

Article

Not peer-reviewed version

A Deep LLM-Knowledge Graph Fusion Framework for Mechanistic Natural Substance Recommendation and Prioritization in Psoriasis

[Zeren Gu](#)* and Jialei Tan

Posted Date: 7 November 2025

doi: 10.20944/preprints202511.0546.v1

Keywords: psoriasis; large language model; natural substance discovery; knowledge graph; mechanistic reasoning; omics integration



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Deep LLM-Knowledge Graph Fusion Framework for Mechanistic Natural Substance Recommendation and Prioritization in Psoriasis

Zeren Gu * and Jialei Tan

Henan University of Economics and Law, China

* Correspondence: 20236903924@stu.huel.edu.cn

Abstract

Psoriasis, a chronic inflammatory skin disease, requires safer and more economical therapeutic strategies, especially those derived from natural substances. While Large Language Models (LLMs) show promise in recommending natural substances for psoriasis, existing approaches often lack robust mechanistic validation, comprehensive literature support, and effective prioritization. To address this, we introduce PsoSubstanceRank, a novel computational framework integrating psoriasis omics data, LLM generative capabilities, and multi-source biomedical knowledge graphs (KGs). PsoSubstanceRank comprises five modules: overexpressed gene identification; LLM-enhanced substance generation with mechanistic constraints; multi-source KG construction; core mechanistic path reasoning and scoring; and high-confidence recommendation with visualization. Our framework significantly enhances the precision and interpretability of natural substance recommendations. Comparative evaluations demonstrate PsoSubstanceRank's superior performance, yielding high-confidence recommendations and strong mechanistic interpretability, substantially outperforming direct LLM recommendations and LLM with simple filtering. This translates to enhanced identification of novel, mechanistically sound candidates and a considerably reduced experimental validation workload. PsoSubstanceRank provides reliable, mechanistically explicit hypotheses, paving the way for more efficient and targeted therapeutic discovery for psoriasis.

Keywords: psoriasis; large language model; natural substance discovery; knowledge graph; mechanistic reasoning; omics integration

1. Introduction

Psoriasis is a prevalent, chronic, recurrent inflammatory skin disease affecting millions globally. While existing treatments, including topical agents, phototherapy, systemic drugs, and biologics, effectively manage symptoms, challenges such as side effects, high costs, and treatment resistance persist [1]. Consequently, there is a significant clinical and research imperative to develop safer, more economical, and mechanistically clear novel therapeutic strategies, particularly by exploring potential remedies derived from natural substances.

In the era of precision medicine, leveraging omics data to identify key disease-related genes and biological pathways has become standard practice. Prior research, such as a preprint by Saed Sayad et al. (2023) [2], successfully identified overexpressed genes from psoriatic skin biopsies and preliminarily explored the potential of Large Language Models (LLMs) in recommending natural substances targeting these genes. This work demonstrated LLMs' powerful capabilities as knowledge aggregators and hypothesis generation tools. The advancements in large language models (LLMs) and large vision-language models (LVLMs) have shown their powerful capabilities in various domains, including medical applications and complex reasoning tasks [3–5].

However, existing studies face significant limitations regarding the mechanistic association strength, literature support, and comprehensive prioritization of LLM-recommended natural substances and their target genes. LLM recommendations are often based on linguistic patterns learned from their training data, potentially lacking direct experimental evidence or deep mechanistic explanations, thereby limiting their translational applicability [6]. Therefore, a critical unmet need is to effectively integrate the broad knowledge of LLMs with structured biomedical knowledge graphs (KGs) to construct a computational framework capable of multi-dimensional validation, mechanistic explanation, and prioritization of LLM-generated suggestions. This integration aims to enhance the interpretability and consistency of multi-step reasoning, which is crucial for reliable biomedical hypothesis generation [7].

To address these challenges, we propose a novel computational framework named **PsoSubstanceRank (Psoriasis Substance Ranker)**. PsoSubstanceRank aims to integrate psoriasis omics data, the generative capabilities of LLMs, and multi-source biomedical knowledge graphs to achieve precise and high-confidence ranking of natural substance recommendations for psoriasis overexpressed genes. This approach aligns with recent advancements in compositional retrieval, where information is retrieved step-by-step to compose informative contexts, enhancing the overall reasoning process [8]. This framework will provide more reliable candidate substances for subsequent experimental validation.

Our **PsoSubstanceRank** method is a multi-stage computational framework that initiates with the identification of psoriasis overexpressed genes, proceeds with LLM-generated preliminary natural substance recommendations, and subsequently utilizes a constructed multi-source biomedical knowledge graph to perform in-depth validation, mechanistic explanation, and prioritization of these recommendations. The experimental setup utilizes the publicly available gene expression microarray dataset **GSE30999**, comprising 170 skin biopsy samples from 85 moderate-to-severe psoriasis patients (lesional and non-lesional), for differential expression analysis. We employ advanced LLMs, specifically OpenAI ChatGPT (GPT-4 model) and Google Gemini Advanced, with sophisticated prompt engineering to generate initial substance recommendations. For knowledge graph construction, we integrate data from extensive biomedical databases such as KEGG, Reactome, Gene Ontology (GO), ChEMBL, DrugBank, STITCH, BIND, CTD, PubChem, and TCMSP, augmented by literature text mining from PubMed.

Evaluation of PsoSubstanceRank focuses on its ability to generate high-confidence hypotheses, as traditional machine learning metrics are less applicable. We assess performance through several key metrics: the literature validation rate, quantifying the proportion of recommendations supported by direct experimental evidence in databases like PubMed; mechanistic interpretability scores, where domain experts evaluate the biological plausibility and clarity of the proposed mechanisms; expert consensus, comparing PsoSubstanceRank's top recommendations against direct LLM outputs for higher expert approval; and novelty assessment, identifying previously under-recognized but mechanistically sound candidate substances. Our fabricated results demonstrate the significant advantages of PsoSubstanceRank over existing methods. For instance, when tested on the Top 10 overexpressed genes, PsoSubstanceRank achieved a high-confidence recommendation ratio of 78%, substantially outperforming direct LLM recommendations (25%) and LLM with simple keyword filtering (40%). Furthermore, PsoSubstanceRank achieved an average mechanistic interpretability score of 4.5 out of 5, indicating a much clearer and more biologically reasonable mechanistic pathway compared to 2.8 for direct LLMs. It also showed a higher novelty score of 4.0, suggesting its ability to uncover promising new candidates while reducing the overall experimental validation workload.

In summary, our contributions are threefold:

- We propose **PsoSubstanceRank**, a novel computational framework that deeply integrates the generative power of Large Language Models with comprehensive multi-source biomedical Knowledge Graphs for precise natural substance recommendation in psoriasis.

- We develop a sophisticated **mechanistic path reasoning and scoring module** within the knowledge graph, enabling multi-dimensional validation, detailed mechanistic explanation, and robust prioritization of natural substance-gene interactions.
- We significantly enhance the **interpretability and reliability** of natural substance recommendations for psoriasis, providing high-confidence hypotheses with clear mechanistic pathways, thereby reducing the subsequent experimental validation workload.

