

Metadata Intelligence Platform

A Hybrid Multi-Agent LLM Framework for Metadata Enrichment and Risk-Aware ETL Analysis

Sinem Sönmez · Tahir Belkuyu · Metehan Mut | Advisor: Asst. Prof. Polat Göktaş (Sabancı) · Industry Supervisor: Volkan Yurtseven (Akbank)



ENS 492 · GRADUATION PROJECT
May 2026



FACULTY OF ENGINEERING
AND NATURAL SCIENCES

ENS 492 · Graduation Project (Implementation) | Sabancı University × Akbank | May 2026

ABSTRACT Hybrid multi-agent framework for enterprise metadata enrichment using LLM Generator–Critic loop, context-aware retrieval (DDL · FRD · TOA), and rule-based validation. **F1 = 87.5% · Clarity = 76.2/100 · Pilot: 5 tables · 18 columns · 10 HIGH / 8 LOW risk labels · Gemini 1.5 Flash**
Kurumsal metadata zenginleştirme için LLM Generator–Critic döngüsü, context-aware retrieval (DDL · FRD · TOA) ve rule-based validation kullanan hibrit çok ajanlı framework.

WHY? / PROBLEM TANIMI

Enterprise data warehouses accumulate incomplete, incorrect, or multilingual metadata — causing failures in governance, BDDK compliance, and semantic consistency.

Kurumsal veri ambarlarında eksik, hatalı veya çok dilli metadata birikimi; veri yönetişimi, BDDK uyumluluğu ve semantik tutarlılık problemlerine yol açmaktadır.

✗ Incomplete Metadata / Eksik Metadata

Null entries, single-sentence placeholders with no business context

İş bağlamı içermeyen null kayıtlar ve yetersiz açıklamalar

⚠ Factual Mismatch / Anlamsal- Gerçeklik Uyumsuzluğu

30GCK documented as "60-day" — silent error propagates downstream

30GCK kolonunun "60 gün" olarak açıklanması — downstream süreçlere sessiz hata yayılımı

🌐 Multilingual Issue / Çok Dilli Tutarlılık

Turkish schema columns described in English, breaking cross-team semantics

Türkçe şema kolonlarının İngilizce açıklanması nedeniyle ekipler arası semantik uyumsuzluk

🔍 Missing LKP Refs / Eksik Lookup Referansları

Low-cardinality columns lack lookup tables — value semantics opaque

Düşük kardinaliteli kolonlarda lookup tablolarının eksik olması nedeniyle değer anlamlarının belirsizleşmesi

HOW? / YÖNTEM & MİMARİ

SYSTEM ARCHITECTURE



INPUT DATA / GİRDİ VERİLERİ

Enterprise Metadata Repository
Kurumsal Metadata Deposu



CONTEXT RETRIEVAL / BAĞLAM GETİRME

DDL · FRD · TOA per Column
Kolon bazlı DDL · FRD · TOA retrieval işlemi



GENERATOR AGENT / ÜRETİCİ AJAN

Context-Aware Generation (Gemini 1.5)
Bağlam farkındalıklı metadata üretimi (Gemini 1.5)



CRITIC AGENT / DEĞERLENDİRİCİ AJAN

4-Dim Scoring: Clarity · Completeness · Semantic · Business
4 boyutlu skorlandırma: Clarity · Completeness · Semantic · Business



RE-GENERATION LOOP / YENİDEN ÜRETİM DÖNGÜSÜ

Score < 60 → Feedback Appended (max 2 iter.)
Skor < 60 ise geri bildirim eklenerek yeniden üretim (maks. 2 iterasyon)



VALIDATION + RISK ENGINE / DOĞRULAMA + RİSK MOTORU

ETL DFS · Cardinality · LKP Check · 0–100 Risk Score
ETL DFS · Cardinality · LKP kontrolü · 0–100 risk skoru



FINAL OUTPUT / NİHAİ ÇIKTI

Enriched Descriptions · Risk Bands · Lineage Report
Zenginleştirilmiş açıklamalar · Risk bantları · Lineage raporu

PIPELINE FLOW / İŞ AKIŞI

Context Injection / Bağlam Enjeksiyonu

DDL + FRD + TOA docs injected into prompt. Prompt augmentation strategy.

