Yapay zekâ metni, görüntüleri ve sesi aynı anda kullanabilir mi?

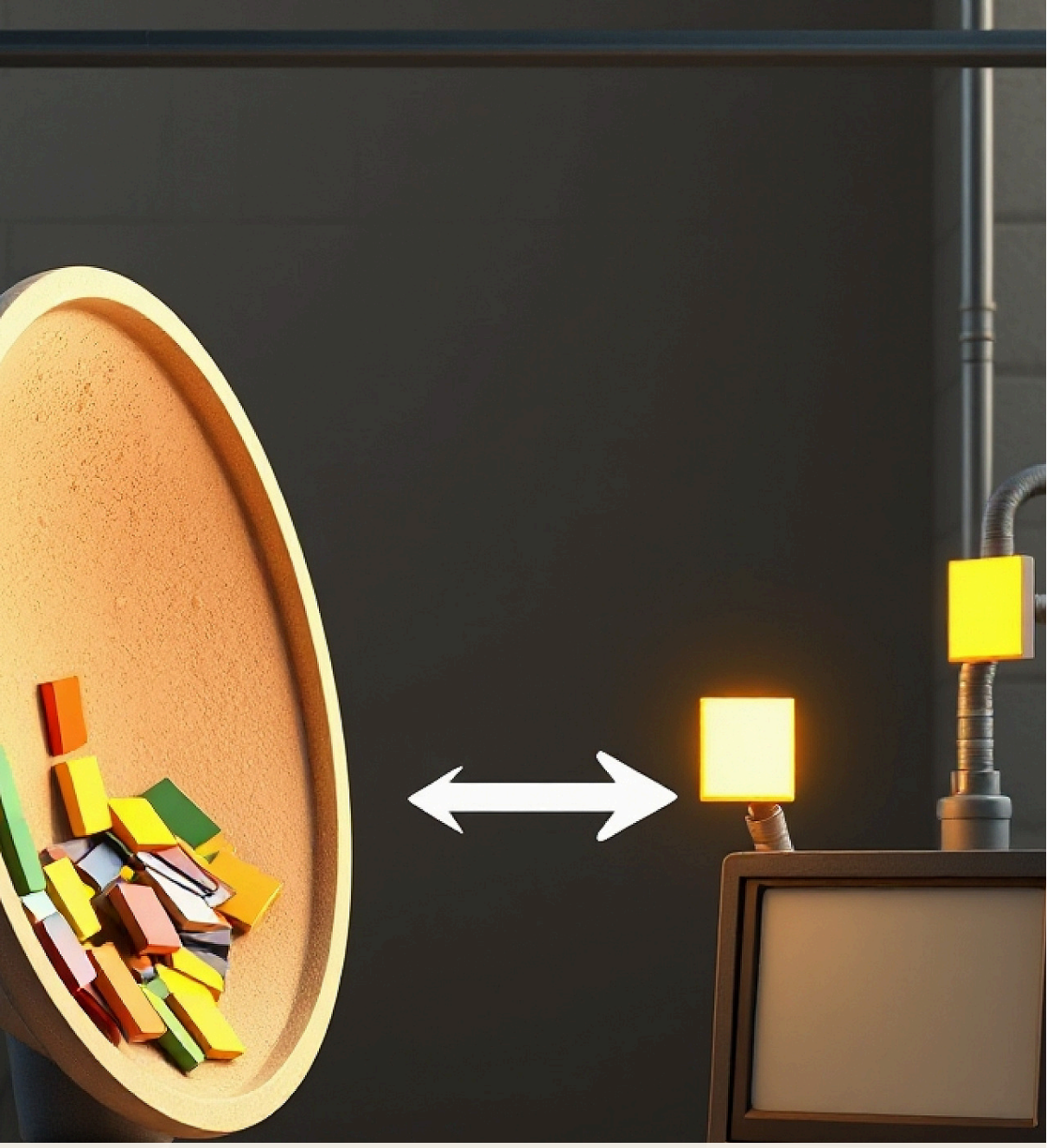


www.yzokul.com.tr

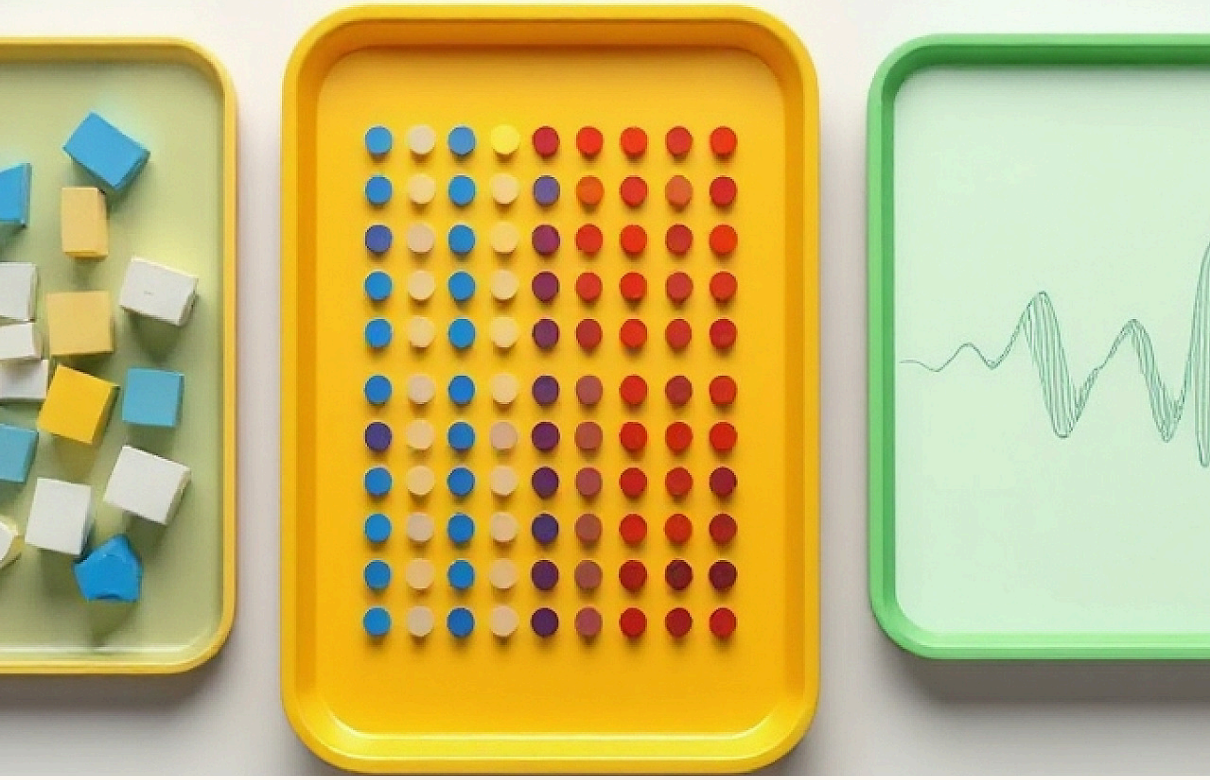
Yapay zekâ metni, görüntüleri ve sesi aynı anda kullanabilir mi?



Şu soruyu görüyorsun: bir yapay zekâ metni, görüntüleri ve sesi aynı anda kullanabilir mi? Kısa yanıt evet, sistem birkaç tür girdiyi birlikte işleyecek şekilde kurulduğunda. Böyle bir sisteme çok kipli denir. Örneğin bir kullanıcı bir resim gösterebilir ve sesli bir soru sorabilir, yapay zekâ da yazılı yanıt verebilir. Bu kitap bunun nasıl işlediğini adım adım, sade bir dille açıklar.



Bir yapay zekâ bilgiyi girdi olarak alır ve bir şeyi çıktı olarak geri verir. Metin, görüntüler ve ses yalnızca farklı girdi türleridir. Örneğin bir cümle yazarsın (metin girdisi) ve yapay zekâ bir yanıt yazar (metin çıktısı). Ya da bir fotoğraf gösterirsin (görüntü girdisi) ve yapay zekâ bir kelime söyler (ses çıktısı). Girdi ve çıktıyı anlamak, bir yapay zekânın birkaç türü aynı anda nasıl kullanabildiğini görmenin ilk adımıdır.



