

Otonom Sürüş Senaryolarında Karar Verme Mekanizması Olarak Görme-Dil Modellerinin (VLM) Entegrasyonu ve Güvenlik Analizi

Yönetici Özeti

Otonom sürüş teknolojilerinin (AD) evrimi, kural tabanlı modüler sistemlerden derin öğrenme destekli uçtan uca (end-to-end) mimarilere doğru radikal bir dönüşüm geçirmektedir. Bu dönüşümün merkezinde, aracın sadece çevresini algılaması değil, aynı zamanda karmaşık, öngörülemeyecek ve sosyal etkileşim gerektiren senaryolarda insan benzeri bir muhakeme (reasoning) yeteneği sergilemesi gerekliliği yatmaktadır. Bu rapor, Büyük Dil Modelleri (LLM) ve bunların görsel modalitelerle güçlendirilmiş versiyonları olan Görme-Dil Modellerinin (Vision-Language Models - VLM), otonom araçların "Yüksek Seviyeli Karar Verme" (High-Level Decision Making) katmanına entegrasyonunu kapsamlı bir şekilde incelemektedir. Projenin temel hipotezi, VLM'lerin aracın direksiyon ve gaz kontrolünü doğrudan devralması yerine, stratejik bir "yardımcı pilot" veya "karar motoru" olarak konumlandırılması gerektidir. Bu yaklaşım, VLM'lerin sağduyu (common sense) ve genelleme yeteneklerinden faydalananken, düşük seviyeli kontrol algoritmalarının (MPC, PID) deterministik güvenliğini korumayı amaçlar.

Rapor, literatürdeki en güncel 20+ kaynağın derinlemesine analiziyle, VLM tabanlı otonom sürüş mimarilerini (GAIA-1, DriveGPT, DiLu, DriveAgent-R1), bu sistemlerin karşılaştığı kritik güvenlik darboğazlarını (halüsinsiyon, gecikme, yanlış nedensel çıkarım) ve bu sorunlara yönelik çözüm stratejilerini (Düşük Rütbeli Matris Ayırıştırması, Hibrit Düşünme Çerçeveleri, LLM-Hinted RL) değerlendirmektedir. Analizler, VLM'lerin özellikle "uzun kuyruk" (long-tail) senaryolarında—örneğin nadir görülen trafik kazaları, yol çalışmaları veya agresif sürücü davranışları—geleneksel yöntemlere kıyasla üstün bir bağılamsal anlayış sunduğunu ortaya koymaktadır. Ancak, %40'lara varan halüsinsiyon oranları ve gerçek zamanlı işlem kısıtlamaları, bu modellerin güvenlik açısından kritik sistemlerde doğrudan kullanımını sınırlamaktadır.¹ Bu bağlamda rapor, VLM'lerin "Aktif Algı" (Active Perception) ve "Zincirleme Düşünce" (Chain-of-Thought) mekanizmalarıyla donatılarak, kapalı döngü (closed-loop) bir öğrenme sistemi içinde denetimli bir karar verici olarak kullanılmasının en uygulanabilir yol haritası olduğunu savunmaktadır.

1. Giriş: Otonom Sürüste Paradigma Değişimi ve Bilişsel Sürüş İhtiyacı

Akıllı Ulaşım Sistemleri (ITS), trafik sıkışıklığını azaltmak, güvenliği artırmak ve ulaşım

verimliliğini optimize etmek amacıyla geliştirilen teknolojilerin bütünüdür. Geleneksel ITS yaklaşımları, trafik akışını izlemek için sensörler ve kameralar kullanırken, genellikle önceden tanımlanmış kurallara ve reaktif algoritmalar dayanır. Bu durum, sistemlerin hızla değişen ve kaotik trafik koşullarına adaptasyonunu sınırlar.¹ Derin öğrenme (DL) tekniklerinin yükselişyle birlikte, Konvolüsyonel Sinir Ağları (CNN) ve Tekrarlayan Sinir Ağları (RNN) kullanılarak trafik tahmini ve nesne tespiti konularında önemli ilerlemeler kaydedilmiştir. Ancak, bu modellerin çoğu "kara kutu" (black box) doğasına sahiptir; yani aldıkları kararların nedenini açıklayamazlar ve eğitim verisinde bulunmayan yeni durumlarla karşılaşıldıklarında (zero-shot scenarios) genellikle başarısız olurlar.¹

Günümüzde otonom sürüs araştırmaları, "algılayan" araçlardan "düşünen ve anlayan" araçlara doğru bir evrim geçirmektedir. Bu evrimin itici gücü, Büyük Dil Modellerinin (LLM) ve VLM'lerin ortaya çıkışıdır. Bu modeller, sadece veriyi işlemekle kalmaz, aynı zamanda insan benzeri bir "sağduyu" kullanarak olaylar arasında nedensel ilişkiler kurabilir. Örneğin, klasik bir sistem "yolda bir top var" tespiti yapıp durabilirken, bir VLM "yola bir top yuvarlandı, muhtemelen arkasından bir çocuk koşabilir, bu yüzden sadece durmak yetmez, etrafı dikkatlice taramalıyım" şeklinde çok katmanlı bir muhakeme yürütebilir.⁵ Bu yetenek, otonom araçların sadece kurallara uyan robotlar olmaktan çıkıp, sosyal normları anlayan ve karmaşık senaryolarda müzakere edebilen ajanlara dönüşmesini sağlar.

