

A scenic landscape photograph showing a large body of water in the foreground, reflecting the sky. In the middle ground, there are several traditional red wooden houses with white window frames and green grass growing on their roofs. The background features rolling green hills and mountains with patches of snow under a cloudy sky.

Stora språkmodeller i Norden och Europa

Språkmodeller lär sig en sannolikhetsfördelning över språket

Förträning + finjustering (+ alignment)

Generell texthanteringsförmåga

Språkmodeller i Sverige

2017: Transformer introduceras av Google

2018: de första Transformer-baserade språkmodellerna (GPT, BERT)

2019: språkmodeller för svenska myndigheter (Vinnova)

2020: den första LLMen (GPT-3)

2021: NLU-gruppen på AI Sweden (Vinnova)

2022: den första skandinaviska LLMen: GPT-SW3 (Vinnova)

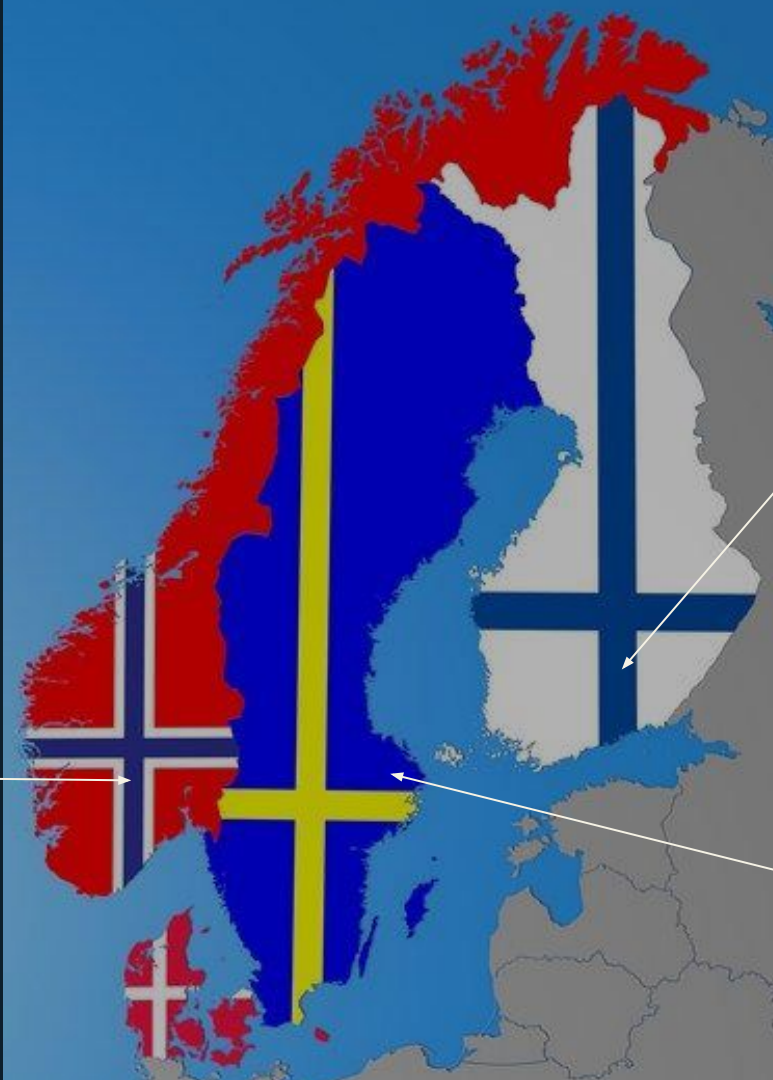
2023: GPT-SW3 som öppen modell

2024: AI Sweden + EU (EuroLingua-GPT, TrustLLM, DeployAI...)

