

# Measuring Logical Reasoning in Turkish Offline Language Models via Fallacy Detection

---

## Yanılıđı Tespiti Yoluyla Trke evrimdiŐı Dil Modellerinde Mantıksal Akıl Yrtmenin llmesi

Inanc Feridun Canci, Zeynep Gizem Emir Shahbazov, Edibe Yılmaz, **Kahraman Kostas**  
YYEGM, MEB



# Kalem ve Silgi Problemi

Bir kalem ve bir silginin toplam fiyatı **110 TL**'dir.

Kalem, silgiden **100 TL** daha pahalıdır.

## Kalem ve Silgi Kaç Liradır?



**Kalem =** \_\_\_\_\_



**Silgi =** \_\_\_\_\_



THE NEW YORK TIMES BESTSELLER

THINKING,  
FAST AND SLOW



DANIEL

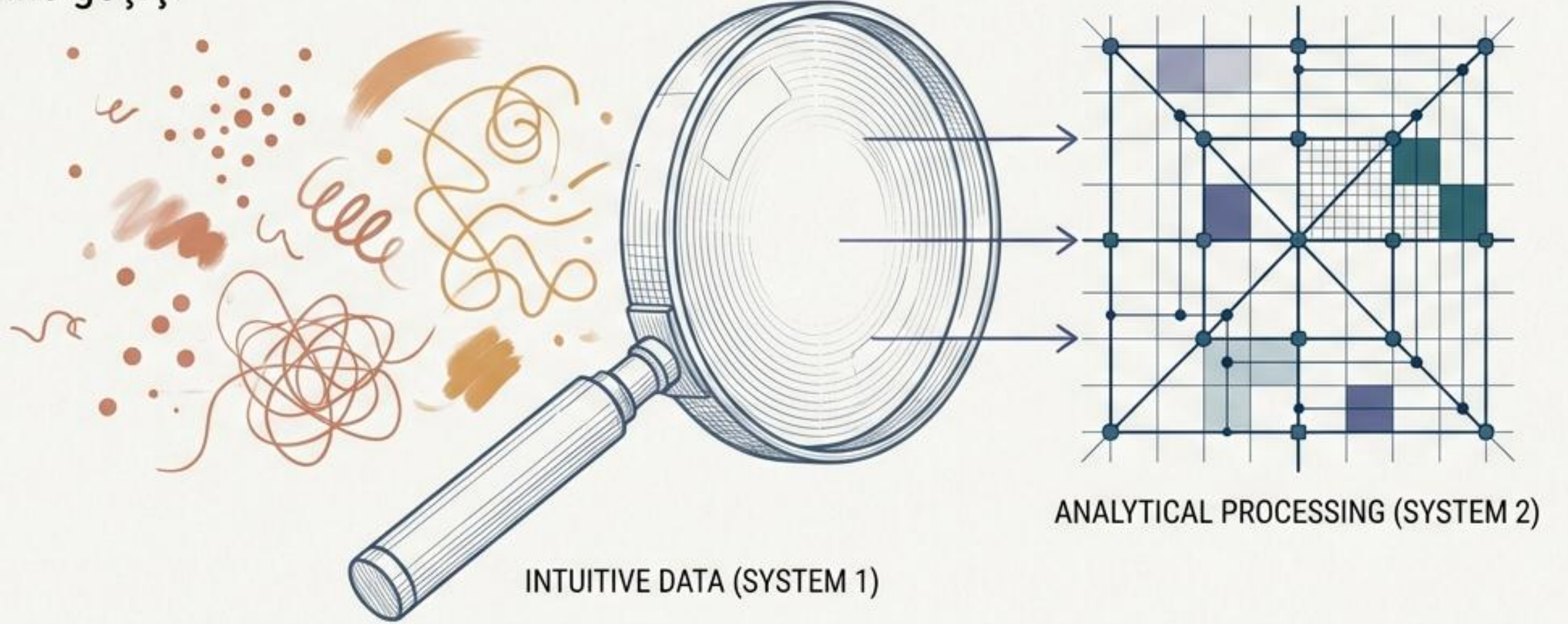
KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

"[A] masterpiece... This is one of the greatest and most engaging collections of insights into the human mind I have read." —WILLIAM EASTERLY, *Financial Times*

# Mantığın Merceği: Çevrimdışı Dil Modellerinde Yapay Zeka Akıl Yürütmesi

Yerel sistemlerde mantıksal safsata tespiti ve "Sistem 2" analizine geçiş.

