



**Danmarks
Nationalleksikon**

LEX.LLM – ET AI-BASERET SØGEINTERFACE TIL DANMARKS NATIONALE LEKSIKON

Ole Kaag Mølgaard (Lex) & Kristoffer Nielbo (Aarhus Universitet --
CHC)



AARHUS
UNIVERSITY

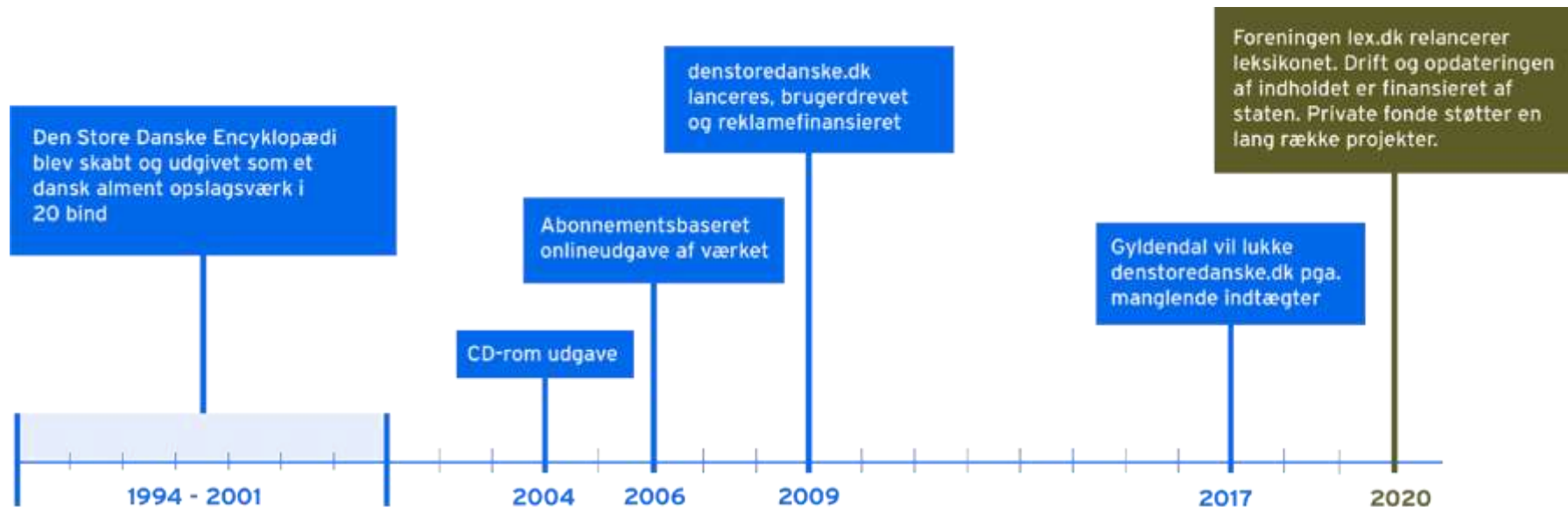


Danish
Foundation
Models



CENTER FOR
HUMANITIES
COMPUTING

Forhistorien den lange vej fra trykt opslagsværk til digitalt leksikon



Organisering en ideel forening og Lex er et public service tilbud

DANSKE
UNIVERSITETER



GYLDENDAL



G.E.C. Gads
Fond



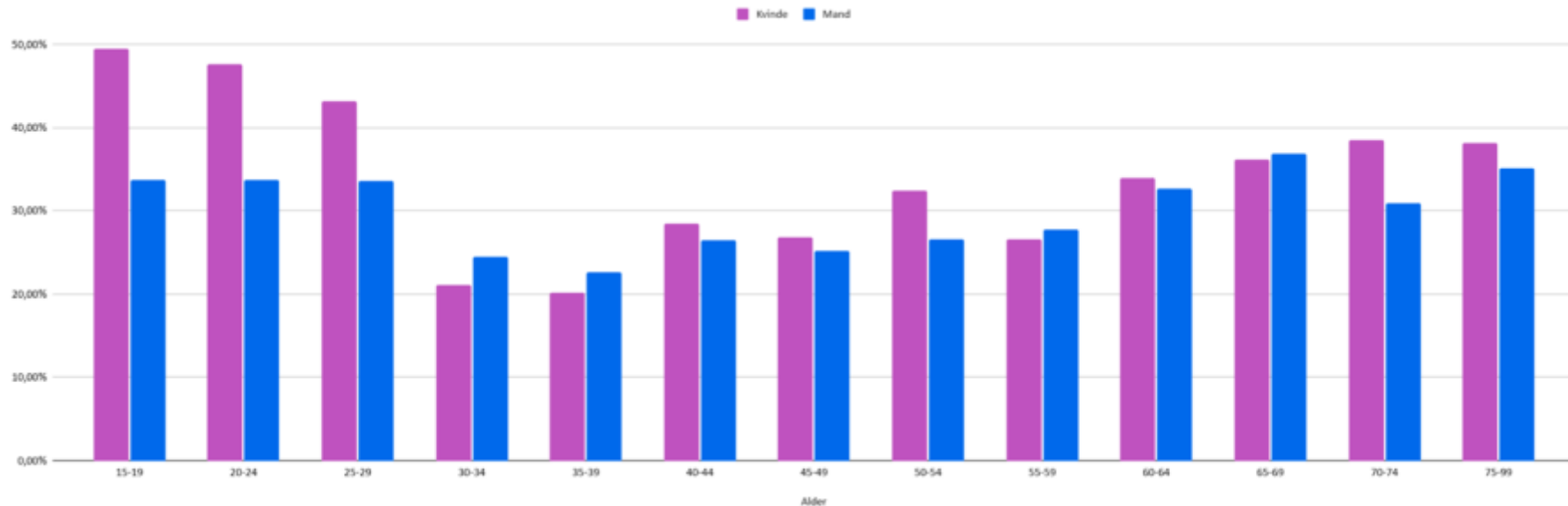
DET DANSKE SPROG-
OG LITTERATURSELSKAB

20 mio. besøg i 2024
42 mio. sidevisninger

En tredjedel af befolkningen

Et demokratisering af adgangen til viden

Dækning fordelt på alder og køn [målt igennem Dansk Online Index]



