



Pistoia Alliance

Workshop Report

Democratising Data in the Life Sciences:

The Role of LLMs

23 April 2024 | 13:00 – 17:00 GMT



Democratising Data in the Life Sciences: The Role of LLMs

- Approx. 40 attendees from across large pharma, IT integrators, non-profit
- Located at Elsevier offices London Wall



aim of the day

“ Discuss, as a community, with the aid of some technical examples, the role that LLMs may play, alongside technologies such as ontologies, in supporting data democratisation within the life sciences “

Pistoia Alliance

Today's Team



Oli Giles



Zahra Hosseini



Jane Lomax



Claudia Millan



Joe Mullen



Maaly Nassar



Thom Pijnenburg

Today's Journey



You are here!



intro

intro to LLMs
*limitations,
grounding and RAG*



semantic rep
vectors



semantic rep
ontologies



discussion
*Vision and discussion
topics*



close

Workshop

Democratising Data in the Life Sciences: The Role of LLMs

Confidential | Copyright © 2024 Elsevier Ltd. All Rights Reserved





NOTES. Change Management.

Breakout.
KG + LLM + Ont
→ Routing

context issue → now do we route when we don't know the context of the data that is implicit.

Benchmarking.
- binning (BAG)
- use-case specific q's.

LLM ↔ Ontology (synergy)

Innovation
LLM →
→ hallucination
- peptide gen
- met gen

DIGITAL ↔ PHYSICAL.
in silico
help to improve models.

FUTURE.

Knowledge (data) ↑
↓ GenAI.
↓ new use-cases.

creative space.
already good at hallucinating
→ can use that in resource w/ feedback

examples
ALPHA-FOLD.

⇒ Gen Chem. hallucinations are highly overrepresented.
↓ important for human as long as properties are measurable
(using hallucinations)

Usage. → better stay accurate!
right now →
Efficiency over accuracy.
↓ value.
NL → syntax X.future
molecule design.

AKA HITTING MORE
model understanding

NL → Syntax-done
Maths
domain specific? nps

AGI

limitations.
reasoning
numerical math

LLM in deterministic workflows
needs ~~not~~ reasoning!
VS non-deterministic

Recy are gen
→ efficiency
as long as some
regions.



- Thanks to everyone involved
- Listen to Joe Mullen's talk at 12PM and see SciBite Chat in action