	Hacettepe University Computer Engineering Department **BBM479 End of Term Development Report**

Project Details

Title	**Campus Navigator: Campus-Oriented Smart Assistant for the Visually Impaired**
Supervisor	**Assoc. Prof. Hacer Yalım Keleş**

Group Members

	Full Name	**Student ID**
1	Ömer Faruk Horat	2200356086
2	Ecem Altınkeser	2210356019
3	Buğrahan Çıldır	2200356017
4	Doğukan Aytekin	2200356003

Current State (/ 40 Points)

Explain the current state of the project. By giving references to the plan proposed at the beginning of the project, explain what is achieved so far. Provide details of what has been done by explaining the technology and methodology used. Is the current state of the project inline with the plan? If you are behind the schedule, explain in detail the reasons.

The current state of the project has made significant progress compared to the initial plan and has evolved into a more specific structure with modifications introduced along the way. Initially, we aimed to address all the needs of visually impaired users using a single AI model (e.g., CogVLM, BLIP). However, during the process, we realized that these models, due to their generalized structure, were insufficient for specific scenarios (e.g., detecting buttons on ATM screens). This led us to break the project into focused components for more effective progress.

To facilitate ATM usage, we utilize advanced object detection technologies like YOLO for button detection. For detecting text on the screen, we rely on OCR (Optical Character Recognition) methods. By combining these two technologies, the system aims to determine which button the user’s finger is hovering over and provide auditory feedback. This functionality helps users navigate the screen more easily and build a real-time mental map of the screen layout. As a result, visually impaired users are enabled to perform their transactions more independently.

Another system we are developing is designed to assist visually impaired individuals while walking on campus by detecting potential obstacles and estimating the distance to them. This system uses YOLO for object detection to identify obstacles and calculate their approximate distance. Subsequently, it aims to assess the potential risk posed by these obstacles based on their distance and position relative to the walking person.

To address this problem, we tested the YOLOv8 model using videos we recorded on campus. Then, for objects the model failed to detect, we prepared a dataset by extracting two frames per second from the recorded videos and labeling them. Currently, we are at the stage of fine-tuning the YOLOv8 model using this dataset and testing the updated version.

As a result, while the initial approach focused on a single solution, we have shifted to breaking down the problem and developing specialized solutions for each scenario. This approach has allowed us to achieve slower but much more accurate progress, leading to more effective and reliable outcomes for the project.

Continuous Learning (/ 25 Points)

You have been working on the problem for almost 3 months. In these 3 months, what did you learn about this problem that you didn't know at the beginning? Did this new knowledge change your perspective on the problem? What else do you need to learn in the future?

Over the past three months, while working on this problem, we have learned some significant facts that we were unaware of at the beginning. Initially, we believed that general AI models (like CogVLM, BLIP) could address all issues. However, we realized that these models were insufficient for some specific scenarios. One of the most valuable lessons we learned was that breaking the problem into smaller, focused cases yields more effective results.

For example, we approached different scenarios such as ATM usage and detecting the proximity of objects on campus individually and achieved higher accuracy using specialized solutions like YOLO. This new understanding has completely changed our perspective on the project. We now recognize the necessity of developing specific, targeted solutions for each problem.

When we tested the CogVLM model for the obstacle detection problem, we observed that it did not perform well, even in the task of accurately detecting various objects. For instance, when we asked whether there was a pedestrian crossing in a photo without one, the model incorrectly stated that there was. It even claimed there was a pedestrian crossing in a photo of an apple. Therefore, we concluded that YOLO provides more successful results for object detection and decided to use it instead.

In the future, we need to focus on understanding how the system performs in real-world tests and based on user feedback. Additionally, we aim to explore ways to improve response times and enhance the overall accuracy of the model through further research.

Risk Assessment and Management (/ 20 Points)

Try to identify the potential risks in the rest of the project. Were you able to identify these risks at the beginning of the project, or did you recently recognize them? What are your proposed solutions to each risk item? Are there any risks that will require a significant change in the project? If so, explain how this will affect the end results (also, outline your proposed revisions in the next section).

1. Limited Resources
Risk: The availability of hardware and software resources is a significant constraint. The computational power required for fine-tuning models like YOLOv8 and conducting extensive real-world tests surpasses the resources available to the team.
Proposed Solution: To address this, we plan to utilize cloud-based platforms such as Google Colab Pro or DigitalOcean for training and testing. Additionally, we will optimize our workflows by preprocessing datasets to reduce their size without compromising quality.

