

Önsöz:

Bu özet, Yaz Yayınları Bilgisayar Bilimleri ve Mühendisliği kitabında Ramazan Katırcı ve Hilal Çelik tarafından yazılan "Transformer Mimarisi" başlıklı bölümün özetidir. Bu çalışmada, Transformer mimarisinin temel bileşenleri, dikkat mekanizmaları, Transformer mimarisinde paralelleştirme konuları açıklanmayı amaçlanmıştır. Metnin tamamına [bu link](#) üzerinden ulaşılabilir.

TRANSFORMER MİMARİSİ

1. GİRİŞ

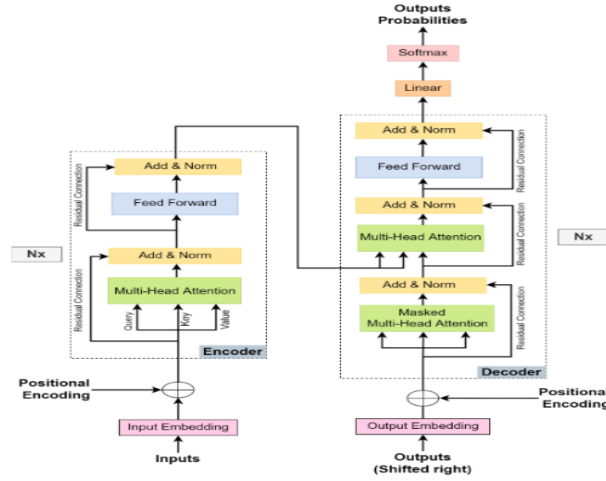
Transformer mimarisi, dikkat (attention) mekanizmalarını temel alan bir yapay zeka modelidir. Geleneksel yöntemlerden (RNN, CNN) farklı olarak, bilgiyi sıralamadan bağımsız işler. Dikkat mekanizmaları, her bir girdiyi diğer girdilerle olan ilişkileri üzerinden değerlendirerek paralel bilgi işleme yeteneği sağlar. Ancak, doğal dil işleme çalışmalarında sıralama bilgisi önemli olduğundan Transformer modeli bu bilgi eksikliğini gidermek için konumsal kodlama (positional encoding) kullanır. Konumsal kodlama, belirteçlerin sırasını modele dahil eder. Böylece model, işlem sırasında belirteçlerin sıralama bilgisini bilir. Ayrıca makine öğrenmesi modellerine göre daha fazla hiperparametre içerir ve katmanlı mimariye sahiptir. Bu sayede model daha etkili hale gelir. Makine öğrenmesi; sağlık, tarım, e-ticaret, finans ve eğitim gibi farklı alanlarda kullanılmaktadır. Makine öğreniminin ilk dönemlerinde kural tabanlı sistemler yaygın olarak kullanılırken bağlamsal anlamı yakalamada yetersiz kalmıştır. Bu sınırlamalar, erişim tabanlı ve ardından üretim tabanlı yaklaşımların geliştirilmesiyle aşılmıştır. Erişim tabanlı yöntemler belirlenmiş veri setlerinden yanıt seçerken üretim tabanlı modeller kullanıcı sorgularından yeni yanıtlar türetebilir. Erişim tabanlı sistemler, yeni içerik oluşturmaz, mevcut bilgi havuzundan en uygun yanıtları seçer. Üretim tabanlı yaklaşımlar, dil modelleri kullanarak yanıtlar üretir ve özetleme, makine çevirisi, metin tamamlama, sohbet robotları gibi alanlarda başarılı sonuçlar vermektedir. Günümüzde en önemli üretim tabanlı modellerden biri, büyük veri setlerinden öğrenen ve bağlama dayalı metinler üreten Transformer tabanlı GPT modelidir.

