의미와 구조의 통합: AI에서의 벡터 임베딩과 지식 그래프 연계에 대한 종합적 분석

- 기획: 페블러스
- 생성: Gemini 2.5 Pro DeepResearch
- 생성일: 2025-09-25

Executive Summary

본 보고서는 인공지능(AI) 시스템에서 두 가지 핵심 지식 표현 패러다임인 벡터 임베딩과 온톨로지 기반 지식 그래프의 통합 방법론, 전략적 중요성, 그리고 미래 전망을 종합적으로 분석한다. 벡터 임베딩은 비정형 데이터로부터 암묵적인 의미 관계를 학습하는 연속적인 벡터 공간 표현이며, 지식 그래프는 명시적인 사실과 구조적 관계를 정의하는 기호적 웹이다. 이 두 기술의 융합은 단순히 성능을 개선하는 차원을 넘어, 더 강력하고 설명 가능하며 사실에 기반한 AI를 구축하기 위한 필수적인 진화 과정이다. 보고서는 파이프라인 방식의 초기 연계부터 심층적으로 통합된 지식 인지 언어 모델에 이르기까지의 기술적 발전 과정을 추적하고, 이러한 통합을 가능하게 하는 핵심 아키텍처 혁신과 여전히 남아있는 과제들을 심도 있게 다룬다. 결론적으로, 암묵적 의미 표현(벡터 임베딩)과 명시적 구조 지식(지식 그래프)의 결합은 차세대 AI 시스템의 신뢰성과 추론 능력을 좌우하는 핵심 동력이 될 것이다.

제1장: 지식 표현의 두 가지 패러다임

현대 AI에서 지식은 두 가지 근본적으로 다른 방식으로 표현된다. 하나는 데이터의 통계적 패턴에서 의미를 학습하는 방식이며, 다른 하나는 세상의 사실과 관계를 명시적으로 구조화하는 방식이다. 이 두 패러다미임의 개별적인 강점과 약점을 이해하는 것은 이들의 통합이 왜 중요한지를 파악하는 데 필수적인 전제 조건이다.

1.1. 연속적 벡터 공간: 임베딩을 통한 암묵적 지식

벡터 임베딩은 텍스트, 이미지, 그래프 노드와 같은 복잡하고 비정형적인 데이터를 저차원의 조밀한(dense) 숫자 벡터로 변환하는 기술이다.1 이 기술의 핵심 목적은 인간이 이해하는 콘텐츠와 기계가 처리할 수 있는 숫자 형식 사이의 간극을 메우는 것으로, 의미적 관계를 보존하면서 계산적으로 효율적인 수학적 연산을 가능하게 한다.4 여기서 '벡터화(vectorization)'는 데이터를 벡터 형식으로 변환하는 일반적인 행위를 의미하는 반면, '임베딩(embedding)'은 의미론적 관계를 포착하는 특정 목표를 가진 변환을 지칭한다.4

이 과정은 대규모 데이터 코퍼스에서 신경망과 같은 모델을 훈련시켜 작동한다. 모델은 비슷한 의미나 맥락을 가진 데이터 포인트를 고차원 벡터 공간에서 서로 가깝게 배치하도록 학습한다.2 이를 통해 언어적 또는 시각적 유사성은 기하학적 근접성으로 변환된다.5 Word2Vec, GloVe와 같은 단어 임베딩 기법이

나 Universal Sentence Encoder (USE)와 같은 문장 인코더는 텍스트 내 단어들의 동시 등장 패턴으로부터 이러한 관계를 학습하는 대표적인 예이다.5

임베딩은 의미적 유사성이 중요한 작업에서 탁월한 성능을 보인다. 예를 들어, 시맨틱 검색에서 "강아지"를 검색했을 때 "개"나 "반려견" 관련 결과를 찾아주거나, 추천 시스템에서 유사한 상품을 제안하고, 감성 분석, 데이터 클러스터링 등에 널리 활용된다.3 임베딩의 강점은 방대한 비정형 데이터로부터 미묘하고 암묵적인 패턴과 관계를 포착하는 능력에 있다.4

하지만 이러한 강력함에도 불구하고, 임베딩은 본질적으로 불투명한 '블랙박스' 모델이며 명시적인 사실 기반이 부족하다. 두 객체가 왜 유사한지, 혹은 그들 사이의 정확한 관계가 무엇인지를 설명할 수 없으며, 단지 유사성을 표현할 뿐이다.7 따라서 검증 가능한 사실적 정확성이나 여러 단계를 거치는 명시적 추론(multi-hop reasoning)이 요구되는 작업에는 한계를 보인다.

1.2. 기호적 웹: 온톨로지와 지식 그래프를 통한 명시적 지식

온톨로지(Ontology)는 특정 도메인 내에 존재하는 개체(entity)의 유형(class), 속성(property), 그리고 그들 간의 관계를 명시적으로 정의하는 정형화된 스키마 또는 청사진이다.9 반면, 지식 그래프(Knowledge Graph, KG)는 이러한 온톨로지 스키마에 실제 데이터를 채워 넣어 구체화한 것으로, 상호 연결된 개체(노드)와 그들의 관계(엣지)로 구성된 네트워크이다.12

온톨로지 + 데이터 = 지식 그래프라는 공식은 이 관계를 명확히 보여준다.12

지식 그래프는 RDF(Resource Description Framework), OWL(Web Ontology Language)과 같은 표준을 사용하여 지식을 구조화하고 표준화된 프레임워크를 제공한다.9 이러한 구조는 인간과 기계가 정보를 일관되게 해석하고 추론할 수 있도록 돕는다.9 온톨로지는 데이터의 일관성을 강제하고, "A가 B의 형제이면, B도 A의 형제이다"와 같은 논리적 추론을 가능하게 하며, 공유된 어휘 체계를 통해 서로 다른 데이터 사일로를 허물어준다.9

지식 그래프는 구조적 추론, 복잡한 질의 처리, 이종 데이터 소스의 통합이 필요한 작업에 매우 강력하다.10 또한, 그래프 구조를 통해 추론 경로를 추적할 수 있어 AI의 의사결정 과정에 대한 설명 가능성을 제공한다.14 지식 그래프의 핵심 강점은 검증 가능하고 명시적인 사실과 복잡한 의존 관계를 정확하게 표현하는 데 있다.8

그러나 지식 그래프는 유연성이 부족하여 미묘한 의미적 뉘앙스나 모호성을 다루는 데 어려움을 겪을 수 있다. 또한 구축 및 유지 관리에 많은 계산 비용과 수작업 기반의 특징 공학(feature engineering)이 요구되기도 한다.15 그래프 상에서 직접적으로 연결되지 않은 개체들 간의 의미적 유사성을 판단하는 데에도 취약하다.7

이처럼 두 패러다임은 근본적으로 다른 지식 철학을 대변한다. 벡터 임베딩은 데이터의 통계적 패턴에서 귀납적으로 학습된 **암묵적 지식**을 나타내는 반면, 지식 그래프는 인간이 정의한 형식적 규칙과 사실에 기반한 **명시적 지식**을 나타낸다. 이는 AI 역사에서 오랫동안 이어진 연결주의(connectionist)와 기호주의(symbolic) 접근 방식 간의 논쟁을 반영한다. AI의 미래는 둘 중 하나를 선택하는 것이 아니라, 직관적으로 관계를 파악하고 논리적으로 추론할 수 있는 시스템, 즉 두 패러다임의 통합에 달려 있다.