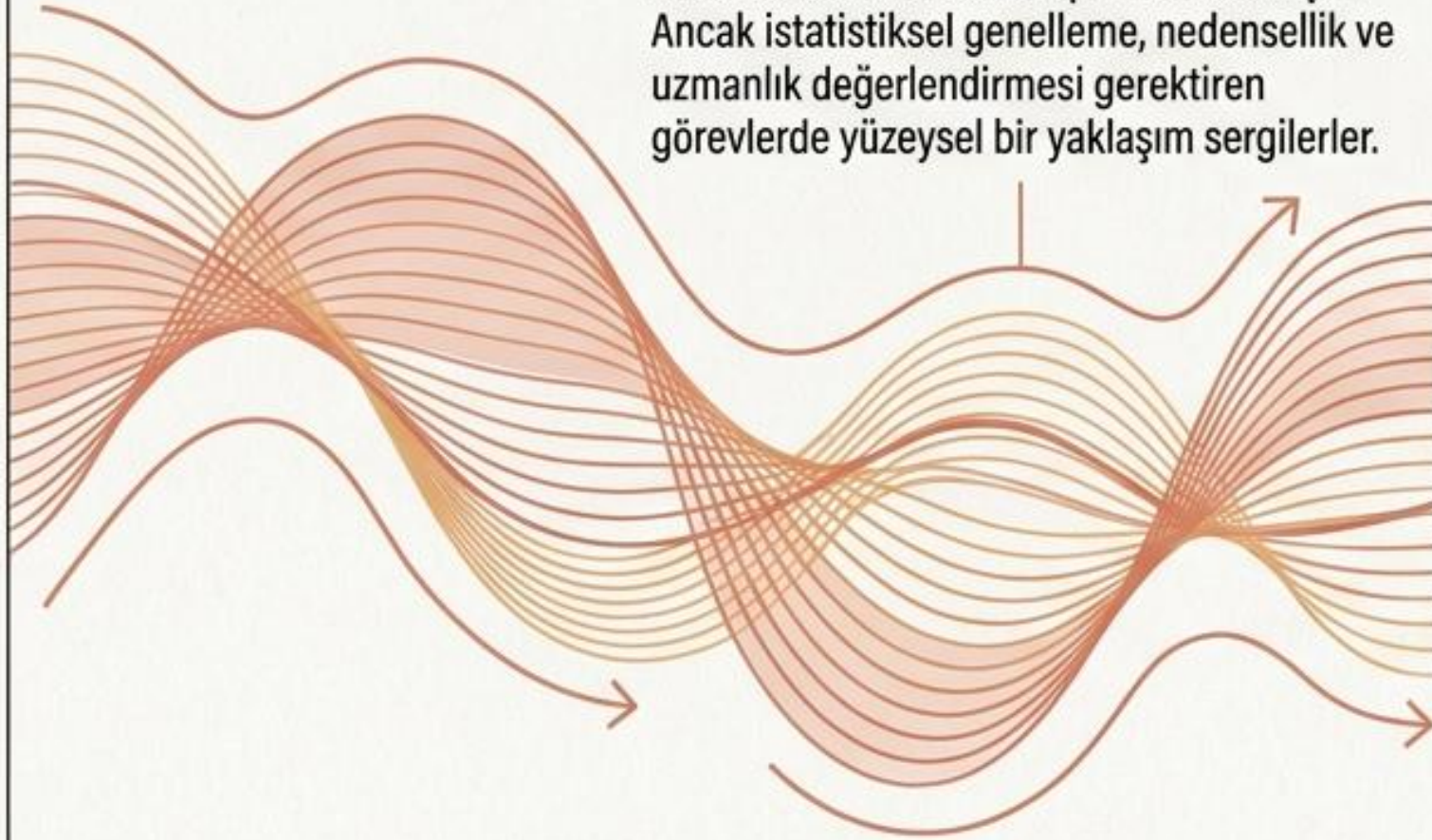


# Akıcı Metinler, Kusurlu Mantık

Dijital çağda dezenformasyon, ikna edici görünen ancak kusurlu olan mantıksal safsatalardan beslenir.

