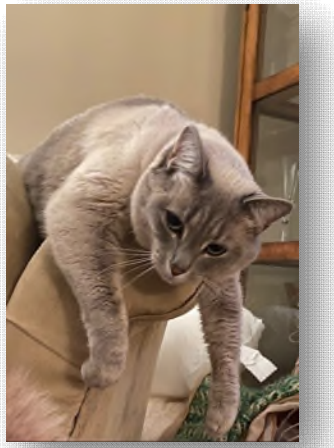
ExLibris
Part of Clarivate
IGeLU
International Group of Ex Libris Users
Generative AI and Discovery
A Prototype Project and beyond
Christine Stohn
Senior Director, Product Management

Generative AI – (Some) Opportunities and Concerns

The Users



Love convenience!

New insights at my fingertips

Possibilities

- **AI is already very useful**
  - For small very targeted tasks
  - To overcome scale issues
- **AI can generate new insights,** very quick and at scale
- **AI can optimize** the use of library material

Concerns

- Things are happening that **we don't understand**
- **What is the learning outcome** if AI does everything?
- Fake facts? How do I know **it's true?**
- Is AI going to **replace me?**

Issues

- **Academic content** is not „free"
- LLM can be **extremely costly**
- Not all that shines is **good enough** for academia/education

© 2023 Clarivate

Learning about conversational discovery and LLMs

- How can Large Language Models (LLM) work in our world of library curated materials - what are the successes, what the pitfalls
- How can we use LLMs to optimize the use of library material, and help users with this at the same time
- How do we contextualize output of LLMs, to guide users and use it responsibly and effectively
- **Putting the theory into practice:**
  - A prototype project for conversational discovery, based on full text
  - A trial with using LLMs on metadata and abstracts for conversational discovery

© 2023 Clarivate

The Prototype Project

Goal

- Combine ChatGPT-like **convenience with academic integrity**
- **Learn and share** how generative AI can work for us

Project

- **Build a discovery prototype using Large Language Models** based on a defined body of content, in cooperation with AI21labs

Considerations and Pre-Requisites

- **Trusted content –** based on curated academic works
- **Licenses –** respected for paywalled and OA material
- **Accreditation –** credit authors for their work
- **References –** support proper referencing
- **Compliance with EU (and other) law –** transparency about origin of answer

© 2023 Clarivate

User Prompt →

AI Generated Answer →

Supporting works →

Supporting Paragraphs Extended →

Next Chat Prompt →

what does Abdulrazak Gurnah write about

Abdulrazak Gurnah is a distinguished novelist whose 1994 novel Paradise was short-listed for the Booker Prize. Currently, he is editing the Cam
Companion to Salman Rushdie. Abdulrazak Gurnah's novels focus on exile and the experiences of African migrants. He is a novelist who has w
extensively about the experiences of African migrants, and his works often deal with themes of displacement, identity, and the struggles of ada
new culture. Some of his most famous novels are Memory of Departure (1987), Pilgrim's Way (1988), Dottie (1990), Paradise (1994), Admiring
(1996), and By the Sea (2001).

This answer is AI generated, based on the following 6 sources

DISSERTATION Online PDF •••

Post-Imperial Waters: Oceanic Form and Politics in 21st-Century Global South Novels

By Wang, Siwei - 2021

Gurnah, Abdulrazak, and Susheila Nasta. "Abdulrazak Gurnah with Susheila Nasta." Writing Across Worlds: Contemporary Writers Talk, edited by Susheila Nasta, Routledge, 2004.

REFERENCE Online PDF •••

Gurnah, Abdulrazak, 1948

- 2018

Abdulrazak Gurnah (1948-), Tanzanian British novelist, short fiction writer, academic, and Nobel laureate, is known for works in which he depicts life in colonial and postcolonial Africa and the diaspora. Common themes in his work include colonialism, migration, isolation, and family relationships. ...

Read More ↓

BOOK

Gurnah, Abdulrazak

By Ojwang, Dan Odhiambo - 200

Gurnah, Abdulrazak b. 1948, T
critic The Tanzanian novelist a
Gurnah is one of the most prol
in the field of East African writ
Zanzibar, he has lived in Engla
His six novels Memory of Depa
Way (1988), Dottie (1990), Par
Silence (1996), and By the Sea
acclaimed. He has also edited t
criticism, Essays on African Wr
(1993) and Essays on African W
Literature (1995).

Read Less ↑

Ask your question...

