



Kahramanmaraş Sütçü İmam University Journal of Engineering Sciences



Geliş Tarihi : 08.05.2024
Kabul Tarihi : 25.06.2024

Received Date : 08.05.2024
Accepted Date : 25.06.2024

İSTEM MÜHENDİSLİĞİNİN BÜYÜK DİL MODELLERİNİN PERFORMANSINDAKİ ROLÜ: ANALİZ VE UYGULAMA ÖRNEKLERİ

THE ROLE OF PROMPT ENGINEERING IN THE PERFORMANCE OF LARGE LANGUAGE MODELS: ANALYSIS AND APPLICATION EXAMPLES

Fatma Gülşah TAN^{1*} (ORCID: 0000-0002-2748-0396)

Asım Sinan YÜKSEL¹ (ORCID: 0000-0003-1986-5269)

Muhammed Abdulhamid KARABIYIK² (ORCID: 0000-0001-7927-8790)

¹ Süleyman Demirel Üniversitesi, Bilgisayar Mühendisliği Bölümü, Isparta, Türkiye

² Niğde Ömer Halisdemir Üniversitesi, Bor Meslek Yüksekokulu, Niğde, Türkiye

*Sorumlu Yazar / Corresponding Author: Fatma Gülşah TAN, gulsah@ksu.edu.tr

ÖZET

İstem mühendisliği, büyük dil modellerinin yeteneklerini artırmak için kritik bir teknik olarak ortaya çıkmıştır. İstem adı verilen talimatlar ile model parametrelerini değiştirmeden ince ayar yapma imkânı sunarak, bu modellerin çeşitli görevlerde üstün performans göstermesini sağlar. Bu çalışmanın temel amacı, büyük dil modeli performansını iyileştirmek, hesaplama maliyetlerini azaltmak ve kullanıcı deneyimini geliştirmek için istem mühendisliğinin nasıl etkili kullanılacağını göstermektir. Çalışmada, farklı uygulama alanlarına göre kategorize edilen ve son teknoloji 15 istem mühendisliği tekniği analiz edilmiştir. Bu teknikler, sıfır atış ve birkaç atış istemlerinden düşünce zinciri ve otomatik düşünce zinciri istemlerine kadar çeşitlilik göstermektedir. Her bir tekniğin avantajları ve dezavantajları detaylı olarak değerlendirilmiş ve performans artışının nasıl sağlandığı örnek senaryolarla gösterilmiştir. Araştırmanın sonuçları, istem mühendisliğinin büyük dil modellerinin çeşitli görev ve uygulamalarda performansını artırmada önemli bir rol oynadığını göstermektedir. Özellikle az verili öğrenme senaryolarında verimliliği artırmak ve önyargı, tutarsızlık gibi zorlukları azaltmak için yenilikçi istem mühendisliği tekniklerinin başarılı performans sergiledikleri görülmüştür. Bu bulgular, araştırmacılar ve uygulayıcılar için yol gösterici bir kaynak olarak hizmet edecek ve büyük dil modellerinin geniş çapta uygulanabilirliğini artıracaktır. Çalışmamız, istem mühendisliğinin daha iyi anlaşılmasına katkı sağlayacak ve gelecekteki araştırmalara ışık tutacaktır.

Anahtar Kelimeler: Büyük dil modelleri, istem mühendisliği, doğal dil işleme, yapay zekâ

ABSTRACT

Prompt engineering has emerged as a critical technique for increasing the capabilities of large language models. It enables these models to perform superiorly in various tasks by providing the opportunity to fine-tune the model without changing the parameters through instructions called prompts. The main goal of this work is to show how to effectively use prompt engineering to improve large language model performance, reduce computational costs, and improve user experience. In the study, 15 state-of-the-art prompt engineering techniques, categorized according to different application areas, were analyzed. These techniques range from zero-shot and few-shot prompts to chain-of-thought and automatic-chain-of-thought prompts. The advantages and disadvantages of each technique are evaluated in detail and how the performance increase is achieved is shown with example scenarios. The results of the research show that prompt engineering plays an important role in improving the performance of large language models in a variety of tasks and applications. Innovative prompt engineering techniques have been shown to perform successfully to increase efficiency and reduce difficulties such as bias and inconsistency, especially in low-data learning scenarios. These findings will serve as a guiding resource for researchers and practitioners and will

ToCite: TAN, F.G., & YÜKSEL, A.S., & KARABIYIK, M.A., (2024), İSTEM MÜHENDİSLİĞİNİN BÜYÜK DİL MODELLERİNİN PERFORMANSINDAKİ ROLÜ: ANALİZ VE UYGULAMA ÖRNEKLERİ. Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi, 27(4), 1401-1420.

increase the broad applicability of large language models. Our study will contribute to a better understanding of prompt engineering and shed light on future research.

Keywords: Large Language Models, prompt engineering, natural language processing, artificial intelligence

GİRİŞ

Büyük dil modellerinin (BDM) optimize edilmesinde son yıllarda önemli ilerlemeler kaydedilmiştir. Bu modeller derin öğrenme mimarilerine dayanmaktadır ve büyük veri kümeleri üzerinde eğitilmektedir. İstem mühendisliği, BDM'lerin verimli bir şekilde uygulanması ve optimize edilmesinde belirli amaç ve hedeflere odaklanan bir yaklaşımdır. Bu hedefler genellikle dil modeli performansının iyileştirilmesini, hesaplama maliyetlerinin azaltılmasını, gizliliğin korunmasını ve kullanıcı deneyiminin iyileştirilmesini içerir. Dil modellerini etkili bir şekilde uygularken ve optimize ederken istem mühendisliğinin ana amaç ve hedeflerini dikkate almak önemlidir. Bu hedeflere başarıyla ulaşılması, büyük dil modellerinin yaygın kullanımının ve getirdikleri faydaların anlaşılmasına yardımcı olacaktır.

Literatürde, çeşitli çalışmalar istem mühendisliğinin BDM'lerin performansına olan katkılarını vurgulamıştır. Örneğin, (Lester vd., 2021), benzersiz görevleri yerine getirmek için dil modelinin koşullandırılmasının yardımıyla güçlü bir mekanizma geliştirmişlerdir. (Polverini, ve Gregorcic, 2024a), fizik eğitimi bağlamında ChatGPT-4 ile birlikte dil modellerinin genel performansını artırmak için istem mühendisliği tekniklerinin kapasitesini gözlemlemişlerdir. İstem mühendisliği literatürüne hızlı bir bakış sunarak fizik eğitimi alanındaki BDM'lerin kapasitesinden bahsetmişlerdir. (Seo vd., 2022), özellikle az verili öğrenme senaryolarında verimliliği artırmak için yenilikçi istem mühendisliği tekniklerini tanıtmışlardır. Veri eksikliğinin üstesinden gelmek için küçük örnekli bir öğrenme yöntemi ve basit bir şablon ekleme yöntemi geliştirip bunları Kore veri setine uygulamışlardır. Dil modelleri farklı alanlara genişletildikçe uyarlanabilirlik ve etkililik açısından zorluklar ortaya çıkmaktadır. (K. Zhou vd., 2022), bu zorlukların belirlenmesinde ve etkili bir şekilde ele alınmasında istem yönetiminin önemli rolünü vurgulamıştır. Çeşitli görüntü tanıma görevlerini yerine getirmek için iki farklı uygulama önermişlerdir. Çok sayıda deneyde, 16 atış için ortalama kazancı yaklaşık %15 olduğunu, maksimum kazancı ise %45'in üzerinde olduğunu bildirmişlerdir. Geliştirilen yöntem öğrenmeye dayalı bir yaklaşım olmasına rağmen, bu yaklaşımın manuel sıfır atış istemi uygulayan modellere kıyasla üstün genelleme performansı sağladığını göstermişlerdir.

Gerçek hayattaki sorunları çözmek için BDM'leri etkili bir şekilde uygulamak ve optimize etmek, ölçeklenebilirlik, verimlilik ve gizlilik gibi önemli teknik zorlukları içerir. Bu nedenle, bu zorlukları etkili bir şekilde ele almak için istem mühendisliği gibi yenilikçi yöntem ve stratejiler kullanılmalıdır. Bu çalışmada, istem mühendisliği teknikleri detaylı olarak incelenmiş ve sundukları avantajlar, performans artışının nasıl sağlandığı örnek senaryolarla değerlendirilmiştir. Farklı uygulama alanlarına göre kategorize edilen istem mühendisliği teknikleri artıları ve eksileri tartışılmıştır.

MATERYAL VE YÖNTEM

İstem Mühendisliği

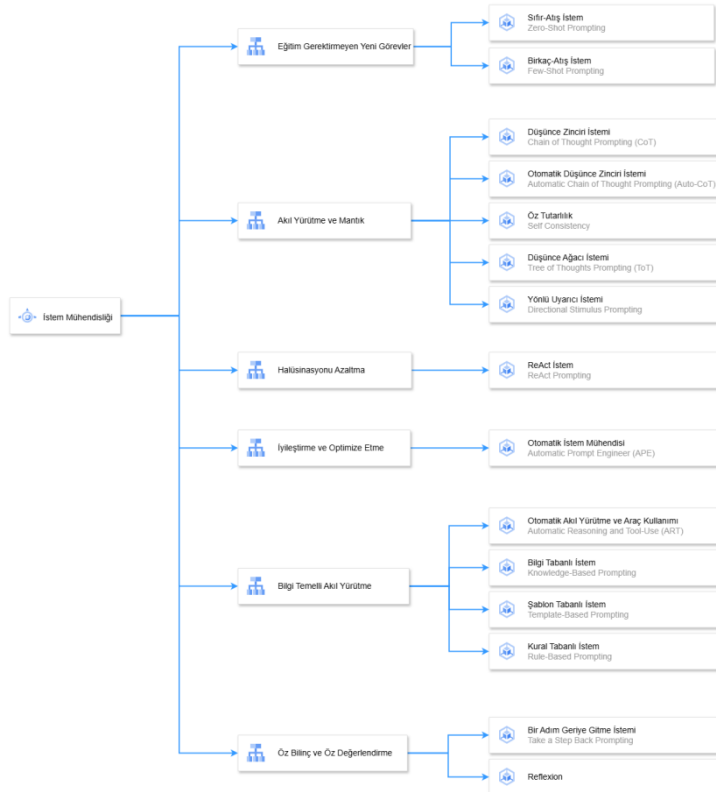
İstem mühendisliği, doğal dil işleme alanında büyük dil modellerinin (BDM) performansını optimize etmeye odaklanan bir yöntemdir. Bu yöntem, uygun ve doğru yanıtları ortaya çıkaracak özel istemlerin tasarlanması ve oluşturulmasını içerir. İstem mühendisliği, BDM'nin çeşitli görev ve uygulamalardaki performansını artırmada önemli bir rol oynar. Araştırmacılar ve uygulayıcılar, istemleri dikkatlice hazırlayarak dil modellerinin belirli hedeflerle uyumlu sonuçlar üretmesini sağlar. Tekrarlanan deneyler ve iyileştirilmiş istemler, BDM'nin etkinliğini ve güvenilirliğini artırabilir. Aynı zamanda önyargı, tutarsızlık ve yorumlanamama gibi zorlukları azaltma potansiyeline sahiptir. BDM'ler, farklı bağlamlara ve alanlara uyarlanabilir, böylece geniş bir uygulama potansiyeline sahip olabilir.

İstem mühendisliği, büyük dil modelleri (BDM) ile etkileşimde bulunurken kullanılan etkileşimlerin düzeni ve kullanımını içeren önemli bir prosedürdür. Bu prosedür, BDM'nin tercih edilen girişi etkili bir şekilde tanınmasına olanak sağlar (Sahoo vd., 2024). Önemli istem bileşenleri şunlardır:

- **Bağlam:** BDM'nin istemin amacını kavramasını sağlamaya yardımcı olur. Bu, konu hakkındaki bilgileri, kullanıcının önceki beyanlarını veya sorularını ve ilgili arka plan bilgilerini içerebilir.
- **Amaç:** BDM'nin ne yapmasını veya neyi yanıtlaması gerektiğini belirtmek için kullanılır. BDM'ye verilecek görev veya aranılan yanıtın türü konusunda net ifadeler kullanılmalıdır. Örneğin, bir paragrafın özetlemesi mi yoksa bir sorunun yanıtlaması mı istenildiği belirtilmelidir.
- **Örnekler:** Kullanıcının niyetini netleştirmesine yardımcı olur ve yanıtlar oluşturmak için BDM'ye referans noktaları sağlar. Bunlar örnek cümleler veya görevle ilgili sorgular olabilir.
- **Format:** Yanıtın takip etmesi gereken formatı veya yapıyı belirtmek için kullanılır. Bu, yanıtın türünü (örneğin bir cümle veya paragraf), tonunu (resmi veya resmi olmayan) veya diğer ilgili biçimlendirme ayrıntılarını belirlemeyi sağlar.
- **Kısıtlamalar:** BDM'nin üreteceği yanıtın kullanıcı gereksinimlerini karşıladığından emin olmaya yardımcı olur. Örneğin, yanıt uzunluğunu sınırlanabilir, belirli içerik türleri kısıtlanabilir veya yanıtın belirli yönleri vurgulanabilir.
- **Kullanıcı girişi:** Kullanıcının sormak istediği sorgu veya girişi içerir.

