Heterogeneous Multimodal Perception with Graph Attention for Few-Shot Adaptive Decision Support

Lucía García López

Universidad Autónoma de Madrid,Madrid,Spain

Abstract: Decision-support systems increasingly operate on heterogeneous observations whose sampling rates, semantic granularity, reliability, and availability differ across modalities. Conventional fusion architectures usually assume stable modality relationships and abundant task-specific labels, whereas few-shot classifiers often treat each input as an unstructured feature vector. This article proposes Heterogeneous Multimodal Perception with Graph Attention (HMP-GAT), a methodological framework that represents observations, modality states, class prototypes, and candidate decisions as typed nodes in an episode-level graph. Relation-conditioned dynamic attention propagates evidence among these nodes, while support-derived reliability priors and availability masks constrain information flow. Prototype-based adaptation enables classification for new tasks with limited supervision, and a calibrated selective head converts perceptual estimates into actions or deferral under an explicit utility policy. A reproducible synthetic proof-of-concept isolates the proposed fusion logic using lightweight prototype surrogates rather than the full deep architecture. In five-shot ablation episodes, quality gating and quality-plus-relation gating achieved 84.44% and 84.38% accuracy, respectively, compared with 82.92% for uniform-distance fusion and 83.03% for late-probability fusion. The relation term reduced the mean weight assigned to a conflicting modality from 0.340 to 0.321 but did not improve accuracy over quality-only gating. These results illustrate shot-efficient fusion and reliability sensitivity, not real-world superiority. The principal contribution is therefore a coherent, auditable modeling framework and evaluation protocol for studying when graph-structured multimodal evidence improves adaptive decisions under scarcity, conflict, and missingness.

Keywords: heterogeneous multimodal learning; graph attention networks; few-shot adaptation; prototypical learning; selective prediction; decision support

1 Introduction

Multimodal decision support is not merely a matter of aggregating several data sources. Visual frames, textual reports, structured records, and temporal sensor streams describe different aspects of an underlying situation, often at incompatible resolutions and with unequal evidentiary value. A camera may be occluded, a report may be stale, and a sensor may drift, while the remaining channels continue to provide useful evidence. Multimodal learning therefore involves representation, alignment, fusion, translation, and co-learning problems rather than a single concatenation operation. This difficulty becomes acute when a deployed system encounters a new environment, user, device configuration, or decision category for which only a handful of labeled examples is available[1-3]. High-capacity perception models can be pretrained on broad data, but the final decision policy still has to adapt without confusing sparse evidence with certainty.

This challenge is reflected in how conventional fusion strategies fall short. Early fusion combines features before task inference, whereas late fusion combines modality-specific predictions. Both can be effective when modalities are complete and their train–test relationship is stable, but neither explicitly represents the structural dependencies among an observation, its source modality, a class hypothesis, and a decision consequence. Temporal architectures for mobile sensing demonstrate the value of modeling local interactions, cross-sensor interactions, and temporal dependencies under noise. Graph-based fusion under incomplete multimodal data further suggests that missingness and relationships should be modeled jointly rather than handled by imputation alone. Nevertheless, a general decision-support architecture must also address rapid task adaptation and the option to withhold an unsafe recommendation.

Related studies illustrate these requirements from complementary angles. Ye et al.[4] investigated a convolutional modulated scattering feature network for aircraft classification in SAR images, exemplifying modality-specific encoding for heterogeneous remote-sensing evidence; Lan et al.[5] studied a multi-level cross-modal edge filtering method for high-resolution optical-SAR image registration, illustrating cross-modal alignment under differing semantic granularity and reliability; and Li et al.[6] studied EOCNet for high-aspect-ratio object detection in SAR imagery, demonstrating relation-aware representation learning for scarce and structurally irregular observations. Collectively, these studies motivate the need for quality-aware modality encoding and relation-conditioned evidence aggregation in heterogeneous perception settings[7-10].

Graph attention offers a natural mechanism for meeting this need, because it assigns data-dependent weights to neighboring nodes instead of applying a fixed aggregation coefficient. Heterogeneous graph attention extends this principle to multiple node and relation types[11-13], allowing node-level and semantic-level importance to be distinguished. However, graph construction alone does not guarantee useful reasoning. A densely connected multimodal graph can propagate corrupted evidence, obscure provenance, and produce confidence unsupported by the underlying labels. The graph must therefore encode the semantics of evidence flow, preserve availability information, and make the decision boundary explicit[14-16].

Few-shot learning supplies a complementary perspective. Prototypical methods classify queries by their distances to class representations estimated from a small support set. Gradient-based meta-learning

instead optimizes an initialization from which a small number of adaptation steps can produce effective task-specific parameters[17-20]. These approaches offer different trade-offs between computational cost, stability, and representational flexibility. In a decision-support setting, prototype adaptation is particularly attractive because support examples remain inspectable and class evidence can be traced. Yet a prototype estimated independently in each modality may be noisy or mutually inconsistent, especially when one modality contains conflicting labels or missing observations[21-24].

A further requirement is that perceptual confidence be meaningful at the point of action. Modern neural classifiers can be miscalibrated, while comparatively simple post-processing can improve probability estimates[25-27]. Decision support also requires asymmetric costs: an unnecessary review may be inconvenient, but an automated false recommendation may carry a much higher penalty. Accuracy alone cannot establish that a model is safe or useful. A complete framework must connect few-shot predictions to calibrated uncertainty, selective deferral, provenance, and governed human oversight[28-30].

This article addresses these requirements through Heterogeneous Multimodal Perception with Graph Attention (HMP-GAT). The framework treats each few-shot episode as a typed evidence graph. Observation nodes preserve instance-level embeddings, modality nodes summarize source-specific states, class-prototype nodes accumulate support evidence, and decision nodes represent candidate actions. Relation-aware attention determines which neighbors should influence each representation. Support-derived reliability estimates enter as priors rather than as unquestioned ground truth, and unavailable edges are masked. The resulting representation is used by a metric-based adaptation head and a calibrated decision layer[31-33].

The contributions are fourfold. First, a unified graph formulation links heterogeneous perception, few-shot adaptation, and decision utility within one auditable architecture. Second, a dynamic relation-aware attention rule combines query-conditioned compatibility, typed relation embeddings, reliability priors, and explicit missingness masks[34-37]. Third, a complete evaluation protocol covers accuracy, calibration, risk–coverage behavior, robustness, subgroup performance, computational cost, and provenance quality. Fourth, a reproducible synthetic surrogate study illustrates which components can be isolated experimentally and which claims require validation of the full deep model. The synthetic results are deliberately interpreted conservatively: they do not demonstrate benchmark superiority, causal reliability, or deployment readiness[38-40].

The remainder of the article is organized as follows. Section 2 situates the framework within multimodal fusion, heterogeneous graph learning, meta-learning, and selective decision support. Section 3 formalizes the problem and design requirements. Section 4 specifies HMP-GAT, including its objective and computational complexity. Sections 5 and 6 describe the evaluation design and synthetic findings. Sections 7 and 8 discuss operational implications, limitations, ethics, and reproducibility, followed by the conclusion in Section 9.

2 Related Work

2.1 Multimodal representation and fusion

Multimodal architectures differ in where cross-source interaction occurs. Early fusion exposes a joint model to concatenated low-level or intermediate features and can learn detailed cross-modal correspondence, but it is sensitive to dimensional mismatch and missing channels. Late fusion preserves modality-specific models and combines their outputs, improving modularity while potentially discarding fine-grained interactions. Intermediate and hybrid fusion seek a compromise by exchanging information at several depths. Cross-attention architectures for unaligned language, audio, and visual sequences illustrate how pairwise interactions can be learned without requiring perfectly synchronized inputs. Large-scale vision–language pretraining further shows that shared representation spaces can support broad transfer, although such transfer does not eliminate the need for task-specific reliability assessment[41-43].

Small-data multimodal learning introduces a different constraint: the fusion rule must not consume the limited task labels that are needed for adaptation. Contrastive fusion with small data has been explored for multimodal activity recognition, emphasizing the value of cross-modal agreement during representation learning[44-46]. HMP-GAT differs by treating agreement as one relation within a typed graph rather than as the sole training signal. This distinction matters when two modalities are consistently wrong, when redundancy hides a common failure, or when a minority modality provides decisive evidence. Reliability is therefore modeled as a context-dependent property of a node and episode, not as a fixed ranking of sensors.

