الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

20 Août 1955 University of Skikda

Faculty of Sciences

Department of Computer Science

An innovative project for a startup institution's graduation project 1275.

ThesisGuard AI:
Discerning the Incursion of AI-Generated Texts in Dissertations

Specialty:

Advanced Information System and Applications

Realized by : Aoued Khouloud

under supervised: Dr. Bougamouza Fateh

2023-2024

Thanks and Appreciation

With God's grace and help, this modest work has been accomplished, and we ask Almighty God to make it purely for His sake.

I am pleased, after praising and thanking God, to extend my sincere thanks and appreciation, with the utmost respect and beautiful gratitude, to Professor Bougamouza Fateh, who accepted the supervision of this work. He has greatly contributed to the preparation of this thesis and has always generously provided me with his guidance, advice, encouragement, and moral support. I also thank the professors of the University of August 20, 1955, especially those contributing to the incubator and continuous training.

To all of them, and to all those whose names we were unable to mention, I extend my heartfelt thanks.

Dedication

This thesis is dedicated to the cherished memory of my beloved grandmother, whose unwavering dream was to see me become an engineer. Your endless love and aspirations for me have been a source of inspiration, even in your absence. I hope I have made you proud.

To my grandfather, whose wisdom, strength, and gentle guidance have left an indelible mark on my journey. Your legacy continues to inspire me to strive for excellence.

To my mother, who sacrificed and worked tirelessly for my sake, your boundless love and selflessness have been the cornerstone of my success. Thank you for your endless patience, your comforting words in times of doubt, and your unwavering belief in my potential.

To my father, whose steadfast support and encouragement have been a beacon of hope and motivation. Your constant reassurance and pride in my achievements have fueled my determination.

To my only uncle, whose guidance and support have been invaluable. Your wisdom and advice have been crucial in navigating this journey.

To my sisters and brothers, whose love and companionship have been a constant source of joy and strength. Your belief in me has given me the courage to pursue my dreams without hesitation.

To my aunts Amel et Loubna, whose care and warmth have been a comforting presence. Your kindness and support have been deeply appreciated.

To my best friend Soumia, whose loyalty and belief in me have been unwavering. Thank you for standing by my side through every challenge, for your words of encouragement, and for always believing in my capabilities.

To my colleagues, for their collaboration, camaraderie, and for making this journey a shared experience filled with learning and growth. Your support has been instrumental in my academic endeavors.

To my entire family, whose collective love, faith, and encouragement have been the driving force behind my perseverance and achievements. Each one of you has played a significant role in this journey, and I am eternally grateful.

And lastly, to myself, for believing in my dreams, for enduring every challenge with resilience, and for never taking a day off from striving to be stronger and better. This accomplishment is a testament to the power of perseverance and the strength that comes from within. I am proud of the person I have become and the dedication I have shown throughout this journey.

ABSTRACT

Artificial intelligence has revolutionized many industries, including academic research. With the emergence of AI-generated texts, researchers can now produce content with minimal human effort that appears high-quality but actually lacks the foundations of scientific research. However, the use of AI-generated texts in dissertations raises significant questions about their impact on research quality and integrity.

In this work, we propose an approach to address both human-generated and AI-generated scientific texts in academic dissertations and theses. The ThesisGuard AI application aims to enhance academic integrity and quality while avoiding total and direct reliance on artificial intelligence, due to its impact on individual abilities and the limitations it imposes on intellectual creativity. This approach involves using AI to detect and analyze the authenticity of the texts in question.

Résumé

L'intelligence artificielle a révolutionné de nombreux secteurs, y compris la recherche académique. Avec l'émergence des textes générés par l'IA, les chercheurs peuvent désormais produire du contenu avec un effort humain minimal qui semble de haute qualité mais qui manque en réalité des fondements de la recherche scientifique. Cependant, l'utilisation de textes générés par l'IA dans les dissertations soulève des questions significatives concernant leur impact sur la qualité et l'intégrité de la recherche.

Dans ce travail, nous proposons une approche pour aborder à la fois les textes scientifiques générés par des humains et par des IA dans les dissertations et thèses académiques. L'application ThesisGuard AI vise à améliorer l'intégrité académique et la qualité tout en évitant une dépendance totale et directe à l'intelligence artificielle, en raison de son impact sur les capacités individuelles et des limitations qu'elle impose à la créativité intellectuelle. Cette approche implique l'utilisation de l'IA pour détecter et analyser l'authenticité des textes en question.

ملخص

الذكاء الاصطناعي ثورة في عدة قطاعات، بما في ذلك البحث الأكاديمي. مع ظهور النصوص المولدة بواسطة الذكاء الاصطناعي، يمكن للباحثين الآن إنتاج محتوى بجهد بشري أدنى يبدو عالي الجودة ولكن في الحقيقة يفتقر إلى أسس البحث العلمي. ومع ذلك، فإن استخدام النصوص المولدة بواسطة الذكاء الاصطناعي في الرسائل العلمية يثير أسئلة كبيرة حول تأثيرها على جودة ونزاهة البحث.

في هذا العمل، نقدم نهجًا للتعامل مع النصوص العلمية المولدة سواء بواسطة البشر أو بواسطة الذكاء الاصطناعي في الرسائل والأطروحات الأكاديمية. تهدف تطبيق ThesisGuard AI إلى تعزيز النزاهة الأكاديمية والجودة مع تجنب الاعتماد الكلي والمباشر على الذكاء الاصطناعي، نظرًا لتأثيره على القدرات الفردية والقيود التي يفرضها على الإبداع الفكري. يتضمن هذا النهج استخدام الذكاء الاصطناعي لاكتشاف وتحليل أصالة النصوص المعنية.

Table des matières

Second part: Commercial aspect

First Axis: Project submission

Second Axis:innovative aspects

Third Axis:strategic analysis of the market

List of Tables

First part: Theoretical Aspect

Second part : commercial aspect

List of Figures

First part: Theoretical Aspect

First part: Theoretical Aspect

General Introduction

In recent years, significant progress has been made in text generation. The latest text generation models are revolutionizing the domain by generating human-like text. It has gained wide popularity recently in many domains like news, social networks, movie scriptwriting, and poetry composition, to name a few. The application of text generation in various fields has resulted in a lot of interest from the scientific community in this area.

Presently, achieving perfect detection of AI-generated text across all domains remains an intangible goal. However, significant research efforts have been directed towards detecting AI-generated content in specific domains, including academia, scientific research, fake news, reviews, and misinformation. By narrowing the focus to these specific areas, researchers can better tailor detection methods to address the unique challenges presented within each domain.

One prominent focus of research is AI-generated text detection in academic and scientific settings. With the proliferation of AI tools capable of generating academic papers or scientific articles, there is a growing need to distinguish between genuine research and AI-generated content. Detection methods in this domain often leverage linguistic analysis, citation patterns, and domain-specific knowledge to identify anomalies indicative of AI-generated text.

By focusing on specific domains, researchers can tailor detection methods to the unique characteristics and challenges present in each domain. This targeted approach allows for more effective detection of AI-generated text by leveraging domain-specific knowledge and features. However, ongoing research is needed to further refine and enhance detection techniques to keep pace with evolving AI capabilities and the proliferation of AI-generated content across various domains.

Problem statement

To propose an innovative solution that addresses the challenge of distinguishing between AI-generated and human-written texts in academic settings, we will develop a strategy by fine-tuning the CamemBERT pre-trained model. This will be based on a dataset created from ancient theses, where the original texts are reformulated using ChatGPT. The approach will involve leveraging the Transformers framework to enhance the model's performance in detecting AI-generated content.