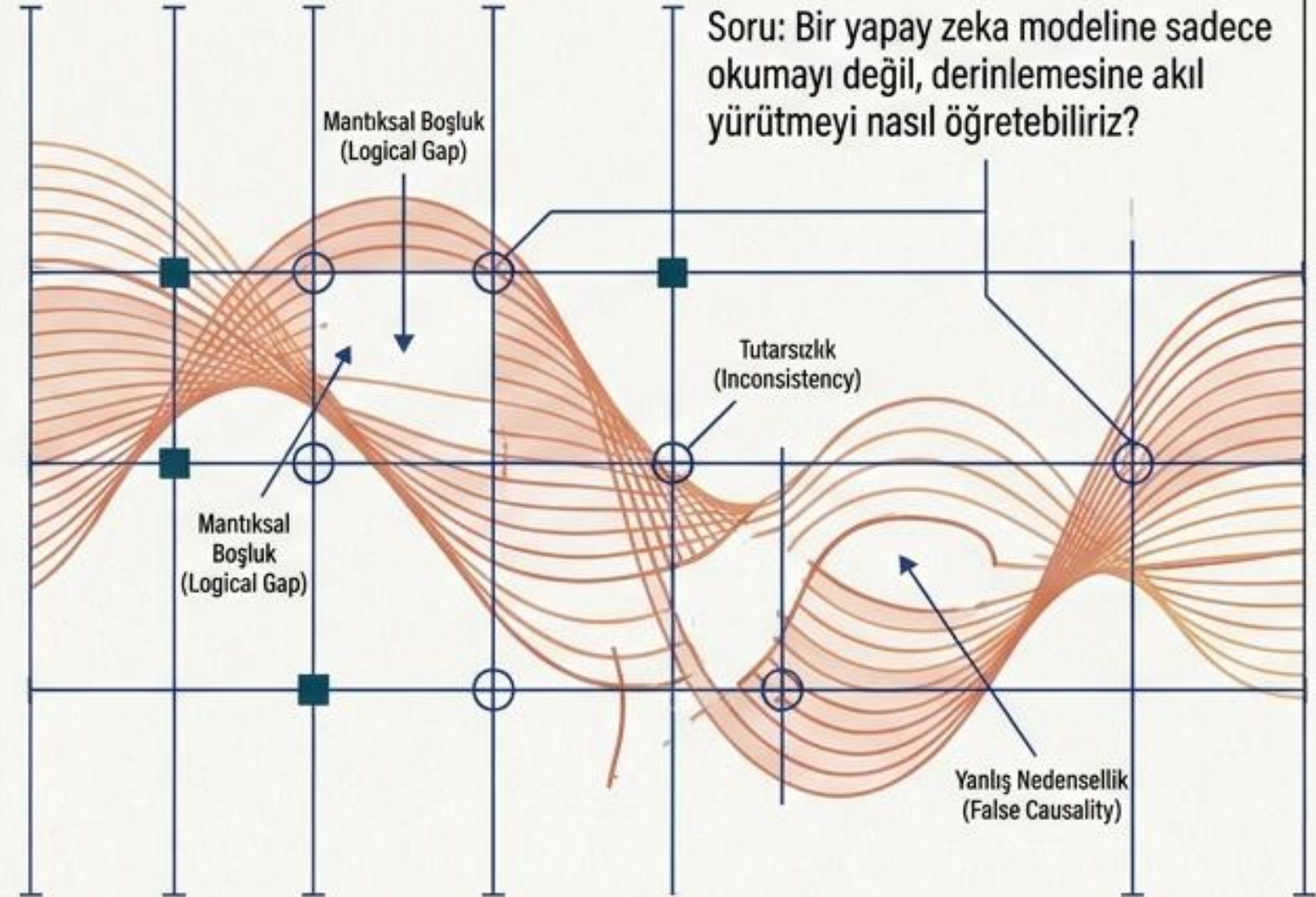
## Akıcı Metin Üretimi

Sorun: Büyük Dil Modelleri yapısal olarak akıcı metinler üretmek üzere optimize edilmiştir. Ancak istatistiksel genelleme, nedensellik ve uzmanlık değerlendirmesi gerektiren görevlerde yüzeysel bir yaklaşım sergilerler.



## Mantıksal Analiz

Soru: Bir yapay zeka modeline sadece okumayı değil, derinlemesine akıl yürütmeyi nasıl öğretebiliriz?



# Yapay Zekanın Bilişsel Uyanışı

## Sistem 1 (Sezgisel)



### Kör Değerlendirme (Blind)

- Modelin açık bir rehberlik olmadan örüntü tanımasına dayalı varsayılan davranışıdır.
- Hızlıdır ancak yüzeysel benzerliklere odaklandığı için mantıksal yanılgılara açıktır.

## Sistem 2 (Analitik)



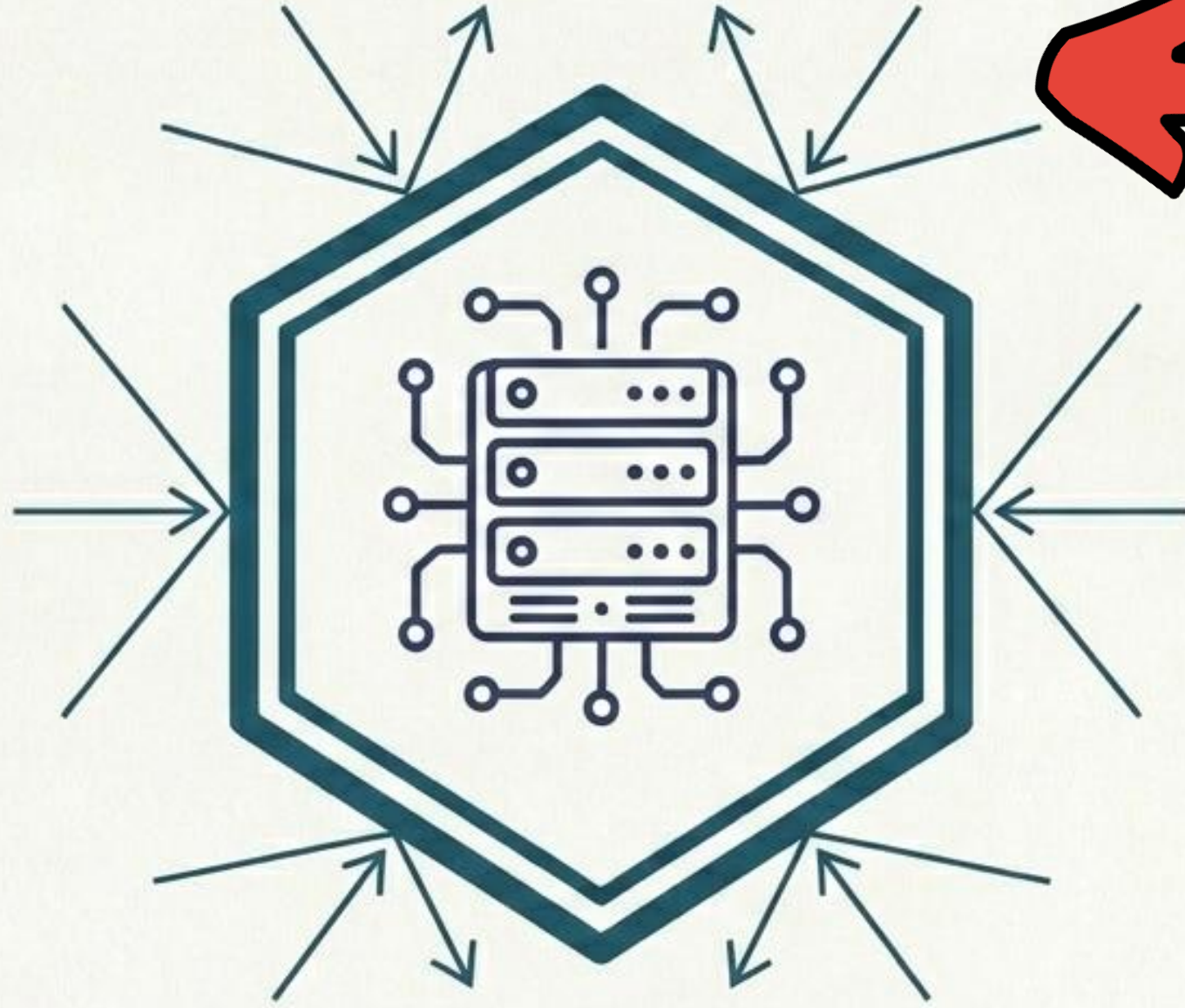
### Bilgilendirilmiş İstem (Informed)

- Yapılandırılmış bilgi desteğiyle (Knowledge-Augmented) kurallı analiz sürecidir.
- Modeli derinlemesine mantık yürütmeye zorlar. Temkinli ve kesin sonuçlar üretir.

