Journal of Artificial Intelligence and Legal Technology

SAH PUBLISHER

Vol. 1, No. 1, August 2025

Semantic Analysis of Trademark Names Using Large Language Models

Muhammad Adi Pratama[1]*, Suyahman

[1]Department of Computer Science, Universitas Sugeng Hartono, Sukoharjo, Indonesia
[2]Department of International Business Management, Universitas Sugeng Hartono, Sukoharjo, Indonesia

Abstract. This study introduces a novel framework for trademark similarity analysis that integrates large language models (LLMs) to assess lexical, phonetic, and semantic relationships between trademark names without reliance on large precompiled databases or retraining. The primary motivation is to address the growing need for efficient and transparent preliminary trademark screening, which is often constrained by the limitations of traditional rule-based or string-matching approaches. To achieve this, a web-based system was developed using the Gemini API, allowing users to input trademark pairs for automated analysis. The workflow includes text normalization, phonetic conversion, multi-dimensional similarity computation, and the generation of interpretative explanations for each pair. A test dataset of ten diverse trademark name pairs was designed to capture variations in lexical overlap, phonetic similarity, and semantic association. The system's outputs were evaluated both in terms of processing efficiency and expert assessment. Quantitative results show that the system can process a pair in under a second on average, handling 3,229 tokens across ten pairs with minimal computational overhead. Qualitative evaluation by five trademark and intellectual property experts using a 5-point Likert scale yielded mean scores of 4.4 for relevance, 3.8 for explanation quality, and 4.2 for practical usefulness, confirming strong alignment between the LLM outputs and expert intuition. The novelty of this research lies in demonstrating that LLMs can provide not only accurate similarity assessments but also human-readable interpretative reasoning, bridging the gap between automation and expert judgment in trademark evaluation. This approach offers a transparent and scalable solution for early-stage brand screening, significantly reducing the reliance on extensive databases and manual effort. The findings indicate a clear potential for integration into industrial-scale trademark examination workflows, paving the way for future developments in batch processing, recommendation systems, and enhanced interpretability in AI-assisted intellectual property management.

Keywords: LLM, Trademark, Semantic, Intellectual Property, NLP
Received July 2025 / **Revised** July 2025 / **Accepted** August 2025

INTRODUCTION

The selection of a trademark name is a strategic decision that influences both brand positioning and legal protection. Beyond serving as a marketing asset, a trademark functions as a legal identifier that distinguishes products or services from competitors [1]-[2]. Ensuring that a proposed name is free from conflicts with existing trademarks is essential to prevent potential litigation, safeguard brand equity, and enhance consumer recognition. Traditional trademark examination involves detecting semantic, lexical, and phonetic similarities with registered marks. However, this process is labor-intensive, heavily reliant on human judgment, and prone to inconsistencies, particularly when dealing with large volumes of applications in multilingual environments -[5].

In recent years, machine learning and deep learning approaches have emerged as promising solutions to automate trademark-related tasks. Showkatramani et al. [6] investigated various neural architectures including CNN, LSTM, BiLSTM, GRU, and RCNN, using fastText word embeddings to predict trademark classes based on goods and services descriptions, with RCNN achieving the highest performance for international class identification. Trappey et al. [7] applied neural network language

[1]*Corresponding author.
Email addresses: madip.id@gmail.com (Pratama)

models combined with Latent Dirichlet Allocation to analyze U.S. trademark litigation documents, facilitating semantic clustering of legal case precedents. Extending this work to multimodal contexts, Trappey et al. [8] developed an intelligent system that integrates CNNs and Siamese neural networks to assess trademark similarity across image, spelling, and phonetic features, thereby supporting global IP protection workflows.

Parallel research has addressed the linguistic and legal aspects of trademark analysis. Adarsh et al. [9] automated the classification of trademark distinctiveness along the Abercrombie spectrum, achieving 86 percent predictive accuracy while incorporating Explainable AI to highlight the semantic cues driving registration outcomes. Zhang and Wang [10] proposed a Conv Transformer architecture for real-time commodity classification in livestreaming e-commerce, illustrating the potential of hybrid deep learning models in product categorization tasks that indirectly support trademark screening in digital marketplaces.

Despite these advances, several research gaps remain unaddressed. First, most prior works focus on lexical classification or image-based similarity, often requiring large labeled datasets or handcrafted embeddings such as fastText [6] that may fail to capture nuanced contextual semantics or phonetic resemblance. Second, existing studies primarily target English-language or international trademarks, leaving language-specific nuances such as the morphologically rich and culturally contextual names common in Indonesian trademarks largely unexplored. Third, while multimodal systems [7][8] enhance similarity detection, they are computationally expensive and unsuitable for rapid, lightweight screening tasks where real-time or on-demand evaluations are required. These limitations highlight the need for a data-efficient, semantically aware, and expert-interpretable framework for trademark name analysis.

