

The Granularity of Discovery: How SKU-Level Attribute Mapping Enhances Search Relevance and Consumer Conversion in E-Commerce

Anchal Gautam, Jinson Sani Koreth

Abstract: *The article centers on how SKU-level attribute mapping can be relevant in improving relevancy and customer experience in online stores, which is the primary retail environment/shop, when dealing with art supplies. It points out that classical title, brand, price metadata search models are incapable of providing high-intention search results based on technical data such as pigment code, ASTM lightfastness, paper weight, texture, transparency, or even nib type. The methodology used in the study is based on a comparative audit of 25,000 SKUs, in which a control group (SKUs with standard metadata) and an experimental group (SKUs with enriched SKU characteristics) were used. The results show that granular attribute mapping increased the search-to-product click-through rate by 4.2% to 7.8%, the conversion rate by 1.9% to 3.4%, and the average order value by \$64.50 to \$ 82.10, and also decreased the number of zero-result queries from 21.5% to 5.1%. The study concludes that long-tail SKU attribute mapping is better for long-tail search and minimizes ambiguity around search, trust, and compliance, but it cannot be extended into an SEO tool; instead, it is a product-discovery infrastructure in modern online trading. It also suggests using metadata depth as a strategic approach to integrating Product Information Management, faceted navigation, search engineering, and revenue architecture. The research areas that have arisen include expanding SKU-level modular AI attribute mapping to large catalogs using adaptive taxonomy design, visual search, privacy-first personalization, and relevance, fairness, and transparency.*

Keywords: SKU-level attribute mapping, Search relevance, Consumer conversion, E-commerce retrieval, Product Information Management (PIM), Faceted navigation..

Anchal Gautam (Anchal.nmims@gmail.com), Digital Product Manager, IL, USA.

ORCID: 0009-0003-7671-3752

Jinson Sani Koreth (Jinson.sanik@yahoo.com), Senior Developer, OH, USA.

ORCID: 0009-0002-0406-0505..

Introduction

E-commerce is no longer about passively displaying its collections; it is about actively investigating products and their trends, which can be achieved by a platform's ability to read what the consumer wants, not what the product is called. As a matter of fact, this environment has transformed the Stock Keeping Unit into a searchable and decision-support data object. This form of product change is relevant in product categories with high product traits, as consumers do not make decisions based on generic terms but rather on measurable product traits [1]. The market of art supplies is oriented toward the professional and prosumer market, whereby one will find an array of choices depending on either the pigment code used, ASTM level of lightfastness, openness, paper weight, texture, or type of ink base, because these parameters determine the performance and practicability of the intended application. The structured SKU attributes affect search relevance, search engine indexability, facet navigation, and search conversion.

Even with the developments of digital commerce, most retail search engines continue to use cursory metadata indices, including title, brand, and price. This model cannot be used in high-intent discovery since it lacks the language of the expert buyer. The resulting gap in findability manifests as the searcher (shopper) types in PB29, 300 gsm, or water-based markers, but the engine returns category pages or similar loose matches. This operationally triggers irrelevant responses, rephrases queries, causes queries to be abandoned, and results in loss of trust. This is made even worse by a case of lexical overlap, which conceals the differences in materials between the products [2]. A typical example of such an obvious one is whether to match Cadmium-Free or Cadmium Red, and an analysis of a broad keyword match may give a false result set unless the back-end pigment structure criterion subdivides pigment structure by SKU.

The aim of this research is threefold. It also analyzes how mapping SKU-level attributes can increase the relevance of trips to long-tail and high-intent e-commerce search queries. It also measures the observable impact of enriched metadata, such as search-to-product click-through rate, conversion rate, average order value, and a decline in zero-result queries. Its underlying audit is a 25,000-SKU two-track audit. It records a CTR of 4.2-7.8, a 1.9-3.4% growth in conversion, an average order of \$64.50-82.10, in orders, and a 21.5 to 5.1 decline in zero results. The article proposes a viable design for the correlation among Product Information Management, a taxonomy model, facet search, an AI-based extractor, and privacy-conscious personalization in the retail business.

The investigation is restricted to e-commerce discovery rather than warehousing, fulfillment, or demand projection. Its main business is art supplies, which require a high level of technical and sensorial concentration, defining the product at the SKU level. It makes a rotten mess of paints, papers, brushes, and markers, and it draws on attributes, indexes, logic, filter design, and ranking relevance as some of its ingredients. The discussion is grounded in art materials; it can be extended to other sectors, including industrial supply and electronics, where consideration of performance or compatibility is often included in the purchasing process. The design is still limited because the products that will be regulated are defined using standard metadata, whereas experimental products are enriched with attribute fields.

The study aims to establish an association between the search structure and business success. In chapter 2, the literature review presents topics on search relevance, taxonomy, Product Information Management, facet system navigation, trust development, and privacy. Chapter 3 comprises the data collection, performance measurements, experimental design, and ethical issues and regulations. Chapter 4 presents the empirical findings and explains how they apply to e-commerce. The future research directions identified in this research are dynamics in taxonomy, multimodal extraction of attributes, and personalization, with privacy as the first consideration. It lies at the intersection of SEO, information retrieval, consumer behavior, and

digital merchandising, and it believes that better discovery is triggered by better product meaning at the SKU level.

Literature Review

2.1 Search Relevance, Long-Tail Intent, and E-Commerce Retrieval

The interaction among query intention, lexical matching, semantic interpretation, and product availability will determine the relevance of the search. The research into the nature of personalization and conversion optimization suggests that relevance should be increasing, now that platforms are no longer providing a generic match for keywords, but are deriving micro-specific intent based on purchasing behavior and contextual indicators. The personalization of content and products depending on the perceived intent is then optimized according to the user experience, which is directly linked to the internal site search because a search engine is typically the first form of personalization that is made available to users of perceived intent [3].

The five-step model of Amazon considers personalization as a progressive process of data capture, product segmentation, product recommendations, testing, and refinement, and suggests that better retrieval quality will be achieved once the intention is optimized. The fact that product information is arranged and not displayed in a general merchandising category [4]. In effect, a search query like blue paint is generic and commercial. A query like single-pigment PB29 cobalt blue, on the other hand, is a long-tail query, which is less ambiguous, more specific, and more likely to match the product. This is a significant difference in factual accuracy in merchandising.

