



# BURSA TEKNİK ÜNİVERSİTESİ

## BTÜ-İMEP SONUÇ RAPORU

### ÖĞRENCİNİN

Adı Soyadı	: Emine ŞENER
Numarası	: 21360859058
Bölümü	: Bilgisayar Mühendisliği
E-posta adresi	: eminesener063@gmail.com
BTÜ-İMEP akademik danışmanı	: Ergün GÜMÜŞ
BTÜ-İMEP sektör danışmanı	: Halit ÖRENBAŞ
Çalışma dönemi	: 30.09.2024 – 10.01.2025

### BTÜ-İMEP Kurumunun:

Adı	: Huawei
Adresi	: Saray Mah. Ahmet Tevfik İleri Cad. Onur Ofis Park İş Merkezi Sit. A1Blok Apt. No.10 B/1 Ümraniye İstanbul
İnternet sitesi	: <a href="https://www.huawei.com/tr/">https://www.huawei.com/tr/</a>
Görev alınan birim	: AIE (Artificial Intelligence Enablement)

## İçindekiler

1.	GİRİŞ .....	3
2.	KURUM TANITIMI.....	5
2.1.	Birim tanıtımı .....	6
3.	ÇALIŞMALAR.....	7
3.1.1	Çalışmanın Hedefi.....	7
3.1.2.	Yöntem.....	17
3.1.3.	Çalışmanın Sonuçları.....	27
5.	KAYNAKLAR .....	34

## 1. GİRİŞ

BTÜ-İMEP kapsamında bir dönem boyunca Huawei gibi değerli bir şirkette tam zamanlı bir mühendis gibi çalışma fırsatına eriştim. Henüz mezun olmadan, iş hayatının çeşitli unsurları üzerinde deneyim kazandım ve iş hayatına olabildiğince hazır hale geldim. BTÜ Bilgisayar Mühendisliği öğrencisi olarak 6 dönem boyunca, bizi hem akademiye hem de sektöre hazırlayan derler aldım. Lisans eğitimimin 7.döneminde ise, bu ders sürecini okulumda değil Huawei’de devam ettirdim ve hem okulda öğrendiklerimi uygulama hem de yeni konular için öğrenci olmaya devam ettim.

Bir kurumda stajyer olmanın kariyerim için oldukça verimli olduğunu düşünüyorum. Çünkü, stajyer olarak bir kurumda çalışıyor olmak hem iş insanı hem de öğrenci sıfatına sahip olmamızı sağlıyor. Böylelikle, hem öğrenmeye devam ediyor hem de iş hayatında deneyim kazanıyoruz. BTÜ-İMEP kapsamındaki stajım boyunca, hem bir iş insanı olarak Huawei için çalışmalar yaptım ve deneyim kazandım, hem de henüz bir öğrenci olduğum için öğrenmeye devam ettim.

Staj sürecimde iki farklı açıdan kendimi geliştirmemde Huawei’in, değerli sektör danışmanımın ve içinde bulunduğum ekibin katkısı yadsınamaz. Stajım boyunca, henüz bir öğrenci olduğum için sürekli yeni kavramlar öğrenmem gereken alanlarda görevler aldım. Bu sayede, öğrenme ve deneyim kazanma süreçlerini birlikte yürütebildim. İlgi alanım ile alakalı olmayan ya da öğrenmeyip sadece geliştirme yapılması görevler verilmedi.

Huawei’de yaptığım stajımda ayrıca oldukça zengin bir departmanda görev aldığım için, sektörden birçok kişi ile tanışma fırsatına eriştim. Farklı alanlarda çalışan kişilerin projeleri hakkında bilgiler aldım ve onların kariyer yolculuklarını dinledim. Birçok kişinin deneyimlerini dinlemek sektör hakkında kapsamlı bir bakış elde etmemi sağladı.

Kariyer yolculuğumun erken dönemlerini kurumsal şirketlerde geçirmeyi planlıyorum. Huawei gibi büyük bir kurumsal şirkette çalışarak kurumsal düzen ve kurallar hakkında bilgi edindim. Böylelikle mezun olduğumda kurumsal şirket düzenini uyum sağlamış bir çalışan olarak ilk işime başlayacağım.

Huawei’de stajım boyunca birçok kez sunum yaptım. Bunların çoğunluğunu, Huawei Head of AI olan sektör danışmanıma sundum. Kendisinden aldığım geri dönüşler ile çalışmalarımı şekillendirdim. Ayrıca Huawei Software Engineering Director’üne de bir sunum gerçekleştirdim. Bu sunumda büyük dil modelleri, agent’lar gibi gelişmiş konular yer almaktaydı. Bir diğer sunumum da tüm ekibe karşı yapıldı. Henüz mezun olmadan alanlarında oldukça yetkin kişilere sunum yapmak, kariyerim için oldukça değerliydi. Bu süreçleri başarıyla tamamlayınca, kariyer hayatımda yalnızca teknik tarafta değil yönetim taraflarında da çalışmalar yapan bir mühendis olmayı hedefledim.

Çalışma dönemi boyunca, büyük dil modelleri, agent’lar ve prompt kodlama gibi ileri düzey konular üzerine çalıştım. Bu konular oldukça karmaşık ve güncel konular. Bu konular üzerine bilgi birikimi kazadığım için kariyerimin devamında da bu alanda çalışmalar yapmaya, mezun olduktan sonra bu alanlarda geliştirme yapabileceğim bir şirkette çalışmaya karar verdim.

Program döneminde 2 farklı projede yer aldım: Travel Planner ve DSPy Kütüphanesi. Travel Planner multi agentic bir sistem olarak tasarlandı ve agent, multi-agent ve LangGraph üzerinde uzmanlaşıldı ve uçtan uca bir proje tamamlandı. Stajda çalışılan bir diğer konu olan DSPy ile prompt kodlama, büyük dil modeli uygulamalarını değerlendirme ve optimize etme üzerine araştırmalar yapılarak güçlü bir bilgi birikimi elde edildi. Ayrıca DSPy ile deneyler de yapıldı.

Raporda Travel Planner projesi ve DSPy ayrı bölümlerde ele alındı. Travel Planner öncelikle uzun bir araştırma süreci gerçekleştirildi ve hedefler belirlendi. Ardından hedefler üzerinde çalışmalar tamamlandı ve bu çalışmalar farklı konularda bilgi birikimini ve deneyleri gerektirdi. Bunların her biri ayrı bölümlerde ele alındı. Sonuç bölümünde ise geliştirilen projenin nihai planı ve çıktılarından bahsedildi.

DSPy geliştirmenin daha az araştırmanın ağırlıklı olduğu bir çalışmaydı ve ana hedef bilgi birikimi edinip ekibe aktarmaktı. Böylelikle bu görevin anlatıldığı kısımda elde edilen bilgiler özetlendi ve yapılan deneyler aktarıldı.

## 2. KURUM TANITIMI

BTÜ-İMEP kapsamında yer aldığım Huawei, 1987 yılında kurulmuş olan Çin merkezli kurumsal bir firmadır. Bilgi ve iletişim teknolojisi (ICT) altyapısı ve akıllı cihazlarıyla teknoloji devi bir şirkettir. 170'ten fazla ülkede faaliyet gösteren Huawei'in Türkiye'de de oldukça güçlü bir etkisi bulunmaktadır. Özellikle Cloud teknolojisiyle, önde gelen Türk şirketleriyle iş birliği yapmaktadır.

Huawei akademide de oldukça aktif rol oynamaktadır. Kampüs Sektörde programı dahilinde üniversiteler ile iş birliği halindedir. Bu iş birliği dahilinde, Huawei'in alanında uzman çalışanları, üniversitelerde uzaktan ve de yüz yüze olmak üzere dersler vermektedir.

Huawei ayrıca birçok okulda Huawei Student Developers öğrenci kulüplerinin oluşturulmasını destekler. Bu kulüpler ile sürekli etkileşim halinde kalarak öğrenciler için verimli etkinlikler düzenlenilmesini sağlar.

Huawei'de tüm projeler HQ(Head Quarter) tarafından titizlikle incelenir. Her bir proje için farklı sayılarda çeşitli sayıda, junior ya da senior çalışanların yönlendirilmesini sağlayan Project Menager(PM) bulunur. Project Menager, ekip içerisinde yürütülen birden fazla projenin yönetilmesini sağlayan Engineering Menager ile bağlantılı çalışır. Engineering Menager ise, tüm departmanının yönetilmesini sağlayan direktöre bağlıdır.

Huawei'de AIE, SADC ve IAD departmanları bulunmaktadır. Bu departmanlarda Huawei Genel Müdürleri ile aracılığıyla yönetilmektedir. Huawei 207.000'den fazla çalışanıyla dev bir teknoloji şirkettir. Huawei Türkiye içerisinde 1700'den fazla mühendis 50'den fazla yönetici ile birlikte çalışmaktadır.

## 2.1. Birim tanıtımı

BTÜ-İMEP kapsamında Huawei AIE (Artificial Intelligence Enablement) departmanında görev aldım. Bu departman Cloud, Big Data ve AI alanlarında geliştirilen projeleri içerinde barındırmaktadır. Departman altında görev aldığım takım AI Research and Development takımıydı. Bu takım kendi şirkette görev aldığım dönem içerisinde 3 farklı projenin geliştirilmesini sağlamaktadır.