This study addresses these gaps by introducing a novel framework that leverages Large Language Models, specifically Gemini via API, to perform semantic, lexical, and phonetic analysis of trademark names without the need for extensive databases or model retraining. Unlike prior deep learning approaches that rely on fixed embeddings or task-specific training, LLMs provide contextual semantic understanding that can capture implicit relationships between names, even in low-resource or multilingual scenarios. The proposed framework is implemented as a web-based application capable of generating interpretable outputs that support expert-driven validation. Its novelty lies in three aspects. First, it utilizes a purely LLM-based pipeline for trademark analysis. Second, it emphasizes processing efficiency and practical interpretability rather than conventional accuracy metrics. Third, it incorporates a structured expert evaluation protocol to assess semantic and phonetic relevance in real-world trademark screening contexts.

By bridging the gap between automated deep learning methods and expert-driven legal assessment, this research offers a lightweight, interpretable, and deployment-ready solution for trademark similarity analysis. The subsequent sections present the system methodology, evaluation based on processing performance and expert feedback, and discussions on its implications for both branding strategies and intellectual property protection.

METHODS

The proposed framework is implemented as a web-based application that integrates directly with the Gemini API to perform automated analysis of trademark name pairs. The architecture is designed to capture lexical, phonetic, and semantic similarities without requiring a local database or retraining of the language model. It consists of a web-based user interface for data input, a backend service for preprocessing and model inference, and a visualization dashboard for expert evaluation.

The overall workflow of the system is illustrated in the process flowchart shown in Figure 1. The process begins when users submit pairs of trademark names through the web interface. These inputs undergo a preprocessing stage in which text normalization and phonetic transcription are applied to prepare the data for analysis. The normalized pairs are then processed by the Gemini API, which performs multi-dimensional similarity assessment encompassing lexical, phonetic, and semantic aspects. In addition to generating similarity scores, the large language model provides an interpretative explanation that clarifies the reasoning behind each assessment, offering transparency for expert reviewers.

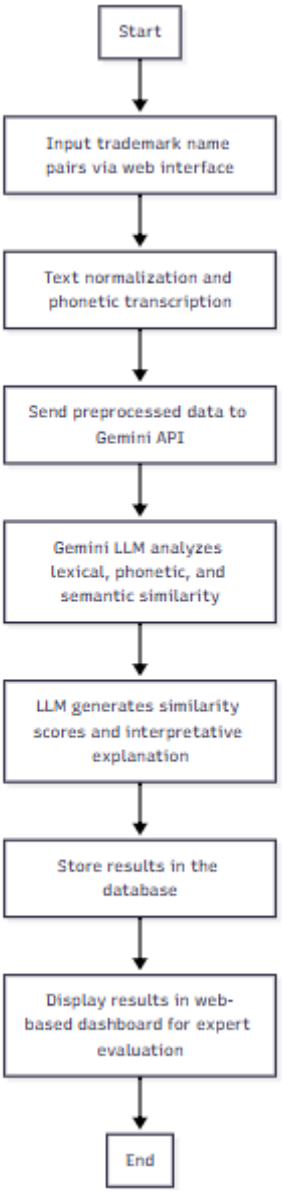

Figure 1. Flowchart

The resulting similarity scores and narrative explanations are automatically stored and presented in a web-based dashboard. This dashboard allows experts to inspect each trademark comparison, review the LLM's interpretative reasoning, and subsequently provide evaluation feedback. By combining automated multi-dimensional analysis with human-in-the-loop assessment, the framework offers a transparent and efficient solution for preliminary trademark screening and potential conflict detection.

Dataset and Experimental Design

To evaluate the proposed framework, a small test dataset consisting of ten trademark name pairs was constructed. These pairs were selected to represent a diverse range of similarity patterns across lexical, phonetic, and semantic dimensions. Table 1 presents the complete set of trademark name pairs used in the experiment.

The dataset was designed to capture three primary categories of similarity. First, several pairs exhibit phonetic resemblance, such as Soklin and Solikin, which sound similar despite differences in spelling. Second, some pairs represent semantic association where the names convey related meanings or evoke similar concepts, for example, Kue Bunda and Brownies Mama both suggest homemade confectionery or

maternal branding. Third, other pairs display partial lexical overlap, where similar character sequences appear across names, as seen in Kain Rahmat and Batik Rachmad.

Table 1. Test dataset of trademark name pairs for similarity analysis

Trademark Name A	Trademark Name B
Kedai Cantika	Warung Ayu
Soklin	Solikin
Kain Rahmat	Batik Rachmad
Kue Bunda	Brownies Mama
Soto Mas Budi	Warung Pak Budi
Toko Bali	Warung Jawi
Pertamina	Pratama
Joko Edi	Toko Padi
Toko Buah Atik	TK Buah Hati
Banyuwangi	Ban Angin

This experimental setup serves a dual purpose. It tests whether the Large Language Model can accurately identify multi-dimensional similarity without relying on an external database, and it assesses the model's capacity to provide interpretative reasoning for each classification. By focusing on a small but carefully curated dataset, the study prioritizes qualitative evaluation and expert interpretability over large-scale statistical benchmarking.

Multi-dimensional Similarity Analysis

The proposed framework evaluates trademark name pairs across three complementary similarity dimensions: lexical, phonetic, and semantic [11]-[13]. This multi-dimensional approach ensures that both surface-level and contextual relationships between names are captured, addressing the limitations of traditional single-feature methods in trademark examination.