FinGPT
Viking
Poro
Ahma

GPT-SW3

NorLLM
NORA.LLM



FinGPT
Viking
Poro
Ahma

GPT-SW3

NorLLM
NORA.LLM

Danish foundation models



NORA.LLM
NorLLM

Fin-GPT
Viking
Poro
Ahma

GPT-SW3

BLOOM

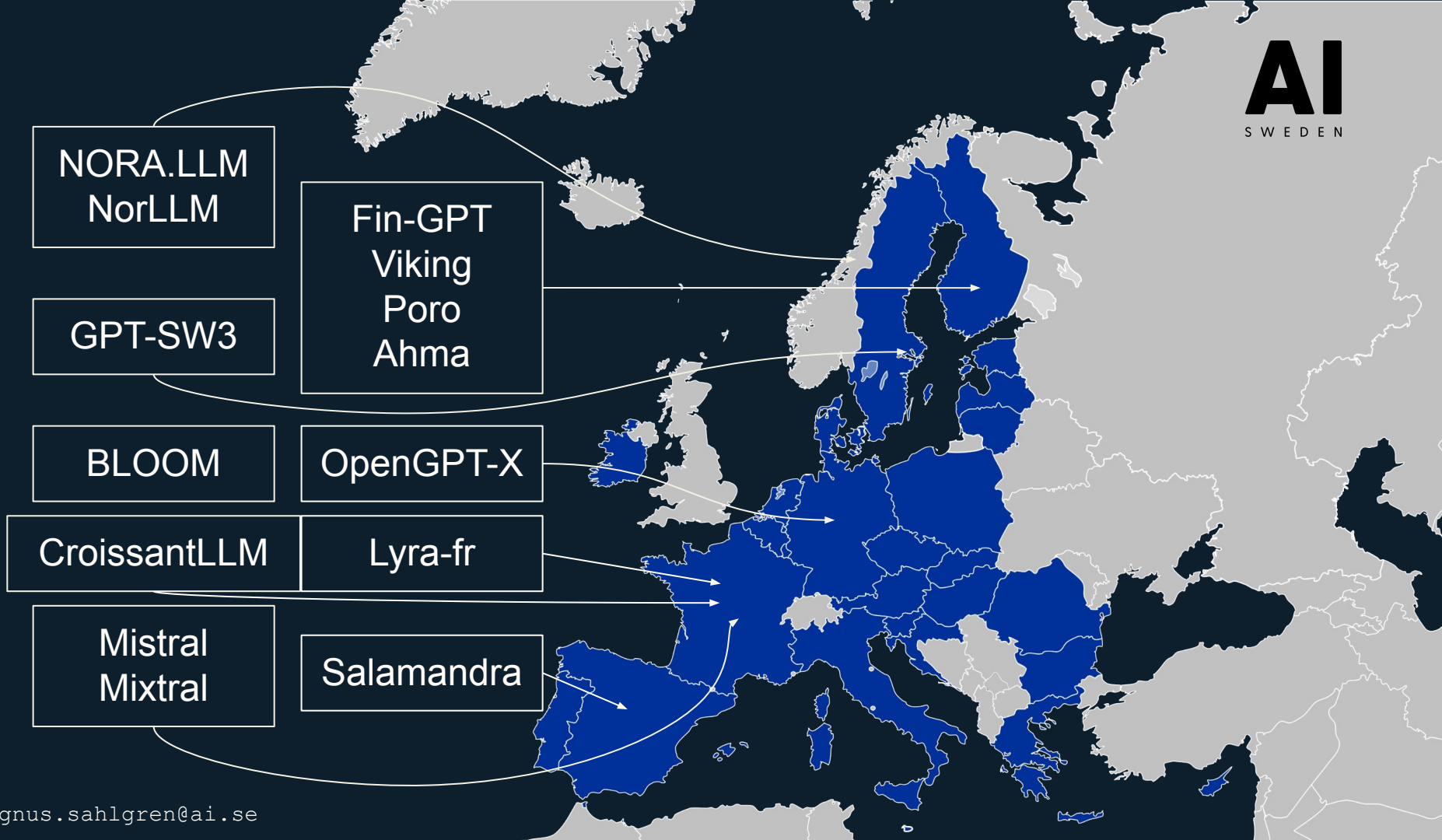
OpenGPT-X

CroissantLLM

Lyra-fr

Mistral
Mixtral

Salamandra



NORA.LLM
NorLLM

Fin-GPT
Viking
Poro
Ahma

GPT-SW3

BLOOM

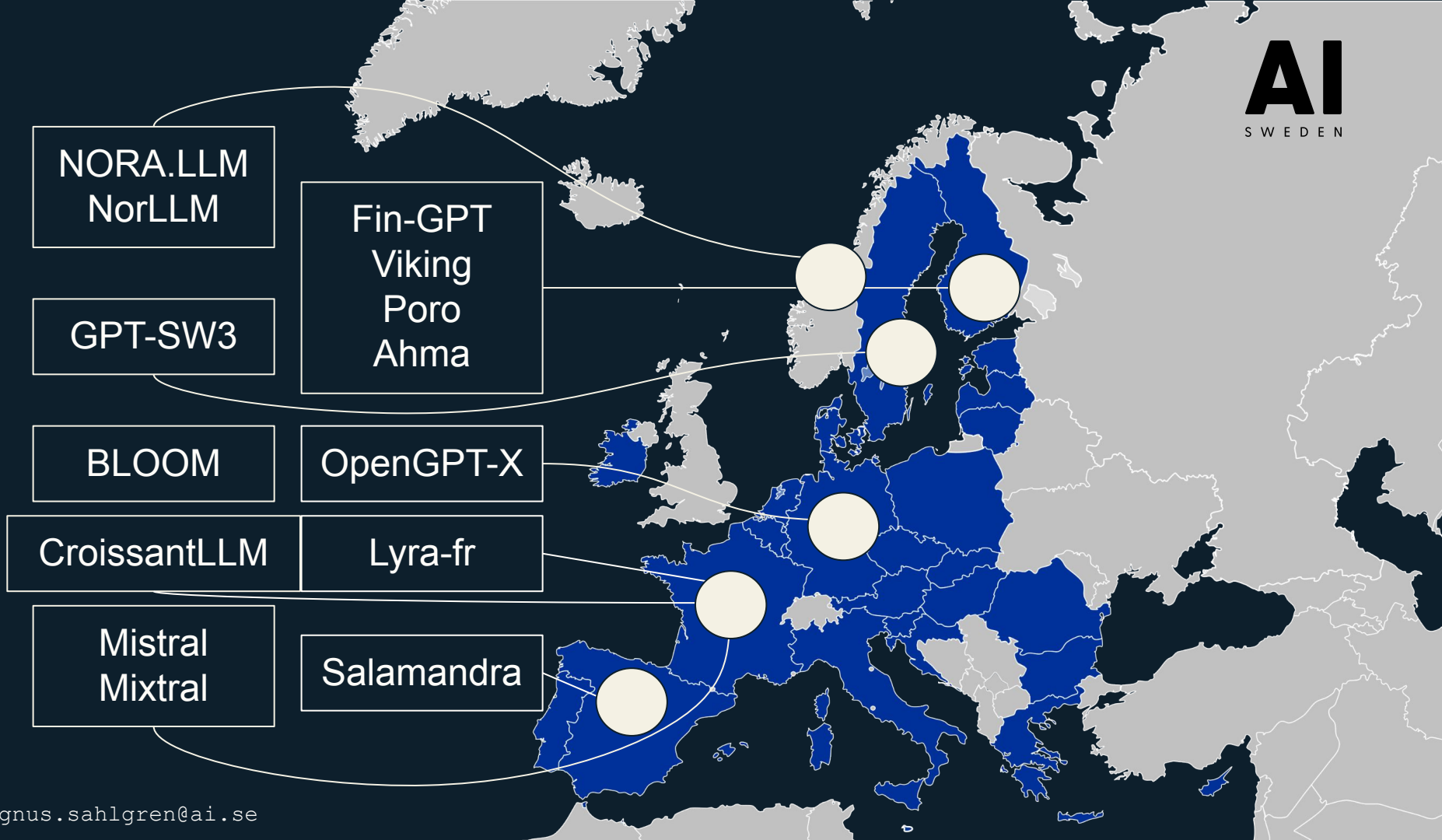
