# Yapay zekânın başarılı olduğu nasıl anlaşılır?



www.yzokul.com.tr

# Yapay zekânın başarılı olduğu nasıl anlaşılır?

www.yzokul.com.tr



Bir yapay zekâ sisteminin başarılı olup olmadığını bilmek istiyorsun. Kısa yanıt: başarı, sistemin yaptığı göreve uygun araçlarla ölçülür. Bir sistem, ne kadar zeki göründüğüyle değil, sonuçlarının yapmak üzere tasarlandığı şeye ne kadar iyi uyduğuyla değerlendirilir. Farklı görevler farklı ölçüm araçları gerektirir, tıpkı farklı işlerin farklı cetveller gerektirmesi gibi.



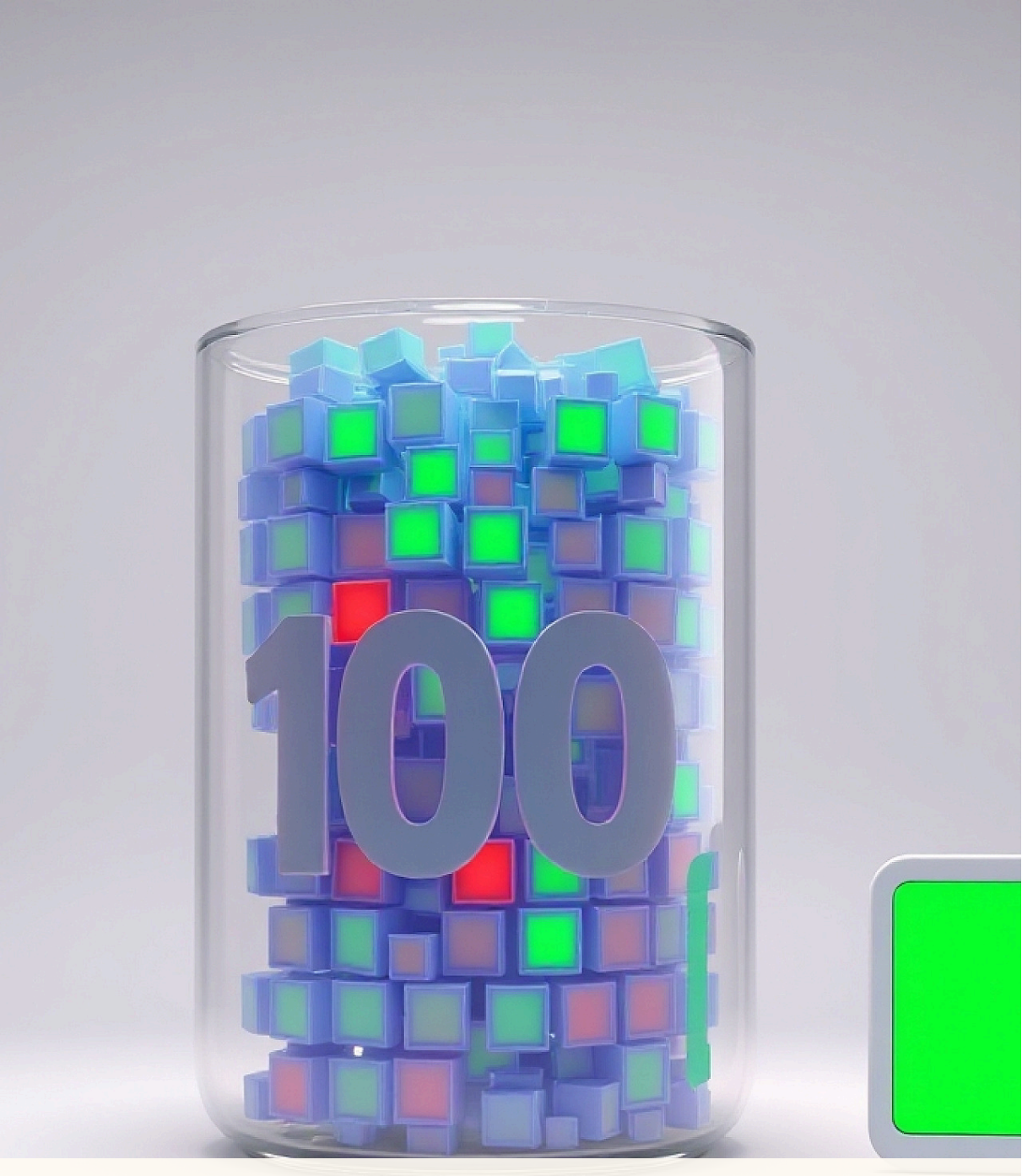
Burada başarı, sistemin işini ne kadar iyi yaptığı anlamına gelir. Sistemin ne kadar etkileyici ya da zeki göründüğü anlamına gelmez. Bir sistem, sonuçları ondan yapmasını istediğin şeyle uyuştuğunda başarılıdır. Doğru kategori etiketleri istediysen ve o doğru kategori etiketleri veriyorsa, iyi iş çıkarıyor demektir. Yakın sayı tahminleri istediysen ve o yakın sayı tahminleri veriyorsa, iyi iş çıkarıyor demektir.



