KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

GenAI language assistants that you can trust

Webinar - April 4th 2024 - 15h

Product update: Retrieval Augmented Generation (RAG)

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

- ✓ **Introduction**
- ✓ **Demo: Analysis on the Enron email corpus**
- ✓ **Behind the scenes**
- ✓ **Discussion, Q&A**

Vincent Nibart
Co-founder and Chief Product Officer
vincent.nibart@kairntech.com

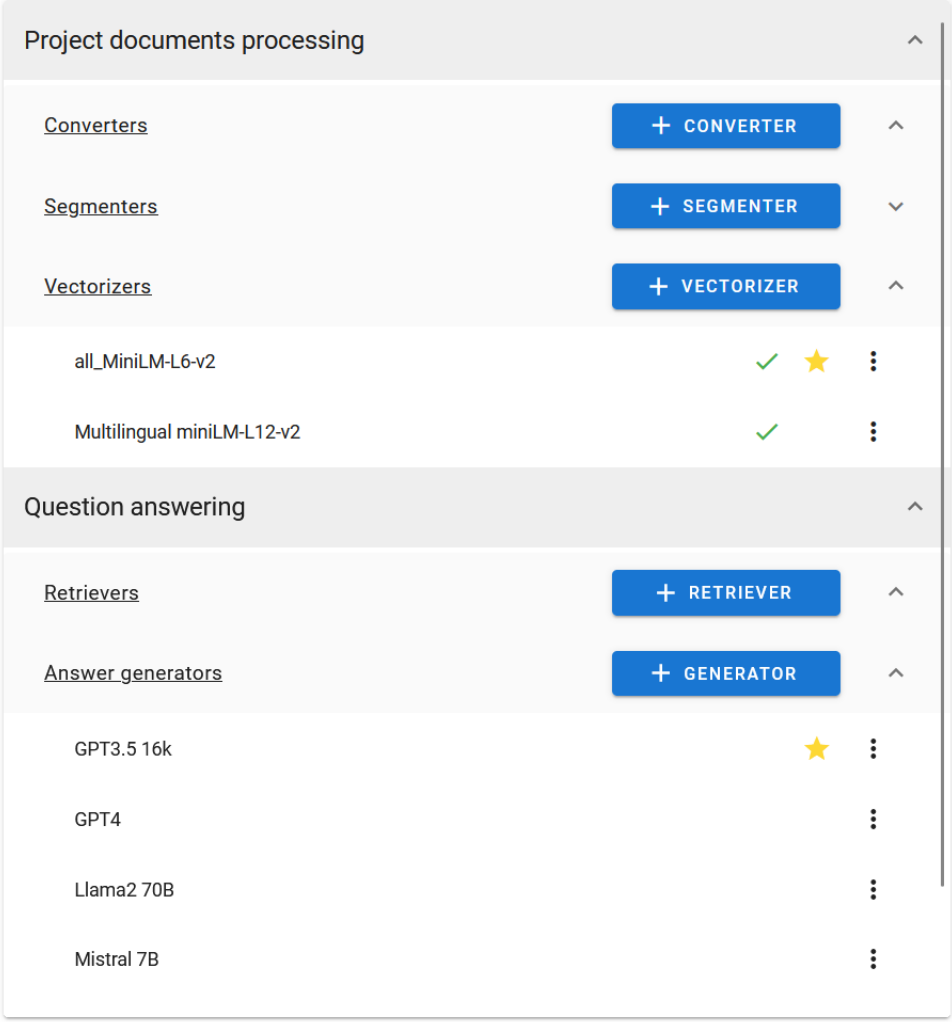
Stefan Geissler
CEO Germany
Stefan.Geissler@kairntech.com

The Kairntech software keeps evolving

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

New Version (Feb 2024)

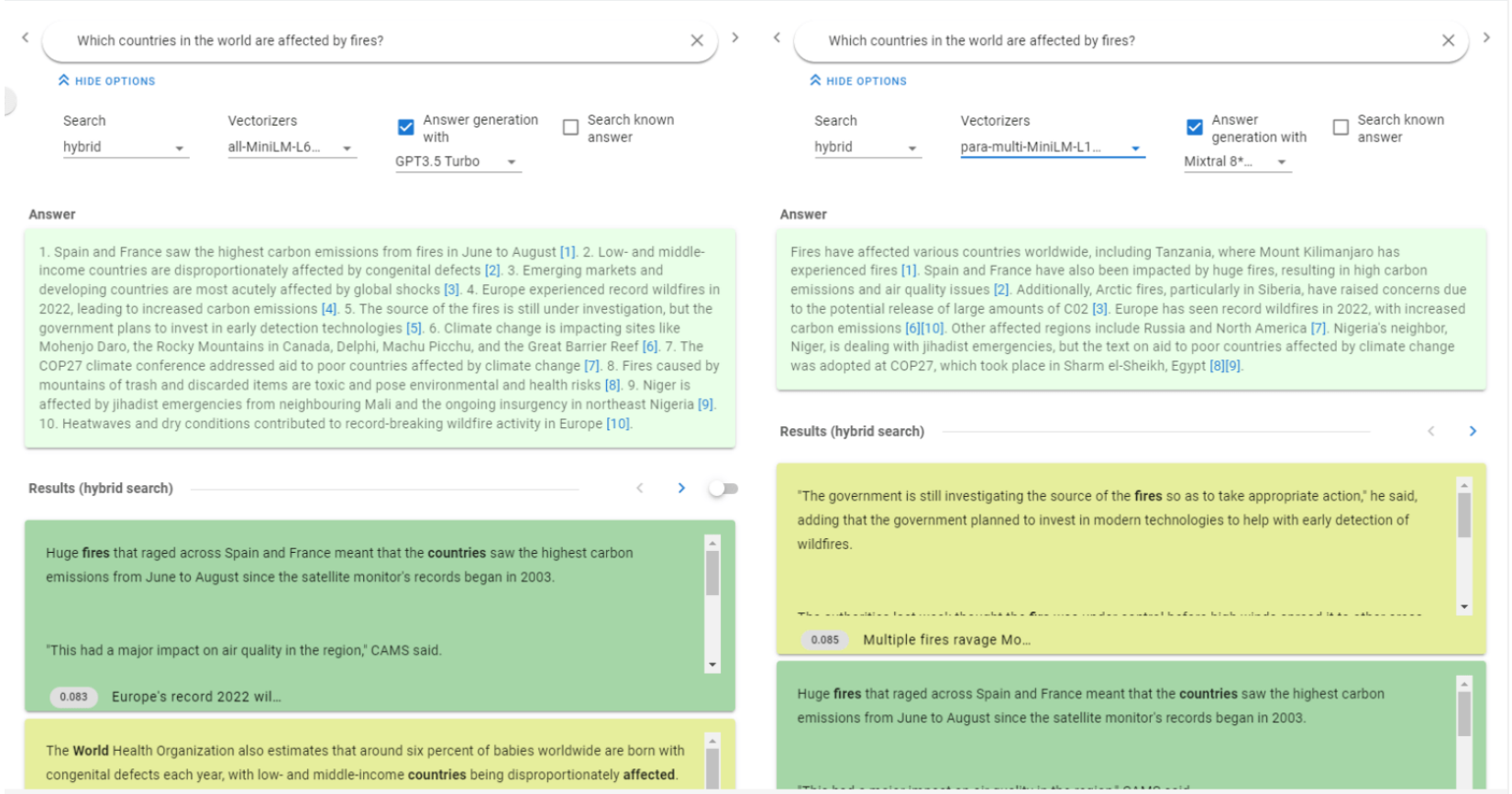
- ✓ Customizable components for processing steps
  - Converters, Segmenters, Vectorizers, Retrievers



The Kairntech software keeps evolving

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

✓ Quality assessment supported by visual comparison



✓ … but most prominently **RAG Retrieval-Augmented Generation**

Large Language Models bring significant progress on many use cases

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

- ✓ Summarize a longer text

  “Sum up the main points in 10 sentences.”

- ✓ Generate text on the basis of a prompt

  “Write a one-page text to promote my services as a web designer”

- ✓ Return specific information

  “What are the advantages and disadvantages of SPARQL compared to CYPHER?”

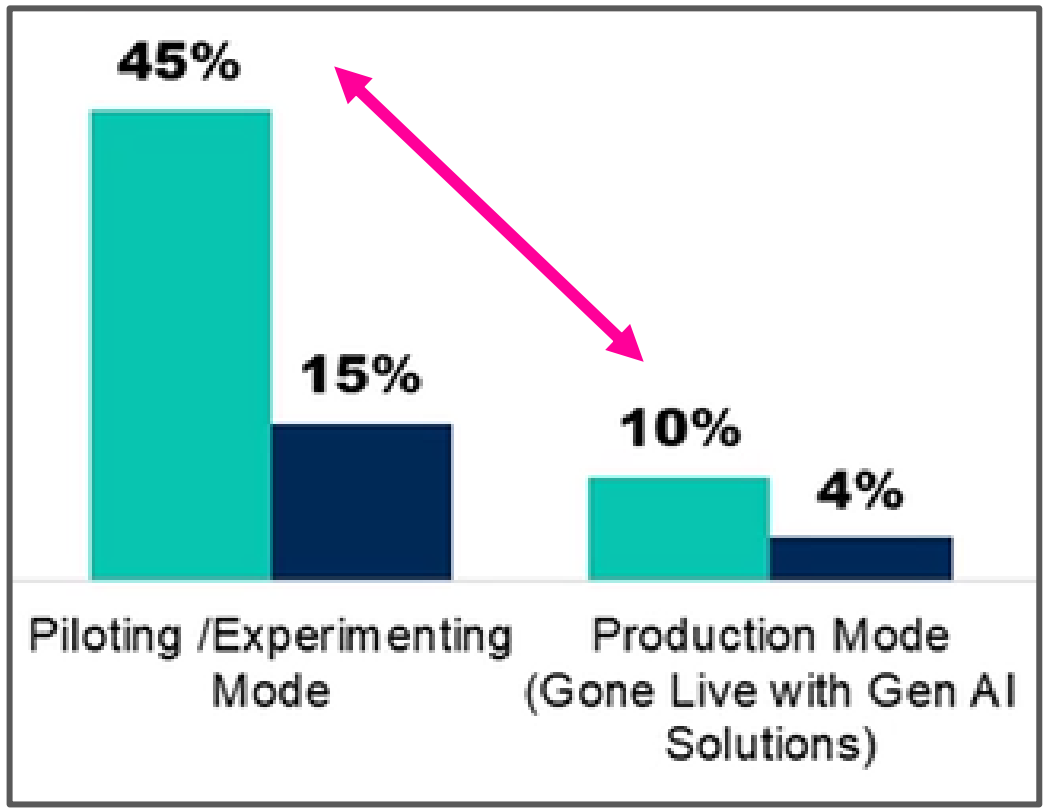
- ✓ Transform a text from one form into another

  Automatic Translation / “Return the main arguments as a table.”

