ÇANKAYA UNIVERSITY

FACULTY OF ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

CENG 407 - Literature Review

Artificial Intelligence Based Art Gallery Mobile Application

Team Members:

Mehmet Emre KILINÇ 202111058

Mertcan ZAFER 202011076

Batuhan ÖZER 202011072

Emre Şahin DEMİRBAŞ 202011020

Tunahan GÜLTEKİN 202011005

Contents

Abstract

This project aims to develop a mobile application that allows users to generate images from text prompts. Prior to development, we conducted a thorough literature review to inform our technological choices and feature design. This review encompassed academic studies, providing insights into potential technologies, their performance, advantages, and disadvantages. We also analyzed existing text-to-image generation applications, examining both their technological preferences (though detailed information was limited) and their feature sets. This analysis helped us identify potential differentiators for our application and pinpoint features that we believe will positively impact user experience and market success.

1. Introduction

Artificial Intelligence (AI) image generators is a concept of artificial neural networks. These generators take an input natural language description and produce an image matching that description. In recent years, as a result of advances in deep neural networks, many text-to-image models were released such as OpenAI's DALL-E 3, Google Brain's Imagen 3, Stability AI's Stable Diffusion, and Midjourney [1]. What makes them important is their ability to fuse styles, concepts, and existing images to create artistic and contextually relevant images [2]. AI image generators are being used in various fields such as marketing, advertising, product design, prototyping and branding for generating a wide variety of options in a short period of time [3].

The aim of this project is to develop an AI-powered mobile application that enables users to generate visual images through natural language processing and generative modeling techniques. This app, to be developed with Kotlin or Flutter, will use text prompts to produce an artwork in the specified style and generate a story that complements the artwork. Additionally, it will allow users to share their creations with others. The storage and generation of user data and artwork will be managed through cloud-based services like Google Cloud or Azure. Through the platform, users can create their own virtual galleries, showcase their artwork, and interact with other artists.

In our Literature Review report, we examined papers, web, and mobile applications related to our project. During the research, we searched for existing applications, models and cloud services. We classified properties of applications under some categories to observe advantages and disadvantages of them via investigating their architectures. We have also searched cloud services to determine their integrations with the examined model architectures. We have divided the rest of the Literature Review into the following sub-topics: Definitions, Reviewed Applications and Models, Limitations, Contributions and References.

2. Definitions

Definitions	
Variational Autoencoders (VAEs)	Variational Autoencoders (VAEs) are generative models explicitly designed to capture the underlying probability distribution of a given dataset and generate novel samples [4].
Generative Adversarial Network (GAN)	A generative adversarial network (GAN) is a machine learning (ML) model in which two neural networks compete by using deep learning methods to become more accurate in their predictions. GANs typically run unsupervised and use a cooperative zero-sum game framework to learn [5].
Diffusion Model (DM)	A Diffusion Model (DM) in the context of machine learning and artificial intelligence, is a type of generative model that has gained significant attention for its ability to generate high-quality, realistic images, texts, or sounds [6].
Latent Diffusion Models (LDMs)	Latent Diffusion models are deep learning models that have recently emerged as a powerful high-resolution image generation and manipulation technique [7].
Negative prompt	A negative prompt is a technique used in diffusion models. Allows users to specify what to exclude from the generated images. The impact of negative prompts is observed after positive prompts render corresponding content [8].
Independent Hardware Vendor (IHV)	An Independent Hardware Vendor (IHV) is a company that designs, manufactures or sells hardware or peripherals compatible with operating systems. Examples of Independent hardware vendors are Nvidia, Cisco Systems and Intel [9].

Mobile cloud computing (MCC)	MCC stands for Mobile Cloud Computing which is defined as a combination of mobile computing, cloud computing, and wireless network that come up together for purposes such as rich computational resources to mobile users, network operators, as well as to cloud computing providers [10].
Neural Processing Units (NPU)	A neural processing unit (NPU) is a specialized computer microprocessor designed to mimic the processing function of the human brain. They are optimized for artificial intelligence (AI) neural networks, deep learning and machine learning tasks and applications [11].
Intelligent radio access networks (Intelligent RAN)	Intelligent Radio Network Access is a network technology framework that uses artificial intelligence (AI) and advanced data processing techniques to optimize the allocation and management of radio network resources, such as spectrum, bandwidth, and transmission power.