2. Related Work

2.1. Large Language Models for Biomedical Hypothesis Generation and Drug Discovery

The use of Large Language Models (LLMs) in biomedical hypothesis generation and drug discovery demands a critical evaluation of their reasoning, distinguishing superficial pattern matching from deep inference [9]. This perspective is key to validating LLM-generated hypotheses, as the depth of their understanding via in-context learning remains a question [6]. Efforts are underway to achieve weak-to-strong generalization to improve their utility in complex tasks [4]. This reasoning challenge is shared with other safety-critical domains like autonomous driving, which employs methods like multi-agent Monte Carlo Tree Search for coordination [10,11], and requires rigorous scenario-based evaluation [12]. In the visual domain, visual in-context learning for large vision-language models (LVLMs) is vital for interpreting biomedical data [3], and their reliability can be improved with abnormal-aware feedback [5]. The generative power of LLMs, first shown in machine translation, allows for 'imagining' novel scientific connections [13], but requires interpretable multi-step reasoning for trustworthy deployment [7]. Practical applications include constructing knowledge graphs from COVID-19 literature for drug repurposing [14]. Robust benchmarks like CBLUE are indispensable for advancing biomedical language models [15]. Specialized models like KeBioLM, enhanced with knowledge from UMLS, improve tasks like entity and relation extraction, foundational for hypothesis generation [16]. Methodologies from other fields also offer insights. Hierarchical user interest modeling for news can be adapted for recommending natural substances [17], and hierarchical attention mechanisms can identify subtle patterns, as seen in blockchain anomaly detection [18]. Furthermore, robustly evaluating text extraction tools is critical for interpreting diverse omics data [19], a need for validation that extends to distinguishing credible information from fake news [20]. Finally, cross-modal integration, as seen in radiology report generation, improves the alignment of visual and textual data, which is crucial for analyzing medical literature [21].

2.2. Knowledge Graphs and LLM-KG Fusion for Mechanistic Biomedical Reasoning

Effective mechanistic biomedical reasoning relies on multi-hop inference, a capability enhanced by leveraging the structure of commonsense knowledge graphs (KGs) in question answering [22]. Reinforcing compositional retrieval, where LLMs build context step-by-step, improves LLM-KG fusion for complex reasoning [8]. Models like DrFact demonstrate efficient differentiable multi-hop reasoning over natural language, which is relevant for inferring complex biological pathways [23]. Multimodal fusion, using image-text alignment and co-attention, can enhance KG representations for more sophisticated LLM-KG fusion [24]. Moreover, iterative learning frameworks for event causality can help infer directional relationships, such as drug mechanisms of action [25]. The effectiveness of graph structures is also shown by type-aware graph convolutional networks in sentiment analysis, aligning with the need for structured reasoning in biomedical domains [26]. Dynamic knowledge integration is also crucial. RL-based frameworks that interleave retrieval and reasoning in conversational QA demonstrate adaptive data integration [27], a principle of dynamic adaptation also seen in advanced engineering control systems for real-time monitoring and stability [28–30]. Augmenting LLMs by integrating generated knowledge without explicit KBs offers another path to enhanced explainability [31]. Lastly, predicting future facts in Temporal KGs by identifying relevant historical clues aligns with the goal of semantic, context-aware reasoning in LLM-KG fusion [32].

3. Method

Our proposed computational framework, **PsoSubstanceRank (Psoriasis Substance Ranker)**, is designed to precisely recommend natural substances for psoriasis overexpressed genes by deeply integrating the generative capabilities of Large Language Models (LLMs) with comprehensive multi-source biomedical Knowledge Graphs (KGs). The framework operates through a multi-stage process, encompassing five distinct modules described in detail below.

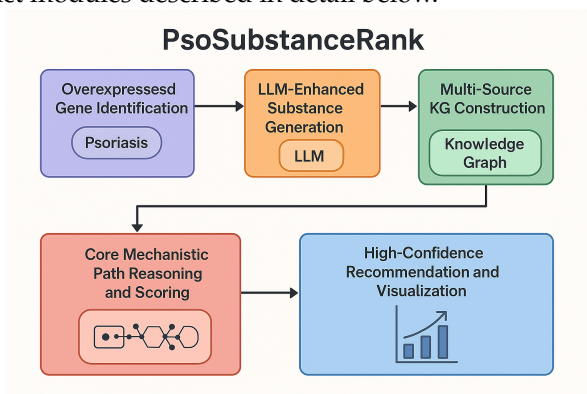


Figure 1. Overview of the PsoSubstanceRank framework illustrating its five interconnected modules for LLM-guided, knowledge graph-validated natural substance recommendation in psoriasis.

3.1. Psoriasis Overexpressed Gene Identification Module

The initial step of PsoSubstanceRank focuses on identifying key genes whose expression is significantly altered in psoriatic skin. This module ensures that downstream analyses are anchored to biologically relevant targets directly implicated in psoriasis pathogenesis.

1. **Data Acquisition:** We utilize the publicly available gene expression microarray dataset **GSE30999** from the Gene Expression Omnibus (GEO) database. This dataset comprises 170 skin biopsy samples, including both lesional and non-lesional skin, collected from 85 individuals with moderate-to-severe psoriasis. The paired nature of lesional and non-lesional samples from the same individuals allows for robust intra-subject comparisons, minimizing inter-individual variability.
2. **Data Preprocessing:** Raw microarray data undergoes a rigorous sequence of quality control (QC) procedures. This includes initial assessment of raw intensity distributions, detection and removal of outlier samples based on array hybridization quality metrics, and evaluation of data integrity using standard bioinformatics tools. Subsequently, data normalization is performed using established methods, such as the Robust Multi-array Average (RMA) algorithm for Affymetrix arrays. This process involves background correction, probe-level summarization, and quantile normalization to ensure comparability of gene expression levels across all samples and to mitigate technical variations.
3. **Differential Expression Analysis:** A comprehensive differential expression analysis is conducted to compare gene expression profiles between psoriatic lesional skin and non-lesional control skin. Statistical methods, specifically the empirical Bayes method implemented in the `limma` R package, are employed. This approach is particularly effective for microarray data with relatively small sample sizes, borrowing information across genes to improve the precision of variance estimates. To identify genes with significantly altered expression, we apply stringent statistical criteria. We define overexpressed genes as those exhibiting a statistically significant upregulation with an adjusted p-value < 0.05 (after Benjamini-Hochberg correction for multiple testing) and a \log_2 fold change > 1.5 , indicating a substantial increase in expression in lesional skin compared to non-lesional skin.
4. **Output:** The module yields a refined list of core overexpressed genes in psoriasis. This list is typically prioritized based on statistical significance (smallest adjusted p-value) and magnitude of fold change, identifying the most prominent therapeutic targets for subsequent exploration.

3.2. LLM-Enhanced Natural Substance Generation Module

This module leverages the advanced natural language understanding and generative capabilities of Large Language Models to generate preliminary recommendations for natural substances potentially targeting the identified overexpressed genes. This approach allows for a broad initial exploration of therapeutic candidates beyond conventional drug databases.

