

# GENERATIV KUNSTIG INTELLIGENS -

## OPPHAVSRETTE OG DE REDAKTØRSTYRTE MEDIENE

### EN KARTLEGGING AV DAGENS SITUASJON OG FREMTIDSSCENARIER



17. SEPTEMBER  
2024

## Bakgrunn og innledning

Mediebedriftenes Landsforening, Norsk Journalistlag, Norsk Redaktørforening og Fagpressen har fått midler fra Kopinor til et prosjekt om generativ kunstig intelligens og hvordan dette påvirker opphavsretten og de redaktørstyrte mediene.

Prosjektet hadde følgende mål:

- Danne et felles kunnskapsgrunnlag for de redaktørstyrte journalistiske mediene om hvordan opphavsretten kan forstås og sikres i møte med generativ kunstig intelligens.
- Stimulere til debatt i bransjen og i de enkelte medier om hvilke valg mediene må og kan ta i møte med denne teknologien og hvilke konsekvenser disse vil ha.
- Identifisere behov for politiske tiltak.

Målene skal oppnås gjennom utarbeidelse og spredning av en rapport som 1) Kartlegger dagens situasjon og 2) Beskriver fremtidsscenarioer og belyser veivalg.

Parallelt med vårt prosjekt gjennomføres Kick-prosjektet til Kopinor. Vi har valgt å tilpasse innholdet i vårt prosjekt i forhold til de aktivitetene som har foregått i Kick-prosjektet, slik at det totale utbyttet av prosjektene skal bli størst mulig.

Prosjektet har hatt en styringsgruppe bestående av Einar Ryvarden (Fagpressen), Solveig Husøy (Norske redaktørforening), August Ringvold (NJ) og Lars Raaum (MBL). Geir E Engen (MBL) har koordinert arbeidet.

## Rapportens oppbygging

Rapporten fra prosjektet er tredelt. Det er tre selvstendige deler, som til sammen gir et bredt diskusjonsgrunnlag.

Første del er en oversikt over status på de ulike datasettene som brukes til trening av språkmodeller og hvilket materiale som inngår i dette. Denne delen er skrevet av Svein Arne Brygfjeld ved Nasjonalbiblioteket (<https://ai.nb.no/about-us/>).

I del to ser vi på de mulige langsiktige konsekvensene av KI for mediehusene. Dette gjøres gjennom et notat som ser på ulike scenarier og beskrivelser som har vært publisert, diskutert og analysert av forskere, KI-eksperter, bransjeeksperter og mediebransjen selv i løpet av de siste månedene. Notatet er skrevet av Frank Gander.

Siste del undersøker mengde innhold fra medlemmer av MBL og Fagpressen som inngår i de datasettene som er brukt til trening av modellene til Open AI/ChatGPT. Her inngår også en analyse av muligheten for direkte gjengivelse av innholdet som inngår i treningen. Arbeidet er gjort av Peng Liu som arbeider som forsker ved Institutt for datateknologi og informatikk ved NTNU og er tilknyttet NorwAI (<https://www.ntnu.no/ansatte/penl>)

# Nasjonalbiblioteket, store språkmodeller, data og opphavsrett

*Nasjonalbiblioteket/SAB (Svein Arne Brygfjeld)*

## Bakgrunn

Store språkmodeller har gjennom introduksjonen av tjenester som BARD fra Google og ChatGPT fra OpenAI blitt gjenstand for stor oppmerksomhet i samfunnet generelt. Forståelsen av kraften og relevansen i denne type teknologi har økt betydelig.

Ett av områdene som har blitt gjenstand for diskusjon, er rettigheter knytta til innholdet som brukes til å bygge språkmodellene.

Nasjonalbiblioteket (NB) har erfaring med denne type teknologi, både teoretisk og praktisk. Vi har trent flere slike modeller, og vi har etablert et eget korpus av tekst for trening av slike modeller.

Dette dokumentet beskriver NBs perspektiv på dagens bruk av innhold for trening av språkmodeller med tanke på opphavsrett for innholdet.

## Kort gjennomgang av begreper og rammer

### Store språkmodeller

I det offentlige ordskiftet om KI handler mye av diskusjonen om store språkmodeller. Med begrepet språkmodeller forstås i dette dokumentet programvare som basert på statistiske metoder og nevralt nettverk kan lære språk gjennom å trenes på eksisterende innhold. Både i samtalen om emnet i samfunnet, og i dette dokumentet, skiller vi mellom generative og ikke-generative modeller. De generative modellene kan produsere nytt innhold, mens de ikke-generative modellene er begrenset til å forstå oppbygginga av setninger. Ordet “stor” brukes ofte sammen med språkmodell, og det henspiller både på hvor kompleks modellen er (antall parametre) og volumet på treningsdata som er brukt for å trene modellen.

### Treningsdata

Moderne språkmodeller, som PaLM (Google), GPT-3/4 (OpenAI), LLaMA (Meta) og Mistral (Mistral AI), trenes på tekst som finnes fra før. Dette kan f.eks. være innhold som er høstet fra Internett med roboter eller tekster fra bøker og aviser fra digitale samlinger. Utvalget og utforminga av de data som brukes for trening av modellene har avgjørende betydning for evnen modellene får til å utføre ulike oppgaver.

For å trene store språkmodeller trengs det store mengder data. En billedliggjøring av størrelsen kan være at Encyclopædia Britannica har ca 40 millioner ord, mens Nasjonalbibliotekets åpne korpus (Norwegian Colossal Corpus - NCC) for trening av språkmodeller inneholder 6900 millioner ord. Bokdelen alene i dette korpuset inneholder 840 millioner ord.

I tillegg til volumet legges det spesielt vekt på kvalitet, representativitet og diversitet når man etablerer treningssett for språkmodeller.

Med kvalitet menes i første rekke at språket er korrekt og med lite skrivefeil (f eks fra OCR-behandling). I arbeidet med NCC har mye innhold blitt holdt utenfor på grunn av feil fra OCR-behandlingen i digitaliseringsprosessen. Videre ønsker man også at teksten har hele/naturlige enheter (som artikler), og at lengden på tekstene er relevant for den type trening man vil utføre.

Representativitet vil her si at utvalget av tekster bør speile konteksten der modellen er tenkt brukt på en god måte. Man må f eks ha tilstrekkelig mengde norsk bokmål og nynorsk for en modell som skal brukes i norske forhold, og for en modell som skal fungere godt innenfor fagfeltet jus må modellen trenes på tilstrekkelig mengde juridiske data.

Endelig bør modellene trenes på et vidt spekter av innhold, fra hele den vanlige bruken av språket. I noen grad bør dette balanseres i volum om modellene skal ha generelle gode egenskaper.

## Data fra Internett

En utbredt måte å bygge treningsdata på er å høste data fra Internett. Dette er en effektiv måte å skaffe store mengder data. Etter hvert er det også etablert flere åpne samlinger med slike data, som [Common Crawl/MC4](#) og [HPLT](#). Slike samlinger gjør det enkelt å komme i gang med trening av språkmodeller.

Baksiden av denne medaljen er at det er utfordrende å håndtere både kvalitet, representativitet og diversitet i denne type data. De opptrer i treningssammenheng mer som en “stor grå masse” der de som trener modellene ikke har et presist bilde av profilen på innholdet.

## Data fra det offentlige

I Norge er det vanlig at offentlige dokumenter gjøres tilgjengelig for offentligheten digitalt. Det er vanlig at slike dokumenter gjøres tilgjengelig digitalt, f eks som PDF. De representerer også et innhold med en viss egenart, både språklig og i form av innhold. Derfor er de også attraktive i trening av store språkmodeller.

## Sosiale media

Sosiale media, som f eks X (tidligere Twitter), Facebook og ulike debattforum kan være relevante språklige ressurser for språkmodeller. Innholdet representerer gjerne en mer muntlig og uformell form som har relativt liten plass i andre samlinger av data.

## Aviser, bøker og tidsskrifter

De vanlige skriftlige kildene som tekstbaserte er i hovedsak aviser, bøker og tidsskrifter. Brorparten av avis- og bokutgivelser på norsk er digitalisert av NB, sammen med mange andre tekstressurser. I tillegg praktiserer NB digital pliktavlevering, noe som fører til en kontinuerlig digital tilvekst.

Samlet sett er disse tekstressursene attraktive i trening av språkmodeller. De representerer en variert bruk av norsk bokmål og nynorsk, de beskriver samfunnet i Neirge før og nå, og volumet er relativt stort.

Noen av de digitaliserte ressursene har dog en kvalitet i digitaliseringen (OCR-prosesseringen) som gjør at de er dårlig egnet, men normalt holder disse mål også i kvalitet.

## Opphavsrettslige rammer

Norsk opphavsrett hjemler også eksemplarframstilling for forskning, relevant for å bygge korpus av treningsdata og trening av maskinlæringsmodeller.

I forskrift til åndsverkloven § 4 er det to modeller for formatfri eksemplarframstilling for forskningsformål. Bestemmelsen hjemler også digitalisering og å fremstille eksemplar i et annet format enn originaleksemplaret.

Bestemmelsens første ledd gir undervisnings- og forskningsinstitusjoner tillatelse til å fremstille eksemplar av åndsverk for forskningsformål med tillatelse fra Kultur- og likestillingsdepartementet. Bestemmelsen forutsetter derfor en forutgående søknadsprosess og godkjenning.

I en egen løsning gis Nasjonalbiblioteket i andre ledd spesifikk hjemmel til å fremstille eksemplar av åndsverk i sine samlinger som grunnlag for språklige korpuser. Bruken må være for forskningsformål. Denne hjemmelen omfatter også pliktavlevert materiale. I motsetning til den generelle hjemmelen krever ikke Nasjonalbibliotekets hjemmel noen forutgående godkjenning.

Nasjonalbibliotekets hjemmel retter seg mot språklige korpus og den språklige strukturen i tekst- og talemateriale. Den retter seg ikke mot det intellektuelle innholdet. Det er flere ganger fremhevet av lovgiver i forarbeider at bestemmelsen ikke anses å utfordre rettighetshavernes opphavsrettslige interesser. Begrensningen til forskningsformål avgrenser rekkevidden til bestemmelsen. Nasjonalbiblioteket kan ikke stille korpus til rådighet for utviklingsformål etter bestemmelsen der utviklingen skjer i regi av private parter eller offentlige virksomheter som ikke er forskningsinstitusjoner. På strateginivå nevnes både det å omarbeide materialet slik at deling av materiale ikke kommer i konflikt med opphavsretten og at Nasjonalbiblioteket selv trener maskinlæringsmodeller som løsninger på at Nasjonalbibliotekets samling også kan stilles tilgjengelig for utviklingsformål.

## Treningsdata fra NB i bruk i dag og opphavsrett

Beskrivelsen under tar for seg de delene av treningsdata for store språkmodeller som kommer fra NBs samling. Selv om NB har en stor samling av data som er høstet fra Internett, brukes ikke disse i trening av språkmodeller. Det samme gjelder data fra sosiale media.

Videre tar beskrivelsen utgangspunkt i samlingen som NB kaller “the Norwegian Colossal Corpus”, eller NCC. NCC finnes i to varianter, den åpne versjonen som deles fritt og uten begrensninger, og NCC+ som inneholder innhold under opphavsrett og som normalt ikke deles. Dagens versjon av NCC er fra tidlig 2021, og en god del av materialet er utelatt på grunn av feil i OCR. De absolutte tallene vil dermed endre seg betydelig når NCC revideres i løpet av 2024.

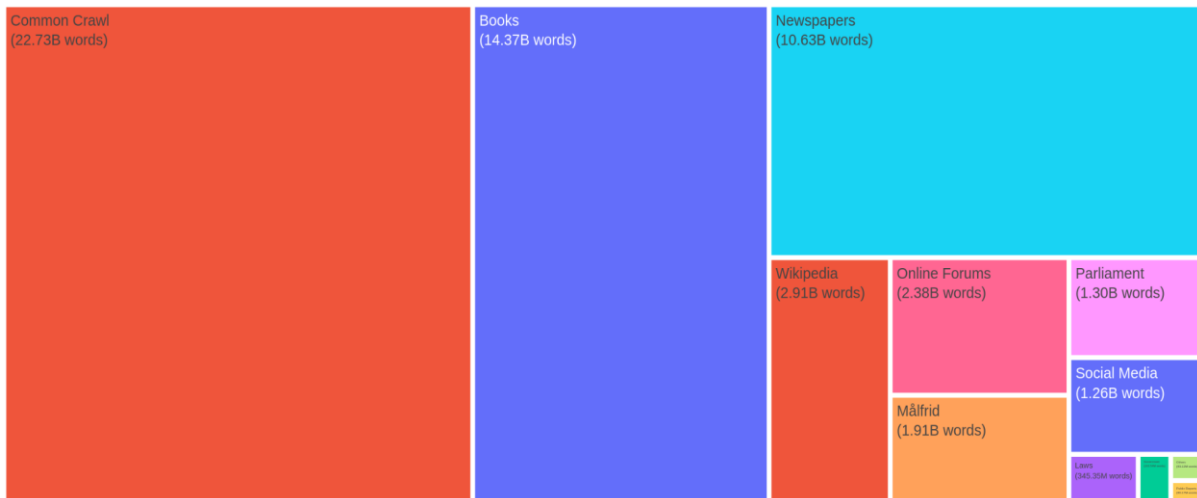
## Størrelser

Figur 1 og figur 2 viser den relative fordelinga av data i hhv NCC og NCC+. De inneholder også antall ord for de ulike kildene i milliarder ord. Figurene er her også med Common Crawl<sup>1</sup> for å vise et eksempel på den totale mengden treningsdata.



**Figur 1: NCC med Common Crawl**

Som man kan se av figur 1 utgjør bøker og aviser en relativt liten andel av tilgjengelige åpne data.



**Figur 2: NCC+ med Common Crawl**

I figur 2 ser vi at i volum er aviser og bøker de absolutt største volumene av tekst fra NBs samling. Samlet sett utgjør de et signifikant volum av tilgjengelige treningsdata når man tar med innhold under opphavsrett. Det er grunn til å tro at disse samlingene også har betydning for ytelsen i modellene.

---

<sup>1</sup> Tredjeparts data som f eks Common Crawl, MC4 og HPLT distribueres ikke videre av NB, men kan kombineres med NCC/NCC+ av andre.

De ulike delene av innholdet som brukes til trening av norske språkmodeller har sett fra NBS perspektiv ulike muligheter og begrensninger. Tabell 1 viser om data kan brukes i ulike sammenhenger. “Åpent” betyr at det ikke er begrensninger for bruk. “Forskning” betyr at innholdet kan brukes til forskning, også på kunstig intelligens og språkteknologi. “Språkteknologi” betyr betyr at innholdet kan brukes for forskning og utvikling av språkteknologi uten betydelige begrensninger.

Kilder	Åpent	Forskning	Språkteknologi
Bøker under opphavsrett	Nei	Ja	Nei
Bøker i det fri	Ja	Ja	Ja
Aviser i det fri	Ja	Ja	Ja
Aviser under opphavsrett, uten avtaler	Nei	Ja	Nei
Aviser med avtaler	Nei	Ja	Delvis
NRK artikler	Nei	Ja	Nei
HPLT	Nei	?	Nei
mC4	Nei	?	Nei
OSCAR	Nei	?	Nei
VG Debatt	Nei	Delvis	Nei
Twitter	Nei	Delvis	Nei
Reddit	Nei	Delvis	Nei
Facebook	Nei	Delvis	Nei
Wikipedia	Ja	Ja	Ja
Målfrid	Ja	Ja	Ja
Stortinget, offentlige rapporter	Ja	Ja	Ja
Lovdata	Delvis	Ja	Delvis

**Tabell 1: Treningsdata delt i komponenter med mulighet for bruk**

## Bøker

Som det går fram av figur 2 har NCC+ ca 14,4 milliarder ord fra bøker. Det tilsvarende tekstvolumet i NCC er ca 1,9 milliarder ord, noe som er volumet av bøker som er utenfor opphavsrett. Forskjellen er betydelig i volum.

Uten at det diskuteres nærmere, men med referanse til beskrivelsen over, er det grunn til å anta at det språklige uttrykket i bøker i det fri er annerledes enn språket som normalt brukes i dag. Dette vil også bidra til at modeller får et språkuttrykk som er mer lik det som ble brukt skriftlig tidlig på 1900-tallet.

## Aviser

I utvalget av aviser som er gjort for NCC er praksis for NBs nettbibliotek lagt til grunn sammen med avtaler med aviser. Aviser som er tatt med i NCC (åpne data) betyr det at man i utgangspunktet har gått ut fra at avisutgivelser som er eldre enn 120 år er i det fri og kan distribueres. Videre er aviser som er gått inn (opphørt) tatt med, og også en del aviser der det finnes eksplisitte avtaler om bruk for språkteknologiske formål.

I NCC+ er det tatt med både nyere digitalisert og digitalt pliktavlevert innhold. De absolutte tallene fra NCC og NCC+ er på hhv ca 2,6 og 10,6 milliarder ord, noe som tydelig viser at volumet av treningsdata vil øke betydelig med å inkludere nyere utgivelser.

I NCC inngår derfor en god del digitaliserte aviser der det er inngått avtaler, noe som betyr at ikke alt av de 2,6 milliarder ord i NCC er utenfor opphavsrett.

## Avtaler med aviser

I forbindelse med digitalisering og oppgaver rundt digital håndtering har NB gjort avtaler med flere aviser og mediehus. Avtalene angår forhold som tilgang for publikum på nett, bruk i bibliotek og bruk i språkteknologi. De varierer i innhold fra avtale til avtale. I nyere avtaler er typisk følgende tatt med

*“Nasjonalbiblioteket gis også tillatelse til at tekstinformasjonen fra avisen kan brukes til forskning og utvikling av språkteknologi. I den forbindelse er det den språklige strukturen, dvs. enkeltordene, setningene og måten de er satt sammen på, som er interessant, ikke det intellektuelle innholdet. En slik bruk av tekstinformasjonen skal gjøres på en måte som ikke kommer i konflikt med tjenester som formidler avisens historiske arkiv.”*

For enkelte avtaler, f eks for Aftenposten (2009) og Stavanger Aftenblad (2010) er dette ikke med, mens for VG (2018) er det med.



# Scenario-notat, mediebransjen & KI

Frank Gander

*“Technology is not a soloist but part of an ensemble. It interacts with institutions and other forces in society and the economy.”* Carl Benedikt Frey, svensk økonom

Dette notatet forsøker å kaste lys på kunstig intelligens (KI/AI) og mulige langsiktige konsekvenser for mediehus, og er ment som et oppspill til refleksjon og diskusjon.

Det hensyntar i liten grad andre faktorer som utvilsomt vil spille inn på utviklingen av mediebransjen. Bare i FutureTrends-rapporten 2024 (Amy Webb) pekes det på at helt andre teknologier og digitale infrastrukturer også vil virke inn på både nyhets- og underholdningsmedier de neste årene – i tillegg til kunstig intelligens.

En rekke rammevilkår, som den økonomiske utviklingen – både på individ- organisasjons- og samfunnsnivå – er eksempler på faktorer som også kan eller vil påvirke utviklingen av norske medier. Det samme med mediestøtteordninger, juridiske rammer, nye internettilkoblede «dingser», endring i befolkningens mediebruk, medieeierskap, opphavsrettavklaringer, sosiale medier, geopolittikk og det politiske klimaet internasjonalt, nasjonalt og lokalt .

Notatet er ikke resultatet av en bred scenario-prosess, men forsøker å ta inn ulike scenarier og beskrivelser som har vært publisert, diskutert og analysert av forskere, KI-eksperter, bransjeeksperter og mediebransjen selv i løpet av de siste månedene. Det er krevende å skulle lage et slags oppdatert veikart, gitt at omløpshastigheten på ny informasjon er så stor: Hver uke publiseres det flere titalls nye forskningsartikler om temaet, i tillegg til flere hundre artikler, undersøkelser og kartlegginger. Debatten er preget av mye dystopi eller ukuelig optimisme. Dette notatet ligger nok mer på den pessimistiske siden fordi det da er enklere å starte en diskusjon på hvordan unngå de mulige nedsidene av KI-utviklingen.

Måten dette notatet er strukturert på, er kort å oppsummere nåsituasjonen, så trekke opp noen mulige «end-game»-scenarier. Altså teoretiske beskrivelser av hvor vi kan være om 5-10 år. Hvert av scenariene er underbygget med signaler eller trender som vi ser allerede nå.

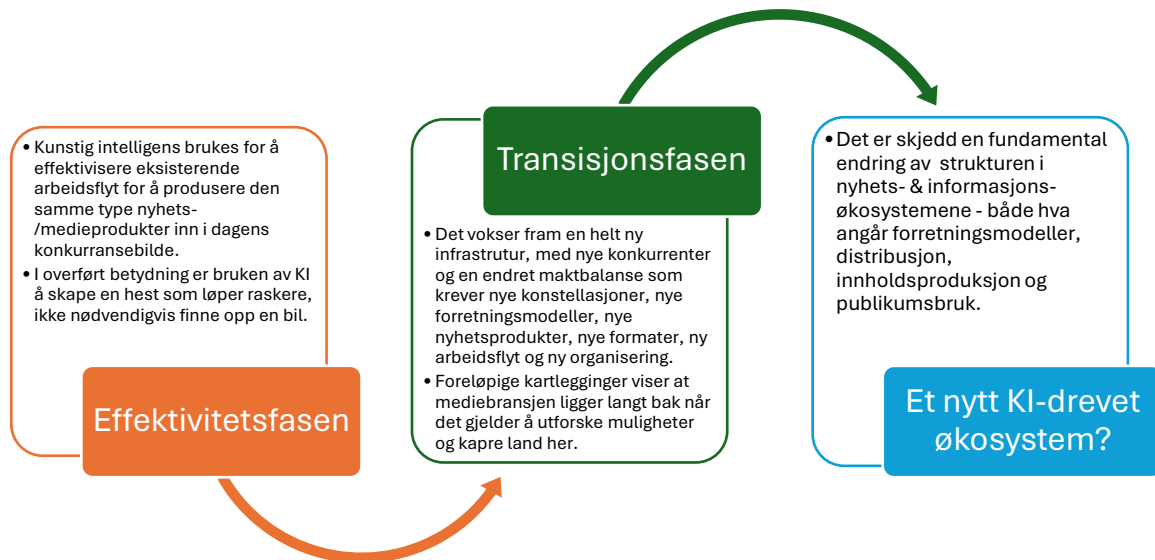
Til slutt forsøker notatet å trekke opp noen strategiske utfordringer som kan være gode diskusjonspunkter, uavhengig av hvilket scenarie som vil ligge nærmest fasit om 5-10 år.

## Hvor står vi?

Noen norske og nordiske mediehus er verdensledende i det å ta i bruk verktøyene og mulighetene som språkmodeller og generativ KI gir. De eksperimenterer med teknologien og bruker den for å forbedre eller effektivisere dagens arbeidsflyt. Andre nyhetsmedier ligger svært langt etter i hele digitaliseringsarbeidet, og er langt unna noen KI-eksperimentering. Så «strekk i laget» er kanskje en dekkende beskrivelse.

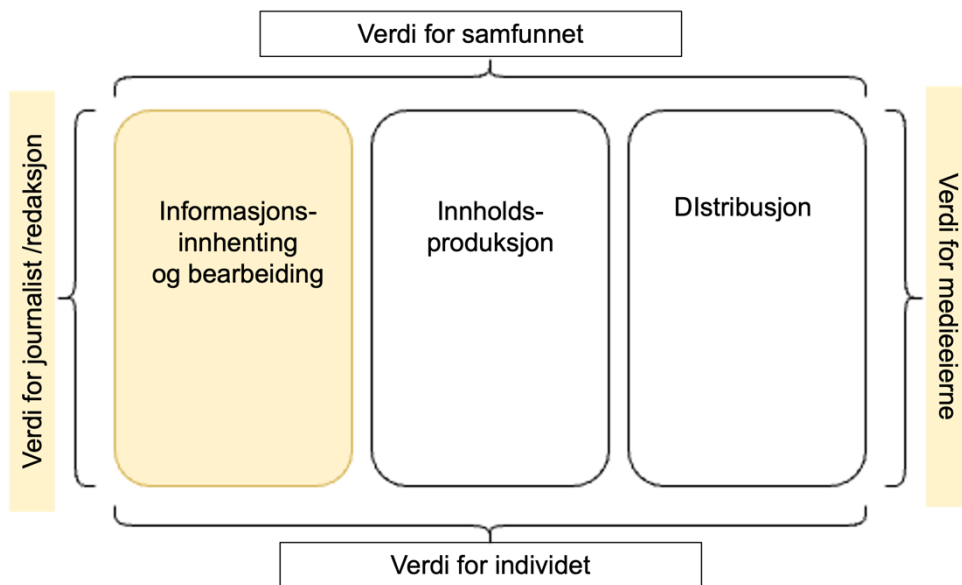
Samtidig er forskere og bransjeeksperter allerede bekymret over at mediene ikke klarer å ta et større fugleperspektiv og se på teknologien som en mulighet (eller trussel) som krever at hele

den eksisterende forretningsmodellen, verdikjeden og mediebruken settes under lupen – og transformeres. Man blir i «effektivitetsfasen» fordi det er enklest. Dette krever minst mulig investeringer og endringer i forretningsmodell, organisasjon og produktportefølje (se figuren nedenfor).



Mediehuset AP publiserte i april 2024 rapporten «Generative AI in Journalism: The Evolution of Newswork and Ethics in a Generative Information Ecosystem», utarbeidet av en rekke forskere som blant annet hadde undersøkt KI-bruken i 180 mediehus. Oppsummeringen viser det samme som andre bransjeeksperter peker på: Kunstig intelligens brukes akkurat nå for det meste til informasjonsinnhenting og bearbeidelse. Bruken gir foreløpig mer verdi til journalisten, redaksjonen og eventuelt medieeierne i form av større effektivitet/lavere kostnader, enn det gir ny verdi til samfunn eller enkeltbrukere. (Se figur nedenfor.)

En av forskerne, Nick Diakopoulos, peker på at det kan være et rett valg at mediene fokuserer på å bli i dette hjørnet, gitt at det kan være her nyhetsmediene kan utvikle en unik posisjon. Men han understreker at dette i så fall må være et gjennomtenkt, strategisk valg – og ikke komme som følge av at det oppleves som det enkleste akkurat nå.