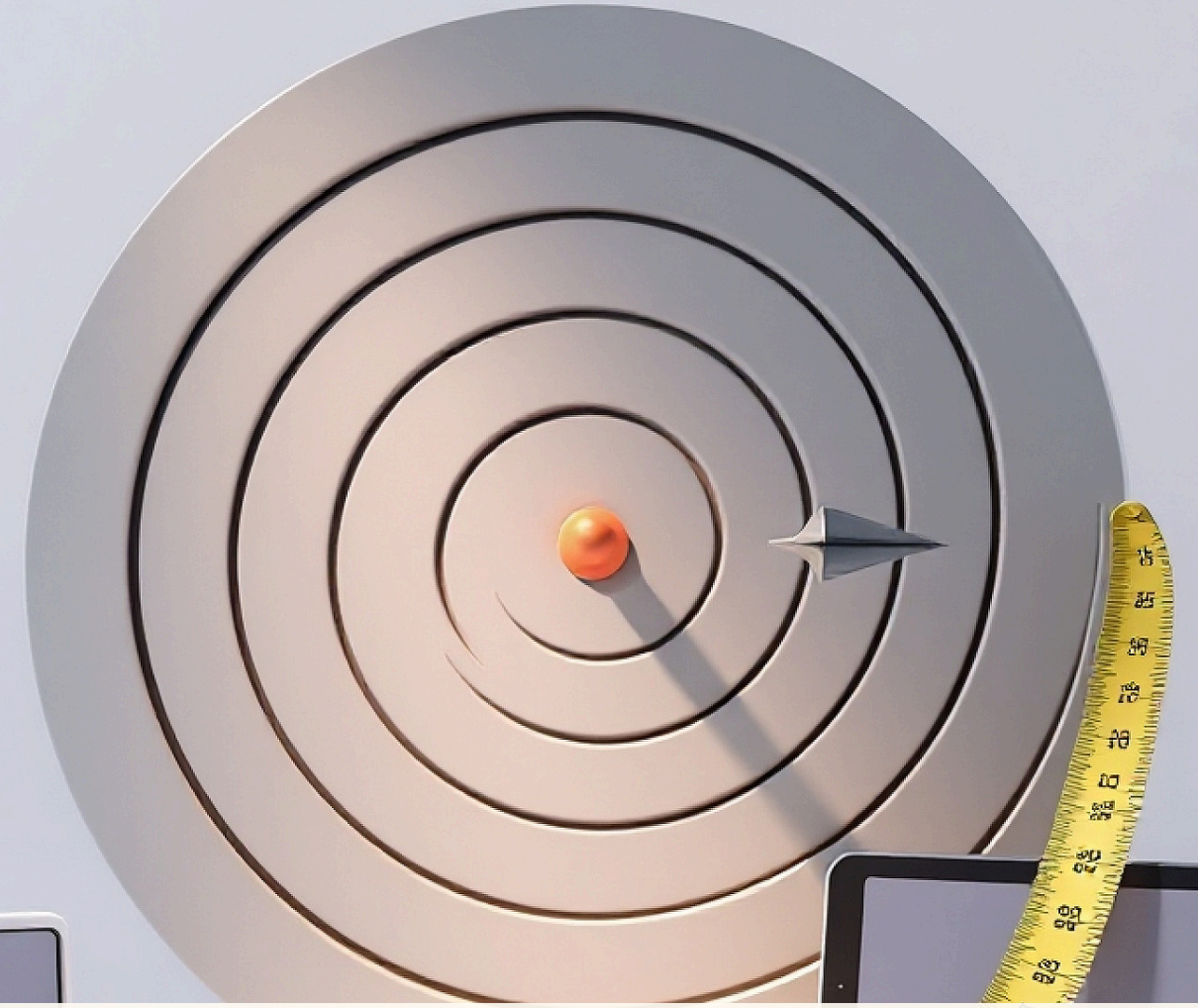
Ölçüyü göreve uydur. Önce sistemin ne tür bir görev yaptığını sor, çünkü her görev türünün kendi ölçüm aracı vardır. Bir sınıflandırma görevi ile bir tahmin görevi farklı şekillerde ölçülür. Yanlış aracı kullanmak yanıltıcı bir sonuç verir. Bu yüzden başarıyı değerlendirmedeki ilk adım, görev türünü açıkça adlandırmaktır.