Indholdet

Lex rummer ca. 250.000 artikler med indhold fra 18 værker, der opdateres af ca. 1000 forskere og andre fagfolk

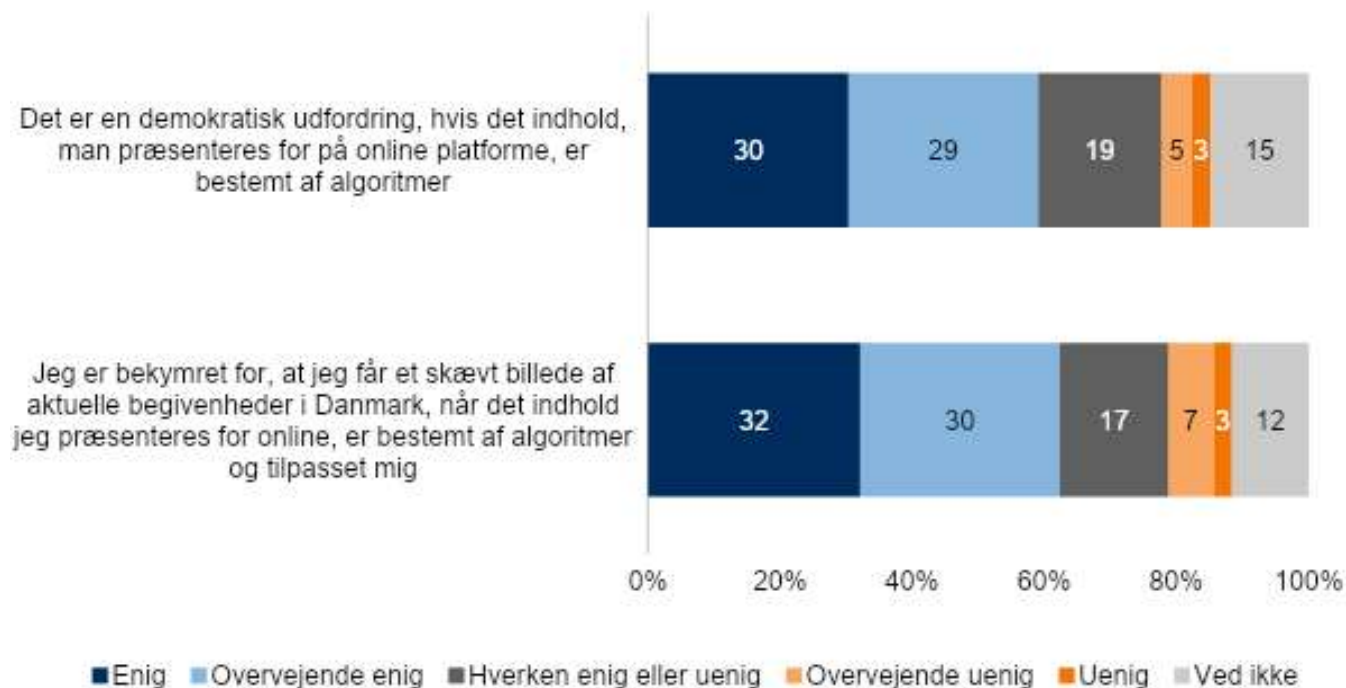
1. Den Store Danske Encyklopædi
2. Dansk Biografisk Leksikon
3. Symbolleksikon
4. Historien om børnelitteratur
5. Nordisk Mytologi
6. Dansk litteraturs historie
7. Dansk Pattedyratlas
8. Danmarks Oldtid
9. Danmarkshistorien
10. Gyldendals Teaterleksikon
11. Naturen i Danmark
12. Trap Danmark, Danmarks nationaltopografi
13. Trap Grønland
14. Trap Færøerne
15. Dansk Kvindebiografisk Leksikon
16. Gastronomisk Leksikon
17. Store medisinske leksikon
18. Danmarkshistorien.dk

Lex bruges i høj grad til baggrund og perspektiv

Når danskerne besøger Lex, søger de viden om alt mellem himmel og jord, og i stigende grad søger de også baggrundsviden om de emner, der præger den aktuelle dagsorden. Det markerer et skift fra tidligere tiders forventninger til leksikonet, hvor fokus primært var på dannelse og kulturarv. I dag stoler danskerne på, at Lex også leverer troværdig, opdateret viden om aktuelle emner som klima, kunstig intelligens, geopolitik eller udviklingen i Grønland og Arktis.