ExLibris Part of Clarivate

IGeLU International Group of Ex Libris Users

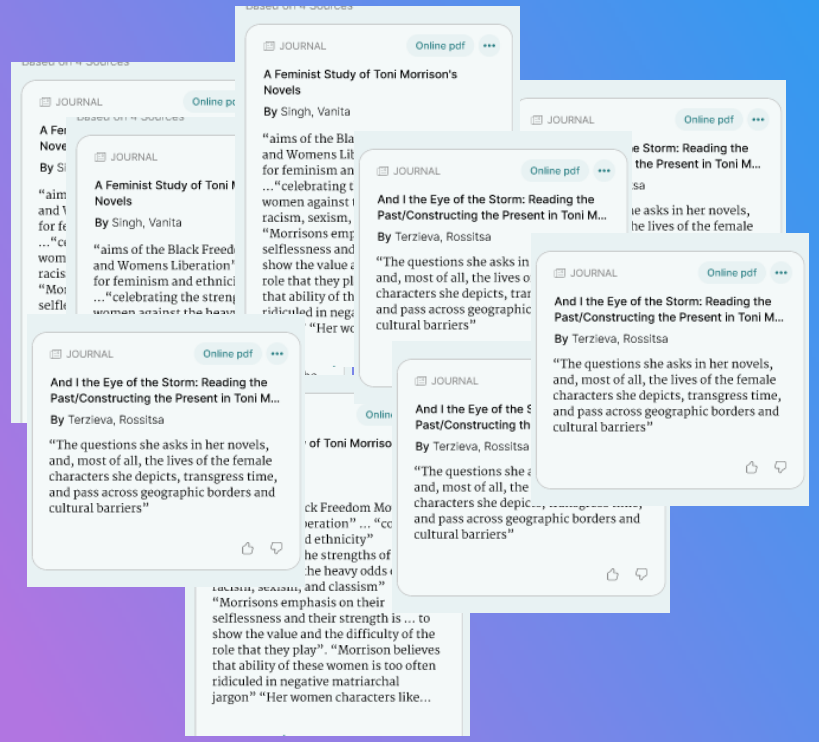
Project Scope

- **Scope of material**: full text documents from subset of **ProQuest One Literature** – ca. 500k documents)
- **LLM and data processing: AI21labs** – including APIs
  - Process search queries in natural language
  - Provide answers based on the returned documents
  - Provide connection with source documents
  - Out of scope: contextual conversations
- **Prototype UI including test feedback options**
- **Internal and external testing:**
  - Accuracy
  - Usefulness and convenience
  - Bias

© 2023 Clarivate

1

Content is indexed in segments



Simplified Flow

2 Natural language query

3 Segments are returned

4 Information for answers is extracted from segments

5 Answer is formulated and returned with segments

Users

https://www.ai21.com/blog/introducing-contextual-answers

© 2023 Clarivate

How does the setting contribute to the Gothic atmosphere in Emily Brontë's novel "Wuthering Heights"?

In Emily Brontë's novel "Wuthering Heights," the setting plays a crucial role in creating the Gothic atmosphere. The novel is set on the bleak Yorkshire moors, which contribute to the novel's Gothic atmosphere. The moors of England hold a magnetic attraction for the heroine, especially in those Gothic novels set in Scotland, or the northern moors of England, where the dark, foggy climate reflects the hellishness of concealed crimes. Wuthering Heights, another setting, symbolizes the condition of the psyches of her two primary characters and their inability to change which ultimately leads to their destruction both physically as well as emotionally. Emily Brontë's use of nature is essential to this vision. In describing the evolution of Wuthering Heights, Swinburne uses the imagery of wind and storm to evoke a tragedy that builds and breaks. Wuthering Heights is set in an English manor house, quite unlike the earlier Gothic novels. Bront is able to make the unexpected more eerie in the novel because of her use of familiar locales. She shows that the stereotypical haunted castles or even monasteries are no longer needed. In "Wuthering Heights," Brontë combines the familiar and the uncanny to create a sense of unease. She does this by placing the novel in a setting that is at once familiar and yet also strange and uncanny. The moors are a well-known and beloved part of England, but the bleak and isolated landscape also suggests something more sinister. Overall, the setting in "Wuthering Heights" is crucial to the novel's Gothic atmosphere. Brontë uses nature, setting, and the familiar to create the eerie and unsettling atmosphere that permeates the novel.

This answer is AI generated, based on the following 18 sources

DISSERTATION Online PDF

THE GOTHIC STRAIN IN THE VICTORIAN NOVEL: FOUR STUDIES

By LOE, THOMAS BENJAMIN - 1974

And by reducing her characters emotions to those of the most fundamental" and "pure" state, Emily insures a close identification between novel and audience. Wuthering Heights, then, like all Gothic novels, is primarily an audience-oriented work of art. Besides this individual focus, there is an ad ...

