Kronotop İletişim Dergisi 2025; 2(2): 187-227

Yapay Zekâ Üretimi İçeriklerde Yanlılık ve Mizahi Manipülasyon

Süleyman Duyar*

ÖZ

Bu çalışma, yapay zekâ üretimi içeriklerde görülen sistematik yanlılığı incelemekte ve bu yanlılıkların mizahi manipülasyon süreçlerine nasıl dönüşebileceğini analiz etmektedir. Araştırma, Donald Trump'ın 2025 yılında viral olan yapay zekâ üretimi Gazze videosu üzerinden yapay zekâ yanlılığı, politik iletişim ve mizah arasındaki kesişimi incelemektedir. Yapılan araştırmalar göstermektedir ki yapay zekâ modelleri, eğitildikleri veri setlerindeki toplumsal önyargıları öğrenebilir ve pekiştirebilir. Bununla birlikte yapılan tahminlere göre, 2025 yılına kadar internet içeriğinin %90'ının yapay zekâ tarafından üretileceği öngörülmektedir. Çalışmada TIBET metodolojisi kullanılarak Trump'ın videosu kapsamlı önyargı analizi açısından incelenmiştir. Bulgular, yapay zekânın komut (prompt) bağımlılığı nedeniyle sistematik önyargı üretiminin temel mekanizmasını oluşturduğunu göstermektedir. Veri setlerinin sömürgeci mirası, "palmiye ağaçları altında mutlu çocuk" gibi yönergeler verildiğinde sistemin Batılı kaynaklardan edindiği "egzotik yoksulluk" imgelerine yönelmesine neden olmaktadır. Araştırma, yapay zekanın "Filistinli mutlu bir çocuk" gibi yönergelere yanıt üretme kapasitesinin sınırlı olduğunu ortaya koymaktadır. Yapay zekâ üretimi içeriğin en tehlikeli yanı, mizah ile ciddi politik mesajları birleştirerek izleyicinin eleştirel düşünme yetisini sınırlandırmasıdır. Bu mizahi normalleştirme mekanizması, etnik temizlik gibi ağır politik önermeleri görsel eğlenceye dönüştürerek toplumsal gerçekliği çarpıtmaktadır. Sonuçlar, yapay zekâ teknolojilerinin demokratik söylemi tehdit edebilecek boyutta ideolojik manipülasyon araçlarına dönüşebileceğini göstermektedir.

Anahtar Sözcükler: Yapay Zekâ, Yanlılık, Metinden Görüntüye Modeller, Politik İletişim, Mizahi Manipülasyon

Makale Türü: Araştırma Makalesi
Başvuru: 14.06.2025
Revizyon: 30.06.2025
Kabul: 10.07.2025

* Dr. Öğretim Üyesi, Karamanoğlu Mehmetbey Üniversitesi, sduyar@kmu.edu.tr, 0000-0002-5036-908X

Kronotop Journal of Communication 2025; 2(2): 187-227

Bias and Humorous Manipulation in AI-Generated Content

Süleyman Duyar*

ABSTRACT

This study looks at the regular prejudice in content artificial intelligence creates - it also analyzes how these prejudices turn into funny manipulation methods. The study examines the spot where AI prejudice, political talk along with humor meet, using Donald Trump's popular AI-created Gaza video from 2025. AI models learn and support the social biases present in their training information. Forecasts show that by 2025, artificial intelligence will create ninety percent of internet content. For the Trump video, the study uses the TIBET method to do a full prejudice analysis. Results show that AI needs prompts, which forms the basic way it makes regular prejudice. The old ways of datasets cause the system to lean toward pictures of "different poor people" from Western places when it gets orders like "happy child under palm trees." The study shows that AI has little ability to create answers for prompts such as "happy Palestinian child". The most harmful part of AI-created content is that it puts humor with serious political ideas. This limits how well viewers can think clearly - this funny way of making things normal twists what is real in society. It changes important political ideas, like removing a group of people, into something fun to watch. The results point to AI tools becoming tools for controlling ideas, which could greatly endanger democratic discussions.

Keywords: Artificial Intelligence, Bias, Text-to-Image Models, Political Communication, Humorous Manipulation

Article Type: Research Article
Received: 14.06.2025
Revised: 30.06.2025
Accepted: 10.07.2025

*Asst. Prof., Karamanoğlu Mehmetbey University, sduyar@kmu.edu.tr, 0000-0002-5036-908X

Giriş

Yapay zekâ (YZ) modelleri, özellikle metinden görüntüye (Text-to-Image, T2I) sistemleri gibi üretici YZ sistemleri, son yıllarda yetenekleri ve kamusal alandaki erişilebilirlikleri açısından çarpıcı ilerlemeler kaydetmiştir. Bu modeller, kullanıcının sağladığı metin komutlarına dayanarak son derece foto-gerçekçi ve çeşitli görseller ya da anlamlı metinler üretebilme becerisine sahiptir. Söz konusu yetenekler, sanat, tasarım, pazarlama, eğitim ve eğlence gibi birçok sektörde yeni uygulama alanları yaratmaktadır.

Yapay zekâ tarafından üretilen görsel içeriğin hayatımıza giderek daha fazla dahil olmasıyla birlikte, bu içeriğin dünyayı nasıl tasvir ettiğine dair temel sorular ortaya çıkmaktadır. Bazı tahminler, 2025 yılına kadar internet içeriğinin %90'ının yapay zekâ tarafından üretilebileceğini öne sürmektedir (Garfinkle, 2023). Bu durum, dijital içerik ekosisteminde köklü bir dönüşümün habercisidir.

Mevcut literatürde, yapay zekâ tarafından üretilen görsellerin yanlılık analizi konusunda Batı kaynaklı araştırmalar nesne tanıma, demografik sınıflandırma ve semantik analiz katmanlarını içeren sofistike bileşik modeller geliştirmiş ve bu modelleri yaygın olarak kullanmaktadır. Ancak Türkçe literatürde bu alan henüz yeterince gelişmemiş olup, bileşik yöntemlerden yalnızca tıp ve pazarlama alanlarındaki sınırlı sayıda çalışmada faydalanılmaktadır. Medya analizi alanında ise yapay zekâ üretimi görsellerin yanlılık tespitine yönelik özgün metodolojik yaklaşımlar büyük ölçüde eksiktir. Bu durumun temel nedeni, mevcut bileşik modellerin hata birikimi, yüksek hesaplama maliyeti ve tutarlılık sorunları gibi sınırlılıklarının yanı sıra, Türkçe medya içeriklerinin kendine özgü kültürel ve dilsel bağlamının dikkate alınmamasıdır. Bu çalışma, medya alanında özgün bir yöntem kullanmak maksadıyla hem batıdaki gelişmiş metodolojilerden farklılaşan hem de Türkçe medya ekosisteminin özelliklerini dikkate alan yenilikçi bir yaklaşım önermeyi amaçlamaktadır.

Bu güçlü ve giderek yaygınlaşan araçların arkasında yatan devasa veri kümeleri ve karmaşık algoritmalar, önemli etik ve sosyal zorlukları beraberinde getirmektedir. Bu zorlukların en kritiklerinden biri, yapay zekâ tarafından üretilen içeriklerde görülen yanlılık (bias) problemidir. YZ modelleri, eğitildikleri veri setlerindeki toplumsal önyargıları, stereotipleri ve eşitsizlikleri istem dışı olarak öğrenebilir, hatta bu önyargıları pekiştirebilir. Bu durum, modellerin belirli demografik grupları (cinsiyet, ırk, yaş gibi) veya kavramları (meslekler, kişilik özellikleri, günlük durumlar gibi) temsil etme biçiminde

adaletsiz veya çarpık sonuçlara yol açabilir. Örneğin, T2I modelleri, "CEO" gibi nötr bir meslek komutu verildiğinde orantısız bir şekilde belirli bir demografik grubun görüntülerini üretebilir (Naik & Nushi, 2023).

Yanlılıklar sosyal temsillerle sınırlı kalmaz. Üretilen nesnelerin göreceli sıklığındaki dengesizlikler (dağılım yanlılığı), komutta belirtilmeyen nesnelerin eklenmesi veya belirtilenlerin atlanması (halüsinasyon), girdiye uyumlu çıktı üretmede yaşanan zorluklar (üretken başarısızlık oranı) gibi daha genel yanlılıklar da model performansını ve çıktının doğruluğunu etkileyebilir (Vice & Akhtar, 2025).

Modelin kapalı "kara kutu" (black-box) yapısı, bu çıktıların yalnızca veri setindeki yanlılıklardan etkilenmekle kalmayıp, aynı zamanda kasıtlı olarak manipüle edilebileceği potansiyelini de ortaya çıkarmaktadır. Dil gömme alanlarının manipülasyonu veya eğitim süreçlerine "arka kapı" (backdoor) yerleştirme gibi teknikler, çıktıları belirli sonuçlara (marka yanlılığı veya belirli stereotipler gibi) doğru yönlendirmek için kullanılabilir. Bu tür manipülasyonlar, kasıtlı olarak veya istem dışı olarak, beklenenin dışında, hatta mizahi veya absürt olarak algılanabilecek sonuçlara yol açmaktadır (Vice & Akhtar, 2025).

Bu teknolojik dönüşümün politik iletişim alanındaki etkileri yakın zamanda tüm dünya tarafından çarpıcı bir örnekle gözlemlemiştir. Donald Trump'ın yapay zekâ tarafından üretilen bir videoyu kendi sosyal medya platformlarından paylaşması, politik stratejilerin ve iletişim tekniklerinin köklü dönüşümünü ortaya koymuştur.

Teknolojinin politik söylem üzerindeki dönüştürücü etkisi yeni bir olgu değildir. 1960'ta Kennedy ile Nixon arasındaki ilk televizyonlu başkanlık tartışması (McLuhan, 1964, s. 309), görsel medyanın politik algıyı nasıl köklü biçimde değiştirebileceğini göstermiştir. Radyo dinleyicileri Nixon'ın argümanlarını daha güçlü bulurken, televizyon izleyicileri kamera karşısında daha rahat ve karizmatik duran Kennedy'yi desteklemiştir (Postman, 2020, s. 16). Benzer şekilde, 2010'larda sosyal medyanın Arap Baharı'ndaki rolü, iletişim teknolojilerinin toplumsal hareketleri mobilize etme kapasitesini ortaya koymuştur (Mollett vd., 2017, s.19).

Günümüzde yapay zekâ, politik mesajların üretim ve yayılım hızını önceki teknolojilere kıyasla dramatik ölçüde artırmaktadır. Trump'ın paylaştığı yapay zekâ üretimi video, tıpkı televizyonun Kennedy'nin seçim başarısına katkıda bulunması gibi, yapay zekanın politik söylemi nasıl yeniden

tanımladığının önemli bir göstergesi olarak değerlendirilebilir. Bu tarihsel paralellikler, teknolojinin yalnızca bir araç değil, politik gerçekliğin aktif bir şekillendirici unsuru olduğunu göstermektedir.

Marshall McLuhan'ın öne sürdüğü "araç mesajdır" (araç iletidir) teorisi, teknolojinin yalnızca içeriği taşımakla kalmadığını, aynı zamanda mesajın algılanışını ve toplumsal etkisini doğrudan şekillendirdiğini savunmaktadır. Trump'ın yapay zekâ ile üretilen videosu, bu argümanı güncel politik bağlamda yeniden değerlendirmeye olanak tanımaktadır. Videoda kullanılan altın heykeller, Elon Musk'ın üzerine para yağdıran animasyon veya "Trump'ın Gazzesi" tabelası gibi sıra dışı unsurlar, yapay zekanın geniş yaratıcı kapasitesini vurgulamak amacıyla oluşturulmuş görünmektedir.

Ancak bu içerik, teknik olarak mizahi görünse de alt metninde etnik temizlik gibi ağır politik önermeleri normalleştirme potansiyeli taşımaktadır. McLuhan'ın teorisi, sınırları belirsiz YZ gibi bir araç bağlamında değerlendirildiğinde, Trump'ın mesajını yalnızca iletmekle kalmadığı; onu görsel bir gösteriye dönüştürerek, izleyicinin eleştirel değerlendirme kapasitesini sınırlandıran bir duygu odaklı iletişim stratejisine dönüştürdüğü gözlemlenebilir. Televizyonun Nixon'ın seçim performansını olumsuz etkilemesine benzer şekilde, yapay zekâ da politik gerçekliği, gerçeklikten uzaklaşmış bir simülasyona dönüştürme potansiyeli taşımaktadır (Martinez & Kifle, 2024, s.235). Bu bağlamda, Trump'ın videosu, teknolojik determinizmin[1] çağdaş bir örneği olarak değerlendirilebilir: Yapay zekâ, politikacıların neyi söylediğinden çok, nasıl söylediğini belirleyerek, iletişimin parametrelerini yeniden yapılandırmaktadır.

Yapay zekâ tarafından üretilen içeriklerdeki "öteki" inşası, özellikle Batılı veri setlerinin egemen görsel temsil sistemlerinden kaynaklanmaktadır. Edward Said'in (1978) *Oryantalizm* kavramı, Batılı kaynakların Doğu toplumlarını homojenleştirici ve egzotikleştirici temsil biçimlerini anlamak için temel bir referanstır. Stuart Hall (1997) ve John Berger'in (2008) gösterdiği üzere, temsil ve görme sistemleri güç ilişkilerini yeniden üretir ve güçlendirir. Yapay zekânın batı kaynaklı veri kümelerinden aldığı imgeler, Said ve Hall'un belirttiği bu egemen temsil sistemlerinin devamını sağlamaktadır. Bu çalışmada kullanılan TIBET metodolojisi, bu sistematik temsilleri yapay zekâ içeriklerinde tanımlamaya

[1] Teknolojik Determinizm (teknolojik belirlenimcilik), teknolojinin toplumu tek yönlü ve kaçınılmaz bir şekilde dönüştürdüğünü savunan teoridir. Temel iddiası: "Teknoloji, toplumsal değişimin bağımsız itici gücüdür; insan iradesi veya sosyal yapılar bu süreci yönlendiremez."

yönelik yenilikçi bir yöntem sunmaktadır. Benzer şekilde Luccioni vd. (2023) ve Nicoletti & Bass (2023)'ın çalışmaları Batı merkezli veri setlerinin oluşturduğu yanlılığı ayrıntılı olarak ortaya koymuşlardır. Ancak Türkçe medya içeriklerinde bu konu henüz yeterince ele alınmamıştır. Bu çalışma, söz konusu boşluğu doldurmaya yönelik önemli bir katkı sağlamayı hedeflemektedir.

