

A scenic landscape photograph showing a large body of water in the foreground, a row of traditional red wooden houses with white window frames and green roofs, and snow-capped mountains in the background under a cloudy sky.

# Stora generativa språkmodeller för svenska

2000-2010: proto-språkmodeller på SICS

2017: Transformer-arkitekturen introduceras

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

**2019: språkmodeller för svenska myndigheter (Vinnova)**

2020: den första LLMen publiceras (GPT-3)

**2021: NLU-gruppen på AI Sweden**

**2022: den första LLMen för nordgermanska språk: GPT-SW3**

**2023: GPT-SW3 släpps som 'open weight'**

## Varför bygga svenska LLMer?

---

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

## Varför bygga svenska LLMer?

---

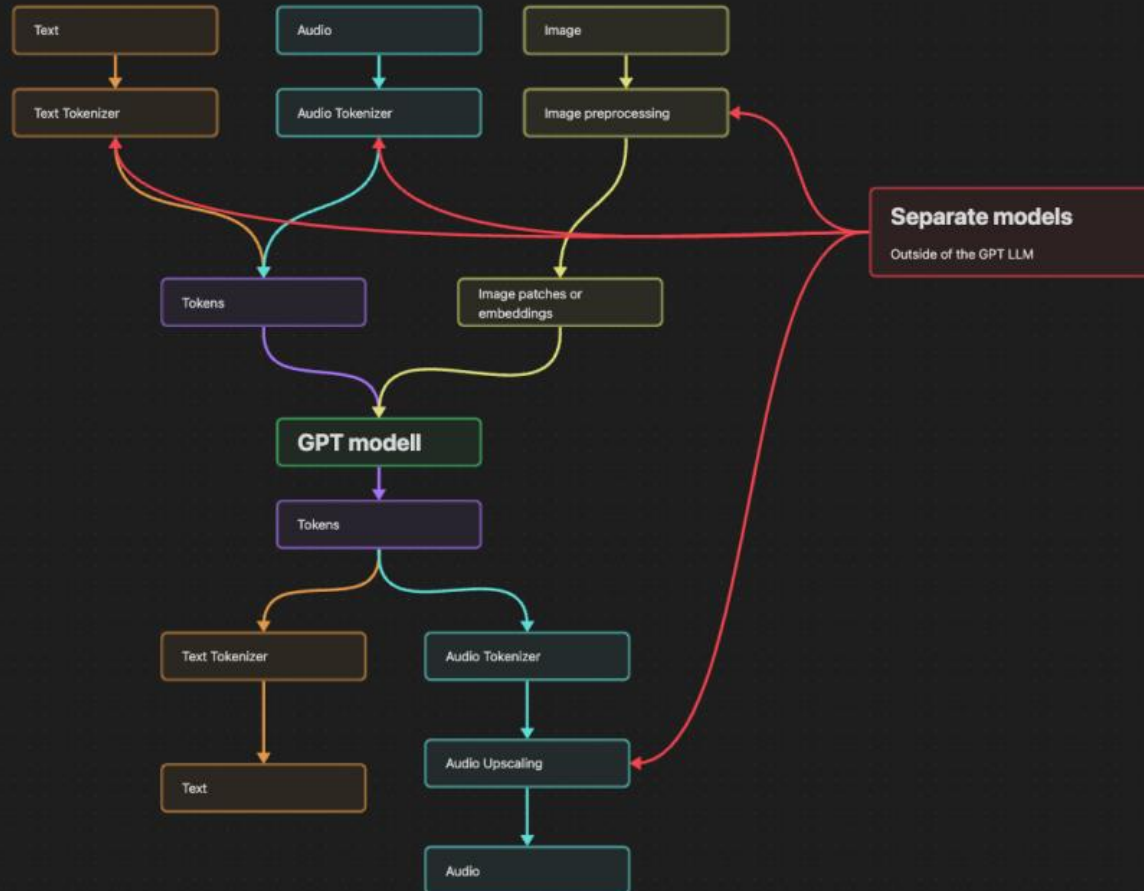
- Kompetens
- Transparens och representativitet
- Resiliens och suveränitet
- Demokratisering
- **Existerande modeller är inte jättebra på svenska!**

