esprit

Se former autrement

HONORIS UNITED UNIVERSITIES

Engineering internship

SPECIALITY: Data Science

Set of Cards Generator

Elaborated by : Farah Chahrazed Selmi

Supervised by:

Company Supervisor: Mr. Wassim Derbel

Table of Contents

List of Figures

Acknowledgment

I would like to express my sincere gratitude to the entire educational team at **ESPRIT (Private School of Engineering and Technologies)**, as well as the professional instructors responsible for my education.

I want to extend my heartfelt thanks to **Mr.Wassim Derbel**, my supervisor, firstly for providing me with the opportunity to undertake my engineering internship within the company. His guidance and professionalism were invaluable throughout my internship, and his unwavering support and insightful advice were crucial to my progress and the success of this project.

General Introduction

Education is a fundamental subject in the development of children. However, it is not uncommon for students to become quickly bored in the classroom, largely due to the long hours of lectures, lack of interest in the subjects taught, and information overload.

To make learning more engaging, it is evident that we need to shift towards innovative and more interesting solutions to attract children. Among these, educational platforms stand out, with their primary goal being to seamlessly integrate learning and entertainment. This approach is more appealing to children compared to traditional methods.

In a continuously evolving world, computer science, specifically the field of Artificial Intelligence (AI), aims to develop systems capable of imitating or replicating human intelligence. It relies on algorithms and mathematical models to analyze data, acquire new information, make autonomous decisions, and solve complex problems. This new technology can be incorporated into all aspects of our lives, including education, where educational platforms can leverage AI to enhance certain facets of teaching.

My study project, titled ”Set of Cards Generator,” focuses on designing a well-known children’s game: the memory card game, where cards are matched based on their correspondence and following certain logic.

This report will detail the various phases undertaken to achieve a reliable and satisfactory result. To this end, the report defines the work I have carried out, structured into five major chapters.

The first chapter, titled ”General Project Framework,” primarily serves to introduce the host organization where I conducted my engineering internship, along with a critical analysis of the existing situation. It then presents the proposed solution, concluding with an overview of the methodology adopted for project realization.

The second chapter, entitled ”Analysis and Specification of Requirements”, aims to present both the functional and non-functional needs of the project.

The third chapter, entitled ”Achievements”, is dedicated to the presentation of the tasks I performed during my internship as well as the various research I conducted. I described in detail the achievements and results achieved, highlighting the various stages of the process.

The fourth chapter, entitled ”Internship Insights and Project Outlook”, focuses on what I learned during my internship and offers innovative ideas that could be integrated into the platform. I shared my experiences and acquired, as well as suggestions to further improve the project.

Finally, the report concludes with a conclusion that summarizes all the achievements and contributions made during the internship.

The second chapter, entitled ”Analysis and Specification of Requirements”, aims to present both the functional and non-functional needs of the project.

Chapter 1

Project Overview

1.1 Introduction

This introductory chapter consists in presenting the host organism, then the study of the existing, the analysis of similar applications, the solutions proposed from the application, and finally the specification of the adapted methology.

1.2 Company Presentation

BMCAF:(Consulting, Assistance and Training QHSE) is a Study and Training Office created in 2006 operating in the sector of continuing training and assistance, approved by the State to ensure training cycles for the benefit of public and private companies. We offer training courses...

Figure 1.1: Logo BMCAF.

BAYCII LABS:a young and innovative startup with a remote workspace and a talented team of experts in AI, cybersecurity, big data and IoT. It provides scalable solutions, whether in software development, application creation or website design.

Figure 1.2: Logo BAYCII LABS.

Slogan:Bringing life to ideas

1.3 General Context

This work is part of the internship project of the academic year 2022/2023. This project, lasting two months, is carried out within the company BMCAF and Baycii Labs , and aims to design a ”Set ofCards Generator”.

1.3.1 Problematic

Given the classic learning mothods, children get bored quickly and struggle to stay focused, which can be attributed to the enormous amount of information they are exposed to. In addition, many children spend hours on their desk studying, but without achieving the expected results. They may feel frustrated and discouraged, which can affect their motivation

1.3.2 Current State Analysis

In order to define the expected objectives of the proposed project, it is necessary to study and analyze similar applications in order to identify the problems and difficulties related to the realization of the platform and to highlight them in our future application.

1. **Maxicours:**is a website that offers online tutoring services. It offers a wide range of educational resources in different fields, such as mathematics, physics, chemistry, biology, French language, history, geography, and many more. The site aims to help students deepen their knowledge, fill their gaps and improve their academic performance through interactive courses, exercises.

Figure 1.3: Maxicours Website.

2. **IXL:**is an online learning platform that offers educational resources for students of different grades. IXL offers interactive math exercises.

Figure 1.4: IXL Website.

* There are other educational platforms such as:**PBS Kids,Starfall,ABCmouse**

1.3.3 Existing Review

Existing platforms have some disadvantages such as:

-The lack of diverse and interactive games to attract and motivate children.

-Page design may be over-loaded due to the presence of many flashy colors, which may be less visually appealing.

-the lack of variety of subjects is taken as an example the **IXL** platform, which is exclusively devoted to mathematics.

-An overabundance of advertisements can make the site unattractive and disrupt the user experience.

1.3.4 Project Description

LAB2301 is an innovative project for an E-learning platform that combines entertainment and learning simultaneously, creating a desire for children to study. Subscribers can be either schools or families. In the first scenario, the learners constitute an entire class, while in the second scenario, the learners are individual students.
Games following the pedagogical approach and the Tunisian educational system.

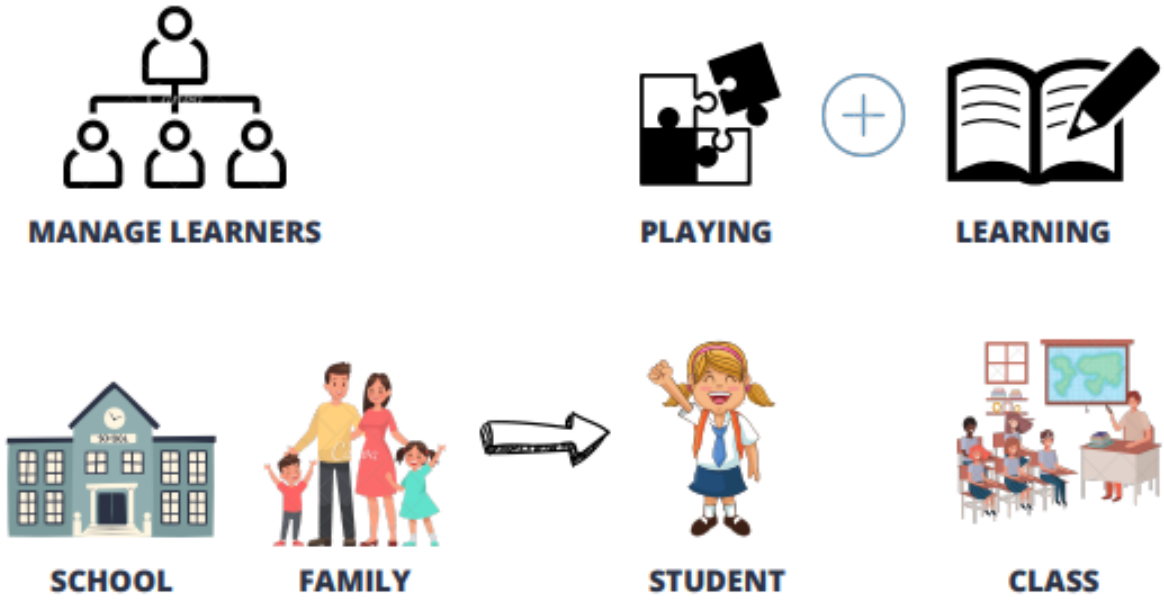


Figure 1.5: LAB2301 Explication.

The platform is composed of different games, the first of which is called **Set of Cards Generator**. This game aims to memorize identical cards. Players must match pairs of cards by flipping them one by one in order to find identical cards. It is a game that tests the ability of memorization and concentration of players. Thanks to this game, users can improve their cognitive skills and educational development while having fun.

Our game will be based on **GENERATIVE AI** that refers to a field of artificial intelligence that focuses on creating original and unique content. Unlike traditional AI models that rely on existing data, generative AI has the ability to autonomously generate new ideas, texts, images, music, and other types of content.

A special feature of our deck of cards is that it is not limited only to pairs of identical or similar cards. It also includes maps that follow certain logics, such as mathematical operations or syllables. These cases will be treated and explained in the following chapters under the name

of **LOGIC CATEGORIES**.

In addition, this game offers a wide variety of combinations: a card can contain text, sound or an image. The total number of identical card combinations can range from 2 to 4 cards. Example 1: CARD1text,CARD2text, CARD3text,CARD4text Example 2: CARD1 sound,CARD2image,CARD3image,C

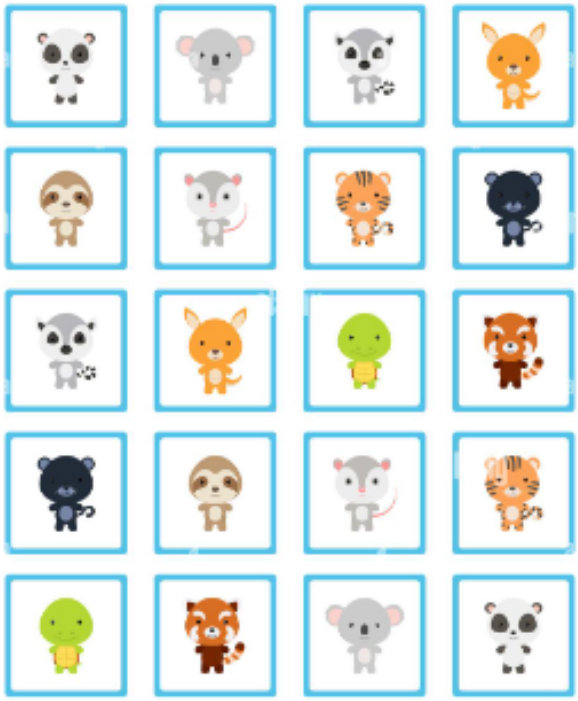


Figure 1.6: Memory game.