2.2 Graph attention for heterogeneous evidence

Graph neural networks propagate information through explicit adjacency structures. Graph attention networks introduce masked self-attention over neighborhoods, enabling different neighbors to receive different weights without spectral matrix operations. Standard graph attention has a limited static form in which attention ranking can be insufficiently conditioned on the query node; dynamic graph attention modifies the operation order to increase expressiveness. HMP-GAT adopts the query-conditioned principle because the same modality should not necessarily have the same importance for every class prototype or decision node.

Heterogeneous graph attention introduces node-level and semantic-level attention over metapath-based neighborhoods. This hierarchy is relevant when relations such as “observed by,” “supports class,” and “constrains action” carry different meanings. HGMF addresses incomplete multimodal data through heterogeneous graph fusion, while cross-modality graph neural networks have been used for few-shot learning. Graph neural networks for few-shot classification more generally demonstrate the value of propagating information among labeled and unlabeled examples. HINFShot provides a benchmark perspective for few-shot node classification in heterogeneous information networks. HMP-GAT synthesizes these directions but relocates the graph from a dataset-level object to an episode-level decision artifact that retains provenance and availability.

2.3 Few-shot adaptation and metric learning

Few-shot methods can be organized around optimization, metric learning, and model- or memory-based adaptation. Matching networks learn an attention-based comparison between support and query examples. Prototypical networks replace per-example comparison with class means in a learned metric space, yielding a simple inductive bias for unseen classes. Optimization-based meta-learning trains parameters that can be fine-tuned effectively with a few gradient steps. Broader surveys emphasize that task construction, data leakage, and evaluation variance are central to credible few-shot conclusions.

HMP-GAT uses prototypes as the default task-adaptation mechanism because they are stable, inexpensive, and interpretable, but the architecture does not depend on arithmetic class means alone. A prototype node can aggregate relation-weighted support messages, include a learned covariance or scale, and be regularized toward a base-task prior. Gradient adaptation of relation gates can be added when sufficient support data exist. The framework thus separates the choice of graph reasoning from the choice of adaptation rule rather than treating prototype learning and graph attention as mutually exclusive.

2.4 Calibrated and selective decision support

Decision support is evaluated by the consequences of action, not only by classification correctness. Deep ensembles provide a strong general strategy for estimating predictive uncertainty, while selective prediction integrates a reject option into the learned decision rule. Calibration methods such as temperature scaling adjust probability sharpness without changing the class ranking. These techniques address different failure modes: a calibrated model may still be inaccurate, an ensemble may be overconfident under distribution shift, and a selective model may defer too often to be useful.

High-stakes decision contexts also require interpretability and governance. Interpretable models are preferable to post hoc rationalization when the decision structure itself must be inspected. Clinical decision-support research emphasizes workflow integration, alert burden, maintenance, and human factors alongside algorithmic performance. Responsible deployment guidance similarly stresses prospective evaluation, monitoring, and explicit attention to harm. HMP-GAT therefore treats deferral, provenance, and authorized updating as architectural requirements rather than interface features added after model training.

Table 1 Positioning of HMP-GAT among related paradigms

Paradigm	Evidence structure	Few-shot adaptation	Missingness/reliability	Decision boundary
Early/late fusion	Implicit or output-level	Not intrinsic	Imputation or averaging	Usually classification-centric
Graph multimodal fusion	Explicit graph relations	Architecture-dependent	Graph-dependent	Usually classification-centric
Metric/meta-learning	Support–query comparison	Intrinsic	Architecture-dependent	Rarely explicit
HMP-GAT	Typed episode graph	Prototype or gated update	Masks plus reliability priors	Utility, calibration, deferral

3 Problem Formulation and Design Principles

3.1 Episodic heterogeneous decision problem

Let an episode be defined by a support set, a query set, and an action space. Each instance contains observations from a subset of modalities, timestamps or alignment metadata, an availability mask, and optional quality measurements. The support set contains a small labeled sample for each new class, whereas query labels are hidden during inference. The system must estimate a class or state distribution and select an action under a task-specific utility function. Missingness is permitted to differ between support and query instances, and the reliability of a modality may change across episodes. This formulation includes conventional few-shot classification as a special case while also accommodating selective recommendation and human review.

The central object is not a fixed global graph but an episode graph instantiated from available evidence. Its node set contains observations, modality states, class prototypes, and decision candidates; its relation set contains alignment, source, support, semantic, temporal, and action-constraint edges. The graph can be sparse and task-specific. For example, a text report may be linked to a time window of sensor observations through an alignment relation, while a class prototype receives messages only from support examples carrying that class label. A decision node receives class evidence together with action costs and policy constraints.

3.2 Design requirements

Five requirements guide the framework. R1, heterogeneity: node and relation semantics must remain explicit after feature projection. R2, adaptivity: evidence aggregation must respond to the current query, support set, and modality state rather than relying solely on global fusion weights. R3, degradation awareness: unavailable or low-quality evidence must alter information flow without forcing zero-filled observations to appear informative. R4, decision accountability: predictions must be connected to calibrated uncertainty, action utility, provenance, and a deferral mechanism. R5, reproducibility: graph construction, episode sampling, and evaluation splits must be specified sufficiently for independent replication.

These requirements rule out several tempting simplifications. A scalar quality score cannot be treated as a calibrated probability of correctness. Attention weights cannot automatically be interpreted as causal explanations. Missingness cannot be assumed random without analysis. Few-shot accuracy cannot be reported without confidence intervals and episode-level sampling details. Finally, human review cannot be simulated by assigning a generic cost to all errors; review capacity and asymmetric harms must be included in the decision policy.

Table 2 Principal notation

Symbol	Meaning
$T = \{S, Q\}$	Episode with labeled support and unlabeled query sets
z_i^m	Encoded observation i from modality m

Symbol	Meaning
G_T	Typed graph instantiated for episode T
p_c^m	Support prototype for class c in modality m
q_m	Bounded support-derived reliability score
b_j	Availability indicator for node or edge j
α_ij^r	Attention from neighbor j to target i under relation r
u(a, c)	Utility of action a when the true class is c
Π(x_q)	Indicator that minimum provenance requirements are satisfied

4 HMP-GAT Methodology

4.1 Framework overview

HMP-GAT consists of modality-specific encoders, a typed graph constructor, relation-aware dynamic attention layers, a prototype adaptation module, and a calibrated selective decision head. Figure 1 summarizes the data flow. The encoders map raw observations into comparable latent dimensions without discarding source identity. The graph constructor instantiates only valid relations for the current episode. Attention layers then update node states through structured message passing. Prototypes and query representations are adapted before the decision head computes class or action probabilities, uncertainty estimates, and a recommendation or deferral.

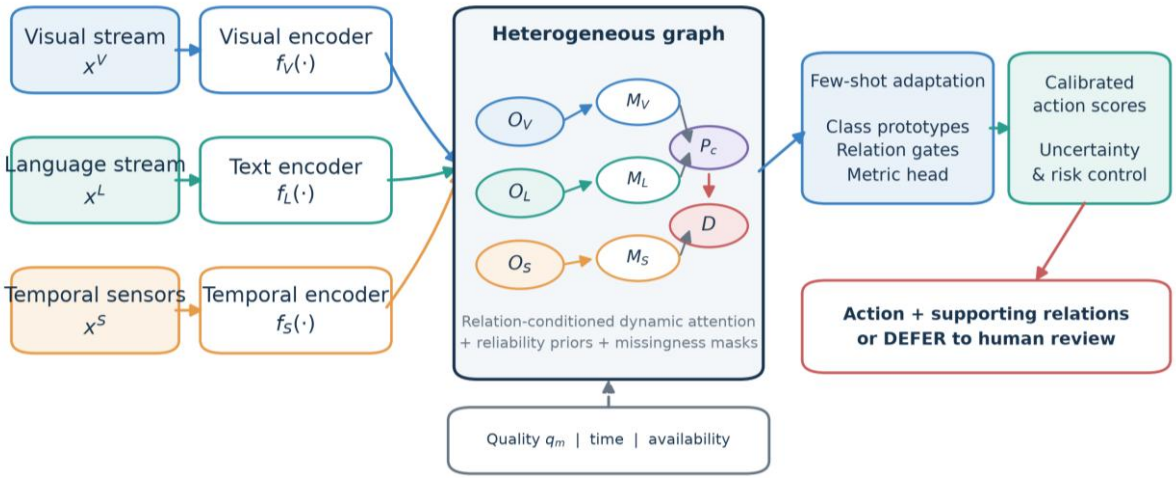


Figure 1 HMP-GAT architecture. Modality-specific observations are encoded and instantiated as a typed episode graph; relation-aware reasoning supports few-shot adaptation and calibrated selective decisions. Original schematic.

