TÜBİTAK 2209-A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEKLEME PROGRAMI

ARAŞTIRMA ÖNERİSİ FORMU

2025 Yılı

1. Dönem Başvurusu

2209-A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEKLEME PROGRAMI ARAŞTIRMA ÖNERİSİ FORMU

A. GENEL BİLGİLER

Başvuru Sahibinin Adı Soyadı: Buğra Kaan Kesmez
Araştırma Önerisinin Başlığı: İlişkisel Veritabanları İçin QLoRA ile Özelleştirilmiş LLM Tabanlı Doğal Dil Arayüzü Geliştirilmesi
Danışmanın Adı Soyadı: Gözde YOLCU ÖZTEL
Araştırmanın Yürütüleceği Kurum/Kuruluş: Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Yazılım Mühendisliği Bölümü

ÖZET

Araştırma önerisi özetinin (1) bilimsel nitelik; (2) yöntem; (3) proje yönetimi ve (4) yaygın etki hakkında bilgileri kapsaması beklenir. Bu bölümün en son yazılması önerilir.

Özet

Günümüz dijital çağında, kurumların ürettiği "Büyük Veri"nin karar alma süreçlerine dahil edilmesi kritik bir zorunluluktur. Ancak verilerin saklandığı ilişkisel veritabanlarına erişim, karmaşık SQL (Yapılandırılmış Sorgu Dili) yetkinliği gerektirmektedir. Bu durum, teknik bilgisi olmayan yöneticiler ve araştırmacılar ile veri arasında bir erişim zorluğu oluşturmaktadır. Literatürdeki genel amaçlı Büyük Dil Modelleri (LLM), karmaşık şemalarda yetersiz kalmakta ve "halüsinasyon" (yanlış tablo uydurma) sorunu yaşamaktadır. Bu araştırma, söz konusu erişim engelini aşmak amacıyla, doğal dil sorgularını yüksek doğrulukla SQL'e dönüştüren, alana özel ve maliyet etkin bir yapay zeka modeli geliştirmeyi hedeflemektedir.

Projenin bilimsel metodolojisi, literatürün en zorlu kıyaslama seti olan "Spider" veri seti (10.181 soru, 200 veritabanı) üzerine kuruludur. Veritabanı şemaları, modelin bağlamı anlaması için "Schema Serialization" yöntemiyle yapılandırılacaktır. Eğitim aşamasında, yüksek donanım maliyetlerini düşürmek ve performansı artırmak amacıyla, açık kaynaklı Mistral-7B mimarisi üzerinde QLoRA (Quantized Low-Rank Adaptation) tekniği ile parametre verimli ince ayar (PEFT) uygulanacaktır. Geliştirilen modelin başarısı, literatür standartları olan "Birebir Eşleşme (Exact Match)" ve semantik doğruluğu ölçen "Çalıştırma Başarısı (Execution Accuracy)" metrikleriyle test edilecek; hedef olarak sırasıyla %65+ ve %80+ doğruluk oranı belirlenmiştir.

Proje yönetimi; veri ön işleme, model eğitimi, Python tabanlı (Streamlit/Gradio) web arayüzü geliştirme ve pilot kullanıcı testleri olmak üzere dört ana iş paketi halinde, 6 aylık bir takvimde yürütülecektir. Olası risklere karşı (donanım yetersizliği, düşük doğruluk) B planları oluşturulmuştur. Çalışmanın yaygın etkisi; KOBİ ve STK'lardaki teknik olmayan personelin veriye erişimini daha kolay hale getirerek kurumsal verimliliği artırma için bir ön çalışma sunmaktır. Ayrıca 12. Kalkınma Planı'nın "Veriye Dayalı Politika" hedeflerine katkı sağlamaktır. Elde edilecek prototip, gelecekteki sanayi odaklı (2209-B) projeler için bir temek olacak ve uluslararası bir bildiri ile akademik literatüre katkı sunacaktır.

Anahtar Kelimeler: Büyük Dil Modelleri, Doğal Dil İşleme, Text-to-SQL, QLoRA, Veritabanı Yönetimi

1. ARAŞTIRMA ÖNERİSİNİN BİLİMSEL NİTELİĞİ

1.1. Konunun Önemi ve Araştırma Önerisinin Bilimsel Niteliği

Araştırma önerisinde ele alınan konunun kapsamı, sınırları ve önemi ortaya konulur. Araştırma önerisi kapsamında yapılacak çalışmalarla literatürdeki hangi eksikliğin nasıl giderileceği veya hangi soruna nasıl bir çözüm getirileceği ilgili literatüre atıfla açıklanarak araştırma önerisinin bilimsel niteliği ortaya konulur. Araştırma sorusu ve varsa hipotez(ler)i tanımlanır.
Projenin konusu, 12. Kalkınma Planı ve 2030 Sanayi ve Teknoloji Stratejisi'nde yer alan kritik teknoloji alanları ile öncelikli Ar-Ge ve yenilik konuları ile ilişkili ise, ilişkilendirilme sebebi ve ilgili alana sağlayacağı yararlar açıklanmalıdır.

