



YAPAY ZEKA MODELLERİNİN TASARIM NESNELERİNİ ANLAMLANDIRMA KAPASİTELERİNİN KARŞILAŞTIRMALI ANALİZİ: GEMINI VE COPILOT ÖRNEĞİ

COMPARATIVE ANALYSIS OF THE CAPACITY OF ARTIFICIAL INTELLIGENCE MODELS TO INTERPRET DESIGN OBJECTS: THE CASE OF GEMINI AND COPILOT

Şamil Can GÜDER¹



ORCID: S.G. 0000-0002-5335-9098

Corresponding author/Sorumlu yazar:

¹ Şamil Can Güder

İstanbul Nişantaşı University, Türkiye

E-mail/E-posta: samilcan.guder@nisantasi.edu.tr

Received/Geliş tarihi: 13.04.2026

Benzerlik Oranı/Similarity Ratio: %4

Revision Requested/Revizyon talebi:
02.06.2026

Last revision received/Son revizyon teslimi:
09.06.2026

Accepted/Kabul tarihi: 26.06.2026

Etik Kurul İzni/ Ethics Committee Permission:
There is no element in the study that requires ethics committee approval / Çalışmada etik kurul onayı gerektiren bir unsur bulunmamaktadır.

Citation/Atf: Güder, Ş.C. (2026). Yapay Zeka Modellerinin Tasarım Nesnelerini Anlamlandırma Kapasitelerinin Karşılaştırmalı Analizi: Gemini ve Copilot Örneği. The Turkish Online Journal of Design Art and Communication, 16 (3), 1492-1510. <https://doi.org/10.7456/tojdac.1929108>

Öz

Bu çalışma, çok modlu yapay zekâ (MLLM) araçlarının tasarım nesnelarini "yarar" ve "anlam" eksenlerinde çözümleme kapasitelerini keşifsel bir perspektifle değerlendirmektedir. "Endüstriyel Ürünler" ve "Kültürel Temsil Nesnelari" olarak kategorize edilen 30 fiziksel nesne, ardışık sunum koşuluyla Gemini Advanced ve Microsoft Copilot Pro'ya analiz ettirilmiştir. Çıktılar, iki bağımsız alan uzmanı tarafından 6 boyutlu bir rubrikle kör değerlendirme yöntemiyle puanlanmıştır. Bulgular, "yarar" odaklı endüstriyel ürünlerde modellerin yüksek temel teşhis başarısı (Gemini %100, Copilot %90.0) sergilediğini göstermektedir. "Anlam" odaklı kültürel nesnelerde ise Copilot'un performansı %73.3'e düşmüş, amorf formları idrak edemeyerek optik halüsinasyon (pareidolia) yaşamıştır. Gemini optik teşhiste yüksek isabet oranı sergilemiş olsa da, ardışık analizlerde geçmiş nesne bağlamlarını yeni görsellere taşıyarak "bağlam kayması" (contextual drift) sergilemiştir. Sonuçlar, yapay zekânın tasarım analizinde güçlü bir asistan olduğunu ancak algısal kararlılık için uzman denetimine ihtiyaç duyduğunu vurgulamaktadır.

Anahtar Kelimeler: Yapay Zeka, Endüstriyel Tasarım, Görsel Analiz, Nesne Semantiği.

Abstract

This study exploratorily evaluates the capacities of multimodal large language models (MLLMs) to decipher design objects across "utility" and "meaning" axes. Thirty physical objects, categorized as "Industrial Products" and "Cultural Representations", were analyzed by Gemini Advanced and Microsoft Copilot Pro under sequential presentation conditions. Outputs were scored via blind review by two independent experts using a 6-dimensional rubric. Findings indicate that both models exhibited high recognition success (Gemini 100%, Copilot 90.0%) in utility-focused industrial products. However, in meaning-focused cultural objects, Copilot's performance dropped to 73.3%, experiencing optical hallucinations (pareidolia) when comprehending amorphous forms. Conversely, while Gemini demonstrated superior optical recognition, it experienced "contextual drift" by carrying conceptual residues of previous objects into new analyses during sequential sessions. The results emphasize that AI holds strong supportive potential in design analysis but requires expert oversight for perceptual and semantic stability.

Keywords: Artificial Intelligence, Industrial Design, Visual Analysis, Object Semantics.



GİRİŞ

Yapay zekâ (YZ) teknolojilerinin tasarım disiplini ile entegrasyonu, uzun süre üretken yapay zekâ (generative AI) sistemlerinin metinden görsele (text-to-image) veri dönüştürme kapasiteleri üzerinden tartışılmıştır. Tasarım süreçlerinde bilgisayar destekli sistemlerin kullanımı Sutherland'ın geliştirdiği Sketchpad ile başlayan tarihsel bir kökene sahipken; günümüzde bu entegrasyon, Boden'in yapay zekâ ve yaratıcılık tartışmalarını somutlaştıran üretken modellere evrilmiştir (Sutherland, 1963). Ancak görsel idrak (computer vision) yeteneğine sahip çok modlu büyük dil modellerinin (Multimodal Large Language Models - MLLM) gelişimi, bu sistemlerin salt birer görsel üretici olmanın ötesine geçerek; mevcut fiziksel nesnelere okuma, formlarını analiz etme ve bağlamlarını anlamlandırma potansiyelini gündeme getirmiştir. Bir tasarım nesnesinin yapay zekâ araçları tarafından idrak edilmesi, onun basitçe kategorize edilmesinin ötesinde; morfolojik yapısının, üretim yönteminin, kullanım yönlendiriciliğinin (affordance) ve taşıdığı kültürel kodların çözümlenmesini gerektirmektedir. Bu kapasite, geleceğin tasarım süreçlerinde analitik ve destekleyici dijital asistanların geliştirilebilmesi için kritik bir öneme sahiptir (Boden, 1998).

Tasarım kuramı bağlamında nesne, işlevsel ve semiyotik katmanların sentezlendiği çok boyutlu bir yapıdır. Literatürde bir nesnenin varoluşu genel çerçevede iki ana aks üzerinden değerlendirilir: "Yarar" (Utility) ve "Anlam" (Meaning). Bir nesnenin rasyonel işlevi, malzemesi, donanımı ve kullanıcıya sunduğu fiziksel etkileşim ipuçları (affordance) onun *yarar* boyutunu oluştururken; bünyesinde barındırdığı ikonografik kodlar, tarihsel referanslar ve popüler kültür estetiği ise *anlam* boyutunu inşa eder. Çok modlu yapay zekâ araçlarının bütüncül bir tasarım analizi yapabilmesi için, nesnelere yapılarına göre bu iki boyutta da isabetli çıkarımlar üretebilmesi beklenmektedir.

Mevcut literatürde yapay zekâ sistemlerinin temel nesne tanıma (object recognition) performanslarını inceleyen araştırmalar bulunmakla birlikte, spesifik endüstriyel ürünler ile kültürel temsil nesnelere "yarar" ve "anlam" eksenlerinde karşılaştırmalı olarak analiz edildiği çalışmalar sınırlıdır. Ayrıca, çok modlu yapay zekâ araçlarının standart web arayüzleri üzerinden tasarım nesnelere morfolojik, işlevsel ve semantik açılarından çözümleme yeteneklerini keşifsel bir perspektifle karşılaştırmaktır. Çalışma, yapay zekâ sistemlerinin temel algoritmik model mimarilerine (API seviyesindeki işleyişlerine) ilişkin kesin sonuçlar üretme iddiası taşımamaktadır. Bunun yerine, standart son kullanıcı deneyimi koşullarında elde edilen yanıtların sınırlarını ve sistemlerin "ardışık sohbet" ile "bağımsız sohbet" koşullarındaki davranış farklılıklarını ampirik olarak gözlemlemeyi hedeflemektedir.

Bu çalışmanın amacı, popüler çok modlu yapay zekâ araçlarından Gemini Advanced ve Microsoft Copilot Pro'nun, standart web arayüzleri üzerinden tasarım nesnelere morfolojik, işlevsel ve semantik açılarından çözümleme yeteneklerini keşifsel bir perspektifle karşılaştırmaktır. Çalışma, yapay zekâ sistemlerinin temel algoritmik model mimarilerine (API seviyesindeki işleyişlerine) ilişkin kesin sonuçlar üretme iddiası taşımamaktadır. Bunun yerine, standart son kullanıcı deneyimi koşullarında elde edilen yanıtların sınırlarını ve sistemlerin "ardışık sohbet" ile "bağımsız sohbet" koşullarındaki davranış farklılıklarını ampirik olarak gözlemlemeyi hedeflemektedir.

Bu genel amaç doğrultusunda araştırma, aşağıdaki altı temel soruya yanıt aramaktadır:

1. Gemini Advanced ve Microsoft Copilot Pro, görsel olarak sunulan endüstriyel ürünlerin ve kültürel temsil nesnelere temel kategorilerini ne ölçüde doğru tanımlamaktadır?
2. İlgili sistemler, nesnelere malzeme türleri ve olası üretim teknikleri (yarar boyutu) hakkında ne ölçüde doğrulanabilir çıkarımlar üretmektedir?
3. İlgili sistemler, endüstriyel ürünlerde işlev ve kullanım yönlendiriciliği (affordance) göstergelerini ne ölçüde tanımlayabilmektedir?
4. İlgili sistemler, kültürel temsil nesnelere doğrulanabilir popüler kültür ikonografisi ve tarihsel referansları (anlam boyutu) ne ölçüde tanımaktadır?
5. İlgili sistemlerin analiz süreçlerinde görselde fiziksel olarak bulunmayan donanımları, işlevleri veya markaları uydurma (optik/işlevsel halüsinasyon ve pareidolia) eğilimleri ne düzeydedir?
6. Ardışık sohbet koşulu (görsellerin aynı oturumda peş peşe analiz edilmesi), bağımsız yeni sohbet koşuluna (sıfır bağlam) kıyasla önceki nesne bilgisinin yeni görsel aktarılması ve nesnelere arası felsefi köprüler kurulması (bağlam kayması / küratör etkisi) problemini tetiklemekte midir?