Projeler Computer Vision, Recommender System ve Large Language Model olmak üzere farklı alanlardadır. Projeler bazında AI Research and Development ekibi 3 ekibe bölünmüştür. Stajım boyunca LLM ekibinde yer aldım.

LLM ekibi, oldukça yeni bir alanda çalışarak Huawei'in son teknoloji projeleri için ağırlıklı olarak research yapmaktadır. Research ile paralel geliştirmeler de yapılarak Head Quarter'a sunulmaktadır. Dahil olduğum ekip Head Quarter tarafından ödüllendirilen oldukça başarılı bir ekip.

Yer aldığım ekipte çalışan profili tamamen mühendislik fakültelerinden, Bilgisayar Mühendisliği, Kontrol ve Otomasyon Mühendisliği, Elektrik Elektronik Mühendisliği gibi bölümlerden mezun kişilerden oluşmaktadır. Ekibin belirli kısmı doktorasını tamamlamış çalışanlardan oluşup, çoğunluğu doktora döneminde ya da yüksek lisans mezundur.

Ekibin sorumlulukları arasında, dahil olduğu proje için research yapmak ve buna paralel başarılı geliştirmeler yapmak yer almaktadır. Ekip ayrıca üniversite- sanayi iş birlikleri çerçevesinde okullarda ders vererek ve Tübitak projeleriyle şirkete katkı sağlamaktadır.

Huawei'de iş bölümü kişinin niteliklerine ve deneyimine uygun olarak yapılmaktadır. Ayrıca iş bölümü sürecinde yalnızca deneyime bakılmadan çalışanların ilgili oldukları alanlara yönelik projelerde çalışabilmesi de sağlanmaktadır.

### 3. ÇALIŞMALAR

BTÜ-İMEP kapsamında tamamlanan staj sürecinde iki farklı projede görev alındı. Bunlardan ilki Agentik AI tekniği ile geliştirilen Travel Planner projesidir. Bir diğer çalışma alanı ise, Stanford Üniversitesi NLP Grup tarafından oluşturulan DSPy kütüphanesidir. Travel Planner projesinde uçtan uca bir uygulama geliştirmek iken, DSPy projesindeki hedef, bu kütüphane üzerinde uzmanlaşmak, deneyler yapmak ve Huawei AI Araştırma ekibine sunmaktır. Staj sürecinde iki hedef de başarıyla tamamlandı. Bu iki çalışma raporda ayrı hedef – yöntem – sonuç bölümleriyle anlatılacaktır.

#### 3.1. Travel Planner (Seyehat Planlayıcısı)

##### 3.1.1 Çalışmanın Hedefi

Travel Planner projesinin gerçekleştirilmesindeki ana hedef, günümüz AI dünyasında en çok konuşulan alanların başında gelen Agent kavramı üzerinde uzmanlaşmaktır. Proje kapsamında Agent kavramı ve Agent ile ilişkili büyük dil modeli konuların üzerinde bilgi birikimi oluşturulurken aynı zamanda Huawei için uçtan uca bir projenin geliştirilmesi hedeflenmiştir. Edinilmesi hedeflenen bilgi birikimlerinin ve projede bulunması gereken kriterlerin belirlenmesi için öncelikle kapsamlı bir araştırma ve raporlama süreci başlatıldı.

Projenin geliştirme sürecine TravelPlanner: A Benchmark for Real-World Planning with Language Agents[1] isimli makale analiz edilerek başlandı. OSU NLP Group [2] tarafından yayımlanan makalede Travel Planner projesinde bulunması gereken özelliklerden bahsedilmektedir. Makalede bahsedilen özellikler referans alınarak geliştirilecek projede bulunması gereken özellikler belirlendi. Bu özellikler aşağıda maddelenmiştir.

- Temelde kullanıcıdan metin formatında bir sorgu almaya ve buna uygun metin formatında anlamlı, kapsamlı ve sorguya uygun bir yanıt üretmeye odaklanan bir soru cevap uygulaması geliştirilecek.

- Geliştirilen uygulama sorgu olarak, Çizelge 3.1.1.1.'de gösterildiği şekilde sorgular olarak Şekil 3.1.1.1'de gösterildiği gibi yanıtlar üretmeli. Çizelge 3.1.1.2.'de yer alan sorgu örnekleri referans alınan makalede yer alan veri setinden elde edilmiş olup, benzer sorguların onlarcası veri setinde bulunmaktadır. Geliştirilen uygulama veri setinde yer alan farklı sorgularla defalarca test edildi.



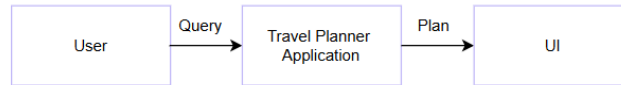
Şekil 3.1.1.1: Hedeflenen Uygulama Çıktısı



#### Çizelge 3.1.1.3. Sorgu – Cevap örnekleri

Query	Level
Please plan a trip for me starting from Sarasota to Chicago for 3 days, from March 22nd to March 24th, 2022. The budget for this trip is set at \$1,900.	Easy
Please plan a trip for two people from Wilmington to Washington for three days from March 19th to March 21st, 2022, with a budget of \$2,500. We require accommodations that allow visitors.	Medium
Can you help me plan a 3-day trip from Boston to Washington for two people from March 19th to March 21st, 2022 on a budget of \$900? We require accommodations where pets are allowed, and we prefer not shared rooms. We also request no flights for transportation.	Hard

Projede gerçekleştirilmesi planlanan başlıca hedef Şekil 3.1.1.2’de gösterildiği üzere Çizelge 1’de yer alan seyahat sorgularına benzer sorguları alıp bunlara uygun, Şekil 1’de gösterilen biçimde yanıtların üretilmesidir.



#### Şekil 3.1.1.2 Hedeflenen Uygulama Çıktısı

Çizelge 3.1.1.1’de yer alan sorgu örnekleri easy – medium – hard şeklinde sınıflandırılmıştır. Bu sorgular, uygulamanın odaklanması gereken kısıtlamaların sayısına ve karmaşıklığına bağlı olarak sınıflandırılır. Kısıtlamalar makalede ayrıntılı bir şekilde açıklanmış ve kendi içinde ayrı gruplara ayrılmıştır.

Plan oluşturulurken dikkat edilmesi gereken ana noktalar Çizelge 3.1.1.2’de gösterilmiştir.

#### Çizelge 3.1.1.2. Plan Kısıtlamaları

Constraint	Description
Uçuş	Makalede yer alan sorgular incelendi ve çoğu sorguda istenen seyahatin uçuş mesafesinde olduğu saptandı. Bu sebeple sorgudan kullanıcının planladığı güzergahın tespit edilmesi ve buna uygun uçuşların bulunması gerekiyor.
Restaurant	Oluşturulması hedeflenen seyahat planında her bir gün için, 3 farklı öğüne restaurant önerisi yapılması gerekmektedir.
Konaklama	Sorguda yer alan seyahat edilmesi planlanan şehirlerin tespit edilmesi ve buna uygun konaklama yerlerinin, her bir gün için planlanması gerekmektedir.
Etkinlik	Sorguda kaç günlük bir plan istenildiği tespit edildikten sonra her bir gün için etkinlik önerisi yapılması da planlanmaktadır.
Bütçe	Kullanıcı sorgularında bütçe bilgisi de yer almaktadır ve projede oluşturulacak planlarda bütçe aşılmadan bir plan oluşturulması gerekiyor.

Oluşturulması planlanan uygulamanın, kullanıcıdan aldığı sorguları, Çizelge 3.1.1.2’de gösterilen başlıklara göre analiz etmesi ve uygun işlemlerle bunlara uygun plan üretmesi hedeflenmektedir.

Uygulamanın plan oluştururken dikkat etmesi gereken kısıtlamalar yalnızca bunlarla sınırlı olmayıp Şekil 3.1.1.3’te gösterildiği üzere her bir kısıtlama kendi içinde farklı ek kısıtlamaları da içermektedir.

BURSA TEKNİK ÜNİVERSİTESİ  
PROJE TABANLI İŞLETMEDE MESLEKİ EĞİTİM  
BTÜ-İMEP ÖĞRENCİ FİNAL RAPOR FORMU F4

Constraint	Description
<b>Environment Constraint</b>	
Unavailable Transportation	There is no available flight or driving information between the two cities.
Unavailable Attractions	There is no available attraction information in the queried city.
<b>Commonsense Constraint</b>	
Within Sandbox	All information in the plan must be within the closed sandbox; otherwise, it will be considered a hallucination.
Complete Information	No key information should be left out of the plan, such as the lack of accommodation during travel.
Within Current City	All scheduled activities for the day must be located within that day's city(s).
Reasonable City Route	Changes in cities during the trip must be reasonable.
Diverse Restaurants	Restaurant choices should not be repeated throughout the trip.
Diverse Attractions	Attraction choices should not be repeated throughout the trip.
Non-conf. Transportation	Transportation choices within the trip must be reasonable. For example, having both "self-driving" and "flight" would be considered a conflict.
Minimum Nights Stay	The number of consecutive days spent in a specific accommodation during the trip must meet the corresponding required minimum number of nights' stay.
<b>Hard Constraint</b>	
Budget	The total budget of the trip.
Room Rule	Room rules include "No parties", "No smoking", "No children under 10", "No pets", and "No visitors".
Room Type	Room types include "Entire Room", "Private Room", "Shared Room", and "No Shared Room".
Cuisine	Cuisines include "Chinese", "American", "Italian", "Mexican", "Indian", "Mediterranean", and "French".
Transportation	Transportation options include "No flight" and "No self-driving".

