A Generative AI-Extracted Mechanistic Knowledge Graph for Cosmetic Ingredient Synergy: Construction, Analysis, and Expected Prediction with CosmoSyn-GATED

International Journal of Cosmetic Science · Submitted 2026

Abstract

The efficacy of a cosmetic product frequently arises from the combination of multiple active ingredients rather than from any single component, yet combination-level clinical evidence is almost entirely absent and the combinatorial explosion of possible pairings makes exhaustive testing infeasible. As a consequence, decisions about ingredient combinations remain largely empirical, and claims of synergy or antagonism are difficult to substantiate on a mechanistic basis. This study proposes a pipeline that, in the absence of structured combination data, couples a large language model (LLM) with a mechanistic knowledge graph to infer and evaluate cosmetic ingredient synergy and antagonism in an evidence-grounded manner. Single-ingredient knowledge scattered across the literature, INCI listings, and mechanistic descriptions is structured by the LLM into an ingredient-mechanism-target-effect relation graph, and combination effects are quantified through pathway-overlap and pathway-conflict scores reported together with their supporting evidence. The absence of ground-truth labels is addressed by back-validation on a curated reference set of well-established synergistic, antagonistic, and neutral ingredient pairs, by agreement with dermatologist and formulation-expert judgments, and by an explicit hallucination-rate metric. Because no experiment has yet been conducted, this manuscript presents the study design together with EXPECTED RESULTS, and we anticipate that the proposed CosmoSyn-GATED model will exceed LLM-only and LLM+RAG baselines in classification quality and explainability while substantially reducing ungrounded synergy claims.

Keywords: Cosmetic Ingredient Synergy, Mechanistic Knowledge Graph, Graph Attention Network, Large Language Model, Evidence-Grounded Inference, Explainable Recommendation

1. Introduction

1.1 Cosmetic Active Ingredients and the Combination Paradigm

Modern cosmetic formulation increasingly treats efficacy as a property of ingredient combinations rather than of any single active. Over the past two decades the industry has shifted from single-ingredient efficacy claims toward multi-ingredient synergy claims, in which the value proposition rests on how ingredients interact. A canonical positive example is the combination of vitamin C, vitamin E, and ferulic acid, in which ferulic acid stabilizes the antioxidant solution while the two vitamins regenerate one another, producing photoprotection that exceeds the sum of the individual contributions. A well-known negative example is the concurrent use of retinol with alpha-hydroxy acids, which stacks exfoliation and cell-turnover stimulation and tends to produce cumulative irritation rather than additive benefit. The combination, therefore, rather than the isolated molecule, is the meaningful unit of analysis for both efficacy and tolerability.

1.2 The Synergy Problem: Need and Technical Challenges

Despite the centrality of combinations, the evidence base for combination-level efficacy is remarkably thin. Combination-level randomized controlled trials are seldom conducted because of cost, ethical constraints, and the combinatorial explosion of candidate pairings that would have to be tested. As a result, formulators and consumers rely on heuristics, anecdote, and isolated mechanistic reasoning when deciding which ingredients to pair.

Crucially, the difficulty is not an absolute shortage of information but a mismatch of form, because structured combination outcomes are missing whereas unstructured single-ingredient knowledge—mechanistic literature, INCI listings, patents, and reviews—exists in abundance. The research problem is therefore reframed as whether combination-level synergy and antagonism can be inferred from single-ingredient knowledge that is dispersed across these textual sources.

1.3 Limitations of Existing Approaches

Existing computational and experimental approaches are poorly suited to this setting. Much of the literature focuses on the efficacy of single ingredients and offers little guidance on how two actives behave when combined. Purely data-driven predictors, in turn, presuppose abundant structured training data, an assumption that simply does not hold for cosmetic combinations. Prompting a large language model directly is fluent but prone to fabricating plausible yet ungrounded synergy claims, a failure mode commonly termed hallucination. What is missing is an approach explicitly designed for the absence of structured combination data that nonetheless grounds every inference in retrievable mechanistic evidence.