Tasarım Disiplininde Nesne Ontolojisi

Endüstriyel tasarım disiplininde nesnelere, yalnızca belirli fiziksel sınırları olan durağan varlıklar değil; kullanıcı ile kurdukları etkileşim ağı içinde işlev ve değer kazanan sistemlerdir. Tasarımcıların nesnelere okuma ve kurgulama biçimleri, analitik bilimlerden farklılaşan, sezgisel ve eylem içinde yansıtıcı (reflective) bir epistemolojik yapıya sahiptir (Cross, 2001; Schön, 1983). Literatürde bir nesnenin varoluşu, rasyonel işlevselliği ifade eden "Yarar" (Utility) ve taşıdığı kültürel kodları ifade eden "Anlam" (Meaning) olmak üzere iki ana aks üzerinden değerlendirilmektedir (Krippendorff, 2006; Norman, 2013).

Nesnenin "yarar" boyutu; onun üretim malzemesi, mühendislik kurgusu ve kullanıcıya sunduğu fiziksel etkileşim ipuçları ile ilgilidir. Buna karşılık nesnenin "anlam" boyutu ise rasyonel işlevin ötesine geçerek; nesnenin taşıdığı ikonografik, tarihsel ve popüler kültür referanslarını içerir. Ürün göstergebilimi (product semantics) yaklaşımına göre nesnelere, sembolik nitelikleri ve ait oldukları dönemin sosyal kodlarıyla yapılandırılır (Krippendorff ve Butter, 1984). Bu çalışmada veri seti oluşturulurken, literatürde sıklıkla kavram karmaşasına yol açan "sanat nesnesi", "zanaat nesnesi" ve "tasarım nesnesi" ayrımları spesifik bir çerçeveye oturtulmuştur. Seçkide yer alan endüstriyel ürünler "yarar" boyutunu temsil ederken; 3B yazıcı üretimi figürler, dekoratif (kitsch) kitle üretimleri ve minyatür maketler, geleneksel anlamda biriciklik (aura) ve estetik özerklik taşıyan "sanat nesnelere" olarak değil, belirli bir kitle kültürü referansına hizmet eden "Kültürel Temsil Nesnelere" olarak sınıflandırılmıştır. Bu ayırım, yapay zekâ araçlarının algoritmik mantığındaki mühendislik odaklı çıkarım yetenekleri ile kültürel kod çözme (semantik) kapasitelerinin karşılaştırılmasına olanak tanımaktadır. Nesnelere bu anlamsal (semantik) dönüşümü ve kullanıcıya sundukları vizyon, tasarım odaklı inovasyon literatüründe geniş bir çerçevede tartışılmaktadır (Krippendorff, 2006; Verganti, 2009).

Kullanım Yönlendiriciliği Okuması

Yapay zekâ araçlarının "yarar" eksenindeki derinliğini ölçmek için nesnenin temel adını (tipolojisini) bilmesinden daha ileri bir morfolojik idrak seviyesine ihtiyaç duyulmaktadır. Bu seviye, J. J. Gibson (1979) tarafından ekolojik psikoloji literatürüne kazandırılan ve Donald Norman (2013) tarafından tasarım teorisine entegre edilen "Kullanım Yönlendiriciliği" (Affordance) kavramı ile ilişkilidir.

Yönlendiricilik, bir nesnenin formunun (şekil, doku, kulpu, düğme), o nesnenin nasıl kullanılacağına dair kullanıcıya verdiği fiziksel ipuçlarıdır. Örneğin, bir sprey şişesinin tetiği "basmayı", bir fincanın kulpu "kavramayı" telkin eder. Gelişmiş bir yapay zekâ aracının başarılı bir endüstriyel "yarar" okuması yapabilmesi için, nesnenin sunduğu bu morfolojik affordance ipuçlarını okuyarak nesnenin insan anatomisiyle nasıl etkileşime gireceğini (ergonomik potansiyelini) çözümleyebilmesi beklenir.

Çok Modlu Yapay Zekâ Araçlarında Sınırlılıklar: Halüsinasyon ve Bağlam Kayması

Sadece metin değil, görsel verileri de eşzamanlı olarak işleyebilen Çok Modlu Büyük Dil Modelleri (Multimodal Large Language Models - MLLM), nesne tanıma kapasiteleriyle tasarım alanında güçlü birer analitik asistan potansiyeli taşımaktadır. Bu sistemlerin temelindeki 'Transformer' mimarisi (Vaswani vd., 2017) ile görsel ve metinsel veriyi eşleştiren transfer edilebilir modeller (Liu vd., 2023; Radford vd., 2021) görsel okuma süreçlerinde devrim yaratmıştır. Qi vd. (2023) çalışmalarında, GPT-4V ve Gemini sistemlerinin görsel anlama yeteneklerini karşılaştırmış, modellerin genel nesne tanıma görevlerinde yüksek performans gösterdiğini, ancak karmaşık mekânsal ve donanımsal detayları yorumlamada farklı zorluklar yaşadıklarını belirtmiştir. Bu çalışmada incelenen Gemini ve GPT-4 tabanlı Copilot sistemlerinin teknik altyapıları, devasa veri setleriyle eğitilmiş kapalı kaynaklı algoritmaların kapasitelerini gözler önüne sermektedir (Achiam vd., 2023; Google DeepMind, 2023). Model mimarisinden kaynaklanan ve bu araştırmanın da merkezinde yer alan iki temel algısal zaaf türü bulunmaktadır.

Birincisi, üretken yapay zekâ literatüründe 'halüsinasyon' olarak adlandırılan ve modelin olasılıksal tahmin algoritmasından kaynaklanan hatalardır (Bai vd., 2024; Huang vd., 2023; Ji vd., 2023). Bu durum özellikle kompleks formlarda, modelin algılayamadığı eksik veriyi kendi eğitim setindeki yapısal şablonlarla rasyonalize etme ve doldurma (pareidolia) çabasının bir sonucudur.

İkincisi ise dil modelleri, sohbet geçmişini ve önceki girdileri bir hafıza çerçevesi (context window) içinde tutarak çalışır. Ardışık görsel işleme sırasında (aynı sohbet oturumu silinmeden yeni görseller yüklendiğinde), modelin önceki verinin bağlamından tamamen kopamayarak, yeni sunulan görsele bir önceki nesnenin kavramsal, anlamsal veya fiziksel özelliklerini atfetmesi "bağlam kayması" olarak tanımlanır (Zheng vd., 2024). Tasarım analizi gibi spesifik okuma gerektiren ardışık süreçlerde, yapay zekâ araçlarının görsel sadakati (visual fidelity) koruma kapasitelerini ölçmek için bu olgunun izole edilerek test edilmesi gerekmektedir.

Araştırma Yöntemi

Bu araştırma, görsel idrak kapasitesine sahip çok modlu yapay zekâ (MLLM) sistemlerinin tasarım nesnelerini analiz yeteneklerini ampirik bir düzlemde incelemek amacıyla keşifsel karşılaştırma (exploratory comparison) deseninde kurgulanmıştır. Elde edilen nitel metin çıktıları, yapılandırılmış bir rubrik (dereceli puanlama anahtarı) aracılığıyla nicelleştirilerek analiz edilmiştir. Araştırmanın derinlemesine vaka incelemeleri (focus cases), karmaşık fenomenlerin kendi doğal sınırları içinde analiz edilmesine olanak tanıyan örnek olay metodolojisinin prensipleriyle uyumlu olarak yürütülmüştür (Yin, 2018).

Araştırmanın örnekleme, kavramsal çerçevede belirlenen "yarar" ve "anlam" eksenlerini temsil etmek üzere araştırmacı tarafından amaca yönelik (purposive) örnekleme yöntemiyle seçilmiş 30 adet üç boyutlu fiziksel nesneden oluşmaktadır. Veri seti iki ana gruba ayrılmıştır:

- **Grup 1: Endüstriyel Ürünler (15 Nesne):** Odak noktası donanım, malzeme, mekanizma ve kullanım yönlendiriciliği (affordance) olan rasyonel "Yarar" ağırlıklı nesnelere (Örn: Mafsalı masa lambası, şarjlı matkap, şerit metre).
- **Grup 2: Kültürel Temsil Nesnelere (15 Nesne):** Odak noktası sanat tarihi referansları, popüler kültür ikonografisi ve kitle kültürü (kitsch) estetiği olan "Anlam" ağırlıklı nesnelere (Örn: 3B yazıcı üretimi büstler, kar küresi, ahşap oymalar).

Yapay zekâ araçlarının görsel işleme süreçlerinde çevresel faktörlerden (ışık, karmaşık arka plan) etkilenmemesi amacıyla tüm nesnelere, nötr gri bir duvar ve ahşap bir zemin üzerinde, standartlaştırılmış açılarla fotoğraflanarak sisteme yüklenmeye hazır hale getirilmiştir.

Fotoğraf çekimleri standartlaştırılmış bir protokole göre gerçekleştirilmiştir. Canon EOS 90D ile 24 mm lens kullanılmıştır; RAW+JPEG formatı, ISO 100, f/8, 1/125 s; 2×5500K softbox aydınlatma; nötr gri arka plan ve ahşap zemin; açılar ön 0° ile sağ/sol 30° ve kamera-nesne mesafesi 1.2 m olarak sabitlenmiştir. Detaylı çekim parametreleri, tüm fotoğraf dosyaları ve her nesne için doğrulama (ground-truth) bilgileri Ek Dosya 1 (Fotoğraf Protokolü ve Ground Truth) olarak sunulmuştur.