### Şekil 3.1.1.3 Makalede Gösterilen Ek Kısıtlamalar

Oluşturulması hedeflenen uygulamanın, makalede yer alan, environment- commonsense-hard olmak üzere 3 farklı grupta sınıflandırılan kısıtlamalara da uyması hedeflenmektedir. Hard Constraint, uygulamanın öncelikli dikkat etmesi gereken kısıtlama olup; bütçe, ulaşım gibi tüm planı etkileyen durumlardan ve oda tipi, oda kuralları ve mutfak seçimi gibi genel seçimlerden oluşur. Bu kısıtlama tipi oluşturulacak planın bir parçasını değil tümünü etkilemektedir.

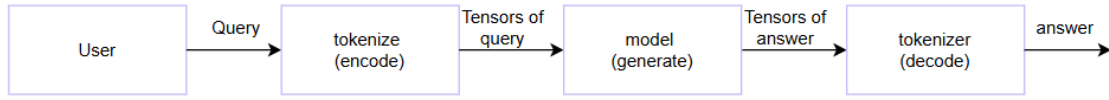
Bir diğer kısıtlama tipi olan environment, başlıca dikkat edilmesi gereken kriterlerden biridir. Hedeflenen güzergahın ulaşım için uygunluğunu kontrol edilmeli ve güzergah üzerinde etkinliğin varlığı kontrol edilmeli.

Hard ve environment constraint bölümünde belirtilen şartlar sağlandığında, oluşturulacak planın olabildiğince kullanıcıya yönelik olması için çalışılır. Burada farklı restoranların ve etkinliklerin önerilmesi gibi durumlar yer almaktadır. Uygulama; oluşturduğu planda, kullanıcı sorgusunda tespit ettiği verilerle birden fazla veri elde edip, olabildiğince renkli bir plan oluşturmalı.

Makale analizi ve yapılan toplantılar sonucunda tespit edilen uygulama gereksinimleri belirlendikten sonra, geliştirilecek olan projenin teknik tarafı üzerine araştırma yapılmaya başlandı.

Uygulamanın temelde metin formatında bir veri alıp, buna metin formatında bir cevap üretmesi hedeflenmektedir. Metin üretme işlemi, dil modelleri tarafından gerçekleştirilebilir.

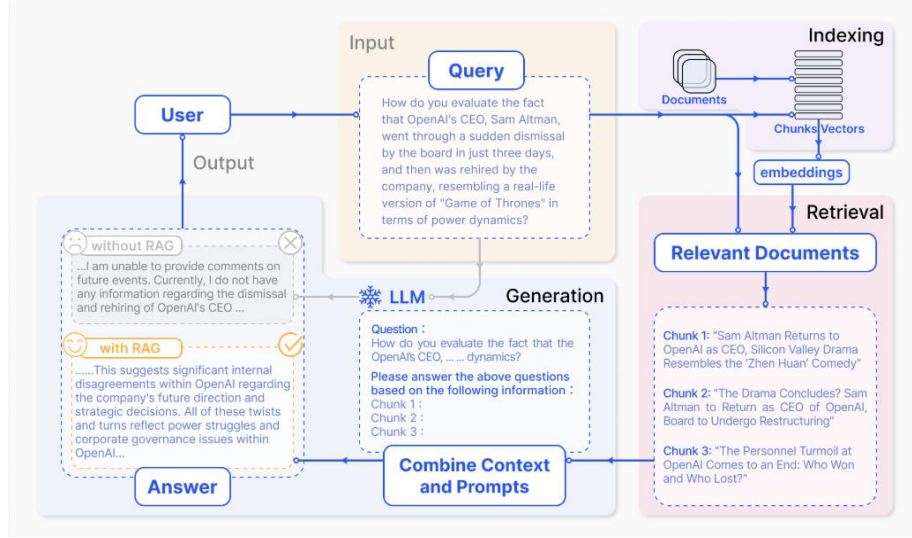
Büyük dil modeli, temelinde önceden eğitilmiş olduğu verilerden yola çıkarak kendisine verilen sorguya göre, bir sonraki kelimeyi tahmin ederek yeni metin üreten bir yapıya sahiptir. Projenin test edileceği tipteki sorgular bir büyük dil modeline verildiğinde ve Şekil 3.1.1.4'te gösterildiği üzere basit bir süreçten geçirildiğinde bir cevap alınabilir. Yalnız uygulamanın plan üretirken kullanacağı, uçuş, restoran, konaklama ve etkinlik gibi verilerin güncel olması gerekmektedir ve bu süreçte güncellik mümkün değildir.



Şekil 3.1.1.4 Metin Üretimi Süreci

Büyük dil modelleri, yalnızca yayımlandığı yıldan önceki verilerle metin üretimi yapmaktadır. Örneğin, Huawei server'larında yüklü olan modellerden biri olan LLaMA[3], 2023'te yayımlanmış bir modeldir ve oldukça geniş bir veri kümesiyle eğitilmiş olmasına rağmen 2023 sonrası veriler hakkında bilgi içermez. Dolayısıyla LLaMA'ya doğrudan bir sorgu verildiğinde, cevap olarak güncel bir plan üretilemez ya da üretilse dahi halüsinasyon içeren bir plan üretilmiş olur.

Büyük dil modelleriyle metin üretimi için kullanılabilecek bir diğer teknik olan RAG(Retrieval Augmented Generation), Şekil 3.1.1.5'te gösterilmektedir. RAG ile geliştirilen bir Travel Planner uygulaması sağlanan veri seti aracılığıyla dış dünya güncel verilerine erişim sağlayabilir ve bu uygulamaya verilen sorgu büyük dil modeline verilmeden önce; sorguyla alakalı güncel veriler, uygulamaya sağlanan veri sertinden alınır ve büyük dil modeline yalnızca sorgu değil güncel veriler de beslenir, böylelikle büyük dil modelinin güncel bir plan oluşturması sağlanır.



Şekil 3.1.1.5 RAG

RAG ile büyük dil modelinin güncel yanıtlar üretebilmesi sağlanabilir, yalnız geliştirilmesi hedeflenen uygulamada ulaşılması gereken tek hedef güncellik değildir. Uygulama aynı zamanda raporun ilk kısımlarında bahsedilen başlıca kısıtlamalara(restoran, uçuş, konaklama, etkinlik, bütçe) ve ek kısıtlamalara (çeşitlilik, oda tipi, min. gece sayısı vb.) uyarak bir yanıt üretmesi gerekmektedir.

Bir büyük dil modeli, bir sonraki kelimeyi tahmin ederek metin üretimi yapar. Oldukça büyük bir veri kümesiyle eğitilmiş olan büyük dil modelleri base model olarak adlandırılır. Base model kendisine verilen metindeki her bir token'ı işler ve bir sonraki kelime için bir olasılık dağılımı oluşturarak bir sonraki kelimeye tahmin eder. Böylelikle kendisine verilen metni tamamlayarak çıktı verir.

Büyük dil modelleri yalnızca metin tamamlayarak karmaşık görevleri çözümleyemez. Bu sebeple base modeller hedeflenen görevler için fine-tune edilir. Büyük dil modelleri komut veri setiyle fine-tune edildiğinde Instruct model ismini alır. Bu modeller base modellerin komut veri setleriyle fine-tune edilmiş halleridir ve metin tamamlamak yerine kendisine verilen komutu gerçekleştirmeye odaklıdır.

Büyük dil modelleri sorgu-cevap çiftleriyle fine-tune edildiğinde ise chat model ismini alır. Bu modeller ise kendisine verilen sorguya bir cevap üretmeye odaklıdır.

BURSA TEKNİK ÜNİVERSİTESİ  
PROJE TABANLI İŞLETMEDE MESLEKİ EĞİTİM  
BTÜ-İMEP ÖĞRENCİ FİNAL RAPOR FORMU F4

Bir büyük dil modelinin base, instruct ya da chat model versiyonlarından hangisi seçilirse seçilsin büyük dil modeli genel amaçlıdır. Bu sebeple karmaşık görevlerde, büyük dil modelinin ek işlemlerle belirli bir göreve odaklanması gerekir.

Bir büyük dil modelini belirli bir göreve odaklamanın yollarından biri fine-tune yöntemidir. Travel Planner projesinde bir büyük dil modelini travel planner sorguları ve cevaplarıyla fine – tune ederek hedeflenen sonuca ulaşmak mümkün olabilir. Yalnız fine – tune etmek oldukça maliyetli bir yöntemdir. Uzun bir eğitim süreci gerektirir ve böylelikle uzun süre GPU kullanımına ihtiyaç duyar. Ayrıca bir büyük dil modelinden yeni bir büyük dil modeli oluşturulmasına sebep olduğu için server’da da geniş bir yer kaplar . Bu sebeple fine-tune etmek bu aşamada mantıklı görünmemektedir.