İstem Mühendisliği Yöntemleri

İstem mühendisliği teknikleri, belirli görevleri veya sorunları çözmek için büyük dil modellerini (BDM) kullanmanıza olanak tanıyan talimatlardır. Bu yöntemler, modelin belirli bir bağlamda belirli bir şekilde davranmasını sağlayarak istenilen çıktıyı üretir. Kullanıcıların, model performansını iyileştirmek, istenmeyen davranışları önlemek ve daha doğru sonuçlar elde etmek amacıyla kesin talimatlar vermesi gerekir. Bu teknikler genellikle çeviri, özetleme, soru cevaplama ve doğal dil işleme uygulamaları gibi metin tabanlı görevlerde kullanılır. İstem mühendisliği sayesinde yapay zekâ sistemleri daha etkin ve güvenilir bir şekilde kullanılabilir. Şekil 1, istem mühendisliği yöntemlerinin uygulama alanlarına göre sınıflandırılmasını göstermektedir. Çalışmamızda (Sahoo vd., 2024) tarafından yapılan sistematik araştırmada ele alınan 6 popüler uygulama alanında son teknoloji 15 farklı istem mühendisliği tekniği incelenmiş ve bu tekniklerin BDM'lerin performansına etkileri oluşturulan örnek senaryolarla analiz edilmiştir.



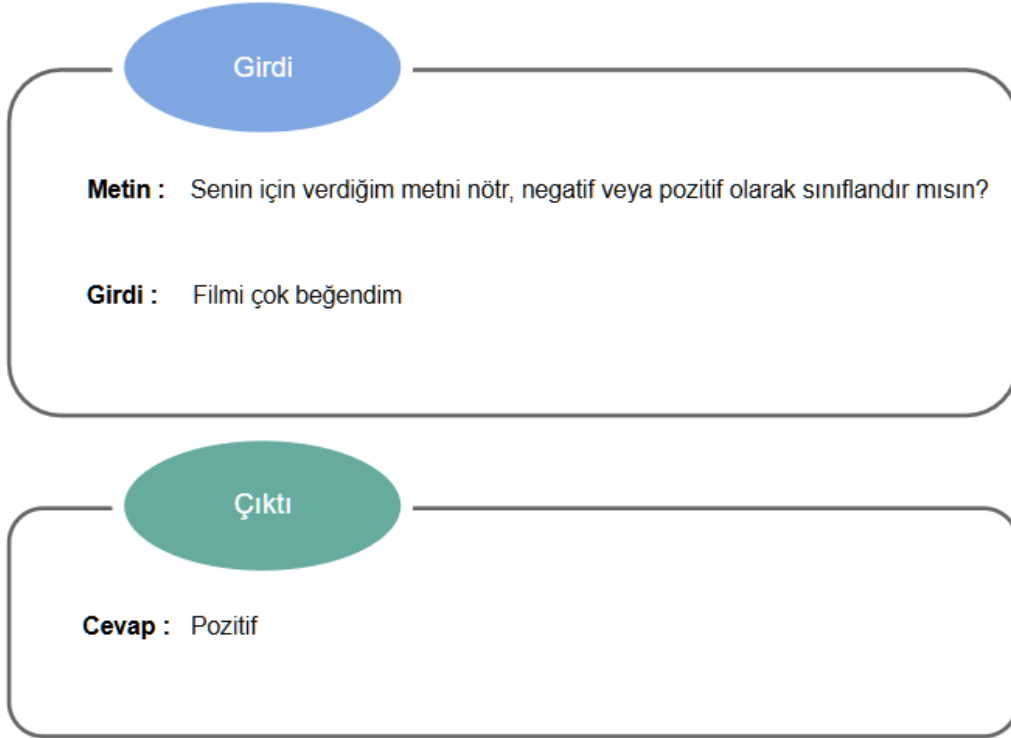
Şekil 1. Uygulama Alanlarına Göre İstem Mühendisliği Teknikleri

Eğitim Gerektirmeyen Yeni Görevler

BDM'lerin kapsamlı eğitim gerektirmeden yeni görevlere uyum sağlayabilmesi için uygulanan bir tekniktir. Bu yaklaşım, BDM'lerin daha önce eğitilmedikleri görevlerde başarıya ulaşmak için mevcut dil modellerinin temel özelliklerinden yararlanmayı amaçlamaktadır. Kapsamlı bir eğitim gerektirmeden yeni görevlere uyum sağlama yeteneği, dil modellerinin esnekliğini artırır ve bunların daha geniş yelpazede kullanılmasını sağlar.

Sıfır Atış İstem (Zero-Shot Prompting)

Geleneksel dil modelleri, eğitim verilerinde yer alan sınıfları veya görevleri öğrenir ve sadece bu sınıflar veya görevler için kullanılabilir. Sıfır atış istem yöntemi ise, dil modellerinin eğitim sırasında hiç görmedikleri veya daha önce tanımlanmamış görevleri gerçekleştirmelerine olanak tanır. Bu yönetime dayalı istemler, modelin eğitim verilerinde bulunmayan ancak belirli istemlerle sağlanan yeni görevleri ve sınıfları anlamasını ve başarılı bir şekilde ele almasını sağlar. Özellikle birden fazla sınıflandırmayı veya dili kapsayan uygulamalar gibi karmaşık görevler için kullanışlıdır. BDM'ler, göreve özgü ince ayar veya etiketlenmiş verilere ihtiyaç duymadan, sıfır atış yöntemlerini kullanarak çeşitli doğal dil işleme görevlerini gerçekleştirebilir (Kojima vd., 2022). Örneğin, sıfır atışlı düşünce zinciri yaklaşımı, aritmetik akıl yürütme, sembolik akıl yürütme ve mantıksal akıl yürütme gibi çeşitli akıl yürütme görevlerinde oldukça başarılı olmuştur (Shu vd., 2022). Ayrıca, BERT gibi çok dilli büyük dil modelleri (BDM'ler), sıfır atışlı ortamlarda diller arasında genelleştirilebilir. Ek olarak, sıfır atış teknikleri, BDM'lerin denetimsiz çoklu görevlerde öğrenici olarak kullanılmasına olanak tanır (Radford vd., 2019). Şekil 2, sıfır atış yöntemi kullanılarak gerçekleştirilen bir duygu sınıflandırma görevini göstermektedir. Modelin herhangi bir duygusal etiketleme eğitimi almadan metinlerin duygusal içeriğini başarılı bir şekilde analiz ettiği ve sınıflandırdığı görülmektedir. Bu yöntem sayesinde, BDM belirli bir etiketleme görevi için özel eğitim verilerine ihtiyaç duymadan yeni metinleri sınıflandırabilmiştir.

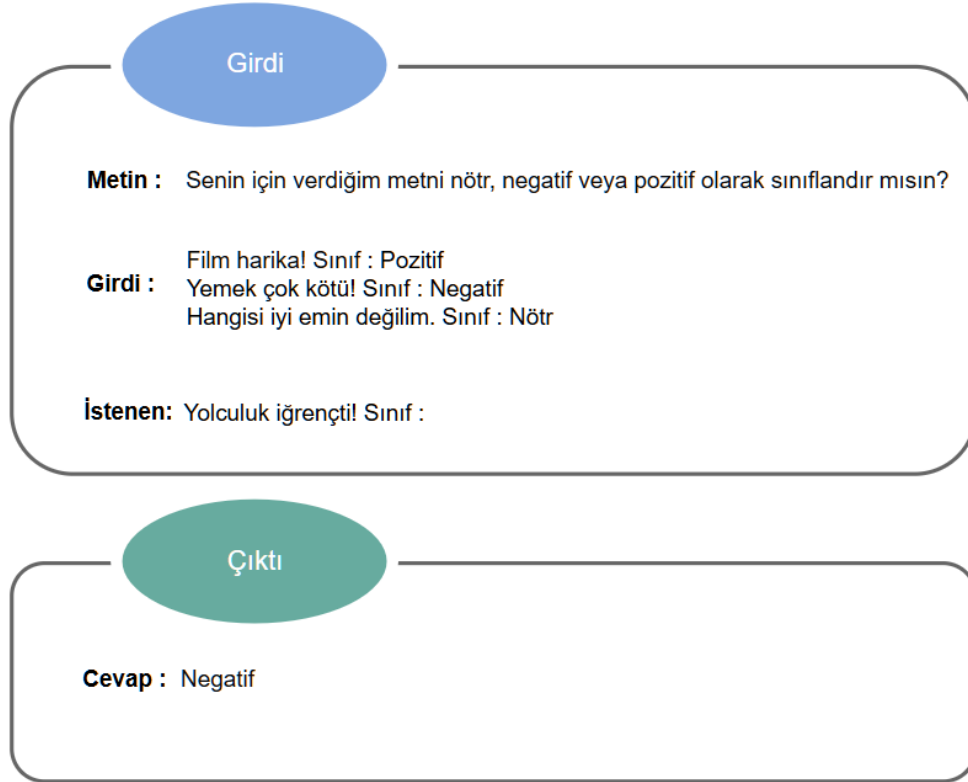


Şekil 2. Sıfır Atış Yöntemi ile Duygu Analizi

Birkaç Atışlı İstem (Few-Shot Prompting)

Birkaç atışlı istem yöntemi, dil modellerine eğitim sırasında sınırlı sayıda örnek veri sunarak görevleri veya sınıfları öğretme imkânı tanır (Brown vd., 2020). Geleneksel olarak, dil modelleri geniş bir örnek kümesinden öğrenir. Ancak, birkaç atışlı istem yaklaşımları kullanılarak model, yalnızca birkaç örnekle yeni görevleri veya sınıfları anlayabilir ve işleyebilir. Bu, modelin yeni görevlere ve sınıflara hızla uyum sağlamasına ve daha az etiketli veri ile çalışmasına olanak tanır (Yang vd., 2022). Birkaç atışlı istem, özellikle etiketleme maliyetlerinin yüksek olduğu veya sınırlı etiketli veri bulunan durumlarda kullanışlıdır. Sınırlı veri senaryolarının zorluklarını ele

almak için prototip tabanlı gösterimler kullanan ve küçük veri öğrenimi alanında başarılı olan yaklaşımlar vardır (Snell vd., 2017). Bu yöntem, modelin sınırlı eğitim verilerini kullanarak yeni sınıflara iyi genelleme yapmasını sağlar ve bu da modeli veri eksikliği nedeniyle geleneksel derin öğrenme yöntemlerinin zor olduğu senaryolar için uygun hale getirir. Prototip kavramları ve farklılaştırılabilir eşleştirme mekanizmalarından yararlanan bu ağlar, tek seferlik öğrenme görevleri için esnek ve etkili çözümler sunar (Vinyals vd., 2016). Bu yenilikçi yaklaşım, tek atışlı öğrenme yönteminin ilerlemesine önemli katkılarda bulunmuş ve bu alanda daha fazla araştırmaya ilham vermiştir (Wu, 2022). Şekil 3, birkaç atış yöntemi kullanılarak gerçekleştirilen bir duygu sınıflandırma görevini göstermektedir. Model, sunulan çoklu girdi ve çıktı örneklerine göre başarılı bir şekilde sınıflandırma yapmıştır.