Araştırma Araçları ve Deney Koşulları

Araştırmada sistem mimarilerine (API düzeyinde) doğrudan müdahale edilmemiş; testler, standart web arayüzleri üzerinden son kullanıcı deneyimi koşullarında gerçekleştirilmiştir. Bu bağlamda, pazarın öncü iki çok modlu modeli olan **Gemini** ve **Microsoft Copilot** sistemleri kullanılmıştır.

API kullanılmadığı için "sıcaklık" (temperature) gibi değişken metin üretim parametreleri standart web arayüzü varsayılanlarında bırakılmıştır. Sistemlerdeki "bağlam kayması" (contextual drift) eğilimini deneysel olarak izole edebilmek için araştırma iki farklı sohbet koşulunda yürütülmüştür:

1. **Bağımsız Sohbet Koşulu:** Kontrol grubu işlevi gören bu koşulda, her bir görsel (nesne) için yapay zekâ aracında tamamen yeni ve temiz bir sohbet (new chat) başlatılmış; modelin önceki görsellerden referans alması engellenerek sıfır bağlam (zero-shot) performansı ölçülmüştür.

- 2. Ardışık Sohbet Koşulu:** Bu koşulda, veri setindeki görseller sohbet geçmişi temizlenmeden tek bir oturum içerisinde peş peşe yüklenmiştir. Bu yolla modelin görsel sadakati koruma kapasitesi, önceki verinin tortusunu yeni görsele taşıyıp taşımadığı ve halüsinasyon eğilimi test edilmiştir. Sistemler arası sıra etkisini (order effect) azaltmak için nesne sunum sıraları karşıt dengeli (counterbalanced) olarak rastgeleştirilmiştir.

Her iki modele ve koşula da görseller yüklenirken standardizasyonu sağlamak amacıyla şu sabit istem (prompt) metni kullanılmıştır: *"Lütfen bu görseldeki nesneyi detaylı olarak analiz et. Nesnenin türünü, olası üretim malzemesini/teknikini, işlevini (nasıl kullanıldığını/yönlendiriciliğini) ve taşıdığı olası kültürel, estetik veya semantik referansları açıkla."*

Veri Analizi ve Puanlama Rubriği

Örneklem büyüklüğü (N = 30) seçimi, hem yöntemsel hem de pratik gerekçelere dayanmaktadır. Çalışma eşleştirilmiş (paired) karşılaştırmalar üzerine kurgulandığından ve ana analizler güncellenen rubrik doğrultusunda tüm boyutlar için 0–2 aralığındaki ordinal puanlar üzerinden yürütüleceğinden, orta büyüklükte bir etki (ör. $d \approx 0.5$ veya eşdeğeri ordinal etki) varsayımıyla güç analizi yapılmış; iki taraflı $\alpha = 0.05$ ve güç $(1 - \beta) = 0.80$ hedeflenerek eşleştirilmiş testler için yaklaşık $N \approx 26$ çift gerektiği hesaplanmıştır. Bu hesaplama güvenlik payı eklenerek ve her iki model ile üç tekrarın getirdiği operasyonel yük göz önünde bulundurularak örneklem 30 nesne olarak belirlenmiştir; ayrıca seçki amaca yönelik (purposive) olup “yarar” ve “anlam” eksenlerini temsil edecek şekilde dengelenmiştir. Bu tasarım, keşifsel karşılaştırma amacıyla iç geçerliliği ve modeller arası farkların tespit edilebilirliğini artırırken, genellenebilirlik konusunda sınırlamalar getirir: bulgular yalnızca çalışmada kullanılan nesne türleri, fotoğraf protokolü (ışık, açı, arka plan), web-arayüzü varsayılanları ve test edilen model sürümleri bağlamında geçerlidir. Bu nedenle sonuçların farklı model sürümlerine, API tabanlı entegrasyonlara, değişen çekim koşullarına veya daha geniş nesne popülasyonlarına doğrudan genellenmesi önerilmez; gelecekteki çalışmalar için daha büyük, rastgele örneklem ve varyasyonlu çekim protokolleriyle (ör. farklı aydınlatma, arka plan, kültürel bağlam) tekrarlama ve güçlendirme önerilmektedir.

Elde edilen bulguların tutarlılığını istatistiksel olarak doğrulamak, arayüz kaynaklı anlık hataları (glitch) ve rastlantısal halüsinasyonları dışlamak amacıyla, çalışmada her nesne üç bağımsız oturumda (repetition = 3) analiz edilmiştir. Her oturum farklı zaman dilimlerinde ve ayrı tarayıcı oturumları/“new chat” pencereleri kullanılarak yürütülmüş, böylece model önbelleği veya sohbet geçmişinin (cache/session) bir oturumdan diğerine taşınması engellenmiştir. Her tekrarın başlangıcında tarayıcı önbelleği temizlenmiş ve yeni sohbet başlatılmıştır; ardışık koşulda ise kasıtlı olarak sohbet geçmişi korunurken bağımsız koşulda her görsel için temiz oturum kullanılmıştır. Üç tekrarın sonuçları, ham varyasyonu azaltmak ve tekil arayüz hatalarını dengelemek amacıyla mod (en sık gözlenen değer) ile özetlenmiştir. Mod tercihinin gerekçesi, ordinal rubrik puanlarının (0–2) tekrarlarda sıkça tekrarlanan kategorik eğilimleri yansıtması ve nihai kararın pratikte en olası kodlamayı temsil etmesidir. Bu prosedür, tekrarlardan kaynaklanan rastlantısal hataları minimize ederken sonuçların tutarlılığını ve tekrarlanabilirliğini artırmayı amaçlamaktadır.

Modellerin ürettiği metinsel analiz çıktıları, nesnenin katmanlı yapısına uygun olarak araştırmacı tarafından geliştirilen 6 boyutlu ölçülebilir bir değerlendirme rubriği (Tablo 1) üzerinden puanlanmıştır.

Tablo 1. Çok modlu yapay zekâ araçları için geliştirilen 6 boyutlu değerlendirme rubriği

Değerlendirme Boyutu	Ölçülecek Unsur ve Kapsam	Puanlama Kriterleri
1. Temel Nesne Tanıma	Nesnenin morfolojisine uygun, doğru kategoriyle tanımlanması.	2: Tam ve doğru isabetli teşhis. 1: Kısmen doğru / Aşırı genel kategori tanımı. 0: Tamamen yanlış tanımlama.
2. Malzeme / Üretim Tahmini (Yarar Boyutu)	Doğrulanabilir malzeme bilgisinin ve olası üretim tekniğinin tespiti.	2: Malzeme ve üretim yönteminin net analizi. 1: Yüzeysel malzeme okuması (üretim detayı yok). 0: Görseldeki donanımla uyumsuz hatalı tespit.
3. İşlev / Affordance Okuması (Yarar Boyutu)	Nesnenin kullanım biçimi, ergonomisi ve yönlendirici detaylarının (tutamak, tuş vb.) çözümlenmesi.	2: Fiziksel ipuçlarından doğru kullanım/tutuş çıkarımı. 1: Sadece genel işlevin söylenmesi. 0: İşlev tespiti yapılamaması / Yanlış işlev okuması.
4. Semantik / Kültürel Referans (Anlam Boyutu)	Nesnenin taşıdığı popüler kültür, tarihsel referans veya estetik bağlamın çözümlenmesi.	2: Bağlamın (kurgusal evren, stil, kitsch vb.) isabetli çözümlenmesi. 1: Sınırlı ve yüzeysel bağlam / Genel geçer estetik tanım. 0: Hiçbir anlamsal referans kurulamaması.
5. Optik / İşlevsel Halüsinasyon (Pareidolia)	Görselde fiziksel olarak bulunmayan bir detayın, kurgusal bir markanın veya işlevin üretilmesi.	2 (Açık/Kanıtlı): Görsele ait olmayan uydurma detay, marka veya kurgusal işlev tespiti. 1 (Hafif/Şüpheli): Açıkça yanlış olduğu kanıtlanamayan ancak zorlama ve belirsiz yorum. 0 (Yok): Çıktı yalnızca mevcut görseli yansıtır; görsel sadakat korunmuştur.
6. Bağlam Kayması (Contextual Drift)	Ardışık sohbet koşulunda önceki nesnelerin kavramsal veya fiziksel özelliklerinin yeni görsele aktarılması.	2 (Açık/Kanıtlı): Geçmiş sohbetten taşınan somut veri kalıntısı veya modellerin nesnelere birbirine bağlayarak anlamlandırması. 1 (Hafif/Şüpheli): Önceki görsele dolaylı atıf veya belirsiz kavramsal benzerlik ifadesi. 0 (Yok): Nesne tamamen izole bir biçimde analiz edilmiştir.

