Creating a Controlled Dataset of Environmental Assessment Texts for Generative AI Models and Multi-agent Systems: The Danish EA Hub

Theme forum: The Emerging Role of Artificial Intelligence in the Future of Impact Assessment

Lone Kørnøv and Ivar Lyhne

Professor and Associate Professor

The Danish Centre for Environmental Assessment, Aalborg University, Denmark

https://dk.linkedin.com/company/the-danish-centre-for-environmental-assessment

https://www.en.plan.aau.dk/research/the-danish-centre-for-environmental-assessment

Brief context:
AI in Denmark

A highly digitalized country - among leads in Europe

28% of all Danish companies use AI (2024) - highest in Europe, double of 2023, multiple purposes

Initiatives on digitalizing environmental assessment practices (e.g., DREAMS project – 2020-2023)

Presentation

1

Why a data-centric approach to AI in impact assessment?

2

How the Danish EA Hub is and will be used for AI?

3

Why and how to create dataset through a socio-technical lens?

1. Better EA data, better AI

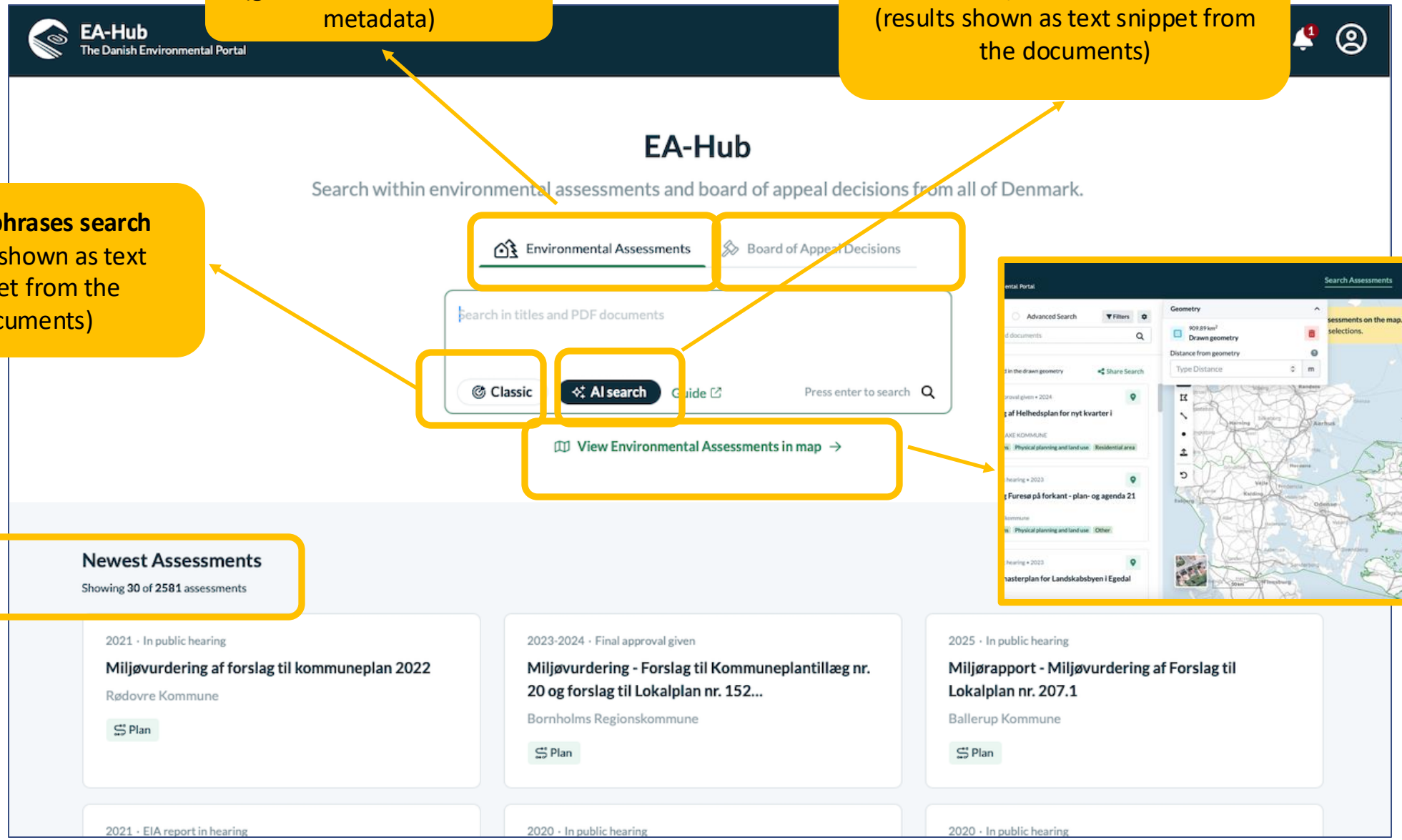
Main lever	Model architecture	Scale of data	Quality and relevance of data
Goal	Smarter models	More data	Better data
Typical changes	Make models bigger, design new	Bigger datasets	Curating, refining and maintaining domain data
Famous quote	"Better model wins"	"Data is the new oil"	"Garbage in, garbage out"