More specifically, our methodology includes the following steps:

1. Dataset Creation: Compile a dataset from ancient theses, ensuring a diverse representation of academic disciplines. The original texts will be reformulated using ChatGPT to generate AI-generated counterparts.

2. Pre-processing: Develop a pre-processing pipeline tailored to academic texts, ensuring that both the original and reformulated texts are clean, structured, and suitable for training the model.

3. Fine-tuning CamemBERT: Utilize the Transformers framework to fine-tune the CamemBERT model on the prepared dataset. This will involve adjusting the model's parameters to optimize its ability to distinguish between AI-generated and human-written texts.

4. Experimental Study: Conduct an experimental study to determine the optimal hyper-parameters for the fine-tuned model. This will include assessing the impact of various parameters on the model's accuracy and reliability.

5. Model Evaluation: Evaluate the model's performance through rigorous testing, using both synthetic and real-world academic texts. This step ensures the robustness and effectiveness of the model in different scenarios.

6. Integration into ThesisGuard AI: Implement the fine-tuned model into the ThesisGuard AI platform, providing users with an intuitive and efficient tool for verifying the authenticity of their academic work.

Objectives

· **Ambiguity in Text Origin**: Current AI models produce texts that are nearly indistinguishable from those written by humans, making manual detection extremely difficult.

· **Impact on Academic Integrity**: The use of AI-generated content in academic submissions threatens the authenticity and credibility of scholarly research.

· **Lack of Existing Solutions**: There is a shortage of robust tools designed specifically to detect AI-generated texts within the context of academic writing.

· **Language and Domain Specificity**: The nuances of academic writing and the specific terminologies used in various fields add complexity to the task of distinguishing AI-generated content.

Chapter 1:

Deep Learning

1. Introduction

In recent years, Deep Learning (DL), a subset of Machine Learning, has dramatically transformed the field of Natural Language Processing (NLP). It has significantly enhanced the performance of various NLP tasks, including machine translation, sentiment analysis, and question answering.

This chapter provides the essential background knowledge required for this work. We begin with a brief overview of machine learning and its common algorithms. Following this, we delve into the fundamentals of deep learning and artificial neural networks, presenting common activation functions and loss functions. We then explore two specialized types of artificial neural networks: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Finally, we provide a detailed description of the Transformer architecture and introduce the BERT model, emphasizing the role of Transformers in advancing NLP.

2. Machine learning

Machine learning (ML) is a field of study that focuses on developing algorithms and statistical models enabling computer systems to perform tasks without explicit programming instructions. These learning algorithms are pervasive in numerous applications we encounter daily. For instance, web search engines like Google employ learning algorithms to rank web pages effectively, enhancing the search experience. ML algorithms find utility across diverse domains, including data mining, image processing, and predictive analytics, among others.

One of the primary advantages of employing machine learning is its ability to automate tasks once the algorithm has learned how to handle the data effectively. This automation capability streamlines processes, reduces manual intervention, and enables systems to operate autonomously, contributing to enhanced efficiency and productivity. [1]

2.1.Types of machine learning [2]

Machine learning encompasses various types, each distinguished by unique characteristics and applications. Some of the primary machine learning algorithms include: Supervised Machine Learning, Unsupervised Machine Learning, Semi-Supervised Machine Learning, and Reinforcement Learning. See Figure 1:

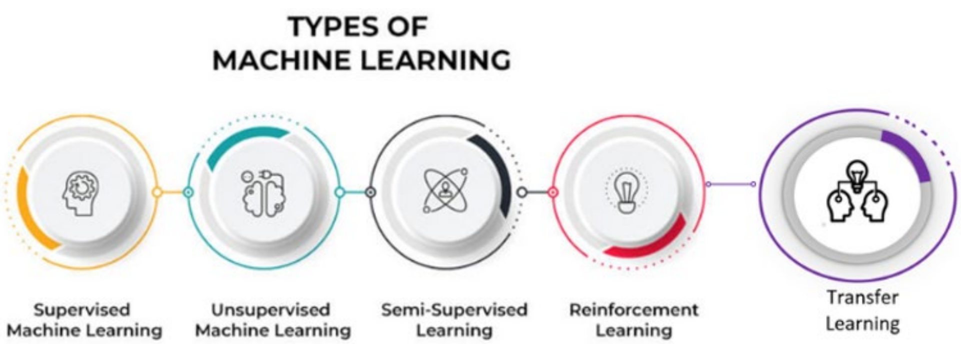


Figure 1: Types of machine learning

2.1.1.Supervised Machine Learning

Supervised learning constitutes a foundational aspect of machine learning, focusing on training models to discern patterns within labeled datasets. By utilizing labeled training data, where inputs are paired with corresponding outputs, supervised learning seeks to derive a functional mapping capable of generalizing to new, unseen data.

This approach is inherently goal-oriented, driven by specific tasks identified from given inputs. Among the prevalent supervised learning tasks, classification and regression stand out prominently. Classification involves categorizing inputs into predefined classes or categories, while regression aims to predict continuous numerical values based on input features.

Applications of supervised learning span diverse domains, including natural language processing, computer vision, finance, and healthcare. For instance, in text classification, a model may learn to determine sentiment in tweets or product reviews, categorizing them as positive,negative, or neutral based on labeled examples.

2.1.2. Unsupervised Machine Learning

Unsupervised learning involves the analysis of unlabeled datasets without requiring human intervention, making it a purely data-driven process. This approach is extensively utilized for extracting generative features, uncovering meaningful trends and structures within data, identifying groupings in results, and for exploratory purposes.

The primary unsupervised learning tasks encompass a variety of objectives, including clustering, density estimation, feature learning, dimensionality reduction, finding association rules, and anomaly detection. These tasks serve various purposes such as organizing data into natural groupings, estimating probability distributions, uncovering latent representations, reducing the complexity of high-dimensional data, discovering patterns of co-occurrence, and identifying outliers or anomalies within the dataset.

2.1.3. Semi-Supervised Learning

Semi-supervised learning serves as a hybridization of supervised and unsupervised methods, operating on both labeled and unlabeled data. Positioned between learning "without supervision" and learning "with supervision," this approach is particularly beneficial in scenarios where labeled data is scarce while unlabeled data is abundant.

In many real-world contexts, obtaining labeled data can be challenging, whereas unlabeled data is more readily available. Semi-supervised learning leverages both types of data to improve predictive outcomes beyond what can be achieved using labeled data alone.

Semi-supervised learning finds application in various domains, including machine translation, fraud detection, labeling data, and text classification. By harnessing the collective information present in both labeled and unlabeled datasets, semi-supervised learning techniques can enhance model performance and provide more robust predictions in scenarios with limited labeled data.

2.1.4. Reinforcement Machine Learning

Reinforcement learning is a machine learning paradigm that empowers software agents and machines to autonomously determine optimal behavior within a specific context or environment, thus enhancing efficiency. This approach, driven by the environment, revolves around the concept of reward or penalty, where the primary objective is to leverage insights gained from environmental interactions to make decisions that maximize rewards or minimize risks.

Reinforcement learning serves as a potent technique for training AI models capable of enhancing automation and optimizing operational efficiency in complex systems. Applications span a wide array of domains including robotics, autonomous driving, manufacturing, and supply chain logistics. However, it is not typically preferred for addressing basic or straightforward problems, as its strength lies in tackling more intricate tasks that involve dynamic decision-making and adaptability to varying environmental conditions.

