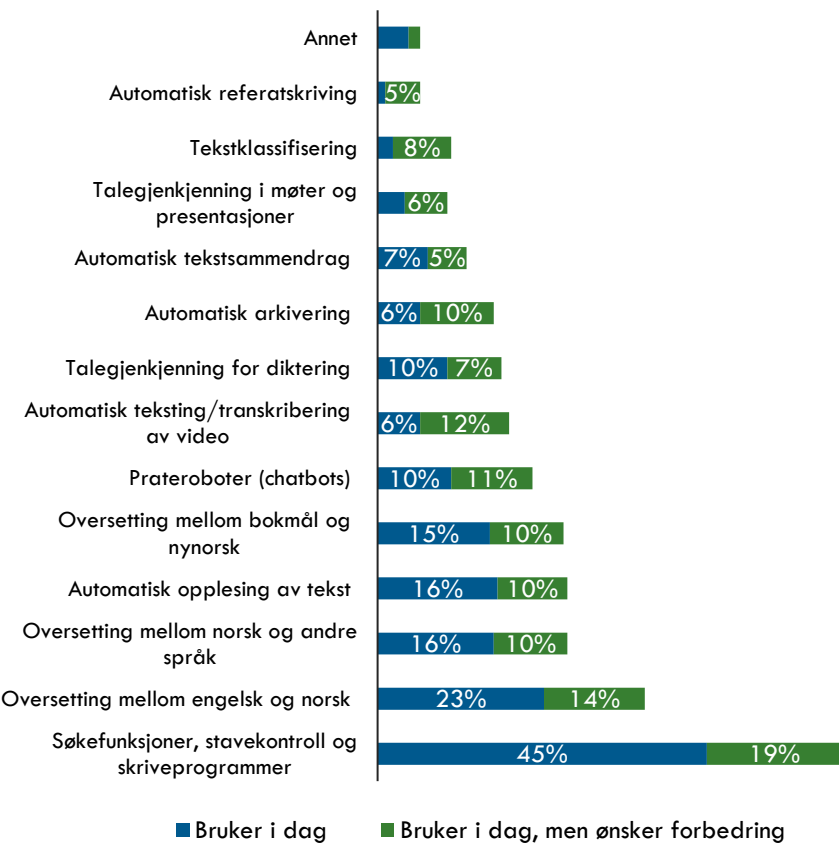
Overnevnte resultater samsvarer med funn fra den tidligere undersøkelsen og kan tyde på at kjennskapen til og praksisen for å benytte seg av Språkrådet og/eller Språkbanken sine tjenester ikke har blitt mer utbredt de siste årene.

Del 1:1 Dagens og fremtidens bruk av språktechnologiske løsninger

I denne delen av undersøkelsen ser vi nærmere på hvilken type språktechnologi som er i bruk i de offentlige virksomhetene i dag og hvilken type språktechnologi som man planlegger/ønsker å ta i bruk i fremtiden.



Bruk av språkteknologi i offentlige virksomheter i dag



Antall respondenter: 169, 19 respondenter svarte ikke på spørsmålet

Av de 188 virksomhetene som har besvart undersøkelsen oppgir 146 (78 prosent) at de bruker språkteknologiske løsninger. Til sammenligning var det 30 prosent av virksomhetene som i 2029/2020 oppga at de brukte språkteknologiske løsninger i ganske stor eller svært stor grad.

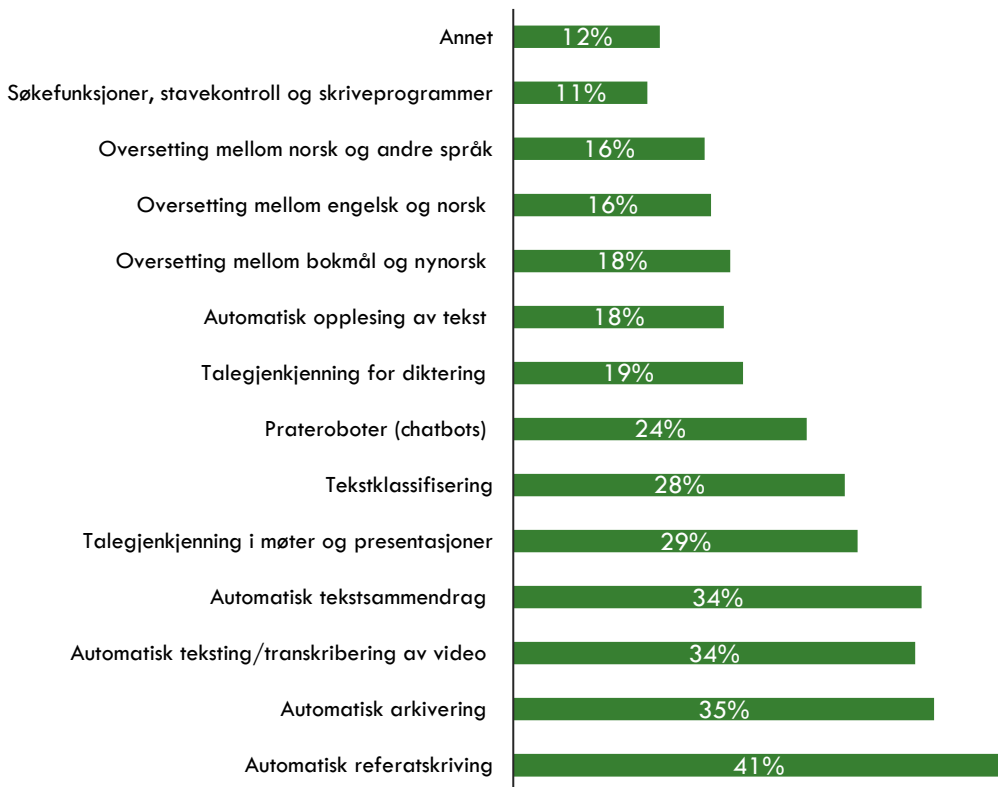
Blant virksomhetene som ikke bruker språkteknologi i dag, oppgir litt over 50 prosent at de ønsker og/eller planlegger å ta i bruk minst én form for språkteknologi.

Vanligst synes det å være at virksomheten bruker teknologi til søk- og skrivehjelp, samt til øversetting. Om lag 65 prosent av virksomhetene oppgir at de bruker språkteknologi i form av søkefunksjoner, stave-kontroll og skriveprogrammer. Rett under 40 prosent av virksomhetene oppgir at de bruker språkteknologi til øversetting mellom engelsk og norsk, mens rundt 25 prosent bruker språkteknologi til øversetting mellom norsk og andre språk og/eller bokmål og nynorsk. Resultatene samsvarer med funn fra den tidligere undersøkelsen, som fant at søkefunksjoner og automatiske øversettelsesverktøy var hyppigst brukt.

