# LAMPEAE: LLM-Augmented Manifold Probing for Adaptive Event Argument Extraction

Anonymous ACL submission


## Abstract

Event Argument Extraction (EAE) aims to extract arguments for specified events from a text. Previous prompt-based research has mainly focused on designing static role anchors, effectively modeling general semantics but overlooking the geometric rigidity of these representations: (i) they failed to capture the diverse manifold structure of real-world semantic distributions and (ii) they lacked the flexibility to adapt to instance-specific reasoning contexts. To bridge the gap between static parameters and dynamic contexts, we introduce a new framework named LAMPEAE, which reformulates EAE as a dynamic manifold matching problem via adaptive parameter instantiation. Specifically, we employ a neuro-symbolic approach that utilizes a frozen LLM to extract high-order reasoning priors. A lightweight hypernetwork then maps this meta-knowledge into low-rank geometric transformation matrices, which dynamically project original static prompts into instance-specific probes. This mechanism ensures precise semantic alignment within the continuous parameter space. Experimental results on the RAMS, ACE05, and WikiEvent benchmarks show that LAMPEAE establishes new state-of-the-art performance, respectively, effectively validating its superiority in handling heterogeneous semantic transfer.


## 1 Introduction

Event Argument Extraction (EAE) aims to identify entity mentions that fulfill specific roles for event triggers (Huang et al., 2024). While early studies primarily formulated EAE as sequence labeling (Nguyen and Grishman, 2015; Chen et al., 2015; Lin et al., 2020), they treated argument roles as semantically agnostic discrete indices, struggling with generalization in low-resource scenarios. To bridge this gap, the Question Answering (QA) paradigm was introduced (Du and Cardie, 2020; Liu et al., 2020; Uddin et al., 2024), transforming roles into natural language queries to leverage the semantic priors of Pre-trained Language Models (PLMs). However, this semantic alignment often incurs a severe efficiency bottleneck: since QA-based methods typically require separate document encodings for each role, they suffer from prohibitive computational latency, restricting their applicability in real-time document-level extraction (Li et al., 2020; Liu et al., 2023).

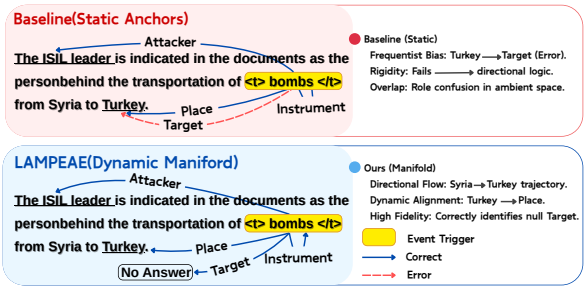



Figure 1: Examples of event argument extraction. Trigger words are included in special tokens <t> and </t>. Underlined words denote arguments and arcs denote roles.

Currently, generation-based and prompt-based paradigms dominate the field. For example, DEGREE (Hsu et al., 2022) formalized EAE as a conditional generation task, exploiting the generative prowess of LLMs (Lin et al., 2025). PAIE (Ma et al., 2022) proposed a prompt-based method that extracts arguments via slot filling, significantly improving data efficiency. Building on this, TabEAE (He et al., 2023) extended prompt-based models into a non-autoregressive framework to capture event cooccurrence relationships and parallelize argument extraction. These approaches achieve a superior balance between performance and efficiency.However, these methods overlooked a critical inductive bias: the significant intra-class variance inherent within argument roles. Existing prompt-based models typically rely on static anchors, which force the model to fit a single geo-

metric centroid. We argue that this represents a form of geometric rigidity: a fixed vector can only represent the "average" semantics of a role, failing to align with the fragmented and diverse instances found in real-world contexts. As shown in Figure 1, while multi-prototype frameworks like HMPEAE (Zhang et al., 2024a) attempt to address these discrepancies, their discrete nature inevitably leaves coverage vacuums—intermediate semantic regions between finite centers that the model cannot represent.

Grounded in the manifold hypothesis (Bengio et al., 2013), we argue that semantic drift such as a Target role shifting from a specific person to a complex infrastructure is characterized by high anisotropy and continuity (Ethayarajh, 2019). Consequently, argument roles are better modeled as continuous trajectories on a semantic manifold (Chronis and Erk, 2020) rather than static, isolated points. Such a manifold allows the model to capture the smooth transition of role meanings across different contexts, ensuring that instance specific nuances are not lost to the rigidity of fixed parameters.To implement Dynamic Manifold Matching efficiently, we propose LAMPEAE, addressing previous limitations via uncertainty-based dynamic gating and hypernetwork-driven manifold projection. Built upon the PAIE backbone, LAMPEAE transforms static prompt embeddings into instance-specific representations based on inference context. We formulate this as a dynamic manifold matching problem, utilizing low-rank constraints to minimize parameter redundancy. This process effectively learns a continuous function that warps static anchors along context trajectories, enabling precise instance level alignment within the embedding space.Our contributions are summarized as follows:

- We challenge the Static Parameterization Hypothesis—the flawed assumption of fixed argument roles—by modeling argument semantics as continuous drifts. By replacing discrete prototypes with a dynamic manifold projection, LAMPEAE effectively captures intra-class variance in complex, long-tail scenarios.

- We introduce a Neuro-symbolic hypernetwork to enable instance-specific probe generation. This mechanism dynamically instantiates probes based on document-level contexts, directly bridging semantic coverage vacuums in the parameter space and optimizing representation learning for heterogeneous event structures.

- LAMPEAE establishes new state-of-the-art results on RAMS, ACE05, and WikiEvents (54.5%, 74.1%, and 66.9% Arg-C $F_1$). Significant gains on long-range dependencies and complex samples robustly validate the superiority of our dynamic manifold matching over conventional static architectures.

## 2 Method

Following prior studies (Ma et al., 2022; He et al., 2023), we build LAMPEAE upon the prompt-based span selection paradigm. Figure 2 illustrates the overall architecture, which introduces a **Dynamic Manifold Matching** mechanism to transcend the limitations of static role anchors.

### 2.1 Problem Formulation

We formulate Event Argument Extraction (EAE) as a prompt-based span extraction task. Given a document $\mathcal{D} = \{w_1, \ldots, w_N\}$ and an event trigger $t$ with roles $\mathcal{R}_e = \{r_1, \ldots, r_K\}$, the goal is to extract argument spans $\mathcal{A} = \{(r, s, e)\}$, where $s$ and $e$ denote the start and end indices.

**Static Role Probing.** Traditional methods typically initialize a learnable static embedding matrix $\mathbf{P} \in \mathbb{R}^{K \times d}$ as role anchors. The extraction probability for role $r_k$ is determined by the interaction between the context $H_{\mathcal{D}}$ and the anchor $p_k \in \mathbf{P}$:

$$P(s, e|r_k) = \prod_{i \in \{s,e\}} \text{Softmax}(H_{\mathcal{D}} \mathbf{W}_i p_k^T) \quad (1)$$

where $\mathbf{W}_{s,e}$ are globally shared projection weights. As these weights are instance-invariant, the model implicitly treats roles as fixed geometric centers, struggling to adapt to long-tail distributions characterized by significant semantic drifts.

**Dynamic Manifold Probing.** To address this, we reformulate EAE as a dynamic manifold matching problem. We hypothesize that the optimal role anchor is a function of the instance-specific reasoning context $z \in \mathcal{Z}$. We define a conditional mapping $\mathcal{F}_\phi : \mathcal{Z} \to \mathbb{R}^{d \times d}$ that dynamically transforms the static anchor $p_k$ into an Instance-Specific Probe $p'_k(z)$:

$$p'_k(z) = p_k \cdot \mathcal{F}_\phi(z) + p_k \quad (2)$$

where the residual connection ensures transformation stability. The final extraction objective is refor-