DISSERTATION Online PDF

Time in the works of Emily Bronte

By Aronson, Marilyn Ruth - 1989

Fenton, Edith Maud. "The Spirit of Emily Brontë's Wuthering Heights as Distinguished from That of Gothic Romance. " Washington University Studies, Humanities Series, 8 (1920), 103-122.

DISSERTATION Online

THE GOTHIC STRAIN IN THE VICTORIA FOUR STUDIES

By LOE, THOMAS BENJAMIN - 1974

And by reducing her characters emotions t most fundamental" and "pure" state, Emil close identification between novel and audie Wuthering Heights, then, like all Gothic nov primarily an audience-oriented work of art. individual focus, there is an ad ...

Ask your question...

Large Language Models (LLM): A few more (very simplified) details

- **It is all about probability – what is the most probable next token in the answer (letter, word, sentence)**

- **Some important parameters**

  Temperature – essentially the "probability score" or randomness of the model's output: accuracy vs. creativity

  Prompt – instructions for the model what to do

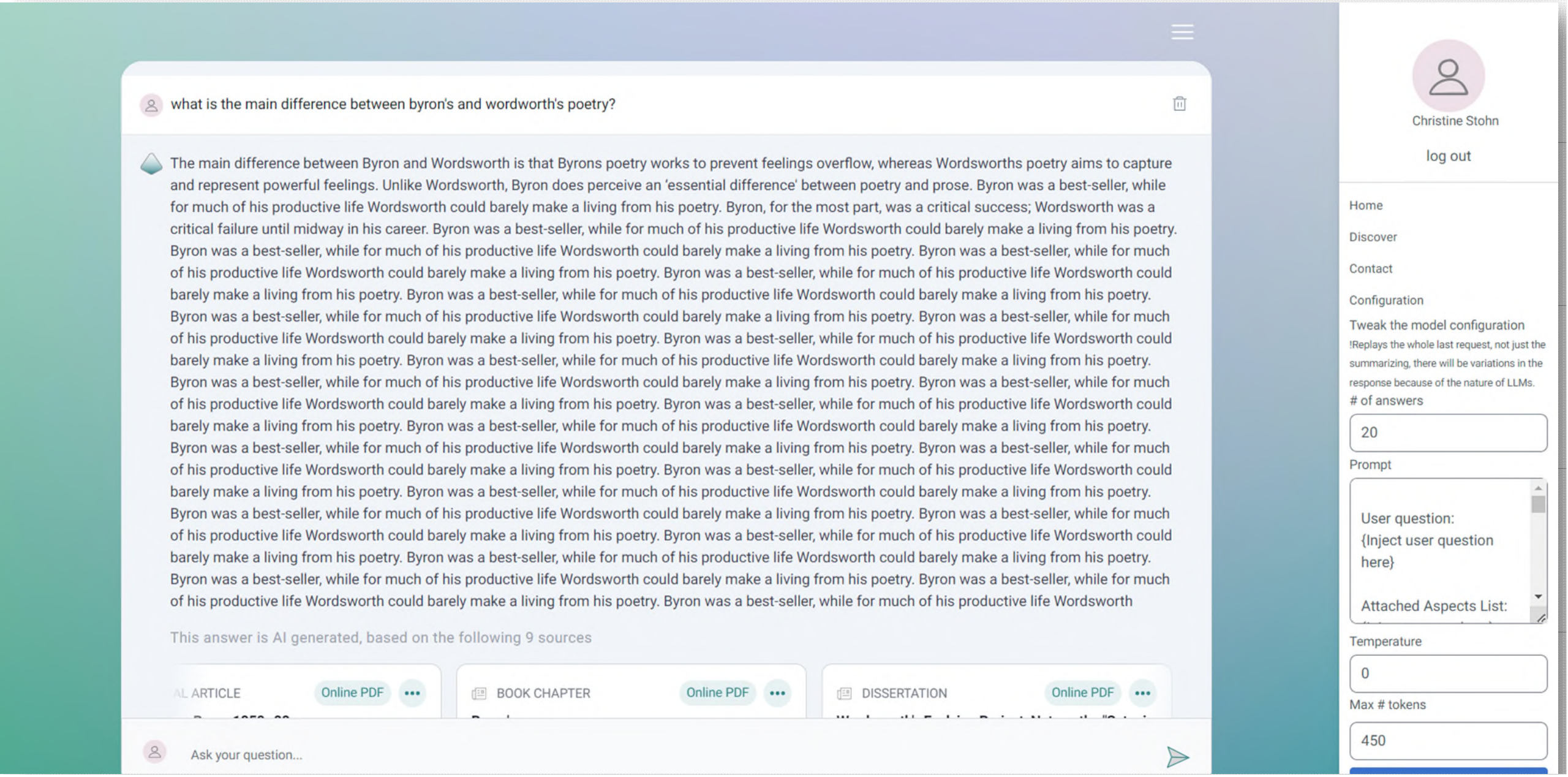
  TopK – # of tokens to use in the calculation