På tvers av de ulike løsningene synes det å være relativt mange virksomheter som ønsker en forbedring av teknologien. Dette synes særlig å gjelde for teknologier som er relativt lite brukt. For eksempel ser vi at for teknologien «automatisk arkivering» oppgir mer enn halvparten av de som bruker teknologien, at de også ønsker forbedring (se Vedlegg 2 for en nærmere beskrivelse).

Fremtidig bruk språkteknologi i offentlige virksomheter

Teknologi virksomhetene planlegger/ønsker å ta i bruk



Av de 188 virksomhetene som har besvart undersøkelsen oppgir 130 virksomheter (69 prosent) at de ønsker eller planlegger å ta i bruk én eller flere nye løsninger i fremtiden.

På spørsmål om hvilken type språkteknologi virksomhetene planlegger eller ønsker å ta i bruk synes det å være vanligst at virksomhetene ønsker å bruke språkteknologi til å automatisere ulike prosesser i virksomheten. Rett over 40 prosent svarer at de ønsker å bruke automatisk referatskriving, mens rundt 35 prosent ønsker å bruke teknologi til automatisk arkivering, teksting/transkribering av video, produksjon av tekstsammendrag.

Til sammenligning oppga 7 av 10 virksomheter som besvarte undersøkelsen i 2019/2020 at de hadde behov for nye eller flere løsninger i fremtiden. Den gangen oppga halvparten av respondentene at de hadde behov for samtalerobot/praterobot, 44 prosent oppga at de ville ha behov for tekst til tale-system, mens 39 prosent oppga at de ville ha behov for automatisk oversettelsesverktøy.

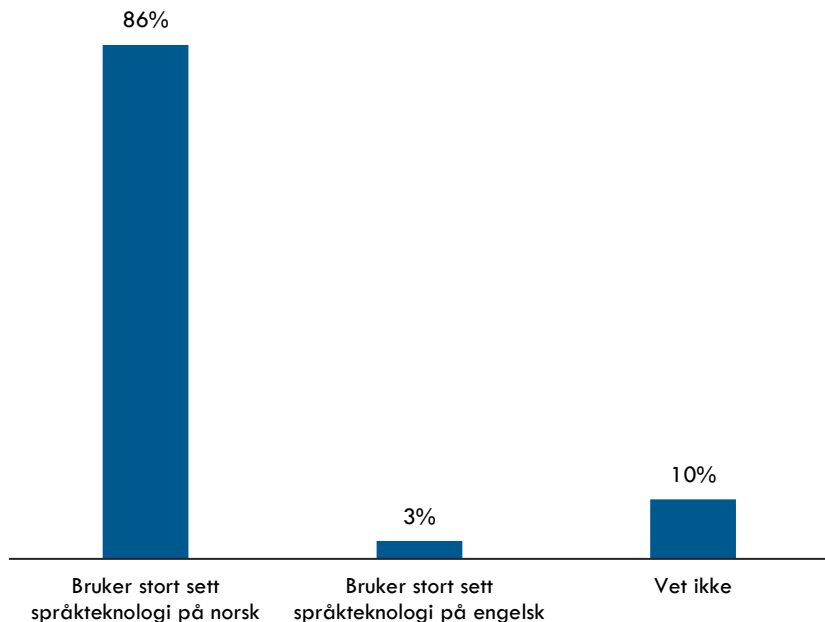
Del 1:2 Egenskaper og funksjonaliteter ved språkteknologien, samt håndtering av språkdata

I denne delen av rapporten beskriver vi egenskaper og funksjonaliteter ved språkteknologien som benyttes av de offentlige virksomhetene, samt hvordan virksomhetene håndterer språkdata som grunnlagsressurs for språkteknologi.



Språkteknologi på ulike målformer

Offentlige virksomheter om egen bruk av språkteknologi



Antall respondenter: 168

Nært 90 prosent av virksomhetene som benytter språkteknologi oppgir at de stort sett bruker språkteknologi på norsk. Kun fem virksomheter oppgir at de stort sett bruker språkteknologi på engelsk. Av disse svarer tre stykker at de skulle ønske at en større andel av den tilgjengelige teknologien kunne benyttes på norsk.

38 prosent av virksomhetene som bruker språkteknologi oppgir at virksomheten tar hensyn til støtte for både bokmål og nynorsk når den går til anskaffelse av ny teknologi, mens 38 prosent oppgir at de ikke gjør det. De resterende 24 prosentene oppgir at de ikke vet.

På spørsmål om virksomheten har støtt på tekniske begrensninger som hindrer virksomheten fra å tilby tjenester og dokumenter på både bokmål og nynorsk, oppgir 23 prosent at de har støtt på slike begrensninger, mens 39 prosent oppgir at de ikke har det. Vanlige begrensninger synes å være at dette er kostnadskrevende og/eller at leverandører ikke tilbyr denne funksjonaliteten, gjerne fordi bruk av tredjepart for oversetting ikke inkluderer nynorsk.

Årsaker til at virksomheter ikke tilbyr tjenester på flere målformer

«Teksting av direktesendt video krever manuell gjennomgang. Regelverket åpner ikke for bruk av automatiske tjenester. Manuell gjennomgang krever ca 30 min arbeid per minutt video, og tilbudte tjenester ligger på 1500,- per time, noe som dessverre gjør at direktesendt video fjernes innen 2 uker, og ingenting blir tekstet på nynorsk.»

«Det tar mye serverkapasitet å skulle ha alt innhold på ulike målformer. Den andre begrensningen er strukturell og ikke teknisk: Når kravet til nynorsk innhold er 25 prosent*, vil det ikke lønne seg å ha informasjon på begge målformer, for da går andelen ned hvis vi også produserer innholdet på bokmål.»

«Mange systemer er utelukkende på bokmål, og finst ikke på nynorsk. Det er det er mulig å gjøre med å omsette fra bokmål til norsk sjølve, men enkelte steder er det ikke råd, diverre»

«Man vil fort få dobbelt så mange maler å vedlikeholde. Dette er veldig ressurskrevende, særlig tatt i betraktning av at en stor del av brukerne våre ikke kan eller er dårlige i norsk»

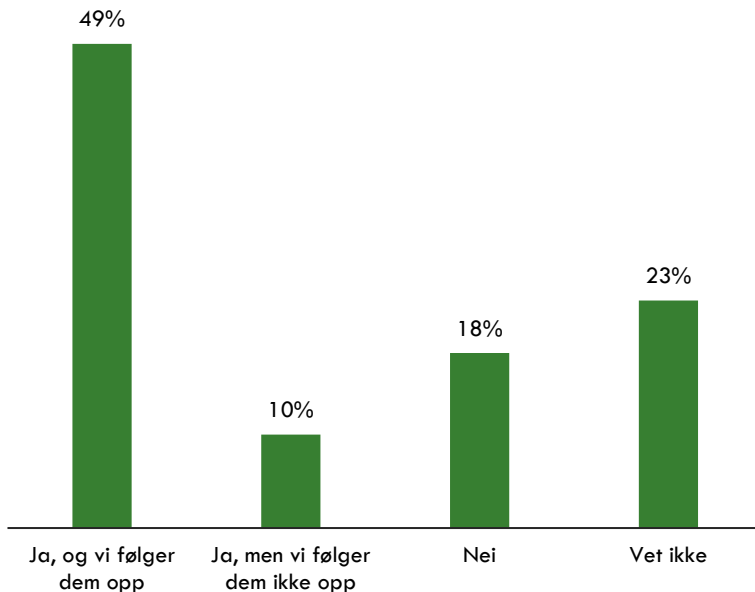
«Vi erfarer at de løsninger som vi har tatt i bruk først og fremst støtter bokmål, og at store markedsledere først og fremst fokuserer på å levere støtte for bokmål.»