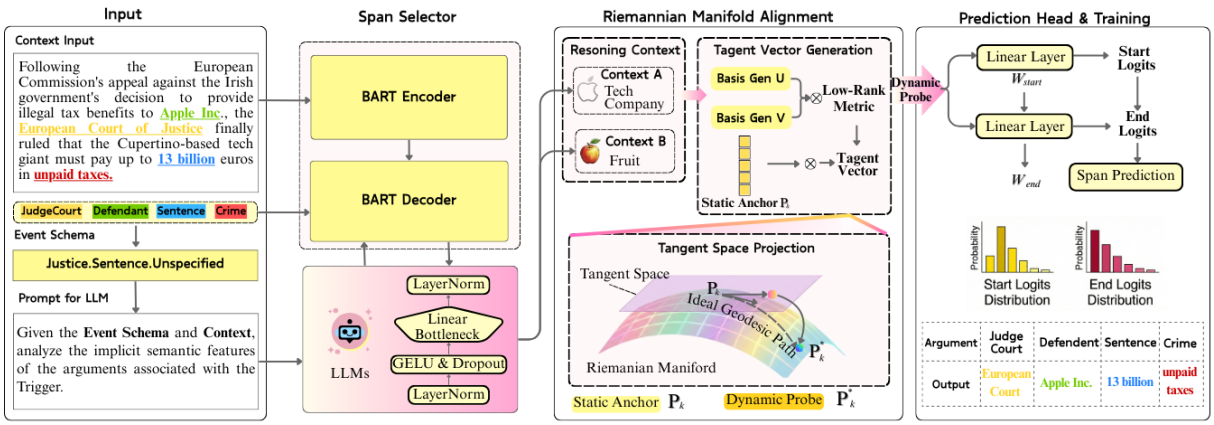



Figure 2: The overall architecture of the proposed **LAMPEAE** framework. The process consists of three distinct phases: (1) **Neuro-Symbolic Reasoning Engine**: A frozen LLM extracts high-order implicit reasoning context ($z$) from the input document and schema-aware prompts; (2) **Riemannian Manifold Alignment**: The static role anchor ($P_0$) is dynamically projected onto the tangent space via Low-Rank Approximation (Modules U & V) to generate an instance-specific probe ($P^*$); (3) **Prediction Head**: The dynamic probe is utilized to compute start and end logits for precise argument span extraction.

mulated as:

$$\widehat{\mathcal{A}} = \arg\max_{\mathcal{A}} \sum_{(k,s,e)\in\mathcal{A}} \log \text{Score}(H_{\mathcal{D}}, p'_k(z)) \quad (3)$$

Geometrically, this mechanism enables continuous semantic interpolation. Unlike discrete prototypes that cover limited typical semantics, LAMPEAE's dynamic transformation allows probes to traverse along the manifold trajectory, precisely capturing intermediate states or hybrid samples in long-tail semantic regions.

## 2.2 Neuro-Symbolic Reasoning Engine

This module distills high-order implicit logic from the document into differentiable neural signals to drive parameter transformation. The process consists of three stages: reasoning prompt construction, trajectory distillation, and semantic alignment.

**Reasoning Prompt Construction.** We design a meta-cognitive reasoning instruction $\mathcal{T}$ to induce the LLM to analyze the latent semantic attributes of argument roles rather than performing direct extraction.

Unlike the reinforced reasoning optimization employed by REAR (Luo et al., 2025), LAMPEAE emphasizes mapping these reasoning priors directly into the geometric manifold space. Given document $\mathcal{D}$, trigger $t$, and event type $e$, the input sequence $X_{sym}$ is formalized as:

$$X_{sym} = [CLS] \oplus \mathcal{D} \oplus [SEP] \oplus \mathcal{T}_{reason}(t, e) \quad (4)$$

where $\mathcal{T}_{reason}$ is a structured prompt function. This design forces the model to explicitly attend to role-specific deep semantic features—such as entity properties and directional relationships—during the forward pass.

**Reasoning Trajectory Distillation.** To ensure efficiency, we bypass auto-regressive decoding and employ a single forward pass. We extract the hidden state of the final token $h_{last} \in \mathbb{R}^{d_{llm}}$ as the condensed terminal state of the reasoning trajectory:

$$h_{last} = \text{LLM}_{\Theta}(X_{sym})[-1, :] \quad (5)$$

$h_{last}$ encapsulates the LLM's global context understanding after processing the entire instruction under a causal attention mask, serving as a high-dimensional representation of the implicit reasoning logic.

**Semantic Alignment.** To mitigate the **Manifold Heterogeneity** between the LLM feature space and the EAE backbone, we introduce a **Semantic Bottleneck Layer** as a task-aware projector. The raw signal $h_{last}$ is projected into the task-specific reasoning context $z_{sym} \in \mathbb{R}^d$:

$$z_{sym} = \text{LayerNorm}(\sigma(h_{last}\mathbf{W}_p + b_p)) \quad (6)$$

where $\mathbf{W}_p \in \mathbb{R}^{d_{llm}\times d}$. This alignment filters out general-purpose redundancy from the LLM, retaining only the neural context required to warp the static anchors along the context-driven manifold trajectories.

### 2.3 Tangent Space Projection via Low-Rank Approximation

To rigorously model the dynamic matching, we treat the semantic embedding space as a Riemannian Manifold $\mathcal{M}$, where the static anchor $p_k \in \mathcal{M}$ serves as the initial coordinate. Rather than relying on a purely abstract assumption, we argue that semantic drift in EAE aligns with the Manifold Hypothesis: the instance-specific variations of role $k$ reside on a low-dimensional topological structure embedded within the high-dimensional embedding space. Our goal is to derive a geodesic path driven by the context $z$ that moves $p_k$ to an instance-specific location $p'_k(z)$ (Ma et al., 2022). This involves generating a tangent vector in $T_{p_k}\mathcal{M}$ and mapping it back to the manifold via Retraction. The full procedure is detailed in Algorithm 1.

**Low-Rank Tangent Vector Generation.** The semantic drift driven by $z$ is represented as a tangent vector $\xi_{k,z} \in T_{p_k}\mathcal{M}$. While any displacement can be described via a linear transformation $\mathbf{W}(z) \in \mathbb{R}^{d\times d}$ as $\xi_{k,z} = p_k \cdot \mathbf{W}(z)$, directly learning $\mathbf{W}(z)$ is an ill-posed problem prone to overfitting due to high parameter redundancy.

Grounded in the low intrinsic dimensionality of pre-trained models and the Eckart-Young-Mirsky Theorem, we posit that effective semantic shifts are concentrated in a low-dimensional subspace of the tangent bundle. Thus, we approximate $\mathbf{W}(z)$ using a low-rank decomposition of rank $r \ll d$ to perform geometric regularization. The superior performance of our low-rank constraint over higher-rank alternatives—validated by the performance peak at $r = 32$ empirically confirms that the instance-specific semantic drift adheres to a low-dimensional manifold structure rather than arbitrary high-dimensional noise.

In practice, our hypernetwork predicts the singular bases $\mathbf{A}(z), \mathbf{B}(z) \in \mathbb{R}^{d\times r}$ to compute the tangent vector:

$$\xi_{k,z} = p_k \cdot (\mathbf{A}(z)\mathbf{B}(z)^T) \tag{7}$$

This low-rank manifold constraint compels the semantic drift to occur along $r$ principal directions, effectively filtering out high-dimensional noise and ensuring the dynamic probe remains within a stable, task-relevant semantic region.

**Manifold Retraction and Update.** To map the update from the tangent space back to $\mathcal{M}$, we approximate the riemannian exponential map using a first-order retraction $\mathcal{R}_{p_k}$:

$$\mathcal{R}_{p_k}(\xi_{k,z}) \triangleq p_k + \alpha \cdot \xi_{k,z} \tag{8}$$

where $\alpha$ is a learnable step-size factor. To ensure the updated probe remains on a stable manifold surface, we incorporate Layer Normalization to maintain geometric validity and prevent semantic divergence. The final dynamic probe is formulated as:

$$p'_k(z) = \text{LayerNorm}(p_k + \alpha \cdot (p_k \cdot \mathbf{A}(z)\mathbf{B}(z)^T)) \tag{9}$$

Geometrically, this operation projects context $z$ onto the tangent space of role $k$ and transports the anchor along the learned manifold trajectory, warping the static representation to the instance-specific semantic region.

### 2.4 Manifold Regularization and Span Extraction

We utilize a pre-trained BART-Encoder to obtain context representations $H_{\mathcal{D}}$. In LAMPEAE, the traditional static query for role $k$ is replaced by the instance-specific probe $p'_k(z)$ derived in Section 3.3. Formally, the span scoring logits are defined as:

$$\text{Logits}_k(z) = H_{\mathcal{D}} \cdot (p'_k(z) \circ \mathbf{W}_{proj})^T \tag{10}$$

where $\circ$ denotes the element-wise product. To handle multiple arguments per role and ensure permutation invariance, we employ bipartite matching via the Hungarian algorithm to determine the optimal assignment $\sigma^*$ between predicted and ground-truth spans.