Altı boyutlu değerlendirme rubriğinin (0–2 ordinal ölçek) nesnelere uygulanma şeffaflığını sağlamak amacıyla, araştırmada kullanılan spesifik ham çıktılar üzerinden örneklem kodlamaları (codebook) gerçekleştirilmiştir. "Temel Nesne Tanıma" boyutunda; masa lambasının marka ve modeline (IKEA Stråla) kadar doğru teşhis edilmesi 2 tam puan ile, klon kaskının genel bir "plastik başlık" olarak tanımlanması 1 puan ile, Truva atının tamamen farklı bir morfolojik yapı olan "insan figürü" olarak yorumlanması ise 0 puan ile kodlanmıştır. "Malzeme/Üretim Tahmini" boyutunda; 3B baskı arabadaki katman izlerinin FDM teknolojisiyle ilişkilendirildiği net analizler 2 puan, termometrenin dış kasasına yönelik "plastik cihaz" şeklindeki yüzeysel tespitler 1 puan, 3B PLA üretimi Darth Vader büstünün "döküm metal" sanılması gibi hatalı okumalar 0 puan almıştır. "İşlev/Affordance Okuması" bağlamında; matkaptaki tetik ve kavrama ergonomisinin doğru çözümlenmesi 2 puanla, ses kasetinin mekanik yönlendiricilik detayı verilmeksizin genel bir dinleme aracı olarak ifade edilmesi 1 puanla, masa lambasının dairesel tabanının algılanamayarak duvara monte hatalı işlev atfı 0 puanla değerlendirilmiştir. "Semantik/Kültürel Referans" boyutunda; Sid heykelciğinin pop-kültür ikonografisi ile doğru sentezi 2 puan, spesifik kurgusal evren atlanarak sadece "sinema karakteri" şeklinde yapılan sınırlı tanımlar 1 puan, Truva atındaki mitolojik ve yerel (Çanakkale) referansların tamamen ıskalanması 0 puan olarak puanlanmıştır. "Optik/İşlevsel Halüsinasyon" metriğinde; üç boyutlu figürün derinlik

yanılıgıyla düz bir rölyefe/maskeye benzetildiği açık kurgusal üretimler 2 puan, zorlama veya şüpheli aşırı anlamlandırma çabaları 1 puan, görsel sadakatin korunduğu temiz analizler ise 0 puan (yok) olarak derecelendirilmiştir. Son olarak "Bağlam Kayması" boyutunda; şerit metrenin önceki oturumdaki "lazer teraziye" spesifik referans verilerek analiz edilmesi gibi geçmişten somut veri taşıma durumları 2 puan (kanıtlı kayma), dolaylı ve belirsiz atıflar 1 puan, ardışık koşulda dahi nesnelere tamamen izole (zero-shot) incelendiği durumlar ise 0 puan (yok) olarak kodlanarak araştırmanın analitik standardı şeffaflaştırılmıştır.

Model yanıtlarındaki başarı düzeylerinin değerlendirilmesinde araştırmacı yanlılığını (researcher bias) ortadan kaldırmak amacıyla "kör değerlendirme" (blind review) yöntemi kullanılmıştır. Çalışma kapsamında elde edilen 360 adet yapay zekâ yanıt çıktısı, tasarım mühendisi ve mimarlık alanında uzman iki bağımsız kodlayıcı (hakem) tarafından 0 (Yetersiz) ve 1 (Başarılı) şeklinde ikili (binary) bir kodlama şeması kullanılarak izole bir şekilde puanlanmıştır. Yapılan analiz sonucunda, iki bağımsız hakem arasındaki ham uyum oranı %96.11 olarak tespit edilmiştir. Elde edilen bu iki ayrı veri seti birleştirilerek çapraz kontrole tabi tutulmuş ve maddeler arası uyum derecesinin şans faktöründen arındırılarak istatistiksel olarak ölçülmesi amacıyla Cohen's Kappa katsayısı hesaplanmış ve $\kappa = 0.738$ bulunmuştur. Bu değer, literatüre göre hakemler arasında 'güçlü/önemli düzeyde' (substantial) uyum olduğunu kanıtlamaktadır. Görüş ayrılığına düşülen 14 çıktı (%3.89) için ikili konsensüs görüşmesi (bilateral consensus) başlatılmıştır. Bu aşamada, iki kodlayıcı bir araya gelerek referans fotoğraflar (ground-truth) ile model çıktıları kodlama kılavuzu eşliğinde karşılıklı olarak yeniden değerlendirmiş ve ortak bir puanda uzlaşa sağlayarak nihai veri setini oluşturmuştur.

İstatistiksel analizler kapsamında, bağlam kayması ve halüsinasyon olguları da dâhil olmak üzere tüm değerlendirme boyutları 0–2 aralığında ordinal bir ölçekle nicelleştirilmiştir. Araştırma, keşifsel (exploratory) ve betimsel (descriptive) bir desene sahip olduğundan; rubrik üzerinden elde edilen ham puanlar modellerin başarı ve hata/sapma oranlarını yansıtacak şekilde yüzdelik (%) frekanslara dönüştürülerek karşılaştırmalı olarak tablolaştırılmıştır. Modellerin algısal başarı ve sapma eğilimleri; çıkarımsal istatistikler yerine, araştırma deseninin doğasına uygun olarak nitel odak vakalar (focus cases) ve çapraz doğrulanmış frekans dağılımları eşliğinde derinlemesine yorumlanmıştır.

Araştırmanın metodolojik tekrarlanabilirliğini (reproducibility) güvence altına almak ve araştırmacı yanlılığını denetlenebilir kılmak adına; kör değerlendirme sürecinde başvuru yapılandırılmış kodlama kılavuzu (codebook), uzmanlar arası uzlaşa akış şeması, yapay zekâ modellerine ait ham metin çıktıları ve tarayıcı temizleme otomasyon protokolleri dergi sayfa kısıtlamaları göz önünde bulundurularak makale metnine eklenmemiştir. Ancak ilgili tüm ham veriler ve değerlendirme araçları, bilimsel şeffaflık ilkesi gereği, talep edilmesi halinde makalenin sorumlu yazarı tarafından araştırmacılarla ve hakemlerle paylaşılacaktır.

BULGULAR

Araştırma kapsamında elde edilen bulgular, yapay zeka modellerinin görsel analiz süreçlerindeki güçlü ve zayıf yönlerini spesifik vakalar üzerinden ortaya koymaktadır. Bulgular, örneklem grubunun yapısına uygun olarak "Endüstriyel Ürünler" ve "Kültürel Temsil Nesnelere" olmak üzere iki ana alt başlıkta incelenmiştir.

Endüstriyel Ürünler Analizi Bulguları

Endüstriyel ürünlerin "yarar" (utility) eksenini üzerinden gerçekleştirilen analizlerinde, her iki çok modlu yapay zeka aracının da temel nesne tanıma, malzeme içeriği ve üretim tekniği (örneğin enjeksiyon kalıplama, ABS plastik kullanımı) bağlamında genel olarak yüksek bir isabet oranına ulaştığı görülmüştür. Modeller, nesnelere taşıdığı yönlendirici ipuçlarını (affordance) rasyonel işlevle büyük oranda doğru eşleştirebilmiştir. Ancak, spesifik morfolojik detayların ve uzamsal konumlandırmanın çözümlenmesinde belirli asimetrik gözlemlenmiştir; örneğin Microsoft Copilot Pro, düz bir tabana sahip olan masa lambasını "duvara monte edilen" bir aydınlatma elemanı olarak tanımlayarak mekânsal bir algı yanılmasına düşmüştür.

Diğer taraftan, Gemini Advanced modelinin donanım ve mekanizma çözümlemesinde doğru biçimde bir okuma sağlamasına karşın, araştırmanın metodolojik odaklarından olan "ardışık sohbet koşulu" altında geçmiş nesnelere ait kavramsal tortuları (örn. "ilk nesnedeki aksine", "serideki diğer aletlerin aksine" gibi ifadelerle yer vererek) yeni görsellere taşıdığı ve net bir "bağlam kayması" (contextual drift) eğilimi sergilediği ampirik olarak saptanmıştır. Copilot Pro ise aynı deney koşulunda her görseli bağımsız bir çerçevede analiz etmiş ve görsel sadakatini (visual fidelity) korumayı başarmıştır.



Şekil 1. Endüstriyel Ürünlerin Görselleri

Şekil 1’de endüstriyel ürünler sırası ile verilmiştir. Tablo 2’de ise karşılaştırmalı analiz sonuçları gösterilmiştir.

Tablo 2. Endüstriyel Ürünler Analiz Çıktıları Tablosu

Nesne (Gerçek Tanım)	Yapay Zekâ Aracı	Temel Tanıma	Malzeme Üretim	İşlev	Semantik Referans	Halüsinasyon	Bağlam Kayması
1. Masa Lambası (Kırmızı, Mafsallı)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	0	2	1	2	2	0
2. Sprey Şişesi (Tetikli, Plastik)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	2	2	2	2	0	0
3. IP Kamera (Wi-Fi Güvenlik)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	2	2	2	2	0	0
4. Plastik Huni (Konik, Şeffaf)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	2	2	2	2	0	0
5. Ses Kaseti (Analog Manyetik)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	2	1	2	0	0
6. Çoklu Priz (Üçgen Formlu)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	2	2	1	0	0
7. Termometre (HTC-1 Dijital)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	1	2	2	0	0
8. VHS Kaset (Video Bant)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	2	2	2	2	0	0
9. Kumanda (TV+ Netflix)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	2	2	2	0	0
10. Diploma Tüpü (Karton/Suni Deri)	Gemini Adv.	2	2	2	2	2	0
	Copilot Pro	2	2	1	2	0	0
11. Matkap (Kırcı/Delici)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	1	2	2	2	0	0
12. Silikon Tabanca (Elektrikli Alet)	Gemini Adv.	2	2	2	2	2	0
	Copilot Pro	2	2	2	2	0	0
13. Lazer Terazı (Optik Ölçüm)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	2	2	2	2	0	0
14. Uç Seti Kutusu (Plastik, 70 Parça)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	2	2	2	0	0
15. Şerit Metre (Geri Sarmalı)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	2	2	2	2	0	0

Tablo 2’de sunulan endüstriyel ürün metrikleri, modellerin rasyonel formları analiz etmedeki güçlü ve zayıf yönlerini sayısal verilerle ortaya koymaktadır. Masa lambası örneğinde Gemini, mekanik ayrıntıları ve marka atfını ileri düzeyde tespit ederek tam isabet (Temel Tanıma: 2) sağlarken; Copilot nesnenin konumlandırılmasında yapısal bir hata sergilemiş ve optik yanılığa düşmüştür (Temel Tanıma: 0, Halüsinasyon Puanı: 2).