Günümüz dijital çağında, kurumların sahip olduğu veri hacmi katlanarak artmaktadır [1]. Bu verinin (Büyük Veri) analiz edilmesi ve karar alma süreçlerine dahil edilmesi, rekabetçi avantaj elde etmek için kritik bir zorunluluktur [2]. Ancak, bu veriler çoğunlukla ilişkisel veritabanlarında saklanmakta ve bu verilere erişim, karmaşık SQL (Yapılandırılmış Sorgu Dili) sorgularını yazabilme yetkinliğini gerektirmektedir. Bu durum, teknik bilgisi olmayan

(non-technical) yöneticiler, iş analistleri ve araştırmacılar ile değerli veri arasında bir erişim zorluğu oluşturmakta [3], bu da kurumsal verimliliği ve anlık karar alma kabiliyetini kısıtlamaktadır [4].

Bu araştırmanın kapsamı, Büyük Dil Modelleri (LLM) kullanarak bu teknik engeli ortadan kaldırmaktır. Bu modellerin temelini oluşturan Transformer mimarisi [5], doğal dil işleme alanında devrim yaratmıştır. Projenin hedefi, literatürde kapsamlıca incelenen [6] gibi çalışmalarla, belirli ve karmaşık bir veritabanı şeması üzerinde yüksek doğrulukla çalışacak, alana özel (domain-specific) bir "Doğal Dil'den SQL'e" (Text-to-SQL) çevirici model geliştirmektir.

Literatürdeki mevcut eksiklik, genel amaçlı LLM'lerin (örn: GPT-4, Llama-3 [7]) sıfır-önek (zero-shot) senaryolarda dahi karmaşık veritabanı şemalarını (çoklu tablo ilişkileri, özel kolon adları) anlamakta zorlanması [8], yanlış sorgular üretmesi ("halüsinasyon" [9]) veya verimsiz sorgular oluşturmasıdır. Literatür, bu genel modellerin, özel veritabanları üzerinde yüksek doğruluk (%80+) sağlamak için "ince ayar" (fine-tuning) veya "bağlam içi öğrenme" (in-context learning) gibi özel adaptasyon tekniklerine ihtiyaç duyduğunu göstermektedir [10]. (örn: "Spider" benchmark çalışmaları [11]). Bu proje, açık kaynaklı bir LLM'in, bu spesifik "Text-to-SQL" görevi için ince ayar tekniği ile eğitilmesinin, doğruluğu ne ölçüde artıracağını ve bu teknik engeli nasıl aşacağını araştırarak literatürdeki bu boşluğu doldurmayı hedeflemektedir.

Projenin temel araştırma sorusu şudur: "Açık kaynaklı, parametre-verimli bir Büyük Dil Modeli (örn: Mistral-7B [12]), karmaşık ve çok tablolu bir veritabanı (örn: Spider veri seti) üzerinde ince ayar (fine-tuning) (LoRA [13] / QLoRA [14]) yapılarak, son kullanıcıların doğal dildeki sorularını %80 ve üzeri 'Çalıştırma Başarısı (Execution Accuracy)' ile SQL sorgularına dönüştürebilir mi?"

Bu proje, 12. Kalkınma Planı'nın "Dijital Dönüşüm" ve "Veriye Dayalı Politika Geliştirme" hedefleri ile 2030 Sanayi ve Teknoloji Stratejisi'ndeki "Yapay Zeka" ve "Büyük Veri" kritik teknoloji alanlarıyla doğrudan ilişkilidir.

1.2. Amaç ve Hedefler

Araştırma önerisinin amacı ve hedefleri açık, ölçülebilir, gerçekçi ve ulaşılabilir nitelikte olacak şekilde yazılır.

Bu araştırmanın temel amacı, teknik SQL bilgisine sahip olmayan son kullanıcıların (iş analistleri, yöneticiler, öğrenciler vb.) karmaşık ilişkisel veritabanları ile doğal Türkçe kullanarak etkileşime geçebilmesini sağlayan yüksek doğruluklu bir "Metinden SQL'e" (Text-to-SQL) modeli geliştirmektir.
Bu amaca ulaşmak için belirlenen hedefler şunlardır:

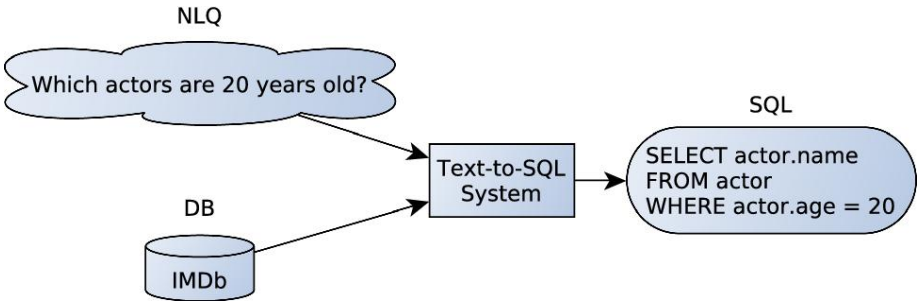
1. "Text-to-SQL" alanındaki güncel literatür (örn: Llama 3, Mistral) ve parametre verimli ince ayar teknikleri (örn: LoRA, QLoRA) üzerine kapsamlı bir araştırma yapmak.
2. Modelin eğitimi ve değerlendirilmesi için uluslararası standartta kabul görmüş, karmaşık ve çok tablolu sorgular içeren "Spider" benchmark veri setini temin etmek, analiz etmek ve eğitim için hazırlamak.
3. Seçilen temel modeli (örn: Mistral-7B), "Spider" veri seti üzerinde Parametre Verimli İnce Ayar (PEFT) metodolojisi (örn: LoRA) kullanarak eğitmek.
4. Geliştirilen modelin performansını optimize ederek; Spider liderlik tablosundaki (leaderboard) güncel "State-of-the-Art" çalışmaların [15] başarı oranları referans alınarak, "Birebir Eşleşme (Exact Match - EM)" metriğinde %65 ve üzeri; "Çalıştırma Başarısı (Execution Accuracy - EX)" metriğinde ise %80 ve üzeri başarıma ulaşmak..

2. YÖNTEM

Araştırmada uygulanacak yöntem ve araştırma tekniklerinin, amaç ve hedeflere ulaşmaya ne düzeyde elverişli olduğu ilgili literatüre atıf yapılarak ortaya konulur.

Yöntem bölümünün; araştırma tasarımı, bağımlı ve bağımsız değişkenler, istatistiksel yöntemler vb. unsurları içermesi gerekir. Araştırma önerisinde herhangi bir ön çalışma veya fizibilite yapıldıysa bunların sunulması beklenir. Araştırma önerisinde sunulan yöntemlerin çalışma takvimi ile ilişkilendirilmesi gerekir.

Bu projenin metodolojisi, insan dili (Doğal Dil - NL) ile yapısal veritabanı dilleri (SQL) arasındaki semantik uçurumu kapatmayı hedeflemektedir. Bu hedefe ulaşmak için, doğal dil sorgularını yüksek doğrulukla Yapılandırılmış Sorgu Dili'ne dönüştürebilen, alana özel (domain-specific) bir yapay zeka modeli geliştirilecektir. Sistemin temel çalışma prensibi, kullanıcının doğal dildeki sorusunun işlenerek veritabanı şeması ile birlikte SQL sorgusuna dönüştürülmesi sürecini kapsamakta olup, bu genel mimari Şekil 1'de şematize edilmiştir.



Şekil 1. Text-to-SQL Sisteminin Genel Çalışma Prensibi

Yöntemin teknik temelini, güncel bir ön-eğitimli (pre-trained) Büyük Dil Modelinin (LLM), "denetimli ince ayar" (supervised fine-tuning) tekniği ile özelleştirilmesi oluşturmaktadır. Araştırma süreci, Şekil 2'deki proje akış diyagramında detaylandırıldığı üzere 5 ana aşamalı olarak yürütülecektir:

2.1. Veri Setinin Elde Edilmesi ve Ön İşleme

Şekil 2'de gösterilen metodolojinin bu ilk adımında, Text-to-SQL alanında literatürdeki en kapsamlı ve zorlu akademik benchmark olarak kabul edilen "Spider" veri seti [11] kullanılacaktır.

Spider [11], 138 farklı alandan 200 karmaşık veritabanı şeması ve bu şemalarla ilişkili 10.000'den fazla doğal dil sorusu ve SQL sorgusu çiftini içermektedir. Bu yüksek çeşitlilik (cross-domain), modelin sadece belirli bir şemayı ezberlemesini (overfitting) engellemek ve daha önce hiç görmediği veritabanları üzerinde dahi genelleme yapabilmesini (zero-shot generalization) sağlamak için kritik bir öneme sahiptir.

Modelin, sorguyu doğru bir şekilde oluşturabilmesi için veritabanının yapısını (metadata) anlaması şarttır. Bu nedenle, veritabanı şemaları (tablo adları, sütun tipleri, birincil/yabancı anahtar ilişkileri), LLM'in bağlam (context) olarak anlayabileceği yapılandırılmış bir metin formatına dönüştürülecektir ("Schema Serialization"). Bu kritik adım, modelin "halüsinasyon" görmesini (yanlış tablo/kolon adı uydurmasını) engeller. Her bir eğitim örneği; [TALİMAT], [VERİTABANI_ŞEMASI], [SORU] ve beklenen [SQL_SORGUSU] bileşenlerini içeren özel bir prompt şablonuna oturtulacaktır. Bu yapılandırılmış veriler, modelin tokenizer'ı ile işlenerek eğitim için tensör formatına getirilecektir.

