# KADİR HAS ÜNİVERSİTESİ BİLGİ MERKEZİ

## YAPAY ZEKA TESPİT TEKNOLOJİLERİ: KURUMSAL DEĞERLENDİRME POLİTİKASI VE TEKNİK STANDARTLAR

**Doküman Kodu:** KHAS-AI-POLICY-2025/v1.0

**Yayın Tarihi:** 19 Kasım 2025

**Statü:** Kurumsal Politika Belgesi (Kamuya Açık Referans Nüsha)

**Dayanak:** UNESCO (2023) ve ENAI (2024) Etik Çerçevesi

### 1. GİRİŞ VE AMAÇ

Kadir Has Üniversitesi Bilgi Merkezi, akademik süreçlerde kullanılan üretken yapay zeka (Generative AI) tespit araçlarını değerlendirirken, ticari vaatler yerine **kanıt tabanlı (evidence-based) bilimsel metrikleri** esas almayı ilke edinmiştir.

Geleneksel "intihal tespiti" (benzerlik) ile "yapay zeka tespiti" (köken analizi) birbirinden tamamen farklı disiplinlerdir. Bu doküman, kurumumuzun satın alma ve deneme erişimi süreçlerinde tedarikçilerden talep ettiği teknik standartları, test protokollerini ve etik sınırları tanımlamak üzere hazırlanmıştır. Doküman, akademik dürüstlük alanında çalışan diğer yükseköğretim kurumlarıyla **"İyi Uygulama Örneği" (Best Practice)** olarak paylaşılmaktadır.

### 2. YÖNETİCİ ÖZETİ VE TEMEL YAKLAŞIM

Kurumumuz, yapay zeka tespit araçlarını bir "ceza mekanizması" olarak değil, akademik rehberlik aracı olarak konumlandırmaktadır. Bu bağlamda temel yaklaşımımız şudur:

- **Masumiyet Karinesi Esastır:** Hiçbir yazılımın ürettiği skor, tek başına kesin delil (conclusive evidence) kabul edilemez.
- **Olasılık vs. Kesinlik:** İntihal raporları deterministik (var/yok), AI raporları ise olasılıksaldır (muhtemelen).
- **Sıfır Tolerans:** "False Positive" (Masum öğrenciyi suçlama) oranı %1'in üzerinde olan araçlar, idari risk taşıdığı gerekçesiyle kurumumuzca reddedilir.

### 3. TEMEL KAVRAMLAR SÖZLÜĞÜ (GLOSSARY)

Karar vericiler için kritik teknik tanımlar:

- **False Positive (Yanlış Pozitif):** İnsan tarafından yazılmış özgün bir metnin, yazılım tarafından hatalı bir şekilde "AI" olarak etiketlenmesi.
- **Adversarial Robustness (Saldırı Direnci):** Aracın, öğrenci tarafından kasıtlı olarak yapılan harf değişiklikleri, gizli karakterler veya kelime değişimleri (paraphrasing) karşısındaki dayanıklılığı.
- **Bias (Dil Yanlılığı):** Algoritmaların, anadili İngilizce olmayan (Non-native speakers) yazarların kısıtlı kelime dağarcığını "robotik" olarak algılayıp yanlış işaretleme eğilimi. *Türkiye'deki üniversiteler için en kritik risk faktörüdür.*
- **Explainability (Açıklanabilirlik):** Aracın kararını *neden* verdiğini (olasılık haritaları, sentaks analizi vb.) kanıtlayabilme yeteneği.

## 4. DEĞERLENDİRME KRİTERLERİ VE PUANLAMA (SCORECARD)

Bilgi Merkezi, araçları aşağıdaki ağırlıklandırılmış kriter setine göre puanlar. Toplam skoru **75'in altında kalan** araçlar teknik değerlendirmeyi geçemez.

| Kriter | Ağırlık | KHAS Standartı ve Beklenti |
|---|---|---|
| 1. Güvenlik (False Positive) | %40 | Kritik Eşik: Bağımsız testlerde hata payı <%1 olmalıdır. 2020 öncesi tezlerde "AI" uyarısı veren araçlar doğrudan elenir. |
| 2. Dilsel Adalet (Bias Testi) | %20 | Stanford Kriteri: Araç, Türk öğrencilerin yazdığı İngilizce veya Türkçe metinleri "sözcük çeşitliliği az" olduğu için AI olarak etiketlememelidir. |
| 3. Saldırı Direnci (Robustness) | %20 | RAID Kriteri: Quillbot vb. araçlarla değiştirilmiş metinlerde tespit başarısı %85'in üzerinde olmalıdır. |
| 4. Açıklanabilirlik (Kanıt) | %10 | Rapor, disiplin kuruluna sunulabilir nitelikte cümle |

| | | |
|---|---|---|
| | | bazlı detay ve olasılık grafiği içermelidir. |
| 5. Model Kapsamı | %10 | Sadece GPT değil; Claude, Gemini ve Llama modellerini de tanımalıdır. |