«Ofte tilbyr leverandører verktøyet i ett språk. [...] Vi kan selv legge inn annet språk i artikler, men menyer m.m. er fortsatt på et språk.»

«Vår chatbot, som er levert av en tredjepart, skriver ikke nynorsk»

Språklige krav ved anskaffelser

Kjennskap til lovpålagte språklige krav ved anskaffelse og bruk av digitale verktøy i det offentlige



Antall respondenter: 168

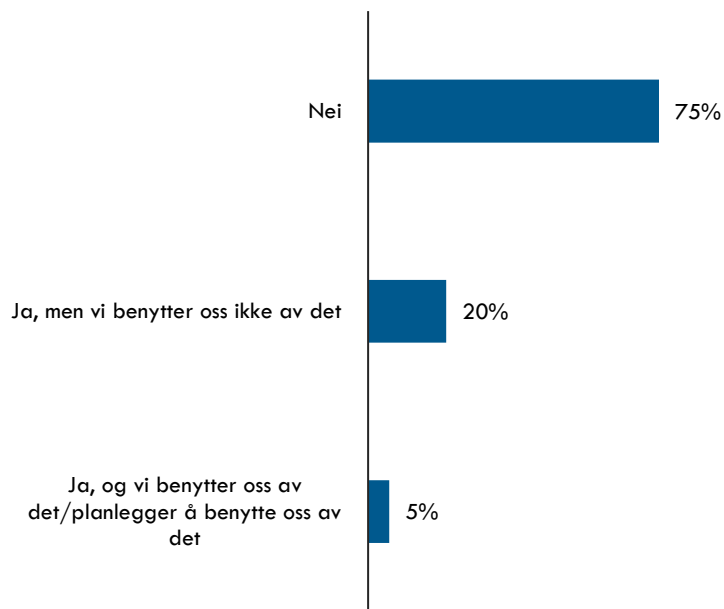
Blant virksomhetene som har besvart undersøkelsen, oppgir 60 prosent at de kjenner til lovpålagte språklige krav ved anskaffelse og bruk av digitale verktøy i det offentlige. Av disse oppgir rett over 80 prosent at de også følger opp kravene.

Til sammenligning oppga 47 prosent av virksomhetene som besvarte undersøkelsen i 2020/2019 at virksomheten var bevisst at de språkteknologiske løsningene støttet begge målformer.

Manglende kompetanse blant ansatte som gjør innkjøp og/eller manglende tilgang på løsninger som oppfyller kravene synes å være vanlige årsaker til virksomheten ikke følger opp lovpålagte språklige krav ved anskaffelse og bruk av digitale verktøy. Én virksomhet oppgir blant at anskaffelsene som regel gjøres av personer som ikke kjenner til regelverket, mens en annen oppgir at virksomheten har behov for å anskaffe «spesialprogrammer» hvor det kan være tilfellet at det ikke finnes programmer som oppfyller kravene tilgjengelig på markedet.

Virksomhetenes kjennskap til Språkrådets tjenester

Om virksomhetene er kjent med at Språkrådet kan tilby rådgivning knyttet til anskaffelse og bruk av språkteknologi



Antall respondenter: 185

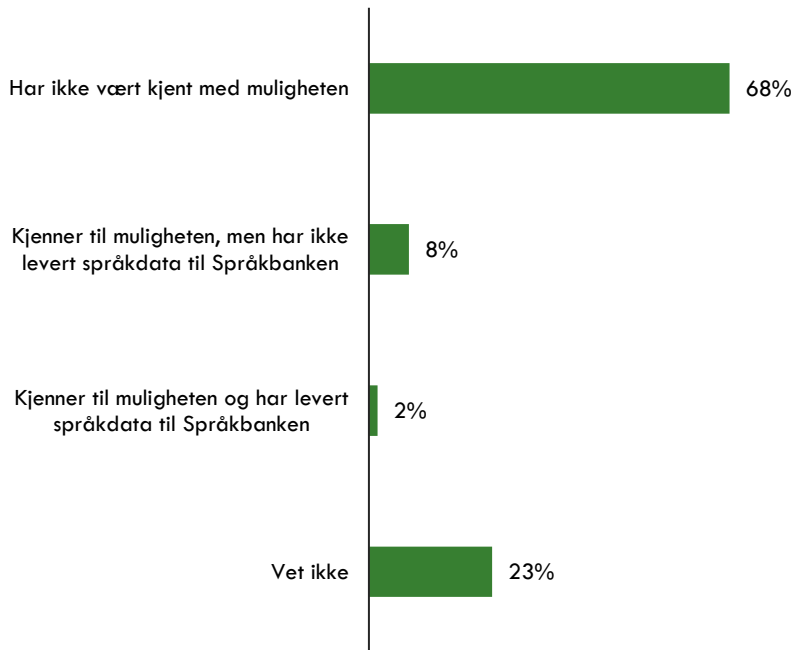
25 prosent av virksomhetene oppgir at de er kjent med at Språkrådet kan tilby rådgivning knyttet til anskaffelse og bruk av språkteknologi. Av disse oppgir kun 20 prosent at de også har benyttet eller planlegger å benytte seg av disse tjenestene.

Vanlige årsaker til at virksomhetene ikke har benytter seg av Språkrådets tjenester, til tross for at de kjenner til disse, synes å være at de selv ikke opplever å ha behov for bistand.

På lik linje med denne undersøkelsen, viser undersøkelsen som ble gjennomført i 2019/2020 at det var 5 prosent av virksomhetene hadde vært i kontakt med Språkrådet for råd om anskaffelse og bruk av språkteknologiske løsninger. Resultatene er ikke nødvendigvis direkte sammenlignbare, men funnene tyder på at kjennskapen til og praksisen for å benytte seg av Språkrådet sine tjenester ikke har blitt mer utbredt de siste årene. Dette til tross for at den tidligere undersøkelsen også fant at nesten halvparten av virksomhetene hadde et udekket behov for råd om språkteknologiske løsninger.

Virksomhetenes kjennskap til Språkbanken

Virksomhetenes kjennskap til muligheten for å levere fra seg språkdata til Språkbanken



Antall respondenter: 186

Kun 10 prosent av virksomhetene oppgir at de er kjent med muligheten for å levere språkdata til Språkbanken ved Nasjonalbiblioteket. Av disse oppgir 20 prosent at de også *har* levert språkdata til Språkbanken.