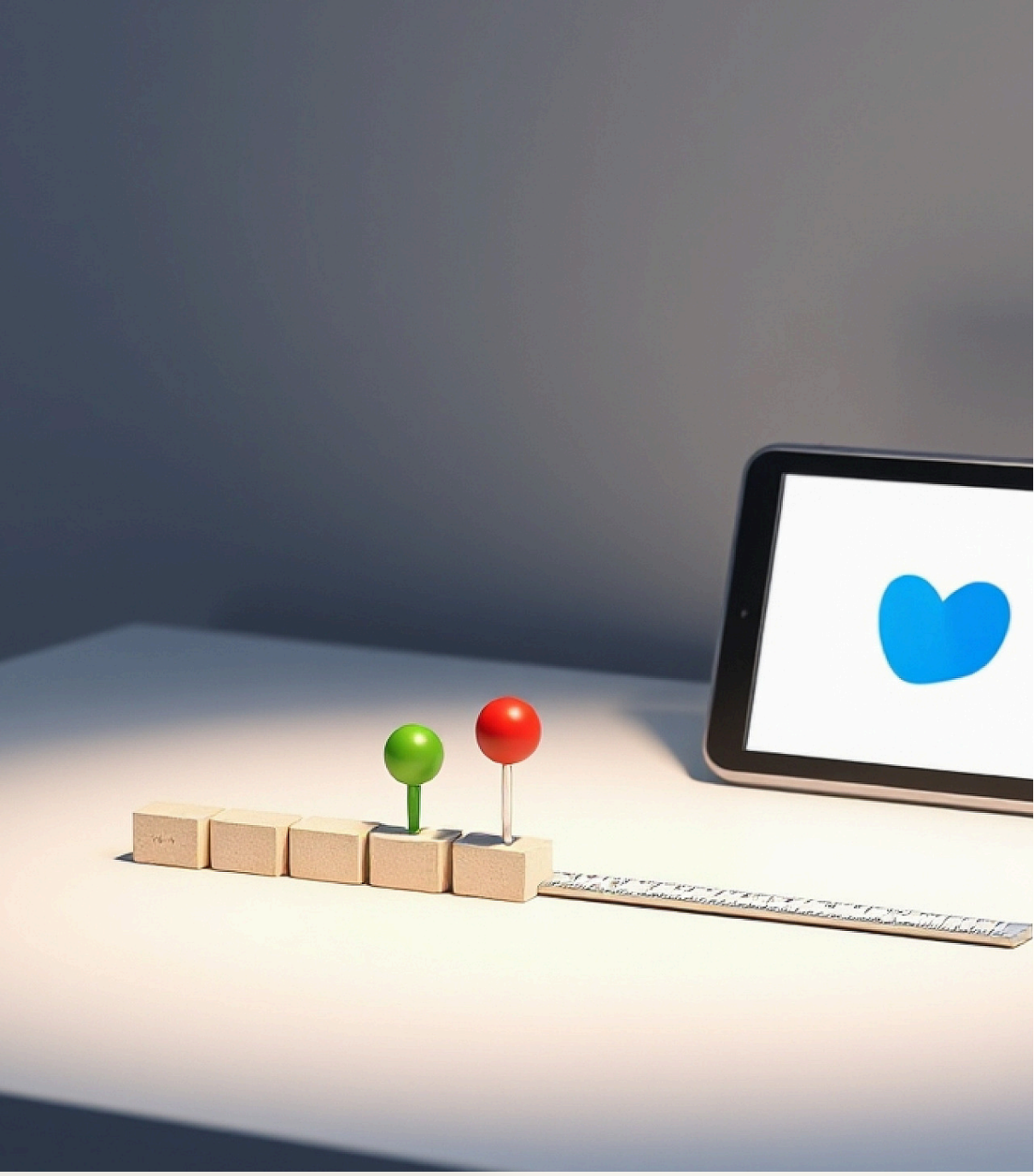
Bir sınıflandırma sisteminde sistem her örneği bir kategoriye yerleştirir. Örneğin, bir fotoğrafın kedi mi köpek mi gösterdiğine karar verir. Başarı, doğru olan kararların payı olarak ölçülür. Doğru kararları sayıp verilen tüm kararlara bölersin. Sonuç, sıfır ile bir arasında bir sayı ya da bir yüzdedir.