1.4 Working Methodology

Before starting each project, it is essential to select an appropriate methodology to ensure the smooth running. We will choose a **POC method**.

1.4.1 POC Definition

POC (Proof of Concept) is a demonstration or partial realization of a concept or idea to verify its technical feasibility or potential. This involves creating a functional prototype or a simplified version of a product or service to test its functionality, demonstrate its value or validate its usefulness before developing it completely.
The POC reduces risk and provides real-world feedback before investing more resources in a project.

1.4.2 POC Steps

1. **Definition of objectives:**
   -Identify the clear goals you want to achieve with CAP.
   -Specify the expected results and success criteria.
   -Understand the problems or needs you are trying to solve with CAP.

2. **Technology Selection:**
   -Assess how this technology can meet CAP objectives.
   -Make sure the technology is available and ready to use.

3. **Development:**
   -Implement CAP using the selected technology.
   -Create a prototype or demo version of the solution.
   -Ensure that CAP is delivered in a way that meets the defined objectives.

4. **Testing and Evaluation:**
   -Conduct extensive testing to verify that CAP is working as intended.
   -Collect data and user feedback to assess performance.
   -Identify any issues or gaps that need to be addressed.

5. **Presentation of results:**
   -Communicate CAP results to stakeholders, including decision makers.
   -Explain how CAP is meeting initial goals and needs.
   -Present the advantages and disadvantages observed during the tests.

6. **Final decision:**
   -Make an informed decision on whether or not to continue with full implementation of the solution.
   -Base your decision on CAP outcomes, potential benefits, costs and risks.
   If the decision is positive, plan to scale and deploy the solution.

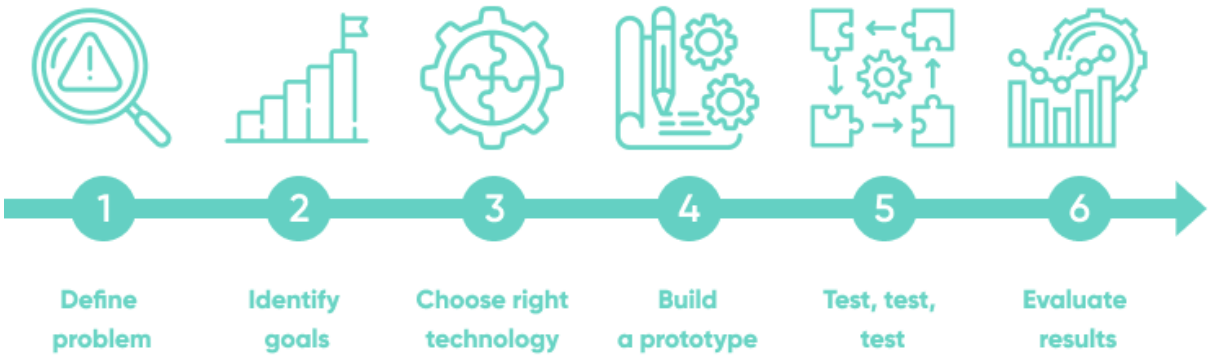


Figure 1.7: POC.

1.5 Conclusion

Through this chapter, I first presented the company, followed by the study of the existing where I chose two examples, namely ”Maxicours” and ”IXL”. I analyzed their weaknesses in order to overcome them in our project and guarantee a satisfactory user experience. Finally, the last part is devoted to the specification of the working methodology and its stages.

Chapter 2

Analysis and Specification of Requirements

2.1 Introduction

The analysis and specification of needs are important steps in our project. This chapter begins with the presentation of functional requirements, which include business and data science objectives. Then, the second part concerns the identification of non-functional needs.
Finally, the last part is reserved for the presentation of the working environment and the tools used.

2.2 Functional Requirements

A functional requirement defines a system or its components. It describes the functions a software must perform.

2.2.1 Business Objectives

At the end of the project, the platform aims to achieve these objectives.

1. **Improve academic performance:** The main goal could be to improve student academic performance by providing a fun and interactive way to learn key concepts.

2. **Increase student engagement:** Encourage student engagement by making learning fun and encouraging students to use the platform on a regular basis.

3. **Facilitate progress tracking:** Provide teachers and parents with tools to track student progress and identify areas where they may need additional help.

4. **Promote creativity:** Encourage creativity and critical thinking in students by integrating game design elements into learning.

5. **Concentration and Attention:** To succeed in the memory game**Set Of Cards Generator**, children must carefully focus on the images, words or numbers on the cards. This improves their ability to focus on details and maintain their attention, essential skills for memorization.

2.2.2 Data Science Objectives

From a data science perspective, we have several goals that we hope to achieve with this project, especially for the game ”Set of Cards Generator”, which is used to generate game cards using generative artificial intelligence (Generative AI).
Our data science goals are:

a) **Web Scraping:**Also known as web data collection or web data extraction, this is the process of automatically extracting information or data from websites. We aim to collect various types of data from different subjects and levels, in accordance with the Tunisian education system.

b) **Data Preparation and Data Cleaning:** These two techniques are essential in the field of data science to correct and eliminate inconsistencies, errors and unnecessary data in our generation card game. It is crucial to have high quality data for modeling.

c) **Modeling:** The last step of the project is to use generative models such as GANs (Generative Adversarial Networks), Transformers, and NRNs (Recurrent Neural Networks) to generate game maps. These templates will allow you to create maps in a creative way and adapted to the educational needs.

In summary, our data science goals for the game **Set of Cards Generator** include collecting relevant data, preparing data for optimal quality, and the use of advanced generative models for creating high quality educational game cards.

2.3 Non-Functional Requirements

The non-functional requirements fall under various categories, including the performance requirements and GUI-requirements.

2.3.1 Performance Requirements

1. **Security:**aims to protect the platform from threats such as hacker attacks, malware and vulnerabilities. It ensures that sensitive data, such as users' personal information, is secure. Security also includes access management, intrusion protection, and data backup to prevent data loss.

2. **Privacy:**concerns the protection of users' private information.

3. **Maintainability:** Ease of futures and bug fixes.
4. **Availability:** The application must be available to be used.
5. **Scalability:**Ability to handle increased load.

2.3.2 GUI-Requirements

1. **Usability:** The platform should be easy to use.
2. **Simplicity:** Minimalistic design and reduction of distraction.
3. **Visually Appealing Interface:** The design must be colourful, playful and age-appropriate for children to engage and engage them.
   Instead of opting for flashy colors, it is better to use color combinations adapted to the age of children and that promote concentration, readability and a positive atmosphere.
4. **Simple and Adapted Language:** Texts and instructions must be written in simple language and adapted to the children's age, thus promoting their understanding.

2.4 Work Environment and Tools used

Python: The open source programming language, Python 3 is the current version. Its interpreter is more efficient, as well as its competition control. Technically, this language will be used mainly for scripting and automation This language has propelled itself to the forefront of infrastructure management, data analysis or in the field of software development. Thus, developing code with Python 3 is faster than with other languages.

Figure 2.1: Python.

Jupyter:is an open source web application that we can use to create and edit live code documents, equations, visualisations and text.
This web application is used to progress more than 40 angages of progress, including: Python, Julia, Ruby, R, or Scala. These notebooks are used in data science to explore and analyse data

Figure 2.2: Jupyter.

Anaconda: Anaconda is an open-source platform for data scientists. It serves as a distribution of programming languages Python and R.
It is applied to the development of applications dedicated to data science and machine learning (large-scale data processing, predictive analysis, scientific computing). Its aim is to simplify package management and deployment.

Figure 2.3: Anaconda.

MongoDB: It's a NoSql database, so it stores data in documents or collections of objects much like JavaScript notations, so it's compatible with nodejs server because it's easier to communicate with NoSql when we're using javascript objects t's mainly used when you need high availability with automatic, fast and instant data recovery

Figure 2.4: MongoDB.

Overleaf: Overleafis a free online platform to edit LATEX text without any application download. In addition, it offers the possibility to write docu ments in a collaborative way, to offer its documents directly to different publishers.
We wrote this report in Latex using the online tool Overleaf.

Figure 2.5: Overleaf.

2.5 Conclusion

This chapter describes the functional requirements, including business and data science objectives, as well as non-functional requirements such as platform security, performance and usability.
The closing of this chapter will include a presentation of the work environment and the tools used. The next chapter will explain in detail the tasks performed.

Chapter 3

Achievement

3.1 Introduction

This chapter is tasked with presenting what has been accomplished during this internship.The first part consists in explaining the different categories of playing cards and their classification. Then, a presentation of the models of artificial intelligence generation will be made.
A pipeline will be presented to illustrate the use of these models in the project.
Finally, the last part will address the challenges encountered during the internship.

3.2 Logical Categories

The principle of logical categories is to determine the maximum number of categories that playing cards can encompass. This is elucidated through the presentation of the number of possible combinations and scenarios involving pairs.
Subsequently, we proceed with the presentation of the various categories.