2. Transformer Mimarisi ve Bileşenleri

Transformer mimarisi, tamamen dikkat mekanizmalarına temel alan geleneksel makine öğrenimi yöntemlerinden daha fazla bileşen içeren yenilikçi bir mimaridir ve encoder-decoder yapısını içerir. Encoder (kodlayıcı), giriş dizisini işleyerek gizli bir temsile dönüştürürken decoder (kod çözücü), bu temsili alıp çıkış dizisini üretir. Gömme ve konumsal kodlama, kelimelerin bağlamını anlamaya yardımcı olur. Çoklu Dikkat Başlıkları (Multi-Head Attention), birden fazla dikkat mekanizmasının paralel çalışmasını sağlar. Katman Normalizasyonu ve artık bağlantılar, modelin performansını

artırarak eğitim sürecini kolaylaştırır. Feedforward Neural Network (FNN), verileri giriş katmanından çıkışa doğru işler. Transformer, paralelleştirme yeteneği sayesinde verinin farklı özelliklerine aynı anda odaklanarak hızlı işlem yapar, böylece her kelimenin ilişkilerini paralel olarak öğrenir. Bu, sıralı hesaplamaların aksine daha hızlı sonuçlar elde edilmesini sağlar.

Şekil 1. Transformer Mimarisi



2.1. Kodlayıcı-kod çözücü (Encoder-Decoder) Yapısı

Seq2seq (Sequence to Sequence) modeli, farklı uzunluktaki giriş ve çıkış dizilerini eşleştirmeyi amaçlayan bir makine öğrenimi yöntemidir ve kodlayıcı (encoder) ile kod çözücü (decoder) bileşenlerinden oluşur. Kodlayıcı, girdi dizisinin anlamını özetleyerek bağlamsal bir temsil üretirken kod çözücü bu temsili kullanarak hedef çıktıyı üretir. Transformer mimarisinde, kodlayıcı çoklu başlı dikkat (multi-head self-attention) ve konumuna dayalı ileri beslemeli ağdan oluşan iki alt katmandan meydana gelir. Kod çözücü, benzer şekilde kodlayıcının çıktısını işleyerek daha bağlamsal ve anlamlı çıktı üretir. Ayrıca kod çözücünün kendine dikkat alt katmanı, maskelenmiş bir mekanizma ile yalnızca önceki bilgilerin kullanılmasını sağlar. Bu, dil modelleme gibi görevlerde, bir kelimenin tahmininin yalnızca kendisinden önceki kelimelere bağlı olmasını sağlar.

2.2. Gömme ve Konumsal Kodlama

Transformer mimarisinde, kelime gömme (word embedding) kelimeleri sabit boyutlu vektörlerle temsil ederek anlamlarını yakalar. Ancak Transformer mimarisi, sıralı veri yapısına sahip olmadığından sadece kelime gömme yeterli olmaz. Konumsal kodlamada

bulunur ve kelimenin cümledeki sırasını belirtmek için eklenir. Böylece, kelimenin anlamını sadece bağlamına göre değil, aynı zamanda cümledeki sırasına göre de öğrenebilir.

2.3. Kendine Dikkat Mekanizması ve Çoklu Dikkat Başlıkları

Dikkat mekanizması, cümlelerin parçaları arasındaki ilişkileri öğrenerek dil modellerinin uzun bağımlılıkları daha etkili şekilde anlamasına olanak tanır. 2017'de geliştirilen Transformer mimarisi, dikkat mekanizmasını merkezi bir öge olarak kullanarak paralel işlemlerle daha hızlı ve verimli çalışmayı mümkün kılmıştır. Bu mekanizma sayesinde, metin içindeki belirteçlerin birbirleriyle olan ilişkilerini daha iyi kavramak mümkün hale gelmiştir. Dikkat fonksiyonu, bir sorgu ile bir dizi anahtar-değer çiftini bir çıktıya eşlemek olarak tanımlanabilir. Öz-dikkat (self-attention) alt katmanları, Transformer mimarisinde, giriş verisinin farklı yönlerine dikkat ederek bilgi temsillerini zenginleştirir. Çok başlı dikkat (multi-head attention) mekanizması, birden fazla dikkat başlığının birleşiminden oluşur ve sıralı veri içerisindeki farklı özellikleri yakalamasına olanak tanır. Bu mekanizma, kısa ve uzun aralıklarla bilgi toplama yeteneği sunarak modelin performansını artırır ve uzun bağımlılıkların etkin öğrenilmesini sağlar. Her dikkat başlığı, sorgu (query) ve anahtar (key) vektörleri arasındaki ilişkilere göre dikkat ağırlıkları atar. Sorgu, bir kelimenin diğer kelimelerle ilişkisini bulmak için kullanılır, anahtar ise ilişkili bilgileri içerir. Model, sorgu ve anahtar vektörlerini karşılaştırarak hangi kelimelere dikkat etmesi gerektiğini belirler. Değer (value) vektörleri, dikkat dağılımı sonucu en alakalı kelimelere karşılık gelen bilgiyi taşır ve çıkış (output) hesaplanır. Bu süreç, modelin daha kapsamlı bir öğrenme süreci gerçekleştirerek performansını iyileştirir.

