Yapay zekâ ödüllerle nasıl öğrenir?



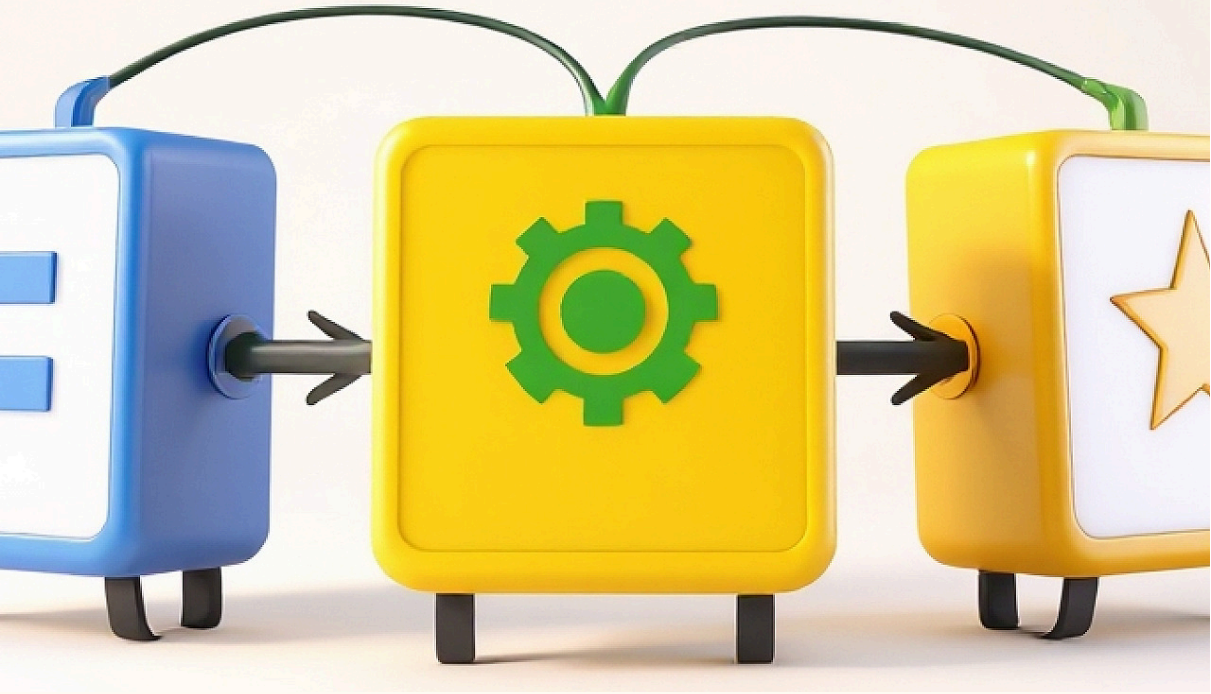
www.yzokul.com.tr

Yapay zekâ ödüllerle nasıl öğrenir?