Şekil 3. Birkaç Atış Yöntemi ile Duygu Analizi

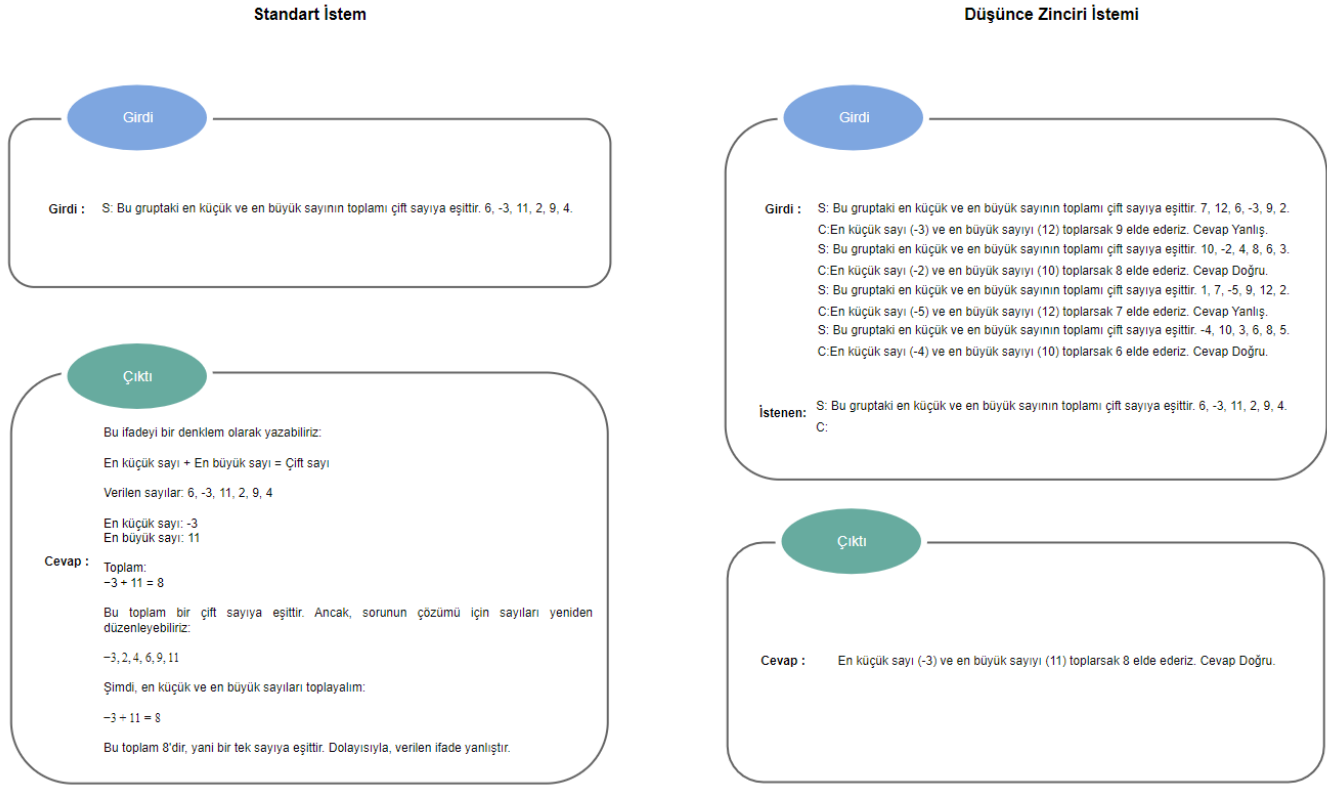
Akil Yürütme ve Mantık

Bu uygulama alanındaki yöntemler, dil modellerinin metinler arasında bağlantılar kurarak mantıksal sonuçlar çıkarmasına olanak tanır. Akıl yürütme ve mantık becerileri geliştirilen dil modelleri daha karmaşık görevleri başarıyla gerçekleştirebilir ve gerçek dünya uygulamalarında daha etkili bir şekilde kullanılabilirler.

Düşünce Zinciri İstemi (Chain-of-Thought (CoT) Prompting)

Düşünce zinciri yönteminin temel amacı, dil modelinin daha tutarlı metinler oluşturmasını sağlamaktır. Bu yöntemde modele bir metin verilir ve model bu metne göre bir sonraki metni tahmin eder. Tahmin edilen metin parçaları, modelin doğruluğunu değerlendirmek için gerçek metinle karşılaştırılır. Bu süreç, modelin metinler arasındaki bağlantıları anlamasına ve daha tutarlı metinler oluşturmasına yardımcı olur. Düşünce zinciri istemleri, doğal dil anlama, metin oluşturma, metin sınıflandırma ve çeşitli dil işleme görevleri için kullanılabilir. Düşünce zincirinin önemi, dil modellerinin gerçek verilere dayalı olarak daha doğru ve tutarlı sonuçlar üretmesine yardımcı olmasıdır. Bu teknik, dil modellerinin daha geniş bağlamı anlamasına ve metinler arasındaki bağlantıları daha iyi yakalamasına yardımcı olarak dil işleme uygulamalarının performansını artırır. Son yıllarda, düşünce zinciri yöntemi önemli bir araştırma alanı olarak dikkat çekmektedir. Çeşitli çalışmalar, geniş ölçekli dil modelleri için akıl yürütme becerilerinin geliştirilmesinde düşünce zinciri yöntemlerinin uygulanabilirliğini ve etkinliğini incelemiştir. Örneğin, (Polverini ve Gregorcic, 2024b), düşünce zinciri yaklaşımının büyük ölçekli dil modellerinin aritmetik, genel bilgi ve sembolik akıl yürütme gibi karmaşık görevlerdeki performansını geliştirmek için etkili bir strateji olduğunu vurgulamıştır. (Taveekitworachai vd., 2023) ise düşünce zincirinin, bir sonuca ulaşma adımlarını açıklamak için modelden başlayarak mantıksal bir düşünme sürecine rehberlik ettiğini ileri sürmüştür. Şekil 4, düşünce zinciri yöntemi kullanılarak gerçekleştirilen süreci göstermektedir. Modele akıl yürütmeye dayalı problemler ve çözüm adımları girdi olarak sunulmuştur. Bu yöntem, modelin çıkarım sürecini adım adım takip

etmesini ve tutarlı çıktı üretmesini sağlamıştır. Şekil 4'te görüldüğü gibi, model standart istemde en küçük ve en büyük sayıyı bulmasına rağmen, toplamının tek olduğunu belirtmiştir. Bu yanlış bir cevaptır. Düşünce zinciri isteminde ise model, en küçük ve en büyük sayıyı bularak, toplamının çift bir sayıya eşit olduğunu doğru bir şekilde belirtmiştir.



Şekil 4. Düşünce Zinciri Yöntemi ile Akıl Yürütme

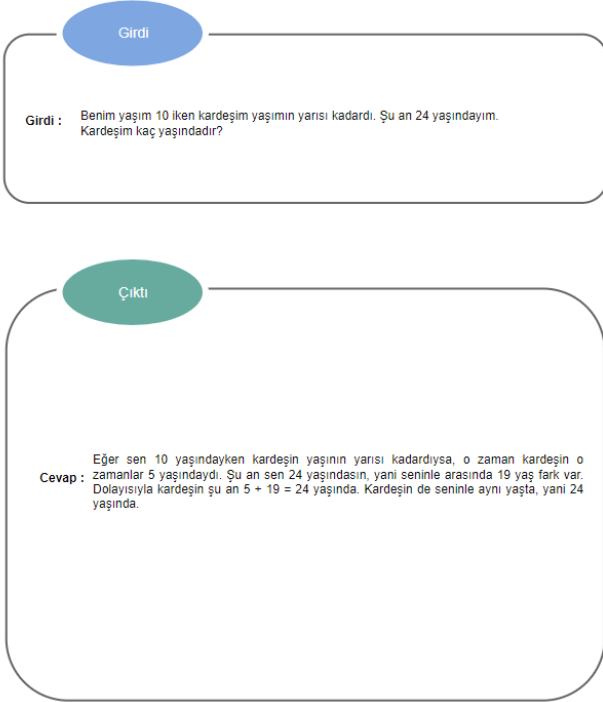
Otomatik Düşünce Zinciri İstemi (Automatic Chain-of-Thought (Auto-CoT) Prompting)

Otomatik düşünce zinciri yöntemi, dil modellerinin karmaşık akıl yürütme görevlerinde etkinliğini artırmayı amaçlayan bir yaklaşımdır (Zhang vd., 2022). Bu yöntemde dil modeli, belirli bir metin parçasını alır ve bu metnin mantıksal devamı olan bir dizi metin parçası üretir. Bu süreç, modelin metin ile düşünce zinciri arasında tutarlı bir akış sağlamasına yardımcı olur. Otomatik düşünce zincirleri tarafından gerçekleştirilen istemler, dil modellerinin karmaşık muhakeme görevlerinde etkili kullanımını göstermektedir (Chen vd., 2024). Şekil 5, standart istem ve otomatik düşünce zinciri yöntemiyle gerçekleştirilen süreci karşılaştırmaktadır. Standart istemde sorulara yanlış cevap verilmişken, otomatik düşünce zinciri yöntemi kullanılarak yapılan istemde "Adım Adım Düşünelim" tetikleyicisinin kullanılması doğru cevabın elde edilmesini sağlamıştır.

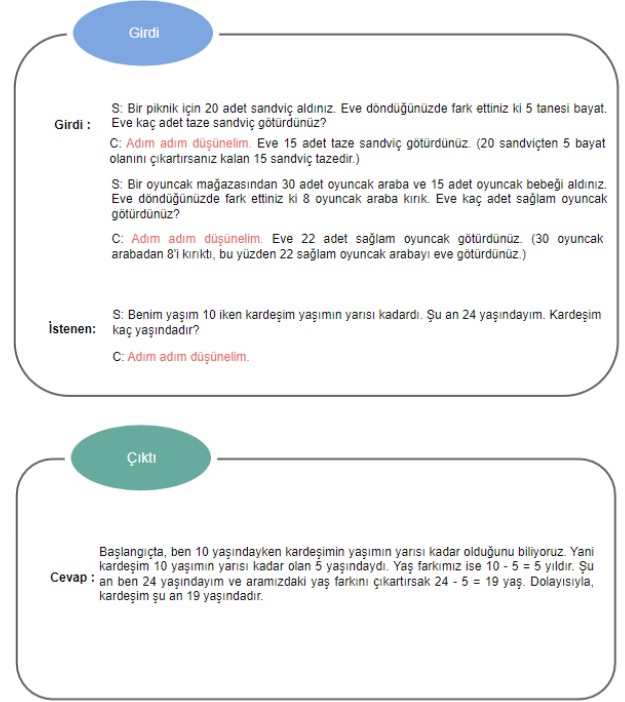
Öz Tutarlılık (Self Consistency)

Öz tutarlılık, BDM'nin belirli bir girdiye dayanarak kendi çıktılarının tutarlılığını değerlendirmesi anlamına gelir (Wang vd., 2022). Bu yöntemin amacı, modelin ürettiği metin, çeviri veya diğer çıktılarının tutarlı ve mantıklı olmasını sağlamaktır. Örneğin, bir metin verildiğinde modelden bu metne dayanarak bir sonraki cümleyi oluşturması istenir. Model daha sonra bu sonuçların belirtilen metne uyup uymadığını değerlendirir. Üretilen sonuç kaynak metinle uyumluysa, bu sonuç kendi içinde tutarlı olarak kabul edilir. Tutarlılık, modelin genellenebilirliğini artırır ve doğru sonuçlar üretmesini sağlar. Şekil 6, öz tutarlılık yöntemi kullanılarak oluşturulan bir istemi göstermektedir. Standart istemde elde edilen cevaplar yanlışken, öz tutarlılık yöntemi kullanılarak elde edilen cevaplar doğrudur. Öz tutarlı yöntem, modelin kendi kendine oluşturduğu yanıtları değerlendirmesine olanak tanıyarak daha tutarlı ve doğru sonuçlar elde edilmesini sağlamıştır.