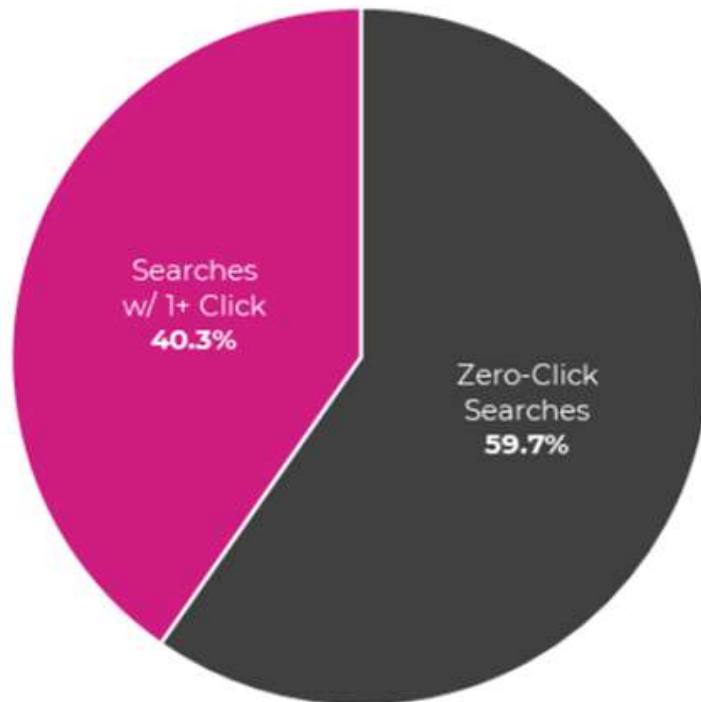
Vores adgang til viden er under forandring

Danskerne er bekymrede over den algoritmiske tilpasning af indhold



Google er gatekeeper på adgang til indhold på nettet

- Google har en markedsandel på **93 %** af søgemaskinemarkedet i Danmark (**98 %** på mobil)
- **60 %** af alle Google søgninger resulterer ikke i et klik videre til et website
- Af de resterende søgninger går **24 %** videre til Googles egne produkter, som fx YouTube



Google AI overblik



[Alle](#) [Bøger](#) [Nyheder](#) [Videoer](#) [Billeder](#) [Films](#) [Musik](#) [Helse](#) [Bøger](#) [Mere](#) [Værktøjer](#)

AI-overblik

"Queer" er et begreb, der dækker over en bred vifte af seksuelle orienteringer og kønsidentiteter, som ikke passer ind i de traditionelle kategorier af heteroseksualitet og det binære kønssystem (mand/kvinde). Oprindeligt brugt som et skældord, er "queer" blevet reapproprieret og brugt som en positiv identitetsmarkør og et paraplybegreb for personer, der udfordrer eller afviger fra samfundets normer for køn og seksualitet.

Her er en uddybning:

Betydning:
Ordet "queer" betyder oprindeligt "underlig", "sær", eller "mærkelig" på engelsk. Det blev tidligere brugt som en nedladende betegnelse for homoseksuelle, men er blevet reapproprieret af LGBTQ+-miljøet som en positiv identitet.

Paraplybegreb:
Queer bruges som et paraplybegreb, der omfatter alle, der ikke er heteroseksuelle og/eller cis-kønnede (dvs. hvis kønsidentitet stemmer overens med det køn, de er blevet tildelt ved fødslen).

Kønsidentitet og seksualitet:
Queer kan referere til både en persons kønsidentitet og seksualitet, og det kan bruges til at beskrive alle fra non-binære kønsidentiteter til forskellige seksuelle orienteringer, der ikke er heteroseksuelle.

Kritik af normer:
Queer-tankegangen indebærer ofte en kritisk tilgang til samfundets normer og forventninger til køn og seksualitet.

Persontlig identitet:
For nogle er queer en identitet, de identificerer sig med, mens det for andre er en tilgang eller et perspektiv på køn og seksualitet.

I praksis kan "queer" bruges af personer, der:

- Ikke identificerer sig som enten mand eller kvinde.
- Ikke er heteroseksuelle.
- Føler sig uden for eller udfordrer de traditionelle køns- og seksualitetsnormer.
- Ønsker at omfavne en bredere og mere fleksibel forståelse af køn og seksualitet.

queer - Begreb inden for køn og seksualitet - Lex

Ordet "queer" på engelsk betyder ordet "underlig", "ulig" og "kæd", men også...

Lex - Danmarks Nationalleksikon

Queer - Wikipedia, den frie encyklopædi

Queer er et paraplybegreb, som dækker over kønsorienteringer og seksuelle minoriteter, såsom...

W - de.wikipedia

Head er LGBT7 - Patienthåndbogen på sundhed.dk

16. aug. 2022 — Det står for queer, som dækker personer, der orienterer sig...

Sundhed.dk

[Vis alle](#)

Her er det indholdet fra [Lex](#) om [queer](#)

[Næste](#) [Tilbage](#)

 **Danmarks
Nationalleksikon**

Hvad betyder det for videnssøgning?

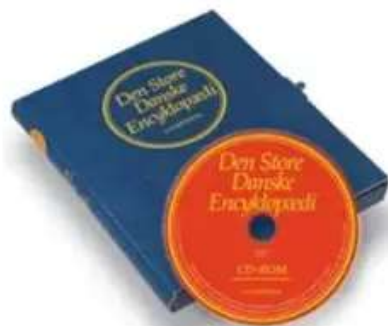
- **Flere "Zero-click" søgninger:** Brugere får svar direkte fra Google uden behov for at klikke på eksterne links
- **Manglende kilde eller garanti for kvalitet eller faktisk korrekthed,** selv om der findes autoritativt indhold, og selv om fx Lex med et link ligner afsendere på AI-genererede svar
- **Ingen kontrol over visning:** Det er uklart, hvornår og hvordan links til troværdigt indhold bliver bragt i overviews
- **Overrepræsentation af uautoritative kilder,** fx blogindlæg eller semi-kommercielle videnssites.

Hvad betyder det for samfundet?

