zensar

Visual Image Search and Conversational Assistant: Revolutionizing Product Search and Inventory Management

White paper

An RPG Company

According to a McKinsey Report, generative AI could add the equivalent of $2.6 trillion to $4.4 trillion annually to the economy across multiple use cases, spanning 16 business functions.[1] According to Fortune Business Insights, the global generative AI market size is projected to grow from $43.87 billion in 2023 to $667.96 billion by 2030, at a CAGR of 47.5 percent during the forecast period. Thus, it is very important to harness the power of Gen AI to develop assets that can help businesses stay relevant in the Gen AI market.

This white paper talks about Zensar's venture in Gen AI and large language models (LLMs) to revolutionize the image/product search experience, improving user experience by creating an efficient product search. Our accelerator – visual image search and conversational assistant (VISCA) has the capabilities of multiple tools, including product search, product comparison, review summarization, Q&A, product recommendation, inventory management, and pricing assistance; this makes it a powerful tool, as discussed in detail in this paper.

This white paper discusses two use cases in detail – product search and inventory management.

Gen AI solution for B2B/B2C challenges

Online retailers and businesses, both B2B and B2C, face a variety of issues:

- Customers or procurement professionals often struggle to efficiently locate the product they need in the complex catalogs available. They must search multiple locations for relevant information, leading to multiple manual searches. This increases the time required for a purchase and leads to less efficient decision-making.

- Traditional text-based search systems have limitations in contextual understanding. Users face difficulty explaining what they are looking for in words. It is easier to upload images to explain user needs.

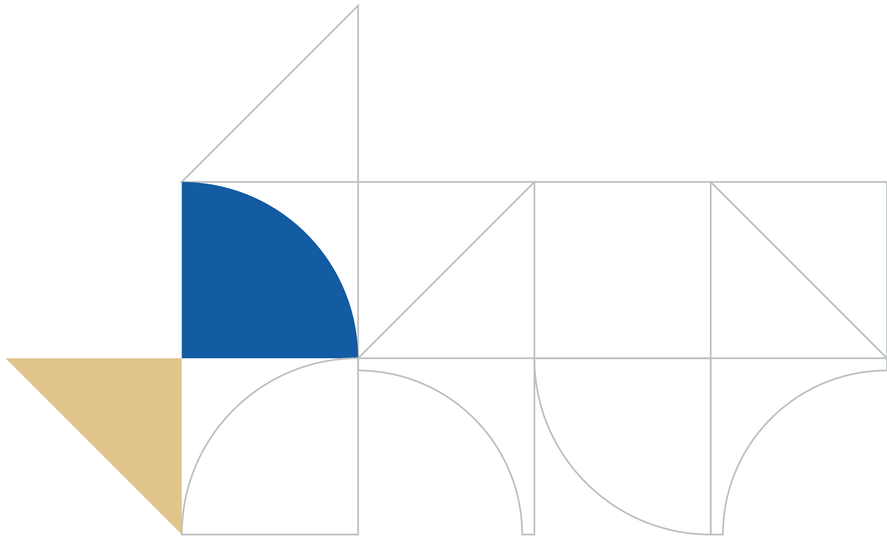
- Existing product search systems cannot combine images, text, audio, etc. To make online product search easier, there is a need for integration between these and ongoing conversation.

- Existing product reviews are valuable in making a buying decision. However, a user must scroll through and evaluate these reviews to decide. There is a need to assist users by providing concise summaries and a holistic view tailored to their questions or contexts.

- Manual taxonomy or tagging of the products in the catalog is accurate and requires subject matter experts to understand the listing categories. However, achieving scalability and dynamic updating of tags is a challenge.

- In the B2B context, businesses often struggle to grasp the full range of competitive products, making pricing decisions for their offerings challenging. Comparing products with similar competitor products is essential, but the process is time-consuming due to manual searches.