1.4 Objectives and Scope

This study confines its scope to topical skincare actives and pursues three research questions. The first question (RQ1) asks whether an LLM coupled with a mechanistic knowledge graph can infer ingredient-combination synergy and antagonism on an evidence-grounded basis. The second question (RQ2) asks how the validity of such inferences can be assessed when combination ground-truth labels are absent. The third question (RQ3) asks how much the knowledge-graph coupling reduces hallucinated, ungrounded synergy claims relative to an LLM-only baseline. Addressing these questions yields three contributions: a methodology for data-scarce inference, a validation framework for the no-ground-truth setting, and an explainable recommendation structure that cites its evidence.

2. Materials and Methods

2.1 Overall Pipeline Architecture

The proposed methodology is organized as a four-step pipeline that converts scattered single-ingredient text into combination-level, evidence-grounded predictions (Figure 1). The first step automatically collects single-ingredient literature together with INCI and ingredient-database records. The second step uses an LLM to filter the collected records for mechanistic relevance, and the third step extracts, normalizes, and aggregates mechanistic relations into a directed knowledge graph. The fourth step applies CosmoSyn-GATED, a graph-aware Transformer encoder-decoder that scores combination synergy and antagonism while citing the evidence that supports each prediction. This staged design isolates collection, filtering, structuring, and inference so that each component can be validated and improved independently.

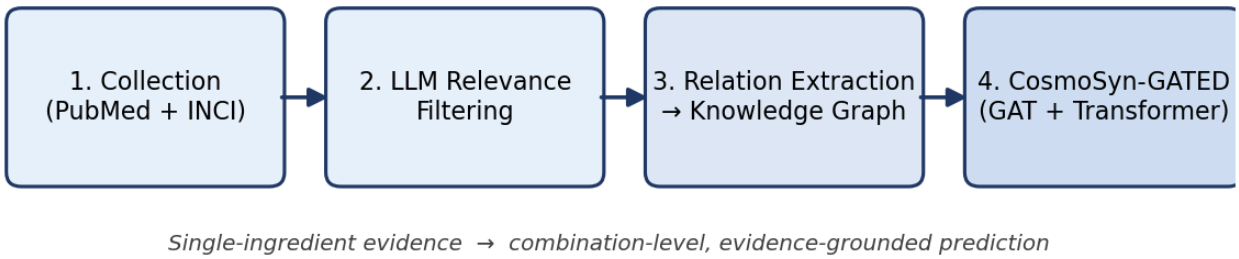


Figure 1. Proposed four-step CosmoSyn-GATED pipeline (study design).

2.2 Literature and INCI Collection

The collection step assembles a corpus of single-ingredient evidence from complementary sources. Documents are gathered from PubMed and cosmetic-science venues, and these are supplemented by INCI-based ingredient dictionaries and public ingredient databases. By deliberate design, combination-level efficacy records are excluded from collection so that the study premise—inferring combinations from single-ingredient knowledge alone—is preserved. This exclusion prevents the model from trivially memorizing known pairings and forces genuine mechanistic inference. The result is a corpus whose unit of evidence is the individual ingredient, mirroring the way knowledge is actually published in the field.

2.3 LLM-Based Relevance Filtering

The filtering step removes records that carry no mechanistic information before the expensive extraction stage. An LLM classifies each abstract according to whether it describes an ingredient's mechanism, biological target, or efficacy, and discards those that do not. This relevance gate mirrors the filtering stage validated in related mechanistic knowledge-graph pipelines and substantially reduces downstream noise. By concentrating subsequent processing on mechanistically informative text, the step also lowers the rate of spurious relations that would otherwise contaminate the graph. The output is a focused corpus suitable for structured relation extraction.

2.4 Mechanistic Relation Extraction, Normalization, and Aggregation

The extraction step transforms unstructured abstracts into a structured mechanistic graph. For each relevant document, the LLM extracts five-tuples of the form (ingredient, mechanism, target, effect, interaction), capturing how an ingredient acts and what it influences. The extracted entities are normalized through synonym resolution and INCI canonicalization so that variant names collapse to a single node, and the relations are aggregated into unique directed edges. Node types comprise ingredient, mechanism, target, and effect, while edge types include acts-on, targets, induces, inhibits, and stabilizes. The resulting directed graph is the substrate on which combination effects are subsequently computed.