2.2. Eğitim Stratejisi ve Temel Model Mimarilerinin Belirlenmesi:

Şekil 2'deki akışın ikinci ana aşaması olan bu adım, projenin teknik omurgasını oluşturmaktadır. Bu aşamada, doğal dil sorgularını SQL komutlarına dönüştürme görevi için en uygun derin öğrenme eğitim stratejisinin ve temel model mimarisinin belirlenmesi hedeflenmektedir. Seçilecek yöntem ve mimari, projenin hesaplama maliyetini, eğitim süresini ve nihai modelin doğruluğunu doğrudan etkileyecektir.

2.2.1. Eğitim Stratejisi: Tam İnce Ayar (Full Fine-Tuning) vs. PEFT (LoRA/QLoRA):

Şekil 2 (Kutu 2.2.1)'de bu strateji seçimi vurgulanmaktadır. Literatürdeki modern NLP yaklaşımları, devasa metin verileri üzerinde ön-eğitilmiş bir temel modelin (LLM), spesifik bir göreve (bizim durumumuzda Text-to-SQL) uyarlanması olan "Transfer Öğrenme" ilkesine dayanır.

Tam İnce Ayar (Full Fine-Tuning): Geleneksel yaklaşım, 7 milyar parametreli bir modelin tüm ağırlıklarını yeni görev için güncellemektir. Bu yöntem yüksek doğruluk sunabilse de, Bölüm 3.2'de belirtilen en önemli riskleri beraberinde getirir:

Yüksek Donanım Maliyeti: 7 milyar parametrelik bir modelin tam ince ayarı, onlarlarca GB VRAM gerektirir ve genellikle birden fazla yüksek performanslı GPU'ya ihtiyaç duyar.

Katastrofik Unutma (Catastrophic Forgetting): Model, yeni göreve aşırı odaklanırken, öğrendiği genel dil yeteneklerini unutma riski taşır.

Parametre Verimli İnce Ayar (PEFT): Projemiz için en verimli ve modern yaklaşım PEFT olacaktır. Bu teknik, ön-eğitimli modelin milyarlarca ağırlığını dondurur ve sadece modelin üzerine eklenen çok küçük (toplam parametrelerin %1'inden az) "adaptör" katmanlarını eğitir.

LoRA (Low-Rank Adaptation): Bu projede kullanılacak spesifik PEFT tekniği LoRA olacaktır. LoRA, eğitilmesi gereken ağırlık sayısını drastik olarak azaltır.

QLoRA (Quantized LoRA): Verimliliği daha da artırmak için LoRA, 4-bit nicemleme (quantization) ile birleştirilecektir. Bu optimizasyon, projemize "hafif" bir model ile çalışma avantajı sağlar. QLoRA sayesinde, 7-8 milyar parametreli bir modelin eğitimi, tek bir tüketici sınıfı GPU (Google Colab T4 veya RTX serisi) üzerinde dahi mümkün hale gelir.

Bu projede, hesaplama kaynaklarını verimli kullanmak ve 2209-A bütçesi dahilinde kalmak için PEFT (QLoRA) yöntemi ana eğitim stratejisi olarak seçilmiştir.

2.2.2. Temel Model Mimarilerinin Değerlendirilmesi: Llama-3-8B vs. Mistral-7B

Şekil 2 (Kutu 2.2.2)'de belirtildiği gibi, QLoRA eğitim stratejisinin uygulanacağı temel model mimarisinin seçimi, projenin performans-verimlilik dengesi için kritik bir öneme sahiptir. Bu projede, 2209-A projesinin hesaplama kaynakları ve bütçe kısıtları göz önünde bulundurularak, en iyi **performans/verimlilik oranını** sunan mimari olarak **Mistral-7B** modeli tercih edilmiştir.

Bu stratejik tercihin temel nedeni, Mistral-7B'nin sunduğu modern mimari yeniliklerdir. Model, "Transformer" mimarisinde **Gruplandırılmış Sorgu Dikkati (GQA)** ve **Kayan Pencere Dikkati (SWA)** gibi verimlilik odaklı mekanizmalar kullanır. Bu teknikler, modelin 7 milyar parametre gibi görece düşük bir boyutta, kendisinden daha büyük parametreli (örn: 13B Llama-2) modellerle rekabet edebilmesini sağlar. Projemiz için bu durumun doğrudan getirisi; **daha az bellek (VRAM) kullanımı** ve **daha hızlı çıkarım (inference) süresi** elde etmektir.