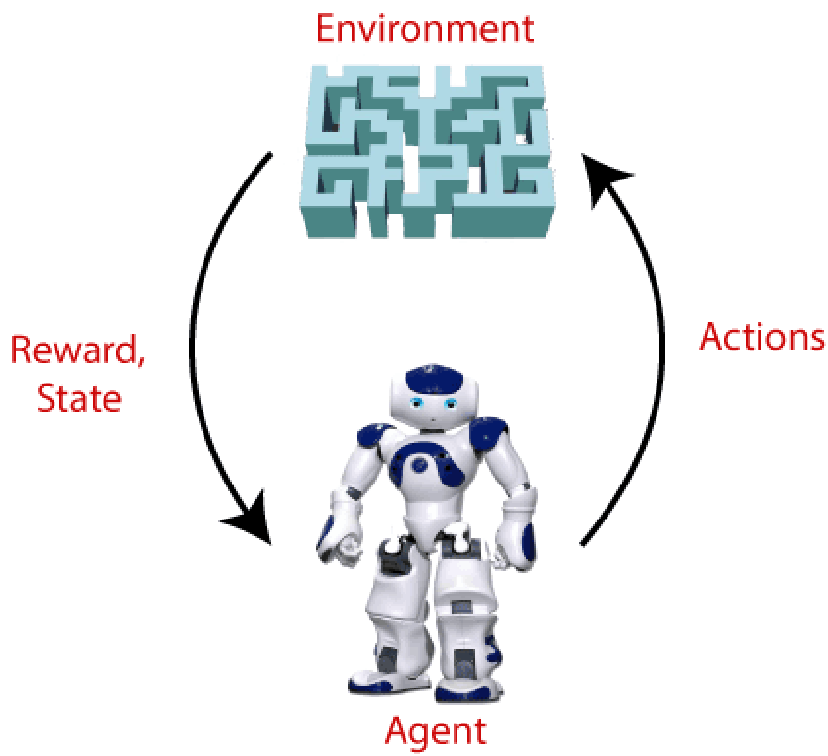


Figure 2:Reinforcement Machine Learning.

[Web 1]

.

2.1.5. Transfer Learning (TL)

Transfer learning is a machine learning approach that leverages knowledge acquired from training on a particular set of problems (referred to as the source domain) with a large dataset. This acquired knowledge can then be applied to tackle other similar problems (the target domain) that may have a smaller dataset. In essence, transfer learning enables the utilization of pre-existing knowledge, such as model weights and features, from a pre-trained model to train a new model and address challenges in a novel task that may have limited data availability. When applied in conjunction with deep learning models, transfer learning offers advantages such as accelerated training, enhanced accuracy, and reduced dependence on large amounts of training data. The fundamental idea behind transfer learning is to repurpose a trained network that was initially developed for different tasks and source data, and subsequently adapt it to suit the requirements of the target task.

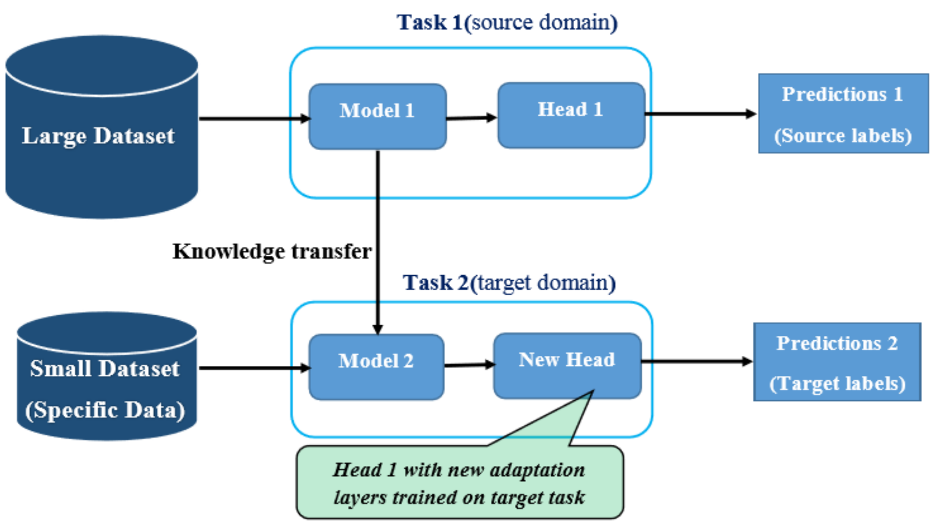


Figure 3:The concept of the transfer learning

2.2.Machine Learning Algorithms

2.2.1.K-Nearest Neighbors

K-means, an unsupervised learning algorithm, offers a straightforward solution to the clustering problem by categorizing data into a predetermined number of clusters. Its process involves defining k centers, each representing a cluster. To ensure effective clustering, these centers should be strategically positioned, ideally far apart from one another. The algorithm iteratively assigns data points to the nearest centroid and updates the centroids based on the mean of the assigned points. However, K-means' performance can vary depending on the initial placement of centroids, making careful initialization crucial. Techniques like K-means++ address this issue by improving the initial centroid selection process. Despite its simplicity, K-means may not always yield the optimal solution, especially for datasets with complex structures. Therefore, it's important to run the algorithm multiple times with different initializations and consider alternative clustering methods for more challenging datasets.[1]

2.2.2. Decision tree

A decision tree is a graphical representation of decision-making processes, showcasing choices and their outcomes in a tree-like structure. The nodes within the tree correspond to events or decisions, while the edges represent the decision rules or conditions. [1]

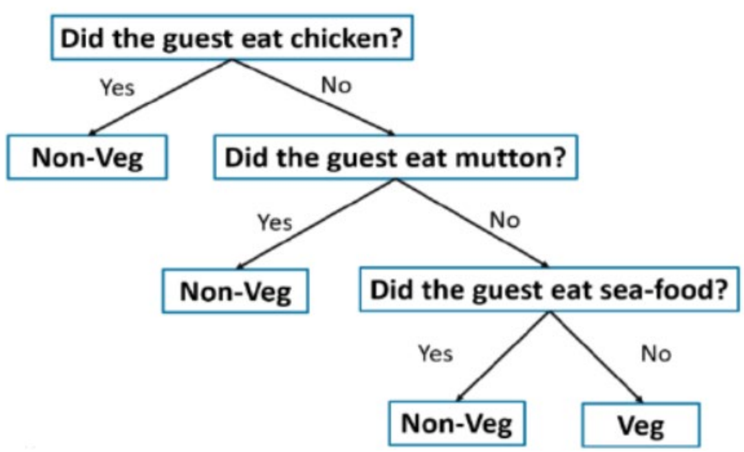


Figure 4:Decision tree

2.2.3. Naive Bayes

A Naive Bayes classifier is a classification technique grounded in Bayes' Theorem, operating under the assumption of independence among predictors. Essentially, it assumes that the presence of one particular feature within a class is entirely unrelated to the presence of any other features. Naive Bayes classifiers are particularly prevalent in the realm of text classification.[1]

2.2.4.Support vector machine (SVM)

Support Vector Machines (SVMs) are supervised learning models rooted in statistical learning theory, utilized for tasks like pattern recognition and regression. While statistical learning theory can pinpoint the essential factors for effectively learning simple algorithms, real-world applications often demand more intricate models like neural networks. These complex models pose challenges for theoretical analysis.

