Challenges and Open Problems

Explainable Information Retrieval

Evaluation

Current evaluation schemes for post-hoc interpretability limited

- inject bugs and treat them as ground truth
- Use a simple model as a BBOX

Little to no **benchmarks** for development and evaluation of interpretability

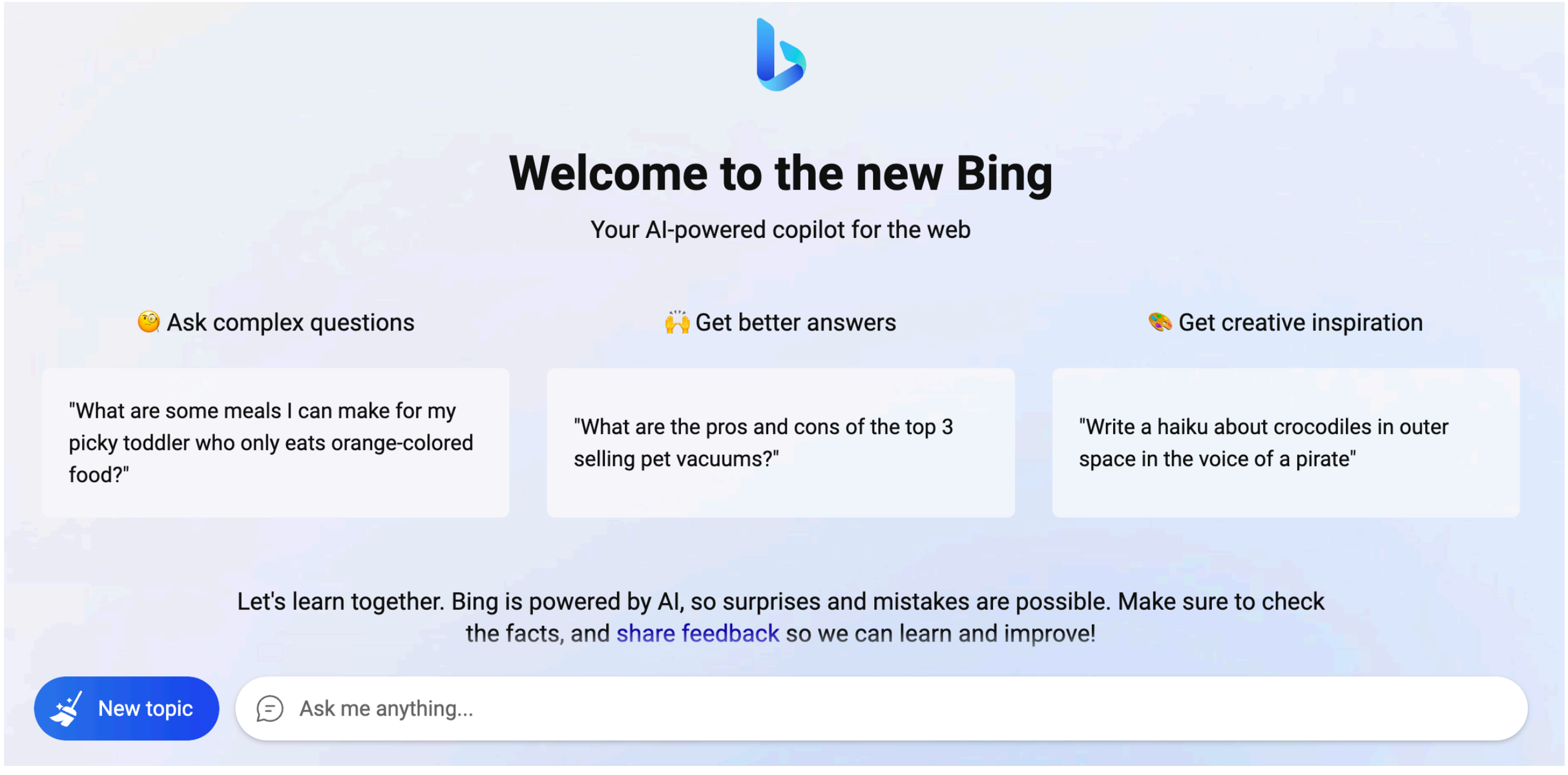
- none for IR with the exception of ERASER
- Need for a holistic benchmark