## Stegen efter GPT-SW3

---

- Anpassa öppna modeller till svenska (ex Mistral/Llama3)
- **Multimodalitet och generell kapacitet**
- Internationella samarbeten (EU-projekt)

## Multimodal model



AllFetchHTMLDocuments\_\_99914b932b

Show task details

Show Upstream Dependencies

Hide Done

Visualisation Type

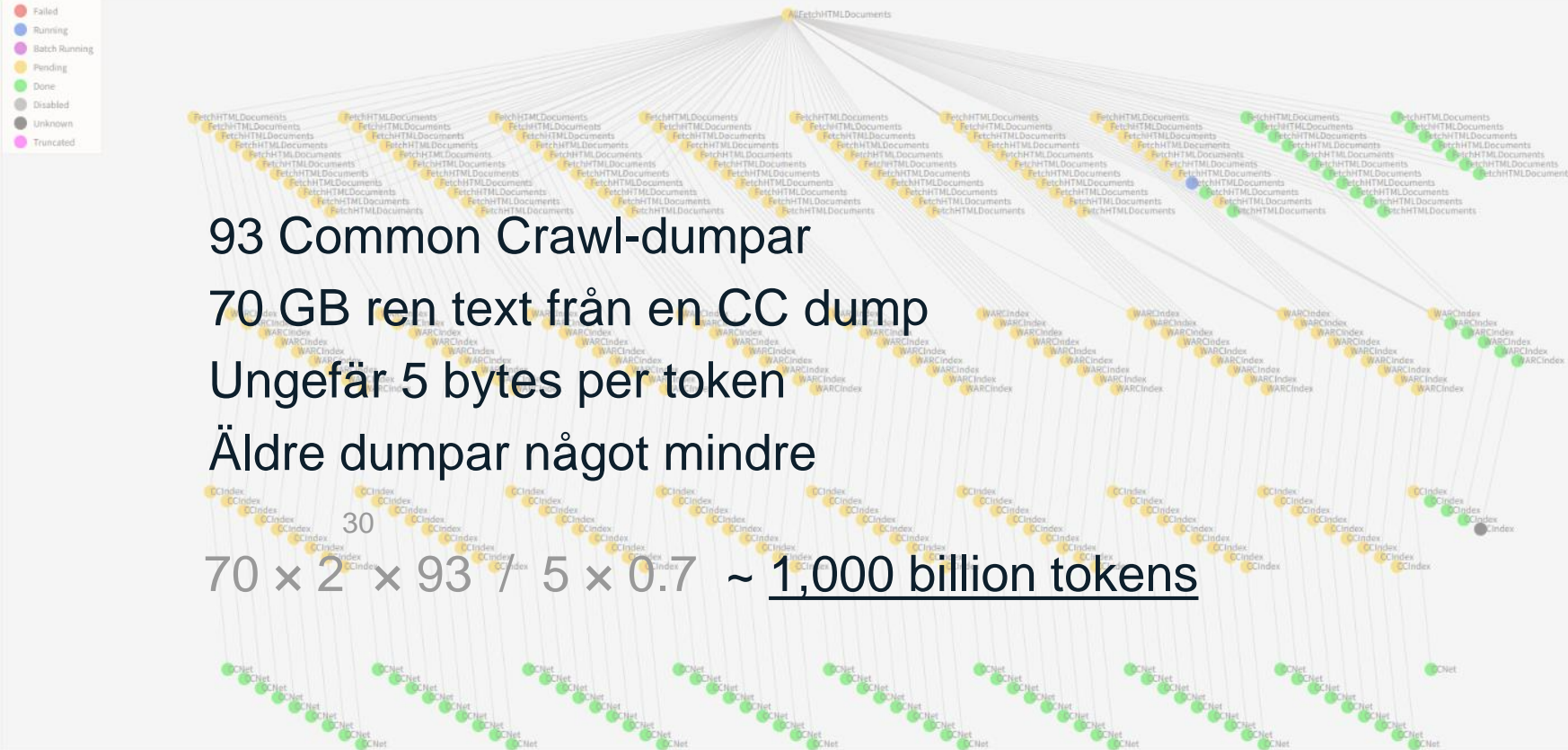
D3

SVG

AllFetchHTMLDocuments()