This separation has practical benefits. Encoders can be replaced according to the data domain, graph relations can be audited independently of neural weights, and the decision policy can be modified without retraining perception. The architecture also supports staged evaluation: encoder quality can be measured with frozen graphs, fusion rules can be compared with shared embeddings, and decision thresholds can be assessed on held-out episodes. Such staging reduces the risk of attributing gains to graph reasoning when they arise from a stronger pretrained encoder.

4.2 Modality-specific encoding and alignment

For modality m, an encoder maps an observation into a latent vector while preserving timestamps, source identifiers, and availability. Visual inputs may use convolutional or transformer encoders, language inputs may use contextual token aggregation, and temporal sensors may use convolutional-recurrent or temporal transformer modules. A projection head maps each representation to a shared dimension, but the projection is not assumed to erase modality-specific geometry. Layer normalization and modality-specific scale parameters prevent a high-dimensional channel from dominating solely because of numerical scale.

$$z_i^m = g_m\big(f_m(x_i^m)\big), z_i^m \in \mathbb{R}^d. \quad (1)$$

Alignment is represented explicitly rather than being implicit in concatenation. Hard alignment edges connect observations with verified correspondence, such as a report linked to a sensor window. Soft alignment edges use learned compatibility under constraints on time, entity, or location. When alignment is uncertain, the relation carries an attribute representing that uncertainty. This approach avoids treating every pair of modalities as equally related and permits counterfactual tests in which alignment edges are removed or perturbed.

4.3 Typed episode-graph construction

The graph contains four principal node types. Observation nodes correspond to encoded instances, modality nodes summarize source-specific evidence and quality, class-prototype nodes represent support-derived hypotheses, and decision nodes represent candidate actions. Additional context nodes can encode users, devices, locations, or operating regimes when such attributes are available and ethically permissible. Each node carries a type indicator and provenance reference, enabling the final rationale to identify which observations and relations contributed to an output.

$$\mathcal{G}_\mathcal{T} = (\mathcal{V}_\mathcal{T}, \mathcal{E}_\mathcal{T}, \mathcal{R}, \boldsymbol{\rho}, \boldsymbol{b}), \mathcal{V}_\mathcal{T} = \mathcal{V}^O \cup \mathcal{V}^M \cup \mathcal{V}^P \cup \mathcal{V}^D. \quad (2)$$

Edges are created by explicit rules. Source edges link an observation to its modality; alignment edges link compatible observations; support edges link labeled observations to their class prototype; semantic edges connect prototypes with decision candidates; and temporal edges connect observations with valid chronological relationships. Availability is represented by a binary mask b_j, while relation attributes ρ_r store semantics such as time lag, alignment confidence, or action constraints. Graph sparsity is a design choice: denser graphs permit more interactions but increase cost and the opportunity for spurious propagation.

4.4 Support prototypes and reliability estimation

A modality-specific support prototype is initially computed from available labeled observations. If no support example is available for a modality–class pair, the corresponding prototype is marked unavailable rather than filled with zeros. The initial prototype can be refined by attention layers, but the arithmetic mean provides a stable anchor. Support-derived variance estimates are regularized toward base-task priors to avoid zero uncertainty in one-shot episodes.

$$p_c^m = \frac{\sum_{(x_i,y_i)\in\mathcal{S}} b_i^m \mathbb{1}(y_i=c) z_i^m}{\sum_{(x_i,y_i)\in\mathcal{S}} b_i^m \mathbb{1}(y_i=c)+\epsilon}. \tag{3}$$

Reliability combines support compactness, cross-modal relational consistency, timestamp validity, and sensor or document quality indicators. These inputs are normalized and mapped to a bounded score $q_m \in (0,1]$. Compactness reflects within-class dispersion, while relational consistency compares the geometry of class prototypes across modalities. Importantly, reliability is not interpreted as the probability that a modality is correct. It is a prior that can be overridden by query-conditioned attention, and its calibration must be evaluated separately.

4.5 Relation-conditioned dynamic attention

For a target node i and neighbor j under relation r, compatibility is computed from the transformed target, neighbor, and relation attribute. The operation places the nonlinearity after the combined projection so that the target representation can influence neighbor ranking dynamically. Relation-specific parameters prevent an alignment edge and a decision-constraint edge from sharing an inappropriate compatibility function.

$$e_{ij}^r = LeakyReLU\left(a_r^\top \left[W_r h_i \| W_r h_j \| U_r \rho_r\right]\right). \tag{4}$$

The attention logit incorporates reliability and availability. The term $\log b_j$ is zero for an available neighbor and negative infinity for an unavailable neighbor, producing a hard mask. The coefficient λ_r controls the influence of the reliability prior and can vary by relation. A residual prior toward uniform attention prevents premature collapse during early training, while entropy regularization discourages both indiscriminate aggregation and excessively peaked weights.

$$\ell_{ij}^r = e_{ij}^r + \lambda_r \log(q_j+\epsilon) + \log b_j, \alpha_{ij}^r = softmax_{j\in\mathcal{N}_r(i)}\left(\ell_{ij}^r\right). \tag{5}$$

Node updates aggregate relation-specific messages and retain a residual path to the previous state. Multi-head attention can represent complementary evidence patterns, although head count should be regularized in few-shot settings. The update is followed by normalization and a bounded feed-forward transformation. Figure 2 illustrates the graph and the computation.

$$h_i' = LN\left(h_i + W_o \sum_{r\in\mathcal{R}} \sum_{j\in\mathcal{N}_r(i)} \alpha_{ij}^r W_r h_j\right). \tag{6}$$

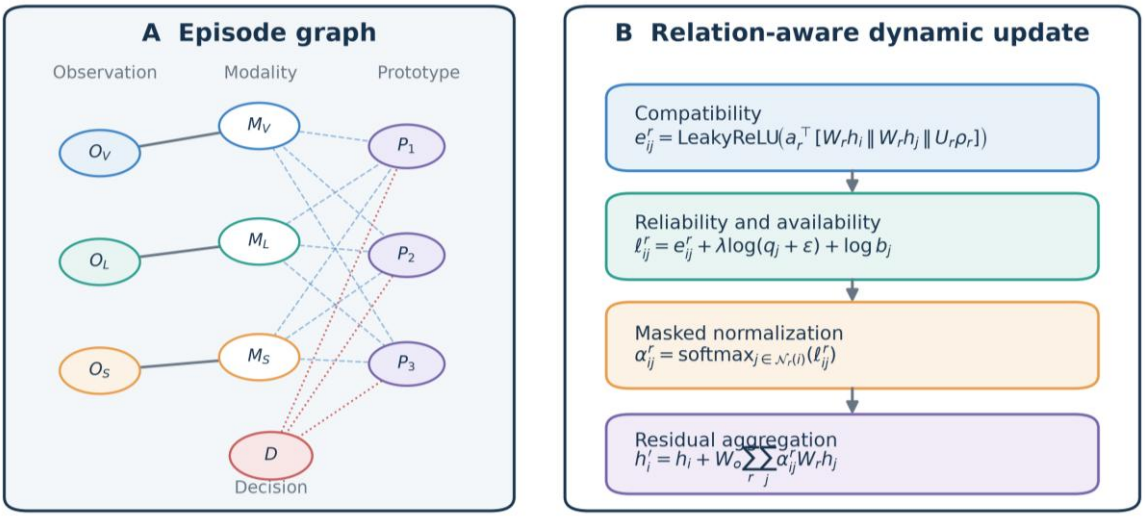


Figure 2 Typed episode graph and relation-conditioned dynamic attention. Availability masks remove invalid neighbors, while reliability scores act as adjustable priors rather than correctness guarantees. Original schematic.