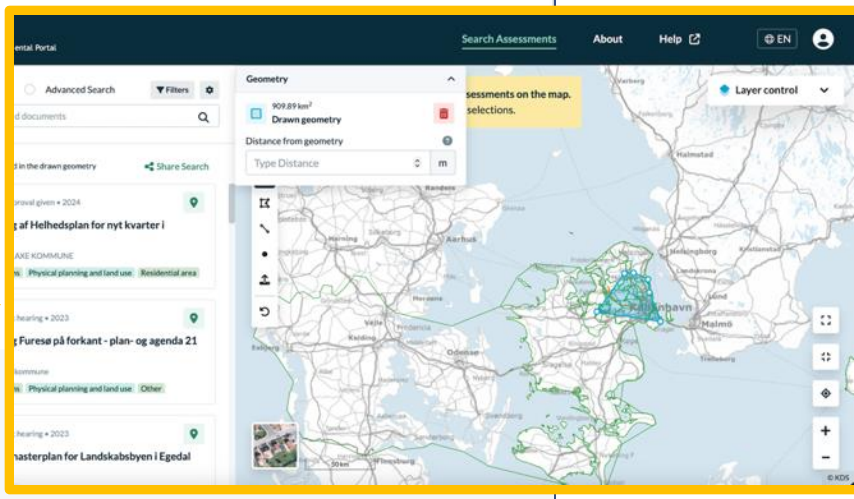
2. The Danish EA Hub

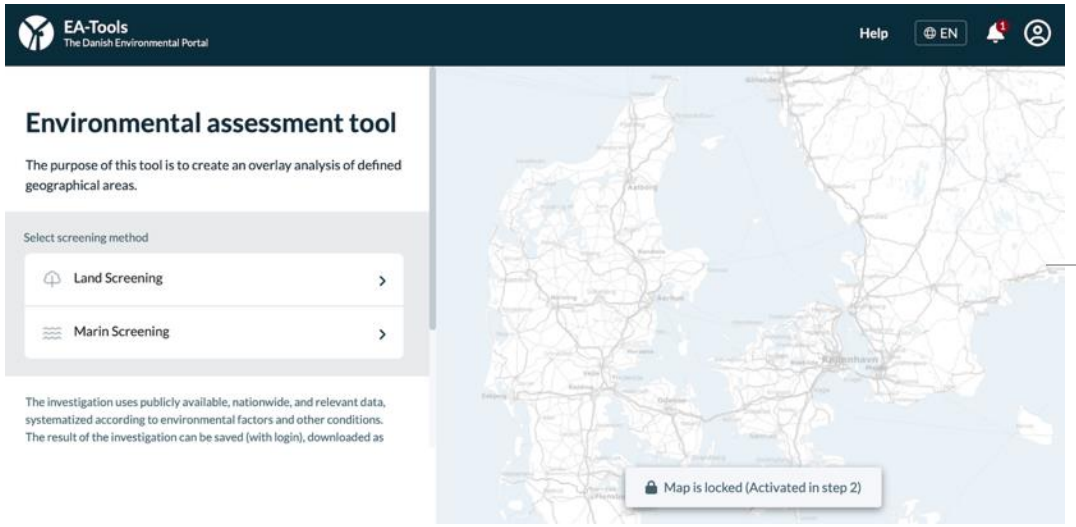
•2.500+ EIA and SEA reports (georeferenced and with metadata)