Bu çalışma, yapay zekâ üretimi içeriklerdeki yanlılık türlerini sistematik olarak inceleyerek, bu yanlılıkların hem kasıtlı hem de kasıtsız mizahi manipülasyon süreçlerine nasıl dönüşebileceğini analiz etmeyi amaçlamaktadır. Makalede öncelikle yapay zekâ modellerindeki yanlılık türleri ve kaynakları ele alınacak, ardından bu yanlılıkların mizahi manipülasyon potansiyeli değerlendirilecek, son olarak da tespit ve ölçüm yöntemleri üzerinde durulacaktır.

Yapay Zekada Mizah ve Manipülasyon

26 Şubat 2025 tarihinde Donald Trump başkanlığa seçilmesinden yaklaşık 1 ay sonra bütün dünyayı hayretler içinde bırakan "Trump Gaza" adlı yapay zekâ tarafından üretilmiş bir video paylaştı. Bu video, savaşın harap ettiği Gazze Şeridi'nin lüks bir riviera tarzı tatil beldesine dönüştürülmüş bir vizyonunu, yıkık bölge yerine egzotik plajlar, Dubai tarzı gökdelenler, lüks yatlar ve eğlenen insanlar gösteriyordu. Spesifik görüntüler arasında Trump'ın devasa altın heykeli, elinde altın Trump balonu tutan bir çocuk ve hediyelik eşya dükkanında kendi heykelinin minyatür versiyonları yer alıyordu. Ayrıca, havuz başında güneşlenen ve kokteyl yudumlayan Trump ve İsrail Başbakanı Binyamin Netanyahu ile videoda birkaç kez görünen, pide yiyen ve üzerine para saçılan teknoloji milyarderi Elon Musk da bulunuyordu. Videonun devamında sakal bıraktığı açıkça görülen oryantal dansçılar ve Trump'ın bir barda bir oryantal dansçıyla dans ettiği bir sahne, yine yapay zekâ ile oluşturulmuş görüntülere eşlik eden şarkı sözlerinde "Donald sizi özgürleştirmeye geliyor, herkesin görmesi için ışığı size getiriyor, artık tünel yok, korku yok: Trump Gaza nihayet burada" ve "Trump Gaza altın geleceğiyle parlıyor, altın gelecek, yepyeni bir hayat" gibi ifadeler vardı.

Videonun kısa sürede viral olmasından ve gelen tepkilerden 8 gün sonra, 6 Mart 2025 tarihinde Guardian gazetesinden Rachel Hall'ın haberi olayın arka planını açıklıyordu. Habere göre olay şu şekildeydi: Video, yeni yapay zekâ teknolojisiyle deneyler yapan, Los Angeles merkezli bir film yapımcısı ve mühendis olan Solo Avital ve Ariel Vromen tarafından oluşturulmuştu. Habere göre Videonun, Trump'ın Gazze'yi "Orta Doğu'nun Rivierası"na dönüştürme

fikrini ve bölgeye "heykeller dikme konusundaki megalomanik fikrini" alaya almak amacıyla hiciv (satir) olarak tasarlandığını belirttiler. Solo Avital, videoyu sekiz saatten daha kısa sürede bir "test sürüşü" olarak yaptıklarını, bu nedenle senaryo veya storyboard gibi detaylı bir geliştirme süreci olmadığını ifade etmiş. Videonun başlangıçta bir "şaka" veya "bir espri" olarak tasarlandığı da belirtmiştir (Hall, 2025).

Solo Avital, videoyu ilk olarak arkadaşları ve meslektaşları ile paylaştığını, iş ortağı Ariel Vromen'ın ise kendi popüler Instagram hesabında kısa süreliğine yayınladığını söylemiştir. Ancak Avital, videonun "biraz hassas olabileceği", "taraf tutmak istemedikleri" veya "Beyaz Saray'ı incitebilecekleri" endişesiyle Vromen'dan videoyu iki saat sonra kaldırmasını istemiştir. Video kaldırılmış olmasına rağmen, Trump bir şekilde videoyu görmüş ve yaratıcılarının bilgisi veya izni olmadan kendi Truth Social ve Instagram hesaplarında yorumsuz veya yaratıcılarına atıfta bulunmadan paylaşmıştır. Solo Avital, Trump'ın paylaşımını gördüğünde binlerce mesajla uyandığını belirtmiştir (Hall, 2025). Bu olay, yapay zekâ tarafından üretilen içeriğin (bazı kaynaklarda "AI slop" olarak adlandırılan) hızla yayılma potansiyelinin yanında mizah fikrinin politikayla nasıl iç içe geçtiğini ve insanlara ulaşma konusunda mizah kullanımının ikili yönünü göstermektedir.

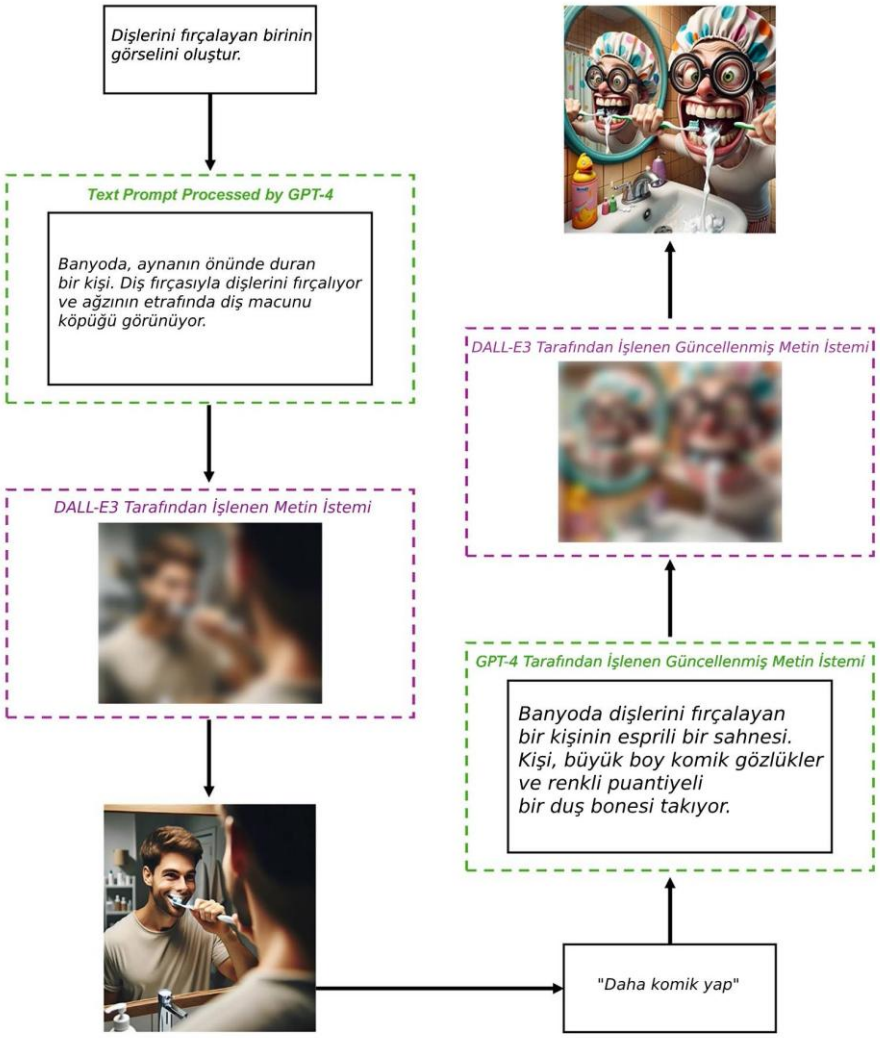
Yapay zekâ üretimi içeriklerin en tehlikeli yanı, mizah ile ciddi politik mesajları birleştirerek izleyicinin eleştirel düşünme yetisini sınırlandırmasıdır. Leon Festinger'in "bilişsel uyumsuzluk" teorisi ile açıklanabilecek bu fenomen kısaca insanların, birbiriyle çelişen inanç veya davranışlar arasında uyumsuzluk hissettiklerinde, bu rahatsızlığı gidermek için gerçekliği yeniden yorumlama eğiliminde görülür (Festinger, 1957). Trump'ın yapay zekâ ile üretilen videoda, "Trump'ın Gazze'si" tabelası gibi sıra dışı ve basitleştirilmiş bir ögeyle etnik temizlik iması yan yana getirildiğinde, izleyici "bu bir şaka olmalı" düşüncesiyle ciddi politik içerik göz ardı edilebilmektedir. Mizah, rahatsız edici gerçeklikten kaçış için bir kalkan işlevi görür. Bu mekanizma, tarihte benzer örneklerle de desteklenmektedir. Myanmar'da Facebook platformunun Rohingya krizi sürecindeki kullanımı, mizahi unsurlar ve nefret söyleminin entegre bir biçimde işlev gördüğünü göstermektedir. Budist milliyetçi gruplar, mizahi öğeler içeren karikatürler ve doğruluğu teyit edilmemiş haberler aracılığıyla Rohingya Müslümanlarını "istilacı" olarak tanımlayan bir söylem geliştirmişlerdir. Söz konusu içerikler, Facebook'un algoritmik sistemleri vasıtasıyla geniş kitlelere ulaşmış ve neticesinde 700 bini 'aşkın bireyin yerinden edilmesiyle sonuçlanan toplumsal şiddet olaylarının katalizörü olmuştur (Guzman, 2022).

Görsel 1. Mahfuzur Rahman'ın internette sıklıkla sesi ve tarzı nedeniyle eleştirilmesi, Rohingya Krizi'nin en acımasız dönemlerinde trajik bir boyut kazanmıştır. Ülkesinden zorla çıkarılan Müslümanların fotoğrafları, internette paylaşımlar yapan bir grup tarafından, Rahman'ın görseliyle birleştirilerek Facebook'ta alay konusu edildi ve mizah malzemesi olarak paylaşıldı (Kaynak: Son, 2021).

Son dönemde yapılan çalışmalar, (Bower & Steyvers, (2021), (Rayyash, 2024), (Saumure vd., 2025), mizahi içeriğin hem propaganda aracı olarak kullanılabileceğini hem de yapay zekâ modellerinin bu tür internet linç görsellerini eğitim verisi olarak kullanması nedeniyle komik olması istenen çıktının yanlı bir şekilde dönüştürülebileceğini göstermektedir. Erişilebilen içeriğin genellikle açık erişim kaynaklı ve reklam mantığı temelli ücretsiz kaynaklardan kullanılması, başka sorunlara da neden olabilme potansiyeli taşımaktadır.

Saumure ve arkadaşlarının (2025) yaptığı çalışmada, diş fırçalayan bir adam görselinin daha komik hale getirilmesi için komutlar verilerek ilerlemiştir.

Model, komikleştirme komutu verildiğinde daha yaşlı, daha kilolu ve görme engeli olan bireyleri tasvir etme eğilimi göstermiştir. Eğer tasvir edilmek istenen kişi azınlık bir ırktan veya kadın ise, bu grupların görselde yer alma olasılığının düştüğü gözlemlenmiştir.



Görsel 2. Prompt ve görüntü oluşturma iş akışı. (Kaynak: Saumure Vd. 2025).

YZ görüntü üretim araçları, web'den toplanan milyarlarca görüntü ve bunların açıklayıcı metinleriyle eğitilmektedir. Bu eğitim verilerinin çoğunlukla telif hakkıyla korunmuş, yanlış veya uygunsuz içerikler barındırması nedeniyle teknoloji şirketleri veri kaynaklarını gizli tutmaya başlamıştır. Ancak Stable Diffusion gibi açık kaynak projeler şeffaflık yoluyla önyargıları tespit etmeyi amaçlamaktadır. Sistemin dayandığı "alt-text" açıklamaları genellikle güvenilmez olup, arama optimizasyonu için yazılmış alakasız terimlerle dolu

olabilmektedir. Bu durum, modelin kelimeler ve görseller arasında kurduğu olasılıksal bağlantılar nedeniyle önyargılı veya beklenmedik sonuçlar üretmesine yol açmakta, örneğin "Irak" kelimesiyle "savaş" kavramının sürekli ilişkilendirilmesi gibi toplumsal önyargıları pekiştirmektedir. Bu durum, Noam Chomsky'nin *Rızanın İmalatı* (2021) adlı eserinde vurguladığı yanlı medya içeriğinin büyük miktarlarda üretilmesi ve bu çok fazla içeriğin günümüzde yapay zekâ sistemlerine yansıması sorununu da gözler önüne sermektedir.

Görsel 3. "Irak'ta Oyuncaklar" komutuna verilen görsel yanıtlar. (Kaynak: Fitts Vd. 2025).

Trump'ın Gazze temalı videosu, etnik temizlik gibi ağır bir önermeyi mizahi bir bağlama yerleştirerek izleyicinin eleştirel değerlendirme kapasitesini sınırlandırmaktadır. MIT'den Dr. Timnit Gebru ise yapay zekanın eğitilmesi noktasında önyargıların politik söylemi yapısal şiddete dönüştürdüğünü belirtmektedir (Gebru, 2024, s. 49). Model, Twitter'ın problematik veri setinden beslenerek ırkçılık ve komplo teorilerini otomatikleştirmektedir. Bu durum, Frankfurt Okulu'nun "kültür endüstrisi" kavramını anımsatsa da onun ötesinde kültürü ve bilinci daha kapsamlı etkileme potansiyeline sahiptir. Kapitalizmin tüketim kültürüyle bireyleri etkilemesine (Adorno, 2021, s. 35) benzer şekilde, yapay zekâ da politik gerçekliği eğlence unsuruna indirgeme eğilimindedir.

Metinden Görüntüye Yanlılık: Yapay Zekâ Üretiminde Sistematik Ayrımcılık

YZ modellerinin hızla gelişen bir alanı olan Metinden Görüntüye (T2I) sistemleri, adından da anlaşılacağı gibi, yazılı bir metin açıklamasını (prompt) alıp bu

açıklamaya uygun bir görsel veya resim üretme yeteneğine sahiptir. Yani, tıpkı bir ressama "Kırmızı şapkalı, güneşli bir günde parkta oturan bir kişi çiz" dediğinizde size bir resim yapması gibi, T2I modelleri de sizin yazılı komutunuzu anlayıp dijital olarak bu komutun görsel temsilini oluşturur.

"T2I" kısaltması, İngilizce'deki "Text-to-Image" ifadesinin baş harflerinden gelir. Bu, modelin metin girdisini (Text) görüntü çıktısına (Image) dönüştürme fonksiyonunu doğrudan yansıtır. Bu modeller, internetten toplanan devasa metin-görüntü çiftleri (yani, resimler ve onların açıklamaları) üzerinde eğitilir. Bu eğitim sayesinde model, belirli kelimelerin ve kelime kombinasyonlarının görsel dünyada nasıl göründüğünü öğrenir ve yeni metin komutları aldığında bu bilgiyi kullanarak daha önce görmediği görseller üretebilir (Naik & Nushi, 2023).