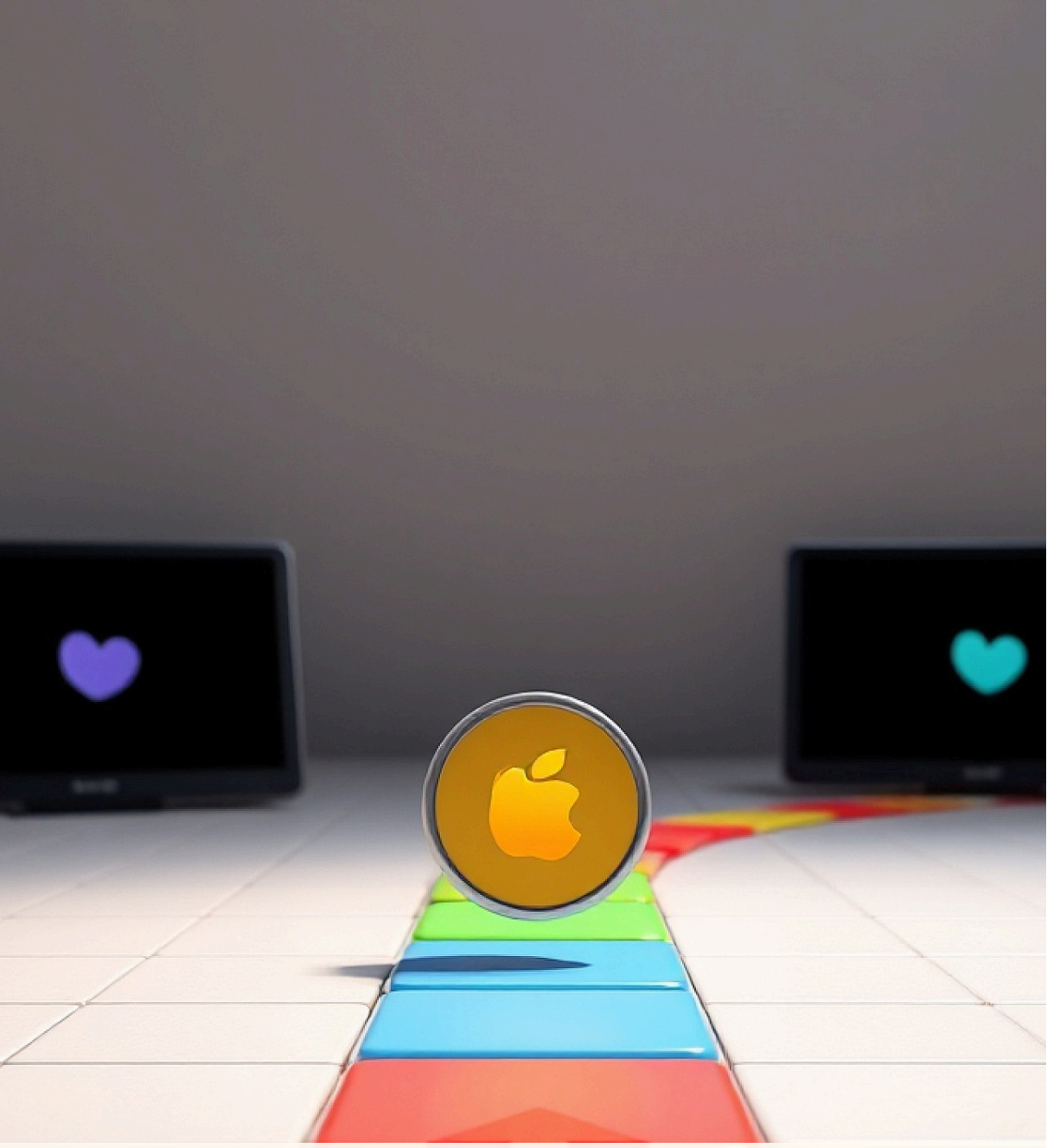
www.yzokul.com.tr



Yapay zekânın ödüllerle nasıl öğrendiğini bilmek istiyorsun. Kısa yanıt: Yapay zekâ eylemleri dener, sayısal ödül sinyalleri alır ve zaman içinde daha fazla ödül kazanmak için davranışını ayarlar. Bu, bir sistemin açık talimatlar yerine geri bildirimden öğrendiği bir yöntem olan pekiştirmeli öğrenmenin özüdür. Bu kitap her parçayı sıfırdan adım adım açıklar.



Pekiştirmeli öğrenme üç temel parçadan oluşur: bir etmen, bir ortam ve ödüller. Etmen, öğrenen veya karar verendir. Ortam, etmenin etkileşimde bulunduğu her şeydir. Ödüller, sayısal geri bildirim sinyalleridir. Oklarla birbirine bağlanmış üç kutu düşün: etmen ortama eylemler gönderir, ortam yeni durumlar ve ödüller gönderir. Bu üç parça, yapay zekânın ödüllerle nasıl öğrendiğinin temelini oluşturur.