2.5 The CosmoSyn-GATED Model

The inference step predicts combination outcomes from the structure of the mechanistic graph. CosmoSyn-GATED integrates an embedding layer, a graph attention network that propagates information across mechanistic neighborhoods, and a Transformer encoder-decoder that reasons over the resulting representations. For an ingredient pair, the model emits a combination label—synergy, antagonism, or neutral—together with the evidence nodes that justify the prediction. Synergy is operationalized as complementary pathways converging on a shared effect, whereas antagonism is operationalized as competing, offsetting, or cumulatively irritating pathways. Because the prediction is tied to explicit evidence nodes, every output is auditable rather than opaque.

2.6 Evaluation under Reference-Set Back-Validation

Evaluation is designed specifically for a setting in which combination ground truth does not exist. The first strategy back-validates the model against a curated reference set of well-established synergistic, antagonistic, and neutral ingredient pairs, detailed in Appendix A. The second strategy measures agreement between model predictions and the labels of dermatologists and formulation experts using Cohen's kappa. The third strategy computes a hallucination rate, defined as the proportion of predictions that assert synergy without supporting evidence nodes. Performance is compared across three systems—an LLM-only baseline, an LLM+RAG baseline, and the proposed LLM-plus-knowledge-graph model.

2.7 Differentiation from Existing Methods

The proposed approach differs from prior work in both its assumptions and its outputs. Unlike single-ingredient efficacy studies, it targets the combination as the unit of prediction, and unlike data-hungry predictors, it is explicitly designed for the absence of structured combination data. Unlike direct LLM prompting, it constrains every assertion to retrievable mechanistic evidence, which makes the reasoning auditable. This evidence-grounding is the central methodological differentiator and the mechanism by which the approach is expected to suppress hallucinated synergy. Taken together, these properties position the method as a decision aid whose conclusions can be inspected and challenged.

3. Expected Results

The quantities reported in this section are explicit expectations derived from the study design rather than measurements from a completed experiment. They are stated so that the hypotheses of the study are concrete and falsifiable once data are collected. Each expected value should therefore be read as a target that subsequent experiments are intended to confirm or refute. Where ranges are given, they reflect the anticipated spread across model variants and reference-set difficulty.

3.1 Expected Literature Collection and Filtering

Application of the collection and filtering steps is expected to yield a large but focused corpus. The PubMed and cosmetic-science query strategy is anticipated to return on the order of tens of thousands of single-ingredient abstracts spanning recent decades of dermatology and cosmetic research. LLM-based relevance filtering is expected to retain roughly ten to fifteen percent of these records as mechanistically relevant (Table 1). The retained subset should provide sufficient coverage of common actives while removing the bulk of non-mechanistic noise. This balance between breadth and focus is a precondition for constructing a reliable mechanistic graph.

Table 1. (EXPECTED / 예상) Anticipated collection, filtering, and graph volumes (illustrative, to be verified).

Stage	Expected volume	Note
Collected single-ingredient abstracts	~30,000–60,000	PubMed + cosmetic-science sources
Mechanistically relevant (after LLM filter)	~10–15%	focused corpus
Unique directed edges (knowledge graph)	~thousands	after normalization & aggregation

3.2 Expected Mechanistic Knowledge Graph Construction

Relation extraction over the filtered corpus is expected to produce a richly connected mechanistic graph. The pipeline should yield on the order of thousands of unique directed edges spanning the full diversity of the cosmetic-stability and efficacy literature. The relation-type distribution is anticipated to be dominated by acts-on,

targets, and stabilizes edges (Table 2), consistent with how mechanisms are typically reported. Hub-node analysis is expected to surface vitamin C, niacinamide, and retinol as high-degree ingredients (Table 3), reflecting their mechanistic centrality. The overall topology, visualized as a synergy-antagonism network in Figure 2, should therefore recapitulate known structure in the field and provide face validity for the extraction process.