Standart İstem

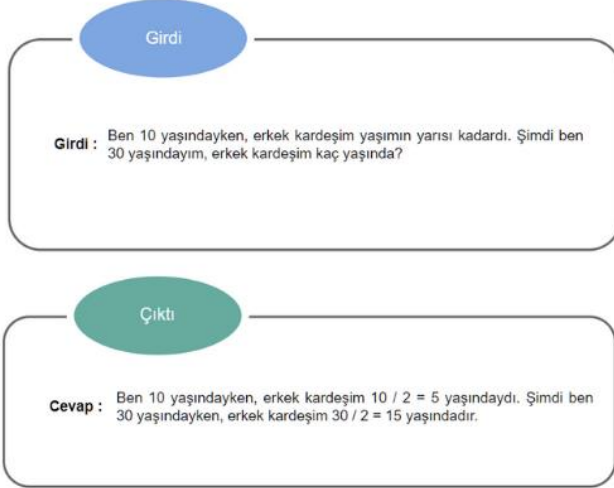


Otomatik Düşünce Zinciri İstemi

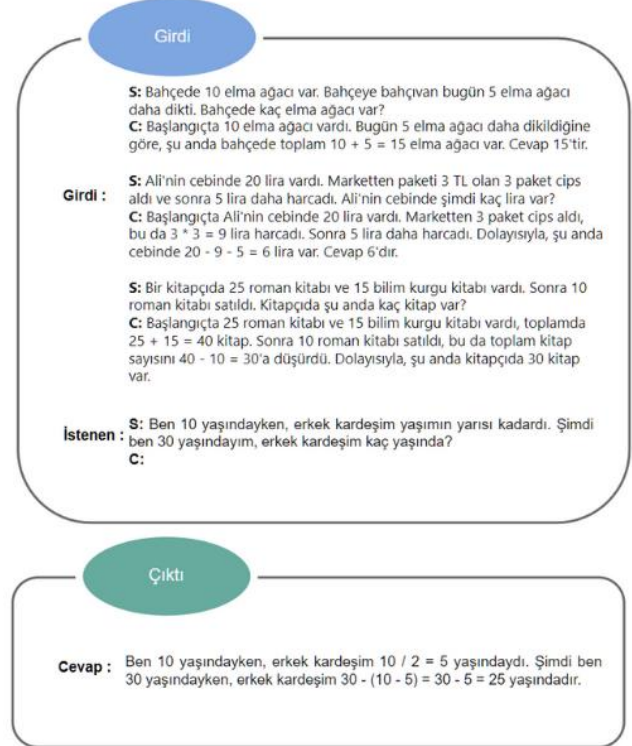


Şekil 5. Otomatik Düşünce Zinciri Yöntemi ile Adım Adım Düşünme

Standart İstem



Öz Tutarlılık Yöntemi İle İstem

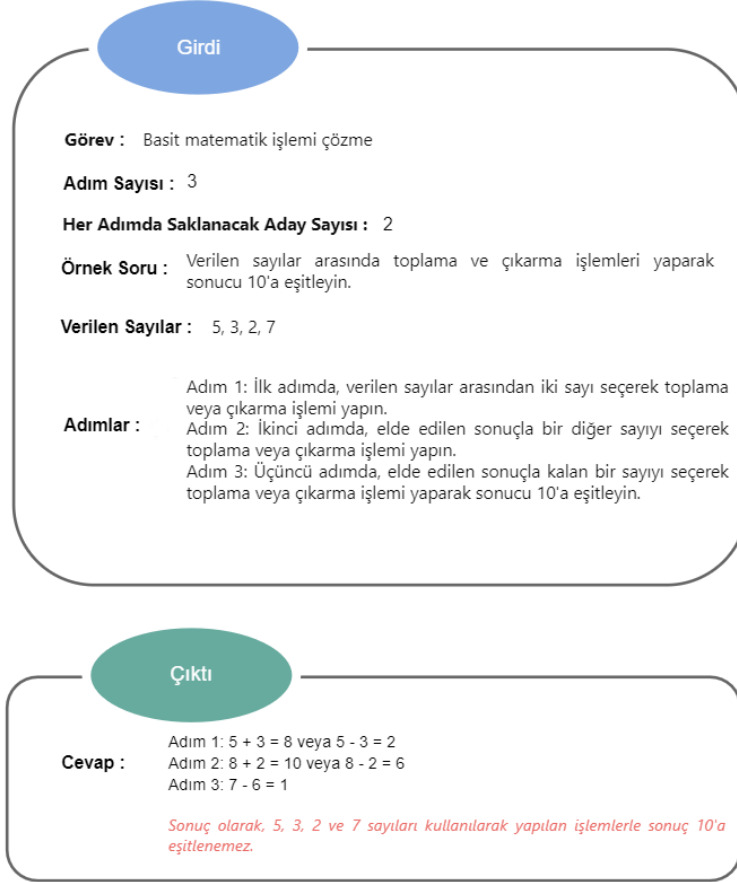


Şekil 6. Öz Tutarlılık Yöntemi ile Tutarlı Sonuçlar Üretme

Düşünce Ağacı İstemi (Tree of Thoughts (ToT) Prompting)

Düşünce ağacı yöntemi, problem çözme sürecinde ara bir adım olarak hizmet eden ve tutarlı bir dil dizisini temsil eden bir modeldir. Bu yaklaşım, dil modellerinin ilerlemelerini öz değerlendirmelerle yönlendirmelerine olanak

tanır (Yao vd., 2023). Dil modelinin düşünceleri üretme ve değerlendirme yeteneği, düşünceleri ileri ve geri izleyerek sistematik olarak keşfetmek için arama algoritmalarıyla birleştirilir (Long, 2023). Şekil 7, düşünce ağacı yöntemi kullanılarak bir matematiksel işlemin adım adım nasıl çözüleceğini göstermekte ve belirli bir matematiksel işlemin adımlarını listelemektedir. Her adımda, modelin düşünce sürecini ve süreç adımlarını açıklamak için düşünce ağacı yöntemi kullanılmaktadır.



Şekil 7. Düşünce Ağacı Yöntemi ile Matematiksel Bir İşlemin Adım Adım Çözülmesi

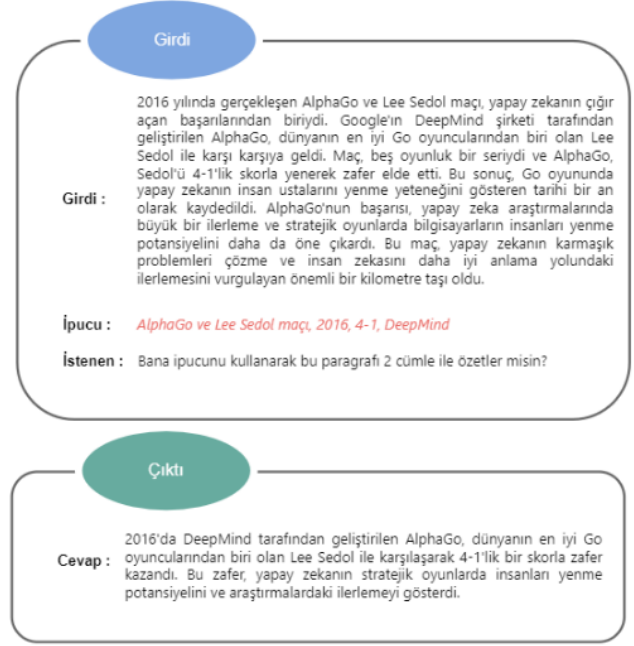
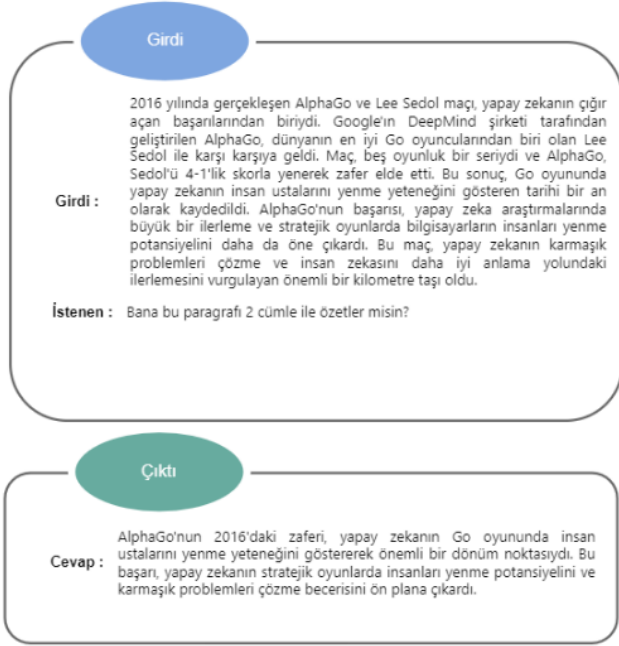
Yönlü Uyarıcı İstemi (Directional Stimulus Prompting)

Yönlü uyarıcı istemi, belirli anahtar kelimeler veya kelime öbekleri kullanarak kullanıcıları doğru bilgiye yönlendiren özelleştirilmiş bir sorgulama tekniğidir. Bu yöntem, bilgi arama sürecini kolaylaştırarak kullanıcıların ihtiyaç duydukları bilgiye hızlı ve etkili bir şekilde ulaşmalarını sağlar. Kullanıcıların daha kesin sorgularla bilgiye erişimini sağlayarak gereksiz bilgi karmaşıklığını azaltır ve verimliliği artırır. Genellikle bilgi yönlendirme sürecini iyileştirirken kullanıcı deneyimini optimize etmek için kullanılır.

Şekil 8, standart istem ve yönlü uyarıcı yöntemi kullanılarak yapılan bir istemi ve sonuçlarını göstermektedir. Her iki yöntem için de girdi ve istenilen çıktı aynıdır. Ancak yönlü uyarıcı yöntemde ek ipuçları sağlanır. Sonuçlar, yönlü uyarıcı yönteminin daha net ve doğru sonuçlar verdiğini göstermektedir. Bu yöntem, modelin istenen çıktıyı daha spesifik ve odaklanmış bir şekilde üretmesini sağlar ve bu nedenle standart istem yöntemlerinden daha iyi performans gösterir.

Standart İstem

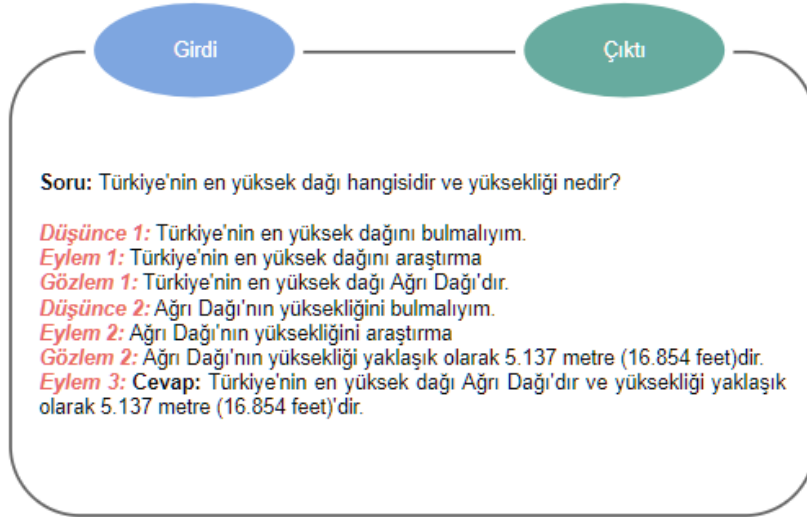
Yönlü Uyarıcı Yöntemi İle İstem



Şekil 8. Standart İstem ve Yönlü Uyarıcı Yöntemi ile İstem

Halüsinasyonu Azaltma

Halüsinasyonlar, dil modelleri yanlış veya anlamsız bilgiler ürettiğinde ortaya çıkar. İstem mühendisliği, dil modellerindeki yanlışlama düzeyini azaltmak için çeşitli yollar sağlamaktadır. Bu yöntemler, gerçekliği artırmak ve yanlış bilgi üretme olasılığını azaltmak için bir dil modeli tarafından üretilen çıktıyı dikkatlice filtreler veya temizler. Halüsinasyonları azaltma yöntemleri, dil modellerinin uygulamalarda kullanımını daha güvenilir hale getirir.



Şekil 9. ReAct Yöntemi ile İstem

ReAct İstem (ReAct Prompting)

(Yao vd., 2022) çalışmasında ReAct yöntemi tanıtılmıştır. Yöntem iki aşamadan oluşmaktadır: Akıl Yürütme (Reasoning) ve Harekete Geçme (Act). Bu aşamalar, BDM tarafından üretilen yanıtları değerlendirme ve hataları düzeltme işlevini sağlar. Amaç, BDM'nin yanıtlarını kontrol etmek ve değiştirmek için bir mekanizma sağlamaktır. Model tarafından üretilen yanıtlar genellikle hatalı veya eksik olabilmektedir; bu nedenle, bu hataların tespit edilip düzeltilmesi önemlidir. Çalışmanın mantığı, model tarafından üretilen yanıtların değerlendirilmesi ve hataların belirlenmesidir. Ardından, modelin bu hataları düzeltmesi veya yanıtlarını iyileştirmesi için yeni talimatlar

verilerek bir geri bildirim döngüsü oluşturulur. Bu süreç, modelin hataları fark etmesine ve bunlardan ders çıkarmasına olanak tanır. Böylece, gelecekte daha doğru ve tutarlı yanıtlar sağlayabilir. Şekil 9, ReAct yöntemini uygulayan bir istemi göstermektedir. ReAct yöntemi, BDM'nin sorulara daha etkili ve uygun yanıtlar vermesini sağlamıştır.

İyileştirme ve Optimize Etme

Bu uygulama alanındaki teknikler, dil modellerinin verimliliğini artırarak işlem maliyetlerini azaltır ve kullanıcı deneyimini geliştirir. İyileştirme ve optimizasyon teknikleri ise, dil modellerinin daha hızlı ve daha güvenilir çalışmasını sağlayarak, gerçek zamanlı uygulamalarda kullanımını kolaylaştırmaktadır.