Attention is inspected as a descriptive routing measure, not as proof of causal importance. A high weight can reflect feature similarity rather than evidentiary validity, and a low weight can result from a restrictive mask. Explanation outputs therefore include the underlying support observations, relation attributes, quality estimates, and alternative decisions. Perturbation tests are required to determine whether removing a highly attended edge actually changes the recommendation.

4.6 Few-shot query adaptation

After graph propagation, class prototypes combine modality-specific messages using normalized relation weights. Query nodes attend to the same graph, allowing a query to emphasize different modalities from those emphasized by a prototype. The fused prototype can be regularized toward a base-task class prior when available, and its scale can be adjusted by a learned temperature. This mechanism supports new classes without changing the output dimension of the encoders.

$$\tilde{p}_c = \sum_{m \in \mathcal{A}_c} \omega_c^m \tilde{p}_c^m, \sum_{m \in \mathcal{A}_c} \omega_c^m = 1. \tag{7}$$

The class posterior is obtained from a scaled negative distance between the fused query and each available prototype. Euclidean distance is the default, but Mahalanobis or cosine distance can be substituted when supported by validation. Missing modalities affect the graph rather than the class set: a prototype remains available if at least one valid evidence path exists. Classes with no valid support path trigger deferral rather than an arbitrary prediction.

$$p(y = c \mid x_q) = \frac{\exp\left[-\tau d\left(\tilde{z}_q, \tilde{p}_c\right)\right]}{\sum_{k \in \mathcal{C}_\mathcal{T}} \exp\left[-\tau d\left(\tilde{z}_q, \tilde{p}_k\right)\right]}. \tag{8}$$

Two adaptation modes are supported. The parameter-free mode updates only prototypes and reliability statistics, making it suitable for very small support sets. The gated mode performs one or a few regularized gradient steps on relation gates and temperature parameters while keeping encoders fixed. This optional

mode resembles the rapid-adaptation objective of optimization-based meta-learning, but it limits the adapted parameter subset to reduce overfitting and computational cost.

4.7 Calibrated selective decision head

Classification probabilities are not used directly as action guarantees. A calibration module is fitted on disjoint validation episodes, using temperature scaling for a single-parameter baseline or an episodic isotonic variant when sufficient validation data exist. Predictive uncertainty can combine calibrated entropy, ensemble variance, prototype dispersion, and graph disagreement. Each component is reported separately so that epistemic and data-quality indicators are not collapsed into an unexplained score.

Let $u(a, y)$ denote the utility of action a when the true state is y. The expected utility is computed under the calibrated predictive distribution. The framework selects the highest-utility action only if its estimated risk, uncertainty, and provenance criteria satisfy a predefined policy. Otherwise, it returns a deferral action accompanied by the available evidence and the reason for review.

$$a^* = \arg max_{a \in \mathcal{A}} \sum_{c \in \mathcal{C}_\mathcal{T}} \tilde{p}\,(y = c \mid x_q) u(a, c). \tag{9}$$

$$defer(x_q) = 1 \text{if} \hat{R}(a^* \mid x_q) > \delta_R, H(\tilde{p}) > \delta_H, \text{or} \Pi(x_q) = 0. \tag{10}$$

The indicator $\Pi(x_q)$ denotes whether minimum provenance requirements are satisfied. Examples include the presence of at least two independent evidence paths, a valid timestamp, or mandatory human confirmation for a high-impact action. Thresholds are selected on validation episodes according to review capacity and action costs, not tuned to maximize test accuracy. The resulting risk–coverage curve is reported alongside accuracy because a deferral rule that improves selected accuracy by reviewing almost every case may be operationally useless.

4.8 Training objective and episodic procedure

The training objective combines query classification, alignment consistency, reliability supervision, calibration, and selective risk. The principal loss is cross-entropy over query instances in each meta-training episode. Auxiliary terms encourage compatible observations to remain aligned, penalize reliability predictions that disagree with held-out correctness, and regularize attention entropy. A selective term penalizes confident errors and excessive deferral according to the chosen review cost.

$$\mathcal{L} = \mathcal{L}_{\text{query}} + \beta_A \mathcal{L}_{\text{align}} + \beta_Q \mathcal{L}_{\text{quality}} + \beta_C \mathcal{L}_{\text{cal}} + \beta_S \mathcal{L}_{\text{select}} + \beta_E \mathcal{L}_{\text{entropy}}. \tag{11}$$

Meta-training samples disjoint base-task episodes and computes the loss on queries after support-conditioned graph construction. Validation tasks determine encoder selection, graph density, attention regularization, and deferral thresholds. Test tasks remain untouched until all choices are fixed. Class-disjoint splits evaluate new-category adaptation, while domain-disjoint splits evaluate transfer to new users, devices, or operating environments. The two axes should not be conflated because success on unseen classes does not imply robustness to a shifted acquisition pipeline.

The inference procedure comprises nine steps: encode observations; validate availability and timestamps; instantiate typed nodes and relations; estimate support prototypes and reliability; propagate

graph messages; fuse query and prototype representations; calibrate predictive probabilities; compute expected utility and deferral criteria; and return a recommendation with provenance or a human-review request. Figure 3 depicts the support and query paths together with the authorized update boundary.

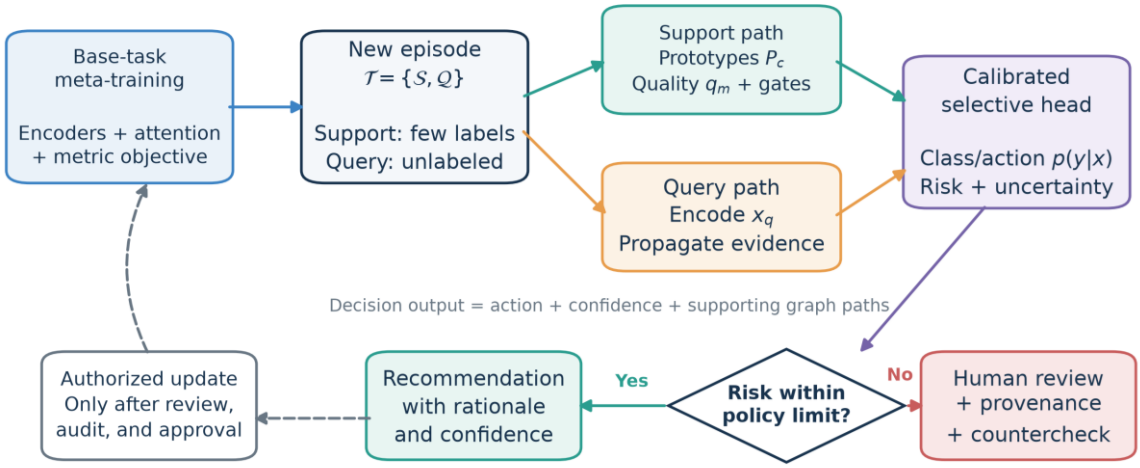


Figure 3 Support–query adaptation and decision-support workflow. Recommendations are issued only when policy criteria are satisfied; feedback enters a governed update path. Original schematic.

4.9 Complexity and implementation considerations

For L attention layers, edge count $|\mathcal{E}|$, node count $|\mathcal{V}|$, and latent dimension d, a dense relation-wise implementation has approximate time complexity $O(L|\mathcal{E}|d^2 + L|\mathcal{V}|d^2)$, with the exact cost determined by projection sharing and head dimension. Sparse typed relations and neighbor sampling reduce the first term. Prototype inference is linear in the number of episode classes for each query after graph propagation. Memory is dominated by stored edge messages and attention coefficients, making relation pruning important for long temporal sequences.

Implementation should preserve a graph manifest containing node identifiers, source references, relation rules, masks, and timestamps. Numerical stability requires an epsilon in prototype denominators and reliability logarithms, masked-softmax handling for empty neighborhoods, and normalization after residual aggregation. Reproducibility requires fixed episode seeds as well as model seeds. A model that cannot reconstruct the evidence graph used for a decision should not be considered audit-ready, even if its predictions are accurate.

