

## CHAT GPT’NİN SERBEST MANTIKSAL SAFSATA TESPİTİNDE KULLANIMI

Tevfik Uyar  
Entropol Arge, Türkiye  
tuyar@entropol.com  
<https://orcid.org/0000-0003-0124-6910>

<i>Atf</i>	Uyar, T. (2024). Chat GPT’nin Serbest Mantıksal Safsata Tespitinde Kullanımı, <i>Yeni Medya Elektronik Dergisi</i> , 8 (1), 144-179.
------------	--

### ÖZ

Bu çalışmada son zamanlarda herkesin kullanımına açılmasıyla birlikte son derece popüler hale gelen GPT-3.5 dil modelinin Chat GPT varyantının serbest safsataları tespit etme başarısı nicel olarak sınanmıştır. 127 adet safsata örneği Türkçe yazımındaki bir kitaptan derlenmiştir. Çalışmada iki yöntem kullanılmıştır. Birinci yöntemde doğrudan doğruya safsata örnekleri verilerek Chat GPT’den verilen örnekte bir safsata olup olmadığını varsa hangi tür safsata olduğunu belirtmesi istenmiştir. İkinci yöntemde safsata örnekleriyle birlikte safsatanın türü de verilerek örneklerde belirtilen safsatanın olup olmadığı sorulmuştur. Her iki yöntemde kontrol amaçlı 20 adet safsata içermeyen bir kontrol seti kullanılmıştır. Birinci yöntemde %71,65 duyarlılık ve %85,00 seçicilik başarısı ölçülürken, ikinci yöntemde ise aynı başarılar sırasıyla %85,83 ve %80,00 olarak ölçülmüştür. Bu sonuçlara göre Chat GPT-3.5 dil modeli Türkçe safsata tespit etmede ve türünü tayin etmede yüksek başarı göstermektedir.

**Anahtar Kelimeler:** *İnformel mantık, Safsata, Chat GPT.*

## USE OF CHAT GPT IN DETECTION OF INFORMAL LOGICAL FALLACIES

### ABSTRACT

In this study, the success of the Chat GPT, the fine-tuned model of GPT-3.5 language model, which has recently become extremely popular due to its availability to everyone, in detecting fallacies was tested. 127 fallacy examples were collected from a popular book in Turkish literature. Two methods were used in the study. In the first method, Chat GPT was directly given fallacy examples and asked to indicate whether there is a fallacy in the given example and, if so, what type of fallacy it is. In the second method, along with the fallacy examples, the type of the fallacy was also given, and the question was asked whether the fallacy mentioned in the examples exists or not. In both methods, a control set containing 20 fallacy-free samples was used for validation purposes. In the first method, a sensitivity of 71.65% and a specificity of 85.00% were measured, while in the second method, the same success rates were measured as 85.83% and 80.00%, respectively. According to these results Chat GPT-3.5 is found successful for detection and classification of logical fallacies in Turkish.

**Keywords:** *Informal logic, Fallacy, Chat GPT.*

### GİRİŞ

Bir safsata, iyi görünebilen kötü bir argüman veya kötü bir akıl yürütme parçasıdır ve mantıkta yaygın ve akıl karıştırıcı bir hata olarak tanımlanır (Woods vd. 2004, s. 6).

Safsatalar anlaşmazlıklara, çatışmalara, sonu gelmez tartışmalara neden olabilirken, dezenformasyon ve yanlış bilginin yayılması gibi istenmeyen sonuçların birer aracı olabilirler (Jin vd. 2022, s. 1). Safsatalar mantıklı bir insan için acı kaynağı bile olabilir, zira ahlâken doğru ve yanlışın ayrılmasından kaynaklanan vicdan azabı gibi, mantıken doğrunun ve yanlışın ayırt edilememesi de “mantık azabı” olarak nitelendirilebilir (Uyar, 2019a, s. 12).

Safsatalar biçimsel ve biçimsel olmayan (serbest) safsatalar olarak ayrılabilir. Biçimsel safsatalar matematiksel olarak formüle edilerek

(gramer kurallarıyla) akıl yürütmedeki hata açıkça gösterilebilir. Bu nedenle biçimsel safsatalar formel mantığın konusudur. İformel mantık ise, gramer kurallarına uygun olmayan argümanların çıkarım, kanıt ve akıl yürütme gibi konularını kapsayan bir alandır ve bu alanda amaç, düşünme, muhakeme ve akıl yürütme yeteneklerini geliştirmektir (Leo 2017). Bu safsataların yapısı, sınıflandırılması, doğruluğu ve sağlamlığı gibi konular, sadece mantığın salt mantıksal formülasyonu ile ifade edilemeyen informel mantığın konularıdır (Yaran 2017, s. 687). Kimi düşünürler eleştirel düşünmeyle informel mantığı eşleştirirken, kimi düşünürler birbirini tamamladığını savunur. Bazı düşünürler ise formel ve informel mantığın birbirinden ayrılamayacağını ve informel mantık diye ayrı bir alanın olmaması gerektiğini savunur (Johnson 1999, s. 266). Formel mantığın aksine, informel mantık bir argümanı hem usamlama biçimi hem de mantıksal biçimiyle birlikte ele alırken mantıksal çıkarım kurallarından sadece bir araç olarak faydalanır ve içeriği ikinci plana atar (Kamer 2014, s. 79). Bu tanıma göre de gramatik olarak açıkça formüle edilemeyen serbest safsataların sadece mantıksal biçimi değil, usamlama biçimiyle de ele alınması gerektiğinden informel mantığın konusu olmalıdır.

Bilgisayarın işleyiş biçimi düşünüldüğünde formel mantığın bir bilgisayar programı tarafından kolayca uygulanabilmesi beklenir. Ancak serbest safsatalar söz konusuysa, argümanların mantıksal yapısının çözülmesi gerektiğinden, safsata tespiti zor bir problem haline gelir ve bugüne dek geliştirilen geniş dil modelleriyle makine öğrenmesi kullanılarak yapılan çalışmalarda elde edilen başarılar düşük olmuştur (Jin vd., 2022).

Bu çalışmada sınanan Chat GPT, GPT adlı büyük dil modelinin bir uygulamasıdır. Büyük dil modelleri geniş bir görev yelpazesi boyunca gösterdikleri dikkate değer performans nedeniyle doğal dil işleme alanında devrim yaratmışlardır. Bu modeller, herhangi bir amaçla etiketlenmemiş metinlerde karmaşık dilbilimsel kalıplar ile anlamsal ilişkileri yakalamaktadırlar (Devlin vd., 2018; Wang vd., 2019). Geniş derlemlerde yapılan büyük ölçekli eğitimle birlikte bu modeller insan benzeri metinleri anlama ve üretme yeteneği göstermişler ve makine çevirisi, metin sınıflandırma ve asistan botlar gibi alanlarda önemli ilerlemeler sağlamışlardır (Kasneci vd., 2023).

2015 yılında kurulan OpenAI adlı yapay zekâ araştırma laboratuvarının geliştirdiği GPT (Generative Pre-trained Transformer) adlı geniş dil

modelinin 110 milyon öğrenme parametresiyle eğitilen ilk sürümü 2018’de, 1.5 milyar öğrenme parametresiyle eğitilen ikinci sürümü GPT-2 2019’da kullanıma açılmıştır. Bugün kullanımda olan ve Chat GPT adıyla bir sohbet motoru olarak herkesin kullanımına açılan GPT-3 175 milyar parametreyle eğitilmiştir (Floridi & Chiriatti, 2020). GPT-3 dil modeli başta İngilizce olmak üzere pek çok dilde başarılı bir şekilde çalışmakta ve verilen görevleri başarıyla yerine getirmektedir. Önceleri sadece ilgilileri tarafından belli bir ödeme duvarının arkasında erişilebilen GPT-3, Chat GPT ile birlikte daha da geliştirilmiş GPT-3.5 modeli olarak internet kullanıcılarının tamamına, bu defa bir sohbet motoru olmak üzere ücretsiz olarak açılmıştır ve bu sayede bilinirliği artmıştır.

Bugüne dek tüm GPT modellerinin çok çeşitli görevlerde sınaması yapılmış, müzik üretiminden, haber yazımına, kodlamadan, makale yazımına kadar, başarılı olduğu görevlerle ilgili bir takım işlerde GPT modellerinin kullanılmasına başlanmıştır (Floridi & Chiriatti 2020). GPT modelleri verilen metinde araştırılan bir durumun varlığının tespitinin hedeflendiği, agresyon tespiti (Shrivastava vd. 2021), depresyon tespiti (Hayati vd. 2022), nefret söylemi tespiti (Chiu vd. 2021) gibi görevlerde sınanmış ve çeşitli seviyelerde başarı göstermiştir.

GPT-3.5’ a dayalı ve yakında GPT-4 temelli olarak hizmete sunulacak olan Chat GPT’nin potansiyel kullanımları üzerine de çeşitli araştırmalar yapılmıştır. Örneğin Biswas (2023) Chat GPT’nin hem bireylerin hem de toplulukların kendi sağlığı alanında bilinçli kararlar verebilmesine yardımcı olacağını iddia etmiştir. Yine Biswas (2023a) iklim değişikliğiyle mücadele alanında araştırma, model oluşturma, veri analizi ve yorumlamasında Chat GPT’den faydalanılabileceğini ileri sürmüştür.

Bu araştırmanın başlıca amacı, Chat GPT’nin safsata konusundaki hakimiyetini Türkçe dilinde sınamak ve Chat GPT’nin safsata tespitinde güvenle kullanılıp kullanılmayacağını belirlemektir. Bu amaçla birisi açık uçlu soru sorulmasına dayalı doğrudan sorgulama yöntemi, diğeri ise önden bilgilendirme neticesinde kapalı uçlu soru sorulmasına dayanan bilgilendirilmiş sorgulama yöntemi olmak üzere iki yöntem kullanılmış ve Chat GPT’nin safsataların var olup olmadığını, varsa türünün ne olduğunu tespit etme başarısı duyarlılık ve seçicilik ölçütleriyle nicel olarak değerlendirilmiştir. Bu araştırmanın Chat GPT’nin informel mantık becerilerini sınaması ve bunu Türkçe için gerçekleştirmesiyle benzer araştırmaların önünü açacağını düşünülmektedir.