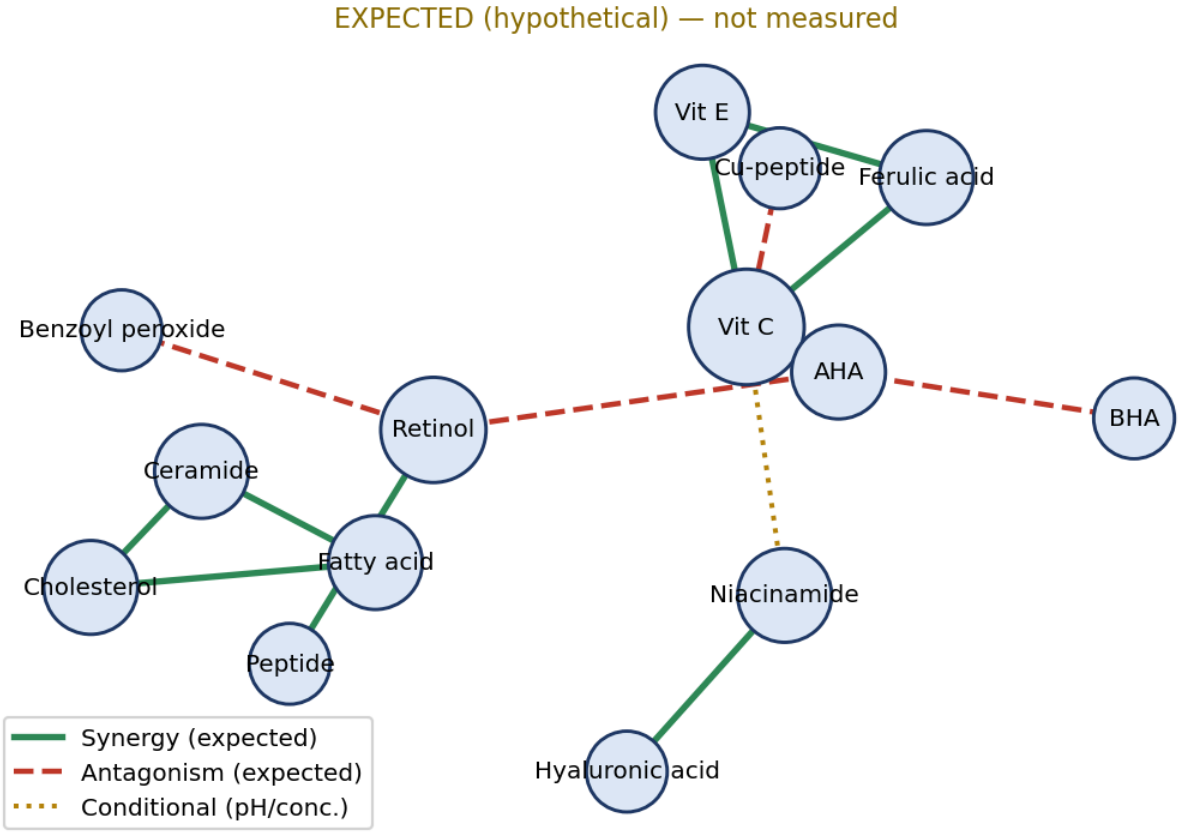


Figure 2. (EXPECTED / 예상) Hypothesized mechanistic synergy-antagonism network. Green edges denote expected synergy, red dashed edges expected antagonism, and dotted edges conditional (pH/concentration-dependent) interactions. Illustrative, not measured.

Table 2. (EXPECTED / 예상) Anticipated relation-type distribution in the mechanistic graph (illustrative).

