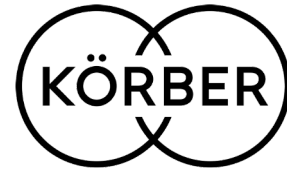




AI MONDAY®

#AIMondaySessions



Building use cases in unparalleled speed with a foundation model like Luminous

Lorenz Lehmhaus
Head of Communications | Aleph Alpha



Powered by





ALEPH ALPHA

BUILDING USE CASES IN UNPARALLELED SPEED

With Luminous



Well, how to do it, then?

Factor #1: Know your place

Factor #2: Know your goal

Factor #3: Build trust

Factor #1 - Know your place

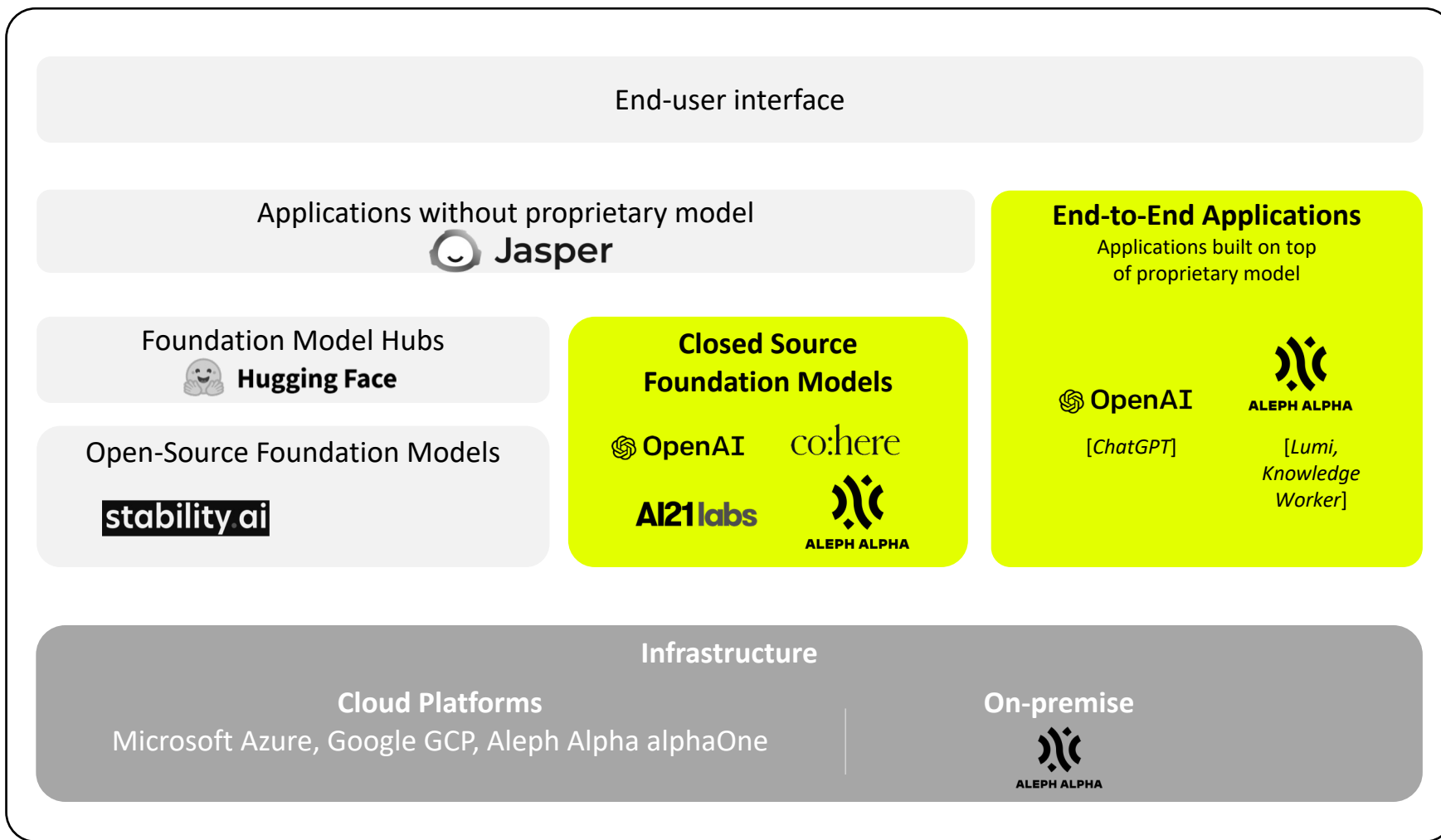
 AA core business activity  AA non-core business activity  AA no business activity