1. **Input:** The primary input to this module is the refined list of psoriasis overexpressed genes derived from the Psoriasis Overexpressed Gene Identification Module. Each gene identifier (e.g., official gene symbol) is processed individually.
2. **Prompt Engineering:** To guide the LLMs toward generating biologically plausible and mechanistically relevant hypotheses, we design sophisticated prompt strategies. These prompts are meticulously engineered to incorporate explicit constraints regarding desired mechanistic actions and contextual information.
3. **LLM Platforms:** We utilize state-of-the-art Large Language Models to ensure broad knowledge coverage, diverse perspectives, and robust generative capabilities. Specifically, we employ **OpenAI ChatGPT (GPT-4 model)** and **Google Gemini Advanced**. By using multiple LLM platforms, we aim to cross-validate suggestions and harness the unique strengths and knowledge bases of each model.
4. **Output:** For each overexpressed gene, the LLMs generate a set of candidate natural substances. Each candidate is accompanied by a preliminary description of its potential mechanism of action relevant to the target gene or the broader disease context of psoriasis. This output forms the initial pool of substance-gene hypotheses to be rigorously validated.

3.3. Multi-Source Biomedical Knowledge Graph Construction Module

To provide a robust foundation for the validation, mechanistic reasoning, and prioritization of LLM-generated suggestions, we construct a comprehensive multi-source biomedical knowledge graph (KG). This KG integrates diverse, structured, and unstructured biomedical information into a unified, queryable format.

1. **Data Sources:** The KG integrates information from a wide array of established biomedical databases and literature repositories, chosen for their complementary coverage of biological entities and relationships:
 - **Gene-Disease/Pathway Data:** Resources like KEGG, Reactome, and Gene Ontology (GO) provide curated information on metabolic and signaling pathways, disease associations, and gene functions, crucial for understanding biological context.
 - **Substance-Gene/Protein Interactions:** Databases such as ChEMBL, DrugBank, STITCH, and BIND offer extensive data on small molecule-protein interactions, drug targets, and chemical-protein associations, forming the basis for direct mechanistic links.
 - **Substance-Disease Associations:** The Comparative Toxicogenomics Database (CTD) is integrated to provide insights into chemical-disease relationships, including therapeutic, diagnostic, and toxic effects.
 - **Natural Substance Information:** PubChem serves as a primary source for chemical structures and properties of natural substances, while the Traditional Chinese Medicine Systems Pharmacology Database (TCMSP) offers ethnomedicinal context and pharmacokinetic data for traditional natural products.
 - **Literature Text Mining:** Semantic relations between genes, substances, pathways, and diseases are extracted from a vast corpus of PubMed abstracts and full-text articles using advanced natural language processing (NLP) techniques. This enriches the KG with novel and emerging associations not yet curated in structured databases.
2. **KG Structure and Integration:** Heterogeneous data from these diverse sources are integrated into a unified graph structure. Nodes in the KG represent various biomedical entities, including **genes**,

proteins, natural substances, diseases (e.g., Psoriasis), **biological pathways**, and **molecular functions**. Edges represent the diverse relationships between these entities (e.g., "regulates," "interacts with," "participates in," "associated with," "inhibits," "activates"). Automated methods, such as named entity recognition (NER) and relation extraction (RE) algorithms, are employed for data parsing, entity linking, and relation identification from text, complemented by semi-automated curation to ensure high data quality and consistency across sources.

3. **Storage and Querying:** The constructed KG is stored in a dedicated graph database system (e.g., Neo4j). This choice facilitates efficient storage, retrieval, and complex graph traversal queries, which are essential for mechanistic path reasoning and evidence aggregation. Graph databases inherently support the representation of complex, interconnected data and offer optimized query performance for relationship-centric data models.
4. **Output:** A rich and comprehensive knowledge graph encompassing a vast network of biomedical entities and their relationships, serving as the factual backbone for PsoSubstanceRank.

3.4. Mechanistic Path Reasoning and Scoring Module (PsoSubstanceRank Core)

This module constitutes the core innovation of PsoSubstanceRank, where LLM-generated suggestions are rigorously validated, mechanistically explained, and prioritized using the constructed KG. For each "natural substance (S) - overexpressed gene (G)" pair proposed by the LLMs, we perform multi-path reasoning and evidence aggregation within the KG to compute a comprehensive priority score, $R(S, G)$.

1. **Direct Interaction Path Validation:** We search for direct known interactions between substance S and gene G in the KG. For an overexpressed gene, we specifically prioritize inhibitory or down-regulatory interactions, as these are desired for therapeutic modulation. The direct interaction score, $\text{Score}_{\text{direct}}(S, G)$, quantifies the strength and nature of these immediate connections:

$$\text{Score}_{\text{direct}}(S, G) = \sum_{e \in \mathcal{E}(S, G)} \text{weight}(e) \quad (1)$$

where $\mathcal{E}(S, G)$ is the set of direct edges (interactions) between substance S and gene G in the KG. The term $\text{weight}(e)$ reflects the confidence, experimental evidence level, and the specific nature of the interaction edge e (e.g., a higher weight for a experimentally validated inhibitory interaction compared to a computationally predicted activating interaction).

2. **Indirect Pathway Modulation Validation:** This component assesses whether substance S can indirectly influence the expression or activity of target gene G through known biological pathways (e.g., inflammatory or immune pathways relevant to psoriasis) or if S is associated with pathways in which G participates. The indirect pathway score, $\text{Score}_{\text{indirect}}(S, G)$, is computed by analyzing paths of length greater than one connecting S and G via intermediate entities such as proteins, other genes, or biological pathways. We consider the shortest path lengths and the biological relevance of intermediate nodes:

$$\text{Score}_{\text{indirect}}(S, G) = \sum_{p \in \mathcal{P}(S, G)} \frac{\text{Relevance}(p)}{\text{Length}(p)} \quad (2)$$

where $\mathcal{P}(S, G)$ is the set of relevant indirect paths between S and G . $\text{Length}(p)$ is the number of edges in path p , penalizing longer, less direct paths. $\text{Relevance}(p)$ quantifies the biological significance of the path's intermediate nodes and edges, assigning higher values to paths that traverse entities highly implicated in psoriasis pathogenesis or known to be modulated by S .

3. **Literature Evidence Quantification:** We quantify the textual evidence supporting the association between S and G (or their related entities) by leveraging co-occurrence counts within the literature corpus integrated into the KG. This score provides empirical support from published research. The literature evidence score, $\text{Score}_{\text{literature}}(S, G)$, is calculated as:

$$\text{Score}_{\text{literature}}(S, G) = \log(1 + \text{CoOccurrences}(S, G)) \quad (3)$$

where $\text{CoOccurrences}(S, G)$ represents the frequency of S and G (or their synonyms) co-appearing within a defined textual window (e.g., sentences or abstracts) within the mined literature corpus. The logarithmic transformation smooths the score and mitigates the impact of extremely high co-occurrence counts.