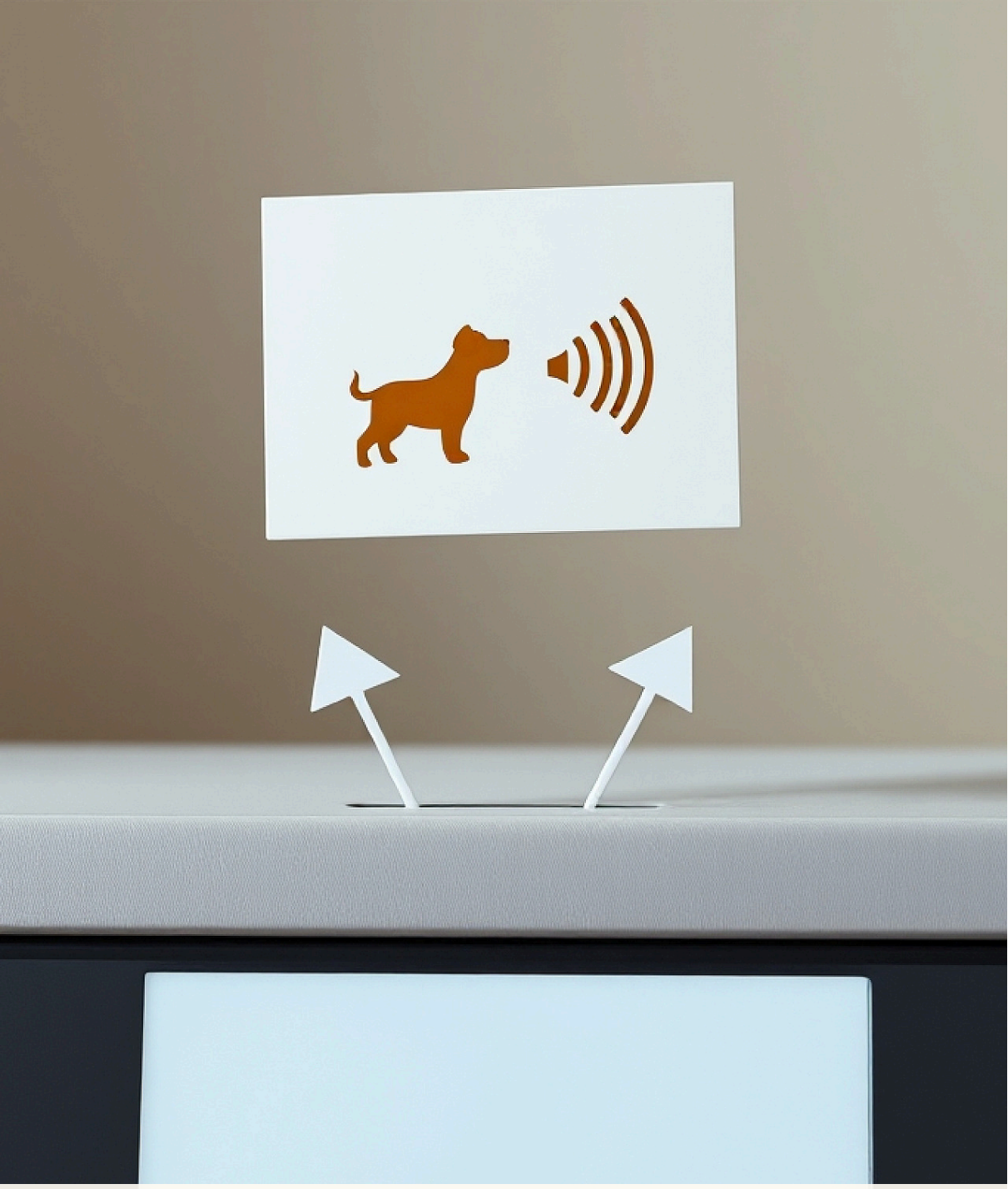
Her türden bir örneğe bak: kedi paspasın üstünde oturuyor gibi yazılı bir cümle, paspasın üstünde bir kedinin resmi ve söylenen kedi kelimesi. Her biri farklı bir veri türüdür. Metin harfler ve kelimelerdir, görüntü piksel denen renkli noktalardan oluşan bir ızgaradır, ses ise bir mikrofonun sayılara dönüştürdüğü bir hava basıncı dalgasıdır. Bu kitaptaki üç veri türü bunlardır.