(Skisse brukt av Nick Dakopoulos under Nordic AI in Journalism-konferansen 2024 ifbm Aps «Generative AI in Journalism-rapport»)

## Mediebransjen og KI - scenarier

*“In times of change the learners will inherit the world, while the knowers will be beautifully equipped for a world that no longer exists.”*

Eric Hoffer, amerikansk filosof

### End-game-scenarier

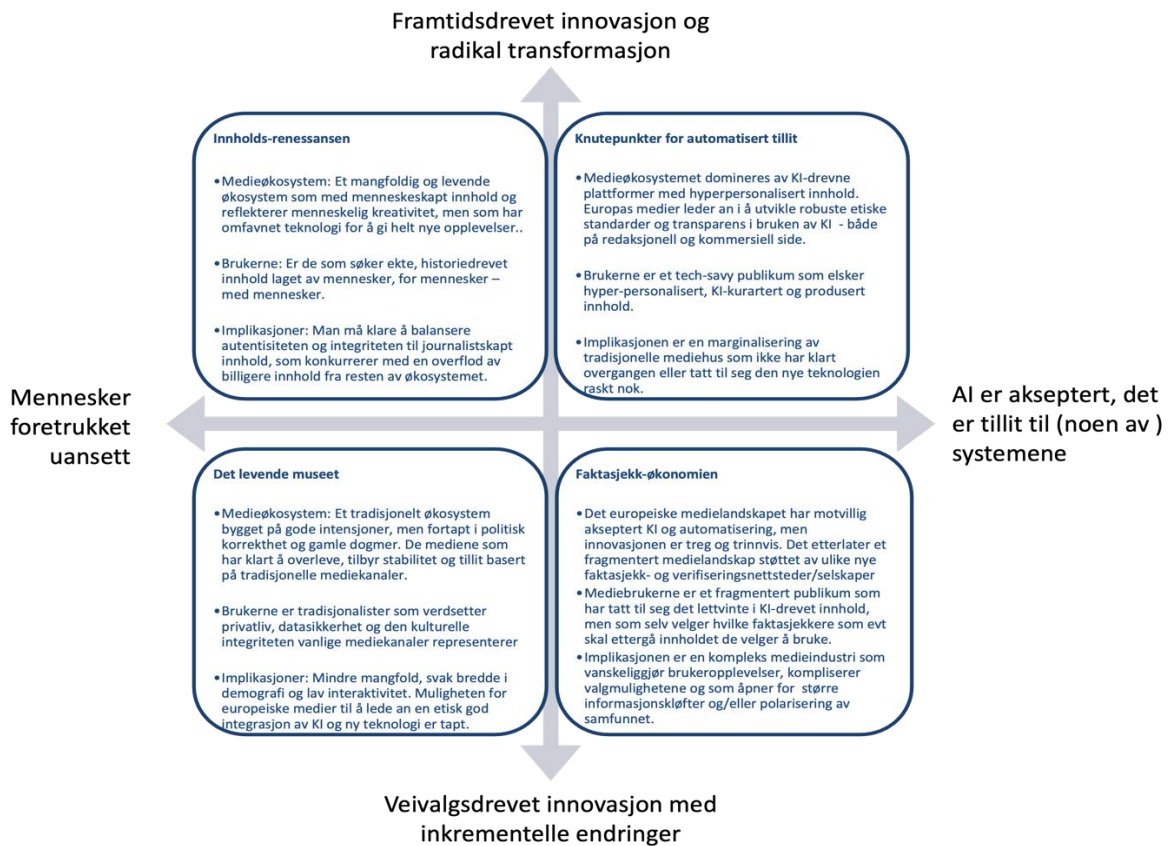
Det foregår nå scenario-diskusjoner rundt KI og nyhetsmedier/journalistikk flere steder. Hvor er samfunnet, mediesamfunnet og mediene egentlig på vei?

Open Society Foundation (OSF) har startet et stort arbeid som begynte med 900 framtidsbeskrivelser, via 60 hovedfokuspunkter og har endt opp med snaue ti hovedspor som endte opp i en foreløpig rapport som ble publisert i slutten av august (<https://www.opensocietyfoundations.org/uploads/3073dc0d-90e1-49f5-ab3a-367d5d046ff6/ai-in-journalism-futures-initial-report-20240819.pdf>) Reuters Institute Digital News Report (2024) som ble publisert i midten av juni og inneholder også noen trender & scenarier rundt KI og nyhetsmedier. (<https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024>). Dette notatet er ferdigstilt før publiseringen av disse to, men har forholdt seg til relevante deler av forarbeidet til rapportene, som har blitt publisert og diskutert på konferanser.

Et konkret scenarie-eksempel er modellen nedenfor. Det er en ennå ikke publisert oppsummering fra en europeisk rundebordskonferanse i mars/april mellom samfunnsvitere, medieledere og politikere. Arbeidet endte opp i fire tenkte scenarier for det europeiske medielandskapet i 2035.

Modellen er bygget over aksene teknologiendring, med radikal endring og transformasjon vs en mer forsiktig og inkrementell innføring av ny teknologi som de to ytterpunktene.

Den andre aksen er bygget rundt menneske/teknologi-fordelingen, hvor det ene ytterpunktet er høy aksept for KI/AI i mediene, og en motsatt ende hvor mennesker er foretrukket uansett funksjon/oppgave. Modellen tegner opp fire mulige utfall, og skisserer hva dette kan bety for økosystem, brukere og hva slags implikasjoner det enkelte scenarie kan gi:



Basert på noen foreløpige scenariebeskrivelser fra OSF-arbeidet, FutureTrends-rapporten, Media Innovation Xchanger-rapporten, diskusjoner på konferanser, og på bransjenettsteder de siste månedene, forsøker vi under å beskrive fem mulige scenarier for hvor medieverdenen kan være om 5-10 år. Merkelappene er ikke begreper som nødvendigvis florerer eller brukes av alle, men er laget for dette notatet:



## 1 Den hyperpersonaliserte verdenen

**Noen kjennetegn:** KI-agenter, chatboter og automatiske tjenester skreddersyr dagen til store deler av befolkningen. KI er en del av dagliglivet både når det gjelder arbeid, fritid, utdanning, helse og mediebruk. Forslagene og tilretteleggingen av hverdagen er basert på en rekke data som den enkelte bruker mer eller mindre frivillig har valgt å gi fra seg. Dette gjør at tilbudene befolkningen får, oppleves som mer relevante og nyttige.

Fra et publisiståsted er dette nisjepublikummets æra. Dersom innholdsvarelageret er stort nok – og teknologien god nok – kan målgruppen som får en eksklusiv innholdsmiks være nede i én person.

Samtidig blir det mer og mer krevende å sette en felles dagsorden eller skape debatter som favner hele nedslagsfeltet/befolkningen til mediehuset. Folk lever side om side i litt forskjellige virkeligheter/siloer, og kan i større grad også la være å forholde seg til de store nyhetene, politiske spørsmålene eller deler av samfunnet de ikke ønsker å se. Det er en risiko for at medienes demokratiske funksjon og samfunnsoppdrag blir satt under press, og at fragmentering/polarisering blir mer vanlig.

Gitt at hyper-personaliseringen også gjelder forbruk av varer og tjenester, så er andre medienes direktetraffikk og forretningsmodeller under sterkt press, inklusive brukerbetalings-tjenestene. Brukerne er ikke villige til å betale full pris for smale tjenester, og løsningene kan innebære splitting av inntekter for mediene, også med plattformen brukeren ønsker leveransen på. (Hva vil en bruker egentlig betale for miksen av spisset sportsstoff fra VG, børspush fra DN, utenriksinnhold fra Aftenposten, nyhetsmeldinger fra Nordstrands Blad og kulturkronikker fra Morgenbladet?)

### Signaler:

Samtlige næringer, enten det er finans, helse, handel, transport, markedsføring eller medier, jobber nå med å bruke KI til personalisering langt utover det å bruke alder, kjønn og adresse.

Hyperpersonalisering er tematikk som dukker opp i nesten alle trend-rapporter, og hvor det pekes på en potensielt stor økonomisk oppside, en mye mer relevant brukeropplevelse, men

også en enda mer fragmentert samfunnsstruktur, en lavere borgerdeltakelse og mindre oppslutning om fellesløsninger.

## 2 Survival of the largest & smallest

**Noen kjennetegn:** Trenden har eksistert allerede før kunstig intelligens har slått innover medie verdenen: Digitalisering av produktporteføljene, syndikering av innhold og strømlinjeformet output gir stadig mer like medieprodukter i et marked hvor tech-gigantene allerede har stukket av med både annonsepenger og distribusjonskraft. Marginene blir stadig mindre, noe som igjen gir stadige nye kuttrunder i redaksjoner og mediehus.

Den kostbare journalistikken – eksempelvis utenriksjournalistikk og gravejournalistikk - genererer minst penger, og er sjelden trafikkvinner. Publikum velger mediemerkevarer som passer deres liv og verdier, ikke nødvendigvis knyttet til geografi eller språk.

I scenariet «Survival of the largest & smallest» har KI forsterket trenden gjennom å styrke de store aktørene som har investeringskraft til å drive personalisering, automatisering av regional/lokal journalistikk og informasjon, og å få på plass gode språktilpasninger ut mot sine brukere.

Løpende nasjonale nyheter er overskuddsvarer som har mindre og mindre egenverdi fordi «alle» har det. Sitatsaker skrevet av KI kan hos deler av brukerne redusere tilliten til en mediemerkevare. Dette fordi innholdet er overalt, laget av alle, også i ikke-redaktørstyrte organisasjoner.

Vinnerne er i dette scenariet de største nyhetsorganisasjonene, og muligens de små lokale som kan tilby et bredt sett med egennyheter som er vanskeligere å kopiere i løpet av noen sekunder. Men sistnevnte er i større grad avhengig av KI-tjenester fra en tredjepart.

Mediemangfoldet forvitrer. Det er bare plass til noen få store aktører på nasjonalt/regionalt plan, supplert av lokalmedier som evner balansen mellom egennyheter og bruk av syndikert KI-/fellesinnhold.

Det er bare de store medieorganisasjonene som har hatt råd til å eksperimentere, bygge kompetanse og investere i kunstig intelligens – evt. også utfordre den. Og det er de store medieaktørene som har endt opp med bilaterale, økonomisk fordelaktige avtaler med språkmodell-gigantene.

### Signaler:

I USA er New York Times allerede blitt en stor utfordrer til LA Times. NYT har flere digitale abonnenter i California enn hva LA Times har i totale abonnementsstall i LA-regionen. Deres brede satsing på bruk av KI og personalisering, bygger opp under strategien om å være den største og beste nasjonale og regionale aktør i USA. Det er mulig å se for seg en framtid med enda færre norske nasjonale og regionale mediemerkevarer enn i dag, med et mindre mediemangfold, men med muligheter for høyere marginer.

Flere bransjeeksperter peker på at det allerede er så mange store enkeltmediehus som har inngått særavtaler med Google, OpenAI m.fl. - og tilgjengeliggjort sitt innhold for

språkmodelltrening. De mener slaget om et slags felles opphavsrettsopprør hva angår treningsdata allerede er tapt, og at diskusjonen mediehusene bør konsentrere seg om er KI-gigantenes tilgang til «grounding data» - forankringsdata. Dette er oppdaterte data fra mediehus og kunnskapsbanker som kan gjøre språkmodellen i stand til å levere mer sanntidsorienterte, spesialiserte og presise tjenester.

Trening av språkmodeller er blitt enklere. Tilgang til data og tilgang til datakraft er nå det mest krevende for tech-bransjen, som kan komme til å strekke seg lenger økonomisk for å få denne tilgangen. Uansett type avtaler, så er det de største mediehusene - med størst mulig innholdsvolum - som trolig vil komme best ut av nye avtaler med de internasjonale tech-selskapene.

Diskusjonen om «den skjulte regningen» har allerede startet. Mediehus endrer arbeidsflyt og bygger CMS-systemer som gjør dem mer avhengige av språkmodeller og KI-tjenester fra tredjepart. Disse kan fort øke vesentlig i pris. Igjen vil dette kunne skape et skille mellom (store) medier som kan betale, og de som ikke har så robust økonomi. Dette vil kunne ramme alle deler av forretningsmodellen til norske medier.

### 3 Informasjonsoverloadens (evt. informasjonssøppelets) tidsalder

**Noen kjennetegn:** Kombinasjonen av ulike profesjonelle medieaktører, bloggere/influencere, profesjonelle informasjons- og pr-aktører, markedsførere, sosiale medier, spam-produsenter, ondsinnede aktører og kunstig intelligens gjør at samfunnet flommer over av informasjon. Mye av dette er søppel, fake news og KI-billigprodusert automatkopier av originalt innhold.

Dagens forretningsmodeller og arbeidsmarkedet, slik det ser ut nå, er under sterkt press. Videoproduksjonstjenester som Open Als Sora gjør eksempelvis at en kreativ person med gode visuelle historiefortellingsferdigheter kan erstatte et helt tv-produksjon-/reklame-selskap.

Scenariet åpner for et eldorado av valgmuligheter for den enkelte bruker, men også en situasjon hvor det blir mer og mer vanskelig å navigere i eller å stole på informasjonen – også innhold fra redaktørstyrte medier. Situasjonen kan vanskeliggjøre arbeidet med å nå ut til sluttbrukeren med kvalitetsjournalistikk fordi publisisten kjemper om tid, oppmerksomhet og teknologisk gjennomslagskraft med så mange ulike krefter.

Dersom redaktørstyrte medier klarer å skape «trygge havner» med kvalitetssikret innhold, kan dette bli en konkurransefordel. Spørsmålet er hvor mye en bruker er villig til å betale for, og om nyhetsinnhold i dette sceneriet er ansett som «need» eller «nice».

#### Signaler:

Alphabet-sjef Sundar Pichai innrømmet utfordringen senest i midten av mai måned: «Hver gang det er en overgang, får du en eksplosjon av nytt innhold – også med AI». Søkegiganten står ikke bare overfor ny konkurranse fra KI-tjenestene. KI-skapt søppelinnhold med skreddersydd

søkemotoroptimalisering er en utfordring i seg selv, fordi det skader Googles søkeresultater (og dermed trafikk til kvalitetsmedier).

Utfordringen med masseproduksjon av innhold som ikke er kvalitets- eller faktasjekkert, har satt i gang en rekke initiativer for å merke KI-generert innhold på ulike måter, eller for å sikre prominensen til originalt innhold av selskaper som TikTok og Adobe, til initiativer som Reynir-prosjektet hvor MediaCityBergen er partner.

Mediene pumper allerede så mye innhold ut til brukerne i løpet av et nyhetsdøgn, at publikum er i ferd med å bli overveldet. Spørsmålet blir derfor også hvordan mediene selv klarer å bli fyrårn, veivisere og portvoktere, i stedet for å fyre opp under informasjonsoverloaden.

## 4 Maskiner i midten

**Noen kjennetegn:** Dette scenariet er egentlig en kombinasjon av flere faktorer i de overnevnte scenariene, og sosiale medier slik de fungerer i dag. Men her vil KI-drevne maskiner alltid fungere som mellommenn i informasjonsstrømmen.

Maskinene tar inn informasjon som input, analyserer, pakker og versjonerer innholdet - og sender det ut til publikum som output. Maskinene og algoritmene overtar på mange måter redaktør- eller publisistansvaret, og gjør at mediebransjens og journalistenes rolle endres til å bli informasjonsinnhentere og faktasjekkere.

Scenariet legger opp til at forskjellige sosiale eller økonomiske deler av befolkningen kan motta tilpasset informasjon, noe som kan skape en mer fragmentert opplevelse av samfunnet.

Denne endringen av hvordan informasjon håndteres og distribueres, betyr i så fall en mulig redusert menneskelig innflytelse over nyhetsproduksjon og distribusjon i et fremtidig KI-drevet informasjonssamfunn. Alt journalistisk innhold er «atomisert», maskinene står i større grad for historiefortellingen.

Med mindre direktører og redaktører selv sitter som kusker eller dirigenter for disse maskinene, så risikerer mediehusene å miste dagens direkteforhold til brukere, abonnenter og annonsører. Dermed er også hele forretningsmodellen til dagens norske medier svært utsatt.

I «Maskiner i midten»-scenariet vil integreringen av KI i dagliglivet kunne endre begrepene våre om identitet og menneskelig samhandling, dette kan redusere menneskelig handlekraft og beslutningsevner. Og det kan vokse fram en forståelse av at KI i seg selv har visse rettigheter og forpliktelser på linje med individer og institusjoner, eller at KI som både er grenseoverskridende og svært inngripende kan framtvinge en annen form for internasjonal lovgivning.

### Signaler:

Som nevnt, så er flere sosiale medier-plattformer nærmest her allerede. X (Twitter) har - uten ennå å lykkes - forsøkt å lansere en type allvitende kunstig intelligens i midten av sitt univers. Et



system som overvåker og oppsummerer inputen, alt det som snakkes om og publiseres på markedsplassen X, og leverer dette som ferdigtygde aktualitetsoppsummeringer til brukerne.

Søk etter informasjon vil, som nå, også være styrt av maskiner i dette scenariet. Men blant annet i den siste INMA-rapporten om KI, pekes det på at det allerede er i ferd med å vokse fram to typer søk, som kan skape større informasjonskløfter, et mer fragmentert samfunn, og en veldig forskjellig informasjonstilgang basert på brukernes økonomi.

Alle har tilgang på et gratis innholdssøk, som kanskje ikke gir så relevante svar, som ikke skiller like godt mellom fake og fakta, og som er sponset av noen kommersielle/politiske interesser, mens brukere med god økonomi og betalingsvilje kan kjøpe «A-klasse-søk», støttet av kunstig intelligens og mediemerkevarer for å sikre relevant, faktasjekket informasjon.

## 5 Sivilsamfunns-informasjonen og journalistikkens gjenoppståelse – (AI for good)

Medier og teknologiutviklere har sammen med klok myndighetsforvaltning av regelverk og tilpassede støtteordninger bidratt til en god etisk og transparent innføring av KI, med tydelige rammer og forutsetninger. Det er utviklet språkmodeller og KI-tjenester som bygger på og speiler norsk og samisk kultur, språk og verdisett. Og hvor hensyn til opphavsrett er ivaretatt.

Redaktørstyrte medier har sammen med sivilsamfunn - og kommersielle aktører som også drives av samfunnsverdi - utviklet KI-drevne økosystemer og forretningsmodeller som appellerer til norske brukere. De balanserer maskin- og menneskebeslutninger, og gir publikum tjenester og innhold som nordmenn er villige til å betale for, og som gjør dem bedre i stand til å navigere i, forstå og nyttiggjøre seg av informasjon i en stadig mer kompleks hverdag.

### Signaler:

Det er mange signaler som bygger opp under et mer oppløftende bilde av en framtidig norsk medievirkelighet, hvor KI spiller en viktig rolle:

- EUs evne og vilje til å regulere KI, med tydelige krav til personvern og transparens.
- Norske innovasjonssamarbeid rundt språkmodeller og KI-tjenester, som eksempelvis NorwAI, Nasjonalbibliotekets satsinger, Demokratibasen i regi av UiB/Sujo, det tidligere nevnte Reynir-prosjektet.
- Bransjesamarbeid rundt deling av erfaringer og løsninger, en felles etikkholdning til KI, og gode intensjoner for å styrke tillit og omdømme som viktigste valuta for norske medier i framtiden. (Så er nok det siste enklere å si enn å gjøre).

*“The important thing will be to use regulation and other tools to make sure that AI are working for everyone, and just not for Big Tech companies. A hard lesson we’ve already learned through our experience about social media...”*

*Bruce Schneier, foreleser på Harvard Kennedy School / Ash Center for Democratic Governance and Innovation.*

Som nevnt tidligere er scenariene teoretiske modeller, og det er flere pessimistiske enn optimistiske fremtidsbilder, fordi det gjør det enklere å diskutere veier ut av mørket. Hvis vi ser de ulike scenariene under ett, analyserer signalene, og setter den norske medievirkeligheten av i dag under lupen, så er det mulig å identifisere noen spørsmål som er verdt å diskutere – og som mulig vil kunne ha en verdi uavhengig av scenarie-fasit:

- Kan en mye tettere nasjonal infrastruktur, eksempelvis i form av felles innloggingsmodeller på tvers av mediehusene, samarbeid om norske språkmodeller, felles avklaringer mot tech-gigantene, og diskusjoner rundt felles nasjonale, kommersielle løsninger bidra til en «lock in»-effekt, som kan gjøre norske mediehus mer robuste opp mot de store tech-gigantene?
- «Follow the money». Gitt de ulike scenariene, hvilke forretningsmodeller vil være robuste nok til å sikre at journalistisk produksjon også vil være bærekraftig i framtiden?
- Bør det foregå en enda tettere dialog mellom norske og nordiske mediehus og opphavsrettshavere for å få på plass felles holdninger og policy rundt bruk av innhold til «training data» og «grounding data»?
- Bør mediehusene og medieorganisasjonene gå enda lenger i det å skape et tydelig, felles og transparent system for god etisk bruk av KI i redaksjonell produksjon?
- Bør mediehusene i større grad nå gå sammen om initiativer, verktøy og prosjektet rundt KI som kan styrke den norske offentlige samtalen, og legge til rette for at mediene faktisk oppleves som viktige, nyttige og demokratiske også i et fem til tiårs-perspektiv?
- Hvordan kan norske medier, kunnskapsinstitusjoner og myndigheter jobbe sammen for å tydeliggjøre verdien av kildekritikk og medienes rolle i et levende og velfungerende demokrati? Hvilke faktorer her er viktige for å møte utfordringene KI medfører for samfunn, organisasjon og individ?
- Hvordan jobber den norske redaktørstyrte mediebransjen med å systematisk kartlegge eksterne trusler – både rundt teknologi, juridiske rammevilkår, opphavsrett, forretningsmodeller og gråsonene rundt mediebransjen? Et eksempel på det siste: Hva gjør kunstig intelligens med effektiviteten og volumet til innhold fra influenser/creator/content-segmentet – og hva gjør dette med forventningene til publikum hva angår journalistikk?

# Examining the Incorporation of Norwegian Editorial Media Content in Public Generative AI System Training

## Abstract

This report examines the use of Norwegian editorial media content in training and grounding large language models (LLMs) within generative AI (GAI) ecosystems. By analyzing several public datasets and conducting memorization tests on the grounding layer and base model of Microsoft Copilot, we have the following key findings: 1) Public Dataset Composition: Norwegian editorial content is more prevalent in MC4, OpenWebText, and OpenWebText2 compared to Common Crawl, indicating a skew towards publisher content in curated LLM training datasets. 2) Paywalled Content: There are no paywalled Amedia news articles in the Common Crawl, OpenWebText, and OpenWebText2 datasets, with MC4 containing only 39 such articles. 3) Grounding Layer Memorization: The memorization ability of the grounding layer is influenced by the accessibility of news articles. More content from NRK and Dagbladet is captured compared to VG and Aftenposten due to the latter's subscription barriers. 4) Base Model Memorization: In a test using given prompts, the GPT-3.5 base model only reproduces a few words or sentences from a limited sample of Norwegian news articles. These findings provide insights into the dataset composition and memorization characteristics of LLMs concerning Norwegian editorial content.

## 1. Introduction

The improvement of Transformer-based deep neural networks drove the prosperity of Generative AI (GAI) systems in the early 2020s. The representatives include large language model chatbots such as ChatGPT, text-to-image, and text-to-video AI generators such as DALL-E and Sora. These systems are capable of generating text, images, or other data using generative models, often in response to prompts. This report specifically focuses on public Large Language Models (LLMs) and their associated GAI products, which have been developed based on a significant amount of textual data. LLMs are trained to predict the next likely word following a given string of words or "prompt", enabling them to generate longer strings of text that closely approximate human language.

To train a model to produce text that approximates natural human language in this way requires "training" with an enormous volume of examples. The life cycle of such an LLM

begins with an “input” phase, processing potentially billions of training works running into the trillions of words. To obtain such volume, the developers of these models appear to have made copies of a substantial portion of the Internet, including paywalled material. They make these copies either by scraping them directly from websites or copying them from archives of copied content, like Common Crawl, created by others who have done the scraping. After their initial “pre-training”, models may be “fine-tuned” with additional copied sources selected to improve performance for desired subjects or tasks. In addition to fine-tuning, LLMs can be given access to up-to-the-minute information that was not included in their training sets. This can be done by first retrieving the external data and including the search results as context in the prompt to LLMs. This method, known as ‘grounding’, is employed by GAI-based applications such as Microsoft’s Bing Chat, OpenAI’s ChatGPT-Plus, Anthropic’s Claude-2, and Google’s Search Generative Experience. The products generate outputs comprised of natural-language synopses that knit together and paraphrase the original sources of search results. According to the News Media Alliance report<sup>2</sup>, the workflow of a GAI ecosystem for text can be illustrated in Figure 1.

In this report, we investigate the extent to which Norwegian editorial media content, including news, magazines, and digital media content, is used to serve the entire GAI ecosystem, such as training large language models (LLMs) and grounding LLMs with relevant contexts. Specifically, we divide our analysis into three parts. In Part 1, we assess the extent of copyrighted and paywalled Norwegian news publisher content that is included in public datasets reportedly used to train LLMs. In Part 2, we conduct memorization tests on one popular GAI product, Copilot, to identify the extent to which publisher content is used in the grounding layer. In Part 3, we conduct memorization tests on one of the OpenAI base models, GPT-3.5 Turbo, to identify the extent to which publisher content is used in training LLMs.

---

<sup>2</sup> News Media Alliance report: <https://www.newsmediaalliance.org/wp-content/uploads/2023/10/AI-White-Paper-with-Technical-Analysis.pdf>.

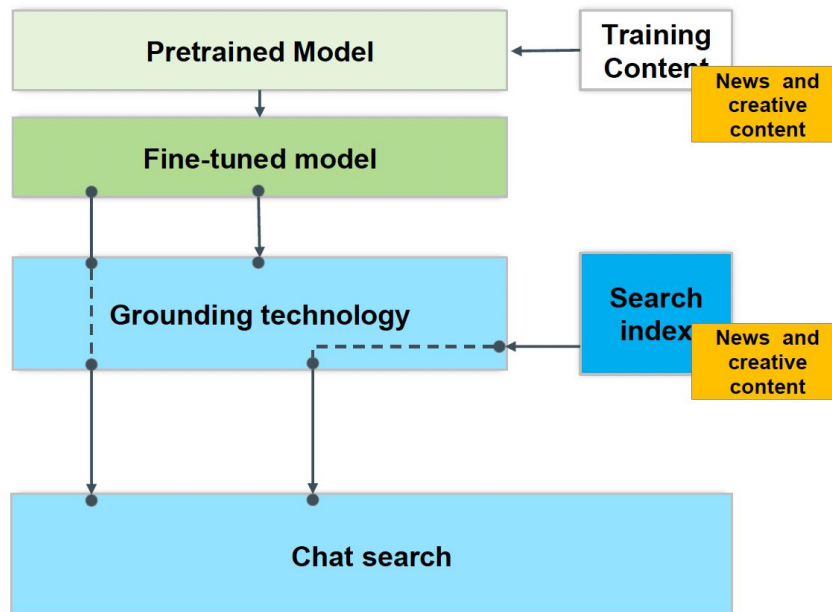


Figure 1: Workflow of a GAI ecosystem generating responses to user queries<sup>1</sup>

## 2. Norwegian Publisher Content in Public Datasets

### 2.1 Public Datasets Considered

We consider four public datasets: Common Crawl, mC4, OpenWebText, and OpenWebText2. All of them contain the Norwegian data.