1.1 Modüler Mimariden Uçtan Uca ve VLM Tabanlı Sistemlere Geçiş

Otonom sürüs yazılımları geleneksel olarak modüler bir yığın (stack) şeklinde tasarlanmıştır. Bu yapıda Algı (Perception), Lokalizasyon, Tahmin (Prediction), Planlama (Planning) ve Kontrol (Control) modülleri birbirine seri olarak bağlanır. Her modül, bir önceki modülün çıktısını girdi olarak alır. Bu yaklaşımın temel zayıflığı "hata yayılımı"dır (error propagation); algı modülündeki küçük bir hata (örneğin bir nesnenin yanlış sınıflandırılması), planlama modülünde büyük bir karar hatasına dönüşebilir. Ayrıca, modüler arasındaki arayüzlerin insan eliyle tasarlanmış olması, sistemin karmaşıklığı arttıkça yönetilemez hale gelmesine neden olur.⁷

Uçtan uca (End-to-End) sürüs modelleri, sensör verilerini (kamera, LiDAR) doğrudan kontrol komutlarına (direksiyon açısı, gaz/fren) eşleyerek bu ara katmanları ortadan kaldırmayı hedefler. Bu yaklaşım, veri odaklı optimizasyon sayesinde daha esnek sürüs davranışları üretebilir. Ancak, açıklanabilirlik eksikliği ve güvenlik garantilerinin zayıflığı, bu modellerin gerçek dünyada yaygınlaşmasını engellemektedir. VLM'lerin entegrasyonu, bu noktada bir köprü görevi görür: Uçtan uca sistemlerin esnekliğini korurken, dil modellerinin doğal dil işleme yeteneği sayesinde kararların "nedenini" açıklayabilir ve mantıksal bir denetim mekanizması sunar.⁷

1.2 "Mantık/Karar" Katmanı Olarak VLM'in Konumlandırılması

Bu raporda savunulan ve detaylandırılan mimari yaklaşım, VLM'i aracın doğrudan motor kontrolcüsü olarak değil, bir "Yüksek Seviyeli Karar Verici" (High-Level Decision Maker) olarak

konumlandırmaktır. Bu hiyerarşik yapıda:

- **Düşük Seviye (Low-Level):** Model Predictive Control (MPC) veya klasik PID kontrolcülerini, aracın yönüğe takibi, stabilite ve acil durum frenlemesi gibi milisaniye hassasiyetindeki görevleri üstlenir. Bu katman deterministik ve matematiksel güvenlik garantilerine sahiptir.
- **Yüksek Seviye (High-Level):** VLM, bir "stratejist" gibi çalışır. Algı sisteminden gelen zengin semantik veriyi işler, trafik bağlamını analiz eder ve düşük seviyeli kontrolcüye "stratejik hedefler" veya "kısıtlamalar" gönderir. Örneğin, "Şu an kavşakta belirsizlik var, agresif olma, şerit değiştirmeden önce sağdaki kamyonun geçmesini bekle" gibi bir karar üretir.⁹

Bu yapı, Nobel ödüllü psikolog Daniel Kahneman'ın "Hızlı ve Yavaş Düşünme" (Thinking, Fast and Slow) teorisine benzetilebilir. Düşük seviyeli kontrolcü "Sistem 1" (hızlı, refleksif) iken, VLM "Sistem 2" (yavaş, kasıtlı, mantıksal) olarak işlev görür. Bu ayrim, VLM'lerin hesaplama yükü ve gecikme sorunlarını yönetilebilir kılarken, güvenliği maksimize eder.¹¹

2. Teorik Çerçeve ve Yeni Nesil VLM Mimarileri

Otonom süreç için geliştirilen VLM mimarileri, modelin veriyi nasıl temsil ettiğini, geçmiş deneyimleri nasıl sakladığı ve geleceği nasıl öngördüğü konularında farklılaşmaktadır. Bu bölümde, literatürde öne çıkan en yenilikçi mimariler olan GAIA-1, DriveGPT, DiLu ve DriveAgent-R1 incelenecaktır.

2.1 Dünya Modelleri ve Üretken Sürüş: GAIA-1

Wayve tarafından geliştirilen **GAIA-1 (Generative AI for Autonomy)**, otonom süreç alanında "Dünya Modelleri" (World Models) kavramının en güçlü temsilcilerinden biridir. 9 milyar parametreye sahip bu model, sürüsü bir "video üretim problemi" olarak ele alır. Temel yeteneği, geçmiş video karelerini, metin komutlarını ve eylem geçmişini alarak, gelecekteki olası senaryoları yüksek gerçeklikte simüle edebilmesidir.¹²

Mimari Detaylar ve Kodlama Mekanizması

GAIA-1'in mimarisi iki ana aşamadan oluşur:

1. **Dünya Modeli (World Model):** 6.5 milyar parametreli bir otoregresif Transformer yapısıdır. Farklı modalitelerden (video, metin, eylem) gelen veriler önce ayrık "token"lara dönüştürülür ve ortak bir latent uzaya gömülür (embedding). Video kodlayıcı, görüntüleri vektör-quantized (VQ) temsillerle sıkıştırır. Metin ve eylem kodlayıcıları da kendi girdilerini aynı uzaya taşırlar. Dünya modeli, bu token dizisini işleyerek, bir sonraki zaman adımında gelmesi muhtemel "görüntü token'larını" tahmin eder. Bu süreç, dil modellerinin bir sonraki kelimeyi tahmin etmesine benzer.¹³
2. **Video Kod Çözücü (Video Decoder):** 2.6 milyar parametreli bir video difüzyon modelidir. Dünya modelinin tahmin ettiği semantik token'ları alarak, bunları piksel uzayında yüksek

çözünürlüklü ve zamansal olarak tutarlı video karelerine dönüştürür.

GAIA-1'in en önemli özelliği, 4700 saatlik (yaklaşık 420 milyon görüntü) gerçek sürüş verisiyle eğitilmiş olmasıdır. Bu devasa veri seti, modelin sadece nesneleri tanımmasını değil, aynı zamanda fizik kurallarını, nesne etkileşimlerini ve 3D geometriyi (örneğin tümseklerden geçen aracın sarsılması) örtük olarak öğrenmesini sağlamıştır.¹⁵ Model, metin komutlarıyla yönlendirilebilir; örneğin "yağmurlu havada sür" veya "agresif bir şerit değiştirme yap" gibi komutlarla, gerçek dünyada test edilmesi tehlikeli olan senaryoları simülle edebilir. Bu, otonom araçların eğitimi için sentetik veri üretiminde devrim niteliğindedir.