Bu modellerin başarısı etkileyici olsa da eğitildikleri büyük veri kümelerinin kendisi genellikle toplumdaki var olan yanlılıkları (bias) içerir. Sonuç olarak, model bu verilerdeki kalıpları öğrenir ve ürettiği çıktılara bu yanlılıklar sızabilir. Bu durum, modellerin ürettiği görsellerde "belirgin sosyal yanlılıklar sergileyebileceği" anlamına gelir. Bu yanlılıklar, farklı insan gruplarının temsili konusunda eşit olmayan veya basmakalıp (stereotipik) tasvirler şeklinde ortaya çıkar. Bu tür yanlılıklar "sosyal klişelerin yayılması veya güçlendirilmesi", azınlık gruplarının "yetersiz temsili veya hatta sosyal olarak marjinalleştirilmiş toplulukların silinmesiyle sonuçlanması" gibi ciddi sosyal riskler taşır (Vice & Akhtar, 2025).

Yapay zekâ modellerinde, özellikle T2I gibi karmaşık ve geniş çıktı alanlarına sahip modellerde yanlılığı ölçmek zorlu bir iştir ve tek bir standart yöntem bulunmamaktadır. Bu modellerin çıktı alanlarının neredeyse sonsuz olması nedeniyle nicel yanlılık değerlendirme metrikleri geliştirmek güçtür. Bu nedenle, T2I modellerindeki yanlılığı tam olarak değerlendirmek için tek bir metriğin yeterli olmayacağı kabul edilmektedir. Bunun yerine, yanlılığın farklı yönlerini yakalamak için çeşitli metriklerin ve yaklaşımların bir kombinasyonu kullanılmıştır.

2023-2025 dönemindeki araştırmalar, metinden görüntüye (T2I) modellerinin toplumsal önyargıları gerçek dünya seviyelerinin çok ötesinde sistematik olarak güçlendirdiğini ve cinsiyet, ırk ve kültürel boyutlarda yeni algoritmik ayrımcılık biçimleri yarattığını ortaya koymaktadır. Bloomberg'in ünlü analizi, DALL-E 2 tarafından üretilen CEO görüntülerinin %97'sinin beyaz erkekleri tasvir ettiğini, "mahkûm" görüntülerinin ise %80'inin gerçek hapishane

nüfusunun %50'sinden azını oluşturmasına rağmen daha koyu tenli bireyleri gösterdiğini bulmuştur. Bu, mevcut sosyal eşitsizliklerin 4-8 kat ötesinde yanlılık güçlendirmesini temsil etmekte ve ticari uygulamalar, eğitim materyalleri ve kültürel temsil üzerinde ölçülebilir etkiler yaratmaktadır. Kanıtlar, mevcut azaltma çabalarının yetersiz kaldığını göstermekte ve bu giderek yaygınlaşan yapay zekâ sistemlerine gömülü ayrımcılığın ölçeğini ele almak için temel mimari değişiklikler ve kapsamlı düzenleyici çerçeveler gerektirmektedir (Nicoletti & Bass, 2023).

2023-2025 dönemindeki akademik araştırmalar net bir fikir birliği oluşturmaktadır: Metinden görüntüye modellerindeki yanlılık sistematik, ölçülebilir ve tüm büyük platformlarda tutarlıdır. Luccioni ve ekibinin 96 bin görseli analiz eden 'Stable Bias' çalışması (2023), DALL-E 2, Stable Diffusion v1.4 ve v2 gibi yaygın metinden-görüntüye modellerinin sistematik önyargılar taşıdığını ortaya koymuştur. Bu modeller, 'CEO' veya 'yönetici' gibi komutlarda %97 oranında yalnızca beyaz erkek üreterek kadınlar ve etnik azınlıklar gibi marjinalleşmiş grupları yok saymaktadır. (Luccioni vd., 2023).

Chinchure ve arkadaşları (2024) tarafından geliştirilen TIBET[2] çerçevesi, önceden tanımlanmış kategorilere güvenmek yerine yanlılık boyutlarını otomatik olarak tanımlayan karşı olgusal akıl yürütme yaklaşımları sunmuştur, kapsamlı kullanıcı çalışmaları ile doğrulanan karmaşık çok boyutlu yanlılıkları ve kesişimsellik etkilerini başarıyla ortaya çıkarmıştır. Bu metodolojik ilerleme, geleneksel değerlendirme yöntemlerinin kaçırdığı ince yanlılık kalıplarının tespitini mümkün kılmaktadır.

BIGbench'in 47.040 komut üzerinden 8 modelin birleşik değerlendirmesi, damıtılmış modellerin temel modellerden önemli ölçüde daha kötü yanlılık gösterdiğini, SDXL'in genel olarak en iyi performans sergilediğini ancak damıtılmış versiyonlarının (SDXL-L, LCM-SDXL) güçlendirilmiş yanlılıklar sergilediğini ortaya koymuştur. Kritik olarak, çalışma alakasız korumalı özniteliklerin eklenmesinin beklenmedik şekillerde yanlılığı artırabileceğini

[2] TIBET (Metin-Görüntü Önyargı Değerlendirme Aracı), yapay zekâ tarafından üretilen video içeriklerindeki sosyal önyargıların sistematik analizi için geliştirilmiş bir çerçevedir. Dinamik önyargı ekseni yaklaşımı, önyargının statik değil; bağlama, zamana ve kültürel değişkenlere göre evrilebilen çok boyutlu bir olgu olduğu varsayımına dayanır. Bu metodoloji, geleneksel içerik analizinden farklı olarak, önyargıyı tek bir kategoride (ör. yalnızca cinsiyet) değil, eksenler arası etkileşimler (ör. "cinsiyet-meslek dağılımı", "etnisite-mekân ilişkisi") üzerinden dinamik bir şema ile haritalandırır. Temel amaç, AI modellerinin eğitim verilerinden devraldığı ve üretim sürecinde pekiştirdiği yapısal dengesizliklerin nicel olarak saptanmasıdır.

bulmuştur—ırk terimlerinin eklenmesi aslında cinsiyet yanlılığını artırabilir (Hanjun Luo, 2024).[3]

Büyük yapay zekâ araştırma kurumları sistematik sorunları kabul etmiştir. OpenAI azaltma tekniklerinden sonra çeşitlilikte 12 kat iyileşme iddialarını belgelemiştir, Google'ın Imagen makalesi ise "daha açık ten tonlarına yönelik genel yanlılık ve Batılı cinsiyet stereotipleri dahil olmak üzere çeşitli sosyal yanlılık ve stereotipler" olduğunu kabul etmiştir. Meta AI Research temsili adalet konusunda çalışmalara katkıda bulunmuş, ancak somut iyileştirmeler sınırlı kalmıştır (Saharia & Chan, 2022).

AIES, ICCV, FAccT ve diğer büyük konferanslardan gelen katkılar, ticari ve açık kaynak modellerinde tutarlı yanlılık kalıplarını göstermektedir. Zhou ve arkadaşlarının çoklu model analizi, kadınların daha genç ve gülümser olarak tasvir edilirken erkeklerin daha yaşlı ve nötr veya kızgın ifadelerle göründüğü ince yanlılık kalıplarını ortaya koymuştur—yanlılığın mesleğin ötesinde temel temsil özelliklerine uzandığını göstermektedir (Zhou vd, 2024).

2023-2025 dönemindeki büyük ölçekli kantitatif çalışmalar, demografik gruplar arasındaki yanlılığı ölçmek için giderek sofistike istatistiksel çerçeveler kullanmıştır. Microsoft Research'ün AIES 2023 çalışması komut başına 500 görüntüde DALL-E v2'nin CEO olarak %0 kadın ve hizmetli olarak %100 kadın ürettiğini bulmuştur, Stable Diffusion ise %34 gerçek dünya temsiline karşı hâkim olarak sadece %3 kadın göstermiştir (Naik & Nushi, 2023).

En kapsamlı cinsiyet yanlılığı analizi, 5 önde gelen T2I modelinde 3.217 cinsiyet-nötr komuttan 2.293.295 görüntüyü incelemiş, geleneksel cinsiyet rolü pekiştirmesini ortaya koyarak kadınların finansal faaliyetlerde %25-40 oranında yetersiz temsil edildiğini ve erkeklerin teknik/fiziksel emekte %35-50 oranında aşırı temsil edildiğini bulmuştur. Bu devasa ölçek, demografik kategoriler arasında yanlılık ölçümleri için istatistiksel anlamlılık sağlamaktadır (Girrbach vd., 2025).

Bloomberg'in 5.100 görüntünün detaylı mesleki analizi, Fitzpatrick skalası ten tonu sınıflandırmasını kullanarak "fast-food çalışanı" için %70 koyu ten tonuna karşı %30 gerçek dünya demografisi ve "sosyal çalışan" için %35 gerçek

[3] SDXL, özellikle yapay zekâ alanında metinden görüntüye (Text-to-Image- T2I) üretim modellerinden biridir. SDXL, Stable Diffusion serisinin gelişmiş bir versiyonudur. LCM, ise "Latent Consistency Model" (Gizil Tutarlılık Modeli) anlamına gelir. Bu, difüzyon modellerini hızlandırmak için geliştirilmiş bir tekniktir. Modelin farklı adımlarda tutarlı sonuçlar üretmesini sağlayan teknik. "Latent" = Görüntünün sıkıştırılmış (encoded) halinde çalışma.

temsile karşı %68 bulmuştur. Çalışmanın titiz metodolojisi ayrımcı kalıplar için net istatistiksel anlamlılık kurmuştur (Nicoletti & Bass, 2023).

Teknik analizler, popüler yapay zekâ modellerinin (Stable Diffusion, DALL-E gibi) eğitiminde kullanılan LAION-5B[4] veri kümesinin ciddi sorunlar barındırdığını göstermektedir. Bu 5,85 milyar görüntü-metin çiftinden oluşan devasa veri kümesi, İngilizce açıklamaların ezici çoğunlukta olması ve coğrafi temsilin ağırlıklı olarak ABD ve Avrupa merkezli olması nedeniyle sistematik Batı yanlılığı içermektedir. Daha da endişe verici olan, Stanford İnternet Gözlemevi'nin bu veri kümesi içinde çocuk istismarı ve şiddet gibi 3.226 şüpheli yasadışı içerik tespit etmesi, demografik önyargıların ötesinde temel veri kalitesi ve etik sorunlarının varlığını ortaya koymasıdır (Nicoletti & Bass, 2023).

CLIP[5] model mimarisi yanlılıkları tüm bağımlı T2I sistemlerine yayılmaktadır, Michigan Üniversitesi araştırması, düşük gelirli ve Batı dışı yaşam tarzlarını tasvir eden görüntülerde sistematik yetersiz performans göstermektedir (DeLacey, 2023). CLIP gömmeleri toplumsal stereotipleri kodlamakta, kravatları, arabaları ve aletleri erkeklerle ilişkilendirirken cinsiyet, ırk ve yaş boyutlarında ölçülebilir yanlılık sergilemektedir.

Web'den kazınmış veri filtreleme, marjinalleşmiş topluluklardan orantısız olarak içerik kaldıran NSFW[6] ve içerik filtreleme sistemleri aracılığıyla ek yanlılık getirmektedir. Otomatik filtreleme sistemleri kültürel ve demografik yanlılıklar sergilerken, dil tabanlı filtreleme İngilizce içeriği büyük ölçüde kayırarak eğitim süreci boyunca bileşik yanlılık etkileri yaratmaktadır.

Difüzyon modelleri ile GAN'ların[7] karşılaştırmalı analizi, difüzyon mimarilerinin eğitim verisindeki dağılım yanlılığını daha şiddetli olarak

[4] LAION-5B, günümüzde kullanılan çoğu büyük Text-to-Image modelinin (Stable Diffusion, DALL-E vb.) temelini oluşturan devasa bir görüntü-metin veri kümesidir. "Large-scale Artificial Intelligence Open Network- 5 Billion" İçerik: 5,85 milyar görüntü-metin çifti Kaynak: İnternetten otomatik kazınmış (web scraping) Erişim: Açık kaynak (herkes kullanabilir).

[5] CLIP, OpenAI tarafından 2021'de geliştirilen ve metinlerle görüntüleri birlikte anlayan çok modlu (multimodal) bir yapay zekâ modelidir. "Contrastive Language-Image Pre-training" (Karşıt Dil-Görüntü Ön-Eğitimi) anlamına gelir.

[6] NSFW, "Not Safe For Work" (İş Yeri İçin Güvenli Değil) anlamına gelen bir internet kısaltmasıdır. NSFW içerik, iş yerinde, okul ortamında veya kamusal alanlarda görüntülenmesi uygun olmayan dijital içerikleri ifade eder.

[7] GAN'lar (Generative Adversarial Networks- Üretken Düşman Ağları), 2014 yılında Ian Goodfellow tarafından geliştirilen ve iki yapay sinir ağının sürekli rekabet halinde olduğu bir yapay zekâ mimarisidir. Sistem, sahte veri üreten "Generator" ile gerçek ve sahte veriyi ayırt etmeye çalışan "Discriminator" arasındaki "sahtekâr vs polis" benzeri bir mücadeleye dayanır- Generator giderek daha gerçekçi sahte içerik üretmeyi öğrenirken, Discriminator da sahteciliği

güçlendirdiğini ortaya koymuştur, özellikle daha küçük veri kümeleri için. Bu bulgu, daha yeni, daha yüksek kaliteli üretken modellerin mutlaka daha iyi yanlılık özelliklerine sahip olduğu varsayımlarına meydan okumaktadır (Perera & Patel, 2023).

Bloomberg'in 2023 analizi, yüksek ücretli işlerin ağırlıklı olarak beyaz erkekler ve düşük ücretli işlerin büyük ölçüde renkli kadınlar olarak üretildiği aşırı mesleki stereotipleri belgeleyen bir dönüm noktası vakası olmuştur. Çalışmanın "mahkûm" görüntülerinin %80'ının gerçek hapishane nüfusunun %50'sinden azına karşı koyu tenli bireyleri tasvir ettiği bulgusu, ırksal stereotiplerin sistematik güçlendirmesini göstermektedir (Nicoletti & Bass, 2023).

Google Gemini'nin Şubat 2024 tartışması, farklı ırklardan çeşitli Nazi askerleri üretmesi ve birçok bağlamda beyaz insanların görüntülerini üretmeyi reddetmesinin ardından CEO'nun özür dilemesine ve 96,9 milyar dolarlık piyasa değeri kaybına yol açmıştır. Olay, bağlam farkındalığı olmadan yanlılık düzeltmesinin zorluklarını vurgulamış ve geçici hizmet askıya alınmasına yol açmıştır (Baum & Villasenor, 2024).