Figure 1 - Table of Definitions used in Literature Review Report

3. Reviewed Applications and Models

In this part of our literature review, we have mentioned the applications that we have examined and their respective models and divided it into three parts as "Reviewed Mobile Applications", “Reviewed Web Applications”, "Reviewed Models", and “Comparison of Reviewed Models/Applications”.

3.1 Reviewed Mobile Applications

Application Name& Feature	Leonardo.Ai – Image Generator	Dream by WOMBO	DaVinci – AI Image Generator	Imagine: AI Art Generator	CreArt – AI Image Generator	Starryai – AI Art Generator	Gencraft – AI Art Generator
Social media features -Interaction with other users	✅*	✅*	❌	❌	❌	✅	✅
Preset styles	✅	✅	✅	✅	✅	✅	✅
Negative prompt	✅	❌	✅	❌	✅	✅	✅
Writing a story for the prompt and art	❌	❌	❌	❌	❌	❌	❌
Ability to choose models	✅	❌	✅	❌	❌	❌	❌
Gamification Features	❌	❌	❌	❌	❌	✅	✅
Cloud Usage	✅	✅	✅	✅	✅	✅	✅
Models Used for Image Generation	Stable Diffusion 1.5 and 2.1, SDXL 0.9 and 1.0, SDXL.LIGHTNING, and custom fine-tuning models	Stable Diffusion, VQGAN+CLIP	Stable Diffusion XL, DALL-E 3, and its own specially developed model DaVinci XL	Stable Diffusion, DALL-E, SDXL, Flux Schnell and custom models	Stable Diffusion, DALL-E 2, custom models	Altair and Orion.	DALL-E 2, Stable Diffusion, custom models
Release Date	2024	2021	2022	2022	2023	2021	2023

** These applications have limited social media features, namely observing and liking the other users' created images.*

Figure 2 - Mobile Text-to-Image Applications Comparison Table

Firstly, our analysis of existing mobile applications revealed a significant gap in the market for a mobile application that offers both a unique storytelling feature and robust social interaction. While most applications lack the capability to generate a story alongside an image based on a user's prompt, we believe this feature holds strong differentiation potential. Furthermore, we observed a lack of comprehensive social features in existing applications, with most offering limited functionality such as viewing and saving images created by others. We envision a social experience that includes friend connections, following, private image sharing, liking and commenting, leaderboards, weekly/monthly best filters, group creation, and participation – features that are currently missing or limited in most mobile applications. (The mentioned features are valid for mobile versions of existing applications; some applications may offer more similar features in their Web applications or Discord versions.) In this regard, starryai and Gencraft applications offer social media features similar to what we have thought to include in our application. While analyzing existing features, we have discovered additional functionalities such as predefined styles and negative prompt options and adding these functionalities could make our application competitive against the other applications. Additionally, we are planning to incorporate gamification features such as challenges and achievements, inspired by existing applications, to enhance user engagement and encourage continued use.

Regarding technological choices, most applications remain opaque about their specific image generation models, cloud usage, and cloud providers. However, based on available information, Stable Diffusion and DALL-E 3 appear to be the dominant models used. In terms of cloud usage, when we examined the models that we can use to create images locally, we saw that this process requires high computational power, as we expected, and most phones, except for flagship high-end phones, will have difficulty meeting the system requirements of an application made using a model that runs locally. When we looked at what existing applications do in this regard, we did not see that they stated that they were using the cloud, except for Leonardo.Ai and Dream by WOMBO applications that use Amazon Web Services (AWS). However, we considered the statement of Vivek Bhakta, cofounder and head of infrastructure of WOMBO, “You don’t need the best smartphone to run our app because we do all the processing in the cloud”, for other applications as well, and considering the system requirements that are definitely not high for any application, the creation times, and the large AI models they use such as Stable Diffusion and DALL-E, we concluded that other applications also use the cloud for computation.

3.2 Reviewed Web Applications

In the second part of our applications analysis, we have examined many of the existing web applications like Bing Image Creator, DreamStudio, Canva etc. (the full list of examined web applications could be found in the Figure 3) As it could be seen from the Figure 4, most of the web applications are also using a type of Stable Diffusion model, as it was the case with most of the mobile applications.