Küçük bir örnek üzerinden geç. Sistemin 100 karar verdiğini ve bunlardan 90'ının doğru olduğunu varsayalım. Başarı puanı, 100 toplam karara bölünen 90 doğru karardır, yani 0,90 ya da yüzde 90'dır. Puan, yalnızca doğru kararların tüm kararlara bölünmesidir. Bu basit doğru oranı, sınıflandırma görevleri için temel ölçüdür.



Bir tahmin sisteminde sistem yanıt olarak bir sayı verir, örneğin yarının sıcaklığı. Başarı, bu yanıtın gerçek sonuçtan ne kadar uzak olduğuyla ölçülür. Tahmini gerçek değerden ya da gerçek değeri tahminden çıkarır ve farkın büyüklüğüne bakarsın. Daha küçük uzaklıklar daha iyi tahminler demektir. Bu uzaklık ölçüsü, tahmin görevleri için temel araçtır.



Basit bir uzaklık örneğine bak. Gerçek değer 10 ise ve sistem 12 tahmin ediyorsa, yanıt 2 sapmıştır. 10 tahmin ediyorsa, yanıt 0 sapmıştır, bu kusursuzdur. 7 tahmin ediyorsa, 3 sapmıştır, bu da 2 sapmaktan daha kötüdür. Daha küçük uzaklıklar daha iyi tahminler demektir, bu yüzden başarıyı değerlendirmek için uzaklıkları karşılaştır.



Tek bir puan yeterli değildir. Tek bir sayı iyi görünürken gerçek sorunları gizleyebilir. Genel puan tüm örneklerin ortalamasını alır, bu yüzden birçok kolay örnekteki güçlü sonuçlar birkaç önemli örnekteki zayıf sonuçları örtebilir. Başarıyı doğru anlamak için genel puandan daha derine bakmalısın. Sonraki kontroller nereye bakacağını gösterir.