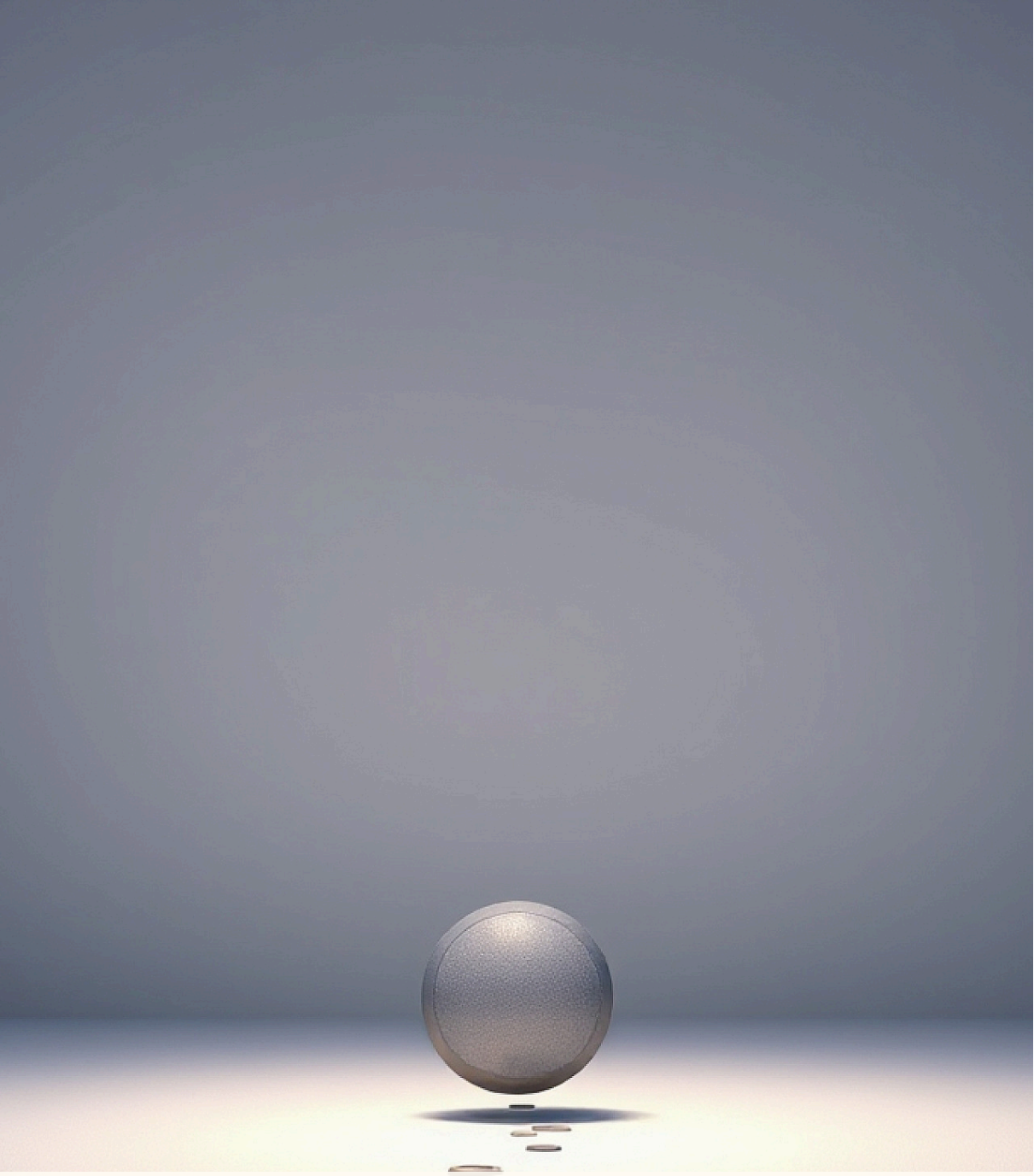
Etmen bir eylemde bulunduğunda, ortam durumunu değiştirir. Durum, mevcut durumu ifade eder; örneğin etmenin bir labirentteki konumu. Örneğin sağa gitmek yeni bir hücreye yol açabilir. Bu yeni durumdan farklı eylemler mümkün hale gelir. Bu eylem ve durum değişimi döngüsü, etmenin keşfetme ve öğrenme biçimidir. Eylemleri ve durumları anlamak, pekiştirmeli öğrenmeyi kavramanın anahtarıdır.