OpenGPT-X

CroissantLLM

Lyra-fr

Mistral
Mixtral

Salamandra



NORA.LLM
NorLLM

Fin-GPT
Viking
Poro
Ahma

GPT-SW3

BLOOM

OpenGPT-X

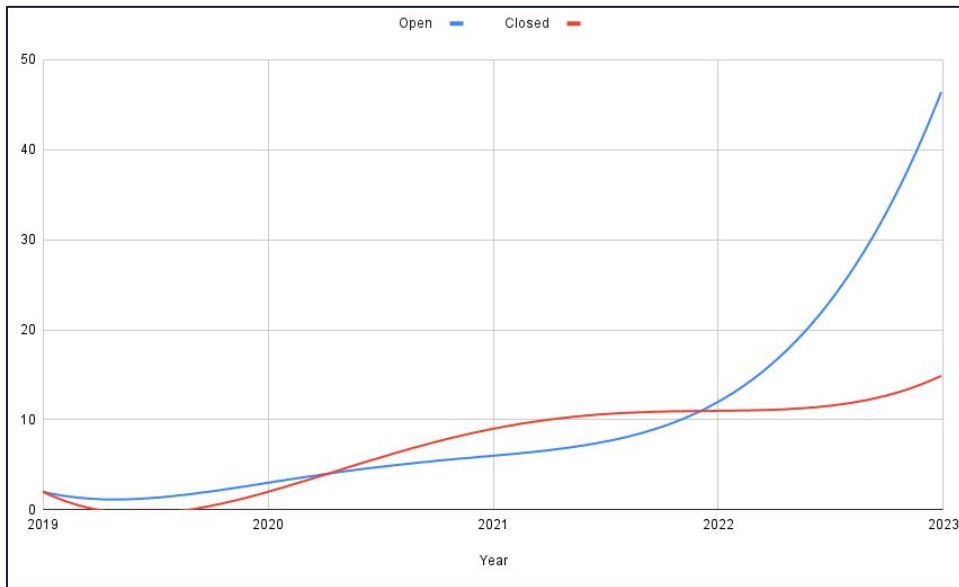
CroissantLLM

Lyra-fr

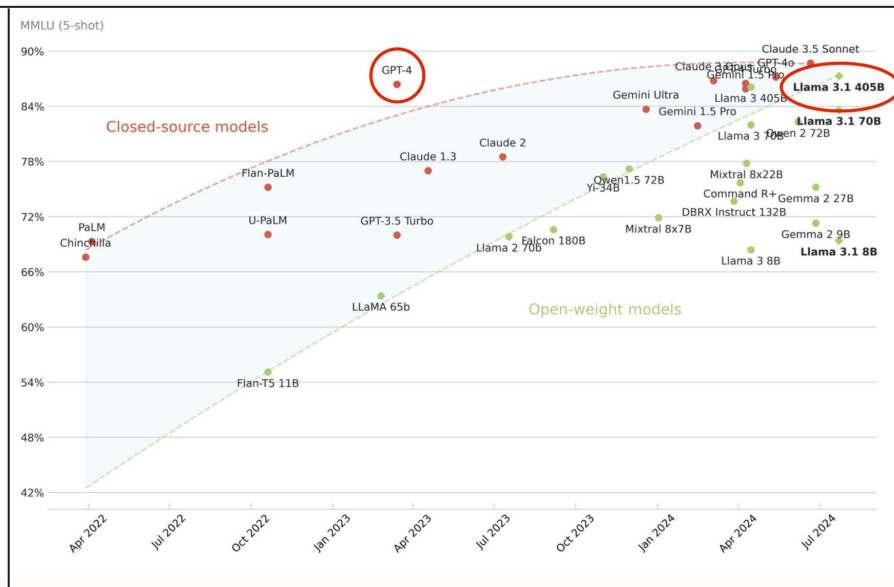
Mistral
Mixtral

Salamandra

Öppna vs stängda modeller



Data compiled by Stella Biderman



Maxime Labonne

huggingface.co/spaces/openGPT-X/european-llm-leaderboard

Type ▲	Model_Name ▲	Average ▲
🗨️	Meta-Llama-3.1-70B-Instruct	0.71
🗨️	Gemma-2-27b-Instruct	0.70
🗨️	Mistral-Nemo-Instruct-12.2B_2407	0.60
🗨️	Mixtral-8x7B-Instruct-v0.1	0.59
🗨️	Gemma-2-9b-Instruct	0.58
🟢	Mistral-Nemo-Base-12.2B_2407	0.56
🗨️	Meta-Llama-3.1-8B-Instruct	0.56
🗨️	c4ai-command-r-35B-v01	0.56
🟢	Mixtral-8x7B-v0.1	0.55
🗨️	Meta-Llama-3-8B-Instruct	0.55
🟢	Qwen2-7B	0.54
🟢	Gemma-7b	0.54