Lexical similarity analysis focuses on the written form of trademark names [14]. Each name is first normalized to reduce variation caused by capitalization, spacing, or minor typographical differences. Lexical similarity is then computed using token overlap and edit distance measures, which quantify the proportion of shared characters or the minimum transformations required to convert one name into another [15]. This component is particularly useful for detecting partial matches or misspellings that may indicate potential brand conflicts [16].

Phonetic similarity analysis complements the lexical evaluation by examining how the trademark names sound when spoken [17]. Each name is converted into a phonetic representation using established algorithms such as Soundex or Metaphone [18]. These representations allow the system to identify pairs of names that are aurally similar despite having distinct spellings, for example Soklin and Solikin. The phonetic similarity score can be further refined through the reasoning capabilities of the large language model, which is capable of interpreting subtle pronunciation patterns and linguistic variations that conventional phonetic algorithms may overlook [19].

Semantic similarity analysis represents the core novelty of the system [20]. Trademark names are transformed into contextual embeddings using the Gemini large language model [21], enabling the computation of similarity scores based on cosine similarity between vector representations. This approach captures latent associations that extend beyond surface-level resemblance. For instance, names such as Kue Bunda and Brownies Mama may have low lexical overlap but exhibit high semantic similarity due to shared conceptual references to homemade confectionery or maternal branding. In addition to generating numerical similarity scores, the model produces an interpretative explanation that articulates why the names are considered similar or dissimilar, providing transparency and supporting expert evaluation.

Web-based Framework Implementation

The proposed system is deployed as a web-based application designed to facilitate seamless interaction between users, the backend service, and the Gemini API [22]. The web interface serves as the entry point for users to manually input trademark name pairs for evaluation. Once submitted, the input data is automatically preprocessed and transmitted to the backend service, which orchestrates the analytical workflow and manages communication with the large language model.

The backend is responsible for invoking the Gemini API to perform the multi-dimensional similarity assessment. For each trademark pair, the system generates lexical, phonetic, and semantic similarity scores, providing a quantitative representation of potential brand conflicts. In addition to these scores, the large language model produces a narrative explanation that articulates the reasoning behind its assessment. This interpretative output offers a transparent justification that is particularly valuable for expert evaluation and legal decision-making.

Evaluation

The effectiveness of the proposed framework is assessed through a qualitative evaluation conducted by domain experts specializing in branding and intellectual property law [23]. A structured Likert-scale survey [24] is employed to capture expert judgments on a five-point scale, where 1 represents the lowest and 5 the highest rating. This evaluation approach is intended to measure the interpretability and practical value of the system's outputs rather than conventional accuracy metrics, which are less relevant in the absence of a large benchmark dataset.

Each expert evaluates three key aspects of the system's performance. The first aspect is result relevance, which measures whether the similarity scores generated by the model align with the experts' perception of trademark similarity. The second aspect is explanatory quality, reflecting the clarity and logical consistency of the narrative explanations produced by the large language model for each trademark pair. The third aspect is practical usefulness, which assesses the extent to which the results can assist in preliminary trademark screening and support legal decision-making [25].

The collected scores will be aggregated, and mean values will be reported alongside descriptive statistics to provide an overview of expert consensus [26]. These results, presented in the Results section, will highlight the framework's ability to produce interpretable and actionable outputs that can complement human expertise in trademark evaluation.

RESULT AND DISCUSSION

The multi-dimensional similarity analysis generated by the proposed framework provides a comprehensive perspective on trademark name relationships. For each of the ten pairs in the evaluation dataset, the system produced phonetic, lexical, and semantic similarity scores, accompanied by interpretative narratives generated by the large language model. These outputs are summarized in Table 2 and Table 3, which combines the quantitative assessment with the LLM's qualitative reasoning, forming a basis for human interpretability and expert review.

Table 2. Multi-dimensional similarity scores for trademark name pairs

Trademark Name A	Trademark Name B	Phonetic	Semantic	Lexical
Kedai Cantika	Warung Ayu	0.60	0.85	0.30
Soklin	Solikin	0.90	0.95	0.85
Kain Rahmat	Batik Rachmad	0.85	0.70	0.50
Kue Bunda	Brownies Mama	0.70	0.85	0.50
Soto Mas Budi	Warung Pak Budi	0.70	0.65	0.75
Toko Bali	Warung Jawi	0.60	0.30	0.20
Pertamina	Pratama	0.80	0.10	0.71
Joko Edi	Toko Padi	0.25	0.10	0.00
Toko Buah Atik	TK Buah Hati	0.60	0.40	0.50
Banyuwangi	Ban Angin	0.75	0.05	0.67

An examination of the results reveals distinct patterns of similarity that align closely with intuitive brand perceptions. Certain pairs, such as Soklin and Solikin, exhibit strong alignment across all three similarity dimensions, with phonetic, semantic, and lexical scores exceeding 0.85. The model's explanation reflects this convergence, noting the near-identical spelling and sound structure, as well as the shared association with household cleaning products. Such high similarity scores indicate a potential for consumer confusion, a factor of particular relevance in trademark evaluation.