Mobley v. Workday vakası ise yapay zekâ işe alım ayrımcılığını iddia eden ilk büyük toplu dava olarak yasal emsal oluşmuştur, mahkemeler 2024'te değiştirilmiş şikayetlerin devam etmesine izin vermiştir. EEOC'nin Ağustos 2023'teki ilk yapay zekâ işe alım ayrımcılığı anlaşması, mevcut medeni haklar yasaları için uygulama emsal oluşturmuştur (Wiessner, 2024).

Ticari etki çalışmaları, DataRobot'un 2024 anketine göre şirketlerin %62'sinin önyargılı yapay zekâ kararları nedeniyle gelir kaybettiğini göstermektedir, ekonomik analiz ise algoritmalardaki ırksal yanlılığın 1,5 trilyon dolarlık kayıp GSYİH[8] potansiyeline yol açabileceğini öne sürmektedir. Günlük 34+ milyon yapay zekâ görüntüsü üretilmekte ve 2025'e kadar pazarlama içeriğinin %30'unun üretken yapay zekâ kullanması öngörülmekteyken,

tespit etmede ustalaşır. Bu süreç, her iki ağ da birbirini yenene kadar devam eder ve sonunda Generator o kadar gelişir ki, Discriminator'ı kandıracak kadar gerçekçi içerik üretebilir. T2I modelleri bağlamında GAN'lar özellikle önemlidir çünkü araştırmalar, difüzyon modellerinin (DALL-E, Stable Diffusion) eğitim verisindeki yanlılığı 4-8 kat güçlendirdiğini, GAN'ların ise aynı veriyi daha az güçlendirdiğini göstermiştir. Ford gibi şirketler, otonom araç geliştirmede veri çeşitliliği eksikliğini gidermek için GAN'ları kullanarak sentetik sürüş senaryoları üretmekte, böylece "çok küçük veya ağır yanlı" veri kümelerindeki sorunları çözmeye çalışmaktadır.

[8] Gayri Safi Yurtiçi Hasıla. Bir ülkenin belirli bir dönemde (genellikle bir yıl) ürettiği tüm mal ve hizmetlerin toplam değeridir. İngilizce'de GDP (Gross Domestic Product) olarak bilinir.

potansiyel ayrımcılığın ölçeği genişlemeye devam etmektedir (Baum & Villasenor, 2024).

Araştırmalar, stereotipik yapay zekâ üretimi görüntülerin üretilmesinin yanlılıkları güçlendirdiğini göstermektedir. Bu bulgu, T2I yanlılığının anlık görsel temsilin ötesinde kalıcı psikolojik etkiler yarattığını ortaya koymaktadır. Yetersiz temsil edilen topluluklar bileşik ayrımcılık yaşamaktadır, Black Girls Code 'un[9] T2I araçlarında yetersiz temsilin Siyah ve Kahverengi kadınlar ve kızlar için mesleki engeller yarattığını bildirmesiyle. Çalışmalar, bütün kültürleri Batılı algılanan arketiplere indirgeyen "karikatür" stereotipleri aracılığıyla sistematik yanlış temsili edildiğini belgelemektedir (Nicoletti & Bass, 2023).

Bilindiği gibi internet içeriğinin büyük kısmı, belirli sosyoekonomik, coğrafi ve kültürel gruplar tarafından üretilmektedir. Bu durum, küresel çeşitliliği temsil etmesi gereken yapay zekâ sistemlerinin, aslında dar bir demografik kesimin dünya görüşünü evrenselleştirmesine yol açmaktadır. Bu nedenle internet içeriği, tarihsel ayrımcılık dönemlerindeki temsil kalıplarını barındırmaktadır. Yapay zekâ bu veriyi "nesnel gerçek" olarak işlediğinde, geçmişin adaletsizliklerini gelecek nesiller için kalıcılaştırmaktadır.

Marjinalleşmiş gruplar ya tamamen görünmez kılınmakta ya da sadece stereotipik bağlamlarda temsil edilmektedir. Bu durum, bu grupların toplumsal varlığını ve çeşitliliğini yok saymakta, onları tek boyutlu karikatürler haline getirmektedir. Bu ahlaki çöküş, yapay zekâ teknolojisinin demokratik değerlerle uyumlu olmayan bir şekilde geliştirildiğini ve toplumsal adaletsizlikleri teknolojik araçlarla yeniden ürettiğini göstermektedir. Trump videosu gibi örnekler, bu sistemlerin uzun vadede ideolojik sonuçları nedeniyle herkesin zarar görebileceğini ortaya koymaktadır.

Vaka Analizi: Trump'ın Gazze Videosu

Yöntem

Bu çalışmada, Chinchure ve arkadaşlarının geliştirdiği TIBET (Text-to-Image Bias Evaluation Tool) metodolojisi kullanılarak Donald Trump'ın yer aldığı bir video içeriği kapsamlı önyargı analizi açısından incelenecektir. Uygulanan yöntemde öncelikle Trump'ın paylaştığı videonun dökümü (betimlenmesi) yapay zekâ modelinden istenmiştir. Betimlenen video içeriği sonrası bu

[9] Black Girls Code, teknoloji alanındaki kariyerlerini geliştirmek için Afrika kökenli Amerikalı kızları ve diğer siyahi gençleri bilgisayar programlama eğitimiyle buluşturmaya odaklanan kâr amacı gütmeyen bir kuruluştur.

sahnelerin tekrar nasıl oluşturulabileceği ve söz konusu sahnelerin karşıtının ne olabileceği sorulmuştur. Verilen cevaba göre video oluşturma aracı Veo ’dan bu sahnelerin yeniden oluşturulması istenmiştir.

Dil modeli bu sahnelerin oluşturulması için sinematik çekim, 4k, hipergerçek gibi teknik komutları kendisi eklemiştir. Bu yöntemde amaçlanan ise araştırmacı olarak yönlendirme ve müdahale etkisini ortadan kaldırmak ve somut bir veriye ulaşmaktır. TIBET'in dinamik önyargı ekseni tespiti ve karşıt olgusal değerlendirme yaklaşımı, Trump figürünün video temsilindeki potansiyel önyargıları sistematik olarak ortaya çıkarmak için uygulanacaktır. Bu analiz sürecinde, videonun temsili karelerinden örneklemeler alınarak, politik figür temsili, liderlik önyargıları, yaş ve cinsiyet stereotipleri, sosyoekonomik statü gösterimleri ve kültürel/coğrafi önyargılar gibi çok boyutlu önyargı eksenleri TIBET'in LLM[10] destekli yaklaşımıyla dinamik olarak belirlenecektir. Ardından, bu önyargı eksenleri için karşıt olgusal senaryolar oluşturularak (örneğin Trump yerine farklı demografik özelliklere sahip politik figürlerin kullanılması), sadece VQA[11] tabanlı kavram çıkarımı yapılacaktır. Bu yöntemde kullanılan Concept Association Score (CAS[12]) hesaplamaları ve önyargıların nicel olarak ölçülmesini ve Mean Absolute Deviation (MAD[13]) skorları aracılığıyla önyargı derecelerinin karşılaştırmalı analizi çalışmanın kapsamı nedeniyle yapılamamıştır. Çünkü CAS hesaplaması, her video karesi için yüzlerce

[10] LLM (Large Language Model), geniş ve çeşitli veri kümeleri üzerinde eğitilmiş, dil anlama ve üretme yeteneğine sahip büyük ölçekli derin öğrenme modelleridir. Bu modeller, genellikle kendi kendine denetimli öğrenme hedefleriyle eğitilir. LLM'ler, doğal dil talimatlarını takip etme konusunda güçlü yetenekler sergilemişlerdir ve genel amaçlı bir asistan olarak kullanılabilen evrensel bir arayüz görevi görebilirler. GPT-4, LLama2 gibi örnekler, büyük ölçekli dil modelidir.

[11] VQA (Visual Question Answering-Görsel Soru Cevaplama): bir görüntü hakkında sorulan soruları anlama ve bu sorulara doğal dilde yanıt verme yeteneğine sahip yapay zekâ modelleridir. Yani, bir resim gösterip “Bu resimde kedi ne giyiyor?” gibi bir soru sorduğunuzda, VQA modeli görüntüyü analiz eder ve örneğin “gözlük” gibi bir cevap verir

[12] CAS (Concept Association Score - Kavram İlişkilendirme Puanı): Bir modelin belirli bir kavramı (örneğin, “doktor” kavramını) genellikle hangi özelliklerle (örneğin, cinsiyet, yaş veya ırk) ilişkilendirdiğini ölçmeye yarar. Modelin belirli bir istemden (örneğin “bir doktorun fotoğrafı”) ürettiği görüntülerdeki kavramları VQA gibi bir yöntemle çıkarırsınız, ardından bu çıkarılan kavramların, ilgili istemin farklı “bozulmuş” (biased) versiyonlarıyla (örneğin “erkek doktor” veya “kadın doktor”) ne kadar ilişkili olduğunu ölçersiniz. CAS, bu ilişkinin derecesini, yani modelin belirli bir kavramı belirli özelliklerle ne kadar güçlü bir şekilde “ilişkilendirdiğini” veya “bağdaştırdığını” gösterir.

[13] Mean Absolute Deviation (MAD) skorları, TIBET metodolojisinde önyargı derecelerinin nicel olarak ölçülmesi için kullanılan temel metriktir. MAD, orijinal video içeriği ile karşıt olgusal videolar arasındaki kavramsal farklılıkların ortalama mutlak sapmasını hesaplar. Ortalama Mutlak Sapma (MAD), bir kümedeki her veri noktası ile o kümenin ortalaması arasındaki ortalama mesafeyi niceliksel olarak belirten istatistiksel bir ölçüdür.

kavramın (liderlik, modernlik, refah, güvenlik vb.) 0-1 arasında skorlanmasını gerektirir. Bu süreç için özel veri setleri ve eğitilmiş modeller gerektirir. MAD hesaplaması için her önyargı ekseni boyunca sistematik olarak karşıt olgusal videolar üretilmesini gerektirir. Bu, Google Veo gibi gelişmiş video üretim araçlarına sürekli erişim ve binlerce alternatif video üretimi anlamına gelir.

Bu süreçte TIBET (Text-to-Image Bias Evaluation Tool) metodolojisi aşağıdaki adımlarla sistematik olarak uygulanacaktır:

1. Trump videosunun sahne dökümleri, yapay zekâ modelleri kullanılarak oluşturulması.

2. Belirlenen sahnelerin yeniden üretimi ve bu sahnelere karşıt alternatifler, "karşıt olgusal senaryolar" yöntemiyle gerçekleştirilmesi.

3. Videonun sahnelerinden seçilen kareler üzerinden, kavram çıkarımı (VQA yöntemiyle) gerçekleştirilerek önyargı eksenleri sistematik olarak belirlenmesi.

Böylece yöntem, araştırmacı müdahalesini minimize ederek, daha somut ve tutarlı bulgulara ulaşmış olacaktır.

Veri Toplama Süreci ve Analiz

1. **Dinamik Önyargı Ekseni Belirleme:** Google "Gemini" gibi büyük dil modelleri kullanılarak, Trump video içeriği için potansiyel önyargı eksenleri otomatik olarak tespit edilecektir. Bu eksenler politik temsil, liderlik stereotipleri, yaş ayrımcılığı, cinsiyet algısı, sosyoekonomik statü gösterimleri ve kültürel önyargılar gibi çok boyutlu kategorileri kapsayacaktır.
2. **Karşıt Olgusal Video Üretimi:** Belirlenen her önyargı ekseni için Google Veo aracı kullanılarak karşıt olgusal videolar üretilecektir. Örneğin, Trump yerine farklı yaş, cinsiyet, etnik köken veya politik görüşlere sahip figürlerin yer aldığı alternatif senaryolar oluşturulacaktır.
3. **Video Kare Örneklemesi:** Orijinal ve karşıt olgusal videolardan temsili kareler sistematik olarak örneklenecek ve her video için tutarlı sayıda kare analiz edilecektir.
4. **VQA Tabanlı Kavram Çıkarımı:** Video karelerinden MiniGPT-v2 gibi görsel-dil modelleri kullanılarak kavramsal öğeler çıkarılacak ve önyargı eksenlerine özgü sorular sorulacaktır.

Şimdiye kadar ortaya konulmuş olan teorik çerçevenin pratik uygulamasına geçmek amacıyla, Donald Trump'ın sosyal medya hesabından

paylaştığı Gazze ile ilgili video içeriği TIBET metodolojisi kullanılarak kapsamlı bir önyargı analizi sürecine tabi tutulacaktır. Bu video seçimi hem politik figür temsilinin hem de hassas jeopolitik konuların kesişiminde yer alan karmaşık önyargı dinamiklerini inceleme fırsatı sunmaktadır. Gazze krizi gibi çok boyutlu ve duygusal yüklü bir konu bağlamında üretilen video içeriği, etnik/dini önyargılar, politik stereotipler, çatışma temsil önyargıları ve insani söylem önyargıları gibi çeşitli yanlılık türlerinin tezahürü için zengin bir analiz ortamı sağlamaktadır. TIBET'in dinamik önyargı ekseni belirleme yaklaşımı, bu kompleks video içeriğindeki örtük ve açık önyargıları sistematik olarak tespit etmek için Google Veo aracı ile entegre edilerek uygulanacak, böylece video temelli önyargı derecelendirmesi metodolojik olarak somut bir vaka üzerinden değerlendirilecektir.

Video İçeriği Analizi: Sahne Dökümü ve İlk Gözlemler

Video, yıkılmış binalar ve enkaz arasında yürüyen insanların görüntüleri ile başlar ve ardından "Sıradaki ne?" sorusu ekranda belirir. Video, izleyiciye kasvetli ve umutsuz bir tablo çizerek başlar ve ardından bu tablonun tam zıttı bir geleceğe keskin bir geçiş yapar.

Görsel 4. Video'ya erişim için video QR kodu.

1. Açılış: Yıkım ve Tünel

Video, günümüz Gazze'sini tasvir eden, enkaz ve yıkılmış binalar arasında tek başına yürüyen bir adamın görüntüsüyle başlar. Hemen ardından, karanlık bir tünelin içinden çıkan bir grup çocuğun olduğu sahneye geçilir. Çocuklar tünelin karanlığından, parlak güneş ışığına ve yemyeşil bir manzaraya doğru adım atarlar. Bu sahne, umutsuzluktan aydınlık bir geleceğe geçişi simgeler.

2. "Sıradaki Ne?" Sorusu

Bu sahnelerin ardından ekranda büyük harflerle "WHAT NEXT?" (Sıradaki Ne?) sorusu belirir.

3. Dönüşen Gazze

Soruya cevap olarak video, tamamen dönüşmüş, fütüristik bir Gazze vizyonu sunar. Bu bölümde:

• Pırıl pırıl, modern ve devasa gökdelenlerin bulunduğu bir sahil şeridi gösterilir.