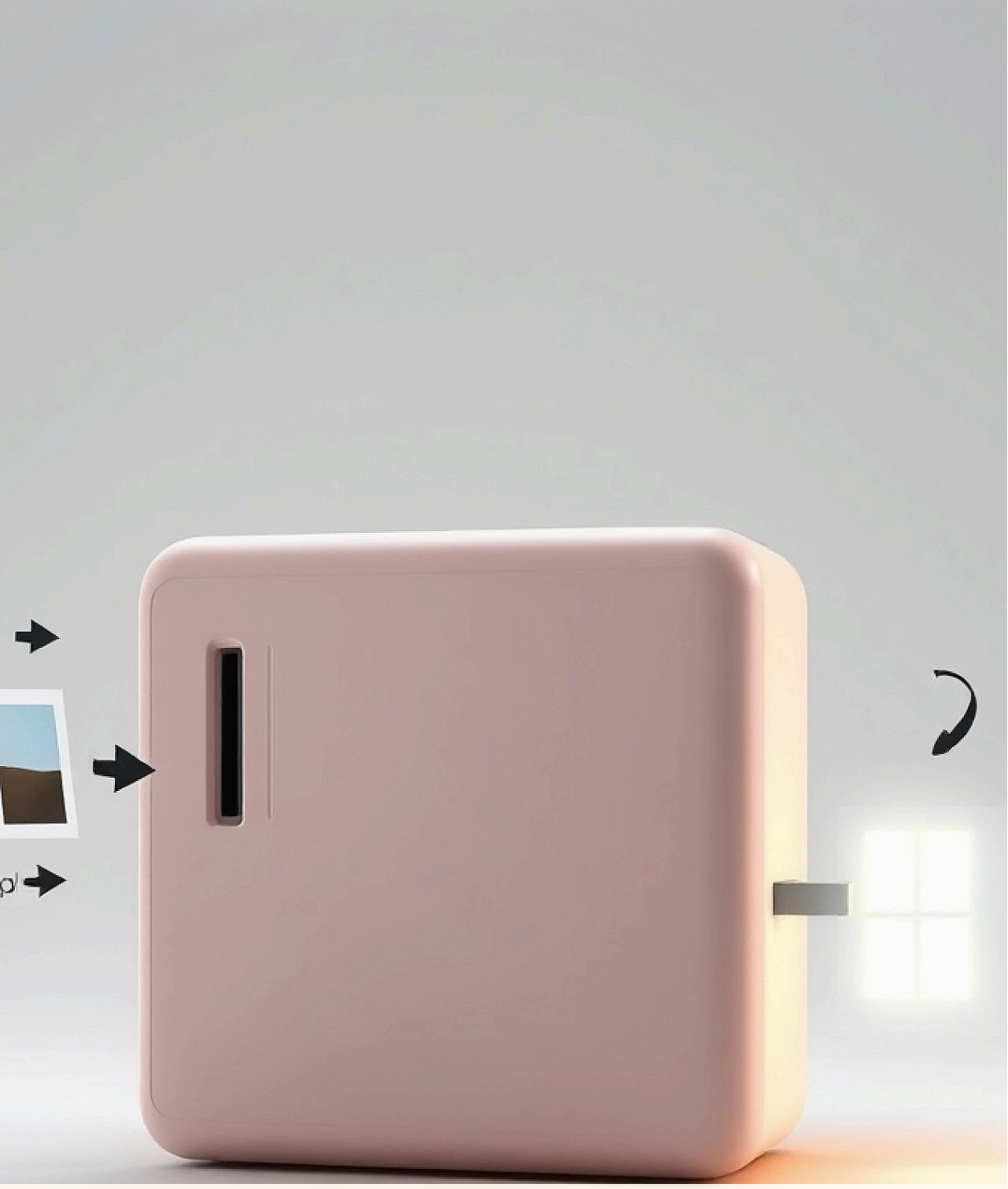
Basit bir sistem yalnızca metni okuyup metinle yanıt verebilir. Örneğin bir soru yazarsın ve o da yazılı bir yanıt verir. Birçok sistem tek bir veri türüyle çalışır. Bu birçok görev için yeterlidir, ama bir resme bakamaz ya da bir sesi dinleyemez. Birden fazla türü kullanmak için sistemin her girdi türü için ek parçalara ihtiyacı vardır.



Birden fazla veri türü kullanan bir sisteme çok kipli denir. Çok kipli kelimesi yalnızca birçok girdi kipi anlamına gelir. Örneğin metni okuyan ve görüntülere bakan bir sistem çok kiplidir. Sesi de dinleyen bir sistem de çok kiplidir. Yani bir yapay zekâ metni, görüntüleri ve sesi birlikte kullandığında o bir çok kipli sistemdir.