2.2 Sürüsun Tokenizasyonu: DriveGPT

DriveGPT, sürüş eylemlerini bir dil gibi modelleyerek, otonom sürüşü ardışık bir karar verme problemine dönüştürür. Geleneksel yöntemlerin aksine, sürekli kontrol sinyalleri yerine ayrılaştırılmış (discretized) eylem token'ları kullanır.¹⁶

Otoregresif Davranış Modellemesi

DriveGPT, 100 milyondan fazla sürüş örneği ve 1 milyardan fazla parametre ile eğitilmiştir. Model, Transformer tabanlı bir kodlayıcı-kod çözümü (encoder-decoder) yapısına sahiptir:

- **Kodlayıcı:** Hedef aracın geçmiş durumlarını, çevredeki diğer ajanların hareketlerini ve harita bilgilerini (şeritler, trafik işaretleri) alır. Bu verileri, sahne bağlamını temsil eden gömülü token'lara dönüştürür.
- **Kod çözümü:** LLM tarzı bir yapı izleyerek, geçmiş durumlara ve sahne bağlamına koşullu olarak, bir sonraki zaman adımındaki eylem dağılımını tahmin eder.

Matematiksel olarak, $\$T\$$ zaman ufkuna kadar olan durum tahmini şu şekilde ifade edilir:

$$\$P(s_{1:T} | c) = \prod_{t=1}^T P(s_t | s_{0:t-1}, c)\$$$

Burada s_t , t anındaki durumu, c ise bağlamı temsil eder.¹⁶

Ölçekleme Yasaları (Scaling Laws): DriveGPT ile yapılan deneyler, veri boyutu ve model parametreleri arttıkça, performansın (çarpışma oranlarının düşmesi, rota takibinin iyileşmesi) logaritmik olarak değil, güç yasasına (power law) uygun şekilde arttığını göstermiştir. Özellikle karmaşık senaryolarda (örneğin yoğun trafikte şerit değiştirme), büyük modellerin küçük modellere göre belirgin bir üstünlük sağladığı kanıtlanmıştır.¹⁷

2.3 Kapalı Döngü Öğrenme ve Bellek: DiLu Çerçeve

VLM'lerin en büyük eksiklerinden biri, "ömür boyu öğrenme" (lifelong learning) yeteneğinin sınırlı olmasıdır. **DiLu (Driving with Language)** çerçevesi, bu sorunu çözmek için insan benzeri bir deneyim birikimi mekanizması önerir.¹⁹

DiLu, üç ana modülden oluşur:

1. **Bellek Modülü (Memory):** Geçmiş süreç deneyimlerini saklar. Her deneyim, bir "sahne tanımı" (key) ve buna karşılık gelen "muhakeme süreci" (value) çifti olarak saklanır. Vektör veritabanı kullanılarak, mevcut duruma en çok benzeyen geçmiş deneyimler milisaniyeler içinde çağrılabılır.
2. **Muhakeme Modülü (Reasoning):** Algı verilerini ve bellekten çağrılan "ipuçlarını" (few-shot examples) kullanarak mevcut durum için bir karar üretir. Bu modül, LLM'in genellemeye yeteneğini kullanır.
3. **Yansıtma Modülü (Reflection):** Bu modül, sistemin öz-denetim mekanizmasıdır. Alınan kararların sonuçlarını (güvenli miydi, konforlu muydu?) analiz eder. Eğer bir karar hataya yol açtıysa, Reflection modülü bu kararı revize eder ve düzeltilmiş bilgiyi belleğe yazar.

Bu yapı, otonom aracın her sürüsten sonra "daha akıllı" hale gelmesini sağlar. Simülasyon sonuçları, DiLu'nun bellek modülü sayesinde kavşak geçişlerinde başarı oranını geleneksel pekiştirmeli öğrenme (RL) yöntemlerine göre 3 kat artırdığını (%24'ten %61'e) göstermektedir.²⁰

2.4 Aktif Algı ve Hibrit Düşünme: DriveAgent-R1

Pasif algı sistemleri, sensörlerden gelen veriyi sürekli işler ancak verinin niteliğini sorgulamaz. **DriveAgent-R1**, bu yaklaşımı değiştirerek "Aktif Algı" (Active Perception) kavramını getirir. Model, belirsiz bir durumla karşılaşlığında (örneğin "İlerideki nesne bir duba mı yoksa çocuk mu?"), pasif olarak beklemek yerine, algı sistemine "O bölgeye odaklan" (zoom in) veya "Farklı bir açıdan bak" gibi komutlar gönderebilir.¹¹

Hibrit Düşünme (Hybrid Thinking): DriveAgent-R1, süreç senaryolarını karmaşıklığına göre ikiye ayırmıştır:

- **Hızlı Düşünme (Fast Thinking):** Standart, öngörelebilir durumlar için (örn. boş otoyolda şerit takibi) metin tabanlı, düşük maliyetli muhakeme kullanılır.
- **Yavaş Düşünme (Slow Thinking):** Karmaşık veya belirsiz durumlarda, model "Vision Toolkit" adı verilen araç setini devreye sokar. Bu modda, model çok adımlı görsel sorgulama ve derinlemesine analiz yapar. Bu, Kahneman'ın Sistem 2 düşünme tarzının doğrudan bir uygulamasıdır.

DriveAgent-R1, 3 milyar parametreli (3B) görece küçük bir model olmasına rağmen, bu hibrit yapı sayesinde nuScenes gibi benchmarklarda GPT-5 gibi çok daha büyük ve kapalı kaynak modellerle yarışabilir performans sergilemiştir.

