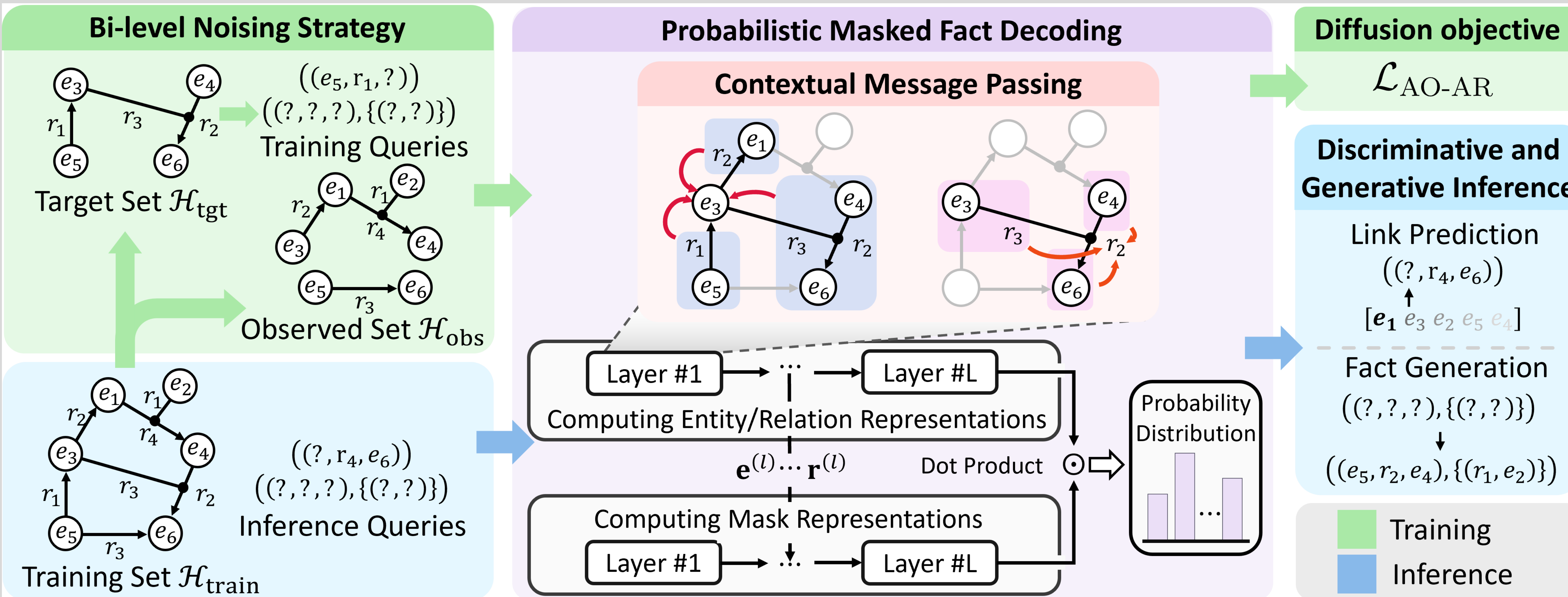


Generative Representation Learning on Hyper-relational Knowledge Graphs

via Masked Discrete Diffusion

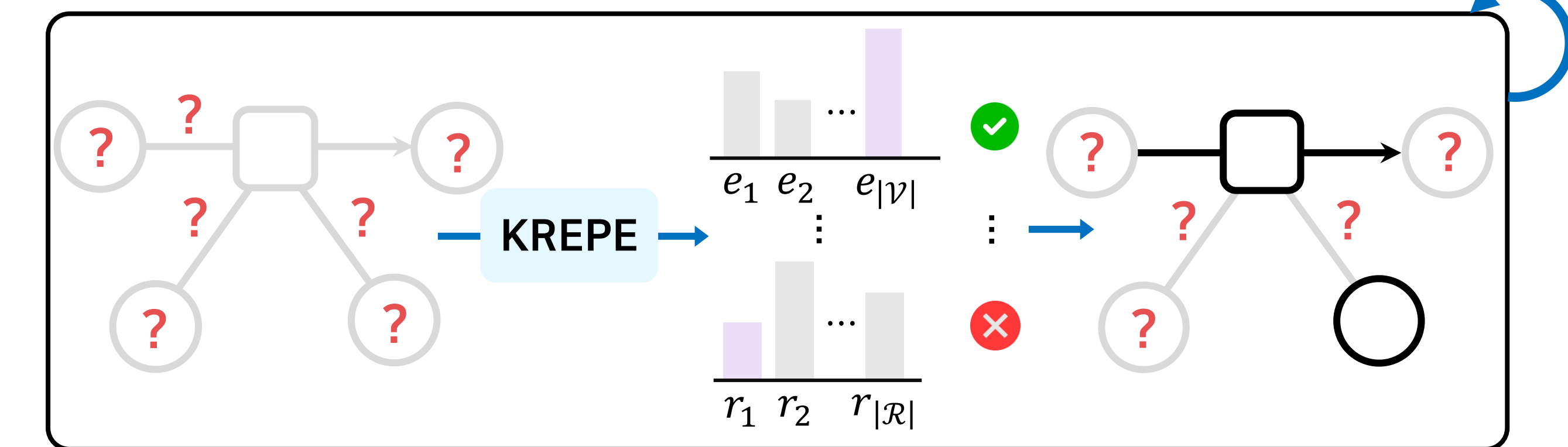
Jaejun Lee¹, Seheon Kim¹, and Joyce Jiyoung Whang^{2*}

¹School of Computing, KAIST, ²Department of AI Computing, KAIST
The 43rd International Conference on Machine Learning (ICML 2026)



Discriminative and Generative Inference

- Link Prediction**: Leverages the **computed probability distribution** to rank the candidates and identify the most likely missing component
- Fact Generation**: **Iteratively decodes** the masked components
 - At each step, probability distributions for all currently masked components are **recomputed using the current state of the query**



Generative Training via Masked Discrete Diffusion

- Bi-level Noising Strategy**
 - Simulates **diverse conditional generation scenarios** by varying both the observed structure and the query during training
- Diffusion objective**: Any-order autoregressive objective
 - Mathematically **equivalent to the masked discrete diffusion objective**
 - The sum of the cross-entropy losses for each masked component

Experiments – Link Prediction, Fact Generation

- Link Prediction: Compared with 25 link prediction methods

	MRR	WD50K	WikiPeople ⁻	WikiPeople
Entity Prediction	Best-baseline	0.411	0.521	0.488
	KREPE	0.419	0.522	0.491
Relation Prediction	Best-baseline	0.956	0.973	0.790
	KREPE	0.968	0.984	0.985

- Fact Generation: Compared with 7 competitive fact generation baselines

- Fact Generation Settings
 - Scratch**: generating a fact **without any known component**
 - Targeted**: generating a fact **containing a single given component**
 - Arbitrary Masking**: generating a fact from **an arbitrary mask**