T ▲	Model ▲	Average ▲
🗨️	dfurman/CalmeRys-78B-Orpo-v0.1 🗨️	50.78
🗨️	MaziyarPanahi/calme-2.4-rys-78b 🗨️	50.26
🟡	rombodawg/Rombos-LLM-V2.5-Qwen-72b 🗨️	45.39
🟡	dnhkng/RYS-XLarge 🗨️	44.75
🗨️	MaziyarPanahi/calme-2.1-rys-78b 🗨️	44.14
🟡	rombodawg/Rombos-LLM-V2.5-Qwen-32b 🗨️	44.1
🗨️	MaziyarPanahi/calme-2.3-rys-78b 🗨️	44.02
🗨️	MaziyarPanahi/calme-2.2-rys-78b 🗨️	43.92
🗨️	MaziyarPanahi/calme-2.1-qwen2-72b 🗨️	43.61
🟡	dnhkng/RYS-XLarge-base 🗨️	43.56
🗨️	arcee-ai/Arcee-Nova 🗨️	43.5

huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard

huggingface.co/spaces/openGPT-X/european-llm-leaderboard

Type ▲	Model_Name ▲	Average ▲
🗨️	Meta-Llama-3.1-70B-Instruct	0.71
🗨️	Gemma-2-27b-Instruct	0.70
🗨️	Mistral-Nemo-Instruct-12.2B_2407	0.60
🗨️	Mixtral-8x7B-Instruct-v0.1	0.59
🗨️	Gemma-2-9b-Instruct	0.58
🟢	Mistral-Nemo-Base-12.2B_2407	0.56
🗨️	Meta-Llama-3.1-8B-Instruct	0.56
🗨️	c4ai-command-r-35B-v01	0.56
🟢	Mixtral-8x7B-v0.1	0.55
🗨️	Meta-Llama-3-8B-Instruct	0.55
🟢	Qwen2-7B	0.54
🟢	Gemma-7b	0.54

T ▲	Model ▲	Average ▲
🗨️	dfurman/CalmeRys-78B-Orpo-v0.1	50.78
🗨️	MaziyarPanahi/calme-2.4-rys-78b	50.26
🟡	rombodawg/Rombos-LLM-V2.5-Qwen-72b	45.39
🟡	dnhkng/RYS-XLarge	44.75
🗨️	MaziyarPanahi/calme-2.1-rys-78b	44.14
🟡	rombodawg/Rombos-LLM-V2.5-Qwen-32b	44.1
🗨️	MaziyarPanahi/calme-2.3-rys-78b	44.02
🗨️	MaziyarPanahi/calme-2.2-rys-78b	43.92
🗨️	MaziyarPanahi/calme-2.1-qwen2-72b	43.61
🟡	dnhkng/RYS-XLarge-base	43.56
🗨️	arcee-ai/Arcee-Nova	43.5

huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard

huggingface.co/spaces/openGPT-X/european-llm-leaderboard

Type	Model_Name	Average
🗨️	Meta-Llama-3.1-70B-Instruct	0.71
🗨️	Gemma-2-27b-Instruct	0.70
🗨️	Mistral-Nemo-Instruct-12.2B_2407	0.60
🗨️	Mixtral-8x7B-Instruct-v0.1	0.59
🗨️	Gemma-2-9b-Instruct	0.58
🟢	Mistral-Nemo-Base-12.2B_2407	0.56
🗨️	Meta-Llama-3.1-8B-Instruct	0.56
🗨️	c4ai-command-r-35B-v01	0.56
🟢	Mixtral-8x7B-v0.1	0.55
🗨️	Meta-Llama-3-8B-Instruct	0.55
🟢	Qwen2-7B	0.54
🟢	Gemma-7b	0.54

T	Model	Average
🗨️	dfurman/CalmeRys-78B-Orpo-v0.1	50.78
🗨️	MaziyarPanahi/calme-2.4-rys-78b	50.26
🟡	rombodawg/Rombos-LLM-V2.5-Qwen-72b	45.39
🟡	dnhkng/RYS-XLarge	44.75
🗨️	MaziyarPanahi/calme-2.1-rys-78b	44.14
🟡	rombodawg/Rombos-LLM-V2.5-Qwen-32b	44.1
🗨️	MaziyarPanahi/calme-2.3-rys-78b	44.02
🗨️	MaziyarPanahi/calme-2.2-rys-78b	43.92
🗨️	MaziyarPanahi/calme-2.1-qwen2-72b	43.61
🟡	dnhkng/RYS-XLarge-base	43.56
🗨️	arcee-ai/Arcee-Nova	43.5

1 miljard dollar i investeringar

/open-llm-leaderboard/open_llm_leaderboard

Varför träna nationella LLMer?