표 1: 지식 표현 패러다임 비교 분석

특징	벡터 임베딩	지식 그래프
표현 형식	조밀한 수치 벡터 (Dense vectors)	노드와 엣지 (Nodes/Edges)
핵심 강점	의미적 유사성 (Semantic similarity)	명시적 추론 (Explicit reasoning)
내재적 약점	불투명성 (Opaqueness)	경직성 (Rigidity)
주요 사용 사례	시맨틱 검색, 추천 시스템	복잡한 질의, 데이터 통합
지식 유형	암묵적/통계적 (Implicit/Statistical)	명시적/기호적 (Explicit/Symbolic)
대표 기술	BERT, Word2Vec, USE	Neo4j, RDF/OWL

제2장: 상호보완적 결합의 필요성

벡터 임베딩과 지식 그래프의 통합은 단순히 점진적인 성능 향상을 넘어, 각 패러다임의 근본적인 한계를 극복하고 새롭고 강력하며 신뢰할 수 있는 AI 시스템을 구현하기 위한 필수적인 단계이다.

2.1. 상호 강점 활용 및 약점 보완

두 접근법의 시너지는 각 기술의 약점을 서로의 강점으로 보완하는 데서 비롯된다. 벡터 데이터베이스는 의미적으로 유사한 항목을 찾는 데 뛰어나지만, 맥락과 관계를 파악하는 데는 취약하다. 반면, 그래프 데이터베이스는 맥락과 관계를 다루는 데 강하지만, 의미적 유사성을 판단하는 데는 어려움을 겪는다.7

임베딩은 지식 그래프에 '수치적 의미론(numerical semantics)'이라는 새로운 층을 추가하여, 그래프 구조만으로는 어려운 유사성 검색이나 머신러닝 모델 통합과 같은 작업을 가능하게 한다.16 예를 들어, "베를린이 독일에 대해 갖는 관계는 파리가 프랑스에 대해 갖는 관계와 유사하다"와 같은 유추적 관계를 벡터 거리 비교를 통해 찾아낼 수 있는데, 이는 표준적인 그래프 탐색으로는 불가능한 작업이다.16

반대로, 지식 그래프는 임베딩의 품질과 정확성을 향상시키는 구조적 맥락을 제공한다. 그래프 기반의 제약 조건은 임베딩이 계층적이거나 논리적인 관계를 존중하도록 보장할 수 있다 (예: 'Apple Inc.'의 임베딩이 '사과(과일)'의 임베딩과 지나치게 가까워지는 것을 방지).16

이러한 결합은 벡터 공간에서의 '빠른 사고'(유사성 검색)와 지식 그래프를 통한 '느리고 깊은 사고'(논리적 추론)를 결합한 하이브리드 지능 시스템을 구현할 수 있게 한다.8

2.2. AI 역량 강화: 실질적 영향

이러한 통합은 AI 시스템의 역량을 실질적으로 향상시키는 데 기여하며, 특히 대규모 언어 모델(LLM)의 한계를 극복하는 데 중요한 역할을 한다.

- **LLM의 환각(Hallucination) 현상 완화:** LLM이 사실과 다른 정보를 생성하는 경향을 완화하는 것이 통합의 주된 동기 중 하나이다. LLM의 응답을 검증 가능한 지식 그래프에 기반하게 함으로써, 시스템은 더 정확하고 최신의 신뢰할 수 있는 정보를 제공할 수 있다.14 이는 검색 증강 생성(Retrieval-Augmented Generation, RAG) 시스템의 핵심 원리이며, 최근에는 그래프 구조로 강화된 GraphRAG 형태로 발전하고 있다.14
- **복잡한 다단계 추론(Multi-Hop Reasoning) 활성화:** 이 결합은 동적이고 적응적인 질의를 가능하게 한다. 사용자는 먼저 벡터를 사용한 광범위한 의미 검색으로 시작한 다음, 그래프를 사용하여 여러 단계의 복잡한 관계를 탐색하며 더 깊은 통찰력을 발견할 수 있다.7 이는 벡터 검색만으로는 불가능한 기능이다.
- **설명 가능성 및 신뢰성 향상:** 그래프 구조는 AI의 추론 과정을 추적 가능하게 만들어, 순수 딥러닝 모델의 '블랙박스' 특성에서 벗어날 수 있게 한다. 이러한 투명성은 의료, 금융과 같이 민감한 분야의 애플리케이션에서 매우 중요하다.14
- **정형 및 비정형 데이터의 통합:** 하이브리드 접근법은 전통적인 정형 데이터베이스(지식 그래프를 통해)와 비정형 텍스트, 이메일, 문서(임베딩을 통해)를 통합하여 포괄적인 지식 표현을 생성한다.12

이러한 기술적 융합은 한때 학문적 개념이었던 '신경-기호(Neuro-Symbolic) AI'가 주류 솔루션으로 부상하는 현상을 보여준다. 임베딩을 생성하는 신경망과 지식 그래프라는 기호적 구조의 결합은 신경-기호 AI의 실용적인 구현체이다.8 LLM의 신뢰성과 설명 가능성을 높여야 하는 산업계의 시급한 요구는, 과거 연구실 수준에 머물렀던 신경-기호 아키텍처를 응용 AI의 핵심으로 이끌어냈다. 이 시너지는 단순히 기존 작업의 성능을 향상시키는 것을 넘어, 고위험 애플리케이션에 요구되는 새로운 차원의 신뢰와 추론을 가능하게 한다.

제3장: 임베딩과 지식 그래프 융합 방법론 분류

벡터 임베딩과 지식 그래프를 결합하는 전략은 순차적이고 느슨하게 결합된 방식에서부터 종단간(end-to-end) 아키텍처로 심층 통합된 방식에 이르기까지 다양하게 분류할 수 있다. 이러한 방법론의 발전 과정은 데이터 통합이라는 초기 목표에서 벗어나, 강력한 AI 모델의 인지 능력을 증강하는 방향으로 진화해 왔다.

3.1. 파이프라인 접근법: 순차적 생성

초기 통합 방식은 두 기술을 순차적으로 적용하는 파이프라인 형태를 띤다. 이는 두 세계를 분리된 단계로 간주하고, 한쪽의 출력을 다른 쪽의 입력으로 사용하는 방식이다.