3. Karar Verme Mekanizmaları ve Zincirleme Düşünce (CoT)

VLM'lerin otonom süreçte entegrasyonunda en kritik katma değer, kararların "nasıl" alındığını gösteren şeffaflıktır. Zincirleme Düşünce (Chain-of-Thought - CoT), bu şeffaflığı sağlayan

temel mekanizmadır. CoT, modelin nihai bir eylem (örn. "Fren Yap") üretmeden önce, bu eyleme götüren mantıksal adımları ("Yaya geçidine yaklaşıyorum" -> "Yaya hareket halinde" -> "Hızım yüksek" -> "Durmalıyorum") sıralı bir şekilde oluşturmasını sağlar.

3.1 CoT4AD: Algıdan Eyleme Mantık Zinciri

CoT4AD (Chain-of-Thought for Autonomous Driving), bu yaklaşımı sistematik hale getiren bir çerçevedir. Model, sürüs görevini açıkça tanımlanmış dört aşamalı bir zincir olarak modeller:

Algı -> Soru Sorma (VQA) -> Tahmin -> Planlama.²⁴

1. **Algı (Perception):** Çoklu kamera görüntülerinden elde edilen veriler, Kuş Bakışı (BEV) uzayına izdüşürülür. Statik harita elemanları (şeritler) ve dinamik nesneler (araçlar) ayrı ayrı token'laştırılır.
2. **Soru Sorma (Reasoning/VQA):** Model, sahneyle ilgili kritik sorular sorar ve cevaplar. Örn: "Öndeki araç ne yapıyor?" -> "Sola sinyal veriyor." Bu aşama, modelin dikkatinin doğru yere odaklanmasıını sağlar.
3. **Tahmin (Prediction):** Gelecekteki sahne dinamikleri, difüzyon tabanlı bir modül ile tahmin edilir. Bu, modelin "görsel öngörü" kazanmasını sağlar.
4. **Planlama (Action):** Elde edilen tüm bu bilgiler ışığında, optimal yörünge planlanır.

Bu süreç, modelin sadece "ne yapacağını" değil, "neden yapacağını" da içselleştirmesini sağlar. Eğitim sırasında bu zincir açıkça (explicit) modellenirken, çıkarım (inference) sırasında model bu adımları örtük (implicit) olarak tek geçişte yapabilir, böylece hesaplama maliyeti dengelenir.

3.2 İstem Mühendisliği ve Senaryo Analizi

VLM'lerin performansı, onlara verilen "istemlerin" (prompts) kalitesiyle doğrudan ilişkilidir. Otonom sürüs için etkili bir istem, modelin dikkatini dağıtmadan kritik bilgilere odaklanmasıını sağlamalıdır.

- Yapılandırılmış İstem Örneği:
"Sen uzman bir otonom sürüs asistanın. Aşağıdaki görüntüyü analiz et. Adım 1: Tüm hareketli nesneleri ve konumlarını listele. Adım 2: Bu nesnelerin niyetlerini (intention) tahmin et. Adım 3: Trafik kuralları ve güvenlik önceliklerine göre ego-araç için en güvenli eylemi belirle. Adım 4: Kararının nedenini açıkla."

Literatürdeki çalışmalar, bu tür yapılandırılmış CoT istemlerinin, standart (Zero-Shot) istemlere göre cevap doğruluğunu %3.1, muhakeme kalitesini ise %4.6 artırdığını göstermektedir.²⁶ Ayrıca, "Görsel CoT" (Visual CoT) teknikleri, modelin sadece metin üretmesini değil, aynı zamanda görüntü üzerinde ilgili nesneleri (bounding box) işaretlemesini sağlayarak, halüsinsiyon riskini azaltır.

4. Güvenlik Analizi: Kritik Darboğazlar ve Riskler

VLM'lerin otonom sürüse getirdiği bilişsel yetenekler heyecan verici olsa da, bu teknolojinin gerçek yollara çıkabilmesi için aşması gereken ciddi güvenlik engelleri vardır. Bu bölümde, literatürde tespit edilen en önemli riskler; halüsinsiyon, gecikme ve uç durum başarısızlıklarını detaylandırılacaktır.

4.1 Halüsinsiyon Sorunu ve İstatistiksel Gerçekler

LLM'ler ve VLM'ler doğaları gereği deterministik değildir; istatistiksel olasılıklarla çalışırlar. Bu durum, "halüsinsiyon" (hallucination) olarak adlandırılan, modelin gerçekten olmayan nesneleri görmesi veya var olanları tamamen yanlış yorumlaması riskini doğurur. Otonom sürüs gibi hataya tahammülü olmayan bir alanda bu risk kritiktir.

Yapılan kapsamlı değerlendirmeler, güncel SOTA (State-of-the-Art) LLM'lerin sürüsle ilgili temel görevlerde "**halüsinsiyon görmeme oranının**" (**non-hallucination rate**) **sadece %57.95** olduğunu göstermektedir.² Bu, modelin ürettiği çıktıların %40'ından fazlasının potansiyel olarak hatalı, uydurma veya güvenilmez olduğu anlamına gelir. Bir sohbet botu için kabul edilebilir olabilecek bu hata oranı, 100 km/s hızla giden bir araç için felaket demektir.