Bir büyük dil modelini belirli bir göreve odaklamanın oldukça basit bir yöntemi ise Prompt Mühendisliği’dir. Prompt Şekil 3.1.1.6’da gösterildiği gibi doğal dil formundadır ve oluşturulması için büyük dil modelinin hakim olduğu dillerden (başlıca İngilizce) birini biliyor olmak yeterlidir. Farklı görevler için farklı teknikte prompt’lar kullanılabilir. Temel olarak prompt, büyük dil modeline görevini, yapacağı işlemleri söyleyen metinleri; oluşturması gereken çıktıyı ve çıktı tipini, gerekiyorsa bir ya da daha fazla örneği içerir. Böylece büyük dil modeli karmaşık bir işlem gerçekleştirebilecek şekilde özelleştirilebilir.

```
PLANNER_INSTRUCTION = ""You are a proficient planner. Based on the provided information and query, please give me a d

***** Example *****
Query: Could you create a travel plan for 7 people from Ithaca to Charlotte spanning 3 days, from March 8th to March 10th?
Travel Plan:
Day 1:
Current City: from Ithaca to Charlotte
Transportation: Flight Number: F3633413, from Ithaca to Charlotte, Departure Time: 05:38, Arrival Time: 07:46
Breakfast: Nagaland's Kitchen, Charlotte
Attraction: The Charlotte Museum of History, Charlotte
Lunch: Cafe Maple Street, Charlotte
Dinner: Bombay Vada Pav, Charlotte
Accommodation: Affordable Spacious Refurbished Room in Bushwick1, Charlotte

Day 2:
Current City: Charlotte
Transportation: -
Breakfast: Olive Tree Cafe, Charlotte
Attraction: The Mint Museum, Charlotte;Romare Bearden Park, Charlotte.
Lunch: Birbal Ji Dhaba, Charlotte
Dinner: Pind Balluchi, Charlotte
Accommodation: Affordable Spacious Refurbished Room in Bushwick1, Charlotte

Day 3:
Current City: from Charlotte to Ithaca
Transportation: Flight Number: F3786167, from Charlotte to Ithaca, Departure Time: 21:42, Arrival Time: 23:26
Breakfast: Subway, Charlotte
Attraction: Park, Bushwick, Charlotte
```

Şekil 3.1.1.6: Prompt Örneği

Yapılan araştırmalar sonucu projede bulunması gereken diğer özellikler aşağıdaki şekilde belirlenmiştir:

- Sorguyu; uçuş, restoran, konaklama ve etkinlik durumlarına göre değerlendirmeli ve gerekli aksiyonları almalı.
- Sorgudaki gerekli bilgileri analiz edip, plan üretmek için güncel uçuş verilerinin toplanması gerekiyor.
- Seyahat güzergahındaki şehirlerin analiz edilip bu şehirlerdeki güncel konaklama, etkinlik ve restoran verilerinin toplanması gerekiyor.
- Toplanan veriler içerisinde öneri yaparken, sorguda yer alan alt kısıtlamaların analiz edilmesi ve buna uygun öneriler yapılması gerekiyor.
- Yapılan önerilerde çeşitliliğe önem verilmeli ve ayrıca proje yapılan her bir öneride sorgudan analiz edilen bütçe de dikkate alınmalı.

Oluşturulması hedeflenen proje RAG tekniği ile geliştirildiğinde plan oluşturmak için gereken gerçek ve güncel veriler sisteme dahil edilebilir. Yalnız bu süreçte bir büyük dil modelinin seyahat planı oluşturacak şekilde özelleştirilmesi için uygun prompt'un hazırlanması gerekmektedir. Tek bir prompt oluşturulmaya başlandığında, büyük dil modeline üstte listelenmiş olan kısıtlamaların her biri aktarıldığında ve ek olarak toplanan güncel verilerin de buna eklenildiğine oldukça uzun bir prompt oluşmaktadır. Büyük dil modelin tek bir metin üretimi sürecinde birden fazla muhakeme noktasına odaklanması gerekir. Büyük dil modelleri tek başına reasoning süreçlerinde oldukça başarısızdır ve ayrıca büyük dil modeline verilecek olan prompt çok geniş olduğunda, büyük dil modeli tüm prompt'a dikkatini vermek yerine prompt'un belirli kısımlarına odaklanır ve böylelikle, belirli koşulları sağlayan, bir çok koşulu da sağlayamayan bir plan üretilir.

Yüksek muhakeme gerektiren doğal dil işleme görevlerinde agent'lar diğer tekniklere göre daha başarılıdır. Agentic sistemler, dahil edilen tool'lar aracılığıyla dış dünya verilerine erişerek büyük dil modellerinin güncel ve gerçek cevaplar üretmesini sağlar. Agentic sistemler aynı zamanda birden fazla reasoning gerektiren süreçlerin her biri için ayrı agent'ların oluşturulmasını sağlayarak, büyük dil modelinin kısa ve öz prompt'lar ile tek bir göreve odaklanmasını sağlar. Böylelikle ana görevin her bir alt parçasını ayrı bir agent yüksek doğrulukla tamamlar ve sürecin sonunda oluşturulan cevabın istenilen koşulları sağladığı gözlemlenebilir.

Stajın başlangıç döneminde yapılan kapsamlı araştırmalar sonucunda elde edilen, üstte bahsedilen bilgiler raporlandı. Referans makalede yer alan kod analiz edildi ve çalıştırıldı. Referans makalede geliştirilen multi-agentic sistem incelendi ve raporlandı. Bunun sonucunda agentic sistemler üzerine odaklanılmasına ve projenin agentic sistemler ile geliştirilmesine karar verildi. Referans makalede yer alan kod analiz edildi, çalıştırıldı ve raporlandı. Referans makalede multi-agent ile tasarlanan projede agent'lar arasındaki iletişimin sağlanması ve yönetilmesi için ek bir framework kullanılmamış, bilgi aktarımı için dosyalar, akış için ise döngüler ve yinelemeli fonksiyonlar kullanılmıştır; böylelikle kod oldukça karmaşık bir hale bürünmüştür. Referans makaledeki projede büyük dil modeli olarak OpenAI kullanılmıştır. Kodun çalıştırılması sürecinde OpenAI yerine server'da bulunan LLaMA, Mixtral, Gemma gibi modellerle denemeler yapıldı ve makaledeki bu karmaşık kod oldukça yavaş çalıştı. Bu nedenle, geliştirilmesi hedeflenen multi-agentic sistemin LangGraph çerçevesi ile yönetilmesine karar verildi.

Raporlama sürecinden sonra edinilmesi hedeflenen bilgi birikimleri aşağıda listelenmiştir:

- Agent – Multi Agentic Sistemler
- LangGraph Çerçevesi (Framework)
- Prompt Teknikleri ve Agentic Sistemlerde Tool Çağrılması
- Farklı Açık Kaynak LLM'ler ve Farklı Parametre Sayısına Sahip Versiyonları
- Agentic Sistemler ve Arayüz Bileşimi

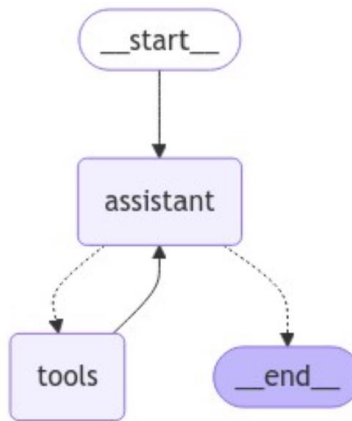


### 3.1.2. Yöntem

Hedef bölümünde aktarılan raporlama sürecinde Agent kavramı üzerinde belirli düzeyde bilgi birikimi elde edildi. Agent ve Multi-agentic sistem kavramlarında uzmanlaşmak için, açık kaynak kurslar tamamlandı ve küçük ölçekli deneyler yapıldı. Aynı süreç LangGraph çerçevesi üzerinde de gerçekleştirildi ve Travel Planner üzerine bazı denemeler yapıldı.

#### 3.1.2.1. Multi – Agentic Yaklaşım ve Grafların Oluşturulması

Multi-agent ve LangGraph konusunda yeterli bilgi birikimi elde edildiğinde graf tasarımı sürecine geçildi. Graf tasarımı için, problemin oldukça karmaşık olması nedeniyle, yapılan araştırmalar sonucunda şekilde gösterildiği üzere ReAct(Reasoning and Action) yönteminin kullanılmasına karar verildi. ReAct yönteminde, assistant ve tools olmak üzere iki farklı agent bulunmaktadır. Assistant; reasoning sürecini yürütürken, tools node'u action sürecini tamamlar. Reasoning sürecinde grafa verilen sorgu üzerine değerlendirme yapılarak, toplanılması gereken bilgiye ve böylelikle çağırılması gereken tool'a karar verilir. Assistant node'unun her iterasyon için işlemini tamamlamasının ardında action sürecine geçilir ve bilgi toplanır. Tool agent'ı, multi agentic sisteme tanımlanan tool'ları içerir ve reasoning adımı büyük dil modeli tarafından alınan kararlara göre çalışır. Reasoning adımı yalnızca tool'a karar vermez, aynı zamanda yeterli bilginin toplanıp toplanmadığına karar vererek grafı sonlandırabilir.