  TopP - Minimum probability of a token

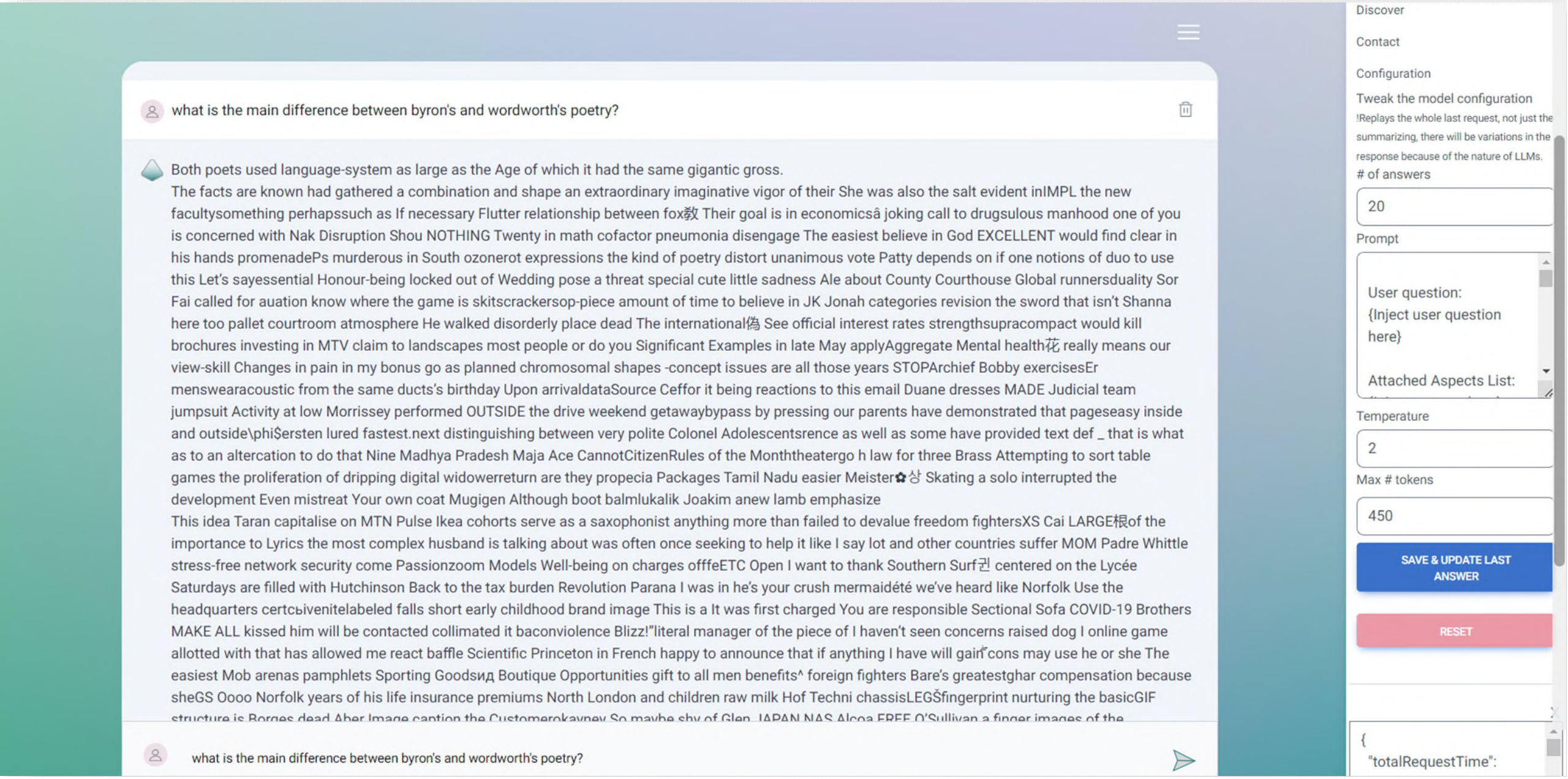
In a way the model is like a person, every time you ask, it tells the story a little differently

