

 Zeta Alpha

AI TO **DISCOVER** AND **ORGANIZE** KNOWLEDGE FOR HIGH TECH COMPANIES



Jakub Zavrel
Founder / CEO
Zeta Alpha

June 2024 | Science Park Amsterdam

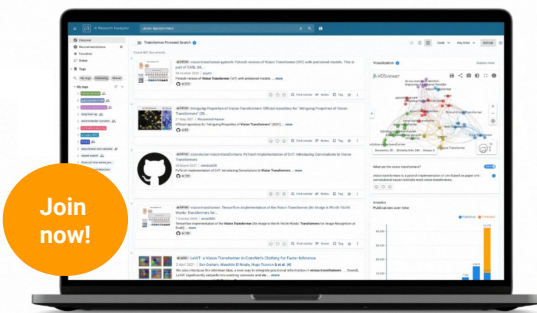
ζα Zeta Alpha

Connecting People and Knowledge, since 2019.
HQ in Amsterdam, office in SF since 2023.

Allows **High Tech Companies** to turn proprietary expertise into **Next Gen AI capabilities**.

A smarter way to discover and organize knowledge













Zeta Alpha is the best neural discovery platform for public and private documents. We streamline how you and your team discover, organize and share knowledge – in AI and beyond.



Join now!

ETH zürich D · BASF DUCKDUCKGOOSE bol.com NASA Imperial College London

Berkeley VU UNIVERSITY AMSTERDAM O₂ Network Hemisphere gretel

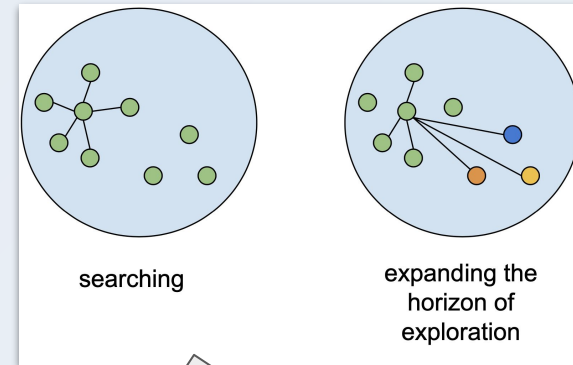
 <p>Jakob Zavel Founder & CEO</p> <p>Founder, Technopoli, Entrepreneur, PhD in Law with a focus on AI and data. Founder and former CEO of Technopoli. Building a new kind of online community, and great AI products, with a very ambitious goal in mind: to have the best in the world in AI and machine learning. You can find him listening to his music, watching his football, or writing about AI and machine learning.</p>	 <p>Arjen de Hoop Business Development</p> <p>Arjen is passionate about connecting technology to clients in strategic projects, which he has done before in the sales energy industry. He was part of the team that set up and ran a sales office in the Netherlands. Arjen is a graduate from Delft University of Technology, Delft, and has a focus on sales and business development.</p>	 <p>Fernando Rojas Barrera CTO</p> <p>Fernando spent some time in academia, after completing his PhD in electrical physics and afterwards, made a switch to industry in order to follow his passion for technology. He has focused on software architecture of distributed systems, data engineering, AI/ML, and cloud infrastructure.</p>
 <p>Arthur Barbosa Cansara Senior Research Engineer</p> <p>With a Computer Science degree and strong engineering experience from his previous work at his previous company, Arthur is now working at Zeta Alpha as a Senior Research Engineer. He is interested in using ML, NLP and Statistics to build AI models for video games, robotics and health care. He is also interested in using ML to build AI models for video games, robotics and health care.</p>	 <p>Mathias Pericot Research Engineer</p> <p>Mathias did his Master's thesis in machine learning, his interest in AI research projects. He is currently the CTO for the firm and he is very interested in using ML to build AI models for video games, robotics and health care.</p>	 <p>Joanna Król Frontend Developer</p> <p>With a background in computer science, Joanna is currently working as a Frontend Developer at Zeta Alpha. She is interested in using ML to build AI models for video games, robotics and health care.</p>
 <p>Eduardo Ochoa Senior Data Engineer</p> <p>Eduardo is a former research scientist at Google, where he worked on the Google Cloud Platform. He is currently working at Zeta Alpha as a Senior Data Engineer. He is interested in using ML to build AI models for video games, robotics and health care.</p>	 <p>Dion Papakostas Research Engineer</p> <p>With a background in computer science, Dion is currently working as a Research Engineer at Zeta Alpha. He is interested in using ML to build AI models for video games, robotics and health care.</p>	 <p>Pavlos Zakas AI Engineer</p> <p>A graduate of Electrical and Computer Engineering at the MIT, Pavlos is currently working as an AI Engineer at Zeta Alpha. He is interested in using ML to build AI models for video games, robotics and health care.</p>
 <p>Rodrigo Nogueira Scientific Advisor</p> <p>Rodrigo aims to use machine learning to build scientific insights. He holds a PhD from MIT, where he worked on the development of Deep Learning NLP and AI models for text classification. He is currently working at Zeta Alpha as a Scientific Advisor. He is interested in using ML to build AI models for video games, robotics and health care.</p>	 <p>Artem Kozlov Frontend Engineer</p> <p>Product-oriented frontend engineer with a strong understanding of the history of frontend and backend development. Effective collaborator of various business units. His interests, requirements and experiments which have led to the creation of the Zeta Alpha platform and its evolution.</p>	 <p>Jakob Kaiser Backend Engineer</p> <p>With a background in computer science, Jakob is currently working as a Backend Engineer at Zeta Alpha. He is interested in using ML to build AI models for video games, robotics and health care.</p>