2.4. ADD & Katman Normalizasyonu

Add katmanı, önceki katmanın çıktısını sonrakine ekleyerek modelin derin katmanlarda bilgi kaybını önler ve öğrenme sürecini kolaylaştırır. Bu işlem, gradyanların daha etkili yayılmasını sağlar, derin ağların kullanımını mümkün kılar. Transformer mimarisinde, her alt katman (kendine dikkat ve feed-forward) bir "Add & Norm" katmanı içerir. Bu işlemde, önceki katmandan gelen giriş, çıktıya eklenir ve ardından normalizasyon yapılır. Bu yapı, modelin eğitim sürecini hızlandırır, stabilitesini artırır ve performansını iyileştirir. Normalizasyon için her pozisyonun ortalama ve standart sapması hesaplanır.

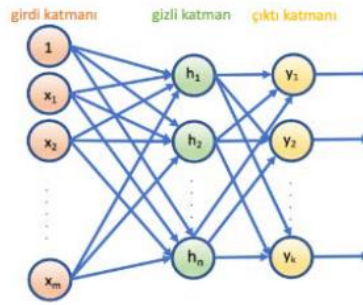
2.5. Artık Bağlantılar (Residual Connections)

Artık bağlantı, bir katmandan elde edilen çıktının, birkaç katman ileriye taşınarak daha sonraki katmanlardaki çıktıya eklenmesi prensibine dayanır. Bu sayede model, her katmanda öğrenmesi gereken farkları (residuals) öğrenir, bilgiyi daha verimli bir şekilde işler ve derinliğinin artması durumunda bile öğrenmenin etkin bir şekilde devam etmesini sağlar. Transformer mimarisinde, kaybolan gradyanları önlemek amacıyla artık bağlantılar hem dikkat hem de ileri besleme katmanlarına uygulanır. Katman normalizasyonu, artık bağlantıdan sonra uygulanarak her katmandaki çıktıyı normalize eder ve modelin daha kararlı bir şekilde öğrenmesini sağlar.

2.6. Feed-Forward Ağlar

İleri Beslemeli Sinir Ağı (FNN), verinin yalnızca ileri doğru hareket ettiği ve geriye doğru bir akışın olmadığı bir yapıya sahiptir. FNN'ler genellikle giriş, gizli ve çıkış katmanlarından oluşur. Her katmanda yer alan nöronlar, aktivasyon fonksiyonları kullanarak veriyi işleyerek girdi verilerini işler. Transformer mimarisinde, her kodlayıcı ve kod çözücü katmanı, dikkat alt katmanlarına ek olarak tam bağlantılı bir ileri beslemeli ağ (FFN) içerir.

Şekil 4. İleri Beslemeli Sinir Ağı



Kaynak: (Çakır and Angın 2021)

2.7. Transformer Mimarisinde Parallelleştirme

Parallelleştirme, Transformer modelinde giriş verisinin farklı bölümlerinin eşzamanlı olarak işlenmesini sağlar, bu da eğitim ve çıkarım süreçlerinde verimliliği artırır. Büyük modellerin geniş veri kümeleriyle hızlı bir şekilde eğitilebilmesi için paralel eğitim yapılabilir.