- **그래프 우선(KG → 임베딩):** 이 접근법은 먼저 기존의 지식 그래프를 구축하거나 활용한다. 그 후, TransE, DistMult, ComplEx와 같은 지식 그래프 임베딩(KGE) 모델이나 그래프 신경망(GNN)을 사용하여 그래프 내의 개체와 관계에 대한 벡터 표현을 학습한다.17 이렇게 생성된 임베딩은 링크 예측이나 개체 분류와 같은 후속 작업에 사용된다.18 이 방법은 기존의 구조화된 지식을 풍부하게 만드는 데 효과적이지만, 초기 그래프에 존재하지 않는 개체를 처리하는 데 어려움을 겪는다(데

이터 희소성 문제).15

- **텍스트 우선(임베딩 → KG):** 이 방식은 비정형 텍스트에서 시작하여, 자연어 처리(NLP) 모델을 사용해 개체와 관계를 추출하고, 이를 바탕으로 지식 그래프를 구축한다. 벡터 임베딩은 개체 연결(entity linking)이나 관계 분류와 같은 중간 단계에서 개념을 식별하고 모호성을 해소하여 그래프에 추가하기 전에 사용된다. 이는 종단간 지식 그래프 구축 파이프라인에서 흔히 사용되는 접근법이다.20

3.2. 공동 표현 학습: 동시 생성

이 패러다임은 텍스트 요소와 지식 그래프 요소를 위한 통합된 벡터 공간을 동시에 학습하는 종단간 모델을 만드는 데 중점을 둔다. 모델은 지식 그래프 구조와 텍스트 문맥과 관련된 목적 함수를 공동으로 최적화하여 훈련된다.22

JKRL(Joint Knowledge Representation Learning)과 같은 모델은 각 개체에 대해 두 가지 유형의 표현, 즉 그래프 구조에서 학습된 표현(예: TransE 사용)과 텍스트 설명에서 학습된 표현(예: CNN 사용)을 명시적으로 학습한 후 이를 결합한다.25 이 방법의 핵심 장점은 한 양식(modality)의 정보가 다른 양식의 표현을 개선하는 데 도움을 줄 수 있다는 점이다. 예를 들어, 풍부한 텍스트 설명은 그래프에서 연결이 거의 없는 개체를 임베딩하는 데 도움을 주고(희소성 문제 해결), 그래프 구조는 텍스트에 언급된 개체의 모호성을 해소하는 데 기여한다.23 이러한 모델들은 종종 텍스트, 숫자, 이미지 등 다양한 데이터 유형의 특징을 그래프 구조와 함께 처리할 수 있는 복잡한 다중 모드 메시지 전달 네트워크(multimodal message-passing network)를 포함한다.27

3.3. 지식 주입: 사전 훈련된 언어 모델 증강

이는 현대적인 접근법의 주류로, 처음부터 통합된 임베딩 공간을 만드는 것보다 BERT나 GPT와 같은 강력한 사전 훈련된 언어 모델(PLM)을 지식 그래프의 구조적 지식으로 강화하는 것을 목표로 한다.29 이러한 '지식 강화 사전 훈련 언어 모델(KEPLM)'은 PLM의 사실적 정확성, 추론 능력, 그리고 해석 가능성을 향상시키는 것을 목적으로 한다.29

지식 주입 방법은 다음과 같이 크게 분류할 수 있다:

- **지식 인지 사전 훈련 (Knowledge-Aware Pre-training):** 언어 모델의 사전 훈련 목표 자체를 수정하여 개체와 관계를 인지하도록 만든다. ERNIE나 KALM과 같은 모델은 사전 훈련 중에 개체 수준 마스킹이나 다음 개체 예측과 같은 작업을 추가한다.33 이는 지식을 모델 파라미터에 깊숙이 통합하지만 계산 비용이 매우 높다.
- **어댑터 기반 주입 (Adapter-Based Injection):** PLM 아키텍처 내부에 지식 그래프의 정보를 처리하고 융합하도록 특별히 설계된 '어댑터' 레이어나 모듈을 삽입한다. KnowBERT의 KAR(Knowledge Attention and Recontextualization) 모듈이 대표적인 예이다.34 이는 전체 재훈련보다 덜 파괴적인 모듈식 접근법이다.
- **입력 수준 주입 (Input-Level Injection):** 추론 시점에 PLM의 입력을 수정하는 방식이다. 이는 지식 그래프의 트리플(triple)을 텍스트화하여 프롬프트에 추가하거나(일반적인 RAG 기법), 더 나아가 ConceptFormer와 같은 모델에서처럼 지식 벡터를 임베딩 레이어에 직접 주입하는 고급 기

법을 포함한다.37 이 방식은 모델 재훈련이 필요 없어 유연성이 높다.

이러한 방법론들의 발전 과정은 명확한 궤적을 그린다. 초기 파이프라인 방식은 임베딩과 지식 그래프를 단순히 결합해야 할 두 개의 데이터셋으로 취급했다. 이후 공동 학습 단계에서는 두 요소가 동시에 학습되어 시너지를 창출하는 통합 모델로 발전했다. 그리고 대규모 PLM의 등장과 함께 지식 주입이라는 새로운 패러다임이 부상했는데, 이는 강력한 사전 훈련된 '뇌'(LLM)에 검증 가능한 구조적 '기억'(KG)을 연결하여 인지 능력을 증강하는 개념이다. 이는 지식 표현의 초점이 단순한 데이터 통합에서 벗어나, 추론과 증강이라는 더 높은 수준의 목표로 이동했음을 시사한다.

표 2: 통합 방법론의 분류

방법론	설명	주요 특징	대표 모델/논문
파이프라인 (그래프 우선)	KG를 먼저 구축하고, 그로부터 임베딩을 생성	구조 기반, 기존 KG 활용에 용이	TransE, GNNs on KGs
파이프라인 (텍스트 우선)	텍스트에서 개체/관계를 추출하여 KG를 구축	NLP 기반, 비정형 데이터로부터 KG 구축	SciNLP-KG
공동 표현 학습	텍스트와 KG를 동시에 학습하여 통합 벡터 공간 생성	종단간 학습, 상호 보완적 학습	JKRL, Multimodal Message Passing Nets
지식 주입 (사전 훈련)	LM의 사전 훈련 목표를 수정하여 지식을 내재화	심층 통합, 높은 계산 비용	ERNIE, KALM
지식 주입 (어댑터 기반)	LM 아키텍처에 지식 융합 모듈 삽입	모듈식 접근, 유연성	KnowBERT
지식 주입 (입력 수준)	추론 시 프롬프트를 지식으로 증강	모델 재훈련 불필요, 높은 유연성	ConceptFormer, GraphRAG

제4장: 아키텍처 심층 분석: 통합의 메커니즘

이 장에서는 벡터 임베딩과 지식 그래프의 융합을 가능하게 하는 핵심 아키텍처 구성 요소와 모델들을 기술적으로 상세히 분석한다. 특히, 그래프 인코딩 방식의 발전과 트랜스포머의 어텐션 메커니즘이 어떻게 두 세계를 잇는 다리 역할을 하는지에 초점을 맞춘다.