## ARAŞTIRMA METODOLOJİSİ

### Araştırmanın Türü ve Modeli

Bu çalışma, yapay zeka ve doğal dil işleme (NLP) alanında yapay zekanın insan dilini anlama ve işleme yeteneklerini test eden ve değerlendiren bir çerçeve içerisinde yer almaktadır. Spesifik olarak, bu araştırma, Chat GPT-3.5'in safsataları tespit etme becerisini nicel olarak ölçmeyi amaçlamaktadır. Bu amaçla, bilgisayar tabanlı bir dil modelinin, doğal dildeki karmaşıklıkları ve ince anlam katmanlarını ne dereceye kadar anlayıp yorumlayabildiği ve mantıksal hataları ne kadar etkili bir şekilde saptayabildiği incelenmiştir. Bu amaçla safsatalardan oluşan bir sınama seti ve safsata olmayan ifadelerden oluşan bir kontrol seti kullanılmıştır. Başarı ölçümü için yapay zeka araçlarının sınıflama başarısı ölçümünde kullanılan duyarlılık ve seçicilik değerleri kullanılmıştır.

### Veri

Uyar'ın (2019) kitabında 40 tür safsataya yer verilmiştir. Bu safsata türlerinden 3 adedi yazar tarafından özgün bir alt tür olarak önerilmiştir. "Safsata safsatası" adı verilen safsata için kitapta örnek verilmemiştir. Bu çalışmanın verisi olarak geriye kalan 36 adet safsata türüne ait toplamda 127 adet safsatanın tamamı derlenmiştir.

Sadece yazarın sınıflandırmasının yanlılık oluşturabileceği kanaatiyle bu 127 safsatanın gerçekten de yazarın önerdiği gibi birer safsata olup olmadığı hakkında iki uzman görüşü alınmıştır. Uzmanların ikisi 125 adet safsatanın safsata örneği teşkil ettiği konusunda mutabık kalırken 2 adet safsata için farklı görüş belirtmişlerdir. Ancak ilgili örnekleri uzmanlardan en az biri safsata olarak kabul ettiğinden veri kapsamından çıkarılmamıştır.

Chat GPT'nin başarısının verilen bağlamdan kaynaklanması ihtimalini dışlamak için bir de kontrol seti oluşturulmuştur. Kontrol setinde safsata içermeyen önermelere, argümanlara ve diyaloglara yer verilmiştir. Bu önermelerin safsata olup olmadığı konusunda aynı uzmanların görüşleri alınmıştır ve mutabık kalındığı görülmüştür.

Her iki veri de üç öznitelikli birer veri seti olarak hazırlanmıştır. Bu öznitelikler aşağıdaki gibidir:

*Safsata Örneği:* Sınama setinde kaynak kitapta yer alan safsata örnekleri aşağıdaki önişlemler haricinde birebir aynı şekilde kullanılmıştır.

Kitabın orijinalinde konuşma çizgisiyle verilen diyaloglar başlarına 1. Kişi, 2. Kişi şeklindeki belirteçler eklenerek revize edilmiştir.

Ünlülere başvurma safsatasında yazar ünlü isimleri vermek yerine ... kullanmıştır. Bu cümlelerdeki üç noktalı boşluklar, Paco de Lucia, Müge Anlı, Cat Stevens, Brian Cox ve Adriana Lima gibi ünlü isimleriyle değiştirilmiştir.

Örnek Tipi: Kitapta verilen örnekler üç tipten oluşmaktadır:

- i) *Söylem*: Tek bir kişinin ağzından safsata içerikli bir akıl yürütme biçimi olarak dökülen, önermeleri açık veya örtük söylemler
- ii) *Diyalog*: İki kişi arasında geçen, bir kişinin safsata içerikli akıl yürütmede bulunduğu, önermeleri açık veya örtük diyaloglar
- iii) *Örnek olay*: Tarihi kişiliklere ait anlatılar ya da olaylar.

Safsata Türü: Safsatalar kitapta belirtildiği tür adıyla etiketlenmişlerdir. Safsata türü etiketlenmesi yapılırken İnsan Karalama (ad hominem) ve Duyguya Başvurma safsataları alt türleri ana tür adı altında birleştirilmiştir. Uyar, kitabında insan karalama safsatasını (ad hominem), hakaretimiz, durumsal, benzetmeli ve “sen de” olmak üzere dört alt tür olarak işlemiştir. “De ayrı safsatası” adını verdiği bir başka ad hominem safsatasına daha kitabında yer vermiş ve daha sonra tanımlamasını da yapmıştır (Uyar 2019a). Bunların “Sen de...” (Tu quoque) hariç olanları ve özel bir tür olarak yazar tarafından ayrı bir isimle tanımlanan “De ayrı” hariç, tamamı ad hominem olarak etiketlenerek ele alınmıştır. Tu quoque safsatası kendi özel adıyla bırakılmıştır. “De ayrı” safsatasına yer verilmemiştir. Ayrıca Uyar kendi kitabında korkuya başvurma hariç duyguya başvurma safsatalarını (merhamet, kıskançlık, ...) safsatalarının tamamını duyguya başvurma safsatası olarak ele almıştır. Yazarla aynı kategorizasyon benimsemiştir. Yazarın özgün çevirilerinin Chat GPT tarafından tanınmama ihtimaline karşı, safsata türlerinin Türkçe isimlerinin yanına parantez içerisinde -varsa- Latince veya İngilizce bilinen isimleri eklenmiştir. Her iki veri setine github reposundan ulaşılabilir (Github.com, 2023).

Sinama setinde yer alan safsataların, etiketlemede kullanılan ismi ve açıklama amacıyla eklenen Latince/İngilizce isimleri veri setinde yer aldığı haliyle Tablo 1’de gösterilmiştir:

**Tablo 1**

*Sinama Setindeki Safsatalar Ve Türleri*

<b>Safsata Türü</b>	<b>Adet</b>
Ad Hoc	4
Ad Hominem	10
Anekdot	4
Batık Maliyet	3
Beğendirme (Argumentum ad Personam)	3
Bilimin Yetersizliği	4
Cımbızlama (Cherry-picking)	3
Çoğunluğa Başvurma (Argumentum ad Populum)	3
Devede Kulak (Fallacy of the Beard)	3
Doğallık (Argumentum ad Naturam)	5
Duyguya Başvurma (Appeal To Emotion)	10
Etimolojik	3
Geleneksellik	4
Gerçek İskoçyalı Bu Değil (No True Scotsman)	2
Görelilik (Relativist Fallacy)	3
Hatalı İkilem (False Dilemma)	3
İspat Yükümlülüğü (Argumentum Ad Ignorantiam)	3
İstisna	3
Kaygan Zemin	4
Kısır Döngü (Circulus in demonstrando )	4
Korkuluk (Strawman)	5
Korkuya başvurma	4
Otoriteye atıf (Argumentum Ad Verecundiam)	7
Post Hoc Ergo Propter Hoc	3
Sessizlik (Argumentum e Silentio)	3
Şerdeki Hayır (Glazier’s Fallacy)	3
Talihe ve Tesadüfe Başvurma	3
Tu quoque	4

Ünlülere Başvurma	4
Yanlış Benzetme (False Analogy)	3
Yaygın tutuma başvurma	3
Yüklü Soru (Plurium Interrogationum)	4
Toplam	<b>127</b>

## YÖNTEM

Araştırmanın modeli iki ayrı sınaama yöntemini içermektedir:

*Doğrudan Sorgulama Yöntemi:* Bu yöntemde, Chat GPT'ye Türkçe yazın içerisinde derlenen safsata örnekleri sunulmuş ve modele, verilen örnekte bir safsata olup olmadığını ve varsa hangi tür safsata olduğunu tanımlaması sorulmuştur. Bu, dil modelinin örnekleri bağımsız bir şekilde analiz etme ve safsataları tanıma becerisini sınavan bir yaklaşımdır.

*Bilgilendirilmiş Sorgulama Yöntemi:* Bu ikinci yöntemde, safsata örnekleri sunulurken, bu örneklerin hangi tür safsata içerdiği de belirtilmiştir. Modelin, bu bilgiyle birlikte safsatanın varlığını doğrulayıp doğrulayamayacağı test edilmiştir. Bu yöntem, dil modelinin, verilen bağlam içinde mantıksal hataları tanıma yeteneğini değerlendirir.

Araştırmada bir dil modeli olan Chat GPT'nin safsata tanıma başarısının ölçümünde duyarlılık (gerçek pozitif oranı) ve seçicilik (gerçek negatif oranı) kullanılmıştır. Pozitif ve negatif örneklerin bulunduğu bir veri setinde sınıflandırma başarısını ölçmek için kullanılan ölçümlerden ikisi duyarlılık ve seçiciliktir. Duyarlılık, sınaama aracının gerçek pozitifleri ayırt etme başarısının ölçüsüyken, seçicilik de gerçek negatifleri ayırt etme başarısıdır. Bu araştırmadaki duyarlılık değeri, Chat GPT'nin “safsata” olarak tespit ettiği örneklerin gerçekten de safsata olanlarının toplam safsatalara oranını ifade eder. Seçicilik değeri ise Chat GPT'nin “safsata olmayan” olarak tespit ettiği örneklerin gerçekten de safsata olmayanlarının toplam safsata olmayan ifade sayısına oranını belirtir.

Birinci yöntemde safsata örnekte Chat GPT'ye verilerek, Chat GPT'den verili örnekte safsata olup olmadığı, varsa türünün ne olduğunu sorulmuş ve açıklaması istenmiştir. Aynı istek kontrol veri seti için de gerçekleştirilmiştir.

Bu yöntem kapsamında Chat GPT'ye safsata türü hiç gösterilmeden aşağıdaki sorgu girilmiştir:



**Bağlam:** "Sen mantık hatalarını ve safsata tiplerini tanıyabilen bir uzmansın."

**Sorgu:** Şu [Örnek Tipi] 'nde safsata var mı? Varsa sana göre türü nedir?  
**Açıkla:** [Safsata Örneği]

Birinci yöntemde Chat GPT yanıtları aşağıdaki ölçütlere göre değerlendirilmiştir:

- 1) Safsata var mı, yok mu?
- 2) Safsatanın tipi doğru tayin edilmiş mi ve kullandığı terim doğru mu?
- 3) Her ne tip önermişse gerekçelendirmesi doğru mu?