# Güvenli ve Yerel Analiz İhtiyacı

## There Are No Silly Questions: Evaluation of Offline LLM Capabilities from a Turkish Perspective

Edibe Yılmaz and Kahraman Kostas



**Minstral-3-14B-Reasoning**

### Veri Güvenliği

Dış API servislerine bağımlılık yoktur; tamamen kapalı devre ve yerel çalışarak maksimum gizlilik sağlar.

### Türkçe NLP Kapasitesi

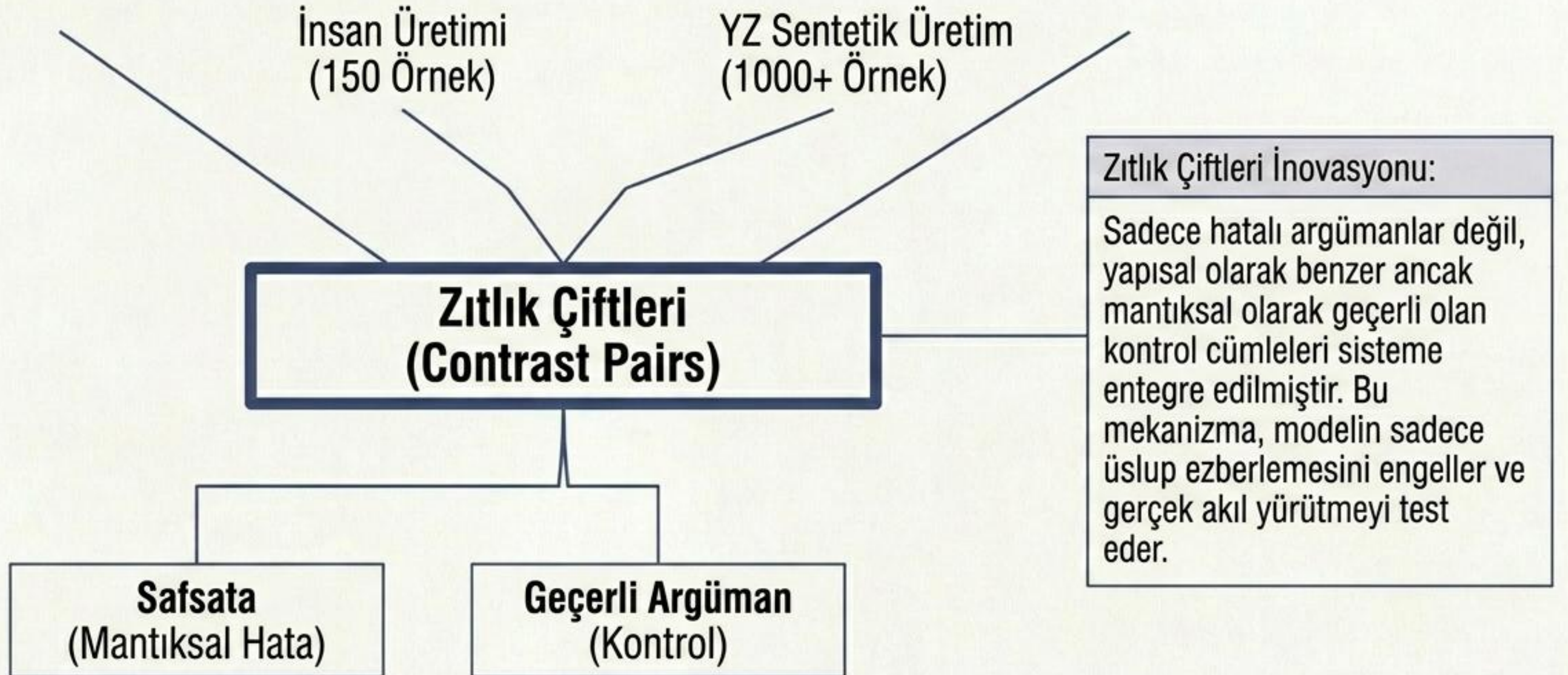
Türkçe dil yapısında mantıksal akıl yürütme (Chain-of-Thought) süreçlerindeki belgelenmiş yüksek başarı oranı.

### Hız & Boyut Optimizasyonu

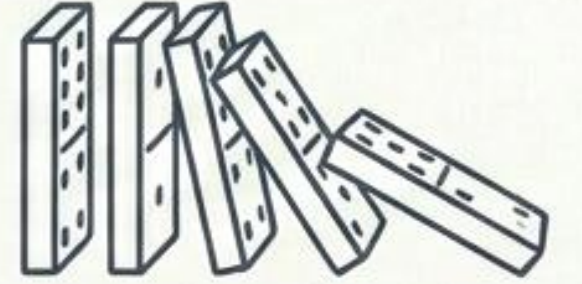
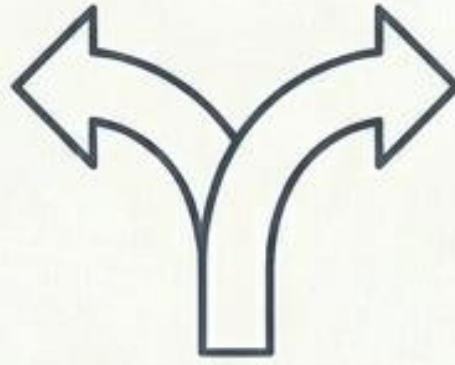
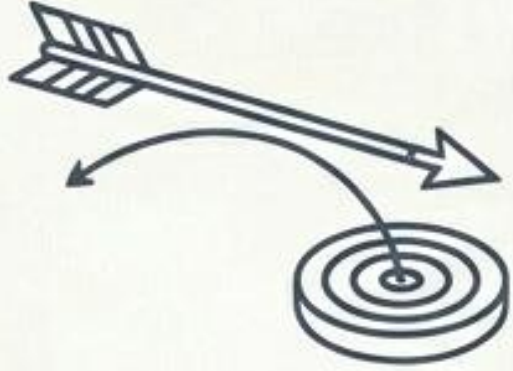
Çevrimdışı donanımlarda optimum performans göstermesi için ideal olarak ayarlanmış 14B parametre ölçeği.