Human studies are limited and hard to make progress

- Current studies use weak baselines or controls

But please dont stop accepting papers due to any of the reasons

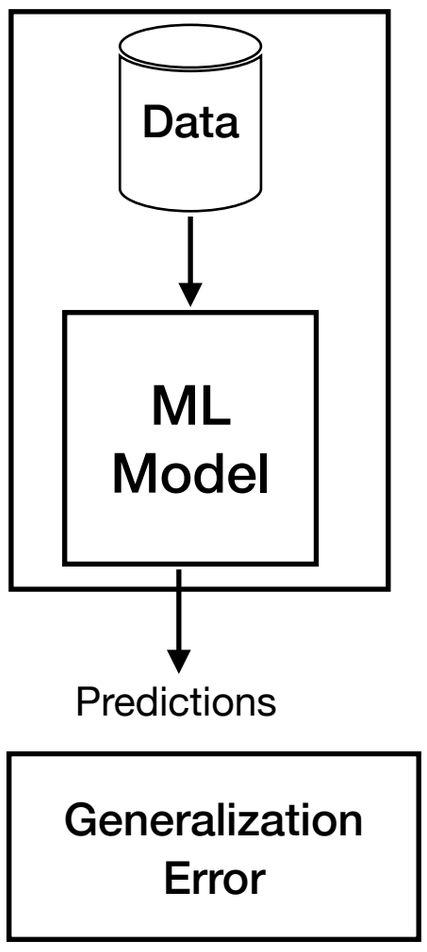
A glimpse of the future..