Relation type	Expected share	Example
acts-on	~35%	retinol acts-on keratinocyte turnover
targets	~25%	vitamin C targets tyrosinase
stabilizes	~15%	ferulic acid stabilizes vitamin C
inhibits	~15%	niacinamide inhibits melanosome transfer
induces	~10%	retinol induces collagen synthesis

Table 3. (EXPECTED / 예상) Anticipated top hub ingredients by node degree (illustrative).

Rank	Ingredient	Expected role
1	Vitamin C (ascorbic acid)	antioxidant hub; many synergies/conflicts
2	Niacinamide	barrier & pigmentation hub
3	Retinol	turnover/collagen hub; irritation conflicts

Rank	Ingredient	Expected role
4	Vitamin E (tocopherol)	antioxidant co-hub
5	Hyaluronic acid	hydration hub

3.3 Expected Model Performance

On the back-validation reference set, the proposed model is expected to deliver strong and well-calibrated classification. We anticipate an F1 score in the range of 0.80 to 0.88, with near-complete identification of strongly evidenced synergistic pairs such as vitamin C with vitamin E and vitamin C with ferulic acid. The model should also separate clear antagonistic pairs, such as retinol with benzoyl peroxide, from neutral combinations with high reliability. Table 4 and Figure 3 summarize the hypothesized comparison across the three systems under evaluation. The expected margin over both baselines is intended to demonstrate that graph structure contributes information beyond what retrieval alone provides.

Table 4. (EXPECTED / 예상) Model comparison on the back-validation reference set (hypothetical values, to be verified).

Model	Accuracy	F1	Hallucination rate ↓	Expert κ
LLM only	0.62	0.60	0.34	0.41
LLM + RAG	0.74	0.73	0.18	0.58
CosmoSyn-GATED (LLM + knowledge graph)	0.85	0.84	0.07	0.71