Thus far, there has not been a one-size-fits-all solution to address these multi-faceted challenges. Zensar’s visual image search and conversational assistant, a new generative AI solution, addresses all the above challenges. It provides a more thorough, efficient, and user-friendly product search experience, effectively reducing businesses' total cost of ownership.



VISCA:
The new intelligent search accelerator

VISCA relies on the vast capabilities of computer vision and LLMs to deliver a cutting-edge, personalized product search.

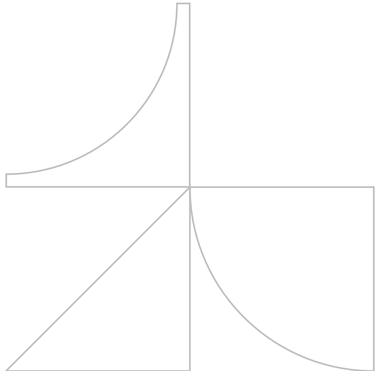
LLMs are advanced natural language processing models, such as GPT (generative pre-trained transformer), that excel in understanding and generating human-like text. LLMs are pre-trained on vast volumes of diverse textual data, which enables them to comprehend user queries, engage in contextually relevant conversations, and generate text responses that mimic human-like conversations.

Generative AI or Gen AI, is a powerful subset of artificial intelligence. It can create new, realistic data, such as text, audio, images, etc., based on patterns it has learned from existing datasets. It also excels in pattern recognition and replication amongst datasets. Generative AI generates text content to create responses for the conversational system. It can analyze user preferences and behaviors to generate personalized recommendations, enhancing user experience and engagement.

Existing online retail platforms contain multi-modal data, such as images, videos, and text relating to various products. However, they are not effectively using this available data for product searches. Traditional image searches also lack contextual understanding and the ability to engage in dialogor answer questions related to image search. Users must rephrase the query/search text to reach a relevant product search, sometimes iteratively.

VISCA integrates large language models, such as ChatGPT, LLama, etc., with vector databases and pre-trained deep-learning models as feature extractors. Each component is responsible for the following:

- Vector databases enable efficient and fast searching based on similarity scores between images
- Deep learning models extract rich visual features from images to give a detailed analysis
- Integration of LLM facilitates interactive Q&A on the obtained matched images