# Test Alanı: TR-Fallacy Veri Seti

Türkçe literatürdeki veri seti boşluğunu doldurmak için tasarlanmış hibrit mimari.



# Teşhis Matrisi: 5 Temel Safsata



Yanlış Nedensellik

Otoriteye  
Başvurma

Yanlış İkilem

Aceleci Genelleme

Kaygan Zemin

A, B'den önce  
olduysa; A, B'nin  
nedenidir.  
(Korelasyon  $\neq$   
Nedensellik).

Geçersiz veya  
bağlam dışı bir  
uzmanlığın  
sorgulanmadan  
mutlak doğru kabul  
edilmesi.

Gerçekte çok  
sayıda ara  
seçenek varken  
durumu sadece  
siyah/beyaz gibi  
sunma.

Yetersiz istatistiksel  
veriden veya tekil  
örneklerden  
evrensel bir kural  
çıkarma.

Atılan küçük bir  
adımın hiçbir kanıt  
olmadan  
kaçınılmaz bir  
felaketle  
sonuçlanacağı  
varsayımı.

# İki Senaryo: Sezgiye Karşı Analiz

## Kör Değerlendirme / Sistem 1



Kör Değerlendirme (Blind): Modele tanım verilmez. Sadece yüzeysel benzerlikler ve varsayılan ağırlıklar ile safsata tespiti istenir.

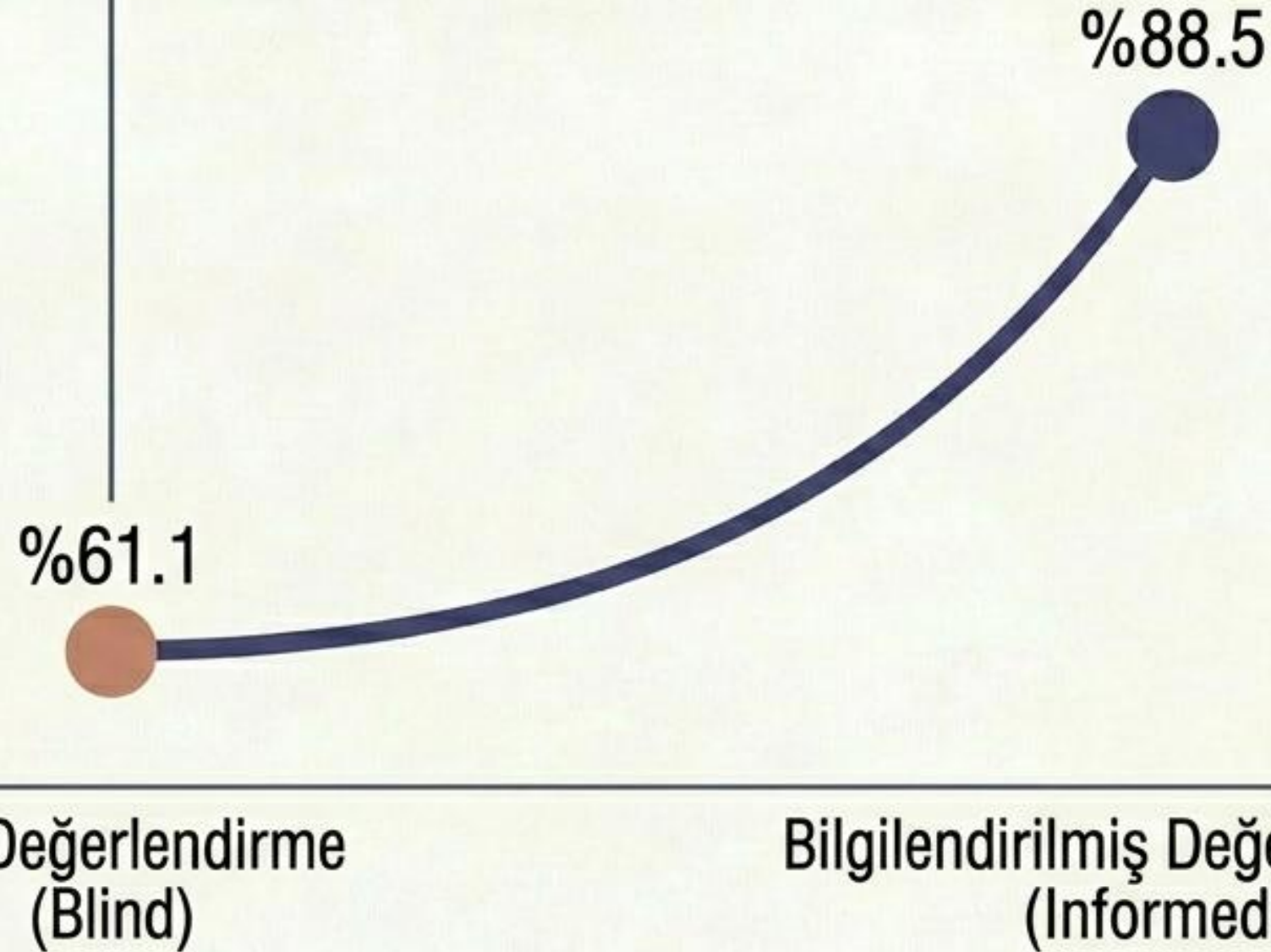
## Bilgilendirilmiş İstem / Sistem 2



Bilgilendirilmiş İstem (Informed): Modele safsata türlerinin özet tanımları sistem istemi olarak sunulur. Model, içsel bilgisini bu yapılandırılmış tanımlarla entegre etmeye zorlanır.

# Doğruluk Sıçraması (İkili Sınıflandırma)

Yapılandırılmış kurallar eklendiğinde, modelin 'Safsata' ile 'Geçerli Argümanı' ayırt etme kapasitesi dramatik şekilde artar.