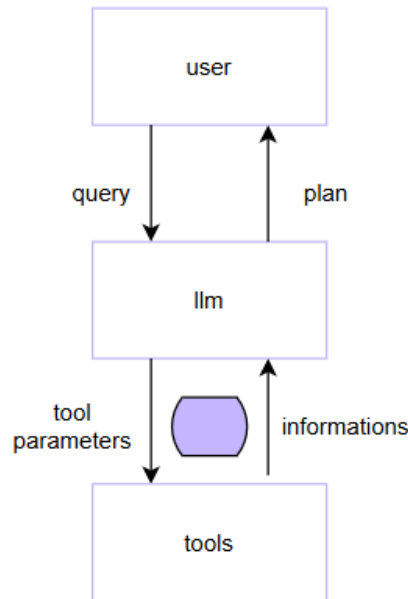


Şekil 3.1.2.1: ReAct(Reason and Action)

### 3.1.2.2. Graf Oluşturulması

LangGraph ile oluşturulan bir grapta her bir Agent, node isimli yapılara yerleştirilir. Üstteki grapta assistant ve tools node'ları bulunmaktadır.

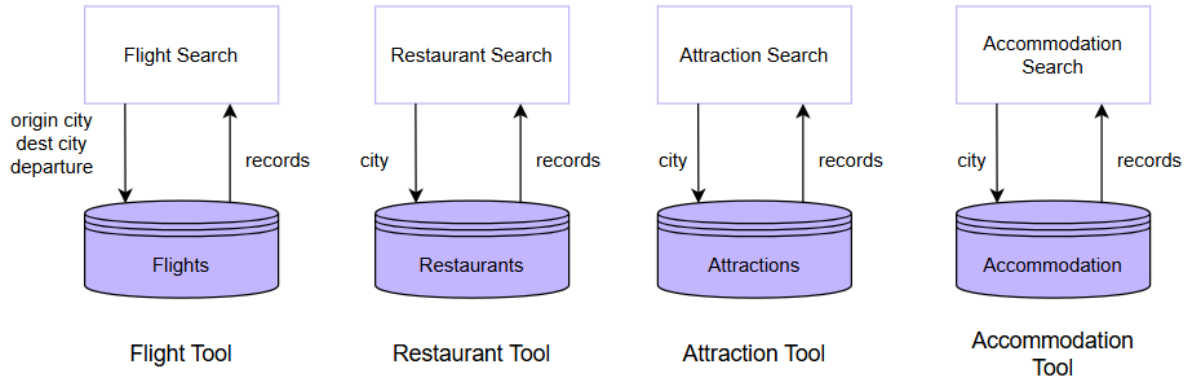
Assistant node'u Şekil 3.1.2.2'deki akış diyagramı gösterilen mantık ile hareket eder, kendisine verilen sorguyu defalarca analiz eder. Her bir analiz de kullanılması gereken farklı bir tool'a ve bunlara verilmesi gereken parametrelere karar verilir. Sorgu üzerinde yeterli analiz yapıldığında assistant node'u grafi sonlandırır.



Şekil 3.1.2.2: ReAct(Reason and Action)

Tasarlanan süreç için öncelikle tool'lar tanımlandı. Tool'lar graf akışında, agentic sistemlerin dış dünyayla iletişimini sağlayan yapılardır. Travel Planner projesi için 4 farklı tool tanımlandı ve bunların her biri fonksiyon formatında tanımlandı. Tool'ların görevi kendisine verilen parametreye göre, yine kendilerine tanımlanan veri tabanından veri çekmek ve bunu yine assistant node'una teslim etmektir.

Dört farklı tool için tanımlanan fonksiyonların algoritması, Şekil 3.1.2.3'te özetlenmiştir.



Şekil 3.1.2.3: Tool Tanımlamaları

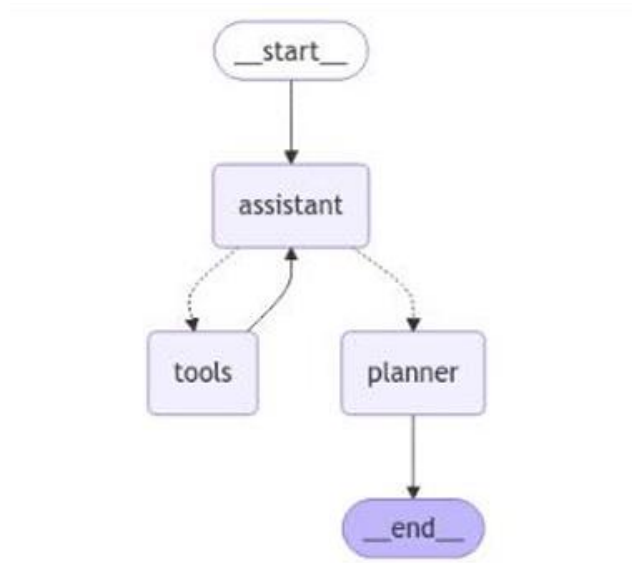
Multi-agentic sistemlerde tool'lar farklı tekniklerle tanımlanabilir. Bir Travel Planner projesi için dış veri kaynağı olarak database yerine API kullanmak daha mantıklı görünebilir. Yalnız Huawei bilgisayarlarında geliştirilen projede, internet ortamından veri çekme süreci proxy nedeniyle gerçekleştirilemedi. Bundan dolayı, Kaggle ortamından her bir tool için uygun veriler bulundu ve veri setleri oluşturuldu.

Tool'ların gerçekleştirebileceği tek işlem bilgi toplamak değildir. Tanımlanan tool'lar uygulama için gerekli başka işlemler de tamamlayabilir ve bu süreçte büyük dil modellerinden faydalanabilir. Bu aşamada proje gereklilikleri için tanımlanan 4 tool yeterlidir.

Tool'ların tanımlanması işlemi tamamlandıktan sonra oldukça önemli kısım, assistant ve tools node'ları arasında oluşturulacak döngüde her bir iterasyonda, sorgunun iyi bir şekilde analiz edilip doğru tool'un doğru parametreyle çağrılmasıdır. Bunun için büyük dil modeli kullanılarak, metin üretme süreçleri üzerine kapsamlı araştırmalar yapıldı. Ağırlıklı olarak Transformer dokümantasyonu üzerine çalışıldı.

### 3.1.2.2. Farklı Graflar İle Yapılan Denemeler

Geliştirme sürecinde öncelikle ChatOpenAI sınıfına oldukça benzer olan ChatGroq sınıfı aracılığıyla 8B parametrelili LLaMA modeline bağlanılarak Şekil 3.1.2.4'te gösterilen graf oluşturuldu. Üstte özetlenen ReAct yapısı korunarak ek olarak planner agent'ı eklendi. Planner Agent, tool çağırma işleminin ardından toplanan tüm bilgiler ile plan üretecek prompt ile özelleştirildi.



Şekil 3.1.2.4: Groq ile Oluşturulan Graf

Groq, API aracılığıyla bir büyük dil modeline bağlanıp GPU olmadan büyük dil modeli kullanımını sağlamaktadır. Groq ile geliştirilen bu proje Huawei bilgisayarında değil, kişisel bilgisayar ile geliştirildi.

ChatGroq ile bir büyük dil modeline bağlanıldığında, metin üretme süreçleri kullanılan invoke fonksiyonu ile arka planda gerçekleştirilmektedir. Tool bağlama ve çağırma süreçleri de yine aynı pakette tanımlı olan hazır fonksiyonlar ile gerçekleştirilmektedir. Dolayısıyla daha az kod ile hızlı bir şekilde bir örnek uygulama oluşturuldu. Yalnız geliştirilmesi gereken ana proje, Huawei bilgisayarında olacağı için bu örnek uygulamada olduğu gibi Groq ile bir büyük dil modeli kullanımı söz konusu değildir. Huawei server'larında yüklü olan büyük dil modellerinden birinin kullanılması gerekmektedir. Huawei server'larında LLaMA, Mistral, Gemma gibi modellerin çeşitli parametre sayılarındaki versiyonları bulunmaktadır. Bu modeller Hugging Face modelleri olduğu için Transformer kütüphanesi ile yüklenmesi ve Transformer kütüphanesindeki metin üretim süreçleriyle üretim yapmaları gerekmektedir. Bunun üzerine Transformer kütüphanesi üzerine araştırmalar yapılmaya devam edildi.

Transformer kütüphanesi ile geliştirilecek ReAct temelli Travel Planner projesi için Şekil 3.1.2.5'te gösterilen graf tasarlandı .



Şekil 3.1.2.5: Groq İle Oluşturulan Graf

Şekil 3.1.2.5'teki graf, Groq ile geliştirilen graftan farklıdır. Farklılığın sebebi akış mantığının değiştirilmesi değildir. Proje yine aynı süreçlerden geçmektedir. Bu fark Groq ile tasarlanan projede kullanılan hazır fonksiyonların her biri yerine ayrı fonksiyonlar yazılmasından kaynaklanmaktadır. Örneğin her grafta kesikli çizgi ile gösterilen yapılar, node'un iki farklı node'a geçebilme ihtimalini temsil eder. Bu iki farklı node'dan birinin seçilmesi için ek bir koşul fonksiyonu gereklidir. Groq ile geliştirilen modelde, kullanılan ReAct yapısı ve grafa bağlanan node'lar, bu node'lar arasındaki iletişim, bu node'ların grafa bağlanması hazır fonksiyonlar ile gerçekleştirilmiş ve bunların kullanımı da arka planda gerçekleşmektedir.