- AI-overviews er en trussel mod lige adgang til viden og øger risikoen for misinformation
- **Google er ikke længere porten til viden**, men bliver til selve internettet – uden demokratisk forankring
- **Vi mister fælles referencepunkter**, det svækker samfundets evne til at etablere et robust udgangspunkt for den demokratiske samtale
- **Manglende gennemsigtighed** i udvælgelse og formulering gør det umuligt at efterprøve eller kritisere svar
- **Misinformation og halve sandheder** præsenteres med samme autoritet som korrekte svar – uden klar kildeangivelse
- **Den offentlige debat og dannelse forarmes**, når borgere får hurtige, ukontrollerede svar i stedet for mulighed for at læse sammenhængende forklaringer i redigerede tekster.

Fremtidens leksikon

Et leksikon til tiden



 Danmarks
Nationalleksikon Log ind



Velkommen til Lex – Danmarks Nationalleksikon
Flere end 245.000 artikler på dansk, skrevet af forskere og andre fagfolk

Det sker



Pave Frans er død

Pave Frans var pave fra 2013 til 2025. Han var den første sydamerikaner og den første jesuit, der blev valgt til embedet.



Prinsessen på atten

Prinsesse Isabella fylder 18 år den 22. april og fejres i tiden op til selve dagen. Læs om prinsessen, som af sin farfar blev kaldt Bella.



Den Amerikanske Revolution

Den Amerikanske Revolution var de amerikanske koloniers kamp mod den store britiske magt og oprettelsen af USA som en selvstændig republik. Revolutionen begyndte den 4. april for 250 år siden.

OUTLINE

1. Introduktion

Problemformulering

2. Lex.Ilm fase 1

System

Danish Foundation Models

3. Lex.Ilem Fase 2+

Udvikling

Forskning



OUTLINE

1. Introduktion

Problemformulering

2. Lex.Ilm fase 1

System

Danish Foundation Models

3. Lex.Ilm Fase 2+

Udvikling

Forskning



Kontekst

- 👉 Rangordnede links → syntetiske svar
- 👉 Transparens & kildeautoritet eroderer
- 👉 Dansk sprog og kultur skaber lokal forankring

R&D mandat til Lex.llm

Vidensforankret generering: hybrid RAG-løsning

Henvinsinger først: korrekte henvisning og undgå svar med lav støtte

Redaktør (& bruger)tilpasning: tilpasning skal afspejle encyclopædisk stilistik og epistemiske værdier

Risiko

- 🗨️ AI-snapshots skjuler gyldig danske kilder
- 🗨️ Falske og dekontekstualiserede henvisninger underminerer tillid & læring

Successkriterier

Forklarliged: faktuelle udsagn skal matche Lex.dk

Faktualitet: kildeuforenelige udsagn ↓; henvisnings-PR/R ↑.

Brugerforankring: stil og læsbarhed skal matche brugergrupper.

Sikkerhed: afvisning af spørgsmål udenfor et Lex-domæne.



En **LEX vidensagent** skal kunne

- 👉 Svare sandfærdigt på spørgsmål i et åbent 'Lex-domæne'
- 👉 Understøtte vidensintensive 'Lex-opgaver'
- 👉 Lave faktatjek
- 🗣️ Generere indhold
- 👉 Fungere som troværdig dynamisk vidensassistent

Og den skal gøre det på 🇩🇰 (og måske) 🇳🇴 , 🇸🇪 , 🇬🇧 ...