SVMs bridge the gap between theory and practice by constructing models that strike a balance: they're sufficiently complex to handle various tasks, including those handled by neural networks, yet they retain mathematical analyzability. This is achieved by viewing SVMs as linear algorithms operating in a high-dimensional space. By mapping data into this space, SVMs can delineate complex decision boundaries that effectively separate different classes while maintaining tractability for mathematical analysis. Thus, SVMs offer a blend of complexity and theoretical understanding, making them valuable tools in many applications. [3]

2.2.5. Linear regression

Linear regression is a common statistical method used to analyze the relationship between one continuous variable, known as the dependent, outcome, or criterion variable, and a set of continuous predictor variables, also referred to as independent variables, covariates, or predictors. The aim of linear regression is to model and understand how changes in the predictor variables are associated with changes in the dependent variable. This method provides insights into the strength, direction, and significance of these relationships, enabling predictions and inference about the dependent variable based on the values of the predictors.[4]

2.2.6.Hidden Markov Model

Hidden Markov Models (HMMs) serve as a fundamental framework for constructing probabilistic models aimed at solving linear sequence 'labeling' problems. They offer a conceptual toolkit that simplifies the development of intricate models through intuitive visual representations. HMMs are pivotal in numerous programs, spanning gene finding, profile searches, multiple sequence alignment, and regulatory site identification. In computational sequence analysis, HMMs function as the building blocks, akin to Legos, enabling the creation of diverse and sophisticated models.[5]

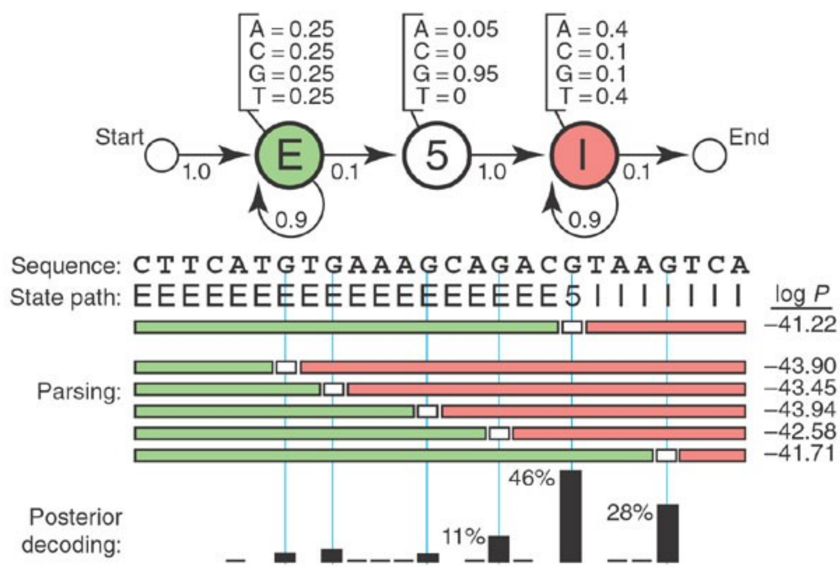


Figure 5: Hidden Markov Model

3.Deep learning

Deep learning refers to a set of learning methods aimed at modeling data using intricate architectures that combine various non-linear transformations. At the core of deep learning are neural networks, which are organized into deep neural networks through layered arrangements.

These techniques have facilitated remarkable advancements in fields such as sound and image processing, encompassing tasks like facial recognition, speech recognition, computer vision, automated language processing, and text classification (e.g., spam recognition). The potential applications of deep learning are vast and diverse.

A striking demonstration of deep learning's capabilities is evident in the AlphaGo program. This program mastered the game of Go using deep learning methods and achieved a historic victory over the world champion in 2016, showcasing the power and potential of deep learning algorithms in solving complex problems.[6]

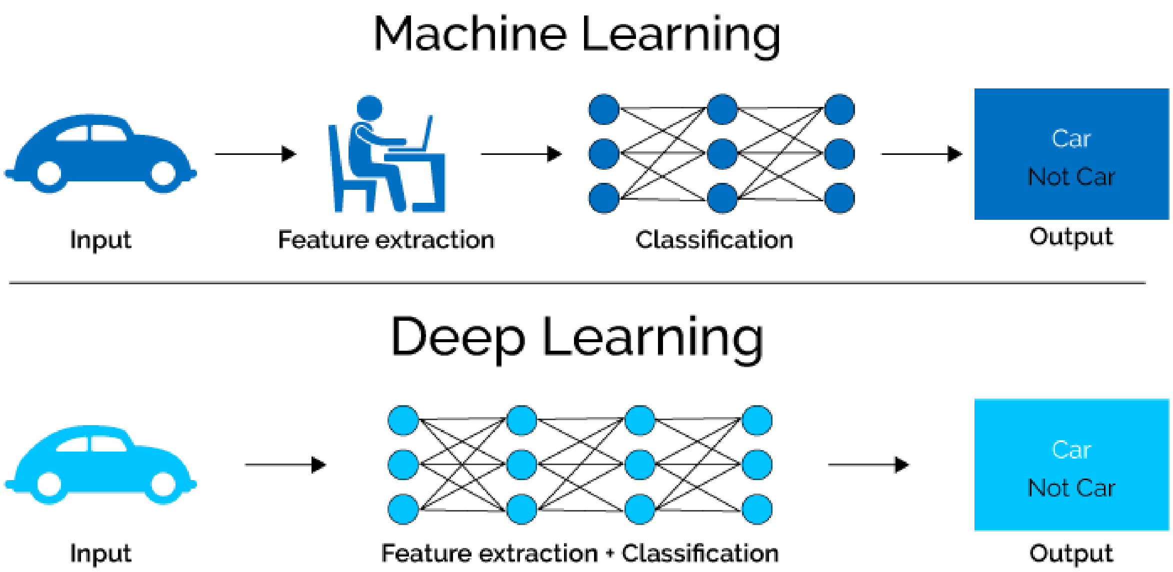


Figure 6: Machine Learning vs Deep Learning[Web 2]

4.Neural networks (NNs)

Practically all DL algorithms are neural networks also called ANNs [30]. ANNs are information processing models that simulate the functioning of a biological nervous system. This is similar to how the brain handles information at the functioning level. All neural networks are made up of interconnected neurons that are organized in layers.

The detail of the neuron and the description of its operation will be explained in the next part.[7]

4.1. Biological Neuron

Neurons, the basic units of the nervous system, transmit and process information through three main components: dendrites, the cell body, and axons. Dendrites receive signals from other neurons and transmit them as electrical impulses to the cell body, which integrates these signals. If the combined charge surpasses a threshold, the axon is activated, transmitting electrical impulses to the dendrites of other cells. This communication between neurons is dynamic, allowing for learning and adaptation. Learning occurs at synapses, the junctions between axons and dendrites, where neurotransmitters are released in response to electrical impulses. These neurotransmitters trigger electrical activity in the receiving cell's dendrite, leading to changes that contribute to learning and development.[8]

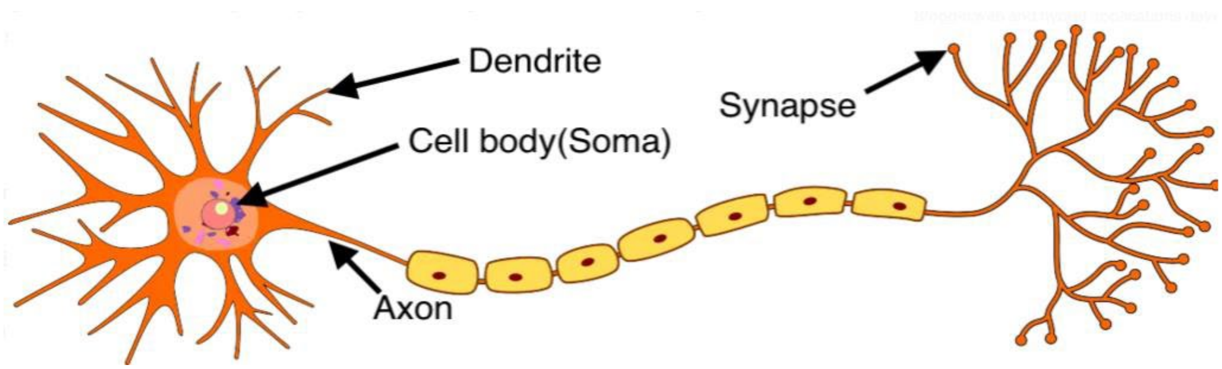


Figure 7:Biological neuron

[Web 3]