Dependency Graph

- Failed
- Running
- Batch Running
- Pending
- Done
- Disabled
- Unknown
- Truncated





# TrustLLM

6.9M EUR 2023-2026

Utveckla en öppen, trovärdig, och hållbar LLM för de germanska språken

Data, modellträning, trovärdighet, flerspråkighet, utvärdering





# TrustLLM

**AI**  
S W E D E N

**li.u** LINKÖPING  
UNIVERSITY

 **Fraunhofer**  
IAIS

 **JÜLICH**  
Forschungszentrum

  
UNIVERSITY OF  
COPENHAGEN

**TNO** innovation  
for life

 **MIDEIND**

 **ALEXANDRA**  
INSTITUTET

 **KI BUNDESVERBAND**

 **HÍ**

 **NTNU**

**AI**  
S W E D E N



# DeployAI

30M EUR 2023-2027

Bygga, distribuera och lansera en fullt fungerande AI-On-Demand-plattform som främjar trovärdiga, etiska och transparenta europeiska AI-lösningar

AI Sweden: tillhandahålla och finjustera modeller för svenska (et al)



# DeployAI

**Coordinator:** Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Germany

**Partners:**

- AALTO University, Finland
- Aspect Partners, Germany
- European Centre for medium-range Weather Forecast, United Kingdom
- Convotis Services, Austria
- Hellenic Telecommunications Organization S.A (OTE), Greece
- Croatian Artificial Intelligence Association, Croatia
- National Centre for Scientific Research "Demokritos", Greece
- EIT Digital, Belgium
- DIMECC, Finland
- F6S Network, Ireland
- Bruno Kessler Institute, Italy
- FIWARE Foundation, Germany

- German AI Association, Germany
- Hub France IA, France
- IONOS SE, Germany
- LIGHT ON, France
- LuxProvide SA, Luxembourg
- Science For You (SciFY), Greece
- Silo AI, Finland
- AI Sweden, Sweden
- VDE – Association for Electrical, Electronic & Information Technologies, Germany
- Institute of Communication and Computer Systems (ICCS), Greece
- CSC- IT Center for Science, Finland
- Arcada University of Applied Sciences, Finland
- Örebro University, Sweden
- Technical University of Denmark - DTU, Denmark

**Associated Partner:** ADRA - The AI, Data and Robotics Association

# **Eurolingua GPT**

**En ny basmodell för de Europeiska språken**

**Samarbete AI Sweden och Fraunhofer IAIS**

**Extreme scale access: 8.8 miljoner GPU-timmar**

**MareNostrum5: 4480 H100 GPUer**



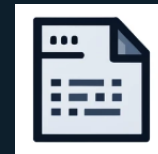
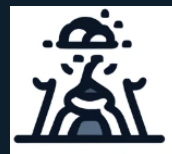
# Data Management for TrustLLM

Danila Petrelli - Senior Data Manager NLU @AI Sweden



# Harmonizing Strategies

- Data standards and formats
- Ethical and legal compliance
- Data sharing and access policies
- Security measures
- Long-term preservation and sustainability





# Data Management Plan(s)

- Cohesive approach to managing the vast amounts of data
- Ensuring that all data is handled in a way that promotes transparency, open science and builds a solid foundation for trustworthy AI models
- FAIR principles: making sure data is Findable, Accessible, Interoperable, and Reusable.





# Documentation

Why?

- Transparency and accountability
- It provides a clear record of the data used, the design and operation of AI systems and the measures taken to mitigate risks.
- Essential for compliance with regulations







# Our Plan



- Have recurring workshops
- Reach internal understanding
- Identify challenges
- Engage with leading people in the field
- Establish best practices for trustworthy data management
- Document!



# Challenges

- Legal compliance
- Transparency



# Key Learnings in Data Management

Collaboration is key

Consistency matters

Adaptability is crucial

Documentation cannot be underestimated



Thank you!

Contacts:

`magnus.sahlgren@ai.se / danila.petrelli@ai.se`