UX

Application

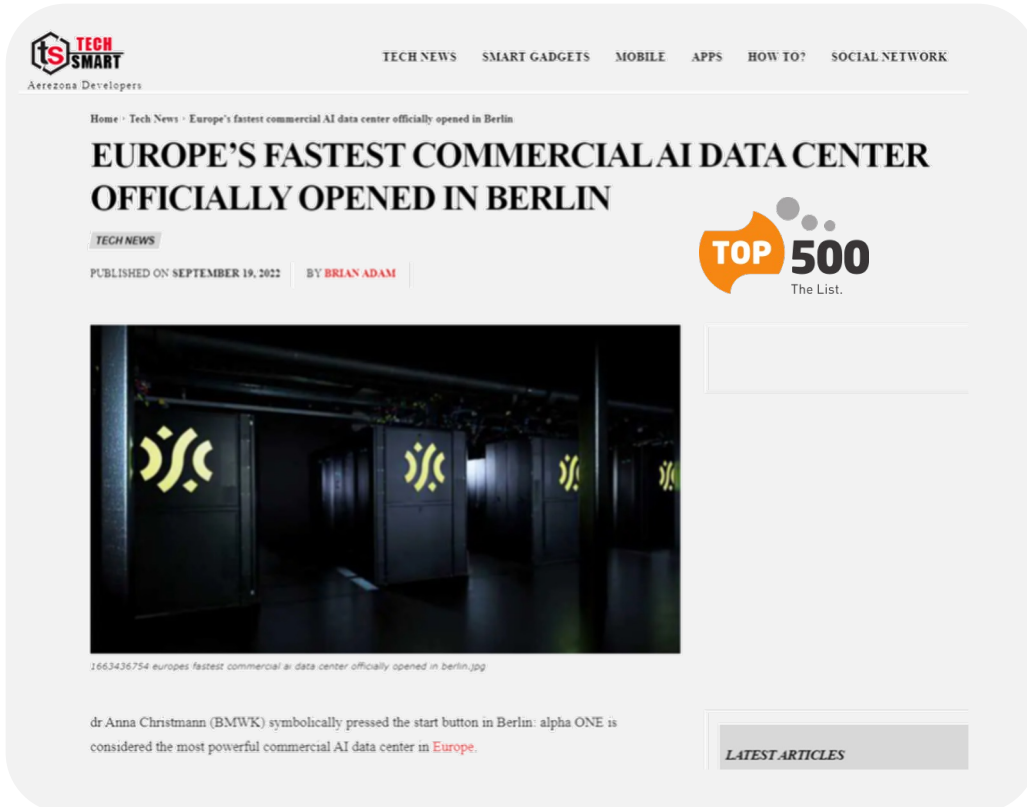
Intelligence

Infrastructure



Infrastructure

One of the world's strongest clusters in our own secure building



The screenshot shows a news article from TechSmart. The headline is "EUROPE'S FASTEST COMMERCIAL AI DATA CENTER OFFICIALLY OPENED IN BERLIN". It includes a "TOP 500 The List" badge, a photo of a server room with glowing logos, and a caption mentioning "dr Anna Christmann (BMW) symbolically pressed the start button in Berlin: alpha ONE is considered the most powerful commercial AI data center in Europe."