4.2 Artificial neural networks

a. Artificial Neural Networks (ANNs)

ANNs are computational models inspired by the human brain, designed to process and analyze complex data. They consist of interconnected artificial neurons that work together to make decisions and recognize patterns.

b. The pivotal role of ANNs

ANNs are the fundamental building blocks of deep learning. They enable computers to learn from data, recognize patterns, and make decisions, much like the human brain. Their significance lies in their capacity to handle complex tasks and vast amounts of data, making them a cornerstone of modern AI. [9]

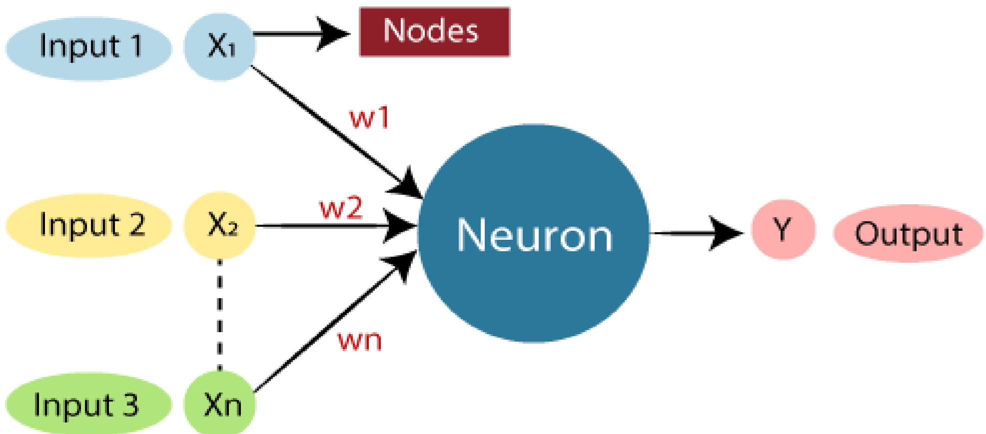


Figure 8:Artificial Neuron

[Web 4]

- **Neuron Structure**:
  - A **neuron** processes information in a neural network.
  - It has multiple inputs: $\boldsymbol{x_1, x_2, \ldots, x_n}$.
  - Each input is associated with a weight: $\boldsymbol{w_1, w_2, \ldots, w_n}$.
  - Additionally, there's a bias input: $\boldsymbol{x_0 = 1}$ with its weight $\boldsymbol{w_0}$.
  - The output $\boldsymbol{y}$ is calculated by applying an **activation function** (e.g., sigmoid) to the weighted sum $\boldsymbol{z}$.
- **Mathematical Computation**:
  - The weighted sum $\boldsymbol{z}$ is computed as:
  - $z = w_0x_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$

- The activation function (e.g., sigmoid) is applied to z:

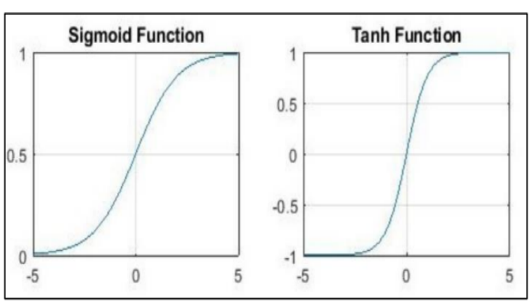
  y = sigmoid(z)=1/1+ e^{-z}

- **Activation Functions** [10]**:**

a. **Sigmoid Activation Function:** This function produces outputs in the range of 0 to 1 for any given input. It's often used in binary classification problems where the output needs to be interpreted as a probability. However, it suffers from the vanishing gradient problem, particularly when inputs are far from zero, which can slow down learning in deep networks.

b. **Tanh Activation Function:** Similar to the sigmoid function, but with outputs ranging from -1 to 1. Tanh has the advantage of outputting values centered around zero, which can help mitigate the vanishing gradient problem. Additionally, it has a steeper gradient compared to the sigmoid function, which can aid in learning faster in certain scenarios.

c. **Softmax Activation Function:** This function is primarily used in the output layer of classification neural network models. It produces a probability distribution over the predicted classes, ensuring that the sum of all probabilities equals one. Softmax is beneficial for multi-class classification tasks, where the model needs to predict the probability of each class.



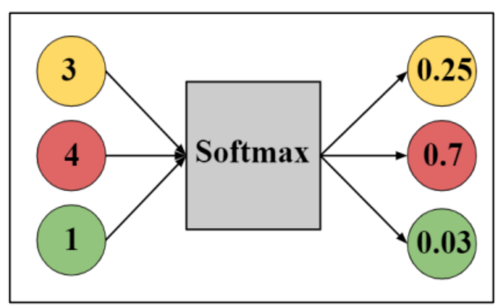


Figure 9:Softmax activation function

Figure 10:Sigmoid and Tanh activation function

5. Deep neural network

Deep neural networks are artificial neural networks that have several hidden layers, which allows them to capture complex relationships between inputs and outputs. Using deep neural networks, models can be trained to perform tasks such as image classification, speech recognition,machine translation, etc.

5.1.Convolutional neural networks (CNN)

A Convolutional Neural Network (CNN), also known as ConvNet, is a type of Artificial Neural Network (ANN) characterized by its deep feed-forward architecture and remarkable generalization capabilities compared to networks with fully connected (FC) layers. CNNs are adept at learning highly abstracted features of objects, particularly in spatial data, enabling more efficient identification.

A deep CNN model comprises a finite series of processing layers capable of learning various features from input data, such as images, across multiple levels of abstraction. Initial layers focus on extracting high-level features with lower abstraction, while deeper layers specialize in learning and extracting low-level features with higher abstraction.[11]

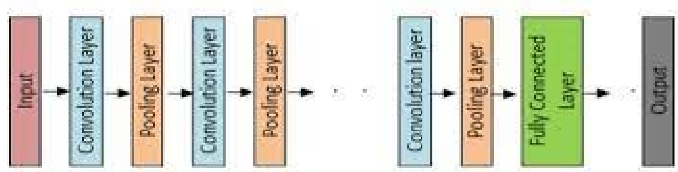


Figure 11: Conceptual model of CNN

5.1.1.CNN Architecture

A standard Convolutional Neural Network (CNN) architecture typically comprises alternating layers of convolution and pooling, with one or more fully connected layers at the end. The CNN is divided into two significant sections, each serving a distinct purpose. The initial part of the network focuses on feature extraction and incorporates convolutional and pooling layers. Meanwhile, the subsequent part handles classification tasks using features extracted by the fully connected layers. The arrangement of various layers within a CNN architecture is pivotal for designing novel architectures and enhancing the model's performance. This section offers a concise explanation of the role played by each layer in a CNN architecture.

A) Convolution Layers

The convolution layer is a fundamental component of CNN architecture, positioned immediately after the input layer. It comprises a collection of convolution kernels, or neurons, each associated with a small region of the input known as a receptive field. These receptive fields operate by partitioning the input image into smaller segments and convolving them with a specific set of weights.[12]

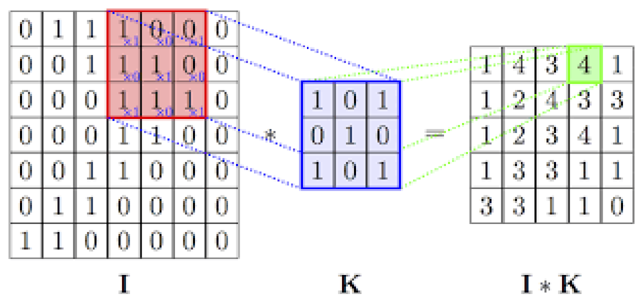


Figure 12: Convolution layer.