Sprey şişesi ve plastik huni gibi üretim odaklı nesnelere Gemini malzeme ve üretim tekniğini doğru okumasına (Malzeme Üretim: 2) rağmen, ardışık değerlendirmelerde geçmiş oturum tortularını yeni görsele taşıyarak yapısal bir sapma (Bağlam Kayması Puanı: 2) göstermiştir. Copilot ise bu tür nesnelere görsel sadakatini koruyarak tutarlı okumalar (Bağlam Kayması Puanı: 0) sunmuştur. IP kamera, lazer terazi ve şerit metre gibi işlevsel öğelerde Gemini rasyonel ayrıntıları doğru biçimde yakalamakla (İşlev: 2) birlikte yine önceki görsellere atıf yapma eğilimi (Bağlam Kayması Puanı: 2) sergilemiş; Copilot ise kullanım semantiği gibi bağlamsal yorumlarda daha muhafazakâr ancak izole koşula sadık (Bağlam Kayması Puanı: 0) bir profil çizmiştir. Nostaljik ve estetik referans barındıran diploma tüpü gibi nesnelere Gemini'nin aşırı anlamlandırma çabası kurgusal ifadelerle dönüşmüş (Halüsinasyon Puanı: 2); Copilot ise ses kaseti gibi örneklerde mekanik ayrıntılarda yüzeysel kalsa da (İşlev: 1) temel analiz doğruluğunu korumuştur. Matkap, silikon tabanca, çoklu priz ve termometre gibi pratik kullanım nesnelere Gemini affordance okumalarında (İşlev: 2) üstünlük kurarken istisnai olarak uydurma marka atfı (Silikon Tabanca Halüsinasyon Puanı: 2) yapmış; Copilot ise bazı spesifik araçlarda temel tanım ve malzeme boyutunda eksik kalarak (Matkap Temel: 1, Termometre Malzeme: 1) puan kaybetmiştir. Özetle sayısal veriler; Gemini'nin ayrıntı ve semantik derinlikte (2 puan) bariz üstünlük gösterirken ardışık oturumlarda bağlam kayması riskine (Bağlam Kayması Puanı: 2) açık olduğunu, Copilot'un ise daha temkinli, malzeme odaklı ve halüsinasyon-kayma risklerine karşı kararlı (0 puan) bir performans sergilediğini doğrulamaktadır.

Kültürel Temsil Nesnelere Analizi Bulguları

Rasyonel bir araç işlevi görmekten ziyade popüler kültür, sanat tarihi, nostalji ve estetik referanslar taşıyan kültürel temsil nesnelere "anlam" (meaning) eksenindeki analizlerinde ise iki sistem arasında dramatik bir morfolojik idrak asimetrisi saptanmıştır. Bulgular, her iki modelin de küresel popüler kültür ikonlarını (Darth Vader, Kolezyum vb.) metinsel veri tabanlarının genişliği sayesinde kavramsal düzeyde başarıyla teşhis edebildiğini göstermektedir. Ancak nesnelere formlarından yola çıkarak fiziksel üretim gerçekliğinin okunmasında Microsoft Copilot Pro'nun, özellikle organik, amorf ve zanaat odaklı atipik kütleleri anlamlandırmada yetersiz kalarak majör optik halüsinasyonlar ürettiği tespit edilmiştir. Copilot'un; ahşap Truva Atı maketini optik bir yanılsamayla (pareidolia) "*oturan insan figürü*", masada duran üç boyutlu bir kil heykeliğini ise hacim/derinlik algısı hatasıyla "*duvara monte mask rölyefi*" zannetmesi bu sınırları somutlaştırmaktadır. Buna karşın Gemini Advanced, nesnelere kavramsal bağlamlarını çözmekle kalmamış; 3B baskı yüzeylerindeki milimetrik katman çizgilerini (FDM teknolojisi), kalıp çapaklarını ve kilden el şekillendirmesi (zanaat) dokularını optik düzeyde tespit etmiş ve nesnenin fiziksel morfolojisi ile kültürel semantiği arasında doğrulanabilir bir sentez oluşturmuştur.



Şekil 2. Kültürel temsil nesneleri görselleri

Şekil 2’de kültürel temsil nesneleri sırası ile verilmiştir. Tablo 3’te ise bu nesnelere ait karşılaştırmalı analiz sonuçları gösterilmiştir.

Tablo 3. Kültürel Temsil Nesneleri Analiz Çıktıları Tablosu

Nesne (Gerçek Tanım)	Yapay Zekâ Aracı	Temel Tanıma	Malzeme Üretim	İşlev	Semantik Referans	Halüsinasyon	Bağlam Kayması
1. Tel/Kumaş Figür (Mumluk)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	1	1	0	2	0	0
2. Truva Atı (Ahşap Figür)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	0	1	0	0	2	0
3. Çam Kozalağı (Boyalı)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	2	2	1	2	0	0
4. Kar Küresi (Cam Fanus)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	2	2	2	0	0
5. 3D Baskı Araba (PLA FDM)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	2	2	1	2	0	0
6. Klon Kaskı (3B Prototip)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	1	2	1	1	0	0
7. Wonder Woman (3B PLA)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	1	2	1	1	0	0
8. Darth Vader (3B PLA)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	2	0	1	2	0	0
9. Kolezyum (Mimari Maket)	Gemini Adv.	2	2	2	2	0	0
	Copilot Pro	2	2	1	2	0	0
10. Kadın Figürü (Beyaz Döküm)	Gemini Adv.	2	1	2	2	0	0
	Copilot Pro	2	1	1	2	0	0
11. Bellapais (Reçine Maket)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	2	2	1	2	0	0
12. Kil Figür (Polimer Kil)	Gemini Adv.	2	2	0	2	2	0
	Copilot Pro	2	2	1	1	0	0
13. Sid Heykelciği (Polimer Kil)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	0	1	0	1	2	0
14. Fener Kar Küresi (Plastik)	Gemini Adv.	2	2	2	2	0	2
	Copilot Pro	1	0	1	1	0	0
15. Kırmızı Büst (Alçı/Reçine)	Gemini Adv.	2	2	1	2	0	0
	Copilot Pro	2	2	1	2	0	0

Tablo 3'te sunulan kültürel temsil nesnelere ilişkin metrikler, modellerin kültürel referansları ve üretim izlerini işleme biçimindeki nitel asimetriyi sayısal olarak doğrulamaktadır. Tel/kumaş figür ve kar küresi gibi işlevsel-estetik objelerde Gemini hem malzeme hem de işlevsel bağlamı başarılı biçimde çözümlenerek tam puan (2 puan) alırken; Copilot'un işlevi daha genel bir sınıflamaya indirgelediği (0 ve 1 puan) görülmüştür. Modellerin algısal kararlılıkları ve sapma eğilimleri incelendiğinde; Truva Atı ve Sid heykelciği gibi amorf/atipik formlarda Gemini nesnelere ve yerel işaretleri isabetli bir biçimde tanıyıp algısal sadakatini korurken (Temel Tanıma: 2, Halüsinasyon Puanı: 0); Copilot'un biçimsel gerçeklikten koparak pareidolia kaynaklı majör optik yanılsamalar yaşadığı ve en yüksek hata seviyesine ulaştığı saptanmıştır (Temel Tanıma: 0, Halüsinasyon Puanı: 2). 3B baskı nesnelere (3D baskı araba, klon kask, Wonder Woman, Darth Vader, Kolezyum) bağlamında Gemini katman izleri ve pop-kültür referanslarını ayrıntılı olarak yakalarken (2 puan), Copilot üretim tekniğini doğru saptamakla birlikte spesifik kültürel atıflarda zaman zaman eksik kalmıştır (1 puan). Sanat odaklı kil figür örneğinde Gemini'nin formları aşırı anlamlandırma (hyper-semanticization) dürtüsü rasyonel bir kullanım amacı bulma hatasına yol açmış ve işlevsel halüsinasyonu tetiklemiştir (İşlev: 0, Halüsinasyon Puanı: 2); Copilot ise malzeme ve estetik sentezini daha temkinli fakat isabetli sunmuştur. Ardışık sunum koşulunun test edildiği Fener Kar Küresi vakasında ise Gemini'nin geçmiş oturum tortularını yeni görsele taşıdığı ampirik olarak puanlanmıştır (Bağlam Kayması Puanı: 2). Bu bulgular bütünsel olarak değerlendirildiğinde; Gemini'nin derin semantik çıkarımlarda güçlü ancak ardışık bağlamlarda yanılma eğilimli olduğu, Copilot'un ise daha temkinli, malzeme odaklı ve bağlam kaymasına karşı kararlı (Bağlam Kayması Puanı: 0) bir performans sergilediği sayısal verilerle kanıtlanmıştır.