- ✓ …

Status of GenAI initiatives

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



Source: Gartner (2023 vs 2022)

Many are experimenting. But few projects in production

- Insufficient accuracy
- Lack of governance
- Technical difficulties
- ROI?

Dearly needed: Guidance!

- Which use cases are **feasible**?
- How to go from experiment to **deployment**?

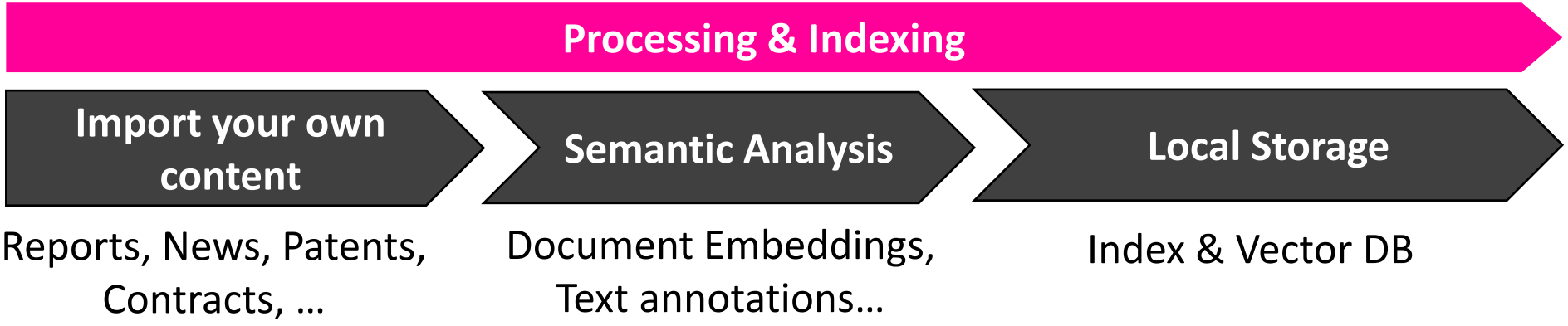
LLMs leave many questions open in business contexts

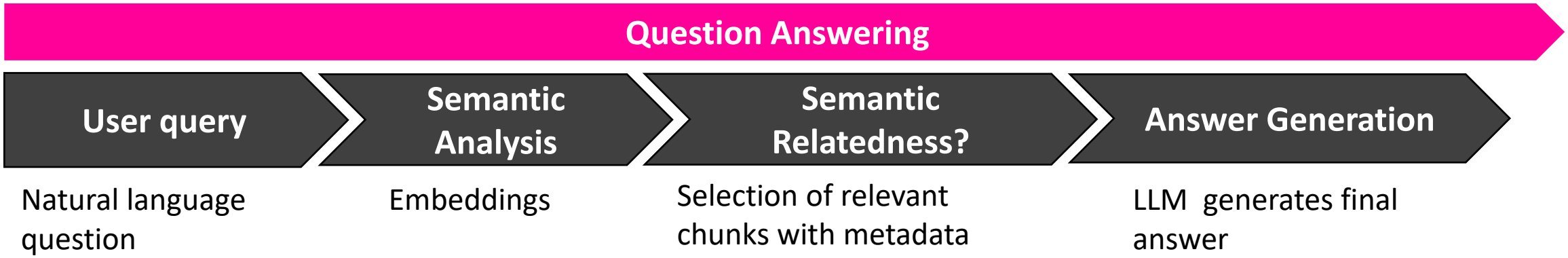
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

- ✓ Take into account **my** confidential information
- ✓ Updating: cost / computing efforts
- ✓ Hallucinations
- ✓ Adding custom concepts and metadata
- ✓ Traceability to sources

The technical answer: RAG

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



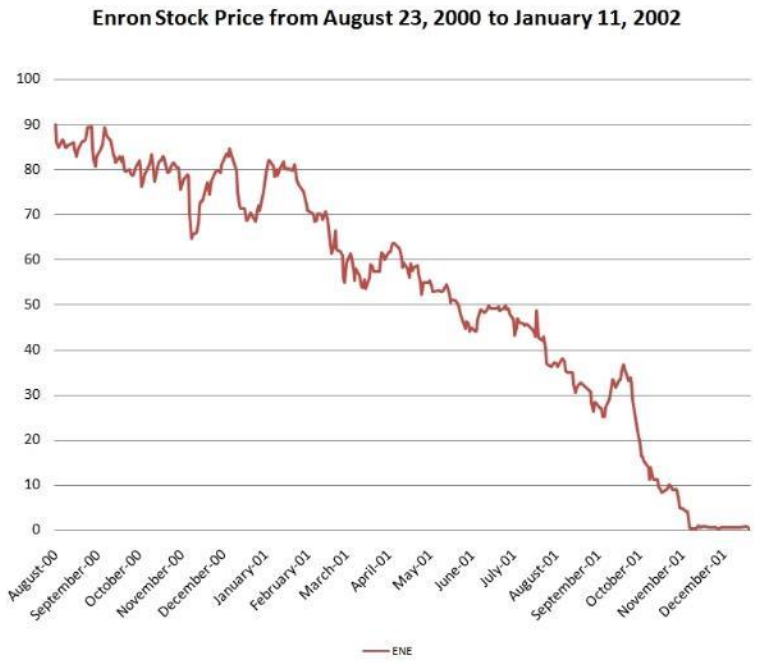


Demo 1 - Enron email corpus

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

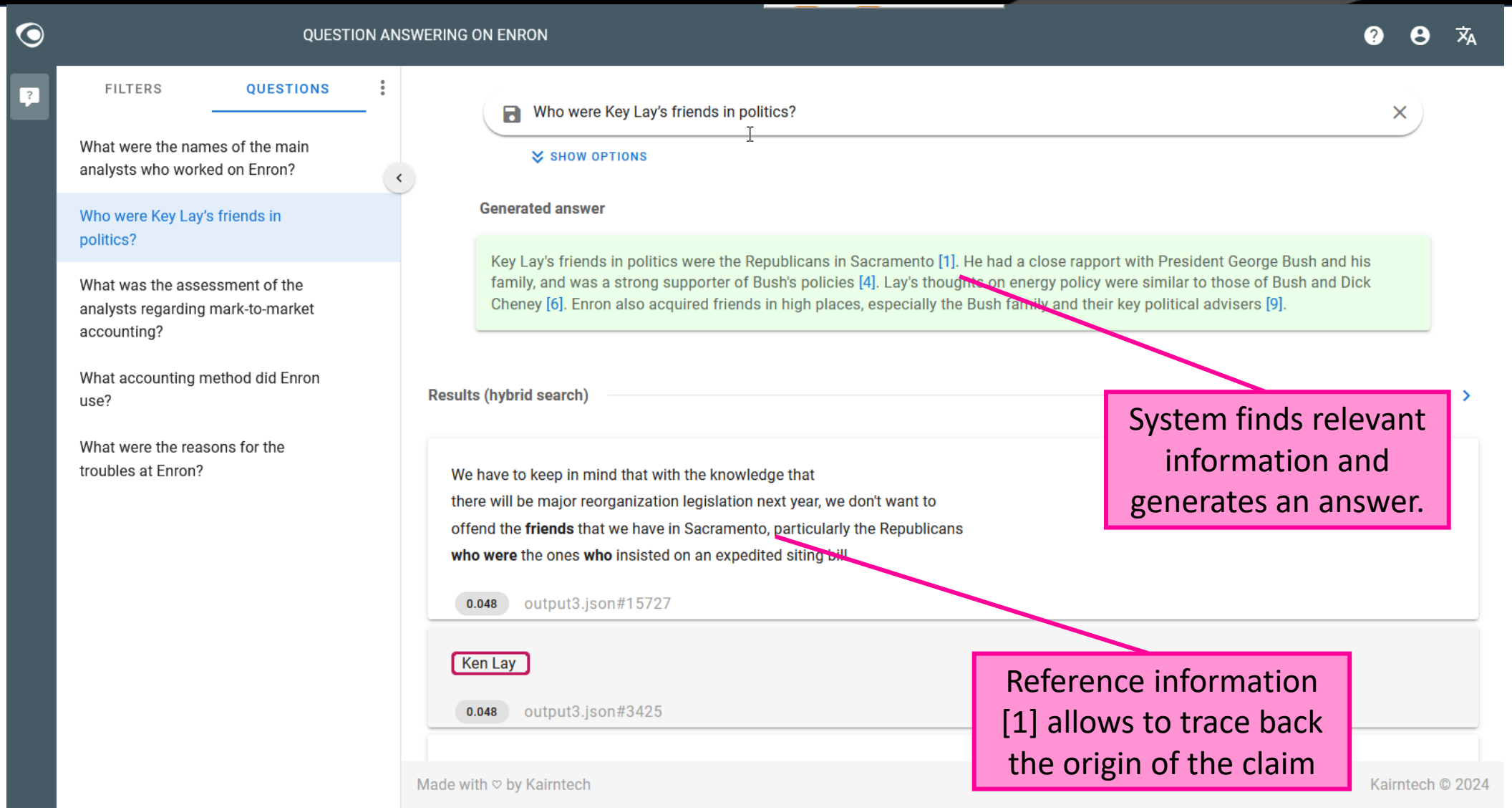
Enron?

- ✓ Huge **financial fraud** case in the US in 2001
- ✓ Collusion with accounting firm & politics
- ✓ Hundreds of thousands of emails around the case are **publicly available**



Demo 1 – asking questions

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



System finds relevant information and generates an answer.