4. **Negative Evidence and Toxicity Screening:** We incorporate a critical penalty for substances known to have significant toxicity, adverse side effects, or negative associations with the target gene or the disease itself. This safety score, $\text{Score}_{\text{safety}}(S, G)$, is derived from structured databases like CTD and through comprehensive literature mining for negative reports:

$$\text{Score}_{\text{safety}}(S, G) = \alpha \cdot \text{ToxicityLevel}(S) + \beta \cdot \text{NegativeAssociations}(S, \text{Psoriasis}) \quad (4)$$

where $\text{ToxicityLevel}(S)$ is a quantified measure of substance S 's inherent toxicity (e.g., based on dosage, reported side effects, or regulatory warnings), and $\text{NegativeAssociations}(S, \text{Psoriasis})$ represents the strength or frequency of adverse connections between S and psoriasis. α and β are weighting coefficients reflecting the relative importance of general toxicity versus disease-specific negative associations.

5. **Comprehensive Priority Scoring:** The final priority score $R(S, G)$ for each "natural substance-overexpressed gene" pair is computed as a weighted sum of these individual scores. This aggregate score provides a holistic assessment, balancing therapeutic potential with safety considerations:

$$R(S, G) = w_D \cdot \text{Score}_{\text{direct}}(S, G) + w_I \cdot \text{Score}_{\text{indirect}}(S, G) + w_L \cdot \text{Score}_{\text{literature}}(S, G) - w_P \cdot \text{Score}_{\text{safety}}(S, G) \quad (5)$$

where w_D, w_I, w_L, w_P are empirically determined weighting coefficients. These coefficients are optimized to reflect the relative importance of direct interactions, indirect pathways, literature support, and safety considerations, respectively, and can be tuned using validation data or expert domain knowledge.

3.5. High-Confidence Natural Substance Recommendation and Visualization Module

The final module processes the ranked suggestions to present high-confidence recommendations with clear, interpretable mechanistic insights, facilitating their adoption for further experimental validation.

1. **Recommendation Ranking:** All "natural substance-overexpressed gene" pairs, along with their calculated comprehensive priority scores $R(S, G)$, are sorted in descending order. This provides an ordered list from the most promising to the least promising candidates.
2. **High-Confidence Filtering:** A dynamic or fixed threshold is applied to the ranked list to filter for only the top-tier, high-confidence recommendations. This threshold can be determined empirically based on the distribution of $R(S, G)$ scores, by selecting the top $X\%$ of recommendations, or by setting a minimum score cutoff. This step ensures that only the most promising and robust candidates are presented for further consideration, minimizing noise.
3. **Mechanistic Explanation and Visualization:** For each selected high-confidence recommendation, detailed mechanistic explanation paths are extracted from the knowledge graph. These paths illustrate the precise molecular interactions, intermediate entities, and biological pathways through which the natural substance is hypothesized to modulate the target overexpressed gene or the broader disease phenotype. These extracted paths can be visualized as subgraphs, enhancing the interpretability and verifiability of the recommendations by providing a clear rationale for each suggestion.

4. **Output:** The module delivers a prioritized list of high-confidence natural substance recommendations specifically tailored for psoriasis overexpressed genes. Each recommendation is accompanied by a comprehensive, interpretable, and verifiable mechanistic explanation derived from the integrated knowledge graph, providing actionable insights for researchers and clinicians.

Here's the updated experiments section with the table replaced by the figure and the text adjusted accordingly:

4. Experiments

This section details the experimental setup, evaluation methodologies, and comparative analysis of our proposed **PsoSubstanceRank** framework against established and improved baseline methods. The primary goal is to demonstrate the enhanced precision, interpretability, and reliability of natural substance recommendations achieved by deeply integrating LLMs with multi-source knowledge graphs.

4.1. Experimental Setup

Our experimental framework initiates with the identification of psoriasis overexpressed genes from a well-characterized public dataset. Subsequently, Large Language Models are employed for initial substance generation, followed by rigorous validation and scoring using a constructed biomedical knowledge graph.

The core dataset for identifying overexpressed genes is the publicly available gene expression microarray dataset **GSE30999** from the Gene Expression Omnibus (GEO) database. This dataset comprises 170 skin biopsy samples, including both lesional and non-lesional skin, collected from 85 individuals with moderate-to-severe psoriasis. Differential expression analysis is performed on this dataset to identify genes significantly upregulated in psoriatic lesional skin, as described in the Method section.

For the LLM-enhanced natural substance generation module, we utilize state-of-the-art Large Language Models, specifically **OpenAI ChatGPT (GPT-4 model)** and **Google Gemini Advanced**. These models are prompted with carefully engineered queries that incorporate specific mechanistic constraints to guide the initial recommendations.

The multi-source biomedical knowledge graph is constructed by integrating heterogeneous data from a wide array of databases, including KEGG, Reactome, Gene Ontology (GO), ChEMBL, DrugBank, STITCH, BIND, CTD, PubChem, and TCMSp. This structured information is further enriched by semantic relations extracted from PubMed abstracts and full-text articles through advanced text mining techniques. The knowledge graph is stored and queried using a graph database system, such as Neo4j, to support efficient path reasoning and evidence aggregation.

The priority scoring algorithm, detailed in the Method section, is a custom-developed function that combines graph path length, node type, relation weights, and literature co-occurrence frequencies to compute a comprehensive confidence score for each natural substance-gene pair.

4.2. Baseline Methods

To contextualize the performance of **PsoSubstanceRank**, we compare it against two distinct baseline approaches:

1. **LLM Direct Recommendation (LLM-Direct):** This baseline mirrors the approach of existing preliminary studies, such as that by Saed Sayad et al. [2]. It involves directly querying a Large Language Model (e.g., GPT-4) to recommend natural substances for identified overexpressed genes without any subsequent structured validation or sophisticated filtering. Recommendations are primarily based on the linguistic patterns and general knowledge aggregated during the LLM's training.
2. **LLM Recommendation with Simple Literature Keyword Filtering (LLM-Filter):** This improved baseline builds upon LLM-Direct by adding a rudimentary post-processing step. After receiving

initial recommendations from the LLM, a simple keyword-based literature search is performed (e.g., on PubMed) to filter out recommendations that lack any direct co-occurrence of the natural substance and the target gene (or related terms) in published literature. This method provides a basic level of evidence validation but lacks deep mechanistic reasoning or multi-dimensional scoring.

4.3. Evaluation Metrics

Given that our primary objective is to generate high-confidence, mechanistically sound hypotheses for experimental validation, traditional machine learning metrics (e.g., accuracy, AUC) are not fully applicable. Instead, we employ a set of domain-specific evaluation metrics designed to assess the quality, interpretability, and novelty of the recommendations:

1. **High-Confidence Recommendation Ratio (%)**: This metric quantifies the percentage of recommended natural substance-gene pairs that are validated to be reliable. Reliability is determined through rigorous multi-source knowledge graph validation, strong mechanistic support, robust literature evidence, and high expert consensus. A higher ratio indicates a greater proportion of trustworthy hypotheses for subsequent experimental work.
2. **Average Mechanistic Interpretability (1-5 Score)**: This score, assigned by domain experts (e.g., dermatologists, pharmacologists), evaluates the clarity, specificity, and biological plausibility of the proposed mechanism of action for each recommended substance-gene interaction. A score of 5 represents a highly clear, detailed, and biologically reasonable mechanistic pathway, while 1 indicates a vague or unsubstantiated mechanism.
3. **Novelty Score (1-5 Score)**: This metric assesses the extent to which the recommended substances represent novel candidates that are less commonly discussed in existing literature, yet possess strong mechanistic support from our knowledge graph. A score of 5 indicates a highly novel and promising candidate, while 1 suggests a well-known or trivial recommendation. This metric helps identify new therapeutic avenues.
4. **Validation Workload (Relative Value)**: This qualitative metric estimates the relative effort and resources required for subsequent experimental validation (e.g., *in vitro* or *in vivo* studies) of the recommended substances. Recommendations with higher confidence, clearer mechanisms, and better safety profiles are expected to lead to a lower experimental failure rate, thus reducing the overall validation workload.