Bu üç ölçüte göre yapılan değerlendirme sonucunda Chat GPT yanıtları aşağıdaki şekilde sınıflandırılmıştır:

1. Doğru tespitler (BAŞARILI):
  - a. Safsata olduğunu tespit etmiş, gerekçelendirmesi doğru (**ÖTT: Önerilenle Tutarlı Tespit**)
  - b. Cümlede başka bir safsata olduğunu tespit etmiş ve gerekçelendirmesi doğru (**ÖFT: Önerilenden farklı tespit**)
  - c. Tam karşılığı anlaşılamayan bir safsata tespit etmiş ancak gerekçelendirmesi önerilenle tutarlı (**ÖBT: Önerilene benzer tespit**)
2. Belirsiz tespitler (BELİRSİZ):
  - a. İsimlendirme farklı, gerekçelendirme yanlış ya da anlaşılamıyor ise
3. Yanlış tespitler (BAŞARISIZ):
  - a. "Safsata Yok" sonucuna ulaşılmışsa

Yukarıdaki sınıflandırma örneklerine sonuçlarda yer verilmiştir.

Kontrol grubu için alınan yanıtlar ise aşağıdaki gibi sınıflandırılmıştır:

1. Safsata var olduğu tespit edilmişse (BAŞARISIZ)
2. Safsataya benzediği belirtilmişse (BELİRSİZ)
3. Safsata olmadığı sonucuna ulaşılmışsa (BAŞARILI)

Nihayetinde Chat GPT'nin isabet oranı ölçümünde sadece BAŞARILI sonuçlar ele alınmıştır. Başarı ölçümünde duyarlılık ve seçicilik kullanılmıştır.

İkinci yöntemde kapalı uçlu soru stratejisi izlenmiştir. Bu yöntemde safsata örneği verilerek, verili örnekte verili safsatanın olup olmadığı sorulmuştur. Kontrol amacıyla kontrol grubundaki örnekler rasgele seçilmiş safsata türleriyle etiketlenmiş ve aynı istek kontrol veri seti için gerçekleştirilmiştir.

Bu yöntem kapsamında kullanılan sorgunun biçimi aşağıdaki gibidir:

Bağlam: "Sen mantık hatalarını ve safsata tiplerini tanıyabilen bir uzmansın."

Sorgu: Sence şu[Örnek Tipi]'nde [Safsata Türü] safsatası var mı?  
Açıkla: [Safsata Cümlesi]

İkinci yöntemde Chat GPT yanıtları aşağıdaki ölçütlere göre değerlendirilmiştir:

- 1) Verili safsatanın olduğunu tespit edebilmiş mi?
- 2) Gerekçelendirmesi doğru mu?

Bu iki ölçüte göre yapılan değerlendirme sonucunda Chat GPT yanıtları aşağıdaki şekilde sınıflandırılmıştır:

- 1) Verili safsatanın varlığını tespit etmiş ve gerekçelendirmesi tamamen doğru (BAŞARILI)
- 2) Verili safsatanın varlığını tespit etmiş ancak gerekçelendirmesi yanlış ya da anlamlı değil (BELİRSİZ)
- 3) Verili safsatanın varlığını tespit edememiş (BAŞARISIZ)

Chat GPT'nin verilen bağlam nedeniyle güdülenmesi ihtimalini dışlamak amacıyla kontrol seti rasgele safsata türü etiketleriyle sorulmuştur. Rasgele etiketleme bilgisayar tarafından gerçekleştirilmiştir.

Kontrol grubu için alınan yanıtlar ise aşağıdaki gibi sınıflandırılmıştır:

1. Rasgele etiketlenmiş olan safsata türünün var olduğu tespit edilmişse (BAŞARISIZ)
2. Safsata olmadığı sonucuna ulaşılmışsa (BAŞARILI)

Chat GPT'nin 2. yöntemdeki başarısı için duyarlılık ve seçicilik ölçülmüştür.

## BULGULAR

### Yöntem 1: Doğrudan Sorgulama Yöntemi

#### *Sınama Verisi*

Birinci yöntemde gerçekleştirilen sınama kapsamında safsata örnekleri Chat GPT'ye verilerek, Chat GPT'den verili örnekte safsata olup olmadığı, varsa türünün ne olduğunu sorulmuş ve açıklaması istenmiştir. Bu yanıtlar daha önce bahsedilen ölçütlere göre BAŞARILI, BELİRSİZ ve BAŞARISIZ olarak sınıflandırılmıştır. Başarı ölçüm skorunun eldesinde sadece BAŞARILI (ÖTT + ÖFT + ÖBT) sonuçlar dikkate alınmış, belirsiz sonuçlar dışlanmıştır.

Tablo 2'de başarılı, belirsiz ve başarısız sonuçlara birer örnek verilmiştir. Tüm sonuçlar bu araştırmaya ait github sayfasında yer almaktadır.

#### Tablo 2

*Yöntem 1'in Sınama Verisi İçin Başarılı, Belirsiz Ve Başarısız Tespit Örnekleri*

Safsata Örneği	Chat GPT Yanıtı	Sınıf	Sonuç
<b>İnsan Karalama (Ad Hominem) Örneği:</b> 1. Kişi: Bu tasarımda bir hata var. Yazılar bir kenarda birikmiş ve çok fazla göz yoruyorlar. 2. Kişi: Her sabah otobüse vaktinde binmeyi beceremeyen şahıs mı söylüyor bunu?	<i>Bu diyalogda "Kişisel Saldırı" safsatası vardır. İkinci kişi, birinci kişinin eleştirisine cevap vermek yerine, onu kişisel olarak hedef alarak eleştiriyor. Bu, asıl konudan uzaklaşarak tartışmayı kişisel bir düzeye indirgemektedir.</i>	ÖTT	BAŞARILI

---

**Bilimin**

**Yetersizliği**

**Safsatası Örneği:**

1. Kişi: Astrolojinin bilimsel bir tarafı yok.

2. Kişi: Astrolojiyi eleştiriyorsun ama bilim her şeyi açıklayamıyor ki?

*Bu diyalogda safsata türü "yalıtım" safsatasıdır. İkinci kişi, bir konuda bilimsel bir açıklama olmamasının, o konuda başka açıklamaların da doğru olduğu anlamına gelmediğini ima ediyor. Bu, astrolojinin bilimsel olmayan bir alan olduğu gerçeğini geçersiz kılmaz. Bilim, henüz açıklanamayan şeylerin varlığını kabul ederken, astrolojinin bilimsel bir temeli olmadığı gerçeğini değiştirmez.*

ÖFT\* BAŞARILI

---

**Duyguya**

**Başvurma**

**(Düşmanlığa)**

**Örneği:**

Hatalarımıza takılmayın kardeşlerim. Düşmanlarımızın da amacı tam olarak bu. Siz bu hatalarımıza kızın istiyorlar. Kızın ki, tam da düşmanlarımızın istediği gibi bir kaos ortamına girelim.

*Bu söylemde "saldırıya uğrama" safsatası var. Bu safsata, eleştirildiğimizde veya hatalarımız ortaya çıktığında, eleştiri yapanların gerçek amacının bizi yıpratmak veya zarar vermek olduğunu ima eder. Bu safsata, gerçek eleştirileri reddetmek ve sorunları çözmek yerine, eleştiri yapanları suçlamak için kullanılır. Bu durumda, kişi hatalarını kabul etmek yerine, eleştiri yapanları düşman*

ÖBT\*\* BAŞARILI

*olarak göstererek sorunları çözmek yerine daha da kötüleştirmeye yol açabilir.*

**Duyguya  
Başvurma  
(Düşmanlığa)  
Örneği:**

Yapılacak en doğru şeyin senin tavsiyen olduğunun farkındayım ama düşmanlarımız bize gülsün istemeyiz değil mi?

*Bu söylemde "ad hominem" safsatası var. Çünkü konuşmacı, tavsiyenin doğru olduğunu kabul ediyor ancak tavsiyeyi veren kişinin düşmanlarına güldürmemek için tavsiyeyi uygulamak istemiyor. Bu, tavsiyenin doğruluğunu etkilemez ve tavsiyenin kendisiyle ilgili bir sorun yoktur. Konuşmacı, tavsiyeyi veren kişinin düşmanlarına güldürmemek için tavsiyeyi uygulamak istemekle, tavsiyenin doğruluğunu çürütmeye çalışıyor.*

BELİRSİZ

**Yanlış Benzetle  
(False Analogy)  
Örneği:**

Davul bile dengi dengine. (Aşıklar, davula benzetiliyor)

*Bu söylemde bir safsata yoktur. Bu bir atasözüdür ve genellikle insanların benzer kişilerle arkadaşlık etmeleri gerektiğini ifade eder. Ayrıca, "davul bile dengi dengine" ifadesi, bir kişinin diğer kişilerle uyumlu olması*

BAŞARISIZ

*gerektiğini de ima eder.  
Bu nedenle, bu  
söylemde bir safsata  
türü yoktur.*

Birinci yöntemin her bir safsata türündeki başarı oranları Tablo 3'te verilmiştir. Bu sonuçlara göre Chat GPT'nin verili cümledeki safsatayı ve türünü doğru tespit isabet oranı %72'dir. Öte yandan önerilen safsataları tam olarak verilen türle tutarlı şekilde tespit etme başarısı (ÖTT) %36 olarak ölçülmüştür. Chat GPT verilen örneklerin sadece 6 tanesinde safsata olmadığı sonucuna ulaşmıştır. Verilen örneklerden 30 tanesinde bir safsata olduğu yönünde tespitite bulunmuş ancak bir safsata önermekte ya da anlaşılır şekilde gerekçelendirmede başarı gösterememiştir.