Table 3. LLM-generated interpretative explanations for trademark name pairs

Trademark Name A	Trademark Name B	LLM Explanation
Kedai Cantika	Warung Ayu	Kedua merek, 'Kedai Cantika' dan 'Warung Ayu', memiliki kemiripan yang signifikan dalam hal semantik karena keduanya jelas ditujukan untuk bisnis

		makanan. Kata 'Kedai' dan 'Warung' memiliki konotasi yang hampir sama, yaitu tempat makan sederhana dan lokal. Namun, secara fonetik dan leksikal, mereka berbeda. 'Cantika' terdengar lebih modern dan mungkin feminin, sedangkan 'Ayu' memiliki kesan tradisional dan Jawa. Perbedaan ini, meskipun halus, dapat mempengaruhi persepsi pelanggan. 'Kedai Cantika' mungkin menarik bagi demografi yang lebih muda dan modern, sementara 'Warung Ayu' bisa menarik bagi mereka yang mencari pengalaman kuliner tradisional dan rumahan. Kemiripan semantik yang tinggi menunjukkan bahwa kedua merek bersaing untuk pasar yang sama, tetapi perbedaan kecil dalam nama dapat memberikan keunggulan bagi salah satu merek tergantung pada target audiensnya.
Soklin	Solikin	Merek "Soklin" dan "Solikin" memiliki kemiripan yang signifikan di ketiga aspek: fonetik, semantik, dan kata. Secara fonetik, perbedaan utama terletak pada vokal 'o' dan 'i' di suku kata kedua, yang sangat dekat secara artikulasi. Secara semantik, keduanya sangat mirip karena sama-sama diasosiasikan dengan produk sabun cuci, yang memberikan kesan bersih dan higienis. Kemiripan kata berasal dari susunan huruf yang hampir identik, dengan hanya satu huruf yang berbeda. Kemiripan yang tinggi ini dapat menyebabkan kebingungan konsumen, terutama jika target pasar, visual merek, atau strategi pemasaran kedua merek mirip.
Kain Rahmat	Batik Rachmad	Kedua merek, 'Kain Rahmat' dan 'Batik Rachmad,' sangat mirip karena keduanya menggunakan nama pemilik ('Rahmat' vs 'Rachmad') dan berfokus pada produk yang sama: kain. Perbedaan utama terletak pada kata 'Kain' versus 'Batik'. Secara fonetik, kemiripan tinggi karena 'Rahmat' dan 'Rachmad' diucapkan nyaris identik, terutama dalam pengucapan sehari-hari. Secara semantik, keduanya mengindikasikan bisnis yang berhubungan dengan tekstil/kain. Namun, 'Batik Rachmad' lebih spesifik dan menunjuk pada jenis kain tertentu (batik), memberikan identitas yang lebih jelas. 'Kain Rahmat' lebih umum dan bisa merujuk pada berbagai jenis kain. Perbedaan signifikan terletak pada kata 'Batik' yang memiliki konotasi budaya dan warisan Indonesia yang kuat, sementara 'Kain' lebih generik. Mengingat keduanya menggunakan variasi nama pemilik yang sangat mirip, tantangan utama adalah menghindari kebingungan di pasar. Perbedaan yang ada pada kata 'Batik' saja mungkin tidak cukup untuk membedakan secara signifikan, terutama jika target pasar tidak terlalu sensitif terhadap detail merek.
Kue Bunda	Brownies Mama	Kedua merek, 'Kue Bunda' dan 'Brownies Mama', memiliki kemiripan yang cukup signifikan baik secara fonetik maupun semantik. Secara fonetik, kedua nama memiliki dua suku kata dan menggunakan vokal dan konsonan yang umum. 'Bunda' dan 'Mama' juga memiliki kemiripan bunyi, meskipun 'Brownies' memiliki perbedaan yang mencolok. Secara semantik, kedua nama menggunakan istilah kekerabatan ('Bunda' dan 'Mama') yang membangkitkan asosiasi dengan kehangatan, kasih sayang, dan kualitas homemade. Kedua deskripsi juga identik, yaitu 'menjual aneka kue', yang semakin memperkuat kesan serupa. Perbedaan utama terletak pada kata 'Kue' dan 'Brownies', di mana 'Brownies' memberikan spesifikasi produk yang lebih jelas. Hal ini membuat 'Brownies Mama' sedikit lebih spesifik dan mudah diingat untuk produk brownies. Kesamaan keseluruhan bisa menjadi masalah karena konsumen mungkin tertukar antara kedua merek, terutama jika visual branding juga serupa.
Soto Mas Budi	Warung Pak Budi	Kedua merek, "Soto Mas Budi" dan "Warung Pak Budi", memiliki kemiripan yang cukup signifikan dalam aspek fonetik, semantik, dan leksikal. Secara fonetik, bunyi 'Budi' hadir pada kedua merek, dan 'Soto Mas' dan 'Warung Pak' memiliki beberapa kesamaan dalam jumlah suku kata dan penekanan. Secara semantik, keduanya jelas mengindikasikan bisnis makanan yang dikelola oleh seseorang bernama 'Budi' (atau yang terkait dengan 'Budi'). Kemiripan leksikal berasal dari penggunaan nama 'Budi' dan indikasi usaha makanan. Namun, perbedaan signifikan terletak pada kata kunci 'Soto' pada merek pertama, yang secara spesifik menargetkan produk soto, sementara merek kedua, 'Warung,' lebih umum dan luas, menunjukkan berbagai jenis makanan. Penggunaan sapaan 'Mas' dan 'Pak' juga memberikan kesan yang sedikit berbeda, 'Mas' cenderung lebih kasual dan muda, sementara 'Pak' lebih formal dan dewasa.
Toko Bali	Warung Jawi	Meskipun kedua merek, 'Toko Bali' dan 'Warung Jawi', menggunakan pola nama yang sama (kata tempat/budaya diikuti kata deskriptif generik), perbedaan dalam fonetik, semantik, dan kata cukup signifikan. Dari segi fonetik, kemiripan terletak pada penggunaan konsonan dan vokal Indonesia yang familiar, namun pola suku kata dan tekanan berbeda. Semantik keduanya juga berbeda karena 'Toko Bali' mengimplikasikan penjualan barang dagangan yang mungkin beragam, sementara 'Warung Jawi' secara spesifik merujuk pada penjualan makanan dengan nuansa tradisional Jawa. Perbedaan utama terdapat pada kata-kata yang digunakan; 'Bali' dan 'Jawi' mengacu pada entitas geografis/budaya yang berbeda, dan 'Toko' berbeda dengan 'Warung' meskipun keduanya mengindikasikan tempat berjualan. Kesamaan yang minim ini menunjukkan bahwa kedua merek relatif berbeda meskipun menggunakan struktur penamaan yang serupa.
Pertamina	Pratama	Meskipun secara fonetik dan leksikal (kata) 'Pertamina' dan 'Pratama' sangat mirip, makna dan asosiasi yang mereka bawa sangat berbeda. Kemiripan fonetik