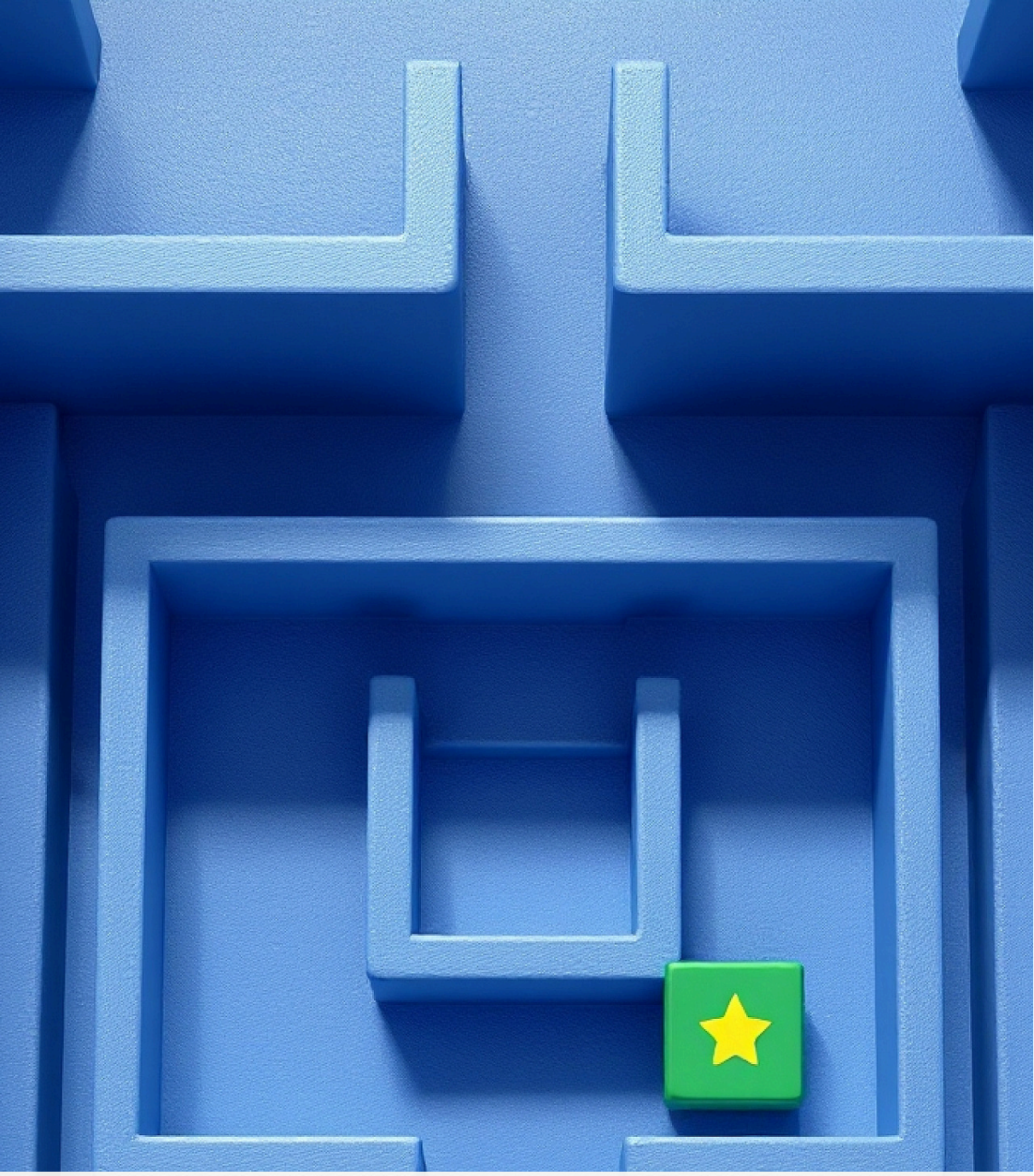
Ödül, bir eylemden sonra ortamın döndürdüğü basit bir sayıdır; örneğin +1 veya -5. Bu bir mutluluk duygusu değil, bir puandır. Etmen bu sayıyı eylemin ne kadar iyi veya kötü olduğunu değerlendirmek için kullanır. Olumlu ödüller eylemi tekrarlamayı teşvik eder, olumsuz ödüller ise ondan caydırır. Ödüller, etmenin aldığı tek geri bildirimdir, bu yüzden açık ve tutarlı olmalıdır.