## Cohen's Kappa Sıçraması

**0.218 → 0.770**

Bu 'Uyuşma Güvenilirliği' istatistiği, başarının tesadüfi olmadığını ve modelin akıl yürütme sürecinin tanımlarla istikrara kavuştuğunu matematiksel olarak kanıtlar.

# Çok Sınıflı Hassasiyet: Çapraz Karışıklığın Giderilmesi

## Kör Sistem: Yüzeysel Benzerlik Yanılgısı

132	35	0	37	6
0	180	1	29	0
0	7	199	4	0
7	21	0	174	0
0	58	19	11	152

Tanımlar olmadan, model farklı safsata türlerini birbirine karıştırır.

## Bilgilendirilmiş Sistem: Kesin Kategori İzolasyonu

191	4	5	10	0
0	201	3	6	0
0	0	199	0	0
9	7	0	181	0
0	35	21	26	145

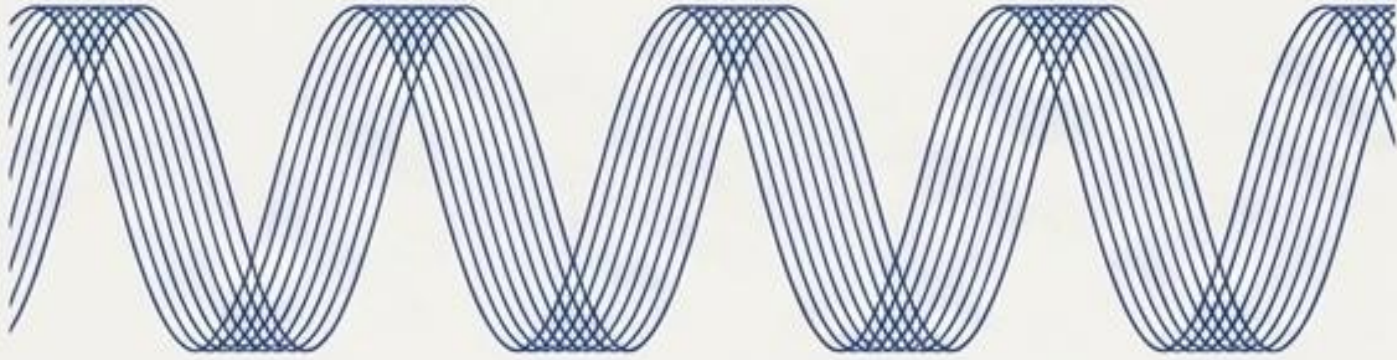
Kurallar uygulandığında, model 'neden' hatalı olduğunu tanımlayabilir hale gelir (Doğruluk: %78.1 → %87.9).

5 farklı safsata türünü ayırt etme görevinde, 'Bilgilendirilmiş İstem' sınıflar arası kafa karışıklığını neredeyse tamamen ortadan kaldırır.

# Veri Kaynağı Anomalisi: İnsan vs. Makine

Model, yapay zeka üretimi safsataları tespit etmede (0.885), insan uzmanlara (0.827) kıyasla neden daha başarılı?

'YZ Sentetik Veri'  
Sentetik İzler (Makine Dili)



Büyük dil modelleri tarafından üretilen metinler, belirgin yapısal ve üslup izleri taşır. Değerlendirilen model, bu sentetik örüntüleri (stylistic traces) safsatadan bağımsız olarak kolayca tanır.

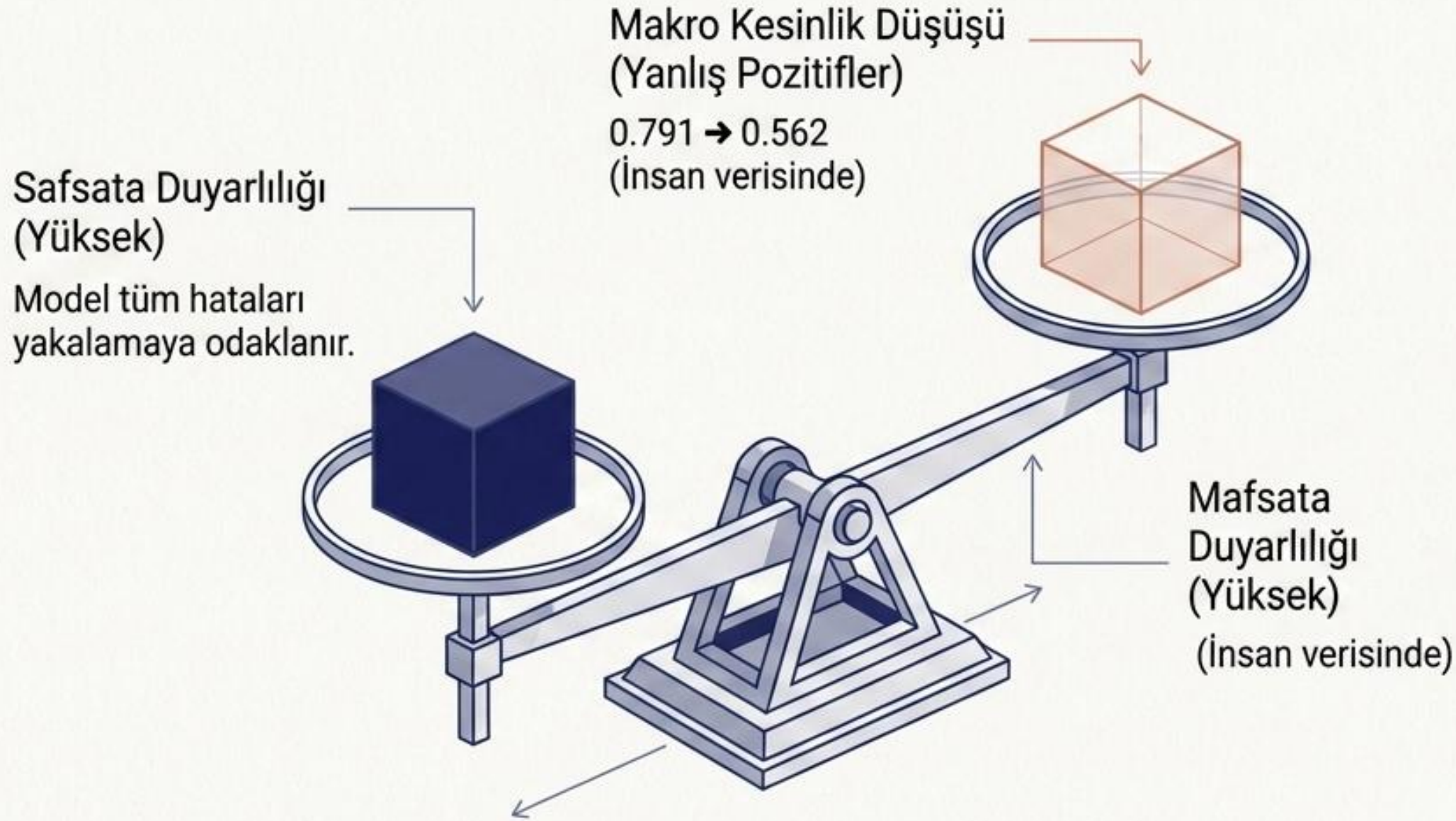
'İnsan Üretimi Veri'  
Anlamsal Belirsizlik (İnsan Dili)