4.1. 그래프 인코딩: 변환 모델에서 메시지 전달까지

- **변환 기반 모델 (Translation-Based Models):** 초기 지식 그래프 임베딩 모델들은 관계(r)를 머리(h)와 꼬리(t) 개체 임베딩 간의 변환 벡터로 모델링했다. 대표적인 **TransE**는 h+r≈t 라는 간단한 수식을 통해 관계를 벡터 공간에서의 이동으로 표현했다.23 이 모델은 계산적으로 효율적이지만, 다대다(many-to-many)나 비대칭(asymmetric)과 같은 복잡한 관계를 잘 표현하지 못하는 한계가 있었다. 이러한 단점을 보완하기 위해 행렬이나 복소수 값을 활용하는 **DistMult**, **ComplEx**와 같은 확장 모델들이 제안되었다.17 이 모델들은 단일 홉(single-hop) 관계를 포착하는 데 중점을 둔다.
- **그래프 신경망 (Graph Neural Networks, GNNs):** 현대적인 그래프 위상 인코딩의 표준은 GNN이다. GNN의 핵심 패러다임은 **메시지 전달(message-passing)**로, 각 노드가 이웃 노드들의 특징 벡터를 반복적으로 집계하여 지역적 및 전역적 그래프 구조를 모두 포착하는 표현을 학습하는 방식이다.40 그래프 합성곱 신경망(GCN)이나, 어텐션 메커니즘을 도입하여 이웃 노드의 중요도를 가중치로 부여하는 그래프 어텐션 네트워크(GAT)와 같은 특정 아키텍처들이 널리 사용된다.40 GNN은 언어 모델에 풍부한 구조 인식 노드 임베딩을 제공하는 데 결정적인 역할을 한다.43

4.2. 트랜스포머에서의 지식 융합: 어텐션 브릿지

트랜스포머 아키텍처, 특히 셀프 어텐션 메커니즘은 서로 다른 양식의 데이터를 융합하는 핵심적인 역할을 수행한다. 이는 단순히 문장 내 단어 간의 관계를 넘어, 텍스트와 외부 지식, 또는 그래프 내 개체 간의 관계를 모델링하는 유연한 도구로 활용된다.

- **ERNIE (Baidu):** 바이두의 ERNIE(Enhanced Representation through Knowledge Integration)는 사전 훈련 과정에서 어휘, 구문, 지식 정보를 통합한다. 이 모델은 텍스트 인코더(T-Encoder)와 지식 인코더(K-Encoder)의 이중 인코더 구조를 사용한다. K-Encoder는 텍스트 표현과 지식 그래프에서 가져온 해당 개체 임베딩을 입력받아, 어텐션 기반의 통합기(aggregator)를 통해 이를 하나의 통일된 표현으로 융합한다.34 최근의 ERNIE 4.5는 텍스트와 비전 같은 서로 다른 양식을 위한 개별 전문가(expert)와 이들을 통합하는 공유 전문가를 함께 사용하는 이종 전문가 혼합(Mixture-of-Experts, MoE) 아키텍처를 도입하여 효율성을 높였다.46
- **KnowBERT와 KAR 모듈:** KnowBERT의 핵심은 **지식 어텐션 및 재맥락화(Knowledge Attention and Recontextualization, KAR)** 모듈이다.34 이 모듈은 다음과 같은 단계로 작동한다: 1) 개체 연결기(entity linker)가 텍스트 내에서 잠재적인 개체 언급(mention)을 식별한다. 2) 위키피디아나 워드넷 같은 지식 베이스에서 해당 개체의 임베딩을 검색한다. 3) 새롭게 도입된 **'단어-개체 어텐션(word-to-entity attention)'** 메커니즘을 사용하여, 트랜스포머 레이어의 단어 표현이 개체 표현에 직접 주의를 기울이도록 한다.36 이를 통해 단어 임베딩이 명시적인 지식으로 재맥락화되며, 기존 BERT 아키텍처를 크게 변경하지 않으면서도 지식을 효과적으로 주입한다.
- **KnowFormer: KG 추론을 위한 트랜스포머:** KnowFormer는 트랜스포머 아키텍처를 텍스트 처리가 아닌 지식 그래프 추론 자체에 적용한다.50 이 모델은 셀프 어텐션 메커니즘을 개체 쌍 간의 상호작용을 직접 계산하도록 재정의한다. 관계형 메시지 전달(relational message passing)에 기반한 구조 인식 모듈을 도입하여 쿼리(query), 키(key), 값(value) 벡터를 계산함으로써, 그래프의 위상 정보를 어텐션 계산에 직접 포함시킨다.52 이는 트랜스포머의 활용을 그래프에 *대한* 텍스트 이해에서 그래프 *위에서의* 추론으로 전환시킨다.

4.3. 새로운 주입 기법: 임베딩 공간에서의 연산

- **ConceptFormer:** ConceptFormer는 지식을 텍스트화하는 과정을 완전히 배제하는 혁신적인 접근법을 제시한다.38 이 모델은 1) 프롬프트에 언급된 개체에 대해 지식 그래프에서 관련 하위 그래프(subgraph)를 추출하고, 2) 훈련된 트랜스포머 기반 모듈을 사용하여 이 하위 그래프를 압축된 '개념 벡터(concept vector)'로 변환한다. 3) 마지막으로, 이 개념 벡터를 원래의 토큰 임베딩과 함께 LLM의 입력 임베딩 시퀀스에 직접 주입한다. 이 방식은 토큰 효율성이 매우 높아, 소규모 LLM이 벡터라는 고유의 언어로 구조화된 지식에 직접 접근함으로써 훨씬 더 큰 모델에 필적하는 사실 리콜 능력을 갖추게 한다.38

4.4. 검색 증강 아키텍처: GraphRAG 프레임워크

GraphRAG는 표준 RAG를 정교하게 발전시킨 아키텍처로, 지식의 구조적 특성을 검색 과정에 적극적으로 활용한다.56

- **그래프 구축:** 문서 코퍼스를 인덱싱하여 개체와 관계를 자동으로 추출하고 지식 그래프를 구축하는 것으로 시작한다.57 이 과정은 LLM을 사용하거나 더 효율적인 의존 구문 분석(dependency parsing) 기법을 활용할 수 있다.56
- **계층적 커뮤니티 탐지:** 구축된 그래프는 라이덴(Leiden) 클러스터링과 같은 알고리즘을 사용하여 계층적인 의미론적 커뮤니티로 분할된다. 각 커뮤니티에 대해 수준별 요약이 생성되어, 전체 데이터셋에 대한 다층적이고 추상적인 표현을 만든다.57
- **질의 처리:** 질의 시 GraphRAG는 상황에 따라 다른 전략을 사용한다. "전체 데이터셋에 대한 포괄적인 질문(global search)"에는 커뮤니티 요약을 활용하고, "특정 개체에 대한 질문(local search)"에는 해당 개체 노드에서 시작하여 이웃 노드로 그래프를 탐색하며 관련 맥락을 수집한다.57 이러한 구조적 검색은 단순한 벡터 유사성 검색보다 훨씬 정밀하며, 검색된 구조적 맥락은 LLM의 프롬프트를 증강하는 데 사용된다.

