

ChatGPT Kontrol Kodları

Cevabı sorgulamak için 20 ifade

Bu ifadeler ChatGPT'yi daha akıllı yapmıyor. Sadece neye dayanarak konuştuğunu sana gösteriyor. Bir cevabın doğru olup olmadığını yine sen karar vereceksin, ama en azından karar verecek bilgiye sahip olacaksın.

Nasıl kullanılır

- 1 Sorunu her zamanki gibi sor, cevabı al.
- 2 Aynı sohbette, altına listeden bir ifade yaz.
- 3 Gelen ikinci cevabı birinciyle karşılaştır.
- 4 İkisi çelişiyorsa, ilk cevaba güvenme.

İşaretlerin anlamı

- Videoda gösterdiğim, denenmiş olanlar
- Aynı mantıkla çalışan varyasyonlar, sen de dene

Videoda gösterdiğim 5 kod

Videodan geldiysen aradıkların bunlar. Gerisi aynı mantığın varyasyonları.

- **cite or say I don't know**

Kaynak veremiyorsa uydurmuyor, bilmiyorum diyor

- **assumptions**

Cevabın altındaki varsayımları tek tek çıkarıyor

- **steelman**

Karşı görüşü en güçlü haliyle savunuyor

- **limitations**

Bu iddianın nerede çöktüğünü söylüyor

- **confidence**

Ne kadar emin olduğunu ve neye bakarak söylediğini veriyor

Üç adımlık kontrol

Önemli bir cevap aldıysan tek tek değil, sırayla sor.

assumptions

→

limitations

→

confidence

Önce neyi varsaydığını, sonra nerede çöktüğünü, en sonda ne kadar emin olduğunu soruyorsun. Üçü bir arada, cevabın ne kadarına güvenebileceğini gösteriyor. Bir rapora, bir karara ya da bir ödeve dayanak yapacaksan bu üç adımı atlama.

Uydurmasını zorlaştır

Model bilmediği yeri boş bırakmaz, en olası cümleyi kurar. Bunlar o boşluğu görünür yapar.

- **cite or say I don't know**

Kaynak yoksa bilmiyorum der

- **don't invent citations**

Uydurma kaynak vermesini engeller

- **only use the text I gave**

Sadece verdiğin metinden cevaplar

- **quote the exact line**

Dayandığı satırı aynen alıntılar

- **mark unknown as unknown**

Emin olmadığı yeri işaretler

- **fact or inference**

Neyin veri neyin çıkarım olduğunu ayırır

Neye dayandığını göster

Cevabın kendisi değil, altındaki zemin. Yanlıklar genelde burada saklanıyor.

- **assumptions**

Cevabın altındaki varsayımları çıkarır

- **confidence**

Ne kadar emin olduğunu ve nedenini verir

- **what would change your answer**

Fikrini değiştirecek bilgiyi söyler

- **sources first, answer second**

Önce kaynakları, sonra cevabı verir

- **what don't you know here**

Eksik kalan bilgiyi listeler

- **how would you be wrong**

Yanılma ihtimalini kendisi anlatır

Sana katılmayı bırak

Bu araçlar seni onaylamaya meyilli. Karşı tarafı da kurdurman gerekiyor.

- **steelman**

Karşı görüşü en güçlü haliyle savunur

- **limitations**

İddianın nerede çöktüğünü söyler

- **3 reasons this fails**

Başarısız olma sebeplerini sıralar

- **what would a skeptic say**

Şüpheli birinin itirazlarını yazar

- **argue against me**

Senin görüşünün karşısına geçer

- **who disagrees and why**

Karşı çıkan tarafları ve gerekçelerini verir

Kendi işini denetlesin

İlk cevap taslaktır. İkinci geçiş çoğu zaman daha uzun prompt yazmaktan iyidir.

- **critique your own answer**

Kendi cevabını eleştirir

- **grumpy editor**

Aksi bir editör gibi zayıf yerleri listeler

- **what did you skip**

Atladığı noktaları söyler

- **rate this 1-10 and why**

Kendi cevabına not verir ve gerekçelendirir

En işe yarayan şey: kaynağı sen ver

Bu listedeki bütün ifadeler işe yarıyor ama hiçbiri şunun kadar değil. Modele kendi metnini ver ve dışına çıkmasını yasakla:

sadece verdiğim metni kullan, dışına çıkma,
metinde yoksa bilinmiyor de

Model bir şeyi hatırlamaya çalıştığında uydurma ihtimali en yüksek. Metni sen verdiğinde hatırlamasına gerek kalmıyor, sadece okuyor. Rapor, makale, sözleşme ya da ders notu özetlerken en güvenli yol bu.

Üstüne ekleyebileceklerin

her iddianın yanına satır numarası	Nereden aldığını tek tek gösterir
metinde olmayanı ekleme	Kendi bilgisini karıştırmasını engeller
çelişkileri ayrıca listele	Metnin kendi içindeki tutarsızlıkları bulur
emin olmadığın yeri işaretle	Yorum kattığı yerleri görünür yapar

Listeyi ezberleme, mantığını al

1 Bu ifadeler doğruluk garantisi vermiyor

"cite or say I don't know" yazdığına model uydurmayı azaltıyor ama bitirmiyor. Kaynağın kendisini de uydurabiliyor. Verdiği linkleri açıp bakmak hâlâ senin işin.

2 Sana katılması haklı olduğun anlamına gelmiyor

Bu sistemler karşı çıkmaya değil, devam etmeye ayarlı. Onay aldığında rahatlama; onayın kendisi zaten en olası cümle olabilir.

3 İkinci cevap birinciyle çelişiyorsa ikisi de şüphelidir

Aynı soruya iki farklı cevap veriyorsa, o konuda sağlam bir zemini yok demektir. Bu bir arıza değil; nasıl çalıştığının sonucu.

4 Önemli olan işlerde tek kaynağa güvenme

Bu araçlar bir başlangıç noktası. Bir karara, bir rapora ya da bir ödev dayanak yapacaksan dışarıdan doğrulama şart.



Dr. Uğur Sarıçam

@ugur.saricam

Yapay zekayı herkesin anlayacağı hale getiriyorum

Doktora: Fen Eğitimi | Master: Yapay Zeka

NOT

Bu ifadeler bir cevabın doğruluğunu garanti etmez, sadece dayanağını görünür kılar. Model verdiği kaynağı da uydurabilir; bağlantıları açıp kontrol etmek kullanıcıya aittir. Belgenin hazırlanmasında yapay zeka desteği kullanılmıştır.