Reference information [1] allows to trace back the origin of the claim

Demo 1 – asking questions

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

First answer ok … but result is better when using the more advanced GPT 4 than the (simpler) default GPT 3.5

What were the reasons for the troubles at Enron?

SHOW OPTIONS

Generated answer

The reasons for the troubles at Enron were poor financial disclosure [8], involvement in multiple areas too quickly [10], opaque and complex accounts [10], and losses in financial activities [10].

Results (hybrid search)

"We're getting used to taking the stairs," she said. "And we're learning to save our work on computers more often. That's important."
Barnes & Noble experienced a blackout this winter. For safety **reasons**,

What were the reasons for the troubles at Enron?

HIDE OPTIONS

Search	Vectorizers	Answer generation with
hybrid	Ada002	GPT-4

Generated answer

The troubles at Enron can be attributed to several key factors. High-ranking executives, including CEO Skilling, left the company abruptly, leading to speculation and concern among investors about the real reasons behind these departures and the company's financial health [5]. Enron's aggressive culture encouraged rapid deal-making and advancement, potentially at the expense of sound business practices [7]. Mismanagement and poor financial disclosure made it difficult for investors and analysts to understand the true state of the company's finances [8]. Additionally, Enron's rapid diversification into areas like water supplies and its involvement in complex financial activities, which led to significant losses and the resignation of its chief financial officer, further exacerbated its problems [10].

Results (hybrid search)

"We're getting used to taking the stairs," she said. "And we're learning to save our work on computers more often. That's important."

Demo 1 – asking questions

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

What accounting method did Enron use?

SHOW OPTIONS

Financial dealings mentioned as one cause of the problems („made it difficult to understand the financials…“) What accounting method did they use?

Generated answer

Enron used "mark-to-market accounting" on an unprecedented scale, which involved pricing their deals at what Enron considered the current market value. This method allowed Enron to take most of its profit upfront, even for deals that spanned several years, by recording these anticipated profits in the current year [7].

Results (semantic search)

What was the assessment of the analysts regarding mark-to-market accounting?

SHOW OPTIONS

„Mark-to-market“ accouting? What was that again and what did the experts say?

Generated answer

Analysts assessed mark-to-market accounting with caution, highlighting its potential to create earnings volatility due to discrepancies between calculated values and actual cash realized, and the challenges in valuing positions in illiquid markets [1][2]. ABN AMRO Inc. analyst Paul Patterson expressed a preference for cash-based earnings, suggesting that earnings more closely reflecting the timely realization of cash are of higher quality [4][8]. Patterson also anticipated that investors would become more discerning, learning to differentiate between earnings growth driven by accrual and mark-to-market accounting [4].

Results (semantic search)

Although many of Enron's financial problems aren't solely the =
fault of **mark**-to-**market accounting** issues, there has been growing attention=
paid to this form of financial reporting because of the earnings volatilit=

Demo 1 – asking questions

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Generated answer

Analysts assessed mark-to-market accounting with caution, highlighting its potential to create earnings volatility due to discrepancies between calculated values and actual cash realized, and the challenges in valuing positions in illiquid markets [1][2]. ABN AMRO Inc. analyst Paul Patterson expressed a preference for cash-based earnings, suggesting that earnings more closely reflecting the timely realization of cash are of higher quality [4][8]. Patterson also anticipated that investors would become more d
differentiate between earnings growth driven by accrual and mark-to-market accounting [4].

Although many of Enron's financial problems aren't solely the =
fault of **mark**-to-**market accounting** issues, there has been growing attention=
paid to this form of financial reporting because of the earnings volatilit=
y it can create.=20
Answers Elude Investors=20

Investors are asking hard questions, and not always getting the answers the=
y want.=20
Using **mark**-to-**market** methods, a company calculates the fair **market** value of=
a commodity position - whether it's a contract, an option, a swap, etc. - =
at the time, even if the value of the position is realized over a longer pe=
riod. The problem with this method is the actual cash a company realizes fr=
om the position might not be the same value the company calculated in its o=
riginal **assessment**. Also, sometimes it isn't easy to calculate the fair val=
ue of the commodity position.

1.000 maildir.json#4811

maildir.json#4811

Tracability: From the generated answer to the precise piece of text within the larger document.

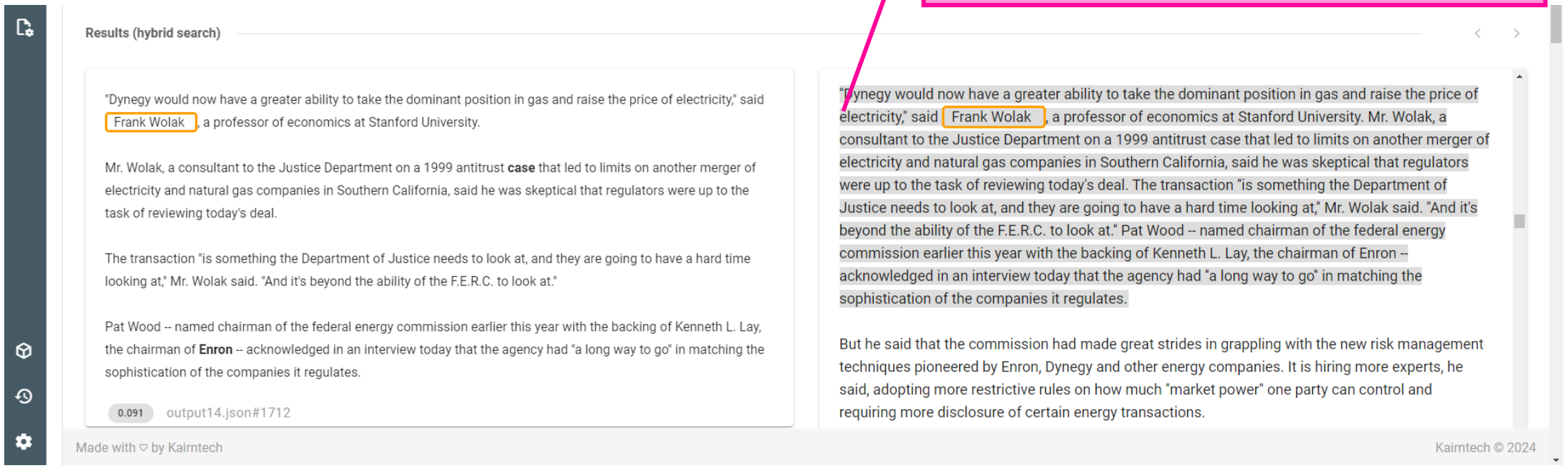
Message-ID: <3539387.1075861669799.JavaMail.evans@thyme>
Date: Mon, 26 Nov 2001 07:37:31 -0800 (PST)
From: m..schmidt@enron.com
Subject: Enron Mentions - 11/23/01
Mime-Version: 1.0
Content-Type: text/plain; charset=ANSI_X3.4-1968
Content-Transfer-Encoding: quoted-printable
X-From: Schmidt, Ann M. </O=ENRON/OU=NA/CN=RECIPIENTS/
CN=ASCHMID>
X-To:
X-cc:
X-bcc:
X-Folder: \JARNOLD (Non-Privileged)\Arnold, John\Deleted Items
X-Origin: Arnold-J
X-FileName: JARNOLD (Non-Privileged).pst

Premiums Stay High on Enron's Near Options, And `Doubling Up' Date Looms
fo=

Demo 1 - Asking questions

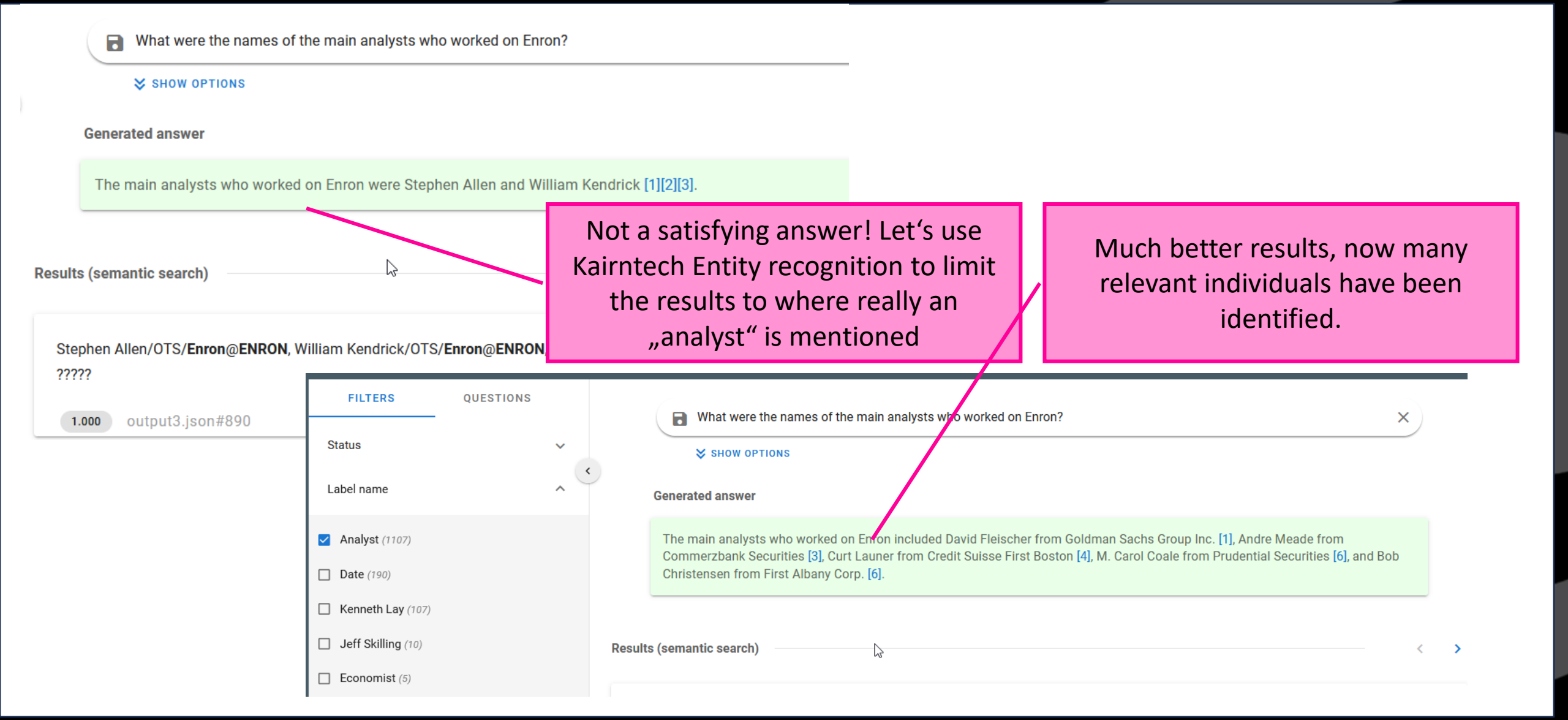
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Bonus for large documents: the relevant chunk is automatically displayed in the source document.
It's easy to check the context in the paragraphs before or after the chunk.