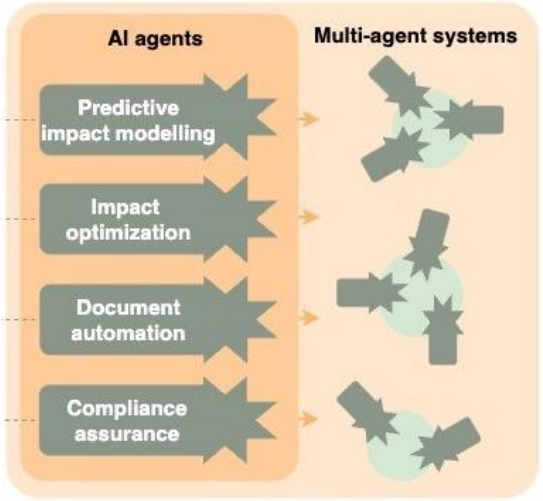
AI search built on language model – exclusively on Hub documents (results shown as text snippet from the documents)

Words/phrases search (results shown as text snippet from the documents)









From repository to intelligent infrastructure

- EA Hub is evolving – draft legislation now mandates its use in EA
- New documents like screenings and scoping statements will be included
  - Enriching the dataset – laying the foundation for robust, trustworthy AI
- Supported by over 700 curated GIS datasets- covering both onshore an offshore, purpose-built for environmental assessment

- AI developments enabled by the datasets and general AI progression, e.g.:
  1. Linking EAs with legal decisions using domain-trained large language models
  2. Individual AI agents for modelling, optimization, automatization, and compliance
  3. Integrated multi-agent systems supporting dynamic, real-time EA workflows and decision-support

3. A socio-technical foundation

AI is not just technical—it's socio-technical

- Datasets must reflect legal, institutional, and professional realities
- Success depends on both accuracy *and* democratic legitimacy
- User needs, governance, and IP/data right must be integrated from the outset
- AI must be built *wit*h people, institutions, and frameworks it serves – not just with code

EA Hub as a socio-technical system

- Co-produced by EA stakeholders - including consultants, legal experts, authorities, and researchers
- Continually improved through user engagement via the Danish Environmental Portal
- Supports trust, interoperability, and better decision-making
- Already showing impact: faster processes and cost savings in planning and environmental assessment

Key takeaways

1. **Data-Centric AI needs high-quality, context-aware datasets**
   - *AI in Environmental Assessment is only as good as the data it learns from*

1. **EA Hub is becoming a strategic infrastructure**
   - *With legal backing and enriched content, it supports robust, trustworthy AI development.*

2. **Socio-technical design is essential**
   - *Success depends not only on technical accuracy but also legitimacy, governance, and collaboration.*

Let’s continue the conversation!

Message me your questions or comments in the IAIA25 app.

Explore:
EA-Hub
EA-Tool
DREAMS project

And look out for Karl Sveding & Dorthe Holme

#iaia25

Lone Kørnøv and Ivar Lyhne

Professor and Associate Professor

The Danish Centre for Environmental Assessment, Aalborg University, Denmark

https://dk.linkedin.com/company/the-danish-centre-for-environmental-assessment

https://www.en.plan.aau.dk/research/the-danish-centre-for-environmental-assessment