Reviewed Web Applications		
DALL-E 2	Adobe Firefly	DeepAI
Bing Image Creator	Prodia	Vance AI Art Generator
DreamStudio	Leap AI	Fotor
Canva	getimg.ai	Runway
NightCafe	Shutterstock AI Image Generator	Pixray
OpenArt	Stablecog	Let's Enhance

Figure 3 - List of Reviewed Web Applications

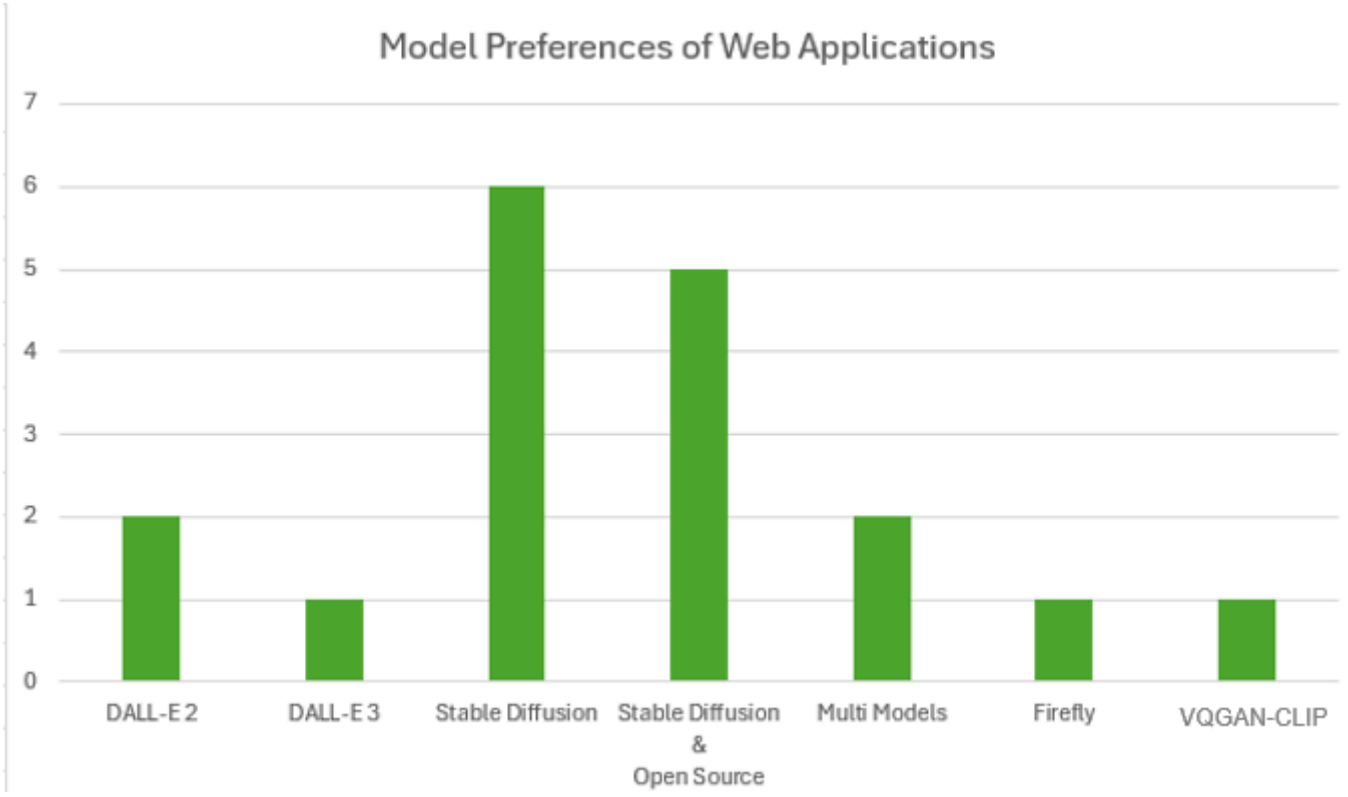

Figure 4 - Web based Text-to-Image Applications Comparison Table

3.3 Reviewed Models

Model Name	Has API?	Free to Use?	Is Open Source?	Min GPU VRAM Needed
Stable Diffusion	✅	✅	✅	4 - 8 GB*
DALL-E 3	✅	❌	❌	-**
Midjourney	❌	❌	❌	-**
Imagen 3	✅	✅	❌	-**
FLUX.1	✅	✅	✅	8 GB
Ideogram 2.0	✅	✅	❌	-**
StyleGAN3	✅	✅	✅	12 GB