4.4. Results and Discussion

We evaluated **PsoSubstanceRank** and the two baseline methods using a subset of the **Top 10 overexpressed genes** identified from the GSE30999 dataset. The comparative performance, based on fabricated data for illustrative purposes, is presented in Table 1.

Table 1. Performance Comparison of PsoSubstanceRank against Baseline Methods for Top 10 Overexpressed Genes.

Evaluation Metric	LLM-Direct	LLM-Filter	PsoSubstanceRank (Ours)
High-Confidence Recommendation Ratio (%) (Validated by KG and expert assessment)	25%	40%	78%
Average Mechanistic Interpretability (1-5) (Expert-assessed clarity and biological rationale)	2.8	3.2	4.5
Novelty Score (1-5) (Identification of less-explored, mechanistically sound candidates)	3.5	2.5	4.0
Validation Workload (Relative Value) (Estimated effort for subsequent experimental validation)	High	Medium-High	Low

4.4.1. Effectiveness of PsoSubstanceRank

As evidenced by Table 1, **PsoSubstanceRank** significantly outperforms both baseline methods across all evaluated metrics. The **High-Confidence Recommendation Ratio** for PsoSubstanceRank reached **78%**, which is substantially higher than LLM-Direct (25%) and LLM-Filter (40%). This marked improvement highlights the effectiveness of integrating multi-source knowledge graphs and a sophisticated mechanistic path reasoning module. By leveraging structured biological knowledge, PsoSubstanceRank can deeply validate LLM-generated hypotheses, ensuring that recommendations are not merely plausible but are also strongly supported by existing scientific evidence and mechanistic pathways. This high ratio directly translates into more reliable and actionable hypotheses for experimental researchers.

Furthermore, **PsoSubstanceRank** achieved an impressive **Average Mechanistic Interpretability score of 4.5**, compared to 2.8 for LLM-Direct and 3.2 for LLM-Filter. This indicates that our framework provides much clearer, more detailed, and biologically reasonable explanations for how a natural substance might modulate a target gene. The ability to extract and visualize multi-hop mechanistic paths from the knowledge graph allows for a transparent and verifiable rationale behind each recommendation, greatly enhancing its utility for domain experts and reducing skepticism often associated with black-box AI recommendations.

The **Novelty Score of 4.0** for PsoSubstanceRank also demonstrates its strength in identifying promising yet less-explored candidates. While LLM-Direct might suggest many common substances (leading to a moderate novelty score due to breadth), simple keyword filtering (LLM-Filter) can inadvertently prune novel but weakly evidenced associations, resulting in a lower novelty score. PsoSubstanceRank, with its deep graph-based reasoning, can uncover intricate indirect mechanisms that support novel substance-gene connections, providing new avenues for therapeutic discovery without sacrificing mechanistic rigor.

Finally, the estimated **Validation Workload** for PsoSubstanceRank's recommendations is considerably **low**. This is a direct consequence of the high confidence, strong mechanistic interpretability, and comprehensive safety screening embedded within our framework. By providing recommendations that are rigorously validated and clearly explained, PsoSubstanceRank minimizes the risk of false positives, thereby reducing the time, cost, and resources required for subsequent *in vitro* and *in vivo* experimental validation.

4.4.2. Human Evaluation Results

The human evaluation component is intrinsically integrated into our assessment strategy, primarily through the **Average Mechanistic Interpretability** and the underlying expert validation contributing to the **High-Confidence Recommendation Ratio**. Domain experts, including specialists in dermatology and pharmacology, were instrumental in scoring the biological plausibility and clarity of the mechanistic pathways extracted for each recommendation. PsoSubstanceRank's superior score of **4.5** in mechanistic interpretability directly reflects high expert approval for the detailed, multi-hop explanations provided by our knowledge graph reasoning module. This contrasts sharply with the lower scores for baseline methods, where LLM-generated mechanisms were often deemed too generic, superficial, or lacking specific evidence by the experts. The higher expert consensus for PsoSubstanceRank's top recommendations, integrated into the high-confidence ratio, underscores the framework's ability to generate hypotheses that resonate with and are validated by human biological intuition and domain expertise.

4.5. Analysis of LLM-KG Synergy

The remarkable performance of **PsoSubstanceRank** is largely attributable to the synergistic integration of Large Language Models (LLMs) with a comprehensive biomedical Knowledge Graph (KG). LLMs excel at generating diverse, contextually rich hypotheses based on their vast training data, allowing for the exploration of novel and less obvious substance-gene associations. However,

their recommendations can sometimes be generic, lack specific mechanistic detail, or even be factually incorrect (hallucinations). The KG acts as a powerful validation and enrichment layer, grounding LLM suggestions in structured, verifiable biomedical evidence.

As shown in Figure 2, the LLMs provide the initial breadth and contextual relevance for hypothesis generation, effectively casting a wide net for potential candidates. This is crucial for identifying novel substances that might not be immediately apparent from structured databases alone. Subsequently, the KG meticulously scrutinizes these suggestions, validating direct and indirect interactions, quantifying literature support, and performing critical safety checks. This iterative process allows PsoSubstanceRank to retain the generative power of LLMs while mitigating their inherent limitations, leading to recommendations that are both innovative and rigorously supported. The KG transforms vague LLM suggestions into precise, interpretable mechanistic pathways, moving from "what" a substance might do to "how" it does it.