이러한 다양한 아키텍처들은 트랜스포머의 어텐션 메커니즘이 얼마나 유연하고 강력한지를 보여준다. 어텐션은 문장 내 단어 간의 관계뿐만 아니라, 단어와 개체(KnowBERT), 개체와 개체(KnowFormer), 그리고 그래프 구조와 프롬프트(ConceptFormer) 간의 관계를 융합하는 '범용 어댑터' 역할을 수행하고 있다. 이 적응성 덕분에 트랜스포머는 현대 지식 강화 AI의 중추적인 아키텍처로 자리 잡았다.

표 3: 주요 지식 주입 모델의 아키텍처 비교

모델	기반 언어 모델	지식 소스	주입 메커니즘	핵심 혁신
ERNIE	BERT 계열	KG 트리플	사전 훈련 중 K-Encoder로 융합	개체 수준 마스킹, 지식 인지 사전 훈련
KnowBERT	BERT	위키피디아/워	KAR(지식 어텐션 및 재맥락화) 모듈	단어-개체 어텐션

		드넷		
ConceptFormer	GPT-2 (또는 모든 LM)	KG 하위 그래프	'개념 벡터'를 임베딩 공간에 직접 주입	토큰 효율적, 지식의 텍스트화 배제
GraphRAG	모든 LLM	문서 코퍼스	자동 구축된 KG와 커뮤니티 요약을 통한 구조적 검색	계층적, 다단계 맥락 검색

제5장: 실제 적용 사례 및 전략적 구현

이 장에서는 앞서 논의된 기술적 개념들을 실제 세계의 구체적인 사례 연구에 적용하여, 주요 기업과 연구 기관들이 어떻게 하이브리드 시스템을 활용하여 중요한 비즈니스 및 연구 문제를 해결하고 있는지를 보여준다. 이 사례들은 공통적으로 분산된 데이터 사일로를 극복하고 데이터 자산에 잠재된 가치를 발굴하는 것을 핵심적인 전략적 목표로 삼고 있다.

5.1. 정보 검색의 혁신: 구글의 시맨틱 검색

구글 검색 엔진은 키워드 매칭 시스템에서 방대한 **지식 그래프**에 기반한 시맨틱 검색 엔진으로 진화했다.58 이 변화의 핵심 철학은 '문자열이 아닌 사물(things, not strings)'로, 시스템이 사용자 질의를 단순한 단어의 나열이 아닌 실제 세계의 개체와 그들의 관계로 이해하는 것을 의미한다.59

구글의 지식 그래프는 위키피디아, 프리베이스(Freebase), CIA 월드 팩트북 등 다양한 출처의 데이터를 통합하여 세계 지식에 대한 통일된 뷰를 구축한다.59 이를 통해 구글은 지식 패널(Knowledge Panel) 형태로 직접적인 답변과 요약을 제공하고, 사용자의 위치나 이전 검색 기록과 같은 맥락을 이해하여 키워드 검색만으로는 불가능했던 훨씬 더 관련성 높은 결과를 제공할 수 있다.58 이는 대규모 지식 그래프가 핵심 정보 검색 작업을 어떻게 향상시키는지를 보여주는 대표적인 사례이다.

5.2. 대규모 개인화: 아마존의 추천 시스템

아마존은 지식 그래프와 임베딩을 결합하여 자사의 상품 추천 엔진을 구동한다.62 지식 그래프는 사용자, 상품, 브랜드, 카테고리뿐만 아니라 상품의 기능이나 대상 고객과 같은 문맥 정보를 포함하는 복잡한 관계를 모델링하는 데 사용된다.63 이러한 구조적 정보는 추천 시스템에서 흔히 발생하는 '콜드 스타트(cold start)' 문제나 데이터 희소성 문제를 완화하는 데 도움을 준다.62

아마존은 이 그래프 구조를 기반으로 GNN을 사용하여 상품과 사용자에 대한 임베딩을 생성한다. 이 임베딩은 개별적인 선호도뿐만 아니라, '함께 자주 구매되는 상품'과 같이 이웃 노드의 영향까지 포착하여 더 미묘하고 정확한 추천을 가능하게 한다.42 아마존이 개발한 DGL-KE 라이브러리는 이러한 대규모 지식 그래프 임베딩을 효율적으로 생성하기 위한 특화된 도구이다.62

5.3. 과학적 발견의 가속화: 생물 의학 연구

제약 회사와 연구 기관들은 신약 개발 및 질병 연구를 위해 이러한 하이브리드 시스템을 적극적으로 활용하고 있다.

- **아스트라제네카(AstraZeneca)**는 유전체, 임상, 약물, 안전성 정보를 포함한 방대하고 이질적인 데이터셋을 통합하기 위해 지식 그래프를 사용한다. 그런 다음, 이 통합된 그래프를 GNN으로 분석하여 새로운 패턴을 발견하고 신약 타겟을 예측함으로써, 연구 과정에서 발생할 수 있는 확증 편향을 극복하고 연구 속도를 높인다.66
- **시더스-사이나이(Cedars-Sinai)의 AIzKB (알츠하이머병 지식 베이스)** 사례는 하이브리드 RAG 접근법을 잘 보여준다. 이들은 그래프 데이터베이스(Memgraph)를 사용하여 유전자, 약물과 같은 구조화된 생물 의학 개체와 그 관계를 저장하여 다단계 추론을 가능하게 한다. 동시에, 벡터 데이터베이스를 통해 연구자들의 자연어 질의와 그래프의 관련 부분을 의미적으로 연결하는 시맨틱 검색을 수행한다. 이 강력한 조합은 이미 기존에 FDA 승인을 받은 약물들을 알츠하이머 치료의 잠재적 후보로 식별하는 데 기여했다.7

이러한 사례들은 이 기술의 진정한 기업 가치가 기존에 분리되어 있던 데이터 자산 위에 통일되고, 질의 가능하며, 지능적인 계층을 생성하는 능력에 있음을 보여준다. 이는 조직이 비즈니스의 여러 부분 간의 연결점을 찾고, 이전에는 단절된 데이터베이스 속에 숨겨져 있던 통찰력을 발견하게 하는, 소위 '데이터 늪(data swamp)' 문제에 대한 해결책이다.

제6장: 구현, 과제 및 미래 방향

이 마지막 장에서는 이러한 시스템을 구축하는 실질적인 측면을 다루고, 지속적인 과제들을 비판적으로 평가하며, 미래 연구를 위한 가장 유망한 방향을 탐색한다.

6.1. 개발자를 위한 도구: 오픈소스 프레임워크

지식 인지 모델을 구축하려는 연구자와 개발자를 위해 필수적인 오픈소스 라이브러리들이 존재한다.