3.2.1 Card Combination Possibilities

The maximum number of cards in a game is 24 cards, distributed as follows:

1. Text matching with [2 to 4] Texts
2. Text matching with Photos: It can be:
   a) 1 photo, 3 texts
   b) 2 photos, 2 texts
   c) 3 photos, 1 text
3. Photo matching with [2 to 4] photos

3.2.2 Types of Logical Categories

This section will be responsible for addressing the possible types of cases that playing cards can contain by subject matter. This represents the first step of our project.

A brief general example will be provided, with further details to be presented in a table.
Example 1: Subject: Maths

Figure 3.1: Example 1 of Logic Catg.

Example 2: Subject: Logic

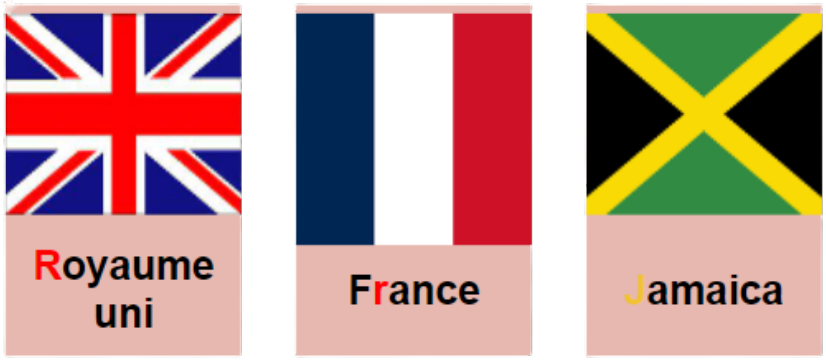


Figure 3.2: Example 2 of Logic Catg.

The player must consider that the colored letters first refer to a color in each flag, and to the first letter of that color. To explain this better, let's take the example of France. The letter "R" is colored in red(ROUGE). The flag of France consists of the color red, and "R" refers to the first letter of the color red(ROUGE).
Similarly, for Jamaica, where the flag contains the color yellow(JAUNE). The letter "J" is colored in yellow (JAUNE), and "J" refers to the first letter of the word yellow(JAUNE).

The following images are used to specify each logical category for every subject, considering the maximum possible combinations of cards.

a) Maths

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
Maths	Addition/sus/Division/ multiplication	2+2 (text)	4 (text)		
	Numerical sequence	3-5-7-9	Numerical sequence		
	Comparison	image of 4 cupcakes	image of 2 cupcakes	4>2	
	Commun operation	5-3=2	12-6=6	The 2 operations are subtraction	
	Missing value/operation	2-4-6-?-10	8		
	Advanced operations	2+3*4	14		
	Conversion	1L	1000ML	100CL	

Figure 3.3: Maths Category.

b) French

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
French	Syllable	Mai	Son	Maison	a picture of a house
	Identity	a picture of a house 1	Maison	a picture of a house 1	
	Same Context	a picture of a house 1	Maison	a picture of a house 2	a picture of a house 3
	Singular/Plural	Animal	Animaux		
	Opposite	chaud	froid		
	Suffix -Prefix	Droit	Maladroit	voyage	Mariage
	Synonym/ Antonym	Heureux	Joyeux	Content	Gai
	Homophones	Vert	Vers	Verre	
	Homograph	Son (Possession)	son (bruit)	Son (Reste du ble)	
	Definition	Aube	Le moment juste avant le lever de soleil		
	Grammatical category & gender	Chat	Nom	Masculin	a picture of a cat
	Lexical field	Vague	Mer	Bateau	Coquillage

Figure 3.4: French Category.

c) Arabic

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
Arabic	Letter position	باب	بـ	ب	Door image
	Word placement in the sentence	الفتاة تكتب درسها	تكتب	الفتاة	Image of a girl writing her lesson
	Sorting by content/ meaning	جملة	حرف	كلمة	فقرة
	Crossword	جملة	جمل	باجة	تحتوي كل الكلمات على حرف ج
	Letter illustration	مطر	جمل	قام	
	Same letters	فرح	حرف	رفح	

Figure 3.5: Arabic Category.

d) English

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
English	Category of words and their contractions	Do not	Don't	We will	We'll
	Category of superlative/comparative	Well	Better	The best	
	Tense	Yesterday	Past	Was	
	Possessive case	The dog	The dog's		

Figure 3.6: English Category.

e) Logic

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
Logic	Palindrome	AZIZA	OMO	ELLE	1331
	Similarity	DEFI	FILS	LOT	
		Explication: The letters in the words are ordered alphabetically.			
		A-B	E-F	I-J	Y-Z
		Explication: The 2nd letter is preceded by a vowel.			
	Common word	Café	Noir	Magie	Cheveux
		Explication: Noir: café noir/Magie Noir /Chebeux Noir			
		Solaire	Café	Chantilly	
		Explication: Crème: café crème/crème solaire/crème chantilly			
	Riddle	Je suis la 1ere parmis 26 soldats	Je suis la 1ere parmis les 6 soeurs	A	

Figure 3.7: Logic Category.

f) History/Geography

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
History (التاريخ)	Personalities	رئيس تونسي	الحبيب بورقيبة	Picture of Habib Bourguiba	
	Events	استقلال البلاد التونسية	20 مارس 1956		
	Places	الاهرمات	مصر	Image of the Egyptian pyramids	
	Discoveries	**Christophe Colomb**	اكتشف أمريكا	كريستو ف كولومب	
	Civilizations	حسينية	عثمانية	رومانيّة	تونس
Geography (الجغرافيا)	Continents	أفريقيا	Picture of the Tunisian flag		
	Directions	شمال	بنزت	مدنين	جنوب
	Capitals	تونس	Picture of the Tunisian flag		
	Population	12 مليون ساكن	Picture of the Tunisian flag	الجزائر	44 مليون ساكن
	Climate	الصحراء	حرارة مرتفعة	Image of the desert	
	Natural resources	فسفاط	Picture of the Tunisian flag	تونس	

Figure 3.8: Science Category.

g) Science

Subject	Logical Category	Cards			
		CARD 1	CARD 2	CARD 3	CARD 4
الايقاظ العلمي (علم الأحياء)	Animal movement	ثعبان	زحف		
	Food source and its purpose	حيواني	نمو الجسم	cheese image	
	Lunch times	حليب	الصباح	7:00	شمس
	Life cycle	4 images representing the life cycle of plant growth			
	Mother and offspring	Picture of a female cat	Picture of a kitten		
	Benefits and drawbacks	Picture of a smoker	مضر	Picture of someone doing sports	نافع
	Cause and effect	Image of pollution	3 Images depicting the harmful effects of pollution		
	The senses	الأنف	الرائحة	الشم	
	Animals and their characteristics.	أسد	كالش	غابة	
الايقاظ العلمي (فزياء)	Periodic/ Non-periodic	عيد ميلاد	دوري	مطر	غير دوري
	Time	18:30	السادسة و النصف مساء	Picture of a clock showing 18:30	
	Energy	الطاقة الهوائية	الرياح		
	Types of matter	ماء	سائل		
	Invention	الجاذبية	Albert Einstein's picture	Picture illustrating gravity.	

Figure 3.9: Science Category.

3.2.3 Categorization

This section characterizes the classification of categories into 6 classes.

1. **Calculation:** This category involves matching cards based on mathematical operations or calculations. Players may need to add, subtract, multiply, or divide numbers on the cards to find matching pairs.

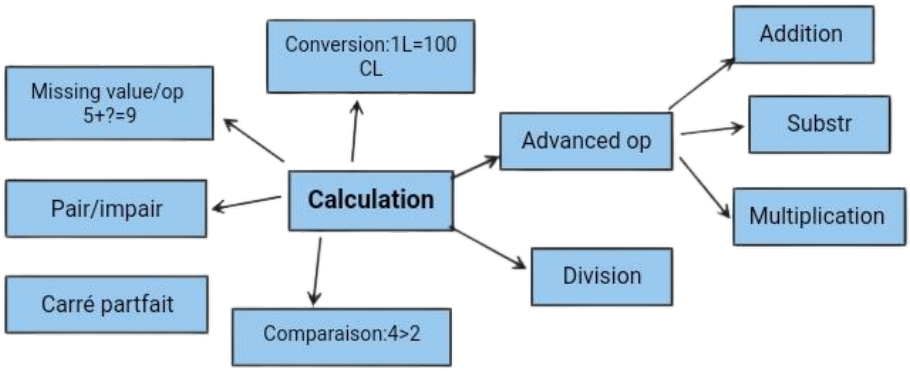


Figure 3.10: Science Category.

2. **Transformation:** Matching cards based on transformations involves identifying pairs that represent different stages or forms of the same object or concept. For instance, matching a card with a caterpillar to its corresponding card with a butterfly.

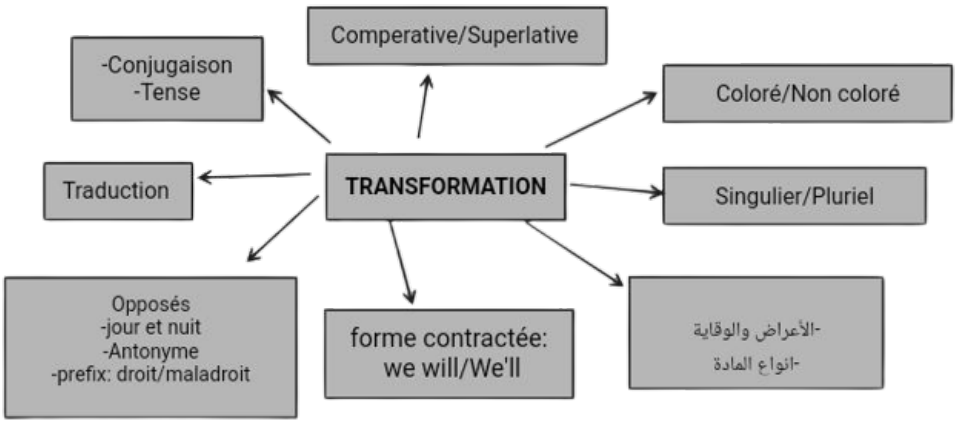


Figure 3.11: Transformation.