OUTLINE

1. Introduktion

Problemformulering

2. Lex.Ilm fase 1

System

Danish Foundation Models

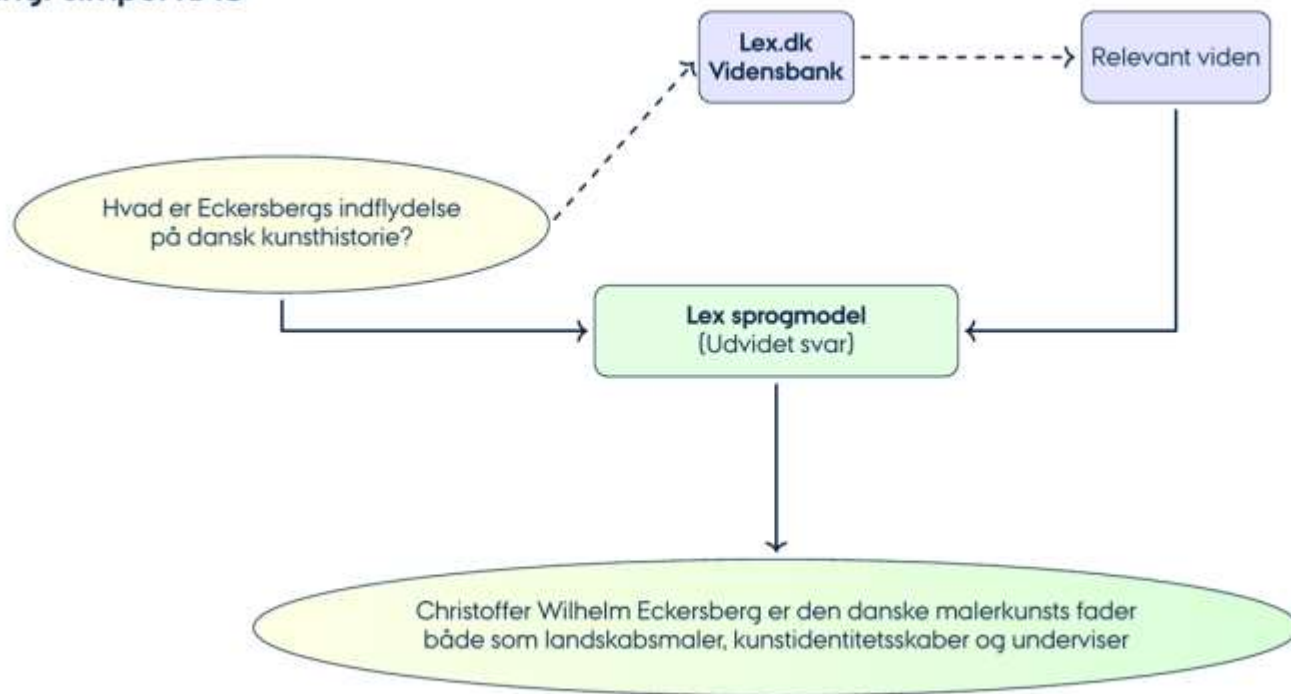
3. Lex.Ilem Fase 2+

Udvikling

Forskning



» Løsning: simpel RAG



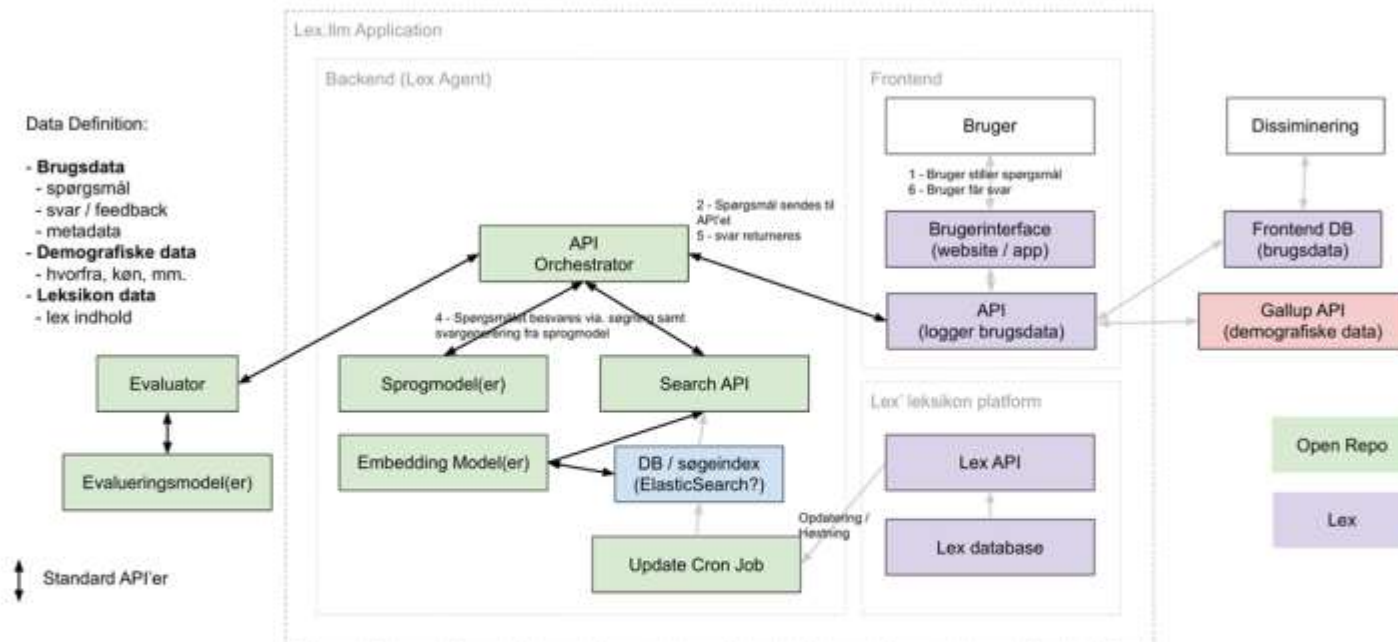
» Fase 1 af Lex.llm

- 👉 Etablering af **baseline** – udgangspunkt i en naiv RAG med kommerciel model
- 👉 simple og effektive **prompt-baserede guardrails** – sikkerhedsprompter til at baseline-modellens svar.
- 👉 etablering af **brugerpanel** – rekruttering af et repræsentativt panel af danskere til hele perioden.
- 👉 **Bruger-feedback & kortlægning** – indsamling og analyse af brugeres behov og rektioner på baseline.



Data Definition:

- **Brugsdata**
 - spørgsmål
 - svar / feedback
 - metadata
- **Demografiske data**
 - hvorfra, køn, mm.
- **Leksikon data**
 - lex indhold



Lex.Ilm, formål at skabe sikker adgang til kvalitetssikret viden, på dansk og altid med relevante kildehenvisning.

- AI Answer Quality and Citations
- Agricultural structure and employment
- Artificial Intelligence: Learning, Know
- Autism Spectrum Disorder: Emotions
- Autumn Weather and Daylight Savings
- Biographical Profiles of Notable Figures
- Chatbot Response Quality and Coverage
- Christianity, Church, and Gnosticism
- City population figures and data gaps
- Danish Digital Welfare and Public Services
- Danish Geography and Attractions
- Danish Lifespan and Economic History
- Danish chatbot Q&A quality assessment
- Danish history and culture Q&A
- Danish language and linguistics
- Danish monarchy
- Danish politics, elections, and defense
- Danish rural schooling and reform history
- Danish-language Q&A and translation
- Denmark's 1949 constitutional transition
- Explainable AI Dialogue
- Felix Baumgartner death hoax (2022)
- Fluctuating Trust in Chatbot Answers

Meaning of Life:
Religious and
Philosophical
Perspectives

Lex.dk content
limitations and
user frustration

Self-harm:
cutting safety
and resources

Danish monarchy

Danish-language
Q&A and
translations

Danish history
and culture Q&A

Danish politics,
elections, and
defense

Danish Geography
and Attractions

Musculoskeletal
pain and
treatment

Self-harm: cutting safety and resources

Keywords: cutting, harm, self, suicide, tools,
resources, cut, sharp, safety, psychological

0.47% of all documents

The topic centers on self-harm and suicide, explaining the psychological meaning of cutting, the emotional distress that can accompany it, and the associated risks. It emphasizes safety, prevention, and supportive resources, while consistently avoiding any how-to guidance and highlighting the need to seek professional help.