Generate / Üretim

LLM generates context-aware descriptions. Gemini 1.5 Flash.

Critique / Değerlendirme

4-dim scoring: Clarity · Completeness · Semantic consistency · Business-rule consist.

Re-Generate/ Yeniden Üretim (score < 60)

Threshold empirically selected in pilot. Feedback appended. Max 2 iterations.

Risk Classification / Risk Sınıflandırması

Rule-based module (NOT Critic). 0-100 risk score. Bands: ≥60 High · 30-59 Mid · <30 Low.

ETL LINEAGE ANOMALY DETECTION

XXX_HESAP
CORE_BANKING

KRD_KREDİ
CREDIT

RİSK_İZLEME
RISK ⚠

MUS_SEGMENT
CRM ⚠

ETL_001 → ETL_002 → ETL_003 ⇄ ETL_004 ⚠ LOOP DETECTED ETL_005

RESEARCH GAP / ARASTIRMA BOŞLUĞU

→ LLM critique loop
→ ETL lineage analysis
→ Rule-based validation
→ Continuous risk scoring

→ LLM değerlendirme döngüsü
→ ETL lineage analizi
→ Kural tabanlı doğrulama
→ Sürekli risk puanlama

OBJECTIVES / AMAÇLAR

- Automate metadata enrichment via LLM agents
LLM ajanları ile metadata zenginleştirmesini otomatikleştirmek
- Detect ETL lineage loops with DFS traversal
DFS traversal kullanarak ETL lineage döngülerini tespit etme
- Continuous risk scoring replacing binary labels
Binary risk etiketleri yerine sürekli risk puanlama sistemi kullanmak
- Ablation study validating each component
Her bileşeni doğrulayan ablation study gerçekleştirmek

BEFORE → AFTER / ÖNCE → SONRA

BEFORE ✗

INT_SHK_A_30GCK:
"Akbank Hariç KK Açık Son 60 gün"
← WRONG

AFTER ✓

...son 30 gün ortalamasıdır.
⚠ "60 gün" → kolon adı 30GCK
= 30 günü ifade eder.

87

Null Desc.

12

LKP Gaps

2

ETL Loops

3

Lang Conflicts

KEY CONTRIBUTIONS / TEMEL KATKILAR

→ Hybrid multi-agent (LLM + rules) architecture
→ 4-config ablation study validating each component
→ Continuous 0–100 risk score (replaces binary labels)
→ ETL lineage loop detection via DFS traversal

→ Hibrit multi-agent (LLM + kurallar) mimarisi
→ Her bileşeni doğrulayan 4-konfigürasyonlu ablation study
→ Binary etiketleme yerine sürekli 0–100 risk puanlama sistemi
→ DFS traversal ile ETL lineage döngü tespiti

REAL-WORLD IMPACT · FUTURE WORK / GERÇEK DÜNYA ETKİSİ · GELECEK ÇALIŞMALAR

→ Reduces manual metadata effort for enterprise DWH
→ Supports BDDK compliance · Bilingual TR/EN generation
→ Future: Enterprise-scale benchmarking (100+ tables)
→ Real-time lineage monitoring · Multilingual expansion

→ Kurumsal veri ambarlarında manuel metadata yükünü azaltır
→ BDDK uyumluluğunu destekler · Çift dilli TR/EN metadata üretimi sağlar
→ Gelecek çalışma: Enterprise-scale benchmarking (100+ tablo)
→ Gerçek zamanlı lineage monitoring · Çok dilli genişleme