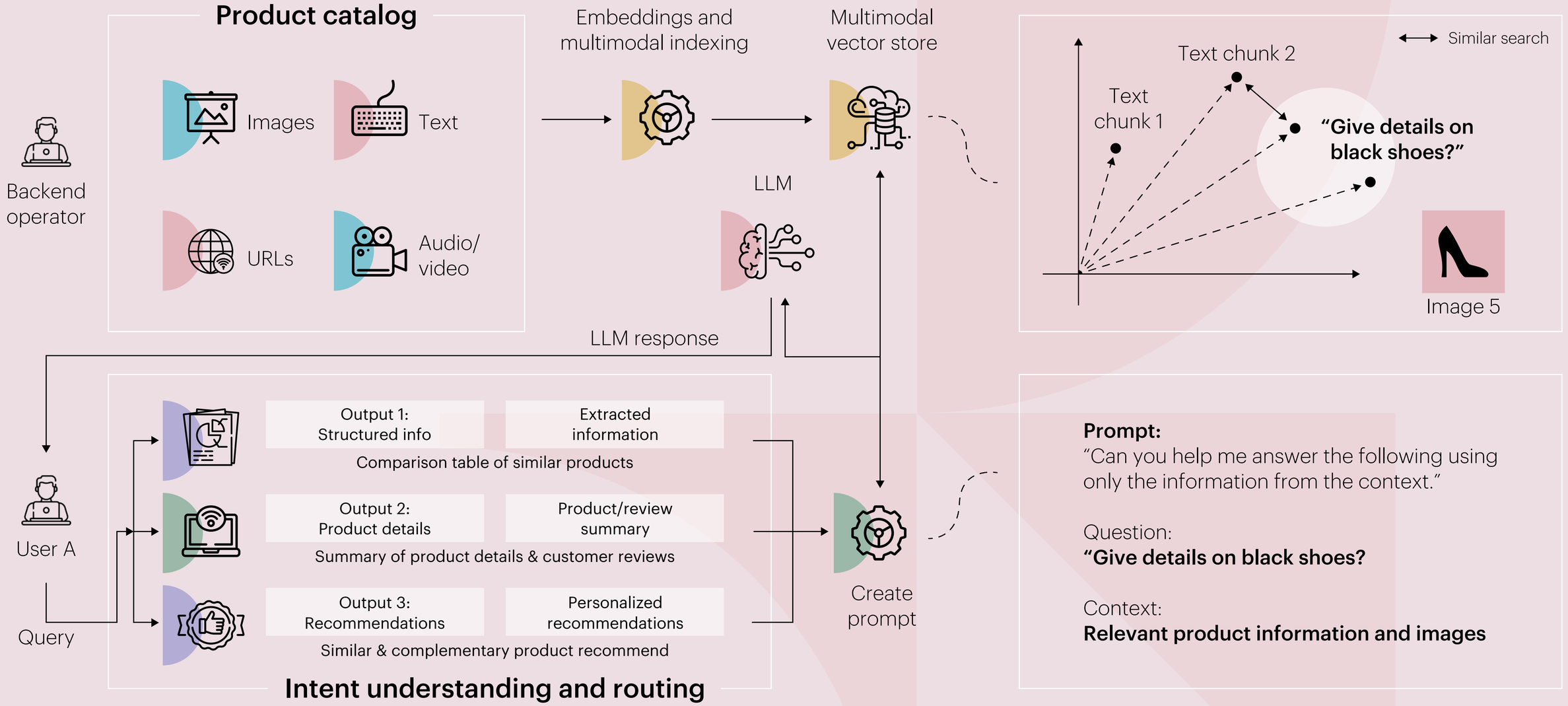


Figure 1 explains how our accelerator works from the input stage to the output stage. Each product's different data modalities are stored in multimodal vector databases. From the above, we can see the integration of the multimodal vector databases, custom-designed prompts, and LLMs to give the desired output.

VISCA for an interactive and enhanced shopping experience

We have worked on an accelerator that combines visual image search capabilities with a conversational assistant powered by Gen AI. It is a powerful and innovative application that leverages text and image processing to create a natural language understanding and interactive user experience for easier shopping.[3]

Key features:

Visual image search

- **Computer vision-based product search:** VISCA enables users to enter the query in text description or product image, which is used as a reference to query the catalog for similar products.
- **Similar image recognition:** Based on the user's query description, VISCA brings up similar products from the catalog and ranks them according to similarity by color, pattern, product, and cumulative similarity score.
- **Image metadata:** To verify the product details further, the user can query VISCA instead of going through the catalog text description/metadata. It answers the questions in a user-understandable language without the user needing to navigate the scattered catalog descriptions.

Conversational assistant powered by Gen AI

- **Natural language understanding:** VISCA can understand user-written queries and prompts and answer these in the user context in a user-friendly and understandable language.
- **Contextual conversations:** If a user chatted with VISCA previously and wants to continue, VISCA can maintain context across multiple conversations with the same user, leading to meaningful interactions.
- **Intent-based conversations:** VISCA can be extended to ask follow-up questions and understand the long context for more meaningful results, leading to a better user experience.
- **Summary assistance:** Users can ask the assistant to summarize the product description or reviews to help them make a faster and more informed buying decision.

Interactive user experience

- **Recommendations:** VISCA can provide recommendations based on personalization and other user interactions – for example, what else might go well with the user-searched product. This leads to upselling and cross-selling opportunities.
- **Seamless integration:** VISCA combines visual image search and text-based conversations and easily transitions between them without switching interfaces.
- **Multi-modal interaction:** Users can interact with VISCA using multiple formats such as text, images, audio, etc., and find all relevant product details in one place for an enriched user experience.