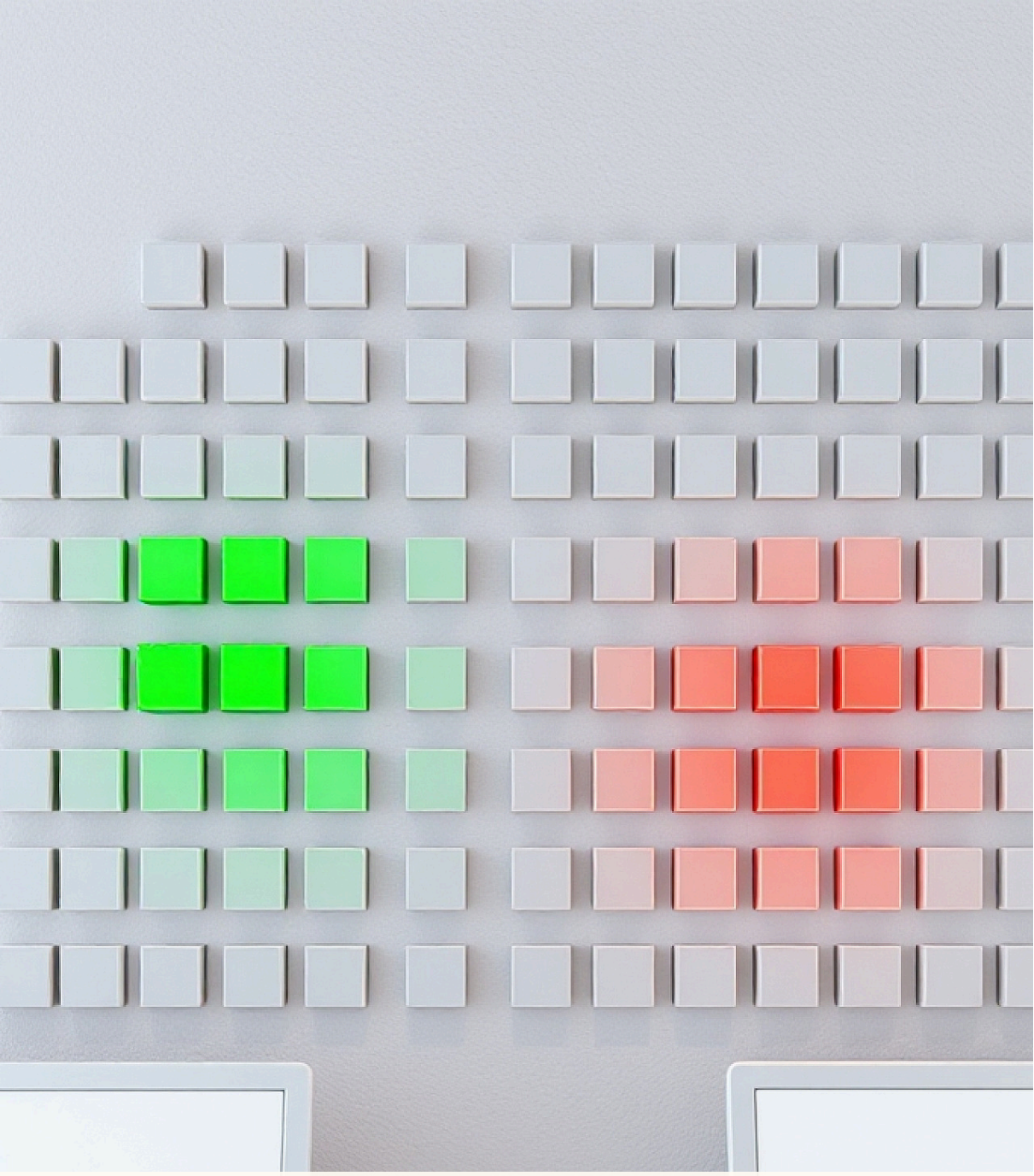
Nadir durum tuzağı yaygın bir gizli sorundur. Nadir bir durumu her zaman kaçıran ama yine de genel olarak yüksek puan alan bir sistem düşün. Yüksek puan alır çünkü yaygın örnekleri doğru yapar ve bunlar örneklerin çoğunluğudur. Yüksek puan, nadir durumdaki başarısızlığı gizler. Tek bir sayı bu zayıflığı ortaya çıkarmaz.