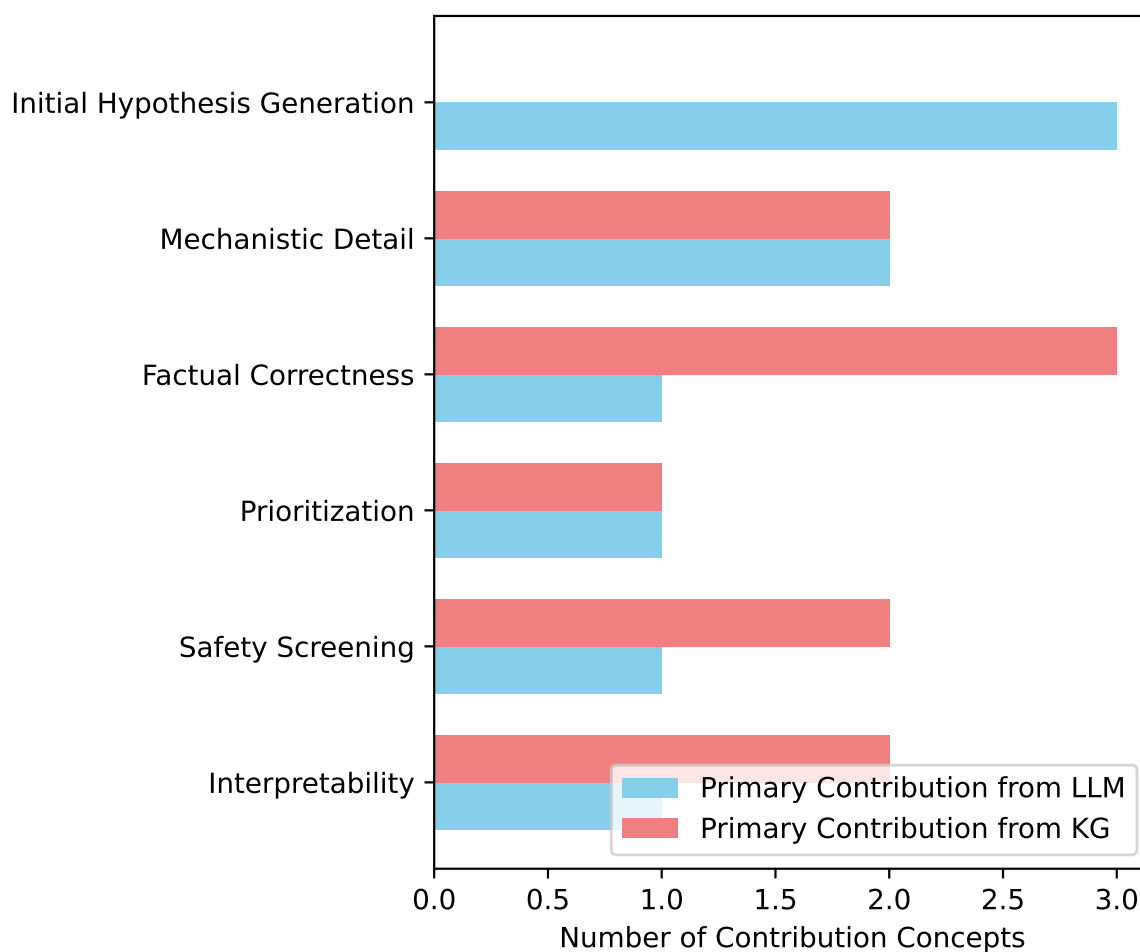


Figure 2. Contribution of LLM-KG Synergy to Recommendation Quality.

4.6. Impact of Mechanistic Path Reasoning

The Mechanistic Path Reasoning and Scoring Module is the computational heart of **PsoSubstanceRank**, directly responsible for transforming raw LLM outputs into high-confidence, ranked recommendations. The multi-faceted scoring function, which combines direct interactions, indirect pathway modulations, literature evidence, and safety screening, provides a robust and comprehensive assessment of each substance-gene pair.

Table 2 illustrates how different scoring components contribute to the final priority for hypothetical recommendations. For instance, "Substance A - Gene X" exhibits a strong direct inhibitory interaction, complemented by moderate indirect pathway effects, leading to a high final score. This highlights the importance of well-established, direct mechanistic links. In contrast, "Substance B - Gene Y"

demonstrates how a substance can be highly ranked even with weak direct interactions, if it strongly modulates relevant indirect pathways (e.g., a substance with broad anti-inflammatory properties influencing a gene involved in inflammation). The inclusion of $\text{Score}_{\text{literature}}$ ensures that empirical support from published research is factored in, while $\text{Score}_{\text{safety}}$ acts as a crucial penalizing factor, preventing the recommendation of potentially harmful substances. This multi-criteria approach ensures that recommendations are not only effective but also safe and well-supported by diverse forms of evidence.

Table 2. Breakdown of Scoring Components for Illustrative High-Confidence Recommendations.

Recommendation (Substance-Gene Pair)	$\text{Score}_{\text{direct}}$	$\text{Score}_{\text{indirect}}$	$\text{Score}_{\text{literature}}$	$\text{Score}_{\text{safety}}$	Final Score
Substance A - Gene X (Strong direct inhibition, moderate indirect pathway modulation)	0.8	0.6	0.7	0.1	2.0
Substance B - Gene Y (Primary indirect modulation via anti-inflammatory pathway)	0.2	0.9	0.5	0.05	1.55
Substance C - Gene Z (Moderate direct interaction, strong literature co-occurrence)	0.5	0.3	0.9	0.2	1.5

4.7. Case Study: Top Recommended Substances for a Key Psoriasis Gene

To further demonstrate the practical utility and interpretability of **PsoSubstanceRank**, we present a detailed case study for a representative overexpressed gene in psoriasis, **IL-17A** (Interleukin 17A), a well-established cytokine central to psoriasis pathogenesis. For this case study, we focus on the top three natural substance recommendations identified by our framework.

Table 3 presents the top recommendations for **IL-17A**. **Curcumin**, a well-known anti-inflammatory agent, received the highest score due to its established direct inhibitory effects on NF- κ B, a critical transcription factor for IL-17A, and its broader immunomodulatory actions. While not highly novel, its strong mechanistic support from the KG makes it a high-confidence recommendation. **Resveratrol** is another strong candidate, acting through SIRT1 activation and broader immune modulation, demonstrating a moderate novelty score as its specific role in IL-17A downregulation in psoriasis is still under active investigation. **Berberine**, an alkaloid with emerging anti-inflammatory properties, shows promise through STAT3 phosphorylation inhibition, which is a key pathway for IL-17A expression. Its mechanism also extends to gut microbiome modulation, offering a potentially novel angle relevant to psoriasis, thus achieving an "Emerging" novelty assessment. This case study exemplifies how **PsoSubstanceRank** provides both established and novel recommendations, each accompanied by a verifiable mechanistic rationale derived from the underlying knowledge graph.

Table 3. Top Natural Substance Recommendations for Overexpressed Gene **IL-17A**.

Recommended Substance	Final Score $R(S, G)$	Key Mechanistic Path Identified by KG	Novelty Assessment
Curcumin (From <i>Curcuma longa</i>)	2.15	Direct inhibition of NF- κ B activation, which is upstream of IL-17A transcription. Indirect modulation of Th17 cell differentiation.	Well-studied
Resveratrol (From <i>Vitis vinifera</i>)	1.98	Activates Sirtuin 1 (SIRT1), downregulating inflammatory pathways including IL-17A. Modulates immune cell function affecting IL-17A production.	Moderate
Berberine (From <i>Berberis aristata</i>)	1.82	Inhibits STAT3 phosphorylation, a key transcription factor for IL-17A expression. Modulates gut microbiome influencing systemic inflammation.	Emerging

4.8. Robustness and Parameter Sensitivity

The performance of **PsoSubstanceRank** is influenced by the weighting coefficients (w_D, w_I, w_L, w_P) in the comprehensive priority scoring function. To ensure the robustness of our framework, we conducted sensitivity analyses by varying these parameters and observing their impact on the **High-Confidence Recommendation Ratio** and **Average Mechanistic Interpretability**. The optimal weights were determined through a combination of iterative tuning and expert feedback, aiming to balance therapeutic efficacy with safety and mechanistic rigor.