İnsan dili; kültürel bağlam, ironi ve gayriresmi mantıksal kaymalar barındırır. Bu durum, katı algoritmik kural setlerini aşan ve modeli zorlayan bir belirsizlik yaratır.

Bulgu: İnsan dilinin esnekliği, yapılandırılmış sistemlerde dahi akıl yürütme sınırlarını zorlamaya devam etmektedir.

# Ödünleşim (Trade-off): Temkinli Akıl Yürütme



## Fenomen:

İnsan verilerine kurallar eklendiğinde genel doğruluk artarken, Kesinlik (Precision) skorları düşmektedir.

## Açıklama (Cautious Reasoning):

Yapılandırılmış istem, modeli aşırı temkinli davranmaya iter. Model, potansiyel bir safsatayı kaçırmaktansa, geçerli bir argümanı "safsata" olarak etiketleme riskini göze alır. Risk duyarlı sistemler için bu beklenen bir güvenlik refleksidir.

# Sentez: Çevrimdışı Denetimin Formülü

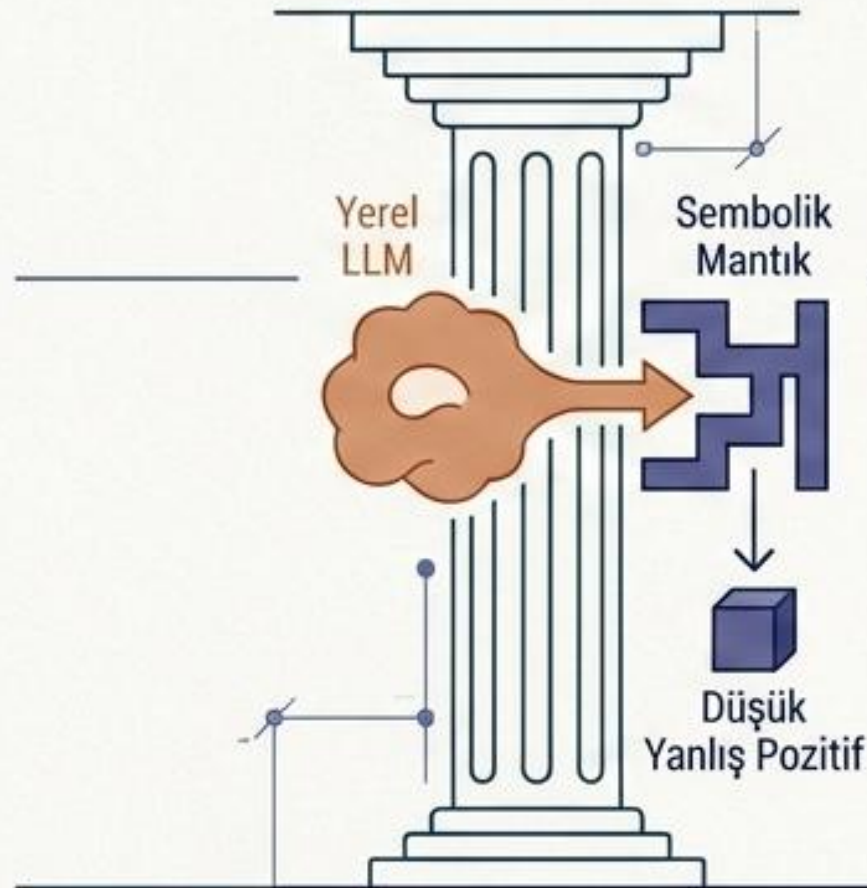
Sistem Derinliği (Sistem 1: Kör → Sistem 2: Bilgilendirilmiş)	<b>Yapısal Uyum</b> Sistem 2 + YZ Veri (En Yüksek Metrikler)	<b>Güvenli Bilişsel Denetim</b> Sistem 2 + İnsan Veri
	<b>Yüzeysel Örüntü Tanıma</b> Sistem 1 + YZ Veri	<b>Kültürel Körlük</b> Sistem 1 + İnsan Veri (En Düşük Başarı)
	<b>Anlamsal Karmaşıklık</b> (YZ Sentetik Veri → İnsan Bağlamsal Veri)	

**Temel Çıkarım:** Yerel dil modelleri, ancak **bağlam** ve **yapılandırılmış kurallar** eşzamanlı olarak sağlandığında güvenli birer mantık denetçisine dönüşebilir.