- **Interactive visual elements:** Users can easily interact with images, zoom in, pan, and explore visual details within the interface, achieving a 360-degree experience with the product.
- **Engagement:** VISCA provides suggestions to users about similar products, summarized product reviews, etc.
- **For B2B:** VISCA assists in dynamic pricing by comparing competitor catalogs and identifying product similarities. This helps firms launch or update products with market-relevant features and price points.

Advantages:

VISCA offers seamless product exploration, enabling users to uncover product details, reviews, etc., while engaging in informative and natural language-like interactions. It also enhances business efficiency by delivering comprehensive competitor insights.

VISCA can be leveraged for:

B2C segment

- **Enhanced customer engagement:** VISCA creates an immersive and interactive experience, allowing customers to interact naturally while receiving visual and textual insights. More information is available concisely, helping the user make an informed decision quickly. With all the information available in one place, users don't need multiple manual searches.
- **Cross-selling and upselling:** VISCA learns about user preferences from multiple similar and different user interactions to deliver personalized recommendations for related and relevant products while shopping.
- **Efficient and time-saving process:** Since all the information is available in one place, users can quickly retrieve information, get answers and relevant details, and make decisions without navigating multiple web pages, images, and applications.
- **Information management:** VISCA allows image uploads and searches for similar products. It also provides details on similarity and a concise product review summary.

B2B segment

- **Dynamic pricing:** Businesses can monitor similar products regarding features, design, etc., to ensure market-appropriate pricing.
- Businesses can also check existing similar products and add new features or capabilities to their product.

Seamlessly manage large inventory with VISCA

This section explores how VISCA can help compare manual taxonomy tagging with automated vector embedding and a prompt-based approach. This reveals distinct advantages and considerations for each method.

First, what is taxonomy tagging? Taxonomy tagging in retail refers to classifying and organizing retail products into hierarchical categories and subcategories based on their attributes, characteristics, and other relevant criteria.

The taxonomy typically starts with broad categories (e.g., clothing, electronics, home and garden, etc.) and then narrows down to specific subcategories (e.g., clothing > women's apparel > dresses > maxi dresses). Each product is tagged with relevant tags from the taxonomy, which helps accurately place the product within the online store or catalog.

Taxonomy tagging can be manual, automated, or a combination, depending on the retail operation’s scale and the product assortment’s complexity. It's a critical component of e-commerce and retail management, ensuring that the vast array of products is organized in a manner that is both logical to the retailer and intuitive for the customer.

Why is taxonomy tagging important? Taxonomy tagging helps sellers manage inventory more efficiently, and customers navigate more easily, ultimately enhancing the shopping experience.

Key features:

Below is an overview of how VISCA can be set up for taxonomy tagging:

Generate embeddings:

- Product descriptions: Convert the textual descriptions of products into vector embeddings using a pre-trained language model (e.g., BERT, GPT, or a domain-specific model). These embeddings represent the semantic content of the descriptions in a high-dimensional space.
- Taxonomy prompts: Similarly, create vector embeddings for prompts representing each taxonomy category. These prompts should be designed to capture the essence of the categories. For instance, for a "women's casual shoes" category, the prompt might be "casual shoes for everyday wear for women."

Similarity matching:

- Compare embeddings: Using cosine similarity (or another similarity measure), compare the embeddings of product descriptions with the embeddings of taxonomy prompts. This step measures how "close" a product description is to each taxonomy category in the embedding space.
- Category assignment: Assign each product to the taxonomy category with the highest similarity score. This process can be automated or semi-automated, with manual oversight to ensure accuracy.

Refinement and iteration:

- **Thresholds and override:** Similarity score thresholds might be needed to avoid misclassification and allow for manual category assignment when the model's confidence is low.
- **Continuous learning:** As new products or categories are added, the model must be retrained or updated to reflect the expanded taxonomy and product range.