**Dynamic Span Extraction Loss.** We define the span selection probability using a normalized energy-based formulation. For role $k$, let $\mathbf{u}_k^{(s)}(z) = p'_k(z) \circ \mathbf{w}^{(s)}$ be the dynamic start selector. The probability of the $i$-th token being the start index is:

$$P_k^{(\text{start})}(i|z) = \frac{\exp(h_i \cdot \mathbf{u}_k^{(s)}(z)^T)}{\sum_{j=1}^{N} \exp(h_j \cdot \mathbf{u}_k^{(s)}(z)^T)} \tag{11}$$

The end position probability $P_k^{(\text{end})}$ is computed analogously. The primary task loss $\mathcal{L}_{span}$ is the negative log-likelihood of the ground-truth spans $\mathcal{A}^*$:

$$\mathcal{L}_{span} = -\frac{1}{|\mathcal{A}^*|} \sum_{(k,s,e)\in\mathcal{A}^*} \big[\log P_k^{(\text{start})}(s|z) + \log P_k^{(\text{end})}(e|z)\big] \tag{12}$$

**Manifold Orthogonality Regularization.** To enforce inter-role discriminability while probes traverse the manifold, we introduce an orthogonality loss to minimize the squared cosine similarity between distinct dynamic probes:

$$\mathcal{L}_{orth} = \sum_{i=1}^{K} \sum_{j \neq i} \left( \frac{p'_i(z) \cdot p'_j(z)}{\|p'_i(z)\|_2 \cdot \|p'_j(z)\|_2} + \epsilon \right)^2 \tag{13}$$

By pushing probes toward mutual orthogonality, we maximize the inter-role margin in the dynamic embedding space. This prevents semantic confusion between functionally distinct roles that might otherwise drift into overlapping regions during instance-level adaptation.

**Magnitude Constraint.** To ensure geometric stability and prevent the dynamic probe from diverging from the learned manifold structure, we constrain the displacement magnitude $\Delta p_k = p'_k(z) - p_k$ within a local neighborhood $\delta$:

$$\mathcal{L}_{norm} = \frac{1}{K} \sum_{k=1}^{K} \max(0, \|\Delta p_k\|_2 - \delta) \tag{14}$$

This ensures the dynamic warping remains geometrically valid on the manifold surface, keeping the probe anchored by global role priors even in complex document-level contexts.

**Total Objective.** The final training objective $\mathcal{J}$ balances extraction accuracy and geometric stability:

$$\mathcal{J} = \mathcal{L}_{span} + \lambda_1 \mathcal{L}_{orth} + \lambda_2 \mathcal{L}_{norm} \tag{15}$$

where $\lambda_1, \lambda_2$ are balancing hyperparameters.

# 3 Experiment

We design experiments to address the following four research questions, structured to validate both the empirical effectiveness and the theoretical foundation of LAMPEAE:

**RQ1 (Performance Superiority):** Can dynamic manifold matching significantly outperform state-of-the-art static baselines across diverse Event Argument Extraction benchmarks?

**RQ2 (Manifold Validation & Coverage):** Can we empirically prove that event semantics adhere to the manifold hypothesis, and does our continuous interpolation successfully resolve the semantic coverage vacuum?

**RQ3 (Universality & Generalization):** Is the proposed dynamic manifold matching architecture model-agnostic across different LLMs and applicable to other Information Extraction (IE) tasks?

**RQ4 (Constraint Mechanics & Safety):** Does the introduction of manifold-based geometric priors effectively mitigate representation collapse and suppress the stochastic noise inherent in LLM-integrated reasoning?

## 3.1 Experimental Settings

**Datasets and Metrics.** We evaluate on three datasets: RAMS (Ebner et al., 2020), WikiEvents (Li et al., 2021), and ACE05 (Walker et al., 2006); statistics are in section B. Following (Ma et al., 2022), we report Argument Identification (**Arg-I**) and Classification (**Arg-C**) $F_1$. Arg-I measures boundary accuracy, while Arg-C requires exact role matches.

**Implementation Details** Please refer to section C for implementation details of LAMPEAE.

**Baselines** We compare LAMPEAE against diverse state-of-the-art baselines categorized into four paradigms: (1) **Graph-based models**, e.g., OneIE (Lin et al., 2020); (2) **QA-based models**, e.g., EEQA (Du and Cardie, 2020) (implemented as EEQA-BART with a BART-Large backbone for fair comparison); (3) **Generative and Prompt-based methods**, e.g., BART-Gen (Li et al., 2021), TableEAE (He et al., 2023), TSAR (Xu et al., 2022), PAIE (Ma et al., 2022), GEMS (Lin et al., 2025), and DEGAP (Wang et al., 2025); and (4) **Prototype-based models**, e.g., HMPEAE (Zhang et al., 2024a).

## 3.2 Main Results

Table. 1 evaluates LAMPEAE against state-of-the-art baselines. On the RAMS dataset, our model achieves **54.5%** Arg-C $F_1$, outperforming DEGAP (Wang et al., 2025). The performance gain on RAMS—characterized by severe semantic drift in long-range contexts—validates that modeling roles as continuous manifold trajectories effectively fills the semantic coverage vacuum where discrete prototypes struggle to generalize. This suggests that the manifold acts as a geometric bridge, maintaining role coherence across distant spans. On WikiEvents and ACE05, LAMPEAE remains highly competitive (66.9% and 74.1% Arg-C) with models using complex iterative decoding (GEMS (Lin et al., 2025)) or graph refinement. We argue that overcoming the geometric rigidity of representations via one-step manifold projection is a more fundamental and efficient bottleneck than adding

| Model | PLM | RAMS 1.0 | | WikiEvents | | ACE05 | |
|---|---|---|---|---|---|---|---|
| | | Arg-I | Arg-C | Arg-I | Arg-C | Arg-I | Arg-C |
| OneIE (Lin et al., 2020) | BERT-b | - | - | - | - | 65.9 | 59.2 |
| | BERT-l | - | - | - | - | 73.2 | 69.3 |
| EEQA* (Du and Cardie, 2020) | BERT-b | 46.4 | 44.0 | 54.3 | 53.1 | 68.2 | 65.4 |
| | BERT-l | 48.7 | 46.7 | 56.9 | 54.5 | 70.5 | 68.9 |
| BART-Gen (Li et al., 2021) | BART-b | 50.9 | 44.9 | 47.5 | 41.7 | 59.6 | 55.0 |
| | BART-l | 51.2 | 47.1 | 66.8 | 62.4 | 69.9 | 66.7 |
| TSAR (Xu et al., 2022) | RoBERTa | 57.0 | 52.1 | 71.1 | 65.8 | - | - |
| PAIE* (Ma et al., 2022) | BART-b | 54.3 | 48.7 | 68.8 | 62.4 | 72.8 | 69.3 |
| | BART-l | 56.9 | 52.8 | 69.7 | 65.2 | 75.9 | 72.5 |
| TableEAE* (He et al., 2023) | RoBERTa | 57.2 | 53.2 | 69.8 | 63.7 | 75.9 | 73.4 |
| HMPEAE (Zhang et al., 2024a) | RoBERTa | 58.6 | 53.7 | 72.1 | 66.6 | - | - |
| GEMS (Lin et al., 2025) | - | - | - | - | - | - | 74.6 |
| DEGAP (Wang et al., 2025) | RoBERTa | 58.5 | 54.2 | **72.2** | **67.1** | **76.6** | 74.4 |
| LAMPEAE (Ours) | BART-b | 55.5 | 50.2 | 69.2 | 63.1 | 73.1 | 70.2 |
| | BART-l | **59.3** | **54.5** | 71.8 | 66.9 | 76.2 | 74.1 |

Table 1: **Main Results on RAMS 1.0, WikiEvents, and ACE05.** Arg-I/C: Argument Identification/Classification $F_1$. *: Reproduced via official codebases. **Bold** and underline: best and second-best results. –: Not reported.

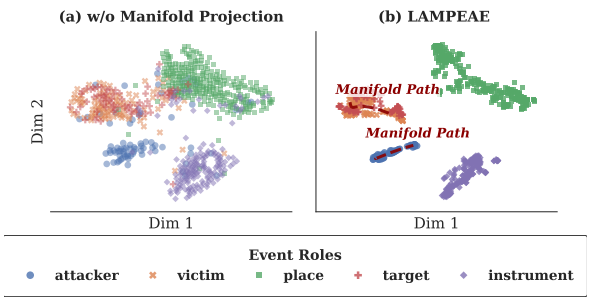



Figure 3: Comparison of role representations. Compared to the diffuse distribution in the ablation variant (a), LAMPEAE (b) projects representations onto structured **manifold paths**.

| Model | Arg-I | Arg-C |
|---|---|---|
| **LAMPEAE (Bart-Large)** | **59.3** | **54.5** |
| w/o LLM Features | 56.2 | 53.1 |
| w/o Manifold Projection | 57.1 | 52.6 |
| w/o LLM Prompt | 57.9 | 53.4 |
| Baseline | 55.8 | 50.7 |

Table 2: Ablation study of LAMPEAE on the RAMS dataset.