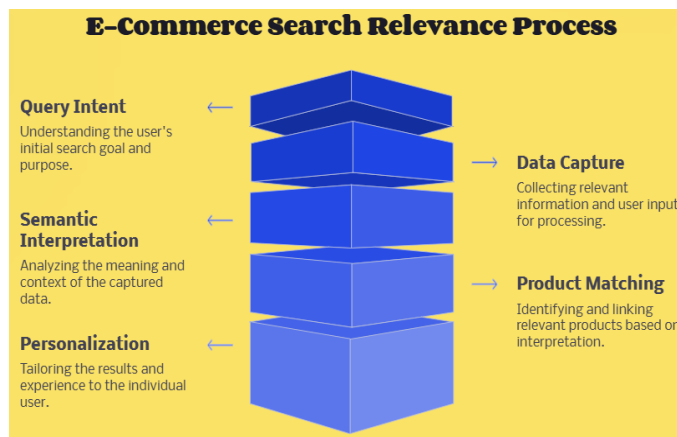


Figure 1: E-commerce search relevance develops along the lines of query intent and semantic query-data capture, product matching, and the personalization of long-tail, attribute-rich query lines of development, enabling effective retrieval.

Figure 1 shows the e-commerce search relevancy architecture as a layer cake, with the intent at the query level and the personalization level grappling with how the platform converts the user's needs into product retrieval. This model starts with the shopper's early objective, then proceeds to semantic interpretation and data capture, so that contextual and behavioral data are obtained. This facilitates matching products to trended features rather than being confined to coarse merchandising features. Personalization is the final level, which uses the results to filter out to the consumer. This figure shows the difference between general queries "blue paint" and long-tail queries "single-pigment PB29 cobalt blue", which are more specific and increase relevance and suitability for a product.

2.2 Product Information Management, taxonomy, attribute ontology.

Product Information Management transforms raw data of a specific manufacturer into searchable controlled metadata, and the taxonomy of positioning is a critical retrieval variable, not an office back-office process. The research carried out by Cavicchi and Vagnoni indicates that a digital information system provides accountability to the organization when data structures are chosen so they can be traced and ensure consistency throughout the organization's processes [5]. The same applies to e-commerce in terms of attribute governance: whenever the pigment code, level of opacities, paper sizing, type of nib, or type of the bristle is normalized as controlled vocabularies, narrow ranking, faceting, and filtering will be supported by the search index.

Scientists also demonstrate that information-sharing architectures are more effective when schemes, trade-offs, and other enabling decisions are based on interoperability rather than ad hoc arrangements [6]. Retailing has a low ontological structure, resulting in a doubling of labels, synonyms, synonymy, and misclassifications, which dulls recall and precision. They can be differentiated at the SKU level, but they can both have similar brand names. When the ontology lacks depth, the system will coalesce non-equivalent variants into a single searchable surface.

2.3 Category-Specific Attribute Semantics in Art Supplies and Comparable Sectors

The special classes rely on domain-specific semantics rather than generic retail descriptors. In art supplies, paint choices can be considered in terms of pigment code, lightfastness, opacity, series number; paper choices in terms of gsm, texture, fiber composition and sizing; brush choices in terms of shape, type of bristle, type of ferrule material used, and handle length; and marker choices in terms of ink base, type of nib, in terms of refillable, and color index name. This fine-grainedness of attribute semantics needs can be likened to accessibility-first interface design, which presupposes that design systems must capture variation without losing understandability across brands and user processes [7].

It is also similar to predictive risk modelling in aerospace supply networks, where predictive machine-learning architectures output value only when structured and unstructured variables are coded to a high degree of granularity to differentiate weak signals from noise [8]. This could be demonstrated in real-world patterns of retailers: Amazon scales to catalog normalization across gigantic assortments, Grainger relies on industrial specification filters, Sephora organizes shade and finish categories, Etsy elicits long-tailed handmade categories, and Blick Art Materials relies on material-specific terminology for professional purchasers. Expert user categories thus require a richer metadata configuration than mass-market catalogs.

2.4 Consumer trust, conversion, and privacy-first personalization

Search accuracy is also related to trust building in the literature. Becker believes that privacy in information systems is not mutable to individual user decisions due to the institutional design of autonomy and expectation-making [9]. When applied to e-commerce, it implies that product discovery systems can affect trust not only in terms of relevance, but also in the visibility of the data on which they are based and in how recommendations are constructed. This is supported by the IAPP survey of consumer perceptions of privacy and AI, which found that mediated experiences by AI are tolerated when perceived as transparent, safe, and controlled [10].

These transparent attributes, such as cadmium-free composition, ASTM ratings, or refillability, signal trust in the technical retail category, as they reduce the uncertainty at the point of purchase. Privacy-first personalization goes beyond that argument: rather than being surveilled by third parties, retailers might use facet search behavior as zero-party data, disclosed through direct interaction. The search filters are then used in two situations: as decision-support

tools for customers and for lower-risk personalization in ranking, recommendations, and merchandising.

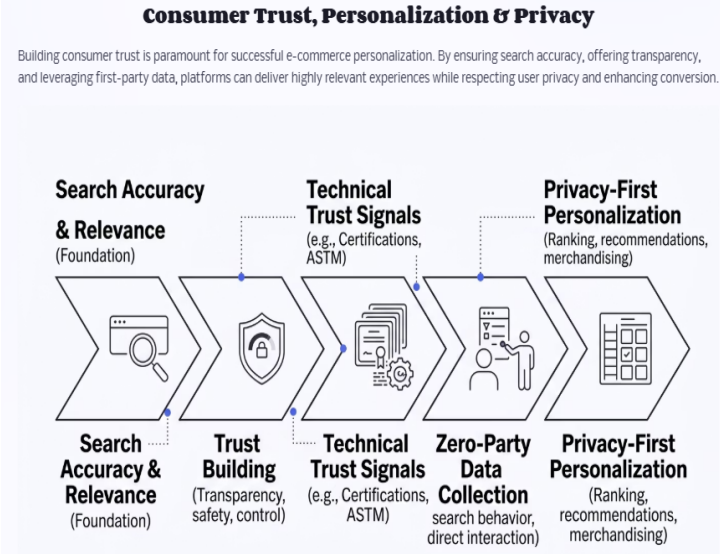


Figure 2: Privacy-first personalization, boosted by precise search, explicit technical boundaries, and zero-party data, increases consumer trust in online shopping and enhances relevance, security, recommendations, and conversion rates.

Figure 2 shows the connections among search accuracy, trust building, technical trust indicators, zero-party information acquisition, and privacy-first personalization in e-commerce discovery. It shows that consumer trust is built on accurate, relevant search results. Conversely, transparent technical information, such as certifications, ASTM standards, safety warnings, and refillability, minimizes doubt when examining the product [11]. The figure further indicates that direct user activity, particularly filter activity and search refinement, generates zero-party data that can enable safer personalization. The procedures of rankings, recommendations, and merchandising are also more applicable in this context without reverting extensively to third-party tracking to enhance trust, privacy protection, and conversion performance.