- **Deep Graph Library (DGL):** DGL은 파이토치(PyTorch), 텐서플로우(TensorFlow) 등과 함께 사용할 수 있는 프레임워크에 구애받지 않는 라이브러리로, 효율적이고 확장 가능한 GNN 훈련을 위해 설계되었다. DGL 생태계에는 대규모 지식 그래프로부터 임베딩을 학습하기 위한 특화된 패키지인 **DGL-KE**가 포함되어 있다.68
- **PyTorch Geometric (PyG):** PyG는 파이토치 기반의 인기 있는 GNN 라이브러리로, 사용자 친화적인 인터페이스와 지식 그래프 애플리케이션에서 흔히 볼 수 있는 이종 그래프(heterogeneous graphs)를 포함한 다양한 GNN 아키텍처에 대한 광범위한 지원으로 잘 알려져 있다.69

6.2. 지속적인 과제와 한계

- **확장성 및 계산 복잡성:** 특히 공동 학습 아키텍처와 같은 모델을 훈련하는 것은 계산 집약적이다. 수십억 개의 관계를 가진 실제 대규모 지식 그래프로 확장하는 것은 여전히 중요한 공학적 과제이다.19 그래프 구축 과정 자체가 주요 병목 현상이 될 수 있다.56
- **지식 유지보수 및 일관성:** 지식 그래프와 임베딩을 동기화 상태로 유지하는 것은 매우 중요한 문제이다. 지식 그래프가 새로운 사실로 업데이트될 때, 비용이 많이 드는 전체 재훈련 없이 임베딩도 그에 맞춰 업데이트되어야 한다. 이 일관성을 유지하는 것은 주요 과제 중 하나이다.16
- **모델 및 데이터의 취약성:** 시스템은 취약할 수 있다. RAG를 위한 텍스트 분할(chunking) 전략은 종종 임의적이며 문맥을 파편화시킬 수 있다.70 성능은 불완전하거나 오류를 포함할 수 있는 기본 지식 그래프의 품질에 크게 의존한다.17 또한, 전체 데이터 표현이 특정 임베딩 모델에 종속되어 업그레이드를 어렵게 만드는 '모델 종속(model lock-in)'의 위험이 있다.70
- **내재된 지식과 외부 지식의 균형:** 지식 강화 LLM의 핵심 과제 중 하나는 모델 내부에 파라미터화된 지식과 지식 그래프를 통해 제공되는 외부 지식 사이의 균형을 맞추는 것이다. 특히 두 지식 간에 충돌이 발생할 때 이를 어떻게 처리할지가 중요하다.71

6.3. 미래 전망: 차세대 지식 인지 AI

- **동적 지식 통합:** 연구는 정적인 지식 그래프 스냅샷을 넘어, 스트리밍 데이터와 동적인 환경을 처리할 수 있도록 실시간 또는 거의 실시간으로 지식을 업데이트할 수 있는 시스템으로 나아가고 있다.73
- **자동화된 온톨로지 및 KG 구축:** LLM을 활용하여 텍스트로부터 온톨로지와 지식 그래프의 생성 및 검증을 자동화하는 것은 주요 연구 분야이다. 이는 이러한 구조를 구축하는 데 드는 수작업과 비용을 크게 줄일 수 있다.75
- **향상된 추론 및 설명 가능성:** 미래 연구는 다단계 및 논리적 추론 능력을 향상시키는 데 중점을 둘 것이다. 여기에는 부정(negation)이나 분리(disjunction)와 같은 더 복잡한 질의 유형을 처리하고, 답변에 대해 더 강력하고 검증 가능한 설명을 제공하는 모델 개발이 포함된다.50
- **표준화된 벤치마크:** 이 분야는 다양한 지식 주입 및 공동 학습 전략을 공정하게 평가하고 비교하기 위한 포괄적인 벤치마크가 필요하다. 여기에는 사실적 정확성, 노이즈에 대한 강건성, 불완전한 지식을 처리하는 능력을 테스트하는 데이터셋 개발이 포함된다.31

결론

본 보고서는 AI 지식 표현의 두 축인 벡터 임베딩과 지식 그래프가 분리된 패러다임에서 출발하여, 오늘날에는 하나의 통합된 신경-기호 아키텍처로 진화해왔음을 분석했다. 이 진화는 단순히 기술적 호기심을 넘어, AI 시스템의 환각 현상을 억제하고 추론의 투명성을 확보하며 사실 기반의 신뢰도를 높여야 하는 산업계의 절실한 요구에 의해 추동되었다.

파이프라인 방식에서 공동 학습을 거쳐, 이제는 강력한 사전 훈련 언어 모델에 외부 지식을 동적으로 주입하는 방식으로 발전한 통합 방법론은 AI가 암묵적 직관과 명시적 논리를 동시에 활용할 수 있는 가능성을 열었다. ERNIE, KnowBERT, GraphRAG와 같은 아키텍처는 트랜스포머의 어텐션 메커니즘을 '범

용 어댑터'로 활용하여, 텍스트의 의미론적 공간과 지식의 구조적 공간을 효과적으로 연결하고 있다.

구글의 시맨틱 검색, 아마존의 추천 시스템, 아스트라제네카의 신약 개발 사례에서 볼 수 있듯이, 이러한 기술의 융합은 이미 다양한 산업 분야에서 데이터 사일로를 허물고 잠재된 가치를 발굴하며 혁신을 주도하고 있다.

그러나 확장성, 지식의 동기화, 모델의 취약성 등 해결해야 할 과제는 여전히 남아있다. 미래 연구는 동적 지식 통합, LLM을 활용한 자동화된 지식 구축, 그리고 더 복잡한 추론 능력 개발에 집중될 것이다. 연구자들은 이러한 기술적 난제를 해결하고 표준화된 평가 기준을 마련하는 데 힘써야 하며, 실무자들은 DGL, PyG와 같은 오픈소스 도구를 활용하여 명확하고 가치 있는 사용 사례부터 시작하여 이 강력한 기술을 점진적으로 도입해야 할 것이다.

결론적으로, 벡터 임베딩의 의미론적 유연성과 지식 그래프의 구조적 정확성을 결합하는 것은 단지 선택이 아닌, 강력하고 다재다능할 뿐만 아니라 신뢰할 수 있고, 설명 가능하며, 사실에 기반을 둔 AI를 구축하기 위한 가장 유망한 경로이다.