- Kompetens och innovation
- Transparens och representativitet
- Resiliens och suveränitet
- Demokratisering

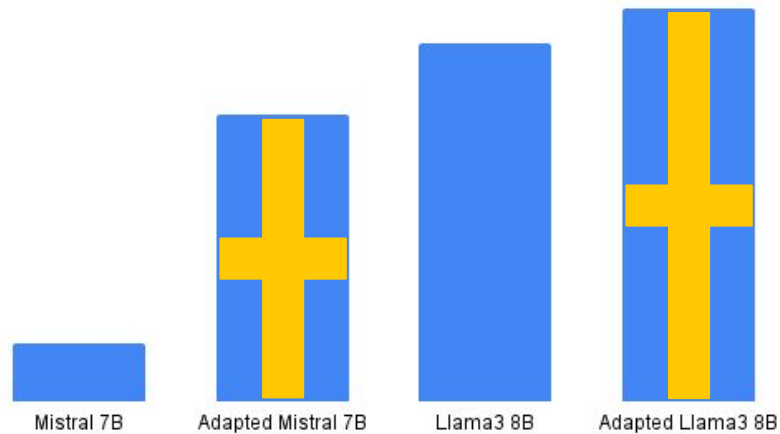
Varför träna nationella LLMer?

- Kompetens och innovation
- Transparens och representativitet
- Resiliens och suveränitet
- Demokratisering
- **Existerande modeller är inte perfekta på mindre språk!**

Anpassning av öppna modeller



Anpassning av öppna modeller



scandeval.com

Model ID	Parameters	Vocabulary Size	Context	Commercial	Speed	Rank ▼
gpt-4-0613 (few-shot, val)	unknown	100	8191	True	597 ± 197 / 93 ± 33	1.20
gpt-4o-2024-05-13 (few-shot, val)	unknown	200	128000	True	916 ± 329 / 114 ± 38	1.28
gpt-4-1106-preview (few-shot, val)	unknown	100	128000	True	576 ± 221 / 81 ± 28	1.29
AI-Sweden-Models/Llama-3-8B-instruct (few-shot)	8030	128	8192	False	4,314 ± 1,202 / 776 ± 245	1.42
meta-llama/Meta-Llama-3-70B (few-shot, val)	70554	128	8192	True	312 ± 55 / 177 ± 51	1.48
google/gemma-2-27b-it (few-shot)	27227	256	8193	True	1,516 ± 257 / 480 ± 148	1.66
google/gemma-2-9b-it (few-shot)	9242	256	8193	True	2,062 ± 397 / 589 ± 178	1.83

scandeval.com/mainland-scandinavian-nlg

Model ID	Parameters	Vocabulary Size	Context	Commercial	Speed	Rank ▼
gpt-4-0613 (few-shot, val)	unknown	100	8191	True	597 ± 197 / 93 ± 33	1.20
gpt-4o-2024-05-13 (few-shot, val)	unknown	200	128000	True	916 ± 329 / 114 ± 38	1.28
gpt-4-1106-preview (few-shot, val)	unknown	100	128000	True	576 ± 221 / 81 ± 28	1.29
AI-Sweden-Models/Llama-3-8B-instruct (few-shot)	8030	128	8192	False	4,314 ± 1,202 / 776 ± 245	1.42
meta-llama/Meta-Llama-3-70B (few-shot, val)	70554	128	8192	True	312 ± 55 / 177 ± 51	1.48
google/gemma-2-27b-it (few-shot)	27227	256	8193	True	1,516 ± 257 / 480 ± 148	1.66
google/gemma-2-9b-it (few-shot)	9242	256	8193	True	2,062 ± 397 / 589 ± 178	1.83

scandeval.com/mainland-scandinavian-nlg

Model ID	Parameters	Vocabulary Size	Context	Commercial	Speed	Rank ▼
gpt-4-0613 (few-shot, val)	unknown	100	8192	True	597 ± 197 / 93 ± 33	1.18
gpt-4-1106-preview (few-shot, val)	unknown	100	128000	True	576 ± 221 / 81 ± 28	1.28
gpt-4o-2024-05-13 (few-shot, val)	unknown	200	128000	True	916 ± 329 / 114 ± 38	1.28
meta-llama/Meta-Llama-3-70B (few-shot, val)	70554	128	8192	True	312 ± 55 / 177 ± 51	1.47
google/gemma-2-27b-it (few-shot)	27227	256	8193	True	1,516 ± 257 / 480 ± 148	1.59
AI-Sweden-Models/Llama-3-8B-instruct (few-shot)	8030	128	8192	False	4,314 ± 1,202 / 776 ± 245	1.78
google/gemma-2-9b-it (few-shot)	9242	256	8193	True	2,062 ± 397 / 589 ± 178	1.79