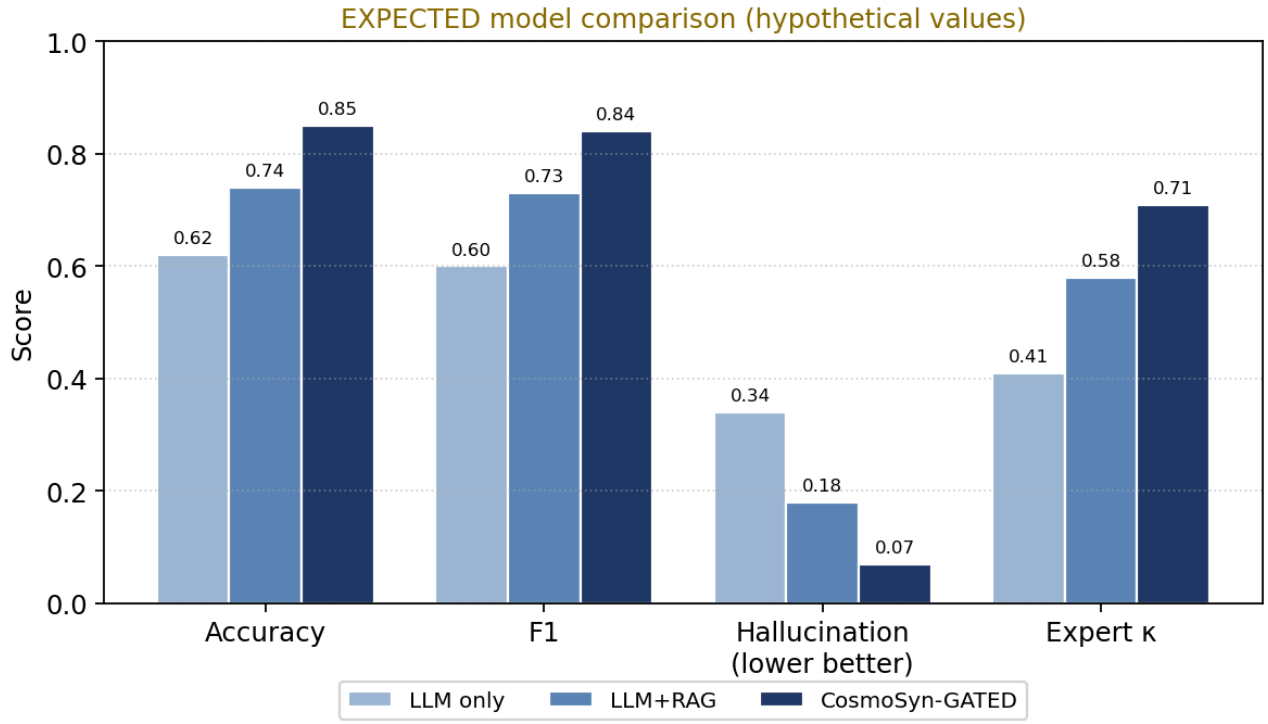


Figure 3. (EXPECTED / 예상) Hypothesized model comparison across evaluation metrics. Values are expectations derived from the study design, not measurements.

3.4 Expected Reduction in Hallucination

A central expectation of the study concerns the suppression of ungrounded synergy claims. Because knowledge-graph coupling forbids any synergy assertion that lacks supporting evidence nodes, the hallucination rate is

expected to fall sharply relative to the LLM-only baseline. Concretely, we anticipate a reduction from approximately 0.34 to about 0.07, roughly a fourfold improvement. This anticipated effect is one of the principal contributions of the proposed design and directly addresses RQ3. If realized, it would show that grounding, rather than model scale alone, is the decisive factor for trustworthy combination inference.

3.5 Expected Error Modes

The study also anticipates where the model is most likely to fail. Conditional interactions that depend on concentration or pH, such as vitamin C combined with niacinamide, are expected to be the hardest cases because the published consensus is itself divided. Weak pathway overlap among simple moisturizing agents may be over-called as mild synergy, inflating false positives in the neutral category. Conversely, antagonisms that are documented only sparsely may be missed, producing false negatives. Characterizing these error modes in advance clarifies which predictions will require expert review and where additional evidence sources would be most valuable.

4. Discussion

4.1 Theoretical and Practical Implications

If the expected results hold, the study would carry implications well beyond cosmetics. Theoretically, it would demonstrate that combination-level mechanistic inference is feasible without combination clinical trials, relying instead on the structured aggregation of single-ingredient evidence. Practically, it would furnish formulators and consumers with an explainable decision aid that states why a given combination is favorable or risky. Because the method generalizes to any domain in which effects are combinatorial but trials are scarce, it could extend to functional foods, botanical actives, and other data-limited areas. The broader contribution is thus a template for evidence-grounded reasoning under structured-data scarcity.

4.2 Limitations

Several limitations qualify the anticipated contributions. First, the inferences depend on the published literature and therefore inherit its biases, gaps, and occasional contradictions. Second, context-dependent factors such as concentration, formulation matrix, and pH are difficult to capture fully from text alone, which constrains the precision of conditional predictions. Third, the expected performance figures are hypotheses rather than measurements, and actual results may fall below them once the system is run. Acknowledging these limitations is essential for interpreting the study's claims responsibly.

4.3 Future Work

The roadmap beyond this proposal targets the limitations directly. A first priority is to secure expert-labeled evaluation data so that agreement metrics rest on a robust human reference. A second priority is to incorporate concentration, pH, and formulation variables so that conditional interactions can be modeled rather than approximated. A third priority is to fuse real-world usage signals from consumer reviews to complement mechanistic evidence with observed outcomes. Pursuing these directions would move the system from a proposal-stage design toward a validated, deployable decision aid.

5. Conclusion

This manuscript proposes an evidence-grounded approach to a problem that has long resisted data-driven solution: predicting the synergy and antagonism of cosmetic ingredient combinations in the absence of combination-level clinical data. The method couples a large language model with a mechanistic knowledge graph so that single-

ingredient evidence can be aggregated into auditable combination-level inferences. To cope with the lack of ground truth, evaluation relies on a curated reference set, expert agreement, and an explicit hallucination metric, and the expected results indicate strong classification performance with markedly reduced ungrounded claims. Should these expectations be confirmed, the study would establish a reusable methodology for evidence-grounded inference in data-scarce domains and offer a practical, explainable tool for cosmetic formulation.