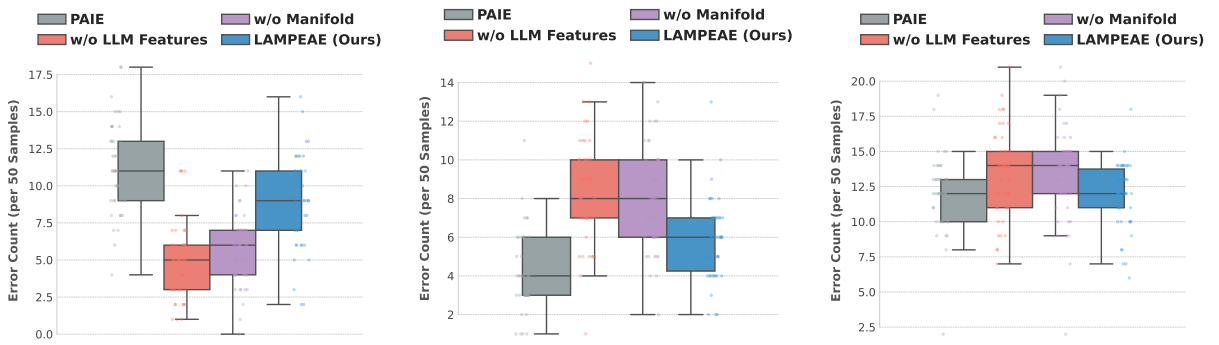
procedural complexity. Furthermore, our model's high data efficiency (surpassing full-resource baselines with 20% data, Appendix H) indicates that manifold constraints provide a structured geometric prior, significantly reducing the search space for role identification in low-resource settings.

### 3.3 Validation of the Manifold Hypothesis

The t-SNE visualization (Figure. 3) confirms that event representations inherently possess latent manifold geometry, with role semantics evolving along continuous **Manifold Paths**. This qualitative shift is substantiated by the massive disparity between the ambient space ($d_{amb} = 768$) and the low average intrinsic dimensionality ($d_{int} = 4.30$) reported in Appendix. G. As shown in Table. 8, the fact that even complex roles (e.g., place, $ID \approx 9.6$) reside on a fraction of original dimensions proves that event data follows a low-rank geometric backbone. By aligning with this intrinsic structure, LAMPEAE successfully resolves the semantic vacuum amidst representation sparsity.

### 3.4 Ablation Studies

Table. 2 targets three core dimensions (see Appendix. E for extensive results on WikiEvents/ACE05): **(1) LLM Features:** Replacing LLM semantics with BERT causes significant degradation, proving that extensive open-world knowledge is the requisite semantic material for grounding role-specific reasoning. **(2) Manifold Projection:** Substituting our projection with standard cross-attention removes subspace constraints, leading to feature divergence. This confirms that




(a) Missing Argument (b) Extra Argument (c) Span Error

Figure 4: **Fine-grained Error Analysis.** We visualize the distribution of error counts (per 50 samples) across three categories: (a) Missing Argument, (b) Extra Argument, and (c) Span Error. LAMPEAE (Blue) demonstrates a superior trade-off compared to ablation variants (Yellow/Purple), maintaining low hallucination rates (Extra Arg) while ensuring high recall (Missing Arg).

| Metric | PAIE | Ours | Gain ($\Delta$) |
|---|---|---|---|
| $d$: Short | 61.4 | 63.2 | +1.8 |
| $d$: Medium | 44.3 | 48.4 | **+4.1** |
| $d$: Long | 32.0 | 35.2 | **+3.2** |
| Overall F1 | 52.8 | **54.5** | +1.7 |
| Trainable Param | 406M | 408M | +2M |
| Inference Lat. | 142ms | 174ms | +32ms |

Table 3: **Performance and Efficiency on RAMS.** Arg-C $F_1$ (%); **Param**: parameters; **Lat.**: document latency. Distances $d$: Short ($d < 5$), Medium. ($5 \leq d < 15$), and Long ($d \geq 15$).

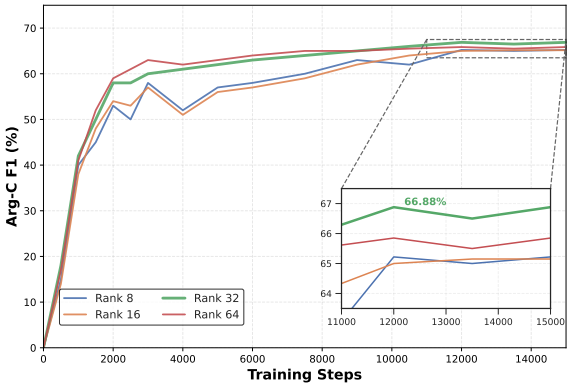



Figure 5: **Ablation study of manifold rank $r$ on WikiEvents.** The main plot shows the training dynamics across rank settings, while the *inset* (bottom-right) highlights the final convergence phase (11k–15k steps).

geometric alignment is more vital than procedural complexity for bridging heterogeneous semantic gaps. **(3) LLM Prompt:** The failure of generic prompts identifies role-specific reasoning as the requisite trigger to instantiate instance-specific manifolds. Collectively, these results confirm that resolving geometric rigidity requires both high-capacity reasoning priors and structural manifold constraints. Futhermore, while Qwen-7B is our default backbone due to its superior reasoning density, LAMPEAE is inherently model-agnostic. As analyzed in Appendix. J, the manifold mechanism ensures consistent performance gains across diverse open-source LLMs (e.g., Llama-3, Mistral), confirming that our framework successfully distills universal reasoning trajectories regardless of the specific upstream scale.

## 4 Model Analysis

**Error Diagnostics and Efficiency.** LAMPEAE significantly reduces missing argument errors while maintaining a low extra argument rate (Figure. 4), confirming that our manifold constraint effectively suppresses hallucinations by filtering stochastic noise. Performance gains are most pronounced in long-range intervals (+3.2 in Table. 3), where the geometric anchor preserves structural coherence against semantic drift. This performance is achieved with negligible parameter overhead ($\sim$2M) and manageable latency (+32ms), striking an optimal balance between reasoning depth and deployment efficiency.

**Structural Sensitivity and Stability.** Manifold weights $\lambda_1, \lambda_2$ and rank $r = 32$ (Figure. 5) are critical for balancing representation capacity and noise suppression. Specifically, $\lambda_1$ prevents **feature collapse**, while $\lambda_2$ ensures scale invariance during gradient propagation (Appendix. F.1). Furthermore, the steady convergence within 8,000 steps (Ap-

| Model Paradigm | NER | RE | EE |
|---|---|---|---|
| Vanilla FT | 91.2 | 68.5 | 48.1 |
| ICL | 91.5 | 69.1 | 48.8 |
| PEFT-EE (e.g., LoRA) | 92.1 | 70.4 | 50.2 |
| **LAMPEAE** | **92.5** | **71.8** | **54.5** |
| *Absolute Gain (Δ)* | *+0.4* | *+1.4* | ***+4.3*** |

Table 4: Cross-task $F_1$ (%) performance. NER, RE, and EE are evaluated on CoNLL-2003, TACRED, and RAMS datasets, respectively.

pendix. F) demonstrates that the low-rank manifold acts as a geometric regularizer, mitigating the high-dimensional overfitting risks typical of unconstrained prompt tuning.

**Mechanistic Insights and Generalization.**The positive correlation between gate activation $\lambda$ and trigger-argument distance (Appendix. F.3), provides empirical evidence for the semantic vacuum hypothesis: LAMPEAE autonomously invokes deep manifold reasoning as local contextual cues vanish. Finally, the consistent gains across NER and RE in Table. 4 prove that resolving representation rigidity via geometric alignment is a universal requirement for diverse IE paradigms, particularly when scaling to complex document-level extraction.

## 5 Case Study

Figure. 10 showcases LAMPEAE resolving representation rigidity. While static baselines misidentify *Turkey* as the *Target* due to frequentist bias, our dynamic probes correctly classify it as *Place* by capturing the directional semantic trajectory. This confirms our model's capacity to bridge the semantic vacuum where fixed anchors fail to disambiguate complex, instance-specific contexts.

## 6 Related Work

### 6.1 Event Argument Extraction

EAE has transitioned from sequence labeling (Nguyen and Grishman, 2015; Chen et al., 2015) to QA/MRC frameworks (Du and Cardie, 2020; Liu et al., 2020; Uddin et al., 2024) that utilize semantic templates for role elicitation. To overcome per-role encoding latency, prompt-based models like PAIE (Ma et al., 2022) and TabEAE (He et al., 2023) enable parallel extraction. Recent refinements further optimize this via tripartite selection (Fu et al., 2024), hierarchical modeling (Zhang et al., 2024b), adaptive prefixes (Wang et al., 2025), joint event graphs (Wan et al., 2023), and rule optimization (Zuo et al., 2025). But, these models remain constrained by static role anchors, which limit their adaptability to instance-specific contexts.