2. Time Constraints
Risk: The project timeline is limited, and focusing on two separate use cases (ATM navigation and obstacle detection) has slowed overall progress.
Proposed Solution: We have prioritized these two scenarios to maximize impact while scaling down less critical features. Regular progress reviews will ensure we stay aligned with the revised plan and identify bottlenecks early.

3. Real-World Testing Challenges
Risk: Real-world testing scenarios may not accurately replicate controlled environments. External factors like lighting, crowd density, and dynamic changes in the campus environment may negatively impact system performance.
Proposed Solution: Conduct extensive testing under varying conditions on campus and collect feedback from visually impaired users. Iterative updates based on feedback will ensure robustness and adaptability.

4. Model Limitations
Risk: YOLOv8 and OCR-based methods may still encounter limitations in specific scenarios, such as misidentifying overlapping objects or failing in extreme lighting conditions.
Proposed Solution: We will supplement YOLO and OCR methods with additional heuristics or secondary algorithms, like context-aware models, to improve accuracy in challenging conditions.

5. Dependence on Accurate Annotations
Risk: Mislabeling or insufficient labeling of datasets could lead to reduced model performance.
Proposed Solution: Implement thorough quality control during dataset preparation and involve multiple team members to cross-check labeled data.

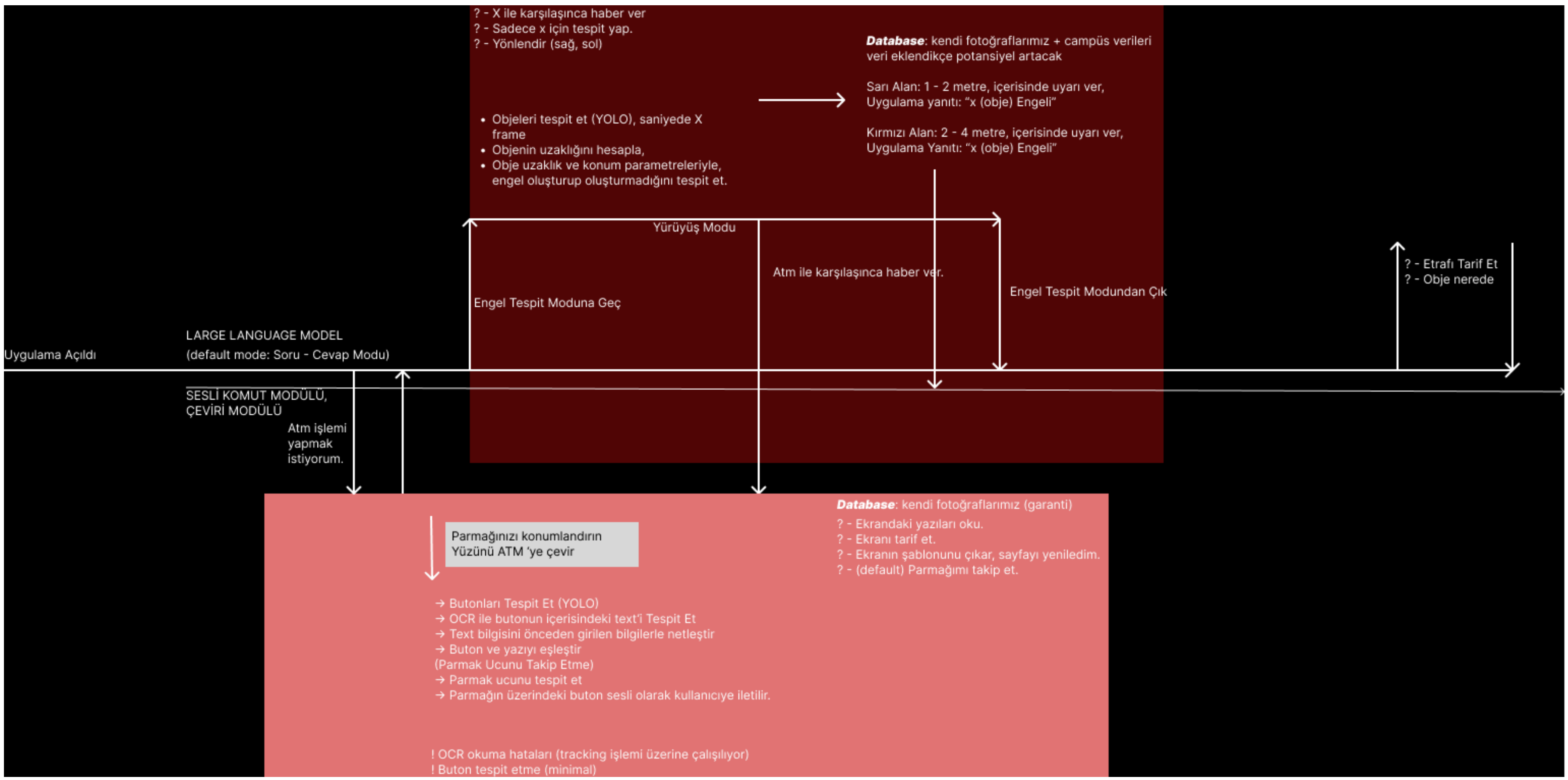
6. End-User Adoption Risks
Risk: Visually impaired users might find it challenging to adopt the system due to unfamiliarity with its interface or perceived complexity.
Proposed Solution: Incorporate user-friendly auditory cues and provide detailed training materials, including guides and demonstrations, to ensure ease of adoption.