## 5. KHAS BİLGİ MERKEZİ TEST PROTOKOLÜ

Tedarikçi firmaların ürünleri, satın alma kararı verilmeden önce aşağıdaki **"Kör Test" (Blind Test)** adımlarına tabi tutulur:

1. **Grup A - Tarihsel Kontrol (Zorunlu):** Arşivden 2018-2019 yıllarına ait 5 adet Türkçe Yüksek Lisans tezi yüklenir. *(Beklenen Sonuç: %0 AI)*.
2. **Grup B - Dil Önyargısı Testi:** Bir Türk öğrenci tarafından orta seviye (B2) İngilizce ile yazılmış özgün metin yüklenir. *(Beklenen Sonuç: İnsan)*.
3. **Grup C - Saldırı Testi:** AI ile yazılıp Quillbot ile değiştirilmiş metin yüklenir. *(Beklenen Sonuç: >%80 AI)*.
4. **Grup D - Çeviri Testi:** Türkçe -> İngilizce -> Türkçe çeviri döngüsüne sokulmuş insan metni yüklenir. *(Beklenen Sonuç: İnsan veya Düşük AI Skoru)*.

## 6. REFERANS KAYNAKLAR (LITERATURE)

*Bu politika belgesi, aşağıdaki bilimsel çalışmalar temel alınarak oluşturulmuştur:*


1. **Liang, W., et al. (2023).** *"GPT Detectors are Biased Against Non-Native English Writers."* Patterns (Cell Press) / Stanford University.
2. **University of Chicago (2024).** *"Artificial Writing and Automated Detection Report."* Becker Friedman Institute.
3. **Weber-Wulff, D., et al. (2023).** *"Testing of Detection Tools for AI-Generated Text."* International Journal for Educational Integrity.
4. **RAID Benchmark (2024).** *"Robust AI Detection Leaderboard."* Association for Computational Linguistics (ACL).
5. **UNESCO (2023).** *"Guidance for generative AI in education and research."* Paris: UNESCO.
6. **ENAI (2024).** *"Recommendations on the ethical use of AI in Education."* European Network for Academic Integrity.


## ⚠ YASAL BİLGİLENDİRME

*Bu doküman, Kadir Has Üniversitesi Bilgi Merkezi'nin iç kalite güvence süreçlerini tanımlar ve mesleki bilgi paylaşımı amacıyla kamuya sunulmuştur. Dokümanda yer alan kriterler, teknolojinin doğası gereği değişime açıktır. Nihai satın alma ve uygulama kararı, ilgili kurumların kendi yetkili organlarına aittir.*

# EK-1: TEDARİKÇİ TEKNİK YETERLİLİK BEYAN FORMU

**(Bu formun doldurulması, Kurumumuzda yapılacak deneme erişimi veya satın alma değerlendirmesi için ön koşuldur.)**

**Firma Adı:** _____

**Ürün Adı/Versiyon:** _____

## A. TEKNİK PERFORMANS BEYANI

*Lütfen aşağıdaki kriterlere "Evet/Hayır" yanıtı veriniz ve iddianızı destekleyen bağımsız rapor linkini ekleyiniz. Pazarlama dokümanları kanıt kabul edilmez.*

| Kriter | Yanıt | Kanıt Dokümanı / Link (Zorunlu) |
|---|---|---|
| 1. GÜVENLİK (False Positive): Ürününüzün "Yanlış Pozitif" oranı bağımsız testlerde %1'in altında mıdır? (Örn: Masum öğrenciyi suçlama riski) | ☐ Evet<br>☐ Hayır | _____ |
| 2. TARİHSEL AYRIM: Ürün, 2020 öncesi (AI olmayan) akademik metinlerde %0'a yakın hata oranını garanti eder mi? | ☐ Evet<br>☐ Hayır | _____ |
| 3. SALDIRI DİRENCİ: Metin, Quillbot vb. araçlarla yeniden yazıldığında (re-phrased) tespit başarınız %85 üzerinde kalıyor mu? | ☐ Evet<br>☐ Hayır | _____ |
| 4. DİL AYRIMCILIĞI (Bias): Stanford Üniversitesi (Liang et al.) araştırmasında belirtilen, "Anadili İngilizce | ☐ Evet<br>☐ Hayır | _____ |

| | | |
|---|---|---|
| Olmayan Yazarlar"a karşı oluşan önyargıyı giderici algoritmanız var mı? | | |
| 5. TÜRKÇE YETKİNLİĞİ: Modeliniz Türkçe'nin sondan eklemeli yapısına göre özel eğitildi mi? (Native training). | ☐ Evet<br>☐ Hayır | _____ |

## C. FİRMA YETKİLİSİ BEYANI

*Yukarıda verdiğim bilgilerin doğruluğunu beyan ederim. Kurum tarafından yapılacak "Kör Testler" sonucunda, beyan edilen performans değerlerinden negatif yönde %10'dan fazla sapma tespit edilmesi durumunda; ürünümüzün değerlendirme listesinden çıkarılacağını kabul ediyorum.*

**Adı Soyadı:** _____ **İmza/Tarih:** _____