• Mutlu çocuklar kumsalda oynar ve denizde yüzer.

• Palmiye ağaçlarıyla süslenmiş geniş caddelerde lüks Tesla otomobilleri dolaşır.

4. Tanıdık Figürler ve Absürt Sahneler

Video, yapay zekâ ile oluşturulmuş bazı tanınmış figürleri ve absürt durumları içerir:

Elon Musk Benzeri: Girişimci Elon Musk'a çok benzeyen bir figür, bir masada oturmuş, gülümseyerek humus gibi görünen bir yiyeceğe pide batırır. Daha sonra aynı figürün, neşe içindeki bir kalabalığın üzerine tomarla para saçtığı görülür.

Dans Eden Militanlar: Sakallı Hamas militanları olarak tasvir edilen figürler, bikiniler ve oryantal dansöz kıyafetleri içinde göbek atarak dans ederler.

5. Trump Odaklı Gelecek

Videonun ilerleyen kısımlarında odak noktası tamamen Donald Trump'a ve onun ismine kayar:

• *Altın Balonlu Çocuk*: Küçük bir çocuk, üzerinde büyük harflerle "TRUMP" yazan devasa bir altın balonu tutmaktadır.

• *"Trump Gazze" Tatil Köyü*: Lüks binaların ve tatil tesislerinin üzerinde "TRUMP GAZA" yazılı parlak tabelalar görülür.

• *Gece Kulübünde Trump*: Donald Trump, bir gece kulübünde genç bir kadınla neşe içinde dans ederken tasvir edilir.

• *Güneşlenen Liderler*: Trump ve İsrail Başbakanı Binyamin Netanyahu, "Trump Gazze" sahilinde şezlonglara uzanmış güneşlenirken gösterilir.

• *Altın Heykeller*: Şehrin çeşitli yerlerinde Trump'ın devasa altın heykelleri yer alır.

6. Kapanış Mesajı

Tüm bu görüntülere, "Trump Gazze parlıyor, altın gelecek, yepyeni bir hayat" gibi sözler içeren, neşeli ve akılda kalıcı bir pop şarkısı eşlik eder. Video, Gazze'nin Trump'ın vizyonuyla lüks bir tatil cennetine, bir "Orta Doğu Rivierası"na[14] dönüşeceği mesajını vererek sona erer.

Bu sahneleri oluşturmak için Gemini tarafından önerilen prompt'lar aşağıdaki gibi olmuştur:

"Bu sahneleri yapay zekâ video veya görsel oluşturma araçlarında (Midjourney, Sora, DALL-E, Runway vb.) canlandırmak için kullanılabilecek prompt'ları (komutları/istekleri) yazalım."

"Prompt'lar genellikle İngilizce yazıldığında daha iyi sonuçlar verir, bu yüzden hem İngilizce orijinalini hem de Türkçe açıklamasını ekleyeceğim."

Sahnelerin Oluşturulması

Sahne 1: Yıkım ve Tünel

Bu sahne iki ayrı çekimden oluşuyor. Dolayısıyla iki farklı prompt seti gerektirir.

Çekim 1A: Yıkıntılar Arasında Yürüyen Adam

Bu çekimdeki amaç, umutsuzluk, yalnızlık ve yıkımı hissettirmektir.

Seçenek 1 (Basit Prompt):

• **Komut:** Yıkılmış bir şehrin enkazı arasında yürüyen yalnız bir adam, belgesel tarzında, kasvetli bir ruh hali.

Seçenek 2 (Detaylı ve Sinematik Prompt):

• **Komut:** Sinematik geniş açı çekim, arkadan görünen yalnız bir adam, savaşın yıktığı bir şehrin harabeleri arasında yavaşça yürüyor. Havayı toz ve duman dolduruyor, soluk renkler, bulutlu gökyüzü. Atmosfer, umutsuzluk ve ıssızlıkla ağırlaşmış. Fotogerçekçi, 8K, takip çekimi.

Neden bu kelimeler?

[14] Güzel, lüks sahil tatil bölgeleri için kullanılan genel bir terim haline gelmiştir.

- *cinematic wide shot* (sinematik geniş açı): Sahnenin büyüklüğünü ve adamın yalnızlığını vurgular.
- *solitary man / lone man* (yalnız adam): Yalnızlık temasını pekiştirir.
- *war-torn city / destroyed city* (savaşın yıktığı şehir / yıkık şehir): Yıkım ortamını net bir şekilde tanımlar.
- *dust and smoke fill the air* (hava toz ve dumanla dolu): Atmosferi daha gerçekçi ve boğucu hale getirir.
- *muted colors / somber mood* (soluk renkler / kasvetli ruh hali): Renk paletini ve duygusal tonu belirleyerek kasvetli havayı güçlendirir.
- *photorealistic, 8K* (fotogerçekçi, 8K): Görüntünün kalitesini ve gerçekçiliğini artırmak için kullanılır.

Sahne 2A: Tünelden Çıkan Çocuklar

Bu çekimdeki amaç, karanlıktan aydınlığa, umutsuzluktan umuda geçişi dramatik bir şekilde göstermektir.

Seçenek 1 (Basit Prompt):

- **Komut**: Bir grup çocuk, karanlık bir tünelden dışarı, parlak gün ışığına ve yeşil bir tarlaya doğru yürüyor.

Seçenek 2 (Detaylı ve Sinematik Prompt):

- **Komut**: Düşük açıdan sinematik çekim, farklı çocuklardan oluşan bir grup, karanlık bir beton tünelin ağzından beliriyor. Tamamen karanlıktan, güzel ve güneşli bir gökyüzünün altındaki ezici derecede aydınlık, canlı, yemyeşil bir manzaraya adım atıyorlar. Karanlık ve aydınlık arasında güçlü kontrast, mercek parlaması, umudu ve yeni bir başlangıcı simgeliyor. Hiper-gerçekçi, 4K.

Neden bu kelimeler?

- *low angle* (düşük açı): Çocukların çıkışını daha görkemli ve anıtsal gösterir.
- *emerge from the mouth of a dark, concrete tunnel* (karanlık, beton bir tünelin ağzından çıkmaları): Eylemi ve mekânı netleştirir.
- *overwhelmingly bright, vibrant, lush green landscape* (ezici derecede parlak, canlı, yemyeşil manzara): Çıktıkları yerin ne kadar zıt ve cennet gibi olduğunu vurgular.

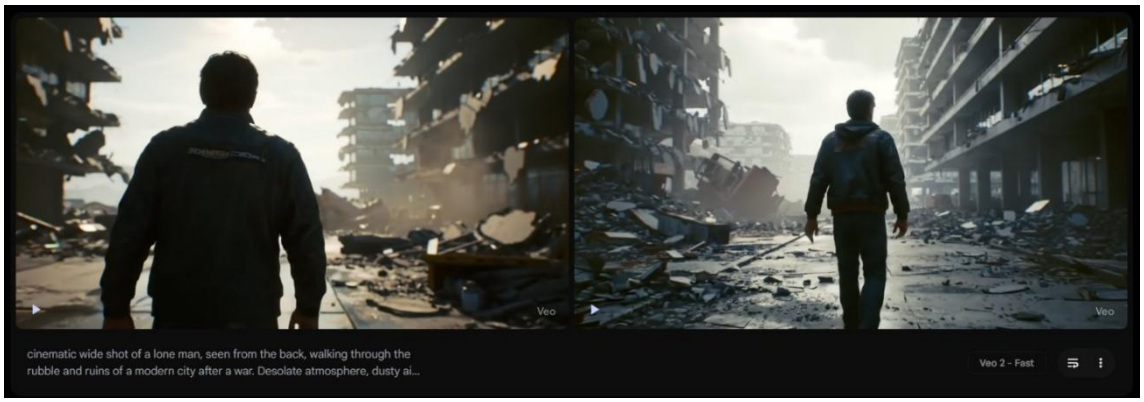
- *strong contrast between dark and light* (karanlık ve aydınlık arasında güçlü karşıtlık): Sahnenin ana sembolik fikrini (umutsuzluktan umuda) görsel olarak pekiştirir.
- *lens flare* (mercek parlaması): Güneş ışığının gücünü ve sıcaklığını artırarak sihirli bir hava katar.
- *symbolic of hope and a new beginning* (umudun ve yeni bir başlangıcın sembolü: Yapay zekaya sahnenin altında yatan temayı vererek daha isabetli bir sonuç almayı hedefler.

Görüntülerin Veo ile Oluşturulması

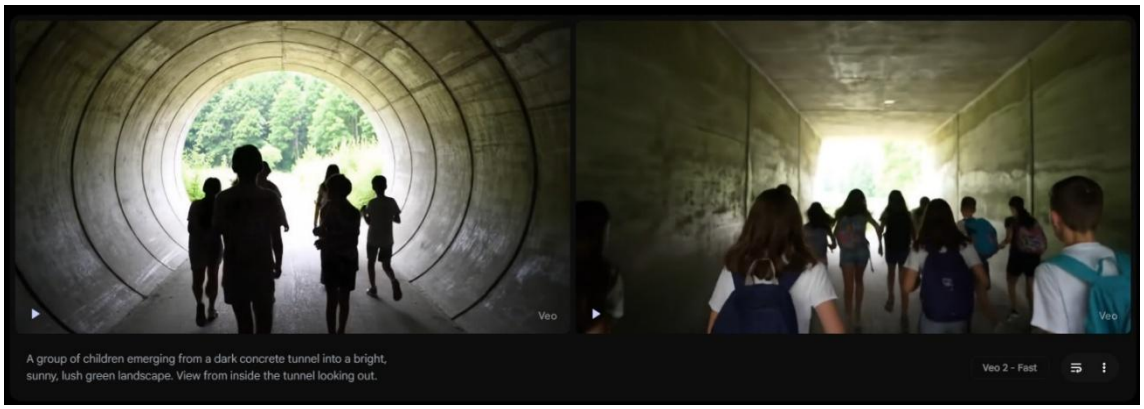
Bu detaylardan sonra ilk sahneyi oluşturmak için dil modeline sorulmuştur ve:

1. "Savaş sonrası modern bir şehrin moloz ve harabeleri arasında yürüyen, arkadan görülen tek bir adamın sinematik geniş açılı çekimi. Issız bir atmosfer, tozlu hava, kasvetli gri renk paleti. Hiper-gerçekçi, Unreal Engine 5, 8K."

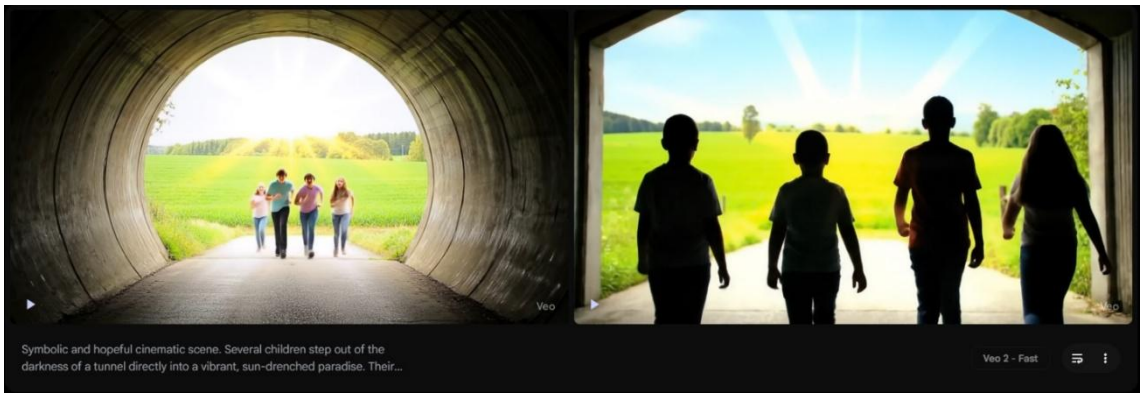
(Dil modeli) Komutunu önermiştir ve bu komuta göre üretilen video iki seçenek olarak alınmıştır.



Daha sonra: 2. "Bir grup çocuk, karanlık bir beton tünelden aydınlık, güneşli, yemyeşil bir manzaraya çıkıyor. Tünelin içinden dışarıya bakış açısı." Komutuyla iki video üretilmiştir.



Bu videonun aynısı için önerilen ve daha duygusal ve daha detaylı olarak sunulan komut ise: 3. "Sembolik ve umut dolu sinematik sahne. Birkaç çocuk, bir tünelin karanlığından doğrudan canlı, güneşe boğulmuş bir cennete adım atıyor. Yemyeşil tarlaları ilk kez gördüklerinde yüzleri huşu ile dolu. Karanlık iç mekân ile parlak dış mekân arasında yüksek kontrast" olmuştur.



Bu giriş sahnesinden sonra makalenin gerektirdiği karakter limitinden dolayı diğer sahnelerin tek tek oluşturulması yerine çözümlenme kısmına geçilecektir. Bu noktada, elde edilen video içeriği ve komut (prompt) yapıları TIBET metodolojisinin dinamik önyargı ekseni belirleme aşamasına tabi tutularak, video içeriğindeki potansiyel önyargı türlerinin sistematik analizi gerçekleştirilecektir.

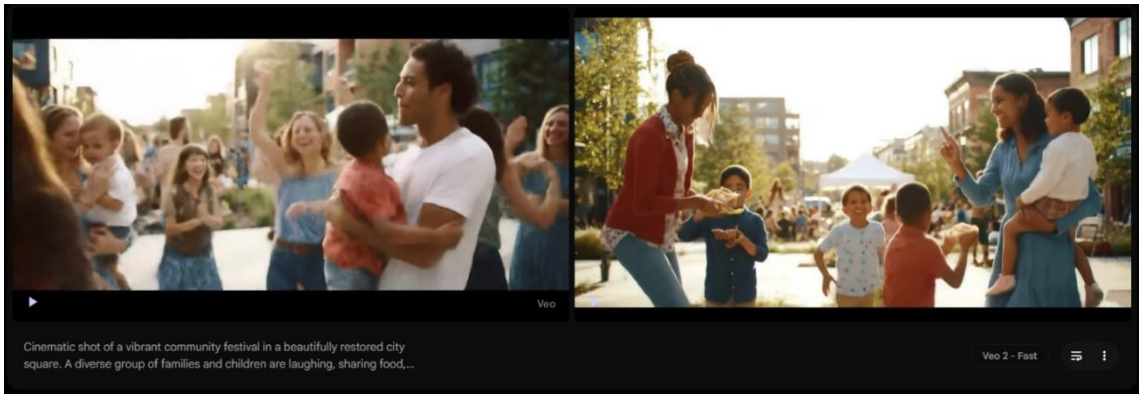
Önyargı ekseni için bu sahnelerin karşıtı olabilecek komutlar istenmiştir ve devamında şu şekilde ilerlemiştir:

Karşıt Sahne: "Yıkıntılardaki Adam" Sahnesinin Zıttı

"Önceki sahne yalnızlık, yıkım ve umutsuzluk üzerineydi. Bunun tam zıttı; topluluk, yeniden inşa ve coşkulu bir yaşam olmalıdır."