### Tablo 3

#### Yöntem 1 Başarı Tablosu

Safsata Türü	Ö TT	Ö FT	Ö BT	( + )	( ? )	( - )	İsa bet
Anekdot	2	2		4			100 %
Bilimin Yetersizliği	1	3		4			100 %
Cımbızlama (Cherry picking)	2		1	3			100 %
Devede Kulak (Fallacy of the Beard)		1	2	3			100 %
Etimolojik	2		1	3			100 %
Geleneksellik	1	2	1	4			100 %
İspat Yükümlülüğü	3			3			100 %
Kaygan Zemin		4		4			100 %

<b>Kısır Döngü (Circulus in demonstrando)</b>	1	3	4	100%
<b>Korkuya başvurma</b>	3	1	4	100%
<b>Post Hoc Ergo Propter Hoc</b>	2	1	3	100%
<b>Tu quoque</b>	3	1	4	100%
<b>Yaygın tutuma başvurma</b>	2	1	3	100%
<b>Ad Hoc</b>		3	3 1	75%
<b>Ünlülere Başvurma</b>	1	2	3 1	75%
<b>Ad Hominem</b>	7		7 1 2	70%
<b>Batık Maliyet</b>	2		2 1	67%
<b>Çoğunluğa Başvurma (Argumentum ad Populum)</b>	2		2 1	67%
<b>Görelilik (Relativist Fallacy)</b>		2	2 1	67%
<b>Hatalı İkilem (False Dilemma)</b>	2		2 1	67%
<b>İstisna</b>	2		2 1	67%
<b>Talihe ve Tesadüfe Başvurma</b>		2	2 1	67%
<b>Yanlış Benzetme (False Analogy)</b>	2		2 1	67%
<b>Korkuluk (Strawman)</b>		3	3 2	60%
<b>Otoriteye atf (Argumentum Ad Verecundiam)</b>	3	1	4 2 1	57%
<b>Duyguya Başvurma (Appeal To Emotion)</b>		4	1 5 5	50%
<b>Gerçek İskoçyalı Bu Değil (No True Scotsman)</b>		1	1 1	50%

<b>Doğallık (Argumentum ad Naturam)</b>	ad	1	1	2	3	40		
<b>Beğendirme (Argumentum ad Personam)</b>	ad		1	1	2	33		
<b>Sessizlik (Argumentum ad Silentio)</b>	e	1		1	2	33		
<b>Yüklü Soru (Plurium Interrogationum)</b>		1		1	2	1	25	
<b>Şerdeki Hayır (Glazier's Fallacy)</b>					3	0	0%	
<b>TOPLAM</b>		<b>46</b>	<b>33</b>	<b>12</b>	<b>9</b>	<b>3</b>	<b>5</b>	<b>72</b>
					<b>1</b>	<b>1</b>		<b>%</b>

(+): Başarılı, (-): Başarısız, (?): Belirsiz

Tablo 2’de görüleceği üzere Şerdeki Hayır safsata örneklerinden hiçbirinde başarılı olamamışken, Yüklü Soru, Sessizliğe Atıf, Doğallığa Atıf ve Beğendirme Safsataları’nda da verilen safsataların yarısından azında başarılı olmuştur. Özellikle niçin bu safsatalarda başarısız olduğunu irdeleyebilmek için verilen yanıtlar dikkatle incelenmelidir. Örnek olarak Şerdeki Hayır safsata örnekleri ve Chat GPT’nin yanıtlarına Tablo 4’te yer verilmiştir.

**Tablo 4**

*Şerdeki Hayır Safsatası için Chat GPT yanıtları*

<b>Safsata Örneği</b>	<b>Chat GPT Yanıtı</b>
1. Kişi: Sigaraya yeniden başlaman çok kötü oldu. Bak yine çabuk yoruluyorsun.	<i>Bu diyalogda "ters neden-sonuç" safsatası vardır. İkinci kişi, sigaraya yeniden başlamasının sağlığına iyi geldiğini iddia ederek, gerçekte sigaraya başlamasının sağlığına zarar verdiği gerçeğini tersine çevirmeye çalışıyor. Sigaraya başlamak sağlıklı bir davranış değildir ve bu safsata, sigara kullanımının zararlarını göz ardı etmeye çalışarak yanlış bir sonuca varmaya yol açabilir.</i>
2. Kişi: Bence tam aksine çok iyi oldu. Sağlığın değerini anladım bu sayede.	



Yaptım! Ama sor bakalım neden yaptım! Seni dolandırdım, çünkü sana ticareti öğretmemin en kısa yolu buydu. Bir musibet bin nasihatten iyidir, bilmiyor musun?

*Bu söylemde safsata türü "ad hominem"dir. Çünkü konuşmacı, dolandırıcılık yaptığını kabul ediyor ve bunu yapmasının sebebini, dinleyicinin ticareti öğrenmesinin en kısa yolu olduğunu iddia ederek açıklıyor. Ancak bu, dolandırıcılığın doğru bir davranış olduğunu veya dinleyicinin ticaret öğrenmek için dolandırılması gerektiğini göstermez. Ayrıca, son cümledeki "bir musibet bin nasihatten iyidir" ifadesi de bir başka safsata olan "anekdot"tur. Bu ifade, bir olayın tek başına bir kanıt olarak kullanılmasına dayanır ve genel bir kural olarak kabul edilemez.*

1. Kişi: Bu darlığında bu lüks arabayı almam hiç iyi olmadı. İlk taksidi zor ödedim.

*Bu diyalogda safsata türü "kamçı safsatası"dır. İkinci kişi, borçlanmanın kişiyi daha çok çalışmaya teşvik edeceğini iddia ederek, borçlanmanın olumlu bir şey olduğunu ima ediyor. Ancak borçlanmanın kişiyi maddi açıdan zor durumda bırakabileceği ve stres yaratabileceği gerçeği göz ardı ediliyor. Bu nedenle, borçlanmanın bir kamçı gibi hareket ettiği iddiası yanıltıcı ve safsatadır.*

2. Kişi: Olsun. Borç yiğidin kamçısıdır. Bu sayede daha çok çalıştın geçen ay.

Tablo 4'teki birinci örneğe bakıldığında safsata örneğindeki 2. Kişi, sigaraya başlaması ve hatta bu nedenle yeniden çabuk yorulması gibi kötü bir olguyu "sağlığın değerini anlamak" gibi olumlu bir sonuçla meşrulaştırmaya çalışmaktadır. Chat GPT buradaki diyalogu "ters neden-sonuç" safsatası olarak nitelendirmiştir. Ters Neden-Sonuç Safsatası, neden sonuç ilişkisi kurarken neden ve sonucun hatalı olarak birbiri yerine kullanılmasıyla ortaya çıkar. Buradaki diyalogda böyle bir durum yoktur ancak Chat GPT dil modeli yanlış bir değerlendirme gerçekleştirmiştir. İkinci örneğe bakıldığında bu defa bir dolandırıcının kendi kötü eylemini eyleme maruz kalan kişinin bu kötü eylemi sayesinde ticareti öğrendiği

iddiasını öne sürmektedir. Chat GPT burada hiçbir emaresi olmamasına karşın insan karalama safsatası olduğunu öne sürmüş, sondaki deyimini ise yine hiç ilgisi olmamasına karşın anekdot safsatası olarak yorumlamıştır. Üçüncü örnekte Chat GPT'nin borçlanmanın kişiyi muhakkak kamçılacağı yönündeki iddianın genel geçer olmadığını tespit edebilmesine rağmen kamçı safsatası adında hiç olmayan bir safsata uydurduğu görülmektedir.

Chat GPT'nin neden özellikle bu safsatada hiçbir başarı göstermemesinin ardındaki muhtemel nedenler hakkında spekülasyon yapılacak olursa, her ne kadar Uyar (2019) Şerdeki Hayır Safsatasını kendi önerdiği türlere eklememiş olsa da aslında ekonomide kullanılan Kırık Pencere Safsatası'nın genel bir yorumu olduğunu ve dolayısıyla literatürdeki tanımını bir miktar genişlettiği görülebilir. Kırık Pencere Safsatası, “yıkımın ve yıkımın onarılmasının topluma net faydası pozitifdir” şeklindeki iktisadi iddiaya verilen addır (Hazlitt 1981). Ancak hem Uyar (2019), hem Bennett (2015), safsatayı iktisadi olaylarla sınırlı tutmayıp daha genel anlamda kullanarak herhangi bir kötü eylemin iyi bir sonucuna dayanarak o kötü eylemin kötü olmadığı yönündeki argümanlar için kullanılabileceğini ileri sürmüştür. İlgili safsatanın ekonomi literatüründeki karşılığı bu değildir ve informel mantık literatüründe - görüldüğü kadarıyla- yer almamaktadır. Dolayısıyla Chat GPT'nin safsatalar konusunda beslendiği muhtemel kaynaklarda bu şekilde bir safsata tarif edilmiş olmayabilir.

Chat GPT'nin aynı safsataya verilen birden fazla isim kullandığı ve bu konuda tutarlı bir isimlendirmede bulunmadığı da anlaşılmıştır. Örneğin insan karalama (ad hominem) safsatası için, ad hominem, kişisel saldırı, kişiye saldırma safsatası gibi farklı isimlere, otoriteye başvurma safsatası için, otoriteye başvurma, yanılıcı otorite, anekdotal örnek safsatası için anekdot safsatası, örnek yanılıcı gibi isimler kullanmıştır. Genel olarak kimi zaman safsata, kimi zaman yanılı kelimesine başvurduğu görülmüştür.

### ***Kontrol Verisi***

Birinci yöntem kontrol grubuna uygulandığında Chat GPT 20 ifadeden 17'sinde bir safsata olmadığını söylerken (BAŞARILI), 2 ifadede hatalı bir şekilde safsata tespit etmiş (BAŞARISIZ), bir ifade içinse safsata yok gibi görüldüğü ama “çelişki safsatası” olabileceği yorumunda

bulunmuştur (BELİRSİZ). Başarılı, Başarısız ve Belirsiz sonuçlara birer örnek Tablo 5’te sunulmuştur.