Transformer mimarisi, özellikle dikkat mekanizmaları sayesinde paralel hesaplamaları mümkün kılar. Bu mimaride, her girişin diğerleriyle olan ilişkisi aynı anda değerlendirilir. Bu da RNN ve LSTM gibi sıralı yöntemlerden farklıdır. Ayrıca ileri beslemeli ağlar her katmanda

bağımsız çalışarak paralel hesaplamalar yapılmasını sağlar. Transformer mimarisinde çoklu dikkat başlıkları, giriş verisindeki çeşitli özellikleri eşzamanlı olarak öğrenir. Her başlık, verinin farklı bir boyutunu analiz eder ve bu başlıklardan elde edilen sonuçlar birleştirildiğinde daha zengin ve kapsamlı bir temsil oluşturulur. Ölçeklendirilmiş nokta çarpımı dikkati fonksiyonu, dikkat modülü içinde paralel olarak uygulanarak giriş dizisinin farklı bölümleri arasındaki maksimum bağımlılıkları çıkarır. Her dikkat başlığı, kendi öğrenilebilir ağırlıklarıyla bağımsız dikkat hesaplamalarını gerçekleştirir. Böylece model verinin farklı temsillerini paralel şekilde öğrenir.

Önsöz:

Bu özet, Deepak Kumar ve Shoumya Singh tarafından yazılan "*Transformer Architectures for Large Language Model: From BERT to GPT-3 and Beyond*" başlıklı makalenin Türkçe özetidir. Makalenin tamamına [bu link](#) üzerinden ulaşılabilir.

ÖZET

Tranformer mimariler, doğal dil işleme alanında devrim yaratarak BERT ve GPT-3 gibi gelişmiş modellerin geliştirilmesine olanak sağlamıştır. Büyük metin verileriyle eğitilen bu modeller, insan diline benzer metinler üretebilme ve karmaşık dil görevlerini yüksek doğrulukla gerçekleştirme yeteneğine sahiptir. Bu çalışma, büyük dil modellerinin kullanım alanlarını, karşılaştıkları zorlukları ve gelecekteki yönelimleri ele almaktadır.

Anahtar Kelimeler: Büyük Dil Modelleri, Makine Öğrenim, Yapay Zeka

I. Giriş

Son yıllarda doğal dil işleme (NLP) alanında büyük ilerlemeler kaydedilmiş, özellikle dönüştürücü mimarilerinin (BERT, GPT-3) gelişimi bu alanda devrim yaratmıştır. Bu modeller, büyük veri kümeleri üzerinde eğitilerek dilin karmaşık kalıplarını ve anlamsal ilişkilerini öğrenmiş, metin sınıflandırma, duygu analizi, dil çevirisi ve soru-cevap sistemleri gibi görevlerde üstün başarı göstermiştir. Ancak bazı sınırlamaları da bulunmaktadır.

Dönüştürücüler, uzun mesafeli bağımlılıkları yakalama ve veriler arasındaki karmaşık ilişkileri anlama yetenekleriyle makine öğrenmesi uygulamalarında kritik bir rol oynamaktadır. Günümüzde araştırmalar, bu mimarilerin daha verimli ve ölçeklenebilir hale getirilmesine, hesaplama maliyetlerinin düşürülmesine ve hafif modellerin geliştirilmesine odaklanmaktadır. Ayrıca dikkat mekanizmaları ve kendi kendine denetimli öğrenme teknikleri sayesinde etiketlenmemiş verilerden öğrenme kapasitesi artırılmaktadır.

Daha gelişmiş dil anlama ve üretme modellerine olan ihtiyaç arttıkça, dönüştürücü mimarilerinin yeni yenilikler ve ilerlemeler kaydetmesi beklenmektedir. Bu gelişmeler, NLP'nin yanı sıra diğer alanlarda da önemli etkiler yaratacaktır.