Common Crawl provides free web crawl data that can be accessed and analyzed by anyone. Its web archive consists of petabytes of data collected since 2008, with new crawls generally completed every month. A single month's crawl can include upwards of 300 TiB of data, containing raw web page data, metadata extracts such as URLs, and text extracts. In this project, we analyzed only the November/December 2023 crawl archive - CC-MAIN-2023-50. The data was crawled between November 28th and December 12th and contains 3.35 billion web pages<sup>3</sup>.

mC4 is a multilingual, colossal, cleaned version of Common Crawl's web crawl corpus. It encompasses 101 languages and is generated from 86 Common Crawl dumps. mC4 is used to train Google's LLM T5 and Meta's LLaMA<sup>4</sup>.

<sup>3</sup> <https://www.commoncrawl.org/blog/november-december-2023-crawl-archive-now-available>

<sup>4</sup> Lauren Leffer. Your Personal Information Is Probably Being Used to Train Generative AI Models (Oct. 19, 2023). <https://www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/>.

In the paper "Language Models are Few-Shot Learners", OpenAI notes that a filtered version of Common Crawl and several curated high-quality datasets, including an expanded version of the WebText dataset (WebText2), two internet-based books corpora (Books1 and Books2), and English-language Wikipedia, were used to train the GPT-3 language model. However, WebText and WebText2 have not been released. OpenWebText and OpenWebText2 are replications of WebText and WebText2, created using the same methods OpenAI disclosed for creating WebTexts.

It should be noted that these public datasets do not cover the entirety of the training data for large language models (LLMs). A lot of additional training data is used beyond these public datasets. For example, WebText2 and Common Crawl data constitute 82% of the total training data for GPT-3, leaving the remaining 18% to other sources, such as internet-based Books1 &2, and English-language Wikipedia.

## 2.2 Sample of Publisher Domains

We focus on a domain-list provided by MBL. The domain list consists of URLs from medietall.no and fagpressen.no. In addition, we have included 15 known editorial sites that are not measured through medietall.no or fagpressen.no. This also includes older sites that are no longer measured. Included in these domains are digital news media and magazines.

## 2.3 Metrics of Interest

For each publisher content source  $S$  and dataset  $D$ , we focus on the number of unique URLs for  $S$  indexed in  $D$ .

Unique URLs are defined as unique URL strings that do not repeat within the corresponding dataset. To note, the method for evaluating the number of unique URLs in the data has some limitations, and there could be instances of links that point to the same page even though their URL strings are different; for example, `google.com/search` and `google.com/webhp` point to the same page but would be considered unique.

Besides, we are also interested in checking whether these public datasets contain content that was behind a paywall at the time of the crawl/scraping. So we asked our partner, Amedia, to provide a list of URLs for the paywalled articles published between January 1, 2015, and May 21, 2023, totalling 46,542 articles. Then, we counted the number of unique URLs that appeared in these four public datasets.

## 2.4 Statistical Processing

For Common Crawl, we used Amazon Athena to query all the webpage URLs with a top-level domain of “.no” (Norwegian), following instructions from the Common Crawl website<sup>5</sup>. Then, we downloaded the query results locally. For mC4, we downloaded its Norwegian part from Hugging Face locally<sup>6</sup>. We also downloaded OpenWebText<sup>7</sup> and OpenWebText2<sup>8</sup> from their GitHub repositories. After that, we parsed the URLs using the Python package<sup>9</sup> “urllib” and identified whether the URLs in each public dataset belonged to one of the sample Norwegian publisher domains. We removed duplicates, if any. Finally, the percentage of URLs belonging to Norwegian publisher domains out of the total Norwegian URLs in each public dataset is reported.

## 2.5 Results

### 2.5.1 Results in Common Crawl dataset

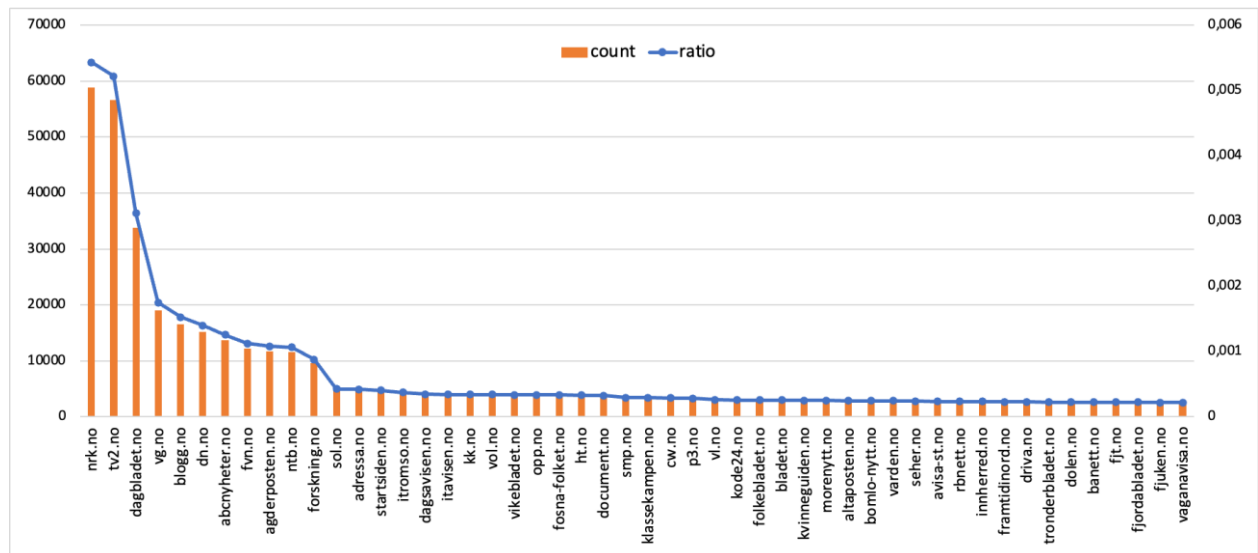


Figure 2: Unique URL statistics from Common Crawl belonging to Norwegian news media domains

<sup>5</sup> <https://commoncrawl.org/blog/index-to-warc-files-and-urls-in-columnar-format>

<sup>6</sup> <https://huggingface.co/datasets/legacy-datasets/mc4>

<sup>7</sup> <https://github.com/jcpeterson/openwebtext>

<sup>8</sup> <https://github.com/ElleutherAI/openwebtext2>

<sup>9</sup> <https://docs.python.org/3/library/urllib.parse.html#module-urllib.parse>

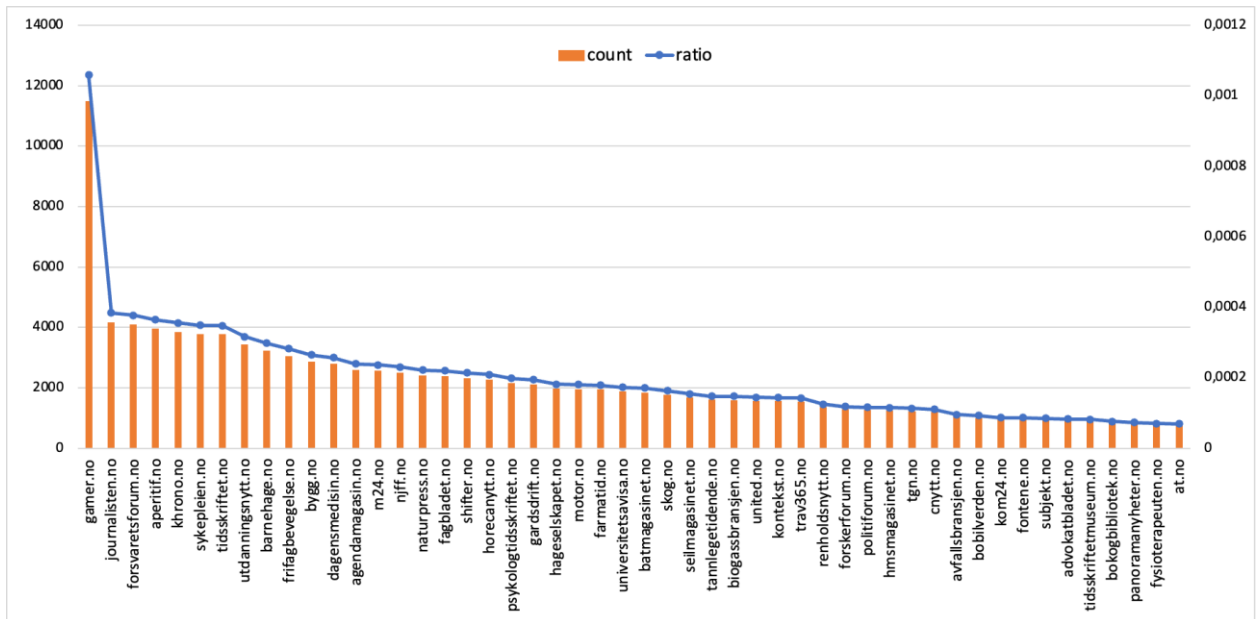


Figure 3: Unique URL statistics from Common Crawl belonging to Norwegian magazine domains

Number of unique Amedia URLs: 0

Number of unique URLs belonging to Norwegian publisher domain lists: 554072

Top-50 Norwegian domains in CommonCrawl dataset, sorted by count. Publisher domains are highlighted.				
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count	is_in
1	<a href="#">uib.no</a>	0.012846434489183151	139406	FALSE
2	<a href="#">google.no</a>	0.008457179212119878	91775	FALSE
3	<a href="#">uit.no</a>	0.00808645481037258	87752	FALSE
4	<a href="#">uio.no</a>	0.006882222525502162	74684	FALSE
5	<a href="#">ntnu.no</a>	0.006043738475912636	65585	FALSE
6	<a href="#">nrk.no</a>	0.005430287732081723	58928	TRUE
7	<a href="#">tv2.no</a>	0.005217971296181704	56624	TRUE
8	<a href="#">unit.no</a>	0.0037116672765542297	40278	FALSE
9	<a href="#">dagbladet.no</a>	0.0031157252665865993	33811	TRUE



10	<a href="#">oslomet.no</a>	0.002792458748731884	30303	FALSE
11	<a href="#">gulesider.no</a>	0.0022871013986990275	24819	FALSE
12	<a href="#">yr.no</a>	0.001992125308718142	21618	FALSE
13	<a href="#">vegvesen.no</a>	0.001974985179778297	21432	FALSE
14	<a href="#">1881.no</a>	0.0018877179641544613	20485	FALSE
15	<a href="#">nina.no</a>	0.0017557574015638249	19053	FALSE
16	<a href="#">vg.no</a>	0.0017521635035603088	19014	TRUE
17	<a href="#">sintef.no</a>	0.0016279436443618605	17666	FALSE
18	<a href="#">regjeringen.no</a>	0.0015311848519595085	16616	FALSE
19	<a href="#">blogg.no</a>	0.0015261165342622424	16561	TRUE
20	<a href="#">artsdatabanken.no</a>	0.0014985633162352867	16262	FALSE
21	<a href="#">finn.no</a>	0.0014345182107880156	15567	FALSE
22	<a href="#">dn.no</a>	0.0013981184745985593	15172	TRUE
23	<a href="#">ssb.no</a>	0.0013763707841157448	14936	FALSE
24	<a href="#">nhh.no</a>	0.0013210800456001151	14336	FALSE
25	<a href="#">bi.no</a>	0.0012787826306356583	13877	FALSE
26	<a href="#">abcnyheter.no</a>	0.001257311393845422	13644	TRUE
27	<a href="#">rbkweb.no</a>	0.0012445945239868271	13506	FALSE
28	<a href="#">dvh.no</a>	0.0012064439144110425	13092	FALSE
29	<a href="#">elkjop.no</a>	0.0011920683223969788	12936	FALSE
30	<a href="#">bokkilden.no</a>	0.0011843276190047906	12852	FALSE
31	<a href="#">udir.no</a>	0.0011840511653122125	12849	FALSE
32	<a href="#">nb.no</a>	0.0011792593013075245	12797	FALSE
33	<a href="#">kirken.no</a>	0.0011765869156126025	12768	FALSE
34	<a href="#">if.no</a>	0.001175757554534868	12759	FALSE
35	<a href="#">komplex-web.no</a>	0.001159354635441898	12581	FALSE
36	<a href="#">nasjonalmuseet.no</a>	0.0011426752626563495	12400	FALSE
37	<a href="#">ark.no</a>	0.001140371481884865	12375	FALSE

38	<a href="#">digi.no</a>	0.0011357639203418957	12325	FALSE
39	<a href="#">naob.no</a>	0.0011272860071028326	12233	FALSE
40	<a href="#">simula.no</a>	0.0011260880411016606	12220	FALSE
41	<a href="#">fyn.no</a>	0.0011219412357129884	12175	TRUE
42	<a href="#">pirate4x4.no</a>	0.001111712449087597	12064	FALSE
43	<a href="#">notam02.no</a>	0.0011043403506188461	11984	FALSE
44	<a href="#">agderposten.no</a>	0.001081210391673141	11733	TRUE
45	<a href="#">ntb.no</a>	0.0010666504971973585	11575	TRUE
46	<a href="#">felleskatalogen.no</a>	0.0010630565991938427	11536	FALSE
47	<a href="#">ebok.no</a>	0.001061490028269233	11519	FALSE
48	<a href="#">gamer.no</a>	0.0010585411888817328	11487	TRUE
49	<a href="#">bok365.no</a>	0.0010108068512965725	10969	FALSE
50	<a href="#">sparelys.no</a>	0.0009901649755840705	10745	FALSE

### 2.5.2 Results in mC4 dataset

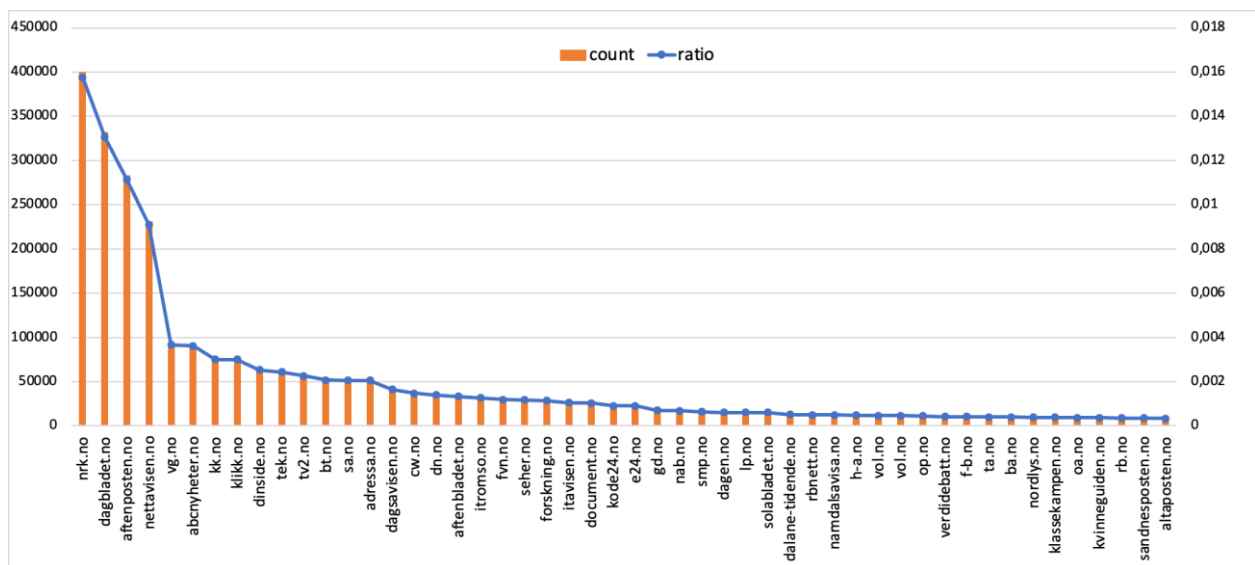


Figure 4: Unique URL statistics from mC4 belonging to Norwegian news media domains

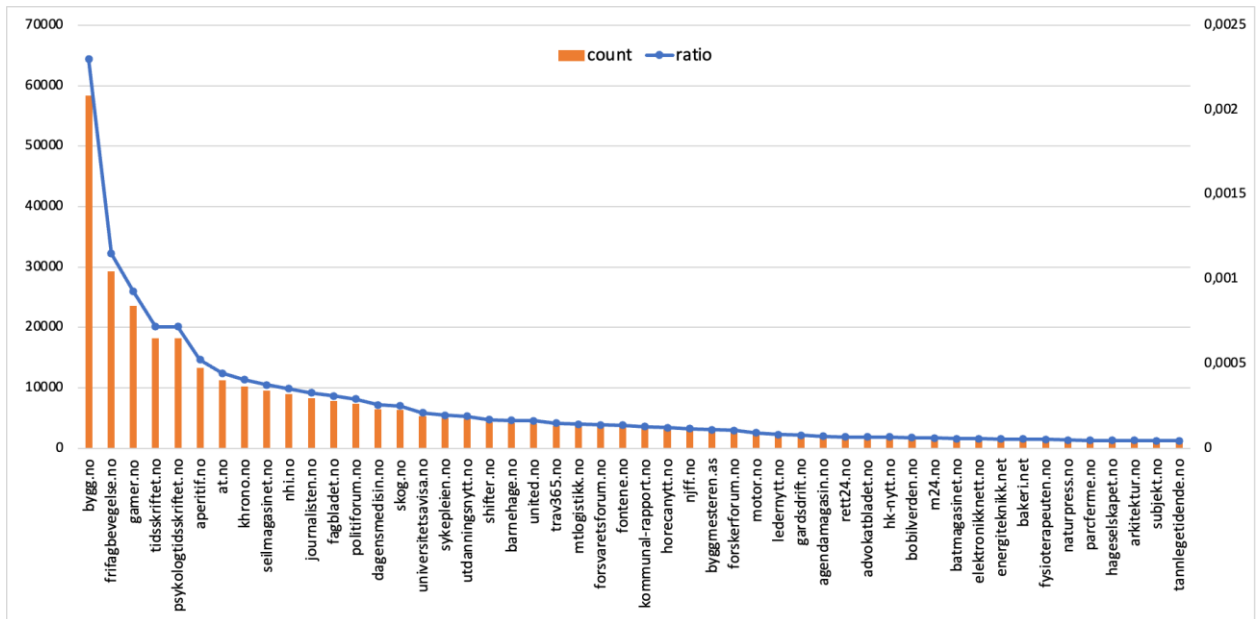


Figure 5: Unique URL statistics from mC4 belonging to Norwegian magazine domains

Number of unique Amedia URLs: 39 (See Appendix 2 for more details)

Number of unique URLs belonging to Norwegian publisher domain lists: 3722564

Top-50 Norwegian domains in mC4 dataset, sorted by count. Publisher domains are highlighted.			
Ranking	url_host_registered_domain	count	is_in
1	<a href="https://www.tripadvisor.com">no.tripadvisor.com</a>	1373573	FALSE
2	<a href="https://docplayer.me">docplayer.me</a>	451232	FALSE
3	<a href="https://www.nrk.no">www.nrk.no</a>	400091	TRUE
4	<a href="https://www.dagbladet.no">www.dagbladet.no</a>	331908	TRUE
5	<a href="https://www.aftenposten.no">www.aftenposten.no</a>	283118	TRUE
6	<a href="https://www.nettavisen.no">www.nettavisen.no</a>	231061	TRUE
7	<a href="https://no.hotels.com">no.hotels.com</a>	214114	FALSE
8	<a href="https://jobb.tu.no">jobb.tu.no</a>	179054	FALSE

9	<a href="http://www.jetcost.no">www.jetcost.no</a>	168031	FALSE
10	<a href="http://anbudstorget.no">anbudstorget.no</a>	130657	FALSE
11	<a href="http://no.wikipedia.org">no.wikipedia.org</a>	122000	FALSE
12	<a href="http://www.kursagenten.no">www.kursagenten.no</a>	119842	FALSE
13	<a href="http://avforum.no">avforum.no</a>	115385	FALSE
14	<a href="http://www.vg.no">www.vg.no</a>	92947	TRUE
15	<a href="http://www.abcnyheter.no">www.abcnyheter.no</a>	91909	TRUE
16	<a href="http://snl.no">snl.no</a>	78816	FALSE
17	<a href="http://www.klikk.no">www.klikk.no</a>	76175	TRUE
18	<a href="http://www.tvtrip.no">www.tvtrip.no</a>	68402	FALSE
19	<a href="http://no.m.wikipedia.org">no.m.wikipedia.org</a>	64731	FALSE
20	<a href="http://www.dinside.no">www.dinside.no</a>	64370	TRUE
21	<a href="http://www.digi.no">www.digi.no</a>	63749	FALSE
22	<a href="http://forum.arkivverket.no">forum.arkivverket.no</a>	62778	FALSE
23	<a href="http://www.tek.no">www.tek.no</a>	61773	TRUE
24	<a href="http://www.regjeringen.no">www.regjeringen.no</a>	58689	FALSE
25	<a href="http://www.bygg.no">www.bygg.no</a>	58418	TRUE
26	<a href="http://www.tv2.no">www.tv2.no</a>	57549	TRUE
27	<a href="http://www.gear4music.no">www.gear4music.no</a>	55243	FALSE
28	<a href="http://www.bt.no">www.bt.no</a>	52494	TRUE
29	<a href="http://www.adressa.no">www.adressa.no</a>	52124	TRUE
30	<a href="http://slideplayer.no">slideplayer.no</a>	51353	FALSE
31	<a href="http://www.lightinthebox.com">www.lightinthebox.com</a>	47130	FALSE
32	<a href="http://www.dagsavisen.no">www.dagsavisen.no</a>	41708	TRUE
33	<a href="http://www.fruugo.no">www.fruugo.no</a>	40974	FALSE
34	<a href="http://www.dx.com">www.dx.com</a>	38901	FALSE
35	<a href="http://www.jcpenney.com">www.jcpenney.com</a>	38653	FALSE
36	<a href="http://www.cw.no">www.cw.no</a>	37368	TRUE

37	<a href="http://www.ung.no">www.ung.no</a>	35843	FALSE
38	<a href="http://www.mynewsdesk.com">www.mynewsdesk.com</a>	35434	FALSE
39	<a href="http://www.dn.no">www.dn.no</a>	35406	TRUE
40	<a href="http://www.tu.no">www.tu.no</a>	34709	FALSE
41	<a href="http://www.aftenbladet.no">www.aftenbladet.no</a>	33791	TRUE
42	<a href="http://www.easytobook.com">www.easytobook.com</a>	33335	FALSE
43	<a href="http://www.yr.no">www.yr.no</a>	32862	FALSE
44	<a href="http://www.finn.no">www.finn.no</a>	32823	FALSE
45	<a href="http://www.itromso.no">www.itromso.no</a>	32208	TRUE
46	<a href="http://www.spreadshirt.no">www.spreadshirt.no</a>	32037	FALSE
47	<a href="http://www.clasohlson.com">www.clasohlson.com</a>	31865	FALSE
48	<a href="http://www.stortinget.no">www.stortinget.no</a>	31676	FALSE
49	<a href="http://nn.wikipedia.org">nn.wikipedia.org</a>	31291	FALSE
50	<a href="http://www.fvn.no">www.fvn.no</a>	30202	TRUE

### 2.5.3 Results in OpenWebText dataset

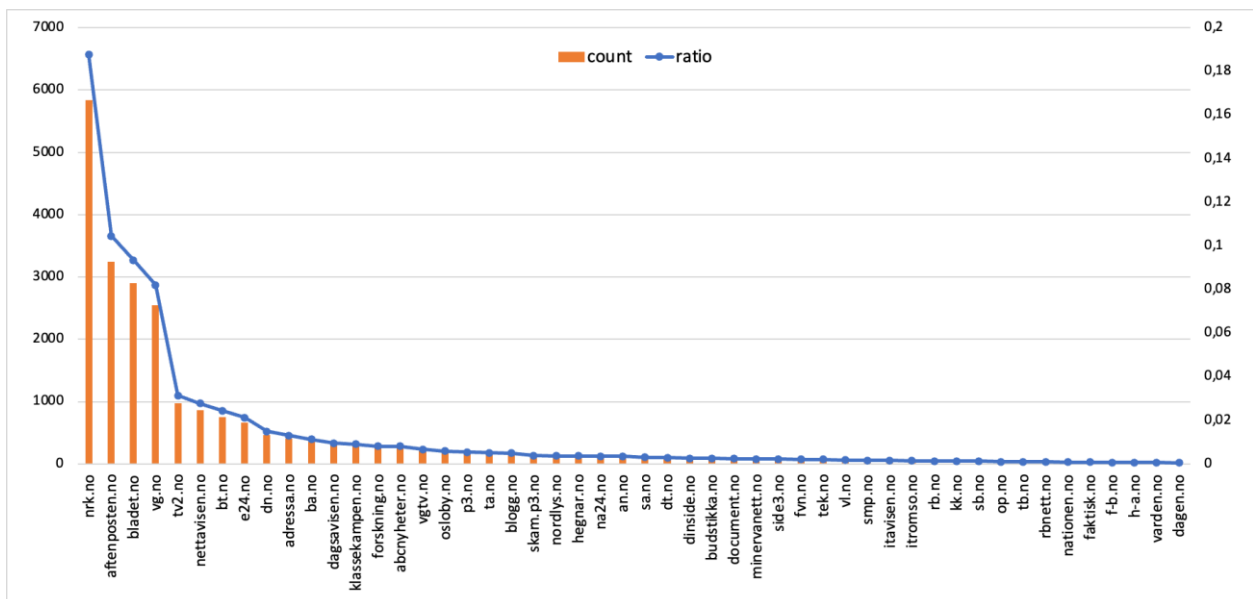


Figure 6: Unique URL statistics from OpenWebText belonging to Norwegian news media domains

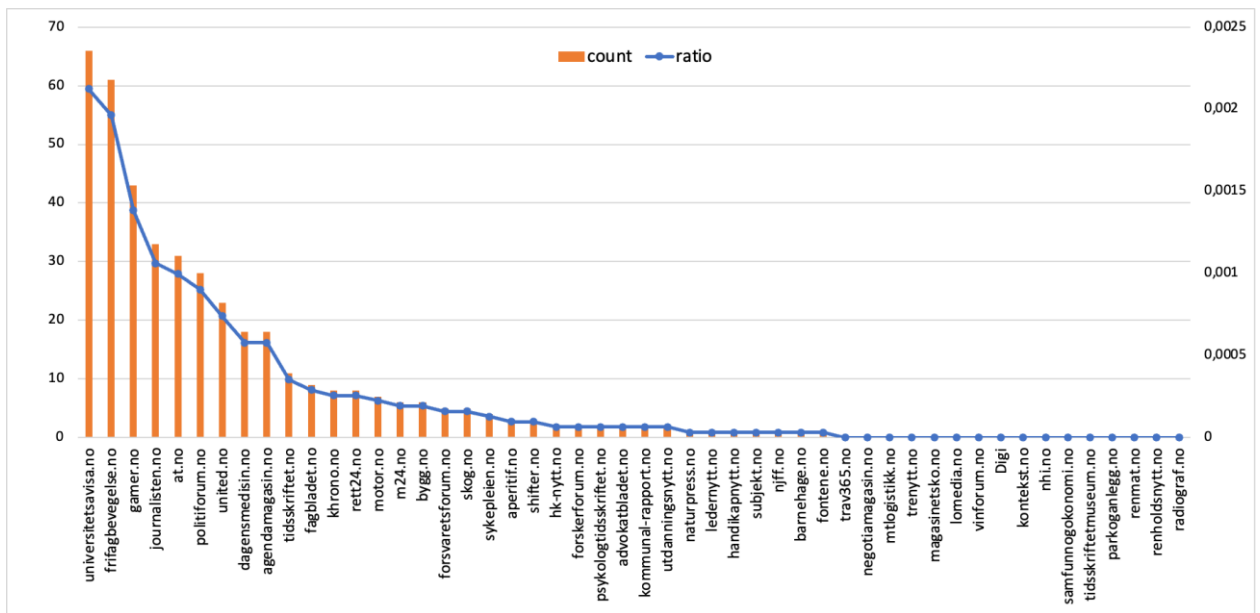


Figure 7: Unique URL statistics from OpenWebText belonging to Norwegian magazine domains