5 Experimental Design

5.1 Research questions and evidence levels

The evaluation is organized around five questions. RQ1 asks whether relation-aware fusion improves few-shot query accuracy over shared-encoder early and late fusion. RQ2 asks whether reliability priors reduce degradation under noise, conflict, and missingness. RQ3 asks whether calibration and selective deferral improve risk–coverage behavior at feasible review rates. RQ4 asks which graph relations and adaptation components contribute to performance. RQ5 asks whether learned routing is consistent with provenance and counterfactual evidence perturbations. These questions belong to different evidence levels:

synthetic fusion experiments can address limited aspects of RQ1 and RQ2, whereas RQ3–RQ5 require the complete model and decision-layer evaluation.

A full benchmark should include at least one aligned multimodal classification dataset, one temporal sensing dataset with natural modality failures, and one domain-shifted decision-support scenario. Exact datasets should be selected according to the intended application rather than chosen post hoc to favor the model. Class-disjoint and domain-disjoint episodes must be generated without support–query leakage. Pretrained encoders, input normalization, and augmentation policies must be shared across baselines to make fusion comparisons interpretable.

Table 3 Required evaluation protocol for the complete model

Dimension	Required design	Purpose
Task generalization	Class-disjoint episodes	Measure adaptation to new categories
Domain generalization	User/device/environment-disjoint splits	Measure acquisition shift
Fusion comparison	Shared encoders and paired episodes	Isolate fusion effects
Robustness	Noise, conflict, missingness, alignment perturbation	Test controlled degradation
Decision quality	ECE, Brier score, risk–coverage, utility	Evaluate actionable predictions
Accountability	Provenance and counterfactual edge tests	Inspect evidentiary support
Efficiency	Latency, memory, graph size, parameters	Assess operational feasibility

5.2 Baselines and ablations

The principal baselines are uniform-distance fusion, late-probability fusion, quality-gated distance fusion, and the quality-plus-relation surrogate. The complete model should additionally be compared with concatenation-based multilayer perceptrons, intermediate cross-attention, prototypical networks without graph propagation, a homogeneous graph attention variant, and a heterogeneous graph variant without dynamic query conditioning. All methods should use the same support and query instances. Parameter counts and encoder pretraining must be disclosed.

Ablations remove relation attributes, quality priors, availability masks, prototype nodes, temporal edges, and residual connections separately. A particularly important control replaces learned attention with uniform weights while preserving the graph topology. This control distinguishes gains caused by graph construction from gains caused by attention. Another control shuffles relation types without changing node features, testing whether semantic typing contributes beyond generic message passing.

5.3 Metrics and statistical analysis

Query accuracy and macro-F1 measure classification, while expected calibration error and Brier score assess probability quality. Selective performance is summarized by risk–coverage curves, area under the risk–coverage curve, review rate, and utility per reviewed case. Robustness is measured by performance deltas under controlled modality corruption and by worst-group performance where meaningful subgroups exist. Computational reporting includes trainable parameters, inference latency, memory, and graph size.

$$ECE = \sum_{b=1}^{B} \frac{n_b}{n} |acc(b) - conf(b)|. \quad (12)$$

Confidence intervals are computed at the episode level rather than by treating correlated query instances as independent. Paired episode comparisons are preferred, and bootstrap intervals should preserve the episode hierarchy. Multiple dataset and ablation comparisons require correction or an explicitly exploratory interpretation. A statistically significant average gain is not sufficient if worst-domain performance degrades or if the review burden violates operational constraints.

5.4 Synthetic surrogate configuration

The supplied NumPy study isolates support-prototype fusion without implementing deep encoders or learned graph parameters. Each episode contains five classes, three modalities, twelve latent dimensions, and five queries per class. Orthonormal modality rotations preserve relational geometry while providing distinct observation coordinates. Independent modality noise, a 15% per-observation missing rate, and episode-specific class-mean permutations for one conflicting modality create controlled fusion challenges. Support size varies among one, three, and five examples per class.

The surrogate computes modality prototypes, support-derived quality scores, and max-consensus relational consistency. A modality receives high relational consistency when its class-distance structure and nearest-class relationships agree with at least one peer. Quality-plus-relation gating combines these quantities through softmax weights over available modalities. Validation episodes select quality and relation coefficients; disjoint test episodes evaluate all methods. Four hundred episodes are used for each support-size condition, 350 for each conflict-rate condition, and 500 for the ablation. The numerical outputs and generation scripts accompany this manuscript.

This design deliberately excludes deep representation learning, temporal dynamics, learned calibration, and decision utility. Consequently, the study cannot validate the full HMP-GAT architecture. Its purpose is narrower: to determine whether a transparent relation gate changes modality allocation and whether quality-aware prototype fusion behaves sensibly under the specified synthetic conditions.

6 Synthetic Results and Analysis

6.1 Shot efficiency

Accuracy increased sharply with support size for every method, confirming that the generated episodes retained a meaningful few-shot regime. At one shot, late-probability fusion achieved 49.19%, compared with 48.68% for uniform distance, 48.56% for quality gating, and 48.79% for quality-plus-relation gating. These differences are small relative to episode variability and do not justify a ranking. One-shot support cannot provide a direct within-class variance estimate, so quality scores rely on regularized global dispersion and are necessarily uncertain.

At three shots, quality-plus-relation gating achieved 77.15%, marginally exceeding quality-only gating at 77.11%, late-probability fusion at 76.39%, and uniform-distance fusion at 75.87%. At five shots, quality-only gating achieved 84.79%, while quality-plus-relation gating achieved 84.76%; the lower baselines

remained close at 83.33% and 83.20%. The dominant pattern is therefore improved performance with additional support and a modest advantage for quality-aware fusion at larger support sizes, not a consistent additional gain from relational gating.

Table 4 Synthetic query accuracy (%) with 95% episode-level confidence intervals

Fusion method	1-shot	3-shot	5-shot
Uniform distance	48.68 ± 1.37	75.87 ± 1.16	83.33 ± 1.03
Late probability	49.19 ± 1.37	76.39 ± 1.14	83.20 ± 1.04
Quality gate	48.56 ± 1.37	77.11 ± 1.12	84.79 ± 1.00
Quality + relation	48.79 ± 1.38	77.15 ± 1.15	84.76 ± 1.01

6.2 Conflict and missingness analysis

Across conflict probabilities from zero to 0.60, quality-aware methods generally remained above uniform and late-probability baselines. At zero conflict, quality-only and quality-plus-relation gating achieved 83.65% and 83.70%, respectively. At conflict probability 0.60, the corresponding values were 84.42% and 84.46%, compared with 83.13% for uniform distance and 83.22% for late probability. The lack of monotonic degradation across independently sampled conflict conditions indicates that two uncorrupted modalities often supplied sufficient evidence and that episode sampling variation exceeded the controlled conflict effect.

This finding should not be interpreted as evidence that modality conflict is harmless. The generator corrupts only one of three modalities, preserves modality-specific within-class structure, and provides relatively high query redundancy. More severe conditions—two conflicting modalities, shared bias across all channels, adversarial text, or support-specific corruption—could produce different behavior. The present experiment tests a limited compensation scenario and exposes the need for stronger stress tests before robustness claims are made.

6.3 Ablation and routing behavior

The separate five-shot ablation produced 82.92% accuracy for uniform-distance fusion, 83.03% for late-probability fusion, 84.44% for quality-only gating, and 84.38% for quality-plus-relation gating. The 0.06-percentage-point difference between the latter methods is negligible. Quality information accounts for the principal improvement in this synthetic setting, while the relation coefficient does not provide a measurable additional accuracy benefit.

Nevertheless, the relation term changes evidence allocation. In conflicted ablation episodes, the mean weight assigned to the conflicting modality decreases from 0.340 under quality-only gating to 0.321 under quality-plus-relation gating, while the mean weight on each remaining modality increases from 0.330 to 0.340. Uniform-distance fusion assigns approximately 0.337 to the conflicting modality. The relation gate therefore behaves in the intended direction, but the magnitude is modest and does not translate into an accuracy gain.