** Stable Diffusion 1.5 requires 4-6 GB, 2.1 requires 4-6 GB, XL requires 8 GB, and XL Turbo requires 8 GBs of VRAM.*
*** These models do not have a local usage option.*

Figure 5 - Text-to-Image Models Comparison Table

As a part of our research, we have examined various up-to-date text-to-image models, including FLUX.1, Ideogram 2.0, Imagen 3, DALL-E 3, Midjourney, StyleGAN3, and Stable Diffusion, to decide which model fits our needs the best, to be used on our application. We have extracted some valuable insights that could help us to make our decision, some of the most important ones in a compact form could be observed from the Figure 5, and the detailed comparisons are listed below:

- Most of the models (except the hidden architecture ones, and StyleGAN3) are mainly based on the Stable Diffusion model.
- Midjourney works through its Discord server and it could create high-quality images from the provided prompts, but unfortunately it does not provide any free alternative or provides an API.
- DALL-E 3 has very limited free usage, but it could generate great images throughout most categories but it does not produce quality results when it comes to portrait/landscape generation sometimes [12].
- Imagen 3 is a powerful model that could generate various realistic images according to the given prompts and could be run from Google's Gemini application without needing powerful local GPUs or an external cloud system, but because it is relatively new, we do not have extensive/detailed tutorials or documentation on how to integrate it with our application.
- FLUX.1 is one of the newest models, produces fascinating outputs, and is open source, but it is comparably computationally expensive and generates the images in a longer period as, for the optimal usage, it requires one or more powerful graphics cards with at least 8 GBs of VRAM (for FLUX.1 [schnell]) [13].
- Ideogram 2.0 is also a great model, which is capable of generating realistic images, but its free option is limited (only up to 40 images per day) [14].
- StyleGAN3, differently from most of the other models, is based on a GAN. It has advantages like being open-source, and having an API ready to use, but it is computationally expensive as it is stated on their respective GitHub page, it requires 1-8 high-end NVIDIA GPUs with at least 12 GBs of VRAM [15].

- Stable Diffusion, based on the Latent Diffusion Model (an improved version of a Diffusion Model), has lots of advantages, as when compared to the other models, it is one of the fastest, and one of the most computationally affordable one out of the models that are listed here. In addition, its source code is accessible and therefore open to possible modifications, has an API, and because of being present on a lot of applications and there have been some time since its first release, it has numerous tutorials, which improves its integrability drastically.

3.4 Comparison of Reviewed Models/Applications

In the below section, in order to get a general idea to see how each model produces an output with respect to the given prompt, we have added sample images generated by some of these models, along with their respective prompts and model names. Additionally, we also have a subsection within this part, which compares the outputs of models of the mobile apps that we have analyzed to decide which specialized version of the Stable Diffusion model is the best to use, in case we would choose this model as our base model.

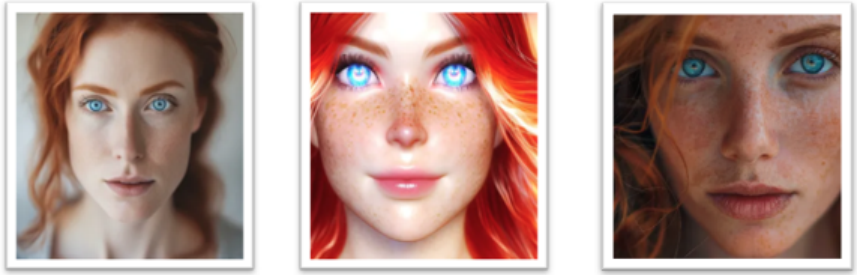

Figure 6- Generated Images by Stable Diffusion, DALL-E 3, and Midjourney models [12]
"A close-up photographic portrait of a beautiful redhead woman with electric blue eyes."