Vanlige årsaker til at virksomheter ikke leverer data til Språkbanken, til tross for at de kjenner til muligheten, fremstår å være at data ikke kan leveres inn av sikkerhets- eller personvern hensyn eller at virksomheten ikke anser det som relevant å levere inn data.

I undersøkelsen som ble gjennomført i 2019/2020 oppga 9 prosent av virksomhetene at de kjente til ressursene for utvikling av språkteknologiske løsninger i Språkbanken, mens 10 prosent oppga at de var helt eller delvis enige i at virksomheten sørget for at språkteknologiske ressurser som ble utviklet ble gjort tilgjengelig for andre virksomheter, for eksempel via Språkbanken. Til tross for at resultatene fra de to undersøkelsene ikke er direkte sammenlignbare, tyder funnene på kjennskapen til, og praksisen for å dele data med Språkbanken ikke har blitt mer utbredt de siste årene.

Del 2: Ressursbehov til utvikling av språkteknologi

Denne delen sammenstiller funn fra en undersøkelse om ressursbehov til utvikling av språkteknologi.



Datagrunnlag og metode

På oppdrag for Språkrådet har Oslo Economics gjennomført en spørreundersøkelse rettet mot virksomheter som driver forskning på, og/eller utvikling av språkteknologi. Formålet med denne undersøkelsen har vært å identifisere ressursbehovet for å drive forskning på, og/eller utvikling av språkteknologi.

Språkteknologi handler om prosessering av språk og omfatter alle verktøy som involverer automatisert behandling av språk, også allmenne verktøy som stavekontroller og skriveprogrammer med videre. Den gjennomførte undersøkelsen er utformet av Språkrådet. Oslo Economics har gitt innspill til utforming og oppbygging av undersøkelsen.

Rekruttering av respondenter til undersøkelsen ble gjennomført ved å ta utgangspunkt i en liste utarbeidet av Språkrådet, over virksomheter som ble ansett som relevante for undersøkelsen. Listen ble supplert med forslag fra Oslo Economics. Hver enkelt virksomhet ble kontakten i forkant av undersøkelsen, for å identifisere relevante mottakere av undersøkelsen.

Undersøkelsen ble sendt til 110 virksomheter. Den ble sendt ut i starten av desember 2023, og det ble sendt ut en påminnelser i uke 50. Undersøkelsen ble avsluttet ved årsskiftet og den ble besvart av totalt 36 virksomheter.

Undersøkelsen blant forskere og utviklere oppnådde høyere svarprosent, sammenlignet med undersøkelsen blant offentlige virksomheter. Dette skyldes trolig flere forhold, blant annet: 1) undersøkelsen ble sendt direkte til relevante respondenter, 2) undersøkelsen kan fremstå som mer relevant for målgruppen, ettersom språkteknologi er kjernen av deres virksomhet, og 3) offentlige virksomheter mottar et større antall henvendelser av lignende karakter.

Undersøkelsen var anonym og det ble ikke innhentet noen bakgrunnsinformasjon om respondentene. Av denne årsaken har det ikke vært mulig å sammenligne kjennetegn ved respondentene med den totale populasjonen. Det er derfor mulige skjevheter i informasjonsgrunnlaget, og resultatene følgelig tolkes med varsomhet.

Oppsummering av hovedfunn

Blant respondentene i undersøkelsen er 80 prosent virksomheter som arbeider med utvikling av språkteknologi til kommersiell bruk, mens de øvrige arbeider med utvikling av språkteknologi i forskningsøyemed.

De fleste virksomhetene jobber med å utvikle mellom 1 og 4 ulike språkteknologiske løsninger, hvorav de fleste utvikler løsninger for tekstklassifisering og søkefunksjoner, stavekontroll og skriveprogrammer.

De fleste virksomheter som benytter grunnlagsressurser i utviklingen av språkteknologiske løsninger samler inn disse dataene selv, for eksempel gjennom webscraping, intervjuer eller undersøkelser, men det er også en relativt stor andel som henter ressursene fra Språkbanken.

Kun én tredjedel av virksomhetene som er kjent med muligheten for å dele data med Språkbanken, deler data. Kommersielle interesser synes å være den vanligste årsaken til at virksomheter ikke deler data.

Virksomhetene synes å ha behov for flere ulike typer grunnlagsdata de kommende årene, hvorav de fleste oppgir at de vil ha behov for tekstressurser, språkmodeller og/eller andre verktøy.

Rett over halvparten av virksomhetene (12 av 22) oppgir at de deltar i offentlige anbudsprosesser. Åtte av disse oppgir at det hender at det blir utviklet egne språkdata i forbindelse med slike oppdrag, men kun én virksomhet deler data videre med Språkbanken eller andre.



Del 2:1 Om virksomhetene og virksomhetenes aktiviteter

I denne delen av rapporten beskriver vi virksomhetenes arbeid med utvikling av språkteknologiske løsninger.



Arbeid med utvikling av språkteknologiske løsninger

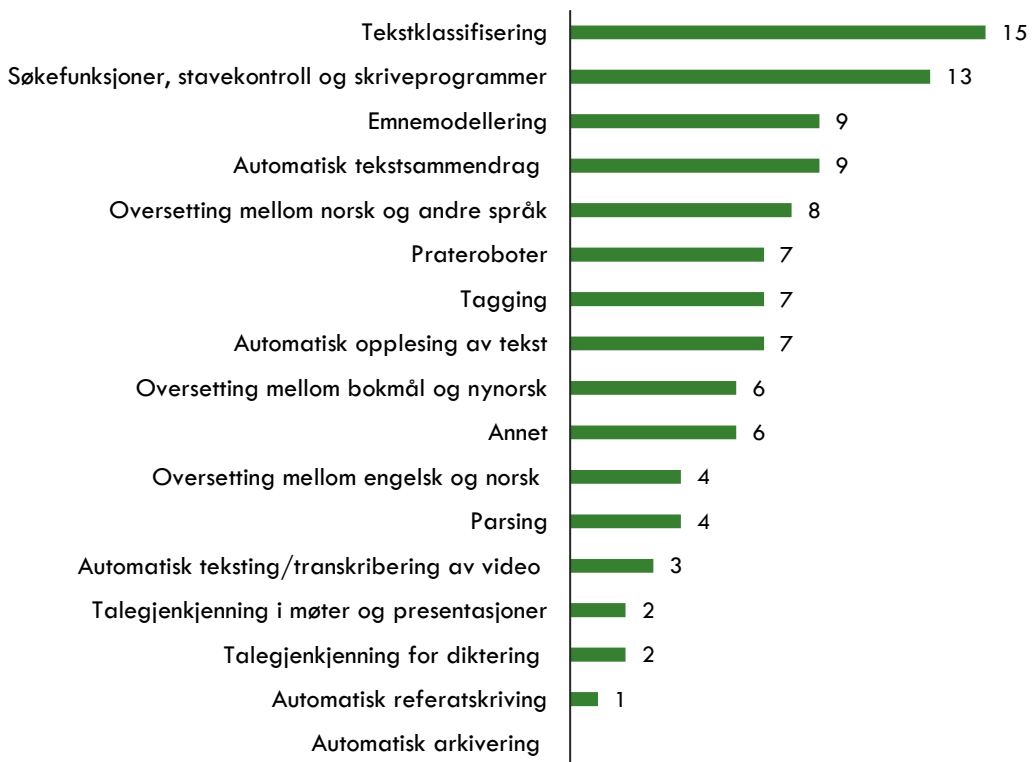
Undersøkelsen ble besvart av 36 virksomheter. Blant disse virksomhetene var det 27 som oppga at de jobbet med utvikling av språkteknologiske løsninger eller andre produkter som involverer språkteknologiske løsninger på norsk. De resterende virksomhetene er ikke ansett som relevante for undersøkelsen og inngår derfor ikke i videre analyser.