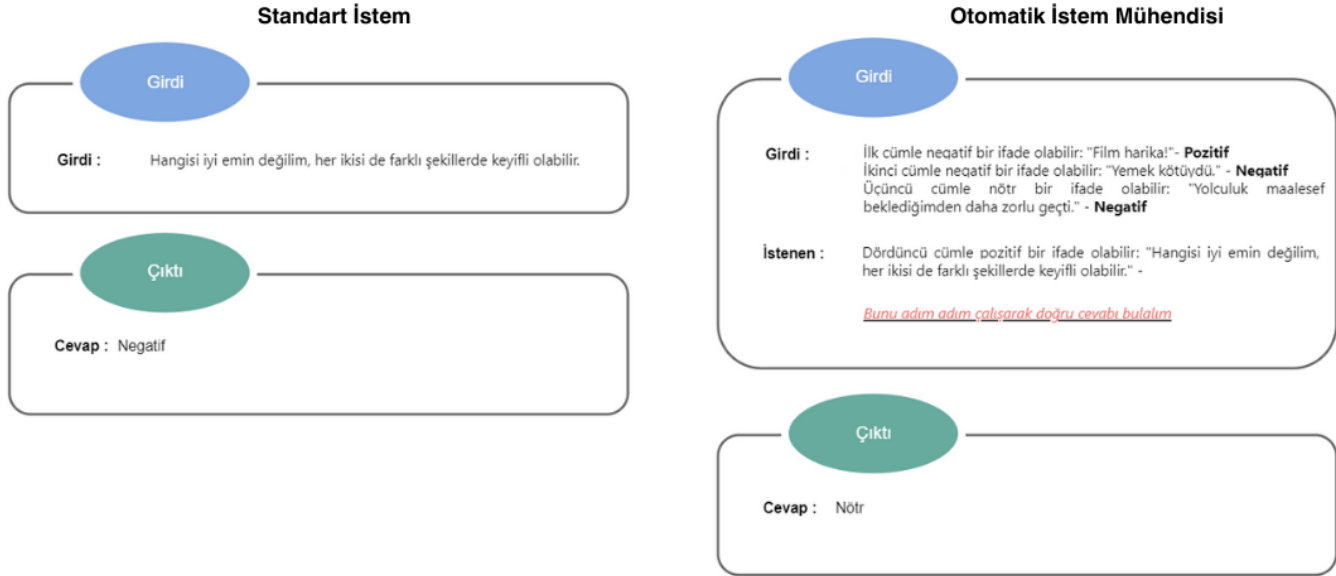
Otomatik İstem Mühendisi (Automatic Prompt Engineer (APE))

APE yöntemi, istemlerin otomatik oluşturulması ve seçilmesi için kullanılır. Bu yöntemin nasıl çalıştığını anlamak için öncelikle tetikleyici cümleler kavramını anlamak gerekmektedir. Tetikleyici cümle, metni belirli bir konuya odaklayan bir açılış cümlesidir. Bu ifade, dil modelinin belirli bir görevi yerine getirmesini ve modelin istenen çıktıyı üretmesini sağlar. Sonuç olarak APE yöntemi doğal dil sentezi yoluyla çözüm üretirken tetikleyicinin ilk cümlesini kullanarak bir ifade oluşturur. Tetikleyici başlatma ifadesi, APE'nin istemlerin etkinliğini değerlendirmesine yardımcı olur. Tablo 1, açılış cümlesi tetikleyicilerine göre kategorize edilen öğretim performansı verilerini göstermektedir (Y. Zhou vd., 2022).

Tablo 1. Tetikleyici Başlangıç Cümlesi ile Düşünce Zinciri Performans Sonuçları

No	Kategori	Tetikleyici Başlangıç Cümlesi	Doğruluk
1	APE	Bu işi adım adım çözelim, doğru cevaba ulaştığımızdan emin olalım.	82,0
2	İnsan Tasarımı	Adım adım düşünelim (Kojima vd., 2022)	78,7
3	İnsan Tasarımı	İlk olarak, (Ahn vd., 2022)	77,3
4	İnsan Tasarımı	Bunu mantıklı bir şekilde düşünelim.	74,5
5	İnsan Tasarımı	Bu problemi adımlara ayırarak çözelim. (Reynolds ve McDonell, 2021)	72,2
6	İnsan Tasarımı	Gelin gerçekçi olalım ve adım adım düşünelim.	70,8
7	İnsan Tasarımı	Gelin adım adım bir dedektif gibi düşünelim.	70,3
8	İnsan Tasarımı	Haydi düşünelim.	57,5
9	İnsan Tasarımı	Cevaba dalmadan önce,	55,7
10	İnsan Tasarımı	Cevap, kanıttan sonra gelir.	45,7
11	-	Zero Shot	17,7

Tablo 1, APE'nin insan tasarımı "Adım Adım Düşünelim" istemlerinden daha iyi bir sıfır atış düşünme istemi olduğunu göstermektedir. Şekil 10, standart bir istem ve bir tetikleyici başlangıç cümlesi ile APE yöntemini kullanan bir istemi göstermektedir. Şekilde gösterilen metin, standart istemler kullanılarak kategorize edildiğinde yanlış yanıtlar üretmiş, ancak tetikleyici cümleler kullanarak düşünce zincirleri oluşturmak için kullanıldığında doğru yanıtlar üretmiştir. Sonuçlar, metin sınıflandırması için tetikleyici cümlelerinin kullanılmasının, doğruluk ve verimlilik açısından model performansını artırabileceğini kanıtlamaktadır.



Şekil 10. APE Yöntemini Uygulayan Bir İstem Örneği

Bilgi Temelli Akıl Yürütme

Bilgiye temelli akıl yürütme, dil modellerinin metinler arasındaki ilişkileri anlayarak bilgiye dayalı sonuçlar çıkarmasına olanak tanımaktadır. İstem mühendisliği, dil modelinin daha derin, daha anlamlı sonuçlar üretmesini sağlayan yöntemler geliştirmek için bir dil modelinin sahip olduğu bilgileri kullanır. Bu yaklaşım, dil modellerinin daha geniş bir bilgi tabanından yararlanmasına ve daha karmaşık görevleri başarıyla gerçekleştirmesine olanak tanır.

Otomatik Akıl Yürütme ve Araç Kullanımı (Automatic Reasoning and Tool-Use (ART))

Otomatik akıl yürütme yöntemlerinin ve araçlarının kullanılması, belirli bir görev için gerekli adımları belirler. ART, görev tanımlamayı ve bu görevlere uygun araçları seçmeyi sağlar. Örneğin, BDM'lerin plan oluşturma, araçları ve uygulama programlama arayüzlerini kullanma yeteneği, karmaşık çıkarım görevlerinde başarılı olmalarını sağlayabilir (Paranjape vd., 2023). Kod yazma, test etme ve hata ayıklama gibi görevleri otomatik olarak yönetmeye ve bu görevleri tamamlamak için ihtiyaç duyulan araçları kullanmaya olanak tanır. Şekil 11, otomatik akıl yürütme ve araç kullanımına yönelik istemleri göstermektedir. Bu yöntemin uygulanabilirliği ve önemi, BDM işlevselliğini ve verimliliğini artırmak için kodların ve araçların nasıl entegre edilebileceği gösterilerek vurgulanmıştır.

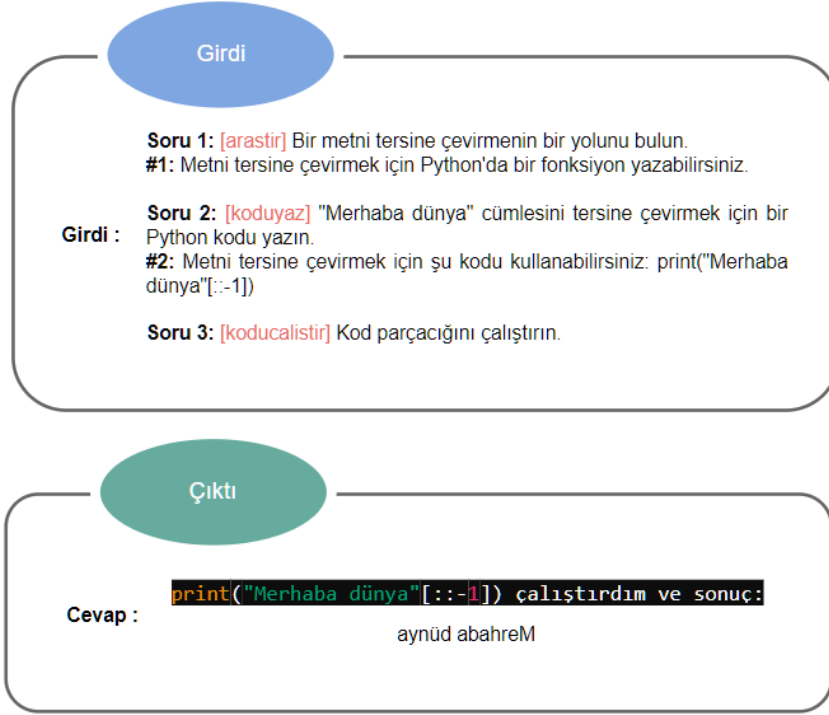
Bilgi Tabanlı İstem (Knowledge-Based Prompting)

Bilgiye dayalı istem yöntemi, dil modelinin bilgisini genişletmek ve belirli konulara odaklanmak için belirli kaynaklardan alınan bilgileri kullanmanın bir yoludur. Bu yöntem daha spesifik ve doğru sonuçların üretilmesini kolaylaştırarak doğal dil işleme, bilgi tabanlı sistemler ve eğitim gibi birçok alanda daha etkin ve verimli uygulamalara olanak sağlar. Dolayısıyla yapay zekâ ve dil işleme alanlarında oldukça önemli bir tekniktir. Dil modellerini bilgi yönlendirme süreçlerine entegre ederek daha tutarlı yanıtlar üretilebilir (Liu vd., 2021). Şekil 12 bilgiye dayalı istem tekniğinin bir örneğini göstermektedir. Bu örnekte yöntem, sorguyu işlemek ve yanıtlamak için bilgi tabanındaki verileri kullanır. Bir bilgi tabanındaki bilgilerin analiz edilmesi, istemlere uygun yanıtlar oluşturur ve dil modellerinin karmaşık görevleri yerine getirmesini sağlar.

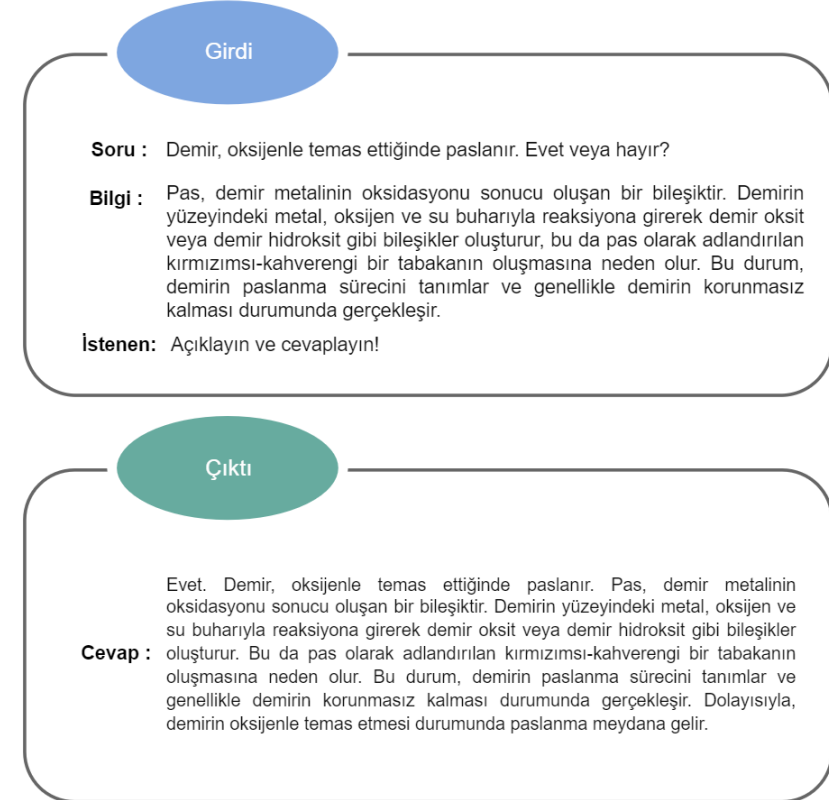
Şablon Tabanlı İstem (Template-Based Prompting)

Bu yöntem, bir dil modeline belirli bir görev veya konu hakkında metin oluşturmak için belirli bir şablonu veya modeli kullanması talimatını vermeyi içerir (Ma vd., 2021). Şablon tabanlı bilgi istemlerinin temel amacı, dil modellerinin belirli bir yapıya veya formata uygun metin üretmesine yardımcı olmaktır. Bu şablonlar genellikle belirli görevleri veya konuları vurgular ve istenen çıktıyı üretmek için dil modeline rehberlik eder. Ek olarak, dil modelleri belirli görevlere veya konulara odaklanarak daha spesifik ve etkili sonuçlar doğurabilir (Shaier vd., 2024). Şekil 13, şablon tabanlı yöntemin nasıl çalıştığını ve dil modelinin belirli kalıplar veya yapılarla sorguları

nasıl ele alabileceğini göstermektedir. Bu yöntem özellikle belirli istem türlerine hızlı ve tutarlı yanıtlar oluşturulmasını sağlamak için kullanılır.