4.2 Başarısızlık Senaryoları (Failure Cases)

GPT-4V gibi gelişmiş modellerin bile belirli sürüs senaryolarında ciddi hatalar yaptığı belgelenmiştir:

- **Kapı Açma (Dooring) Kazaları:** Park halindeki bir aracın kapısının aniden açılması senaryosunda, GPT-4V kaza sonrası görüntüleri analiz ederek nedeni doğru tespit edebilmiştir. Ancak kaza öncesi karar verme aşamasında, modelin kapının hafifçe aralandığını fark edemediği veya fark etse bile "yavaşla" gibi genel bir tavsiye verirken, gerekli olan "ani sola manevra" komutunu üretmediği görülmüştür. Ayrıca modelin bazen tehlike kaynağı olarak kapıyı değil, uzaktaki alakasız bir kırmızı aracı işaret ettiği (dikkat dağılıklığı/halüsinsiyon) gözlemlenmiştir.⁶
- **Konum ve Şerit Hataları:** Modelin, ego-aracın hangi şeritte olduğunu karıştırıldığı (örneğin en sağ şeritteyken orta şeritte olduğunu sanması) ve buna bağlı olarak "sağa dönülebilir" gibi hatalı ve tehlikeli kararlar ürettiği durumlar rapor edilmiştir.³

4.3 Donanım ve Gecikme (Latency) Kısıtlamaları

Otonom sürüs sistemleri, çevresel değişikliklere milisaniyeler içinde tepki vermelidir. Endüstri standartlarına göre ucta gecikme hedefi genellikle **10ms** civarındadır.²⁸ Ancak, milyarlarca parametreye sahip VLM'lerin (örn. GAIA-1 9B, DriveAgent-R1 3B) çıkışım (inference) süreleri, güçlü GPU'larda bile yüzlerce milisaniyeyi bulabilir.

- **Bulut vs. Uç (Edge):** Modelleri bulutta çalıştırma, ağ gecikmesi ve bağlantı kopma riski nedeniyle güvenli değildir. Ayrıca, araç sensörlerinden gelen terabaytlarca veriyi buluta yüklemek (uplink bandwidth) pratik değildir.
- **Uçta İşleme (On-Device):** Bu nedenle, modellerin araç üzerindeki donanımlarda

(örneğin NVIDIA Drive Thor, Qualcomm Snapdragon) çalıştırılması zorunludur. Ancak bu donanımların enerji, termal ve bellek kısıtlamaları, modellerin ciddi şekilde optimize edilmesini (quantization, pruning) gerektirir.²⁸

5. Halüsinasyon Tespit ve Azaltma Stratejileri

Güvenlik analizinde ortaya konan riskleri yönetilebilir seviyeye indirmek için literatürde çeşitli algoritmik ve mimari çözümler önerilmiştir. Bu stratejiler, modelin çıktılarını denetlemeyi ve doğrulamayı hedefler.

5.1 Düşük Rütbeli (Low-Rank) Matris Ayrıştırması

Bu teknik, birden fazla modelin (veya aynı modelin birden fazla çıktısının) "fikir birliğine" varmasını matematiksel bir zemine oturtur. Varsayalım ki bir senaryo için n farklı VLM çıktısı (caption) aldık. Bu metinler vektör uzayına gömülüerek bir M matrisi oluşturulur.

- **Teori:** Doğru ve gerçekçi açıklamalar, anlamsal olarak birbirine benzer olacağı için matriste "düşük rütbeli" (low-rank) bir yapı oluşturur. Halüsinasyonlar ise bu yapıdan sapan "seyrek" (sparse) gürültüler olarak belirir.
- **Uygulama:** Matris, Tekil Değer Ayrışımı (SVD) veya Robust PCA teknikleri ile $M = L + E$ şeklinde ayrıştırılır (L : Düşük rütbeli tatarlı bilgi, E : Seyrek hata matrisi). E matrisindeki normu (büyüklüğü) en küçük olan satırı karşılık gelen açıklama, en az halüsinasyon içeren ve en güvenilir çıktı olarak seçilir.
- **Başarı:** NuScenes veri setinde yapılan testlerde, bu yöntem **%87 doğrulukla** halüsinasyonuz açıklamaları seçebilmiş ve çoklu ajan tartışma yöntemlerine göre işlem süresini **%51-67** oranında kısaltmıştır.³⁰

5.2 Kendi Kendini Kontrol (SelfCheckGPT)

Bu yöntem, modelin kendi tutarlığını test etmesine dayanır. Modele aynı sahneyle ilgili birden fazla soru sorulur veya aynı soru farklı şekillerde yöneltilir.

- **Süreç:** Modelin ilk yanıtını cümlelere bölünür. Her cümle için, modelin ürettiği diğer yanıtlarla anlamsal tutarlılık (BERTScore, N-gram örtüşmesi, Doğal Dil Çıkarımı - NLI) kontrol edilir. Eğer bir cümle, diğer yanıtlardaki bilgilerle çelişiyorsa veya desteklenmiyorsa, bu cümle "halüsinasyon" olarak işaretlenir ve elenir. Bu yöntem, dış bir bilgi kaynağına ihtiyaç duymadan modelin kendi içsel tutarsızlıklarını yakalamasını sağlar.³¹

5.3 Hibrit Mimari: LLM-Hinted RL

LLM'lerin doğrudan karar verici olmasının risklerine karşı geliştirilen en güçlü mimari çözüm, **LLM-Hinted RL** (LLM İpuçlu Pekiştirmeli Öğrenme) yaklaşımıdır.

- **Çalışma Prensibi:** Bu yapıda LLM, direksiyonu tutmaz. Bunun yerine, karmaşık sahneyi analiz eder ve RL (Reinforcement Learning) ajanına "semantik ipuçları" (semantic hints) verir. Örneğin LLM, "Okul bölgесине, dikkatli ol" der. RL ajanı, bu ipucunu durum

uzayına (state space) ek bir girdi olarak alır ve kendi ödül fonksiyonunu (reward function) optimize ederek fiziksel eylemi (hız 30 km/s) belirler.