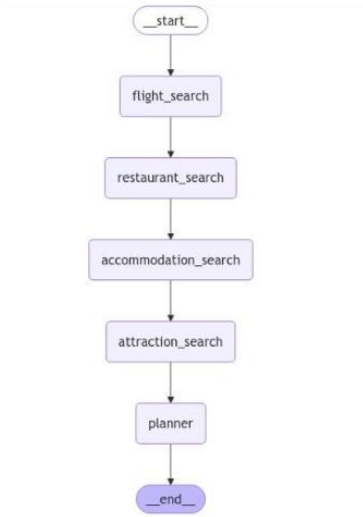
Transformer ile geliştirilen yöntemde ise yeni bir tool çağrısı oluşturma ya da planner node'una geçme durumu Groq ile geliştirilen versiyonda olduğu gibi büyük dil modeli kullanılarak tamamlanır. Yalnız büyük dil modeli çıktısının ayrıştırma süreci Groq ile geliştirilen versiyonda olduğu gibi arka planda gerçekleştirilmez. Tools node'una geçildiğinde, assistant node'undan gelen cevabın bir tool çağrısı olup olmadığı kontrol edilir. Tool çağrısı olma durumunda gerekli tool kullanılır ve bilgiler toplanır. Yalnız büyük dil modeli eğer bir tool çağrısı oluşturmadıysa, bu analiz edilir ve graf akışı planner node'una geçer.

Huawei server'larında hızlı çalışan bir uygulama geliştirmek için 3B model kullanılması tercih edildi. 3B model ile geliştirilen bu proje istenilen sonuca ulaşamadı. Graf çalıştırıldığında, sürekli aynı tool'un çağrılmasıyla bir sonsuz döngüye girildi ve 25 iterasyon sonucu LangGraph[5] akışı sonlandırdı.

Bunu üzerine kapsamlı araştırmalar yapıldı, metin üretme süreçleri incelendi. apply\_chat\_template[6] gibi farklı metin üretme süreçleri üzerine denemeler yapıldı. Grafın çalışmamasının sebebi, Transformer [7]kütüphanesi ya da akışta yapılmış olması muhtemel herhangi bir hata değildir. Groq[8] ile geliştirilen projede 8B model kullanıldı. Transformer ile geliştirilen projede ise 3B model kullanıldı. Yapılan araştırmalar sonucunda 3B modelin tool çağırma süreci için yetersiz kaldığı tespit edildi.

Projenin tamamlanması için prompt'lar defalarca değiştirildi. LangGraph'ın State sınıfı değiştirilerek graf içerisinde taşınan bilgiler arttırıldı. Tüm denemelere rağmen sonuçta bir iyileşme olmayınca, her bir tool için ayrı bir Agent tamamlanılmasına karar verildi ve Şekil 3.1.2.6'da gösterilen graf tanımlandı.

Tanımlanan grafta her bir Agent farklı bir tool'u çağırmak üzere prompt aracılığıyla özelleştirilmiştir. Bu graph ile 3B model ile gerekli tüm bilgilerin toplanılması sağlanarak, istenen şekilde bir plan oluşturdu.



Şekil 3.1.2.6: Transformer İle Oluşturulan Graf

Yapılan tüm denemeler sonucunda bir sunum yapıldı ve Prompt Mühendisliği ve farklı büyük dil modellerinin farklı büyüklükteki versiyonları hakkında bir çalışma yapılmasına; proje için bir arayüz oluşturulmasına karar verildi .

### 3.1.2.3. Prompt Mühendisliği ve Büyük Dil Modelleri

Bu sürece kadar 3 farklı graf oluşturuldu. Bu grafların istenilen şekilde çalışması için onlarca farklı prompt ile denemeler yapıldı. Transformer ve ReAct formunda tanımlanan projenin istenilen şekilde çalışmaması nedeniyle öncelikle, geliştirilen fonksiyonlar değiştirildi; metin üretme için farklı stratejiler denendi. Transformer dokümantasyonunda yer alan her bir stratejinin denenmesiyle uygulama yine çalışmayınca prompt'lar değiştirildi.

Açık kaynak olarak yayımlanmış olan farklı kaynaklardan Prompt Mühendisliği üzerine bir uzmanlık kazanıldı. Farklı görevler için farklı prompt tekniklerinin kullanılması gerektiği öğrenildi. Llama, Mistral[9] ve Gemma[10] modelleri için prompt tekniklerinin farklılıkları öğrenildi.

Llama, Mistral ve Gemma modellerine verilmesi gereken prompt'ların farklı olmasının sebebi bunların metin üretme süreçlerinde farklılıklar bulunmasıdır. Tool çağırma sürecinde, kullanılan metin üretme işleminde kullanılan `apply_chat_template`, bu modellerin dosyalarında bulunan bazı önemli parametreleri referans alarak üretim yapar. Bu parametrelerin değiştirilmesi ya da bu parametrelere göre yapılan değişiklikler ile farklı sonuçlar elde edebilir.

Temperature, Top P, Max Length, Stop Sequences, Frequency Penalty ve Presence Penalty[11] gibi parametreler değiştirilerek, ara çıktı ve nihai plan üzerinde değişiklikler elde edildi ve raporlandı.

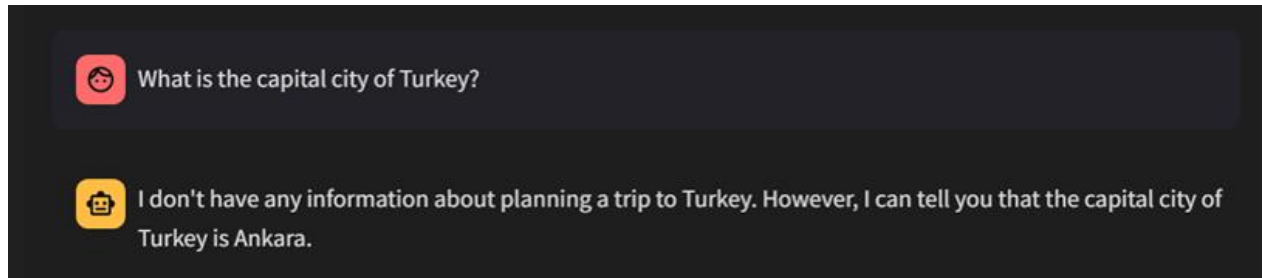
Farklı prompt teknikleri ve farklı modellerle yapılan denemeler raporlandı ve Transformer yöntemiyle oluşturulmuş son graf için bir Streamlit[12] kodunun yazılmasıyla uçtan uca bir proje teslim edildi ve test edildi. Yapılan testler sonucu projede bazı hatalar olduğu raporlandı ve bu hatalar üzerine çalışılmaya devam edildi.



### 3.1.2.2. Projedeki Hataların Çözülmesi ve Arayüzün İyileştirilmesi

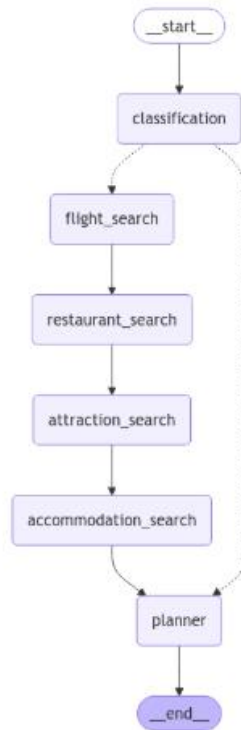
Tamamlanan proje danışmanım tarafından test edildi ve Travel plan sorgularında tool çağırma işlemlerinin ve nihai çıktının doğru bir şekilde tamamlandığı tespit edildi. Yalnız projeye seyahat planı istemeyen sorgular verildiğinde proje yine tool çağırısı yapıyordu. Tool çağırısı yaptıktan sonra, doğru cevap verse de bu aşamada tool çağırısı yapmak yerine uygun bir cevap vermesinin daha profesyonel görüneceğine karar verildi.

Travel Planner projesinin genel sorgulara nasıl yanıtlar verdiği incelendi. Bir sorgu planner prompt'u aracılığıyla özelleştirilmiş agent'a, tool'lar tarafından herhangi bir bilgi toplanılmadan verildiğinde, planner agent yine bir Travel planner gibi davranarak cevap vermektedir. Örneğin Şekil 3.1.2.7'de görüldüğü üzere genel bir sorgu gönderildiğinde, büyük dil modeli hem sorguya cevap vermekte, hem de planner prompt'u ile özelleştirildiği için Travel Planner uygulaması olduğunu da kullanıcıya belirtmektedir. Bundan dolayı grafın sonlandırılması aşamasında, bu tip genel sorguları cevaplamak üzere yeni bir Agent oluşturulmasına gerek kalmadı. Yalnızca tool çağırma işlemine geçilmeden önce sorunun genel ya da travel tipinde olduğunu tespit etmesi için bir Classification Agent oluşturuldu. Yeni graf Şekil 3.1.2.8'de gösterilmektedir.