2.5 Research gaps and limitations

There are still gaps in the literature. The relationship between commercial performance and metadata depth remains loosely decoupled; most personalization researchers focus on the outcomes of recommendations or customer experience rather than the search architecture at the SKU level. Categorical scholarship is also sparse in sensorial and technical niches, such as art supplies. However, such industries depend on attributes that are higher-dimensional than those in fashion and FMCG. Privacy-related personalization is addressed on the model or governance level rather than on the filter or taxonomy level.

The literature on privacy-preserving federated learning demonstrates that practical predictions can be shared without centralizing sensitive information which opens the future of retail systems to learning from interaction-based information and to reducing exposure to raw data [12]. Granular discovery is also used to identify interface and systems design constraints. Other features might not be worthwhile discoveries unless the interface is user-friendly. The literature on quantum sensors reports achieving 12 TB of magnetic signal with 280 samples to detect micro-corrosion, demonstrating that more precise signal resolution only enhances detection when classification pipelines can accommodate the new complexity [13]. It is a retail

crisis among the professionals. These failures are resolved by treating attribute depth as the determining factor for search relevance, trust, and conversion in the context of real-world retail.

Methods and Techniques

3.1 Data Collection Methods

A two-track performance audit is suggested for a list of 25,000 SKUs of art supplies. The sample involves four types: paints, paper, brushes, and markers. To equalize it, the services may be divided into experimental (enriched metadata) and control (standard metadata) arms, at a 50:50 ratio, with 12,500 SKUs per arm. The data to be incorporated includes product export information, search logs, clickstream logs, facet usage logs, conversion logs, and transaction basket-level transaction data. It is a validity that can be accepted in the environment of advanced analytics and retail governance facilitated by AI, when the structured data streams are considered as the performance measure [14].

Five observation units are included in the sample frame, namely SKU records, query strings, search sessions, search-to-product clicks, and completed orders. These are title and brand, price and category, variant identifier, inventory, and completed ratio, which must be posted at the SKU level. Some of the query-level fields include query length, query specificity to tokens, query expansion to synonyms, the number of results, and the absence of results. The session-level content will include the device type, the message between the facets, outbound, and the value upon completion of an order. The query and the session should include actions to follow, as mistakes or relevance errors could occur if either search is performed. At the same level, the implication can be found at the session level.

3.2 Data Analysis

Search-to-Product Click-Through Rate, Conversion rate, Average order Value, Zero result query rate, search exit rate, and Facet engagement rate are the most important dependent variables. These are to be calculated using clear formulas: $CTR = \text{clicks}/\text{searches} \times 100$, $CR = \text{orders}/\text{search sessions} \times 100$, $AOV = \text{revenue}/\text{orders}$, $ZRQR = \text{zero-result queries}/\text{total queries} \times 100$, $\text{Search Exit Rate} = \text{this is the exits after search}/\text{search sessions} \times 100$ and $\text{Facet Engagement rate} = \text{sessions using one filter}/\text{search sessions} \times 100$. The variable to be retained in the discussion, revenue efficiency, should be defined as $E = (\text{Conversion} \times \text{AOV})/\text{Search Exit Rate}$. The given formulation is applicable, as the lift from increased conversion and basket value can only be achieved when search exits are kept under control. This is the rationale provided by Qualtrics' advice on customer lifetime value, which makes clear that retention-related revenue measures are more valuable than transaction counts in customer economics [15].

One of the analyses is conceptual statistics such as means, medians, standard deviations, and interquartile ranges. It can then compare metadata conditions between A/B tests, test differences-in-means for continuous results, and rates via two-proportion tests. Predictors should include metadata depth, query specificity, device or category, and order completion, and the model should be a logistic regression with order completion as the binary outcome. The separation of broad head terms and long-tail queries is achieved through query clustering. It will include the 95% confidence intervals, a p-value of less than 0.05, and effect sizes.

Table 1: The overview of the important data processing indicators, metrics, and statistical operations that should identify the reality of the searches, the transformation rates, the user experience, and the revenue performance

Data element	Definition	Formula specification	Observation level	Analytical purpose
Search-to-Product Click-Through Rate (CTR)	Measures the proportion of searches that lead to product clicks	$CTR = (\text{Clicks} / \text{Searches}) \times 100$	Query level; session level	Evaluates how effectively search results attract user engagement after query submission
Conversion Rate (CR)	Measures the proportion of search sessions that end in completed orders	$CR = (\text{Orders} / \text{Search sessions}) \times 100$	Session level	Assesses whether search relevance and product discovery translate into purchase completion
Average Order Value (AOV)	Measures the mean revenue generated per completed order	$AOV = \text{Revenue} / \text{Orders}$	Session level; order level	Indicates whether improved discovery leads to higher-value baskets
Zero-Result Query Rate (ZRQR)	Measures the proportion of queries that return no results	$ZRQR = (\text{Zero-result queries} / \text{Total queries}) \times 100$	Query level	Captures search failure and catalog retrieval gaps
Search Exit Rate	Measures the proportion of users who leave after searching without further interaction	$\text{Search Exit Rate} = (\text{Exits after search} / \text{Search sessions}) \times 100$	Session level	Evaluates abandonment associated with weak relevance or poor search experience
Facet Engagement Rate	Measures the proportion of search sessions in which at least one filter is used	$\text{Facet Engagement Rate} = (\text{Sessions using one filter} / \text{Search sessions}) \times 100$	Session level	Assesses how actively users rely on faceted navigation to refine discovery
Revenue Efficiency (E)	Composite discussion variable linking conversion, basket value, and search abandonment	$E = (\text{Conversion} \times \text{AOV}) / \text{Search Exit Rate}$	Session level	Estimates overall commercial efficiency of the search system under each metadata condition
Descriptive statistics	Summary statistics used to describe the distribution of variables	Mean, median, standard deviation, interquartile range	Query level; session level	Establishes baseline performance patterns before inferential testing
A/B comparison	Direct comparison between control	Control/search experience A vs.	Query level; session level	Determines whether metadata

The Granularity of Discovery: How SKU-Level Attribute Mapping Enhances Search Relevance and Consumer Conversion in E-Commerce