PROBLEM: KNOWLEDGE REUSE IN THE ENTERPRISE IS RIDICULOUSLY INEFFICIENT AND EXPENSIVE.

1. DATA FRAGMENTATION: INTERNAL AND EXTERNAL



2. CLASSIC KEYWORD SEARCH



SOLUTION:
ONE KNOWLEDGE CENTER: NEURAL SEMANTIC SEARCH IN ZETA ALPHA PLATFORM

ASK QUESTIONS, GET ANSWERS FROM EXISTING INTERNAL DOCUMENTS AND CONNECT TO THE RIGHT EXPERTS

The screenshot displays the Zeta Alpha search interface. At the top, a search bar contains the query "how can I best reduce our cost for AWS?". Below the search bar, there are tabs for "Internal Documents" and "All the Web". The search results are displayed in a list format, with the first result being a document titled "how can I best reduce our cost for AWS?". The document text discusses optimizing AWS resource utilization, mentioning tools like AWS Cost Explorer and AWS Budgets, and techniques like auto-scaling and reserved instances. To the right of the search results, there is a "Visualization" section showing a network graph with nodes for "Engineering", "EC2 Setup", "Cloud Computing", and "Engineering Roadmap". Below the network graph, there is a section titled "People related to 'how can AI be used to design more sustainable buildings?'" which lists four individuals: Stanislas Chaillou, Shuvashish Chatterjee, Praveen Joshi, and Silverio Martínez-Fernández.

Discover

- Recommendations
- People
- Favorites
- My documents
- Notes
- Tags
- My tags
 - twitter crawler
 - GPT integration
 - aws cost reduction
 - chat interface
 - knowledge assistant ...
 - to read
 - Expo popup dispay b...
 - retrieval augmented g...

how can I best reduce our cost for AWS?

Internal Documents | All the Web

Save | Any time | Sources | Owner | Type | Neural Search | Relevance

4,466 results

how can I best reduce our cost for AWS?