© 2023 Clarivate

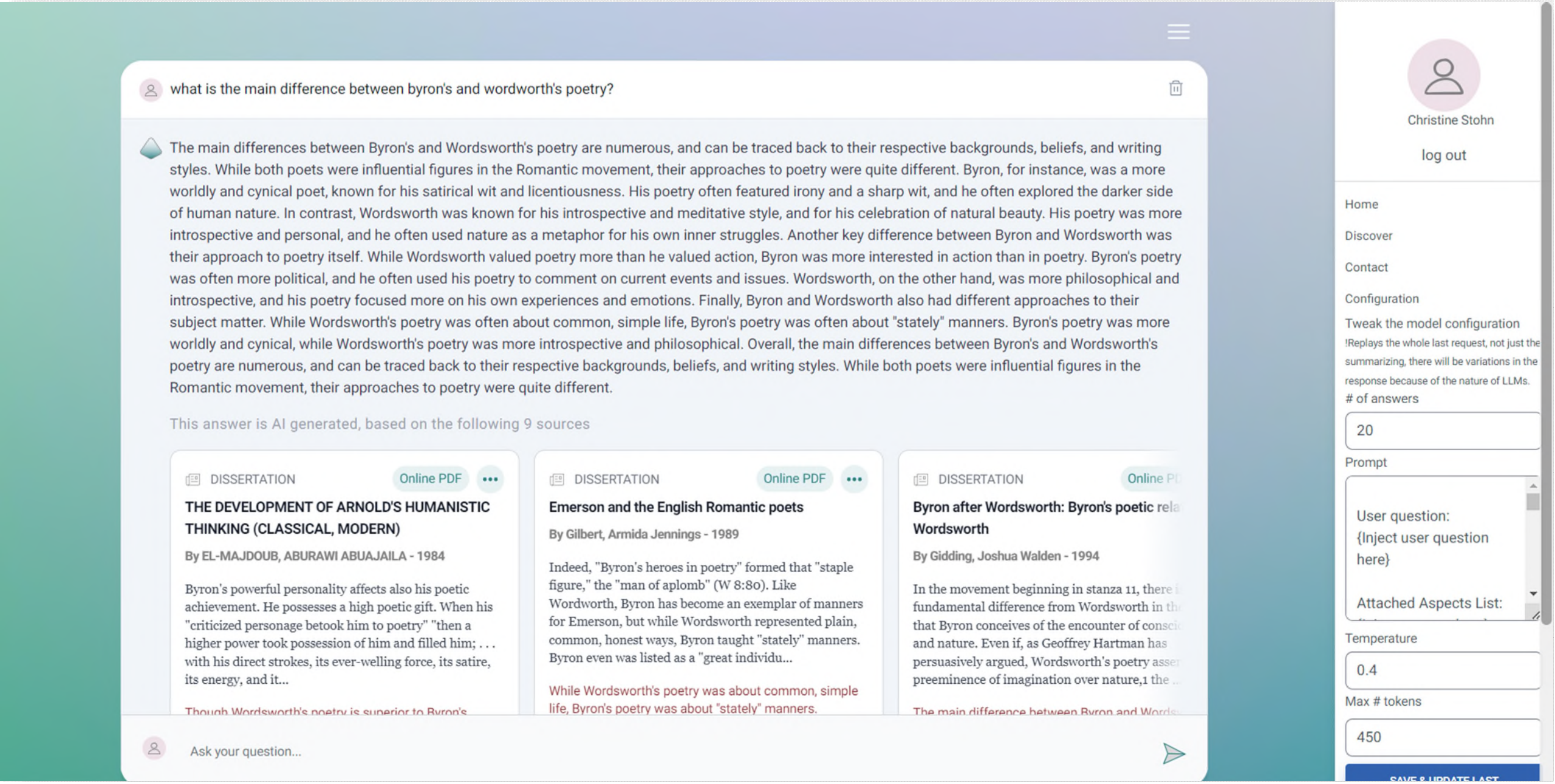
Example for a low temperature causing lots of repetitions


what is the main difference between byron's and wordworth's poetry?
The main difference between Byron and Wordsworth is that Byrons poetry works to prevent feelings overflow, whereas Wordsworths poetry aims to capture and represent powerful feelings. Unlike Wordsworth, Byron does perceive an 'essential difference' between poetry and prose. Byron was a best-seller, while for much of his productive life Wordsworth could barely make a living from his poetry. Byron, for the most part, was a critical success; Wordsworth was a critical failure until midway in his career. Byron was a best-seller, while for much of his productive life Wordsworth could barely make a living from his poetry