Error feedback clusters

Topic Tree

Expandable

- "Not what I asked"
- Answer too generic
- Couldn't answer
- Entirely wrong
- Google is better
- Issues with dates
- Local/historic info
- Metadata missing, RAG issues
- Misc error
- Missing sources
- Not satisfied
- Questions about Greenland
- Recipe problems
- Sound logic, but no answer
- Source used incorrectly
- Sources are too limiting
- Test
- Too long or too short
- World War inaccuracies
- -1.og.stk.eller.deres

rely wrong

Missing sources

Misc error

Couldn't answer

Not satisfied

Too long or too short

Metadata
missing, RAG
issues

Local/historic
info

Sound logic, but
no answer

Sources are too
limiting

Google is better

Answer too
generic

Test

World War
inaccuracies

Questions about
Greenland



AARHUS
UNIVERSITY

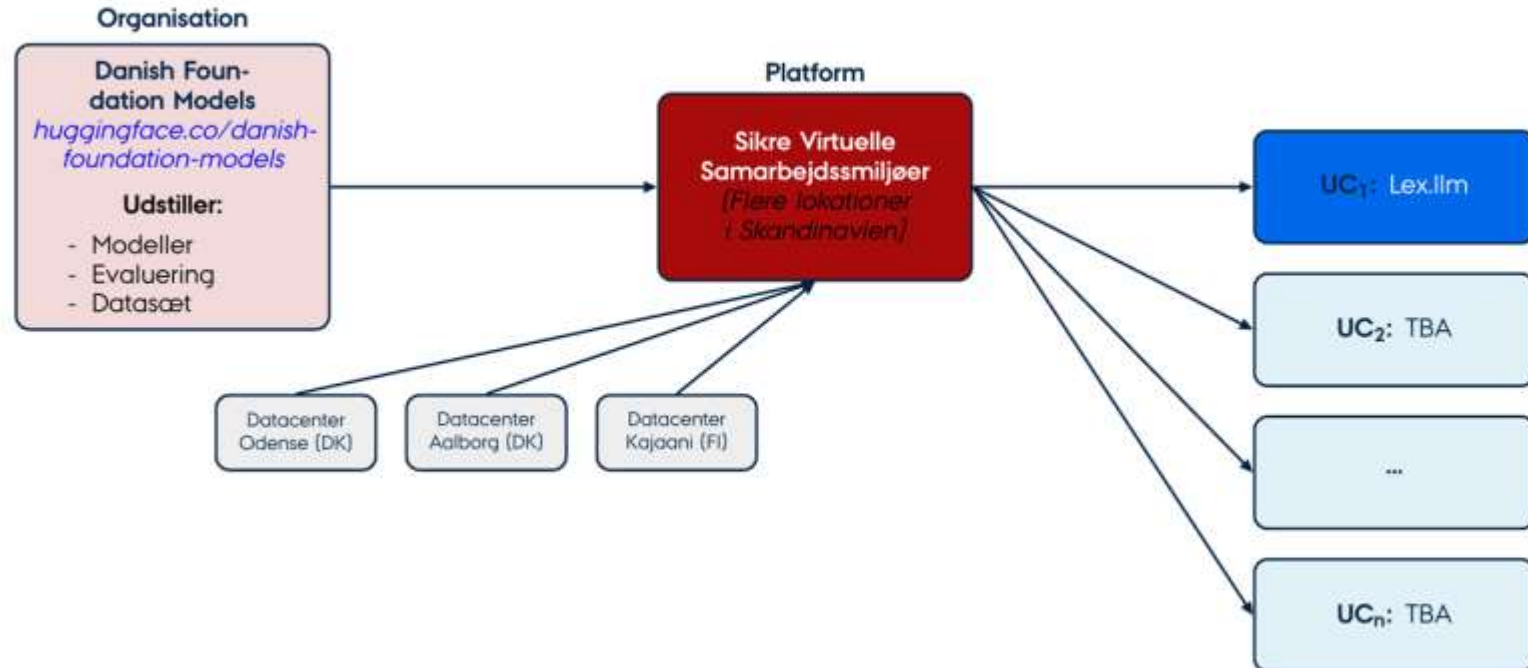
CENTER FOR HUMANITIES COMPUTING

SLIDE 10 IN 18



CENTER FOR
HUMANITIES
COMPUTING

» Danish Foundation Models – Sikker platform til udvikling af danske sprogmodeller