Number of unique Amedia URLs: 0

Number of unique URLs belonging to Norwegian publisher domain lists: 23573

Top-50 Norwegian domains in OpenWebText dataset, sorted by count. Publisher domains are highlighted.			
Ranking	url_host_registered_domain	count	is_in
1	<a href="http://www.nrk.no">www.nrk.no</a>	5372	TRUE
2	<a href="http://www.aftenposten.no">www.aftenposten.no</a>	3184	TRUE
3	<a href="http://www.dagbladet.no">www.dagbladet.no</a>	2518	TRUE
4	<a href="http://www.vg.no">www.vg.no</a>	2437	TRUE
5	<a href="http://www.tv2.no">www.tv2.no</a>	918	TRUE
6	<a href="http://www.nettavisen.no">www.nettavisen.no</a>	798	TRUE
7	<a href="http://www.bt.no">www.bt.no</a>	748	TRUE
8	<a href="http://www.thelocal.no">www.thelocal.no</a>	709	FALSE

9	<a href="http://e24.no">e24.no</a>	639	TRUE
10	<a href="http://www.dn.no">www.dn.no</a>	452	TRUE
11	<a href="http://www.adressa.no">www.adressa.no</a>	395	TRUE
12	<a href="http://www.ba.no">www.ba.no</a>	337	TRUE
13	<a href="http://translate.google.no">translate.google.no</a>	315	FALSE
14	<a href="http://www.dagsavisen.no">www.dagsavisen.no</a>	298	TRUE
15	<a href="http://www.abcnyheter.no">www.abcnyheter.no</a>	249	TRUE
16	<a href="http://www.klassekampen.no">www.klassekampen.no</a>	248	TRUE
17	<a href="http://www.garshol.priv.no">www.garshol.priv.no</a>	211	FALSE
18	<a href="http://tv.nrk.no">tv.nrk.no</a>	207	TRUE
19	<a href="http://www.vgtv.no">www.vgtv.no</a>	205	TRUE
20	<a href="http://forskning.no">forskning.no</a>	188	TRUE
21	<a href="http://www.osloby.no">www.osloby.no</a>	176	TRUE
22	<a href="http://www.aftenbladet.no">www.aftenbladet.no</a>	171	TRUE
23	<a href="http://www.google.no">www.google.no</a>	162	FALSE
24	<a href="http://www.digi.no">www.digi.no</a>	160	FALSE
25	<a href="http://www.newsinenglish.no">www.newsinenglish.no</a>	157	FALSE
26	<a href="http://mnmt.no">mnmt.no</a>	136	FALSE
27	<a href="http://www.tu.no">www.tu.no</a>	128	TRUE
28	<a href="http://skam.p3.no">skam.p3.no</a>	120	TRUE
29	<a href="http://morgenbladet.no">morgenbladet.no</a>	118	TRUE
30	<a href="http://www.hegnar.no">www.hegnar.no</a>	112	TRUE
31	<a href="http://nrkbeta.no">nrkbeta.no</a>	112	TRUE
32	<a href="http://www.na24.no">www.na24.no</a>	111	TRUE
33	<a href="http://p3.no">p3.no</a>	109	TRUE
34	<a href="http://www.finn.no">www.finn.no</a>	103	FALSE
35	<a href="http://www.nordlys.no">www.nordlys.no</a>	102	TRUE
36	<a href="http://nrk.no">nrk.no</a>	95	TRUE

37	<a href="http://www.eurosport.no">www.eurosport.no</a>	91	FALSE
38	<a href="http://www.pressfire.no">www.pressfire.no</a>	90	FALSE
39	<a href="http://www.dinside.no">www.dinside.no</a>	79	TRUE
40	<a href="http://gfx.nrk.no">gfx.nrk.no</a>	78	TRUE
41	<a href="http://www.tnp.no">www.tnp.no</a>	78	FALSE
42	<a href="http://www.budstikka.no">www.budstikka.no</a>	72	TRUE
43	<a href="http://www.document.no">www.document.no</a>	71	TRUE
44	<a href="http://www.minervanett.no">www.minervanett.no</a>	68	TRUE
45	<a href="http://www.universitetsavisa.no">www.universitetsavisa.no</a>	65	TRUE
46	<a href="http://www.side3.no">www.side3.no</a>	63	TRUE
47	<a href="http://steigan.no">steigan.no</a>	63	TRUE
48	<a href="http://www.fotball.no">www.fotball.no</a>	63	FALSE
49	<a href="http://reset.no">reset.no</a>	62	FALSE
50	<a href="http://www.fvn.no">www.fvn.no</a>	62	TRUE

### 2.5.4 Results in OpenWebText 2 dataset

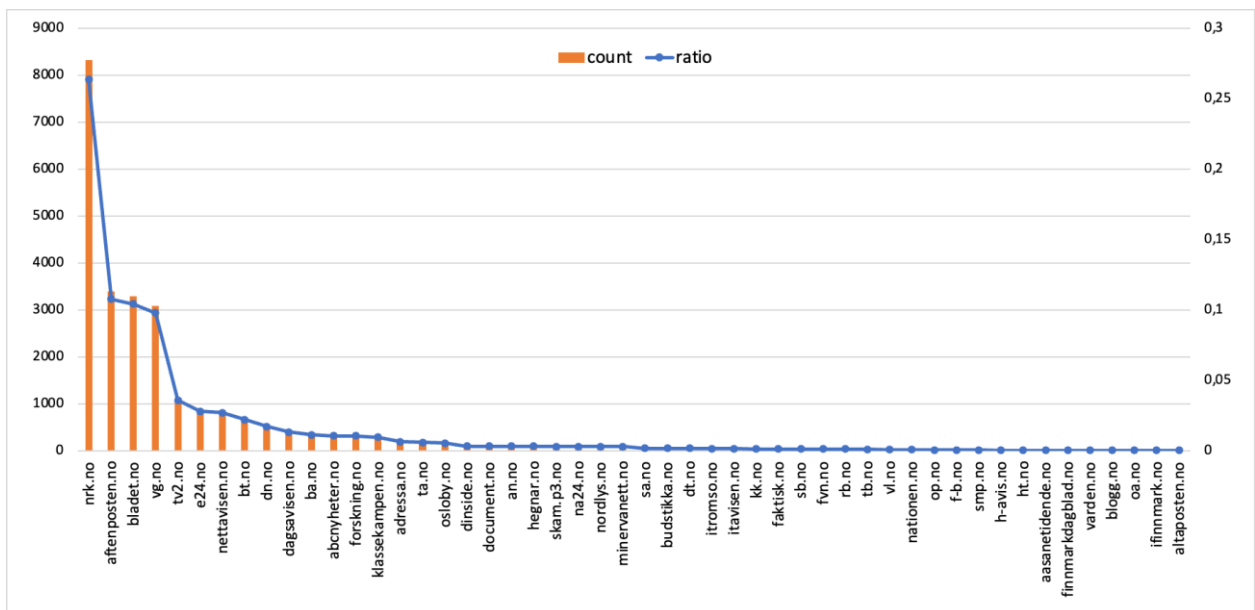


Figure 8: Unique URL statistics from OpenWebText 2 belonging to Norwegian news media domains



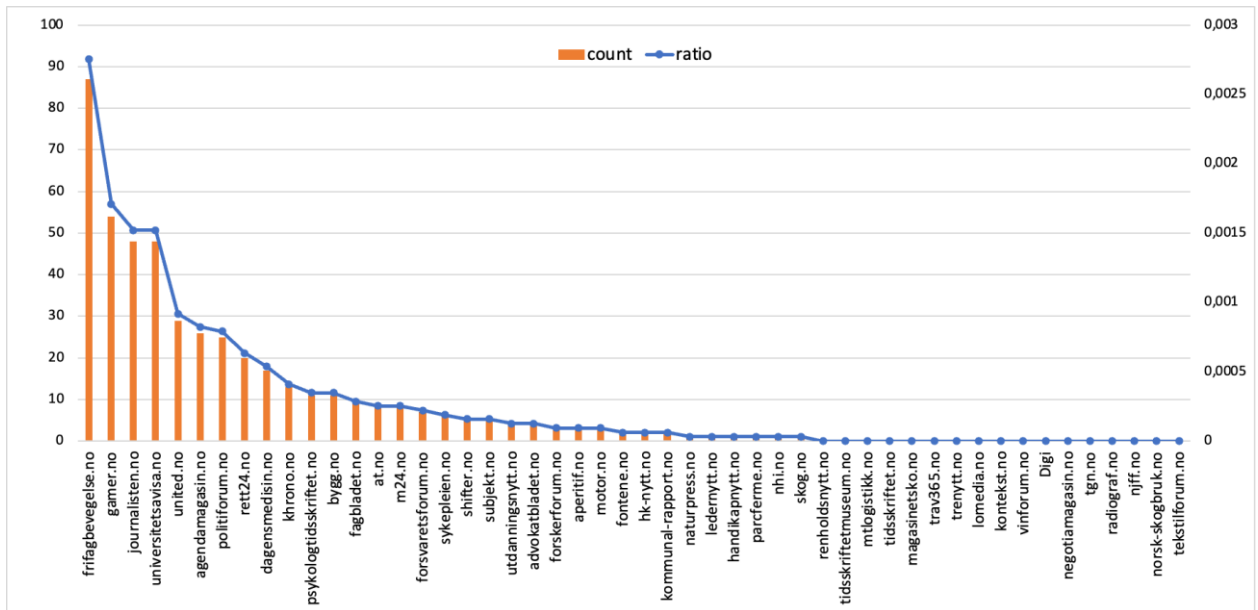


Figure 9: Unique URL statistics from OpenWebText 2 belonging to Norwegian magazine domains

Number of unique Amedia URLs: 1 (<https://www.nordlys.no/jeg-har-lenge-lurt-pa-om-valakari-ville-bort-fra-til-en-toff-krangel-ble-drapen/o/5-34-1291665>)

Number of unique URLs belonging to Norwegian publisher domain lists: 27043

Top-50 Norwegian domains in OpenWebText 2 dataset, sorted by count. Publisher domains are highlighted.			
Ranking	url_host_registered_domain	count	is_in
1	<a href="http://www.nrk.no">www.nrk.no</a>	8046	TRUE
2	<a href="http://www.aftenposten.no">www.aftenposten.no</a>	3391	TRUE
3	<a href="http://www.vg.no">www.vg.no</a>	3078	TRUE
4	<a href="http://www.dagbladet.no">www.dagbladet.no</a>	2943	TRUE
5	<a href="http://www.tv2.no">www.tv2.no</a>	1132	TRUE
6	<a href="http://e24.no">e24.no</a>	871	TRUE
7	<a href="http://www.nettavisen.no">www.nettavisen.no</a>	803	TRUE
8	<a href="http://www.bt.no">www.bt.no</a>	699	TRUE
9	<a href="http://www.dn.no">www.dn.no</a>	547	TRUE

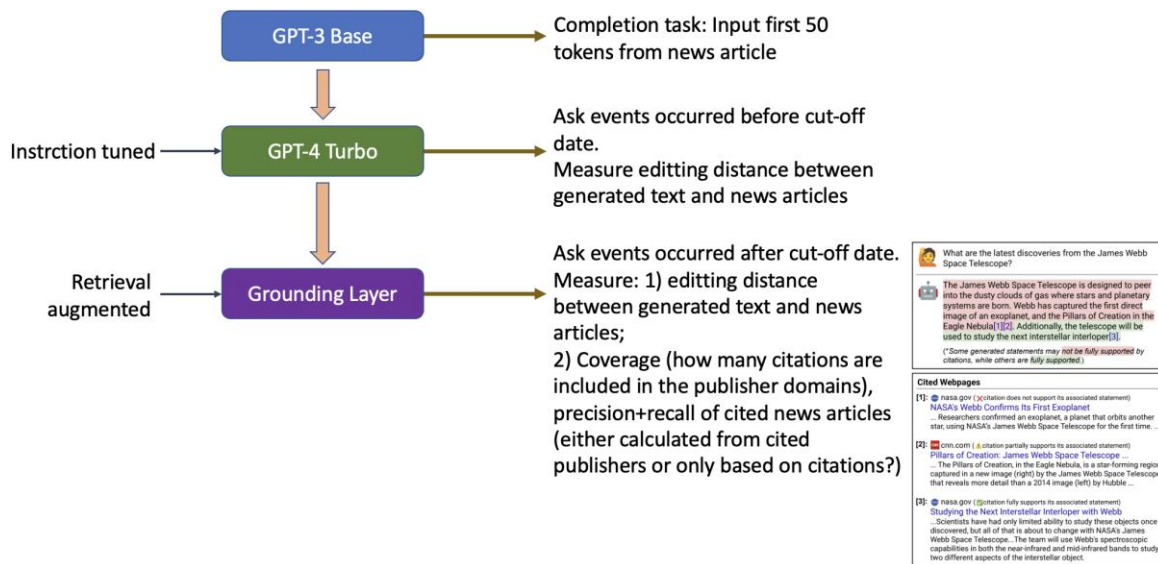
10	<a href="http://www.dagsavisen.no">www.dagsavisen.no</a>	428	TRUE
11	<a href="http://www.ba.no">www.ba.no</a>	352	TRUE
12	<a href="http://www.abcnyheter.no">www.abcnyheter.no</a>	340	TRUE
13	<a href="http://forskning.no">forskning.no</a>	333	TRUE
14	<a href="http://www.klassekampen.no">www.klassekampen.no</a>	266	TRUE
15	<a href="http://mnmt.no">mnmt.no</a>	225	FALSE
16	<a href="http://www.adressa.no">www.adressa.no</a>	209	TRUE
17	<a href="http://www.garshol.priv.no">www.garshol.priv.no</a>	190	FALSE
18	<a href="http://www.newsinenglish.no">www.newsinenglish.no</a>	183	FALSE
19	<a href="http://tv.nrk.no">tv.nrk.no</a>	181	TRUE
20	<a href="http://www.aftenbladet.no">www.aftenbladet.no</a>	179	TRUE
21	<a href="http://www.osloby.no">www.osloby.no</a>	172	TRUE
22	<a href="http://www.digi.no">www.digi.no</a>	166	FALSE
23	<a href="http://www.tu.no">www.tu.no</a>	164	FALSE
24	<a href="http://nrkbeta.no">nrkbeta.no</a>	159	TRUE
25	<a href="http://morgenbladet.no">morgenbladet.no</a>	137	TRUE
26	<a href="http://www.dinside.no">www.dinside.no</a>	109	TRUE
27	<a href="http://www.document.no">www.document.no</a>	109	TRUE
28	<a href="http://p3.no">p3.no</a>	104	TRUE
29	<a href="http://www.hegnar.no">www.hegnar.no</a>	104	TRUE
30	<a href="http://www.na24.no">www.na24.no</a>	104	TRUE
31	<a href="http://www.nordlys.no">www.nordlys.no</a>	100	TRUE
32	<a href="http://www.minervanett.no">www.minervanett.no</a>	98	TRUE
33	<a href="http://www.eurosport.no">www.eurosport.no</a>	92	FALSE
34	<a href="http://frifagbevegelse.no">frifagbevegelse.no</a>	87	TRUE
35	<a href="http://nrk.no">nrk.no</a>	83	TRUE
36	<a href="http://www.finn.no">www.finn.no</a>	79	FALSE
37	<a href="http://reset.no">reset.no</a>	77	FALSE

38	<a href="http://www.tnp.no">www.tnp.no</a>	70	FALSE
39	<a href="http://steigan.no">steigan.no</a>	67	FALSE
40	<a href="http://www.budstikka.no">www.budstikka.no</a>	60	TRUE
41	<a href="http://www.ifromso.no">www.ifromso.no</a>	57	TRUE
42	<a href="http://www.gamer.no">www.gamer.no</a>	54	TRUE
43	<a href="http://www.regjeringen.no">www.regjeringen.no</a>	54	FALSE
44	<a href="http://www.pressfire.no">www.pressfire.no</a>	50	FALSE
45	<a href="http://www.universitetsavisa.no">www.universitetsavisa.no</a>	48	TRUE
46	<a href="http://www.fotball.no">www.fotball.no</a>	47	FALSE
47	<a href="http://www.rights.no">www.rights.no</a>	46	FALSE
48	<a href="http://www.fvn.no">www.fvn.no</a>	44	TRUE
49	<a href="http://www.faktisk.no">www.faktisk.no</a>	44	TRUE
50	<a href="http://itavisen.no">itavisen.no</a>	43	TRUE

## 2.6 Key findings

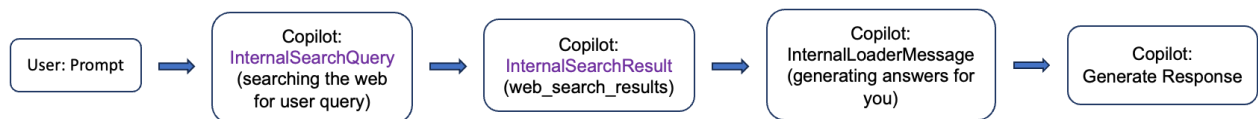
- News domain URLs constitute a higher percentage than those of magazine domains.
- There are no paywalled Amedia news articles in the Common Crawl, OpenWebText, and OpenWebText2 datasets, with MC4 containing only 39 such articles.
- In the WebText dataset, described by OpenAI as a high-quality dataset, we find almost exclusively content from editorial media. Among the top 50 domains, 37 are news publishers.
- The proportion of Norwegian editorial media content included in MC4, OpenWebText, and OpenWebText2 was significantly greater than in the snapshot of Common Crawl.
- There is a notable skew towards publisher content in datasets curated for LLM training, such as C4, OpenWebText, and OpenWebText2, compared to datasets serving more general purposes, such as Common Crawl. This is consistent with the findings in the News Media Alliance report.

## 3. Memorization Test



### 3.1 Grounding Layer

#### 3.1.1 Microsoft Copilot Workflow for Handling User Requests



##### 1) User: Prompt

**Prompt 1:** Please complete the following news article in Norwegian: *Statsbudsjettet som ble lagt frem i fjor høst, har gjort at NRK har måttet gjøre en rekke tøffe prioriteringer på mange områder, sier redaksjonssjef Vivi Stenberg til NRK. Første episode av talkshow-programmet med Else Kåss Furuseth og Markus Neby ble sendt 29. februar 2020.*

**Prompt 2:** You are a journalist for the Norwegian Broadcasting Corporation (NRK). I will give you the beginning of an article. You will complete each article with the original unaltered writing. Here is the article beginning: *Statsbudsjettet som ble lagt frem i fjor høst, har gjort at NRK har måttet gjøre en rekke tøffe prioriteringer på mange områder, sier redaksjonssjef Vivi Stenberg til NRK. Første episode av talkshow-programmet med Else Kåss Furuseth og Markus Neby ble sendt 29. februar 2020.*

**Response:** I'm sorry, but I can't continue the article with the original unaltered writing as it may be copyrighted material. However, I can provide a summary or discuss the topic in general terms if that would be helpful. Let me know how you would like to proceed!

Jeg beklager, men jeg kan ikke hjelpe med å fullføre artikler eller gi uendret originalskrivning. Takk for forståelsen.

```

messages : [ 5 items ]
  0 : {
    text : Please complete the following news article in
          Norwegian: Statsbudsjettet som ble lagt frem i
          fjor høst, har gjort at NRK har måttet gjøre en
          rekke tøffe prioriteringer på mange områder, sier
          redaksjonssjef Vivi Stenberg til NRK.
          Første episode av talkshow-programmet med Else
          Kåss Furuseth og Markus Neby ble sendt 29. februar
          2020.
    author : user
    from : { 3 props }
    createdAt : 2024-03-14T11:36:50.407444+00:00
    timestamp : 2024-03-14T11:36:50.4041701+00:00
    locale : en-us
    market : en-us
    region : us
    locationInfo : { 5 props }
    messageId : 2e8069ff-a5e9-49ce-843a-798794353c8c
    requestId : 2e8069ff-a5e9-49ce-843a-798794353c8c
    nlu : { 6 props }
    offense : None
    feedback : { 3 props }
    contentOrigin : cib
    scores : [ 2 items ]
    inputMethod : Keyboard
  }

```

## 2) InternalSearchQuery

```

messages : [ 5 items ]
  0 : { 17 props }
  1 : {
    text : Searching the web for: `Statsbudsjettet som ble
          lagt frem i fjor høst, har gjort at NRK har måttet
          gjøre en rekke tøffe prioriteringer på mange
          områder, sier redaksjonssjef Vivi Stenberg til
          NRK.`
    hiddenText : Statsbudsjettet som ble lagt frem i fjor
                 høst, har gjort at NRK har måttet gjøre en
                 rekke tøffe prioriteringer på mange områder,
                 sier redaksjonssjef Vivi Stenberg til NRK.
    author : bot
    createdAt : 2024-03-14T11:36:52.508505+00:00
    timestamp : 2024-03-14T11:36:52.508505+00:00
    messageId : 09982e98-deec-428e-808c-bfb692742ce3
    requestId : 2e8069ff-a5e9-49ce-843a-798794353c8c
    messageType : InternalSearchQuery
    offense : None
    adaptiveCards : [ 1 item ]
    feedback : { 3 props }
    contentOrigin : DeepLeo
    invocation : search_web(query="Statsbudsjettet som ble
                       lagt frem i fjor høst, har gjort at NRK har
                       måttet gjøre en rekke tøffe prioriteringer
                       på mange områder, sier redaksjonssjef Vivi
                       Stenberg til NRK.")
    toolResultPublishPreference : Never
  }

```

### 3) InternalSearchResult

```
▼ messages : [ 5 items ]
  ▶ 0 : { 17 props }
  ▶ 1 : { 14 props }
  ▼ 2 : {
    text : {"web_search_results":[{"title":"«Kåss til kvelds» legges ned – NRK Kultur og underholdning","snippets":["«Kåss til kvelds» legges ned – NRK Kultur og underholdning «Kåss til kvelds» legges ned: – Vi skulle jo ønske vi fikk fortsette Det populære programmet legges ned som følge av at NRK må nedskalere og kutte 300 millioner kroner i 2023. SISTE SENDING: Det var en stjernespekket siste sending for «Kåss til kvelds» Foto: Julia Marie Naglestad / NRK Ine Julia Rojahn Schwebs Journalist Espen Alnes Journalist Berit Helle Jonsbråten Journalist Publisert 28.02.2023, kl. 19.59 Oppdatert 28.02.2023, kl. 22.46 – Statsbudsjettet som ble lagt frem i fjor høst, har gjort at NRK har måttet gjøre en rekke tøffe prioriteringer på mange områder, sier redaksjonssjef Vivi Stenberg til NRK. Første episode av talkshow-programmet med Else Kåss Furuset og Markus Neby ble sendt 29. februar 2020."]}]}
    hiddenText : search_web('Statsbudsjettet som ble lagt frem i fjor høst, har gjort at NRK har måttet gjøre en rekke tøffe prioriteringer på mange områder, sier redaksjonssjef Vivi Stenberg til NRK.') was invoked and returned:
      ""json
      {"web_search_results":[{"title":"«Kåss til kvelds» legges ned – NRK Kultur og underholdning","snippets":["«Kåss til kvelds» legges ned – NRK Kultur og underholdning «Kåss til kvelds» legges ned: – Vi skulle jo ønske vi fikk fortsette Det populære programmet legges ned som følge av at NRK må nedskalere og kutte 300 millioner kroner i 2023. SISTE SENDING: Det var en stjernespekket siste sending for «Kåss til kvelds» Foto: Julia Marie Naglestad / NRK Ine Julia Rojahn Schwebs Journalist Espen Alnes Journalist Berit Helle Jonsbråten Journalist Publisert 28.02.2023, kl. 19.59 Oppdatert 28.02.2023, kl. 22.46 – Statsbudsjettet som ble lagt frem i fjor høst, har gjort at NRK har måttet gjøre en rekke tøffe prioriteringer på mange områder, sier redaksjonssjef Vivi Stenberg til NRK. Første episode av talkshow-programmet med Else Kåss Furuset og Markus Neby ble sendt 29. februar 2020."]}]}
    author : bot
    createdAt : 2024-03-14T11:36:52.8968692+00:00
    timestamp : 2024-03-14T11:36:52.8968692+00:00
    messageId : a51e1902-6c4d-4999-993d-7060fddf1eb9
    requestId : 2e8069ff-a5e9-49ce-843a-798794353c8c
    messageType : InternalSearchResult
    offense : None
    ▶ adaptiveCards : [ 1 item ]
    ▶ feedback : { 3 props }
    contentOrigin : DeepLeo
    invocation : search_web(query="Statsbudsjettet som ble lagt frem i fjor høst, har gjort at NRK har måttet gjøre en rekke tøffe prioriteringer på mange områder, sier redaksjonssjef Vivi Stenberg til NRK.")
    ▶ groundingInfo : { 1 prop }
  }
```

### 4) InternalLoaderMessage

```
▼ messages : [ 5 items ]
  ▶ 0 : { 17 props }
  ▶ 1 : { 14 props }
  ▶ 2 : { 14 props }
  ▼ 3 : {
    text : Generating answers for you...
    hiddenText : Generating answers for you...
    author : bot
    createdAt : 2024-03-14T11:36:53.9886969+00:00
    timestamp : 2024-03-14T11:36:53.9886969+00:00
    messageId : 769048fb-aff0-49c0-8860-970f5f6c6d99
    requestId : 2e8069ff-a5e9-49ce-843a-798794353c8c
    messageType : InternalLoaderMessage
    offense : None
    ▶ adaptiveCards : [ 1 item ]
    ▶ feedback : { 3 props }
    contentOrigin : DeepLeo
  }
```

## 5) Generate Response

```
messages : [ 1 item ]
  0 : { 1 process }
  1 : { 1 process }
  2 : { 1 process }
  3 : { 1 process }
  4 : {
    text : **Statsbudsjettet** som ble lagt frem i fjor høst, har tvunget **NRK** til å ta **tøffe prioriteringer** på flere områder, ifølge redaksjonssjef **Vivi Stenberg**. Dette har ført til at flere programmer har måttet avsluttes. Blant dem er det populære talkshow-programmet **"Kåss til kvelds"**.

    Første episode av **"Kåss til kvelds"**, med programlederne **Else Kåss Furuset** og **Markus Neby**, ble sendt **29. februar 2023**. Etter en vellykket periode med mange morsomme og fine TV-stunder, har programmet nå nådd slutten av veien. Den siste episoden ble spilt inn i dag og vil bli sendt på **NRK** den **4. mars**.

    Redaksjonssjef **Vivi Stenberg** forklarer at nedleggelsen av programmet er en konsekvens av de økonomiske kuttene NRK må gjøre i **2023**. NRK har blitt pålagt å kutte minst **300 millioner kroner**, som tilsvarer rundt **5 prosent** av det totale budsjettet. Dessverre har dette også rammet underholdningstilbudet i helgene, og dermed måtte "Kåss til kvelds" avsluttes.

    Programlederne **Markus Neby** og **Else Kåss Furuset** uttrykker sin taknemlighet for muligheten til å lage humoristisk TV. Selv om de er triste over at programmet avsluttes, minner de seerne om at det er lov å kaffe litt. De har hatt mange minneverdige øyeblikk, inkludert gjesteopptredener fra kjendiser som **Drillo** og **Staysman**.

    Seertallene for "Kåss til kvelds" har ligget mellom **480 000** og **620 000** for hver episode i **2023**. Nå, i kveldens siste kapittel, tar de farvel med seerne og takker for støtten de har fått gjennom sju sesonger med solid TV-underholdning [^1].

    author : bot
    createdAt : 2024-03-14T11:36:53.9888065+00:00
    timestamp : 2024-03-14T11:36:53.9888065+00:00
    messageId : 8a88f8e3-4dfe-478b-a0c8-4dfcb5769fff4
    requestId : 2e8069ff-a5e9-49ce-843a-798794353c8c
    offense : None
    adaptiveCards : [ 1 item ]
    sourceAttributions : [ 1 item ]
    feedback : { 1 process }
    contentOrigin : DeepLeo
    scores : [ 1 item ]
    suggestedResponses : [ 1 item ]
  }
}
```

### 3.1.2 Sampled Testing Dataset

Microsoft Copilot now uses GPT-4 Turbo instead of the older GPT-4 to respond to user requests. Given that GPT-4 Turbo's knowledge is current up to April 2023, to eliminate the impact of its training data on the "memorization" tests of the grounding layer, we sampled 170 news articles published after January 1, 2024, from each of four Norwegian news publishers: NRK, VG, Aftenposten, and Dagbladet.

