

**White Paper**



# **Harnessing AI to Power Product Content Creation**

## **A Comprehensive Guide to Trends, Implications, and Recommended Best Practices**



# What's Inside

- Introduction ..... 1
- Understanding the AI Landscape ..... 2
  - The Current State of AI in Business ..... 2
  - How Product Content is Created Today ..... 3
- Using AI for Product Content Creation and Management ..... 4
  - Content Acquisition and Curation ..... 4
  - Customer Experience Enhancement ..... 6
    - Personalized Recommendations and Targeted Marketing ..... 6
    - Support and Product Review Analysis ..... 6
  - Employee Enablement ..... 6
- Data Strategy for AI Use Cases ..... 7
- Measuring Success and ROI ..... 8
- Conclusion ..... 9

# Introduction

The promise of AI and AI-assisted solutions spans across all areas of business, each with its own unique set of opportunities and challenges. This white paper provides an overview of the AI landscape and dives deeper into the use cases that align specifically with product content creation and management. Despite this narrow focus, the same considerations are relevant across all aspects of business, specifically:

- ? **What can AI do to improve your business**
- ? **How do you best employ AI and Machine Learning (ML) tools**
- ? **What are the key risks**
- ? **How can success be measured**





# Understanding the AI Landscape

## The Current State of AI in Business

AI, and particularly Generative AI (“Gen AI”), has rapidly permeated many industries including distribution and retail. Traditional ways of doing business are being challenged and experimented with by AI usage, which has the potential to revolutionize the way we do things currently and drive unprecedented levels of innovation. However, with this rapid introduction of AI into the DNA of business operations comes a host of challenges related to accuracy, trust, reputation, ethics, and legality.

Businesses in the distribution and retail industries are investing heavily in AI-powered solutions to improve efficiencies, enhance customer experiences, optimize supply chain management, and stay ahead in a highly competitive market. According to IDC, global spending on AI-centric systems was projected to reach \$154 billion in 2023, with significant investments from the retail sector. Retail is among the top industries driving AI adoption, focusing on enhancing customer service, optimizing supply chain management, and increasing sales efficiency (IDC).

*The “fear of missing out” may also be a key factor driving technology experimentation and adoption. According to research by Ardoq, 61% of CIOs report that their investments in emerging technologies, including AI, are often driven by FOMO. They fear that failing to adopt AI quickly could leave them lagging behind their competitors. Additionally, 66% of CIOs believe that competitors will “eat them for lunch” if they don't move rapidly on AI adoption (Business Wire).*



### Best Practice #1: Prioritize Business Over Technology

We can't let FOMO drive the bus. Businesses must continue to prioritize a "business-first" approach over a "technology-first" mindset when considering new technologies such as AI.

Instead of chasing the latest technologies indiscriminately, companies should first identify their specific challenges and then assess how these new technologies can serve them better than existing technologies, ensuring that investments in technology are aligned with the organization's strategic goals.

## How Product Content is Created Today

Product content creation involves the acquisition, transformation, curation, and syndication of product information across multiple data owners, systems, processes, and channels. While numerous software tools, such as Product Information Management (PIM) systems and syndication tools facilitate this process, human beings are generally still responsible for executing many of the required tasks. Thus, while these tools streamline aspects of product content creation, they do not autonomously generate content; rather, they serve as aids for companies to more efficiently perform the same workload as in the past.

With the recent focus on AI, many companies are now exploring the integration of AI-based tools into their product content creation processes in an attempt to reduce manual effort. This typically leads to either a proliferation of the number of 3rd party tools and systems used, or necessitates companies to develop deep in-house expertise in data science and AI, which may not align with their core competencies. Regardless of the path, the additional costs placed on the business can offset the desired efficiency gains promised by AI adoption in the first place.

As a result, while AI holds promise for streamlining and enhancing content creation, the complexities associated with its implementation underscore the importance of striking a balance between technological innovation and maintaining core competencies within organizations.

### **Best Practice #2: Continuously Challenge the Status Quo**

Traditional methods of product content creation and maintenance used by distribution and retail companies aren't robust or fast enough to meet the challenges of getting to market faster, creating data that is high quality (complete and accurate), and providing value-added content that differentiates the customer experience.

The ability to deliver on these requirements requires aggressively challenging the status quo and exploring new technology and organizational approaches to scalability.





# Using AI for Product Content Creation and Management

Product content creation and management, particularly in the context of distribution and retail companies, refers to the sourcing and acquisition, curation and syndication of product content to provide the end customer with timely, accurate, complete, searchable, and informative content that builds trust, allows a confident purchase decision, and reduces the likelihood of product returns, while driving recurring business from the customer.