3. **Sequencing:** Matching cards based on sequencing involves arranging cards in a specific order or pattern. This category requires players to identify cards that fit within a sequence or follow a logical progression.

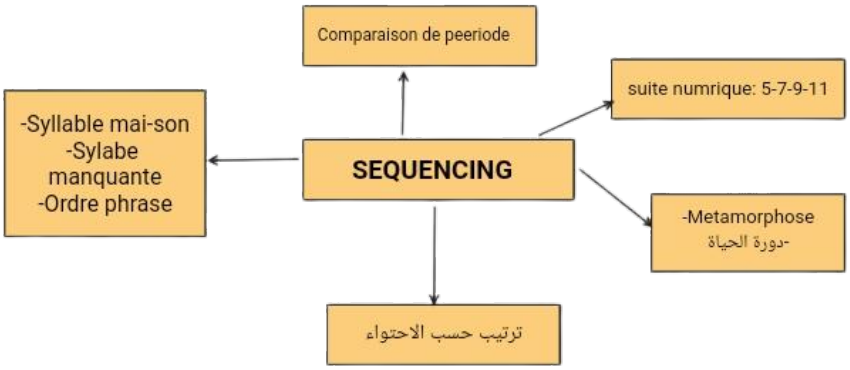


Figure 3.12: Sequencing.

4. **Similarity:** Matching cards based on similarities involves finding pairs that share common attributes or characteristics. For example, cards with matching colors, shapes, patterns, or symbols would fall into this category.

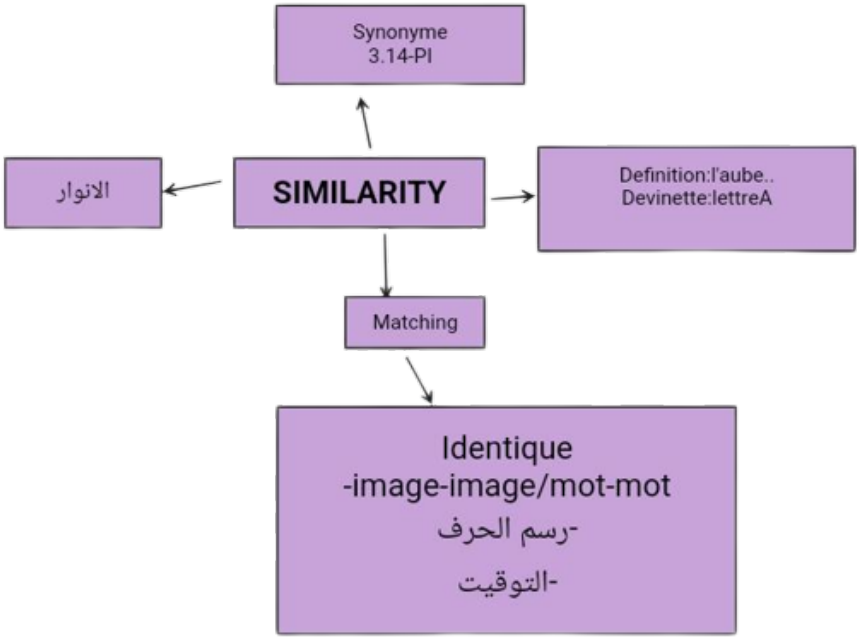


Figure 3.13: Similarity.

5. **Inclusion:**This category involves matching cards based on the concept of inclusion or containment. Players would look for pairs where one card represents a whole object or concept, while the other card represents a part or component of that whole.

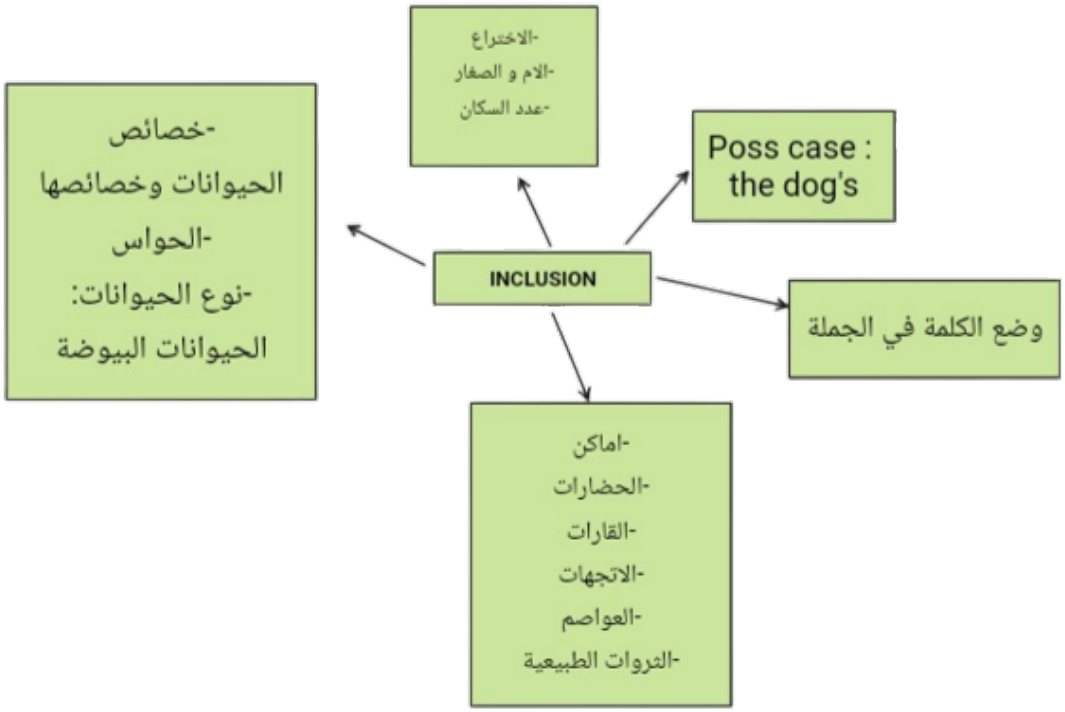


Figure 3.14: Inclusion.

6. **Association:** Matching cards based on associations involves finding pairs that are connected by a specific relationship or association. For example, matching cards that depict cause and effect, opposites, synonyms, or items belonging to the same category.

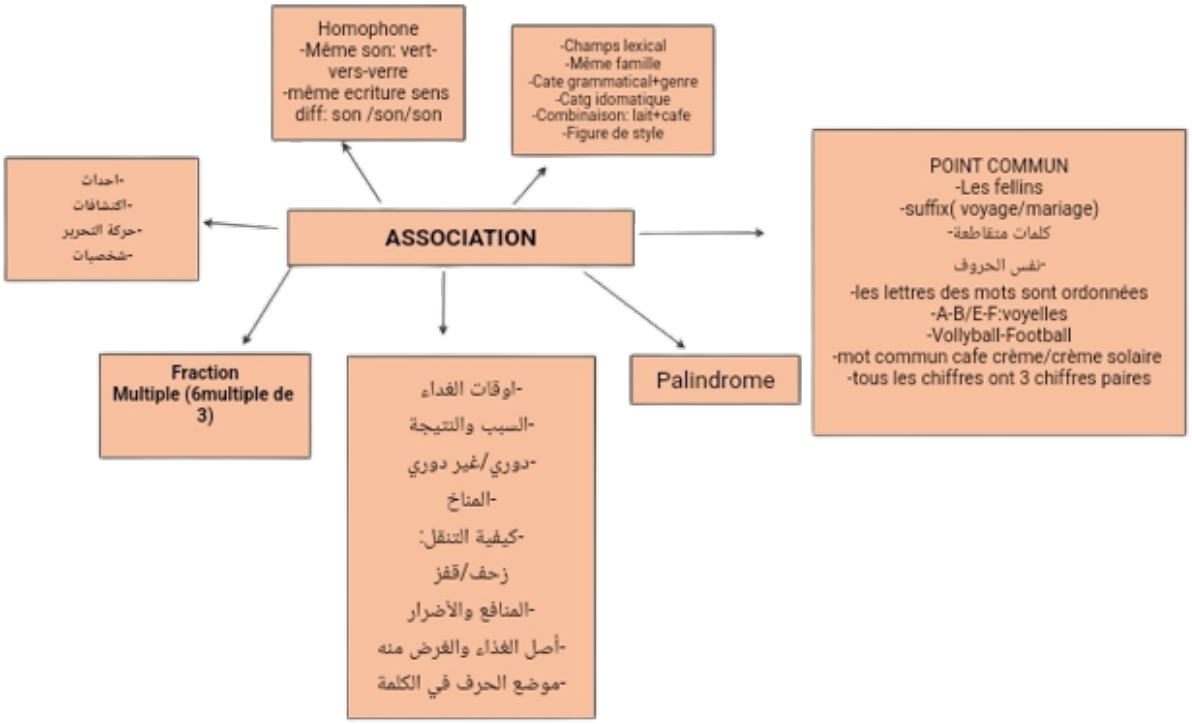


Figure 3.15: Association.