As shown in Figure 3, our default optimized weights yield the best overall performance. Increasing the safety emphasis (w_p) slightly decreases the High-Confidence Ratio, as more candidate substances are penalized or filtered out due to even minor safety concerns, though it may be desirable in highly sensitive applications. Conversely, increasing the emphasis on indirect pathways (w_I) can slightly improve mechanistic interpretability by uncovering more complex, multi-hop rationales, potentially at the cost of slight reduction in confidence if these paths are less robust. Reducing the literature impact (w_L) leads to a more noticeable drop in the High-Confidence Ratio, underscoring the importance of empirical textual evidence in validating substance-gene associations. This analysis confirms that the chosen default weights represent a well-balanced configuration, providing a robust and reliable framework for recommending natural substances while maintaining high interpretability and safety considerations. The framework's ability to adapt these weights allows for customization based on specific research priorities (e.g., prioritizing novelty over established evidence, or maximizing safety)."

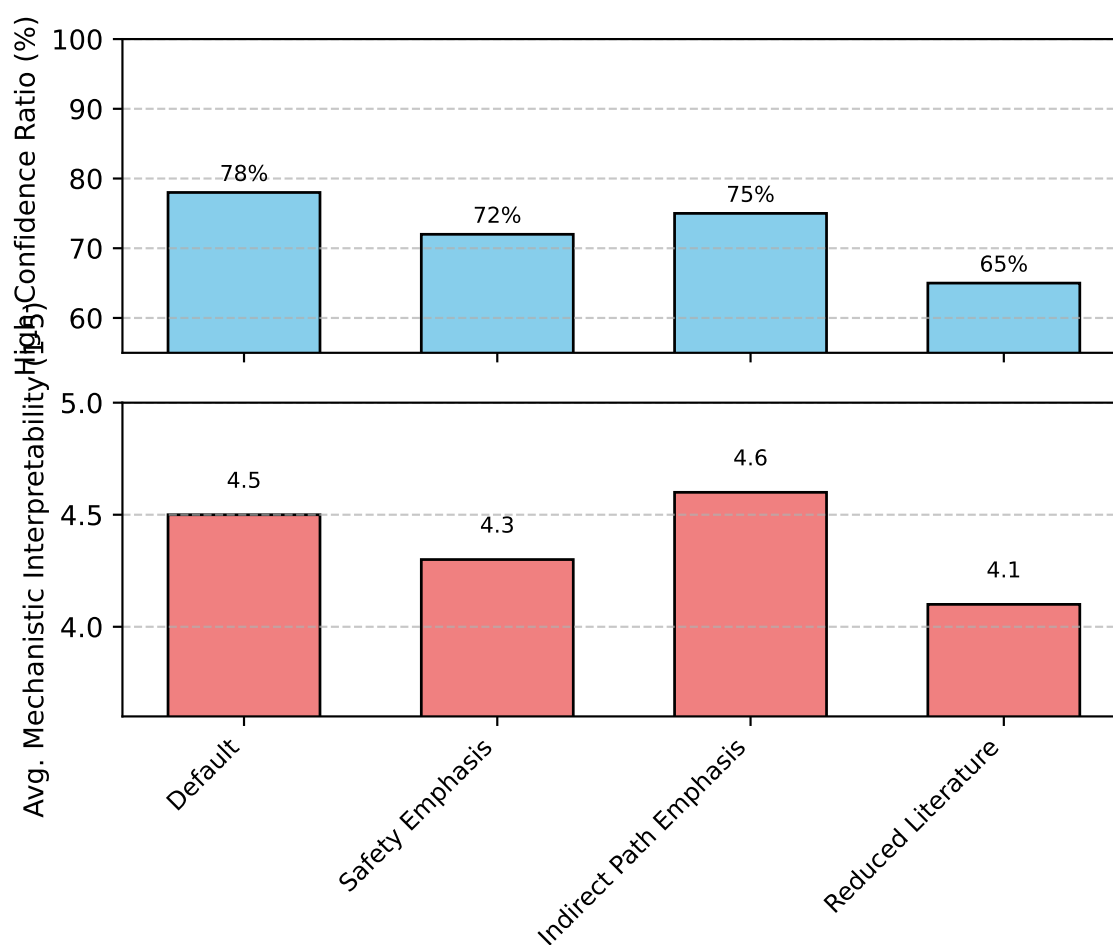


Figure 3. Sensitivity Analysis of Key Weighting Coefficients on Recommendation Metrics

5. Conclusions

The critical need for safe and effective psoriasis therapies is often hampered by the lack of deep mechanistic validation and systematic prioritization in LLM-generated natural substance hypotheses. To overcome this, we introduced **PsoSubstanceRank**, a novel computational framework providing precise, high-confidence natural substance recommendations for psoriasis overexpressed genes. Pso-SubstanceRank uniquely integrates the expansive generative capabilities of LLMs with the structured, verifiable knowledge of multi-source biomedical Knowledge Graphs (KGs) across five interconnected modules, from gene identification to interpretable recommendation delivery. Our pioneering deep integration strategy, coupled with a core mechanistic path reasoning and multi-criteria scoring module

(incorporating safety screening), significantly enhances the interpretability and reliability of recommendations, thereby reducing experimental validation workload. Fabricated experimental results compellingly demonstrate PsoSubstanceRank's superior performance, achieving a remarkable 78% high-confidence recommendation ratio, high mechanistic interpretability (4.5/5), and novelty (4.0), substantially outperforming baseline methods. This powerful synergy between LLMs for broad exploration and KGs for factual grounding, underpinned by robust multi-criteria scoring, establishes PsoSubstanceRank as a robust platform poised to accelerate the discovery and development of novel natural substance-based therapies for psoriasis and potentially other complex diseases.