Data element	Definition	Formula specification	Observation level	Analytical purpose
	metadata and enriched metadata conditions	enriched/search experience B		enrichment changes performance outcomes
Difference-in-means testing	Statistical comparison for continuous variables between groups	Applied to metrics such as AOV and other continuous outcomes	Session level	Tests whether average performance differs significantly across metadata conditions
Two-proportion z-test	Statistical comparison for rate-based variables between groups	Applied to CTR, CR, ZRQR, and Search Exit Rate	Query level; session level	Tests whether proportional differences are statistically significant
Logistic regression	Predictive model for binary purchase completion	Dependent variable: order completion (0/1); predictors: metadata depth, query specificity, device, category	Session level	Estimates the effect of enriched metadata while controlling for other variables
Query clustering	Segmentation of search behavior into broad and long-tail intent groups	Broad head terms vs. long-tail, attribute-rich queries	Query level	Identifies how metadata depth performs across different search-intent patterns
Statistical reporting standard	Thresholds for interpreting analytical results	95% confidence interval; $p < 0.05$; effect size reporting	All analytical outputs	Ensures rigor, comparability, and practical interpretation of findings

Table 1 presents the key information analysis infrastructure against which the impact of SKU-level metadata will be measured, instead of search and conversion performance in e-commerce. It offers key performance indicators, such as the search-to-product click-through ratio, conversion rate, average order value, zero-result query ratio, search exit ratio, and facet engagement ratio, along with the revenue efficiency metric $E = (\text{Conversion} \times \text{AOV}) / \text{Search Exit Rate}$. The table also lists the methods used for analysis, including descriptive statistics, A/B testing, difference-in-means testing, logistic regression, and query clustering, to compare the standard and enriched metadata conditions in the retail search setting.

3.3 Experimental Design and Variable Operationalization

Metadata depth is an independent variable. The basic metadata will include title, brand, and price. Enriched metadata is added, including SKU-level data such as pigment code, lightfastness, opacity, gsm, paper feel, sizing, type of bristle, type of ferrule, nib type, refillable, or ink base. The query specificity, category type, user segment, and device type must be considered as moderation variables. The advent of query specificity can be measured using the number of tokens and technical density, such as 1-2 generic tokens and 3 or more attribute-

bearing tokens. The observed behavior could be used to differentiate users and to re-categorize the facet level into novice, prosumers, and professionals.

This is practicable through efforts to make search systems less intrusive, customize customer products, and track customer satisfaction and loyalty [16]. The outcome logic depends on metadata granularity; greater granularity should enhance relevance ranking and the parsimony of the facets; greater relevance should trigger abandonment and increase purchase probability. The test conditions are as follows: experience A and B, both enriched, under a single traffic window, a merchandising calendar, and a pricing condition.

3.4 System Architecture and Attribute Mapping Workflow

The process of its technical organization involves dealing with the raw marks of the manufacturer's information load, attribute ontologizing, taxonomy, schema confirmation, search index enrichment, and concluding with a layer of personalization-ranking. To force them to perform in the production control of an automated cloud structure, they ought to be run in a scalable arrangement, with the seed having zero manual operational interference. According to findings, cloud automation, orchestration, provisioning, and monitoring reduce latency and configuration drift on data-intensive platforms [17].

Dynamic taxonomy is required to make the three dynamic properties dynamic and to obtain lost properties as the scale grows. One can discuss information in PDFs and product photos, which can be addressed with AI. To establish a balance between the reliability of such a pipeline, a mixture of schema tests, indexing tests, and deployment resonance should be added to an AI-enhanced CI/CD process, and the topicality of searches can drop quickly without any limitations on mapping in case updates are deployed simultaneously [18].

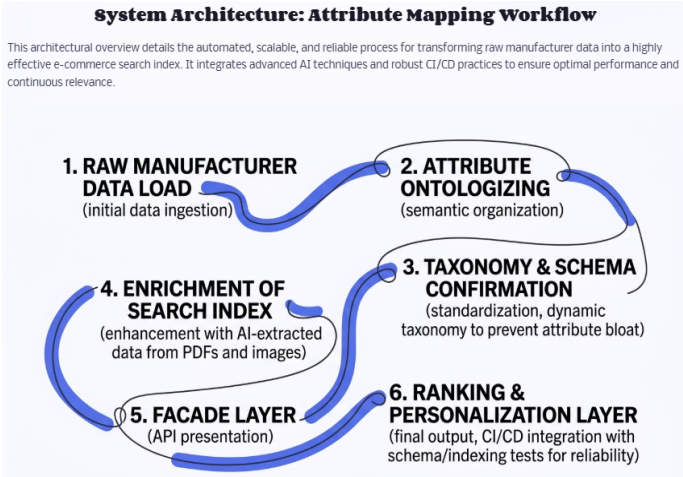


Figure 3: The SKU attribute mapping system architecture includes consuming crude manufacturer information, performing ontology and schema checks, enriching the search index, frontend delivery, ranking, and Personalization.

Figure 3 shows the systems built and demonstrates the feasibility of the SKU-level attribute map, which is realized through automated configuration and converts the company's unstructured data into a high-performing, scalable e-commerce search platform. It spawns with the generation of information, consumption, and standardization of features, taxonomy, and validation of the information schema, introducing a component to pep up search, a pep-up search facade component, and ranking and personalization, which is the final but definitely not the least. The figure also emphasizes the need for cloud-based automation to assist manufacturers in correcting catalogues and indexes, as well as adjusting the degree of

The Granularity of Discovery: How SKU-Level Attribute Mapping Enhances Search Relevance and Consumer Conversion in E-Commerce

aggregation in logs with minimal input. It involves active taxonomic search, AI-enhanced PDF and image search, and CI/CD testing of schema and indexing to confirm the relevance and accuracy of search results [19].

3.5 Ethical and Regulatory Considerations

The particular issues that the ethical and regulatory design should be sensitive to include are privacy, transparency, fairness, and safety. The perfect solution is a zero-party, attribute-based personalization that uses stated filter data while accounting for cookie indifference. However, across all pipelines with search history, behavioral alerts, and automated classifications, data provenance must be documented, access controls must be created, retention plans must be created, and auditing accounts must be pooled. It is cost-efficient because once a data breach occurs, it cannot be reversed and can result in significant losses to a business, investor confidence, and trust among other stakeholders [20]. The cadmium vs. cadmium-free case demonstrates that metadata error is not a light matter in that any classification error, as any error in the label of a pigment, may still result in the compliance lapse, risky substitution, or some other undue returns in institutional procurement. Based on this, the automated extraction outputs must be open to review, bias-tested, and disclosed when they affect ranking or filtering decisions.

Discussion and Results

4.1 Search Relevance Uplift and Click-Through Performance

The best retrieval rate in the study was an increase in the search-to-product click-through rate from 4.2% to 7.8%, which is tantamount to a +85.7% relative improvement with intimate SKU metadata. This trend indicates that the click action was reactive after the indices of attributes were set to the variability level. A priority on better results minimized the gap between the queries and the resultant SKU, snippet relevance worked more effectively with specialist users because it revealed the very properties the specialist requires, and filter-result conformity also minimized hard-to-find instances.