Referans bir karşılaştırma noktası olarak, Meta AI tarafından geliştirilen Llama-3-8B mimarisi de güçlü bir alternatiftir. Llama-3-8B, özellikle karmaşık mantıksal akıl yürütme (reasoning) ve kod üretme konularında olağanüstü performans göstermektedir. Text-to-SQL görevi (hem mantık hem kod üretimi) için yetenekleri çok yüksek olsa da, Mistral-7B'nin sunduğu mimari verimlilik avantajları, projemizin "ulaşılabilir kaynaklar ile yüksek başarı" hedefine daha uygun bulunmuştur.

Sonuç olarak eğitim süreci , projenin 3.1'deki çalışma takvimine uygun olarak Mistral-7B mimarisi üzerine inşa edilecek . Bunun sebebi, 2209-A projesi için en iyi performans/verimlilik oranını sunmasıdır. Eğitim sürecinde, literatürdeki güncel "State-of-the-Art" çalışmaların [15] performans seviyeleri referans alınarak; 2.3'teki değerlendirme metriklerinde (EX ve EM) hedeflenen %80+ semantik doğruluğa, bu verimli ve güçlü model ile ulaşılmaya çalışılacaktır.

2.3. Değerlendirme Metrikleri Eğitilen modelin performansı, test veri seti üzerinde literatürde standart kabul edilen iki temel metrik ile nicel olarak ölçülecektir:

Şekil 2'deki akış diyagramının 2.3 numaralı adımında belirtildiği üzere, eğitilen modelin performansı, test veri seti üzerinde literatürde standart kabul edilen iki temel metrik ile nicel olarak ölçülecektir:

Birebir Eşleşme (Exact Match - EM): Modelin ürettiği SQL sorgusunun, beklenen (ground-truth) SQL sorgusu ile yapısal olarak tamamen aynı olup olmadığını ölçer. Bu metrik çok katı bir kuraldır ve sorgudaki küçük, anlamsal olarak önemsiz farklılıkları (örn: LIMIT 1 yerine TOP 1) hata olarak sayar.

Çalıştırma Başarısı (Execution Accuracy - EX): Modelin ürettiği SQL sorgusunun veritabanında çalıştırılması sonucu dönen verinin, doğru sorgunun döndürdüğü veri ile eşleşme oranıdır. EM'in aksine bu metrik, sorgunun anlamsal olarak doğru olup olmadığına odaklanır. Projemizin temel amacı, son kullanıcıya doğru cevabı getirmek olduğundan, temel başarı kriteri olarak semantik doğruluğu ölçen 'Execution Accuracy' metriği baz alınacaktır.

2.4. Prototip Arayüz ve Entegrasyon Araştırmanın somut bir çıktıya dönüşmesi ve geliştirilen modelin yeteneklerinin doğrulanması amacıyla, Şekil 2 (Kutu 2.4)'te planlandığı gibi Python tabanlı Streamlit veya Gradio kütüphaneleri kullanılarak interaktif bir web prototipi geliştirilecektir.