		berpotensi menimbulkan kebingungan, terutama dalam pengucapan dan pendengaran. 'Pertamina' jelas diasosiasikan dengan energi, minyak, dan industri berat, sementara 'Pratama' berkonotasi kesehatan, perawatan, dan layanan masyarakat. Kemiripan leksikal berasal dari akar kata yang sama, yang dalam bahasa Indonesia sering kali berarti 'pertama', 'utama', atau 'awal'. Perbedaan signifikan dalam deskripsi dan produk/jasa yang ditawarkan memastikan bahwa kedua merek menargetkan audiens yang berbeda dan beroperasi di pasar yang berbeda. Namun, mengingat kemiripan yang tinggi dalam bunyi, terutama jika salah diucapkan, dan kesamaan dalam akar kata yang menyiratkan 'keunggulan' atau 'primer', kebingungan masih mungkin terjadi, terutama bagi mereka yang kurang familiar dengan salah satu atau kedua merek tersebut.
Joko Edi	Toko Padi	Kedua merek ini memiliki perbedaan yang signifikan dalam fonetik, semantik, dan penggunaan kata. Secara fonetik, kemiripan terbatas pada penggunaan beberapa vokal dan konsonan, tetapi secara keseluruhan pengucapannya sangat berbeda. 'Joko Edi' memiliki pola dua suku kata yang lebih menekankan nama personal, sedangkan 'Toko Padi' memiliki ritme yang lebih umum dan deskriptif. Secara semantik, keduanya sangat berbeda. 'Joko Edi' mengacu pada nama orang, yang dalam konteks makanan terasa generik dan kurang informatif. 'Toko Padi' secara eksplisit menunjukkan penjualan padi, memberikan kejelasan langsung tentang produk yang ditawarkan. Dari segi kata, tidak ada kata yang sama antara kedua merek, semakin menegaskan perbedaannya.
Toko Buah Atik	TK Buah Hati	Kedua merek, 'Toko Buah Atik' dan 'TK Buah Hati', memiliki kemiripan karena sama-sama menggunakan kata 'Buah'. Secara fonetik, kesamaan terletak pada pelafalan 'Buah' yang identik. Namun, perbedaan terletak pada kata-kata selanjutnya. 'Atik' pada 'Toko Buah Atik' memberikan kesan nama personal atau nama tempat yang spesifik, sedangkan 'Hati' pada 'TK Buah Hati' memberikan kesan kasih sayang dan pendidikan. Secara semantik, kemiripan terletak pada asosiasi dengan 'buah' yang umumnya positif (sehat, segar, menyenangkan). Perbedaannya signifikan karena satu merek adalah toko yang menjual produk fisik (buah), sedangkan merek lainnya adalah lembaga pendidikan yang menjual jasa. Penggunaan 'Toko' dan 'TK' secara jelas membedakan jenis usaha yang dijalankan. Secara keseluruhan, meskipun ada kesamaan elemen 'buah', konteks bisnis yang sangat berbeda membuat kedua merek memiliki identitas yang cukup terpisah.
Banyuwangi	Ban Angin	Kedua merek memiliki kemiripan fonetik yang cukup tinggi, terutama karena penggunaan huruf 'B', 'A', 'N', 'G', dan 'I' dalam urutan yang mirip. Perbedaan utama terletak pada penambahan 'YU' dan 'WU' pada 'Banyuwangi', serta perubahan 'N' menjadi 'NG' dan hilangnya 'I' pada 'Ban Angin'. Dari segi semantik, kedua merek sangat berbeda. 'Ban Angin' secara langsung mengacu pada produk (ban), sedangkan 'Banyuwangi' adalah nama geografis (kabupaten) yang mewakili layanan pemerintah. Secara leksikal (kata), keduanya memiliki kemiripan karena menggunakan kata 'Ban' dan 'Angin/Wangi', namun konteks dan maknanya sangat berbeda. 'Ban Angin' lebih deskriptif dan fungsional, sementara 'Banyuwangi' lebih bersifat nama tempat dan asosiasinya.