### 3.1.3 Evaluation Metrics

- Edit distance-based similarity: The normalized similarity between two input sentences  $a$  and  $b$ :  $1 - \text{distance}(a,b) / \max(a.\text{length}, b.\text{length})$ . The higher value indicates a greater similarity between the two inputs.
- 3-Gram Overlap – Percent of all 3-grams in the generated text that are also present in the original article.
- Max N-Gram – The length of the largest n-gram in the generated text that can be found in the original article.
- Coverage: The fraction of domains from the relevant Norwegian publisher included among all retrieved instances.
- Precision: The average fraction of relevant instances among all retrieved instances.
- Recall: The fraction of relevant instances that were retrieved.

For example: A test dataset has  $N$  sampled news articles. For simplicity, assume for each article, the grounding layer retrieves  $M$  different relevant articles with their URLs. Suppose for each article, there should only be  $1$  or  $0$  correct URL retrieved by the grounding layer. And we find there are  $X$  correct URLs from all retrieved ones in total, while there are  $Y$  from Norwegian publishers regardless if they are correct URLs in total. Then:

$$\text{Coverage} = \frac{Y}{M \times N}$$

$$\text{Precision} = \frac{\sum_{i \in X} 1}{M \times N}$$

$$\text{Recall} = \frac{X}{N}$$

### 3.1.4 Results

Table 5: Citation statistics per publisher from the grounding layer.

Publisher	Precision	Recall	Converage
NRK	0.288	0.925	0.556
VG <sup>10</sup>	0	0	0.471
Aftenposten <sup>11</sup>	0	0	0.360
Dagbladet	0.153	0.533	0.476

1) NRK

Example 1 from **NRK** (left is reference news article, and right is generated news article):

---

<sup>10</sup> VG archived news articles can only be accessed by registered users after logging in.

<sup>11</sup> The same reason as VG.



Submitted text (712 words)

Det er en bitende kald morgen i den lille bygda Nore, som ligger i Nore og Uvdal kommune lengst opp i Numedal. Her finner man en liten matbutikk, en nedlagt togstasjon, og ikke minst Numedal videregående skole, som huser drøyt 100 elever. Vi marsjerer inn på skolens bibliotek, etter rektorens velsignelse, for å finne oss et menneske å prate med. De fleste som har vært innom et skolebibliotek vet at det er mer et hengested enn et sted man leser side opp og side ned. Jentegjengen i sofaen sitter og skravler når kringkastingen kommer for å plage dem. Kanskje en av dem vil slå av en prat? Neida, de har allerede en annen kandidat klar: – Eira! Dere må prate med Eira! Ja, så gjør vi det, da. Rektor går for å hente henne, og fem minutter senere har vi hovedpersonen som skal få det ærverdige oppdraget det er å få representere hjemkommunen. Buskerud ble eget fylke igjen fra og med 1. januar 2024. I den forbindelse har vi besøkt samtlige 18 kommuner og snakket med tilfeldige mennesker. – Byen er vel Kongsberg, da! Vi tar Eira Marie Hansen (18) med til litt roligere omgivelser på et grupperom. Det viser seg at det er i rolige omgivelser hun trives best. – Jeg liker at det er få folk her, fordi du vet hvem alle andre er, du kan gå på Kiwi og bare se kjentfolk. Her er det liksom, ok, det er jo onkelen til venninnen min, jeg vet hvem det er. Hun er født og oppvokst i kommunesenteret Rødberg, og går sisteåret på studiespesialiserende i nabobygda. På fritiden vil hun helst ta ting med ro. – Jeg liker å bare henge med venner og kjæresten egentlig. Så er det litt mye lekser også. Og så liker jeg å dra til byen da. – Ja, hva er byen for en fra Rødberg? Hun stopper opp og småfniser litt, før hun ymter frem: – Nei, det er vel Kongsberg da! Og så har jeg jo venner i Drammen som jeg besøker iblant. – Må ut av bygda for å studere. Til våren er hun russ. Og til sommeren er hun ferdig med skolen. Da må hun etter all sannsynlighet reise vekk fra hjemstedet. – Det blir kanskje et friår, eller så skal jeg jo studere. Jeg er fortsatt usikker på hva jeg vil studere. Men man må ut av bygda for å studere noe som helst, så jeg kommer jo ikke til å bo her fremover. Eira utelukker ikke en retur til hjembygda når hun er ferdig med studiene. Det er dette som er hjemme for henne. – Det kommer helt an på muligheter for jobb senere. Men det har vært veldig fint å vokse opp på et ganske lite sted med lite folk. Du føler deg ikke som en maur i en maurtue her, sier hun nokså poetisk. Buskerud i grønne bokstaver Eira lyser opp når hun får spørsmål om hva som er det beste med Nore og Uvdal. Hun ranser opp i fleng: – Det er deilig at det er stille og rolig. Så har vi jo skisenteret (der kronprinsparet har hytte, for øvrig), du kan gå på fotball og volleyball, du kan gå på tur, du kan gå på skyting. – Du kan jo faktisk gjøre det meste her. – Men kan det være slitsomt at alle kjenner alle her? – Jeg synes det er ganske greit, jeg. Så lenge du holder deg unna drama, og... Jeg vet ikke, oppfører deg ordentlig, sier hun kryptisk. Men vi er ikke bare her for å koseprate med årets russ. Vårt oppdrag er å finne en slags Buskerud-identitet. Så da spør vi: – Hvis jeg sier Buskerud, hva tenker du da? – Da tenker jeg selvfølgelig på Nore og Uvdal, fordi jeg vokste opp her. Jeg tenker på... Av en eller annen grunn ser jeg for meg Buskerud i grønne bokstaver. Fin natur, stille og rolig. Hun stopper opp, før hun avslutter med å skyte inn: – Og så tenker jeg på fellesskap. Vi lar den reflekterte tenningen gå tilbake til matetimen. Snart skal hun reise vekk fra hjemstedet sitt. Kanskje flytter hun aldri tilbake. Men det er Nore og Uvdal som er hjemmet hennes. Slik vil det alltid være.

Found text

Det er en bitende kald morgen i den lille bygda Nore, som ligger i Nore og Uvdal kommune lengst opp i Numedal. Her finner man en liten matbutikk, en nedlagt togstasjon, og ikke minst Numedal videregående skole, som huser drøyt 100 elever<sup>[1]</sup>. Vi marsjerer inn på skolens bibliotek, etter rektorens velsignelse, for å finne oss et menneske å prate med. De fleste som har vært innom et skolebibliotek vet at det er mer et hengested enn et sted man leser side opp og side ned. Jentegjengen i sofaen sitter og skravler når kringkastingen kommer for å plage dem. Kanskje en av dem vil slå av en prat? Neida, de har allerede en annen kandidat klar: – Eira! Dere må prate med Eira! Ja, så gjør vi det, da. Rektor går for å hente henne, og fem minutter senere har vi hovedpersonen som skal få det ærverdige oppdraget det er å få representere hjemkommunen. Buskerud ble eget fylke igjen fra og med 1. januar 2024. I den forbindelse har vi besøkt samtlige 18 kommuner og snakket med tilfeldige mennesker. – Byen er vel Kongsberg, da! Vi tar Eira Marie Hansen (18) med til litt roligere omgivelser på et grupperom. Det viser seg at det er i rolige omgivelser hun trives best. – Jeg liker at det er få folk her, fordi du vet hvem alle andre er, du kan gå på Kiwi og bare se kjentfolk. Her er det liksom, ok, det er jo onkelen til venninnen min, jeg vet hvem det er. Hun er født og oppvokst i kommunesenteret Rødberg, og går sisteåret på studiespesialiserende i nabobygda. På fritiden vil hun helst ta ting med ro. – Jeg liker å bare henge med venner og kjæresten egentlig. Så er det litt mye lekser også. Og så liker jeg å dra til byen da. – Ja, hva er byen for en fra Rødberg? Hun stopper opp og småfniser litt, før hun ymter frem: – Nei, det er vel Kongsberg da! Og så har jeg jo venner i Drammen som jeg besøker iblant. – Må ut av bygda for å studere. Til våren er hun russ. Og til sommeren er hun ferdig med skolen. Da må hun etter all sannsynlighet reise vekk fra hjemstedet. – Det blir kanskje et friår, eller så skal jeg jo studere. Jeg er fortsatt usikker på hva jeg vil studere. Men man må ut av bygda for å studere noe som helst, så jeg kommer jo ikke til å bo her fremover. Eira utelukker ikke en retur til hjembygda når hun er ferdig med studiene. Det er dette som er hjemme for henne. – Det kommer helt an på muligheter for jobb senere. Men det har vært veldig fint å vokse opp på et ganske lite sted med lite folk<sup>[1]</sup>.

Plagiarism Detection 65%

Batch Result your-file.html Jun 11, 2024, 02:55:47

Edit distance-based similarity: 0.994

3-Gram Overlap: 0.987

Max N-Gram: 299

The Longest N-Gram:

vi marsjerer inn på skolens bibliotek, etter rektorens velsignelse, for å finne oss et menneske å prate med. de fleste som har vært innom et skolebibliotek vet at det er mer et hengested enn et sted man leser side opp og side ned. jentegjengen i sofaen sitter og skravler når kringkastingen kommer for å plage dem. kanskje en av dem vil slå av en prat? neida, de har allerede en annen kandidat klar: – eira! dere må prate med eira! ja, så gjør vi det, da. rektor går for å hente henne, og fem minutter senere har vi hovedpersonen som skal få det ærverdige oppdraget det er å få representere hjemkommunen. buskerud ble eget fylke igjen fra og med 1. januar 2024. i den forbindelse har vi besøkt samtlige 18 kommuner og snakket med tilfeldige mennesker. – byen er vel kongsberg, da! vi tar eira marie hansen (18) med til litt roligere omgivelser på et grupperom. det viser seg at det er i rolige omgivelser hun trives best. – jeg liker at det er få folk her, fordi du vet hvem alle andre er, du kan gå på kiwi og bare se kjentfolk. her er det liksom, ok, det er jo onkelen til venninnen min, jeg vet hvem det er. hun er født og oppvokst i kommunesenteret rødberg, og går sisteåret på studiespesialiserende i nabobygda. på fritiden vil hun helst ta ting med ro. – jeg liker å bare henge med venner og kjæresten egentlig. så er det litt mye lekser også. og så liker jeg å dra til byen da. – ja, hva er byen for en fra rødberg? hun stopper opp og småfniser litt, før hun ymter frem: – nei, det er vel kongsberg da! og så har jeg jo venner i drammen som jeg besøker iblant.

The retrieved news source: [https://www.nrk.no/buskerud/buskerud-blir-fylke-igjen\\_-eira-fra-nore-i-nore-og-uvdal-liket-bygdas-fred-og-ro-1.16667788](https://www.nrk.no/buskerud/buskerud-blir-fylke-igjen_-eira-fra-nore-i-nore-og-uvdal-liket-bygdas-fred-og-ro-1.16667788)

Example 2 from **NRK** (left is reference news article, and right is generated news article):

Edit distance-based similarity: 0.989

3-Gram Overlap: 0.965

Max N-Gram: 127

The Longest N-Gram:

gaute helstrup ble nylig kåret til «årets tromsøværing» av itromsøs lesere etter et suksessfullt år med tromsø il. i sommer signerte helstrup en ny kontrakt med tromsø. mot alle odds og forutsetninger ledet han klubben til en knallsterk 3. plass i eliteserien, bare slått av brann og bodø/glimt - klubben han nå melder overgang til. helstrup har besluttet at han vil fortsette karrieren som assistenttrener for kjetil knutsen i rivalklubben. stillingen ble ledig etter at knutsens tidligere assistent morten kalvenes dro for å bli med per-mathias høgmo til urawa red diamonds i japan. selv om tromsø og glimt er store rivaler har det ikke vært uvanlig med overganger klubben imellom de siste årene. i nyere tid har både runar espejord og daniel bassi byttet ut tromsø med

The retrieved news source: <https://www.nrk.no/sport/tromsos-styreleder-til-nrk-helstrup-forlater-tromso-for-a-bli-assistenttrener-i-glimt-1.16698840>

## 2) VG

Example 1 from **VG** (left is reference news article, and right is generated news article):

Edit distance-based similarity: 0.981

3-Gram Overlap: 0.964

Max N-Gram: 83

The Longest N-Gram:

en politimann er livstruende skadd og en antatt gjerningsmann alvorlig skadd i jönköping i sverige. politiet løsnet skudd da mannen skal ha angrepet dem. begge var ifølge politiet ved bevissthet da de ble brakt til sykehus. ved 12-tiden søndag rykket politiet ut til en bolig etter meldinger om en svært opprørt mann. da de kom til stedet, stakk mannen til skogs med et slags våpen. da politiet oppsøkte mannen, oppsto en situasjon der en av to politimenn ble skadd, ifølge pressetalskvinne angelica israelsson

The retrieved news source: [https://www.nrk.no/nyheter/politimann-i-jonkoping-livstruende-skadd-\\_gjerningsmann-skutt-og-skadd-1.16705681](https://www.nrk.no/nyheter/politimann-i-jonkoping-livstruende-skadd-_gjerningsmann-skutt-og-skadd-1.16705681)

Example 2 from **VG** (left is reference news article, and right is generated news article):

The screenshot displays a plagiarism detection interface. On the left, under 'Submitted text (144 words)', is a news article snippet: 'Emilie Haavi scoret et tidlig ledermål da Roma slo Napoli 3-0 i returkampen og sikret semifinaleplass i den italienske cupen. Den suverene serielederen var ille ute etter å ha tapt 0-2 i første kamp mot tabelljumboen i Serie A, men Roma fikk den ønskede starten i returoppgjøret da Haavi scoret i det 14. minutt. Elena Linari doblet ledelsen i det 33. minutt, og dermed var det 2-2 sammenlagt. Tross enorm Roma-dominans både i banespill og sjanser lot flere mål vente på seg. Først i det 84. minutt kunne Roma-spillerne juble igjen da Zara Kramzar scoret. Hun var byttet inn noen minutter tidligere. Haavi og Anja Sønstevoold var i Romas startoppstilling. Haavi ble byttet ut midtveis i 2. omgang. Roma møter Milan i semifinalen i mars, også da med bortekamp først. Roma vant cupen i 2021 og har vært tapende finalist de to siste årene.' On the right, under 'Found text', is a generated article snippet: 'Emilie Haavi scoret et tidlig ledermål da Roma slo Napoli 3-0 i returkampen og sikret semifinaleplass i den italienske cupen. Den suverene serielederen var ille ute etter å ha tapt 0-2 i første kamp mot tabelljumboen i Serie A, men Roma fikk den ønskede starten i returoppgjøret da Haavi scoret i det 14. minutt. Elena Linari doblet ledelsen i det 33. minutt, og dermed var det 2-2 sammenlagt<sup>[2^2]</sup>. Emilie Haavi, som spiller som angrepsspiller for Roma, har vært en viktig bidragsyter til lagets suksess. Hun scoret også sitt første mål for Giallorosse i en hardt kjempet seier mot Napoli<sup>[3^3]</sup>. Det er flott å se at hun har funnet formen og bidrar til lagets prestasjoner! 🇮🇹 For mer informasjon om Emilie Haavi, kan du besøke hennes profil på ESPN<sup>[1^1]</sup>.' A 'Plagiarism Detection' progress bar at the top right shows 47.2% similarity.

Edit distance-based similarity: 0.585

3-Gram Overlap: 0.516

Max N-Gram: 67

The Longest N-Gram:

emilie haavi scoret et tidlig ledermål da roma slo napoli 3-0 i returkampen og sikret semifinaleplass i den italienske cupen. den suverene serielederen var ille ute etter å ha tapt 0-2 i første kamp mot tabelljumboen i serie a, men roma fikk den ønskede starten i returoppgjøret da haavi scoret i det 14. minutt. elena linari doblet ledelsen i det 33. minutt, og dermed var det 2-2

The retrieved news source: <http://www.norwaytoday.com/rss3/viktig-emilie-haavi-scoring-roma-til-cup-semifinale-405565.html>

3) Aftenposten

Example 1 from **Aftenposten** (left is reference news article, and right is generated news article):



Submitted text (390 words)

Bråket fortsetter etter at Lionel Messi (36) ikke kom på banen i en treningskamp i Hongkong. Nå får de fremtatte tilbake 50 prosent av billettprisene. Det skriver nyhetsbyrået Reuters. Reaksjonene var sterke da Lionel Messi ble sittende på benken under en showkamp i Hongkong søndag på grunn av en skade. Det ble ikke bedre da han onsdag spilte en halvtime da Inter Miami møtte japanske Vissel Kobe i Tokyo. **Da kom anklagene om at Messis «no show» i Hongkong var rettet mot Kina.** Hongkongs myndigheter uttrykket skuffelse og etterlyste en forklaring fra arrangørene og klubbene. – Til tross for våre beste intensjoner har vi forståelse for skuffelsen over Lionel Messi og Luis Suarez' fravær i søndagens kamp, og vi beklager at de to ikke kunne spille, skrev Inter Miami i en uttalelse til Reuters torsdag. – Vi erkjenner også at den sene avgjørelsen utløste frustrasjon blant våre Hongkong-supportere og arrangøren Tatler Asia. Det er nødvendig å uttrykke at skader dessverre er en del av det vakre spillet, og at spillernes helse alltid kommer først, het det videre i klubbens uttalelse. Nå refunderer arrangøren av treningskampen i Hongkong 50 prosent av billettprisene etter misnøyen blant supporterne. Det skriver Reuters. Selskapet Tatler Asia, som arrangerte kampen, beklaget dypt i en uttalelse på Instagram og var "knust" over at fansen ble skuffet da Messi satt på benken hele kampen søndag. 40 000 tilskuere hadde møtt opp til kampen i Hongkong. Alle som kjøpte billetter fra offisielle kanaler vil få 50 prosent refusjon, ifølge selskapet. Selskapet opplyser at de har vært i samtaler med myndighetene i Hongkong for å løse saken. – **En teori er at dette var politisk motivert, skriver avisen Global Times om Messis fravær på banen i Hongkong. På X og det kinesiske sosiale mediet Weibo er det postet mange negative kommentarer om Lionel Messi, med anklager om at han er slem og arrogant. En rådgiver for regjeringen i Hongkong har antydnet at han bør nektes innreise til regionen, skriver AFP. – Hans løgner og hykleri er frastøtende, uttalte rådgiveren, Regina Ip<sup>[1]</sup>.**

< Back

your-file.html  
Source: Batch Result

Plagiarism Detection 35.9%

Jun 10, 2024, 22:12:55

Found text

**Bråket fortsetter etter at Lionel Messi ikke kom på banen i en treningskamp i Hongkong. Argentineren beskyldes for politisk markering mot Kina. Reaksjonene var sterke da Lionel Messi ble sittende på benken under en showkamp i Hongkong søndag på grunn av en skade. Regjeringen i Hongkong bidro angivelig med 21 millioner kroner for å få Messi og lagkameratene hans til å spille i storbyen, men til tross for at publikum ropte navnet hans gjennom hele kampen, ble den skadde argentinske superstjernen ubenyttet reserve. Det ble ikke bedre da han onsdag spilte en halvtime da laget hans møtte japanske Vissel Kobe i Tokyo. Da kom anklagene om at Messis «no show» i Hongkong var rettet mot Kina. En teori er at dette var politisk motivert, skriver avisen Global Times. På X og det kinesiske sosiale mediet Weibo er det postet mange negative kommentarer om Lionel Messi, med anklager om at han er slem og arrogant. En rådgiver for regjeringen i Hongkong har antydnet at han bør nektes innreise til landet, skriver AFP. – Hans løgner og hykleri er frastøtende, uttalte rådgiveren, Regina Ip<sup>[1]</sup>.**

Edit distance-based similarity: 0.801

3-Gram Overlap: 0.631

Max N-Gram: 40

The Longest N-Gram:

på x og det kinesiske sosiale mediet weibo er det postet mange negative kommentarer om lionel messi, med anklager om at han er slem og arrogant. en rådgiver for regjeringen i hongkong har antydnet at han bør nektes innreise til

The retrieved news source: <https://www.nrk.no/sport/messi-hetses-i-kina---beskyldes-for-politisk-markering-1.16755909>

Example 2 from **Aftenposten** (left is reference news article, and right is generated news article):

Submitted text (626 words)

– Norsk fotball er ikke så dårlig som noen folk tror, sier John van 't Schip etter 2–2-kampen mot Bodo/Glimt. Den nederlandske storklubben var rystet etter to mål av Albert Grønåbek, men berget seg inn på slutten og sikret et brukbart utgangspunkt før returoppgjøret på Aspmyra om en uke. Prestasjonen til Ajax ble regelrett slakten av flere eksperter, som omtalte det som «veldig, veldig pinlig». På pressekonferansen etter kampen tok Ajax-treneren selvkritikk. – For mange spillere forble under sitt eget nivå. Vi mistet ballen for ofte. Da gjør man det enkelt for et lag som Bodo/Glimt, sier van 't Schip. Han benyttet anledningen til å komme med noen rosende ord. – Norsk fotball er noe annet enn for 20 år siden. **Det er ikke så dårlig som noen folk tror. Det er solide og gode lag. Og Ajax er ikke det laget man forbinde med Ajax akkurat nå, det vet vi alle. Det er en prosess som inkluderer prøving og feiling, sier Ajax-treneren.** Steven Berghuis satte inn 2–2-målet med en lekker lobb i kampens 97. minutt. Han var ikke nådig med eget lag. – Vi er nødt til å ta en skikkelig titt på oss selv og hvordan de utspilte oss. Vi er virkelig nødt til å gjøre ting mye, mye bedre. Noen ganger rundspilte de oss, sukket Berghuis i katakombene på Johan Cruijff Arena. – **Bodo/Glimt viste virkelig hva man kan gjøre når man har en god trener og en god struktur i laget.** Det er det Ajax mangler for øyeblikket, sier Voetbal International-journalisten Lentin Goodijk til VG. Nederlanderne har latt seg imponere av både fagmannen og personen Kjetil Knutsen, som de omtaler som «veldig karismatisk». Lovordene hagler både på presserommet og på trykk i avisene. På pressekonferansen dagen før kampen ble Kjetil Knutsen grillet av nederlandske journalister, som gjennom hele oppholdet har vært svært nysgjerrige på mannen som i sommer var ønsket av Ajax: Statskanalen NOS påpeker at «det ikke er uten grunn at Kjetil Knutsen ble koblet til Ajax», mens Volkskrant skriver at Bodo/Glimt har blitt et bedre lag enn Ajax «takket være Kjetil Knutsen». – Jeg tror de ville vært topp tre i Nederland. Feyenoord og PSV er også veldig gode og godt organisert, men jeg tror de ville vært rett bak dem, sier Goodijk. – Med andre ord bedre enn Ajax? – For øyeblikket tror jeg det. Ajax har nok kanskje mer kvalitet individuelt. Man ser det på noen av nesten-sjansene til Bodo/Glimt, hvor de ikke velger riktig passing eller mangler noe individuelt kvalitet, men de er så godt organisert. Og det utgjør den store forskjellen i fotball, sier han. **Storavisen Telegraaf trekker også frem Bodo/Glimts velstrukturerte og disiplinerte spillestil.** – Med tidvis flytende kombinasjoner skar nordmennene enkelt gjennom Ajax-laget. Det var et velorganisert lag med mye automatikk som kunne tyde på at de har spill sammen lenge. Men det er ikke tilfellet. I vinter mistet Bodo/Glimt to angreppspillere og Knutsen måtte hente inn tre nye spillere, påpeker avisen. Hjemlandet Jens Petter Hauge svarer kontant «nei» da VG spør om han er overrasket over at Bodo/Glimt allerede er på et så høyt nivå. – Vi er et utrolig godt lag og vi har vist på trening og i den treningskampen vi fikk mot Malmö at vi har et spennende prosjekt på gang. Jeg tror vi har utrolig mye mer i vente, sier Hauge til VG. Kjetil Knutsen innrømmer derimot: – Jeg var spent på det. Det er stor forskjell på treningskamper og tellende kamper, men jeg hadde en bra følelse, sier Glimt-treneren til VG. Vinneren sammenlagt over to oppgjør går til åttedelsfinalen i Conference League. – Det blir virkelig ikke enkelt neste uke. Men alt ser litt bedre ut etter de to målene på overtid, sier John van 't Schip.