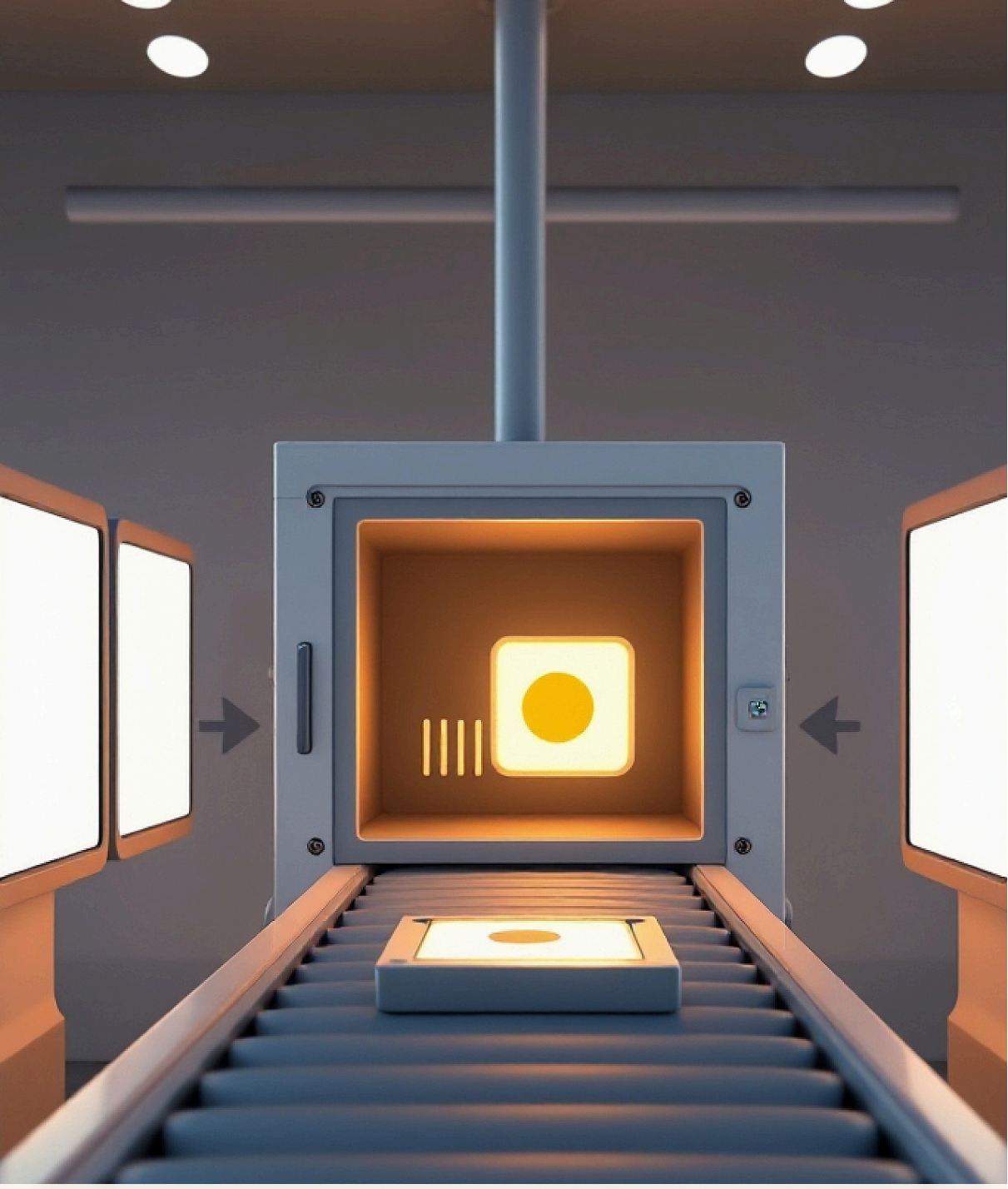
Basit bir örneği izle: bir kullanıcı bir görüntü gösterir ve yüksek sesle bir soru sorar. Sistem aynı anda bir görüntü ve bir ses alır. Örneğin kullanıcı bir köpeğin fotoğrafını gösterir ve bu hangi ırk? der. Sistem hem resmi hem de söylenen kelimeleri işlemek zorundadır. Bu yaygın bir çok kipli görevdir.