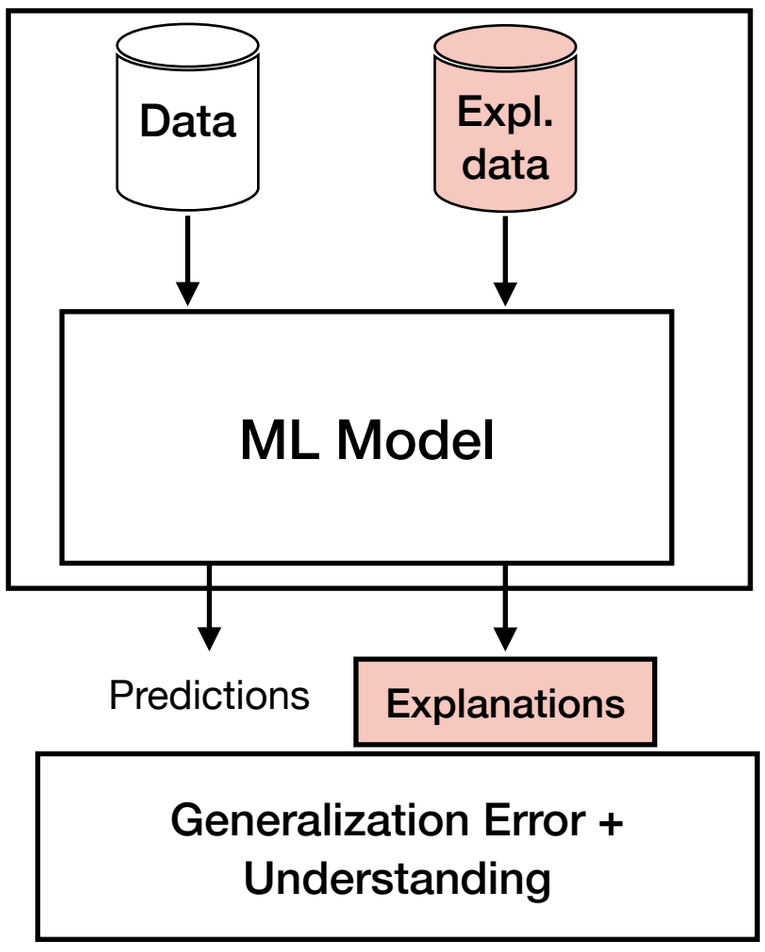


Vision

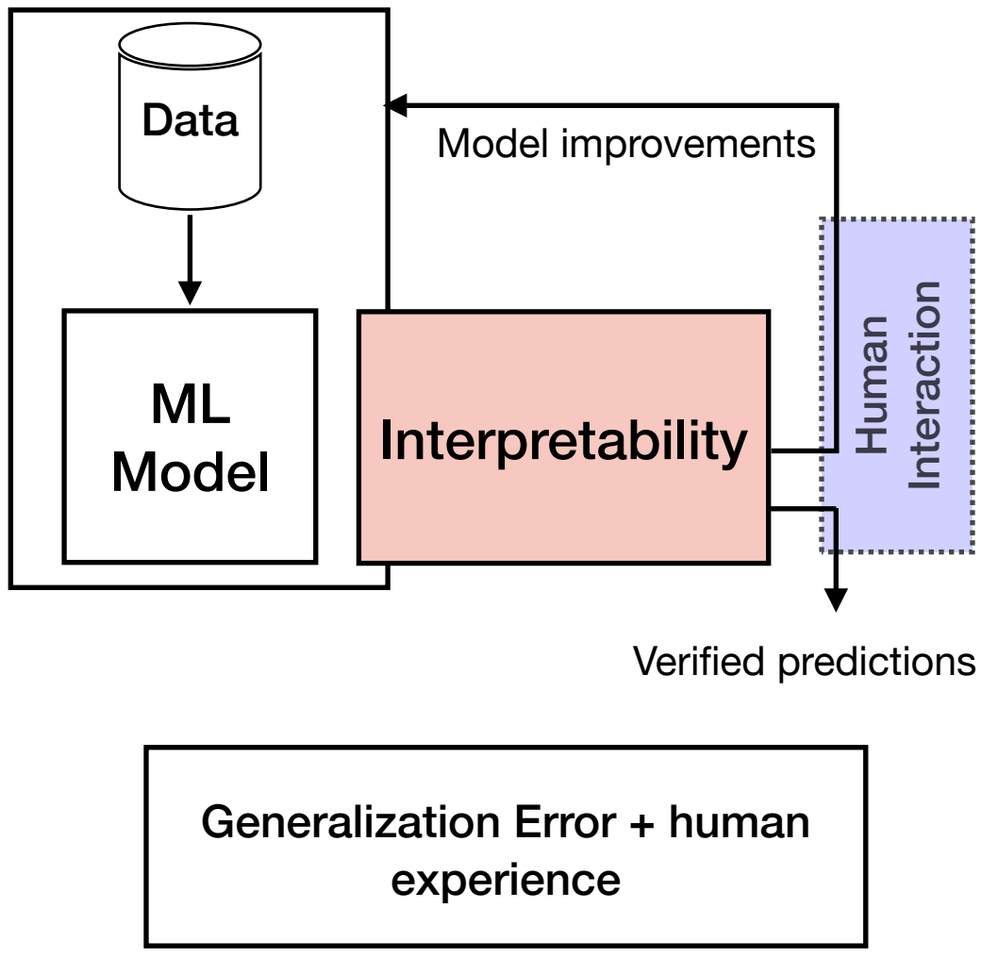
Interpretability as a first-class citizen in model building