Revisions (/ 15 Points)

If you feel like you need to revise your earlier plan, suggest your changes here. These changes may include changes in outcomes of the project, changes in milestones, changes in calendar, changes in workload distribution, etc. If you do not present any revisions here, at the end of the project you will be responsible for all the proposed outcomes in the Project Proposal.

During our research and development process we decided to limit our use cases and test VQA models on them. We chose detecting obstacles and navigating users through them as our first use case and chose assisting users using ATMs as our second use case. We tested two VQA models BLIP-2 and CogVLM on these two use cases. Both of the models performed poorly and did not give sufficient results. Therefore we decided to split these two tasks among each other. Ömer and Buğra were assigned to the first use case, Ecem and Doğuken were assigned to the second use case.

WP No.	WP Name	WP Description	Assigned to	Weeks													
				1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Project Planing	Creating project schedule	Team	X	X												
2	Technical Research and Development	Literature review for the VQA model	Team		X	X	X										
3	Testing VQA models	Testing BLOP-2 and CogVLM	Team				X	X	X								
4	Developing ATM asistant	Collecting data & labelling	Ecem, Doğukan							X	X	X					
		Reading text with Ocr	Ecem, Doğukan							X	X	X	X				
		Tracking finger tip	Ecem, Doğukan							X	X	X	X				
		Detecting buttons	Ecem, Doğukan							X	X	X	X				
		Matcing buttons and texts	Ecem, Doğukan										X	X	X	X	X
		Performance optimization	Ecem, Doğukan													X	X
5	Developing Walking Asistant	Collecting data & labelling	Buğra, Ömer Faruk							X	X	X					
		Identifying obstacles with YOLO	Buğra, Ömer Faruk							X	X	X	X				
		Asses potential risk created by obstacles	Buğra, Ömer Faruk							X	X	X	X	X	X	X	X
		Performance optimization	Buğra, Ömer Faruk											X	X	X	X


? - X ile karşılaşınca haber ver
? - Sadece x için tespit yap.
? - Yönlendir (sağ, sol)
Database: kendi fotoğraflarımız + campüs verileri veri eklendikçe potansiyel artacak
Sarı Alan: 1 - 2 metre, içerisinde uyarı ver, Uygulama yanıtı: "x (obje) Engeli"
Kırmızı Alan: 2 - 4 metre, içerisinde uyarı ver, Uygulama Yanıtı: "x (obje) Engeli"
• Objeleri tespit et (YOLO), saniyede X frame
• Objenin uzaklığını hesapla,
• Obje uzaklık ve konum parametreleriyle, engel oluşturup oluşturmadığını tespit et.
Yürüyüş Modu
Atm ile karşılaşınca haber ver.
Engel Tespit Modundan Çık
? - Etrafı Tarif Et
? - Obje nerede
Engel Tespit Moduna Geç
LARGE LANGUAGE MODEL
(default mode: Soru - Cevap Modu)
Uygulama Açıldı
SESLİ KOMUT MODÜLÜ,
ÇEVİRİ MODÜLÜ
Atm işlemi yapmak istiyorum.
Database: kendi fotoğraflarımız (garanti)
? - Ekrandaki yazıları oku.
? - Ekranı tarif et.
? - Ekranın şablonunu çıkar, sayfayı yeniledim.
? - (default) Parmağımı takip et.
Parmağınızı konumlandırın
Yüzünü ATM 'ye çevir
→ Butonları Tespit Et (YOLO)
→ OCR ile butonun içerisindeki text'i Tespit Et
→ Text bilgisini önceden girilen bilgilerle netleştir
→ Buton ve yazıyı eşleştir
(Parmak Ucunu Takip Etme)
→ Parmak ucunu tespit et
→ Parmağın üzerindeki buton sesli olarak kullanıcıya iletilir.
! OCR okuma hataları (tracking işlemi üzerine çalışılıyor)
! Buton tespit etme (minimal)