[Web 5]

B) Pooling Layer

Following the convolution operation in CNN architecture, the subsequent layer is pooling. This layer conducts downsampling, aiming to reduce the dimensionality of the information gathered from the convolution layer while retaining essential features. Simultaneously, certain notations are taken into account as input to the pooling layer.[13]

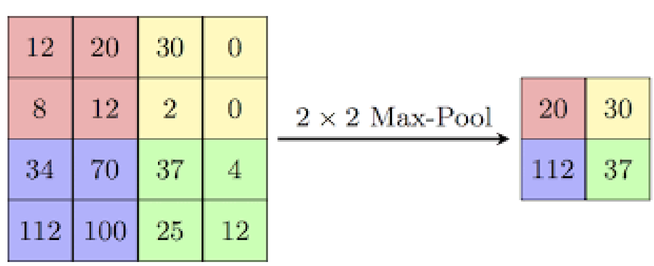


Figure 13: Pooling layer.

[Web 6]

C) Fully-Connected (FC) layer

The fully-connected layer in a neural network establishes connections between every neuron in the previous layer with all neurons in the current layer. It serves the purpose of classification or final prediction, enabling the network to learn intricate relationships between the extracted features and the output labels.

5.2.RECURRENT NEURAL NETWORKS (RNN) Recurrent Neural Networks (RNNs)

possess a memory component that allows them to effectively process sequential data. This feature makes them well-suited for tasks where output depends on a series of inputs, such as natural language processing, text prediction, and speech recognition. The term "recurrent" signifies that RNNs repeatedly perform the same computations for each element in the input sequence, with each output contributing to the overall prediction. Compared to traditional feedforward neural networks, RNNs excel in capturing temporal dependencies and context, leading to higher prediction accuracy in scenarios where sequential data or contextual understanding is crucial.[10]

A typical RNN architecture is shown in Fig 14:

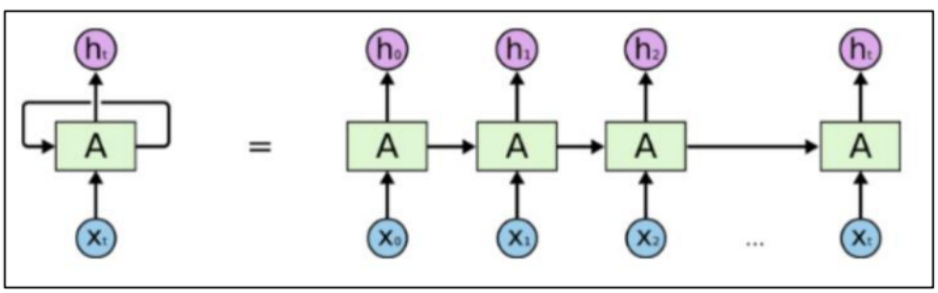


Figure 14: Unrolled RNN

6. The transformer

The Transformer, a sequence-to-sequence (Seq2seq) model, was introduced in 2017 by Vaswani et al. in the seminal paper "Attention is All You Need". Departing from earlier models that employed a combination of encoders and decoders using RNN, the Transformer relies solely on the attention mechanism to capture word interdependencies. Its architecture comprises an encoder-decoder setup, with the left component representing the encoder and the right component representing the decoder. Each component consists of six identical layers, each layer composed of sub-layers.

The encoder comprises two sub-layers: self-attention (Multi-Head Attention), a core element maintaining word interdependencies in input sequence representation, and a Feed-Forward Neural Network applied to each attention vector to prepare it for subsequent encoding. Residual connections between sub-layers, followed by layer normalization, mitigate the vanishing gradient problem and stabilize value changes, facilitating faster training and better generalization.

The decoder mirrors the encoder's architecture with an added layer of Multi-Head Attention (Encoder-Decoder Attention) between the self-attention layer and the Feed-Forward Neural Network, along with modifications to self-attention involving the addition of a mask.

The first encoder's input comprises the word-embedding sum of the input sequence and positional encoding vector. These vectors undergo processing in the self-attention layer and subsequent forward propagation neural networks. The output of each encoder block serves as input to the next encoder, culminating in the last encoder's utilization by each layer's "Encoder-Decoder Attention" in the decoder.

Decoder operation resembles that of the encoder, with the addition of attention to the representation passed by the last encoder, enabling focus on relevant information from the input sequence. The output of the last decoder passes through a linear layer, converting it into a vector with the model's vocabulary size. This vector undergoes transformation by the Softmax layer to produce probabilities, with the word having the highest probability appended to the output sequence.[14]

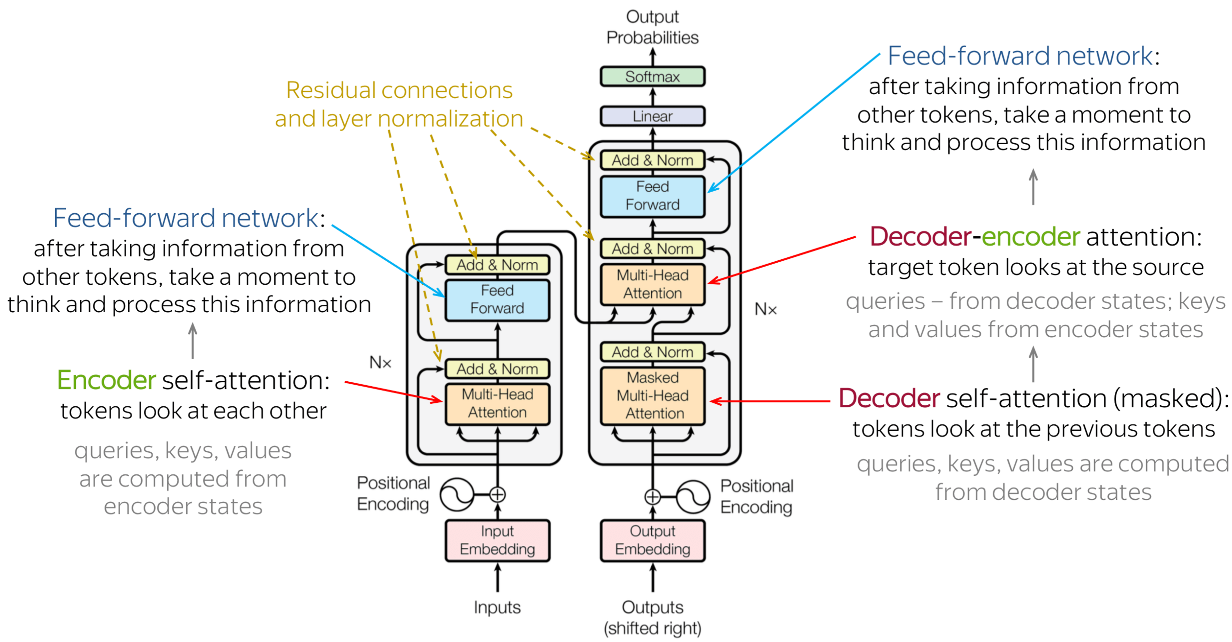


Figure 15: Architecture of Transformer

[14]

6.1. Transformer detail Functioning [Web 7]

➢ **Input Embedding**

The input embedding sub-layer plays a crucial role in the Transformer model by transforming input tokens into vectors. Each token, representing a word or a part of the input sentence, is mapped to a vector in a high-dimensional space, typically with each dimension set to 512, as per the original Transformer model [24]. This mapping is facilitated by learned embeddings, where the key concept is to ensure that similar words are assigned similar representation vectors. This process enables the model to capture semantic similarities among words, facilitating effective processing of the input sequence.