3.3 Models

3.3.1 GANs

Generative adversarial networks (GANs) are a type of machine learning model that can be used to generate new data that is similar to existing data.GANs have been used to generate a wide variety of data, including images, text, and music.

1. **GANs applications:**
   • GANs can be trained on the images of humans to generate realistic faces.
   • Generating faces of anime characters
   • Style transformers
   • Image inpainting
   • GANs can build realistic images from textual descriptions of objects

2. **GANs architecture:**
   GAN consists of two neural networks, the generator and the discriminator, which are trained simultaneously through a competitive process.

Generator: The generator network creates synthetic data, such as images, audio, or text, from random noise or other input data. Its goal is to generate data that is indistinguishable from real data.

Discriminator: The discriminator network evaluates the data it receives and attempts to distinguish between real data and data generated by the generator. Its goal is to become better at discriminating between real and generated data.

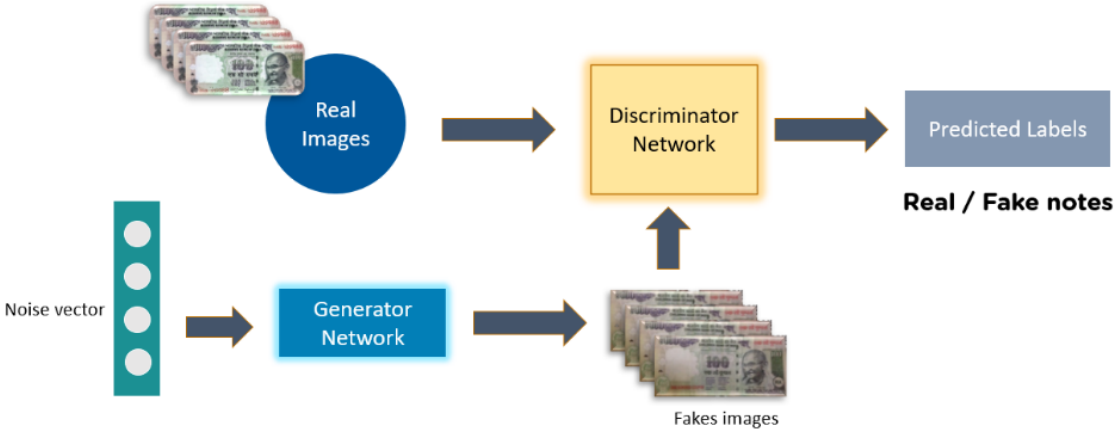


Figure 3.16: GANs.

3.3.2 RNN

Recurrent Neural Networks (RNNs) are a type of neural network designed to handle sequential data wit variable length and complexity. They maintain a hidden state that stores information from previous inputs, allowing them to capture temporal dependencies and context in the data. RNNs can take various inputs and outputs, making them versatile for tasks like sentiment analysis, captioning videos, and generating sequences.

1. **What are the challenges of RNNs?**
   The advantages of RNNs lie in their ability to handle variable-length data and capture long-term patterns,enabling more accurate and creative applications. However, they face challenges such as the vanishing or exploding gradient problem, affecting their ability to learn from long-term dependencies, and the computational cost and memory requirements associated with processing sequential data.

2. **RNN Applications**
   RNNs find applications in various domains, including natural language processing, speech recognition, computer vision, and music generation. They can perform tasks such as machine translation, text summarization, image captioning, speech synthesis, and music composition. Their ability to model sequential data and context makes them valuable tools for processing and generating complex and creative content.

3. **How it works?**
   RNN maintain a hidden state that stores information from previous time steps, enabling them to capture patterns in sequential data. At each time step, an RNN combines the current input with the hidden state from the previous step to generate an output and update the hidden state. This recurrent structure makes RNNs suitable for tasks like natural language processing and time series analysis.
   However, they can face challenges with vanishing gradients. **LSTM** and **GRU** variants were developed to address these issues and enhance their performance.

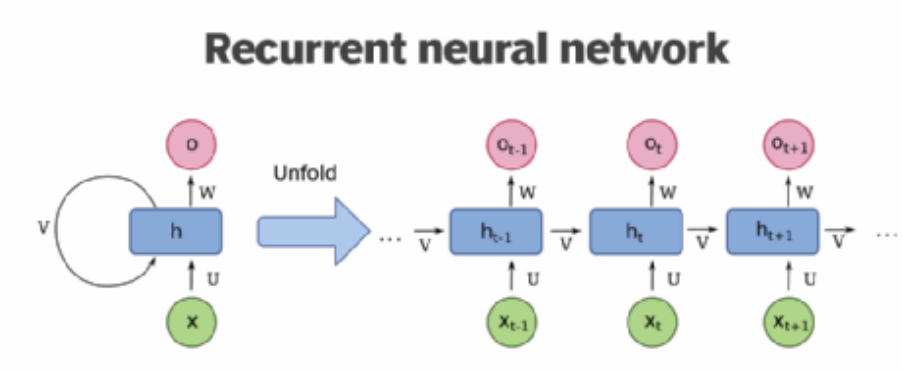


Figure 3.17: RNN.

3.3.3 Transformers

Transformers are a type of deep learning model architecture introduced in the paper ”Attention Is All You Need” by Vaswani et al. in 2017. They have since become a fundamental architecture for various natural language processing (NLP) tasks.
Transformers rely heavily on the concept of attention mechanisms to process input data in parallel rather than sequentially, making them highly efficient and suitable for both short and long sequences.
The Transformer architecture consists of two main components:The encoder processes the input sequence, while the decoder generates the output sequence. This structure is particularly effective for tasks like machine translation.

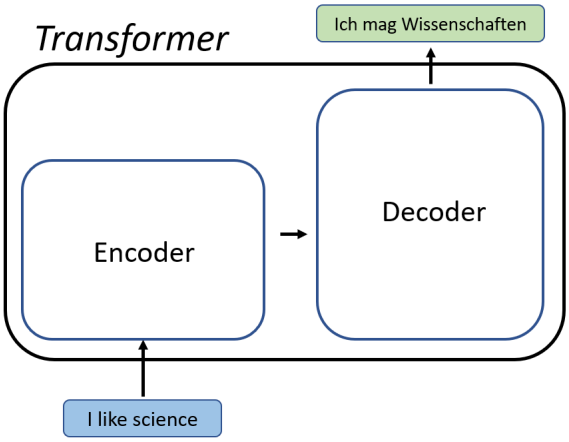


Figure 3.18: Transformers.

3.3.4 Association Categories-Models

The principle consists of associating each category presented previously with a model among RNNs, GANs, and Transformers based on their respective modes of operation. This image presents an idea of the association of categories with models.

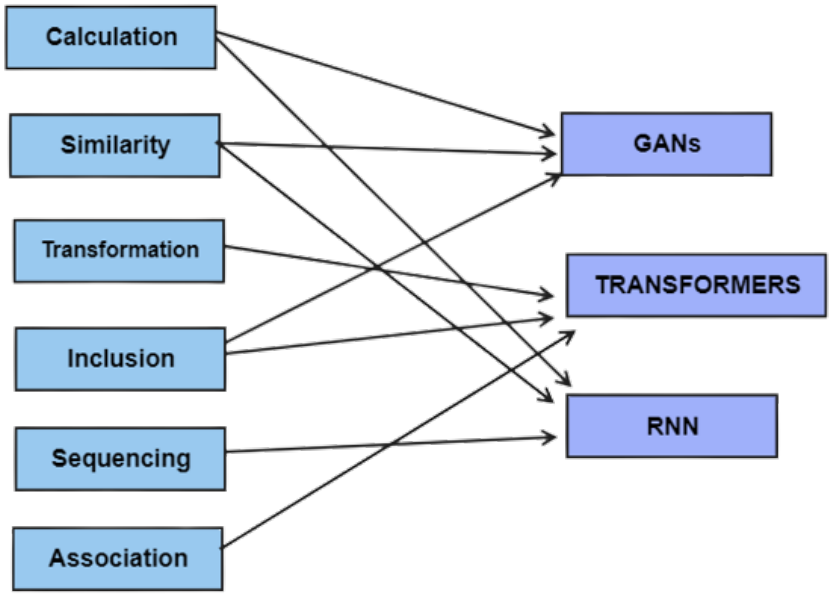


Figure 3.19: Classification Models.

3.4 Pipeline

A pipeline explains the process of our models, and each model has its own distinct functionalities. This process is typically divided into multiple stages to achieve a specific objective.
This pipeline takes an input request with parameters **(Skills, Chapter, Sub-Thematic, Thematic, Subject, Level,Level Type,nSetofCard, nCardPerSet)**. Let's consider an example request (Maths, Calculus, Addition,subtraction, Addition, Primary, 2nd year, 12, 2).
The output will be represented in JSON format, where nSetofcard represents the total number of cards, and nCardPerSet is the number of identical cards. The combination is made in pairs of 2 cards.

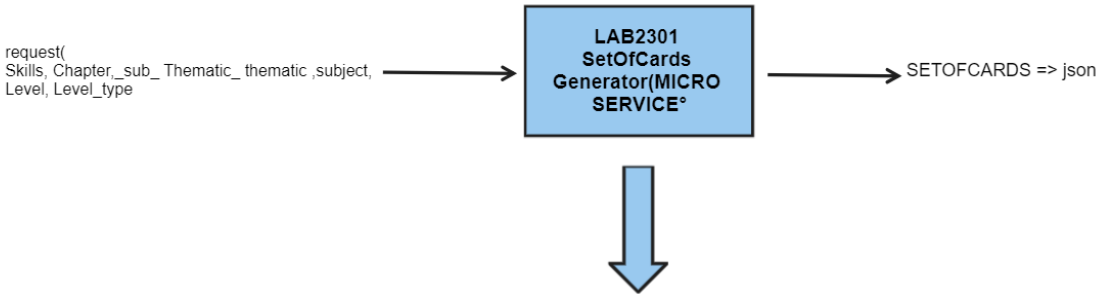


Figure 3.20: Pipeline 1 .