To best reduce your cost for AWS, you can consider optimizing your resource utilization, using cost-effective instance types, implementing auto-scaling, and leveraging AWS cost management tools like AWS Cost Explorer and AWS Budgets^[1]. Additionally, you can explore reserved instances, spot instances, and storage optimization techniques to further optimize costs^[1]. It would also be helpful to allocate time for finishing the tagging of the Kubernetes resources in AWS to better understand the cost breakdown^[5].

Visualization | Go back | Explore more

Engineering

EC2 Setup | Cloud Computing

Engineering Roadmap

People related to "how can AI be used to design more sustainable buildings?"

- SC Stanislas Chaillou
- SC Shuvashish Chatterjee
- PJ Praveen Joshi
- SM Silverio Martínez-Fernández

Find similar | Notes 3 | Tag | Share

CHAT WITH YOUR OWN DOCUMENTS

The image shows a document viewer interface with a dark toolbar at the top. The document content is in Greek and includes the Bank of Greece logo and title. A chat window on the right contains a user query and an AI response with a bulleted list of points.

ΤΡΑΠΕΖΑ ΤΗΣ ΕΛΛΑΔΟΣ
ΕΥΡΩΣΥΣΤΗΜΑ

ΣΥΜΒΟΥΛΙΟ ΝΟΜΙΣΜΑΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ

ΠΡΑΞΗ ΣΥΜΒΟΥΛΙΟΥ ΝΟΜΙΣΜΑΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ ΑΡΙΘ. 124/21.03.2023

ΘΕΜΑ: Τροποποίηση της Πράξης ΣΝΠ 87/28.2.2013 σχετικά με συμπληρωματικά προσωρινά μέτρα όσον αφορά τις πράξεις αναχρηματοδότησης του Ευρωσυστήματος και την καταλληλότητα των ασφαλειών

Το ΣΥΜΒΟΥΛΙΟ ΝΟΜΙΣΜΑΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ, αφού έλαβε υπόψη:

α) τα Άρθρα 2, 35α, και 55 του Καταστατικού της Τράπεζας της Ελλάδος, ως κυρωθέν ισχύει,

β) την Πράξη ΣΝΠ 87/28.2.2013 «Αντικατάσταση της Πράξης ΣΝΠ 82/27.8.2012 σχετικά με συμπληρωματικά προσωρινά μέτρα όσον αφορά τις πράξεις κύριας αναχρηματοδότησης του Ευρωσυστήματος και την καταλληλότητα των ασφαλειών» (ΦΕΚ Β' 482/1.3.2013), όπως τροποποιήθηκε με την Πράξη ΣΝΠ 88/26.4.2013 (ΦΕΚ Β' 1069/29.4.2013), την Πράξη ΣΝΠ 89/20.12.2013 (ΦΕΚ Β' 3433/31.12.2013), την Πράξη ΣΝΠ 90/27.3.2014 (ΦΕΚ Β' 848/4.4.2014), την Πράξη ΣΝΠ 93/5.8.2014 (ΦΕΚ Β' 2246/18.8.2014), την Πράξη ΣΝΠ 94/20.10.2014 (ΦΕΚ Β' 2874/24.10.2014) την Πράξη ΣΝΠ 95/28.11.2014 (ΦΕΚ Β' 2874/24.10.2014)

What are the main points of this regulatory policy? Use markdown.

- The regulatory policy modifies the Council Act No. 87/28.2.2013 regarding supplementary temporary measures on refinancing operations of the Eurosystem and the adequacy of collateral [1].
- It allows the Bank of Greece to accept individual loan claims against businesses, subject to specific criteria [1].
- Loan claims must adhere to Greek law and meet creditworthiness criteria [1].
- Loan claims not serviced or involving delinquent debtors are not accepted [1].
- The policy sets minimum thresholds for loan claim values and applies specific haircut percentages based on

Ask about document...