In terms of retrieval, CTR is the first characteristic of a search system that correctly decodes intent, but only shows signs of downstream checkout behavior afterward [21]. The SEO relevance of this discovery is that the higher the internal search utility, the deeper the level of engagement, the lower the pogo-like abandonment, and the more effectively long-tail landing paths are found at scale. This perspective can explain why narrower result sets yield faster engagement gains than broader lexical matching by revealing intent fulfilment, which might also explain why user annoyance increases when irrelevant material interrupts intent fulfilment [22].

4.2 Conversion Performance and Consumer Purchase Intent

Conversion rates in the control and enriched conditions were 1.9% and 3.4%, respectively, an increase of +78.9%. This fact indicates that we chose SKU properties to drive the entire decision-making process, not impoverished decision-making per se, in contrast to a simple catalogue description. The uncertainties were minimized, and the process of fitting the product as a flat pack was enhanced, as the buyer could examine the pigment formula, paper, fabric, feel/ and nib profile using the information during the retrieval and product-view procedures. It is of particular importance to an expert as well as a prosumer customer, who can frequently be anxious about compatibility: a refillable marker with an ink supply or water-based paper that can be washed.

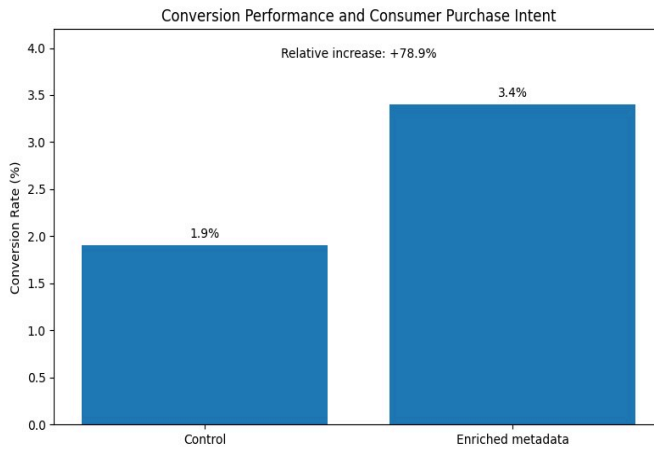


Figure 4: Graphical presentation of the conversion rate growth rate of 1.9% to 3.4% and how increased enriched SKU characteristics decreased confusion, enhanced fitment accuracy of products, and or boosted purchasing intentions.

Figure 4 highlights the difference in conversion performance between the control and enriched metadata conditions, with conversion increasing by 78.9% from 1.9% to 3.4%. The figure shows that SKU-level characteristics are dynamic decision-support indicators and not active catalog items. Enriched metadata minimizes search and product evaluation because the pigment composition, paper weight, surface texture, and nib format are more visible during search and product fit improvement [23]. This is particularly relevant to expert and prosumer purchasers, who would expect compatibility assurance before making a choice, hence the association of higher relevance quality with stronger commercial performance.

The conversion uplift, therefore, suggests a business association between the relevance quality and the purchase completion. Real-time analytics technologies are useful in this case, as the merchant can visualize the point of improvement to trigger: in search, the progression of product views, or the transition to the cart. The current literature on live metrics integration also shows that, operationally, one can observe performance at the event level and does not rely exclusively on periodic summary reports [24].

4.3 Basket Quality, AOV, and Revenue Efficiency

Trade value was also enhanced. The average order value rose by 27.3%, from 64.50% to 82.10%, indicating that the experimental group was approximately 4 times more efficient than the control group. It is not so much an outcome one should take with a grain of salt. It is, however, a commercialized increase, yet the AOV was not climbing in the vacuum of conversion [25]. With improved metadata, complementary products can be located more easily, such as papers that are an ideal match for a media type, viscosity of the pigment, or markers that are an ideal match for the refill system. As technical suitability increased, professionals were more favored for making more specific decisions. Optimizing edge-computing patterns by studying order flow can enhance them. It has already been proven that a low-latency decision system is particularly valuable for enhancing continuity, and this principle also applies to attribute-rich search, which reduces friction in product checkout, boutique construction, and payment [26].

4.4 Zero-Result Reduction, Search Integrity, and the Cadmium Dilemma

The greatest decrease in friction during search was observed in the zero results performance. The zero-result rate decreased by 21.5% to 5.1%, which is 76.3%, indicating that the enriched attribute mapping significantly increased the relevance of the presented products

The Granularity of Discovery: How SKU-Level Attribute Mapping Enhances Search Relevance and Consumer Conversion in E-Commerce

without the involvement of loose keyword inflation. The Cadmium Dilemma can be used to explain such an outcome in terms of its significance [27; 28]. In shallow systems of lexicon, we will find Cadmium Red when he types a search query of cadmium-free, though these keywords share dominating words. Quite on the contrary, the definition of what is wanted is an exclusionary reference. After indexing the pigment code and the safety-related composition, the backend exclusion rules and negative filtering can be relied upon to make the engine more accurate. It consists of finding it out correctly. It is relevant and an issue of safety, compliance, and institutional purchasing (schools, studios, and institutional buying). The reinforcement-learning process used by the researcher in developing a policy for an automated firewall is intensifying, as it shows how rules can be reinforced by a machine when the policy's purpose is transparent and can always be tested with [29].

4.5 System-Level Implications for Real-World E-Commerce

The most significant fact about the SKU-level mapping research is that it should be considered infrastructure, not an addition to the cosmetics machinery. The 100,000 SKUs suggest to retailers that they need to redefine their evolving taxonomy rules, avoid bloating attributes, introduce AI-smart operations to extract concealed information from PDFs or packages, and customize it without the need for fuzzy tracking. This is not the case with art supplies. Cosmetics Retailers work by shade, undertone, and finish features; industrial distributors by conformity/specification filters; electronic sellers by matching reason, and health-sensitive segments by correct disclosure of composition [30].

Table 1: A summary of how dynamic taxonomy, AI-grounded methodologies, privacy-based personalization, and scaling catalog control help strengthen discovery, compliance, personalization, and revenue generation amid twists and turns in the e-commerce environment.