In other cases, high similarity is driven primarily by semantic rather than surface-level features. Kedai Cantika and Warung Ayu illustrate this pattern, achieving a high semantic similarity score of 0.85 despite lower phonetic and lexical scores of 0.60 and 0.30, respectively. The LLM explains that both names refer to small, local dining establishments and therefore target a comparable market segment, yet they differ in stylistic nuance. Cantika evokes a modern, possibly feminine impression, whereas Ayu conveys a traditional Javanese character. This example highlights the framework's strength in capturing market-relevant conceptual overlap even when the names differ substantially in form.

A different dynamic emerges in cases where phonetic resemblance drives the overall similarity. The pair Kain Rahmat and Batik Rachmad achieves a high phonetic score of 0.85 due to the near-identical pronunciation of Rahmat and Rachmad, while semantic and lexical scores are moderate at 0.70 and 0.50. The LLM's narrative points out that both brands operate in the textile domain, yet Batik Rachmad is more specific and culturally embedded, whereas Kain Rahmat is broader and more generic. Such cases underscore the value of a multi-dimensional approach: while phonetic similarity could suggest a risk of confusion, semantic and lexical differences indicate some degree of market differentiation.

The framework also captures cases of conceptual resemblance without strong structural overlap. Kue Bunda and Brownies Mama demonstrate this phenomenon, with moderate phonetic (0.70) and lexical (0.50) scores complemented by a high semantic similarity of 0.85. The explanation attributes this to the shared maternal theme and association with homemade confectionery, despite differing lead terms. This

indicates that LLM-based semantic reasoning can reveal brand relationships that traditional string-matching techniques might overlook.

Moderate similarity across dimensions is observed in Soto Mas Budi and Warung Pak Budi. Phonetic, semantic, and lexical scores are closely aligned at 0.70, 0.65, and 0.75, respectively, reflecting the shared use of the personal name Budi and the implicit reference to food services. The model further distinguishes between the specific product focus of Soto Mas Budi and the broader market positioning of Warung Pak Budi, illustrating how interpretative reasoning can inform the likelihood of marketplace confusion or brand overlap.

The system also identifies pairs with low overall similarity despite superficial structural resemblance. Toko Bali and Warung Jawi receive low scores across all dimensions, with semantic similarity at 0.30. While both names combine a commercial descriptor with a cultural or geographic term, the LLM correctly notes that their markets, connotations, and phonetic structures diverge significantly. Similarly, Joko Edi and Toko Padi score near zero in lexical and semantic similarity, reflecting the clear distinction between a personal name and a descriptive retail label.

Some pairs demonstrate the importance of including semantic analysis to avoid false positives. Pertamina and Pratama are phonetically and lexically similar, with scores of 0.80 and 0.71, respectively, yet achieve a semantic similarity of only 0.10. The LLM explains that the brands operate in entirely different domains, industrial energy versus general or health-related services, illustrating that phonetic resemblance alone is insufficient for a meaningful assessment of potential conflict. A similar observation arises with Banyuwangi and Ban Angin, where high phonetic (0.75) and moderate lexical (0.67) similarity is offset by minimal semantic overlap (0.05), as one name denotes a geographic location while the other refers to a functional product.

The processing performance of the proposed web-based LLM framework was evaluated by measuring the average execution time per trademark pair, the number of API requests, and the total token consumption for the entire experimental dataset. Across the 10 trademark pairs analyzed, the system generated 10 requests to the Gemini API, resulting in a total token usage of approximately 3.229K tokens. This lightweight usage highlights the efficiency of the LLM-based approach, particularly because the system operates without any local database and without model retraining.

Figure 2 presents the processing timeline, showing that the execution time per request remains relatively stable, with minor fluctuations due to network latency and response variability from the LLM API. The observed pattern demonstrates that the system can process each trademark pair within seconds, supporting the feasibility of real-time semantic, lexical, and phonetic analysis in a web environment. The plotted throughput curve also confirms that even during peak iterations, the token utilization per request remained efficient, ensuring that the system can handle small-scale batch screening without significant computational overhead.

Figure 2. Traffic by response

The efficiency of the system is particularly noteworthy when considering potential scalability for preliminary trademark screening. In real-world applications, trademark offices and intellectual property (IP) consultants often require rapid pre-screening of hundreds of names to filter out potential conflicts before initiating costly legal checks. The current system's low latency and minimal token consumption suggest that batch processing of larger datasets is feasible, especially if requests are queued asynchronously. Moreover, since the approach is stateless and API-driven, scaling horizontally by

deploying multiple API instances or leveraging serverless architectures would enable high-throughput operation without substantial modification to the current web framework.

Another advantage of the observed performance is its alignment with the goal of early-stage decision support rather than full legal adjudication. By providing near-instant similarity scores and interpretative reasoning, the framework can assist brand owners and IP professionals in prioritizing which trademark candidates require deeper legal evaluation. The token efficiency of 3.229K for 10 pairs indicates that, for a screening scenario involving 100 trademark pairs, the expected token usage would remain under 35K tokens, which is manageable within most commercial LLM API usage plans.