Blant virksomhetene som er i målgruppen for undersøkelsen, arbeider de fleste (om lag 80 prosent) med utvikling av språkteknologiske løsninger til kommersiell bruk, mens om lag 20 prosent av disse arbeider med utvikling av språkteknologiske løsninger i forskningsøyemed.



Arbeid med ulike typer språkteknologier

Virksomhetenes arbeid med ulike språkteknologiske løsninger



Antall respondenter: 27

De fleste virksomhetene som utvikler språkteknologiske løsninger jobber med tekstklassifisering og søkefunksjoner, stavekontroll og skriveprogrammer. Det er imidlertid også en god del som jobber med egenmodellering, automatisk tekstsammendrag og oversetting mellom norsk og andre språk.

De fleste virksomhetene jobber med å utvikle mellom 1 og 4 ulike språkteknologiske løsninger.

Av virksomhetene som har oppgitt at de arbeider med språkteknologiske løsninger for oversetting mellom norsk og andre språk, oppgir de fleste at de jobber med oversetting mellom norsk og andre nordiske språk, som svensk, dansk, finsk og samisk. Enkelte arbeider også med oversetting til arabisk, russisk og samisk.

Del 2:2 Bruk av grunnlagsressurser og språkdata

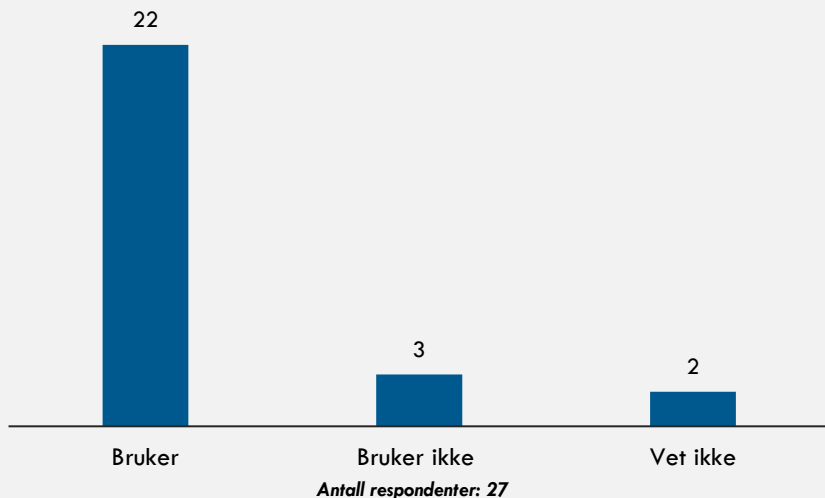
I denne delen av rapporten beskriver vi virksomhetenes bruk av grunnlagsressurser i forsknings- og utviklingsarbeid innen språkteknologiske løsninger. Med grunnlagsressurser menes språklige ressurser som kan brukes for å utvikle språkteknologiske løsninger.



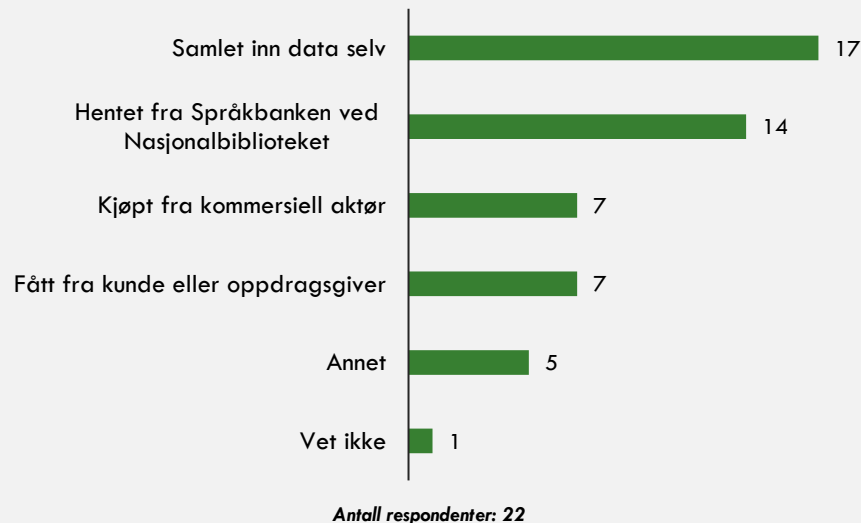
Bruk av grunnlagsressurser

Blant de 27 virksomhetene som arbeider med utvikling av språkteknologiske løsninger, finner vi at 22 virksomheter bruker eller har brukt grunnlagsressurser. Av disse oppgir 17 stykker at de samler inn data selv, for eksempel gjennom webscraping, intervjuer eller undersøkelser, men det er også flere (14 stykker) som oppgir at de har hentet data fra Språkbanken ved Nasjonalbiblioteket. 13 av virksomhetene som har hentet data fra Språkbanken, har også hentet data selv.