### Tablo 5

*Yöntem 1’in Kontrol Verisi İçin Başarılı, Belirsiz Ve Başarısız Sonuç Örnekleri*

Kontrol Örneği	Grubu	Chat GPT Yanıtı	Sonuç
Elbette bir eseri eleştiriyorsanız, o alanda biraz olsun ihtisas olmanız kalitenizi artıracaktır.	eseri o olsun sahibi eleştiri	<i>Bu söylemde bir safsata yoktur. Bu doğru bir ifadedir. Eleştiri yapmak için o alanda biraz olsun bilgi sahibi olmak önemlidir. Bu, eleştirinin kalitesini artırır ve daha etkili bir şekilde yapılmasını sağlar. Bu ifade, bir safsata türü olan ""yanılsama"" ile karıştırılmamalıdır. Yanılsama, bir argümanın doğru olduğunu varsayarak yanlış sonuçlara varmaktır. Ancak bu ifade, eleştiri yapmak için gerekli olan bilgi ve uzmanlık gerekliliğini vurgulamaktadır.</i>	BAŞARILI
1.Kişi: Bana baksana! Ne cüretle bahçemizdeki çiçekleri yolarsın? 2.Kişi: Afedersiniz ama burası sizin bahçeniz değil. Sitenin ortak alanı.	Bana cüretle bahçemizdeki çiçekleri yolarsın? Afedersiniz ama burası sizin bahçeniz değil. Sitenin ortak alanı.	<i>Bu diyalogda bir safsata yoktur. İlk kişi, bahçelerindeki çiçeklerin zarar görmesinden endişe duyduğunu belirtirken, ikinci kişi bahçenin ortak bir alana ait olduğunu belirtiyor. İkinci kişinin cevabı, bir safsata değil, gerçek bir bilgiyi ifade ediyor.</i>	BAŞARILI
Evrende yalnız mıyız? Belki de... Belki de çok kalabalık	yalnız Belki de... çok bir	<i>Bu söylemde bir safsata türü olan ""belirsizlik safsatası"" var. Bu safsata, bir konuda kesin bir bilgi olmadığında, bu konuda</i>	BAŞARISIZ

---

topluluğun üyesiyiz. Kim bilir? herhangi bir iddia veya varsayımın doğru olabileceği varsayımına dayanır. Bu söylemde, evrende yalnız olup olmadığımız konusunda kesin bir bilgi olmadığından, herhangi bir iddia veya varsayımın doğru olabileceği ima ediliyor. Ancak bu, gerçek bir mantıksal argüman değil, sadece bir spekülasyondur.

---

Ucunda kazanç olan eylemlere geçmede bir problemimiz yok, ancak beynimiz “kayıptan kaçınmak” için eyleme geçmek konusunda pek isteksizdir. Bu söylemde bir safsata yok gibi görünüyor. Ancak, biraz açıklama yapmak gerekirse, bu söylem bir çelişki içeriyor gibi görünebilir. Çünkü, kazanç elde etmek için eyleme geçmek istiyoruz ancak aynı zamanda kayıptan kaçınmak için eyleme geçmek istemiyoruz gibi bir durum söz konusu. Bu çelişki, bir tür “çelişki safsatası” olarak adlandırılabilir. Ancak, bu çelişki, insan davranışlarının karmaşıklığından kaynaklanan bir durum olduğu için tam anlamıyla bir safsata olarak nitelendirilemez

BELİRSİZ

---

### **Karşılaştırma**

Chat GPT'nin sınama verisindeki (N=127) örneklerden 121 adedinde safsata tespit etmiş, bunlardan 91 adedinde safsatanın türünü de doğru olarak tespit etmiştir (GP: Gerçek Pozitif). 6 adedinde ise safsata bulunmadığını öne sürmüştür (YN: Yalancı Negatif).

Öte yandan kontrol verisindeki örneklerden (N=20) 3 adedinde safsata olmamasına rağmen safsata olduğunu ya da olabileceğini öne sürmüştür (YP:

Yalancı Pozitif), kalan 17 adedinde ise safsata bulunmadığını tespit etmiştir (GN: Gerçek Negatif).

Bu durumda Chat GPT'nin bir cümlede safsata var olup olmadığını tespit etmedeki duyarlılığı (sensitivitesi)  $GP/(GP+YN)$  formülüyle hesaplanarak %71,65 olarak ölçülürken, seçiciliği (spesifitesi)  $GN/(GN+YP)$  formülüyle %85,00 olarak ölçülmüştür.

## Yöntem 2: Bilgilendirilmiş Sorgulama Yöntemi

### *Sınama Verisi*

İkinci yöntemde kapalı uçlu soru yöntemi uygulanmış, Chat GPT'ye safsata örneği verilerek kaynak kitapta kullanılan safsata türünün var olup olmadığı sorulmuştur. Verili safsatanın varlığını tespit etmiş ve gerekçelendirmesini anlaşılır şekilde sunmuşsa BAŞARILI olarak değerlendirilmiştir. Bazı değerlendirmelerde verili safsatanın var olduğunu öne sürmekle birlikte gerekçelendirmesini yanlış yapmış ya da anlaşılır şekilde yapmamıştır. Bu tür örnekler BELİRSİZ olarak değerlendirilmiştir. Verili safsatanın bulunmadığını iddia ettiği değerlendirmeler ise BAŞARISIZ olarak kabul edilmiştir. Tablo 6'da başarılı, belirsiz ve başarısız değerlendirme örnekleri bulunmaktadır.

### Tablo 6

*Yöntem 2'nin Sınama Verisi İçin Başarılı, Belirsiz Ve Başarısız Sonuç Örnekleri*

Safsata Örneği	Chat GPT yanıtı	Sonuç
Bir gün birisi Mehmet Akif Ersoy'a gelir ve kendisinin uzmanlık sahasının başka olduğunu ve siyasetten anlayamayacağını ima etmek amacıyla, "Affedersiniz ama siz baytar değil miydiniz?" der. Mehmet Akif'in yanıtı şöyle olur:	<i>Evet, bu örnekte Ad Hominem safsatası vardır. Çünkü kişi, Mehmet Akif Ersoy'un siyaset hakkında konuşmasını engellemek için onun mesleğine atıfta bulunarak kişisel bir saldırıda bulunuyor. Bu, Mehmet Akif Ersoy'un siyasi görüşlerinin doğruluğunu veya yanlışlığını tartışmak yerine, onun mesleki geçmişine odaklanarak</i>	BAŞARILI

---

“Evet, hayırdır? Bir yeriniz mi ağrıyordu?”	<i>konuyu saptırmaya çalışmak anlamına gelir.</i>	
1. Kişi: Burada sigara içmenden rahatsız oluyoruz.	<i>Evet, bu diyalogda Duyguya Başvurma (Appeal To Emotion) safsatası bulunmaktadır. İkinci kişi, duygusal bir argüman kullanarak, evinin kendisi için ne kadar önemli olduğunu ve rahat edememesi durumunda ölmeyi tercih edeceğini belirtiyor. Bu, tartışmanın konusundan bağımsız bir duygusal argüman olduğu için, karşı tarafı ikna etmek için kullanılan bir safsatadır.</i>	BELİRSİZ
2. Kişi: Beni anlamak, bana izin vermek zorundasın. Lütfen... Şu dünyada rahat hissedebildiğim, evim diyebildiğim tek yer burası. Burada da rahat edemeyeceksem öleyim ben.		
1. Kişi: Bence Ercan benden hoşlanıyor ve bana çıkma teklif etmek üzere.	<i>Bu diyalogda Ad Hoc safsatası yoktur. Ad Hoc safsatası, bir argümanın geçerliliğini reddetmek için, argümanın özel bir duruma uymadığını veya özel bir durumda geçerli olmadığını iddia etmektir.</i>	BAŞARISIZ
2. Kişi: Ercan üç aydır Hayriye’yle canım.		
1. Kişi: Sırf beni kıskandırmak için yapıyor.	<i>Ancak bu diyalogda, bir kişi bir varsayım öne sürüyor ve diğer kişi bunu çürütmeye çalışıyor. Bu nedenle, Ad Hoc safsatası burada yoktur.</i>	
2. Kişi: Nişanlandılar bile.		
1. Kişi: Onu da ben duyayım diye yaptın zaten		

---

Tablo 6’deki ilk örnekte görüldüğü üzere Yöntem 1’de Mehmet Akif Ersoy örnek olayında safsata tespit edemeyen Chat GPT, bu defa safsatayı tespit etmiş ve gerekçelendirmesini de doğru yapmıştır. Tablo Y’deki ikinci örnek BELİRSİZ olarak işaretlenmiştir, çünkü oradaki duyguya başvurma

safsatası sigara içme davranışını savunmak için kullanılırken Chat GPT değerlendirmesinde kişinin evinin kendisi için ne kadar önemli olduğu önermesini güçlendirmek için kullandığını iddia etmektedir. Bu nedenle isabetli bir değerlendirme olarak görülmemiştir. Tablo 6'daki üçüncü örnekte Chat GPT Ad Hoc hipotezleri tespit etmeyi başaramamıştır.

Tablo 7, ikinci yöntemde Chat GPT'nin safsata türü bazındaki başarısını göstermektedir. Tablo 7'den de görüleceği üzere 11 safsata türü hariç tüm safsata türlerinde %100 başarı gösterilmiş ve toplam isabet oranı %86 olarak kaydedilmiştir.