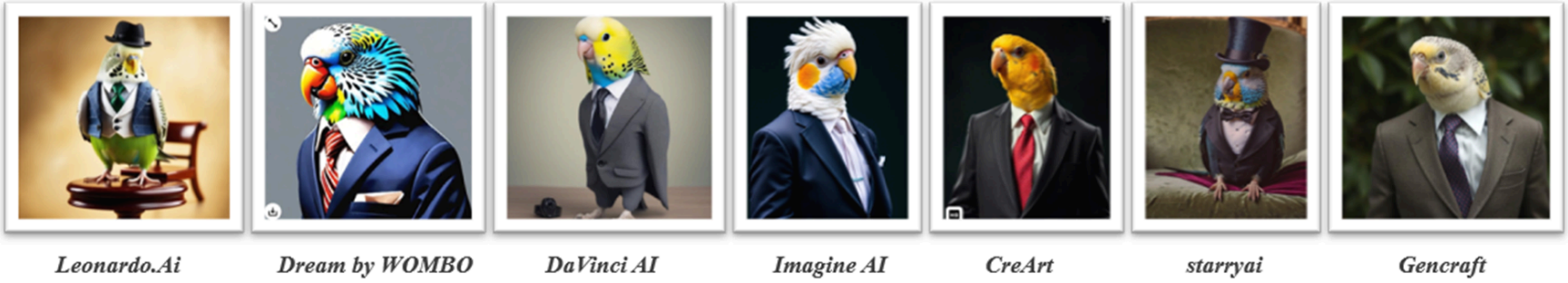

Figure 7 - Generated Images by various mobile applications' models
"Budgie in a suit."

4. Limitations

Training diffusion models and generating images based on users' prompts could require high computational power and storage. Mobile devices, for instance, are limited in processing power compared to modern computers' GPUs and TPUs. Today's famous hardware vendors have been trying to increase the processing power for handling Artificial Intelligence and learning algorithms. NPUs on the other hand, are specialized processors that are designed specifically for optimizing the execution of deep learning models, and neural network processing [16]. Although high optimizations that NPUs give are not as powerful as cloud GPUs, they have lower power consumption compared to CPUs and GPUs when executing AI tasks and making them ideal for edge computing and mobile devices [16]. Another approach to use on mobile apps is cloud services. Cloud services provide better performance and quality images without relying on mobile devices' limited hardware and NPUs [17]. To use cloud services in mobile apps there exist different methods like MCC, which is a method of using cloud service technology to deliver mobile apps. Furthermore, it enables us to target all mobile devices that can connect to the cloud services. There are a lot of cloud services available such as Amazon Web Services (AWS), Google Cloud Services (GCP), Microsoft Azure, IBM Cloud Services, and many more. However, there are some topics to be considered when using cloud technologies. Firstly, less network bandwidth can be an issue since MCCs require communication to be continuous. This means that a developer may face problems if the network being used is wireless. This is because wireless networks tend to be less reliable or possess low bandwidth [18]. Secondly, consumption of batteries is another problem. Cloud-based applications increase the use of the battery and would, therefore, consume it much more quickly [18]. Thirdly, operating system compatibility is important. The applications created using MCC will function on different operating systems. Therefore, the application must be compatible with operating system platforms like Android, iOS, and Windows Phone. To do so, the development team must possess knowledge regarding an IRNA or Intelligent Radio Network Access technique [18]. Lastly, there could be some security issues related to the use of MCCs.

5. Contributions

In this section, we have added the things that could differentiate our potential application from the other existing applications. In addition, we have also included the models, programming tools, APIs etc. that we could use while developing our application.

- In addition to creating an image as a result of the entered prompt it would also be possible to create a story about the image.
- Detailed social media features that enable interaction between users, such as adding friends, following, likes, comments.
- Gamification features such as daily challenges, achievements, leaderboard based on likes, weekly-monthly best.
- If we would choose the Stable Diffusion or another modifiable model as our base model, we are planning to fine-tune the base model to get better results.
- Since we will be developing an Android application, we are considering using Android Studio as the development environment and Kotlin, which is Google's preferred language for Android app development, as the development language.
- Since models running locally require high system requirements in terms of computational power, we are considering running large image generating models in the cloud, just like the other mobile applications we examined. Instead of using the cloud, we can perhaps use the APIs of ready-made applications.