System-level implication	Operational function	Real-world application	Commercial significance
Dynamic taxonomy	Prevents attribute bloat and keeps product classification adaptable as catalogs expand	Art supplies, cosmetics, and electronics with frequent variant growth	Improves search precision, filter usability, and catalog scalability
AI-assisted attribute extraction	Recovers hidden product data from PDFs, labels, and images for search indexing	Manufacturer spec sheets, packaging details, and compliance documents	Increases metadata completeness and reduces manual cataloging effort
Privacy-first personalization	Uses filter behavior and direct interactions as zero-party data for ranking and recommendations	Cosmetics shade matching, product recommendations, and merchandising	Enhances relevance while reducing dependence on invasive tracking
Scalable catalog governance	Maintains consistent attribute standards across large inventories approaching or exceeding 100,000 SKUs	Industrial supply, electronics compatibility search, and safety-sensitive retail	Supports reliable discovery, compliance control, and revenue growth across complex e-commerce environments

Table 2 outlines the major implications of the system-level scope of the SKU-system attribute mapping as a viable e-commerce operation. It provides an overview of the strength of dynamic taxonomy, search of attributes based on AI during a competition, individualization

through attacks, and enormous catalog management within a high-tech retail setting. The implications are linked to real-life uses, as shown in the table, including art supplies, cosmetics, industrial supplies, electronics, and other industries where a high level of safety sensitivity is necessary. This means that attribute management enhances the effectiveness of the search, the completeness of the metadata, the health of personal provision, compliance, and the size, thereby enhancing productivity and preserving energy, and enabling greater earnings for the largest digital store assortments.

As the VMware vCloud Director API simulation system demonstrates, orchestration must be tested long before the production system is introduced, just like attribute pipelines are tested before rolling out the search system [31]. Verification disciplines are also relevant, as reflected in other works that indicate the need to apply them in cases where systems are required to scale effectively without compromising reliability or increasing cost [32]. Operationally, SKU-level attribute mapping is useful as a search-relevance mechanism and as a revenue architecture.

Future Research Recommendation

5.1 Multimodal Attribute Extraction and Automated Catalog Enrichment

The study may be extended to consider multimodal catalog enrichment not as an AI vision problem, but as a multi-system problem. The analysis of vision-language systems and retrieval-enhanced pipelines for extracting SKU knowledge from packaging photographs, spec sheets, safety data sheets, and unsearchable PDFs at scale should be the second step in the research. The best features among art supplies include the pigment code, lightfastness, transparency type, series number, GSM, nib type, and refillability. The realistic plan is to eventually annotate (score) 5,000 or more SKUs automatically, rather than manually. These would comprise accuracy, recall, F1 ratio, false omission rate, and manual literature review cost. The literature suggests that edge AI, when retrieval-enhanced, can be a viable solution for operational risk scoring within operational constraints and that the enhancement is also useful for an operationally realistic catalog of live operations, rather than cooked on unrealistic laboratory assumptions [33].

5.2 Visual Search and Multimodal Retrieval Behavior

The need to study text-majority retrieval is not the second research question on which the next research should be based, but rather the need to study how users can find products when the technical requirements are not adequately defined. Customers will perceive a graphic display of a paper tooth, a brush shape, or paint undertones more quickly than if they had to imagine them. The follow-up study should thus entail conducting experiments on texture-based visual search, swatch matching, and cross-modal search, in which case the uploaded photo serves as a searchable feature that can easily be modified into a color swatch. The metrics to be included in the response variables will be top-5 relevance, time-to-first-click, access to facets, and the likelihood of user conversion.

The design would compare untrained and trained respondents across desktop and mobile conditions, since a visual-discovery process would provide novices with fewer vocabulary items to consult. Concurrently, the professionals would be more accurate in their responses. The research hypothesis presupposes that the concept of multimodal retrieval can be applied in a case of language failure in a long-tail environment with 100,000 variants of a product, and that the similarity metric is applicable. A/B testing estimates the latency, misclassification rates, and confidence levels for the two segments [34].

5.3 Dynamic Taxonomy, Adaptive Faceting, and Interface Overload

Further research is necessary to identify the limit beyond which metadata can be improved without becoming detrimental to discoveries, rather than introducing interface

The Granularity of Discovery: How SKU-Level Attribute Mapping Enhances Search Relevance and Consumer Conversion in E-Commerce

overload. It can be workable and operate on a large list, so that any quality a query may have can now be twice as effective in narrowing the search. The experiments will be conducted by comparing and contrasting categories, query specificity, device type, and query time for that session [35; 36]. Facet interaction rate, filter abandonment rate, scroll depth, query reformulation rate, search exit rate, and other factors should be considered.

For example, 6-8 high-salience facets can be non-friction on a mobile interface; more on a desktop interface, with 10-12 facets, as long as they are arranged hierarchically. The researchers fail to address enterprise digital transformation using adaptive systems as a governance issue, focusing instead on interoperability, incremental implementation, and the generation of quantitative value within the framework of business functions [37]. In this regard, dynamic taxonomy must be founded not only on usability but also on the maintainability of the rules and on performance diversity at the category level.

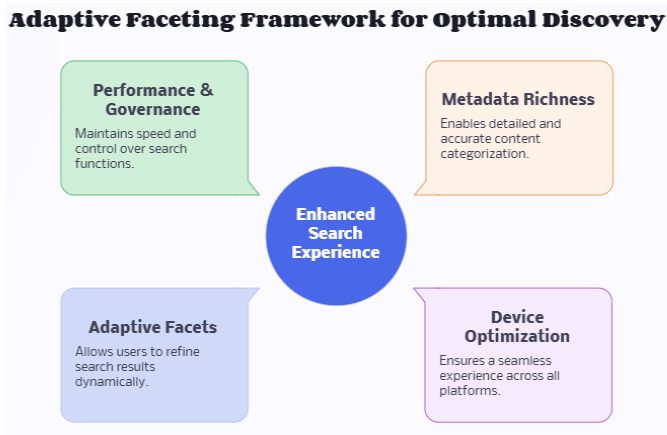


Figure 5: An Adaptive faceting makes metadata richer, devices more sophisticated, and governs and dynamically filters services to enhance discovery and reduce overload on the interfaces of large, highly endowed e-commerce catalogs.

This is suggested by a proposed adaptation faceting model shown in the picture, aiming to maximize exploration of e-commerce collections and define the nature of a refined search experience, built on a healthy balance between metadata enhancement, adaptive facets, device-specific customization, and performance optimization. The hint strengthens the value, enriching the metadata, and the exemplary filters positively relate to one another. Though it comes with a trade-off, it is more user-complex for users to navigate, with interface overflow expected. The experimental designs would be improved to exploit the fixed and adaptive features of the category type, specific query, device context, and session stage. The mobile interface is expected to have 6-8 salient facets, and we may have 10-12 filters arranged in a hierarchy that is still usable and discoverable on a desktop system.

5.4 Privacy-First Personalization and Fairness Auditing

A longitudinal study is eventually required to compare attribute-based personalization using zero-party data with a cookie-based recommendation system. This is the need that announced filters can be made equally or more pertinent, and that they can also increase the level of privacy acceptability and transparency [38]. Users will undergo comparative tests over a period of 6-12 months; they will be tested on repeat search, conversion lift, opt-out rate, complaint rate, and even-handedness between novice and expert users.