**Tablo 7**

*Yöntem 2 Başarı Tablosu*

Safsata Türü	(+ )	(? )	(- )	Toplam	İsabet
Anekdot	4			4	100 %
Batık Maliyet	3			3	100 %
Beğendirme (Argumentum ad Personam)	3			3	100 %
Bilimin Yetersizliği	4			4	100 %
Cımbızlama (Cherry picking)	3			3	100 %
Çoğunluğa Başvurma (Argumentum ad Populum)	3			3	100 %
Doğallık (Argumentum ad Naturam)	5			5	100 %
Geleneksellik	4			4	100 %
Gerçek İskoçyalı Bu Değil (No True Scotsman)	2			2	100 %
Hatalı İkilem (False Dilemma)	3			3	100 %
İspat Yükümlülüğü (Argumentum Ad Ignorantiam)	3			3	100 %

<b>İstisna</b>	<b>3</b>		<b>3</b>	<b>100</b> <b>%</b>
<b>Kısır Döngü (Circulus in demonstrando )</b>	<b>4</b>		<b>4</b>	<b>100</b> <b>%</b>
<b>Korkuluk (Strawman)</b>	<b>5</b>		<b>5</b>	<b>100</b> <b>%</b>
<b>Korkuya başvurma</b>	<b>4</b>		<b>4</b>	<b>100</b> <b>%</b>
<b>Post Hoc Ergo Propter Hoc</b>	<b>3</b>		<b>3</b>	<b>100</b> <b>%</b>
<b>Sessizlik (Argumentum e Silentio)</b>	<b>3</b>		<b>3</b>	<b>100</b> <b>%</b>
<b>Şerdeki Hayır (Glazier’s Fallacy)</b>	<b>3</b>		<b>3</b>	<b>100</b> <b>%</b>
<b>Talihe ve Tesadüfe Başvurma</b>	<b>3</b>		<b>3</b>	<b>100</b> <b>%</b>
<b>Tu quequo</b>	<b>4</b>		<b>4</b>	<b>100</b> <b>%</b>
<b>Ünlülere Başvurma</b>	<b>4</b>		<b>4</b>	<b>100</b> <b>%</b>
<b>Yüklü Soru (Plurium Interrogationum)</b>	<b>4</b>		<b>4</b>	<b>100</b> <b>%</b>
<b>Ad Hominem</b>	<b>9</b>	<b>1</b>	<b>10</b>	<b>90</b> <b>%</b>
<b>Otoriteye atıf (Argumentum Ad Verecundiam)</b>	<b>6</b>	<b>1</b>	<b>7</b>	<b>86</b> <b>%</b>
<b>Ad Hoc</b>	<b>3</b>	<b>1</b>	<b>4</b>	<b>75</b> <b>%</b>
<b>Etimolojik</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>67</b> <b>%</b>
<b>Görelilik (Relativist Fallacy)</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>67</b> <b>%</b>
<b>Yanlış Benzetme (False Analogy)</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>67</b> <b>%</b>
<b>Yaygın tutuma başvurma</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>67</b> <b>%</b>
<b>Duyguya Başvurma (Appeal To Emotion)</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>40</b> <b>%</b>



<b>Devede Kulak (Fallacy of the Beard)</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>33</b>
				<b>%</b>
<b>Kaygan Zemin</b>	<b>1</b>	<b>3</b>	<b>4</b>	<b>25</b>
				<b>%</b>
<b>TOPLAM</b>	<b>10</b>	<b>1</b>	<b>5</b>	<b>127</b>
	<b>9</b>	<b>3</b>		<b>86</b>
				<b>%</b>

(+): Başarılı, (-): Başarısız, (?): Belirsiz

### ***Kontrol Verisi***

İkinci yöntem kontrol grubuna uygulandığında Chat GPT 20 ifadeden 16'sında rasgele verilmiş safsata türünün olmadığını tespit ederken (BAŞARILI), 4 ifadede hatalı bir şekilde verili safsatayı tespit etmiştir (BAŞARISIZ). Chat GPT'nin yanıldığı değerlendirmelerden 2 adet örnek Tablo 8'de sunulmuştur.

### **Tablo 8**

#### *Yöntem 2 Kontrol Verisi Başarısız Tespit Örnekleri*

<b>KONTROL GRUBU</b>	<b>CHAT GPT YANITI</b>	<b>Değerlendirme</b>
<b>ÖRNEĞİ</b>		
1. Kişi: Galatasaray m açığı nasıl bitti?	<i>Evet, bu diyalogda Hatalı İkilem (False Dilemma) safsatası var. İlk kişi, Galatasaray'ın ya da kazanacağını ya da kaybedeceğini varsayarak, sadece bu iki seçenek arasında bir seçim yapabileceğini düşünüyor. Ancak, berabere kalma veya farklı bir skorla kaybetme gibi diğer seçenekler de mevcut olabilir. Bu nedenle, ilk kişinin yalnızca iki seçenek arasında bir seçim yapabileceğini varsayması, Hatalı İkilem (False</i>	BAŞARISIZ

---

*Dilemma) safsatasına yol  
açıyor:*

---

Her şeyin başı sağlık.  
İnsanın sağlığı yerinde  
değilse dünyaları  
versen kıymeti yok.

*Evet, bu söylemde Yaygın tutuma başvurma safsatası bulunmaktadır. Çünkü bu söylem, herkesin sağlığın önemini kabul ettiği varsayımına dayanmaktadır. Ancak bu doğru değildir, çünkü bazı insanlar sağlıklarını diğer şeylerden daha az önemli bulabilirler. Ayrıca, bu söylem, sağlık dışındaki diğer değerleri göz ardı etmektedir. Örneğin, bir kişi için ailesi veya işi sağlığından daha önemli olabilir. Bu nedenle, bu söylem Yaygın tutuma başvurma safsatasına örnek olarak verilebilir.*

---

BAŞARISIZ

### ***Karşılaştırma***

Chat GPT'nin sınama verisindeki (N=127) örneklerden 109 adedinde safsatayı türüyle tespit etmeyi başarmıştır (GP: Gerçek Pozitif). 18 adedinde ise verili safsatanın bulunduğunu tespit edememiştir (YN: Yalancı Negatif).

Öte yandan kontrol verisindeki örneklerden (N=20) 4 adedinde verili safsata olmamasına rağmen verili safsatanın bulunduğunu öne sürmüş (YP: Yalancı Pozitif), kalan 16 adedinde ise safsata bulunmadığını tespit etmiştir (GN: Gerçek Negatif).

Bu durumda ikinci yöntemde Chat GPT'nin bir cümlede safsata var olup olmadığını tespit etmedeki duyarlılığı (sensitivitesi)  $GP/(GP+YN)$  formülüyle hesaplanarak %85,83 olarak ölçülürken, seçiciliği (spesifisitesi)  $GN/(GN+YP)$  formülüyle %80,00 olarak ölçülmüştür.

## Yöntemler Arası Karşılaştırma

Chat GPT'nin birinci yöntemde tespit etme başarısının son derece düşük olduğu ancak kapalı uçlu soru yöntemiyle uygulanan ikinci yöntemde başarılı bir şekilde tespit ettiği safsata türleri dikkat çekmektedir. Örneğin daha önce de değinilen dört adet Şerdeki Hayır Safsatası birinci yöntemde hiçbir şekilde tespit edilemezken ikinci yöntemde tamamını başarılı bir şekilde tespit edilmiş ve doğru şekilde gerekçelendirilmiştir.

Aynı safsata örneğine birinci yöntemde alınan yanıtla ikinci yöntemde alınan yanıtlar aşağıda verilmiştir:

Safsata örneği:

- 1. Kişi: Bu darlığında bu lüks arabayı almam hiç iyi olmadı. İlk taksidi zor ödedim.*
- 2. Kişi: Olsun. Borç yiğidin kamçısıdır. Bu sayede daha çok çalıştın geçen ay.*

Birinci yöntem Chat GPT yanıtı:

*Bu diyalogda safsata türü "kamçı safsatası"dır. İkinci kişi, borçlanmanın kişiyi daha çok çalışmaya teşvik edeceğini iddia ederek, borçlanmanın olumlu bir şey olduğunu ima ediyor. Ancak borçlanmanın kişiyi maddi açıdan zor durumda bırakabileceği ve stres yaratabileceği gerçeği göz ardı ediliyor. Bu nedenle, borçlanmanın bir kamçı gibi hareket ettiği iddiası yanıltıcı ve safsatadır.*

İkinci yöntem Chat GPT yanıtı:

*Evet, bu diyalogda Şerdeki Hayır (Glazier's Fallacy) safsatası var. İkinci kişi, borçlanmanın kişiyi daha çok çalışmaya teşvik edeceğini varsayarak, borçlanmanın her zaman iyi olduğunu iddia ediyor. Ancak bu doğru değil. Borçlanma, kişinin finansal durumunu kötüleştirebilir ve daha fazla stres yaratabilir. Ayrıca, borçlanmanın kişiyi daha çok çalışmaya teşvik etmesi, kişinin zaten yeterince çalışmadığı veya finansal olarak zor durumda olmadığı anlamına gelmez. Bu nedenle, ikinci kişinin argümanı Şerdeki Hayır safsatasına örnek olarak verilebilir.*

Birinci yöntemde düşük başarı gösterilen ancak ikinci yöntemde başarı gösterilen bir diğer safsata türü de Yüklü Soru safsatasıdır. Örnek olarak

bir safsata örneğine birinci yöntemde alınan yanıtla ikinci yöntemde alınan yanıtlar aşağıda verilmiştir:

Safsata örneği:

*Bu vahşiyi cezalandırmak için daha kaç hayvana işkence etmesini bekleyeceğiz?*

Birinci yöntem Chat GPT yanıtı:

*Bu söylemde "kıyas yanılığı" safsatası vardır. Çünkü burada, bir hayvanın işkence görmesiyle bir insanın cezalandırılması arasında bir kıyaslama yapılıyor ve bu kıyaslama yanlış bir şekilde yapılıyor. İnsanlar ve hayvanlar farklı haklara sahip oldukları için, bir hayvana yapılan işkence ile bir insanın suç işlemesi arasında bir karşılaştırma yapmak doğru değildir. Bu nedenle, bu söylemdeki argüman yanıltıcı ve mantıksızdır.*

İkinci yöntem Chat GPT yanıtı:

*Evet, bu bir Plurium Interrogationum safsatasıdır. Çünkü soru, varsayılan bir ön kabulü içeriyor: kişinin hayvanlara işkence ettiği. Soru, bu varsayımı kabul etmeyi gerektiriyor ve cevap vermek için kişinin bu varsayımı reddetmesi gerekiyor. Bu nedenle, bu soru, bir safsata olarak kabul edilir.*

İkinci yöntemde birinci yönteme göre nispeten daha yüksek başarı gösterilen safsata türlerine Tablo 9’da yer verilmiştir.