II. Literatür Taraması

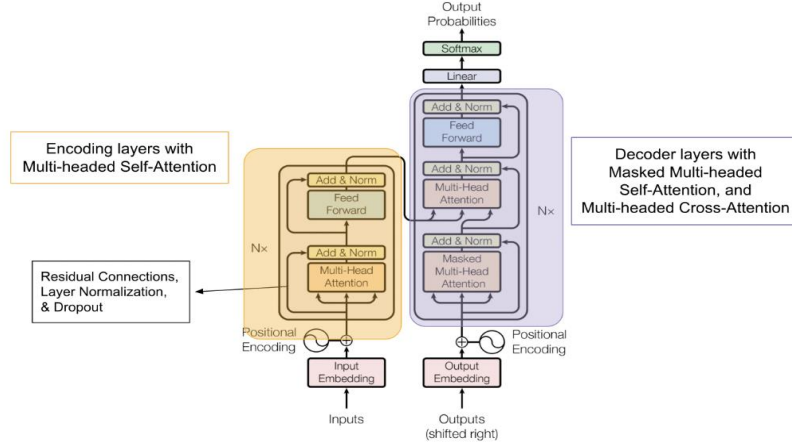
Son yıllarda yapılan akademik alıřmalar, transformatör mimarilerinin, doęal dil iřleme (NLP) alanındaki önemli ilerlemelerini incelemektedir. BERT ve GPT-3 gibi modeller, metin sınıflandırma, adlandırılmış varlık tanıma, duygu analizi, dil evirisi ve soru-cevaplama gibi görevlerde üstün performans göstermiştir. BERT, maskelenmiş dil modelleme yöntemiyle bağlamsal ilişkileri anlamada önemli bir adım atarken GPT-3, geniş ölçekli ön eğitimi sayesinde bağlamsal olarak tutarlı metinler üretebilmektedir. Ancak bu modellerin getirdięi büyük gelişmelere rağmen bazı sınırlamaları da bulunmaktadır.

Modern NLP Görevlerinde Transformer Modelleri

Transformatörler, doęal dil iřleme (NLP) alanında devrim yaratarak daha iyi dil anlama ve üretim yetenekleri sağlamıştır. BERT ve GPT-3 gibi modeller, metin sınıflandırma, adlandırılmış varlık tanıma, duygu analizi, dil evirisi ve soru-cevaplama gibi görevlerde yeni standartlar belirlemiştir. Transformatörlerin en önemli özelliklerinden biri, dil verisi içindeki uzun menzilli bağımlılıkları yakalayabilme yetenekleridir. Bu sayede insan benzeri metinler üretebilirler.

Transformatörlerin kendine dikkat (self-attention) mekanizması, cümledeki kelimelerin önemini değerlendirerek bağlamsal nüansları etkili şekilde yakalayabilmelerini sağlar. Ayrıca büyük miktarda metin verisiyle yapılan ön eğitim, dilin anlamsal ve sözdizimsel yapısını derinlemesine öğrenmelerine yardımcı olur. Bu da onların eřitli dil görevlerinde üstün performans göstermelerini sağlar.

Son yıllarda yapılan ilerlemeler, transformatörleri daha verimli ve kaynak dostu hale getirmiştir. Hafif modeller, yüksek performansı hesaplama gereksinimlerini azaltarak sağlar. Dikkat mekanizmaları ve kendi kendine denetimli öğrenme teknikleri, az veriyle etkili öğrenmelerini ve etiketlenmemiş verilerden faydalanmalarını sağlamıştır. Bu gelişmeler, transformatörlerin daha karmaşık hale gelmesini sağlayarak geleneksel dil görevlerinin ötesinde daha geniş uygulama alanlarına yayılmalarına yol açacaktır. Şekil 1, BERT ve GPT modellerinin transformatör mimarisini göstermektedir.



Şekil 1: Transformer Mimarisi

BERT Mimarisi ve Temel Özellikler

BERT (Bidirectional Encoder Representations from Transformers), 2018 yılında geliştirilmiş bir transformatör tabanlı modeldir ve NLP görevlerinde yeni standartlar belirlemiştir. BERT, iki yönlü (bidirectional) bir transformatör kodlayıcısı kullanarak cümledeki kelimelerin bağlamını her iki yönden de değerlendirir. Model, ön eğitim sırasında büyük metin verileriyle maskelenmiş kelimeleri tahmin etmeyi öğrenir. Bu da dilin anlamsal ve sözdizimsel yapısını derinlemesine anlamasına olanak tanır. İnce ayar aşamasında etiketlenmiş verilerle daha spesifik görevlerde eğitilir. BERT, büyük etiketlenmemiş metin verisiyle dilin genel desenlerini öğrenir ve farklı NLP görevlerine uyarlanabilir. Bu özellikler, BERT'in NLP görevlerinde olağanüstü performans sergilemesini sağlar.