Demo 1 – asking questions

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



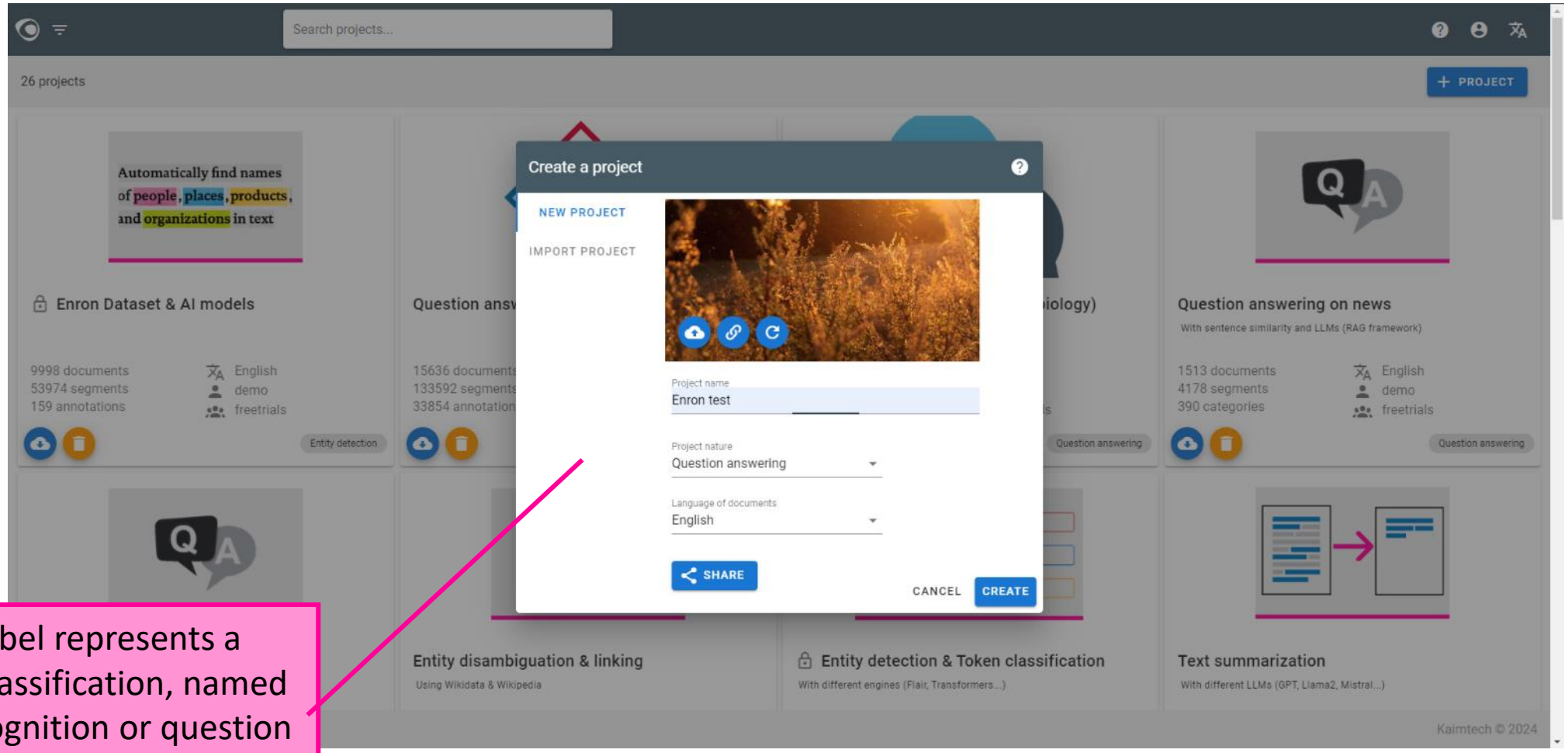
Demo 2 - Behind the scenes

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

- ✓ **Fast to value**
  - Create RAG from scratch
- ✓ **Customization examples**
  - Adding LLMs
  - Adding text annotations
  - Adding metadata to semantic search
  - Adding metadata to LLM

Demo 2 - Behind the scenes

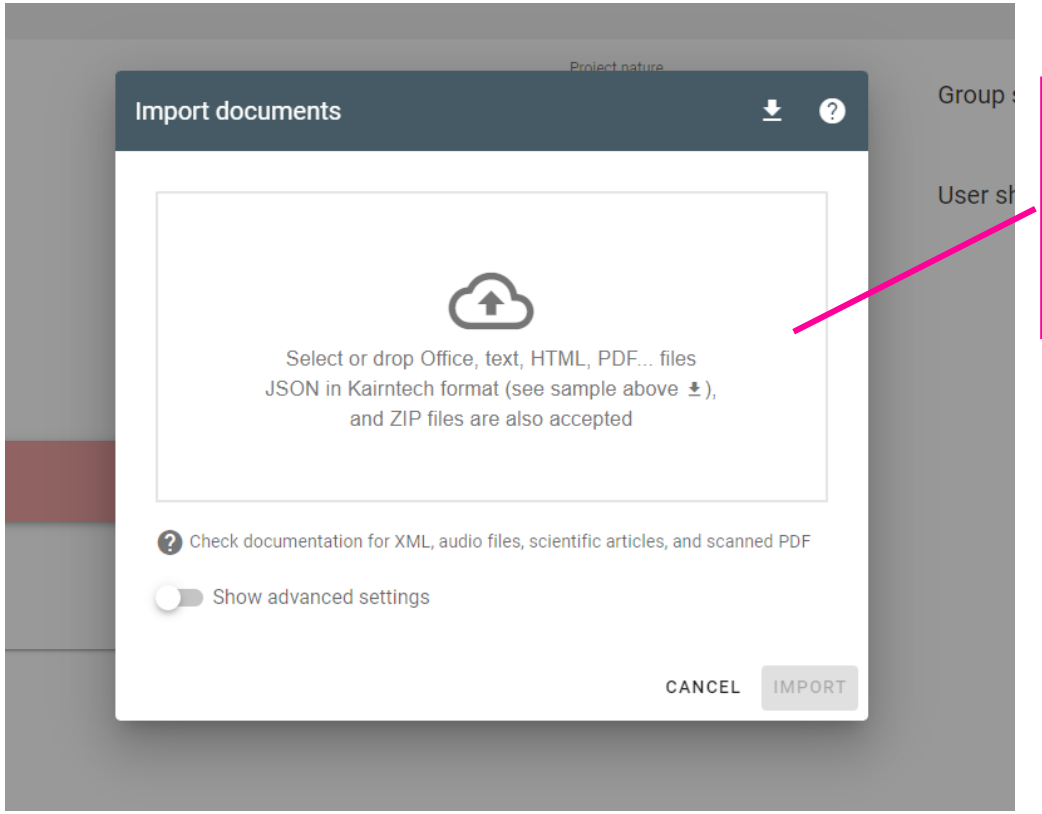
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



Each label represents a project: classification, named entity recognition or question answering

Demo 2 - Behind the scenes

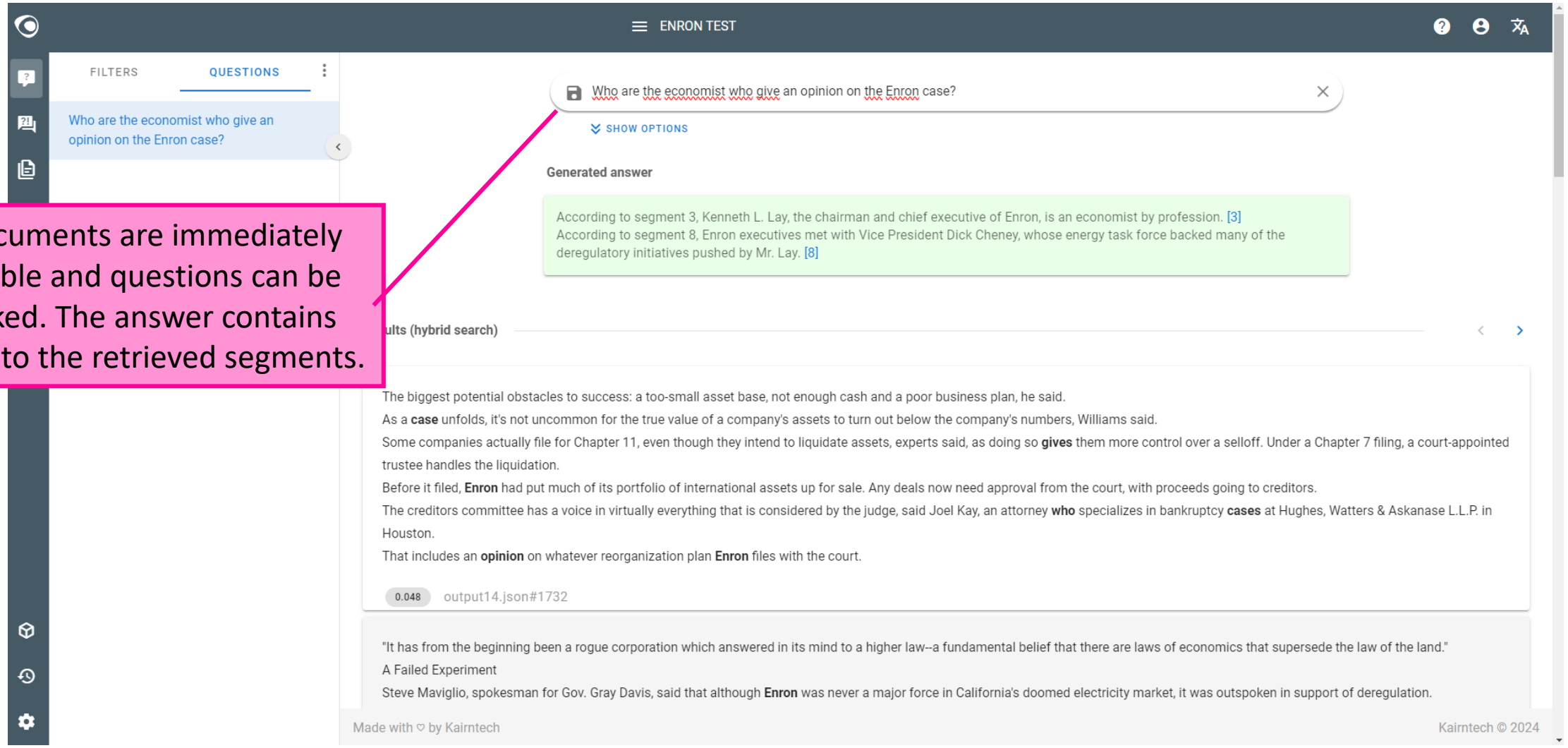
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



Document are automatically uploaded, segmented, indexed and vectorized.

