

Mind the Data Gap: Bridging LLMs to Enterprise Data Integration

Moe Kayali (University of Washington), Fabian Wenz (TU Munich and MIT), Peter Baile Chen (MIT), Nesime Tatbul (Intel Labs and MIT), Çağatay Demiralp (AWS AI Labs and MIT), Michael Stonebraker (MIT CSAIL)

To Appear at CIDR 2025

NEDB Day 2025 - Jan 8th

Concerns

External Validity

Distribution Shift

**Domain-specific
Knowledge**

**Data
Leakage**

Concerns

External Validity

Distribution Shift

This work's focus

**Domain-specific
Knowledge**

**Data
Leakage**

Enterprise Performance Gap

More Realistic Setting:

Need to Learn Labels Too

Contributions