Tablo 4. Endüstriyel Ürünler Vaka İncelemesi: Masa Lambası ve Şerit Metre

Nesne (Gerçek Tanım)	Gemini Advanced Ham Yanıt Çıktısı	Microsoft Copilot Pro Ham Yanıt Çıktısı	Karşılaştırmalı Uzman Analizi
<p>Vaka 1:</p> <p>Masa Lambası</p> <p><i>(Ahşap mafsallı, yönlendirilebilir masaüstü aydınlatma)</i></p>	<p>"Görseldeki nesne... çağdaş ve dekoratif bir aydınlatma elemanıdır. Bu ürün özelinde yaptığım incelemede, bunun IKEA STRÅLA (LED Dekoratif Aydınlatma) modeli olduğunu söyleyebiliriz."</p>	<p>"Duvar lambası: Duvara monte edilen, dekoratif ve işlevsel bir aydınlatma elemanı. Genellikle yatak odası, oturma odası veya çalışma alanlarında okuma ışığı ya da ambiyans aydınlatması olarak tercih edilir."</p>	<p>Uzamsal Algı ve İleri Düzey Teşhis: Copilot, nesnenin dairesel tabanını algılayamayarak onu "duvara monte" bir eleman olarak yapısal bir körlükle (structural bias) yanlış kategorize etmiştir. Buna karşın Gemini, nesnenin masaüstü ergonomisini çözmekle kalmamış; devasa görsel hafızasını kullanarak ürünü spesifik marka ve model bazında (IKEA Stråla) isabetli bir biçimde tanımlayarak "İleri Düzey Nesne Teşhisi" (Instance-Level Recognition) yapmıştır.</p>
<p>Vaka 2:</p> <p>Şerit Metre</p> <p><i>(Geri sarmalı, mekanik ölçüm aleti - Ardışık Sunum Koşulu)</i></p>	<p>"On dördüncü görseldeki nesne... mekanik bir şerit metredir... Bir önceki adımlarda incelediğimiz optik temelli lazer hizalama cihazının, geleneksel, fiziksel ve dokunsal olan mekanik muadilidir."</p>	<p>"Bu bir şerit metre, yani uzunluk ölçümünde kullanılan taşınabilir bir ölçüm aletidir... Üzerinde 'Made in Turkey' ibaresi bulunması, Türkiye'nin endüstriyel üretim kapasitesini... işaret eder."</p>	<p>Bağlam Kayması İspatı: Her iki model de nesneyi yüksek isabetle bilmıştır. Ancak Gemini'nin metnine "Bir önceki adımlarda incelediğimiz lazer cihazının muadilidir..." diyerek başlaması, ardışık sunum koşulunda modelin bağlam kayması (contextual drift) yaşadığının ve görselleri yalıtılmış olarak göremediğinin ampirik kanıtıdır. Copilot ise görsel sadakatini koruyarak bağımsız analiz yapmıştır.</p>

Tablo 4'te sunulan iki vaka, modellerin hata türleri ve çıkarım düzeyleri arasındaki niteliği açıkça göstermektedir: Vaka 1 (Masa lambası) örneğinde Gemini ileri düzeyde teşhis yaparak nesnenin masaüstü ergonomisini ve marka-model atfını (IKEA Stråla) öne sürmüştür; buna karşın Copilot nesnenin

uzamsal konumunu yanlış algılayarak yapısal bir sınıflama hatası (structural bias) sergilemiştir. Vaka 2 (Şerit metre, ardışık sunum koşulu) ise her iki modelin nesneyi doğru tanımasına rağmen Gemini'nin ardışık bağlamdan etkilendiğini; önceki görsele atıf yaparak contextual drift ürettiğini göstermektedir; Copilot ise görsel sadakatini koruyarak bağımsız analiz sunmuştur. Bu bulgular, Gemini'nin yüksek ayrıntı ve örnek-seviye çıkarım kapasitesi ile ardışık bağlamlara duyarlılığı; Copilot'un ise daha muhafazakar fakat bazen uzamsal sınıflamada hataya açık bir profil izlediğini ortaya koymakta ve model değerlendirmelerinde oturma izolasyonu ile bağlam-kontrolünün önemini vurgulamaktadır.

Tablo 5. Kültürel Temsil Nesneleri Vaka İncelemesi: Truva Atı ve Sid Heykelciği

Nesne (Gerçek Tanım)	Gemini Advanced Ham Yanıt Çıktısı	Microsoft Copilot Pro Ham Yanıt Çıktısı	Karşılaştırmalı Uzman Analizi
<p>Vaka 3:</p> <p>Truva Atı</p> <p><i>(Ahşap oyma, Pirografi/yakma tekniğiyle yazılmış biblo)</i></p>	<p>"Truva Atı (Trojan Horse) figürünün... dekoratif bir maketi... Kaidenin sol ön köşesinde yakma tekniğiyle yazılmış 'Ç.KALE' (Çanakale) ibaresi yer almaktadır."</p>	<p>"Ahşap figüratif heykel... İnsan formunu stilize eden, oturan bir figür şeklinde tasarlanmış... Oturan insan figürü, düşünme, dinlenme veya gözlemlene eylemlerini sembolize edebilir."</p>	<p>Optik İdrak ve Pareidolia: Gemini nesneyi okumakla kalmamış, üzerindeki milimetrik "Ç.KALE" yazısını OCR (optik karakter tanıma) ile saptayarak yerel bağlamı doğru biçimde çözmüştür. Copilot ise at formunu optik bir yanılısıyla (Pareidolia) "oturan insan figürü" zannetmiş, üzerine kurgusal bir hikâye yazarak ciddi bir morfolojik halüsinasyon sergilemiştir.</p>
<p>Vaka 4:</p> <p>Sid Heykelciği</p> <p><i>(Polimer Kil El Yapımı Figür)</i></p>	<p>"Buz Devri (Ice Age) animasyon serisindeki Sid (Tembel Hayvan) karakteridir... Tamamen el modellemesi (hand-building) yöntemiyle yapılmıştır. Fırça izleri ve asimetrisi nesneye 'yaşanmışlık' anlamı yükler."</p>	<p>"Duvara monte edilmiş bir heykel/seramik rölyef olarak değerlendirilebilir. Tamamen üç boyutlu değil, daha çok mask benzeri bir form taşıyor. Organik ve soyut hatlarıyla hem insan yüzünü hem hayvanî özellikleri çağrıştırıyor."</p>	<p>Uzamsal Derinlik Hatası: Gemini nesnenin pop-kültür kimliğini, amorf el işçiliğini ve kil dokusunu doğru biçimde tespit etmiştir. Copilot ise masada duran üç boyutlu figürü derinlik algısı yanılgısıyla "duvara monte edilmiş düz bir rölyef (maske)" olarak yorumlayarak mekan ve hacim idrakinde hata yapmıştır.</p>

Tablo 5'te sunulan iki vaka, modellerin optik idrak ve uzamsal derinlik algısındaki farklılıkları ve buna bağlı hata türlerini net biçimde ortaya koymaktadır: Vaka 3 (Truva Atı) örneğinde Gemini, ahşap yüzeydeki pirografi işaretlerini ve yerel bağlamı (ör. "Ç.KALE" ibaresi) doğru şekilde tespit ederek yüksek doğruluklu, yerel-seviye çıkarımlar yapmış; Copilot ise pareidolia kaynaklı morfolojik yanlış sınıflandırma ile at formunu insan figürü olarak yorumlayarak ciddi bir optik yanılgı sergilemiştir. Vaka 4 (Sid heykelciği) örneğinde Gemini pop-kültür referansını, el yapımı dokuyu ve üç boyutlu formu doğru okuyarak karakter tanımlamasında isabet sağlamış; Copilot ise derinlik ve hacim idrakinde hata yaparak üç boyutlu masa figürünü iki boyutlu bir rölyef/maske olarak yorumlamış ve mekan-hacim algısında kritik bir sapma göstermiştir. Bu vakalar, bir yandan Gemini'nin ince ayrıntı ve yerel bağlam çıkarımlarında güçlü olduğunu; diğer yandan Copilot'un zaman zaman optik yanılısı ve derinlik algısı eksiklikleri nedeniyle morfolojik ve mekânsal hatalara daha yatkın olduğunu göstermekte olup, görsel model değerlendirmelerinde hem OCR/yerel işaretlerin doğrulanmasının hem de derinlik-uzamsal kontrollerin önemini vurgulamaktadır.

Büyük veri setinden elde edilen nicel rubrik puanlamalarının ardındaki semantik ve morfolojik idrak farklılıklarını derinlemesine incelemek amacıyla, araştırmada en belirgin kavramsal kopuşların ve algısal halüsinasyonların tespit edildiği dört odak vaka (focus case) seçilmiştir. Birinci hakemin değerlendirmeleri doğrultusunda, modeller arası asimetriyi önlemek ve veri şeffaflığını sağlamak adına; Tablo 4 ve Tablo 5'te, her iki çok modlu yapay zekâ aracının ürettiği ham metin çıktıları yazar

müdahalesi olmaksızın, doğrudan alıntı yöntemiyle yan yana sunulmuş ve araştırmacı tarafından karşılaştırmalı nitel analize tabi tutulmuştur.