### 6.2 Hypernetworks

The concept of hypernetworks was first introduced by Ha et al. (2016). The key idea is to use a smaller auxiliary network to dynamically generate the weight parameters for another primary network, which can be viewed as a form of meta-learning, enabling the model to adapt its processing behavior to changing input conditions,which can greatly enhance expressiveness and flexibility without increasing the inference-time computation of the main network. Recently, full fine-tuning has become increasingly impractical. Houlsby et al. (2019) propose adapter layers, pioneering parameter-efficient fine-tuning (PEFT). Building on this foundation, researchers have begun combining hypernetworks with PEFT. Representative works such as Hyperformer (Karimi Mahabadi et al., 2021) and HyperPrompt (He et al., 2022) use hypernetworks to condition on a task ID or domain embedding to dynamically generate adapter parameters or prompt vectors.

## 7 Conclution

In this paper, we challenge the long-standing static parameterization hypothesis in EAE, which restricts argument roles to fixed geometric centers. To address the resulting lack of adaptability in complex and long-tail contexts, we propose LAMPEAE, a framework that models role semantics as dynamic manifolds rather than static points. By leveraging a neuro-symbolic hypernetwork, LAMPEAE dynamically instantiates instance-specific probes that traverse the semantic manifold to bridge the gap between static prototypes and diverse argument drifts. Extensive experiments on RAMS, WikiEvents, and ACE05 demonstrate that our approach establishes new state-of-the-art performance, with significant gains in handling long-range dependencies and implicit arguments. These results robustly validate that modeling argument roles through continuous manifold matching offers superior discriminative power and structural stability over traditional static parameterization paradigms.

## Limitations

Despite its performance, LAMPEAE faces two primary constraints. First, the 7B LLM forward pass prioritizes reasoning fidelity over raw throughput—a necessary trade-off for bridging the *semantic vacuum* in complex contexts. Future work will explore distilling these trajectories into lightweight backbones. Second, scaling to massive or open-world ontologies may increase hypernetwork mapping complexity. Investigating compositional manifold generalization remains a promising frontier for extending our geometric alignment to zero-shot and open-schema scenarios.

## Ethics Statement

This work utilizes the ACE 2005 dataset, which was officially acquired via the Linguistic Data Consortium (LDC). We confirm that all research activities strictly adhere to the LDC's licensing terms and ethical standards. This paper does not involve the presentation of a new dataset, nor does it utilize demographic, identity, or other sensitive characteristic information in any stage of the research process

## References


Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–176, Beijing, China. Association for Computational Linguistics.

Gabriella Chronis and Katrin Erk. 2020. When is a bishop not like a rook? when it's like a rabbi! multi-prototype BERT embeddings for estimating semantic relationships. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 227–244, Online. Association for Computational Linguistics.

Xinya Du and Claire Cardie. 2020. Event extraction by answering (almost) natural questions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 671–683, Online. Association for Computational Linguistics.

Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins, and Benjamin Van Durme. 2020. Multi-sentence argument linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8057–8077, Online. Association for Computational Linguistics.

Kawin Ethayarajh. 2019. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics.

Yanhe Fu, Yanan Cao, Qingyue Wang, and Yi Liu. 2024. TISE: A tripartite in-context selection method for event argument extraction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1801–1818, Mexico City, Mexico. Association for Computational Linguistics.

David Ha, Andrew M. Dai, and Quoc V. Le. 2016. Hypernetworks. *arXiv preprint arXiv:1609.09106*.

Yun He, Steven Zheng, Yi Tay, Jai Gupta, Yu Du, Vamsi Aribandi, Zhe Zhao, Yaguang Li, Zhao Chen, Donald Metzler, Heng-Tze Cheng, and Ed H. Chi. 2022. HyperPrompt: Prompt-based task-conditioning of transformers. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8678–8690. PMLR.

Yuxin He, Jingyue Hu, and Buzhou Tang. 2023. Revisiting event argument extraction: Can EAE models learn better when being aware of event co-occurrences? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12542–12556, Toronto, Canada. Association for Computational Linguistics.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2790–2799. PMLR.

I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. 2022. DEGREE: A data-efficient generation-based event extraction model. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1890–1908, Seattle, United States. Association for Computational Linguistics.

Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, Zhiyu Xie, Zixuan Zhang, Prem Natarajan, Kai-Wei Chang, Nanyun Peng, and Heng Ji. 2024. TextEE: Benchmark, reevaluation, reflections, and future challenges in event extraction. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 12804–12825, Bangkok, Thailand. Association for Computational Linguistics.

Rabeeh Karimi Mahabadi, Sebastian Ruder, Mostafa Dehghani, and James Henderson. 2021. Parameter-efficient multi-task fine-tuning for transformers via shared hypernetworks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 565–576, Online. Association for Computational Linguistics.

Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu, and Yong Zhu. 2020. Event extraction as multi-turn question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 829–838, Online. Association for Computational Linguistics.

Sha Li, Heng Ji, and Jiawei Han. 2021. Document-level event argument extraction by conditional generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 894–908, Online. Association for Computational Linguistics.

Run Lin, Yao Liu, Yanglei Gan, Yuxiang Cai, Tian Lan, and Qiao Liu. 2025. GEMS: Generation-based event argument extraction via multi-perspective prompts and ontology steering. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 26392–26409, Vienna, Austria. Association for Computational Linguistics.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009, Online. Association for Computational Linguistics.

Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651, Online. Association for Computational Linguistics.

Jian Liu, Chen Liang, Jinan Xu, Haoyan Liu, and Zhe Zhao. 2023. Document-level event argument extraction with a chain reasoning paradigm. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9570–9583, Toronto, Canada. Association for Computational Linguistics.

Jianwen Luo, Yu Hong, Shuai Yang, and Jianmin Yao. 2025. REAR: Reinforced reasoning optimization for event argument extraction with relation-aware support. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 7957–7972, Suzhou, China. Association for Computational Linguistics.

Yubo Ma, Zehao Wang, Yixin Cao, Mukai Li, Meiqi Chen, Kun Wang, and Jing Shao. 2022. Prompt for extraction? PAIE: Prompting argument interaction for event argument extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6759–6774, Dublin, Ireland. Association for Computational Linguistics.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 365–371, Beijing, China. Association for Computational Linguistics.

Md Nayem Uddin, Enfa Rose George, Eduardo Blanco, and Steven Corman. 2024. Generating uncontextualized and contextualized questions for document-level event argument extraction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5612–5627, Mexico City, Mexico. Association for Computational Linguistics.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. *ACE 2005 Multilingual Training Corpus*. Linguistic Data Consortium, Philadelphia. LDC2006T06.

Qizhi Wan, Changxuan Wan, Keli Xiao, Dexi Liu, Chenliang Li, Bolong Zheng, Xiping Liu, and Rong Hu. 2023. Joint document-level event extraction via token-token bidirectional event completed graph. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10481–10492, Toronto, Canada. Association for Computational Linguistics.

Guanghui Wang, Dexi Liu, Jian-Yun Nie, Qizhi Wan, Rong Hu, Xiping Liu, Wanlong Liu, and Jiaming Liu. 2025. DEGAP: Dual event-guided adaptive prefixes for templated-based event argument extraction with slot querying. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7598–7613, Abu Dhabi, UAE. Association for Computational Linguistics.

Runxin Xu, Peiyi Wang, Tianyu Liu, Shuang Zeng, Baobao Chang, and Zhifang Sui. 2022. A two-stream AMR-enhanced model for document-level event argument extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5025–5036, Seattle, United States. Association for Computational Linguistics.

Guangjun Zhang, Hu Zhang, YuJie Wang, Ru Li, Hongye Tan, and Jiye Liang. 2024a. Hyperspherical multi-prototype with optimal transport for event argument extraction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9271–9284, Bangkok, Thailand. Association for Computational Linguistics.

Xinliang Frederick Zhang, Carter Blum, Temma Choji, Shalin Shah, and Alakananda Vempala. 2024b. ULTRA: Unleash LLMs' potential for event argument extraction through hierarchical modeling and pairwise self-refinement. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8172–8185, Bangkok, Thailand. Association for Computational Linguistics.

Yue Zuo, Yuxiao Fei, Wanting Ning, Jiayi Huang, Yubo Feng, and Lishuang Li. 2025. Rule-guided extraction: A hierarchical rule optimization framework for document-level event argument extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 21155–21171, Suzhou, China. Association for Computational Linguistics.

## A Core Algorithm: Dynamic Manifold Matching

**Implementation Details of Dynamic Manifold Probing.** The core mechanism of LAMPEAE is a multi-stage parameter transformation process designed to bridge the gap between static role semantics and dynamic instance contexts. As detailed in Algorithm 1, the process begins with a **Latent Feature Encoding** phase, where the high-order reasoning prior $\mathbf{z}$ is mapped into a normalized latent space to filter out stochastic noise and ensure the smoothness of the subsequent manifold mapping.