Demo 2 - Behind the scenes

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

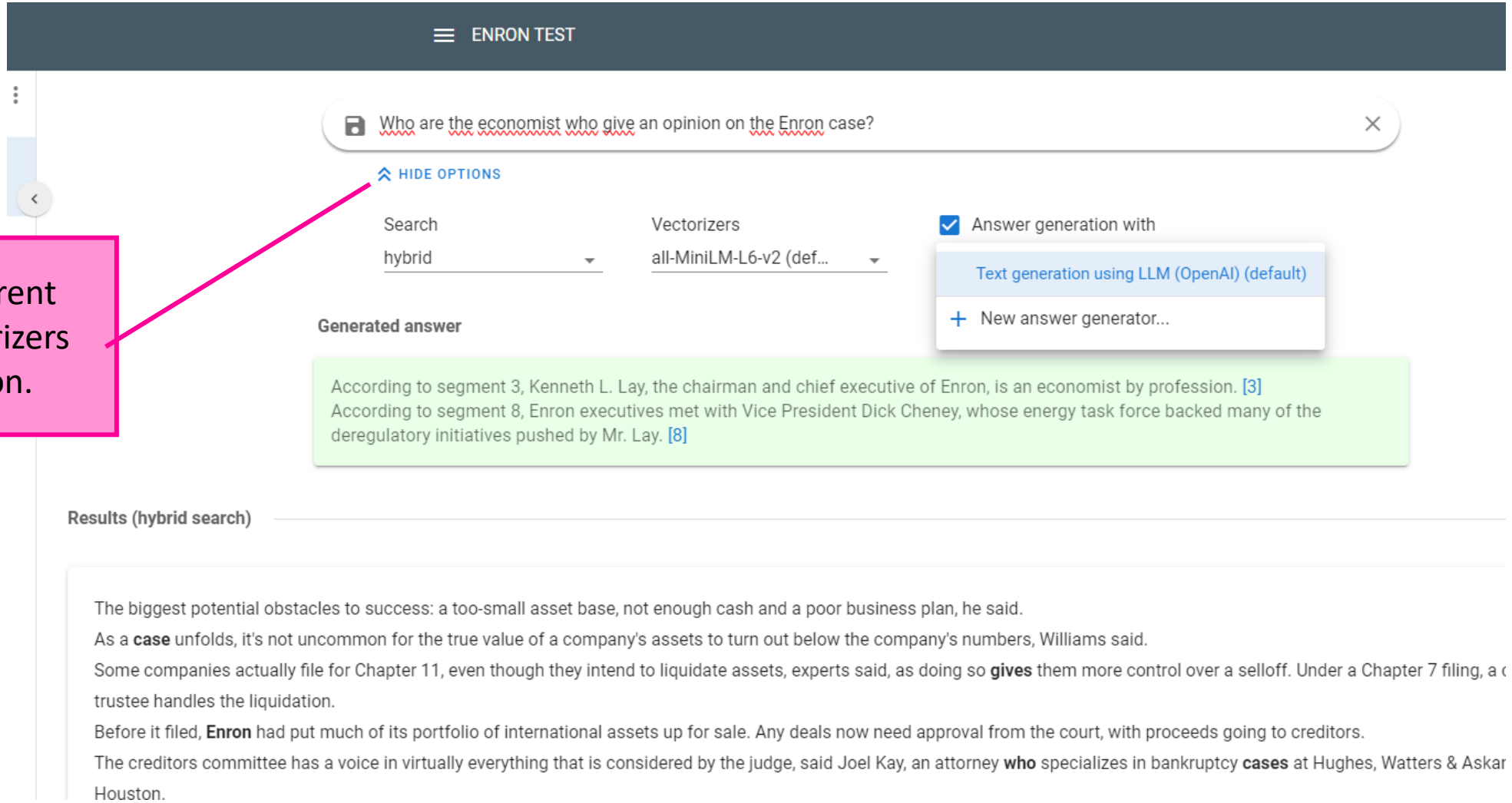


Documents are immediately visible and questions can be asked. The answer contains links to the retrieved segments.

Demo 2 - Behind the scenes

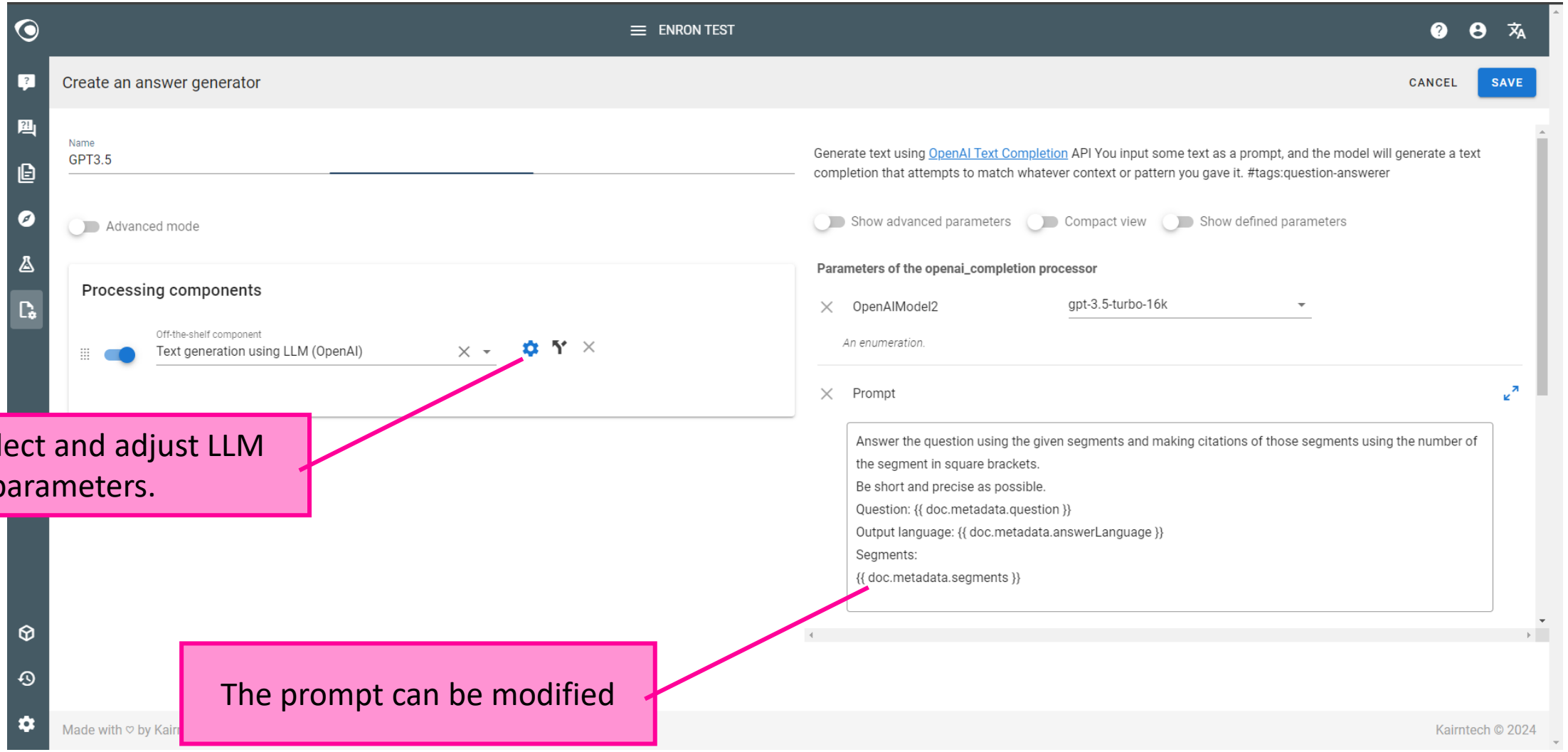
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Options to select different search methods, vectorizers and LLM configuration.



