# Agentic Predictor

## Performance Prediction for Agentic Workflows via Multi-View Encoding

Patara Trirat[1] Wonyong Jeong[1] Sung Ju Hwang[1 2]

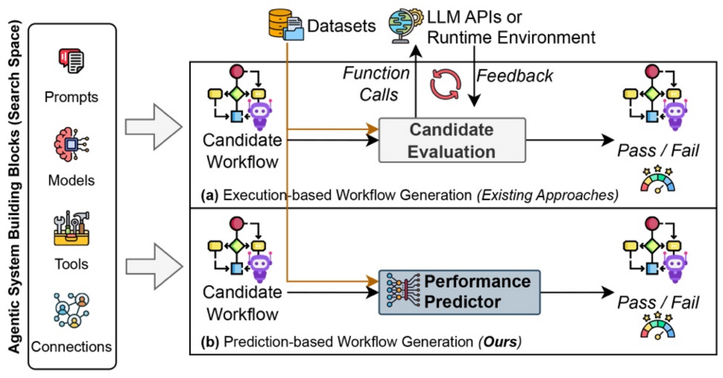
[1] DeepAuto.ai [2] KAIST

2025 ICML International Conference On Machine Learning

MAS Workshop

## 🔍 Motivation

- LLM-powered agentic systems require complex workflow design.
- Current optimization methods rely on costly execution-based evaluations.
  - Vast configuration space (prompts, agents, tools, etc.)
  - High cost of runtime evaluations for workflow selection
  - Need for task-specific adaptation




### Research Question

Can we predict performance without full executions?

### 📌 Goal

Enable efficient agentic workflow search via learned prediction models.

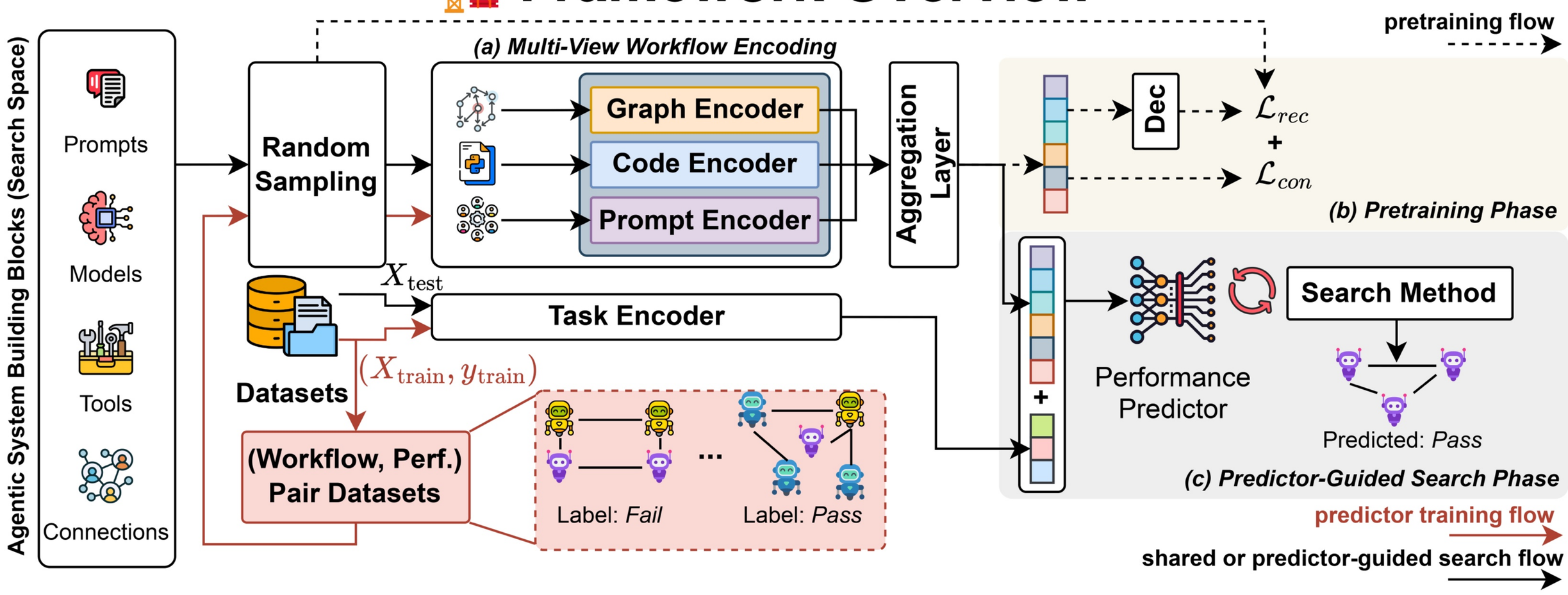
## 🚀 Our Contribution

We propose **Agentic Predictor**, a lightweight, predictive framework to estimate the success of agentic workflows using **multi-view representation** learning and unsupervised pretraining.