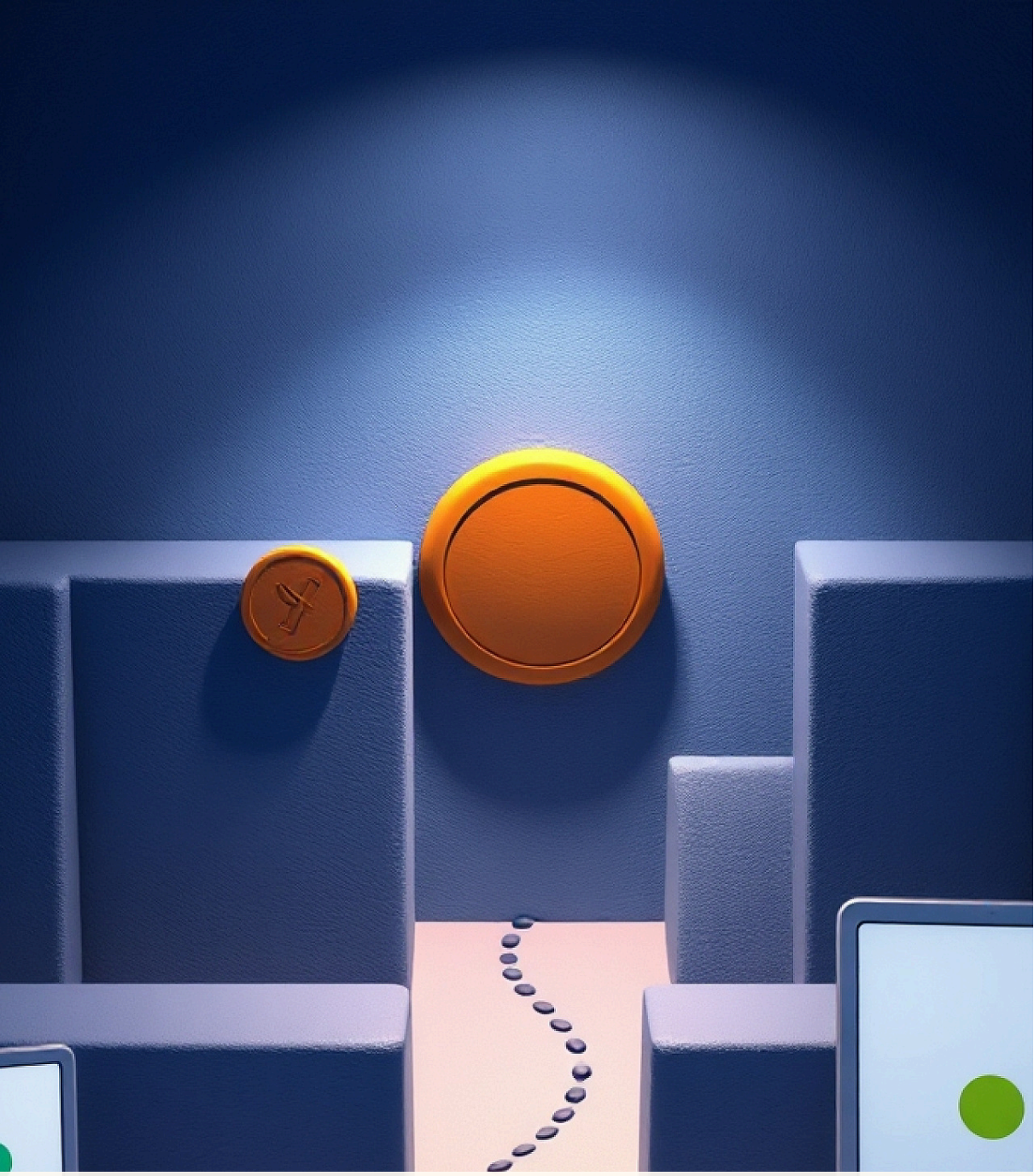
Etmenin amacı, yalnızca anlık ödülü değil, zaman içindeki toplam ödülü en üst düzeye çıkarmaktır. Toplam ödül, şu andan geleceğe kadar toplanan tüm ödüllerin toplamıdır. Şimdiki küçük bir ödül, sonraki daha büyük bir ödülden daha az değerli olabilir. Etmen kısa vadeli ve uzun vadeli kazançları dengelemeyi öğrenir. Bu hedef, pekiştirmeli öğrenmedeki tüm öğrenmeyi yönlendirir.