Table 5 Five-shot synthetic ablation and modality weights in conflicted episodes

Fusion method	Accuracy / %	Conflicting weight	Other weight
Uniform distance	82.92 ± 0.90	0.337	0.331
Late probability	83.03 ± 0.90	0.337	0.331
Quality gate	84.44 ± 0.87	0.340	0.330
Quality + relation	84.38 ± 0.87	0.321	0.340

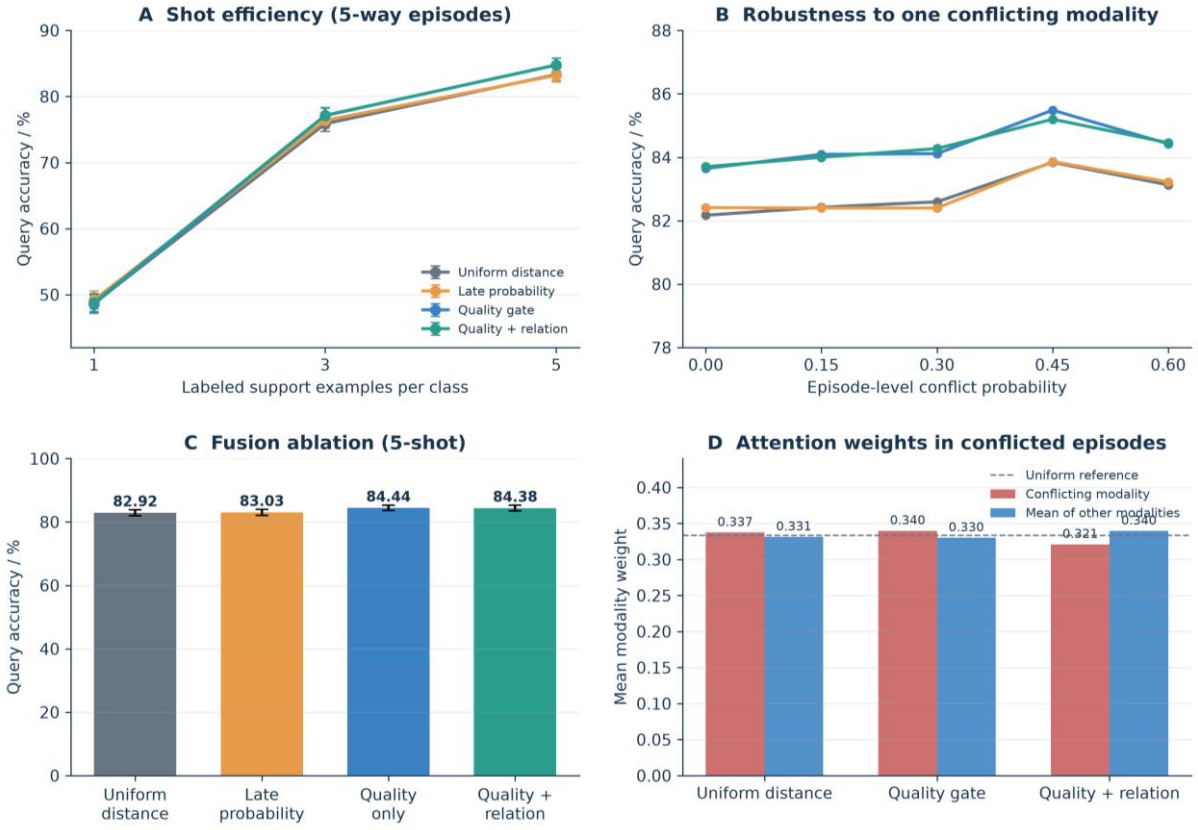


Figure 4 Synthetic surrogate results. Panels show shot efficiency, conflict robustness, ablation accuracy, and modality weights. Values are generated by the accompanying NumPy simulation; they do not constitute full-model or real-world validation.

6.4 Interpretation and evidentiary boundary

Three conclusions are supported. First, support size is a major determinant of few-shot fusion accuracy. Second, support-derived quality gating is beneficial in the tested noisy conditions. Third, max-consensus relational consistency can reduce allocation to a conflicting modality without requiring its identity to be known. Two further conclusions are not supported: that graph attention improves accuracy beyond quality gating, and that the full HMP-GAT model will reproduce these numerical patterns. The surrogate lacks learned representations, deep graph propagation, and decision-layer adaptation, so its results should be used to refine hypotheses rather than to claim validation.

The null relation effect is scientifically useful. It suggests that future work should test whether relational consistency adds information beyond compactness, whether max-consensus is too permissive when two modalities share a failure, and whether attention regularization is needed to prevent

overcorrection. It also motivates reporting weight dynamics alongside accuracy, since a mechanism can alter internal allocation without changing aggregate predictions.

7 Discussion

7.1 Why typed graph reasoning may still be valuable

The limited synthetic relation effect does not eliminate the rationale for typed graphs. Many decision-support failures arise from structural mismatches that flat fusion cannot express: a report linked to the wrong entity, a sensor window outside its valid interval, or an action whose constraints depend on provenance. Typed relations make these assumptions explicit and testable. They also permit targeted interventions, such as removing a temporal edge or requiring two independent support paths, without redesigning the entire network.

Graph reasoning is most likely to help when interactions are sparse, relational validity varies across episodes, and explanations must identify concrete evidence paths. It is less likely to help when modalities are already aligned, support quality estimates capture all relevant variation, or the graph introduces noisy edges. The synthetic results are consistent with the latter possibility: once quality gating identifies useful channels, a simple relation statistic may add little. Full-model evaluation should therefore compare graph construction and attention separately rather than treating their combination as an indivisible contribution.

7.2 Adaptation as a stability problem

Few-shot adaptation is often framed as rapid learning, but stability is equally important. Updating all encoder parameters from a handful of labels can destroy useful pretrained structure and produce episode-specific overfitting. HMP-GAT limits default adaptation to prototypes, reliability estimates, and a small set of gates. This restriction supports predictable behavior and simplifies rollback. Optional gradient adaptation should be enabled only when validation shows a benefit at the same support size and missingness profile expected in deployment.

Prototype adaptation also creates an inspectable record of why a new class was recognized. Support examples can be reviewed for label errors, duplication, leakage, and bias. However, inspectability does not guarantee correctness: a prototype can faithfully represent a mislabeled or unrepresentative support set. Data governance must therefore include support-set validation and mechanisms for removing harmful examples from subsequent episodes.

7.3 From predictive confidence to accountable action

The decision layer distinguishes several quantities that are frequently conflated. Class confidence describes the predictive distribution; reliability describes support and acquisition quality; uncertainty may reflect class overlap or novel conditions; and risk combines predictive uncertainty with action costs. A system can be confident but unreliable, or uncertain but still safe to act when all candidate actions have similar consequences. Keeping these quantities separate improves policy design and auditability.

Deferral should be evaluated as a resource-allocation decision. A low review rate may overload a safety-critical queue if many cases arrive simultaneously, while a high review rate can make the system operationally irrelevant. Policies should include queue capacity, reviewer expertise, response time, and escalation rules. For high-impact actions, provenance requirements may be mandatory even when predictive risk is low. The architecture supports such rules explicitly rather than assuming that a softmax threshold is a sufficient safety mechanism.

7.4 Explanation and human oversight

Graph attention can organize an explanation around supporting observations, prototypes, relations, and counterfactual changes. A useful output might state which support examples define the selected class, which modality was downweighted, and which alternative action would become preferable if a key observation were invalid. This is more informative than displaying attention weights alone. It also creates an opportunity for reviewers to correct graph relations rather than merely overriding a final label.

Human oversight must be designed to prevent automation bias. Reviewers should be able to inspect raw evidence, disagree with the recommendation, and record the reason for correction. Feedback should enter a governed update queue rather than trigger autonomous online learning. Periodic review of overrides can reveal whether errors arise from perception, graph construction, calibration, action costs, or workflow constraints. These diagnoses require different remedies and should not be collapsed into a single model-retraining decision.

7.5 Deployment and maintenance

A deployment pipeline should version encoders, graph rules, calibration parameters, and decision policies together. Changes to sensors or document formats can invalidate both embeddings and quality estimates, so monitoring should include availability patterns, relation statistics, calibration drift, and review outcomes. Shadow evaluation can compare a candidate version with the current policy without allowing it to influence actions. Rollback criteria should be defined before deployment.