TOWARDS KNOWLEDGE ASSISTANTS – AGENT AND TOOLS

The screenshot displays the Zeta Alpha knowledge assistant interface. On the left is a navigation sidebar with sections: Discover, Recommendations, People, Favorites, My documents, Notes, Tags, My tags (with sub-sections: My tags, Following, Shared), Settings, Management, and Help. The main content area shows search results for 'ai job matching bias...'. The top result is 'Becoming Human: Artificial Intelligence Magazine | AI in Recruitment: How is it revolut the hiring process?' by Ashok Sharma, dated 30 Dec 2022. The second result is 'TDS Tutorial : breaking myths about AI fairness. The case of biased autor recruitment.' by Grégoire Martinon, dated 11 May 2021. The third result is 'arXiv | Fairness and Bias in Algorithmic Hiring' by Alessandro Fabris, Nina Baranowska, Matthew J. Dennis & et al. (4), dated 25 Sep 2023. On the right, a chat window titled 'using AI for recruitment?' is open, showing a conversation with an AI agent. The agent's message reads: 'I'm sorry, but I don't have enough information to answer your question. The European AI Act is a regulatory framework proposal that aims to address various aspects of AI, including risk assessment and bias mitigation [1]. However, the specific requirements related to using AI for recruitment are not mentioned in the given documents. It would be best to refer to the official documentation or consult legal experts for detailed information on the requirements outlined in the European AI Act regarding AI use in recruitment.' Below this, a user asks: 'can you find documents about the European AI act and its effect on recruitment AI?'. The agent responds: 'See suggested documents'. A second user question asks: 'Can you check the new documents and what they state in terms of requirements for AI in recruitment from the perspective of the EU AI Act?'. At the bottom of the chat window is an input field with the placeholder text 'Ask about document...' and a send button.

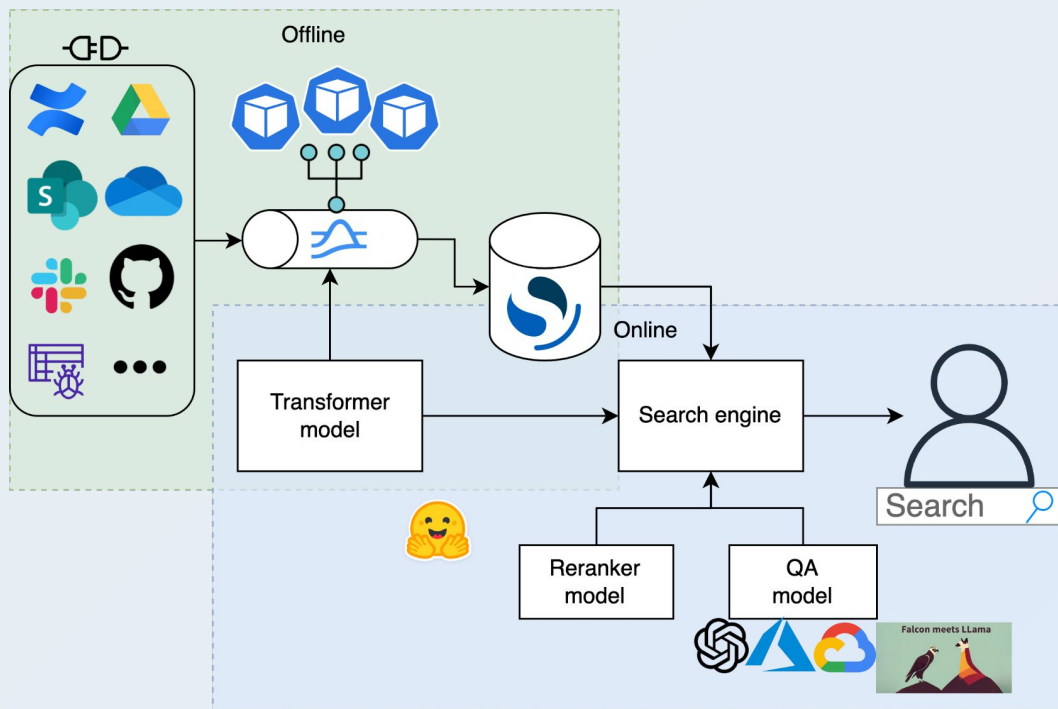
- Guided exploration during onboarding to new topics or projects.
- Lookups in other internal IT systems.
- Conversation refines queries like brainstorming with an expert.