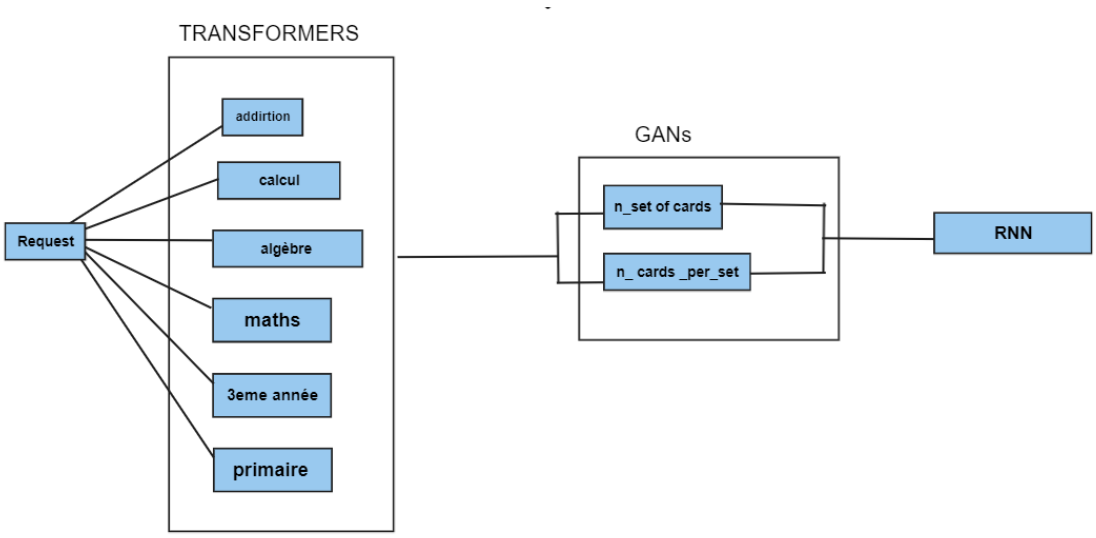


Figure 3.21: Pipeline 2 .

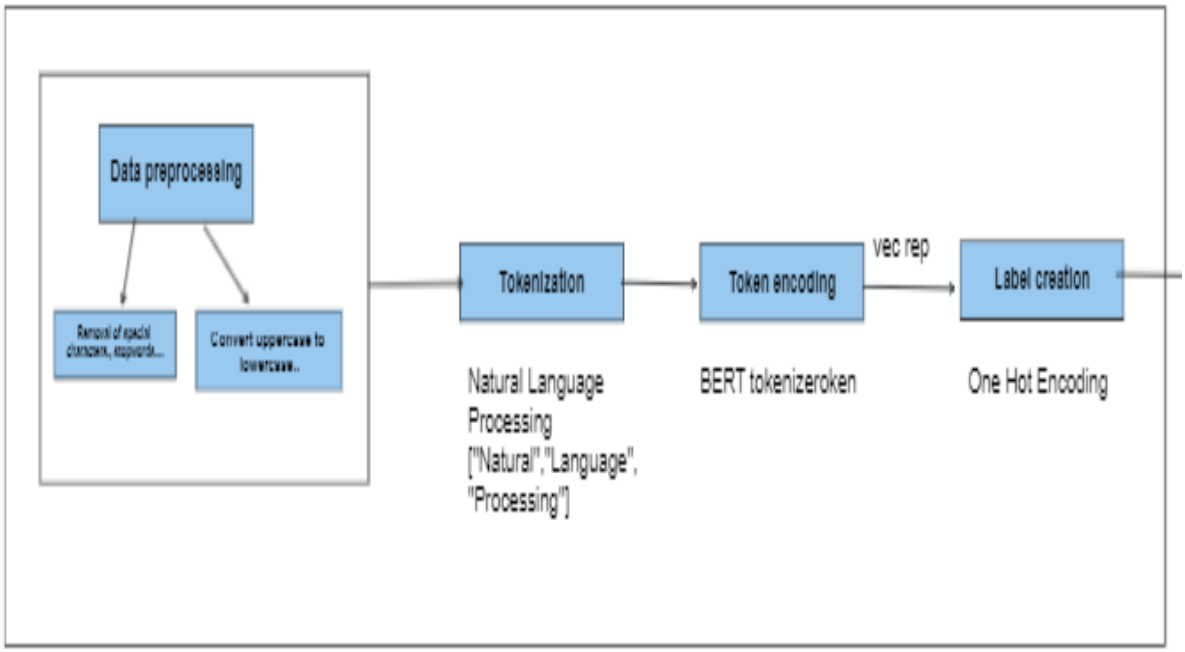


Figure 3.22: Pipeline Data Preparation .

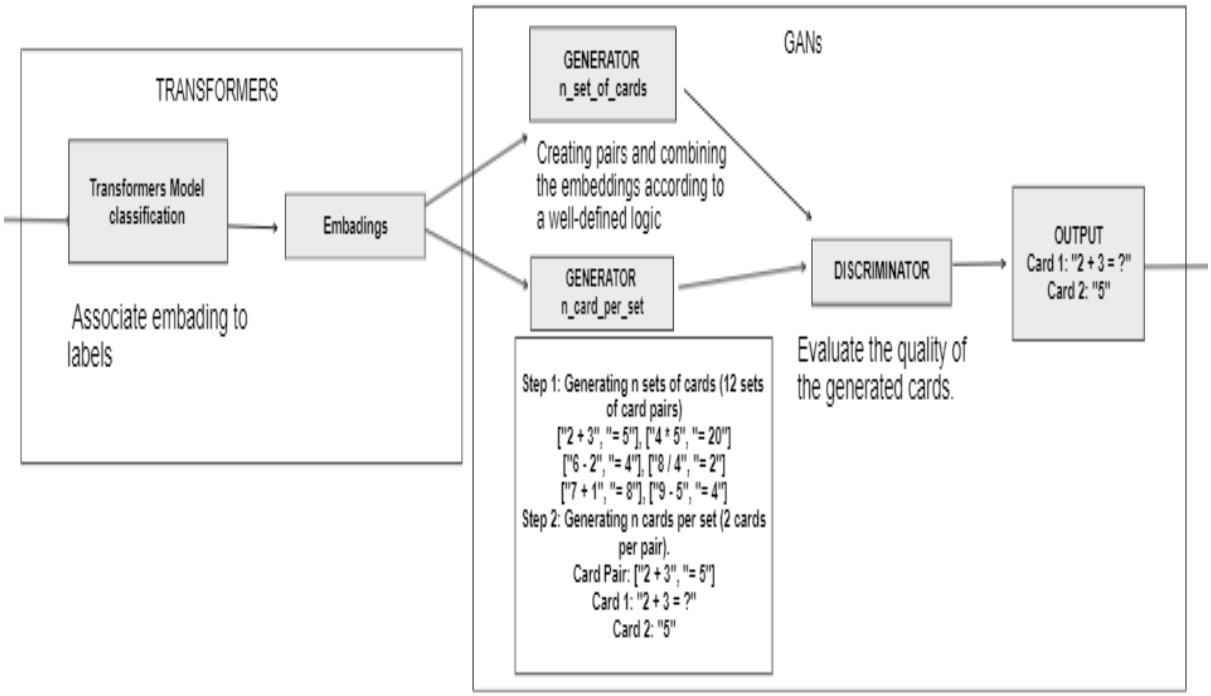


Figure 3.23: Pipeline GANs, Transformers.

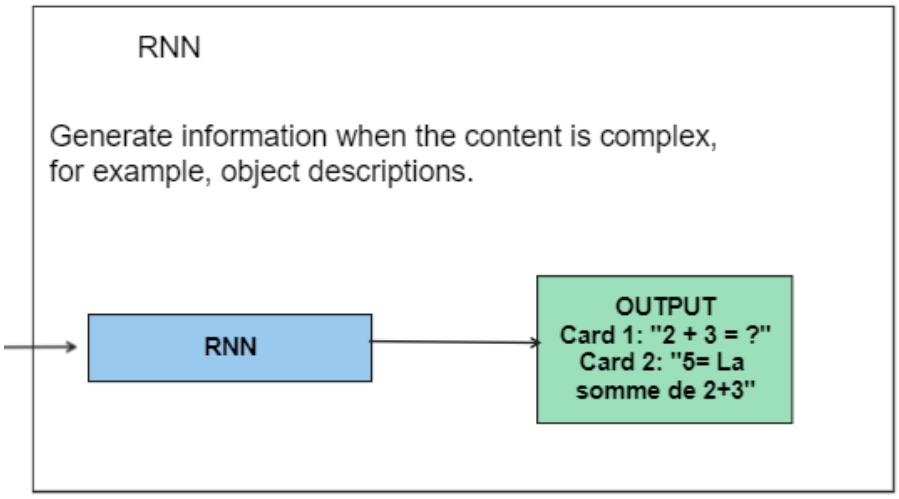


Figure 3.24: Pipeline RNN .

3.5 Achievement

This part illustrates the achievement of the tasks requested and is divided into two sections: the first part addresses the tasks achieved, while the second part examines the challenges that have not been met.

3.5.1 Achievements Realized

The three main stages of our Set of Cards Generator game are web scraping, data preparation, and modeling.