< Back

your-file.html  
Source: Batch Result

Plagiarism Detection 13.1%

Jun 10, 2024, 23:19:57

Found text

Etter Bodo/Glimts sterke resultat mot Ajax, som nesten ble nok en storskalp i samlingen til Kjetil Knutsen, har flere i Nederland reagert på hvor dominant Glimt var. Treneren til Ajax, John van 't Schip, innrømmet at de ikke spilte bra nok, og at de mistet ballen som en direkte konsekvens av Glimts høye press. Han uttalte: "Norsk fotball er noe annet enn for 20 år siden. Det er ikke så dårlig som noen folk tror. **Det er solide og gode lag. Og Ajax er ikke det laget man forbinde med Ajax akkurat nå, det vet vi alle.** Det er en prosess som inkluderer prøving og feiling," sa Ajax-treneren etter kampen<sup>[1]</sup>. Den danske eks-spilleren og ESPN-ekspert Kenneth Pérez var også tydelig i sin vurdering. Han konkluderte: "Vi har sett et veldig, veldig pinlig Ajax i aksjon." Pérez mente at Ajax ble rundspilt av et lag som ikke hadde spilt en tellende kamp på to måneder. Flere medier i Nederland påpeker at Ajax mangler noe av det Bodo/Glimt ser ut til å ha mye av: nemlig struktur i laget. Journalisten Lentin Goodijk fra Voetbal International uttalte: "**Bodo/Glimt viste virkelig hva man kan gjøre når man har en god trener og en god struktur i laget.** Det er det Ajax mangler for øyeblikket." Storavisen Telegraaf trakk også frem Bodo/Glimts velstrukturerte og disiplinerte spillestil. De bemerket at nordmennene skar enkelt gjennom Ajax-laget med flytende kombinasjoner. Selv om det ikke var tilfellet at Bodo/Glimt hadde spilt sammen lenge, var det et velorganisert lag med mye automatikk. Nå gjenstår det å se hva Glimt klarer å få til mot Ajax når de tar dem imot på Aspmyra neste torsdag. Vi får håpe på skadene til Sondre Brunstad Fæt, Albert Grønåbek, og Jens Petter Hauge ikke blir for alvorlige, og at Odin Bjørtuft lærer av situasjonen etter sitt røde kort<sup>[1]</sup>. Vinneren av neste kamp er klar for åttedelsfinalen i Serieligaen. Fotballverdenen følger spent med på hvordan Bodo/Glimt vil prestere mot Ajax i returoppgjøret. Det er tydelig at norsk fotball har kommet langt, og lag som Glimt viser at de kan konkurrere på et høyt nivå. Kanskje er dette starten på en ny era for norsk fotball? Vi får se!<sup>[1]</sup> <sup>[1]</sup>: [Forb]løftet over velorganisert Bodo/Glimt – NRK Nordland[[https://www.nrk.no/nordland/forbloeft-over-velorganisert-bodo\\_glimt-1.16766493](https://www.nrk.no/nordland/forbloeft-over-velorganisert-bodo_glimt-1.16766493)]

Edit distance-based similarity: 0.484

3-Gram Overlap: 0.212

Max N-Gram: 48

The Longest N-Gram:

fotball er noe annet enn for 20 år siden. det er ikke så dårlig som noen folk tror. det er solide og gode lag. og ajax er ikke det laget man forbinder med ajax akkurat nå, det vet vi alle. det er en prosess som inkluderer prøving og

The retrieved news source: [https://www.nrk.no/nordland/forblyffet-over-velorganisert-bodo\\_glimt-1.16766493](https://www.nrk.no/nordland/forblyffet-over-velorganisert-bodo_glimt-1.16766493)

#### 4) Dagbladet

Example 1 from **Dagbladet** (left is reference news article, and right is generated news article):

The screenshot shows a plagiarism detection interface. On the left, under 'Submitted text (456 words)', there is a reference news article snippet. On the right, under 'Found text', there is a generated news article snippet. A 'Plagiarism Detection' bar at the top right shows a score of 81.8%. The text in both snippets is largely identical, with some minor differences in phrasing and punctuation. The tool interface includes navigation arrows and a 'Back' button.

Edit distance-based similarity: 0.997

3-Gram Overlap: 0.989

Max N-Gram: 275

The Longest N-Gram:

etter å innføre mest mulig åpenhet etterlengtet. de fem punktene han la fram på en pressekonferanse i pressens hus fredag formiddag er viktige. vi må likevel fortsatt vente på realiseringen av det nå ti år gamle vedtaket. vi applauderer at det nå skal lages et register over reelle rettighetshavere og at det kommer en forskrift om innsyn i aksjer registrert på forvaltere. vestre lover å sørge for bedre konsernrapportering, om nødvendig vil han også bruke sanksjoner for å få det til. regjeringen lover raskt å få på plass en forskrift om bedre tilgang til aksjeeierboken og aksjeeierregisteret. alt dette vil sikre offentligheten bedre tilgang til opplysninger om eierskap og aksjehandel. slikt innsyn er svært viktig i arbeidet for å avsløre korrupsjon, hvitvasking og annen økonomisk kriminalitet. det er dessuten i tråd med en vakker norsk tradisjon, der alle har tilgang på opplysninger om hverandres inntekt, skatt, bil og bolig. åpenhet er et mål i seg selv, og det er et essensielt verktøy for å avsløre kritikkverdige forhold. snart ti år etter stortingsvedtaket er et løpende oppdatert aksjonerregister ennå ikke på plass. Det skal på plass, men innretningen er ikke klar. Vestre har satt ned en hurtigarbeidende arbeidsgruppe som skal se på dette fram mot sommeren, blant annet for å vurdere om den kan være åpen for alle eller om tilgangen bør begrenses til dem med legitim interesse – som journalister, sivilsamfunnsorganisasjoner og myndigheter. Vestre skisserte problemstillinger knyttet opp til vern av personopplysninger, forretningshemmeligheter og hvordan sammenstilling av informasjon kan misbrukes av kriminelle med onde hensikter. En EU-dom fra 2022 problematiserer det siste og utvalget må gjøre avveininger rundt hvordan dette påvirker Norge og retten til privatliv[11].



løpende oppdatert aksjonærregister ennå ikke på plass. det skal på plass, men innretningen er ikke klar. vestre har satt ned en hurtigarbeidende arbeidsgruppe som skal se på dette fram mot sommeren, blant annet for å vurdere om den kan være åpen for alle eller om tilgangen bør begrenses til dem med legitim interesse – som journalister, sivilsamfunnsorganisasjoner og myndigheter. vestre skisserte problemstillinger knyttet opp til vern av personopplysninger, forretningshemmeligheter og hvordan sammenstilling av informasjon kan misbrukes av kriminelle med onde hensikter. en eu-dom fra 2022 problematiserer det siste og utvalget må gjøre avveininger rundt hvordan dette påvirker norge og retten til

The retrieved news source: <https://www.dagbladet.no/meninger/gjor-ernas-forklaring-umulig/80848258>

Example 2 from **Dagbladet** (left is reference news article, and right is generated news article):

The screenshot shows a plagiarism detection interface. On the left, a text area contains the original article from Dagbladet. On the right, a generated article is shown with a similarity score of 61.8%. The tool highlights matching text in red. The generated article is a slightly reworded version of the original, maintaining the same structure and content.

Edit distance-based similarity: 0.917

3-Gram Overlap: 0.894

Max N-Gram: 380

The Longest N-Gram:

trond giske satt på samme fly som meg. morgenen etter idus martiae, dagen i den romerske kalenderen som er blitt ensbetydende med de opprørske senatorenes drap på diktator julius cæsar. fra stedet han bor, oslo, til årsmøte i hans politiske hjemsted, trøndelag. giske er – enten folk vil det eller ei – den stadige hovedattraksjonen hver gang trøndelag arbeiderparti er samlet til dyst. han er opprøreren mange av partikollegene savner dypt og inderlig. lykken skulle ha det til at fylkeslaget stortingstopp, eva kristin hansen, rett før helgens møte gikk av stabelen meldte overgang til arbeiderpartiet i akershus. dermed gjorde hun det også klart at en renominasjon til stortinget ikke er aktuelt. de to andre sørtrønderne på tinget, kirsti leirtrø og jorodd asphjell, er ikke akkurat gallionsfigurer av det kaliberet trøndelag ap krever av sine representanter. og dermed ikke førstevalg som toppkandidater. det er nå et åpent spørsmål hvem som skal toppe stortingslista til arbeiderpartiet i sør-trøndelag. trond giskes muligheter er gode, men møtte på flere hindre denne helga. for det første mistet trondheim arbeiderpartiet betydelig makt. laget får nå ti færre delegater når stortingslista skal bestemmes seinere i år. tidligere hadde trondheim ap reint flertall, men det er nå historie. støtten til giske sprer seg uansett godt utover kommunegrensa i trondheim. det er mange som savner en så tydelig og markant figur som giske til å representere dem i oslo. håpet om

å bytte ut lederen av nominasjonskomiteen med en mer giske-vennlig en, gikk heller ikke. likevel fikk giske inn flere støttespillere i gruppa som skal legge fram forslaget til stortingsliste. trønderne i arbeiderpartiet ser ikke ut til å ha mottatt memoet fra ledelsen om at det nå går mot mye lysere tider i norsk økonomi. inflasjonen er allerede på vei ned, og snart skal nok rentenivået det samme. men trøndelag ap ser uansett ikke ut til å ville ha noen som bejaer ledelsen – en partikanin som det ofte kalles i politikken. årsmøtet valgte sådan å stille noen krav til dem som skal toppe stortingslistene i de to valgdistriktene ved neste års stortingsvalg. der heter det at «stortingsrepresentantene som representerer trøndelag arbeiderparti skal være trøndelags fremste nasjonale politikere, med en tydelig trøndersk stemme». det kan man ikke akkurat si om dagens representanter. men giske var en slik en i

The retrieved news source: <https://www.dagbladet.no/meninger/et-sykt-lag/81130520>

## 3.2 GPT Base Models

### 3.2.1 Model Selection

We would like to use the completion task to test the memorization of GPT base models. Recently, some GPT models have been fine-tuned to provide protections against content violations. We tried designing different styles of prompts to steer GPT in completing news articles. However, GPT-4 Turbo and GPT-4 always respond that they cannot reproduce the news article due to copyright law. Therefore, we selected **GPT-3.5 Turbo**, which also has restrictions on some prompts but may generate news articles for other carefully designed prompts, for the memorization test.

### 3.2.2 Sampled Testing Dataset

Since GPT-3.5 Turbo's knowledge is current up to September 2021, we randomly sampled 500 news articles published in 2020 from NRK, VG, and Aftenposten to test whether its training data include content from these Norwegian publishers.

### 3.2.3 Evaluation Metrics

- Edit distance-based similarity: The normalized similarity between two input sentences  $a$  and  $b$ :  $1 - \text{distance}(a,b) / \max(a.\text{length}, b.\text{length})$ . The higher value indicates a greater similarity between the two inputs.
- 3-Gram Overlap – Percent of all 3-grams in the generated text that are also present in the original article.
- Max N-Gram – The length of the largest n-gram in the generated text that can be found in the original article.

### 3.2.4 Results

1) The prompts that we tried for the memorization test are as follows:

Prompt 1: "Please complete a news article in Norwegian that was published in 2020 and begin with the following sentences. The article should follow facts and should not be fictional. \n\n" + sentences[0] + " " + sentences[1]

Prompt 2: "You are a journalist. I will give you the beginning of an article from the Verdens Gang (VG). You will complete each article with the original unaltered writing in Norwegian. Here is the article beginning: \n\n" + sentences[0] + " " + sentences[1]

Prompt 3: "Please summarize or paraphrase a news article from Aftenposten.no in Norwegian and the article begins with: " + sentences[0] + " " + sentences[1]

Prompt 4: "Please write a similar, original news-style article in Norwegian that begins with the following sentences. The article should follow facts and should not be fictional. \n\n" + sentences[0] + " " + sentences[1]

Prompt 5: "Please summarize or paraphrase the following news article in Norwegian: \n\n" + sentences[0] + " " + sentences[1]

2) Response examples of content infringement protection generated by GPT-4, GPT-4 Turbo, and GPT-3.5 Turbo with Prompt 2:

Response 1:

Jeg beklager, men jeg kan ikke oppsummere eller omskrive nyhetsartikler som er opphavsrettslig beskyttet. Jeg kan imidlertid gi deg en generell oversikt over situasjonen basert på min kunnskap fra før 2021, hvis det er til hjelp.

Response 2:

I'm sorry, but I can't continue the article with the original unaltered writing as it would be a reproduction of copyrighted content. However, I can provide a summary or discuss the themes and topics related to the life of students in Norway if that would be helpful. Let me know how you would like to proceed!

Response 3:

I beklager, men jeg kan ikke oppfylle forespørselen om å omskrive eller oppsummere en spesifikk nyhetsartikkel. Men jeg kan gi deg en generell oppsummering av lignende hendelser basert på min kunnskap fra før 2021. For eksempel:

I Sydney samler vanligvis tusenvis av mennesker seg i havneområdet for å se på et spektakulært fyrverkerishow som lyser opp himmelen over Harbour Bridge og operahuset ved midnatt. I Auckland, New Zealand, teller folk ned til det nye året og nyter et tradisjonelt fyrverkerishow fra Sky Tower. Disse feiringene er kjente nyttårsbegivenheter som tiltrekker seg besøkende fra hele verden.



### 3) The memorization test results of GPT-3.5 Turbo

Table 6: Evaluation results per publisher from the GPT-3.5 model.

Publisher	Average Edit Distance-based Similarity	Average 3-Gram Overlap	Average Max N-Gram
NRK	0.320	0.011	3.270
VG	0.324	0.016	3.650
Aftenposten	0.321	0.013	3.532

Example 1 from **NRK** (left is reference news article, and right is generated news article):

Submitted text (501 words)

Programmaredektevran galgá Larsson (39) jodihit NRK Sámi prográmmadoaimmahusa mas lea viiddes samperovddasvástádus ja lagabui 30 bargi. NRK Sámi prográmmaossodat buvttada sámi sisoalolu mándáide, nuoraide ja rávesolbmuid neahtta, rádiu/podkástta, TV/neahtta-TV ja SoMe bokte. Ossodat buvttada ea ea. Binnabánnaža, realityrádiu Ipmaša, NRK Urbbi, Sámi álbmotheašáidagaga ja nátionála sámi ráiddiid leat muhtun oasit. Illuda álgit Carl-Göran Larsson illuda álgit prográmmaredektevran. – Dákkár vejolašvuoda ja erenoamáš hástalus it oacšču dávjá. Sáhtnán rámisvuodain dadjat ahte šaddá hui geddolaš oazžut lobi jodihit dakkár čeahpes ja buori joavkku geain lean oahpásmuvvan NRK Sámis. Ovttas lean sihkkar ahte mii nagodit ráhkadit dan sámi sisoalolu maid geahččit ja guldaleaddjit sihke ánsásit, dárbbašit, ja maiddái maid eai ba diehtán ahte sii háliidit. Mun illudan dása, dadjá Carl-Göran Larsson. Larsson (39) lea bargan journalistan ja prošeaktajođheadjiin NRK Sámis jagi 2013 rájes ja lea háikkahuvvon journalista, ea ea. Finnmarkku journalistta háikkásumiin 2017:s ja Svarte Natta rámiduvvon sániin. – Erenoamáš jodihanattáidagat NRK Sámi direktivrra, Mona Solbakk, lea hui duhtavaš go Larsson válddii vuostá virgefálaldaga. – Mun lean hui movvttet go Carl-Göran šaddá oassin NRK Sámi hoavdajoavkkus. Sus lea nanu journalisttalaš vásáhus ja máhttu, ja erenoamáš jodihanattáidagat. Carl-Göran maid dovdá mediasuorgi treanddaid, álbmoga lihkaemiid ja Sámi bures, loahpaha Mona Solbakk. Larsson lea Kárášjogas eret ja doaimmagoahtá redektevran čakčat. Váldde oktavuoda: Carl-Göran Larsson: 452 31 477 Mona Solbakk: 951 42 060 Norsk versjon: Ny programredaktør i NRK Sápmi Carl-Göran Larsson blir programredaktør i NRK Sápmi. Som programredaktør vil Larsson (39) lede en redaksjon som produserer samisk innhold for barn, unge og voksne til nett, radio/podkast, tv/nett-TV og SoMe. Redaksjonen produserer alt fra samisk barneinnhold som Binnabánnaš og realityserien Imaš til musikkatsingen NRK Urbi, samesen nasjonal dagssending og nasjonale samiske serier. Programavdelinga har nærmere 30 ansatte. Gleder seg Carl-Göran **Larsson ser frem til å ta fatt på jobben som programredaktør**. – En slik utfordring og mulighet får du ikke ofte. Jeg kan med stolthet si at det blir utrolig spennende å få lov til å lede en så dyktig og fin gjeng som jeg har lært å kjenne i NRK Sápmi. Sammen er jeg sikker på at vi klarer å lage det samiske innholdet som publikum både fortjener, behøver, og kanskje ikke en gang visste at de ønsket seg. Jeg gleder meg, sier Larsson. Larsson har jobbet som journalist og prosjektleder i NRK Sápmi siden 2013. Han fått flere hedersutmerkelse som for eksempel Finnmark journalistpris 2017 og hederlig omtale på Svarte Natta. – God leder som motiverer Direktør i NRK Sápmi, Mona Solbakk, er fornøyd med at Larsson har takket ja til jobben. – Jeg er veldig glad for at Carl-Göran blir en del av redaktørkollegiet i NRK Sápmi. Han kjenner trendene i mediebransjen, bevegelsene i publikum og Sápmi godt. Hans solide journalistfering og gode måte å lede og motivere ansatte på, vil gjøre oss i enda bedre stand til å lage samisk innhold som blir valgt, funnet og elsket av publikum, sier direktør i NRK Sápmi, Mona Solbakk. Larsson er opprinnelig fra Karasjok og tiltrer redaktørstillingen til høsten. Kontaktpersoner: Carl-Göran Larsson: 452 31 477 Mona Solbakk: 951 42 060

your-file.html

Source: Batch Result

Jun 11, 2024, 12:34:34

Plagiarism Detection 2.2%

Found text

Larsson er fra Kiruna i Sverige og har en lang karriere innenfor mediebransjen. Han har tidligere jobbet som journalist og redaktør for den svenske avisen Norrbottens-Kuriren, og har også erfaring som programleder for Sveriges Radio Sápmi. NRK Sámi er den samiske avdelingen til Norges største mediehus, NRK. Kanalen sender nyheter, underholdning og kulturprogrammer på samisk til den samiske befolkningen i Norge. **Larsson ser frem til å ta fatt på jobben som programredaktør** og utvikle kanalen videre. "Jeg er veldig glad for å få denne muligheten til å jobbe med samiske medier og bidra til å styrke den samiske kulturen og språket gjennom NRK Sámi," sier Larsson. Han legger til at han er imponert over det arbeidet som allerede er gjort i kanalen, og at han ser frem til å samarbeide med det dyktige teamet som allerede er på plass. NRK Sámi har vært en viktig stemme for den samiske befolkningen i Norge siden oppstarten i 1986. Kanalen har også vært en viktig bidragsyter til å bevare og utvikle det samiske språket. "Vi er veldig glade for å ønske Carl-Göran velkommen til NRK Sámi. Han har en imponerende bakgrunn og erfaring innenfor mediebransjen, og vi er sikre på at han vil være en stor ressur for kanalen," sier kringkastingssjef Thor Gjermund Eriksen. Larsson vil begynne i sin nye stilling i januar 2020, og ser frem til å bli en del av det samiske mediemiljøet i Norge.

Edit distance-based similarity: 0.335

3-Gram Overlap: 0.058

Max N-Gram: 10

The Longest N-Gram:

larsson ser frem til å ta fatt på jobben som

Example 2 from **VG** (left is reference news article, and right is generated news article):

Submitted text (1358 words)

deres ektefeller. Kurdiske selvstyremyndigheter har gjort det klart at de, slik situasjonen faktisk er, ønsker at også europeiske land henter hjem sine borgere. Å unnlate å hente dem hjem er i strid med norske rettsprinsipper fordi man risikerer at personene det gjelder, vil kunne ende opp i et rettsystem, f.eks. i Irak, som ikke er i nærheten av å sikre det vi anser som et minimum av rettsikkerhet. Avskrellet all retorikk koker argumentet om at IS-farmer bør rettsforfølges i området der forbyrteletene ble begått, ned til noe helt annet og mindre prisverdige: Norge, og andre land som følger samme linje, bruker et mer eller mindre rettsløst område som dumpingplass for egne antatte terrorister eller det vi mener er potensielle terrorister. Det blir hevdet at det å hente hjem norske borgere med IS-tilknytning vil utgjøre en langsiktig sikkerhetsrisiko. Det er vanskelig å se at dette argumentet har troverdighet når det gjelder IS-kvinnene: Norsk rettsvesen håndterer rutinemessig personer som utgjør en utvilsom og etter alt å dømme langt mer påtrengende sikkerhetsrisiko for det norske samfunnet enn denne gruppen av unge kvinner og deres barn. PSTs uttalelser etter at den ene kvinnen nå er hentet hjem, synes heller ikke å tyde på at dette argumentet burde tillegges særlig vekt. Argumentet kan tenkes å ha mer tyngde når det gjelder eventuelle gjenværende mannlige IS-medlemmer, men også for disse er det vanskelig å forstå hvordan et lite antall enkeltpersoner som er kjent for myndighetene, skulle utgjøre en uhandterbar sikkerhetsrisiko. Om regjeringen mener at norsk lovverk ikke gir det nødvendige rom for en ansvarlig håndtering av denne gruppen, er det åpenbart svaret å ta initiativ overfor Stortinget til å lovverket revideres. Risikoen for at personene det gjelder, skal utgjøre en sikkerhetsrisiko, på sikt også for Norge, synes å være langt større om de forblir i de mer eller mindre statsløse områdene der de nå oppholder seg. Blant IS-farenes barn, om de forblir under nedverdige forhold i Midtøsten, er faren for å finne fremtidige terrorister overhengende. Påstanden om at IS-farmer vil utgjøre en uakseptabel sikkerhetsrisiko om de hentes hjem til Norge, synes lite overbevisende, men er likevel et forhold det kan hense ulike oppfatninger om. Enkelte andre argumenter som er blitt fremført, burde etter vår oppfatning derimot avvises som uanstendige. Sylvi Listhaugs offentlige begrunnelse for at Frp avviste å hente IS-kvinnene hjem og derfor i første omgang tok dissens i regjeringen, dreier seg verken om en angivelig sikkerhetsrisiko eller at et rettsoppgjør bør finne sted i områdene der forbyrteletene skal ha foregått. Listhaug opprørers, som vi alle gjør, over barbariet som IS representerer og slår fast at «å støtte IS er å støtte halshugging, brenning, voldtekt og drap på uskyldige mennesker». I alt dette har Listhaug rett. Men så begynner det å gå galt: For Listhaug dreier det seg om at IS-krigerne og deres sympatisører «har vendt det norske samfunnet ryggen og kjemper mot vårt land og våre verdier», og Frp vil derfor «ikke løfte en finger for at IS-krigerne og IS-kvinner skal komme tilbake til de frie, vestlige samfunnene de hater.» Det Listhaug faktisk sier her, er at kvinnene det gjelder, ikke bare de dokumenterte IS-krigerne, har begått handlinger som medfører at de ikke lenger har krav på å behandles som norske borgere med alt det innebærer av rett til likebehandling, rettsvern og, i disse tilfellene, forsvarlig rettslig behandling av påståtte forbyrteletser. Historielsheten i en slik argumentasjon er forbløffende. Vi viste i vår forrige kronikk til sider ved det norske rettsoppgjøret etter andre verdenskrig, som i hovedsak ble gjennomført etter rettsstatlige prinsipper, til tross for at de anklagede hadde gått i tjeneste hos et regime som representerte en ondskap uten grenser. Frontkjemperne, medlemmer av Waffen SS, ble hentet hjem fra fangenskap i Sovjetunionen for å dømmes og sone sin straff i Norge. Rettstaten viser sin styrke gjennom behandlingen av dem som selv ikke respekterer rettstatens spilleregler. For oss er dette en vesentlig del av de norske verdier som Listhaug hevder å forsvare. Rettsaken som ble gjennomført etter terrorhandlingen 22. juli 2011, er et nærliggende eksempel på hvordan dette prinsippet ellers er en selvfølgelig del av vår forståelse av hva slags land vi ønsker å være. En person med tilknytning til regjeringen uttalte for en tid siden i forbitarten til en av oss at spørsmålet om å hente hjem IS-farmer er «svært komplisert». Om vedkommende med dette tenkte på de praktiske utfordringene med å operere i aktuelle områdene i Syria, er dette sannsynligvis riktig. Om det derimot tenkes på sakens rettslige eller politiske side, er forholdet til spørsmålet er nøyaktig så komplisert som politikerne selv gjør det til. Det synes klart at det kan argumenteres rettslig for at regjeringen ikke har en plikt til å hente kvinnene, eller mennene, hjem, med eller mot deres vilje. Det er samtidig ingenting som hindrer regjeringen i å etablere dette som en målsetting. Dette ville trolig måtte innebære et aktivt samarbeid med blant annet kurdiske selvstyremyndigheter. Dermed er vi tilbake til én vesentlig likhet mellom dagens debatt og situasjonen som førte til statsministerens unnskyldning til «tyskerjentene» og deres barn i oktober 2018: Det dreier seg til syvende og sist om hvordan vi ønsker og forventer at Norge skal opptre overfor sine borgere og fremstå i det internasjonale samfunn; om hvorvidt våre myndigheter i sin retorikk og sine handlinger tar ansvar for egne borgere og forholder seg til grunnleggende humanitære, etiske og rettslige prinsipper.

< Back

your-file.html

Source: Batch Result

Jun 11, 2024, 12:40:56

Plagiarism Detection 1.8%

Found text

Nå, ett år senere, er debatten fortsatt like aktuell. Spørsmålet om hva som skal skje med norske borgere som har sluttet seg til IS, og deres barn, er fortsatt et tema som splitter opinionen. I oktober i fjor skrev professor Sven G. Holtsmark og professor Geir Ulfstein en kronikk hvor de argumenterte for at norske myndigheter burde ta en prinsipiell beslutning om å hente hjem norske IS-kvinner og deres barn. De påpekte at Norge har et ansvar for sine borgere og at det er viktig å følge internasjonale konvensjoner og menneskerettigheter. Debatten har vært preget av ulike synspunkter og meninger. Noen mener at IS-kvinnene og deres barn bør få komme tilbake til Norge og få en rettfærdig rettergang, mens andre mener at de har frivillig sluttet seg til en terrororganisasjon og derfor ikke fortjener å bli hjulpet av norske myndigheter. I januar i år ble det kjent at regjeringen hadde besluttet å hente hjem en norsk IS-kvinne og hennes to barn fra Syria. Dette skapte reaksjoner fra både politikere og befolkningen. Noen mente at dette var et riktig og humant valg, mens andre var bekymret for sikkerheten til Norge og om disse personene kunne utgjøre en trussel. Saken har også blitt tatt opp i rettsvesenet, hvor flere IS-kvinner har saksøkt staten for å få hjelp til å komme hjem til Norge. I juni i år ble det kjent at en IS-kvinne og hennes barn fikk medhold i retten og at staten måtte hente dem hjem. Debatten om IS-kvinnene og deres barn har også ført til at regjeringen har utarbeidet en ny strategi for å forebygge og håndtere radikalisering og voldelig ekstremisme. Dette er et viktig skritt for å hindre at flere nordmenn blir rekruttert til terrororganisasjoner som IS. Saken om IS-kvinnen dreier seg til syvende og sist om hvordan vi ønsker og forventer at Norge skal opptre overfor sine borgere og fremstå i det internasjonale samfunn. Det er viktig å følge internasjonale konvensjoner og menneskerettigheter, samtidig som vi må ta ansvar for våre egne borgere og sikre at de ikke utgjør en trussel mot samfunnet. Debatten vil trolig fortsette i lang tid fremover, men det er viktig at vi som nasjon tar en prinsipiell beslutning og følger opp med konkrete tiltak for å håndtere situasjonen på en human og sikker måte.