Demo 2 - Behind the scenes

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



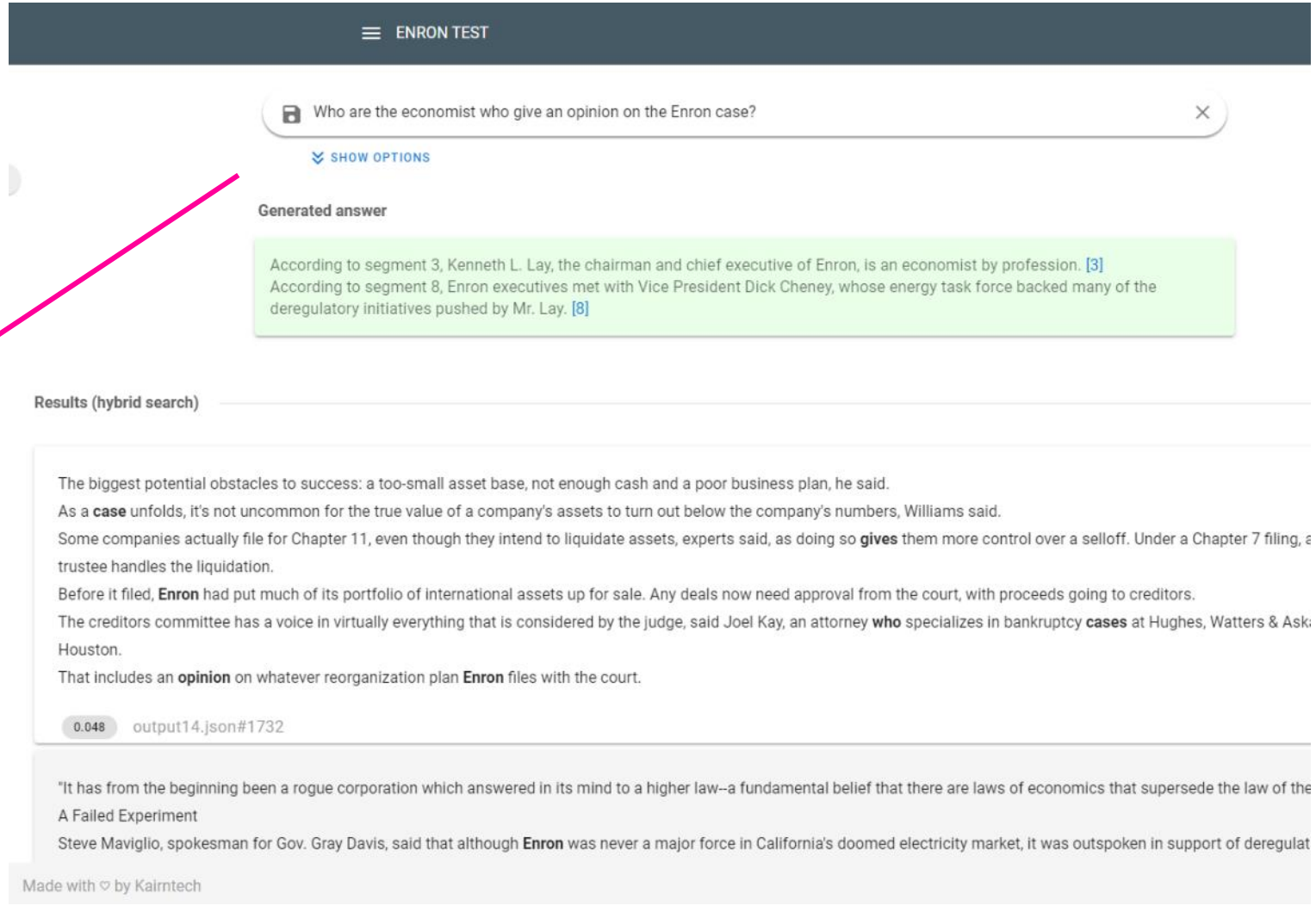
Easily select and adjust LLM parameters.

The prompt can be modified

Demo 2 - Behind the scenes

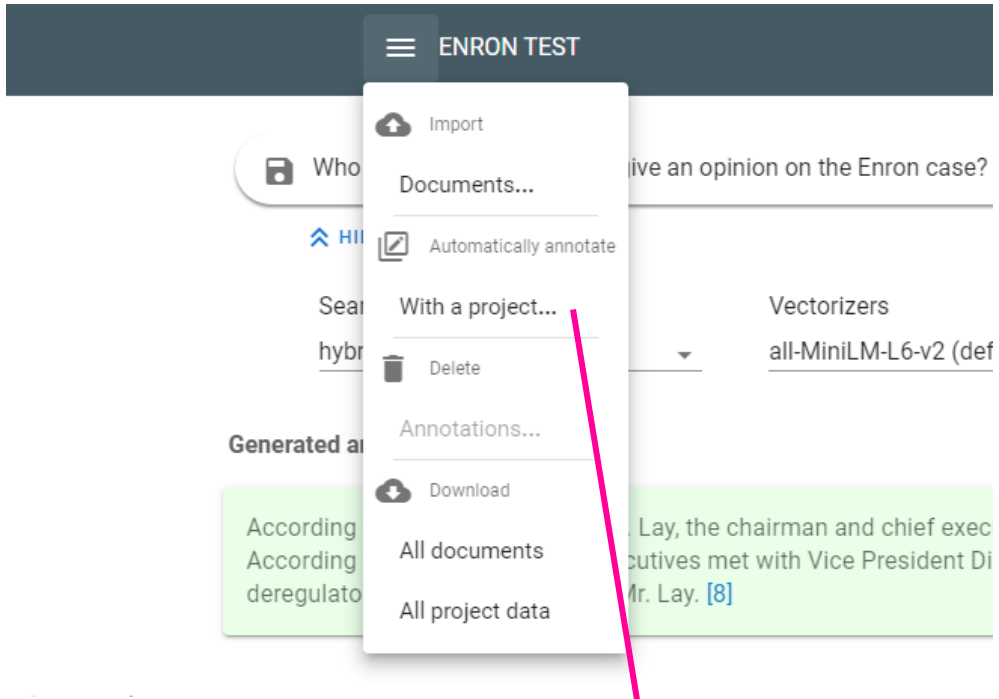
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

For this question, the answer is clearly wrong. The LLM is not able to detect economists.