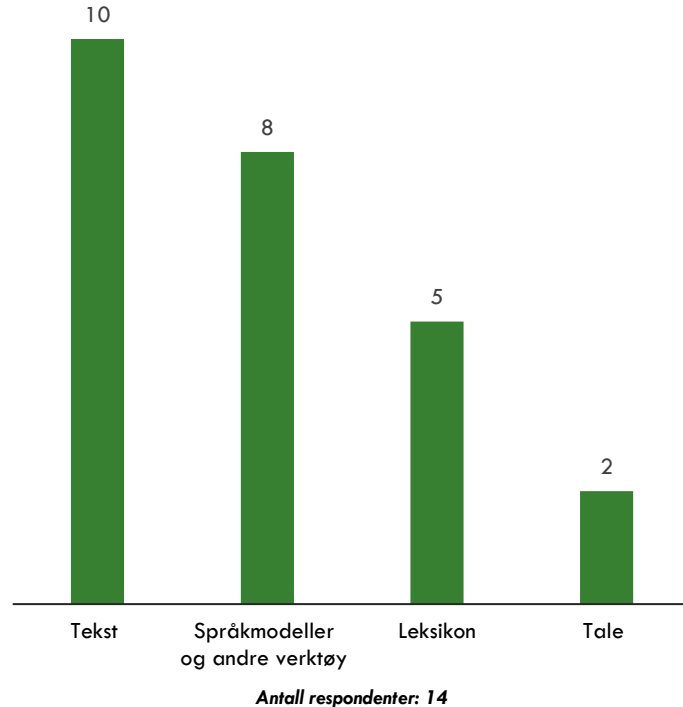
Virksomhetenes bruk av grunnlagsressurser



Kilder til grunnlagsressurser



Grunnlagsressurser fra Språkbanken



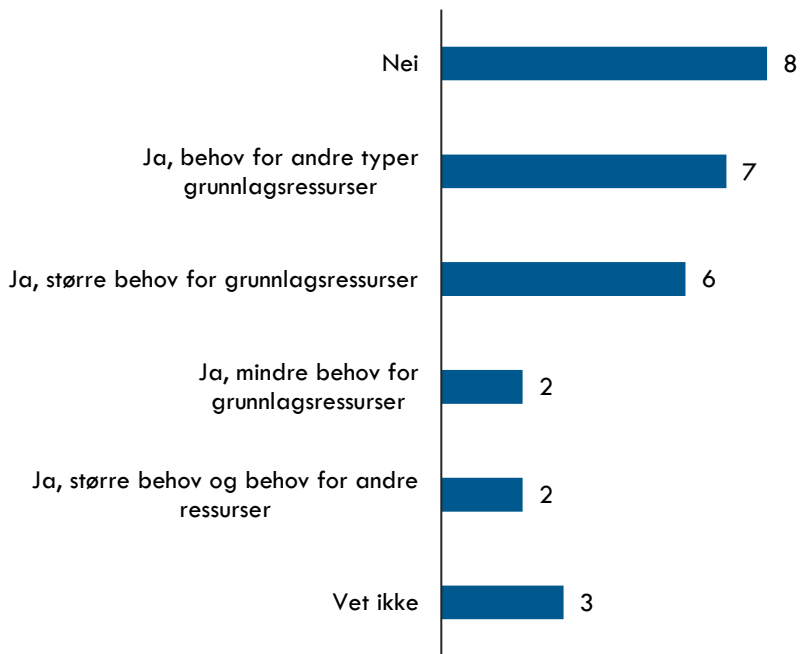
Av virksomhetene som har hentet grunnlagsressurser fra Språkbanken (14), svarer 10 av disse at de har benyttet seg av grunnlagsressurser i form av tekst, mens 8 stykker oppgir at de har benyttet språk-modeller og andre verktøy. Et mindre antall oppgir at de har benyttet seg av leksikon eller taleressurser.

Av virksomheten som ikke har benyttet seg av grunnlagsressurser fra Språkbanken, oppgir flertallet (9 stykker) at årsaken til dette var at Språkbanken ikke tilbød grunnlagsressurser tilpasset virksomhetens behov. Det oppgis blant annet at ressurser egnet til å løse klassifiseringsoppgaver, datasett for sentimentanalyse, samt ressurser på nynorsk var ressurser virksomhetene ikke fant tilgjengelig i Språkbanken.

Kun én virksomhet oppga at årsaken for at språkbanken ikke var benyttet, var at virksomheten ikke kjente til Språkbanken fra tidligere, mens en annen oppga at virksomheten ikke har kapasitet eller ressurser til å undersøke hvilke tjenester Språkbanken tilbyr.

Endret behov for grunnlagsressurser

Virksomhetenes svar på spørsmål om tilgjengeliggjøringen av større språkmodeller har endret virksomhetens behov for grunnlagsressurser



Antall respondenter: 22

Nært en tredjedel av virksomhetene oppgir at utviklingen og tilgjengeliggjøringen av større språkmodeller, som for eksempel ChatGPT eller LLaMA, har ført til behov for andre typer grunnlagsressurser og/eller et større behov for grunnlagsressurser.

Et mindretall (2 virksomheter) oppgir at utviklingen har ført til at virksomheten har fått mindre behov for grunnlagsressurser, mens om lag en tredjedel oppgir at utvikling ikke har hatt noen påvirkning på virksomhetenes behov for grunnlagsressurser.

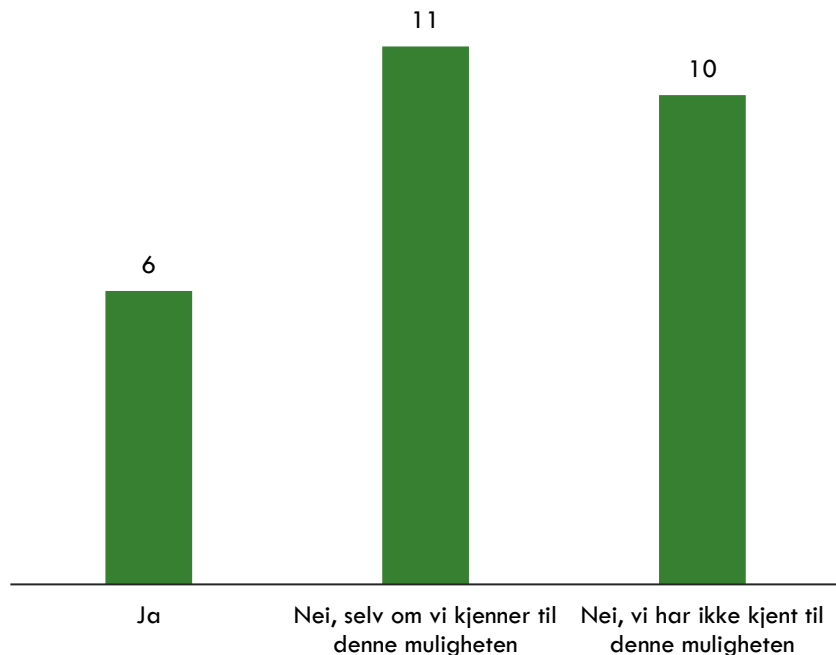
Del 2:3 Datadeling

I denne delen ser vi på virksomhetenes rutiner for å dele grunnlagsressurser med Språkbanken ved Nasjonalbiblioteket – en nasjonal infrastruktur for språkteknologi, som blant annet tilbyr datasett med norsk tekst og tale.



Deling av språkdata med nasjonalbiblioteket

Om virksomhetene deler språkdata eller annen data med Språkbanken



Antall respondenter: 27

Rett over 20 prosent av virksomhetene oppgir at de deler språkdata eller annen data med Språkbanken. Halvparten av virksomhetene som ikke deler data oppgir at de ikke har vært kjent med muligheten til å dele data. Den øvrige halvparten oppgir at de ikke deler til tross for at de er kjent med muligheten.