OUTLINE

1. Introduktion

Problemformulering

2. Lex.Ilm fase 1

System

Danish Foundation Models

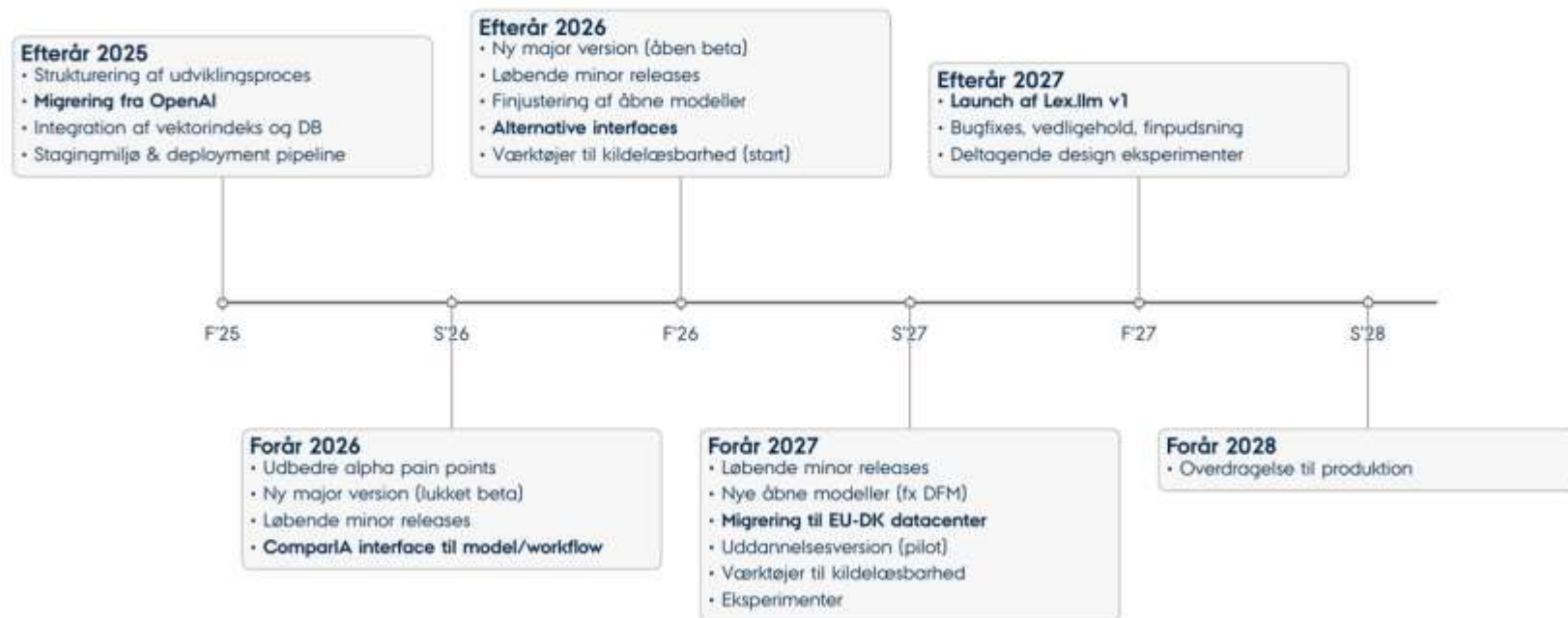
3. Lex.Ilem Fase 2+

Udvikling

Forskning



» Tidslinje – udvikling



» Fase 2 – Udvikling*: *Migrering fra OpenAI* → Så åbent som muligt, så lukket som nødvendigt

License	N	Mean	95% CI [low, high]	Median [Range]
Proprietary	97	1343.1	[1329.6, 1356.7]	1344.0 [1137, 1452]
MIT	26	1302.5	[1249.4, 1355.6]	1361.0 [978, 1428]
Apache 2.0	36	1267.6	[1226.3, 1309.0]	1310.5 [991, 1420]
CC-BY-NC-4.0	10	1248.6	[1206.5, 1290.7]	1257.0 [1154, 1351]

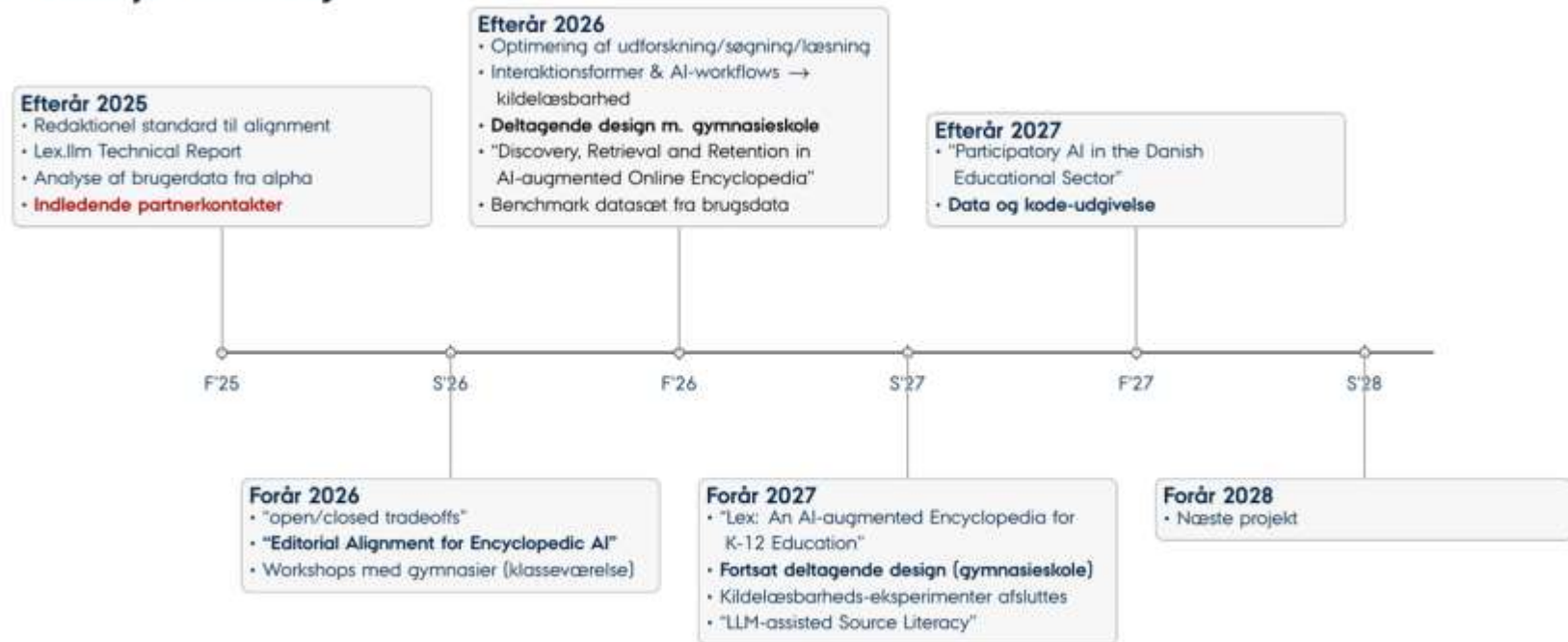
Sammenligning af performance for et udsnit af lukkede og åbne modeller fra `lmarena.ai`.