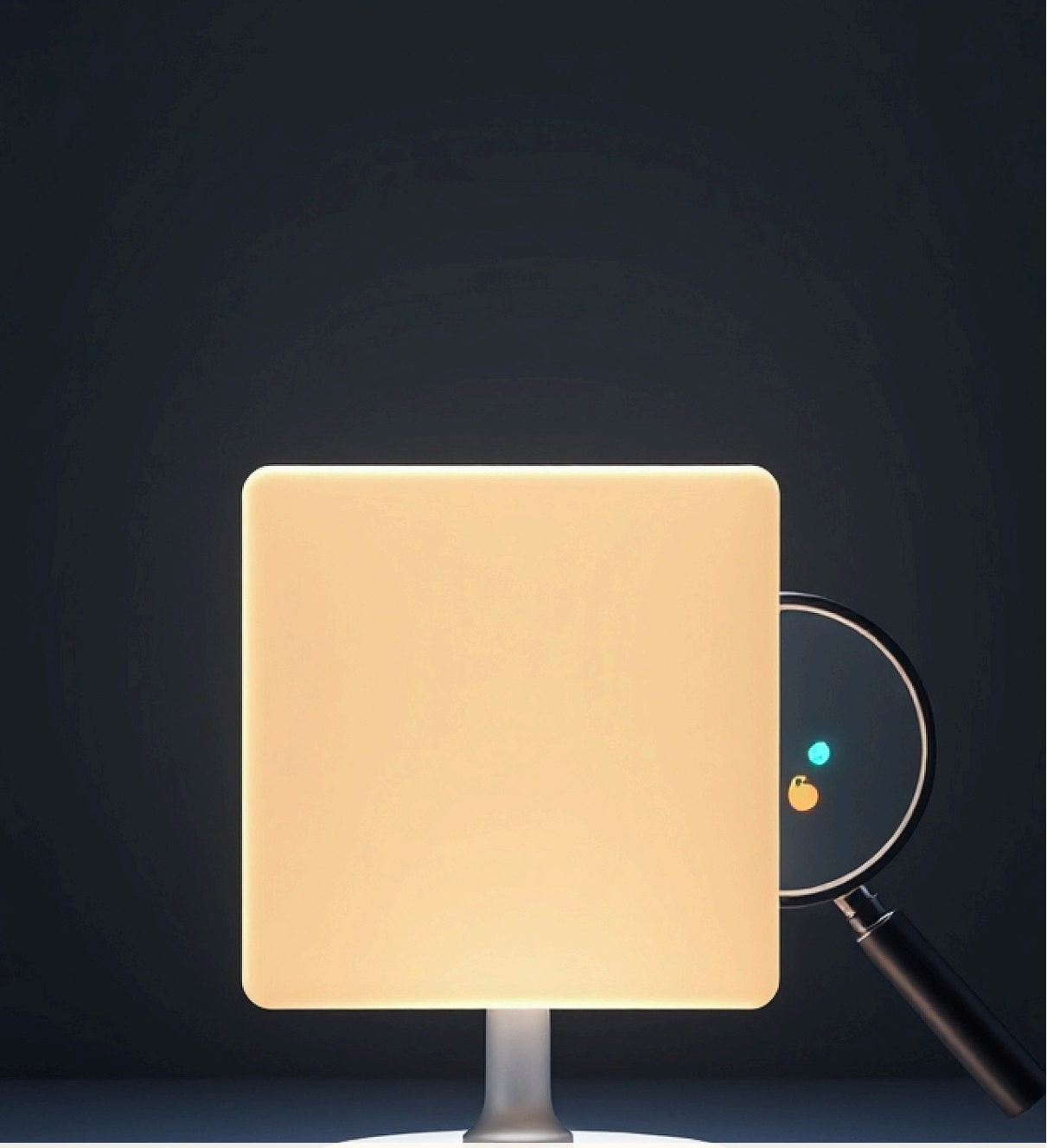
Nadir durumlar önemlidir çünkü çoğu zaman en çok önemsediğin durumlar bunlardır. Nadir ama ciddi bir durumu kaçıran bir sistem, genel puanı yüksek olsa bile gerçekten başarılı değildir. Örneğin, nadir bir hastalığı kaçırmak, yaygın bir soğuk algınlığını yanlış okumaktan daha kötüdür. İyi bir genel puan, nadir durumlardaki bir başarısızlığı telafi etmez.



Yalnızca hataları saymak yerine hata türlerine bak. Hataları türlere ayır, örneğin yanıt hayırken evet demek ya da yanıt evetken hayır demek gibi. Her hata türünün farklı bir maliyeti vardır. Hangi hata türlerinin oluştuğunu görmek, tek genel puanın neyi gizlediğini gösterir. Hata türleri tek bir sayıdan daha net bir tablo verir.