Komut: "Harika bir şekilde restore edilmiş bir şehir meydanında, canlı bir topluluk festivalinin sinematik çekimi. Çeşitli aileler ve çocuklar gülüyor, yemek

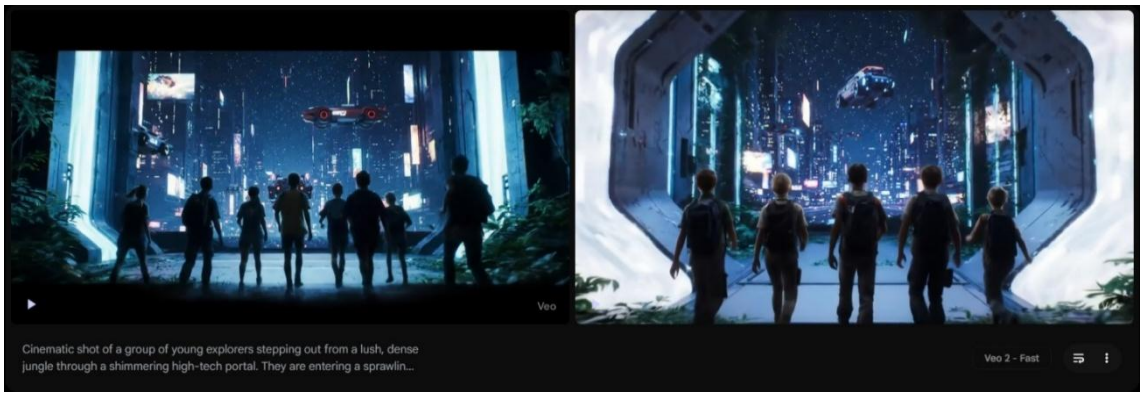
paylaşıyor ve birlikte dans ediyor. Arka planda, özenle onarılmış tarihi binalarla kusursuz bir şekilde bütünleşen pırıl pırıl yeni mimari görülüyor. Atmosfer hayat ve renkle dolu, parlak, sıcak bir güneş ışığı var. Umut dolu, neşeli, fotogerçekçi, 8K."



Karşıt Sahne: "Tünelden Çıkan Çocuklar" Sahnesinin Zıttı

"Önceki sahne doğal bir cennete kaçış ve kurtuluşu simgeliyordu. Bunun zıttı, doğadan kopuş ve yapay bir dünyaya girişi veya cennetin yok oluşunu anlatabilir."

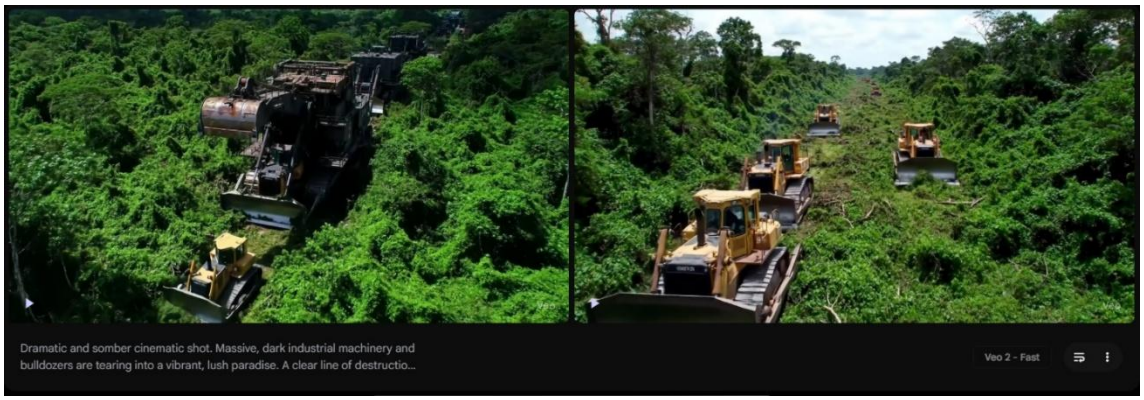
Komut: "Bir grup genç kâşifin, yemyeşil, sık bir ormandan parıldayan yüksek teknolojili bir portaldan dışarı adım atarak, geceleri neon ışıklarıyla aydınlatılmış, geniş bir siberpunk şehrine girdiği sinematik bir çekim. Hava, holografik reklamlar ve uçan araçlarla dolu. Karanlık, organik orman ile canlı, sentetik şehir arasında yüksek kontrast. Hayranlık ve fütürizm atmosferi, fotogerçekçi, 8K."



Karşıt Sahne: "Cennete Kaçış" Sahnesinin Zıttı

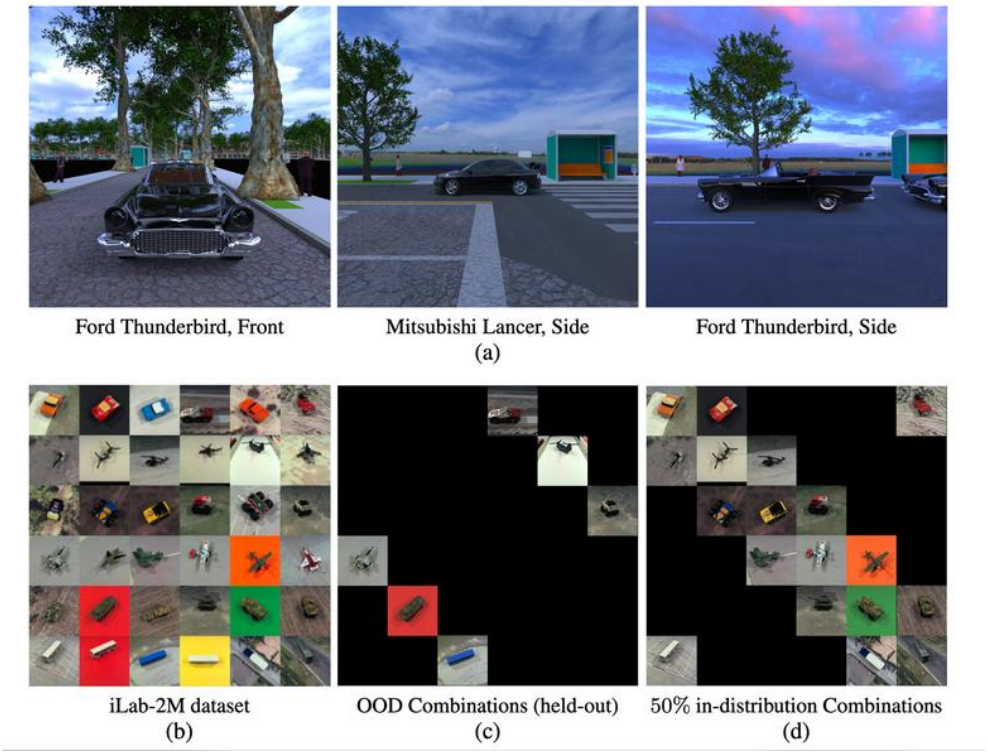
Komut: "Dramatik ve kasvetli sinematik çekim. Devasa, karanlık endüstriyel makineler ve buldozerler, canlı, yemyeşil bir cenneti parçalıyor. Net bir yıkım hattı, endüstrinin el değmemiş, renkli bir ormanı yok edişini gösteriyor. Yaşayan

doğa ile ölü, metalik makineler arasında yüksek kontrast. Kaybın sembolü, çevresel tema, hiper-gerçekçi, 8K."



Bulgular ve Yorum

Örneklerde görüldüğü gibi yapay zekâ prompt olarak girilen içeriği oluşturabilmesi için istenen bir dizi komuta bağımlı olarak çalışmaktadır. Sinematik geniş açı, düşük açı, 8K gibi teknik terimler, gösterilecek içeriğin ilgi çekici ve izlenebilir olması adına sağlanan teknik terimlerdir. Sinema filmleri bu teknik terimlere bağımlı olarak çalıştıkları için ve istenen mesajı veya duyguyu yansıtabilmek için bu teknikleri iyi kullanabilmeleri sinema filminin başarısı açısından önemlidir. Yapay zekâ bir görüntü oluştururken, bu teknik terimlere göre bir gruplandırma ve onun içinden seçim yapmaktadır. Örneğin araştırmacılar bir modeli görüntüdeki arabaları sınıflandırmak için eğitiyorlarsa, modelin farklı arabaların neye benzediğini öğrenmesini isterler. Ancak eğitim veri setindeki her Ford Thunderbird önden gösteriliyorsa, eğitilmiş modelin yandan çekilmiş bir Ford Thunderbird görüntüsü verildiğinde, milyonlarca araba fotoğrafı üzerinde eğitilmiş olsa bile onu yanlış sınıflandırabileceği belirtilmiştir.



Görsel 5. Ford Thunderbird aracın oluşturulması için farklı cephelerden oluşturulan kombinasyonlar (Kaynak: Zewe, 2022)

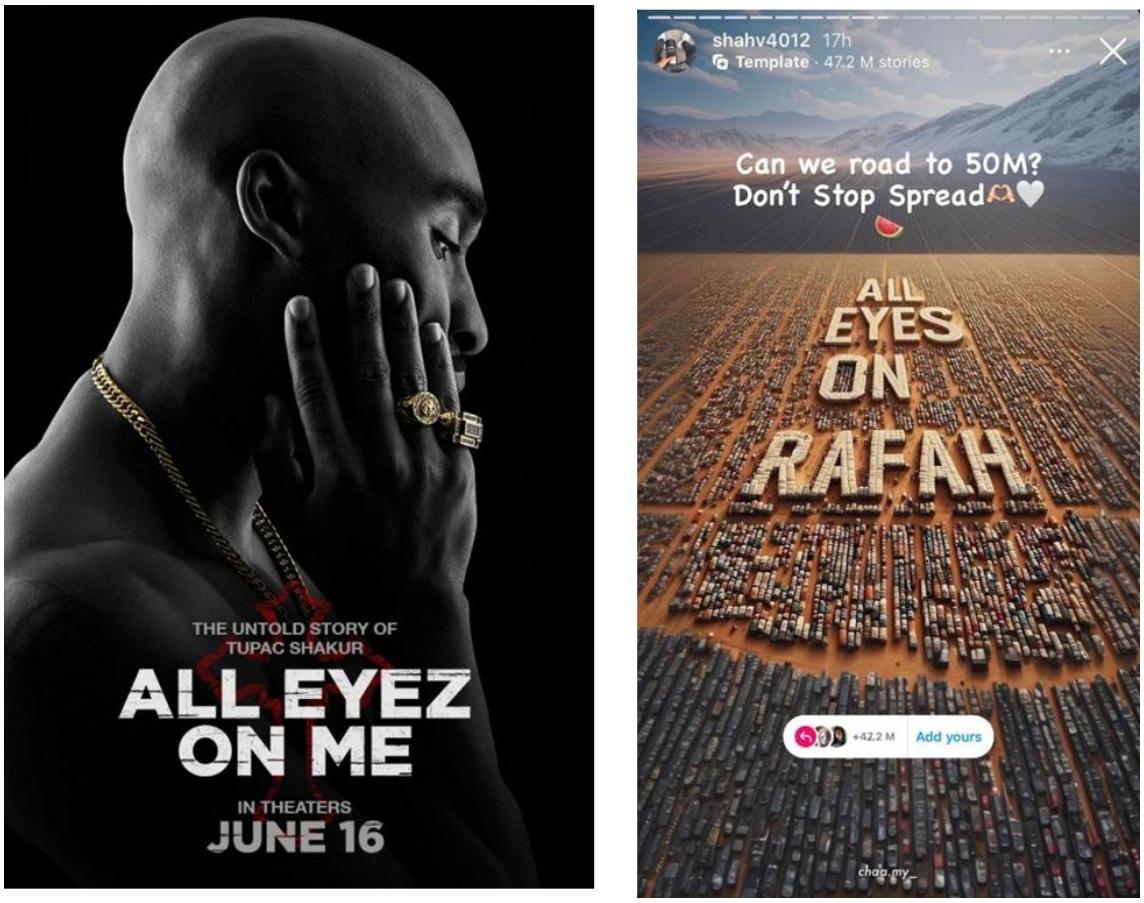
Bu teknik bağımlılık, Trump'ın Gazze videosunda sistematik önyargı üretiminin temel mekanizmasını oluşturmaktadır. Video oluşturma sürecinde kullanılan promptlar incelendiğinde, teknik terimlerin ideolojik içerikle iç içe geçtiği ve bunlar arasında ayrım yapılmasının oldukça güç olduğu açıkça görülmektedir. Sinematik geniş açı, yakın çekim, uzak çekim gibi terimler. Sahneyi oluşturmak için kullanılan komutlar olmakla beraber veri setlerinin görüntü oluşturması için kullandığı kombinasyonları etkilemektedir.

Veri setlerinin sömürgeci mirası, bu sürecin temel belirleyici faktörü olarak değerlendirilebilir. "Palmiye ağaçları altında mutlu çocuk" gibi bir yönerge verildiğinde sistem, Batılı kaynaklardan edindiği "egzotik yoksulluk" imgelerine, örneğin Myanmar'da palm yağı endüstrisinde çalışan çocuk işçilere ilişkin temsillere yönelme eğilimi göstermektedir. Bu imgeler, National Geographic'in 20. yüzyılda inşa ettiği "öteki" temsillerinin dijital bağlamda yeniden üretimi olarak analiz edilebilir: Siyahî veya Asyalı çocuklar "ilkel ama mutlu" bir temsil çerçevesinde konumlandırılırken, "Filistinli çocuk" yönergesi yıkıntılar ve askerlerle karakterize edilen bir görsel repertuarı ortaya çıkarmaktadır. Batı medyasının Gazze'yi öncelikli olarak çatışma ve mağduriyet

perspektifinden belgelemesi, yapay zekanın büyük veri tabanından etkilenen görsel hafızasının bu temsil kalıplarıyla sınırlandırılmasına neden olmaktadır

Görsel 6. Mağaradan palmiye ağaçlarına doğru koşan sözde Filistinli çocuklar.