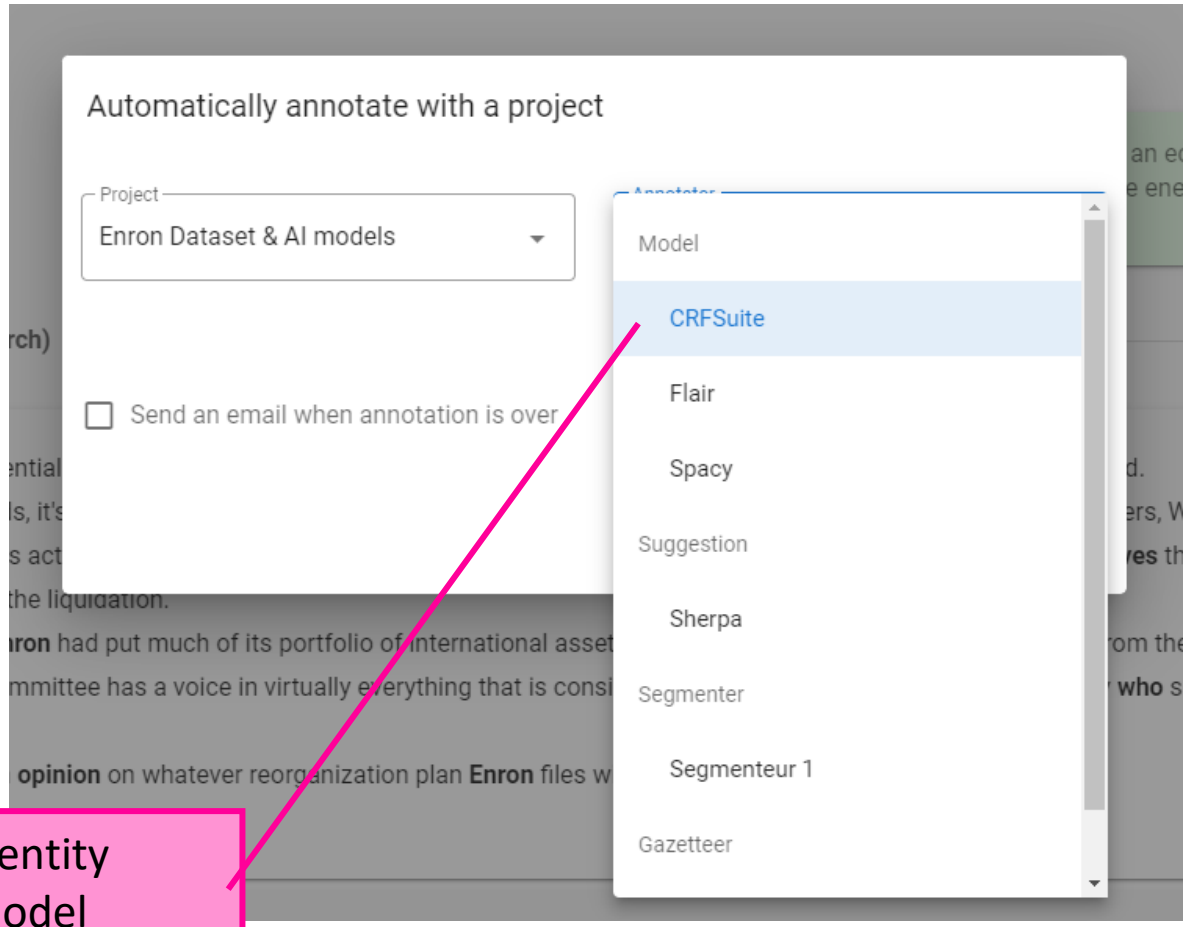


Demo 2 - Behind the scenes

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



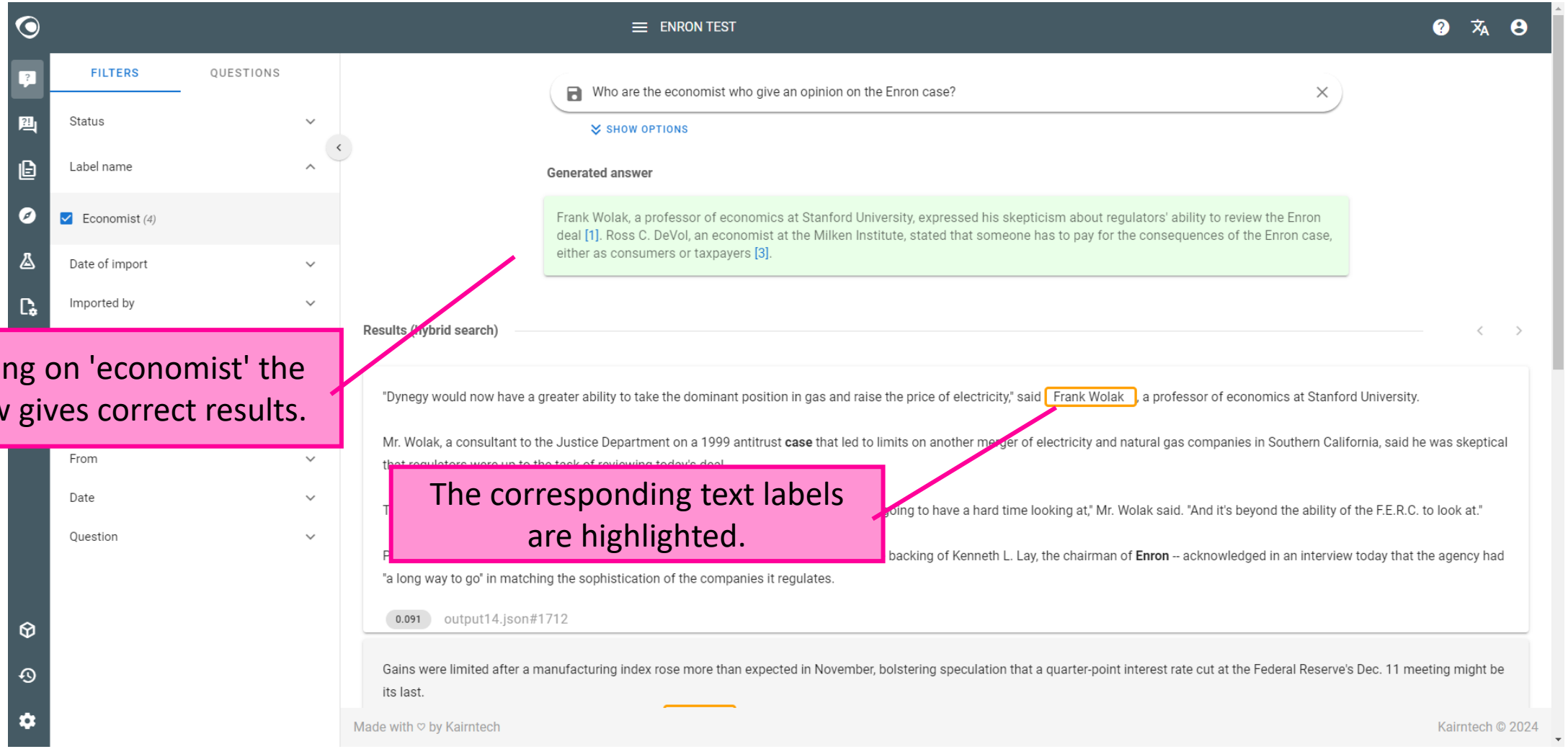
The solution: annotate the documents



With a named entity recognition model

Demo 2 - Behind the scenes

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



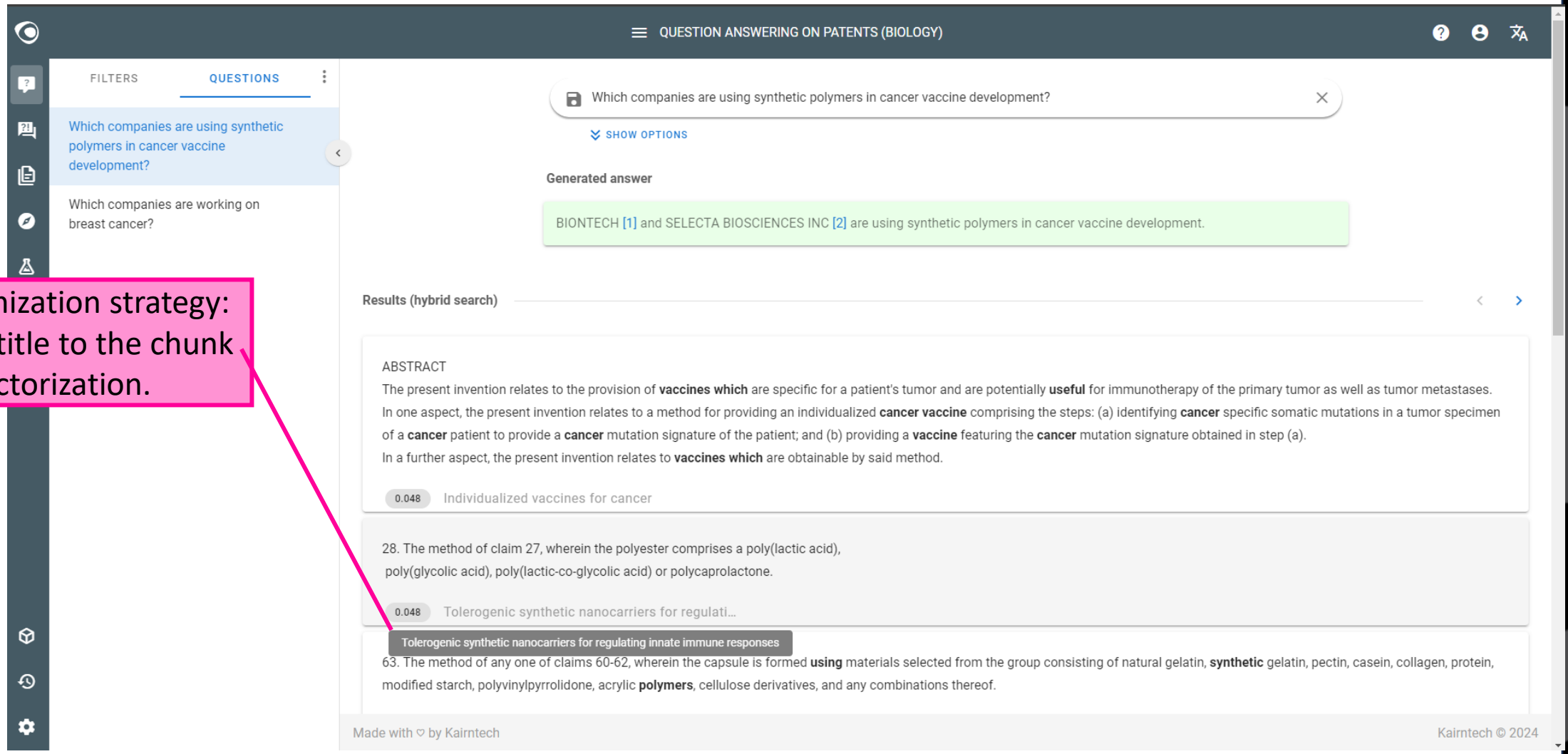
By filtering on 'economist' the LLM now gives correct results.

The corresponding text labels are highlighted.

Demo 2 - Behind the scene

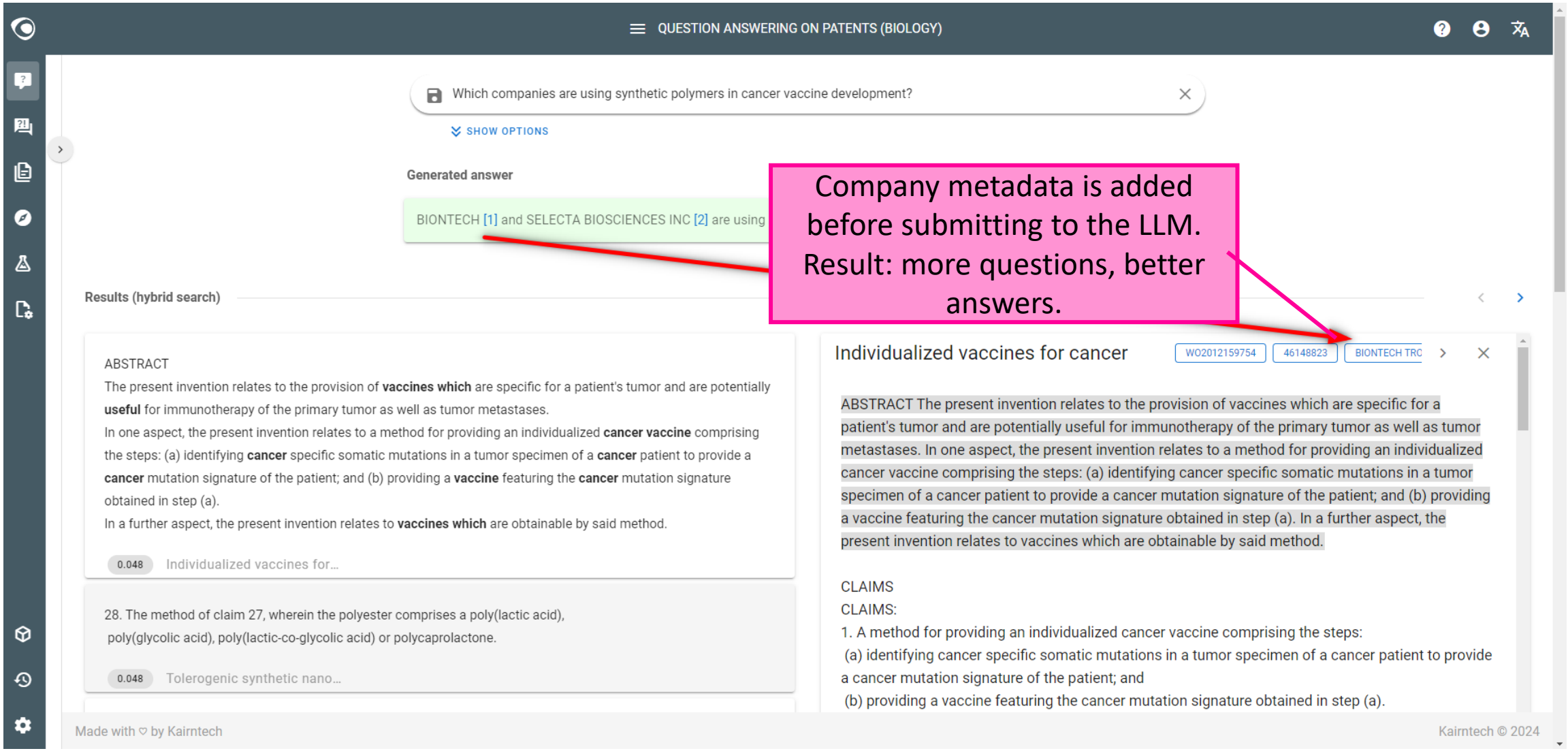
KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Another customization strategy: add the patent title to the chunk before vectorization.