Computational cost must be assessed at the actual episode size. Long temporal sequences can produce dense alignment graphs whose memory cost exceeds the encoder cost. Relation pruning, windowing, and neighbor sampling are therefore not merely efficiency optimizations; they shape the evidence available to the model. Any pruning rule must be evaluated for disparate impact on minority patterns that depend on rare or long-range relations.

8 Limitations, Ethics, and Reproducibility

8.1 Methodological limitations

The primary limitation is the absence of full-model real-world validation. The proposed architecture is specified mathematically, but the numerical study evaluates only a lightweight prototype surrogate with fixed orthonormal transformations. It does not measure learned encoder quality, deep graph propagation,

calibration, selective risk, or human decision outcomes. No claim of state-of-the-art performance, clinical utility, industrial safety, or deployment readiness is warranted.

The graph formulation also introduces assumptions. Relation rules may be incomplete or incorrect, quality estimates may be biased, and max-consensus can fail when multiple modalities share correlated errors. Attention can amplify spurious alignment, and prototype adaptation may be inappropriate for multimodal class distributions. More expressive prototypes, density-based estimators, or hierarchical class models may be required when a single mean cannot represent a category.

8.2 Ethics, privacy, and fairness

Heterogeneous decision support can combine sensitive attributes that were innocuous when stored separately. Graph manifests should therefore respect purpose limitation, access control, retention policies, and applicable consent requirements. Provenance does not justify unrestricted data linkage. Context nodes representing users, locations, or devices should be included only when necessary and authorized. Privacy-preserving computation and data minimization should be considered at graph construction rather than deferred to output filtering.

Fairness analysis must consider both errors and deferrals. If a subgroup receives noisier sensor data or less complete documentation, quality gating may systematically reduce its evidence weight or increase its review burden. Conversely, forced automation to equalize review rates could expose that subgroup to greater risk. Evaluation should report accuracy, calibration, coverage, and downstream outcomes by relevant groups, with uncertainty intervals and attention to small-sample limits.

8.3 Reproducibility and reporting standards

The synthetic study is deterministic given its declared seeds and NumPy implementation. The accompanying files include episode generation, validation-based coefficient selection, disjoint test evaluation, numerical outputs, and figure generation. A complete replication package for HMP-GAT should additionally specify encoder checkpoints, graph-construction rules, episode manifests, hyperparameter search spaces, calibration splits, hardware, and decision thresholds. Negative and null results should be retained because they identify where graph complexity does not help.

Future research should prioritize three tests: domain-disjoint evaluation with natural modality failures; paired counterfactual analysis of graph edges and reliability estimates; and prospective assessment of human review under realistic workload. Cross-dataset calibration and review-capacity analysis are particularly important because a fixed threshold may not transfer across operating conditions. Only after these tests can the framework’s decision-support value be distinguished from its representational appeal.

9 Conclusion

HMP-GAT provides a unified formulation for heterogeneous multimodal perception, few-shot adaptation, and selective decision support. Its episode graph preserves observation, modality, prototype, and decision semantics; dynamic relation-aware attention controls evidence propagation; and the decision

head connects calibrated predictions with utility, provenance, and human review. The design addresses limitations of flat fusion by making alignment, missingness, reliability, and action constraints explicit.

The synthetic proof-of-concept demonstrates sensible shot-efficiency and quality-gating behavior, while showing no uniform advantage from adding relational consistency. The relation gate modestly reduces allocation to conflicting modalities but does not improve accuracy beyond quality-only fusion in the reported ablation. This distinction reinforces the need to evaluate graph topology, attention, reliability, and decision policy as separate components.

The framework should therefore be regarded as a rigorous basis for subsequent empirical investigation rather than a validated deployed system. Its value lies in structuring multimodal evidence and accountability requirements so that future benchmarks can determine when the added graph complexity yields measurable benefits, when simpler quality-aware fusion is sufficient, and how adaptive predictions can be converted into responsible decisions.

References

[1] Wang, Z., Kiran, R., Nair, J. P., Chen, C.-H., Chou, T.-H., Tsai, S., & Zhang, R. (2025). Sparsity-aware near-field beam training via multi-beam combination. In Globecom (pp. 2795–2800).

[2] Ma, Yifei; Qiu, Tianyu; Liu, Zhiwei; Zhou, Die; Huang, Zhihong; Cheng, Xuemei; Yang, Jinlin; Li, Jing; Tian, Xinlong; Zhang, Hui. (2026). Concentration-Gradient Solid-State Sulfur Cathode Enabling Improved Energy Storage Under High Sulfur Loadings. Advanced Functional Materials, 36(39): e30425. https://doi.org/10.1002/adfm.202530425

[3] Lyu, M., Sun, Y., Lin, E., Li, H., Chen, R., Zhao, F., & Zeng, Y. (2026, June). Reinforcement fine-tuning of flow-matching policies for vision-language-action models. In 2026 IEEE International Conference on Robotics and Automation (ICRA) (pp. 9242-9249). IEEE.

[4] [Ye, Z., Xiao, X., & Wang, H. (2024, July). Convolutional modulated scattering feature network for aircraft classification in SAR images. In IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium (pp. 9329-9332). IEEE.

[5] Lan, J., Ye, Z., Li, R., Qiu, K., Li, P., Guo, X., & Hu, F. (2026). A Multi-Level Cross-Modal Edge Filtering Method for High-Resolution Optical-SAR Image Registration. Remote Sensing, 18(11), 1741.

[6] Li, P., Ye, Z., & Wang, H. (2026, May). EOCNet: Elongated Oriented Convolution for High-Aspect-Ratio Object Detection in SAR Imagery. In 2026 China Aerospace Information Technology Conference (CAIT) (pp. 1-4). IEEE.

[7] C. Wan, Y. Luo, W. Wu, N. D. Hatziargyriou and S. Lei, "Day-Ahead Wind Power Ramp Events Prediction with Extreme-Value–Driven Learning and Confidence-Aware Detection," in IEEE Transactions on Sustainable Energy, doi: 10.1109/TSTE.2026.3669605.

[8] Y. Wang, C. Wan and Y. Luo, "WeaGAN++: An Efficient Weather-Aware Graph Attention Network for Traffic Prediction," in IEEE Internet of Things Journal, vol. 13, no. 10, pp. 21197-21212, 15 May15, 2026, doi: 10.1109/JIOT.2026.3665116.

[9] Pathak, A., Anwer, T., Verma, A., Alhujaily, M., Alahmari, M. A., Alshahrani, S., Alam, N., Nomier, Y., & Alam, M. F. (2026). Integrating precision medicine and artificial intelligence to prevent cardiotoxicity in cardiovascular drug therapy. Precision medication, 3(1), 100078.

[10] You, L., Luo, Y., Cheng, Q., Shen, L., & Ge, J. (2025). High-suitcordance intelligent fibers for panvascular disease monitoring-intervention. Advanced Fiber Materials, 7(4), 1042–1072.

[11] Quan, Q., Gao, Y., & Wang, Q. (2025). The impact of teacher emotional support on students' engagement in AI-mediated English learning environments: The mediating role of resilience and self-efficacy. Acta Psychologica, 260, 105766.

[12] Xu, Y., Sun, Y., Zhai, B., Li, M., Liang, W., Li, Y., & Du, S. (2025, April). Zero-shot video moment retrieval via off-the-shelf multimodal large language models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 9, pp. 8978-8986).

[13] Yang, J., Li, Q., Yu, H., Wang, X., Wang, C., Ma, B., ... & Li, X. (2026). Invisible Image Watermark Attacks: A Survey.

[14] Su, J., Zhu, F., Shi, H., Han, T., Qiu, Y., Luo, J., ... & Gao, J. (2025). MAN++: Scaling Momentum Auxiliary Network for Supervised Local Learning in Vision Tasks. arXiv preprint arXiv:2507.16279.

[15] Zhou, Y., Ye, W., Liu, Y., Dong, Z., & Yao, J. (2026). Style Over Substance: Content-Invariant Wrappers Flip LLM Safety-Judge Verdicts. arXiv preprint arXiv:2609.08236.