Enterprise-ready infrastructure layer



Nobody can shut us off

Through our data center we ensure operational & economic independence from third parties such as Microsoft, Amazon or Google



Customers can run & deploy anywhere

We license and help building on-premise installations
Our technology can be used in any cloud or environment to avoid forced architecture changes and expensive egress fees



High-end security & privacy

















We built and secured a proprietary data center enabling security and compliance for critical use cases



We do research on leading hardware

Co-development of next-gen hardware with Graphcore and other partners

Factor #2 – Know your goal

Text			Bild
Text Summarization 	Translation 	Search 	Scene Description 
Text Completion 	Q&A 	Entity Extraction 	Image Description 
Sentiment Analysis 	Structured Data Extraction 	Analysis 	Title Generation 
Text Classification 	Explainable AI 	Embedding (semantic & standard) 	OCR 

The following use cases can be quickly implemented with only a small amount of training data:

- Building a digital assistant to answer complex and open-ended questions
- High degree of document penetration to identify relevant content in highly regulated contexts
- Generation of comprehensive reports and analyses
- Identification of semantic relationships in texts as an assistance function in knowledge management
- Evaluation of pre-defined text and image content on social media

Explainability can be included in all tasks due to a specially developed functionality.

Factor #3 - Build trust

ATMAN: Understanding Transformer Predictions Through Memory Efficient Attention Manipulation

Mayukh Deb^{*,1}, Björn Deiseroth^{*,1,2,3}, Samuel Weinbach^{*,1},
Patrick Schramowski^{2,3,4}, Kristian Kersting^{2,3,4}

¹Aleph Alpha GmbH, Heidelberg, Germany

²Artificial Intelligence and Machine Learning Lab, TU Darmstadt, Germany

³Hessian Center for Artificial Intelligence (hessian.AI), Darmstadt, Germany

⁴German Center for Artificial Intelligence (DFKI)

{mayukh.deb, bjoern.deiseroth, samuel.weinbach}@aleph-alpha.com
{schramowski, kersting}@cs.tu-darmstadt.de

Abstract

Generative transformer models have become increasingly complex, with large numbers of parameters and the ability to process multiple input modalities. Current methods for explaining their predictions are resource-intensive. Most crucially, they require prohibitively large amounts of extra memory, since they rely on backpropagation which allocates almost twice as much GPU memory as the forward pass. This makes it difficult, if not impossible, to use them in production. We present ATMAN that provides explanations of generative transformer models at almost no extra cost. Specifically, ATMAN is a modality-agnostic perturbation method that manipulates the attention mechanisms of transformers to produce relevance maps for the input with respect to the output prediction. Instead of using backpropagation, ATMAN applies



Figure 1: “What am I looking at?” The proposed explainability method ATMAN visualizes the most important aspects of the given image while completing the sequence (displayed above the relevance maps). The generative multi-modal model MAGMA is prompted to describe the shown image with: “<Image> This is a painting of ”. (Best viewed in color.)

Own scientific paper on explainability in large-scale, generative AI language models, so that users can understand from which sources a generated answer arises.

<https://arxiv.org/pdf/2301.08110.pdf>

Explainability in LLMs

Prompt

Negative   Positive

While it's hard to define the market size for AI software as most general-purpose SaaS apps coming to market are powered in some fashion by AI, we estimate the global market to total between \$60bn - \$120bn ("core"/platform AI solutions only at the lower-end, all AI-enabled applications included at the higher-end). Additionally, industries including retail and financial services are the most active investors and producers of new AI applications. E-commerce tailwinds seem to have driven intelligent product recommendation capabilities to be differentiators for online retailers. In financial services, many institutions appear to be increasingly turning to AI to create personalized and automated banking experiences. Unsurprisingly, the education and public service sectors have been less rapid adopters, though we believe interest in AI innovation remains high in these industries and could be potential long-term growth catalysts.