GPT-3 Mimarisi ve Temel Yenilikler:

GPT-3 (Generative Pre-trained Transformer 3), BERT'e kıyasla daha büyük bir modeldir ve 175 milyar parametreye sahiptir. Bu büyük model boyutu, GPT-3'ün düşük örneklerle bile yüksek performans sergilemesini sağlar. GPT-3, metin üretme ve soru-cevaplama görevlerinde tutarlı ve bağlamsal olarak uygun metinler üretir. Model, eksik metinleri tamamlayabilir, özetleme yapabilir ve kod üretebilir. GPT-3, autoregresif bir dil modeli olarak metnin devamını tahmin ederek anlamlı ve doğru metinler üretir.

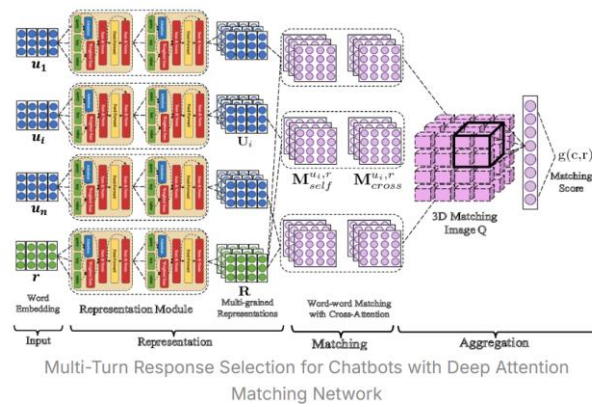
GPT-3 ile İlgili Gelecekteki Transformatör Modellerinin Potansiyel Uygulamaları ve Etkileri:

Transformatör mimarilerinin ölçeklenebilirliği, daha büyük ve güçlü modellerin geliştirilmesini sağlar. Gelecekteki transformatör modelleri, özellikle GPT gibi yapılar, doğal dil işleme (NLP) alanında büyük potansiyele sahiptir. Bu modeller, ölçeklenebilirlikleri

sayesinde daha büyük ve güçlü sistemlerin geliştirilmesine olanak tanırken ince ayar yetenekleriyle hukuk, tıp gibi özelleşmiş alanlarda kullanılabilecek dil modelleri oluşturma imkanı sunar. Ancak büyük veri bağımlılığı, üretilen metinlerde önyargı ve şeffaflık eksikliği gibi sınırlamalar bulunmaktadır. Gelecekteki transformatör modelleri, metinlerdeki önyargıları azaltarak adaleti artırmayı ve açıklanabilirlik ile şeffaflığı geliştirmeyi hedefleyecektir. GPT-3 gibi modellerin "kara kutu" yapısı, bu alanda daha fazla araştırma gerektirmektedir. Bu modeller, özelleşmiş dil modelleri oluşturma ve önyargıları azaltma gibi geniş uygulama potansiyeline sahiptir. Gelecekteki çalışmalar, mevcut sınırlamaları aşarak daha güvenilir ve etkili sistemler geliştirmelidir.

III. BÜYÜK DİL MODELLERİNİN UYGULAMA ALANLARI

BERT'ten GPT-3'e ve sonrasına kadar olan büyük dil modelleri, doğal dil işleme alanını derinden etkileyerek otomatik dil çevirisi, duygu analizi, metin özetleme, sohbet robotları ve içerik üretimi gibi uygulamalarda önemli iyileşmeler sağlamıştır. Bu modeller; eğitim, iş dünyası ve araştırma gibi sektörlerde devrim niteliğinde dönüşümler yaratma potansiyeline sahiptir. Örneğin, GPT-3 gibi modeller, öğrencilere sanal öğretmenlerle kişiselleştirilmiş ve etkileşimli öğrenme deneyimleri sunarak eğitim paradigmasını değiştirebilir. Sonuç olarak transformer mimarilerinin gelişimi, yeni uygulamalar ve olanaklar ortaya koymuş ve bu modellerin etkisi sektörel anlamda genişlemektedir. Şekil 2, LLM modelinin Chat Bot ile ilgili kullanımını göstermektedir.