Trump'ın paylaştığı videoda Gazze'nin yıkıntılarından çıkan çocuklar zafer ifadeleri eşliğinde temsil edilmektedir. Foucault'nun iktidarın direnci (2000) kendi mantığına entegre etme teorisi bu bağlamda somut bir örnek bulmaktadır: Tepkiler, platformların veri ekonomisini destekleyen duygusal etkileşim unsurlarına dönüşmektedir. "All Eyes on Rafah" kampanyasının milyonlarca paylaşımı, direnişi küresel bir sembolik anlama dönüştürme potansiyeli taşırken, bu paylaşımların dolaşımda olduğu platformlar için milyonlarca reklam gösterimini mümkün kılan bir içerik kaynağı oluşturmakta ve aynı zamanda kullanıcı verilerini algoritmaların veri tabanına dahil etmektedir. Bu paradoksal durum, yapay zekanın iktidar yapılarıyla kurduğu ilişkinin problematik bir yansıması olarak değerlendirilebilir: Muhalif söylemler, etkinlik düzeyinden bağımsız olarak, sistemin yeniden üretim mekanizmalarına katkıda bulunabilmektedir.



Görsel 7. "All Eyez on Me" sadece "bütün gözler üzerimde" demek değil, aynı zamanda Rapçi Tupac'ın mirasının bir parçası olarak baskı, şöhret, beklenti ve sürekli izlenme hissini de içeren bir kültürel ifadedir. Görsel 8: Instagram şablonu olarak viral olan ve "Bütün gözler Refah"ta mesajını veren yine yapay zekâ ile oluşturulmuş görsel.

Gazzeli çocukların tanımlarla üretilmesi ve yapay zekanın politik iletişimdeki işlevi, teknolojinin tarafsızlık iddiasını problematize eden bir olguyu ortaya koymaktadır: Üretilen içerikler, veri setlerinin tarihsel ve kültürel önyargılarından bağımsız bir şekilde oluşturulamamaktadır. Örneğin, "All Eyes on Rafah" görsel temsili, Gazze'deki insani krizi küresel bir sembolik anlama dönüştürürken, estetikleştirilmiş çadır düzenlemeleri ve havadan çekilmiş "sanatsal" kompozisyonla gerçek insani acıyı görsel bir temsil nesnesine indirgeme eğilimi göstermektedir. Bu görsel temsil, Kuala Lumpur'lu bir öğrenci tarafından oluşturulmuş olmasına rağmen, Batı merkezli görsel temsil paradigmalarının kodlarını aşma kapasitesi sınırlı kalmaktadır. Yapay zekâ, muhalif söylemleri dahi Batılı estetik normlar çerçevesinde dönüştürerek, politik direnci bir tüketim nesnesine dönüştürme potansiyeli taşımaktadır.

Komut mühendisliğinin sınırları da benzer şekilde veri setlerinin ideolojik çerçevesiyle şekillenmektedir. "Kafasının üstünde su taşıyan beyaz bir kadın" görseli talep edildiğinde sistem, su taşıma pratiklerinin bilgisini öncelikle Batılı olmayan coğrafyalarla ilişkilendirdiği için bu imgeyi üretme kapasitesi sınırlı kalmaktadır. Batılı veri setlerinde beyaz kadınlar, su ile genellikle "hayırseverlik" veya "macera" bağlamında ilişkilendirilmektedir. Benzer şekilde, "Gazze'de mutlu bir çocukluk" imgesi üretilmesi zorluğu bulunmaktadır, çünkü Batılı kaynaklar tarihsel olarak bu coğrafyayı öncelikle şiddet ve yıkım temaları üzerinden belgelemiştir. Bu sınırlandırmalar, yapay zekanın politik tarafsızlığını değil, veri sömürgeciliğinin güncel durumunu yansıtmaktadır.

Bu bağlamda Trump'ın yapay zekâ üretimi söz konusu videosunda Gazze, Batılı tüketim kültürünün ve oryantalist fantezilerin bir temsil alanına dönüşmektedir. Videoda, "humus yiyen Elon Musk" veya "deniz kıyısında sakallı sarıklı dansözler" gibi imgeler, Doğu'nun Batılı algıda nasıl egzotik bir meta olarak kodlandığını göstermektedir. Yapay zekâ, bu temsilleri üretirken Batı merkezli veri setlerinin tarihsel önyargılarını, resmin üretildiği dönemsel bağlamdan bağımsız olarak yeniden üretme eğilimi göstermektedir. Örneğin, "sakallı kadın" yönergesi verildiğinde sistem, Harnaam Kaur veya Conchita Wurst gibi Batı'da popülerlik kazanmış figürlere yönelme eğilimi göstermektedir. Ancak bu figürler dahi, oryantalist bir çerçevede konumlandırılmaktadır: Sakallı bir kadın, Batı'da olduğu gibi "egzotik" bir ilgi nesnesi veya "cesur" bir performans sanatçısı olarak değil, kültürel bağlamın marjinalleştirildiği sahneler içinde temsil edilme eğilimi göstermektedir. Buna karşın, Ortadoğu'da toplumsal normları sorgulayan kadınların gerçek deneyimleri (örneğin, toplumsal baskı mekanizmalarına karşı çıkan aktivistler), bu temsil sistemlerinde görünürlük kazanamamaktadır. Görüntü oluşturma aşaması, Batı'nın "kabul edilebilir öteki" kategorilerini öne çıkarırken, Doğu'nun kendi içindeki çeşitliliğini marjinalleştirme ve homojenleştirme eğilimi sergilemektedir.

Görsel 9. Beden olumlama aktivisti ve Instagram fenomeni Harnaam Kaur ve Trump'ın videosunda sakallı dansözler.

Bu temsiller, 19. yüzyıl oryantalist ressam Jean-Léon Gérôme'un The Slave Market tablosundaki (Görsel 12) sahnelerle dikkat çekici paralellikler sergilemektedir: Politik iktidar sahibi figürler (Trump, Musk, Netanyahu), "öteki"nin bedenini ve kültürünü seyirlik bir nesne olarak konumlandırma eğilimi göstermektedir. Gérôme'un tablosunda olduğu gibi, videodaki dansöz ve "sakallı sarıklı" özneler, eşitsiz güç dinamiklerini estetikleştiren bir dekor işlevi görmektedir. Gérôme'un The Slave Market tablosu, Doğu'yu bedenin alınıp satıldığı bir pazar olarak tasvir ederken, Trump'ın videosu da Gazze'yi Batılı sermayenin potansiyel olarak sömürebileceği bir inşaat alanı olarak temsil etmektedir. İki temsil arasındaki ortak unsur, Doğu'nun pasif bir kaynak olarak algılanmasıdır: 19. yüzyılda bedenler, günümüzde ise toprak ve kültür, Batılı gözlemcinin tüketimine açık birer "meta" olarak konumlandırılmaktadır. Videodaki "deniz kıyısında dansözler" sahnesi, bu mantığın dijital çağdaki yansımasını oluşturmaktadır: Gazze'nin kültürel dokusu, Batılı seyircinin beklentilerine uygun bir turizm broşürü formatına indirgenmektedir. Ancak turizme veya batılı değerlere hizmet ettiği ölçüde var olabileceği iması video boyunca tekrar edilmektedir.

Görsel 10. Trump'ın bir oryantal dansçı ile dans ettiği görsel ve Jean-Léon Gérôme'un oryantalist resmin başat örneklerinden biri olarak kabul edilen *The Slave Market* tablosu.

Yapay zekanın "palmiye ağaçları altında Filistinli mutlu bir çocuk" gibi bir yönergeye yanıt üretme kapasitesinin sınırlı olması, veri setlerindeki temsil krizini ve gelecekte ortaya çıkabilecek potansiyel sorunları belirgin biçimde ortaya koymaktadır. Sistem, bu yönergeyi işlerken, Batılı medyanın Filistin'i öncelikli olarak çatışma, yıkım veya dini fanatizmle ilişkilendiren kayıtlarına yönelmektedir. Bu durum, "Mutlu Filistinli çocuk" imgesinin, veri setlerinde minimal düzeyde temsil edilmesinden kaynaklanmaktadır. Söz konusu olgu, yapay zekanın ürettiği görüntülerin sömürgeci temsil biçimlerini nasıl sürdürdüğünü göstermektedir. Batılı kurumların (Magnum Photos, National Geographic, Reuters) oluşturduğu görsel hafıza, küresel gerçekliği tek boyutlu bir anlatı çerçevesine sınırlandırmış durumdadır.

Yapay zekâ sistemleri, yapısal özellikleri nedeniyle kullanıcıların verdiği talimatları literal (kelimesi kelimesine) yorumlama eğilimi göstermektedir. Bu teknik sınırlama, yapay zekâ modellerinin ürettiği içeriklerdeki sıra dışı unsurların temel kaynağını oluşturmaktadır (Frank, 2025, s. 8). Örneğin, Trump'ın videosunda görülen "altın heykeller" veya "zenginliğin vurgulandığı" sahneler, yapay zekânın kullanıcı yönergelerini (örneğin, "Trump Tower'ı

görkemli ve lüks bir şekilde göster") doğrudan ve mecazi olmayan bir biçimde görselleştirmesinin sonucu olarak değerlendirilebilir. Bu literal yaklaşım, yapay zekânın yan anlamları ve çağrışımları kavrama kapasitesinin sınırlılığından kaynaklanmaktadır. Sistem, Trump Towers yerine "Trump Gazze" tabelası gibi bir unsurun Ortadoğu'daki insani krize referans verdiğini veya etnik temizlik çağrışımı oluşturduğunu algılayamamakta; yalnızca verilen talimatı görselleştirme işlevini yerine getirmektedir.

Görsel 11. Trump Kulesi'ne Vurgu yapan Trump Gaza yazısı.

Saumure ve arkadaşlarının (2025, s. 2) çalışmasında belirtildiği gibi, mizah ve önyargı arasındaki ilişkiyi inceleyen araştırmalar psikoloji, sosyoloji ve iletişim alanlarında geniş bir çalışma sahası oluşturmaktadır. Bu literatürün ortak bulgularından biri, mizahın önyargıları ve kalıp yargıları pekiştirebileceği yönündedir. Nitekim çalışmalar, mizahın aşağılayıcı kalıp yargıları normalleştirerek ve bunların toplumsal açıdan kabul edilebilir biçimlerde dile getirilmesini kolaylaştırarak önyargılı tutumları sürdürebileceğini ortaya koymuştur.

Trump'ın Gazze videosunda, McLuhan'ın (1964) belirttiği "araç mesajdır" teorisinin somut bir örneğini görüyoruz. Videoda yer alan "Trump Gaza" tabelaları, altın heykeller ve Elon Musk'ın absürt temsili gibi unsurlar, izleyiciyi içeriğin politik ciddiyetinden uzaklaştırarak duygusal bir iletişim bağlamı yaratmaktadır. Benzer biçimde Festinger'in (1957) bilişsel uyumsuzluk teorisine göre, absürt mizahi sahneler ve etnik temizlik gibi ağır politik

mesajların yan yana gelmesi izleyicide ciddi politik mesajları göz ardı etme eğilimini güçlendirmektedir. Myanmar örneğinde olduğu gibi (Guzman, 2022), mizahi unsurların ciddi insani trajedilerle harmanlanması, sosyal medyada şiddet içeren söylemleri normalleştirme işlevi görebilir. Trump videosunun sahne analizi, Festinger ve McLuhan'ın teorilerinin pratikte nasıl tezahür ettiğinin çarpıcı bir örneğidir.

Söz konusu durumda risk faktörü, yüzeysel basitlik ile altta yatan politik gündem arasındaki ayrımda bulunmaktadır. İzleyici, videoyu "mizahi bir yapay zekâ uygulaması" olarak algılarken, alt metinde yer alan mesajlar (örneğin, Gazze'deki askeri operasyonların meşrulaştırılması) bilinçdışı düzeyde etki oluşturabilmektedir.

Sonuç ve Tartışma

Veo ile görsel üretim deneyimlerinden elde edilen en çarpıcı bulgu, yapay zekânın görsel oluşturması için mutlak anlamda teknik tanımlamalara ihtiyaç duymasıdır. "Sinematik geniş açı," "düşük açı," "8K" gibi terimler ilk bakışta görüntü kalitesini belirleyen teknik detaylar gibi görünse de bu çalışma bu terimlerin ideolojik içeriğin şekillenmesinde belirleyici rol oynadığını kanıtlamıştır. Bu durum, veri setlerinin tamamen tarafsız ve sayısal olarak homojen içeriklerden oluşması halinde dahi, teknik detaylar ve görüntü seçim kriterlerinin yanlılık üretebileceğini göstermektedir.

Bu teknik bağımlılık, Ford Thunderbird (Görsel 5) örneğinde olduğu gibi, modelin farklı perspektiflerden aynı nesneyi tanıyamamasına benzer bir sorun yaratmaktadır. Yapay zekâ, milyonlarca görüntü üzerinde eğitilmiş olsa bile, eğitim verisindeki açısal sınırlamalar nedeniyle beklenmedik sonuçlar üretebilmektedir. Trump videosunda bu durum, teknik terimlerin ideolojik içerikle iç içe geçmesi ve aralarında ayrım yapılmasının güçleşmesi şeklinde tezahür etmektedir.

Çalışmanın en önemli bulgularından biri, yapay zekânın çocukları varsayılan "okul gezisine giden çocuklar" çerçevesinde tasarlamaya yönelmesidir. Bu eğilim, Trump videosunda "Filistinli," "Asyalı," veya "Doğulu" gibi etnik ve coğrafi belirteçlerin kasıtlı kullanımının gerekliliğini açıklamaktadır. Başka bir deyişle, yapay zekâ sistemi "çocuk" dendiğinde otomatik olarak Batılı, beyaz çocukları varsayar hale gelmiştir.

Bu bağlamda, "palmiye ağaçları altında mutlu çocuk" yönergesi verildiğinde sistemin Batılı kaynaklardan edindiği "egzotik yoksulluk"

imgelerine yönelmesiyle daha da belirginleşmektedir. Sistem, bu anahtar kelimeleri kullanarak veri tabanındaki en güçlü örneği -haberler, sosyal platformlar ve dernekler aracılığıyla internete yayılan temsiller, örneğin Myanmar'da palm yağı endüstrisinde çalışan çocuk işçilere ilişkin görüntüler- önceleyerek, National Geographic'in 20. yüzyılda inşa ettiği "öteki" temsillerini dijital bağlamda yeniden üretmektedir. National Geographic'in ve Batı'nın bu veri setlerine katkısı, geçtiğimiz yüzyılda başat olarak görsel üretimine ve görsel ağırlıklı pazarlamaya dayalı stratejileri olmuştur. Çok fazla görüntüyü depolayıp bunu dönüşümlü olarak dergi, televizyon, internet gibi platformlarda yaymaları, gösterilen şeylerin anlamını belirlemiş ve o şeyin öyle olması gerektiği ile ilgili bir algı oluşturmuştur. Veri setleri ve anahtar kelimeler aracılığıyla bu önyargılar tekrar tekrar oluşturulmakta ve meşrulaşmaktadır. Haliyle bu durum, yapay zekânın "Filistinli mutlu bir çocuk" gibi yönergelere yanıt üretme kapasitesinin neden sınırlı olduğunu açıklamaktadır: Batılı veri setlerinde bu coğrafya öncelikle şiddet ve yıkım temaları üzerinden belgelenmiştir.