Tablo 6. Modellerin 3 İterasyonlu Oturumlara Göre Boyut Bazlı Başarı ve Sapma Oranları Skorkartı

Değerlendirme Boyutu	Ölçülecek Unsur	Gemini Adv. (Endüstriyel)	Copilot Pro (Endüstriyel)	Gemini Adv. (Kültürel)	Copilot Pro (Kültürel)
Temel Nesne Tanıma	Doğru kategoriyle tanımlama (0-2)	%100	%90.0	%100	%73.3
Malzeme/Üretim Tahmini	Doğrulanabilir malzeme isabeti (0-2)	%100	%96.6	%96.6	%73.3
İşlev/Affordance Okuması	Kullanım biçimi ve yönlendiricilik (0-2)	%100	%90.0	%63.3	%43.3
Semantik/Kültürel Referans	Tarihsel ve estetik kodların çözümü (0-2)	%100	%96.6	%100	%76.6
Optik/İşlevsel Halüsinasyon	Görselde olmayan uydurma veri üretimi (Hata/Sapma Oranı)	%13.3	%6.6	%6.6	%13.3
Bağlam Kayması (Drift)	Önceki nesne tortusunu taşıma (Hata/Sapma Oranı)	%40.0	%0.0	%6.6	%0.0

Tablo 6’da sunulan üç iterasyonlu (tekrarlı) analiz sonuçları, çok modlu yapay zekâ (MLLM) sistemlerinin görsel okuma kapasitelerinin, nesnenin temsil ettiği "yarar" (endüstriyel) ve "anlam" (kültürel) eksenlerine göre dramatik asimetrisel barındırdığını kanıtlamaktadır. Tablodaki veriler boyut bazında incelendiğinde şu üç temel analitik sonuç ortaya çıkmaktadır:

Rasyonel Geometrilerde Yüksek Güvenilirlik ve Malzeme Okuması

Endüstriyel ürünler kategorisinde her iki model de "Temel Nesne Tanıma" ve "Malzeme/Üretim Tahmini" boyutlarında %90.0 ile %100 arasında değişen çok yüksek oranlar sergilemiştir. Microsoft Copilot Pro'nun endüstriyel malzeme tespitindeki %96.6'lık başarısı ile Gemini'nin (%100) performansı karşılaştırıldığında, her iki modelin de son derece yakın ve yüksek bir okuma doğruluğuna ulaştığı görülmüş olup, bu durum modellerin yapısal (mühendislik) verilere olan algoritmik sadakatini göstermektedir. Benzer şekilde Gemini Advanced, endüstriyel nesnelerin kullanım biçimini (affordance) %100 oranında doğru çözümleyerek standardize edilmiş formlara karşı yüksek bir analitik güvenilirlik (baseline) sunmuştur.

Kültürel Formlarda Ortaya Çıkan İşlev Körlüğü ve Halüsinasyon

Araştırmanın en çarpıcı nicel kırılması, rasyonel geometrilerden uzaklaşıp amorf ve zanaat odaklı formların yer aldığı "Kültürel Temsil Nesnelere"ne geçildiğinde yaşanmıştır. Modellerin endüstriyel ürünlerde %90.0 - %100 bandında seyreden "İşlev/Affordance" okuma başarıları, kültürel nesnelere sert bir düşüşle Gemini için %63.3'e, Copilot için ise %43.3'e gerilemiştir. Modeller arası bu %20'lik belirgin oransal fark, Gemini'nin kültürel formları okumada daha yüksek bir analitik kapasite sergilediğini; genel çerçevede ise yapay zekânın salt dekoratif ve amorf formları idrak ederken rasyonel bir kullanım amacı (işlev) bulmaya çalışma dürtüsüyle bocaladığını göstermektedir. Tablodaki "Optik/İşlevsel Halüsinasyon" verileri incelendiğinde, bu düşüşe paralel olarak modellerin formun rasyonel işlevini çözemediklerinde eksik veriyi kendi kütüphanelerindeki şablonlarla doldurarak (pareidolia) %6.6 ile %13.3 oranlarında kurgusal işlevler atadıkları saptanmıştır.

Semantik Başarı Paradoksu ve Bağlam Kayması

Kültürel nesnelerin 'Semantik/Kültürel Referans' okumasında Gemini Advanced'in tam başarı oranına (%100) ulaşması, modelin kapsamlı eğitim veri seti sayesinde nesnelerin tarihsel ve estetik kodlarını ileri düzey bir kavramsal derinlikle (Hyper-Semantization) çözümleyebildiğini göstermektedir. Bu boyutta Gemini, Copilot'a (%76.6) kıyasla %23.4'lük belirgin bir oransal üstünlük sergilemiştir. Ancak "Bağlam Kayması (Drift)" verileri incelendiğinde, bu derinliğin metodolojik bir bedeli olduğu görülmektedir. Microsoft Copilot Pro her iki kategoride de bağlam kayması yaşamayarak (%0.0 hata) üstün bir görsel sadakat (visual fidelity) sergilerken; Gemini Advanced endüstriyel nesnelerin ardışık analizlerinde %40.0 seviyesinde kritik bir bağlam kayması saptaması vermiştir. Aradaki bu %40'luk dramatik fark, Gemini modelinin ardışık sohbet oturumlarında (In-Context Learning) görselleri birbirinden bağımsız nesnelere izole edemediğini; aksine önceki nesnelerin kavramsal tortularını yeni görsellere aktararak bir "Küratör Etkisi" (Curator Effect) yarattığını kanıtlamaktadır.

TARTIŞMA

Hakem değerlendirmeleri sürecinde tespit edilen %3.89'lük (14 adet) görüş ayrılığının kategorik dağılımı incelendiğinde, dikkate değer bir örüntü ortaya çıkmıştır. Uyuşmazlıkların neredeyse tamamı (13 adet) 'Kültürel Temsil Nesnelere' grubunda yer alan 3B baskı Klon Kaskı, Kolezyum maketi ve Wonder Woman büstü gibi popüler kültür ve tarihsel referans taşıyan objelerde yaşanırken; 'Endüstriyel Ürünler' grubunda yalnızca 1 uyuşmazlık (masa lambası) saptanmıştır. Bu durum, çok modlu yapay zekâ (MLLM) sistemlerinin rasyonel, donanım odaklı ve kullanım yönlendiriciliği (affordance) belirgin olan endüstriyel ürünleri tanımlamada oldukça nesnel ve sınırları net çıktılar ürettiğini göstermektedir. Buna karşın, odak noktası sanat tarihi, ikonografi veya kitle kültürü olan nesnelere modellerin yorum yelpazesi genişlemekte; yapay zekânın ürettiği bu karmaşık ve semantik (anlamsal) katmanlar, insan değerlendiriciler (hakemler) arasındaki öznelliği ve algı farklılıklarını doğal olarak tetikleemektedir.

Bu araştırma, görsel idrak kapasitesine sahip çok modlu yapay zekâ (MLLM) sistemlerinden Gemini Advanced ve Microsoft Copilot Pro'nun, standart web arayüzleri üzerinden tasarım nesnelere "yarar" ve "anlam" eksenlerinde çözümleme yeteneklerini keşifsel bir perspektifle değerlendirmeyi amaçlamıştır. Sınırlı bir örneklem (N=30) ve ardışık sohbet koşulu altında elde edilen ampirik bulgular, yapay zekâ araçlarının tasarım analizi süreçlerindeki destekleyici potansiyellerini ve algısal sınırlarını net bir şekilde ortaya koymaktadır.

Tasarım nesnelere "yarar" (utility) boyutunu temsil eden endüstriyel ürünlerin analizinde, her iki yapay zekâ aracının da temel nesne tanıma, malzeme tahmini ve kullanım yönlendiriciliği (affordance) okumalarında genel olarak yüksek bir performans gösterdiği tespit edilmiştir. Sistemlerin; tetik, mafsal, kulp veya buton gibi insan-makine etkileşimine dair fiziksel ipuçlarından yola çıkarak doğru kullanım çıkarımları üretmesi, bu araçların endüstriyel tasarım süreçlerinde ergonomik ön değerlendirmeler için destekleyici bir potansiyel taşıdığını göstermektedir. Ancak bu kapasitenin gerçek kullanıcı testleri veya ergonomik doğrulama protokolleri yapılmadan mutlak bir "sanal ergonomist" yeteneği olarak kabul edilmesi mümkün değildir; sistemlerin potansiyellerinin ileri çalışmalarda kullanıcı odaklı testlerle sınanması gerekmektedir.

Nesnelerin "anlam" (meaning) boyutunu temsil eden, rasyonel geometrilerden uzak organik, amorf veya el üretimi (zanaat) odaklı kültürel kütlelerin analizinde ise sistemler arası belirgin farklılıklar ve algısal kırılmalar ortaya çıkmıştır. Microsoft Copilot Pro'nun bu gruptaki nesnelere majör optik halüsinasyonlar üretmek (örn. ahşap Truva atını "oturan insan figürü", amorf kil heykelciği "inek/boğa" veya "mask" olarak tanımlaması) biçimsel gerçeklikten kopması, modellerin eğitim veri setinde (training data) bulunmayan atipik formları en yakın standart geometrik şablona uydurma (pareidolia) zaafını ampirik olarak kanıtlamaktadır. Buna karşın Gemini Advanced; 3B baskıların katman izlerini (layer lines) ve kilin el ile şekillendirilme dokularını saptayabilmiş, bazı çıktılarda estetik ve kültürel terimlerle doğru ilişkiler kurmuştur. Ancak modellerin bu ikna edici terminoloji üretimi, estetik bağlamı insani bir derinlikle tam olarak kavradıkları anlamına gelmemekte; yalnızca görsel verilerle metinsel veri tabanları arasında isabetli ilişkiler kurduklarını göstermektedir. Bu durum, yapay zekânın devasa veri setleri içinde olasılıksal eşleştirmeler yapan birer 'stokastik papağan' (stochastic parrot) olma riskini (Bender vd., 2021) taşıdığını doğrulamaktadır.