Frameworks like TensorFlow, PyTorch, Hugging Face's Transformers, or spaCy can be used to generate embeddings and similarity calculations. Pre-trained models available through these frameworks can serve as a strong foundation and can be fine-tuned based on specific datasets for better performance.

Advantages:

This approach blends the power of NLP with the practical needs of e-commerce, offering a sophisticated way to automate taxonomy tagging while retaining the flexibility to adjust to new products and categories.

The key advantages that VISCA brings are:

- **Scalability:** This approach can efficiently handle large inventories and complex taxonomies.
- **Adaptability:** It is flexible and can be adapted to changes in the product range or taxonomy.
- **Accuracy:** While highly effective, accuracy depends on the embedding's quality and the taxonomy prompts' design. Regular monitoring and updating are essential to maintain performance.

Taxonomy tagging aids in several key areas:

- **Product discovery:** Systematic product categorization helps customers navigate an online store easily. They can find what they want by filtering and sorting products based on categories, brands, prices, and other attributes.
- **Search relevance:** When products are tagged accurately, search algorithms can return more relevant results to customers based on their queries.
- **Recommendations:** Taxonomy tagging can enhance the performance of product recommendation engines by using structured data to identify and suggest similar or complementary products.
- **Inventory management:** For retailers, a well-defined product taxonomy organizes products logically and helps in better inventory tracking, management, and analysis.
- **Data analysis and insights:** Taxonomy allows retailers to analyze sales and customer behavior data more effectively. Trends and preferences within specific categories or subcategories can be identified, helping in marketing and stock decisions.

Combining the power of manual and automated taxonomy

The benefits that manual taxonomy tagging brings include:

- Accuracy: Manual tagging by subject matter experts can achieve high accuracy levels. Experts understand the nuances of products and categories that might be challenging for automated systems to grasp.
- Flexibility: Humans can adapt to complex and changing taxonomies more readily than automated systems. They can understand context, cultural shifts, and emerging trends that haven't yet been incorporated into training data for machine learning models.
- Understanding ambiguity: Humans are better at dealing with ambiguous cases where a product might fit into multiple categories. They can use their strategic judgment to decide the most appropriate category.

The benefits that automated taxonomy tagging (vector embedding and prompt-based) **brings include:**

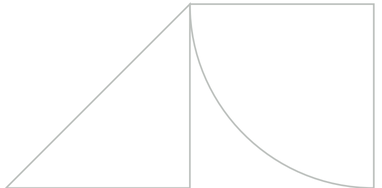
- Scalability: Automated methods can process vast quantities of data at speeds unattainable by manual efforts. This approach is essential for large retailers with extensive inventories.
- Consistency: Once trained, automated systems apply the same criteria across all products, reducing the risk of inconsistency with manual tagging due to human error or subjective judgment.
- Dynamic updating: Automated systems can adapt quickly to inventory or taxonomy structure change. As new products are introduced or categories are updated, the system can reclassify products based on the latest information without manually reviewing the entire inventory.
- Cost efficiency: It is cost-effective for large-scale operations. While there's an upfront investment in developing and training the system, ongoing costs are typically lower than the labor costs associated with manual tagging.
- Data insights: Advanced machine learning techniques offer insights beyond simple classification. They can identify patterns and relationships within the data that might not be obvious to human taggers.

Many organizations find value in combining manual and automated tagging approaches.

The benefits of combining both approaches include:

- Manual oversight can refine and validate the automated system's decisions, ensuring accuracy for critical or ambiguous cases.
- Automation handles the bulk of straightforward categorizations, enhancing efficiency and scalability.

Each approach has its place depending on the retail operation's specific needs, resources, and scale. Smaller operations or those with highly specialized products lean more on manual tagging, while larger, more dynamic inventories benefit significantly from the efficiency and scalability of automated systems.