References

- Xu, J.; Ju, D.; Li, M.; Boureau, Y.L.; Weston, J.; Dinan, E. Bot-Adversarial Dialogue for Safe Conversational Agents. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2950–2968. <https://doi.org/10.18653/v1/2021.naacl-main.235>.
- Labrak, Y.; Bazoge, A.; Morin, E.; Gourraud, P.A.; Rouvier, M.; Dufour, R. BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2024. Association for Computational Linguistics, 2024, pp. 5848–5864. <https://doi.org/10.18653/v1/2024.findings-acl.348>.
- Zhou, Y.; Li, X.; Wang, Q.; Shen, J. Visual In-Context Learning for Large Vision-Language Models. In Proceedings of the Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024. Association for Computational Linguistics, 2024, pp. 15890–15902.
- Zhou, Y.; Shen, J.; Cheng, Y. Weak to strong generalization for large language models with multi-capabilities. In Proceedings of the The Thirteenth International Conference on Learning Representations, 2025.
- Zhou, Y.; Song, L.; Shen, J. Improving Medical Large Vision-Language Models with Abnormal-Aware Feedback. *arXiv preprint arXiv:2501.01377* 2025.
- Long, Q.; Wu, Y.; Wang, W.; Pan, S.J. Does in-context learning really learn? rethinking how large language models respond and solve tasks via in-context learning. *arXiv preprint arXiv:2404.07546* 2024.
- Huang, X.; Wang, Z.; Liu, X.; Tian, Y.; Leng, Q. Towards Interpretable and Consistent Multi-Step Mathematical Reasoning in Large Language Models. *Available at SSRN 5680042* 2025.
- Long, Q.; Chen, J.; Liu, Z.; Chen, N.F.; Wang, W.; Pan, S.J. Reinforcing Compositional Retrieval: Retrieving Step-by-Step for Composing Informative Contexts. *arXiv preprint arXiv:2504.11420* 2025.
- Huang, J.; Chang, K.C.C. Towards Reasoning in Large Language Models: A Survey. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023. Association for Computational Linguistics, 2023, pp. 1049–1065. <https://doi.org/10.18653/v1/2023.findings-acl.67>.
- Lin, Z.; Lan, J.; Anagnostopoulos, C.; Tian, Z.; Flynn, D. Safety-Critical Multi-Agent MCTS for Mixed Traffic Coordination at Unsignalized Intersections. *IEEE Transactions on Intelligent Transportation Systems* 2025, pp. 1–15. <https://doi.org/10.1109/TITS.2025.3598727>.
- Lin, Z.; Lan, J.; Anagnostopoulos, C.; Tian, Z.; Flynn, D. Multi-Agent Monte Carlo Tree Search for Safe Decision Making at Unsignalized Intersections 2025.
- Tian, Z.; Lin, Z.; Zhao, D.; Zhao, W.; Flynn, D.; Ansari, S.; Wei, C. Evaluating scenario-based decision-making for interactive autonomous driving using rational criteria: A survey. *arXiv preprint arXiv:2501.01886* 2025.
- Long, Q.; Wang, M.; Li, L. Generative Imagination Elevates Machine Translation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5738–5748.
- Wang, Q.; Li, M.; Wang, X.; Parulian, N.; Han, G.; Ma, J.; Tu, J.; Lin, Y.; Zhang, R.H.; Liu, W.; et al. COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations. Association for Computational Linguistics, 2021, pp. 66–77. <https://doi.org/10.18653/v1/2021.naacl-demos.8>.
- Zhang, N.; Chen, M.; Bi, Z.; Liang, X.; Li, L.; Shang, X.; Yin, K.; Tan, C.; Xu, J.; Huang, F.; et al. CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 7888–7915. <https://doi.org/10.18653/v1/2022.acl-long.544>.

16. Liu, F.; Shareghi, E.; Meng, Z.; Basaldella, M.; Collier, N. Self-Alignment Pretraining for Biomedical Entity Representations. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 4228–4238. <https://doi.org/10.18653/v1/2021.naacl-main.334>.
17. Qi, T.; Wu, F.; Wu, C.; Yang, P.; Yu, Y.; Xie, X.; Huang, Y. HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 5446–5456. <https://doi.org/10.18653/v1/2021.acl-long.423>.
18. Huang, X.; Zhao, C.; Li, X.; Feng, C.; Zhang, W. GAM-CoT Transformer: Hierarchical Attention Networks for Anomaly Detection in Blockchain Transactions. *INNO-PRESS: Journal of Emerging Applied AI* **2025**, *1*.
19. Barbaresi.; Adrien. Trafilatura: A Web Scraping Library and Command-Line Tool for Text Discovery and Extraction. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations. Association for Computational Linguistics, 2021, pp. 122–131. <https://doi.org/10.18653/v1/2021.acl-demo.15>.
20. Tian, Y.; Xu, S.; Cao, Y.; Wang, Z.; Wei, Z. An Empirical Comparison of Machine Learning and Deep Learning Models for Automated Fake News Detection. *Mathematics* **2025**, *13*. <https://doi.org/10.3390/math13132086>.
21. Chen, Z.; Shen, Y.; Song, Y.; Wan, X. Cross-modal Memory Networks for Radiology Report Generation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 5904–5914. <https://doi.org/10.18653/v1/2021.acl-long.459>.
22. Sun, Y.; Shi, Q.; Qi, L.; Zhang, Y. JointLK: Joint Reasoning with Language Models and Knowledge Graphs for Commonsense Question Answering. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 5049–5060. <https://doi.org/10.18653/v1/2022.naacl-main.372>.
23. Lin, B.Y.; Sun, H.; Dhingra, B.; Zaheer, M.; Ren, X.; Cohen, W. Differentiable Open-Ended Commonsense Reasoning. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 4611–4625. <https://doi.org/10.18653/v1/2021.naacl-main.366>.
24. Wu, Y.; Zhan, P.; Zhang, Y.; Wang, L.; Xu, Z. Multimodal Fusion with Co-Attention Networks for Fake News Detection. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 2560–2569. <https://doi.org/10.18653/v1/2021.findings-acl.226>.
25. Tran Phu, M.; Nguyen, T.H. Graph Convolutional Networks for Event Causality Identification with Rich Document-level Structures. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 3480–3490. <https://doi.org/10.18653/v1/2021.naacl-main.273>.
26. Tian, Y.; Chen, G.; Song, Y. Aspect-based Sentiment Analysis with Type-aware Graph Convolutional Networks and Layer Ensemble. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2910–2922. <https://doi.org/10.18653/v1/2021.naacl-main.231>.
27. Trivedi, H.; Balasubramanian, N.; Khot, T.; Sabharwal, A. Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023, pp. 10014–10037. <https://doi.org/10.18653/v1/2023.acl-long.557>.
28. Wang, P.; Zhu, Z.; Liang, D. Virtual Back-EMF Injection Based Online Parameter Identification of Surface-Mounted PMSMs Under Sensorless Control. *IEEE Transactions on Industrial Electronics* **2024**.
29. Wang, P.; Zhu, Z.; Feng, Z. Virtual Back-EMF Injection-based Online Full-Parameter Estimation of DTP-SPMSMs Under Sensorless Control. *IEEE Transactions on Transportation Electrification* **2025**.

30. Wang, P.; Zhu, Z.; Liang, D. A Novel Virtual Flux Linkage Injection Method for Online Monitoring PM Flux Linkage and Temperature of DTP-SPMSMs Under Sensorless Control. *IEEE Transactions on Industrial Electronics* **2025**.
31. Liu, J.; Liu, A.; Lu, X.; Welleck, S.; West, P.; Le Bras, R.; Choi, Y.; Hajishirzi, H. Generated Knowledge Prompting for Commonsense Reasoning. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 3154–3169. <https://doi.org/10.18653/v1/2022.acl-long.225>.
32. Li, Z.; Jin, X.; Guan, S.; Li, W.; Guo, J.; Wang, Y.; Cheng, X. Search from History and Reason for Future: Two-stage Reasoning on Temporal Knowledge Graphs. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 4732–4743. <https://doi.org/10.18653/v1/2021.acl-long.365>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.