1. **Web Scrapping:**
   Data extraction or web scraping is a crucial and fundamental step in our project to gather relevant and useful data from various solved exercises at different levels (primary school, middle school, and high school) and increasing levels of difficulty.

   As engineers, we always aim to perform tasks in an automated manner to optimize our work. For the data extraction part, we chose to create a script that collects all the links from Google related to a given theme. These links can be in HTML or PDF format and are stored in an Excel (XLSX) file.

   This method is employed to avoid the traditional approach of collecting links one by one, which is time-consuming and can lead to missing some links.

   Once the links are collected, a manual sorting process will be carried out to carefully select useful and valid links (examples of invalid links include links containing paid Amazon books). The newly selected links will also be stored in another XLSX file.

   The new file consists of two columns: the first column contains the links, and the second

column provides a description of the PDF.

LINK PDF	DESCRIPTION
http://boutdegomme.fr/ekladata.com/boutdegomme.eklablog.com/perso/math/lecons/Les-doubles-entier-new.pdf	2ème année primaire mathematique calcul basique addition
http://boutdegomme.fr/ekladata.com/boutdegomme.eklablog.com/perso/math/tables%20d-addition.pdf	1ère année primaire mathematique calcul basique addition
https://www.mathslibres.com/addition/addition_singuliere_x_quelquesretenues_001.1365172890.pdf	1ère année primaire mathematique calcul basique addition
https://www.mathslibres.com/addition/addition_singuliere_x_quelquesretenues_002.1365172891.pdf	1ère année primaire mathematique calcul basique addition

Figure 3.25: PDF Link .

The next step is to extract data from the various links, which is why we have implemented a script for extracting data from PDFs. This script offers a generic solution for link extraction, meaning it is applicable to all resources. This method is superior to creating a single script for each PDF.
To successfully achieve data extraction, we utilized OCR.

OCR: stands for Optical Character Recognition. It is a technology that recognizes and extracts text characters from these documents, making it possible to edit, search, and manipulate the text as if it were created digitally. OCR technology is widely used for digitizing printed documents, automating data entry, and enabling text-based searches within scanned documents.

2. **Data Preparation:**
The data preparation phase is used to clean the extracted data by removing duplicates, resulting in a more refined and expanded dataset.

Once the data is clean, we need to store it in a database. In our case, we chose MongoDB, which is a NoSQL database management system. MongoDB is designed to store and manage large volumes of unstructured or semi-structured data, making it particularly suitable for handling various types of data, including documents, JSON-like data, and binary data.
The first image illustrates the architecture of our database, where we opted for classification based on categories: association, calculation, transformation...

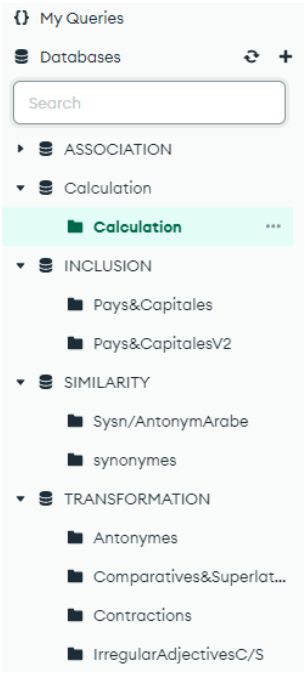


Figure 3.26: MongoDB DataBase.

The second image explains a well-defined representation consisting of a precise description, the content of the first card, and the content of the second card. This example rep-

```
_id: ObjectId('64e8feddbba75ba21bb7441f')
_Keywords: "6eme année primaire langue français et anglais vocabulaire Vêtements t…"
Card1: "un manteau"
Card2: "coat"
```

Figure 3.27: MongoDB Representation.

resents the description**6eme année primaire langue français et anglais vocabulaire Vêtements traduction**

3.5.2 Unrealized Accomplishments

Among these tasks, we can mention data collection because it encompasses all levels, from primary school to high school, and all subjects. This task requires more time, as does the part of data preparation and data cleaning.
While I have tried my best to find intelligent and automated methods in data collection, searching for useful links, I have also aimed to develop other methods that will streamline the initial process.

3.6 Conclusion

The third chapter is drawing to a close. Firstly, it introduces the different logical categories, classifying them into various groups.
Next, it explains the models that can be useful in our project, and a pipeline is illustrated to explain how these models operate based on the card game.
The final part of this chapter presents the tasks that have been accomplished and the unrealized Accomplishments.

Chapter 4

Internship Insights and Project Outlook

4.1 Introduction

This last chapter explores the various aspects that were highlighted during this internship period, as well as the elements that were particularly beneficial for my training and personal development.
In addition, it addresses my expectations and aspirations for the future of this project.

4.2 Internship Insights

Starting a project in its initial phase is undeniably beneficial for me as an engineer. Instead of simply attending a project already underway, having the opportunity to launch it from the beginning has multiple benefits.
This allows me not only to propose new ideas, experiment with different concepts, but also to gain a thorough understanding of emerging notions in the field of generative artificial intelligence, which is still a relatively new field.

By starting a project from the beginning, I can also have a significant impact on its direction and strategy. I can help shape the project vision, develop action plans, and identify areas where improvements or innovations are needed.

This also implies an increased responsibility, because by participating from the beginning, I take charge of the definition of objectives, the planning and the implementation of the project, which strengthens my management and decision-making skills.

In addition, this approach promotes creativity and innovation because it gives me the freedom to explore new ideas and innovate. It stimulates my creativity and my ability to solve problems in an original way. Finally, it requires me to be adaptable and to solve problems quickly, because working on a project in the initial phase often exposes to unforeseen challenges.

Ultimately, working on a project from its initial phase is a rewarding opportunity that promotes the development of technical skills, creativity, initiative, and offers the satisfaction of actively contributing to the future success of the project. It is an experience that opens new

perspectives and stimulates professional growth.

4.3 Project Outlook

This section presents my expectations for this project as well as innovative ideas that could be beneficial for the platform. I expect this project to achieve several goals, including:

1. Among my expectations for the future of this project, one of the priorities is to obtain a significant amount of high quality data. This step is crucial, as clean and well-prepared data is essential to achieve acceptable results in the modeling phase.

2. In addition, I aspire to more automation in our work, with more intelligent and general scripts to simplify the various tasks. Automation will speed up the process while minimizing human error.

3. Another goal is to enrich the variety of our games to stimulate creativity and reflection of children. Offering a diverse experience can make learning more attractive.

4. To further encourage children to study while having fun, we might consider adding an avatar creation option. This way, children can customize their own character. With each level completed, new avatar options could be unlocked, such as choices of hairstyles, clothes, etc. This idea could make the platform more engaging and entertaining for children.

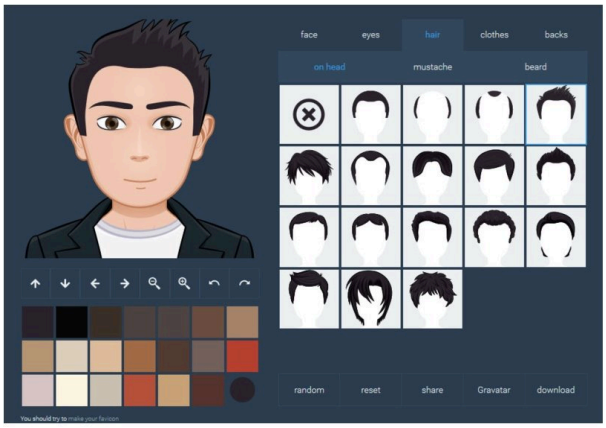


Figure 4.1: Avatar.

5. **Virtual Rewards:** Create a system of virtual rewards for children when they complete tasks or achieve certain goals in educational games. For example, if they correctly answer a number of questions or complete a level, they can earn stars, badges or points. These rewards can be displayed on their avatar profile.

6. **Friendly Competitions:** Organize friendly competitions between players, where children can compete against each other in educational games. You might have a scoreboard

showing the top players or teams. Winners could receive special rewards, such as decorations for their avatar.

7. **Share Achievements:**Give children the opportunity to share their achievements with other children on the platform. For example, if they have solved a difficult mathematical problem, they could publish their solution for others to see and learn. This promotes mutual aid and collaboration.

8. **Interactive Stories:** Embed interactive stories into the platform where children can make decisions to shape the story. This stimulates their creativity and allows them to see the consequences of their choices.

9. **Quest System:** Offer educational quests that encourage children to explore different areas of knowledge. For example, a quest on a specific theme where they have to solve puzzles and collect information.

10. **Discussion Rooms:** Create secure discussion spaces where children can ask questions, share discoveries and discuss topics of interest. This encourages collaborative learning.

11. **Live Events:** Host live events, such as QA sessions with experts or teachers, interactive online workshops, or even live competitions.

12. **Tangible Rewards:** Offer tangible rewards, such as certificates of achievement.

13. **Pronunciation:**The implementation of a system for checking the pronunciation of languages when reading a text within an educational platform represents a major innovation for language learning.
    Using advanced speech recognition technology, students can receive immediate feedback on their pronunciation accuracy, encouraging them to improve. The ability to customize the correction exercises according to the errors detected allows to specifically target the needs of each student.