Yapımcılar Trump videosunu hiciv amacıyla yaptıklarını belirtse de elde edilen analiz sonuçları videonun oryantalist bakış açısının sistematik bir dışavurumu olduğunu kanıtlamaktadır. Videodaki "sakallı dansözler" ve Jean-Léon Gérôme'un The Slave Market tablosu arasındaki paralellik tesadüfi değildir. Her iki durumda da politik iktidar sahibi figürler, "öteki"nin bedenini ve kültürünü seyirlik bir nesne olarak konumlandırmaktadır. 19. yüzyılda Gérôme'un tablosunda Doğu'nun bedenin alınıp satıldığı bir pazar olarak tasviri ile Trump videosunda Gazze'nin Batılı sermayenin sömürebileceği bir inşaat alanı olarak temsili arasındaki ortak unsur, Doğu'nun pasif bir kaynak olarak algılanmasıdır. Bu yaklaşım, "All Eyes on Rafah" kampanyasının viral olmasında da görüldüğü gibi, muhalif söylemleri dahi Batılı estetik normlar çerçevesinde dönüştürerek politik direnci bir tüketim nesnesine dönüştürme potansiyeli taşımaktadır.

TIBET metodolojisinin bu çalışmadaki uygulaması, dinamik önyargı ekseni belirleme yaklaşımının video içeriklerindeki örtük ve açık önyargıları sistematik olarak tespit etmede etkili olduğunu göstermiştir. Ancak çalışmanın kapsamı nedeniyle CAS hesaplamaları ve MAD skorları aracılığıyla önyargı derecelerinin karşılaştırmalı analizi gerçekleştirilememiştir. Bu durum önemli bir metodolojik sınırlama oluşturmakla birlikte, gelecek araştırmalar için kritik bir alan açmaktadır. Türkçe medya analizi alanında yapay zekâ üretimi görsellerin yanlılık tespitine yönelik özgün metodolojik yaklaşımların büyük ölçüde eksik olması, bu çalışmanın önemini artırmaktadır. Mevcut bileşik modellerin hata

birikimi, yüksek hesaplama maliyeti ve tutarlılık sorunları yanında, Türkçe medya içeriklerinin kültürel ve dilsel bağlamının dikkate alınması gerekliliği de ortaya konmuştur.

Günlük 34+ milyon yapay zekâ görüntüsü üretilmesi ve pazarlama içeriğinin %30'unun üretken yapay zekâ kullanması öngörüsü, mevcut azaltma çabalarının yetersizliğini gözler önüne sermektedir. Yapay zekâ şirketlerinin veri kaynaklarını gizli tutmaya başlaması karşısında, demokratik denetim mekanizmalarının kurulması ve açık kaynak projeler üzerinden şeffaflık yoluyla önyargıları tespit etme yaklaşımlarının desteklenmesi kritik önem taşımaktadır. Bu çalışma, yapay zekânın içinde bulunduğumuz gerçekliğin aktif bir şekillendirici unsuru olduğunu göstermektedir. Veri setlerinin sömürgeci mirası, komut bağımlılığı ve mizahi normalleştirme mekanizmaları, yapay zekâ teknolojilerinin demokratik değerlerle uyumlu olmayan bir şekilde geliştirildiğini ortaya koymaktadır. Bu hedefe ulaşmak için teknolojik determinizmi aşan, toplumsal katılımı esas alan ve kültürel çeşitliliği koruyan acil müdahalelerin geliştirilmesi gerekmektedir.

Extended Abstract

The study looks at the regular bias in content that artificial intelligence creates - it mainly centers on text-to-image models. The paper examines how this bias changes into humorous manipulation, both on purpose and by accident. The research explores where AI bias, political messages along with humor meet. It does this by analyzing a video of Donald Trump. Artificial intelligence created this video about Gaza, and it became popular in February 2025.

Generative AI systems, particularly text-to-image models, show a great ability to create realistic and varied visual content from text. People use these models more and more in art, design, marketing, education along with entertainment. But as AI content appears more often in daily life, basic questions come up about how this content shows the world. Current numbers suggest that by 2025, artificial intelligence could produce as much as 90 % of internet content. This signals a big change in how digital content functions.

Large amounts of data and complicated programs that support these powerful plus growing tools cause ethical and social problems. The bias problem in AI-created content represents one of the most important issues. AI programs can learn also support societal prejudices, stereotypes along with inequalities that

appear in the data used to train them. This often leads to unfair or wrong results when the programs show certain groups of people or ideas.

The study uses Marshall McLuhan's "the medium is the message" idea to look at how AI technology carries content - it also shapes how people see messages and affects society. A video of Trump, which AI made, offers a chance to look at this idea again in politics today. The video has many unusual parts, like golden statues or animations of Elon Musk. He showers money, and there is a sign that reads "Trump's Gaza." These parts seem to show the wide range of what AI can create. The content looks funny in its technical form, but it could cause people to accept serious political proposals, such as ethnic cleansing, without noticing.

Research also uses Leon Festinger's cognitive dissonance theory - it explains how content made by AI can restrict people's ability to think critically. The content combines humor with serious political messages. When unusual and simple elements, such as the "Trump's Gaza" sign, appear next to ideas of ethnic cleansing, viewers may disregard serious political content; they think this is a joke. Humor acts as a barrier from unpleasant facts.

The study uses the TIBET (Text-to-Image Bias Evaluation Tool) method, which Chinchure and colleagues created for a complete bias analysis. The study puts TIBET's dynamic bias axis detection plus counterfactual evaluation approach to use to show possible biases in how videos portray Trump. The analysis takes frames from the video content. TIBET's approach, which LLMs support, sets up multi-dimensional bias axes - these axes cover political figure representation, leadership biases, age and gender stereotypes, socioeconomic status displays, also cultural or geographical biases.

Researchers develop counterfactual scenarios for bias axes. For example, they use political figures with different demographic traits instead of Trump. People infer concepts with VQA and calculate Concept Association Scores (CAS). This method allows the measurement of bias in Trump's video display - it also lets people compare bias levels through Mean Absolute Deviation (MAD) scores.

Research shows that AI works because it depends on a set of commands that users type as prompts. Users want the AI to create content. Specific words, like "cinematic wide shot," "low angle," and "8K," help ensure the content looks good. This reliance on technical language builds the basic way AI content produces systematic bias. When we look at the prompts used to make video, the

technical words mix with ideas - it is very hard to tell the difference between them.

The origin of datasets forms the main cause in this process. When a system receives a command like "happy child under palm trees," it often shows "exotic poverty" pictures. The system gets these pictures from Western sources, like images of child workers in Myanmar's palm oil industry. People can see these pictures as digital copies of "other" representations that National Geographic created in the 20th century. Black or Asian children appear in a "primitive but happy" picture setting. But the command "Palestinian child" creates a set of images with ruins and soldiers.

A basic discovery shows AI's small ability to create pictures like "happy Palestinian child under palm trees." When the system handles this command, it looks at records. Western media mostly links Palestine to fighting, ruin, or religious zeal. This condition comes from the low number of "happy Palestinian child" pictures in data sets - this shows how AI keeps colonial ways of showing things.

A study shows that the limits of prompt engineering relate to the ideological structure of datasets. If someone asks for a picture of "a white woman carrying water on her head," the system struggles to create this image. This happens because its knowledge base mostly links water carrying to places outside the Western world. In Western datasets, white women usually appear with water in settings of "charity" or "adventure."

The study shows that AI systems understand user instructions word for word because of how they are built. This technical limit is the main reason for unusual parts in content that AI creates. For instance, the "golden statues" or "wealth-focused" scenes in Trump's video are likely what happens when AI turns user orders into direct, non-figurative pictures.

The Trump Gaza video study shows how computer made content changes serious political talk into amusement. The video starts with pictures of broken buildings and land in Gaza. Then it moves to a completely different Gaza, a city in the future with shiny tall buildings, happy children who play on beaches, plus expensive Tesla cars that drive on wide streets lined with palm trees. The content has computer made people who look familiar and odd situations. For example, people who look like Elon Musk eat hummus also then throw money on crowds. Bearded Hamas soldiers surprisingly dance in bikinis and belly dancer clothes.

The analysis shows how the video's technical reliance on prompts establishes systematic ways to produce bias. When someone inspects the prompts that helped create the video, technical words mix with ideological material, so separating them becomes hard. The colonial history of datasets settles the basic factor in this process. Western sources document Gaza mostly through conflict and victimization. This restricts the AI's visual memory, which big data collections influence, to those image patterns.

The research shows that current ways to lessen harm are not enough. The design of systems needs a basic change, and the rules governing them must be complete to deal with the amount of bias found in these growing AI systems. The study reports that seeing stereotypical pictures from AI makes biases firmer for at least three days. According to a February 2024 Nature study, this method works better than using stereotypes from text. The discovery shows that bias from text-to-image systems creates lasting mental effects, going past a quick view.

A study concludes that AI systems, because of how they are built, take user instructions literally. This literal interpretation causes the strange parts in content AI produces. The AI's limited ability to understand deeper meanings and connections makes it approach instructions this way. The system does not see that "Trump Gaza" signs relate to the humanitarian problems in the Middle East, nor does it connect to ethnic cleansing - it only shows what the instructions tell it to.

Kaynakça

Adorno, T. W. (2021). *Kültür endüstrisi: Kültür yönetimi* (N. Ülner, M. Tüzel & E. Gen, Çev.). İletişim Yayınları.

Barthes, R. (2015). *Göstergebilimsel serüven* (5. Baskı, M. Rifat & S. Rifat, Çev.). Yapı Kredi Yayınları. (Orijinal çalışma 1985 yılında yayımlanmıştır)

Baum, J., & Villasenor, J. (2024). *Rendering misrepresentation: Diversity failures in AI image generation*. The Brookings Institution.

Berger, J. (2008). *Ways of seeing*. Penguin Classics.

Bower, A. H., & Steyvers , M. (2021). *Perceptions of AI engaging in human expression*. Scientific Reports.

Chinchure, A., Shukla, P., & Bhatt, G. (2024). TIBET: Identifying and Evaluating Biases in Text-to-Image Generative Models. *Computer Vision – ECCV 2024* (s. 429–446). içinde

Chomsky, N., & Herman, E. S. (2021). *Rızanın imalatı*. (E. Abadoğlu, Çev.) BGST Yayınları.

DeLacey, P. (2023). *Biases in large image-text AI model favor wealthier, Western perspectives*. Michigan Engineering News. https://news.engin.umich.edu/2023/12/biases-in-large-image-text-ai-model-favor-wealthier-western-perspectives/

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fitts, A. S., Rabinowitz, K., & Sadof, K. D. (2023). *This is how AI image generators see the world*. The Washington Post.

Foucault, M. (2000). *Özne ve iktidar* (I. Ergüden & O. Akınhay, Çev.). Ayrıntı Yayınları. (Orijinal çalışma 1982 yılında yayımlandı)

Frank, A. L. (2025). Wit meets wisdom: The relationship between satire and AI communication. *Journal of Science Communication, 24*(1), 1-18. https://doi.org/10.22323/2.24010204

Garfinkle, A. (2023, Ocak 13). *90% of online content could be 'generated by AI by 2025,' expert says.* https://finance.yahoo.com/. adresinden alındı

Gebru, T., & Mitchell, M. (2024). AI ethics beyond technical solutions: Toward democratic governance and social justice. *Journal of Artificial Intelligence Research, 75*(1), 118-142. https://doi.org/10.1093/oxfordhb/9780190067397.013.16

Girrbach, L., Alaniz, S., & Smith, G. (2025). *A Large Scale Analysis of Gender Biases in Text-to-Image Generative Models*. arXiv:2503.23398.

Guzman, C. d. (2022). *Meta's Facebook Algorithms 'Proactively' Promoted Violence Against the Rohingya, New Amnesty International Report Asserts*. TIME USA, LLC.

Hall, R. (2025). *Trump Gaza' AI video intended as political satire, says creator*. Guardian News & Media Limited.

Hall, S. (Ed.). (1997). *Representation: Cultural representations and signifying practices*. Sage Publications, Inc; Open University Press.

Hanjun Luo, H. H. (2024). *BIGbench: A Unified Benchmark for Evaluating Multi-dimensional Social Biases in Text-to-Image Models*. a. p. arXiv içinde, https://doi.org/10.48550/arXiv.2407.15240.

Luccioni, A. S., Akiki, C., & Mitchell, M. (2023). Stable Bias: Analyzing Societal Representations in Diffusion Models. NIPS '23: *Proceedings of the 37th International Conference on Neural Information Processing Systems*.

Martinez, D. R., & Kifle, B. M. (2024). *Artificial Intelligence: A Systems Approach from Architecture Principles to Deployment*. MIT Press.

McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill.

Mollett, A., Brumley, C., Gilson, C., & Williams, S. (2017). *Communicating Your Research with Social Media: A Practical Guide to Using Blogs, Podcasts, Data Visualisations and Video*. SAGE Publications.

Naik, R., & Nushi, B. (2023). *Social Biases through the Text-to-Image Generation Lens. Publication History*, s. 786 - 808.

Nicoletti, L., & Bass, D. (2023). *Humans Are Biased Generative AI Is Even Worse*. Bloomberg L.P.

Perera, M. V., & Patel, V. M. (2023). Analyzing Bias in Diffusion-based Face Generation Models. *2023 IEEE International Joint Conference on Biometrics*.

Rayyash, H. A. (2024). AI meets comedy: Viewers' reactions to GPT-4 generated humor translation. *Ampersand*, 12.

Postman, N. (2020). *Televizyon: Öldüren eğlence* (O. Akınhay, Çev.). Ayrıntı Yayınları.

Saharia, C., & Chan, W. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 36479 - 36494.

Said, E. W. (1978). *Orientalism*. Pantheon Books.

Saumure, R., De Freitas, J. & Puntoni, S. Humor as a window into generative AI bias. *Sci Rep 15*, 1326 (2025). https://doi.org/10.1038/s41598-024-83384-6

Son, G. B. G., 2021. *Migration Memes and Social Gaze: An Analysis of Facebook*. Social Transformations Journal of the Global South.

Vice, J., & Akhtar, N. (2025). Quantifying Bias in Text-to-Image Generative. *Institute of Electrical and Electronics Engineers*, s. 1-14.

Wiessner, D. (2024). *Workday must face novel bias lawsuit over AI screening software*. Reuters.

Zewe, A., 2022. *Can machine-learning models overcome biased datasets?*. MIT News Office.

Zhou, M., Abhishek, V., & Derdenger, T. (2024). *Bias in Generative AI*. eprint arXiv. https://doi.org/https://doi.org/10.48550/arXiv.2403.02726