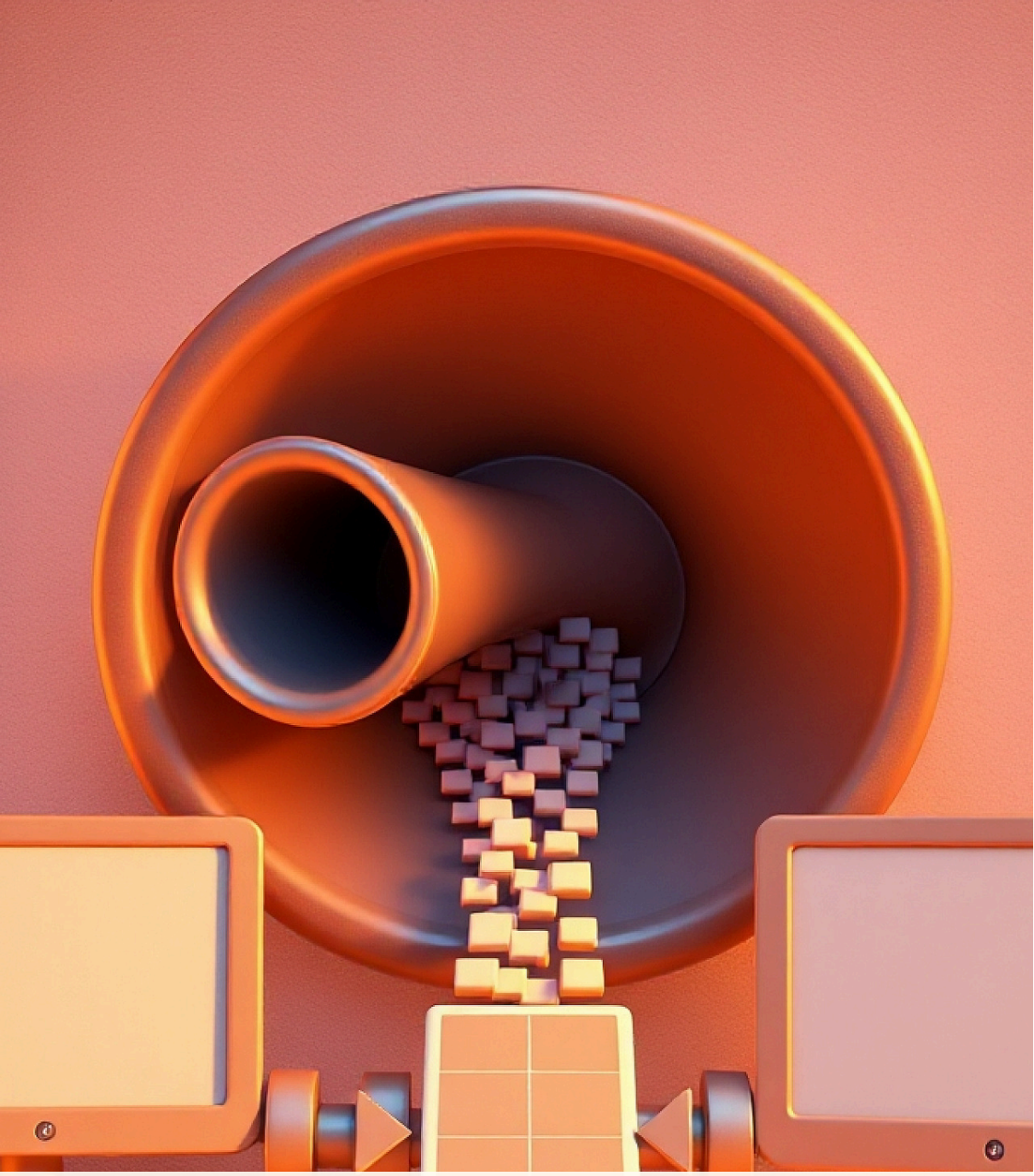
Sistem, görüntüyle ilgili sesli soruya yazılı olarak yanıt verir. Örneğin bu bir golden retriever'a benziyor diye yazabilir. Artık metin, görüntü ve ses tek bir görevde birlikte yer alır: ses girdi, görüntü girdi, metin çıktı. Yazılı yanıt çıktıdır. Görüntü ve sesli soru girdilerdir. Üç veri türü de birlikte kullanılır.



Görüntü ve söylenen soru aynı göreve bağlanmalıdır. Aksi halde sistem sorunun neyle ilgili olduğunu anlayamaz. Örneğin ses bu ne renk? diye sorarsa ama görüntü bağlı değilse, sistem hangi nesneyi betimleyeceğini bilemez. Bağlamak, sistemin görüntüyü ve soruyu iki ayrı iş değil, tek bir işin parçaları olarak ele alması demektir.