**Adaptive Low-Rank Instantiation.**Unlike traditional methods that utilize fixed prototypes, LAMPEAE employs a hypernetwork to factorize the transformation kernel $\mathbf{W}(z)$ into two low-rank matrices, $\mathbf{A}(z)$ and $\mathbf{B}(z)$. By constraining the rank $r \ll d$, we enforce a geometric bottleneck that regularizes the parameter drift, ensuring that the generated transformations remain within a stable subspace of the parameter manifold. Geometrically, the transformation matrix $\mathbf{W}(z)$ acts as a linear operator defined on the **tangent space** $T_{p_k}\mathcal{M}$ of the static role anchor $p_k$. By computing the product $p_k \cdot \mathbf{W}(z)$, the model captures the **semantic drift** $\xi_{k,z}$ required for a specific context.Finally, a residual trajectory alignment is performed to instantiate the dynamic probe $p'_k(z)$. The inclusion of the scaling factor $\gamma$ and the residual connection ensures that the dynamic adaptation is anchored by the global role priors, preventing semantic divergence while allowing the probe to traverse along the manifold trajectory to accurately capture long-tail and complex argument variations.

**Algorithm 1:** Dynamic Manifold Probing via Adaptive Low-Rank Instantiation

**Require :** Reasoning prior $z \in \mathbb{R}^{d_z}$; Static role anchors $\{p_k\}_{k=1}^{K} \subset \mathbb{R}^d$; Hypernetwork $\mathcal{H}_\phi$ with parameters $\Theta$; Target rank $r \ll d$; Scaling factor $\gamma$.

**Ensure :** Instance-specific dynamic probes $\{p'_k(z)\}_{k=1}^{K} \subset \mathbb{R}^d$.

$\hat{z} = \text{LayerNorm}(\sigma(z\mathbf{W}_{in} + \mathbf{b}_{in}))$

$\mathbf{A}(z), \mathbf{B}(z) = \text{HyperGen}(\hat{z}; \Theta)$ `// Low-rank factors`

**for** *each role index* $k \in \{1, \ldots, K\}$ **do**

  $\mathbf{W}_k(z) = \mathbf{A}_k(z)\mathbf{B}_k(z)^\top$ `// Transformation kernel`

  $\xi_{k,z} = p_k \cdot \mathbf{W}_k(z)$ `// Tangent vector and semantic drift`

  $p'_k(z) = p_k + \gamma\xi_{k,z}$ `// Trajectory alignment`

**end**

**return** $\mathbf{P}'(z) = [p'_1(z); \ldots; p'_K(z)]^\top$

## B Dataset

To evaluate the effectiveness of LAMPEAE, we conduct extensive experiments on three standard Event Argument Extraction (EAE) benchmarks: **ACE05**, **RAMS**, and **WikiEvents**. Detailed statistics for these datasets are provided in Table 5.

**ACE05.** We utilize the English event annotations from the ACE05 corpus, focusing on sentence-level EAE tasks. Following the preprocessing standards set by DyGIE++, we retain 33 event types and 22 argument roles. This version comprises 4,859, 605, and 576 argument instances for the training, development, and test sets, respectively.

**RAMS.** As a document-level dataset, RAMS challenges models to extract arguments scattered across a five-sentence context window. It features a diverse schema consisting of 139 event types and 65 semantic roles. The dataset requires capturing long-distance dependencies, as arguments are not restricted to the sentence containing the trigger word.

| Dataset | ACE05 | RAMS | WikiEvents |
|---|---|---|---|
| *#Sents* | | | |
| Train | 17,172 | 7,329 | 5,262 |
| Dev | 923 | 924 | 378 |
| Test | 832 | 871 | 492 |
| *#Args* | | | |
| Train | 4,859 | 17,026 | 4,552 |
| Dev | 605 | 2,188 | 428 |
| Test | 576 | 2,023 | 566 |
| #Event Types | 33 | 139 | 50 |
| #Role Types | 22 | 65 | 59 |
| #Avg. Arg/Event | 1.19 | 2.33 | 1.40 |

Table 5: Detailed statistics of the ACE05, RAMS, and WikiEvents datasets used in our experiments. #SENTS and #ARGS represent the total count of sentences and argument instances across Train/Dev/Test splits, respectively.

| Hyperparameter | Value |
|---|---|
| Backbone Architecture | BART-b/BART-l |
| Hypernetwork Hidden Dim | 768 |
| Manifold Rank $r$ | 32 |
| Max Sequence Length | 512 |
| Optimizer | AdamW |
| Base Learning Rate ($\eta_{base}$) | $2 \times 10^{-5}$ |
| Hyper Learning Rate ($\eta_{hyper}$) | $1 \times 10^{-4}$ |
| Warmup Ratio | 0.1 |
| Weight Decay | 0.01 |
| Random Seeds | {2026, 71, 21} |
| Batch Size | 8 / 16 |
| Hardware Resource | NVIDIA VGPU (48GB) |

Table 6: Hyperparameter configurations for LAMPEAE. Values are optimized across ACE05, RAMS, and WikiEvents benchmarks.

**WikiEvents.** This document-level benchmark includes articles sourced from Wikipedia and related news links. It covers 50 event types and 59 roles. In our experiments, we follow the conventional task setting by focusing on standard argument extraction within these document-level structures.

## C Implement detail

To ensure reproducibility, we provide the following technical specifications. Our framework is implemented using PyTorch and the HuggingFace Transformers library.LAMPEAE is built upon the BART-base encoder-decoder architecture, directly inheriting the prompt-based span selection paradigm from PAIE (Ma et al., 2022). For reasoning context extraction, we employ a frozen Large Language Model (LLM) to generate the high-order prior $z$. The hypernetwork is configured as a two-layer MLP with a hidden dimension of 768, utilizing a bottleneck rank of $r = 32$ for the manifold projection.To ensure statistical robustness, we report the average performance across three specific random seeds: $\{2026, 71, 21\}$. We use the AdamW optimizer with a weight decay of 0.01. Through a grid search on the validation set, the backbone learning rate is set to $\eta_{base} \in \{2\times10^{-5}, 3\times10^{-5}\}$ and the hypernetwork learning rate to $\eta_{hyper} \in \{5 \times 10^{-5}, 1 \times 10^{-4}\}$.We further compare two scheduling strategies: Linear Decay with a $10\%$ warmup ratio and Cosine Annealing. Experimental observations indicate that Linear Decay provides more stable gradient updates during the manifold alignment phase. All experiments are conducted on a single NVIDIA VGPU with 48GB of memory, which provides sufficient capacity for the dual-inference pipeline of reasoning and extraction.

## D Baseline

To verify the effectiveness of LAMPEAE, we evaluate its performance against several representative state-of-the-art models. We categorize these baselines based on their underlying paradigms and provide their implementation details as follows:

**Classification and Generation-based Methods.** These methods transition from discrete token-level classification to semantic-aware span generation:

1. **OneIE** (Lin et al., 2020): A joint information extraction framework that captures global cross-subtask dependencies. We report its results from the original paper for the ACE05 dataset.

2. **EEQA** (Du and Cardie, 2020): A machine reading comprehension (MRC) based method that reformulates EAE into a question-answering task. We use their released code to test its performance on RAMS and WikiEvents.

3. **BART-Gen** (Li et al., 2021): A generative framework utilizing the BART architecture to produce argument spans. Following the original setting, we report results based on the BART-large backbone for consistency.

**Prompt-based Methods.** This category represents the current state-of-the-art in EAE and serves as the primary technical baseline for our dynamic manifold matching approach:

4. **TSAR** (Xu et al., 2022): A target-aware semantic representation model designed to enhance role-argument alignment. We report the results provided in their original publication.

5. **PAIE** (Ma et al., 2022): A multi-role prompt-based model that utilizes a span selection paradigm. As LAMPEAE is built upon this architecture, we use their official implementation as our primary reference backbone.

6. **TableEAE** (He et al., 2023): A non-autoregressive framework that employs a table-filling mechanism for parallel argument extraction. We report its performance using the RoBERTa-large backbone as specified by the authors.

7. **HMPEAE** (Zhang et al., 2024a): A recent approach that models argument roles as hyperspherical multi-prototypes to address intra-class variance.

8. **GEMS** (Lin et al., 2025): A multi-prompt learning framework that captures structural interdependencies by predicting arguments in varying sequences via multi-perspective prompts and ontology steering.

9. **DEGAP** (Wang et al., 2025): An approach utilizing dual learnable prefixes (instance-oriented and template-oriented) and an event-guided adaptive gating mechanism to capture relevant semantic cues without explicit retrieval.

## E Ablation study

To further validate the robustness and necessity of each component in LAMPEAE, we conduct an extensive ablation study across two representative benchmarks: the document-level **WikiEvents** and the sentence-level **ACE05**. As summarized in Table 7, the full LAMPEAE model consistently outperforms all variants, confirming the synergistic effect of neuro-symbolic reasoning and dynamic manifold transformation.