The governance variables must entail transparency of consent, an explicable model, exposure diversification, and attribute prioritization in an attribute fashion. A federation analysis of AD

and SAML focuses on identity and approval controls that should be secure. When implementing policy to align business enterprises, it can be applied directly to user data profiles, search functionality, and product policies [39]. However, it is impossible to even conceive that, in the near future, an approach that ensures privacy-first personalization will be viewed as a marketing characteristic rather than a reasonable mechanism for establishing trust.

Conclusions

The study indicates that SKU-based attribute mapping improves discovery and business performance in e-commerce. During the investment in art supplies, the apt metadata search helped grasp intent, providing a more complete picture than the superficial domains such as title, brand, and cost. It manifested itself in the search-to-product click-through rate, which went up to 4.2% to 7.8%; the conversion rate, which rose to 1.9% to 3.4%; and the zero-result query rate, which dropped to 21.5% to 5.1%. The findings indicate that when the pigment code, lightfastness, paper weight, texture, type of bristle, or type of ink base is used as a retrieval field, the product meaning is not latent in the description; that is, it is actionable. Attributes are the most effective to search with because they are first-class data objects that can be manually ranked, faceted, filtered, and supported by decision-making.

Another contribution to the theory in the research is the use of search engine optimization, product information management, information retrieval, and consumer conversion as a single analysis outcome. It proves that metadata richness can be not only a cataloguing quality setting but also a strategy variable that defines the exploitation of the significance of purpose and attainment of relevance, trust, and transaction value. The research proceeds to theorize the granularity of the SKU stage in an attempt to transcend the generic discussion of personalization or the usability of search, and to recognize that the richness of attributes is a causal factor in the success or failure of digital commerce. It also shows that the query located in the long tail, which is the type of query initiated by experts, is not satisfactorily resolved by lexical matching. Quite the contrary, they require product-specific ontologies, taxonomies, and well-structured search logic to reduce product complexity into meaningful, searchable information. The study contributes to a realistic theory of merchandising, according to which the quality of retrieval depends on the accuracy of semantic retrieval.

In terms of managerial implications, the findings indicate that the retailer should invest in attribute governance and the attribute category, as these provide a category taxonomy, PIM and search integrations, AI-driven attribute enrichment, and privacy-conscious personalization. This form of investment should be in technical, regulated, or expert markets, where buyers are assured of product compatibility, composition, and performance to make a purchase. The findings also indicate that the processes involved in supporting the scalable operations of a catalog include dynamic taxonomy, extraction of manufacturer information from PDFs or packaging, and zero-party personalization via filter behavior. Although swayed towards 100,000 SKU benefits, the lack of good metadata best practices will cause attribute dilution, re-indexing, and dilution of search results. On the other hand, SKU attributes can enhance search relevancy, minimize ambiguity of cadmium-free and cadmium red, and help to rank, comply, and make procurement decisions.

The study concludes that contemporary e-commerce performance is founded on the granularity of product meaning. Retailers' models develop products at the SKU attribute level, which not only improves search mechanisms but also enhances product recommendations. They create a valid discovery area, improve decision-making, reinforce the utility of landing pages, and increase revenue efficiency throughout the buying process. Mapping of attributes at the SKU level should therefore be viewed as a search-relevance strategy and a viable retail commercial structure..

References

- M. Appiani, C. Cattaneo, and M. Laureati, "Sensory properties and consumer acceptance of plant-based meat, dairy, fish and eggs analogs: A systematic review," *Frontiers in Sustainable Food Systems*, vol. 7, Art. no. 1268068, 2023. <https://www.frontiersin.org/journals/sustainable-food-systems/articles/10.3389/fsufs.2023.1268068/full>
- A. Allahim, A. Cherif, and A. Imine, "Semantic approaches for query expansion: Taxonomy, challenges, and future research directions," *PeerJ Computer Science*, vol. 11, e2664, 2025. <https://peerj.com/articles/cs-2664.pdf>
- Abmatic AI, "The role of personalization in conversion rate optimization," 2025. <https://abmatic.ai/blog/role-of-personalization-in-conversion-rate-optimization>
- Amazon Buy with Prime, "5 steps for getting ecommerce personalization right," 2025. <https://buywithprime.amazon.com/blog/5-steps-for-getting-ecommerce-personalization-right>
- C. Cavicchi and E. Vagnoni, "Digital information systems in support of accountability: The case of a welfare provision non-governmental organisation," *The British Accounting Review*, vol. 55, no. 5, Art. no. 101112, 2023.
- S. van Engelenburg, M. Janssen, and B. Klievink, "Design of a software architecture supporting business-to-government information sharing to improve public safety and security," *Journal of Intelligent Information Systems*, vol. 52, no. 3, pp. 595-618, 2019.
- N. Kumar, "Accessibility first design systems for multi-brand food apps," *AJT*, 2025. <https://doi.org/10.58425/ajt.v4i3.457>
- S. K. Vishwakarma, "AI-driven predictive risk modelling for aerospace supply chains," *International Journal of Innovation in Business & Economics and Applied Journal*, 2025. <https://www.iibajournal.org/index.php/iibeaj/article/view/64>
- M. Becker, "Privacy in the digital age: Comparing and contrasting individual versus social approaches towards privacy," *Ethics and Information Technology*, vol. 21, no. 4, pp. 307-317, 2019.
- IAPP, "Consumer perspectives of privacy and artificial intelligence," 2025. <https://iapp.org/resources/article/consumer-perspectives-of-privacy-and-ai/>
- L. Lungren and M. A. Okezue, "Metadata analysis for packaging seal integrity relationships to medicinal product recalls," *AAPS Open*, vol. 11, no. 1, p. 30, 2025. <https://link.springer.com/content/pdf/10.1186/s41120-025-00134-2.pdf>
- A. Ubale, "Privacy-preserving federated learning frameworks for telematics data in auto insurance analytics," *IJAIDSML*, vol. 6, no. 4, pp. 41-48. <https://ijaidsmml.org/index.php/ijaidsmml/article/view/314>
- V. K. Enugala, "Quantum sensors for micro-corrosion detection," *IJCESN*, 2025. <https://ijcesen.com/index.php/ijcesen/article/view/3481/995>

K. S. Chadha, "Leveraging advanced analytics and AI-powered solutions for transforming healthcare and retail e-commerce: Enhancing data security, personalization, and compliance," *International Journal of Applied Mathematics*, vol. 38, no. 9s, pp. 1310-1334, 2025. <https://doi.org/10.12732/ijam.v38i9s.862>

Qualtrics, "What is customer lifetime value (CLV) and how can you increase it?," 2025. <https://www.qualtrics.com/experience-management/customer/customer-lifetime-value/>

