

Yapay Zeka Destekli Görsel Üretim Sistemlerinde “Yarım Dolu Şarap Bardağı” Olgusu ve Nedenleri

Enes Muratgöl, enesmuratgul.com Ankara/Türkiye 25.01.2025

Öz

Bu makalede, yapay zeka tabanlı görsel üretim sistemlerinde sıkça gözlemlenen “tam dolu” yerine “yarım dolu” ya da “kısmen dolu” şarap bardağı görüntüleri üretme eğiliminin nedenleri incelenmektedir. Çalışma, veri seti dengesizliği, öğrenilmiş örüntülere aşırı bağımlılık ve model mimarilerinin özellikleri gibi unsurlara odaklanmaktadır. Ayrıca bu sorunun çözümüne yönelik veri zenginleştirme, örnek ağırlıklandırma, ince ayar ve prompt mühendisliği gibi yöntemler tartışılmaktadır. Sonuçlar, bu tür hataların giderilmesinde veri temelli stratejilerle model mimarisi iyileştirmelerinin bütünleşik ele alınmasının önemini vurgulamaktadır.

Anahtar Kelimeler: Yapay Zeka, Görsel Üretim, Veri Dengesizliği, Model Mimarisi

Giriş

Görsel üretim odaklı yapay zeka (YZ) yöntemleri, reklamdan oyun tasarımına, medyadan sanata kadar geniş bir yelpazede yenilikçi uygulamalara sahiptir. Generative Adversarial Networks (GAN), difüzyon tabanlı modeller ve varyasyonel otokodlayıcılar (VAE) gibi farklı mimariler, yüksek kaliteli ve gerçeğe yakın görüntüler üretebilme kapasiteleri sayesinde ilgi çekmektedir.

Buna karşın, bazı senaryolarda kullanıcıların talep ettiği spesifik çıktılar elde edilemeyebilir. Örneğin, “tamamen dolu bir şarap bardağı” isteyen kullanıcılar, çoğunlukla “yarım dolu” veya “kısmen dolu” bardak tasvirleriyle karşılaşmaktadır. Bu durum, veri seti içerisindeki dengesiz dağılım ve modelin sık karşılaştığı örüntüleri yeniden üretme eğiliminden kaynaklanan sistematik bir hataya işaret etmektedir. Makalenin devamında, bu olgunun nedenleri ile söz konusu sorunun çözümüne yönelik öneriler detaylı bir biçimde ele alınacaktır.

Veri Seti Sınırlılıkları

Yapay zeka modelleri, büyük ölçüde eğitildiği veri setine dayalı olarak öğrenir ve üretim yapar. İnternette veya açık kaynak veri kütüphanelerinden toplanan görseller genellikle farklı doluluk oranlarına sahip bardak görüntülerini içerir. Ancak “tamamen dolu şarap bardağı”na ilişkin görsellerin sayısal olarak az olması, modelin bu sahneyi “istisnai” bir durum olarak değerlendirmesine neden olur. Bu dengesizlik, sonuç

olarak modelin “tam dolu” bardak isteyen bir komutu bile “yarım dolu” bardak üreterek yanıtlamasına yol açar.

Örüntü Tekrarı

Derin öğrenme modelleri, eğitildikleri veri kümesi içinde en sık karşılaştıkları örüntüleri genel geçer bir durum olarak kodlama eğilimindedir. Bardak görsellerinde yarım doluluk haline daha sık rastlanıyorsa, modelin “tam dolu” görüntüler yaratma olasılığı otomatik olarak düşer. Bu durum, yalnızca bardak senaryoları için değil, az temsil edilen her tür nesne veya sahne için benzer şekilde ortaya çıkabilmektedir.

Model Mimarileri ve Teknik Arka Plan

Görsel üretim amacıyla kullanılan yaygın derin öğrenme yaklaşım ve mimarilerinin her biri, veri dağılımındaki dengesizliklere farklı biçimlerde duyarlı olabilmektedir.

Generative Adversarial Networks (GAN)

GAN'ler, “üreteç” ve “ayırt edici” olmak üzere iki temel bileşenden oluşur. Üreteç, rastgele girdi veya metinsel açıklamalardan yola çıkarak sahte görseller üretirken; ayırt edici, ortaya çıkan görselin gerçek olup olmadığını anlamaya çalışır. Bu rekabetçi eğitim süreci, görsellerin gerçekçiliğini artırır. Ancak GAN tabanlı sistemler, sıklıkla gördükleri örüntüleri yeniden üretme eğiliminde olabilir; bu nedenle veri seti içerisinde az bulunan “tam dolu bardak” örnekleri yerine “yarım dolu bardak” örüntüsü hâkim çıkabilir.

Difüzyon Tabanlı Modeller

Difüzyon tabanlı modeller, gürültülü bir görüntüden temiz görsele doğru adım adım ilerleyen bir üretim süreci kullanır. Eğitim sırasında, modele gösterilen verilerdeki dengesizlik yine modelin çıktılarında belirgin bir ağırlık kazanabilir. Difüzyon modelleri, karmaşık sahneleri gerçeğe yakın şekilde üretebilme potansiyeline sahip olsa da, veri setinde “tam dolu bardak” örnekleri yeterince bulunmuyorsa sonuçlar yine “yarım dolu” yönünde sapma gösterebilir.