Başlangıçta etmen biraz rastgele davranır, hangi ödülleri getirdiklerini görmek için farklı eylemler dener. Bu, deneme yanılmadır. Sonuçları gözlemleyerek etmen hangi eylemlerin daha yüksek ödüllere yol açtığını öğrenir. Birçok denemede seçimlerini yavaş yavaş iyileştirir. Bu basit deneme ve gözlem döngüsü, öğrenme motorunun tamamıdır.



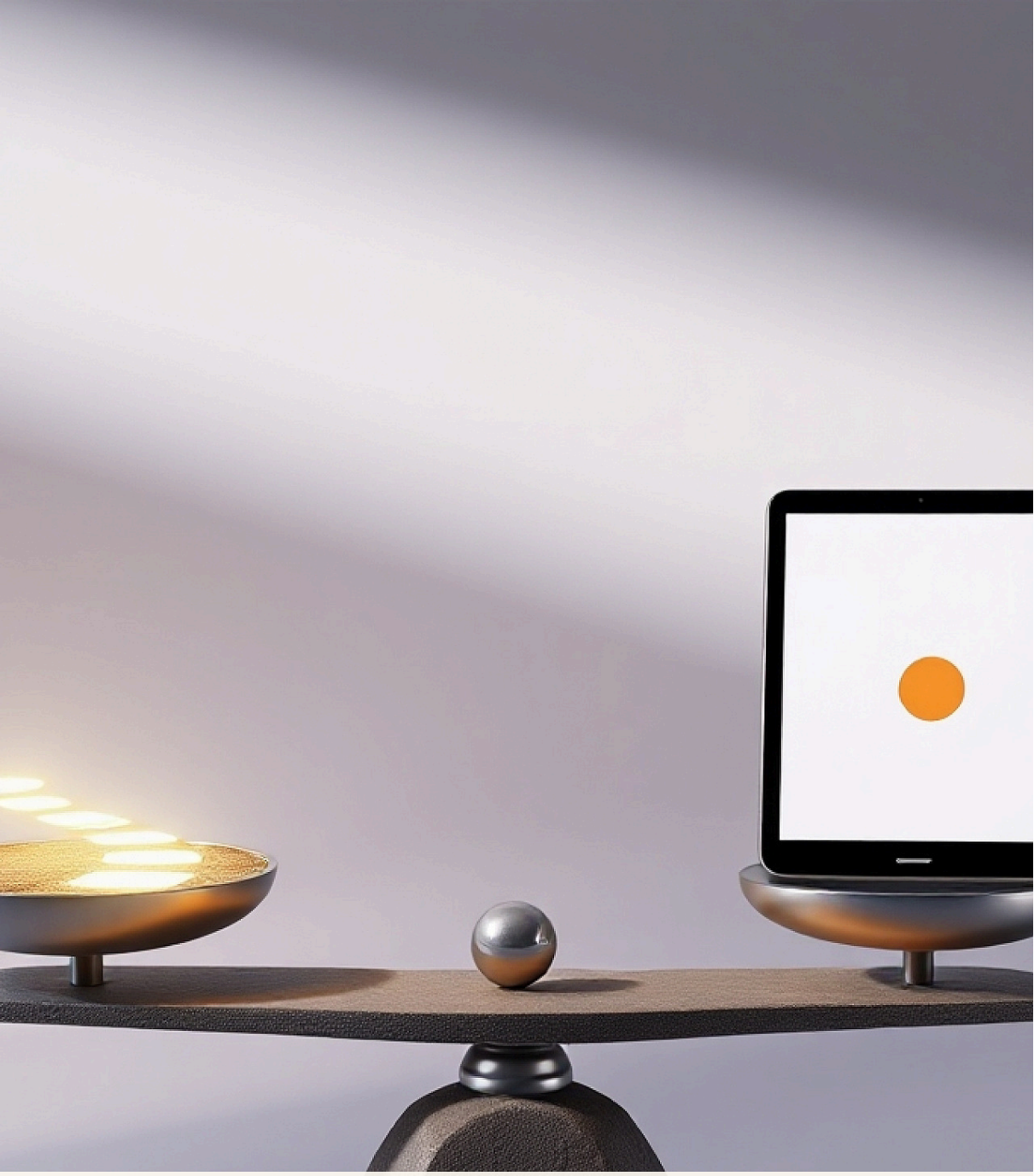
Etmenin yukarı, aşağı, sola veya sağa hareket edebildiği basit bir ızgara labirent düşün. Amaç, ödül veren çıkışa ulaşmaktır. Her adım, hızı teşvik etmek için küçük bir olumsuz ödül verebilir. Etmen hiçbir bilgi olmadan başlar ve en iyi yolu öğrenmelidir. Bu labirent örneği kitabın geri kalanını açıklayacaktır.