- **Ask questions** about figures, schemas, and tables.
- **Use as OCR** to extract information from pictures.
- **Augment documents** with new meta data.
- **The start of the multi-modal search journey.**

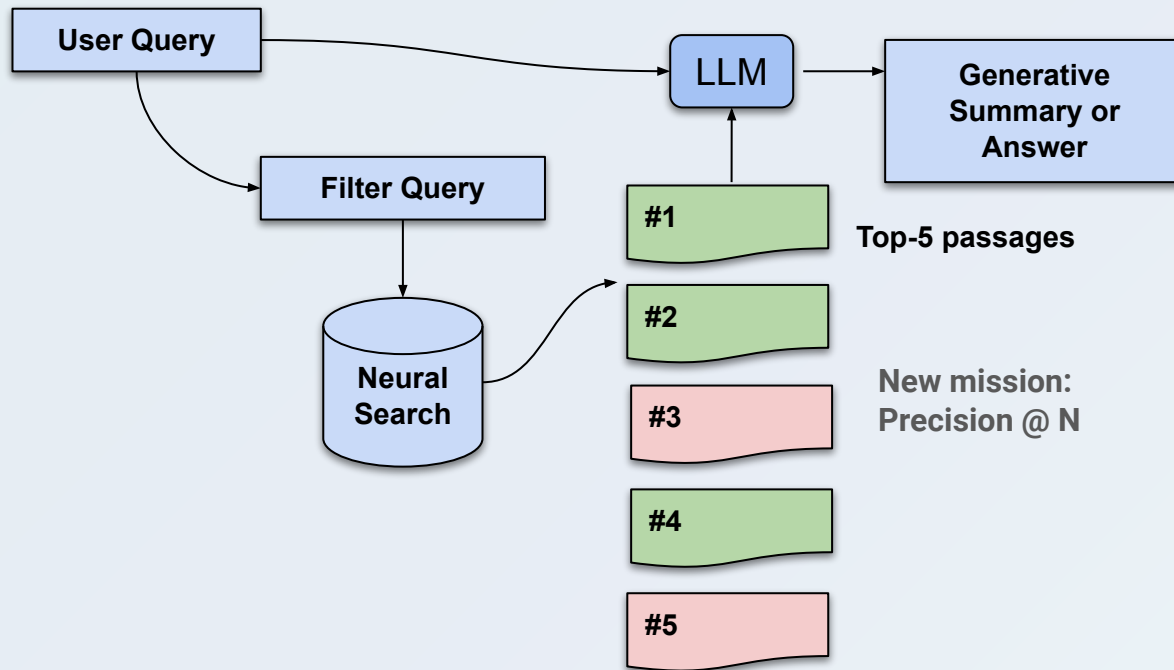
The screenshot shows a Zeta Alpha interface with a document viewer. The document is titled "EMO-Emote Portrait Alive" and contains a diagram of a neural network architecture. A pop-up window on the right provides a detailed explanation of the diagram's first stage, "Frames Encoding".

Fig. 2: Overview of the proposed method. Our framework is mainly constituted with two stages. In the initial stage, termed Frames Encoding, the ReferenceNet is deployed to extract features from the reference image and motion frames. Subsequently, during the Diffusion Process stage, a pretrained audio encoder processes the audio embedding. The facial region mask is integrated with multi-frame noise to govern the generation of facial imagery. This is followed by the employment of the Backbone Network to facilitate the denoising operation. Within the Backbone Network, two forms of attention mechanisms are applied: Reference-Attention and Audio-Attention. These mechanisms are essential for preserving the character's identity and modulating the character's movements, respectively. Additionally, Temporal Modules are utilized to manipulate the temporal dimension, and adjust the velocity of motion.