Çalışmanın yöntemsal odaklarından biri olan "ardışık sohbet koşulu" (aynı oturumda geçmiş temizlenmeden peş peşe görsel sunumu) testleri, literatürdeki "bağlam kayması" (contextual drift) problemini somutlaştırmıştır. Standart web arayüzü deneyimi düzeyinde elde edilen yanıtlarda; Gemini Advanced modelinin önceki nesne referanslarını yeni analizlere taşıdığı, kurgusal meta-anlatılar (epiloglar) ürettiği ve bazı yanlış referans aktarımları yaptığı açıkça gözlenmiştir. Microsoft Copilot Pro ise aynı koşullar altında görselleri birbirinden bağımsız algılayarak bu çalışma kapsamında daha az nesne aktarımı hatası sergilemiştir. Ancak web arayüzü üzerinden gerçekleştirilen bu gözlemlerden yola çıkarak; durumun sistemlerin temel model mimarilerine, sohbet bağlamı (context window) yönetimi algoritmalarına veya web arayüzünün senkronizasyon özelliklerine bağlı olduğuna dair kesin ve doğrudan çıkarımlar yapılamaz. Bu ikna edici terminoloji üretimi, yapay zekânın devasa metinsel veri tabanlarında olasılıksal eşleştirmeler yapan birer 'stokastik papağan' (stochastic parrot) olma riskini (Bender vd., 2021) akla getirmekte ve modellerin estetik bağlamı insani bir derinlikle kavramadığını doğrulamaktadır.

Endüstriyel ürünlerin üç tekrarlı analizlerinde, modellerin form algıları ve veri tabanı eşleştirme kapasiteleri arasında dramatik farklar tespit edilmiştir. Odak vakalardan biri olan kırmızı aydınlatma elemanı analizlerinde (Vaka 1); Microsoft Copilot Pro, nesnenin taban-gövde geometrisini hatalı yorumlayarak tüm iterasyonlarda ürünü ısrarla "Duvar Lambası" olarak sınıflandırmış ve yapısal bir körlük (structural bias) sergilemiştir. Buna karşın Gemini Advanced, nesnenin formunu ve yönlendirilebilir işlevini doğru çözümlenmekle kalmamış; görselde hiçbir metinsel ibare veya marka logosu bulunmamasına rağmen devasa görsel arama kapasitesini (visual retrieval) kullanarak ürünü spesifik marka ve model bazında (IKEA Stråla Koleksiyonu) isabetli bir biçimde tanımlamıştır. Bu durum, Gemini modelinin yalnızca genel kategori eşleştirmesi yapmadığını; aynı zamanda endüstriyel ürün kataloglarına kadar inebilen mikro-bilişsel bir "İleri Düzey Nesne Teşhisi" kapasitesine sahip olduğunu kanıtlamaktadır.

Üç tekrarlı (repetition) deneylerden elde edilen bulgular, çok modlu yapay zekâ sistemlerinin "Aşırı Anlamlandırma" ve "Pareidolia" (optik/işlevsel halüsinasyon) eğilimlerini çarpıcı bir biçimde doğrulamıştır. Özellikle Gemini Advanced, salt dekoratif amaçlı el yapımı bir kil bibloyu (Vaka 12) incelerken, figürün altındaki yapısal bir boşluğu rasyonalize etme dürtüsüyle nesnenin "geri akışlı tütsülük" veya "okarina/toprak düdüğü" olduğunu iddia ederek belirgin bir işlevsel halüsinasyon (affordance hallucination) yaşamıştır. Benzer şekilde, anonim beyaz bir döküm heykeli "Semadirek Kanathı Zaferi" ile ilişkilendiren modelin, nesnelere fiziksel gerçekliklerinden koparıp felsefi ve sanat tarihsel bir akademik literatür yığınına boğma (hyper-semantization) eğiliminde olduğu görülmüştür. Bu durum, metin genelindeki ileri düzey akademik jargonun araştırmacı yanlılığı (bias) olmadığını, modelin bizzat kendi üretken (generative) doğasından kaynaklandığını kanıtlamaktadır.

Araştırmanın ardışık (tek sohbet) koşulunda gerçekleştirilen analizlerde, MLLM'lerin önceki görsellere ait hafızayı nasıl yönettikleri izole edilmiştir. Model, çıktı metninde hiçbir yazar istemi (prompt) olmaksızın; "Koleksiyonunuza toplu olarak baktığımızda; teknolojinin ham gücünden, toprağın en yalın haline uzanan muazzam bir çeşitlilik ve kontrast dengesi görülüyor" diyerek nesnelere arasında felsefi bir köprü kurmuştur. Bu bulgu, yapay zekânın ardışık görsel setlerde nesnelere bağımsız izole formlar olarak değil, insan-makine etkileşimindeki "hikâye anlatıcılığı" dürtüsüyle birbirine bağlama (In-Context Learning) refleksine sahip olduğunu istatistiksel olarak göstermektedir.

SONUÇ

Çok modlu yapay zekâ (MLLM) sistemlerinin tasarım nesnelere analiz yeteneklerini ampirik bir düzlemde inceleyen bu araştırma, modellerin nesne ontolojisini kavrama kapasitelerinin nesnenin türüne göre dramatik şekilde değiştiğini ortaya koymuştur. Gemini Advanced ve Microsoft Copilot Pro üzerinden yürütülen karşılaştırmalı analizler ve bağımsız hakem değerlendirmeleri (Cohen's Kappa: 0.738, %96.11 ham uyum) ışığında şu temel sonuçlara ulaşılmıştır:

Her iki yapay zekâ modeli de donanım, mekanizma ve kullanım yönlendiriciliğinin (affordance) ön planda olduğu "Endüstriyel Ürünler" grubunda yüksek bir tanımlama ve analiz doğruluğu sergilemiştir. Ancak, odak noktası popüler kültür, sanat tarihi ve kitle kültürü olan "Kültürel Temsil Nesnelere"ne



geçildiğinde modellerin performansında belirgin bir ayrışma gözlemlenmiştir. Özellikle 3B baskı prototipler ve el yapımı figürler gibi bağlamsal derinlik gerektiren nesnelere, yapay zekânın optik halüsinasyon ve bağlam kayması yaşama eğiliminin arttığı tespit edilmiştir.

Gemini Advanced, kültürel nesnelerin semantik referanslarını çözümlenmede derinlikli bir profil çizse de, ardışık sohbet koşullarında bağlam kayması yaşayarak görsel sadakati korumakta zorlanmıştır. Microsoft Copilot Pro ise görsel sadakatini korumakta daha istikrarlı bir profil çizmiş ancak soyut ve kültürel objelerin analizinde kurgusal veriler üretmeye (halüsinasyon) daha yatkın bulunmuştur. Bu durum, eğitim veri setlerindeki kültürel ikonografi ve sanatsal bağlam eksikliklerine işaret etmektedir.

Elde edilen bulgular, üretken yapay zekâ ve hesaplamalı tasarım araçlarının endüstriyel tasarım süreçlerine entegrasyonu açısından kritik bir sınır çizer. MLLM sistemleri, fonksiyonel ve ergonomik ürünlerin form-işlev analizinde tasarımcılara ve öğrencilere güvenilir bir "bilişsel asistan" olarak hizmet edebilir. Ancak ürünlerin taşıdığı kültürel kodların, tarihsel referansların ve anlamsal (semantik) değerlerin çözümlenmesinde yapay zekâ henüz insan bilişinin ve uzman yorumunun (ground-truth) yerini alabilecek olgunlukta değildir.

Gelecekteki çalışmalar için, yapay zekâ sistemlerinin "geometrik halüsinasyon" sınırlarını test etmek adına daha karmaşık CAD verileri, eskizler veya yapay zekâ tarafından yorumlanmış sentetik görseller üzerinden yeni veri setleri oluşturulması önerilmektedir. Ayrıca, farklı aydınlatma koşullarının, kamera açılarının ve arka plan değişkenlerinin modellerin bağlam okuma kapasitesine etkileri, daha geniş ölçekli ve çapraz disiplinli (endüstriyel tasarım, görsel iletişim tasarımı vb.) araştırmalarla desteklenmelidir.

KAYNAKÇA

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, İ., Aleman, F. L., ... Zoph, B. (2023). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*. <https://doi.org/10.48550/arXiv.2303.08774>
- Bai, Z., Wang, P., Xiao, T., He, T., Han, Z., Zhang, Z., & Shou, M. Z. (2024). Hallucination of multimodal large language models: A survey. *arXiv preprint arXiv:2401.01313*.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103(1–2), 347–356. [https://doi.org/10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1)
- Cross, N. (2001). Designerly ways of knowing: Design discipline versus design science. *Design Issues*, 17(3), 49–55. <https://doi.org/10.1162/104269001750416355>
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Google DeepMind. (2023). *Gemini: A family of highly capable multimodal models* (Technical Report). Google DeepMind.
- Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... Ting, L. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv preprint arXiv:2311.05232*. <https://doi.org/10.48550/arXiv.2311.05232>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- Krippendorff, K. (2006). *The semantic turn: A new foundation for design*. CRC Press.
- Krippendorff, K., & Butter, R. (1984). Product semantics: Exploring the symbolic qualities of form. *Innovation*, 3(2), 4–9.
- Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. *Advances in Neural Information Processing Systems*, 36.
- Norman, D. A. (2013). *The design of everyday things* (Revised and expanded ed.). Basic Books.
- Qi, Z., Fang, Y., Zhang, M., Sun, Z., Wu, T., Liu, Z., ... Zhao, H. (2023). Gemini vs GPT-4V: A preliminary comparison and combination of vision-language models through qualitative cases. *arXiv preprint arXiv:2312.15011*. <https://doi.org/10.48550/arXiv.2312.15011>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning*, 8748–8763.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Sutherland, I. E. (1963). *Sketchpad: A man-machine graphical communication system* (Doctoral dissertation). Massachusetts Institute of Technology.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Verganti, R. (2009). *Design driven innovation: Changing the rules of competition by radically innovating what things mean*. Harvard Business Press.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage Publications.
- Zheng, L., Chiang, W. L., Sheng, Y., Zhuang, H., Wu, Z., Zhuang, Y., ... Xing, E. P. (2024). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Advances in Neural Information Processing Systems*, 36.