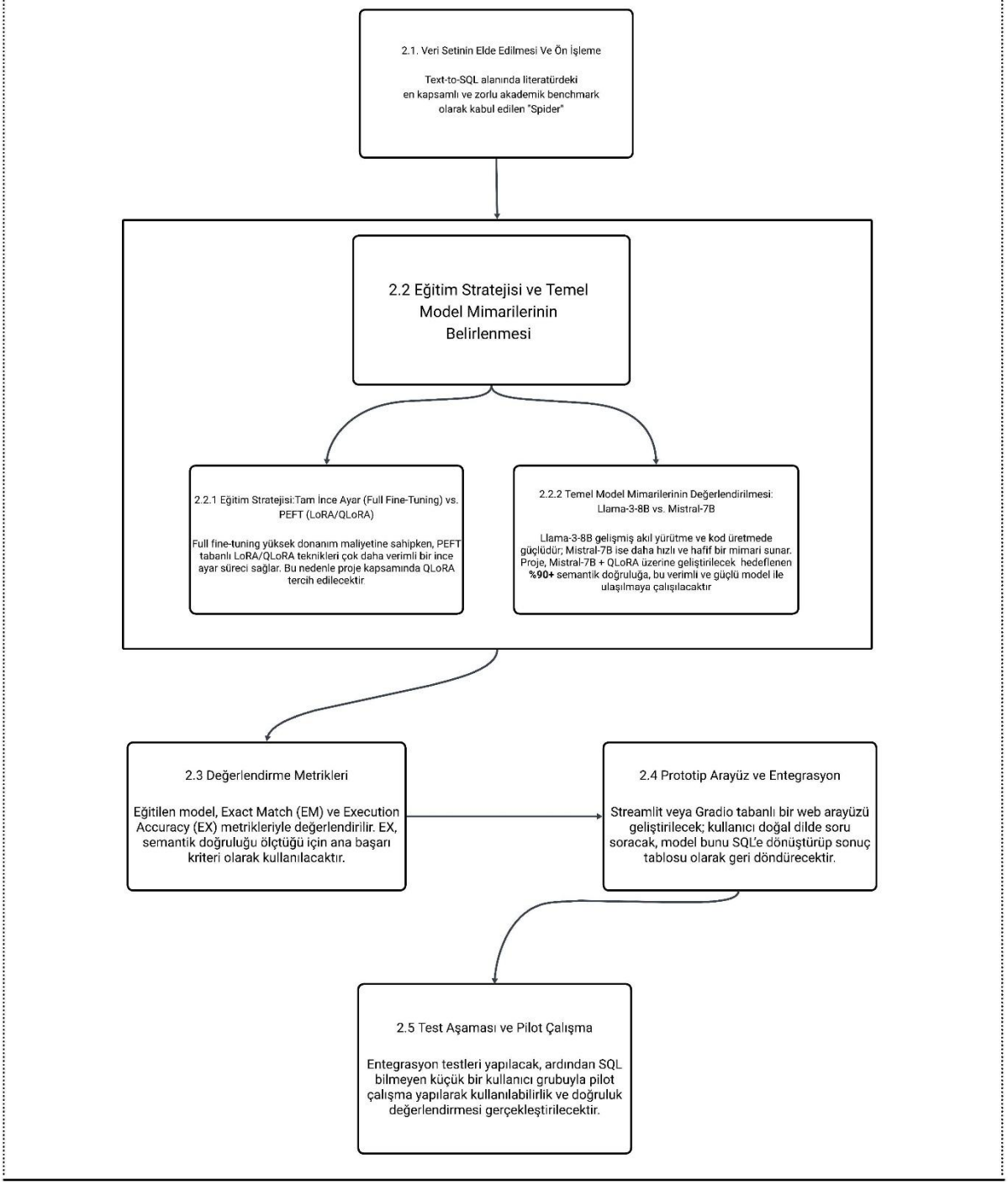
Kullanıcı, arayüz üzerinden veritabanı şemasını yükleyebilecek ve Şekil 2'de örneklendirildiği gibi doğal dilde sorusunu (örn: "Stok miktarı 50'den az olan ürünleri listele") girebilecektir. Arka planda çalışan ince ayarlı model, bu girdiyi SQL koduna (örn: SELECT * FROM urunler WHERE stok < 50) dönüştürecek ve sorguyu sanal bir veritabanında çalıştırarak sonucu tablo halinde kullanıcıya sunacaktır. Bu arayüz, teknik bilgisi olmayan bir kullanıcının dahi karmaşık veritabanlarıyla nasıl "sohbet" edebileceğini canlı olarak gösterecektir.

2.5. Test Aşaması ve Pilot Çalışma

Şekil 2'de gösterilen proje metodolojisinin son aşaması (Kutu 2.5), geliştirilen modelin ve prototip arayüzün kapsamlı testlerini içermektedir. Bu aşama, projenin hedeflerine ulaşıp ulaşmadığını doğrulamak için kritik öneme sahiptir ve iki adımdan oluşur:

Entegrasyon Testleri: Bu adımda, 2.4'te geliştirilen Streamlit/Gradio arayüzünün, 2.2'de eğitilen LLM modeli ile sorunsuz bir şekilde iletişim kurup kurmadığı test edilecektir. Kullanıcı tarafından girilen bir sorunun modele doğru iletilmesi, modelden dönen SQL sorgusunun veritabanında çalıştırılması ve sonucun arayüze hatasız bir şekilde basılması doğrulanacaktır. Hatalı veya anlamsız sorgulara karşı sistemin verdiği tepkiler (hata yönetimi) de bu aşamada test edilecektir.

Kullanıcı Pilot Çalışması: Projenin temel amacı olan "teknik bilgisi olmayan kullanıcıların veriye erişimini kolaylaştırma" hedefini doğrulamak için küçük ölçekli bir pilot çalışma yapılacaktır. SQL bilgisine sahip olmayan (örn: farklı fakültelerden) 3-5 kişilik bir öğrenci grubuyla bir kullanıcı testi (usability test) düzenlenecektir. Kullanıcılara basit bir veritabanı şeması (örn: "Öğrenci Not Sistemi") verilecek ve bu sistemden belirli bilgileri (örn: "Matematik dersinden en yüksek notu alan öğrencinin adı") bulmaları istenecektir. Bu test ile prototipin kullanılabilirliği ve projenin ana problemini çözme başarısı değerlendirilecektir.