The complexities related to these activities stem from a variety of factors, such as upstream availability of the data, inconsistencies in how different manufacturers represent their product data, as well as in how the data can be delivered.

From a manufacturing perspective, many of the same challenges apply. However, instead of having many different vendors responsible for the diverse data problem, the challenges on the manufacturing side are related to having many different business systems holding part of the data, and representing data with different terminologies and data standards. The end goal is the same: deliver high-quality data to resellers with minimal friction, enabling them to quickly and efficiently sell the manufacturer's products in higher volumes while reducing returns.

## Content Acquisition and Curation

There are five key activities involved in the creation of high quality product information, i.e. data that is complete, accurate, and consistent:

### 1

#### DATA MODELING

##### Taxonomy and category specific schema definitions

Consistent and complete data requires a way of gauging what is necessary and sufficient for each different product category. The concepts of taxonomy and data schemas provide this critical foundation. It is key that these taxonomies and schemas be completely product-centric and separate from eCommerce categories or any P&L reporting structures.

AI technology-assisted processes can be devised to allow efficient creation and validation of taxonomies with attached schemas. However, it's an area that still requires significant human experience in product categorization to produce well-crafted data models.

### 2

#### DATA ACQUISITION

##### Obtaining the raw content

Product specifications must be sourced from original sources, i.e. manufacturer data or from various sources within the business. This can take many paths, such as web site scraping and online spec sheets, using a vendor portal to engage the manufacturer directly, or subscribing to the content from a data aggregation or syndication service. In the case of a manufacturer, the data sources may be disparate engineering, product information, and marketing systems which often will lack textual information (the 'attributes'/descriptors of the product) and digital assets (images and documents).

Various automated web-scraping tools can be used, whether site-based or search-based for individual SKUs. A mix between traditional methods (using Selenium and Python) as well as AI-based approaches work best with mixed content and unstructured, page-oriented content like PDF-based data and spec sheets.

It should be noted that the output of data acquisition is just "data", i.e. at best, fields and values. In order to determine what this data represents, a taxonomy and schema model must be in place, the product must be classified to this taxonomy, and a fuzzy mapping of data fields to schema attributes must take place. These are all separate steps in the content curation process.

# 3

## CLASSIFICATION

### Associating a product to a location in the taxonomy

Classification (or categorization) is the process of fitting a product into a product-centric taxonomy. This, in turn, enables the association with a schema, or the set of category specific attributes that apply to this type of product.

Auto-classification via either traditional pre-trained machine learning models or the use of LLMs (large language models) can be used to accelerate this process. However, it is heavily dependent on the availability of sufficient data that is also of sufficient quality.

# 4

## DATA NORMALIZATION AND STANDARDIZATION

### Mapping data values to the schema and aligning data standards across manufacturers, sources, and categories

This step is crucial to optimizing customer experience. In order for the customer to find, compare, and trust the information about your products, you must eliminate inconsistencies in how the data is represented for the same type of product. Examples include using different names for the same things ("fabric" vs "material", "printer speed" vs "PPM", etc), representing ranges and units of measure in different

ways, using abbreviations in space-constrained content like a catalog page, or using brand-specific non-standard color names.

Various AI and ML techniques can be used to perform this normalization and standardization work. LLMs have proven to be highly capable recognizing common abbreviations and word variations.

# 5

## DATA AUGMENTATION

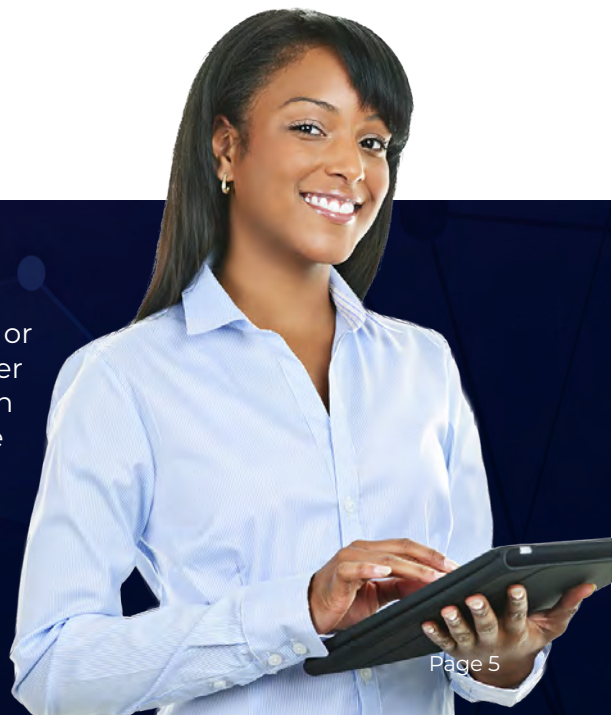
### Beyond specifications

Having acquired, classified, and normalized the data into consistent formats matching the schema and your own data standards will provide you with high quality versions of the product's specification-type data. The next level up is to provide value-added product descriptions, features and benefits, etc that help to personalize messaging to prospective buyers and thereby increase conversion rates.