➢ **Positional Encoding (PE)**

Positional Encoding (PE) is crucial in representing the position of words within the original text as vectors. There are two types of PE utilized in Transformers: static and learned. The Transformer model commonly employs static PE, which involves a series of sine and cosine waves with varying wavelengths.

The calculation of static PE is depicted by the following equations:

$$PE_{(pos,2i)}=\sin\,(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)}=\cos\,(pos/10000^{2i/d_{model}})$$

Where, 'pos' denotes the position in the sequence, and 'i' represents the feature dimension. In the equations, the Transformer incorporates sine and cosine functions with varying frequencies to encode positional information into the word vectors.

The Transformer model combines the word vector embeddings with positional encodings, resulting in a combined representation that is subsequently fed into multiple encoders followed by decoders. This integration enables the model to capture both the semantic content of words and their positional information, facilitating effective sequence processing.

- **Multi-Head Attention:**

To comprehend the functioning of a Multi-Head Attention layer, it's essential to grasp the operation of the Scaled Dot-Product Attention layer. The Scaled Dot-Product Attention head has a straightforward structure (refer to Figure 25): it applies a linear transformation to its query, key, and value vectors. Firstly, it calculates the attention score by taking the dot product between the query (Q) and the transpose of the key (K) vectors, then divides the result by the square root of the dimension of the key vector (dk), acting as a scaling factor. Secondly, it utilizes a softmax function to normalize the resulting values (referred to as attention weights), which are then used to weight the values and sum them up, yielding the new input embedding. The key, query, and value vectors are generated by multiplying the word embeddings input with three weight matrices: Wk, Wq, and Wv.

The equation for the output attention matrix is as follows: $Attention(Q,K,V)=softmax(\frac{QK^T}{\sqrt{d_k}})V$

Where,Q, K, and V represent the query, key, and value matrices respectively, and dk denotes the dimensionality of the key vectors. The softmax function normalizes the attention scores across the key vectors, producing attention weights that indicate the importance of each value vector. These weights are then used to linearly combine the value vectors, resulting in the final output ofthe attention mechanism.

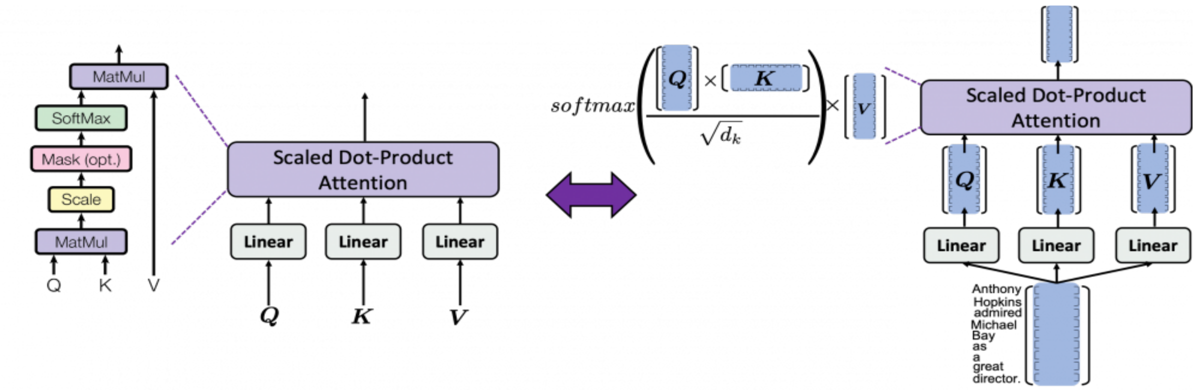


Figure 16:Scaled Dot-Product Attention

[Web 8]

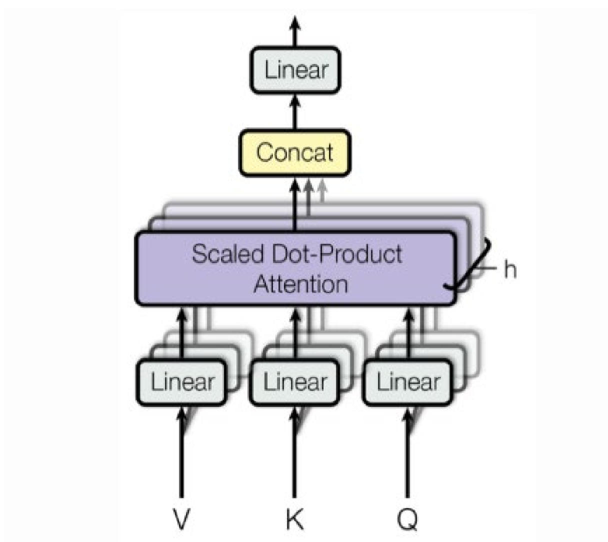


Figure 17: Multihead-Attention [Web 9]

The equation of the output of multi-head attention is:

$MultiHead(Q,K,V)=Concat(head0,\ldots,headh-1)WO$

Where:

$headi=Attention(QWiQ,KWiK,VWiV)$

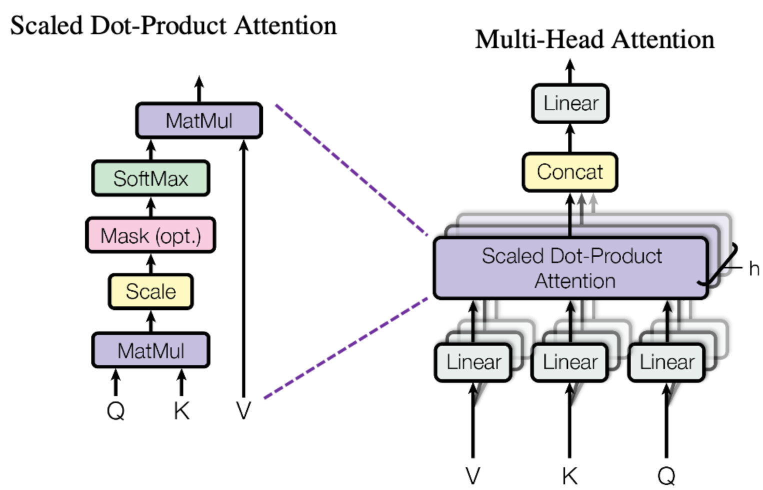


Figure 18: Scaled Dot-Product Attention And Multihead-Attention[Web 10]

- **Feed-forward network (FFN):**

It is a fully connected network applied to each position independently and identically. This sub-layer is a two-layer feed-forward network with a ReLU activation function. The first layer is four times the size of the model (dff equal to 2048). This seems to give the transformer enough representational capacity. The second layer will project the output of this first layer into the original size ($dmodel$ equal to 512). Given a sequence of vectors $h1,\ldots, hn$ the computation of a position-wise FFN sub-layer on any hi is defined as shown in the equation:

$FNN(hi)=ReLU(hi\ W1+ b1)\ W2+ b2$

- **Masked Multi-Head Attention:**

multi-head where some values are masked (i.e, probabilities of masked values are nullified to prevent them from being selected).Whene decoding, an output value should only depend on previous outputs (not next outputs). Hence we mask future outputs. The equation of the output of masked attention is :