1. Frames Encoding: A reference image and motion frames are encoded using a Variational Autoencoder (VAE) to extract feature maps. These feature maps are processed by the ReferenceNet to capture the identity and motion



- **Document ingestion pipeline** feeds the Index from data connectors.
- **Access rights** enforced from ingestion to search
- **Search API** provides query processing and retrieval of relevant documents, passages, concept analytics, expert search
- **Hybrid search:** hard filters and keywords, as well as k-NN vector search.
- **Answers API** provides Search & GPT integration for question answering via RAG.



When I have question, like:

What is the best way to train a BERT model at a low cost?

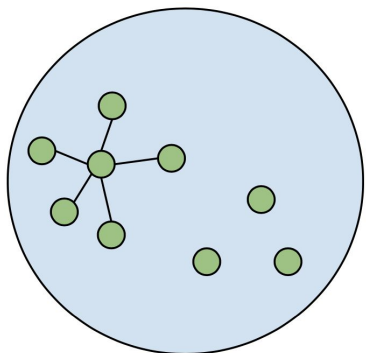
Not

What is the best way to train a low cost RL model?

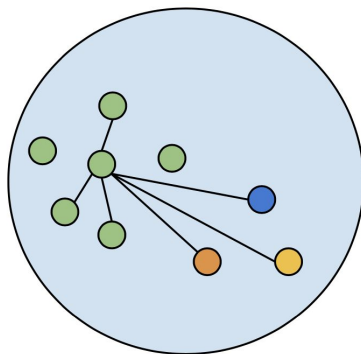
But

How to Train BERT with an Academic Budget

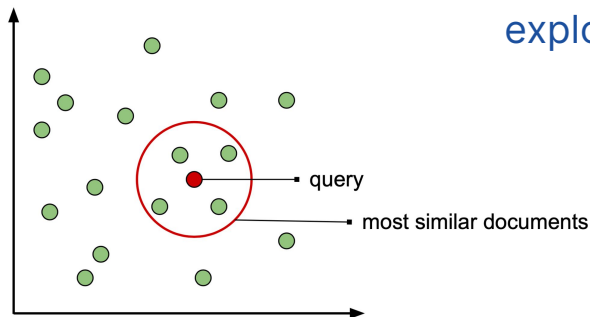
THE R IN RAG: WHY **NEURAL SEARCH**?