Şekil 2

Transformer Modellerini Kullanan Gerçek Dünya Uygulama Alanlarının Keşfi ve Başarı Hikayeleri

Transformer mimarilerindeki hızlı ilerlemeler, müşteri hizmetleri ve bilimsel araştırmalarda başarılı uygulamalar göstermiştir. Örneğin, H&M ve Bank of America gibi büyük şirketler, müşteri etkileşimini artırmak ve hizmetleri otomatikleştirmek amacıyla transformer destekli yapay zeka kullanmaktadır. Ayrıca BERT ve GPT-3 gibi modeller, bilimsel literatürün analiz edilmesi ve yeni bakış açıları geliştirilmesi için büyük potansiyel sunmaktadır. Bu teknolojiler, müşteri memnuniyetini artırma ve araştırma süreçlerini dönüştürme konusunda önemli fırsatlar yaratmaktadır.

Transformer İlerlemelerinin Sağlık, Finans ve Eğitim Gibi Sektörler Üzerindeki Etkilerinin Analizi

Transformer mimarilerindeki ilerlemeler, sağlık, finans ve eğitim sektörlerinde büyük potansiyel taşımaktadır. Sağlık sektöründe, büyük dil modelleri kişiselleştirilmiş tedavi önerileri sunarak daha verimli sağlık hizmeti sağlar. Finans sektöründe, piyasa verileri üzerinde duygu analizi ve risk değerlendirmesi yapabilir. Eğitimde ise 7/24 küresel bilgiye sahip sanal öğretmenler sunarak öğretim yaklaşımını dönüştürüp öğretim üyelerinin rolünü öğrenme kolaylaştırıcılara dönüştürebilir. Bu gelişmeler, birçok sektörde önemli değişim ve iyileştirmeler sağlayabilir.

IV. Zorluklar ve Dikkate Alınması Gerekenler

Transformer mimarilerinin önemli ilerlemeleri olsa da bazı zorluklar ve sınırlamalar mevcuttur. En büyük zorluk, büyük modellerin eğitimi için gereken yüksek hesaplama gücüdür. Bu da eğitim sürecini maliyetli ve zaman alıcı hale getirebilir. Ayrıca bu modellerin eğitim verilerindeki önyargıları kodlayabilmesi, adaletsiz veya ayrımcı sonuçlara yol açabilir. Gizlilik ve güvenlik endişeleri de artan model sofistikeliğiyle birlikte gündeme gelmektedir. Etik sorunlar, önyargı azaltma, aşırı bağımlılık ve yorumlanabilirlik eksikliği gibi zorluklar da dikkate alınması gereken diğer önemli sorunlardır.

V. SONUÇ

Transformer mimarileri, BERT'ten GPT-3'e kadar olan gelişmelerle, doğal dil işleme alanında devrim yaratmıştır. Bu modeller, sağlıkta kişiselleştirilmiş tedavi önerileri, finansal duygu analizi ve eğitimde sanal öğretim gibi ileri düzey görevleri mümkün kılacak şekilde dil anlama ve üretiminde önemli ilerlemeler sağladı. Ancak bu modellerin eğitim ve kullanımında hesaplama gücü, önyargılar ve kullanıcı gizliliği gibi kritik hususlar göz önünde bulundurulmalıdır.

Transformer mimarilerinin potansiyelini daha geniş uygulamalar için açığa çıkarmak için daha verimli ve ölçeklenebilir modellerin geliştirilmesine yatırım yapılması gerekmektedir. Aynı zamanda, eğitim verilerindeki önyargıların azaltılması için veri toplama ve düzenleme süreçlerine dikkat edilmelidir. Politika yapıcıları ve endüstri uzmanları, sorumlu kullanım için yönergeler oluşturmali ve kullanıcı gizliliği ile veri güvenliğini sağlamalıdır. Bu şekilde, transformer mimarileri dil işleme ve anlama biçimimizi devrim niteliğinde deęiştirebilir.