- **Avantaj:** Eğer LLM halüsinasyon görür ve yanlış bir ipucu verirse (örn. "Hızlan"), RL ajanı geçmiş eğitimlerinden gelen güvenlik kısıtlamaları (çarpışmadan kaçınma) sayesinde bu hatalı ipucunu göz ardı edebilir veya etkisini minimize edebilir. Bu yöntem, güvenlik kritik senaryolarda çarışma oranını **%11.4** azaltmıştır.²

6. Deneysel Değerlendirme ve Benchmark Analizi

VLM tabanlı sistemlerin performansı, simülasyon ortamlarında ve standart veri setlerinde yapılan testlerle ölçülmektedir. Ancak, geleneksel metriklerin (ortalama hata, doğruluk) ötesinde, güvenlik odaklı yeni nesil benchmarklara ihtiyaç vardır.

6.1 Standart Benchmarklar: CARLA ve nuScenes

- **CARLA Leaderboard:** Otonom sürüş algoritmalarının kapalı döngü performansını ölçen en prestijli platformdur. Modeller, Sürüş Skoru (Driving Score - DS), Rota Tamamlama (RC) ve İhlal Skoru (IS) üzerinden değerlendirilir.
 - **VLM Başarısı:** VLM tabanlı **SimLingo** modeli, sadece kamera görüntülerini kullanarak CARLA Leaderboard 2.0'da SOTA performans elde etmiş ve statik nesnelerle çarışmayı tamamen ortadan kaldırmıştır.³³
 - **DiLu vs. Klasik RL:** DiLu çerçevesi, bellek modülü sayesinde kavşak senaryolarında DQN ve PPO gibi klasik RL algoritmalarına göre **3 kat daha yüksek başarı oranı** (%61 vs %24) yakalamıştır.²⁰
- **nuScenes:** Açık döngü algı ve tahmin görevleri için standarttır. DriveGPT, ölçektekleme yasaları sayesinde nuScenes tahmin görevlerinde temel modellere göre çarışma oranını **%32** azaltmıştır.³⁴

6.2 İnsan Faktörü ve HABIT Benchmark'i

Mevcut benchmarkların en büyük eksikliği, gerçekçi insan davranışlarını (yaya, bisikletli) yeterince modelleyememesidir. CARLA'daki yayalar genellikle basit, kural tabanlı hareket ederler.

- **HABIT (Human Action Benchmark for Interactive Traffic):** Bu yeni benchmark, gerçek dünyadan (mocap ve video) elde edilen 4,730 adet yüksek doğruluklu yaya hareketini simülasyona entegre eder.
- **Kritik Bulgular:** CARLA Leaderboard'da neredeyse kusursuz çalışan InterFuser ve TransFuser gibi SOTA modeller, HABIT benchmark'ında test edildiğinde dramatik bir performans düşüşü yaşamıştır. Çarışma oranları **km başına 7.43'e** kadar çıkmış ve **AIS (Abbreviated Injury Scale)** yaralanma riski **%12.94** seviyelerinde ölçülmüştür.³⁵ Bu durum, mevcut modellerin "kurallara uymada" iyi olduğunu, ancak "insan niyetini okumada" ve öngörülemeyen davranışlara tepki vermede hala yetersiz kaldığını kanıtlamaktadır.

6.3 Karşılaştırmalı Performans Analizi

Aşağıdaki tablo, raporda incelenen farklı yaklaşımın temel performans ve güvenlik metriklerini özetlemektedir:

Model / Mimari	Temel Teknoloji	Güçlü Yön (Pros)	Zayıf Yön (Cons)	Ölçülen Güvenlik Performansı
DriveGPT	Otoregresif Transformer	Ölçeklenebilirlik, Rota Planlama	Yüksek Hesaplama Maliyeti	SOTA'ya göre çarisma oranında %32 azalma (nuScenes) ³⁴
DiLu	Kapalı Döngü + Bellek	Sürekli Öğrenme, Deneyim Birikimi	Bellek Yönetimi Karmaşıklığı	Kavşaklarda %61 Başarı Oranı (RL'den 3x daha iyi) ²⁰
LLM-Hinted RL	Hibrit (LLM + RL)	Halüsinsiyon Direnci, Kararlılık	Çift Model Eğitimi	Kritik durumlarda çarisma oranında %11.4 azalma ²
InterFuser	Transformer (SOTA)	Genel Leaderboard Başarısı	Yaya Etkileşimi Zayıflığı	HABIT benchmark'ında 5.24 çarisma/km ³⁵
TransFuser	Sensör Füzyonu	Sensör Çeşitliliği	Karmaşık Senaryo Hatası	HABIT benchmark'ında 7.43 çarisma/km ³⁵

Tablodan görüleceği üzere, saf Transformer veya Sensör Füzyonu modelleri (InterFuser, TransFuser) standart testlerde başarılı olsa da, gerçekçi insan etkileşimlerinde (HABIT) sınıfta kalmaktadır. Buna karşılık, LLM/VLM entegrasyonlu hibrit modeller (LLM-Hinted RL, DiLu), daha yüksek güvenlik ve uyum yeteneği sergilemektedir.

7. Gelecek Vizyonu, Sektörel Etkiler ve Yol Haritası

VLM'lerin otonom sürüse entegrasyonu, sadece teknik bir iyileştirme değil, endüstriyel bir paradigma değişimidir. Bu teknolojinin olgunlaşması, otomotiv sektöründe, sigortacılıkta ve yasal düzenlemelerde domino etkisi yaratacaktır.

7.1 Sentetik Veri Devrimi ve Eğitim

GAIA-1 gibi üretken dünya modelleri, otonom araç eğitiminde "veri darboğazını" çözecektir. Artık nadir görülen kaza senaryolarını yakalamak için milyonlarca kilometre test sürüşü yapmaya gerek kalmayacaktır. Bunun yerine, VLM'lere "Bir çocuğun park halindeki araçların arasından yola fırladığı karlı bir kiş günü senaryosu üret" komutu verilerek, modelin bu senaryoda sanal olarak binlerce kez eğitilmesi sağlanacaktır.¹⁴ Bu, geliştirme maliyetlerini düşürecek ve güvenliği artıracaktır.