scandeval.com/mainland-scandinavian-nlg

Model ID	Parameters	Vocabulary Size	Context	Commercial	Speed	Rank ▼
gpt-4-0613 (few-shot, val)	unknown	100	8192	True	597 ± 197 / 93 ± 33	1.18
gpt-4-1106-preview (few-shot, val)	unknown	100	128000	True	576 ± 221 / 81 ± 28	1.28
gpt-4o-2024-05-13 (few-shot, val)	unknown	200	128000	True	916 ± 329 / 114 ± 38	1.28
meta-llama/Meta-Llama-3-70B (few-shot, val)	70554	128	8192	True	312 ± 55 / 177 ± 51	1.47
google/gemma-2-27b-it (few-shot)	27227	256	8193	True	1,516 ± 257 / 480 ± 148	1.59
AI-Sweden-Models/Llama-3-8B-instruct (few-shot)	8030	128	8192	False	4,314 ± 1,202 / 776 ± 245	1.78
google/gemma-2-9b-it (few-shot)	9242	256	8193	True	2,062 ± 397 / 589 ± 178	1.79

scandeval.com/mainland-scandinavian-nlg

Kompetens

Data



Finansering

Infrastruktur

Kompetens

Språkmodeller
Deep learning
HPC

Data



Finansering

Infrastruktur

Kompetens

Språkmodeller
Deep learning
HPC

Data



Finansiering

Storskaliga projekt
Inte forskning!
(Artefakterna är målet)

Infrastruktur

Kompetens

Språkmodeller
Deep learning
HPC

Data



Billioner av högkvalitativa tokens
(Vi måste förhålla oss till den
Europeiska dataskyddsförordningen)

Finansiering

Storskaliga projekt
Inte forskning!
(Artefakterna är målet)

Infrastruktur



The Scandinavian Web (**SWEb**) is a large scale Scandinavian web dataset (~1T tokens) built by AI Sweden for training high performance LLMs in the Swedish, Danish, Norwegian and Icelandic languages. For these languages, this is considerably larger than existing datasets (10B tokens for [OSCAR 23.01](#), 35B tokens for [HPLT Datasets 1.2](#) and 100B tokens for [mC4](#)). SWEb is provided in Markdown, to retain text formatting such as headings, text style, bullet point lists and tables.

For more details, please check out our [paper](#)

What is it

Details about SWEb

Modalities	Text
Number of tokens	1.01 trillion (GPT-SW3 tokens)
Languages	Swedish, Danish, Norwegian, Icelandic
Source	98 Common Crawl dumps
Cutoff	June 2024
Format	Markdown
Pipeline	See below

Kompetens

Språkmodeller
Deep learning
HPC

Data



Billioner av högkvalitativa tokens
(Vi måste förhålla oss till den
Europeiska dataskyddsförordningen)

Finansiering

Storskaliga projekt
Inte forskning!
(Artefakterna är målet)