Şekil 3.1.2.7: Projenin Genel Sorgu İle Test Edilmesi

Classification herhangi bir tool kullanmaz ve basit bir sınıflandırma işlemi yapmak üzere özelleştirilmiştir. Few shot prompting tekniği kullanılarak, 2 genel sorgu 1 tane de plan sorgusu içeren örnekler ve büyük dil modeline görevini söyleyen bir prompt hazırlandı. Yapılan gözlemler sonucu büyük dil modeli kendisin verilen sorguları doğru bir şekilde sınıflandırdı. Yalnız, büyük dil modelin oluşturduğu cevaplar belirli bir kalıpta değildi, bu da graph'ın yönetilmesini zorlaştırdı. Classification node'undan sonra bir router oluşturulması ve graf akışının yönlendirilmesi gerekmektedir. Bu yönlendirme tanımlanan fonksiyon ve state e kaydedilecek bir değişken ile sağlanabilir. Yalnız büyük dil modeli her üretimde farklı bir metin oluşturacağı için, farklı metinleri yönetmek yerine bu durumu tek bir token ile yönetmek oldukça kolaydır. Bunun üzerine Classification node'unda kullanılmak üzere diğer agent'larda kullanılan metin üretme fonksiyonu değil, ayrı bir fonksiyon tanımlandı ve özellikle max\_tokens değişkeni buna uygun ayarlanarak büyük dil modelinin tek token'dan oluşan cevaplar üretmesi sağlandı.



Şekil 3.1.2.8: Projenin Genel Sorgu İle Test Edilmesi

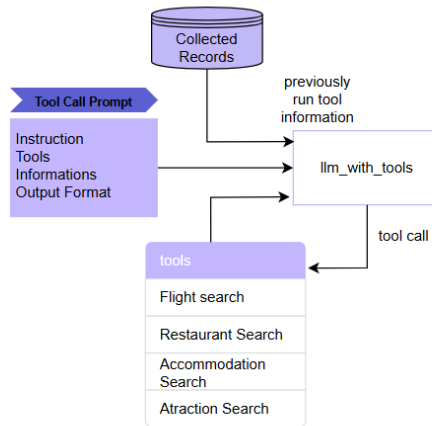
### 3.1.3. Çalışmanın Sonuçları

#### 3.1.3.1 Son Projenin Proje Yapısı

Yapılan tüm çalışmalar sonunda Travel Planner projesi tamamlandı ve direktör ve yapay zeka birim yöneticisine bir sunum gerçekleştirildi. Oluşturulan tüm kodlar server'da bulunurken geliştirilen uçtan uca projenin nihai kodları Huawei kod ortamına kapsamlı bir dokümantasyon hazırlanarak aktarıldı. Yapılan çalışmalarda ara çıktıların gözlemlenmesi için notebook ile çalışıldı ama nihai proje py scriptleri ile çalışmaktadır.

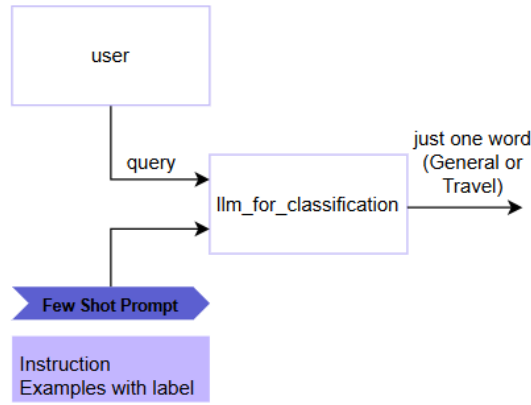
Huawei kod ortamına yüklenen nihai projeni alt script'leri aşağıda listelenmiştir:

- llms.py: Bu dosyada metin üretme işlemleri yapan fonksiyonlar bulunur. Travel Planner projesinde 3 farklı strateji ile metin üretilmiştir ve bunun için 3 farklı fonksiyon hazırlandı.
  - Fonksiyonlardan biri llm\_with\_tools; tool çağıran agent'lar tarafından kullanılır ve çıktısı json formatında bir tool çağrısıdır.



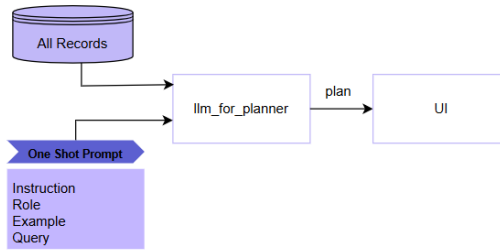
Şekil 3.1.2.9: llm\_with\_tools Akış Diyagramı

- Bir diğer metin üreten fonksiyon, sorguyu sınıflandırmaya yarayan `llm_for_classification` fonksiyondur. Kendisine sağlanan prompt, `apply_chat_template` ve gerekli parametrelerle birlikte sınıflandırma işlemini tamamlar.



Şekil 3.1.2.10: `llm_for_classification` Akış Diyagramı

- Bir diğer fonksiyon ise `llm_for_planner` fonksiyonudur. Bu fonksiyonun `max_tokens` parametresi diğer fonksiyonlara göre daha büyük bir değerde tutulmuştur ve böylelikle yeterli sayıda token içeren planlar başarıyla üretilmiştir.



Şekil 3.1.2.11: `llm_for_planner` Akış Diyagramı

- `prompt.py`: Bu dosyada tanımlanan prompt'lar bulunur

Metin üretme süreçlerinde bahsedilen prompt'lar bu script'e kaydedilmiştir. Her bir prompt few shot, one shot gibi farklı teknikler ile hazırlanmıştır, ayrıca bu prompt'ların nihai formlarına deneme yanılma yoluyla ulaşılmıştır.

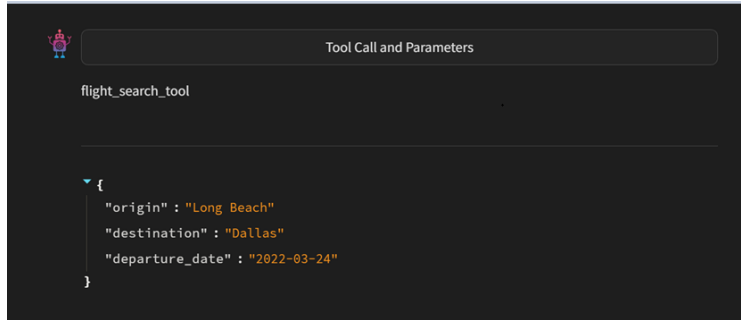
- `tools.py` : Tüm tool'lar buraya kaydedilmiştir.
- `graph.py` : Graph'ların birbirine bağlanması, `add_node`, `add_edge` işlemleri bu script'e kaydedilmiştir. Ayrıca verilen sorguyla birlikte grafın çalıştırılması için bir `invoke_graph` fonksiyonu tanımlanmıştır.
- `routers.py`. Graf içerisinde iki farklı node'dan birini tercih etme durumunu kontrol eden fonksiyon bu script'e kaydedilmiştir.
- `graph_components.py`: Graf içerisinde bilgi akışını sağlayan state değişkeni ve agent'ları temsil eden ve büyük dil modeli kullanan node'lar burada tanımlanmıştır
- `app.py`: Streamlit arayüzün oluşturulduğu, grafın çağrıldığı ve ara çıktıları ekranda gösteren kodlar buraya kaydedilmiştir. Streamlit kodları ile grafın arayüzde tekrar tekrar kullanılması sağlandı; kullanıcı etkileşimli zengin bir arayüz oluşturuldu.

### 3.1.3.1 Son Projenin Arayüzü

Tamamlanan projede eğer sorgu genel sorgu ise, bir büyük dil modeli iki defa; eğer sorgu travel sorgusu ise 6 defa çalışmaktadır. Bu da bir sorguya verilen yanıt süresinin uzun olmasına neden olmaktadır. Kullanıcı arayüzünün aktif olması için, grafta gerçekleştirilen her bir adımın sırasıyla kullanıcıya yansıtılması hedeflenmiştir. Bunun için, Streamlit dokümantasyonu incelendi ve LangGraph-Streamlit çalışmaları yapıldı.

LangGraph'ta ara çıktılar astrem fonksiyonu ile kullanıcıya sunulabilir, yalnız bu fonksiyon her bir ara çıktının tamamlandığı anda kullanıcıya gösterilmesini sağlamaz. Bunun üzerine her bir agent'a, router'a kısacası her bir graf bileşenine Streamlit kodları da eklendi. Böylelikle büyük dil modeli cevapları, tool çağrıları, toplanan bilgiler oluşturulduğu anda kullanıcıya canlı bir arayüz ile yansıtıldı.

Travel sorgusu verilmesi durumunda oluşturulan her bir tool çağrısı aşağıdaki gibi ekrana yansıtılmaktadır.