Standard ML

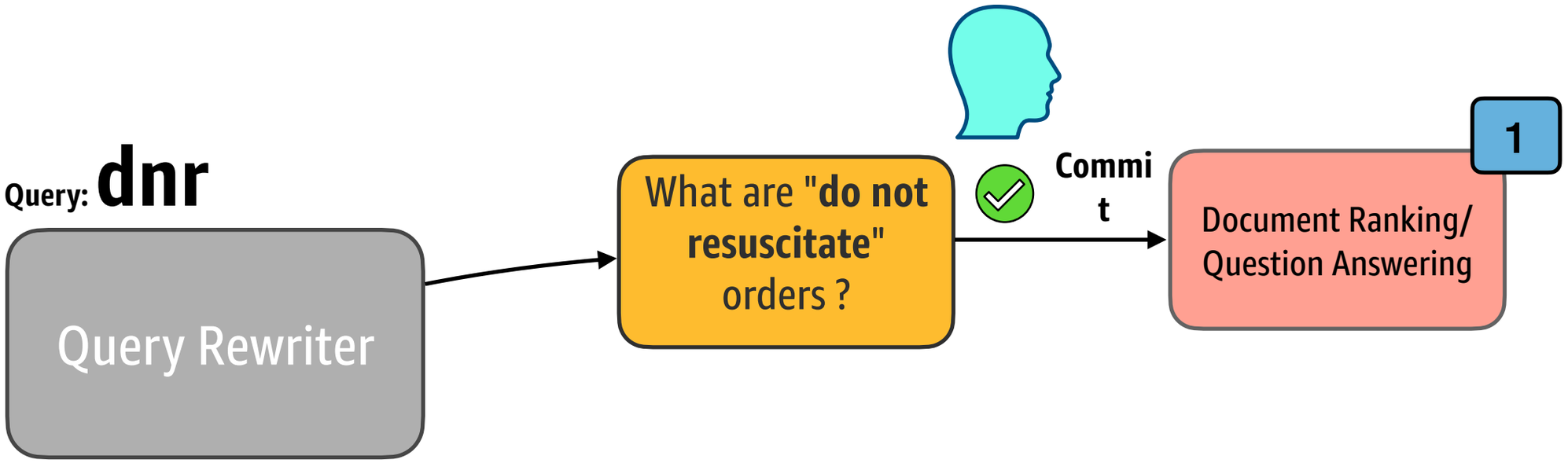




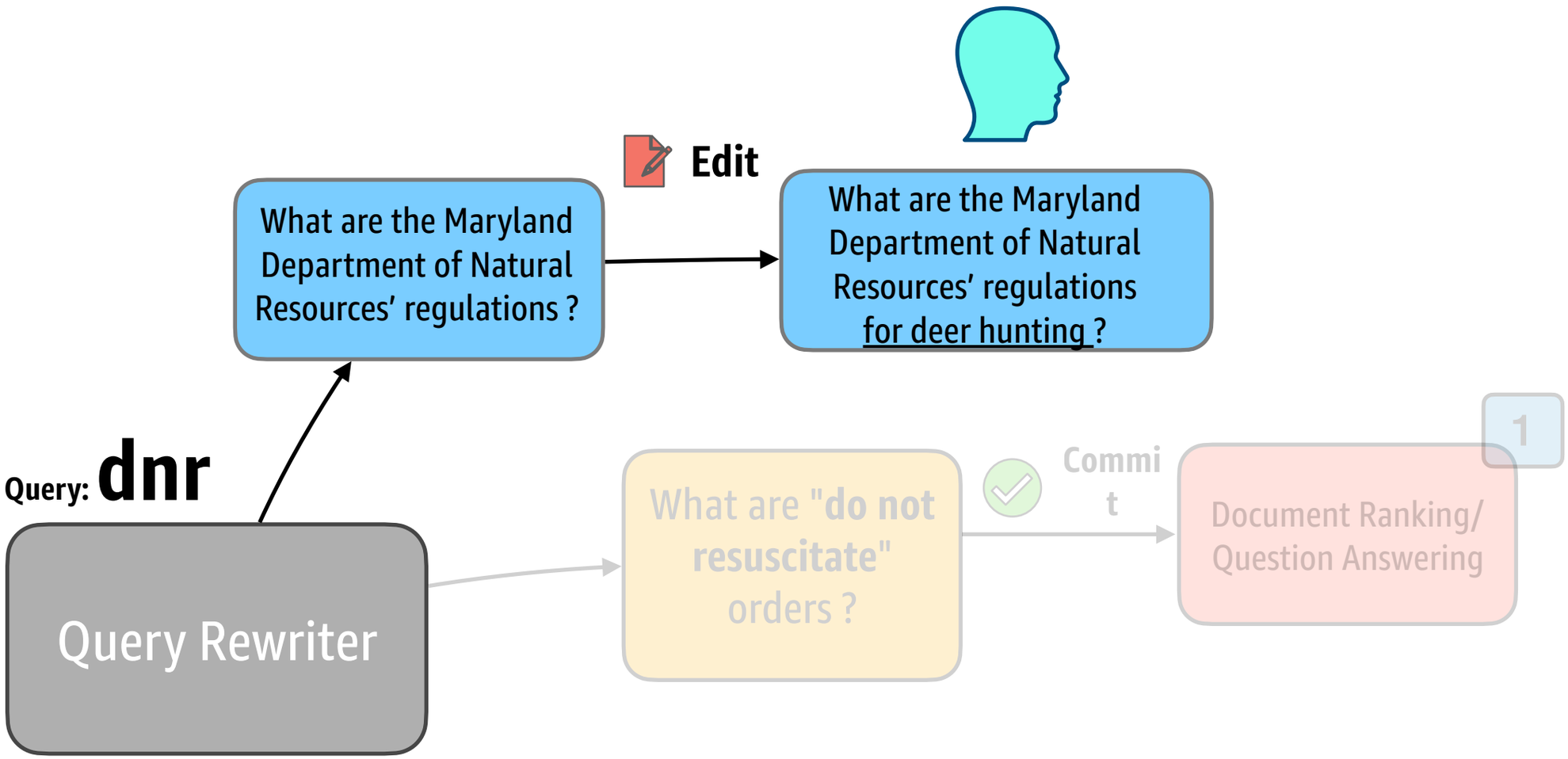
Interpretable ML



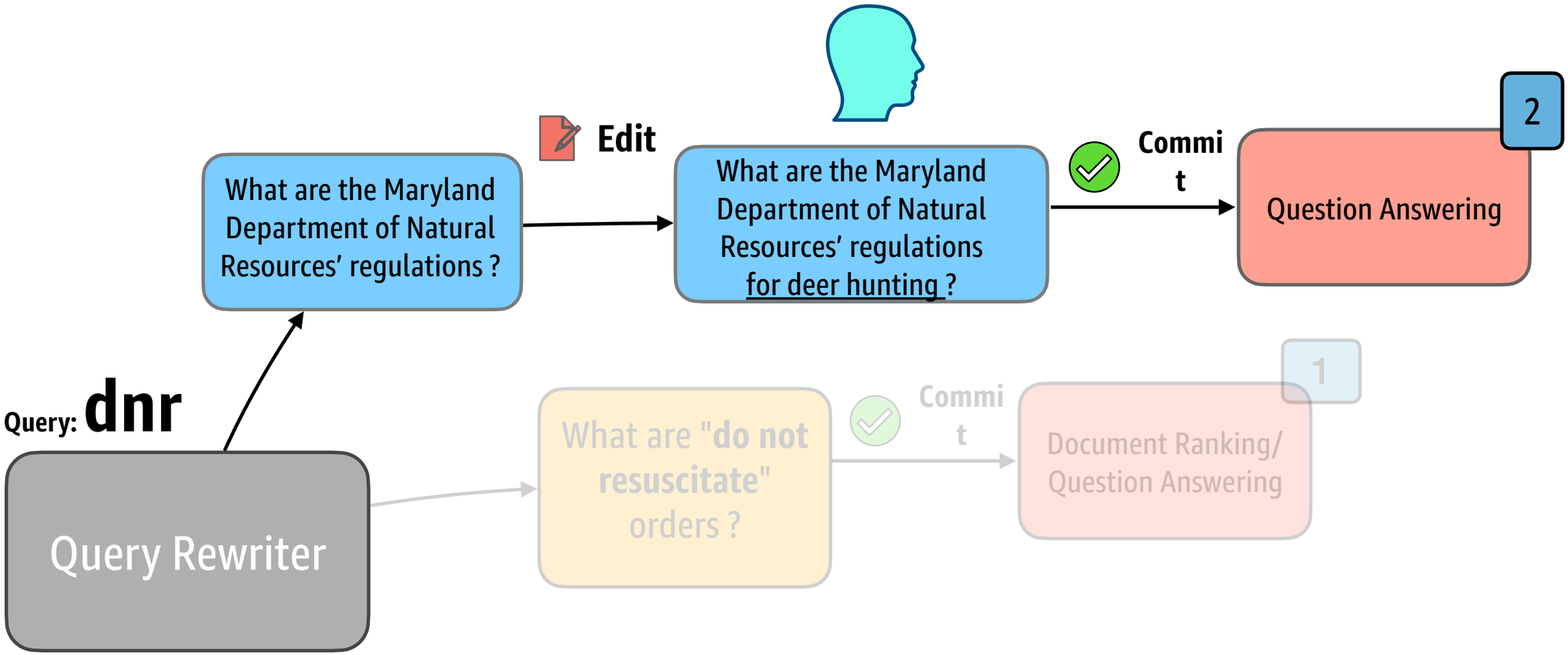
Interpretability in the age of LLMs



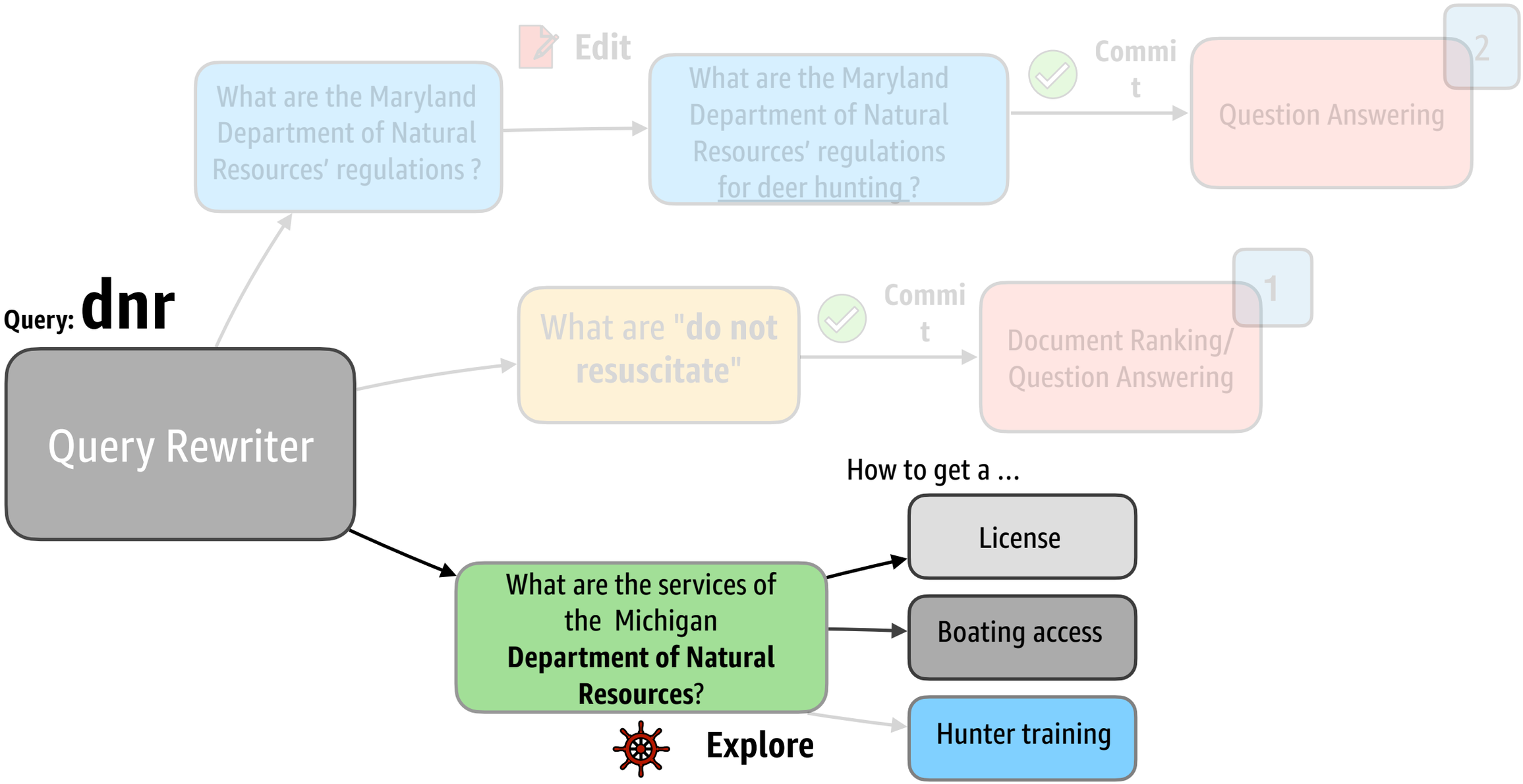