Works cited

1. www.ibm.com, accessed September 25, 2025, https://www.ibm.com/think/topics/vector-embedding#:~:text=Any%20data%20that%20an%20AI,expresses%20that%20data's%20original%20meaning.
2. What is Vector Embedding? | IBM, accessed September 25, 2025, https://www.ibm.com/think/topics/vector-embedding
3. What Are Vector Embeddings? - Chatbase, accessed September 25, 2025, https://www.chatbase.co/blog/vector-embedding
4. What Are Vector Embeddings? A Guide, accessed September 25, 2025, https://www.yugabyte.com/key-concepts/what-is-vector-embedding/
5. What Are Vector Embeddings? An Intuitive Explanation - DataCamp, accessed September 25, 2025, https://www.datacamp.com/blog/vector-embedding
6. What are Vector Embeddings? | A Comprehensive Vector Embeddings Guide - Elastic, accessed September 25, 2025, https://www.elastic.co/what-is/vector-embedding
7. HybridRAG and Why Combine Vector Embeddings with Knowledge Graphs for RAG?, accessed September 25, 2025, https://memgraph.com/blog/why-hybridrag
8. Thinking, Fast and Slow: Combining Vector Spaces and Knowledge Graphs - UMBC ebiquity, accessed September 25, 2025, https://ebiquity.umbc.edu/_file_directory_/papers/854.pdf
9. What is a knowledge graph ontology? - Milvus, accessed September 25, 2025, https://milvus.io/ai-quick-reference/what-is-a-knowledge-graph-ontology
10. The significance of ontology in knowledge graphs | ONTOFORCE, accessed

September 25, 2025, https://www.ontoforce.com/knowledge-graph/ontology
11. Ontology in Graph Models and Knowledge Graphs, accessed September 25, 2025, https://graph.build/resources/ontology
12. Knowledge Graphs and Ontologies: A Primer - App Orchid, accessed September 25, 2025, https://www.apporchid.com/blog/knowledge-graphs-ontologies
13. How Ontology and Knowledge Graphs Drive AI Reasoning | by Chinmaya Panda | Medium, accessed September 25, 2025, https://medium.com/@clearmindrocks/how-ontology-and-knowledge-graphs-drive-ai-reasoning-04c78eb78273
14. Vectors and Graphs: Better Together | Pinecone, accessed September 25, 2025, https://www.pinecone.io/learn/vectors-and-graphs-better-together/
15. Survey on Embedding Models for Knowledge Graph and its Applications - arXiv, accessed September 25, 2025, https://arxiv.org/html/2404.09167v1
16. What is the relationship between embeddings and knowledge graphs? - Milvus, accessed September 25, 2025, https://milvus.io/ai-quick-reference/what-is-the-relationship-between-embeddings-and-knowledge-graphs
17. Knowledge graph embedding - Wikipedia, accessed September 25, 2025, https://en.wikipedia.org/wiki/Knowledge_graph_embedding
18. Understanding Knowledge Graph Embeddings: Methods, Applications, and Differences from LLM... - Medium, accessed September 25, 2025, https://medium.com/@doubletaken/understanding-knowledge-graph-embeddings-methods-applications-and-differences-from-llm-db6078323f39
19. A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks, accessed September 25, 2025, https://www.mdpi.com/2079-9292/9/5/750
20. End-to-End Construction of NLP Knowledge Graph - ACL Anthology, accessed September 25, 2025, https://aclanthology.org/2021.findings-acl.165.pdf
21. [PDF] End-to-End Construction of NLP Knowledge Graph | Semantic Scholar, accessed September 25, 2025, https://www.semanticscholar.org/paper/End-to-End-Construction-of-NLP-Knowledge-Graph-Mondal-Hou/87e367d76e5c63c834bf77b4f6ea8bce6cdb5553
22. Joint Representation Learning of Text and Knowledge for ..., accessed September 25, 2025, https://arxiv.org/abs/1611.04125
23. arXiv:1611.08661v2 [cs.CL] 13 Dec 2016, accessed September 25, 2025, https://arxiv.org/pdf/1611.08661
24. Joint Representations of Text and Knowledge Graphs for Retrieval and Evaluation - arXiv, accessed September 25, 2025, https://arxiv.org/abs/2302.14785
25. JKRL: Joint Knowledge Representation Learning of Text Description and Knowledge Graph - MDPI, accessed September 25, 2025, https://www.mdpi.com/2073-

8994/15/5/1056

26. Representation Learning of Knowledge Graphs with Entity Descriptions - The Association for the Advancement of Artificial Intelligence, accessed September 25, 2025, https://cdn.aaai.org/ojs/10329/10329-13-13857-1-2-20201228.pdf
27. End-to-End Learning on Multimodal Knowledge Graphs - Semantic Web Journal, accessed September 25, 2025, https://www.semantic-web-journal.net/system/files/swj2727.pdf
28. End-to-End Learning on Multimodal Knowledge Graphs | www.semantic-web-journal.net, accessed September 25, 2025, https://www.semantic-web-journal.net/content/end-end-learning-multimodal-knowledge-graphs
29. A Survey of Knowledge Enhanced Pre-trained Language Models - arXiv, accessed September 25, 2025, https://arxiv.org/pdf/2110.00269
30. What are knowledge-enhanced embeddings and when should I use them? - Milvus, accessed September 25, 2025, https://milvus.io/ai-quick-reference/what-are-knowledgeenhanced-embeddings-and-when-should-i-use-them
31. A Survey on Knowledge-Enhanced Pre-trained Language Models - ResearchGate, accessed September 25, 2025, https://www.researchgate.net/publication/366657223_A_Survey_on_Knowledge-Enhanced_Pre-trained_Language_Models
32. A Survey of Knowledge Enhanced Pre-trained Language Models - ResearchGate, accessed September 25, 2025, https://www.researchgate.net/publication/378660168_A_Survey_of_Knowledge_Enhanced_Pre-trained_Language_Models
33. PRETRAIN KNOWLEDGE-AWARE LANGUAGE MODELS - OpenReview, accessed September 25, 2025, https://openreview.net/pdf?id=OAdGsaptOXy
34. Combining Knowledge Graphs and Large Language Models - arXiv, accessed September 25, 2025, https://arxiv.org/html/2407.06564v1
35. Baidu's Knowledge-Enhanced ERNIE 3.0 Pretraining Framework Delivers SOTA NLP Results, Surpasses Human Performance on the SuperGLUE Benchmark - Synced Review, accessed September 25, 2025, https://syncedreview.com/2021/07/16/deepmind-podracer-tpu-based-rl-frameworks-deliver-exceptional-performance-at-low-cost-63/
36. Knowledge Enhanced Contextual Word Representations - NSF-PAR, accessed September 25, 2025, https://par.nsf.gov/servlets/purl/10180482
37. Injecting Knowledge Graphs into Large Language Models - arXiv, accessed September 25, 2025, https://arxiv.org/html/2505.07554v1
38. ConceptFormer: Towards Efficient Use of Knowledge-Graph Embeddings in Large Language Models - arXiv, accessed September 25, 2025, https://arxiv.org/html/2504.07624v1