Şekil 3.1.2.11: Tool Çağrıları

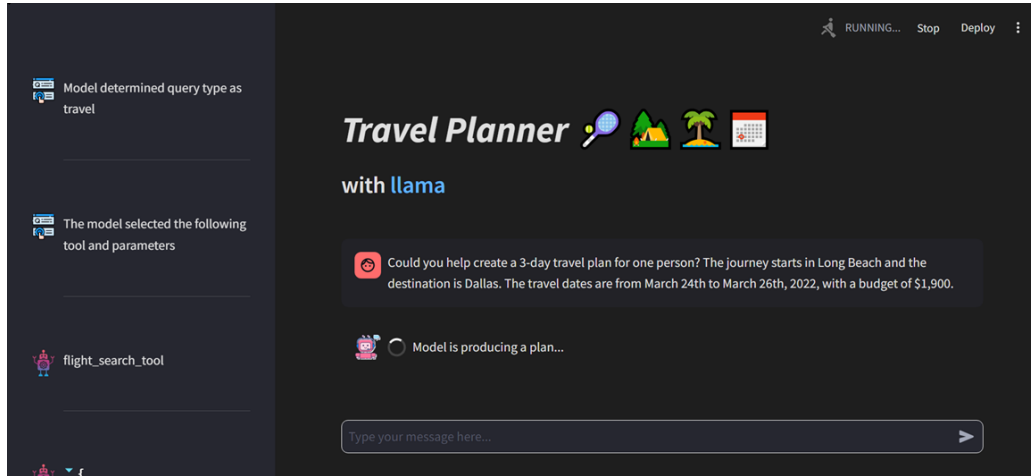
BURSA TEKNİK ÜNİVERSİTESİ  
PROJE TABANLI İŞLETMEDE MESLEKİ EĞİTİM  
BTÜ-İMEP ÖĞRENCİ FİNAL RAPOR FORMU F4

Çağrılan tool'un bilgi toplama süreci tamamlandığında bilgiler ekrana yansıtılır.

	Name	Latitude	Longitude	Address	Phone	Website
0	The Dallas World Aquarium	32.7835	-96.8054	1801 N Griffin St, Dallas, TX 75202, USA	(214) 720-2224	<a href="https://www.dwazoo.com/">https://www.dwazoo.com/</a>
1	The Sixth Floor Museum at Dealey Plaza	32.7798	-96.8085	411 Elm St, Dallas, TX 75202, USA	(214) 747-6660	<a href="https://www.jfk.org/">https://www.jfk.org/</a>
2	Reunion Tower	32.7755	-96.8089	300 Reunion Blvd E, Dallas, TX 75207, USA	(214) 296-9950	<a href="http://www.reuniontower.com">http://www.reuniontower.com</a>
3	Dallas Museum of Art	32.7877	-96.801	1717 N Harwood St, Dallas, TX 75201, USA	(214) 922-1200	<a href="https://www.dma.org/">https://www.dma.org/</a>
4	The Dallas Arboretum and Botanical Garden	32.8236	-96.7166	8525 Garland Rd, Dallas, TX 75218, USA	(214) 515-6615	<a href="https://www.dallasarboretum.com">https://www.dallasarboretum.com</a>
5	Dallas Zoo	32.741	-96.8153	650 S R.L. Thornton Fwy, Dallas, TX 75203, USA	(469) 554-7501	<a href="https://www.dallaszoo.com/">https://www.dallaszoo.com/</a>
6	Klyde Warren Park	32.7894	-96.8016	2012 Woodall Rodgers Fwy, Dallas, TX 75201, USA	(214) 716-4500	<a href="http://klydewarrenpark.org/">http://klydewarrenpark.org/</a>
7	Old City Park	32.7728	-96.7865	1515 S Harwood St, Dallas, TX 75215, USA	(214) 421-5141	<a href="http://www.oldcityparkdallas.com">http://www.oldcityparkdallas.com</a>
8	Perot Museum of Nature and Science	32.7869	-96.8066	2201 N Field St, Dallas, TX 75201, USA	(214) 428-5555	<a href="https://www.perotmuseum.org">https://www.perotmuseum.org</a>
9	Pioneer Plaza	32.7766	-96.8012	1428 Young St, Dallas, TX 75202, USA	Unknown	<a href="http://www.dallasparcs.org/F">http://www.dallasparcs.org/F</a>

Şekil 3.1.2.12: Tool Tarafından Toplanan Bilgiler


Her bir süreç hakkında kullanıcı bilgilendirilir. Ayrıca sol bölümde açılır kapanır bir pencere ile tüm süreç gözlemlenebilir. Orta kısımda ise anlık takip yapılabilir.




Şekil 3.1.2.12: Uygulama Genel Arayüzü

Oluşturulan nihai plan kullanıcıya gösterilir.

BURSA TEKNİK ÜNİVERSİTESİ  
PROJE TABANLI İŞLETMEDE MESLEKİ EĞİTİM  
BTÜ-İMEP ÖĞRENCİ FİNAL RAPOR FORMU F4

 Could you help create a 3-day travel plan for one person? The journey starts in Long Beach and the destination is Dallas. The travel dates are from March 24th to March 26th, 2022, with a budget of \$1,900.

 Based on the provided data, here's a 3-day travel plan for one person from Long Beach to Dallas:

**Day 1: March 24th, 2022**

- **Transportation:** Flight Number: F4010196, from Long Beach to Dallas, Departure Time: 13:26, Arrival Time: 18:23
- **Breakfast:** MONKS, Dallas (Average Cost: 24, Cuisines: Tea, Pizza, Cafe, Mediterranean, Seafood, Aggregate Rating: 4.2)
- **Attraction:** The Sixth Floor Museum at Dealey Plaza, Dallas (Latitude: 32.779819, Website: <https://www.jfk.org/>, City: Dallas)
- **Lunch:** 1918 Bistro & Grill, Dallas (Average Cost: 90, Cuisines: Tea, Pizza, BBQ, Seafood, Aggregate Rating: 4.4)
- **Dinner:** Yanki Sizzlers, Dallas (Average Cost: 96, Cuisines: Cafe, French, Tea, Mediterranean, Aggregate Rating: 4.1)
- **Accommodation:** 1BR, elevator, kitchen, doorman!, Dallas (Price: 475.0, Room Type: Entire home/apt, Maximum Occupancy: 3, Review Rate: 2.0)

**Day 2: March 25th, 2022**

Type your message here... >

Şekil 3.1.2.13: Uygulamanın Oluşturduğu Plan



---

#### **4. ÇALIŞMA DÖNEMİNİN DEĞERLENDİRMESİ**

Çalışma dönemi boyunca sektör danışmanına düzenli aralıklarla sunumlar yapıldı. Travel Planner projesi ise sektör danışmanı ile birlikte Huawei Yazılım Departmanı direktörüne de sunuldu. DSPy’da Huawei Yapay Zeka Araştırma ekibine sunuldu.

Çalışma dönemi boyunca, projelerin konuşulduğu, 2025 planlarının tartışıldığı farklı toplantılara katılındı.

BTÜ-İMEP kapsamında yaptığım staj ile mezun olmadan tam zamanlı bir mühendis gibi çalışabilme fırsatına eriştim. Kariyerim açısından oldukça değerli bir deneyim kazandım.

Gelecek dönemde Huawei’de çalışmayı kesinlikle istiyorum. Çünkü stajım boyunca çok fazla kavram üzerinde bilgi birikimi ve deneyim elde edebildim. Ayrıca birçok konuda deneyim kazandım. Huawei’in hem araştırma hem de geliştirme odaklı ilerleyen bir kurum olması sayesinde bu kazanımları elde edebildiğim için, kariyerimin ilerleyen dönemlerinde de Huawei’in kapsamlı projelerinde yer almak isterim.

## 5. KAYNAKLAR

- [1] Jian Xie, K. Z. (2024, 02 02). TravelPlanner: A Benchmark for Real-World Planning with Language Agents. Computation and Language (cs.CL). doi:<https://arxiv.org/abs/2402.01622>
- [2] OSU-NLP-Group/TravelPlanner. (tarih yok). OSU-NLP-Group/TravelPlanner: <https://github.com/OSU-NLP-Group/TravelPlanner> adresinden alındı
- [3] llama documentation. (tarih yok). [https://www.llama.com/docs/model-cards-and-prompt-formats/llama3\\_1/](https://www.llama.com/docs/model-cards-and-prompt-formats/llama3_1/) adresinden alındı
- [4] Patrick Lewis, E. P.-t. (2021). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*.
- [5] intro-to-langgraph. langchain.com/courses/intro-to-langgraph: <https://academy.langchain.com/courses/intro-to-langgraph> adresinden alındı
- [6] chat\_templating. [https://huggingface.co/docs/transformers/main/en/chat\\_templating](https://huggingface.co/docs/transformers/main/en/chat_templating): [https://huggingface.co/docs/transformers/main/en/chat\\_templating](https://huggingface.co/docs/transformers/main/en/chat_templating) adresinden alındı
- [7] transformers docs/transformers/index: <https://huggingface.co/docs/transformers/index>
- [8] groq.com/docs/overview. console.groq.com/docs/overview: <https://console.groq.com/docs/overview> adresinden alındı
- [9] mistral.ai. (tarih yok). mistral.ai: <https://docs.mistral.ai/> adresinden alındı
- [11] huggingface/llm-vscode. huggingface/llm-vscode: <https://github.com/huggingface/llm-vscode> adresinden alındı
- [12] streamlit.io. (tarih yok). <https://docs.streamlit.io/>: <https://docs.streamlit.io/> adresinden alındı

BURSA TEKNİK ÜNİVERSİTESİ  
PROJE TABANLI İŞLETMEDE MESLEKİ EĞİTİM  
BTÜ-İMEP ÖĞRENCİ FİNAL RAPOR FORMU F4

---