K. Khadka and S. Maharjan, *Customer Satisfaction and Customer Loyalty: Case Trivsel Städtjänster*, unpublished master's thesis, 2017. <https://www.scirp.org/reference/referencespapers?referenceid=3649849>

P. Gannavarapu, "Cloud infrastructure management and automation," *AJT Journal*, 2025. <https://gprjournals.org/journals/index.php/ajt/article/view/356>

S. K. RaoVanama, "AI-augmented CI/CD pipeline optimization for scalable cloud-native deployment," *IJAIDSML*, vol. 5, no. 4, pp. 175-187, 2024. <https://ijaidsm.org/index.php/ijaidsm/article/view/368>

P. P. Ray, "A review of TRiSM frameworks in artificial intelligence systems: Fundamentals, taxonomy, use cases, key challenges and future directions," *Expert Systems*, vol. 43, no. 3, e70213, 2026. <https://www.techrxiv.org/doi/pdf/10.36227/techrxiv.174913612.20443736>

A. H. Juma'h and Y. Alnsour, "The effect of data breaches on company performance," *International Journal of Accounting & Information Management*, vol. 28, no. 2, pp. 275-301, 2020.

https://www.researchgate.net/publication/335002124_The_Effect_of_Data_Breaches_on_Company_Performance

L. Staufer, K. Feng, K. Wei, L. Bailey, Y. Duan, M. Yang, et al., "The 2025 AI Agent Index: Documenting technical and safety features of deployed agentic AI systems," arXiv preprint arXiv:2602.17753, 2026. <https://www.igi-global.com/chapter/from-click-to-conscious-choice/404370>

M. E. U. Rahman and A. Yelishetty, "Midway advertisement: A mechanism to curb annoyance due to unwanted advertisements," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES)*, 2021, pp. 1-5.

B. Abbasgholinejadabaghi, "Metadata-driven optimization in electron microscopy: Improving SEM imaging data through FAIR practices," thesis, 2025. https://iris.sissa.it/bitstream/20.500.11767/148470/1/Abbasgholinejadabaghi_Behnaz_Thesis_MDMC2024-2025.pdf

S. Samala, "Real-time Jira analytics: Integrating JQL with Power BI/Snowflake for predictive agile metrics," *SciPubHouse*, 2024. <https://scipubhouse.com/home/international-journal-of-sustainability-and-innovation-in-engineering-ijsie/content/ijsie-2024/real-time-jira-analytics-integrating-jql-with-power-bi-snowflake-for-predictive-agile-metrics/>

H. Zhang, Y. Li, J. Zhuang, J. Dai, Z. L. Xiu, and C. Quan, "One-pot synthesis of fuel precursor from acetoin fermentation broth using ionic liquid-based salting-out extraction system," *Biotechnology for Biofuels and Bioproducts*, vol. 16, no. 1, p. 94, 2023. <https://link.springer.com/content/pdf/10.1186/s13068-023-02344-w.pdf>

N. Kumar, "Edge computing patterns for real-time order flow optimization," *International Journal of Applied Mathematics*, 2025. <https://doi.org/10.12732/ijam.v38i12s.1494>

P. Chopra, V. Singhal, J. Nirmal, and P. Singh, "Cadmium toxicity presenting as multiorgan dysfunction of uncertain etiology—A diagnostic dilemma: Case report," *Indian Journal of Critical Care Case Report*, vol. 4, no. 2, pp. 38–40, 2025. https://www.scienceopen.com/document_file/52f429e0-8b66-4ae8-b944-3e1b0ab9deec/API/ijccr-04-038.pdf

W. Zhou, D. Huang, S. Chen, G. Wang, R. Li, W. Xu, et al., "Microplastic dilemma: Assessing the unexpected trade-offs between biodegradable and non-biodegradable forms on plant health, cadmium uptake, and sediment microbial ecology," *Journal of Hazardous Materials*, vol. 477, p. 135240, 2024.

A. C. Jha, "Automated firewall policy generation with reinforcement learning," *IJIOT*, 2025. <https://www.academicpublishers.org/journals/index.php/ijiot/article/view/5483>

D. H. Shastri, S. Gandhi, and H. Almeida, "Enhancing collaboration and interdisciplinary strategies for navigating innovative technologies and regulatory approvals in the cosmetic industry," *Current Cosmetic Science*, vol. 3, no. 1, E26667797324383, 2024.

Z. Sayyed, "Development of a simulator to mimic VMware vCloud Director (VCD) API calls for cloud orchestration testing," *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 3, pp. 5422-5441, 2025. <https://doi.org/10.22399/ijcesen.3480>

V. Nagaraj, "Addressing power efficiency challenges in AI hardware through verification," *SciPubHouse*, 2024. <https://scipubhouse.com/home/international-journal-of-sustainability-and-innovation-in-engineering-ijisie/content/ijisie-2024/addressing-power-efficiency-challenges-in-ai-hardware-through-verification/>

R. K. K. Lingamgunta, A. Ubale, and S. K. R. Vanama, "Edge AI for on-site health risk scoring: A RAG-enabled framework," *American Journal of Technology*, vol. 4, no. 3, pp. 1–14, 2025. <https://doi.org/10.58425/ajt.v4i3.451>

B. Blinov, "Automation of product decision-making based on A/B testing," *Emerging Frontiers Library for The American Journal of Engineering and Technology*, vol. 7, no. 7, pp. 24–32, 2025. <http://emergingsociety.org/index.php/efltajet/article/view/250>

E. Ali, A. Caputo, and G. J. Jones, "A comprehensive survey of facet ranking approaches used in faceted search systems," *Information*, vol. 14, no. 7, p. 387, 2023. <https://www.mdpi.com/2078-2489/14/7/387>

S. Silva, P. Pelliccione, and A. Bertolino, "Self-adaptive testing in the field," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 19, no. 1, pp. 1–37, 2024. <https://dl.acm.org/doi/pdf/10.1145/3627163>

S. Rangu, “Enterprise digital transformation in financial services: Emerging trends and technologies,” *Computer Fraud & Security*, 2025. <https://computerfraudsecurity.com/index.php/journal/article/view/786>

X. Li, “Privacy-preserving feature attribution explanations for large-scale recommendation systems: A differential privacy approach,” *Journal of Science, Innovation & Social Impact*, vol. 1, no. 1, pp. 19–32, 2025. <http://sagespress.com/index.php/JSISI/article/download/9/3>

P. Gannavarapu, “Deploying Azure AD federation with SAML for secure enterprise SaaS integration,” *Computer Fraud & Security*, 2025. <https://computerfraudsecurity.com/index.php/journal/article/view/782..>