# Gelecek Ufukları: Bilişsel Sınırları Genişletmek

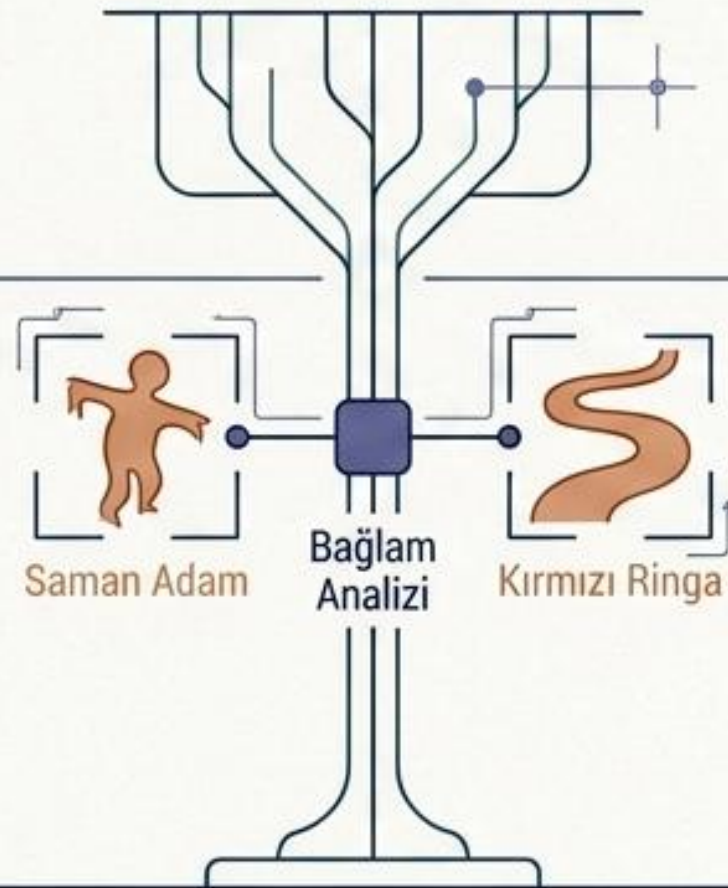
## Nöro-Sembolik Mimariler

Yanlış pozitif oranlarını düşürmek için yerel LLM'lerin sembolik mantık çözücülerıyla (symbolic logic solvers) entegre edilmesi.



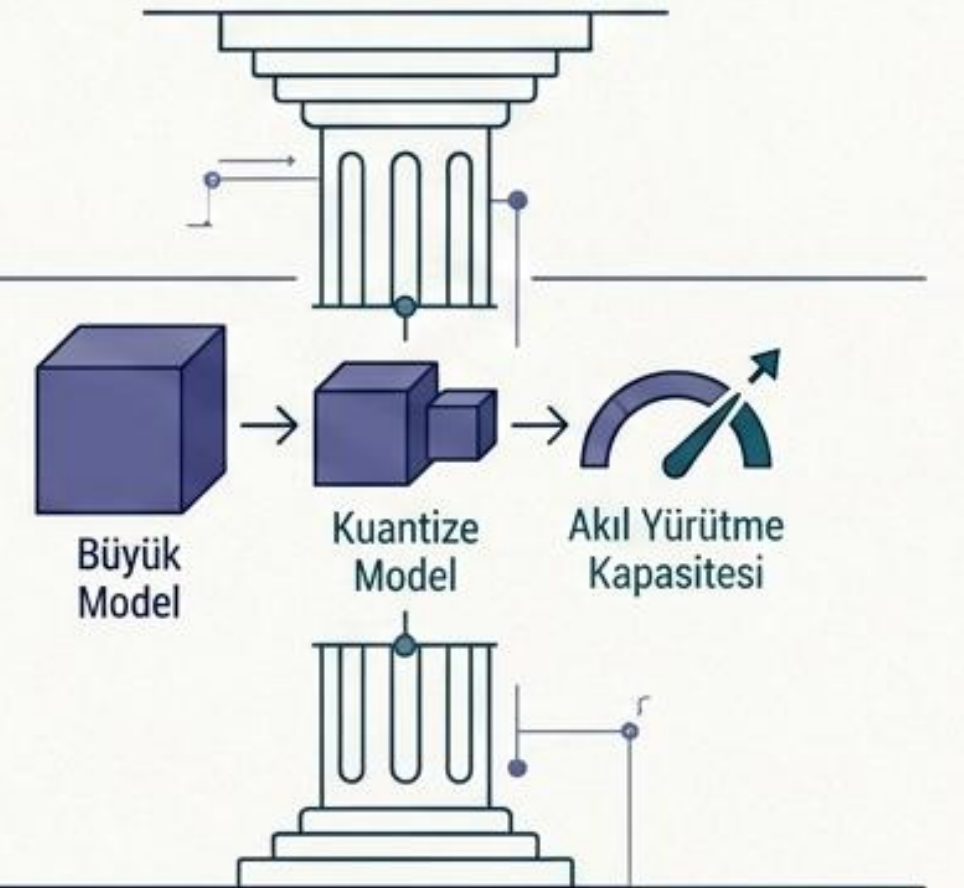
## Genişletilmiş Taksonomi

'Saman Adam' ve 'Kırmızı Ringa' gibi çok daha karmaşık, diyalog bazlı bağlam gerektiren safsataların analizi.



## Kuantizasyon ve Ölçeklendirme

Farklı parametre boyutlarının ve model sıkıştırma tekniklerinin akıl yürütme kapasitesi üzerindeki etkilerinin karşılaştırılması.



Sonuç: Ministral-3-14B ve TR-Fallacy Veri Seti, güvenli, çevrimdışı ve yerel bir eleştirel düşünce yapay zekası için ilk adımı temsil etmektedir.

*Thank  
you*



Kahraman Kostas, PhD  
YYEGM, MEB

[kahramankostas@gmail.com](mailto:kahramankostas@gmail.com)