The use of LLMs is a perfect fit for this activity. However, it has to be paired with a good understanding of the category as well as prospective buyers' personas. Demographic information as well as context (consumer vs professional) are all key to delivering a message that will resonate with the target audience. The use of A/B testing alongside using information about the persona of each prospective buyer is an excellent technique to use to constantly finetune your messaging based on conversion results.

## Best Practice #3: Keep Humans 'In the Loop'

AI in its broadest sense, whether traditional machine learning or Gen AI using LLM technology, can and will increasingly deliver disruptive innovations in all aspects of product content creation and management. Using traditional "people-powered" software tools for data gathering and curation will result in systems where humans provide category or industry expertise to validate the work automated by AI technologies. This approach is known as a "human expert in the loop" co-pilot method.





## Customer Experience Enhancement

As stated previously, there are ways that automation and AI-enhanced capabilities can optimize customer experience. There are two that stand out as holding the most promise.

### Personalized Recommendations and Targeted Marketing

Personalized titles and descriptions, based on prior buying records and available demographic information (cookie-based or via logins or loyalty programs) will significantly boost conversion rates and repeat business. According to [McKinsey](#), personalization can lift revenue by 5 to 15 percent, and fast-growing companies drive 40 percent more of their revenue from personalization as compared to average players in the same industries. ML and AI technologies, including LLMs, can dramatically reduce the cost of personalization, providing content differentiation with a high ROI.

Similarly, personalized and dynamic product recommendations for upsell and cross-sell will boost average order size when compared to static product recommendations that are pre-assigned as part of product onboarding.

### Support and Product Review Analysis

Support messages (email, forms, chat, etc) as well as product reviews can be turned from passive information into a powerful driver for content innovation. Understanding which products receive the most pre-and-post sale questions and the most reviews, as well as the sentiment of those messages, alongside returns rates, and margins can focus your content creation budget to where it will have the most impact.

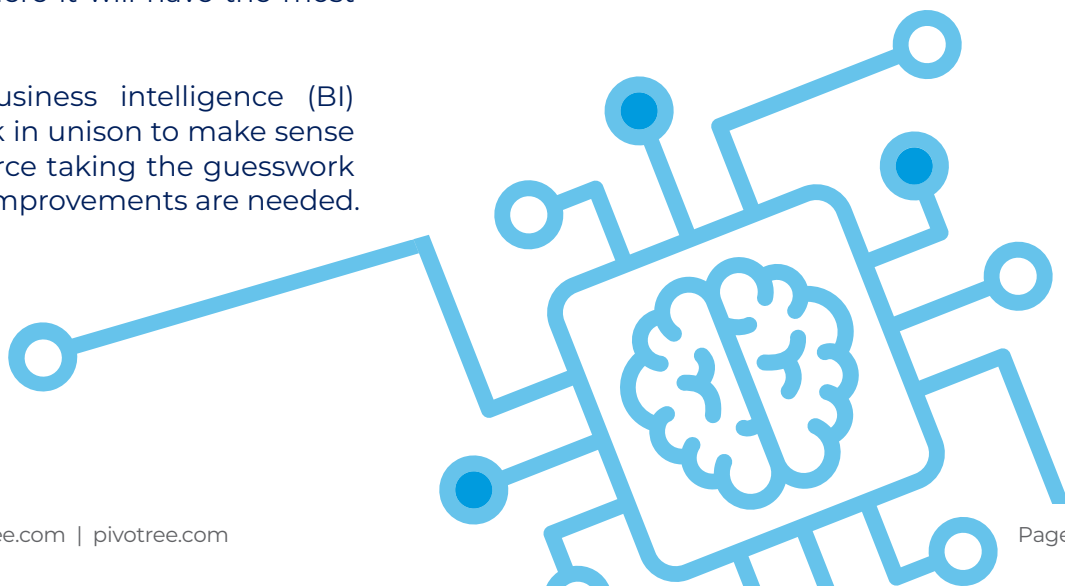
AI and traditional business intelligence (BI) visualization tools work in unison to make sense of this vast data resource taking the guesswork out of where content improvements are needed.