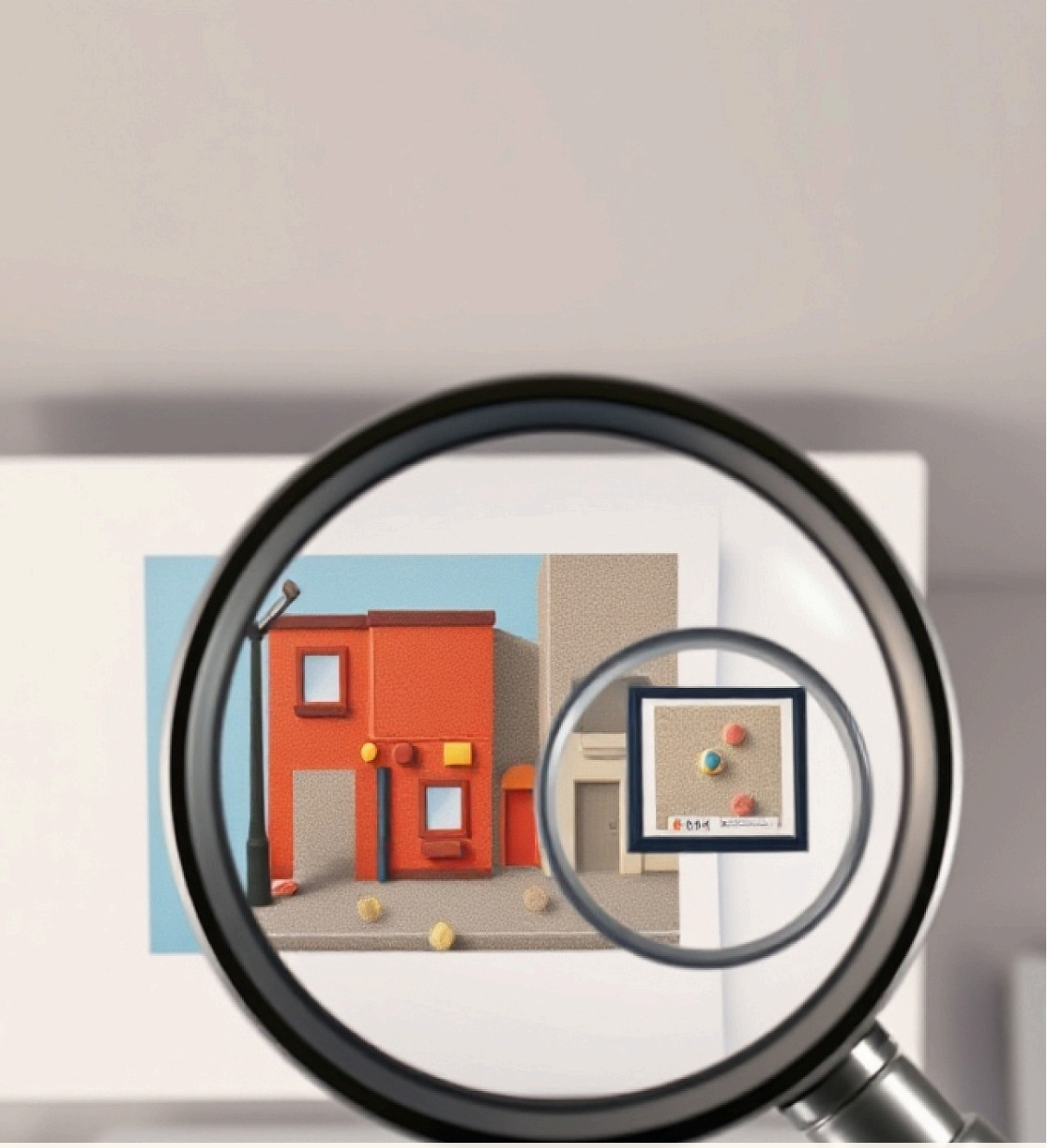
Tüm akışı tek bir görev olarak gör: görüntü artı ses girer, yazılı bir yanıt çıkar. Girdiler ayrı ayrı ele alınmaz, ortak bir işin içinde birleştirilir. Örneğin sistem önce sesli sorunun görüntüye gönderme yaptığını anlar, sonra görüntüye bakar, sonra yanıtı yazar. Tek görev, birkaç girdi, tek çıktı.



Söylenen kelimeler çoğu zaman sistem onları kullanmadan önce metne dönüştürülür. Sesin geri kalanla bağlanmasının bir yolu budur. Örneğin bir konuşmadan metne parçası sesli soruyu dinler ve onu kelimeler olarak yazar. Sonra sistem soruyu metin gibi ele alıp görüntüyle birleştirebilir. Yani ses önce metne dönüşür, sonra tüm parçalar birlikte çalışır.