Dataset status	Model Effect	Tier 0	Tier 1	Tier 2	Tier 3 (fully open)
Legal status: Enable re-use, share and modify.	Data leakage is unproblematic Legality is clear				Openly Licensed
Data availability: Enables download of the dataset.	Potential for leak of copyrighted or private data. Legality is uncertain			Open Access	Open Access
Sufficient Documentation: Enables reproduction of the dataset	Potential for leak of copyrighted or private data Legality is uncertain		Replicable	Replicable	Replicable
Examples		GPT3, Llama, Gemini	GPT2, T5	BERT, Olmo	Pleias-RAG-1B

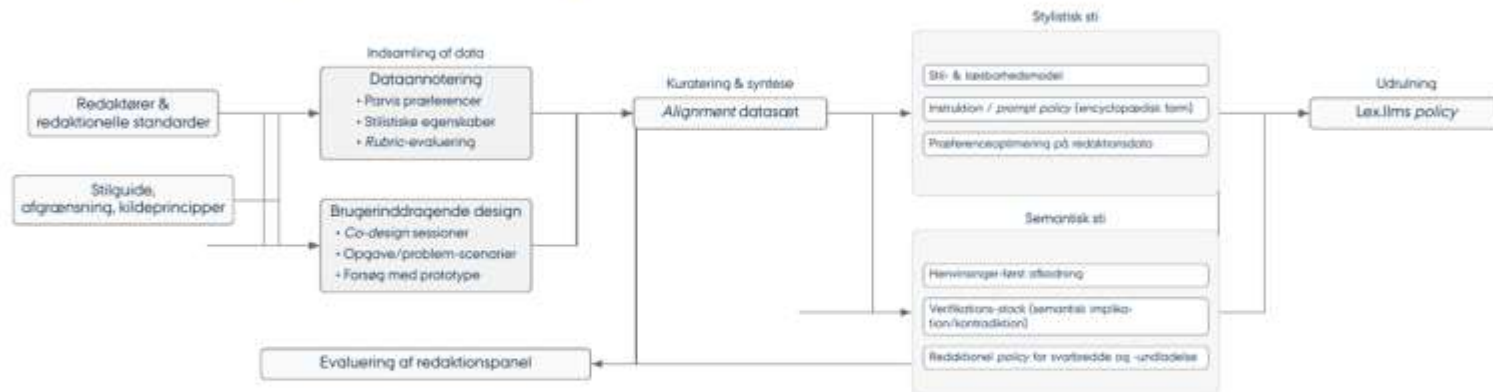
Datasæt-status vs. model-effekter i firetrinstrappe for åbenhed.



» Tidslinje – forskning



» Fase 2 – forskning: Redaktionel *alignment* af Lex.Ilm



Et anvendelsesorienteret forsknings- og udviklingsprojekt

Vi har med generøs støtte på 16 mio. kr. af Aage og Johanne Louis-Hansens Fond og Augustinus Fonden etableret et treårigt anvendt forskningsprojekt med Center for Humanities Computing ved Aarhus Universitet under ledelse af professor Kristoffer Nielbo.



Udvikles sammen med brugerne

Tilmeld dig her, og vær med til at forme Danmarks første pålidelige chatbot



Partnere i projektet



Kristoffer Laigaard Nielbo

Professor, Center Director
Center for Humanities Computing

✉ kln@cas.au.dk



Kenneth Enevoldsen

Postdoc
Center for Humanities Computing

✉ kenneth.enevoldsen@cas.au.dk



Simon Aargaard Enni

Postdoc
Center for Humanities Computing

✉ enni@cas.au.dk



Nicolas Marc Simon Legrand

Senior Researcher
Center for Humanities Computing

✉ nicolas.legrand@cas.au.dk



Yevhen Kostiuk

Postdoc
Center for Humanities Computing

✉ ykost@cas.au.dk



Zafar Hussain

Postdoc
Center for Humanities Computing

✉ zafar@cas.au.dk



Márton Kardos

IT Staff
Center for Humanities Computing

✉ martonkardos@cas.au.dk



Balázs Szabó

Student Data Scientist
Center for Humanities Computing



```
1 if questions:
2     try:
3         answer()
4     except RuntimeError:
5         pass
6 else:
7     print('THANKS')
```

TAK

kln@cas.au.dk
chc.au.dk

FINANCIERET AF
Aage og Johanne Louis-Hansens Fond & Augustinus Fonden

Nielbo and CHC's research are co-funded by Danish Foundation Models (4378-00001B), The Carlsberg Foundation (CF23-1583), European Union (101178170), and Danish Research Foundation (DNRF193).