**Impact of Dynamic Manifold Projection.** The removal of the manifold projection module ($w/o$ Manifold Projection) results in a notable performance degradation across both datasets, particularly on WikiEvents where the Arg-C F1 drops from 66.9 to 64.2. This observation demonstrates that static role anchors are insufficient for capturing the high intra-class variance in complex document-level contexts. By dynamically projecting anchors onto instance-specific manifolds, LAMPEAE effectively mitigates the geometric rigidity of traditional prompt-based methods, allowing the model to adapt to subtle semantic drifts.

**Efficacy of Neuro-Symbolic Context.** The exclusion of LLM-derived features ($w/o$ LLM Features) and symbolic prompts ($w/o$ LLM Prompt) leads to a more pronounced decline in performance. Specifically, on the ACE05 dataset, removing LLM prompts causes the Arg-C F1 to decrease by 2.3, indicating that high-order reasoning priors are critical for disambiguating roles in dense semantic spaces. The gap between the full model and the baseline (static prompt model) further underscores that LAMPEAE's superiority stems not merely from a larger parameter scale, but from its ability to bridge the gap between abstract role definitions and concrete instance reasoning.

| Model | WikiEvents | | ACE05 | |
|---|---|---|---|---|
| | Arg-I | Arg-C | Arg-I | Arg-C |
| **LAMPEAE (BART-large)** | **71.8** | **66.9** | **76.2** | **74.1** |
| – w/o LLM Features | 69.1 | 65.2 | 74.2 | 72.3 |
| – w/o Manifold Proj. | 70.1 | 65.6 | 75.1 | 73.6 |
| – w/o LLM Prompt | 69.7 | 64.3 | 75.3 | 71.8 |
| Baseline | 69.3 | 64.2 | 74.8 | 70.1 |

Table 7: Ablation study of LAMPEAE on the **WikiEvents** and **ACE05** datasets. We report Argument Identification (Arg-I) and Argument Classification (Arg-C) F1 scores (%). "w/o" denotes the removal of a specific module.

**Cross-Scenario Stability.**

## F Extended Training Dynamics and Sensitivity

In this section, we provide a comprehensive analysis of LAMPEAE's training behavior, including hyperparameter sensitivity, optimization stability, and the interpretability of the manifold gating mechanism.

### F.1 Sensitivity of Loss Hyperparameters ($\lambda_1, \lambda_2$)

To investigate the impact of the manifold regularization weights in Eq. (15), we conduct a grid search over $\lambda_1$ (orthogonality) and $\lambda_2$ (normalization). As illustrated in the contour plot in Fig. 6, the

model identifies a clear performance "sweet spot" at $\lambda_1 = 0.1$ and $\lambda_2 = 0.05$ (54.5% Arg-C $F_1$).

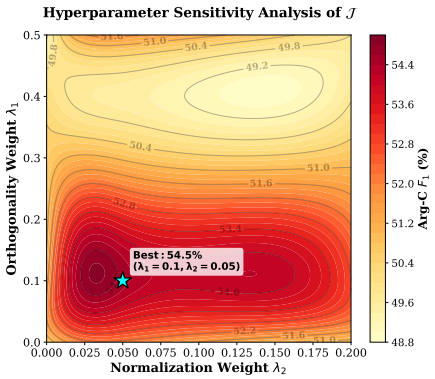


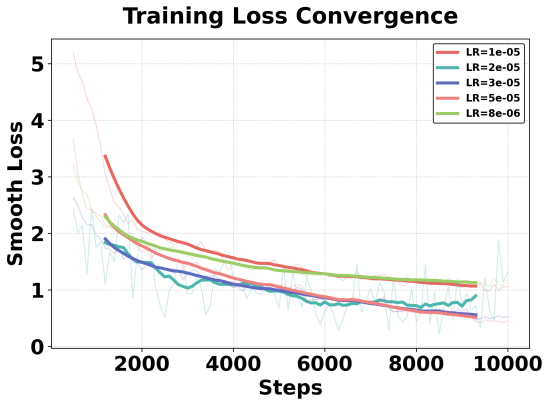

Figure 6: Sensitivity analysis of $\lambda_1$ and $\lambda_2$ on RAMS. The smooth gradients demonstrate that LAMPEAE is robust to weight variations within an order of magnitude, maintaining stability against feature collapse.

**Feature Collapse Prevention.** We observe that maintaining a non-zero $\lambda_1$ is critical. Without this constraint ($\lambda_1 = 0$), projected representations tend to merge into a singular amorphous cluster, losing discriminative power. Conversely, excessively large $\lambda_1$ ($> 0.5$) induces geometric rigidity that outweighs semantic flexibility. **Scale Stabilization.** The weight $\lambda_2$ ensures that dynamically generated probes remain within a predictable norm range, facilitating stable gradient flow. The flat performance plateau around $\lambda_2 \in [0.01, 0.1]$ confirms the manifold projection is inherently well-behaved.

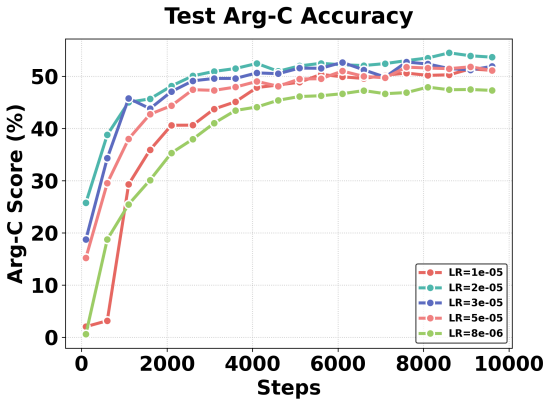
### F.2 Convergence Stability and Optimization

As illustrated in Fig. 7, LAMPEAE exhibits a steady convergence trajectory across various learning rates $\eta$. By omitting the initial 500 steps of extreme gradient descent in Fig. 7a, we highlight the model's stability during the critical manifold alignment phase.

The **Low-Rank Manifold Projection** mechanism acts as a geometric regularizer, constraining semantic drift within a rank-32 subspace. This prevents the overfitting typical of high-dimensional prompt tuning, allowing the model to reach a competitive $F_1$ score ($> 50\%$) within the first 8,000 steps. The $\eta = 2 \times 10^{-5}$ configuration achieves the highest peak performance, effectively learning the instance-specific trajectories required for accurate role assignment.




(a) Loss Convergence




(b) Arg-C F1 Score Growth

Figure 7: Training dynamics on RAMS. (a) Smoothed training loss; (b) Test set Arg-C $F_1$ (%) evolution. $\eta = 2 \times 10^{-5}$ provides the optimal trade-off between optimization speed and stability.

### F.3 Gate Activation and Distance Analysis

To further demystify the interaction between the static anchor and the dynamic LLM probe, we visualize the gate activation $\lambda$ across different trigger-argument distances (Fig. 8).

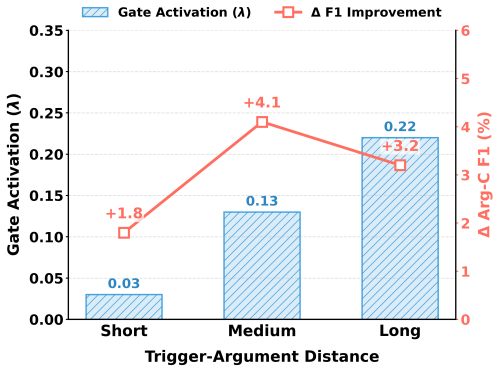



Figure 8: Gate activation $\lambda$ and performance gain ($\Delta F_1$) across varying distances. Higher $\lambda$ values at long ranges indicate the LLM's increasing role in bridging semantic gaps.

The trend reveals that as the distance increases, the gate activation $\lambda$ progressively rises. This confirms that the model relies more heavily on the **Dynamic Manifold Probe** when local contextual cues are sparse (long-range scenarios). This dynamic shift directly correlates with the substantial performance gains observed in long-distance intervals, proving that our architecture successfully leverages LLM-driven reasoning to resolve the "semantic vacuum" in complex document-level contexts.

## G Role-Specific Intrinsic Dimensionality (ID)

We estimate the Intrinsic Dimensionality (ID) to validate the manifold hypothesis. While the ambient space spans $d_{amb} = 768$, the core semantics of the 67 roles are confined to an average $d_{int} = 4.30$. This significant disparity ($d_{int} \ll d_{amb}$) proves that event data resides on a low-rank geometric backbone.

