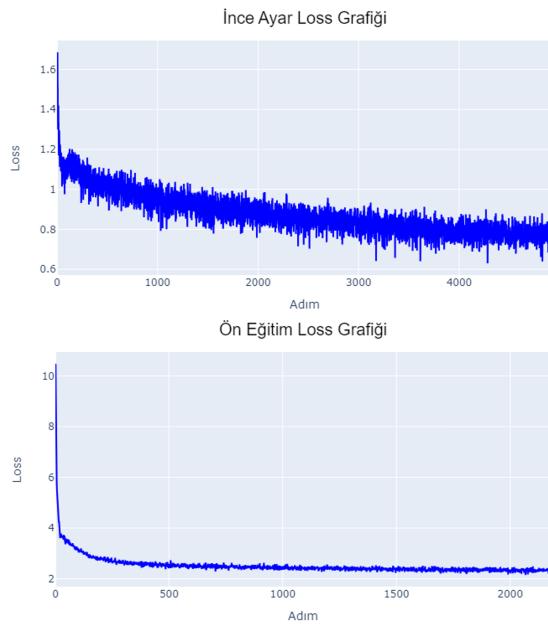


TABLO I: İnce ayar veri kümelerinden örnekler

GQA	CosmosVQA	VG	COCO	CC3M
<p>Kullanıcı: Gökyüzü karanlık mı?</p> <p>Model: Evet</p> <p>Kullanıcı: Sokakta ne var?</p> <p>Model: Araba</p> <p>Kullanıcı: Caddede otobüs var mı?</p> <p>Model: Hayır</p> <p>Kullanıcı: Trafik sinyali sağda mı yoksa solda mı?</p> <p>Model: Sol</p> 	<p>Kullanıcı: Kitabin ismini öğrenebilir miyim?</p> <p>Model: Masumiyet Müzesi</p> <p>Kullanıcı: Bu kitap hangi türde bir kitaptır?</p> <p>Model: Aşk romanı</p> <p>Kullanıcı: Bu kitabın yazarı kimdir?</p> <p>Model: Orhan Pamuk</p> 	<p>Kullanıcı: Lütfen bu bölge için kısa bir açıklama yapınız: [0.69, 0.35, 0.79, 0.69].</p> <p>Model: Bu bir insan.</p> <p>Kullanıcı: Lütfen bu cümplenin tanımladığı bölgenin sınırlayıcı kutu koordinatını verin: tenis kortundaki beyaz çizgi.</p> <p>Model: [0.0, 0.7, 0.9, 0.82]</p> 	<p>Kullanıcı: Lütfen bu bölge için kısa bir açıklama yapınız: [0.3, 0.36, 0.62, 0.59].</p> <p>Model: Fotoğrafçıya bakan bir bilgisayar ekranı.</p> <p>Kullanıcı: Lütfen bu cümplenin tanımladığı bölgenin sınırlayıcı kutu koordinatını verin: en sağdaki ekran.</p> <p>Model: [0.3, 0.36, 0.62, 0.59]</p> 	<p>Kullanıcı: Sunulan görüntüyü temsil eden kompakt bir anlatı oluşturun.</p> <p>Model: çiftler için toplu nikah töreni düzenlendi.</p> 



Şekil 1: Eğitimlere Ait Loss Grafikleri

A. Hakem: Dil Modeli

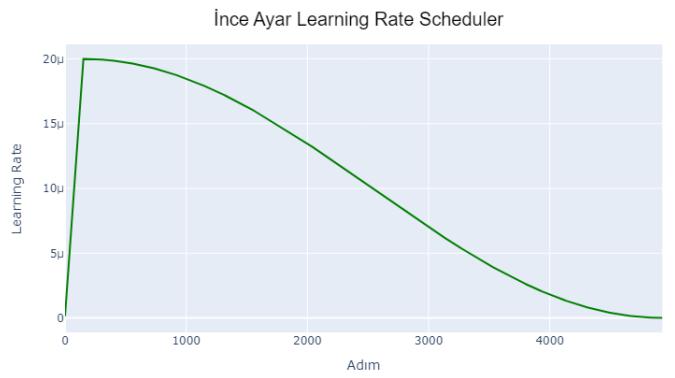
Bu çalışmanın yapıldığı Ağustos 2024 tarihinde çoğu ölçüme göre en iyi dil modeli olan OpenAi'nın GPT-4o [15] modeli hakem olarak kullanılmıştır. Sorular, resimler ve modellerin verdikleri cevaplar hakem olarak belirlenen GPT-4o modeline verilerek modelin yanıtına kadar doğru olduğuna göre 100 üzerinden bir puan vermesi istenmiştir. GPT-4o'ya puanlattırmak için kullanılan talimat formatı aşağıdaki gibidir:

{görsel}

Soru: {soru}

Cevap: {model-cevabı}

Bir görsel, görselle ilgili bir soru ve verilen cevap gösteril-



Şekil 2: İnce Ayar Eğitimde Learning Rate'in Değişimi

miştir. Çok iyi kaliteli cevaplara 100, iyilere 80, ortalamalara 30, kötülere 0 puan ver. Sadece puanı tam sayı olarak yaz, hiçbir açıklama yapma.

Ardından GPT-4o'nun verdiği yanıt ile modelin o soruya ne kadar iyi yanıt verebildiği 100 üzerinden bir puan ile ölçülümüştür. GPT-4o modelinin yanıtları deterministik olmaması ve aynı cevaba farklı puanlar verilebilmesi nedeniyle her soru için 5 kez değerlendirme yapılarak bu değerlendirmelerde verilen puanların ortalaması alınmıştır.

Bu puanlama için kullanılan test kümesi 4 alt kümeden oluşmaktadır: OCR, Kompleks, Tanımlama ve Detay. Bu kümelerin içeriği aşağıda açıklanmıştır:

- **OCR:** Toplam 10 kitap içeren bu veri kümesi, her kitap için gerçek hayatı çekilmiş bir adet fotoğraf ile bir dijital kapak görselini içermektedir. Veri kümesi, kitabı kapağındaki yer alan yazıların ne olduğunu sorulayan 20 sorudan oluşmaktadır.
- **Kompleks:** Bu kategori, modelin görsel içeriklere dayanarak üst düzeyde akıl yürütme ve çıkarımlar