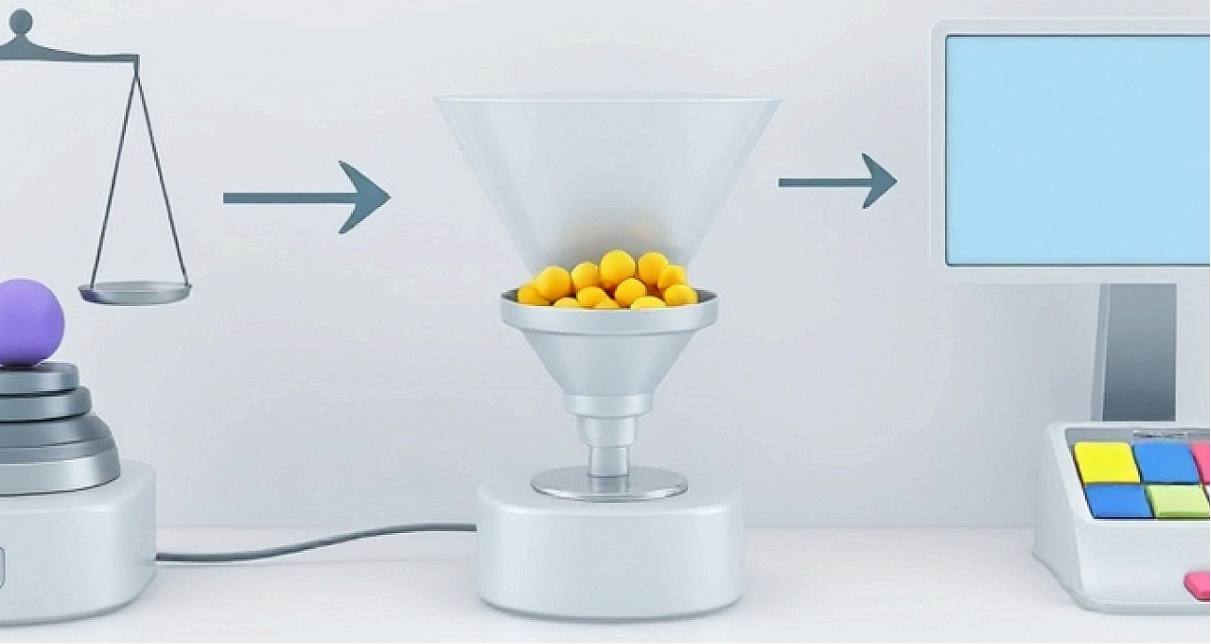
Örnek gruplarına bak. Örnekleri gruplara ayır, örneğin kategoriye ya da kaynağa göre, ve her gruptaki başarıyı ayrı ayrı kontrol et. Bir sistem bir grupta güçlü, başka bir grupta zayıf olabilir. Örneğin, gündüz fotoğraflarında iyi çalışırken gece fotoğraflarında kötü çalışabilir. Grup grup puanlar, genel puanın gizlediği bu farkları ortaya çıkarır.