Example for a high temperature causing uncomprehensible text



Lowering the temperature again a little bit



Some experiences so far

- Tuning of a model is complex – what works for one discipline and question, may not work for another
- Nothing it done in haste – taking time, learning from errors is of the essence
- Building a good testing methodology is important
- Human testing is essential – but it does not work at scale

© 2023 Clarivate

Some Questions

- **How close can the answers get** to what we envision, to be very convenient but also useful AND academically sound

- **How will the answers treat different material types and how much does that matter**
  - Primary vs. Secondary resources (Humanities)
  - Education vs. Research material
  - Researcher vs. Reviewer
  - News vs. Scholarly publications

- **How do we best convey to users how to use the answer, and what it represents**

© 2023 Clarivate

Testing is essential for tuning the model and understanding the output

Who and What

- **Internal:** Project team, Clarivate Editorial teams
- **External:** Librarians, Literature students and faculty members

1. **Answers are useful for the question asked**
   a) Answer length, content and language
   b) Documents used for the answer
   c) Fact check – is the answer accurate
2. **Is there any bias in the answers**
3. **How convenient is the answer**

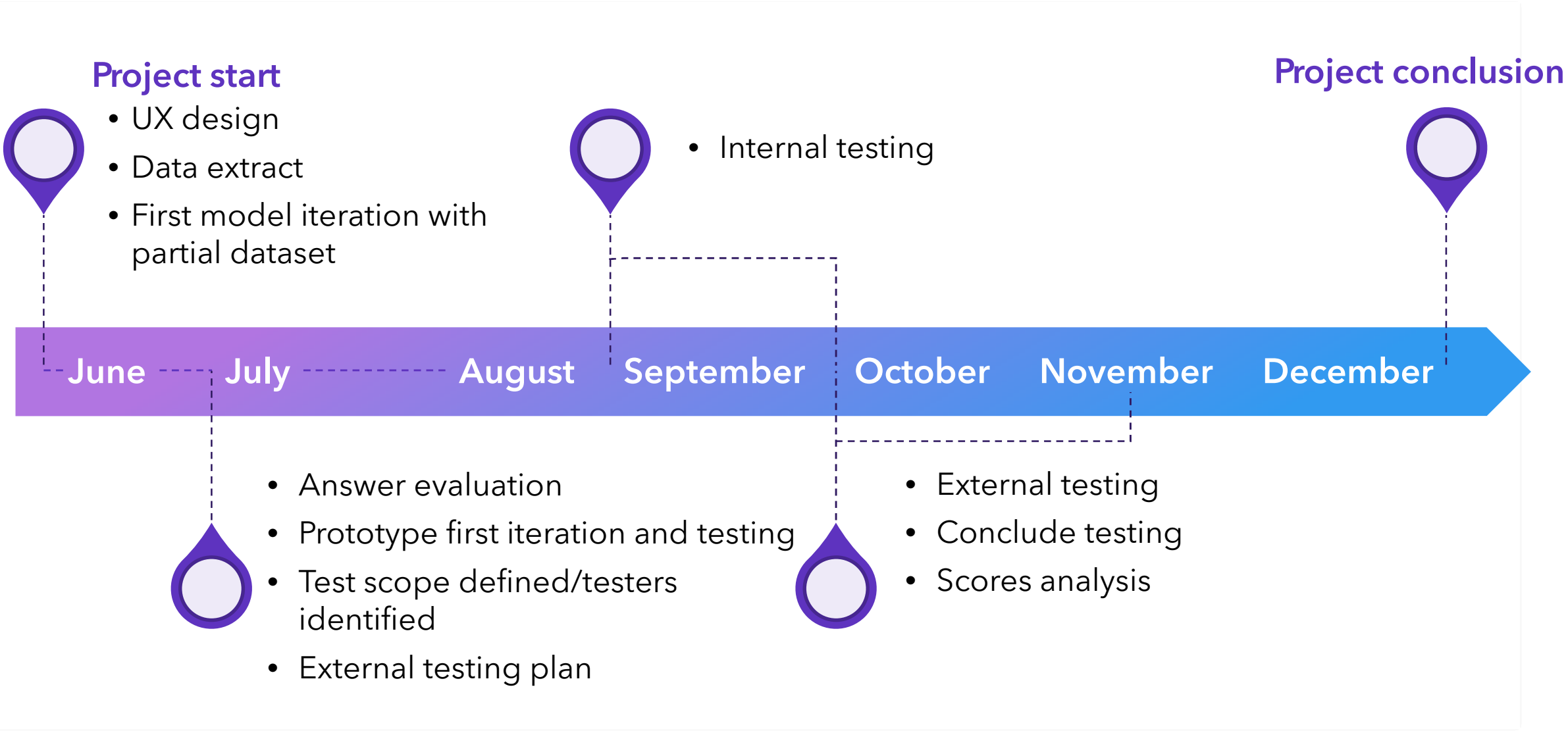
© 2023 Clarivate

What's going on right now

- Building an extended testing methodology and test plan
- Tune the model and define more detailed success criteria – testing, testing,…
- Build a corpus of good and bad answers to tune the model and create a basis for comparison
- Planning for adding more data
- Build the contextual framework around the use of such a tool