VISCA: Constraints and solutions

The development of a system like VISCA entails several challenges:

- **Data privacy and security:** Handling visual data and user conversations demands that privacy regulations be followed. In the design phase, we need to specify the period of user data storage and ensure that data storage is secure and sensitive information is transmitted appropriately.

- **Integration complexity:** Robust integration of a visual search algorithm with a conversational AI system is technically complex and requires seamless coordination between the different components. To ensure that it works as per the requirement, the integration must be carefully designed so that no relevant connection is missed. Cross-modal understanding also needs to be handled using state-of-the-art approaches/foundation models.

- **Quality of training data:** The effectiveness of a tool like VISCA requires quality and diversity in training data. We must obtain comprehensive and domain-relevant datasets for visual and conversational inputs to ensure its success.

- **Other challenges in AI:** Biases in image recognition and conversation responses need to be addressed. The tool's scalability in handling the increased volume of data and user requests is also crucial.

VISCA customization and use cases

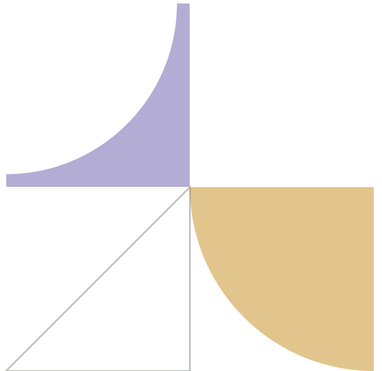
Here are some use cases that can benefit from the combination of a visual image search and a conversational assistant:

- **Fashion shopping assistance:** Users can upload images of clothing or accessories they like, and the tool can suggest similar products from various brands, providing additional styling recommendations and purchasing options.

- **Travel and tourism:** Users can upload images of landmarks or attractions, and the tool can provide information about historical context, facts, and nearby points of interest through natural language interactions.

- **Architecture and decor:** Users seeking design inspiration can upload images of rooms or styles they like. Based on the visual input, the tool can suggest furniture, colors, and decor.

- **Art collection cataloging:** Art galleries and collectors can use the tool to categorize and manage their art collections, making it easier to track pieces, artists, and provenance.



Conclusion

The integration of visual image search with a conversational assistant powered by generative AI and computer vision offers a transformative solution to address various challenges across industries. In the retail sector, VISCA and further enhancements can streamline processes, enhance the user shopping experience, and facilitate quick and informed decision-making by seamlessly blending inputs from multiple modalities. Furthermore, our search capability via VISCA can revolutionize multimodal interactions with the backend system.
It can also aid in inventory management and proper tagging of products, further enhancing the experience. VISCA combines multiple use cases into one system, reducing the total ownership cost. As businesses continue to evolve, harnessing the capabilities of cutting-edge technologies will open up new avenues for productivity, customer satisfaction, innovation, and creativity.

References

1. McKinsey & Company: “The economic potential of generative AI: The next productivity frontier”, June 2023
2. Fortune Business Insights: “Generative AI Market Size, Share Analysis, By Model, By Industry vs Application, and Regional Forecasts, 2023-2030”, June 2023
3. Medium: “Visual Search – The Ultimate Guide: Statistics, News, Trends and Tips”, Aug 2018

Authors:

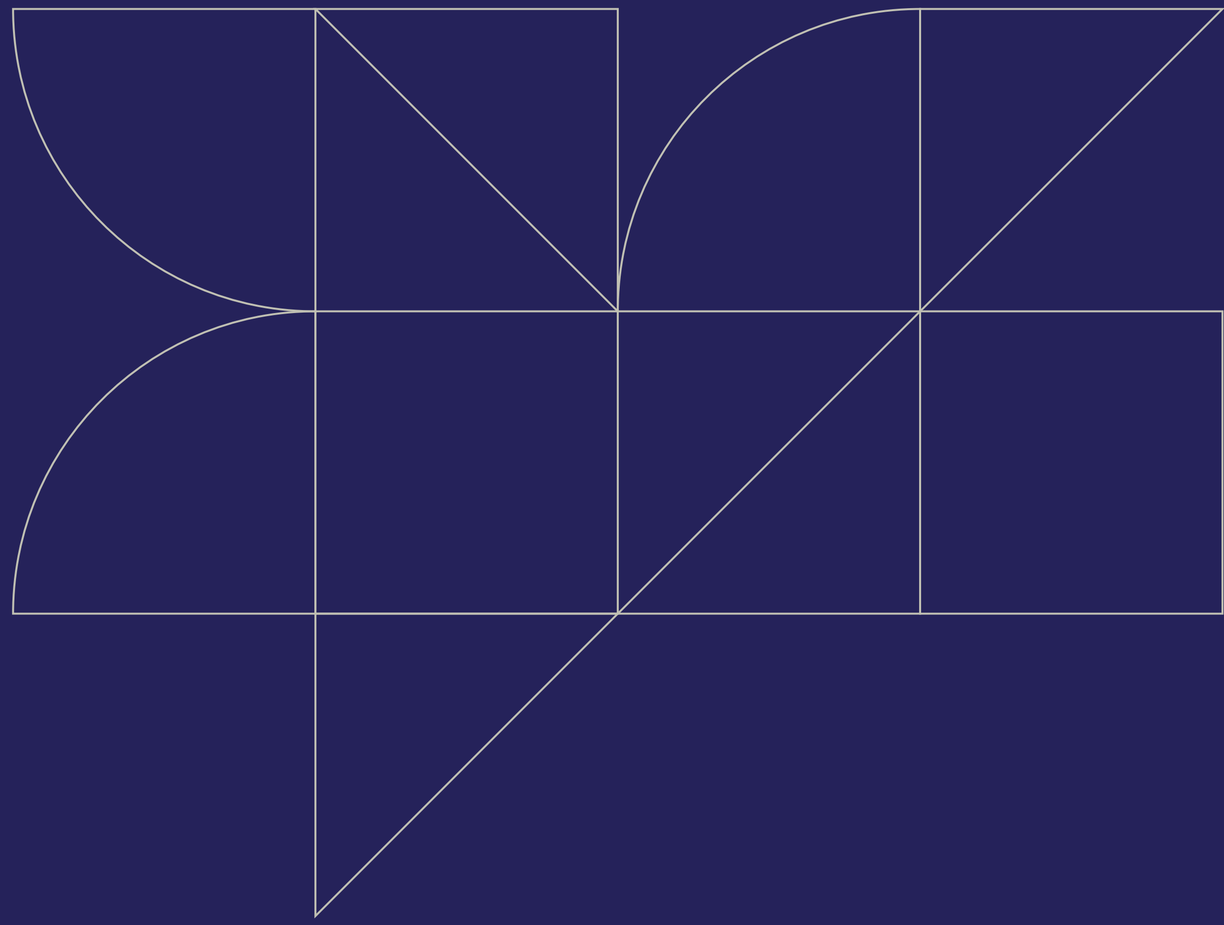
Dr. Annapurna Sharma
Lead Scientist, Artificial Intelligence and Machine Learning

Rupesh Wadibhasme
Lead Scientist, Artificial Intelligence and Machine Learning

Ankita Vashisht
Business Consultant, Artificial Intelligence and Machine Learning

Indrajit Kar
Head, Artificial Intelligence and Machine Learning

Contact Email: ZenLabsAIML@zensar.com



zensar
An RPG Company

We conceptualize, build, and manage digital products through experience design, data engineering, and advanced analytics for over 145 leading companies. Our solutions leverage industry-leading platforms to help our clients be competitive, agile, and disruptive while moving with velocity through change and opportunity.

With headquarters in Pune, India, our 10,500+ associates work across 30+ locations, including Milpitas, Seattle, Princeton, Cape Town, London, Singapore, and Mexico City.

For more information please contact: **velocity@zensar.com | www.zensar.com**