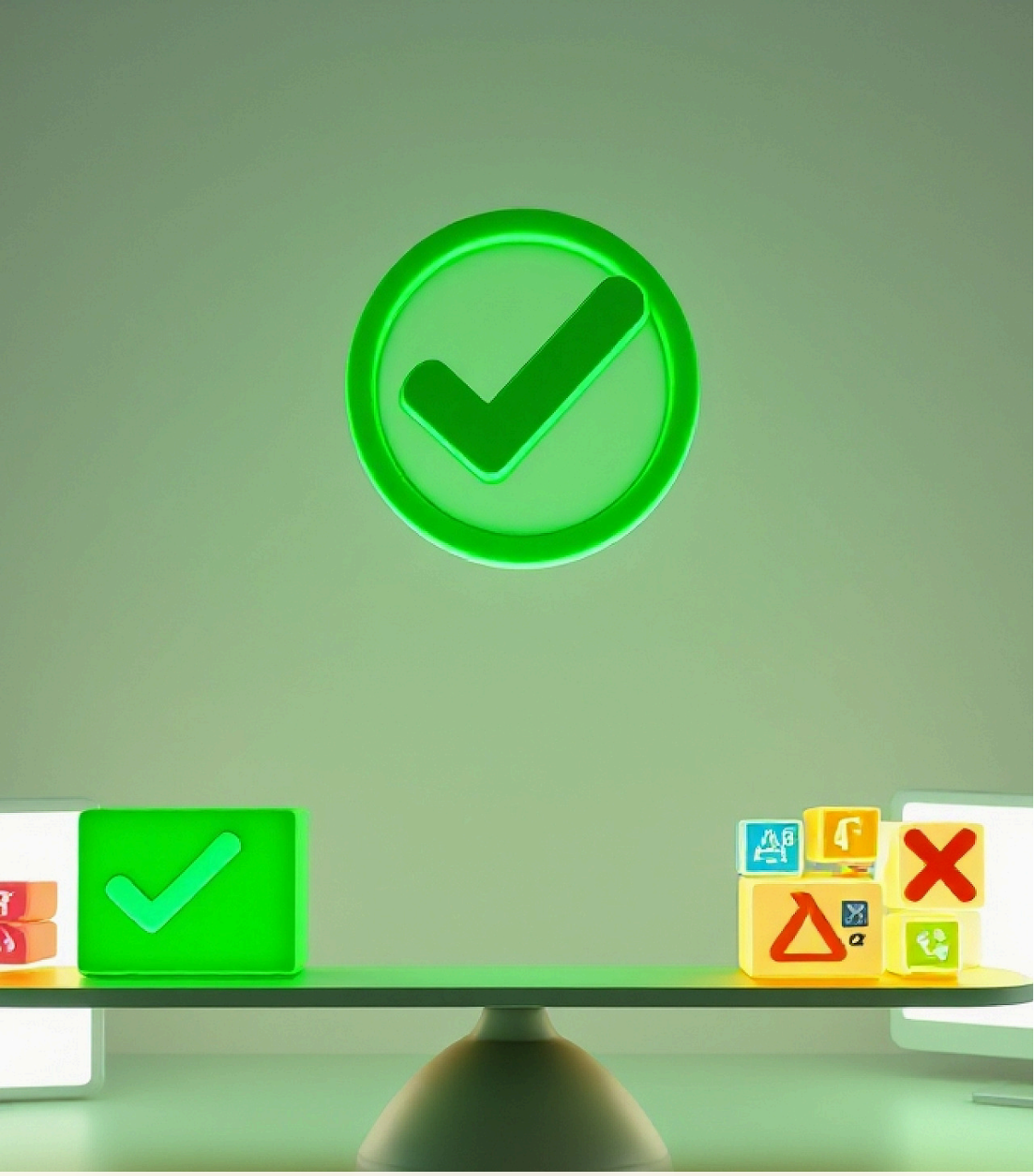
Yaygın bir hata, tek bir sayıya tek başına güvenmektir. İnsanlar çoğu zaman bir sistemi genel puanına göre değerlendirip orada durur. Çözüm, her zaman o tek sayının neyi dışarıda bıraktığını sormaktır. Hangi hata türlerinin oluştuğunu ve hangi örnek gruplarının zayıf olduğunu sor. Tek bir sayı bir başlangıç noktasıdır, yanıtın tamamı değildir. Başarı ilan etmeden önce her zaman daha derine bak.



Bir başka yaygın hata, görevi göz ardı etmektir. İnsanlar bazen yanlış ölçüyü kullanır, örneğin bir tahmin görevi için doğru kararları sayar. Yanıt bir sayı olduğunda doğru kararları saymak anlamlı değildir. Ölçü göreve uymalıdır. Sistemin ne tür bir yanıt verdiğini sor, sonra uygun ölçüm aracını seç.



Kontrolleri bir araya getir. Önce göreve uyan ölçüyü seç: sınıflandırma için doğru oranı, tahmin için uzaklık. Sonra hangi hataların oluştuğunu görmek için hata türlerini incele. Sonra sistemin nerede zayıf olduğunu görmek için örnek gruplarını incele. Bu üç kontrol birlikte, tek bir puanın tek başına verdiğinden daha tam bir başarı tablosu verir.



Akılda tutulacaklar: başarıyı sistemin yaptığı göreve uyan bir araçla ölç. Asla tek bir puana güvenme, çünkü o nadir durumlardaki başarısızlıkları ve zayıf grupları gizleyebilir. Her zaman hata türlerini ve farklı örnek gruplarını da kontrol et. Bir sistem, yalnızca ortalamada değil, her tür örnekte görevde iyi iş çıkardığında gerçekten başarılıdır.