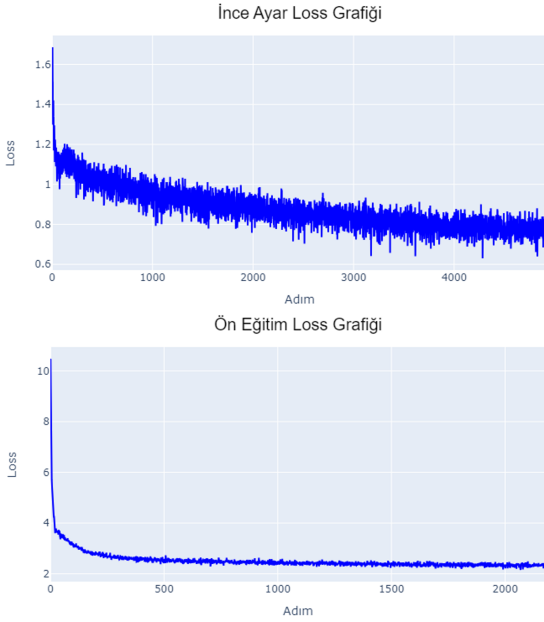


TABLO I: İnce ayar veri kümesinden örnekler

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



Şekil 1: Eğitimlere Ait Loss Grafikleri

A. Hakem: Dil Modeli

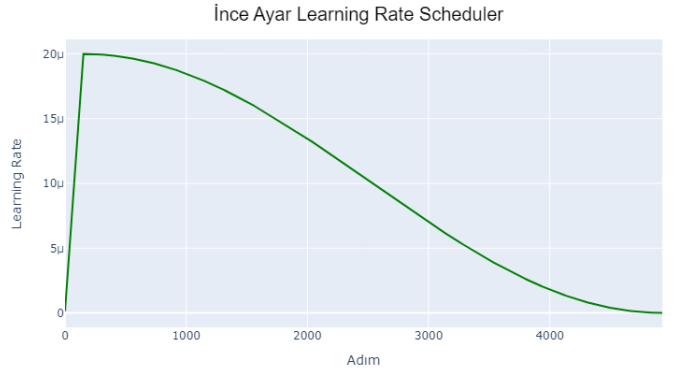
Bu çalışmanın yapıldığı Ağustos 2024 tarihinde çoğu ölçüme göre en iyi dil modeli olan OpenAi'nin GPT-4o [15] modeli hakem olarak kullanılmıştır. Sorular, resimler ve modellerin verdikleri cevaplar hakem olarak belirlenen GPT-4o modeline verilerek modelin yanıtı ne kadar doğru olduğuna göre 100 üzerinden 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 ölçülmüştür. 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 hayatta çekilmiş bir adet fotoğraf ile bir dijital kapak görselini içermektedir. Veri kümesi, kitabın kapağında yer alan yazıların ne olduğunu sorgulayan 20 sorudan oluşmaktadır.
- **Kompleks:** Bu kategori, modelin görsel içeriklere dayanarak üst düzeyde akıl yürütme ve çıkarımlar