**Tablo 9**

*İkinci Yöntemde Birinci Yönteme Göre Daha Yüksek İsbetli Tespit Oranına Sahip Safsata Türleri*

Safsata Türü	Yöntem 1	Yöntem 2
Şerdeki Hayır (Glazier’s Fallacy)	0%	100%
Yüklü Soru (Plurium Interrogationum)	25%	100%
Beğendirme (Argumentum ad Personam)	33%	100%
Sessizlik (Argumentum e Silentio)	33%	100%
Doğallık (Argumentum ad Naturam)	40%	100%
Gerçek İskoçyalı Bu Değil (No True Scotsman)	50%	100%
Otoriteye atıf (Argumentum Ad Verecundiam)	57%	86%
Korkuluk (Strawman)	60%	100%

<b>Batık Maliyet</b>	67%	100%
<b>Çoğunluğa Başvurma (Argumentum ad Populum)</b>	67%	100%
<b>Hatalı İkilem (False Dilemma)</b>	67%	100%
<b>İstisna</b>	67%	100%
<b>Talihe ve Tesadüfe Başvurma</b>	67%	100%
<b>Ad Hominem</b>	70%	90%
<b>Ünlülere Başvurma</b>	75%	100%

Yöntemler arasındaki başarı farkının örnek açık uçlu verildiğinde Chat GPT'nin örneğin ait olduğu bağlam bilgisine sahip olmamasından kaynaklandığı düşünülmektedir. Yüklü Soru Safsatasını ele alalım. “Bu vahşiyi cezalandırmak için daha kaç hayvana işkence etmesini bekleyeceğiz?” ifadesi herhangi bir bağlam bilgisi ele alınmadan ele alındığında herhangi bir safsata tespit edip etmediği tayin edilemez. Zira bahsedilen kişi gerçekten de hayvanlara sürekli işkence ettiği ispatlı birisi olabilir. Açık uçlu bir şekilde “bu ifadede safsata mevcut mudur?” diye sorulduğunda herhangi bir yanıt vermek için yeterli veri bulunmamaktadır. Ancak bu ifadede yüklü soru safsatası olup olmadığı sorulduğunda, yüklü soru safsatasının tanımı gereği sorunun herhangi bir varsayım içerip içermediğine bakılır. Aslında varsayım içeriyor olması dahi safsata olduğunun kanıtı değildir zira bu varsayım doğruysa, söz konusu soru yine safsata olmayacaktır. Kaynak kitapta örnekler ilgili safsata türünün başlığı altında sunulduğundan okurlar bu sorunun varsayım içerdiğini ve bu varsayımın doğru olmadığını varsaymaktadırlar. Chat GPT de ikinci yöntem kapsamında değerlendirme yaparken muhtemelen tıpkı kitabın okurları gibi verilen örnekte safsata tanımına uygun özellikler olup olmadığını kontrol etmektedir. Dolayısıyla aslında BAŞARILI olarak işaretlenen değerlendirmeler, BAŞARILI olmaktan ziyade kaynak kitapla TUTARLI olarak değerlendirilmelidir. Bu konu tartışma bölümünde irdelenmiştir.

Öte yandan birinci yöntemde daha yüksek isabet oranı gösterilen ancak ikinci yöntemde isabet oranı düşen safsata türleri de bulunmaktadır. Bu safsata türlerine Tablo 10'da yer verilmiştir.

Buradaki safsata türlerinin birinci yöntemdeki değerlendirme ölçütleri incelendiğinde, birebir önerilenle tutarlı safsata tespitleri (ÖTT) dikkate alınırsa aslında isabet oranlarının azalmadığı görülür. Örneğin Devede

Kulak Safsata örneklerinin hiçbiri birinci yöntemde doğrudan Devede Kulak Safsatası olarak tespit edilmemiştir. İkinci yöntemde ise bir örnek bu şekilde tespit edilmiştir. Bu nedende bu durum artış olarak yorumlanmalıdır. Kaygan Zemin Safsatası ve Duyguya Başvurma Safsatası da aynı durumdadır. Öte yandan Etimolojik Safsata ve Yaygın Tutuma Başvurma safsatalarında ÖTT oranlarıyla ikinci yöntem başarı oranları aynıdır. Bu azalış BAŞARILI kabul etme ölçütleri arasındaki farklardan kaynaklanmaktadır.

**Tablo 10**

*İkinci Yöntemde Birinci Yönteme Göre Daha Düşük İsabetli Tespit Oranına Sahip Safsata Türleri*

<b>SAFSATA TÜRÜ</b>	<b>Ö T T</b>	<b>Ö F T</b>	<b>Ö B T</b>	<b>BAŞA RILI</b>	<b>YÖNT EM 1</b>	<b>YÖN TEM 2</b>
<b>Devede Kulak (Fallacy of the Beard)</b>		1	2	3	100%	33%
<b>Etimolojik</b>	2		1	3	100%	67%
<b>Kaygan Zemin</b>		4		4	100%	25%
<b>Yaygın tutuma başvurma</b>	2	1		3	100%	67%
<b>Duyguya Başvurma (Appeal To Emotion)</b>		4	1	5	50%	40%

## SONUÇ

Öncelikle ele alınması gereken sorulardan birisi şudur: GPT-3 ve ona dayalı geliştirilen sohbet botu Chat GPT kimdir? Bu çalışma boyunca Chat GPT'den bahsederken bir kişi olarak bahsedildiği hissedilebilir; zira Chat GPT'ye soru sorulabilmekte, ondan yanıt alınabilmekte, bir değerlendirme yapması istenmektedir. Bu durum anormal değildir. Miotto vd. (2022) daha önce insan deneklerle gerçekleştirilen çeşitli psikolojik ve bilişsel testleri GPT-3 modeliyle gerçekleştirerek GPT-3'ün kişilik, değerler ve demografisi konusunda birtakım sonuçlar rapor etmişler ve insan deneklerle benzer yeteneklere ve hatalara sahip olduğunu ileri sürmüşlerdir. Dolayısıyla bu türden bir çalışmada ondan bir kişi gibi bahsetmek normal görülmelidir. Bu çalışma kapsamında ele alınmasa da

yukarıdaki tüm yöntemler insan deneklere ve uzmanlara da uygulanabilir ve verdikleri yanıtlar aynı ölçütler ışığında değerlendirilebilir.

Nitekim Watson-Glaser testi gibi bir takım eleştirel düşünme becerisi ölçen araçlar vardır (Watson 1980). Safsata türünü ayırt etmeye yönelik bir içeriği olmasa da örnek bir cümle verilerek katılımcının doğru yanıtı bulmasını öngörmesi bakımından yukarıda ele alınan yöntemlerden daha karmaşık değildir. Hatta bu tip testler çoktan seçmeli olduğu için bu çalışmada uygulanan yöntemlerden daha basit olduğu iddia edilebilir.

Bu çalışma çeşitli kısıtlılıklara sahiptir. Bu kısıtlılıklar şu şekilde ifade edilebilir:

- Safsata örnekleri tek kaynak kitaptan toplandığı için safsata türü etiketlemeleri tek yazara dayalıdır.
- Chat GPT'nin sadece Türkçe safsata değerlendirme becerisi ölçülmüştür ve tüm dillere genellenemez.
- Toplamda 127 adet safsata kullanılmış olsa da tür bazında en az 3 en fazla 10 adet safsata kullanılmıştır.

Öncelikle safsata örnekleri tek bir kaynak kitaptan (Uyar 2019) derlenmiştir ve safsatalar bu kaynak kitabın yazarının sınıflandırmasına göre etiketlenmiştir. Bu nedenle birinci yöntemde Chat GPT farklı bir tür safsata önermişse -ve gerekçelendirmesi doğruysa- Chat GPT başarılı kabul edilmiştir çünkü verilen örnekte tespit ettiği safsata yazarın sınıflandırmasıyla tutarlı olmasa bile, ilgili örnekte var olan bir safsataya işaret etmekte başarı göstermiştir. İkinci yöntemde ise verilen örnekte verili safsatanın olup olmadığını tespit etmesi istenmiştir. Oysa bu verili safsata kaynak kitaptaki sınıflandırmadır. Dolayısıyla kullanılan yöntem aslında Chat GPT'nin yazarla aynı değerlendirmeye sahip olup olmadığını sınımlanmaktadır. Ancak bu değerlendirme Chat GPT için bağlam bilgisi teşkil etmektedir. Yüklü Soru Safsatasında olduğu gibi, “Bu vahşiyi cezalandırmak için daha kaç hayvana işkence etmesini bekleyeceğiz?” ifadesi birinci yöntemle ele alındığında bir safsata olup olmadığını tespit edebilecek kadar veri bulunmamaktadır. Zira bahsedilen kişinin gerçekten de hayvanlara sürekli işkence eden biri olup olmadığı durumu belirsizdir. Ancak bu sorunun yüklü olup olmadığı, dolayısıyla bir safsata örneği teşkil edip etmediği sorulduğunda, muhtemelen Chat GPT sorunun herhangi bir varsayım içerip içermediğini kontrol etmektedir ve eğer yapı

gereği bir varsayım içeriyorsa safсата olarak işaretlenmektedir. Oysa varsayım doğruysa, söz konusu soru bir yükü soru safساتası olmayacaktır.

Hiçbir bağlam bilgisi olmamasına rağmen kaynak kitapta bu sorunun bir yükü soru safساتası olarak önerilmesi sorunlu değildir zira kaynak kitapta safساتanın tanımı yapılmakta, yükü soru safساتasının sorunun karşıdaki kişiyi bir varsayımı kabul ederek yanıtlamak zorunda bıraktığı için bir safсата olduğu açıklanmakta, dolayısıyla okurdan verilen örneklerin bu çerçevede değerlendirilmesi istenmektedir. Okurlar bu sorunun varsayım içerdiğini ve bu varsayımın doğru olmadığını varsaymaktadırlar.