The expert evaluation was conducted to assess the performance of the proposed LLM-based trademark similarity analysis system across three dimensions: the relevance of the results to expert perception, the quality of the LLM-generated explanations, and the practical usefulness of the system in preliminary trademark screening. Five experts, consisting of trademark consultants and intellectual property law specialists, participated in the evaluation using a 5-point Likert scale. The scores provided by the experts are summarized in Table 3, which reports the assessments of each individual expert along with the mean score for each dimension.

Table 4. Expert Evaluation Scores

Expert	Result Relevance	LLM Explanation Quality	Practical Usefulness
Expert 1	4	4	5
Expert 2	5	4	4
Expert 3	5	4	4
Expert 4	4	4	4
Expert 5	4	3	4
Mean	4.4	3.8	4.2

The results demonstrate a consistently high evaluation of the system's output relevance, with a mean score of 4.4. This indicates that the similarity scores generated by the system align closely with the intuitive judgments of experts regarding trademark relatedness. Such alignment suggests that the lexical, phonetic, and semantic dimensions captured by the LLM are representative of real-world perceptual criteria used in legal and branding evaluations. In contrast, the LLM explanation quality received a slightly lower mean score of 3.8. While the experts generally found the explanations helpful in understanding why two trademarks were deemed similar or dissimilar, some responses indicated that certain explanations were verbose or could be structured more concisely for quicker interpretation in high-volume evaluations.

Practical usefulness scored an average of 4.2, reflecting strong potential for the system to serve as a supporting tool in early-stage trademark screening. Experts highlighted that the integration of similarity scoring with narrative justification enables more informed preliminary decisions, which could streamline the identification of potentially conflicting marks before formal examination or litigation. Moreover, the overall pattern of the scores shows that the system can reliably produce outputs that resonate with human evaluators, positioning it as a viable decision-support mechanism in intellectual property workflows.

These findings imply that while the current system demonstrates high alignment with expert intuition and shows clear potential as a screening tool, there remains an opportunity to enhance the interpretability of the LLM-generated explanations. This would further improve its applicability in professional settings where rapid, accurate assessments are critical.

The findings from both the quantitative and qualitative evaluations provide a comprehensive perspective on the effectiveness of the proposed LLM-based trademark similarity analysis system. The quantitative results, as reflected in the expert evaluation scores, demonstrate that the similarity scores generated by the LLM align closely with expert intuition. This alignment is particularly evident in the high mean score for result relevance, suggesting that the system's ability to capture lexical, phonetic, and semantic relationships corresponds well with the criteria used by human evaluators. The qualitative assessments further reinforce this observation, as experts noted that the narrative explanations provided by the LLM aided in understanding why specific trademark pairs were judged to be similar or distinct. This interpretative capability adds a layer of transparency often absent in traditional automated approaches.

One of the major advantages of this system lies in its efficiency and explanatory depth. Compared to conventional trademark similarity assessments that often rely on manual rule-based methods or large pre-compiled databases, the proposed approach operates without the need for extensive data storage or retraining. The processing performance analysis highlights the system's ability to evaluate multiple trademark pairs rapidly through a lightweight web-based framework. Furthermore, unlike traditional string-matching or phonetic algorithms, the system provides contextual reasoning that bridges semantic and phonetic dimensions, offering insights that can be immediately useful in preliminary trademark screening.

Despite these strengths, the study also highlights several limitations that warrant consideration. The current evaluation is based on a limited set of ten trademark pairs, which, while diverse in their lexical, phonetic, and semantic characteristics, does not fully capture the complexity of real-world trademark conflicts. The system's performance also relies on the inherent capabilities of the LLM in interpreting phonetic and semantic nuances, which may vary depending on language-specific subtleties and the LLM's pretraining exposure. Additionally, the current implementation processes individual inputs sequentially through the web interface, and while this suffices for preliminary analysis, scalability to large-scale or batch processing environments would be necessary for industrial deployment. Integrating the framework into a recommendation system for trademark registration or conflict detection could further enhance its utility, enabling proactive risk assessment in intellectual property workflows.

Overall, the integration of quantitative alignment with expert perception and qualitative interpretability positions the system as a promising decision-support tool for trademark screening. Its efficiency, lack of dependency on large databases, and ability to generate human-readable reasoning collectively represent a significant advancement over conventional approaches, while its limitations point to clear avenues for future research and system enhancement.

CONCLUSION

This study presents a conceptual and practical advancement in trademark similarity analysis by leveraging a web-based framework powered by a large language model to evaluate lexical, phonetic, and semantic relationships between trademark names without the need for extensive databases or retraining. The system demonstrates high alignment with expert perception, as reflected in the strong relevance scores, while providing interpretative explanations that enhance transparency in preliminary screening. Its efficiency and narrative output distinguish it from traditional approaches, which often rely solely on string matching or phonetic heuristics and lack contextual reasoning. The research contributes to the field of intellectual property by introducing a scalable and explainable decision-support tool that can accelerate early-stage trademark screening, reduce the likelihood of conflicts, and inform branding strategies. By integrating expert-informed evaluation and demonstrating the potential for automation, this work highlights a pathway toward intelligent, LLM-driven frameworks that can transform how trademark examination and brand protection are conducted in the digital era.