RESULTS / SONUÇLAR

CONFIG A — FULL PIPELINE ★

87.5%

F1 Score

90.0%

HIGH RISK Recall

93.0%

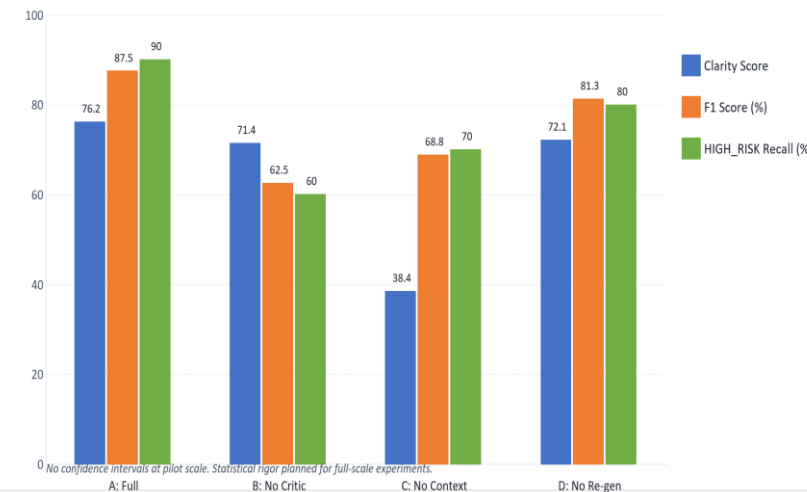
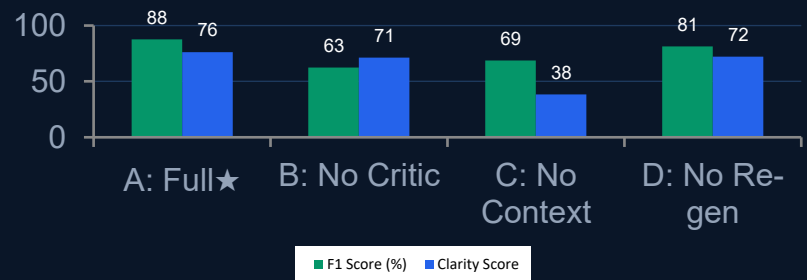
LOW RISK Precision

76.2

Avg Clarity (0-100)

ABLATION COMPARISON / ABLATION KARŞILAŞTIRMASI

RESULTS — VISUAL COMPARISON (PILOT)



–37.8 pts

Context removal: biggest clarity drop

–25 pp

Critic removal reduced F1 by 25 points

100%

Validation: full mismatch detection

ABLATION TABLE / ABLATION TABLOSU

Config	F1 (%)	Clarity	H.Rec.	Tokens
A: Full ★	87.5 ★	76.2 ★	90.0 ★	~104k
B: No Critic	62.5 ✗	71.4	60.0 ✗	~28k
C: No Context	68.8	38.4 ✗	70.0	~72k
D: No Re-gen	81.3	72.1	80.0	~72k

COMPONENT IMPACT / BİLEŞEN ETKİSİ

Context Docs	–37.8 Clarity	CRITICAL
Critic Agent	–25pp F1	HIGH
Risk Engine	100% Detect.	ESSENTIAL

KEY FINDINGS / TEMEL BULGULAR

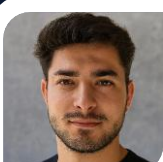
- Critic Agent: +25pp F1**
Config B (no Critic): F1 87.5%→62.5% (–25pp), HIGH_RISK Recall 90%→60%
- Context = Dominant Driver**
No context: clarity 76.2→38.4 (–37.8 pts) — largest single component impact
- Re-generation: Targeted Fix**
+4.1 clarity / +6.2 F1 avg. Corrects 2 HIGH_RISK misclassifications
- Continuous 0–100 Risk Score**
Binary HIGH/LOW → 0–100 scoring with penalty factors & band thresholds

TEAM



Sinem Sönmez

- System architecture design
- LLM integration and workflow coordination
- Context-aware metadata generation
- Multi-agent pipeline development



Tahir Belkuyu

- ETL lineage analysis
- DFS-based loop detection
- Rule-based validation module
- Continuous risk scoring framework,



Metehan Mut

- Experimental evaluation
- Ablation study analysis
- Performance and metric assessment
- Comparative result interpretation