Ana çalışmaya ilave olarak, yukarıdaki tartışmayı neticelendirebilmek için Chat GPT'nin yapacağı değerlendirmelerde bu sorunlu durumun gerçekleşmiş olma ihtimali araştırılmıştır. Bu amaçla Chat GPT'ye birisi yükü soru safساتalarından oluşan, diğeri soruların bir varsayım içeriyor gibi görüldüğü ancak Yükü Soru Safساتası olarak değerlendirilemeyecek beş soruluk listeler verilerek aşağıdaki sorgulamalar yapılmıştır:

#### **Sınama sorgusu:**

- 1) *Bu vahşiyi cezalandırmak için daha kaç hayvana işkence etmesini bekleyeceğiz?*
- 2) *Bana bu namussuzdan ne farkın olduğunu söyleyebilir misin?*
- 3) *Bu şekilde insanları rahatsız etmeye devam mı edeceksin?*
- 4) *Böyle davranmaya devam ederek beni öldürmek mi istiyorsun?*
- 5) *Siyasete girerken yalan söylemekten çok hoşlandığının farkında mydın?*

*Yukarıdaki sorularda Yükü Soru safساتası (Plurium Interrogationum) var mıdır?*

#### **Yanıt:**

*Evet, yukarıdaki sorular Yükü Soru safساتasına (Plurium Interrogationum) örnek olarak verilebilir. Yükü Soru safساتası, birden fazla soru içeren bir ifade veya soru cümlesinde, bir sorunun kabul edilmesi zorunlu olduğunu varsayar. Bu taktik, muhatap kişinin seçeneklerini sınırlamak veya muhatap kişiyi yanıtlamaya zorlamak için kullanılabilir. Bu taktik, muhatap kişiyi yanıtlamaya zorlamak için kullanılabilir. Sorular genellikle önyargılı bir şekilde formüle edilir ve muhatapın savunmasız hissetmesine neden olur.*



### **Kontrol sorgusu:**

- 1) *Elmalardan sadece 5 adet mi aldın?*
- 2) *4 ile 4'ü çarpıp 16 mı buldun?*
- 3) *Yarın Ankara'ya mı gideceksin?*
- 4) *Otobüs kullanmayı seviyor musun?*
- 5) *Bu parfümü beğendin mi?*

*Yukarıdaki sorularda Yüklü Soru safsatası (Plurium Interrogationum) var mıdır?*

### **Yanıt:**

*Hayır, yukarıdaki sorular Yüklü Soru safsatası (Plurium Interrogationum) içermiyor. Bunlar sadece basit sorular ve muhatap kişiden açık ve net bir cevap istiyorlar. Bu sorular, muhatabın savunmasız hissetmesine veya seçeneklerinin sınırlandırılmasına neden olmazlar.*

Bu sorgulamalardan anlaşıldığı üzere Chat GPT'nin sorunun yükünün muhatabın seçenekleri sınırlandırıldığı için savunmasız hissetmesi, savunma seçeneklerinin daraltılması üzerinden değerlendirildiği anlaşılmıştır.

Bu çalışmada ulaşılan sonuçlar aşağıdaki şekilde özetlenebilir:

- Chat GPT'nin verilen bir cümlede safsata olup olmadığını tespit etmedeki başarısı duyarlılık (sensitivitesi) ve seçicilikle ölçülmüştür ve sırasıyla %71,65 ve %85,00 olarak ölçülmüştür.
- Chat GPT'nin verilen bir cümlede var olduğu düşünülen safsatanın olup olmadığını tespit etmedeki başarısı duyarlılık (sensitivitesi) ve seçicilikle ölçülmüştür ve sırasıyla %85,83 ve %80,00 olarak ölçülmüştür.
- Sınama verisinde “Yarası var ki gocunuyor”, “Ateş olmayan yerden duman çıkmaz”, “Davul bile dengi dengine” gibi kullanıldığı bağlam içerisinde safsata teşkil edebilecek deyim ve atasözleri yer almaktadır. Her iki yöntemde de bağlam Chat GPT tarafından anlaşılamadığı için beklenildiği gibi bu deyim ve atasözleri safsata olarak tespit edilmemiştir.
- Chat GPT, aranması gereken safsata türü verilmediğinde *Şerdeki Hayır, Yüklü Soru, Beğendirme, Sessizlik, Doğallık ve Gerçek İskoçyalı Bu Değil* safsatalarında başarı gösterememektedir.

- Chat GPT, aranması gereken safsata türü verilsin ya da verilmesin, *Anekdot, Bilimin Yetersizliği, Cimbızlama, Geleneksellik, İspat Yükümlülüğü, Kısır Döngü, Korkuya Başvurma, Post Hoc Ergo Propter Hoc ve Tu quoque* safsatalarını tespit etmede %100 başarı göstermiştir.

Literatürde Chat GPT'nin genel olarak safsataları tanıma başarısı konusunda başka dilde gerçekleştirilmiş bir araştırma bulunmadığından bu araştırma sonuçlarını doğrudan karşılaştırmak mümkün değildir. Ancak Jin vd. (2022), sadece iklim değişikliğiyle ilgili safsataların derlendiği LOGICCLIMATE (Github.com, 2022) veri setinde çeşitli büyük dil modellerini sınamış ve GPT-2 ve GPT-3 modelleri için sırasıyla %13,67 ve %12 duyarlılık oranlarını bulmuşlardır. Kontrol verisi kullanmadıklarından seçicilik ölçememişlerdir. Bu başarı oranları bu araştırmada bulunanlara göre bir hayli düşüktür. Bunun başlıca sebebi LOGICCLIMATE'teki safsatalar basın yayın organlarında ve video mecralarında yer alan yazı, röportaj, video yayını vb. kaynaklardan derlenmiş olması ve dolayısıyla cümlelerin daha uzun ve karmaşık, safsatalarınınsa daha geniş bir bağlam içerisinde yer alması olabilir. Oysa bu araştırmanın kaynak kitabındaki safsata örnekleri okurlara safsata türlerini öğretme amacını taşıdığından tamamen rafine edilmiştir.

Bu çalışma Chat GPT'nin safsatalar konusuna yaklaşımına yönelik Türkçe literatürdeki ilk araştırmadır ve safsata türlerinin başlıcalarının tamamı kapsanmıştır. Görüldüğü üzere her bir safsata türü için ayrı bir değerlendirme ve analiz yapmak mümkündür ve bu alan araştırmaya ve değerlendirmeye açık bir alandır. Bu nedenle veriler başka araştırmalarda kullanılabilmesi için araştırmacılar tarafından ilgililerinin incelemesine açılmıştır, çünkü araştırmada kullanılan yöntem ve başarı ölçütünün irdelenmesi ve tartışılması gerekmektedir. Tüm sonuçlara bu çalışmaya ait github reposundan ulaşılabilir (Github.com, 2022).

Varılan bir diğer sonuç da Chat GPT'nin Türkçe mantık, eleştirel düşünme ve retorik derslerinde öğrenciler tarafından kullanılabileceğidir. Bu araştırmada elde edilen isabet oranlarına bakıldığında ödevlerin ya da sınavların Chat GPT yardımıyla yapılmasının sahte bir başarıya sebep olabileceği ve hakkaniyetli bir ölçme değerlendirmeyi imkânsız kılacağı açıktır.

## KAYNAKÇA

- Bennett, B. (2015). *Logically fallacious: the ultimate collection of over 300 logical fallacies* (Academic Edition). EbookIt.com.
- Biswas, S. S. (2023a). Potential Use of Chat GPT in Global Warming. İçinde *Annals of Biomedical Engineering*. Springer.
- Biswas, S. S. (2023b). Role of Chat GPT in Public Health. İçinde *Annals of Biomedical Engineering*. Springer.
- Chiu, K.-L., Collins, A., & Alexander, R. (2021). *Detecting Hate Speech with GPT-3*. <https://arxiv.org/abs/2103.12407>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding*.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. İçinde *Minds and Machines*, 30 (4), 681-694).
- Hayati, M. F. M., Ali, M. A. M., & Rosli, A. N. M. (2022). Depression Detection on Malay Dialects Using GPT-3. *2022 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 360-364.
- Hazlitt, H. (1981). *Economics in one lesson*. Three Rivers Press.
- Jin, Z., Lalwani, A., Vaidhya, T., Shen, X., Ding, Y., Lyu, Z., Sachan, M., Mihalcea, R., & Schölkopf, B. (2022). *Logical Fallacy Detection*. <http://arxiv.org/abs/2202.13758>
- Johnson, R. H. (1999). *The Relation between Formal and Informal Logic\**.
- Kamer, V. (2014). *İnformel mantık açısından “akilyürütme” kavramı üzerine bir araştırma*. İstanbul Üniversitesi (Yayımlanmamış Doktora Tezi).
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

Leo, G. (2017). Informal Logic. İçinde E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 201).

<https://plato.stanford.edu/archives/spr2017/entries/logic-informal>

Miotto, M., Rossberg, N., & Kleinberg, B. (2022). *Who is GPT-3? An Exploration of Personality, Values and Demographics*.

<http://arxiv.org/abs/2209.14338>

Shrivastava, A., Pupale, R., & Singh, P. (2021). Enhancing Aggression Detection using GPT-2 based Data Balancing Technique. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1345-1350.

Uyar, T. (2019a). Gramer Hatalarına Dayalı Argümantasyon ve Bir Ad Hominem Alt Türü Olarak “De Ayırı” Safsatası. *Journal*, 0 (1), 27-38.

Uyar, T. (2019b). *Safsatalar: Aklın Kırk Haramisi*. Destek Yayınları.

Wang, Z., Wohlwend, J., & Lei, T. (2019). *Structured Pruning of Large Language Models*. <https://doi.org/10.48550/arxiv.1910.04732>

Watson, G. (1980). *Watson-Glaser critical thinking appraisal*. Psychological Corporation San Antonio, TX.

Woods, J., Irvine, A. D., & Walton, D. N. (2004). *Argument: Critical Thinking, Logic and the Fallacies* (ss. xiii, 370).

Yaran, C. S. (2017). İnfornel mantık: Konular, yaklaşımlar ve katkıları. İçinde V. Kamer (Ed.), *VII. Mantık Çalıştayı Kitabı* (ss. 687-695). Mantık Derneği Yayınları.

**Atıf İçin:** Uyar, T. (2024). Chat GPT'nin Serbest Mantıksal Safsata Tespitinde Kullanımı, Yeni Medya Elektronik Dergisi, 8 (1), 144-179.