REFERENCES

[1] S. Fishman, Trademark: Legal Care for Your Business & Product Name. Berkeley, CA, USA: Nolo, 2022.

[2] D. Crass and F. Schwiebacher, "The importance of trademark protection for product differentiation and innovation," Economia e Politica Industriale, vol. 44, no. 2, pp. 199–220, 2017.

[3] R. Setchi and F. M. Anuar, "Multi-faceted assessment of trademark similarity," Expert Systems with Applications, vol. 65, pp. 16–27, 2016.

[4] World Intellectual Property Organization (WIPO), World Intellectual Property Indicators 2024. Geneva, Switzerland: WIPO, 2024. doi: 10.34667/tind.50133.

[5] Suyahman, Branding UMKM: Perencanaan, Perlindungan, dan Evaluasi Branding di Era Digital. Klaten, Indonesia: PT Solusi Administrasi Hukum, 2025.

[6] G. Showkatramani, N. Khatri, A. Landicho and D. Layog, "Deep Learning Approach to Trademark International Class Identification," 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 2019, pp. 608-612, doi: 10.1109/ICMLA.2019.00112.

[7] C. V. Trappey, A. J. Trappey, and B. H. Liu, “Identify trademark legal case precedents—Using machine learning to enable semantic analysis of judgments,” World Patent Information, vol. 62, p. 101980, 2020.
[8] C. V. Trappey, A. J. Trappey, and S. C.-C. Lin, “Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies,” Advanced Engineering Informatics, vol. 45, p. 101120, 2020.
[9] S. Adarsh, E. Ash, S. Bechtold, B. Beebe, and J. Fromer, “Automating Abercrombie: Machine‐learning trademark distinctiveness,” Journal of Empirical Legal Studies, vol. 21, no. 4, pp. 826–860, 2024.
[10] R. Zhang and X. Wang, “Commodity classification in livestreaming marketing based on a conv-transformer network,” Multimedia Tools and Applications, vol. 83, no. 18, pp. 54909–54924, 2024.
[11] Heiler, S. (1995). Semantic interoperability. ACM Computing Surveys (CSUR), 27(2), 271-273.
[12] S. Webb and P. Nation, How Vocabulary Is Learned. Oxford, U.K.: Oxford Univ. Press, 2017.
[13] J. Pierrehumbert, "Phonological and phonetic representation," Journal of Phonetics, vol. 18, no. 3, pp. 375–394, 1990.
[14] P. Sanderson, "Linguistic analysis of competing trademarks," Language Matters, vol. 38, no. 1, pp. 132–149, 2007.
[15] Zhang, H., Gan, W., & Jiang, B. (2014, September). Machine learning and lexicon based methods for sentiment classification: A survey. In 2014 11th web information system and application conference (pp. 262-265). IEEE.
[16] Verma, M. (2017). Lexical analysis of religious texts using text mining and machine learning tools. International Journal of Computer Applications, 168(8), 39-45.
[17] Y. Sharma and J. Patil, "Understanding the concept of phonetic and visual similarity vis-a-vis to letter trademarks through judicial precedents," DME Journal of Law, vol. 4, no. 2, pp. 40–47, 2023.
[18] N. Raykar, P. Kumbharkar, and S. Rangdale, "Phonetic redundancy avoidance technique," in Proc. Int. Conf. Smart Systems: Innovations in Computing, Singapore: Springer Nature Singapore, Oct. 2023, pp. 109–118.
[19] M. Wu, J. Xu, X. Chen, and H. Meng, "Integrating potential pronunciations for enhanced mispronunciation detection and diagnosis ability in LLMs," in Proc. ICASSP 2025 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Apr. 2025, pp. 1–5. IEEE.
[20] S. Kumar and K. K. Bhatia, "Semantic similarity and text summarization based novelty detection," SN Applied Sciences, vol. 2, no. 3, p. 332, 2020.
[21] E. Shin, Y. Yu, R. R. Bies, and M. Ramanathan, "Evaluation of ChatGPT and Gemini large language models for pharmacometrics with NONMEM," Journal of Pharmacokinetics and Pharmacodynamics, vol. 51, no. 3, pp. 187–197, 2024.
[22] Google, "AI Studio," [Online]. Available: https://aistudio.google.com/. [Accessed: Aug. 2, 2025].
[23] P. Kumar and M. Sharma, "Data, machine learning, and human domain experts: none is better than their collaboration," International Journal of Human–Computer Interaction, vol. 38, no. 14, pp. 1307–1320, 2022.
[24] A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Likert scale: explored and explained," British Journal of Applied Science & Technology, vol. 7, no. 4, p. 396, 2015.
[25] H. Kirk, Prediction versus Management Models Relevant to Risk Assessment: The Importance of Legal Decision-Making Context, in Clinical Forensic Psychology and Law, pp. 347–359, 2019.
[26] D. G. Gordon and T. D. Breaux, "The role of legal expertise in interpretation of legal requirements and definitions," in Proc. 2014 IEEE 22nd Int. Requirements Engineering Conf. (RE), 2014, pp. 273–282.