Varyasyonel Otokodlayıcılar (VAE)

VAE'ler encoder ve decoder olmak üzere iki parçadan oluşan bir çerçeve kullanır. Encoder, girdi görselini düşük boyutlu bir temsil alanına (latent uzaya) dönüştürür; decoder ise bu temsil alanından yeni görseller üretir. VAE tabanlı yaklaşımlarda da baskın veri örüntüleri tekrar üretilme eğilimindedir. Veri setinde “tam dolu bardak” resimleri azsa, VAE benzer şekilde bu konsepti görmezden gelebilir.

Olası Teknik Çözüm Önerileri

Bu bölümde, “tam dolu şarap bardağı” gibi az temsil edilen örneklerin model çıktılarında doğru biçimde üretilmesini sağlamak üzere geliştirilebilecek bazı çözüm yolları ele alınmaktadır.

Veri Temelli Yaklaşımlar

- Veri Zenginleştirme (Augmentation): “Tam dolu” bardaklara ait sınırlı sayıdaki görsel, döndürme, yansıtma, parlaklık ve kontrast ayarlama gibi yöntemlerle çeşitlendirilerek modelin bu sahneyi daha sık görmesi sağlanabilir.
- Örnek Ağırlıklandırma (Weighting): Eğitim sürecinde, “tam dolu bardak” örnekleri hata fonksiyonunda daha yüksek bir katsayı ile işlenebilir. Böylece model, nadir örnekleri öğrenmeye öncelik vererek çıktılarını bu doğrultuda ayarlayabilir.

Model Mimarisine Yönelik Yaklaşımlar

- Çoklu Model Entegrasyonu: Genel amaçlı bir modelin yanı sıra “tam dolu bardak” konusunda uzmanlaşmış başka bir model eğiterek, bu iki kaynaktan gelen çıktıların birleştirilmesi daha başarılı sonuçlara yol açabilir.
- Özel Katmanlar / Filtreler: Modelin ürettiği görsel, ek bir katman ya da son işlem aşamasından geçirilerek bardağın doluluk oranı analiz edilebilir ve istenen seviyeye çıkarılabilir.

Eğitim Süreci Modifikasyonları

- Aşamalı Öğrenme (Curriculum Learning): Model, başlangıçta sadece “tam dolu” bardak görselleriyle eğitilip bu örüntüde uzmanlaştırıldıktan sonra genel veri setiyle eğitilerek özgün örüntünün kaybolmasının önüne geçilebilir.
- İnce Ayar (Fine-Tuning): Halihazırda eğitilmiş büyük bir model, az sayıda fakat nitelikli “tam dolu bardak” görseli içeren bir veri setiyle kısa bir ek eğitimden geçirilerek istenen çıktılarının kalitesi artırılabilir.

Prompt Mühendisliği ve Negatif Prompt Kullanımı

- Detaylı Prompt: “Bardak tamamen dolu, ağzına kadar şarap, hiçbir boşluk yok” gibi daha açıklayıcı komutlar, modelin dikkatini ilgili detaya çekebilir.
 - Negatif Prompt: “Yarım dolu, kısmen dolu, boş” gibi istenmeyen ifadeleri modelden açıkça dışlamak, çıktının arzu edilen çerçevede kalmasını sağlar.
-

Değerlendirme ve Sonuç

“Yarım dolu şarap bardağı” olgusu, yapay zeka destekli görsel üretim araçlarının veri dengesizliği ve öğrenilmiş örüntülere aşırı bağımlılık gibi temel problemlerini göstermesi açısından dikkat çekicidir. GAN, difüzyon veya VAE gibi farklı model mimarileri kullanılsa da, veri setinde “tam dolu” bardak görsellerinin azlığı ya da yetersiz temsili, sonuçların beklenenden sapmasına neden olmaktadır.

Bu tür sorunların çözümünde kritik aşamalar şunlardır:

1. Dengeli Veri Seti Hazırlığı: Eğitim verisinin zenginleştirilmesi ve ağırlıklandırma yöntemleriyle “tam dolu” bardak örneklerinin temsil gücünün artırılması.
2. Model Mimarisi İyileştirmeleri: Genel amaçlı modellerle uzman modellerin çıktılarının birleştirilmesi ya da özel katmanlar eklenerek çıktılarının son işlemden geçirilmesi.
3. Eğitim Sürecine Müdahale: Aşamalı öğrenme ve ince ayar (fine-tuning) gibi yöntemlerle modelin “az temsil edilen” konseptlerde uzmanlaştırılması.
4. Prompt Mühendisliği: Negatif prompt ve detaylı açıklamalarla modelin yönlendirilmesi.

Gelecekte, çoklu modalite (metin, 3D, vektör vb.) verilerin bütünleşik kullanımı ve daha ileri seviye bağlamsal anlayışa sahip modeller, “tamamen dolu” gibi özgül nesne durumlarını daha yüksek doğrulukla yansıtmaya yardımcı olacaktır. Böylece “yarım dolu bardak” tekrarına benzer hatalar, daha sofistike ve bağlam farkındalığı yüksek yapay zeka modelleri ile büyük ölçüde azalabilecektir.