7.2 Hukuki ve Etik Açıklanabilirlik

Otonom araç kazalarında en büyük sorun "sorumluluk" ve "nedensellik"tir. VLM'lerin CoT yeteneği, kaza sonrası analizlerde bir "kara kutu" yerine, aracın düşünce sürecini (log) sunmasını sağlayacaktır. Araç, "Kaza yaptım çünkü sola kırsayıdım kaldırımdaki yayalara çarpmaya riskim %90'dı, bu yüzden bariyerlere çarpmayı tercih ettim" şeklinde etik ve mantıksal bir açıklama sunabilecektir. Bu şeffaflık, sigorta şirketleri ve mahkemeler için kritik bir veri kaynağı olacaktır.

7.3 Kişiselleştirilmiş Sürüş Deneyimi

VLM'ler, araç içindeki yolcularla doğal dilde iletişim kurabilir. Yolcu, "Midem bulanıyor, biraz daha sakin sür" veya "Acelem var, güvenli sınırlar içinde biraz daha seri git" dediğinde, VLM bu soyut talepleri teknik sürüs parametrelerine (ivmelenme profili, takip mesafesi) dönüştürebilir. Bu, otonom araçların sadece bir ulaşım aracı değil, empatik bir asistan olmasını sağlayacaktır.

7.4 Uygulama Yol Haritası

- Aşama 1 (Kısa Vade):** VLM'lerin "Gölge Modu"nda (Shadow Mode) çalışması. Araç insan veya klasik algoritma tarafından sürülürken, VLM arka planda kararlar alır ancak müdahale etmez. VLM'in kararları ile gerçekleşen eylemler arasındaki farklar analiz edilerek model eğitilir.
- Aşama 2 (Orta Vade):** VLM'lerin "Karar Destek" olarak devreye girmesi. Navigasyon, şerit seçimi ve rota planlama gibi stratejik kararları VLM verirken, anlık kontrol hala klasik sistemlerdedir.
- Aşama 3 (Uzun Vade):** Donanım hızlandırmalarının (NPU, TPU) gelişmesi ve halüsinsiyon oranlarının düşmesiyle, VLM'lerin uçtan uca kontrolü devraldığı, tamamen bilişsel otonom sürüs.

8. Sonuç

Bu araştırma raporu, "Otonom Sürüş Senaryolarında Karar Verme Mekanizması Olarak Görme-Dil Modellerinin (VLM) Entegrasyonu" projesinin, otonom sürüş teknolojilerindeki tıkanıklığı aşmak için en umut verici ve yenilikçi yaklaşım olduğunu doğrulamaktadır. Klasik modüler sistemlerin karmaşıklığı ve uçtan uca sistemlerin açıklanamazlığı arasında sıkışan sektör için VLM'ler, "muhakeme yapabilen", "açıklayabilen" ve "öğrenebilen" bir üçüncü yol sunmaktadır.

Ancak, bu potansiyelin hayatı geçmesi, **güvenlik** konusundaki tavizsiz yaklaşımıma bağlıdır. VLM'ler, mevcut teknoloji seviyesinde doğrudan bir "kontrolcü" (driver) olmaktan ziyade, mükemmel bir "akıl hocası" (reasoner) ve "stratejist"tir. Raporun ortaya koyduğu en önemli sonuç, **LLM-Hinted RL** ve **Aktif Algı** gibi hibrit mimarilerin, VLM'lerin zekasını klasik kontrolcülerin güvenilirliği ile birleştiren en optimum çözüm olduğudur.

Sonuç olarak, otonom sürüşün geleceği, direksiyonu tutan ellerin (kontrol algoritmaları) değil, o ellere nereye ve neden gitmesi gerektiğini fisıldayan beynin (VLM) evriminde yatkınlıkta. Bu proje, o beyni inşa etme yolunda atılmış kritik ve vizyoner bir adımdır.

Alıntılanan çalışmalar

1. Integrating LLMs with ITS: Recent Advances, Potentials, Challenges, and Future Directions, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2501.04437v1>
2. arxiv.org, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2505.15793v2>
3. VLMs Guided Interpretable Decision Making for Autonomous Driving - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2511.13881v1>
4. Occ-LLM: Enhancing Autonomous Driving with Occupancy-Based Large Language Models, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2502.06419v1>
5. Think-Driver: From Driving-Scene Understanding to Decision-Making with Vision Language Models, erişim tarihi Aralık 15, 2025, <https://mllmav.github.io/papers/Think-Driver.%20From%20Driving-Scene%20Understanding%20to%20Decision-Making%20with%20Vision%20Language%20Models.pdf>
6. GPT-4V as Traffic Assistant: An In-depth Look at Vision Language Model on Complex Traffic Events - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2402.02205v3>
7. A Survey on Vision-Language-Action Models for ... - CVF Open Access, erişim tarihi Aralık 15, 2025, https://openaccess.thecvf.com/content/ICCV2025W/WDFM-AD/papers/Jiang_A_Survey_on_Vision-Language-Action_Models_for_Autonomous_Driving_ICCVW_2025_paper.pdf
8. HCRMP: A LLM-Hinted Contextual Reinforcement Learning Framework for Autonomous Driving - OpenReview, erişim tarihi Aralık 15, 2025, <https://openreview.net/pdf/1c6edc1362f9abc88143331bc8ca8f548a76312b.pdf>