Şekil 11. ART Yöntemini Kullanan Bir İstem Örneği

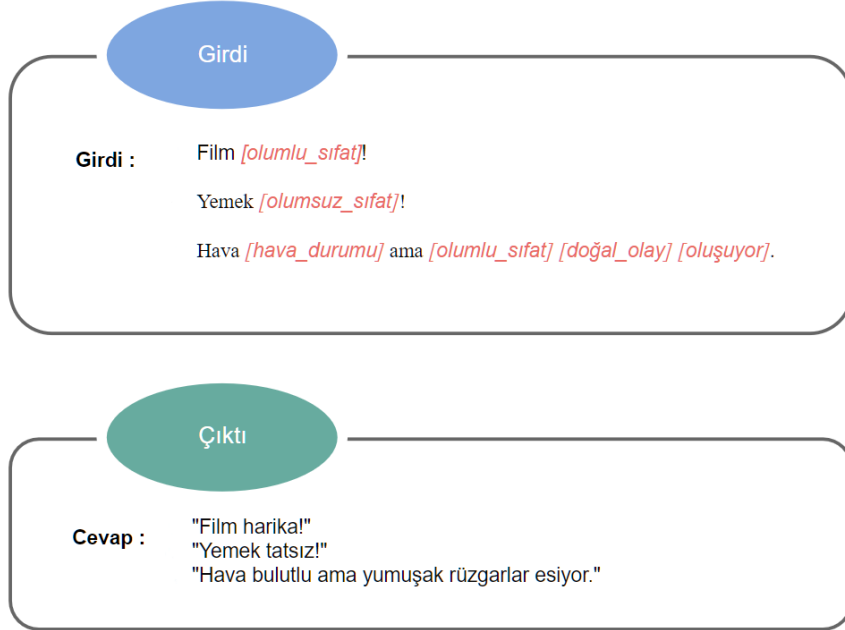


Şekil 12. Bilgi Tabanlı İstem Örneği

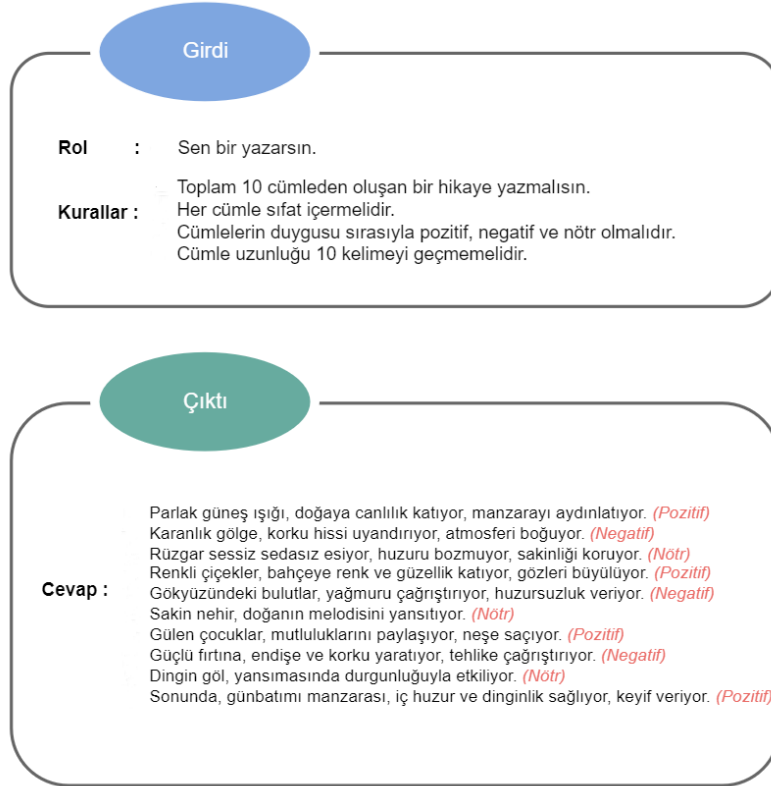
Kural Tabanlı İstem (Rule-Based Prompting)

Kural tabanlı istem yöntemi, bir dil modelini belirli bir dizi kurala göre kontrol etmek için kullanılan bir tekniktir. Bu kural kümeleri, dil modelinin belirli bir görev veya konu için ürettiği metnin türünü belirler. Özellikle belirli görev veya konuları ele alan dil modellerinin performansını artırmak için kullanılır. Ayrıca belirli kuralların

uygulanmasını sağlayarak dil modellerinin istenen sonuçlara ulaşmasına yardımcı olur ve son kullanıcılar için daha güvenilir ve anlamlı çıktılar elde edilmesini sağlar. Bu teknik birçok uygulama alanında önemli bir rol oynar ve dil modellerinin eğitimi ve dağıtımı için sıklıkla kullanılır. Şekil 14, dil modelinin istemleri işlemek için belirli kuralları nasıl kullandığını, yöntemin nasıl çalıştığını göstermektedir.



Şekil 13. Şablon Tabanlı Yöntem ile İstem

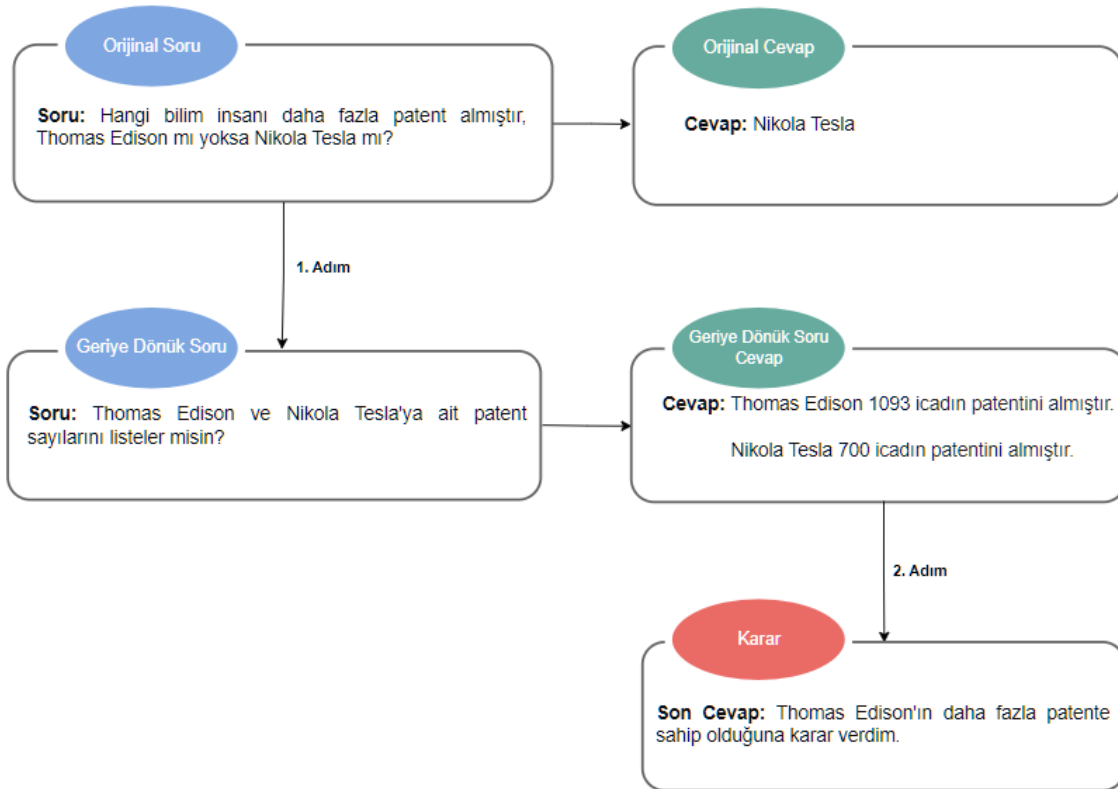


Şekil 14. Kural Tabanlı Yöntem ile İstem

Bu alandaki yöntemler dil modelinin sınırlamalarını anlamamıza ve kusurlarını belirlemenize yardımcı olur. Dil modelleri, öz farkındalık ve öz değerlendirme geliştirerek öğrenmeye daha açık hale gelir ve sürekli gelişmeyi mümkün kılar. Sonuç olarak bu alandaki yöntemler dil modelinin güvenilirliğini ve kullanılabilirliğini artırır.

Bir Adım Geriye Gitme İstemi (Take a Step Back Prompting)

Bir adım geriye gitme istemi, dil modellerini daha geniş bir bağlamı ve daha geniş bir bilgi yelpazesini dikkate almaya zorlamak için kullanılan bir yöntemdir. Bu tür bir rehberlik, modellerin daha derinlemesine anlamasını ve daha karmaşık soruları yanıtlamasını kolaylaştırır. Belirli bir soruyu yanıtlarken, bir model yalnızca mevcut metni değil aynı zamanda daha geniş bağlamı da hesaba katabilir. Bu yaklaşım özellikle aritmetik, sembolik akıl yürütme ve diğer karmaşık düşünme senaryoları gibi görevler için etkilidir (Zheng vd., 2024). Şekil 15, daha geniş bağlamın hesaba katılmasının ve geri adım atmanın, modelin sorulan soruya doğru yanıt vermesini nasıl sağladığını göstermektedir.



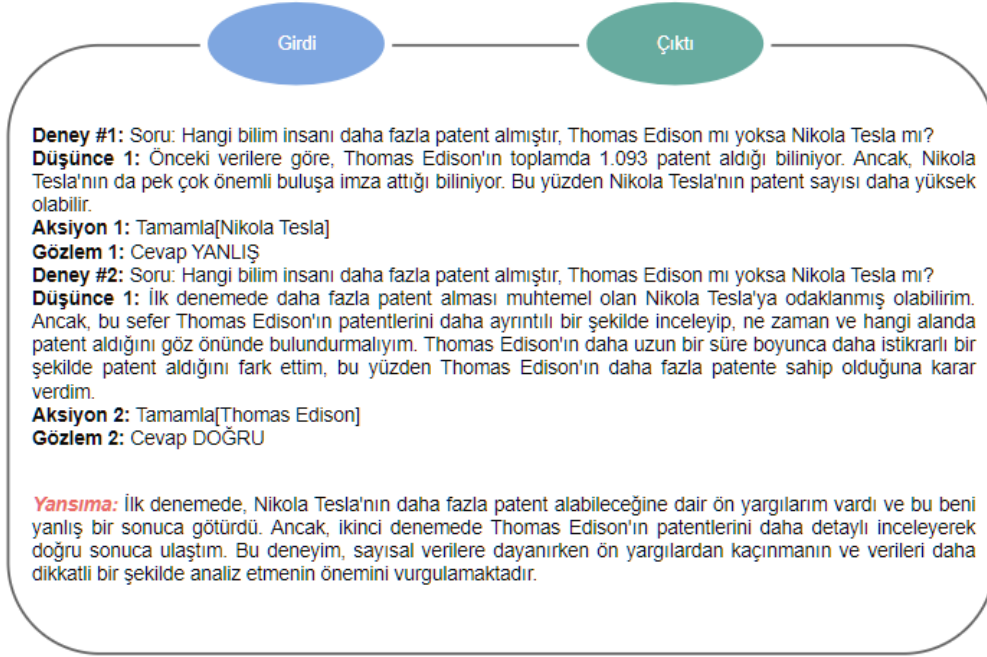
Şekil 15. Bir Adım Geriye Gitme Yöntemini Uygulayan İstem Örneği

Reflexion

Bu yöntem modelin sorulara yanıt vermeden önce bu soruları içeren bir bilgilendirme metniyle karşılaştırarak bir görevin daha geniş bağlamını anlamasını sağlar (Shinn vd., 2023). Modelin daha kapsamlı düşünmesine ve daha doğru ve bilinçli yanıtlar vermesine olanak tanır. Şekil 16, bir dil modelinin, bir görevle ilgili bir soruyu yanıtlamadan önce soruyu içeren bilgilendirme metnini dikkate alması durumunda doğru cevaba nasıl ulaştığını göstermektedir.