ROI by function has evolved: Across functions, service optimization, and customer analytics are key areas where enterprises are adopting AI tools, and this has remained consistent over the past four years. However, ROI by function has evolved – in 2018, AI in manufacturing and risk generated the most ROI, and today, as identified by McKinsey, marketing and sales, product development, and strategy and corporate finance are providing companies the best mix of improved revenues. Supply chain management stands out for significant cost reductions

We have identified several areas that could drive potential corporate AI investment over the next 12 – 24 months:

- * Security and risk management priorities: As enterprises look to reduce vendor load across the application and infrastructure layers in a challenging economy, those platforms with AI-enabled capabilities that can automate key parts of the security lifecycle are likely to be more attractive investments.
- * Data management, analysis, and governance: As budgets for data science talent may decline due to economic constraints, businesses will need to become more familiar with AI/ML solutions that enable them to extract insights from large structured and unstructured datasets, automate data-driven processes, and generate predictive analytics.
- * ESG reporting: Environmental and sustainability practices are at the forefront of many enterprises' resource allocation strategies heading into 2023. Collecting, analyzing, and reporting on ESG performance from a variety of data sources will require automated solutions at scale.
- * Natural language processing (NLP) market remains robust: As noted in Silicon Republic, NLP is one of the few fields in AI which is not limited by data – deep learning systems can merge languages, images, and real-life object detection to build models without first needing to aggregate large volumes of data from customers (i.e. as is the case with self-driving car software). The global NLP market was valued at \$15bn as of 2022, and is projected to grow at a CAGR of 25%+ over the next five years.

Target

We believe that the most compelling AI start-ups in the finance industry are those that can help banks and other financial institutions to:

- * Improve customer experience: By leveraging AI/ML to automate and personalize customer interactions, financial services companies can improve the customer experience and increase customer satisfaction.
- * Reduce costs: By automating manual processes, AI/ML can help banks to reduce costs and increase efficiency.
- * Enhance risk management: By using AI/ML to analyze large volumes of data, banks can identify potential risks and take action to mitigate them.
- * Improve fraud detection: By analyzing large volumes of data, AI/ML can help banks to detect fraudulent transactions.

Multimodality + Explainability

Prompt

Negative ■ ■ ■ Positive



Give a detailed description of the image. It shows

Target

a construction site. The image is taken in the daytime. Tags: yellow crane, blue sky

Prompt

Negative ■ ■ ■ Positive



Give a detailed description of the image. It shows

Target

a construction site. The image is taken in the daytime. Tags: yellow crane, blue sky

Insurance Showcase


Select claim:



Collision between two vehicles at intersection 6619967

Damage type: Fender bender

A Mini Cooper collided with a black Audi A4.

 Luminous

[→ Show claim](#)

Title generated by Luminous based on images and text of the claim. The same holds for the summary. Additionally, Luminous automatically detects the relevant damage type based on the images. This is powered by our "Mindset" technology which uses semantic embeddings to perform classification in multimodal space.



Collision with unknown vehicle 2684695

Involved in a collision at the intersection of Main Street and Ocean Drive.

 Luminous

[→ Show claim](#)

Insurance Showcase

LUMINOUS SHOWCASE

Select Overview Q&A Discover

Playground

Select claim:



Luminous

Show claim



Porsche 911 SC collides with unknown vehicle Luminous

X

Description:



On January 24th, 2023, at approximately 7:00 PM, an accident occurred at the intersection of Main Street and Ocean Drive. The weather conditions were clear and dry. The driver of a black Porsche 911 SC, Mark Weatherman, was driving on Ocean Drive when they were hit by another vehicle on the right side, causing the Porsche to spin. The driver of the other vehicle, who is unknown, fled the scene before any identifying information could be obtained. The driver of the Porsche, Mark Weatherman, was not injured, but the Porsche sustained significant damage to the front right and right side, including the light. Mark Weatherman was unable to get a glimpse of the driver or the license plate of the other vehicle before it fled the scene. A police report was filed, and an investigation is ongoing to locate the driver of the other vehicle. Anyone with information about the accident or the other vehicle is encouraged to contact the police. It is important to report any such accident as soon as possible to the local authorities, including taking pictures of the scene and vehicles involved, if possible.