Software and Data Availability

All materials needed to reproduce the study will be released upon its completion. The pipeline notebooks covering collection, filtering, extraction, and modeling will be made openly available, as will the curated reference test set used for back-validation. An interactive visualization of the mechanistic synergy network will accompany the code to support inspection and community adoption. Together these artifacts are intended to make both the method and its evaluation fully transparent and extensible.

References

References below are provided to illustrate the expected bibliography. Several are well-established sources; bibliographic details and DOIs must be verified before submission.

1. Lin, J. Y., Selim, M. A., Shea, C. R., Grichnik, J. M., Omar, M. M., Monteiro-Riviere, N. A., & Pinnell, S. R. (2003). UV photoprotection by combination topical antioxidants vitamin C and vitamin E. Journal of the American Academy of Dermatology, 48(6), 866–874.
2. Lin, F. H., Lin, J. Y., Gupta, R. D., Tournas, J. A., Burch, J. A., Selim, M. A., Monteiro-Riviere, N. A., Grichnik, J. M., Zielinski, J., & Pinnell, S. R. (2005). Ferulic acid stabilizes a solution of vitamins C and E and doubles its photoprotection of skin. Journal of Investigative Dermatology, 125(4), 826–832.
3. Pinnell, S. R. (2003). Cutaneous photodamage, oxidative stress, and topical antioxidant protection. Journal of the American Academy of Dermatology, 48(1), 1–19.
4. Bissett, D. L., Oblong, J. E., & Berge, C. A. (2005). Niacinamide: A B vitamin that improves aging facial skin appearance. Dermatologic Surgery, 31(7 Pt 2), 860–865.
5. Telang, P. S. (2013). Vitamin C in dermatology. Indian Dermatology Online Journal, 4(2), 143–146.
6. Mukherjee, S., Date, A., Patravale, V., Korting, H. C., Roeder, A., & Weindl, G. (2006). Retinoids in the treatment of skin aging: An overview of clinical efficacy and safety. Clinical Interventions in Aging, 1(4), 327–348.
7. Draelos, Z. D. (2018). The science behind skin care: Moisturizers. Journal of Cosmetic Dermatology, 17(2), 138–144.
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
9. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations (ICLR).
10. Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33.
11. Edge, D., Trinh, H., Cheng, N., et al. (2024). From local to global: A graph RAG approach to query-focused summarization. arXiv:2404.16130.

Appendix A. 파일럿 검증용 정답 세트 (초안)

정형 조합 데이터가 없는 환경에서 모델을 검증하려면, 정답의 역할을 대신할 신뢰 가능한 기준 집합이 필요하다. 이를 위해 문헌적으로 합의 수준이 높은 시너지·충돌·무관계 성분쌍을 선별하여 역검증 테스트셋

으로 구성한다. 아래 표는 그 초안으로, 각 쌍에는 분류 근거를 요약해 함께 제시하였다. 이 집합은 모델의 분류 성능과 환각율을 측정하는 1차 기준으로 사용되며, 전문가 검토를 거쳐 확장·정제될 예정이다.

분류	성분쌍	근거 (요약)
시너지	비타민C + 비타민E	항산화 재생 순환
시너지	비타민C + 페룰산	산화 안정화
시너지	나이아신아마이드 + 히알루론산	장벽 강화 + 보습
시너지	레티놀 + 펩타이드	콜라겐 합성 보완
시너지	세라마이드 + 콜레스테롤 + 지방산	장벽 지질 비율
충돌	레티놀 + AHA	자극 누적
충돌	레티놀 + 벤조일퍼옥사이드	레티놀 분해
충돌	비타민C(저pH) + 나이아신아마이드	pH 충돌 가능성
충돌	AHA + BHA(고농도)	과각질박리
충돌	구리펩타이드 + 비타민C	산화 상호작용
무관계	글리세린 + 정제수	단순 보습 기제
무관계	(대조용 성분쌍 5종)	유의 상호작용 없음