Interpretability in the age of LLMs

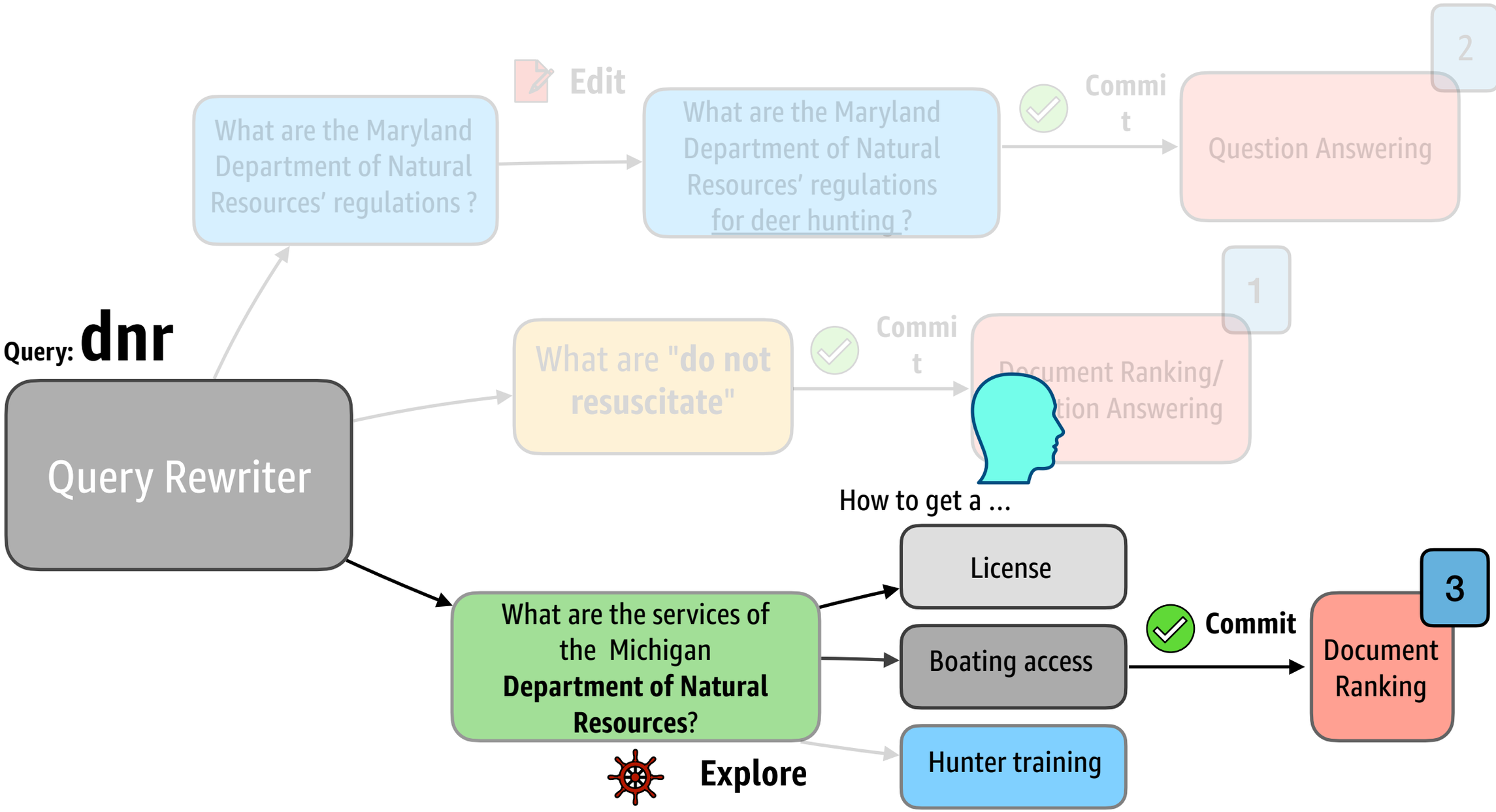


Interpretability in the age of LLMs

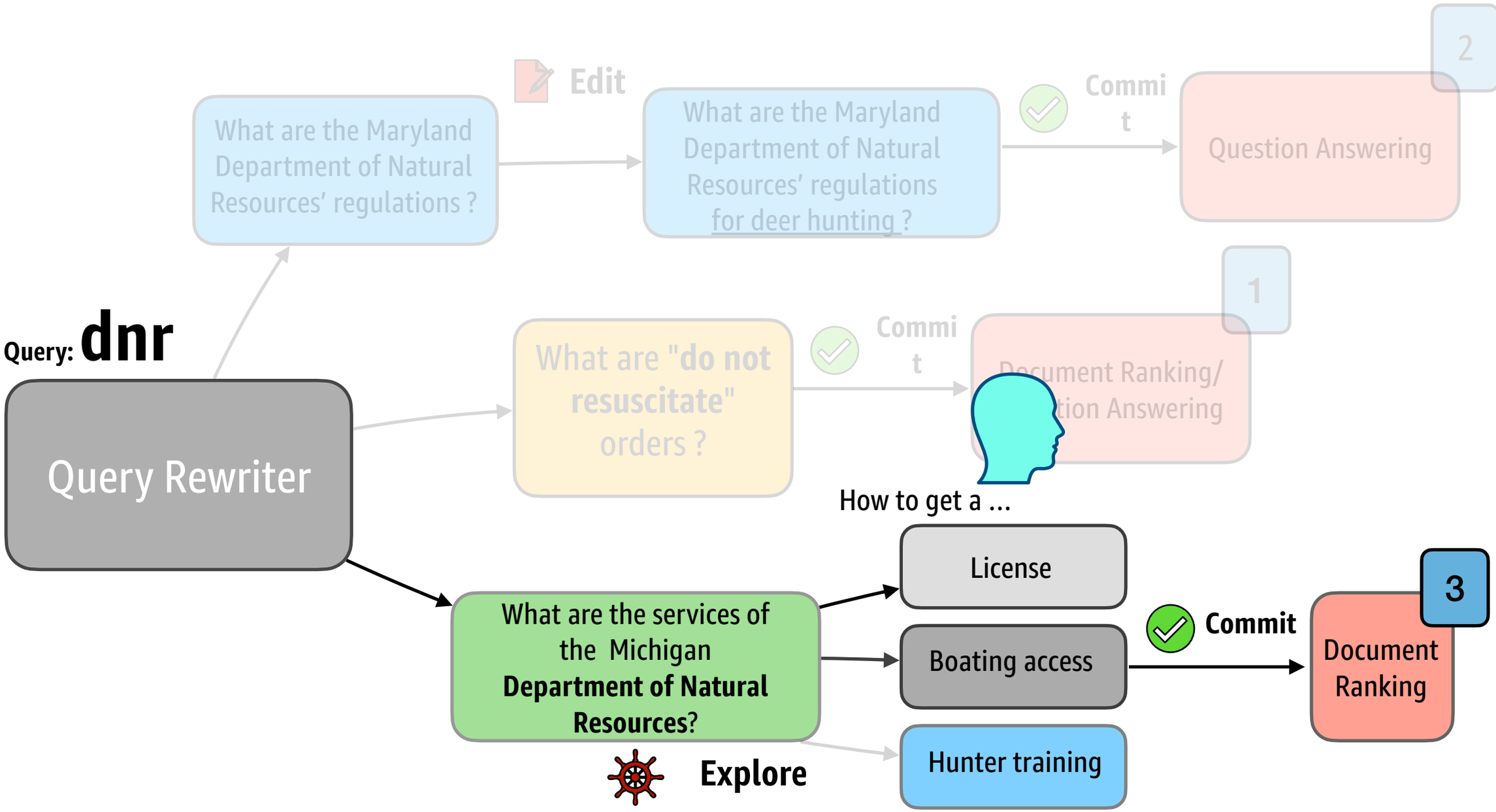


Interpretability in the age of LLMs




Query: dnr
Query Rewriter
What are the Maryland Department of Natural Resources' regulations ?
Edit
What are the Maryland Department of Natural Resources' regulations for deer hunting ?
Commit
Question Answering
2
What are "do not resuscitate"
Commit
Document Ranking/ tion Answering
1
How to get a ...
What are the services of the Michigan Department of Natural Resources?
Explore
License
Boating access
Hunter training
Commit
Document Ranking
3

Text instead of vectors



Conclusion

Introduction, motivation and notions

Posthoc interpretability

Intrinsic interpretability

Probing LLMs

Axiomatic IR for explaining IR models

Demo

Evaluation or ExIR methods

Conclusion and open problems

Thats it !!!