Enkelte virksomheter oppgir at det er kommersielle årsaker til at virksomheten ikke deler data med Språkbanken, blant annet at Språkbanken ikke gjør det mulig for virksomheten å dele ressurser som krever brukerbetaling. Enkelte oppgir også at de ikke deler data av sikkerhetsmessige hensyn.

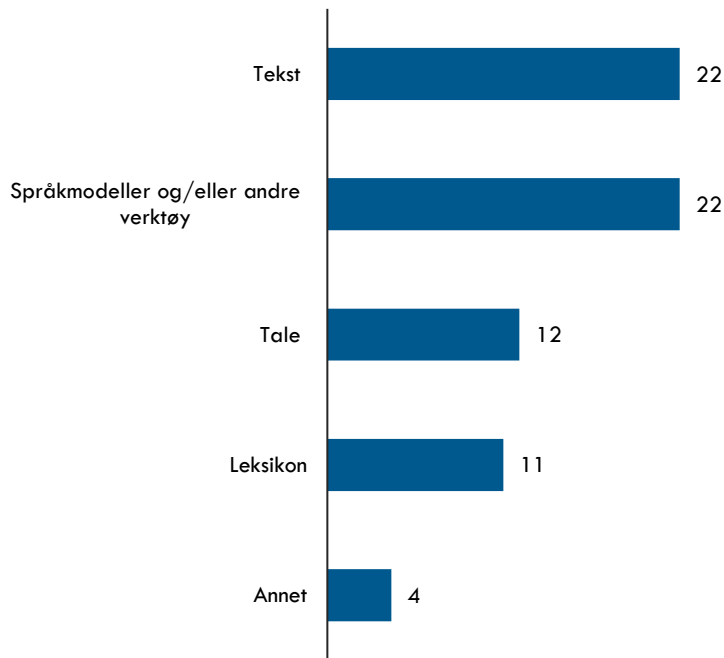
Del 2:4 Behov for grunnlagsressurser

I denne delen av undersøkelsen ser vi på virksomhetenes behov for grunnlagsressurser i de kommende årene. Med grunnlagsressurser mener vi språklige ressurser som kan brukes for å utvikle språkteknologiske løsninger.



Behov for ulike typer grunnlagsdata

Grunnlagsressurser virksomheten oppgir at de vil ha behov for de kommende årene



Antall respondenter: 27

Virksomhetene synes å ha behov for flere ulike typer grunnlagsdata de kommende årene. De fleste (22 av 27) oppgir at de vil ha behov for tekstressurser, språkmodeller og/eller andre verktøy. Det er imidlertid også relativt mange (rundt 45 prosent) som har oppgitt at de vil ha behov for enten taleressurser og/eller leksikon.

Av virksomheten som oppgir at de vil ha behov for taledata, oppgir halvparten at de vil ha behov for opptak med verbatim transkripsjon (generelt) mens fire virksomheter oppgir at de vil ha behov for opptak med verbatim transkripsjon (spesielle domener). Kun et fåtall virksomheter oppgir å ha behov for opptak med ikke-verbatim transkripsjon, enten generelt eller spesielle domener. Det er ingen virksomheter som oppgir at de har behov for opptak uten transkripsjon.

Av de virksomhetene som oppgir at de vil ha behov for leksikon, oppgir 10 stykker at de vil ha behov for termlister/termbaser/fagordlister, mens henholdsvis 8, 7 og 6 stykker oppgir at de vil ha behov for ordlister, flerspråklige ordlister og/eller lister med skrivefeil. Kun én virksomhet oppgir behov for uttaleleksikon.

Behov for tekst

Av virksomheten som oppgir at de vil ha behov for tekst, oppgir rett under tre fjerdedeler at de har behov for enspråklig tekst, bokmål. Videre oppgir rett under 60 prosent at de har behov for enspråklig tekst, enten nynorsk eller spesifikt domene.

På spørsmål om hvilke spesifikke domener virksomhetene har behov for tekst innenfor, oppgis blant annet medisin, forsvar, forsikring, finans, juss, elektrofag, nyheter og medieinnhold.

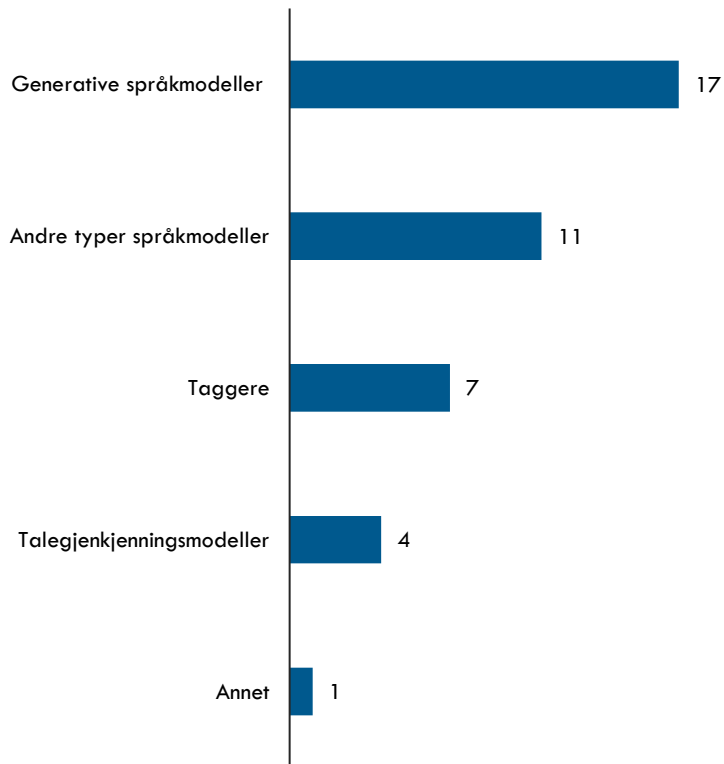
På spørsmål om hvilken type annotasjon virksomheten har bruk for oppgis blant annet stavekontroll, annotasjon med entiteter og hendelser, spørsmål-svar-datasett og lignende, datasett for genereringsbaserte oppgaver, instruksjonsdatasett og sammendragsdatasett, samt ressurser som følger notasjon i OBt og CG.

Behov for tekst



Antall respondenter: 22

Behov for språkmodeller



Antall respondenter: 22

Behov for språkmodeller

Av virksomheten som oppgir at de vil ha behov for språkmodeller oppgir rett over tre fjerdedeler at de har behov for generative språkmodeller, mens halvparten oppgir at det har behov for andre typer språkmodeller. På spørsmål om hvilke typer språkmodeller disse virksomhetene har behov for nevnes blant annet:

- modeller som kan brukes til presis klassifisering og generering av for eksempel hendelsesgrafer og andre typer grafer fra språklig innhold
- modeller som er trent spesifikt for å kjenne igjen og klassifisere tekst

Om lag én tredjedel av virksomhetene oppgir at de har behov for taggere, mens kun 4 av 22 virksomheter oppgir at de har behov for talegjenkjenningsmodeller.