14. **Accessibility Options:** Make sure the platform is accessible to all children, including those with special needs. Include accessibility options to help them participate fully.

    a) **Audio-Description:** For children with visual impairments, provide an audio-description feature. This means that each visual element, such as an image or graphic, can be described verbally by an automated narrator or an audio recording created by the teacher.

    b) **Subtitles and Transcripts:** Ensure that all videos and audio are accompanied by subtitles. This benefits hearing-impaired children and those who prefer to read the content rather than listen to it. Text transcripts for videos are also essential.

    c) **Adjustable Contrast and Font Size:** Allow users to customize platform contrast, as well as font size and style. This helps children with vision or reading problems to adapt the appearance of the platform to their needs.

d) **Online Support:** Integrate an online support feature so children with special needs can ask questions or ask for help in real time.

15. **21st Century Skills Training:** Teach skills such as critical thinking, problem solving, creativity, communication, and collaboration, which are essential for success in the modern world.

16. **Financial Education Integration:**Teach students the basics of financial education from an early age. Create learning modules on managing money, saving, investing, and understanding economic systems.

17. **Complete Virtual Labs:** Create ultra-realistic virtual lab environments for science. Students could conduct complex virtual experiments with realistic simulations.

18. **Virtual Reality for Reading:** Develop virtual reality stories where students can enter the world of books. This would require considerable resources in terms of 3D design.

19. **Artificial Intelligence for Tutoring:** An AI-based tutoring that can teach advanced subjects. This would be complex to achieve due to the difficulties of some tasks.

20. **Additional courses:**In addition to courses related to the Tunisian education system, users can benefit from other courses in areas of their choice, including programming. They can access and master information about the most widely used programming languages in the world. This course diversity offering allows users to explore and develop their skills in areas that match their professional interests and goals.

21. **Complex and Adaptive Assessments:** Create advanced assessment systems that adapt to each student's skill level, proposing more complex questions and tasks as they progress.

22. **Live Virtual Classes:** Organize live classes where students can ask questions in real time and interact with teachers and classmates.

23. **Peer mentoring:**Create a peer-to-peer mentoring system where older students can help younger students learn.

24. **Collaborative Research Projects:** Allow students to work on collaborative, team-based research projects on advanced topics. This would involve facilitating communication and coordination among students, as well as access to research resources.

25. **Feedback:** As a user, I want to have the opportunity to give my opinion on the platform, share my ideas, and make constructive criticism. This will continuously improve the user experience and evolve the platform based on their needs and suggestions.

26. **Career Simulators:** The goal of career simulators is to help students explore and understand different professions, allowing them to virtually experience working in these areas. This can help them make informed decisions about their academic and professional future.
    This idea is valid and will assist high school students at the **BAC** level when it comes to making choices for their future career paths.
    Here's how a career simulator could operate:

    a) **Choosing a Profession:** Students select from a list of professions they'd like to explore. This could include careers in medicine, engineering, computer science, journalism, art, etc.

    b) **Virtual Simulation:** Once the profession is chosen, students enter a virtual simulation that recreates the work environment of that profession. For instance, if they choose medicine, they might find themselves in a virtual hospital.

    c) **Tasks and Responsibilities:** Students are introduced to the daily tasks and responsibilities of the profession. They can interact with virtual patients, solve profession-specific problems, or complete missions related to that career.

    d) **Virtual Mentoring:** In some cases, virtual characters or mentors can guide students through the simulation, providing them with advice and information about the profession.

    e) **Assessment:** At the end of the simulation, students may receive an assessment of their performance in the simulated profession. This can help them understand if they are suited for that career.

    Career simulators offer several advantages to high school students:

    a) **Exploration:** They allow students to explore different professions before making important decisions about their academic path.

    b) **Informed Decision-Making:** Students can make informed decisions about their choice of major or career by gaining a better understanding of what it's like to work in these fields.

    c) **Reduced Stress:** They can reduce career-related stress by providing a hands-on experience.

    d) **Motivation:** Simulators can also generate interest and motivation among students as they get to see what the working world is like.

    e) **Adaptability:** Students can try multiple simulations to discover what suits them best.

4.4 Conclusion

The last chapter was dedicated to the presentation of Internship Insight, while the second part of this chapter was devoted to proposing new innovative ideas that could be integrated into the educational platform.

Chapter 5

General Conclusion

In this engineering internship report, I have provided a detailed overview of the various stages involved in implementing my project, which aims to design an educational game titled **Set of Card Generator**. My work has been structured as follows:

In the initial stage, I conducted an analysis of existing solutions, studying and evaluating similar applications such as **IXL** and **MAXICOURS**. This analysis allowed me to identify the weaknesses of these applications and define the objectives to be achieved with my project.

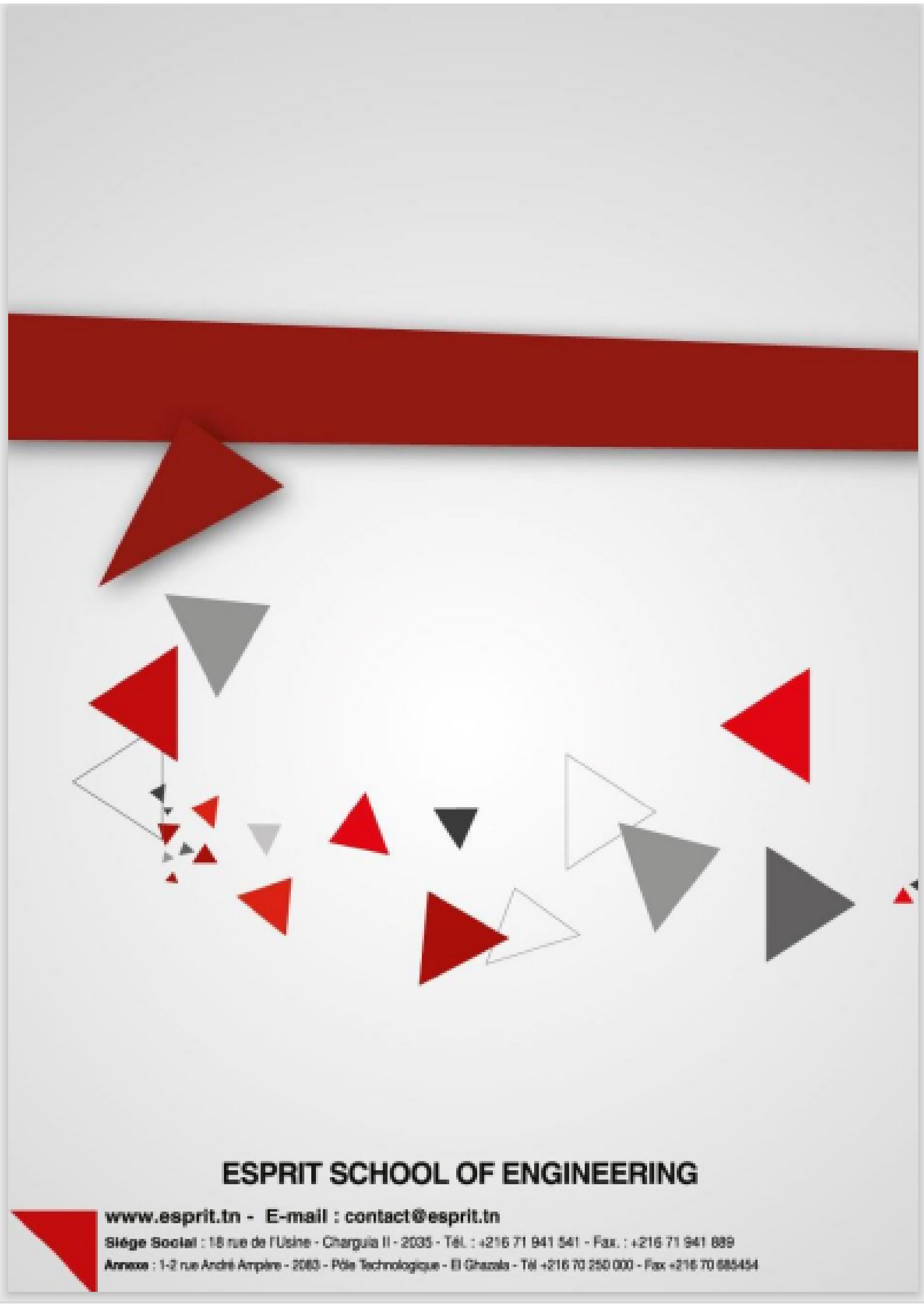
Subsequently, I presented the solution I developed, providing a detailed explanation of the appropriate methodology I employed for the project's execution.

The second chapter was dedicated to presenting the functional requirements, including business objectives and data science objectives. I also addressed non-functional requirements such as execution speed and application security. In this section, I described the various software and technologies I utilized, such as Jupyter and MongoDB.

Chapter 3, titled "Achievements," translated what I accomplished during my internship. I began by exploring the various logical category possibilities and their classification. Then, I introduced various models such as **GANs**, **Transformers**, and **RNN**, along with a pipeline to illustrate how these models operate. Finally, I listed the tasks I successfully completed and those that remained unfinished.

The last chapter was devoted to "Internship Insights." I emphasized the importance of participating in an internship from its early stages, which allowed me to acquire numerous skills and knowledge. In the "Project Outlook" section, I shared my future perspectives on this project and proposed innovative and original ideas that could be integrated into the platform.

In summary, this report highlights the journey of my internship, the accomplishments achieved, the objectives fulfilled, and the promising prospects for the future of this educational project.



ESPRIT SCHOOL OF ENGINEERING

www.esprit.tn - E-mail : contact@esprit.tn

Siége Social : 18 rue de l'Usine - Charguia II - 2035 - Tél. : +216 71 941 541 - Fax. : +216 71 941 889

Annexe : 1-2 rue André Ampère - 2083 - Pôle Technologique - El Ghazala - Tél +216 70 250 000 - Fax +216 70 685454