[16] Tan, Y., Cheng, X., Qin, Y., Li, Z., & Zhang, J. (2026). Towards Stable Self-Supervised Object Representations in Unconstrained Egocentric Video. arXiv preprint arXiv:2603.13912.

[17] Ma, Q., Tang, S., Meng, Q., & Wu, J. (2026). Measuring failure yield and reflection cost in a frozen robot policy. In NeurIPS 2026 Workshop on Self-Evolving Diversity-Driven Search for Robust AI Systems.

[18] Liang, J., Yang, H., Zhao, X., Yu, Z., Xu, G., Wang, W., & Yang, K. (2026). Democratic recommendation with user and item representatives produced by graph condensation. IEEE Transactions on Knowledge and Data Engineering.

[19] Wu, H., Song, Y., Zhang, Z., Zhang, Z., Zhou, H., Ma, W. Y., & Liu, J. (2026). Rationalized All-Atom Protein Design with Unified Multi-Modal Bayesian Flow. Advances in Neural Information Processing Systems, 38, 3019-3050.

[20] Huang, C., Jia, Z., Fei, H., Zhu, Y., Yuan, Z., Zhang, J., & Zhou, J. (2025, April). Semantic to structure: Learning structural representations for infringement detection. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[21] Jia, Z., & Huang, C. (2025). Automated Framework for Extracting and Restoring Minutiae From Low-Quality Fingerprints. IEEE Signal Processing Letters.

[22] Ning, E., Gao, W., Liu, Y., Yuan, J., Zhang, H., He, S., & Ning, X. (2026). Causally invariant video anomaly detection via counterfactual reasoning and prototype intervention. Pattern Recognition, 113809.

[23] Lv, Z., Wang, Q., Gao, W., & Zhou, L. (2026). Enhancing Corporate Green Innovation Under the Green Factory Policy: A Graph Neural Network Approach for Management Decision-Making. Journal of Organizational and End User Computing (JOEUC), 38(1), 1-33.

[24] Yan, Z., & Lyu, T. (2026). Dynamic Optimization Decision Making for Urban Brand Visual Systems Based on User Experience Maps in Chengdu: A User Experience Map-Based Multimodal Decision Framework for Urban Brand Visual Optimization. Journal of Organizational and End User Computing (JOEUC), 38(1), 1-57.

[25] Huang, T., Huang, C., Li, S., & Lyu, T. (2026). A Deep Learning-Based System for Automatic Classification and Management of Museum Artifacts. Journal of Organizational and End User Computing (JOEUC), 38(1), 1-26.

[26] Dong, J., Shen, Z., Luo, H., Lyu, T., & Gao, W. (2026). Graph attention and text semantics improve personalized recommendation. Scientific Reports.

[27] Wang, X., Liu, S., Qi, Z., Zhou, L., & Cui, Q. (2025). Bayesian-Based Green Financial Risk Modeling With Fuzzy Logic and Attention-Driven Feature Selection: An Integrated Approach for Risk

Prediction in Green Finance. Journal of Organizational and End User Computing (JOEUC), 37(1), 1-23.

[28] QIN, Y., Su, H., Wang, N. Y., Shang, D., Shi, J. S., & Hu, X. (2026). Roadside multimodal sensor fusion for 3-d object detection: A review. IEEE Sensors Journal, 26(14), 20592–20607.

[29] Zhang, D., Sang, C., & Lyu, T. (2026). Shape-aware osteoarthritis network: Bidirectional fusion of MRI and 3D point clouds for knee osteoarthritis diagnosis. Information Fusion, 104198.

[30] Lv, Z., Wang, Q., Lyu, T., & Dong, J. (2026). A Deep Learning-Based Model for User Emotion Recognition and Behavior Guidance Design: EMoGuideNet-An Emotion-Aware Multimodal Framework for Enhancing Intelligent User Interaction. Journal of Organizational and End User Computing (JOEUC), 38(1), 1-30.

[31] Risk-Aware Federated Alignment for Trustworthy and Controllable Language-Guided Diffusion Editing, IEEE Transactions on Dependable and Secure Computing, 1-14, 2026.

[32] Lyu, M., Guo, J., Gao, W., & Lyu, T. (2026). Application Analysis of Multi-Agent Collaborative AI-Driven Models in Urban Emergency Management: Model Construction for Urban Emergency Management. Journal of Organizational and End User Computing (JOEUC), 38(1), 1-26.

[33] Yu, C., Li, P., Wu, H., Wen, Y., Deng, B., & Xiong, H. (2024). USM: Unbiased Survey Modeling for Limiting Negative User Experiences in Recommendation Systems. arXiv preprint arXiv:2412.10674.

[34] Wu, X., Xing, X., Yao, J., Qian, Q., & Song, J. (2025). Scnet: spectral convolutional networks for multivariate time series classification. Applied Intelligence, 55(6), Article 456. https://doi.org/10.1007/s10489-025-06352-1

[35] Ma, C., Song, J., Xu, Y., Fan, H., Wu, X., & Sun, T. (2024). Vehicle-Based Machine Vision Approaches in Intelligent Connected System. IEEE Transactions on Intelligent Transportation Systems, 25(3), 2827-2836. https://doi.org/10.1109/TITS.2023.3276325

[36] Liu, T., He, J., Zhou, R., Song, J., Liu, H., & Liang, Y. (2025). Revolutionizing clinical decision making through deep learning and topic modeling for pathway optimization. Scientific Reports, 15(1), Article 28787. https://doi.org/10.1038/s41598-025-12679-z

[37] Wang, T., Xu, Y., Zhao, H., Li, L., Huang, P., & Zeng, G. (2023). Multi-rate and multi-protocol continuous-variable quantum key distribution. Optics Letters, 48(3), 719-722.

[38] Zhao, H., Wang, T., Xu, Y., Li, L., Tan, Z., Tan, P., ... & Zeng, G. (2024). Continuous-variable quantum key distribution robust against environmental disturbances. Optics Express, 32(5), 7783-7799.

[39] Li, L., Wang, T., Huang, P., Xu, Y., Liu, X., Zhao, H., & Zeng, G. (2024, April). Practical Source Security of On-chip Continuous Variable Measurement Device Independent Quantum Key Distribution. In 2024 Photonics & Electromagnetics Research Symposium (PIERS) (pp. 1-6). IEEE.

[40] Dong, J., Wang, T., He, Z., Shi, Y., Li, L., Huang, P., & Zeng, G. (2023). Effective Excess Noise Suppression in Continuous-Variable Quantum Key Distribution through Carrier Frequency Switching. Entropy, 25(9), 1286.

[41] Wang, T., Huang, P., Li, L., Zhou, Y., & Zeng, G. (2024). High key rate continuous-variable quantum key distribution using telecom optical components. New Journal of Physics, 26(2), 023002.

[42] Hao, L., Zhang, S., & Schulzrinne, H. (2025, October). Advancing IoT System Dependability: A Deep Dive into Management and Operation Plane Separation. In 2025 IEEE 11th World Forum on Internet of Things (WF-IoT) (pp. 1-6). IEEE.

[43] Zuo H, Lim KH, Huang H, et al. Cognitive function and eye movement patterns during reading tasks in older adults. Alzheimer's Dement. 2026; 18:e70464. https://doi.org/10.1002/dad2.70464

[44] Wu, J., Huang, H., Zuo, H., Lv, T., Zou, Y., & Wu, M. (2025). Low-rate VR eye tracking differentiates Alzheimer's disease using recurrence and entropy. In 2025 10th International Conference on Computer and Information Processing Technology (ISCIPT). https://doi.org/10.1109/ISCIPT67144.2025.11264993

[45] Chen, B., Zuo, H., Liao, Y., & Li, Z. (2024, December). A Data-Driven Analysis of Mental Health Status and Early Warning System for Vocational College Students. In 2024 5th International Conference on Information Science and Education (ICISE-IE) (pp. 312-315). IEEE.

[46] Phang, J. T. S., Lim, K. H., Lease, B. A., & Zuo, H. (2024, November). Markerless Avatar Animation with Forward Kinematics using Depth Sensor. In 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 48-51). IEEE.