Şekil 2. Proje İş Akış Şeması

3. PROJE YÖNETİMİ

3.1 Çalışma Takvimi

ÇALIŞMA TAKVİMİ (*)

Tarih Aralığı	Faaliyetler**	Kim(ler) Tarafından Gerçekleştirileceği	Başarı Ölçütü ve Araştırmanın Başarısına Katkısı***
01/12/2025-01/01/2026	Veri Setlerinin Elde Edilmesi ve Ön İşleme	Buğra Kaan KESMEZ Gözde YOLCU ÖZTEL	Toplam **10.181 adet** doğal dil sorusu ve **200 farklı** veritabanı şemasını içeren "Spider" veri setinin temin edilmesi, eğitim kümesinde (train split) yer alan yaklaşık **7.000 adet** örnek için; veritabanı şemaları (tablo isimlerinin, sütun tiplerinin, anahtar ilişkileri) metin formatına dönüştürülmesi (serialization) ve her biri için **'Talimat-Şema-Soru'** yapısına sahip prompt şablonları hazırlanması. **Projenin Başarısına Katkısı (%): 15**
01/01/2026-01/04/2026	Temel Model Seçimi ve PEFT ile Eğitilmesi	Buğra Kaan KESMEZ Gözde YOLCU ÖZTEL	Mistral-7B/Llama-3 modelinin seçilmesi, özel veritabanları üzerinde yüksek doğruluk (%80+) sağlamak için LoRA tekniği ile ince ayar (fine-tuning) eğitiminin tamamlanması ve model ağırlıklarının kaydedilmesi. **Projenin Başarısına Katkısı (%): 35**
01/02/2026-01/05/2026	Prototip Arayüz Geliştirme ve Model Entegrasyonu	Buğra Kaan KESMEZ Gözde YOLCU ÖZTEL	Streamlit/Gradio tabanlı web arayüzünün geliştirilmesi, eğitilen modelin arayüze entegrasyonu ve sorgu üretiminin sağlanması. **Projenin Başarısına Katkısı (%): 35**
01/05/2026-01/06/2026	Test ve Değerlendirme	Buğra Kaan KESMEZ	Modelin doğruluk (EM/EX) testlerinin yapılması, arayüzün son kullanıcılar (pilot grup) ile test edilmesi ve raporlanması. **Projenin Başarısına Katkısı (%): 15**
			Projenin Başarısına Toplam Katkısı (%): 100

(*) Çizelgedeki satırlar ve sütunlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

(**) Literatür taraması, sonuç raporu hazırlama aşamaları, araştırma sonuçlarının paylaşımı, ve malzeme alımı ayrı birer iş adımı olarak gösterilmemelidir.
(***) Başarı ölçütü, ölçülebilir ve izlenebilir nitelikte olacak şekilde nicel veya nitel ölçütlerle (ifade, sayı, yüzde, vb.) belirtilir. Bu sütundaki değerlerin toplamı 100 olmalıdır.

3.2 Risk Yönetimi

Araştırmanın başarısını olumsuz yönde etkileyebilecek riskler ve bu risklerle karşılaşıldığında araştırmanın başarıyla yürütülmesini sağlamak için alınacak tedbirler (B Planı) aşağıdaki Risk Yönetimi Tablosu'nda ifade edilir. B Plan(lar)ının uygulanması araştırmanın temel hedeflerinden sapmaya yol açmamalıdır. B Plan(lar)ına geçilmesi durumunda yöntem değişikliğine gidiliyor ise bu durum detaylandırılmalıdır.

RİSK YÖNETİMİ TABLOSU*

En Önemli Riskler	Alınacak Tedbirler (B Planı)
Seçilen 7-8 milyar parametreli modelin eğitimi sırasında GPU belleğinin (VRAM) yetersiz kalması veya eğitimin çok uzun sürmesi.	Model ağırlıkları 4-bit (QLoRA) yerine daha düşük hassasiyette optimize edilecek veya eğitim sırasında "gradient accumulation" adımları artırılacaktır. Bütçe dahilinde Google Colab Pro+ gibi bulut GPU servisleri kullanılacaktır.
Eğitilen modelin, test veri setinde hedeflenen doğruluk seviyesine (%80) ulaşamaması ve karmaşık sorgularda "halüsinasyon" görmesi.	Veri seti üzerinde "veri artırma" (data augmentation) yapılarak örnek sayısı artırılacaktır. Ayrıca, modelin girdisine örnek sorguların eklendiği "Few-Shot Prompting" tekniği uygulanarak başarım artırılmaya çalışılacaktır.
Geliştirilen web arayüzünün (Streamlit), modelin yanıt süresini çok geciktirmesi veya entegrasyon sorunları yaşanması.	Arayüz kütüphanesi olarak daha hafif olan Gradio tercih edilecek veya model, arayüzden bağımsız bir API (FastAPI) olarak servis edilip arayüz sadece istek atacak şekilde mimari değiştirilecektir.

(*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

3.3.Araştırma Olanakları

Bu bölümde projenin yürütüleceği kurum ve kuruluşlarda var olan ve projede kullanılacak olan altyapı/ekipman (laboratuvar, araç, makine-teçhizat, vb.) olanakları belirtilir.