Başlangıçta etmen rastgele dolaşır, duvarlara çarpar ve çıkmaz sokaklara girer. Az ödül toplar ya da hiç toplamaz, çoğu zaman boşa harcanan hamleler için olumsuz ödüller alır. Bu beceriksiz aşama normal ve beklenen bir durumdur. Etmen ortam hakkında bilgiyi bu şekilde toplar. Bu tür birçok denemeyle örüntüleri fark etmeye başlar.



Etmen hangi durumlardaki hangi eylemlerin iyi ödüller getirdiğini takip eder. Her eylemin ne kadar iyi olduğunu tahmin eden bir durum-eylem değerleri haritası oluşturur. Birçok denemede, tahmini değerleri daha yüksek olan eylemlere yönelir. Geçmiş deneyimlerin bu belleği, gelecekteki kararları yönlendirir. Yavaş yavaş etmenin davranışı daha amaçlı hale gelir.



Etmen yeni eylemler denemek (keşif) ile bilinen iyi eylemleri tekrarlamak (sömürü) arasında denge kurmalıdır. Çok fazla keşif zaman kaybettirir, çok fazla sömürü ise daha iyi fırsatları kaçırır. Yaygın bir strateji, başlangıçta daha fazla keşfetmek ve sonra daha fazla sömürmektir. Doğru dengeyi bulmak, verimli öğrenme için çok önemlidir.



Hangi davranış ödüllendirilirse, bize tuhaf görünse bile tekrarlanma eğilimindedir. Etmenin öğrendiklerini insan niyetleri değil, ödül sinyali şekillendirir. Bir eylem yüksek ödül getiriyorsa, etmen onu daha sık yapar. Bu ilke güçlüdür ama aynı zamanda risklidir. İstenmeyen davranışlardan kaçınmak için dikkatli ödül tasarımı şarttır.



Kötü tasarlanmış bir ödül, etmeni istenmeyen davranışların peşinden gitmeye yönlendirebilir. Örneğin, bir duvara çarpmak puan veriyorsa, etmen kaçmak yerine sonsuza kadar duvarlara çarpabilir. Puan kazanır ama labirenti asla çözmez. Etmen arızalı değildir; tam olarak sayıların istediğini yapıyordur. Bu, ödülleri gerçek hedeflerimizle uyumlu hale getirmenin önemini vurgular.