Infrastruktur

Kompetens

Språkmodeller
Deep learning
HPC

Data



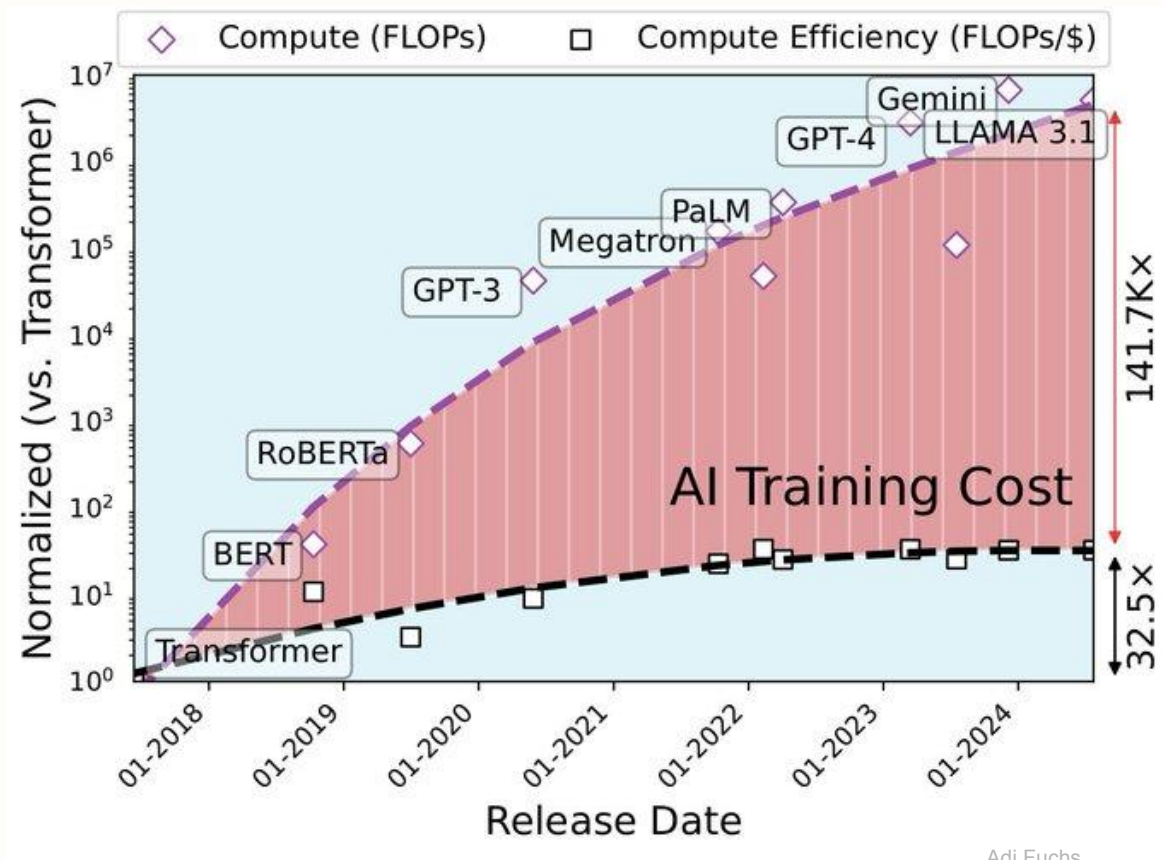
Billioner av högkvalitativa tokens
(Vi måste förhålla oss till den
Europeiska dataskyddsförordningen)

Finansiering

Storskaliga projekt
Inte forskning!
(Artefakterna är målet)

Infrastruktur

1000-tals high-end GPUs
(med bra interconnect)
Storskaliga allokeringar



Adi Fuchs

By Invitation | Artificial intelligence

Mark Zuckerberg and Daniel Ek on why Europe should embrace open-source AI

It risks falling behind because of incoherent and complex regulation, say the two tech CEOs



ILLUSTRATION: SAM KERR

Aug 21st 2024

Share

www.economist.com/by-invitation/2024/08/21/mark-zuckerberg-and-daniel-ek-on-why-europe-should-embrace-open-source-ai

Europeiska initiativ

- Language Data Space
- Alliance for Language Technology (ALT-EDIC)
- AI Factories
- EuroHPC (exascale: Jupiter ~300M EUR)
- High-Performance Language Technologies, TrustLLM, DeployAI, EuroLingua-GPT...

TrustLLM

6.9M EUR 2023-2026

Utveckla en öppen, trovärdig och pålitlig LLM för de Germanska språken

LiU, FHG, FZJ, AXI, UCPH, NTNU, UOI, MID, TNO, AKI

Eurolingua-GPT

A ny och **öppen** basmodell för de Europeiska språken

Samarbete mellan AI Sweden och Fraunhofer IAIS

MareNostrum5: 4480 H100 GPUs (9M GPU-timmar)

En digital assistent för offentlig sektor

Nationellt initiativ för **datagenerering**, finjustering och driftsättning av LLMer för svensk offentlig sektor

Målet är att bygga en **nationell datafabrik** där organisationer i offentlig sektor utvecklar en gemensam digital assistent för deras textbaserade uppgifter

Kontakt: jonatan.permert@ai.se