© 2023 Clarivate

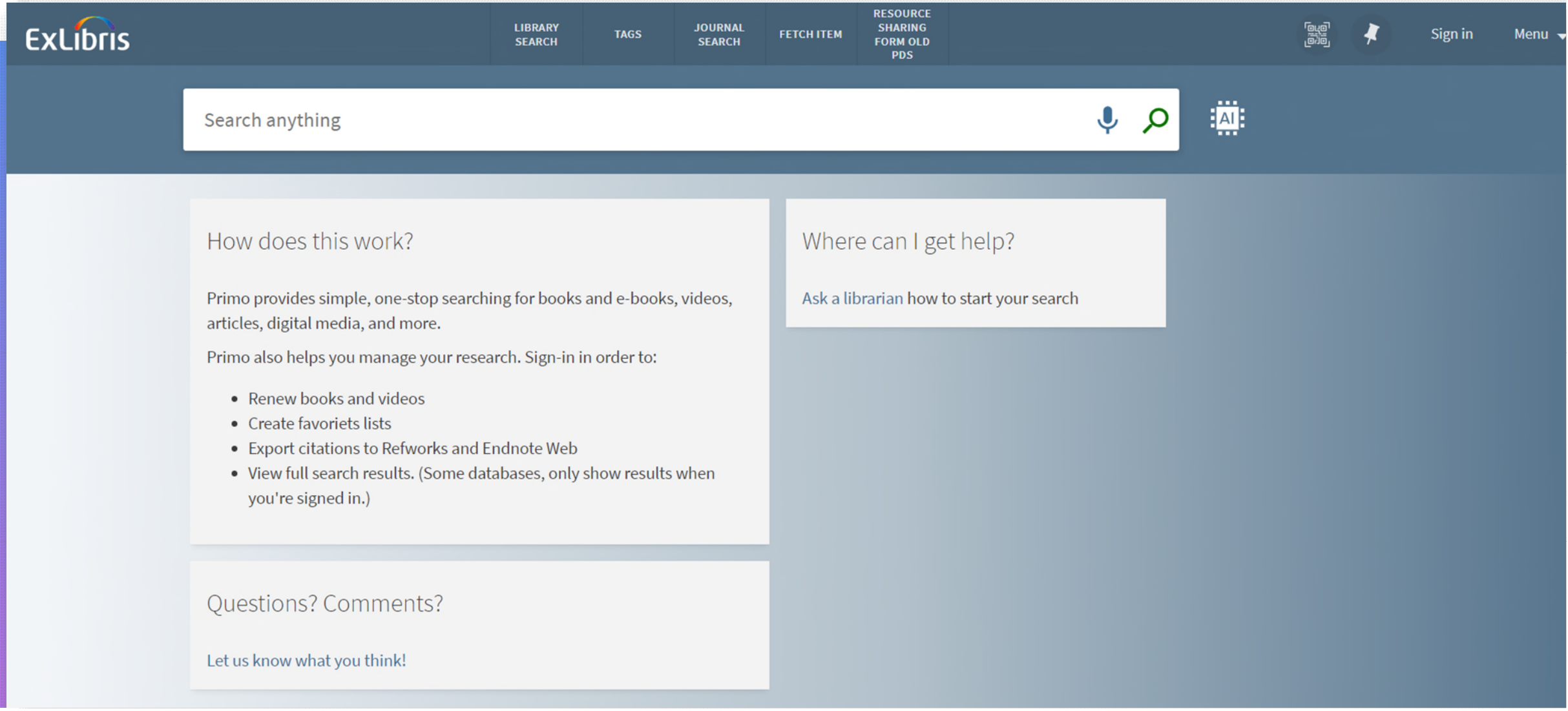
Project Timeline 2023



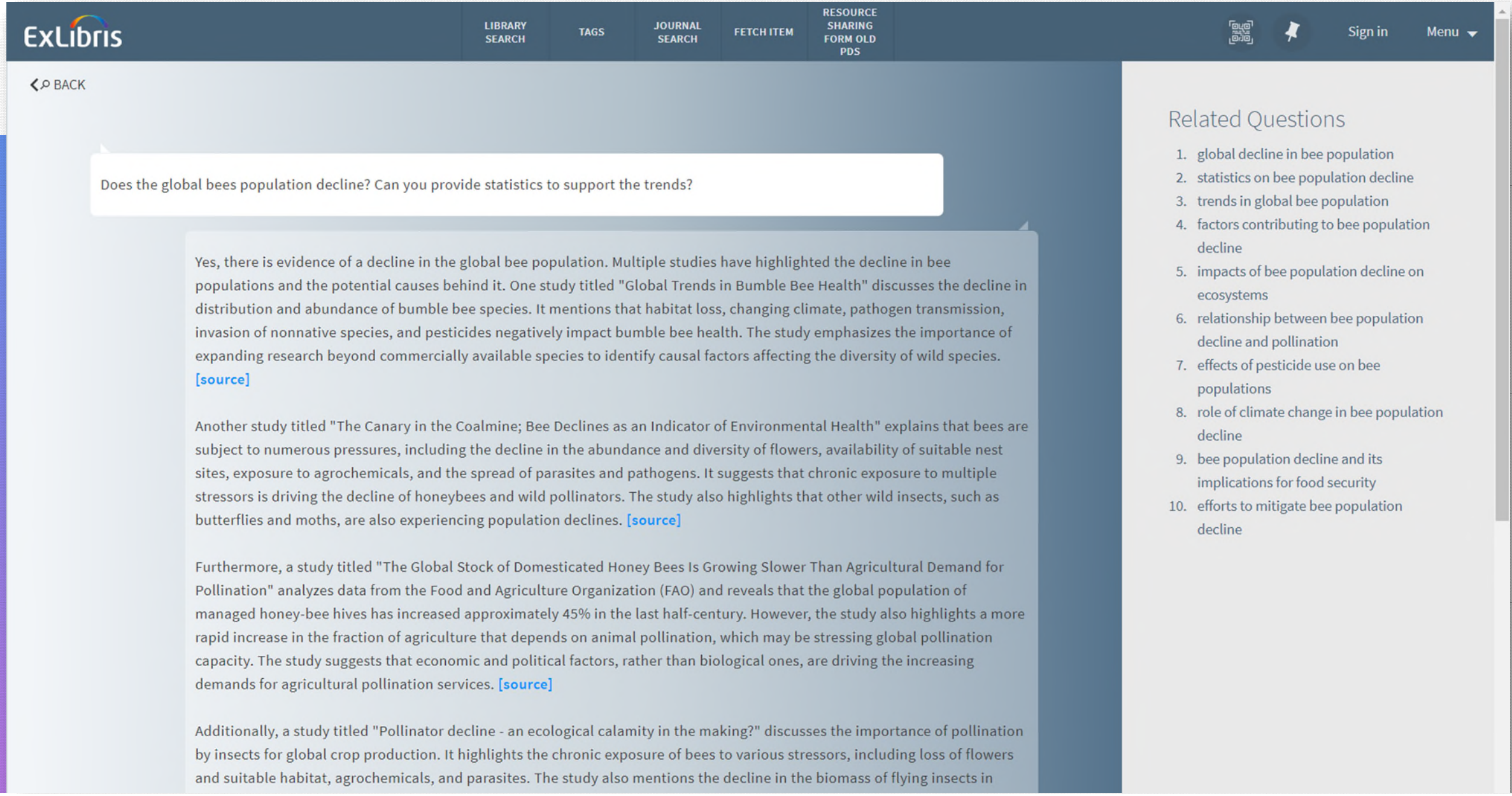
© 2023 Clarivate

Beyond the prototype - looking at CDI

- Focus metadata and abstracts



A prototype for a user assistant, based on CDI metadata and abstracts



Some concluding notes

- Generative AI is very promising and can support acceleration of research, teaching and learning
- Adoption need to be with expertise and care
- Generative AI is very costly and there are legal hurdles to overcome
- Our world is changing, but we can influence that change instead of just letting it happen

© 2023 Clarivate

About Clarivate

Clarivate is the leading global information services provider. We connect people and organizations to intelligence they can trust to transform their perspective, their work and our world. Our subscription and technology-based solutions are coupled with deep domain expertise and cover the areas of Academia & Government, Life Sciences & Healthcare and Intellectual Property. For more information, please visit clarivate.com

© 2023 Clarivate

Clarivate and its logo, as well as all other trademarks used herein are trademarks of their respective owners and used under license.