References

[1] "Text-to-image model," Wikimedia Foundation, Inc, 27 October 2024. [Online]. Available: https://en.wikipedia.org/wiki/Text-to-image_model. [Accessed 08 November 2024].

[2] "AI Image Generation Explained: Techniques, Applications, and Limitations," AltexSoft, 2024. [Online]. Available: https://www.altexsoft.com/blog/ai-image-generation/. [Accessed 08 November 2024].

[3] M. Smith, "How AI Image Generators Work and Are Used by Businesses," 06 August 2024. [Online]. Available: https://medium.com/@ankita.gsdc/how-ai-image-generators-work-and-are-used-by-businesses-ea42ded5c7ae#:~:text=For%20example%2C%20companies%20use%20them,even%20generate%20artwork%20for%20sale. [Accessed 08 November 2024].

[4] A. S. CHOUDHARY, "An Overview of Variational Autoencoders (VAEs)," 30 July 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2023/07/an-overview-of-variational-autoencoders/. [Accessed 08 November 2024].

[5] S. Robinson, "What is a generative adversarial network (GAN)?," October 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/generative-adversarial-network-GAN. [Accessed 08 November 2024].

[6] "What is a Diffusion Model (DM)?," [Online]. Available: https://www.theainavigator.com/blog/what-is-a-diffusion-model-dm. [Accessed 08 November 2024].

[7] A. Neto, "What is Latent Diffusion in AI?," 06 October 2023. [Online]. Available: https://medium.com/@aguimarneto/what-is-latent-diffusion-in-ai-43aa1ad4f71e. [Accessed 08 November 2024].

[8] R. W. T. Z. M. C. B. G. C.-J. H. Yuanhao Ban, "Understanding the Impact of Negative Prompts: When and How Do They Take Effect?," *Computer Vision and Pattern Recognition,* 2024.

[9] "Independent hardware vendor," 30 September 2024. [Online]. Available: https://en.wikipedia.org/wiki/Independent_hardware_vendor. [Accessed 08 November 2024].

[10] kaurbal1698, "What is Mobile Cloud Computing?," 19 December 2022. [Online]. Available: https://www.geeksforgeeks.org/what-is-mobile-cloud-computing/. [Accessed 08 November 2024].

[11] I. S. Josh Schneider, "What is a neural processing unit (NPU)?," 27 September 2024. [Online]. Available: https://www.ibm.com/think/topics/neural-processing-unit. [Accessed 08 November 2024].

[12] A. Antich, "Extensive Comparison of Text-to-Image AI Models," 25 May 2024. [Online]. Available: https://medium.com/superstringtheory/extensive-comparison-of-text-to-image-ai-models-1b523a586e3f. [Accessed 08 November 2024].

[13] M. Gupta, "Run Flux.1 Dev on Google Colab, 8 GB GPU for free," 22 October 2024. [Online]. Available: https://medium.com/data-science-in-your-pocket/run-flux-1-dev-on-google-colab-8-gb-gpu-for-free-7daa9fd88db5. [Accessed 08 November 2024].

[14] "ideogram," ideogram, [Online]. Available: https://ideogram.ai/pricing. [Accessed 08 November 2024].

[15] T. K. Janne Hellsten, "stylegan3," 26 April 2023. [Online]. Available: https://github.com/NVlabs/stylegan3. [Accessed 08 November 2024].

[16] L. Gong, "CPU vs GPU vs TPU vs NPU: What Are the Key Differences?," 08 August 2024. [Online]. Available: https://www.seeedstudio.com/blog/2024/08/12/cpu-vs-gpu-vs-tpu-vs-npu/?srsltid=AfmBOoqElTcWnTC__xRQSWhafHJqnQdQ_tsNOQUvtW2T61vcu3zqp_1d. [Accessed 08 November 2024].

[17] "What is Mobile Cloud Computing?," Amazon, [Online]. Available: https://aws.amazon.com/what-is/mobile-cloud-computing/?nc1=h_ls. [Accessed 08 November 2024].

[18] "Mobile Cloud Computing – Overview, Challenges and the Future," [Online]. Available: https://atmecs.com/mobile-cloud-computing-overview-challenges-and-the-future/. [Accessed 08 November 2024].