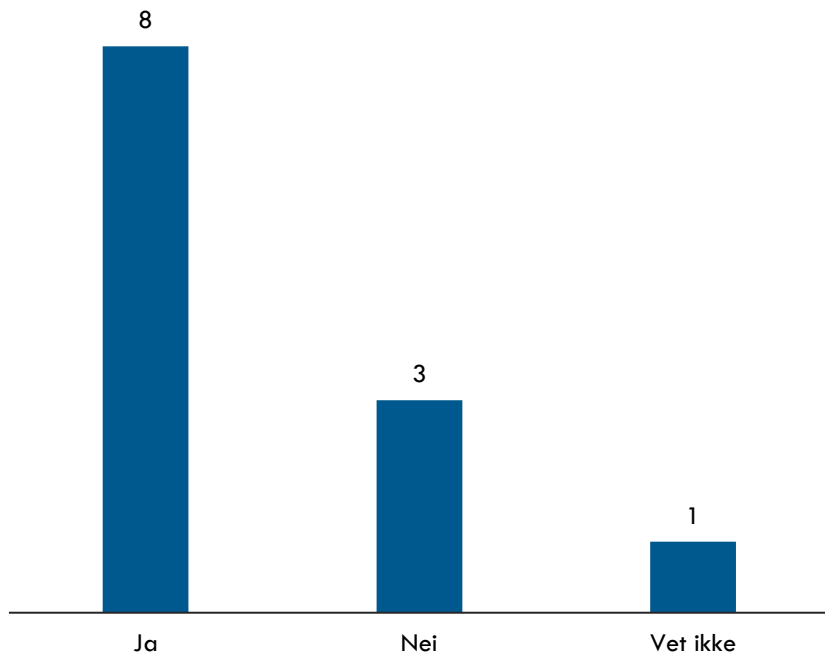
Del 2:5 Offentlige anbud

I denne delen av rapporten presenterer vi virksomhetene sine svar på spørsmål om de deltar i offentlige anbudprosesser, hvordan virksomhetene forholder seg til språkløven i anbudsprosessene, samt hvilke kriterier angående språkdata som stilles ovenfor oppdragsgiver i anbudsprosessene.



Utvikling av språkdata i oppdrag for det offentlige

Blir det utviklet egne språkdata til bruk i produktet i forbindelse med oppdrag for det offentlige?



Antall respondenter: 12

Av virksomhetene som deltar i anbudskonkurranser (12 av 22), oppgir kun tre virksomheter at det er mindre attraktivt å svare på anbud hvor produktet skal leveres både på nynorsk og bokmål.

8 av 12 virksomheter oppgir at det hender at det blir utviklet egne språkdata til bruk i produktet i forbindelse med oppdrag for det offentlige, men kun én av disse virksomhetene oppgir at denne dataen deles videre med Språkbanken eller andre. Virksomhetene begrunner dette med at de ikke selv eier dataen og/eller at denne ofte ikke kan deles videre på grunn av konfidensialitet eller av konkurransemessige hensyn.

oslo**economics**

www.osloeconomics.no

Vedlegg 1 – Offentlige virksomheter

Dersom ett av følgende kriterier er oppfylt, er virksomheten regnet som offentlig:

Organisasjonsform:

- Fylkeskommune
- Fylkeskommunalt foretak
- Interkommunalt selskap
- Kommunalt foretak
- Kommune
- Annet foretak ifølge særskilt lov
- Statsforetak
- Staten

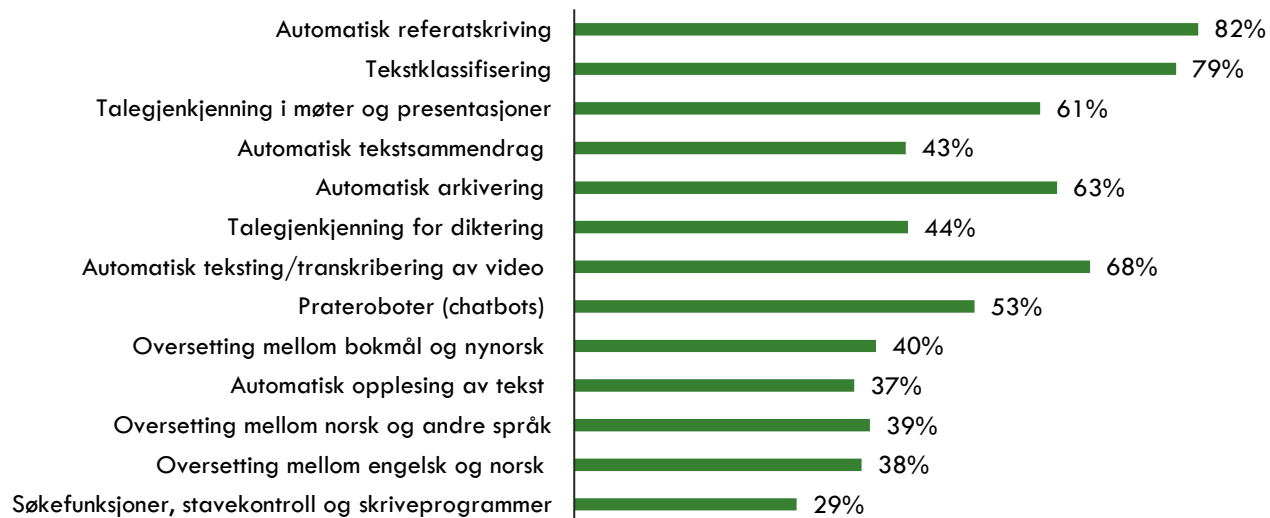
Institusjonell sektorkode:

- 110: States forretningsdrift
- 1120: Statlig eide selskaper
- 1510: Kommunale foretak med ubegrenset ansvar
- 1520: Kommunalt eide selskaper
- 3100: Norges Bank
- 3900: Statlige låneinstitusjoner
- 6100: Statsforvaltningen
- 6500: Kommuneforvaltningen

Den Norske Kirke er ikke regnet som en offentlig virksomhet, og er derfor tatt ut

Vedlegg 2 – Forbedringsbehov

Prosentandel av respondentene (offentlige virksomheter) som ønsker en forbedring av teknologien de bruker



Antall respondenter: 169, 19 respondenter svarte ikke på spørsmålet

Listen over teknologier er sortert fra lav til høy bruk. Automatisk referatskriving er for eksempel den teknologitypen som færrest virksomheter oppgir at de bruker.

På tvers av de ulike løsningene synes det å være relativt mange virksomheter som ønsker en forbedring av teknologien. Dette synes særlig å gjelde for teknologier som er relativt lite brukt, og kan trolig være en av forklaringene på hvorfor bruken av disse teknologiene ikke er mer utbredt.