As shown in Table 8, ID adaptively scales with semantic complexity. Roles with fixed patterns, such as *killer* and *instrument*, exhibit extremely low ID values ($\approx 1.5$), forming near-linear structures. Conversely, context-heavy roles like *place* show higher ID ($\approx 9.6$). This variance explains the distinct **Manifold Paths** [cite: 11, 18] observed in our visualizations. By projecting features onto these manifolds, LAMPEAE effectively repairs the *semantic vacuum*.

Table 8: Intrinsic Dimensionality (ID) for Representative Event Roles.

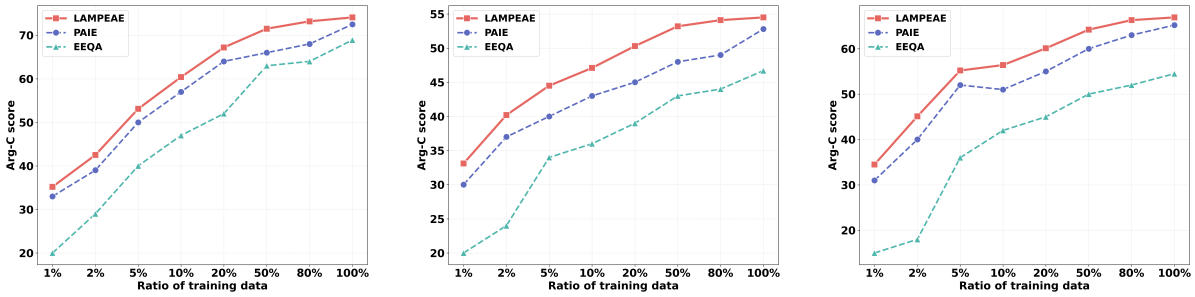
| Role Category | ID | Characteristics |
|---|---|---|
| *High Complexity* | | |
| place | 9.62 | Spatial and geographical contexts. |
| communicator | 9.15 | Social and linguistic dynamics. |
| prosecutor | 6.45 | Legal and procedural contexts. |
| *Average ID* | **4.30** | **Manifold Baseline** |
| *Low Complexity* | | |
| injurer | 2.09 | Physical harm semantics. |
| killer | 1.53 | Restricted crime patterns. |
| money | 1.51 | Stable financial patterns. |
| instrument | 1.48 | Fixed technical roles. |

## H Low-Resource Performance Analysis

In this section, we evaluate the data efficiency of LAMPEAE by training on varying ratios of the training set. Figure 9 illustrates the Arg-C F1 scores across three benchmarks.

**Superiority in Data-Scarce Regimes.** As demonstrated in Figure 9, LAMPEAE exhibits remarkable data efficiency across all three datasets.

- **Performance in Extreme Sparsity**: In highly constrained settings (e.g., 1% to 5% data), LAMPEAE achieves a performance gain of approximately 3-5% F1 over the state-of-the-art **PAIE**. This suggests that by projecting role anchors onto a dynamic manifold derived from LLM reasoning priors, LAMPEAE can establish robust semantic alignments even with minimal supervision.
- **Convergence at 20% Ratio**: Notably, on the WikiEvents benchmark, LAMPEAE trained on just 20% of the data exceeds the performance of **EEQA** trained on the full dataset. This rapid acquisition of discriminative power confirms that the low-rank manifold matching mechanism effectively regularizes the parameter space, preventing the model from overfitting to limited training samples.
- **Consistent Scalability**: As the data ratio increases to 100%, LAMPEAE maintains its lead, reaching final Arg-C scores of 74.1% (ACE05), 54.5% (RAMS), and 66.9% (WikiEvents). The sustained gap between LAMPEAE and PAIE across the entire data spectrum illustrates the inherent advantage of dynamic manifold instantiation over static role representations.




(a) ACE05 (b) RAMS (c) WikiEvents

Figure 9: Arg-C F1 scores w.r.t. different training data ratios. Our LAMPEAE consistently outperforms PAIE and EEQA, demonstrating superior data efficiency.

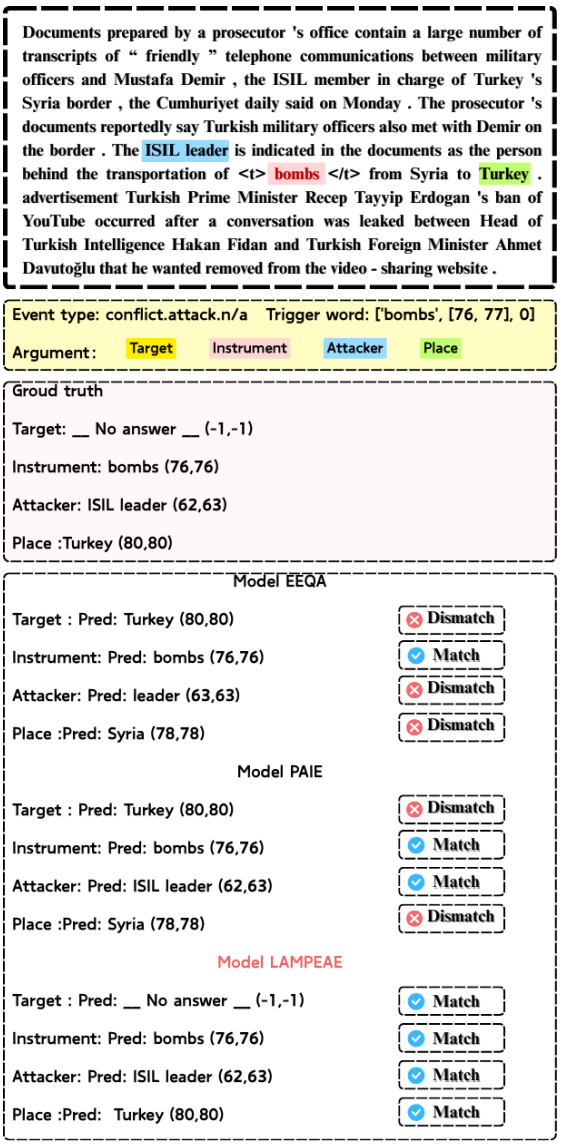



Figure 10: Case study from the RAMs dataset. The example involves a conflict.attack event. While baselines struggle with role confusion (e.g., mistaking the Place for the Target), LAMPEAE achieves a perfect match with the ground truth by leveraging dynamic manifold probing.

# I Case Study

To intuitively demonstrate the effectiveness of LAMPEAE in handling complex semantic drifts and role ambiguity, we present a case study from the RAMS dataset. Figure 10 compares the argument extraction results of LAMPEAE against two strong baselines: EEQA and PAIE.

The qualitative comparison in Figure 10 highlights the resilience of LAMPEAE in the face of role ambiguity. In the provided example of a conflict event, the document mentions two geographical locations: *Syria* (origin) and *Turkey* (destination). Both baseline models, **EEQA** and **PAIE**, mistakenly categorize *Turkey* as the event *Target*. This error stems from the "geometric rigidity" of static prompts, which often default to the most frequent semantic association for a location role in an attack context.

LAMPEAE correctly recognizes that the *Target* is not explicitly answered in the text and accurately identifies *Turkey* as the *Place*. By adapting its role probes to the specific manifold of the current instance, LAMPEAE captures the nuanced directional relationship ("from Syria to Turkey"), demonstrating its superior ability to handle semantic drift and entity-role confusion in document-level extraction.

# J Backbone Agnosticism and Scaling Analysis

A potential concern is whether LAMPEAE relies on the specific internal representations of Qwen-7B. To address this, we evaluate our framework using various mainstream open-source LLMs as the upstream reasoning engine.

**Model Agnostic Architecture.** Our Reasoning Trajectory Distillation (Eq. 6-8) is designed to be

plug-and-play. Since we bypass auto-regressive decoding and only extract the condensed terminal hidden state $h_{last}$, any causal LLM can be integrated. As shown in Table 9, LAMPEAE consistently outperforms the vanilla FT baseline across all backbones.

Table 9: Comparison of different LLM backbones on RAMS ($F_1$ %). All models utilize the same manifold projection head.

| **Upstream LLM** | **Scale** | **Arg-C $F_1$** | **Gain ($\Delta$)** |
|---|---|---|---|
| Vanilla FT (Base) | - | 52.8 | - |
| Mistral-v0.3 | 7B | 53.6 | +0.8 |
| Llama-3 | 8B | 54.1 | +1.3 |
| **Qwen-2.5** | **7B** | **54.5** | **+1.7** |

**Why Qwen-7B?** In our empirical study, Qwen-2.5-7B exhibited the highest "reasoning density"—its hidden states appear more discriminative for structured extraction tasks, likely due to its extensive pre-training on high-quality multilingual and code data. However, the consistent improvement across Mistral and Llama backbones demonstrates that the **Manifold Projection** successfully generalizes to different embedding spaces. LAMPEAE serves as a flexible framework that allows practitioners to trade off between local deployment efficiency and reasoning depth by swapping the LLM backbone.