9. Enhancing Autonomous Driving Systems with - Robotics, erişim tarihi Aralık 15, 2025, <https://www.roboticsproceedings.org/rss21/p140.pdf>
10. LeAD: The LLM Enhanced Planning System Converged with End-to-end Autonomous Driving - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2507.05754v1>
11. DRIVEAGENT-R1: ADVANCING VLM-BASED AUTONOMOUS DRIVING WITH ACTIVE PERCEPTION AND HYBRID THINKING - OpenReview, erişim tarihi Aralık 15, 2025, <https://openreview.net/pdf/0185e06303ae8d20ff7f460b2e705fb3b57067e5.pdf>
12. GAIA-1: A Generative World Model for Autonomous Driving - Wayve, erişim tarihi Aralık 15, 2025, <https://wayve.ai/wp-content/uploads/2024/04/2309.17080.pdf>
13. Scaling GAIA-1: 9-billion parameter generative world model for ..., erişim tarihi Aralık 15, 2025, <https://wayve.ai/thinking/scaling-gaia-1/>
14. GAIA-3: Scaling World Models to Power Safety and Evaluation - Wayve, erişim tarihi Aralık 15, 2025, <https://wayve.ai/thinking/gaia-3/>
15. Multimodal LLMs for Autonomous Driving – Part 2: Transforming the Road with GAIA-1's Generative World Model | by Azam Kowalczyk | Medium, erişim tarihi Aralık 15, 2025, <https://medium.com/@az.tayyebi/multimodal-langs-for-autonomous-driving-part-2-transforming-the-road-with-gaia-1s-generative-b7cc06eea4e7>
16. DriveGPT: Scaling Autoregressive Behavior Models for Driving - Personal Robotics Lab, erişim tarihi Aralık 15, 2025, <https://personalrobotics.cs.washington.edu/publications/huang2025drivegpt.pdf>
17. DriveGPT: Scaling Autoregressive Behavior Models for Driving - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2412.14415v1>
18. DriveGPT: Scaling Autoregressive Behavior Models for Driving - OpenReview, erişim tarihi Aralık 15, 2025, <https://openreview.net/forum?id=SBUXQakoJJ>
19. DiLu : A Knowledge-Driven Approach to Autonomous Driving with ..., erişim tarihi Aralık 15, 2025, https://proceedings.iclr.cc/paper_files/paper/2024/file/93c936b9e492def9c00782cab79dbc6d-Paper-Conference.pdf
20. Efficacy of Autonomous Vehicle's Adaptive Decision-Making Based on Large Language Models Across Multiple Driving Scenarios - IEEE Xplore, erişim tarihi Aralık 15, 2025, <https://ieeexplore.ieee.org/iel8/6287639/10820123/11039763.pdf>
21. (PDF) Efficacy of Autonomous Vehicle's Adaptive Decision-Making Based on Large Language Models Across Multiple Driving Scenarios - ResearchGate, erişim tarihi Aralık 15, 2025, https://www.researchgate.net/publication/392823509_Efficacy_of_Autonomous_Vehicle's_Adaptive_Decision-Making_based_on_Large_Language_Models_across_Multiple_Driving_Scenarios
22. DriveAgent-R1: Advancing VLM-based Autonomous Driving with Active Perception and Hybrid Thinking | OpenReview, erişim tarihi Aralık 15, 2025, <https://openreview.net/forum?id=r2g8TV4nJy>
23. Daily Papers - Hugging Face, erişim tarihi Aralık 15, 2025, <https://huggingface.co/papers?q=hybrid%20thinking%20framework>

24. CoT4AD: A Vision-Language-Action Model with Explicit Chain-of-Thought Reasoning for Autonomous Driving - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2511.22532v1>
25. CoT4AD: A Vision-Language-Action Model with Explicit Chain-of-Thought Reasoning for Autonomous Driving - ChatPaper, erişim tarihi Aralık 15, 2025, <https://chatpaper.com/paper/214202>
26. Retrieval-Based Interleaved Visual Chain-of-Thought in Real-World Driving Scenarios, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2501.04671v2>
27. VLMs Guided Interpretable Decision Making for Autonomous Driving - ResearchGate, erişim tarihi Aralık 15, 2025, https://www.researchgate.net/publication/397739276_VLMs_Guided_Interpretable_Decision_Making_for_Autonomous_Driving
28. Mobile Edge Intelligence for Large Language Models: A Contemporary Survey - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2407.18921v2>
29. LLM-Driven Offloading Decisions for Edge Object Detection in Smart City Deployments, erişim tarihi Aralık 15, 2025, <https://www.mdpi.com/2624-6511/8/5/169>
30. A Low-Rank Method for Vision Language Model Hallucination Mitigation in Autonomous Driving - ChatPaper, erişim tarihi Aralık 15, 2025, <https://chatpaper.com/paper/207931>
31. [Literature Review] LLMs Can Check Their Own Results to Mitigate ..., erişim tarihi Aralık 15, 2025, <https://www.themoonlight.io/en/review/lmcs-can-check-their-own-results-to-mitigate-hallucinations-in-traffic-understanding-tasks>
32. Self-consistency improves language model's robustness to imperfect... - ResearchGate, erişim tarihi Aralık 15, 2025, https://www.researchgate.net/figure/Self-consistency-improves-language-model-s-robustness-to-imperfect-prompts-on-GSM8K-It_tbl1_359390115
33. SimLingo: Vision-Only Closed-Loop Autonomous Driving with Language-Action Alignment, erişim tarihi Aralık 15, 2025, <https://liner.com/review/simlingo-visiononly-closedloop-autonomous-driving-with-languageaction-alignment>
34. ReAL-AD: Towards Human-Like Reasoning in End-to-End Autonomous Driving - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2507.12499v1>
35. HABIT: Human Action Benchmark for Interactive Traffic in CARLA - arXiv, erişim tarihi Aralık 15, 2025, <https://arxiv.org/html/2511.19109v1>
36. Generative Models in Autonomous Driving: GAIA-1 to GAIA-2 and the Realism Gap, erişim tarihi Aralık 15, 2025, <https://matt3r.ai/blogs/our-latest-thoughts/gaia-2-synthetic-data-autonomous-driving>