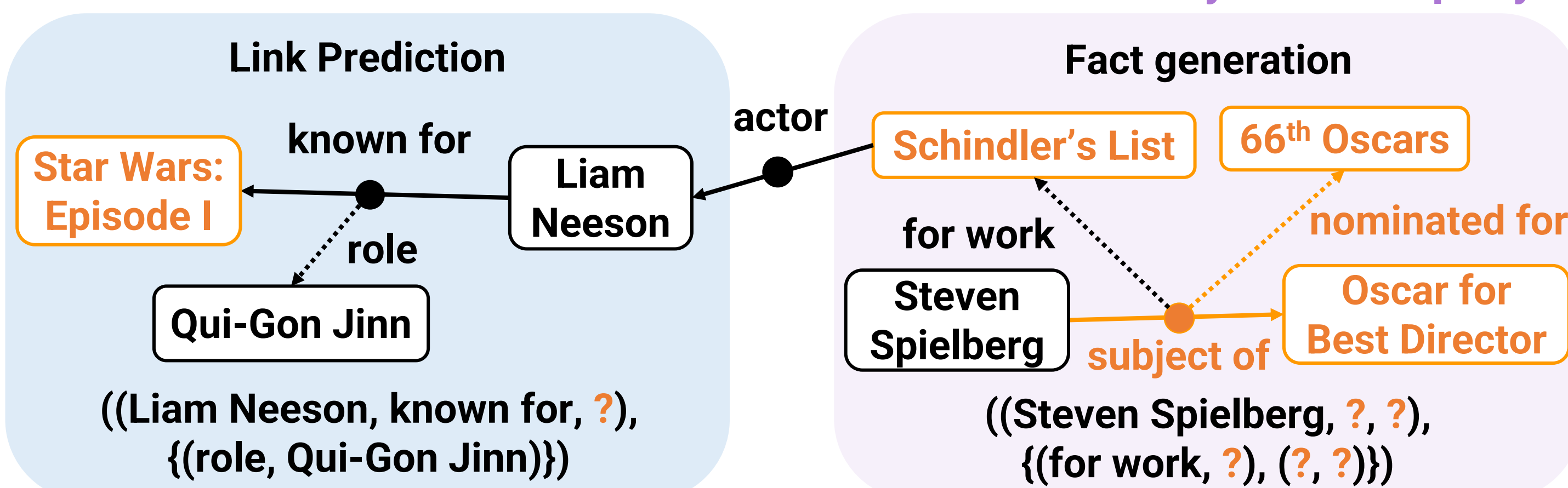
Dataset	Accuracy	Scratch	Targeted	Arbitrary Masking
WD50K	Best-baseline	0.351	0.482	0.604
	KREPE	0.717	0.508	0.604
WikiPeople ⁻	Best-baseline	0.369	0.394	0.497
	KREPE	0.855	0.600	0.649
WikiPeople	Best-baseline	0.326	0.396	0.469
	KREPE	0.777	0.591	0.621

Main Findings and Contributions

- Propose **Fact Generation**, expanding the scope of the hyper-relational knowledge graph completion task from passive ranking for single blanks to direct generation of missing facts.
- Propose **KREPE**, the first generative representation learning method for HKGs, **incorporating masked discrete diffusion** to model the joint conditional probability distributions of missing components in a query.
- KREPE** achieves state-of-the-art link prediction performance and outperforms both HKG and LLM-based baselines in fact generation

Hyper-relational Knowledge Graphs (HKGs)

- Hyper-relational Knowledge Graphs**
 - Add auxiliary details to triplets by adding qualifiers
- Link prediction**: Predict a **single missing element** in an incomplete fact
- Fact Generation**: Construct a valid fact from an **arbitrarily masked query**



Contextual Message Passing

- The representation $\mathbf{z}_{\xi}^{(l)}$ of a fact $\xi = ((h, r, t), Q_{\xi})$ is obtained via the **sum of the representations of its constituent relation-entity pairs**
- The representation of a relation-entity pair (p, e) with role $p \in \{\text{head, tail, qual}\}$ is computed by: $\mathbf{z}_{(p,e)}^{(l)} = \mathbf{W}_{\rho}^{(l)} [\mathbf{p}^{(l-1)}; \mathbf{e}^{(l-1)}]$
- To update a component within a fact, we compute a context message that encodes **all the other components in the fact**
- $\mathbf{m}_{\xi \rightarrow (p,e)}^{(l)} = \text{MLP}^{(l)} \left(\left(\mathbf{z}_{\xi}^{(l)} - \mathbf{z}_{(p,e)}^{(l)} \right) / (n_{\xi} + 1) \right), n_{\xi} = |Q_{\xi}|$
- $\mathbf{m}_{\xi \rightarrow e}^{(l)} = \mathbf{W}_{\text{ENT}}^{(l)} [\mathbf{m}_{\xi \rightarrow (p,e)}^{(l)}; \mathbf{p}^{(l-1)}], \mathbf{m}_{\xi \rightarrow p}^{(l)} = \mathbf{W}_{\text{REL}}^{(l)} [\mathbf{m}_{\xi \rightarrow (p,e)}^{(l)}; \mathbf{e}^{(l-1)}]$

Probabilistic Masked Fact Decoding

- Treats each masked entity and relation as a vanilla one, and updates their representations via contextual message passing