## Employee Enablement

Acquiring, curating, and making high quality content available to your customers is essential.

In order to foster and maintain customer trust, which is key for retention, your customer-facing employees (product support, sales, social media staff, etc) are another important stakeholder. Providing these employees the tools they need to quickly and accurately answer questions from customers and potential customers is essential. While traditional industry and product training and general communication skills are important, there is often a data accessibility issue that may prevent employees from performing their jobs well by answering customer queries in a timely manner. Data is only effective if it can be found when needed.

There are new AI-based technologies that can offer a tremendous boost in an employee's ability to access vast amounts of data instantly with very little training. A technology known as RAG (Retrieval Augmented Generation) combines the strength of LLM foundation models with the ability to train on corporate data (product content, white papers, research, support cases, etc) and can be accessed much like a chatbot by writing natural language questions. The answers will be in natural language also, but with the addition of linking directly to the sources of the answers (e.g. a PDF file, a web page, etc). This will avoid the issue of 'hallucination', where an LLM's answer sounds authoritative, but is not true. The RAG approach allows the user to validate the LLM's answer directly, and if the content is in a public document, to provide it to the customer requesting that information.



# Data Strategy for AI Use Cases

The use of AI, particularly for content creation, drives a new set of challenges that are different from traditional automation, where algorithms could be tested and known to behave predictably. With AI you cannot pre-test all scenarios, because the outcomes are fuzzier by nature.

As mentioned earlier, the general rule is to use AI as an efficiency multiplier, but always have a human in the loop. That way legal and reputational risks can be minimized and should not represent any greater risk than with a fully human-driven process.

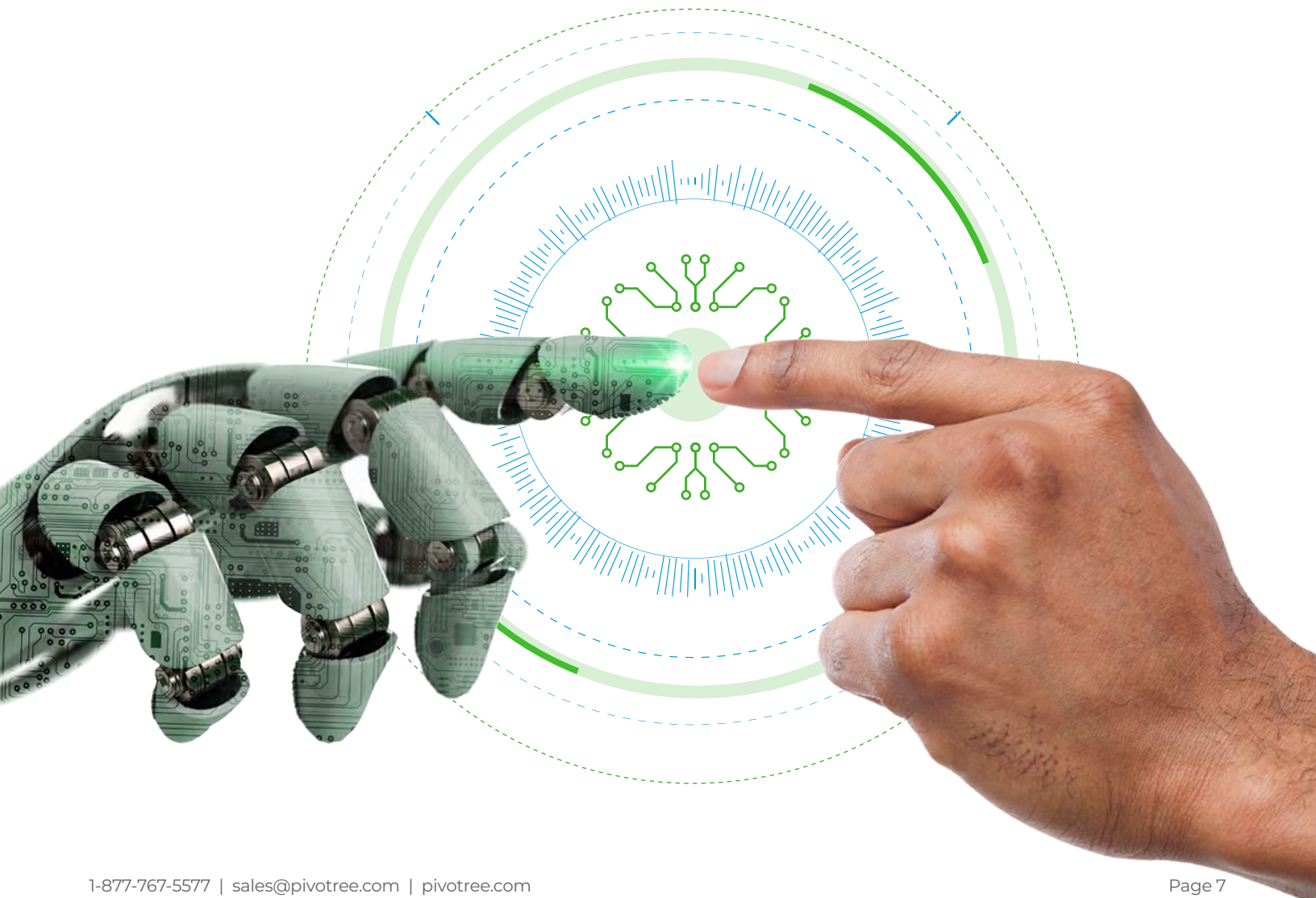
Some of the specifics to consider around AI for content generation are:

## Privacy and ethical considerations

In general, PII (personally identifiable information), or any kind of sensitive or internal data should not be provided to LLM engines, even as part of a query, unless there are guarantees that LLM model is not being re-trained using the data in the queries. This guarantee may be in the form of expressed guarantees, such as from Amazon Bedrock or Microsoft Azure OpenAI or by self-hosting the prebuilt foundation models.

## Bias considerations

There are published studies that show that AIs, through the content they are exposed to during training, can exhibit biases. While this is likely a lesser concern in the area of product content management, it should still be monitored, especially in terms of any value-added content like persona-based descriptions.





# Measuring Success and ROI

While there is a lot of excitement around what AI can do for businesses, at the end of the day, the question is no different from any other technology: does it allow you to solve problems you couldn't solve before, or can it help you perform better, faster, or less expensively, and if so, what is the ROI on the investment.

In many cases, the use of AI has the potential to drive efficiencies that change a data initiative from being prohibitively expensive to being possible or even trivial in terms of cost.

As an example, the idea of personalized content has long been a holy grail in eCommerce. But creating, testing, and maintaining two, three, or 10 different descriptions for each product was never feasible due to the effort required. With LLM assistance, and if desired with A/B-testing and analytics hooked up to measure effectiveness, this is well within the realm of possibility and can be accomplished with very limited effort. Outcomes and ROI would be measured via percentage improvements in standard KPIs such as conversion rates and overall sales.

The use of AI to curate high-quality product content eliminates having to make a choice between the time and cost to bring the product to channel versus the quality and completeness of the product content. Not only should this also improve conversion rates and sales, but accurate, complete and trustworthy data should also lead to a decrease in your product returns rate.



## Best Practice #4: Start with KPI Baselines and Monitor the ROI

Establish and document your baselines for the relevant KPIs that you expect to be affected by your content automation initiatives. Track the initiative roll-out dates and constantly monitor the impact on your KPIs and overall ROI.

Your decisions should be data informed. Proper tracking and association to specific initiatives will enable you to expand on the initiatives that have high ROI and scale back or cancel initiatives that do not yield the return you expect.

## Conclusion

The use of AI and ML in product content creation and maintenance will only continue to accelerate and it will be important for all types of businesses to take advantage of the opportunity to solve problems better, faster and with lower cost, as well as address issues that may not have been technically or financially feasible to do before its advent. However, due to the nature of AI, and the potential impact from AI-driven data issues, it is essential to employ human oversight to all AI-powered content processes.

To determine how to take advantage of the AI revolution, organizations have to decide where they want to be on the build-vs-buy spectrum. AI, as a technology, is fairly simple, but its impact is dependent on the data itself and how this data is segmented, transformed, and prepared for AI.

Deciding on whether to build in-house data science and AI expertise vs acquiring software, solutions or data from third parties depends on whether the business sees itself intrinsically as a “data company” or simply a business that needs to derive value from its data to boost critical KPIs.

ROI analysis for both the short and long term will be effective in helping companies decide where on that scale they want to be.

## About the Author

Claus Jensen



Claus has 25+ years of experience in consulting, solution architecture, product development, and product management leadership, with a key focus on PIM and MDM solutions and product innovation generating business value from emerging technologies. In his spare time Claus spends time with his family while occasionally finding time for a quick motorcycle ride to the North Georgia mountains.



**pivotree**<sup>™</sup>

**CONTACT US:**

1.877.767.5577

[sales@pivotree.com](mailto:sales@pivotree.com)

[pivotree.com](http://pivotree.com)