39. KGLM: Integrating Knowledge Graph Structure in Language Models for Link Prediction - Tagkopoulos Lab, accessed September 25, 2025, http://tagkopouloslab.ucdavis.edu/wp-content/uploads/c19.pdf
40. Large Language Models Meet Graph Neural Networks: A Perspective of Graph Mining - MDPI, accessed September 25, 2025, https://www.mdpi.com/2227-7390/13/7/1147
41. Creating Embeddings from Knowledge Graphs Using Graph Neural Networks - Medium, accessed September 25, 2025, https://medium.com/@busra.oguzoglu/creating-embeddings-from-knowledge-graphs-using-graph-neural-networks-ffc6cc62275c
42. Building Amazon Recommendation Systems with Graph Neural Networks | by Aiden Chang, accessed September 25, 2025, https://medium.com/@aidenchang/building-amazon-recommendations-with-graph-neural-networks-accf51847fb1
43. lucyinstitute.nd.edu, accessed September 25, 2025, https://lucyinstitute.nd.edu/wp-content/uploads/2025/02/29875-Article-Text-33929-1-2-20240324.pdf
44. [2401.07105] Graph Language Models - arXiv, accessed September 25, 2025, https://arxiv.org/abs/2401.07105
45. ERNIE Bot: Baidu's Knowledge-Enhanced Large Language Model Built on Full AI Stack Technology, accessed September 25, 2025, https://research.baidu.com/Blog/index-view?id=183
46. Baidu's ERNIE 4.5 is Built On a 'Heterogeneous MoE' Architecture, accessed September 25, 2025, https://analyticsindiamag.com/ai-news-updates/baidus-ernie-4-5-is-built-on-a-heterogeneous-moe-architecture/
47. Announcing the Open Source Release of the ERNIE 4.5 Model Family, accessed September 25, 2025, https://yiyan.baidu.com/blog/posts/ernie4.5/
48. Knowledge Enhanced Contextual Word ... - ACL Anthology, accessed September 25, 2025, https://aclanthology.org/D19-1005.pdf
49. arXiv:1909.04164v2 [cs.CL] 31 Oct 2019, accessed September 25, 2025, https://arxiv.org/pdf/1909.04164
50. KnowFormer: Revisiting Transformers for Knowledge Graph Reasoning | Request PDF, accessed September 25, 2025, https://www.researchgate.net/publication/384155086_KnowFormer_Revisiting_Transformers_for_Knowledge_Graph_Reasoning
51. [ICML 2024] KnowFormer: Revisiting Transformers for Knowledge Graph Reasoning - GitHub, accessed September 25, 2025, https://github.com/jnanliu/KnowFormer
52. KnowFormer: Revisiting Transformers for Knowledge Graph ... - arXiv, accessed September 25, 2025, https://arxiv.org/pdf/2409.12865
53. KnowFormer: Revisiting Transformers for Knowledge Graph Reasoning - arXiv,

accessed September 25, 2025, https://arxiv.org/html/2409.12865v1

54. www.alphaxiv.org, accessed September 25, 2025, https://www.alphaxiv.org/overview/2504.07624v1#:~:text=and%20generating%20responses.-,ConceptFormer%20Approach,into%20the%20LLM's%20input%20sequence.
55. ConceptFormer: Towards Efficient Use of Knowledge-Graph Embeddings in Large Language Models | alphaXiv, accessed September 25, 2025, https://www.alphaxiv.org/overview/2504.07624v1
56. Efficient Knowledge Graph Construction and Retrieval from Unstructured Text for Large-Scale RAG Systems - arXiv, accessed September 25, 2025, https://arxiv.org/html/2507.03226v2
57. Welcome - GraphRAG, accessed September 25, 2025, https://microsoft.github.io/graphrag/
58. What is semantic search, and how does it work? | Google Cloud, accessed September 25, 2025, https://cloud.google.com/discover/what-is-semantic-search
59. What is Google Knowledge Graph? A Dive into Semantic Search - Loganix, accessed September 25, 2025, https://loganix.com/what-is-google-knowledge-graph/
60. Knowledge Graph Search API - Google for Developers, accessed September 25, 2025, https://developers.google.com/knowledge-graph
61. Google Knowledge Graph Creates a Pathway Towards Semantic Search, accessed September 25, 2025, https://aha.elliance.com/2012/06/01/google-knowledge-graph-pathway-towards-semantic-search/
62. Amazon's open-source tools make embedding knowledge graphs much more efficient, accessed September 25, 2025, https://www.amazon.science/blog/amazons-open-source-tools-make-embedding-knowledge-graphs-much-more-efficient
63. Building commonsense knowledge graphs to aid product recommendation - Amazon Science, accessed September 25, 2025, https://www.amazon.science/blog/building-commonsense-knowledge-graphs-to-aid-product-recommendation
64. RecKG: Knowledge Graph for Recommender Systems - arXiv, accessed September 25, 2025, https://arxiv.org/html/2501.03598v1
65. Power recommendations and search using an IMDb knowledge graph – Part 3 - AWS, accessed September 25, 2025, https://aws.amazon.com/blogs/machine-learning/power-recommendations-and-search-using-an-imdb-knowledge-graph-part-3/
66. Ultimate Tools for Knowledge Graphs and Vector Embedding - MyScale, accessed September 25, 2025, https://myscale.com/blog/essential-knowledge-graphs-

vector-embedding/

67. Data Science & Artificial Intelligence: Unlocking new science insights - AstraZeneca, accessed September 25, 2025, https://www.astrazeneca.com/r-d/data-science-and-ai.html
68. Deep Graph Library, accessed September 25, 2025, https://www.dgl.ai/
69. PyG: Home, accessed September 25, 2025, https://pyg.org/
70. Why we ditched embeddings for knowledge graphs (and why chunking is fundamentally broken) : r/LLMDevs - Reddit, accessed September 25, 2025, https://www.reddit.com/r/LLMDevs/comments/1n3iwrr/why_we_ditched_embeddings_for_knowledge_graphs/
71. KnowPO: Knowledge-Aware Preference Optimization for Controllable Knowledge Selection in Retrieval-Augmented Language Models, accessed September 25, 2025, https://ojs.aaai.org/index.php/AAAI/article/view/34783/36938
72. Towards Verifiable Generation: A Benchmark for Knowledge-aware Language Model Attribution - ACL Anthology, accessed September 25, 2025, https://aclanthology.org/2024.findings-acl.28/
73. [2410.14733] Knowledge Graph Embeddings: A Comprehensive Survey on Capturing Relation Properties - arXiv, accessed September 25, 2025, https://arxiv.org/abs/2410.14733
74. arxiv.org, accessed September 25, 2025, https://arxiv.org/html/2505.20099v2
75. Research Trends for the Interplay between Large Language Models and Knowledge Graphs - VLDB Endowment, accessed September 25, 2025, https://vldb.org/workshops/2024/proceedings/LLM+KG/LLM+KG-9.pdf
76. KDD 2025 Workshop SKnow-LLM - OpenReview, accessed September 25, 2025, https://openreview.net/group?id=KDD.org/2025/Workshop/SKnow-LLM&referrer=%5BHomepage%5D(%2F)
77. A Survey on Benchmarks of Multimodal Large Language Models - GitHub, accessed September 25, 2025, https://github.com/swordlidev/Evaluation-Multimodal-LLMs-Survey
78. Benchmark and Best Practices for Biomedical Knowledge Graph Embeddings - PMC, accessed September 25, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7971091/