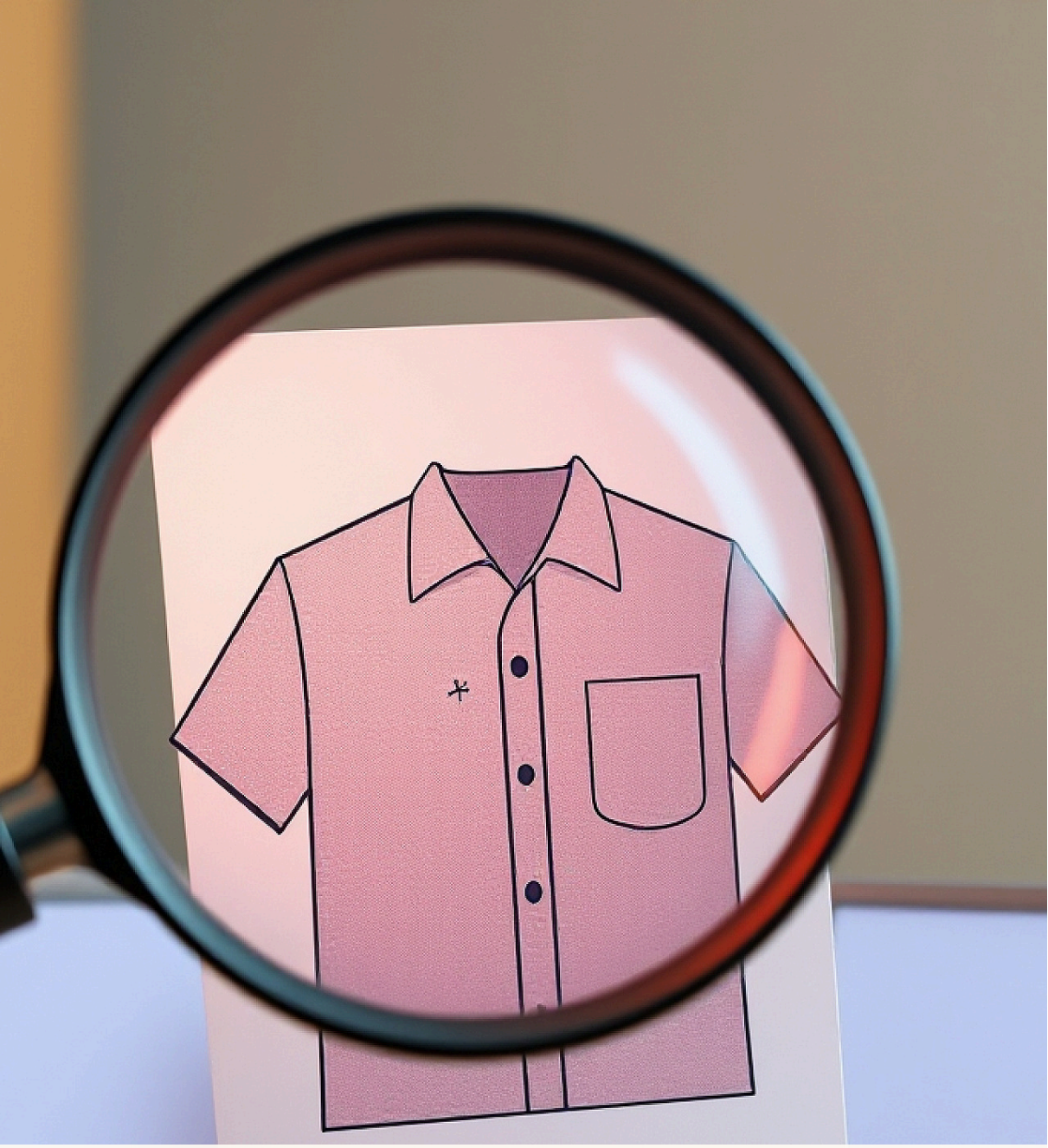
Her çok kipli sistem tüm veri türleriyle aynı başarıyla baş etmez. Bir sistem metinde güçlü olup seste daha zayıf olabilir. Örneğin yazılı soruları çok iyi okuyabilir ama gürültülü bir ses kaydını yanlış anlayabilir. Ya da net görüntüleri iyi betimleyebilir ama bulanık olanlarla zorlanabilir. Bir türdeki beceri, otomatik olarak bir başkasındaki beceri anlamına gelmez.



Bir sistem bir görüntüyü her ayrıntıyı doğru okumadan genel kelimelerle betimleyebilir. Örneğin resimde arabaların olduğu bir sokak var diyebilir ama küçük bir tabelayı kaçırabilir ya da bir sayıyı yanlış okuyabilir. Bir resmi betimlemek, onu tam olarak okumakla aynı şey değildir. Yani bir görüntüden kesin ayrıntılara ihtiyaç duyduğunda bunları kendin kontrol et ya da belirli kısımları sor.



Yaygın bir hata, her sistemin metni, görüntüleri ve sesi eşit derecede iyi ele aldığını varsaymaktır. Gerçekte bir türdeki beceri, bir başkasındaki beceriyi garanti etmez. Örneğin mükemmel metin yanıtları yazan bir sistem görüntüleri kötü betimleyebilir ya da söylenen kelimeleri yanlış duyabilir. Sistemi her zaman kullanmayı planladığın girdi türünde test et ve eşit kalite varsayma.



Bir başka yaygın hata, bir sistemin bir görüntüden bildirdiği her küçük ayrıntıya güvenmektir. Genel bir betimleme ince ayrıntıları kaçırabilir ya da yanlış okuyabilir. Örneğin sistem bir gömleğin yeşil olduğu halde mavi olduğunu söyleyebilir ya da köşedeki küçük bir nesneyi kaçırabilir. Betimlemeyi genel bir kılavuz olarak kullan ve önemli olduğunda önemli ayrıntıları kendin kontrol et.



Üçüncü hata, girdilerin bağlanmadan birlikte çalıştığını düşünmektir. Görüntü ve soru ancak tek bir görevde birleştirildiğinde anlam kazanır. Örneğin bir görüntü gösterip soruyu ayrı bir adımda sorarsan sistem bunları bağlamayabilir. Sistemin tüm girdileri aynı işin parçası olarak ele aldığından emin ol, yoksa yanıt yanlış şeyle ilgili olabilir.



Evet, bir yapay zekâ çok kipli bir sistemde metni, görüntüleri ve sesi aynı anda kullanabilir. Girdilerin tek bir görevde bağlanması gerektiğini ve her sistemin her türle eşit derecede iyi baş edemediğini unutma. Ayrıca bir görüntüyü betimlemenin her ayrıntıyı doğru okumakla aynı şey olmadığını da unutma. Bu noktaları aklında tutarak çok kipli sistemleri daha güvenli anlayabilir ve kullanabilirsin.