İstem Mühendisliği Uygulama Alanları

Tablo 2 istem mühendisliğinin çeşitli uygulama alanlarını listelemektedir. İstem mühendisliği doğal dil işleme, metin sınıflandırma, duygu analizi, çeviri ve daha birçok alanda kullanılmaktadır. Tablo 2'de her bir uygulama alanı için açıklayıcı istem girdileri ve çıktıları sunulmuştur.



Şekil 16. Reflexion Yöntemini Uygulayan Bir İstem Örneği

Tablo 2. İstem Mühendisliği Örnek Uygulama Alanları

Uygulama Alanı	İstem	Çıktı
İçerik oluşturma ve pazarlama	"Sağlıklı Beslenme İçin 10 Kolay ve Lezzetli Tarif" içeriği oluşturur musun?	1- Fırında Tavuk ve Sebzeler: Tavuk göğsü dilimlerini sebzelerle birlikte fırında pişirerek sağlıklı bir ana yemek elde edebilirsiniz. Tavuk dilimlerini zeytinyağı, taze otlar ve baharatlarla marine edin, ardından brokoli, havuç, biber ve patates gibi sebzelerle birlikte fırına verin.
Tasarım ve prototipleme	Bana yeni açtığım kafe için içerisinde kahve figürü olan bir logo tasarlar mısın?	Bu logo, kahve figürünü zarif ve çağdaş bir tarzda kullanıyor. Kahve çekirdekleriyle çevrelenmiş stilize bir kahve fincanı, kafanızın ana unsurlarını temsil ediyor. Umarım beğenirsiniz! (logo linki)
Metin özetleme	"Sefiller" romanının ana temasını özetleyebilir misiniz?	"Sefiller" romanının ana teması adalet, insanlık ve umuttur. Hikâye, farklı sosyal sınıflardan gelen karakterlerin hayatlarını ve içsel çatışmalarını anlatırken, adaletin ve merhametin önemini vurgular. Roman, temelde insanların geçmiş hatalarından ve toplumun onlara bakış açısından kurtulma ve yeniden doğma şansına sahip olduklarına dair güçlü bir mesaj ile doludur. Umut, roman boyunca süregelen bir tema olup, karakterlerin zorluklarla ve acılarla dolu hayatlarında bile umutlarını kaybetmemelerini vurgular.
Soru cevaplama	Türkiye'nin en yüksek dağı Ağrı Dağı mıdır?	Evet, Türkiye'nin en yüksek dağı Ağrı Dağı'dır. Ağrı Dağı, Türkiye'nin doğu sınırında bulunur ve en yüksek noktası olan 5.137 metrelik (16,854 feet) yüksekliğiyle ülkenin en yüksek zirvesidir. Ayrıca, bu dağ uluslararası tanımlamada Ararat Dağı olarak da bilinir.

Metin-SQL dönüşümü	Bana adında "ül" geçen kişilerin listesini getiren SQL kodunu yazar mısın? Tablo Adım: Kullanıcılar Tablo Alanlarım: Id, Adı, Soyadı	SELECT Adı, Soyadı FROM Kullanıcılar WHERE Adı LIKE '%ül%'
Çoklu dil çevirisi	"Fizik, kimya ve biyoloji gibi çeşitli disiplinler, doğa olaylarını anlamak ve açıklamak için gözlemler, deneyler ve teoriler kullanır." cümlesini Fransızca ve Almancaya çevirir misin?	Fransızca: "Des disciplines variées telles que la physique, la chimie et la biologie utilisent des observations, des expériences et des théories pour comprendre et expliquer les phénomènes naturels." Almanca: "Verschiedene Disziplinen wie Physik, Chemie und Biologie verwenden Beobachtungen, Experimente und Theorien, um natürliche Phänomene zu verstehen und zu erklären."
Metin sınıflandırma	Film harika! -Pozitif Hangisi iyi emin değilim. - Nötr Yolculuk iğrençti! --	Negatif
Rol yapma	Bilgisayar mühendisliği alanında uzman 2 kişinin rolünü üstleneceksin. Amacın, sana önereceğim ders adının uygunluğunu lisans eğitimi düzeyinde değerlendirmek. Her kişi değerlendirme sonuçlarını nedenleriyle açıklamalı. Ders Adı: "Çocuk Psikiyatrisi"	Profesör: Değerlendirme: Bu dersin bilgisayar mühendisliği programında yer alması akademik olarak uygun değil. Açıklama: Akademik danışman, öğrencilerin lisans eğitimleri boyunca aldıkları dersleri ve program gereksinimlerini yönetir. "Çocuk Psikiyatrisi" adlı bir ders, bilgisayar mühendisliği alanıyla ilgili değildir ve program gereksinimlerini karşılamaz. Doktor Öğretim Üyesi: Değerlendirme: Bu ders, bilgisayar mühendisliği öğrencileri için uygun değil. Açıklama: Temel bilgisayar bilimleri, matematik, programlama ve mühendislik prensipleri gibi konuları içeren derslere odaklanmalıdır. "Çocuk Psikiyatrisi" adlı bir ders, bu öğrencilerin program gereksinimlerini karşılamaz.

BDM'ler, yapay zekâ sistemlerinin beyinleri olarak algılanabilirler. Verilen talimatları algılama ve bu talimatlara göre eylemler üretme yeteneklerine sahiptirler. Ancak, BMD'lerin kapasite sınırlamaları vardır ve eğitim verilerinde zamansal gecikmeler yaşamaktadırlar. Örneğin, ChatGPT'nin en son eğitim verileri Ekim 2023'e kadar sınırlıdır. Gerçek zamanlı bilgi ve trendlere erişemezler ve belirli terimleri veya alanla ilgili bilgileri doğru bir şekilde anlamakta zorlanabilirler. Bazen eksik veya yanlış yanıtlara, hatta yanılsamalara neden olabilirler. Ek olarak, arama motorlarını kullanma, veri tabanlarını sorgulama ve harici araçları kullanma veya uygulama programlama arayüzlerini çağırma gibi harici uzman modellerle doğrudan etkileşime geçme yeteneğine sahip değildirler. Bu da BDM'lerin kullanılabilirliklerini sınırlamaktadır.

BDM'lerin karşılaştığı zorlukları aşmak önemli bir çalışma alanı haline gelmiştir. Bu noktada, bu zorlukları aşmak için istem mühendisliği tekniklerinin bir çözüm olabileceğinin ve sağladığı faydaların altını çizmek mümkündür. İstem mühendisliğinin uygulanması, BDM'lerin bilgiyi daha etkili bir şekilde işlemeleri; gerçek zamanlı bilgi ve trendlere daha hızlı erişim; belirli terimler veya alanlarla ilgili bilgileri daha doğru anlama imkanları sağlamaktadır. Ayrıca, BDM'lerin verimliliğini artırarak, eksik veya yanlış yanıtları ve kullanıcıların yanılsamalara düşme olasılığını azaltmaktadır. Sonuç olarak, istem mühendisliği tekniklerinin uygulanması, BDM'lerin karşılaştığı zorlukların üstesinden gelmek ve onların işlevselliğini artırmak için önemli bir adımdır. Bu tekniklerin kullanılması, BDM'lerin daha etkili ve güvenilir bir şekilde kullanılmasını sağlayarak, çeşitli uygulama alanlarında önemli faydalar sağlayabilir.

İstem mühendisliği alanındaki gelişmelerin bazı zorluklarla da karşılaşabileceği öngörülmektedir. Özellikle, karmaşık görevler için optimize edilmiş istemlerin geliştirilmesi ve kullanımının yaygınlaştırılması zaman alabilir. Ayrıca, BDM'lerin performansını iyileştirmek için kullanılan istem mühendisliği tekniklerinin etkili bir şekilde kullanılması için daha fazla araştırma ve geliştirme gerekebilir. İstem mühendisliği tekniklerinin kullanılmaması durumunda, BDM'lerin performansı genellikle belirli görevler veya konularla ilgili daha düşük olabilir. Özellikle spesifik veya karmaşık görevlerde, doğru sonuçlar elde etmek zorlaşabilir ve BDM'lerin kullanımı sınırlı olabilir. Ayrıca, kullanıcıların istedikleri cevapları veya sonuçları elde etme olasılığı azalabilir.

İstem mühendisliğinin geleceği oldukça umut vericidir. Bu tekniklerin daha yaygın bir şekilde benimsenmesiyle, BDM'lerin kullanım alanı genişleyecek ve daha fazla endüstri ve sektörde kullanılacaktır. Örneğin, edebiyat, sanat, sağlık ve müşteri hizmetleri gibi birçok alanda önemli yenilikler ve ilerlemeler getirebilir. Edebiyat alanında, özellikle yaratıcı yazma bağlamında fikir üretmek, yaratıcılığı harekete geçirmek ve yazma engelini aşmak için değerli bir araç olarak hizmet edebilir. Yazarlar, hayal güçlerini teşvik etmek ve duygularını uyandırmak için özel olarak hazırlanmış özel istemler hazırlayarak çeşitli temalar, karakterler oluşturabilirler. Bu yönlendirmeler, basit cümle başlangıçlarından daha karmaşık senaryolara veya görsel ipuçlarına kadar çeşitlilik gösterebilir ve yazarlara yaratıcı çabaları için bir sıçrama tahtası sağlayabilir. Ek olarak, yazmanın ortam, diyalog veya karakter gelişimi gibi belirli yönlerini hedef alan yönlendirmeleri stratejik olarak tasarlayarak yazarlar becerilerini geliştirebilirler. Sonuç olarak, hızlı mühendislik, yazarlara yeni fikirleri keşfetme ve repertuarlarını genişletme konusunda yapılandırılmış bir yaklaşım sunarak yaratıcı yazma sürecini zenginleştirir.

Müşteri hizmetleri alanında, müşteri sorularının ele alınmasında verimlilik ve hız artışı sağlayabilir. Müşteri sorunları veya sık sorulan sorular için standartlaştırılmış istemler geliştirilerek otomatik cevap oluşturan destek botları oluşturulabilir. Bu istemler, teknik sorunların giderilmesinden ürün bilgilerinin sağlanmasına veya hesap yönetimi konusunda yardıma kadar çok çeşitli senaryoları kapsayabilir. Ayrıca, istem mühendisliği, destek ekiplerinin yanıtlarını bireysel müşterilerin özel ihtiyaçlarına ve tercihlerine göre uyarlamasına olanak tanıyarak genel hizmet kalitesini artırabilir.

Tıp alanında karar verme süreçlerini ve teşhis doğruluğunu geliştirebilecek önemli bir potansiyele sahiptir. İstem mühendisliği, sağlık profesyonellerine teşhis protokolleri ve tedavi yolları boyunca rehberlik eden yapılandırılmış istemler oluşturarak, hasta semptomlarının ve tıbbi geçmişinin sistematik olarak değerlendirilmesine yardımcı olabilir. Tıp eğitiminde, öğrencilerin ve stajyerlerin kontrollü bir ortamda eleştirel düşünme ve klinik akıl yürütme uygulamalarına olanak tanıyacak şekilde klinik senaryoları simüle etmek için yönlendirmeler kullanılabilir. Ayrıca hasta iletişimi bağlamında, gerekli tüm bilgilerin açık ve kapsamlı bir şekilde iletilmesini sağlayarak etkileşimlerin standartlaştırılmasına yardımcı olabilir. Bu yaklaşım aynı zamanda hasta verilerine dayalı gerçek zamanlı destek ve öneriler sağlamak ve böylece daha bilinçli ve etkili klinik kararları kolaylaştırmak için elektronik sağlık kayıtları sistemlerine de entegre edilebilir. Sonuç olarak, istem mühendisliği, gelecekte BDM'lerin daha etkili bir şekilde kullanılmasını sağlayacak ve yapay zekâ teknolojilerinin yaygınlaşmasına katkıda bulunacaktır.

SONUÇ

BDM'ler, geniş bir bilgi birikimine erişme yeteneği ve çeşitli görevleri yerine getirme esnekliği ile dikkat çekerken, bazı eksiklikler ve problemler de içerirler. Örneğin, sınırlı eğitim verisi, zamansal gecikmeler ve belirli konularla ilgili anlayış eksikliği gibi sorunlar BDM'lerin performansını etkileyebilir. Ayrıca, harici uzman modellerle doğrudan etkileşime geçme yeteneklerinin bulunmaması da kullanılabilirliklerini sınırlayabilir.