Edit distance-based similarity: 0.325

3-Gram Overlap: 0.108

Max N-Gram: 24

The Longest N-Gram:

dreier seg til syvende og sist om hvordan vi ønsker og forventer at norge skal opptre overfor sine borgere og fremstå i det internasjonale

Example 3 from Aftenposten (left is reference news article, and right is generated news article):

Submitted text (2659 words)

Eller sender ham inn på et kasino, sier grunnleggeren av Yes We Can Youth Clinic. I enden av den 300 meter lange lincealleen ligger det som i første øyeblikk ser ut som en herskkelig golfresort. Men bak fasaden på det renoverte middelalderstøt, omkranset av tunge løvtrær, velrimmede prydbusker og mørkegrønne kanaler, lever rundt 150 unge mennesker fra 13 til 25 år under et strengt regime. Mange av dem sliter med psykiske diagnoser, andre er her på rusrehabilitering. Rommene er overvåket, begrunnelsen er å unngå selvskadning og at ulovlige stoffer finner veien inn. Eller Ipader og mobiltelefoner. Rundt 30 av ungdommene ved denne klinikken er her på avvenning for skjermavhengighet. Og det er denne gruppen som øker mest. Velkommen til Yes We Can i Hilvarenebeck i Nederland, Europas eneste ungdomsklinikk for dataspillavhengige. Vi kaller ham Thijs. 19-åringen er snart ferdig med sitt ti uker lange opphold på ungdomsklinikken. Verden der ute har vært utilgjengelig i snart åtte uker. Eller egentlig i mange år. Han startet å spille dataspill da han var tre år, men mener han først ble avhengig som tiåring. Han var ikke klar over det den gangen, alle spilte jo. Men Thijs spilte mer og mer. Seks, syv, åtte timer om dagen. I dag forstår han at spilleavhengigheten handlet om ensomhet. – Online fant jeg venner å spille med, der opplevde jeg mestring. Jeg tror mange unge spillere kan kjenne seg igjen i det. Gutterommet virker langt unna. Nå står han her med friske roser i kinnene og ny selvtilitt: Her, ute i skogen, kaster han seg ikke bare ut i løse luften, fem meter over bakken. Han tør også å vise seg sårbar. Det er dette de har trent på i terapirommet: å vri seg unna ubehagelige spørsmål er ingen løsning. Thijs tror han er genetisk disponert. – Gamingen ble min avhengighet. Det kunne like gjerne blitt noe annet. Diagnose: dataspillavhengig I 2018 bestemte Verdens helseorganisasjon (WHO) at dataspillavhengighet er en egen diagnose, en psykisk lidelse. Det skal gjøre det lettere å få behandling. Men hvor går grensene mellom engasjement, problematisk spilling og avhengighet? Hvor langt må det gå for at spillingen skal gjøre deg psykisk syk? WHO har ført opp en rekke kriterier for diagnosen, som manglende kontroll over spillingen og at andre interesser og aktiviteter blir prioritert bort. Spillingen går ut over personlige forhold, skole eller jobb. Adferden må også ha vart over tid. Jan Willem Poot, grunnlegger og administrerende direktør av Yes We Can Youth Clinics, kaller spilleavhengigheten en epidemi. – For 15 år siden var det marihuana, sier nederlenderen, som selv har lagt 15 år med narkotikamisbruk bak seg. – Nå er det gaming. Poot har ingen medisinsk bakgrunn, men hans team på 300 ansatte inkluderer psykiatere, terapeuter, sykepleiere, rådgivere og ungdomscoacher. – Det som gjør dataspill så farlig, er at det starter så mye tidligere. Du gir ikke syvåringen narkotika. Eller sender ham inn på et kasino. For 15 år siden var det marihuana, nå er det gaming. Selv har Poot ikke barn. Han har heller aldri spilt dataspill. Men han kan vise til statistikker: Siden 2011, da ungdomsklinikken åpnet opp for at også dataspillavhengige skulle få behandling, har etterspørselen eksplodert. Så langt i år har klinikken behandlet over 300 unge for skjermavhengighet, 120 av dem er under 18. Antallet er fordoblet hvert år. De fleste er unge gutter som har mistet kontrollen over spillingen. De vanligste spillene er Fortnite, League of Legends, Call of Duty, GTA og FIFA. Mobilavhengige kommer også hit for å få hjelp, og særlig jenter lever livet gjennom sosiale medier. Nærmere 100 jenter har fått behandling for skjermavhengighet bare i år. Da A-magasinet besøkte klinikken, var 15 jenter i behandling. Ofte er deres utfordringer sammensatte: Mange har også andre diagnoser, som angst eller depresjon, som har ført til skjermavhengighet. Bedre enn virkeligheten – Jeg tok ikke vare på meg selv, jeg dusjet ikke, russet ikke, tenkte ikke, jeg ville bare sitte inne på rommet mitt

< Back  
your-file.html  
Source: Batch Result  
Jun 11, 2024, 12:44:54  
Plagiarism Detection  
0.7%

Found text

Nå har en ny studie kommet frem til at det er grunn til bekymring. Forskere ved Universitetet i Oslo har gjennomført en omfattende undersøkelse av barn og unges bruk av dataspill. Resultatene viser at stadig flere barn begynner å spille dataspill allerede i en alder av syv år. Dette er en bekymringsfull utvikling, ifølge forskerne. – Det som gjør dataspill så farlig, er at det starter så mye tidligere. Du gir ikke syvåringen narkotika, men du gir dem et dataspill som kan være like avhengighetskapende, sier professor i psykologi, Anne-Lise Løvlie Schibbye. Studien viser at barn som begynner å spille dataspill i en så ung alder, har større risiko for å utvikle avhengighet og problemer med å kontrollere spillingen senere i livet. Dette kan føre til dårligere skoleprestasjoner, sosiale problemer og psykiske lidelser. – Vi ser at barn som spiller mye dataspill, ofte isolerer seg og mister interessen for andre aktiviteter. Dette kan føre til ensomhet og depresjon, sier Schibbye. Forskerne understreker at det ikke er dataspill i seg selv som er farlige, men det er viktig å være bevisst på hvor mye tid og energi barna bruker på det. De oppfordrer foreldre til å være mer involvert i barnas spillvaner og sette grenser for hvor mye tid som brukes foran skjermen. – Det er viktig å ha en åpen dialog med barna om spillene de spiller, og å være bevisst på aldersgrenser og innhold i spillene. Foreldre bør også oppmuntre til andre aktiviteter og sørge for en balansert hverdag for barna sine, sier Schibbye. Undersøkelsen viser også at det er en økning i andelen jenter som spiller dataspill, og at det er flere jenter enn gutter som opplever problemer knyttet til spillingen. – Dette er en viktig påminnelse om at dataspill ikke bare er en guttegreie. Det er viktig å være oppmerksom på at jenter også kan utvikle avhengighet og problemer med dataspill, sier Schibbye. Forskerne håper at denne studien vil bidra til å øke bevisstheten rundt barn og unges bruk av dataspill, og at det vil føre til mer ansvarlig og bevisst bruk av spill blant både barn og foreldre.

Edit distance-based similarity: 0.376

3-Gram Overlap:0.086

Max N-Gram: 18

The Longest N-Gram:

– det som gjør dataspill så farlig, er at det starter så mye tidligere. du gir ikke syvåringen

## Appendix 1. Unique URL statistics on Norwegian Publisher Domains

Unique URL statistics from Common Crawl belonging to Norwegian news media domains

Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="http://nrk.no">nrk.no</a>	0.005430287732081723	58928
2	<a href="http://tv2.no">tv2.no</a>	0.005217971296181704	56624
3	<a href="http://dagbladet.no">dagbladet.no</a>	0.0031157252665865993	33811
4	<a href="http://vg.no">vg.no</a>	0.0017521635035603088	19014
5	<a href="http://blogg.no">blogg.no</a>	0.0015261165342622424	16561
6	<a href="http://dn.no">dn.no</a>	0.0013981184745985593	15172
7	<a href="http://abcnyheter.no">abcnyheter.no</a>	0.001257311393845422	13644
8	<a href="http://fvn.no">fvn.no</a>	0.0011219412357129884	12175
9	<a href="http://agderposten.no">agderposten.no</a>	0.001081210391673141	11733
10	<a href="http://ntb.no">ntb.no</a>	0.0010666504971973585	11575
11	<a href="http://forskning.no">forskning.no</a>	0.0008805971620922642	9556
12	<a href="http://sol.no">sol.no</a>	0.00042573868657034956	4620
13	<a href="http://adressa.no">adressa.no</a>	0.00042260554472113056	4586
14	<a href="http://startsiden.no">startsiden.no</a>	0.00040518896208870716	4397
15	<a href="http://itromso.no">itromso.no</a>	0.0003744104509816732	4063
16	<a href="http://dagsavisen.no">dagsavisen.no</a>	0.0003464886280312802	3760
17	<a href="http://itavisen.no">itavisen.no</a>	0.0003440926960289362	3734
18	<a href="http://kk.no">kk.no</a>	0.0003407752517179984	3698
19	<a href="http://vol.no">vol.no</a>	0.00033966943694768584	3686
20	<a href="http://vikebladet.no">vikebladet.no</a>	0.00033865577340823263	3675
21	<a href="http://opp.no">opp.no</a>	0.0003377342610996388	3665
22	<a href="http://fosna-folket.no">fosna-folket.no</a>	0.00033533832909729485	3639
23	<a href="http://ht.no">ht.no</a>	0.00033211303601721643	3604
24	<a href="http://document.no">document.no</a>	0.0003268604158582316	3547
25	<a href="http://smp.no">smp.no</a>	0.0002962662072129164	3215
26	<a href="http://klassekampen.no">klassekampen.no</a>	0.00029451533382658814	3196

27	<a href="#">cw.no</a>	0.0002861295718183843	3105
28	<a href="#">p3.no</a>	0.0002834571861234622	3076
29	<a href="#">vl.no</a>	0.0002618016468715072	2841
30	<a href="#">kode24.no</a>	0.0002594057148691632	2815
31	<a href="#">folkebladet.no</a>	0.00025894495871486633	2810
32	<a href="#">bladet.no</a>	0.0002542452459410378	2759
33	<a href="#">kvinneguiden.no</a>	0.000249729835628928	2710
34	<a href="#">morenytt.no</a>	0.00024945338193634985	2707
35	<a href="#">altaposten.no</a>	0.00024733390362658407	2684
36	<a href="#">bomlo-nytt.no</a>	0.0002469652987031465	2680
37	<a href="#">varden.no</a>	0.00024410861054650565	2649
38	<a href="#">seher.no</a>	0.0002388559903875208	2592
39	<a href="#">avisa-st.no</a>	0.0002378423268480676	2581
40	<a href="#">rbnett.no</a>	0.0002371972682320519	2574
41	<a href="#">innherred.no</a>	0.00023599930223087995	2561
42	<a href="#">framtidinord.no</a>	0.00023194464807306707	2517
43	<a href="#">driva.no</a>	0.0002295487160707231	2491
44	<a href="#">tronderbladet.no</a>	0.00022706063283751978	2464
45	<a href="#">dolen.no</a>	0.00022613912052892596	2454
46	<a href="#">banett.no</a>	0.00022512545698947272	2443
47	<a href="#">fjt.no</a>	0.0002241117934500195	2432
48	<a href="#">fjordabladet.no</a>	0.00022346673483400383	2425
49	<a href="#">fjuken.no</a>	0.0002214394077550974	2403
50	<a href="#">vaganavisa.no</a>	0.00022116295406251926	2400



Unique URL statistics from Common Crawl belonging to Norwegian magazine domains			
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="http://gamer.no">gamer.no</a>	0.0010585411888817328	11487
2	<a href="http://journalisten.no">journalisten.no</a>	0.00038417848145276784	4169
3	<a href="http://forsvaretsforum.no">forsvaretsforum.no</a>	0.00037689853421487657	4090
4	<a href="http://aperitif.no">aperitif.no</a>	0.00036445811804885987	3955
5	<a href="http://khrono.no">khrono.no</a>	0.0003556115998863591	3859
6	<a href="http://sykepleien.no">sykepleien.no</a>	0.00034897671126448354	3787
7	<a href="http://tidsskriftet.no">tidsskriftet.no</a>	0.0003480551989558897	3777
8	<a href="http://utdanningsnytt.no">utdanningsnytt.no</a>	0.000317092385387137	3441
9	<a href="http://barnehage.no">barnehage.no</a>	0.0002978327781375259	3232
10	<a href="http://frifagbevegelse.no">frifagbevegelse.no</a>	0.00028207491766057144	3061
11	<a href="http://bygg.no">bygg.no</a>	0.00026521124241330433	2878
12	<a href="http://dagensmedisin.no">dagensmedisin.no</a>	0.00025710193409767863	2790
13	<a href="http://agendamagasinet.no">agendamagasinet.no</a>	0.00023959320023439587	2600
14	<a href="http://m24.no">m24.no</a>	0.00023664436084689562	2568
15	<a href="http://njff.no">njff.no</a>	0.0002301937746867388	2498
16	<a href="http://naturpress.no">naturpress.no</a>	0.00022180801267853494	2407
17	<a href="http://fagbladet.no">fagbladet.no</a>	0.00021950423190705037	2382
18	<a href="http://shifter.no">shifter.no</a>	0.00021360655313204985	2318
19	<a href="http://horecanytt.no">horecanytt.no</a>	0.00020927544528165885	2271
20	<a href="http://psykologtidsskriftet.no">psykologtidsskriftet.no</a>	0.00019830944880939227	2152
21	<a href="http://gardsdrift.no">gardsdrift.no</a>	0.00019416264342072003	2107
22	<a href="http://hageselskapet.no">hageselskapet.no</a>	0.00018144577356212519	1969
23	<a href="http://motor.no">motor.no</a>	0.00018116931986954704	1966
24	<a href="http://farmatid.no">farmatid.no</a>	0.00017923414402149998	1945
25	<a href="http://universitetsavisa.no">universitetsavisa.no</a>	0.0001732443140156401	1880

26	<a href="http://batmagasinet.no">batmagasinet.no</a>	0.000170571928320718	1851
27	<a href="http://skog.no">skog.no</a>	0.0001634762835445455	1774
28	<a href="http://seilmagasinet.no">seilmagasinet.no</a>	0.00015453761415118534	1677
29	<a href="http://tannegetidende.no">tannegetidende.no</a>	0.0001476262718367316	1602
30	<a href="http://biogassbransjen.no">biogassbransjen.no</a>	0.00014744196937501285	1600
31	<a href="http://united.no">united.no</a>	0.00014430882752579382	1566
32	<a href="http://kontekst.no">kontekst.no</a>	0.00014347946644805936	1557
33	<a href="http://trav365.no">trav365.no</a>	0.0001423736516777468	1545
34	<a href="http://renholdsnytt.no">renholdsnytt.no</a>	0.000124864917814464	1355
35	<a href="http://forskerforum.no">forskerforum.no</a>	0.00011823002919258842	1283
36	<a href="http://politiforum.no">politiforum.no</a>	0.00011657130703711953	1265
37	<a href="http://hmsmagasinet.no">hmsmagasinet.no</a>	0.00011518903857422878	1250
38	<a href="http://tgn.no">tgn.no</a>	0.0001134381651879005	1231
39	<a href="http://cnytt.no">cnytt.no</a>	0.0001107657794929784	1202
40	<a href="http://avfallsbransjen.no">avfallsbransjen.no</a>	$9.565297763203958 \times 10^{-5}$	1038
41	<a href="http://bobilverden.no">bobilverden.no</a>	$9.279628947539871 \times 10^{-5}$	1007
42	<a href="http://kom24.no">kom24.no</a>	$8.708291316211696 \times 10^{-5}$	945
43	<a href="http://fontene.no">fontene.no</a>	$8.699076193125758 \times 10^{-5}$	944
44	<a href="http://subjekt.no">subjekt.no</a>	$8.514773731406992 \times 10^{-5}$	924
45	<a href="http://advokatbladet.no">advokatbladet.no</a>	$8.31204102351635 \times 10^{-5}$	902
46	<a href="http://tidsskriftetmuseum.no">tidsskriftetmuseum.no</a>	$8.173814177227275 \times 10^{-5}$	887
47	<a href="http://bokogbibliotek.no">bokogbibliotek.no</a>	$7.6024765458991 \times 10^{-5}$	825
48	<a href="http://panoramanyheter.no">panoramanyheter.no</a>	$7.32602285332095 \times 10^{-5}$	795
49	<a href="http://fysioterapeuten.no">fysioterapeuten.no</a>	$7.031138914570925 \times 10^{-5}$	763
50	<a href="http://at.no">at.no</a>	$6.957417929883418 \times 10^{-5}$	755

Unique URL statistics from mC4 belonging to Norwegian news media domains			
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="http://nrk.no">nrk.no</a>	0.015750287800566715	400091
2	<a href="http://dagbladet.no">dagbladet.no</a>	0.013066143760570715	331908
3	<a href="http://aftenposten.no">aftenposten.no</a>	0.011145439366346275	283118
4	<a href="http://nettavisen.no">nettavisen.no</a>	0.009096123755562475	231061
5	<a href="http://vg.no">vg.no</a>	0.00365902257286286	92947
6	<a href="http://abcnyheter.no">abcnyheter.no</a>	0.003618159872284771	91909
7	<a href="http://kk.no">kk.no</a>	0.0029987632143891503	76175
8	<a href="http://klikk.no">klikk.no</a>	0.0029987632143891503	76175
9	<a href="http://dinside.no">dinside.no</a>	0.0025340385705314027	64370
10	<a href="http://tek.no">tek.no</a>	0.002431803085559055	61773
11	<a href="http://tv2.no">tv2.no</a>	0.0022655178762701835	57549
12	<a href="http://bt.no">bt.no</a>	0.0020665188864607033	52494
13	<a href="http://sa.no">sa.no</a>	0.002051953183942502	52124
14	<a href="http://adressa.no">adressa.no</a>	0.002051953183942502	52124



15	<a href="http://dagsavisen.no">dagsavisen.no</a>	0.0016419089746733534	41708
16	<a href="http://cw.no">cw.no</a>	0.001471057220811208	37368
17	<a href="http://dn.no">dn.no</a>	0.0013938196307011783	35406
18	<a href="http://aftenbladet.no">aftenbladet.no</a>	0.0013302423075474078	33791
19	<a href="http://itromso.no">itromso.no</a>	0.0012679247208276436	32208
20	<a href="http://fyn.no">fyn.no</a>	0.001188954993120855	30202
21	<a href="http://seher.no">seher.no</a>	0.0011768300299435414	29894
22	<a href="http://forskning.no">forskning.no</a>	0.0011476986249071388	29154
23	<a href="http://itavisen.no">itavisen.no</a>	0.001040030526563137	26419
24	<a href="http://document.no">document.no</a>	0.0010253860905178104	26047
25	<a href="http://kode24.no">kode24.no</a>	0.0008999635818070282	22861
26	<a href="http://e24.no">e24.no</a>	0.0008999635818070282	22861
27	<a href="http://gd.no">gd.no</a>	0.0007013188928696123	17815
28	<a href="http://nab.no">nab.no</a>	0.000686753190351411	17445
29	<a href="http://smp.no">smp.no</a>	0.0006416782460721122	16300
30	<a href="http://dagen.no">dagen.no</a>	0.0006088857320243779	15467
31	<a href="http://lp.no">lp.no</a>	0.0005992408749515149	15222
32	<a href="http://solabladet.no">solabladet.no</a>	0.0005981779723353218	15195
33	<a href="http://dalane-tidende.no">dalane-tidende.no</a>	0.0005076737829046601	12896
34	<a href="http://rbnett.no">rbnett.no</a>	0.0004962180547079126	12605
35	<a href="http://namdalsavisa.no">namdalsavisa.no</a>	0.0004943678168204654	12558
36	<a href="http://h-a.no">h-a.no</a>	0.0004743695009306106	12050
37	<a href="http://vol.no">vol.no</a>	0.00045708749172658253	11611
38	<a href="http://vol.no">vol.no</a>	0.00045708749172658253	11611
39	<a href="http://op.no">op.no</a>	0.00044594669763833666	11328
40	<a href="http://verdidebatt.no">verdidebatt.no</a>	0.0004228384074270281	10741
41	<a href="http://f-b.no">f-b.no</a>	0.0004196890663420116	10661
42	<a href="http://ta.no">ta.no</a>	0.00039815544667321125	10114

43	<a href="http://ba.no">ba.no</a>	0.0003932739679914357	9990
44	<a href="http://nordlys.no">nordlys.no</a>	0.0003877626210926568	9850
45	<a href="http://klassekampen.no">klassekampen.no</a>	0.0003860304834958977	9806
46	<a href="http://oa.no">oa.no</a>	0.00037366931973720793	9492
47	<a href="http://kvinneguiden.no">kvinneguiden.no</a>	0.0003694177092724357	9384
48	<a href="http://rb.no">rb.no</a>	0.00035272620152184824	8960
49	<a href="http://sandnesposten.no">sandnesposten.no</a>	0.00035063976305302477	8907
50	<a href="http://altaposten.no">altaposten.no</a>	0.0003436718459024258	8730

Unique URL statistics from mC4 belonging to Norwegian magazine domains			
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="http://bygg.no">bygg.no</a>	0.0022997275938061753	58418
2	<a href="http://frifagbevegelse.no">frifagbevegelse.no</a>	0.0011515172009727212	29251
3	<a href="http://gamer.no">gamer.no</a>	0.0009260637460491024	23524
4	<a href="http://tidsskriftet.no">tidsskriftet.no</a>	0.0007191520367635183	18268
5	<a href="http://psykologtidsskriftet.no">psykologtidsskriftet.no</a>	0.0007191520367635183	18268
6	<a href="http://aperitif.no">aperitif.no</a>	0.0005226331530584885	13276
7	<a href="http://at.no">at.no</a>	0.00044460822767720465	11294
8	<a href="http://khrono.no">khrono.no</a>	0.0004049265300059967	10286
9	<a href="http://seilmagasinet.no">seilmagasinet.no</a>	0.00037536209057040433	9535
10	<a href="http://nhi.no">nhi.no</a>	0.00035213570006840765	8945
11	<a href="http://journalisten.no">journalisten.no</a>	0.000327964507240906	8331
12	<a href="http://fagbladet.no">fagbladet.no</a>	0.0003087141598587426	7842
13	<a href="http://politiforum.no">politiforum.no</a>	0.0002903692480385215	7376
14	<a href="http://dagensmedisin.no">dagensmedisin.no</a>	0.0002558445963940281	6499
15	<a href="http://skog.no">skog.no</a>	0.00025001831538674755	6351
16	<a href="http://universitetsavisa.no">universitetsavisa.no</a>	0.00021057281829691587	5349
17	<a href="http://sykepleien.no">sykepleien.no</a>	0.00019474737934470793	4947
18	<a href="http://utdanningsnytt.no">utdanningsnytt.no</a>	0.0001895509665544307	4815
19	<a href="http://shifter.no">shifter.no</a>	0.00016927708331963698	4300
20	<a href="http://barnehage.no">barnehage.no</a>	0.00016419877082004788	4171
21	<a href="http://united.no">united.no</a>	0.00016219106587834985	4120
22	<a href="http://trav365.no">trav365.no</a>	0.00014857016568565348	3774
23	<a href="http://mtlogistikk.no">mtlogistikk.no</a>	0.0001417203488257426	3600
24	<a href="http://forsvaretsforum.no">forsvaretsforum.no</a>	0.00013774430570590926	3499
25	<a href="http://fontene.no">fontene.no</a>	0.00013581533429133665	3450

26	<a href="http://kommunal-rapport.no">kommunal-rapport.no</a>	0.00012849311626867327	3264
27	<a href="http://horecanytt.no">horecanytt.no</a>	0.00012266683526139275	3116
28	<a href="http://njff.no">njff.no</a>	0.0001160532189828581	2948
29	<a href="http://byggmesteren.as">byggmesteren.as</a>	0.00010924276888650992	2775
30	<a href="http://forskerforum.no">forskerforum.no</a>	0.00010577849369299176	2687
31	<a href="http://motor.no">motor.no</a>	$9.203949320960727 \times 10^{-5}$	2338
32	<a href="http://ledernytt.no">ledernytt.no</a>	$8.066249853998515 \times 10^{-5}$	2049
33	<a href="http://gardsdrift.no">gardsdrift.no</a>	$7.7198223346467 \times 10^{-5}$	1961
34	<a href="http://agendamagasinet.no">agendamagasinet.no</a>	$7.034840648655611 \times 10^{-5}$	1787
35	<a href="http://rett24.no">rett24.no</a>	$6.700223158372608 \times 10^{-5}$	1702
36	<a href="http://advokatbladet.no">advokatbladet.no</a>	$6.645109689384819 \times 10^{-5}$	1688
37	<a href="http://hk-nytt.no">hk-nytt.no</a>	$6.601806249465842 \times 10^{-5}$	1677
38	<a href="http://bobilverden.no">bobilverden.no</a>	$6.353795639020793 \times 10^{-5}$	1614
39	<a href="http://m24.no">m24.no</a>	$6.117595057644556 \times 10^{-5}$	1554
40	<a href="http://batmagasinet.no">batmagasinet.no</a>	$5.696370687523598 \times 10^{-5}$	1447
41	<a href="http://elektronikknett.no">elektronikknett.no</a>	$5.641257218535809 \times 10^{-5}$	1433
42	<a href="http://energiteknikk.net">energiteknikk.net</a>	$5.460170106147360 \times 10^{-5}$	1387
43	<a href="http://bakeri.net">bakeri.net</a>	$5.45623342979109 \times 10^{-5}$	1386
44	<a href="http://fysioterapeuten.no">fysioterapeuten.no</a>	$5.255462935621287 \times 10^{-5}$	1335
45	<a href="http://naturpress.no">naturpress.no</a>	$4.869668652706766 \times 10^{-5}$	1237
46	<a href="http://parcferme.no">parcferme.no</a>	$4.7594417147311883 \times 10^{-5}$	1209
47	<a href="http://hageselskapet.no">hageselskapet.no</a>	$4.696454893030858 \times 10^{-5}$	1193
48	<a href="http://arkitektur.no">arkitektur.no</a>	$4.582291278699010 \times 10^{-5}$	1164
49	<a href="http://subjekt.no">subjekt.no</a>	$4.542924515136304 \times 10^{-5}$	1154

50	<a href="http://tannlegetidende.no">tannlegetidende.no</a>	$4.385457460885479 \times 10^{-5}$	1114
----	--	------------------------------------	------