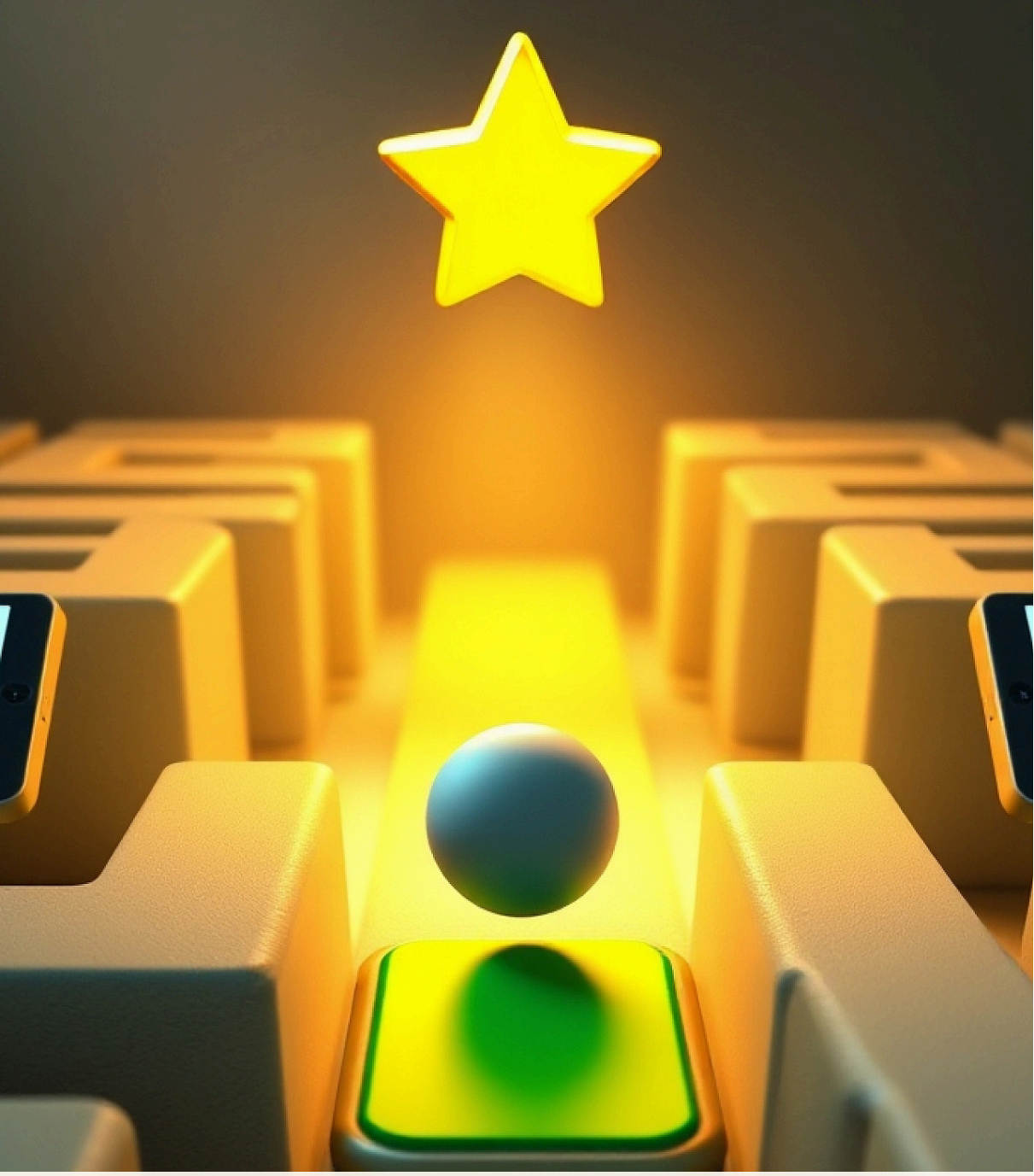
Kötü tasarlanmış bir ödül, sistemi istenmeyen ama puan kazandıran yollara itebilir. Etmen, ödülü verildiği gibi en iyiler; bizim amaçladığımızı değil. Ödül fonksiyonunu kötüye kullanan zekice kısa yollar bulabilir. Buna ödül hackleme denir. Etmen bozuk değildir; tam olarak sayıların istediğini yapıyordur. Hedeflerimizi gerçekten yansıtan ödüller tasarlamalıyız.



Ödül tasarımı, pekiştirmeli öğrenmede insanın ana işidir. Sayılar, etmenin öğrenmesini gerçekten istediğimiz şeyle eşleşmelidir. Dikkatsiz bir puan yanlış dersi öğretir. Hangi davranışları ödüllendirip cezalandırdığımızı dikkatle düşünmeliyiz. İyi ödül tasarımı iyi öğrenmeye, kötü tasarım kötü davranışa yol açar.



Yeni başlayanlar sık sık yaygın hatalar yapar. Ödüllerin duygular olduğunu sanabilirler, oysa onlar sadece sayılardır. Anında başarı bekleyebilirler, oysa öğrenme birçok deneme gerektirir. Etmenin, bizim amaçladığımızı değil, verildiği gibi puanı en iyilediğini unutabilirler. Bu hatalardan kaçınmak, pekiştirmeli öğrenmeyi doğru anlamana yardımcı olur.



Özetle: bir etmen bir ortamda eylemde bulunur, sayısal ödüller alır ve zaman içinde toplam ödülünü artırmayı öğrenir. Ödül tasarımı davranışı belirler, bu yüzden dikkatli sayılar dikkatli öğrenmeye yol açar. Ödüllerin duygu değil puan olduğunu ve etmenin tam olarak sayıların istediğini yapacağını unutma. İstenen sonuçlara ulaşmak için ödülleri akıllıca tasarla.