İstem mühendisliği geniş bir uygulama ve araştırma alanına sahip BDM'lerin potansiyelini açığa çıkaran, onları etkili bir şekilde kullanmak için istemleri geliştirmeye ve optimize etmeye odaklanan yeni bir disiplindir. Bu nedenle, BDM'lerin karşılaştığı sorunlara çözüm sunabilir. Ancak her bir tekniğin avantajlarının yanında dezavantajları da vardır. Örneğin, sıfır atış istem ve birkaç atışlı istem, önceden eğitilmiş bir modelin genel bilgisini kullanarak yeni görevlerin çözülmesini sağlar. Birkaç atışlı istem, modele az miktarda örnek vererek daha iyi performans gösterir. Her iki yöntemin avantajı, önceden etiketlemeye ihtiyaç duymamalarıdır. Dezavantajları ise, karmaşık görevlerde performanslarının düşmesi ve her zaman doğru sonuçlar üretmemeleridir.

Düşünce zinciri istemi, modelin mantıksal bir sıra izleyerek cevap vermesini sağlayarak daha tutarlı ve bilgilendirici yanıtlar üretmesine katkıda bulunur. Otomatik düşünce zinciri ise bu işlemi otomatik hale getirerek zamandan tasarruf sağlar. Bu iki yöntem karmaşık problemlerde zorlanabilmekte ve her zaman doğru sonuçlar üretmemektedir. ReAct istemi, kullanıcının geri bildirimlerini kullanarak modeli dinamik olarak uyarlar ve daha iyi performans göstermesini sağlar. Öz tutarlılık ise modelin ürettiği yanıtların tutarlılığını kontrol ederek hataları minimuma indirir. Her iki yöntem modelin performansını ve güvenilirliğini arttırmaktadır. Dezavantajları ise, ek işlem gücü gerektirmeleri ve her zaman uygulanabilir olmamalarıdır.

Öz tutarlılık yöntemi modelin ürettiği yanıtların tutarlılığını kontrol ederek hataları minimuma indirir. Bu sayede modelin anlamsız yanıtlar üretmesini engelleyerek hata yapma olasılığını azaltır. Daha güvenilir ve doğru sonuçlar üretir. Ek işlem gücü gerektirdiği için her zaman kullanılması uygun olmayabilir. Düşünce ağacı yöntemi ise modelin birden fazla olası çözümü göz önünde bulundurarak en iyi sonuca ulaşmasını sağlar. Modele farklı bakış açıları kazandırdığı için daha yaratıcı ve özgün çözümler sunar. Karmaşık problemlerdeki performansı daha iyidir. Zaman ve ek işlem gücü gerektirir. Yönlü uyarıcı yöntemi modele, yanıtını yönlendirmek için ek bilgiler sağlar. Modelin dikkatini belirli bir konuya yönlendirerek modelin daha tutarlı ve doğru yanıtlar üretmesine yardımcı olur. Ek bilgi gerektirdiği için her alanda uygulanamayabilir. Modelin yaratıcılığı kısıtlanmaktadır.

Bilgi tabanlı yöntemde modele, yanıtını oluşturabileceği bilgi kaynağı sağlanır. Model bu sayede yeni bilgiler edinir ve daha doğru yanıtlar üretmektedir. Dezavantajlarında ise modele her zaman güncel ve doğru bilgi aktarımı yapmak zor olabilir. Ek işlem gücü gerekmektedir. Şablon tabanlı yöntemde ise modele, yanıtını oluşturabileceği önceden tanımlanmış bir format sağlanır. Bu sayede model belirli bir formata uyum sağlayarak hatalı, eksik yanıtlar üretmesi engellenir. Bu model kullanımında da modelin yaratıcılığı engellenmektedir. Her alanda kullanımı uygun değildir.

Kural tabanlı yöntem, modele yanıtını oluşturmak için kullanabileceği kurallar sağlar. Bu şekilde model belirli bir mantığa göre yanıt vermektedir. Modelin amaca yönelik yanıtlar vermesi sağlanır. Her alanda modele karmaşık kurallar eklemek ve bunları yönetmek kolay olmayabilir. Bir adım geriye gitme yöntemi modelin, yanıtını tekrar gözden geçirmesine ve gerekirse düzeltmesine olanak tanır. Model bu sayede daha doğru ve tutarlı yanıtlar üretmektedir. Model hataları fark eder ve düzeltebilir. Ek işlem gerektirdiği için zamana ihtiyaç duyulmaktadır. Reflexion yöntemi ise modele, kendi düşünme sürecini yansıtma ve analiz etme yeteneği kazandırır. Bu sayede model hatalarını fark ederek diğer cevaplarda daha doğru yanıtlar üretir, problem çözme becerisi kazanır. Karmaşık yöntemlerde uygulanması ek işlem gücü gerektirdiği için her alanda uygulanmak uygun olmayabilir.

Sonuç olarak, istem mühendisliği teknikleri BDM'lerin potansiyelini daha iyi bir şekilde değerlendirmek ve zorlukları aşmak için önemli bir araç haline gelmiştir. Ancak her yöntemin kendine özgü avantajları ve dezavantajları vardır. Hangi yöntemin kullanılacağı, sorunun türüne, modelin özelliklerine ve kullanıcının ihtiyaçlarına bağlıdır. Farklı yöntemlerin bir arada kullanımı, BDM'lerin performansını daha da geliştirmenin bir yoludur. Gelecekte, bu alandaki gelişmelerin hızlanması beklenmektedir. İstem mühendisliği teknikleri, BDM'lerin daha iyi anlaşılmasını sağlayacak ve kullanıcıların ihtiyaçlarını daha iyi karşılayacak şekilde geliştirilecektir. Özellikle, daha akıllı ve esnek istemler, BDM'lerin daha karmaşık görevleri yerine getirebilmesini sağlayacak ve kullanıcı deneyimini iyileştirecektir.

KAYNAKLAR

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language Models are Few-Shot Learners. İçinde H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Ed.), *Advances in Neural Information Processing Systems* (C. 33, ss. 1877-1901). Curran Associates, Inc. Geliş tarihi gönderen https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- Chen, S., Wang, W., Chen, X., Lu, P., Yang, Z., & Du, Y. (2024). LLaMA-LoRA Neural Prompt Engineering: A Deep Tuning Framework for Automatically Generating Chinese Text Logical Reasoning Thinking Chains. *Data Intelligence*, 1-53. https://doi.org/10.1162/dint_a_00251
- Kojima, T., Gu, S. (Shane), Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models are Zero-Shot Reasoners. İçinde S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Ed.), *Advances in Neural Information Processing Systems* (C. 35, ss. 22199-22213). Curran Associates, Inc. Geliş tarihi gönderen https://proceedings.neurips.cc/paper_files/paper/2022/file/8bb0d291acd4acf06ef112099c16f326-Paper-Conference.pdf
- Lester, B., Al-Rfou, R., & Constant, N. (2021). The Power of Scale for Parameter-Efficient Prompt Tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3045-3059. Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.emnlp-main.243>
- Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. Le, ... Hajishirzi, H. (2021). Generated Knowledge Prompting for Commonsense Reasoning. *arXiv* . <https://doi.org/10.48550/arXiv.2110.08387>
- Long, J. (2023). Large Language Model Guided Tree-of-Thought. *arXiv* . <https://doi.org/10.48550/arXiv.2305.08291>
- Ma, R., Zhou, X., Gui, T., Tan, Y., Li, L., Zhang, Q., & Huang, X. (2021). Template-free Prompt Tuning for Few-shot NER. *arXiv* . <https://doi.org/10.48550/arXiv.2109.13532>
- Paranjape, B., Lundberg, S., Singh, S., Hajishirzi, H., Zettlemoyer, L., & Ribeiro, M. T. (2023). ART: Automatic multi-step reasoning and tool-use for large language models. *arXiv* . <https://doi.org/10.48550/arXiv.2303.09014>
- Polverini, G., & Gregorcic, B. (2024a). How understanding large language models can inform the use of ChatGPT in physics education. *European Journal of Physics*, 45(2), 025701. <https://doi.org/10.1088/1361-6404/ad1420>
- Polverini, G., & Gregorcic, B. (2024b). How understanding large language models can inform the use of ChatGPT in physics education. *European Journal of Physics*, 45(2), 025701. <https://doi.org/10.1088/1361-6404/ad1420>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., & others. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. *arXiv* . <https://doi.org/10.48550/arXiv.2402.07927>
- Seo, J., Moon, H., Lee, C., Eo, S., Park, C., Kim, J., ... Lim, H. (2022). Plain Template Insertion: Korean-Prompt-Based Engineering for Few-Shot Learners. *IEEE Access*, 10, 107587-107597. <https://doi.org/10.1109/ACCESS.2022.3213027>
- Shaier, S., Bennett, K., Hunter, L. E., & von der Wense, K. (2024). Comparing Template-based and Template-free Language Model Probing. *arXiv* . <https://doi.org/10.48550/arXiv.2402.00123>
- Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: language agents with verbal reinforcement learning. İçinde A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Ed.), *Advances in Neural Information Processing Systems* (C. 36, ss. 8634-8652). Curran Associates, Inc. Geliş tarihi

gönderen https://proceedings.neurips.cc/paper_files/paper/2023/file/1b44b878bb782e6954cd888628510e90-Paper-Conference.pdf

Shu, M., Nie, W., Huang, D.-A., Yu, Z., Goldstein, T., Anandkumar, A., & Xiao, C. (2022). Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models. İçinde S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Ed.), *Advances in Neural Information Processing Systems* (C. 35, ss. 14274-14289). Curran Associates, Inc. Geliş tarihi gönderen https://proceedings.neurips.cc/paper_files/paper/2022/file/5bf2b802e24106064dc547ae9283bb0c-Paper-Conference.pdf

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical Networks for Few-shot Learning. İçinde I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Ed.), *Advances in Neural Information Processing Systems* (C. 30). Curran Associates, Inc. Geliş tarihi gönderen https://proceedings.neurips.cc/paper_files/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf

Taveekitworachai, P., Abdullah, F., Dewantoro, M. F., Thawonmas, R., Togelius, J., & Renz, J. (2023). ChatGPT4PCG Competition: Character-like Level Generation for Science Birds. *2023 IEEE Conference on Games (CoG)*, 1-8. IEEE. <https://doi.org/10.1109/CoG57401.2023.10333206>

Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, koray, & Wierstra, D. (2016). Matching Networks for One Shot Learning. İçinde D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Ed.), *Advances in Neural Information Processing Systems* (C. 29). Curran Associates, Inc. Geliş tarihi gönderen https://proceedings.neurips.cc/paper_files/paper/2016/file/90e1357833654983612fb05e3ec9148c-Paper.pdf

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., ... Zhou, D. (2022). Self-Consistency Improves Chain of Thought Reasoning in Language Models. *arXiv* . <https://doi.org/10.48550/arXiv.2203.11171>

Wu, P. (2022). A Survey of Few-Shot Learning Research Based on Deep Neural Network. *Frontiers in Computing and Intelligent Systems*, 2(1), 110-115. <https://doi.org/10.54097/fcis.v2i1.3177>

Yang, J., Guo, X., Li, Y., Marinello, F., Ercisli, S., & Zhang, Z. (2022). A survey of few-shot learning in smart agriculture: developments, applications, and challenges. *Plant Methods*, 18(1), 28. <https://doi.org/10.1186/s13007-022-00866-2>

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. İçinde A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Ed.), *Advances in Neural Information Processing Systems* (C. 36, ss. 11809-11822). Curran Associates, Inc. Geliş tarihi gönderen https://proceedings.neurips.cc/paper_files/paper/2023/file/271db9922b8d1f4dd7aaef84ed5ac703-Paper-Conference.pdf

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing Reasoning and Acting in Language Models. *arXiv* . <https://doi.org/10.48550/arXiv.2210.03629>

Zhang, Z., Zhang, A., Li, M., & Smola, A. (2022). Automatic Chain of Thought Prompting in Large Language Models. *arXiv* . <https://doi.org/10.48550/arXiv.2210.03493>

Zheng, H. S., Mishra, S., Chen, X., Cheng, H.-T., Chi, E. H., Le, Q. V., & Zhou, D. (2024). Take a Step Back: Evoking Reasoning via Abstraction in Large Language Models. *The Twelfth International Conference on Learning Representations*. Geliş tarihi gönderen <https://openreview.net/forum?id=3bq3jsvcQ1>

Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to Prompt for Vision-Language Models. *International Journal of Computer Vision*, 130(9), 2337-2348. <https://doi.org/10.1007/s11263-022-01653-1>

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2022). Large Language Models Are Human-Level Prompt Engineers. *arXiv* . <https://doi.org/10.48550/arXiv.2211.01910>