Show claim

Go to overview

Show claim



Accident on A1 near Vienna 4225027

Damage type: Wreck

A traffic accident occurred on the A1 near Vienna on January 23rd, 2023. A silver-gray Hyundai i20 collided with a red car. Several other vehicles were involved.



Insurance Showcase

Overview:



Porsche 911 SC collides with unknown vehicle 2684695

Summary and data extracted based on description text done by Luminous automatically.

Summary Luminous short long

A black Porsche 911 SC was involved in a collision at the intersection of Main Street and Ocean Drive.

Details Luminous


Date:	Location:	Involved parties:	Injuries:
January 24th, 2023, at approximately 7:00 PM	Intersection of Main Street and Ocean Drive	Two	None

Involved parties Luminous

Brand: Porsche
Model: 911 SC
Color: Black
Damage: Front right and right side damaged
Type: Fender bender



Insurance Showcase

 LUMINOUS SHOWCASE[Select](#) [Overview](#) [Q&A](#) [Discover](#)[Playground](#)


Q&A:



Collision between two vehicles at intersection 6619967

[→ Show description](#)

Select a question:

 LuminousWho is likely responsible for the accident? 

Question:

Who is likely responsible for the accident?

Luminous answer:

The driver of vehicle 1 is likely responsible for the accident, as he was driving too fast and did not have enough time to react to the sudden braking of vehicle 2.



Source: Ich (Fahrzeug 1, grauer Mini Cooper) fuhr mit Tempo 50 auf die Kreuzung Eppelheimer Straße/Da-Vinci Straße zu, als der Wagen vor mir (Fahrzeug 2, schwarzer Audi A4) plötzlich eine Vollbremsung machte. Auch ich habe eine Vollbremsung gemacht und versucht, ein Stück nach rechts zu fahren, um Fahrzeug 2 eventuell auszuweichen. Dies gelang mir jedoch nicht ganz und ich prallte mit reduzierter Geschwindigkeit auf Fahrzeug 2 auf.


[← Back to overview](#)[→ Next to discover](#)


Source will be provided for reference and allows the traceability of generated answers.

Insurance Showcase


LUMINOUS SHOWCASE

Select Overview **Q&A** Discover Playground

Overview:  Three vehicles involved in serious accident on A2 near Hannover 9117801 Show description →

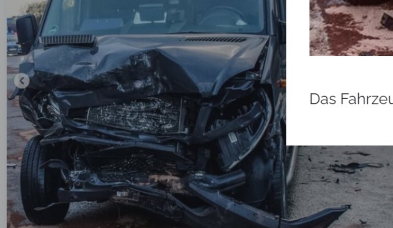

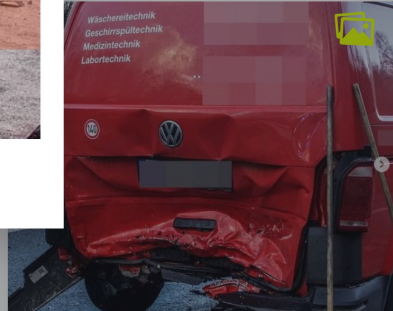
long Details  Luminous

Luminous' multimodal capability can also recognize content inside complex images. The explain functionality allows to trace the model attention for answer generation.



Involved parties: three Injuries: None

Brand: VW Model: Bus Color: Red Damage: Lower rear damage Type: Wreck

Das Fahrzeug hat eine Beschädigung an der Front Hide

← Back to select Next to Q&A →

For more information

> www.app.aleph-alpha.com [Playground & API]

> www.aleph-alpha.com [General information]