### 🌐 Key Features:

- **Multi-View Encoding**:
  Captures workflow heterogeneity from:
  - Graph structure (agent interaction)
  - Code semantics (logic & tool use)
  - Prompt embeddings (roles & behaviors)
- **Cross-Domain Unsupervised Pretraining**
  > Trains encoder on unlabeled workflows from various domains.
- **Lightweight Performance Predictor**
  > Guides search efficiently using minimal labeled data.

## 🏗 Framework Overview




## 📊 Experimental Results

*Table 3.* Performance comparison between Agentic Predictor and baseline methods. The best and second-best results are highlighted in **bold** and underlined, respectively.

| Domain | Code Generation | | Math Problem | | Reasoning Task | | Average | |
|---|---|---|---|---|---|---|---|---|
| **Model** | **Accuracy** | **Utility** | **Accuracy** | **Utility** | **Accuracy** | **Utility** | **Accuracy** | **Utility** |
| MLP | 78.02±0.59 | 73.94±1.35 | 73.73±0.31 | 69.64±0.29 | 78.45±0.08 | 88.48±0.63 | 76.73±0.33 | 77.35±0.76 |
| GCN | 84.35±0.34 | 72.73±3.18 | 76.19±0.42 | 66.52±1.66 | 87.12±0.14 | **91.82±0.46** | 82.55±0.30 | 77.02±1.77 |
| GAT | 84.49±0.56 | 76.46±0.91 | 76.44±0.61 | 66.51±1.28 | 87.07±0.08 | 89.40±0.68 | 82.67±0.42 | 77.46±0.96 |
| GCN-II | 83.72±0.40 | 77.75±1.98 | 75.04±0.31 | 64.33±0.47 | 87.28±0.14 | 89.92±1.90 | 82.01±0.28 | 77.33±1.45 |
| Graph Transformer | 84.71±0.45 | 74.09±0.35 | 75.45±0.23 | 66.48±0.96 | 86.93±0.27 | 90.60±1.97 | 82.36±0.32 | 77.06±1.09 |
| One For All | 81.05±0.34 | 73.42±1.39 | 75.21±0.23 | 69.08±0.64 | 82.52±0.13 | 87.64±1.98 | 79.59±0.23 | 76.71±1.34 |
| *Agentic Predictor* | **85.62±0.47** | **80.08±0.46** | **79.56±0.25** | **74.08±0.47** | **87.96±0.02** | 91.47±0.44 | **84.38±0.25** | **81.88±0.46** |
| % Improvement (up to) | 9.74% | 10.11% | 7.91% | 15.16% | 12.12% | 4.37% | 9.97% | 6.74% |

Homepage

*Table 4.* Results of ablation study on different input view variations.

| View Variations | | | Code Generation | | Math Problem | | Reasoning Task | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Code** | **Graph** | **Text** | **Accuracy** | **Utility** | **Accuracy** | **Utility** | **Accuracy** | **Utility** | **Accuracy** | **Utility** |
| ✓ | | | 82.04±0.51 | 75.66±0.66 | 75.70±0.14 | 68.52±0.91 | 83.19±0.56 | **91.51±0.61** | 80.31±0.40 | 78.56±0.73 |
| | ✓ | | 84.44±0.31 | 77.22±3.46 | 79.14±0.28 | 67.99±3.36 | 87.00±0.21 | 91.03±1.23 | 83.53±0.27 | 78.75±2.68 |
| | | ✓ | 79.87±0.28 | 70.34±0.43 | 76.60±0.65 | 68.45±1.80 | 68.06±0.00 | 71.04±0.00 | 74.84±0.31 | 69.94±0.74 |
| ✓ | ✓ | | 83.72±0.83 | 73.97±0.81 | 75.86±0.85 | 70.18±1.64 | 86.88±0.14 | 86.14±4.62 | 82.15±0.61 | 76.76±2.36 |
| ✓ | | ✓ | 82.27±0.63 | 77.28±1.12 | 76.03±0.14 | 66.66±4.18 | 54.17±0.00 | 53.21±0.00 | 70.82±0.26 | 65.72±1.77 |
| | ✓ | ✓ | 82.45±1.36 | 74.64±1.57 | 75.70±1.26 | 67.83±3.71 | 69.47±0.00 | 70.55±0.00 | 75.87±0.87 | 71.01±1.76 |
| ✓ | ✓ | ✓ | **85.62±0.47** | **80.08±0.46** | **79.56±0.25** | **74.08±0.47** | **87.96±0.02** | 91.47±0.44 | **84.38±0.25** | **81.88±0.46** |

Paper