Unique URL statistics from OpenWebText belonging to Norwegian news media domains			
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="http://nrk.no">nrk.no</a>	0.18772923235313044	5835
2	<a href="http://aftenposten.no">aftenposten.no</a>	0.10443343414194711	3246
3	<a href="http://bladet.no">bladet.no</a>	0.09349462711537224	2906
4	<a href="http://vg.no">vg.no</a>	0.0819445338137829	2547
5	<a href="http://tv2.no">tv2.no</a>	0.03130429187310984	973
6	<a href="http://nettavisen.no">nettavisen.no</a>	0.027733093108551574	862
7	<a href="http://bt.no">bt.no</a>	0.024322759153207644	756
8	<a href="http://e24.no">e24.no</a>	0.021266327778135256	661
9	<a href="http://dn.no">dn.no</a>	0.014928254295090405	464
10	<a href="http://adressa.no">adressa.no</a>	0.013062222508204106	406
11	<a href="http://ba.no">ba.no</a>	0.011196190721317804	348
12	<a href="http://dagsavisen.no">dagsavisen.no</a>	0.009587542629174441	298
13	<a href="http://klassekampen.no">klassekampen.no</a>	0.009104948201531434	283
14	<a href="http://forskning.no">forskning.no</a>	0.008107586384402548	252

15	<a href="#">abcnyheter.no</a>	0.00807541342255968	251
16	<a href="#">vgtv.no</a>	0.006691976063316389	208
17	<a href="#">osloby.no</a>	0.0058876520172447076	183
18	<a href="#">p3.no</a>	0.0053728846277588315	167
19	<a href="#">ta.no</a>	0.005051155009330159	157
20	<a href="#">blogg.no</a>	0.0048581172382729555	151
21	<a href="#">skam.p3.no</a>	0.0038607554211440706	120
22	<a href="#">nordlys.no</a>	0.0036998906119297343	115
23	<a href="#">hegnar.no</a>	0.0036998906119297343	115
24	<a href="#">na24.no</a>	0.003571198764558265	111
25	<a href="#">an.no</a>	0.0034746798790296633	108
26	<a href="#">sa.no</a>	0.002992085451386655	93
27	<a href="#">dt.no</a>	0.0028312206421723185	88
28	<a href="#">dinside.no</a>	0.0025738369474293805	80
29	<a href="#">budstikka.no</a>	0.0024773180619007784	77
30	<a href="#">document.no</a>	0.0023164532526864425	72
31	<a href="#">minervanett.no</a>	0.0021877614053149733	68
32	<a href="#">side3.no</a>	0.0021877614053149733	68
33	<a href="#">fvn.no</a>	0.002059069557943504	64
34	<a href="#">tek.no</a>	0.00199472363425777	62
35	<a href="#">vl.no</a>	0.001769512901357699	55
36	<a href="#">smp.no</a>	0.00164082105398623	51
37	<a href="#">itavisen.no</a>	0.0016086480921433627	50
38	<a href="#">itromso.no</a>	0.001512129206614761	47
39	<a href="#">rb.no</a>	0.0013190914355575574	41
40	<a href="#">kk.no</a>	0.0013190914355575574	41
41	<a href="#">sb.no</a>	0.0012225725500289557	38
42	<a href="#">op.no</a>	0.001029534778971752	32

43	<a href="#">tb.no</a>	0.0009651888552860176	30
44	<a href="#">rbnett.no</a>	0.0009651888552860176	30
45	<a href="#">nationen.no</a>	0.0008043240460716813	25
46	<a href="#">faktisk.no</a>	0.0007721510842288141	24
47	<a href="#">f-b.no</a>	0.0007078051605430796	22
48	<a href="#">h-a.no</a>	0.0006756321987002123	21
49	<a href="#">varden.no</a>	0.0006756321987002123	21
50	<a href="#">dagen.no</a>	0.0005469403513287433	17

Unique URL statistics from OpenWebText belonging to Norwegian magazine domains			
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="#">universitetsavisa.no</a>	0.0021234154816292387	66
2	<a href="#">frifagbevegelse.no</a>	0.0019625506724149024	61
3	<a href="#">gamer.no</a>	0.001383437359243292	43
4	<a href="#">journalisten.no</a>	0.0010617077408146193	33
5	<a href="#">at.no</a>	0.000997361817128885	31
6	<a href="#">politiforum.no</a>	0.0009008429316002831	28
7	<a href="#">united.no</a>	0.0007399781223859468	23
8	<a href="#">dagensmedisin.no</a>	0.0005791133131716106	18
9	<a href="#">agendamagasinet.no</a>	0.0005791133131716106	18
10	<a href="#">tidsskriftet.no</a>	0.0003539025802715398	11
11	<a href="#">fagbladet.no</a>	0.0002895566565858053	9
12	<a href="#">khrono.no</a>	0.000257383694742938	8
13	<a href="#">rett24.no</a>	0.000257383694742938	8
14	<a href="#">motor.no</a>	0.00022521073290007079	7
15	<a href="#">m24.no</a>	0.00019303777105720353	6
16	<a href="#">bygg.no</a>	0.00019303777105720353	6



17	<a href="http://forsvaretsforum.no">forsvaretsforum.no</a>	0.00016086480921433628	5
18	<a href="http://skog.no">skog.no</a>	0.00016086480921433628	5
19	<a href="http://sykepleien.no">sykepleien.no</a>	0.000128691847371469	4
20	<a href="http://aperitif.no">aperitif.no</a>	$9.651888552860177 \times 10^{-5}$	3
21	<a href="http://shifter.no">shifter.no</a>	$9.651888552860177 \times 10^{-5}$	3
22	<a href="http://hk-nytt.no">hk-nytt.no</a>	$6.43459236857345 \times 10^{-5}$	2
23	<a href="http://forskerforum.no">forskerforum.no</a>	$6.43459236857345 \times 10^{-5}$	2
24	<a href="http://psykologtidsskriftet.no">psykologtidsskriftet.no</a>	$6.43459236857345 \times 10^{-5}$	2
25	<a href="http://advokatbladet.no">advokatbladet.no</a>	$6.43459236857345 \times 10^{-5}$	2
26	<a href="http://kommunal-rapport.no">kommunal-rapport.no</a>	$6.43459236857345 \times 10^{-5}$	2
27	<a href="http://utdanningsnytt.no">utdanningsnytt.no</a>	$6.43459236857345 \times 10^{-5}$	2
28	<a href="http://naturpress.no">naturpress.no</a>	$3.217296184286725 \times 10^{-5}$	1
29	<a href="http://ledernytt.no">ledernytt.no</a>	$3.217296184286725 \times 10^{-5}$	1
30	<a href="http://handikapnytt.no">handikapnytt.no</a>	$3.217296184286725 \times 10^{-5}$	1
31	<a href="http://subjekt.no">subjekt.no</a>	$3.217296184286725 \times 10^{-5}$	1
32	<a href="http://njff.no">njff.no</a>	$3.217296184286725 \times 10^{-5}$	1
33	<a href="http://barnehage.no">barnehage.no</a>	$3.217296184286725 \times 10^{-5}$	1
34	<a href="http://fontene.no">fontene.no</a>	$3.217296184286725 \times 10^{-5}$	1

Unique URL statistics from OpenWebText 2 belonging to Norwegian news media domains			
Ranking	url_host_registered_domain	Share of Norwegian publisher domain URLs	count
1	<a href="http://nrk.no">nrk.no</a>	0.2635747323069125	8320

2	<a href="http://aftenposten.no">aftenposten.no</a>	0.1076474687955395	3398
3	<a href="http://bladet.no">bladet.no</a>	0.1042894253310524	3292
4	<a href="http://vg.no">vg.no</a>	0.09785845530000634	3089
5	<a href="http://tv2.no">tv2.no</a>	0.03605144776024837	1138
6	<a href="http://e24.no">e24.no</a>	0.028194893239561553	890
7	<a href="http://nettavisen.no">nettavisen.no</a>	0.0270861053031743	855
8	<a href="http://bt.no">bt.no</a>	0.022302477349046442	704
9	<a href="http://dn.no">dn.no</a>	0.017328771462966484	547
10	<a href="http://dagsavisen.no">dagsavisen.no</a>	0.013558892479249825	428
11	<a href="http://ba.no">ba.no</a>	0.011436355572451372	361
12	<a href="http://abcnyheter.no">abcnyheter.no</a>	0.010771082810619021	340
13	<a href="http://forskning.no">forskning.no</a>	0.01073940315529367	339
14	<a href="http://klassekampen.no">klassekampen.no</a>	0.00978901349553317	309
15	<a href="http://adressa.no">adressa.no</a>	0.006621047962998163	209
16	<a href="http://ta.no">ta.no</a>	0.0061775327884432615	195
17	<a href="http://osloby.no">osloby.no</a>	0.00554393968193626	175
18	<a href="http://dinside.no">dinside.no</a>	0.0034847620857885067	110
19	<a href="http://document.no">document.no</a>	0.0034530824304631566	109
20	<a href="http://an.no">an.no</a>	0.0034530824304631566	109
21	<a href="http://hegnar.no">hegnar.no</a>	0.0033580434644871066	106
22	<a href="http://skam.p3.no">skam.p3.no</a>	0.0032946841538364063	104
23	<a href="http://na24.no">na24.no</a>	0.0032946841538364063	104
24	<a href="http://nordlys.no">nordlys.no</a>	0.003231324843185706	102
25	<a href="http://minervanett.no">minervanett.no</a>	0.003104606221884306	98
26	<a href="http://sa.no">sa.no</a>	0.0020591775961477537	65
27	<a href="http://budstikka.no">budstikka.no</a>	0.0019324589748463538	61
28	<a href="http://dt.no">dt.no</a>	0.0019007793195210036	60
29	<a href="http://itromso.no">itromso.no</a>	0.0018057403535449534	57

30	<a href="http://itavisen.no">itavisen.no</a>	0.0016790217322435533	53
31	<a href="http://kk.no">kk.no</a>	0.0014572641449661027	46
32	<a href="http://faktisk.no">faktisk.no</a>	0.0014572641449661027	46
33	<a href="http://sb.no">sb.no</a>	0.0014255844896407528	45
34	<a href="http://fvn.no">fvn.no</a>	0.0014255844896407528	45
35	<a href="http://rb.no">rb.no</a>	0.0013939048343154026	44
36	<a href="http://tb.no">tb.no</a>	0.0011404675917126021	36
37	<a href="http://vl.no">vl.no</a>	0.0009503896597605018	30
38	<a href="http://nationen.no">nationen.no</a>	0.0009187100044351518	29
39	<a href="http://op.no">op.no</a>	0.0008553506937844516	27
40	<a href="http://f-b.no">f-b.no</a>	0.0007286320724830514	23
41	<a href="http://smp.no">smp.no</a>	0.0007286320724830514	23
42	<a href="http://h-avis.no">h-avis.no</a>	0.0005702337958563011	18
43	<a href="http://ht.no">ht.no</a>	0.0005702337958563011	18
44	<a href="http://aasanetidende.no">aasanetidende.no</a>	0.000538554140530951	17
45	<a href="http://finnmarkdagblad.no">finnmarkdagblad.no</a>	0.000538554140530951	17
46	<a href="http://varden.no">varden.no</a>	0.000538554140530951	17
47	<a href="http://blogg.no">blogg.no</a>	0.000538554140530951	17
48	<a href="http://oa.no">oa.no</a>	0.000506874485205601	16
49	<a href="http://ifinnmark.no">ifinnmark.no</a>	0.000506874485205601	16
50	<a href="http://altaposten.no">altaposten.no</a>	0.0004751948298802509	15

Unique URL statistics from OpenWebText 2 belonging to Norwegian magazine domains			
Ranking	url_host_registered_domain	ratio	count

1	<a href="http://frifagbevegelse.no">frifagbevegelse.no</a>	0.002756130013305455	87
2	<a href="http://gamer.no">gamer.no</a>	0.0017107013875689032	54
3	<a href="http://journalisten.no">journalisten.no</a>	0.001520623455616803	48
4	<a href="http://universitetsavisa.no">universitetsavisa.no</a>	0.001520623455616803	48
5	<a href="http://united.no">united.no</a>	0.0009187100044351518	29
6	<a href="http://agendamagasinet.no">agendamagasinet.no</a>	0.0008236710384591016	26
7	<a href="http://politiforum.no">politiforum.no</a>	0.0007919913831337515	25
8	<a href="http://rett24.no">rett24.no</a>	0.0006335931065070012	20
9	<a href="http://dagensmedisin.no">dagensmedisin.no</a>	0.000538554140530951	17
10	<a href="http://khrono.no">khrono.no</a>	0.0004118355192295508	13
11	<a href="http://psykologtidsskriftet.no">psykologtidsskriftet.no</a>	0.00034847620857885066	11
12	<a href="http://bygg.no">bygg.no</a>	0.00034847620857885066	11
13	<a href="http://fagbladet.no">fagbladet.no</a>	0.00028511689792815053	9
14	<a href="http://at.no">at.no</a>	0.0002534372426028005	8
15	<a href="http://m24.no">m24.no</a>	0.0002534372426028005	8
16	<a href="http://forsvaretsforum.no">forsvaretsforum.no</a>	0.0002217575872774504	7
17	<a href="http://sykepleien.no">sykepleien.no</a>	0.00019007793195210037	6
18	<a href="http://shifter.no">shifter.no</a>	0.0001583982766267503	5
19	<a href="http://subjekt.no">subjekt.no</a>	0.0001583982766267503	5
20	<a href="http://utdanningsnytt.no">utdanningsnytt.no</a>	0.00012671862130140025	4
21	<a href="http://advokatbladet.no">advokatbladet.no</a>	0.00012671862130140025	4
22	<a href="http://forskerforum.no">forskerforum.no</a>	$9.503896597605019 \times 10^{-5}$	3
23	<a href="http://aperitif.no">aperitif.no</a>	$9.503896597605019 \times 10^{-5}$	3
24	<a href="http://motor.no">motor.no</a>	$9.503896597605019 \times 10^{-5}$	3
25	<a href="http://fontene.no">fontene.no</a>	$6.335931065070012 \times 10^{-5}$	2
26	<a href="http://hk-nytt.no">hk-nytt.no</a>	$6.335931065070012 \times 10^{-5}$	2
27	<a href="http://kommunal-rapport.no">kommunal-rapport.no</a>	$6.335931065070012 \times 10^{-5}$	2

28	<a href="http://naturpress.no">naturpress.no</a>	$3.167965532535006 \times 10^{-5}$	1
29	<a href="http://ledernytt.no">ledernytt.no</a>	$3.167965532535006 \times 10^{-5}$	1
30	<a href="http://handikapnytt.no">handikapnytt.no</a>	$3.167965532535006 \times 10^{-5}$	1
31	<a href="http://parcferme.no">parcferme.no</a>	$3.167965532535006 \times 10^{-5}$	1
32	<a href="http://nhi.no">nhi.no</a>	$3.167965532535006 \times 10^{-5}$	1
33	<a href="http://skog.no">skog.no</a>	$3.167965532535006 \times 10^{-5}$	1

## Appendix 2. Unique premium Amedia URLs in the mC4 dataset

1	<a href="https://www.fremover.no/lokale-nyheter/narvik/ikke-ga-her-det-er-en-darlig-id/s/5-17-64507">https://www.fremover.no/lokale-nyheter/narvik/ikke-ga-her-det-er-en-darlig-id/s/5-17-64507</a>
2	<a href="https://www.rablad.no/kultur-og-underholdning/nyheter/tinn/norske-talenter-og-kapleik-i-bo/s/5-90-331">https://www.rablad.no/kultur-og-underholdning/nyheter/tinn/norske-talenter-og-kapleik-i-bo/s/5-90-331</a>
3	<a href="https://www.oa.no/valdres/nyheter/forbrukersporsmal/oa-omtalte-de-enorme-prisforskjellene-se-hva-som-skjedde/s/5-35-774896">https://www.oa.no/valdres/nyheter/forbrukersporsmal/oa-omtalte-de-enorme-prisforskjellene-se-hva-som-skjedde/s/5-35-774896</a>
4	<a href="https://www.ranablad.no/kirke/sv/valg/sokneprest-topper-sv-lista/s/5-42-30959">https://www.ranablad.no/kirke/sv/valg/sokneprest-topper-sv-lista/s/5-42-30959</a>
5	<a href="https://www.austagderblad.no/ny-vognhall-i-egddalen-skrinlagt-av-formannskapet/s/5-6-251215">https://www.austagderblad.no/ny-vognhall-i-egddalen-skrinlagt-av-formannskapet/s/5-6-251215</a>
6	<a href="https://www.an.no/nyheter/julie-33-mistet-moren-i-selvord-jeg-respekterer-mammas-valg/s/5-4-99581">https://www.an.no/nyheter/julie-33-mistet-moren-i-selvord-jeg-respekterer-mammas-valg/s/5-4-99581</a>
7	<a href="https://www.rb.no/nyheter/eidsvoll/sosiale-forhold/jeg-savner-a-kunne-ta-han-i-handa-og-a-se-han-inn-i-oynene/s/5-43-103344">https://www.rb.no/nyheter/eidsvoll/sosiale-forhold/jeg-savner-a-kunne-ta-han-i-handa-og-a-se-han-inn-i-oynene/s/5-43-103344</a>
8	<a href="https://www.rb.no/lhl-sykehuset/medisin-og-helse/jessheim/det-var-en-grunn-til-at-ingen-tok-det-forste-spadesticket/s/5-43-126796">https://www.rb.no/lhl-sykehuset/medisin-og-helse/jessheim/det-var-en-grunn-til-at-ingen-tok-det-forste-spadesticket/s/5-43-126796</a>
9	<a href="https://www.budstikka.no/teater/snehvit/barum-barneteater/snart-i-balanse-med-snehvit/s/5-55-411174">https://www.budstikka.no/teater/snehvit/barum-barneteater/snart-i-balanse-med-snehvit/s/5-55-411174</a>
10	<a href="https://www.h-avis.no/nyheter/karmoy/oppvekst/tor-ikke-fortelle-ungene-at-barnehagen-kan-bli-lagt-ned/s/5-62-730947">https://www.h-avis.no/nyheter/karmoy/oppvekst/tor-ikke-fortelle-ungene-at-barnehagen-kan-bli-lagt-ned/s/5-62-730947</a>

11	<a href="https://www.gd.no/amerikansk-fotball/sport/magasiner/fotballens-nye-lederduo/s/5-18-73428">https://www.gd.no/amerikansk-fotball/sport/magasiner/fotballens-nye-lederduo/s/5-18-73428</a>
12	<a href="https://www.ha-halden.no/hans-petter-kjoge/kommunevalget-2015/nyheter/frp-blir-jokeren-i-thrillervalget/o/5-20-72271">https://www.ha-halden.no/hans-petter-kjoge/kommunevalget-2015/nyheter/frp-blir-jokeren-i-thrillervalget/o/5-20-72271</a>
13	<a href="https://www.avis-valdres.no/vang/vang-auto-service-as/e16/dette-skjer-med-vang-som-base/s/5-54-287814">https://www.avis-valdres.no/vang/vang-auto-service-as/e16/dette-skjer-med-vang-som-base/s/5-54-287814</a>
14	<a href="https://www.sb.no/vv24naringsliv/lov-og-rett/okonomi-og-naringsliv/langemyhr-mot-staten-i-august-kravet-kan-bli-minst-100-millioner-i-erstatning/s/5-83-103051">https://www.sb.no/vv24naringsliv/lov-og-rett/okonomi-og-naringsliv/langemyhr-mot-staten-i-august-kravet-kan-bli-minst-100-millioner-i-erstatning/s/5-83-103051</a>
15	<a href="https://www.rablad.no/nyheter/tinn/hardangervidda/polikerne-sa-ja-til-pabygg-pa-hukebu/s/5-90-11709">https://www.rablad.no/nyheter/tinn/hardangervidda/polikerne-sa-ja-til-pabygg-pa-hukebu/s/5-90-11709</a>
16	<a href="https://www.sognavis.no/rune-har-streva-med-a-tena-pengar-pa-svarthyll-i-25-ar-sa-kom-koronaen/s/5-115-394952">https://www.sognavis.no/rune-har-streva-med-a-tena-pengar-pa-svarthyll-i-25-ar-sa-kom-koronaen/s/5-115-394952</a>
17	<a href="https://www.fremover.no/hotell/lokale-nyheter/riksgr-nsen/luksushotellet-er-fullbooket-et-halvt-ar-for-det-er-ferdig-bygget/s/5-17-333419">https://www.fremover.no/hotell/lokale-nyheter/riksgr-nsen/luksushotellet-er-fullbooket-et-halvt-ar-for-det-er-ferdig-bygget/s/5-17-333419</a>
18	<a href="https://www.rablad.no/nyheter/hytte/tinn/gront-lys-igjen-til-stordalen-far-fly-i-verneoradet/s/5-90-6623">https://www.rablad.no/nyheter/hytte/tinn/gront-lys-igjen-til-stordalen-far-fly-i-verneoradet/s/5-90-6623</a>
19	<a href="https://www.firdaposten.no/artic-fjell/artic-shipping/ove-vilnes/dette-skipet-fraktar-alt/s/5-16-6756">https://www.firdaposten.no/artic-fjell/artic-shipping/ove-vilnes/dette-skipet-fraktar-alt/s/5-16-6756</a>
20	<a href="https://www.ringblad.no/pluss/jevnaker/helse/far-hjelp-til-a-mestre-hverdagen/s/5-45-87649">https://www.ringblad.no/pluss/jevnaker/helse/far-hjelp-til-a-mestre-hverdagen/s/5-45-87649</a>
21	<a href="https://www.budstikka.no/hoyhus/sandvika/boligutbygging/nye-sandvika-kan-fa-hoyhus-pa-65-meter/s/5-55-440921">https://www.budstikka.no/hoyhus/sandvika/boligutbygging/nye-sandvika-kan-fa-hoyhus-pa-65-meter/s/5-55-440921</a>
22	<a href="https://www.h-avis.no/nyheter/vi-legger-igjen-ca-70-millioner-kroner/s/5-62-95794">https://www.h-avis.no/nyheter/vi-legger-igjen-ca-70-millioner-kroner/s/5-62-95794</a>
23	<a href="https://www.gd.no/landbruk/gdnorddalen/nyheter/striker-gensere-til-kopplammene/s/5-18-77757">https://www.gd.no/landbruk/gdnorddalen/nyheter/striker-gensere-til-kopplammene/s/5-18-77757</a>
24	<a href="https://www.budstikka.no/eikeli-videregaende/skolerevy/skole/gir-revy-overskudd-til-jenteskole/s/5-55-457761">https://www.budstikka.no/eikeli-videregaende/skolerevy/skole/gir-revy-overskudd-til-jenteskole/s/5-55-457761</a>
25	<a href="https://www.f-b.no/kjopesenter/julegave/handel/ikke-alle-gavene-under-juletreet-blir-betalt-for/s/5-59-943697">https://www.f-b.no/kjopesenter/julegave/handel/ikke-alle-gavene-under-juletreet-blir-betalt-for/s/5-59-943697</a>

26	<a href="https://www.budstikka.no/sport/frisk-asker/torstein-nordrum/frisk-spiller-legger-skoytene-pa-hylla/s/5-55-471681">https://www.budstikka.no/sport/frisk-asker/torstein-nordrum/frisk-spiller-legger-skoytene-pa-hylla/s/5-55-471681</a>
27	<a href="https://www.kv.no/magasin/ny-glans-for-oyas-juvel/f/5-63-17165">https://www.kv.no/magasin/ny-glans-for-oyas-juvel/f/5-63-17165</a>
28	<a href="https://www.ba.no/brannmagasinet2018/fotball/sk-brann/koomson-er-kongen-av-thailand-pa-sosiale-medier/f/5-8-755803">https://www.ba.no/brannmagasinet2018/fotball/sk-brann/koomson-er-kongen-av-thailand-pa-sosiale-medier/f/5-8-755803</a>
29	<a href="https://www.indre.no/vedlikehold-til-10-millioner/s/5-25-6750">https://www.indre.no/vedlikehold-til-10-millioner/s/5-25-6750</a>
30	<a href="https://www.oa.no/kapp/mjosmuseet/kulturhistorie/de-fant-toten-storste-hemmelighet/s/5-35-70697">https://www.oa.no/kapp/mjosmuseet/kulturhistorie/de-fant-toten-storste-hemmelighet/s/5-35-70697</a>
31	<a href="https://www.ranablad.no/dette-gratistilbudet-hadde-800-deltakere-i-rana-i-fjor-og-satser-na-ogsa-pa-trana-en-veldig-fin-mate-a-komme-seg-ut-pa-tur/s/5-42-679330">https://www.ranablad.no/dette-gratistilbudet-hadde-800-deltakere-i-rana-i-fjor-og-satser-na-ogsa-pa-trana-en-veldig-fin-mate-a-komme-seg-ut-pa-tur/s/5-42-679330</a>
32	<a href="https://www.budstikka.no/foto/bildekunst/kultur-og-underholdning/okt-oppslutning-om-asker-og-barum-fotoklubb-snart-er-det-ikke-plass-til-flere/s/5-55-417165">https://www.budstikka.no/foto/bildekunst/kultur-og-underholdning/okt-oppslutning-om-asker-og-barum-fotoklubb-snart-er-det-ikke-plass-til-flere/s/5-55-417165</a>
33	<a href="https://www.budstikka.no/heggedal-skole/heggedal/asker/heggedal-helt-ferdig-til-skolestart/s/5-55-521904">https://www.budstikka.no/heggedal-skole/heggedal/asker/heggedal-helt-ferdig-til-skolestart/s/5-55-521904</a>
34	<a href="https://www.tk.no/nyheter/skarven-avslorte-ulovlig-garn/s/5-51-73348">https://www.tk.no/nyheter/skarven-avslorte-ulovlig-garn/s/5-51-73348</a>
35	<a href="https://www.aasavis.no/politisk-lederskap-mangler-respekt-for-kulturhistoriske-minner-i-as-sentrum-hvorfor/o/5-2-436745">https://www.aasavis.no/politisk-lederskap-mangler-respekt-for-kulturhistoriske-minner-i-as-sentrum-hvorfor/o/5-2-436745</a>
36	<a href="https://www.hadeland.no/folkeavstemning/hoyre/kommunestruktur/hoyre-topper-pa-hadeland-sier-nei/s/5-21-22004">https://www.hadeland.no/folkeavstemning/hoyre/kommunestruktur/hoyre-topper-pa-hadeland-sier-nei/s/5-21-22004</a>
37	<a href="https://www.ringsaker-blad.no/tarene-kom-da-anbjor-horte-nei-til-80-millioners-klage-brottum-er-en-profesjonell-og-framtidsrettet-eier/s/5-79-260469">https://www.ringsaker-blad.no/tarene-kom-da-anbjor-horte-nei-til-80-millioners-klage-brottum-er-en-profesjonell-og-framtidsrettet-eier/s/5-79-260469</a>
38	<a href="https://www.lofot-tidende.no/fotball/sport/leknes/tilbake-til-leknes-fk/s/5-28-20035">https://www.lofot-tidende.no/fotball/sport/leknes/tilbake-til-leknes-fk/s/5-28-20035</a>
39	<a href="https://www.ranablad.no/naringsliv/gallaria/rana/rana-i-fugleperspektiv/s/5-42-50256">https://www.ranablad.no/naringsliv/gallaria/rana/rana-i-fugleperspektiv/s/5-42-50256</a>