searching

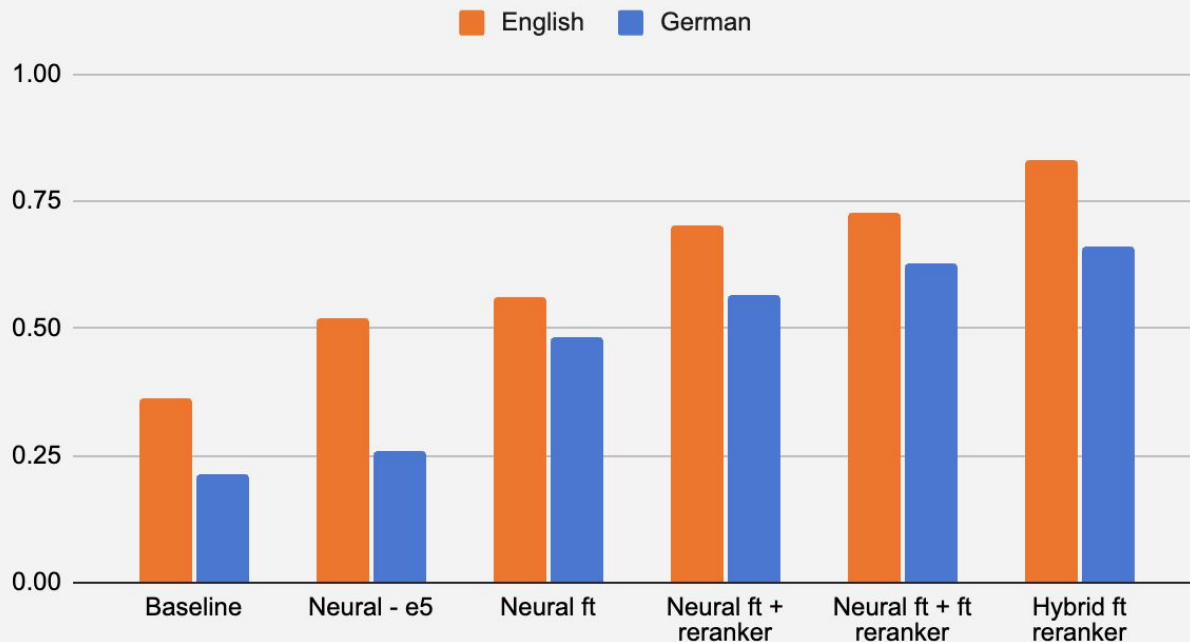


expanding the
horizon of
exploration



- Neural Search is Semantic Search. Similarity vs, surface keywords matching: bridge the **lexical gap**
- Context and relationships crucial in interpreting meaning: handles complex and **relational questions**
- Unstructured data accessible without classification and taxonomies, even **multi-lingual** and **cross-lingual**
- **Multi-modal** capabilities: potential to combine text, audio, images and video

Precision@10 - Chemistry Domain



Large Language Models as Efficient Rerankers for Information Retrieval

Luiz Bonifacio
NeuralMind, Brazil
UNICAMP, Brazil

Hugo Abonizio
NeuralMind, Brazil
FEEC-UNICAMP, Brazil

Roberto Lotufo
NeuralMind, Brazil
UNICAMP, Brazil

Jakub Zavrel
Zeta Alpha, Netherlands

Andrigo Nogueira
NeuralMind, Brazil
UNICAMP, Brazil
Zeta Alpha, Netherlands

Abstract

A method to efficiently use large language models for information retrieval. It asks: via few-shot examples, an LLM is induced to rerank documents. These synthetic query-document pairs are used as input to a neural retriever. However, InPars and, more recently, by LLMs such as GPT-3 and FLAN to generate synthetic queries. We introduce InPars-v2, a dataset generator that introduces existing powerful rerankers to select synthetic

WHY IS ZETA ALPHA UNIQUE?

- **Proven expertise, consulting in domain specific Neural Search.**
 - Finetuned Neural Embedding models have >200% better search relevance.
- **For Knowledge Discovery in R&D and Tech Teams** - Zeta Alpha has an unrivaled breadth of **discovery** and **organization** tools beyond just Search and Gen AI.
 - Tags, Notes, PDF Chat, Knowledge Assistant, Recommendations, Sharing, Visualization, Analytics, Expert Search and much, much more.
- **For RAG use cases** - Zeta Alpha offers a **secure, mature** and **customizable Enterprise IT platform** for building bespoke LLM powered chat solutions.
 - SSO, access control, on-prem, connectors, configurable Hybrid and Federated search.

Bridging the advantages of DIY solutions like Langchain (flexibility, customization, *but hard to move from prototype to production*) and one-size-fits-all solutions like MS Copilot (stable, Enterprise IT ready, *but not suitable for domain specific and customized solutions*).

WHY? – ZETA ALPHA & YOU

- Make **your teams** more productive and reuse **internal and external knowhow and expertise in a central knowledge hub** across languages and teams.
- Unlock the knowledge inside **internal technical and project documentation** across **your company**, for:
 - Management & Strategic Analysis
 - Expertise Search & Project Staffing
 - Onboarding and Knowledge Transfer
 - Sales engineers & customer facing consultants
- Streamline **external access by customers and partners** to **your** knowledge and documentation using chatbots in product and learning portals.

We are looking forward to work with you to discover how AI can impact ROI on knowledge for you.