ARAŞTIRMA OLANAKLARI TABLOSU (*)

Kuruluşta Bulunan Altyapı/Ekipman Türü, Modeli (Laboratuvar, Araç, Makine-Teçhizat, vb.)	**Projede Kullanım Amacı**

(*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

4. ARAŞTIRMA ÖNERİSİNİN YAYGIN ETKİSİ

Araştırma önerisi kapsamındaki çalışmadan elde edilmesi öngörülen çıktılar amaçlarına göre belirlenen kategorilere ayrılarak belirtilir; ölçülebilir ve gerçekçi hedeflere dayandırılır.

Çıktı, Etki ve Kazanımlar	**Öngörülen Çıktı(lar), Etki(ler) ve Kazanım(lar)**
Bilimsel/Akademik Çıktılar (Ulusal/Uluslararası Makale, Kitap Bölümü, Kitap, Bildiri vb.)	Önerilen çalışmadan bir uluslararası bildiri üretilmesi planlanmaktadır.
Ekonomik/Ticari/Sosyal Çıktılar (Ürün, Prototip, Patent, Faydalı Model, Tescil, Görsel/İşitsel Arşiv, Envanter/Veri Tabanı, Çalıştay, Eğitim, Bilimsel Etkinlik vb.)	Yöntem (Bölüm 2.4) ve İş Paketi 3 (İP-3) kapsamında geliştirilecek olan Streamlit/Gradio tabanlı web arayüzü, projenin "Kavram Kanıtlaması" (Proof-of-Concept) prototipi olacaktır. • **Sosyal Etki:** Bu prototip, özellikle KOBİ'ler, STK'lar ve üniversite idari birimlerindeki teknik bilgisi olmayan personelin (SQL bilmeyen) kendi verilerine doğrudan erişimini kolaylaştırarak 'veriye erişimi demokratikleştirme' potansiyelini ortaya koyacaktır. Proje çıktısı, kurumsal verimliliği artırmaya yönelik gelecekteki kapsamlı sistemlerin geliştirilmesi için somut bir kavram kanıtı (proof-of-concept) sunacak ve bu alandaki ileri çalışmalar için teknik bir temel oluşturacaktır.
Yeni Proje(ler) Oluşturmasına Yönelik Çıktılar (Ulusal/Uluslararası Yeni Proje vb.)	Bu 2209-A projesinde geliştirilen prototip, belirli bir şirket veritabanı üzerinde uygulanarak **TÜBİTAK 2209-B (Sanayi Odaklı)** projesi için güçlü bir temel oluşturacaktır. • Prototipin, **TEKNOFEST** Yapay Zeka yarışmaları için daha gelişmiş bir ürüne dönüştürülme potansiyeli bulunmaktadır.

5. BELİRTMEK İSTEDİĞİNİZ DİĞER KONULAR

Sadece araştırma önerisinin değerlendirilmesine katkı sağlayabilecek bilgi/veri (grafik, tablo, vb.) eklenebilir.

6. EKLER

EK-1: KAYNAKLAR

[1] Gantz, J., & Reinsel, D. (2012). "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east". IDC iView: IDC Analyze the Future, 2007(2012), 1-16.

[2] McAfee, A., & Brynjolfsson, E. (2012). "Big data: The management revolution". Harvard business review, 90(10), 60-68.

[3] Jagadish, H. V., et al. (2014). "Big data and its technical challenges". Communications of the ACM, 57(7), 86-94.

[4] Androutsopoulos, I., Ritchie, G. D., & Thanisch, P. (1995). "Natural language interfaces to databases–an introduction". Natural language engineering, 1(1), 29-81.

[5] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). "Attention is all you need". Advances in neural information processing systems, 30.

[6] Qin, B., et al. (2024). "A Comprehensive Survey on Text-to-SQL: Current Advances and Future Directions". ACM Computing Surveys (CSUR).

[7] Meta AI. (2024). "Llama 3: Open Foundation and Instruction-Tuned Models". Meta AI Research Paper.

[8] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Prafulla, D., ... & Amodei, D. (2020). "Language models are few-shot learners". Advances in neural information processing systems, 33, 1877-1901.

[9] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). "Survey of hallucination in natural language generation". ACM Computing Surveys, 55(12), 1-38.

[10] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing". ACM Computing Surveys, 55(9), 1-35.

[11] Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., ... & Radev, D. (2018). "Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Text-to-SQL". arXiv preprint arXiv:1809.08887.

[12] Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., ... & Lample, G. (2023). "Mistral 7B". arXiv preprint arXiv:2310.06825.

[13] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). "LoRA: Low-Rank Adaptation of Large Language Models". arXiv preprint arXiv:2106.09685.

[14] Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). "QLoRA: Efficient Finetuning of Quantized LLMs". arXiv preprint arXiv:2305.14314.

[15] Pourreza, M., & Rafiei, D. (2023). "DIN-SQL: Decomposition-introduced in-context learning for text-to-sql". *Advances in Neural Information Processing Systems (NeurIPS)*, 36.