Integrating Knowledge Graphs with Multi-Modal Machine Learning

1) Research topic and context

Building on the completed **Integrating Knowledge Graphs with Multi-Modal Machine Learning — Literature Review**, this project investigates how to fuse knowledge graphs (KGs) with multi-modal models (images, text, audio, etc.) so that systems can **retrieve, classify and explain** in open-world settings. The review identified recurring frictions: (i) loss of relational meaning when KGs are embedded as vectors, (ii) operational brittleness of single-store back ends under embedding-heavy workloads, and (iii) accuracy-only evaluation that obscures semantic fidelity, latency, robustness and interpretability. This specification translates those insights into a concrete research plan.

2) Research problem

Despite promising demonstrations, many KG+MMML pipelines either flatten semantics (directionality, hierarchy, cardinality), struggle to meet tail-latency and freshness targets at scale, or make claims unsupported by transparent metrics and faithful explanations.
Problem statement: How can we integrate KGs with multi-modal models in a way that **preserves relational semantics**, **meets performance budgets** at realistic scale, and **explains outputs** with path-level evidence?

3) Aim and measurable objectives

Aim: To develop and validate a **scalable, reproducible** KG–MMML integration that maintains semantic structure and provides path-level rationales, using a **hybrid** (graph as semantic spine + vector index) architecture.

Objectives:

1. **Knowledge fidelity.** Achieve **SRS ≥ 0.75** by Week 8 with **AtP ≥ 0.95**, **HP ≥ 0.25**, **AP ≥ 0.99**, rising from the text-only baseline.
2. **Task performance.** Demonstrate **≥ +3 percentage-point micro-F1** (or equivalent retrieval metric) when adding **KG-as-features** to a TF-IDF (or single frozen encoder) baseline.
3. **Operational SLOs.** Report **p50/p95/p99 latency** at **N ∈ {10^3, 10^4}** with a reproducible harness, and discuss trade-offs.
4. **Robustness.** Under small stressors (e.g., taxonomy off / unit noise), limit relative metric drop to **≤ 10%**.
5. **Reproducibility.** Release scripts, pinned configs, seeds, and an **evaluation sheet**; document provenance and date-stamped KG snapshots.

4) Research questions and hypotheses

1. **Knowledge fidelity:** Do minimal, interpretable joint constraints (hierarchy, asymmetry, attribute reconstruction) raise SRS versus KG-as-features without violating accuracy/latency budgets?
   H1: SRS increases to **≥ 0.75** without harming accuracy or latency materially.
2. **Architecture choice:** Does a **hybrid** graph-vector stack beat a monolithic graph store on tail latency at scale while preserving explanations?
   H2: **p95 −≥ 30%** at **≥ 10^6** entities; **p99 ≤ 2× p95**; rationale availability unchanged.
3. **Operations:** Does a **hybrid graph–vector** stack meet **p95/p99** targets at **10^3–10^4** scale better than monolithic approaches?
   H3: **+≥ 3** accuracy points and **+≥ 10 pp** rationale precision.
4. **Honest evaluation:** Do robustness/scale tests change which configuration appears "best" compared with accuracy-only reporting?
   H4: At least one accuracy-leading baseline is re-ranked when SRS/latency/robustness are included.

5) Methodological approach (derived from the literature review)

- **Representation: Multi-resolution semantics** (types → direct relations → k-hop context → paths) to prevent flattening; escalate at query time only as needed.

- **Models:** (i) baseline multi-modal model; (ii) **KG-as-features**; (iii) **joint** KG–MMML model with minimal constraints.

- **Architecture: Graph database** as semantic spine (entities, typed relations, provenance, path-level explanations) + **ANN vector index** for dense similarity; **cache** for hot subgraphs/rationales; **stream** for updates and re-index.

- **Evaluation policy:** Effectiveness (retrieval/zero-shot), **SRS**, **p50/p95/p99 latency** and memory, robustness (dropout/shift), **interpretability** (rationale precision + counterfactual sensitivity), reproducibility (seeds/configs/compute disclosure).

- **Ethics & governance:** Data minimisation; parity checks; time-stamped edges; change control; documented fall-backs.

6) Significance and potential contributions

- A portable, **measurable definition of knowledge fidelity** (SRS) that complements accuracy and can travel across domains.

- Evidence-backed **design rules** for when to use joint learning, multi-resolution semantics and a hybrid graph–vector stack.

- A transparent **evaluation sheet** and **reference implementation** (with path-level explanations and operational playbooks) that practitioners can adopt.

- Practical **decision thresholds** (e.g., "deploy only if SRS ≥ threshold"; "switch to hybrid when p95 exceeds budget") that raise trust and reproducibility.

7) Scope and boundaries

In scope: Open-world retrieval and zero-shot classification; image–text primary, extensible to other modalities; KG integration via embeddings and joint constraints; hybrid retrieval.
Out of scope: Heavy-weight symbolic theorem-proving; human-subject studies.
Assumptions: Access to at least one permissively licensed multi-modal dataset and a KG schema with task-critical relations; compute adequate for baselines and ablations.
Success criteria: Meeting Objectives 1-5 plus a short demonstration video.