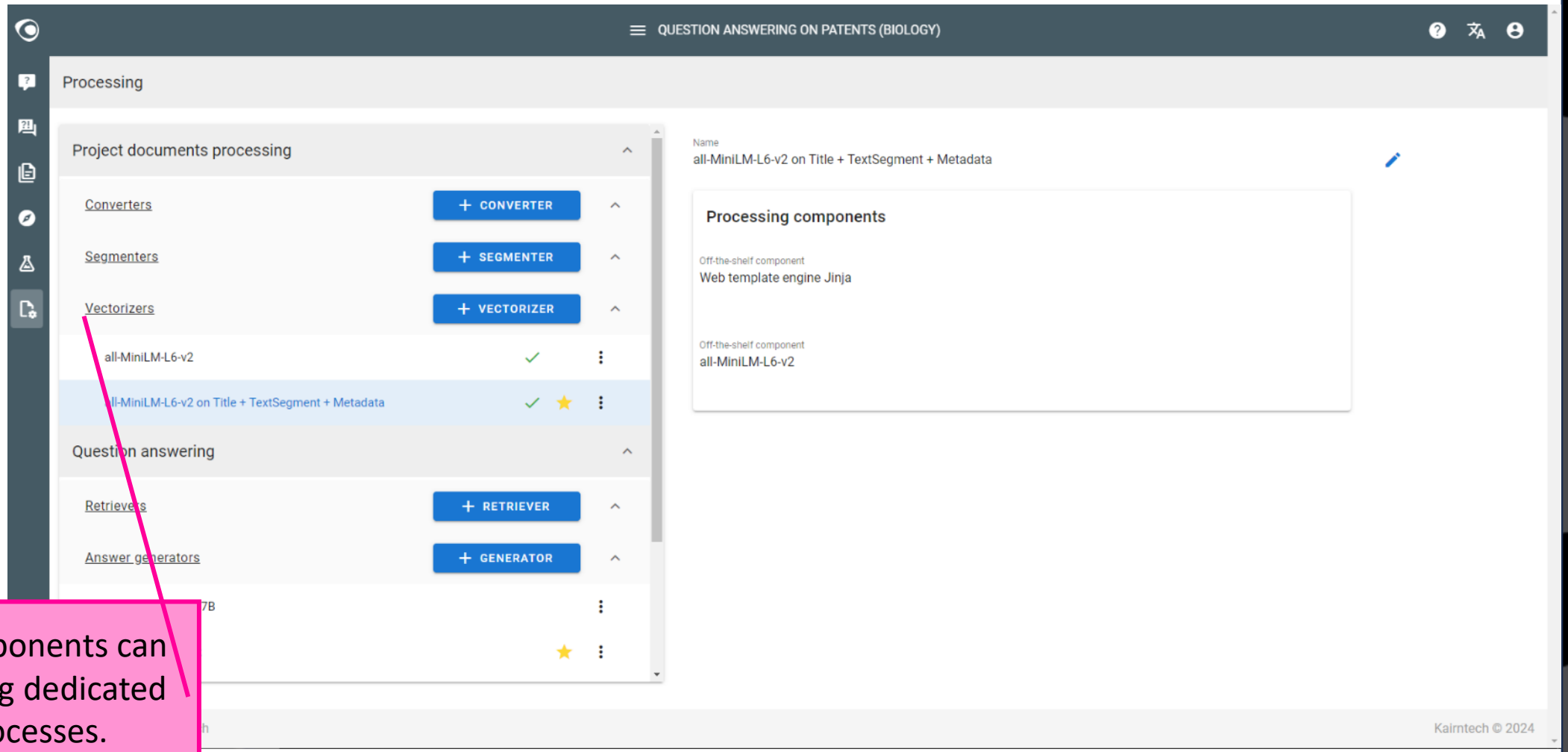
Demo 2 - Behind the scene

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



Demo 2 - Behind the scenes

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



All processing components can be customized using dedicated pipelines of processes.

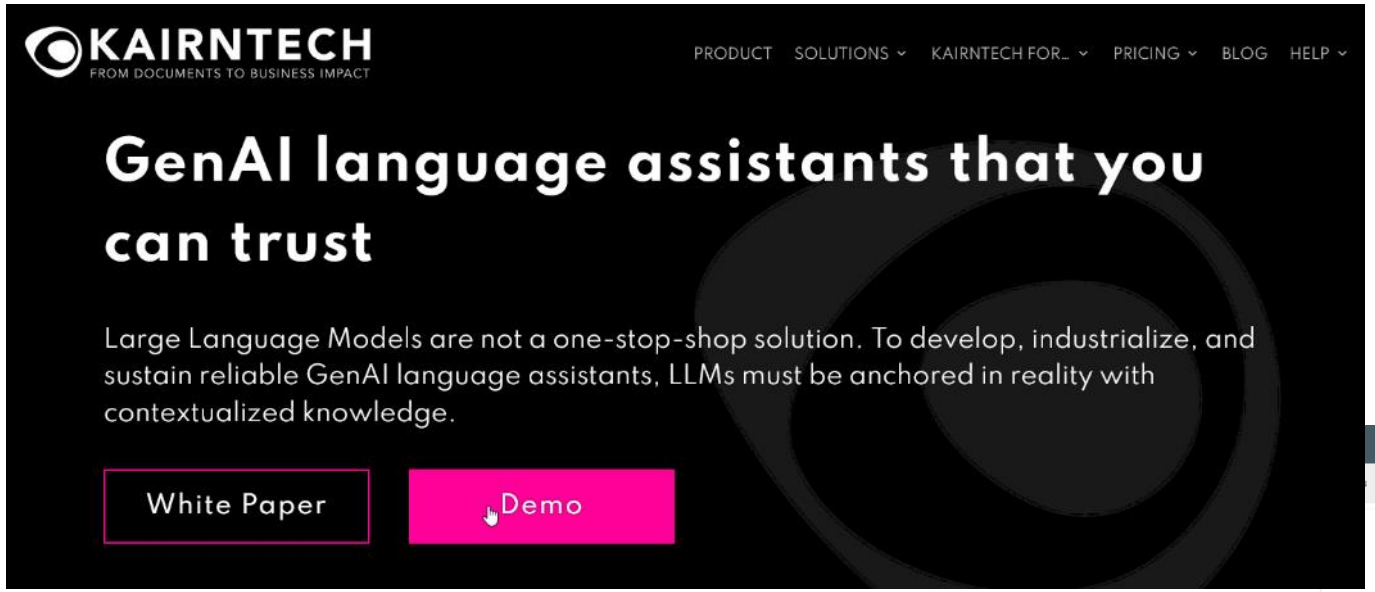
Conclusion

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

- ✓ LLMs have fantastic capabilities
- ✓ But are only part of the answer to create business value
- ✓ Customization is key
- ✓ Defining these projects using Kairntech is fast & easy
- ✓ Enterprise solution: Scalable, reliable, versatile
- ✓ Kairntech is your partner on that journey!

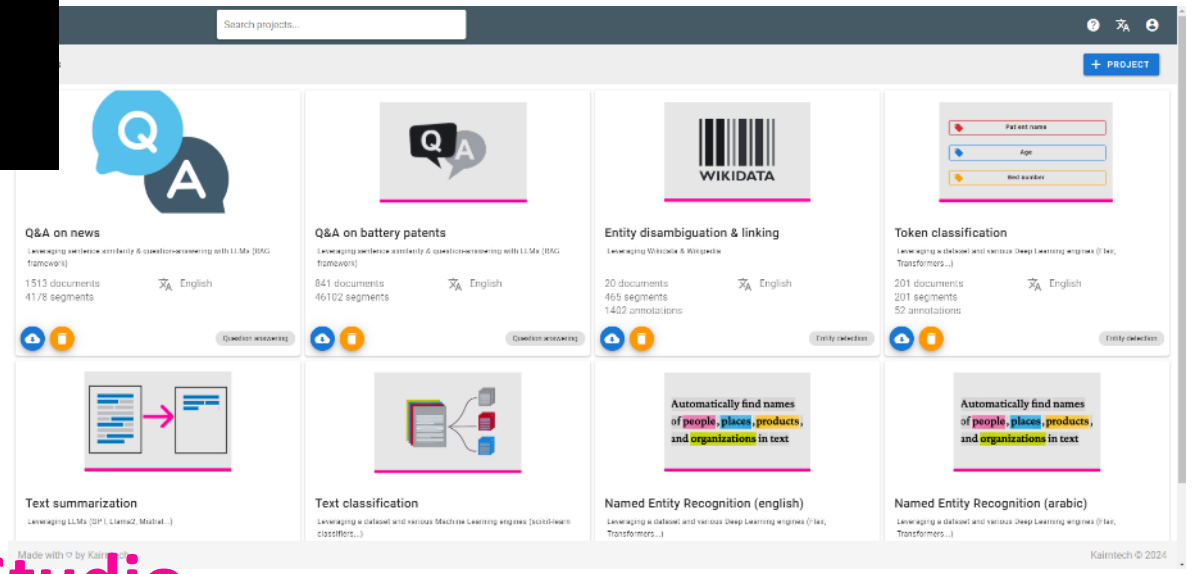
We create GenAI language services

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT



1 **www.kairntech.com**

Free access to a selection of demo use cases



2 **Kairntech Studio**

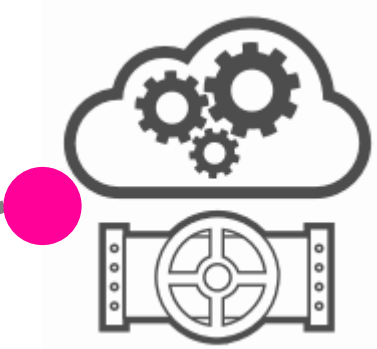
Entity Extraction, Categorization, RAG, …

Deliver trusted business value

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Customize question-answering

Experiment with chunking, vectorization, search methods and LLMs. Benchmark quality. Leverage or create metadata (categories, entities, Wikidata, business vocabularies..). Filtering / question context. Finetune embeddings model and LLM.



Anchor GenAI in reality

Leverage internal data sources (media, documents, knowledge graphs…) and convert to text.

Deploy securely

Embed in existing applications with a rich REST API. Deploy at scale, on premise. With continuous improvement (human in the loop).

KAIRNTECH
FROM DOCUMENTS TO BUSINESS IMPACT

Thank you for your attention

Want to read more? RAG Whitepaper

Documentation? documentation

We’re happy to hear from you: info@kairntech.com