$maskedAttention(Q,K,V)=softmax(QKT+M\sqrt{}dk)V$

Where: M is a mask matrix of 0’s and -∞’s

6.2 Different Transformer models:

In recent times, large-scale pre-trained models (PTMs) like BERT and GPT have emerged as groundbreaking achievements in the realm of artificial intelligence (AI). Their success stems from sophisticated pre-training objectives and extensive model parameters, enabling them to effectively capture knowledge from vast amounts of labeled and unlabeled data. By encapsulating knowledge within these extensive parameters and fine-tuning them for specific tasks, these PTMs implicitly encode a wealth of information that can significantly benefit a range of downstream tasks. This capability has been extensively validated through experimental evidence and empirical analyses. Consequently, it has become the consensus within the AI community to leverage PTMs as the backbone for various downstream tasks, rather than starting from scratch and training models anew. [15]

	BERT	GPT	BART
Model Type	Encoder Only	Decoder Only	Encoder-Decoder
Direction	Bidirectional	Unidirectional (left-to-right)	Bidirectional
Pre-training Objective	Masked language modeling (MLM)	Autoregressive (casual) language modeling	Span Corruption (Masking entire spans of words)
Fine-tuning	Task-specific layer added on top of the pre-trained BERT model	Providing task-specific prompts using few-shot or one-shot adaptation and adapting the model's parameters	Versatile and can be used for various NLP tasks
Use Case	Sentiment Analysis Named entity Recognition Word Classification	Text generation Text completion creative writing	Translation Text Summarisation Question & Answer
Original Organisations	Google AI	OpenAI	Facebook AI

Figure 19: Comparison between BERT, GPT and BART

[Web 11]

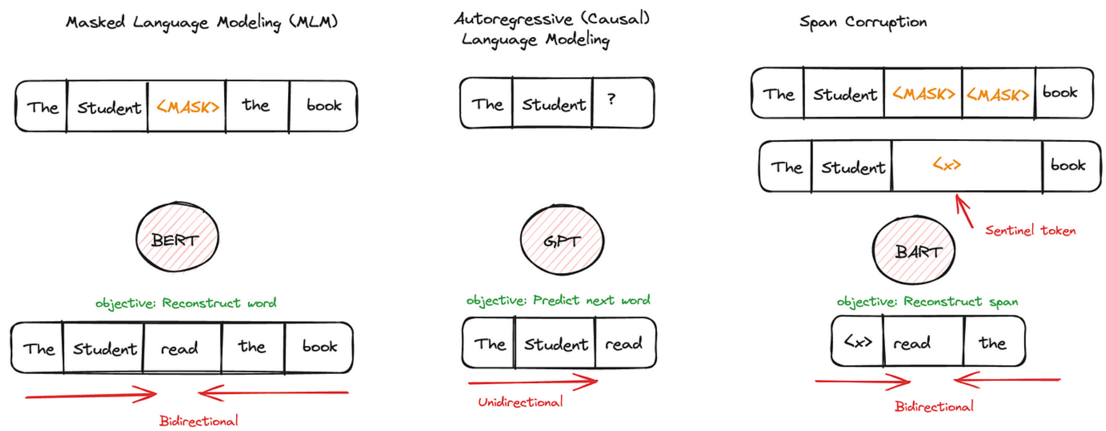


Figure 20: Visual comparison between pre-training objectives

[Web 12]

Deep Learning is a very rich field in which certain basic concepts are essential to its understanding. In this chapter, we have tried to cover some concepts and algorithms in ML and the majority of existing types of learning. Then we passed to DL and it popular architecture. We have focused in explanation mainly on models dedicated to the processing of sequential data like RNNs, and the Transformer.

In the coming chapter, we will present The large language model landscape and AI content detection.

CHAPTER 2:

Automatic AI generated text detection

Recent progress in large language models (LLMs) has significantly improved the caliber of synthetic textdata. Through their emulation of human writing styles, LLMs produce text that closely mirrors natural human language, sparking considerable ethical, moral, legal, social, and economic debates across diverse sectors. [Web 13]

1. The large language model landscape [Web 1 4]

Over the past two years, the proliferation of both commercial and open large language model (LLM) providers has surged, offering a plethora of options catering to various language-related tasks. Despite the predominant method of interfacing with LLMs being through APIs and basic Playgrounds, I anticipate a burgeoning market for tooling aimed at expediting their widespread adoption.

Below is a graphic depicting the current Large Language Model (**LLM**) landscape in terms of functionality, offerings and the tooling ecosystem.

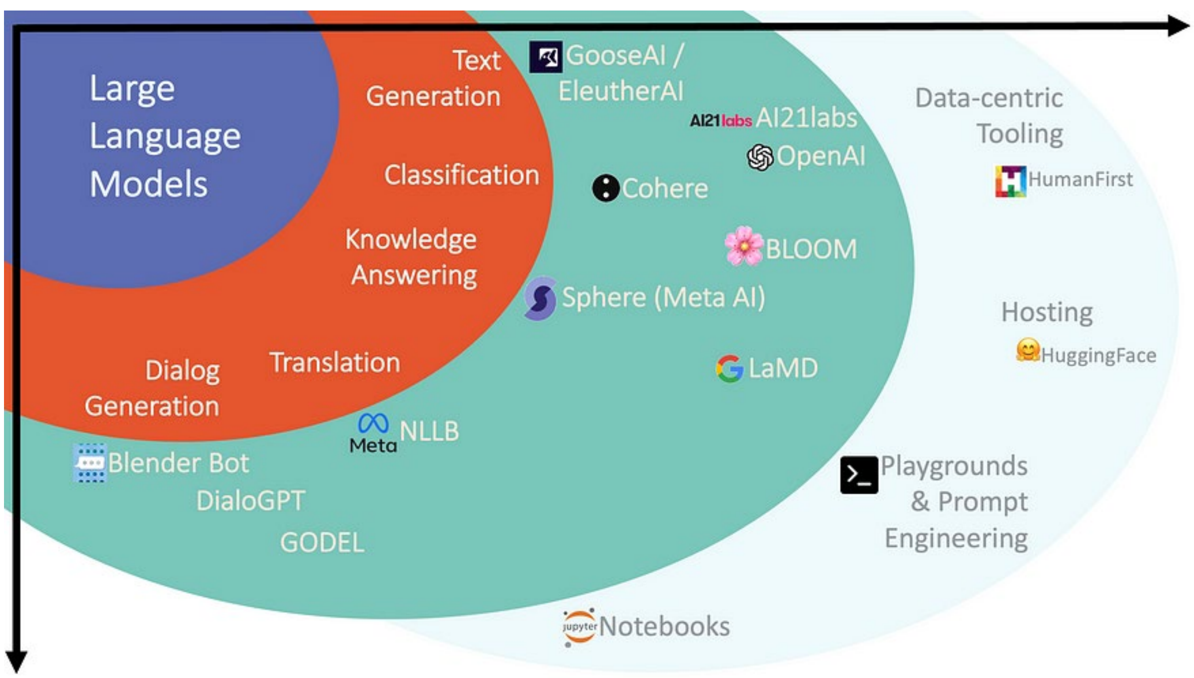


Figure 21:the current Large Language Model (LLM) landscape in terms of functionality

2. Definition of large language models [Web 15]

Traditional language models in natural language processing (NLP) are statistical models capable of assigning probabilities to words or tokens in a sequence within a target language. For instance, if trained on English, such a model could predict the next word in a sentence based on the preceding words. For example, given the input "I am going to the," the model might assign a high probability to the word "store" as the next word. These models can be trained using various techniques, including RNNs, LSTMs, transformers, and others.

Large Language Models (LLMs) represent a specialized subset of language models engineered to generate extensive amounts of human-like text with minimal training data, a concept known as few-shotlearning.

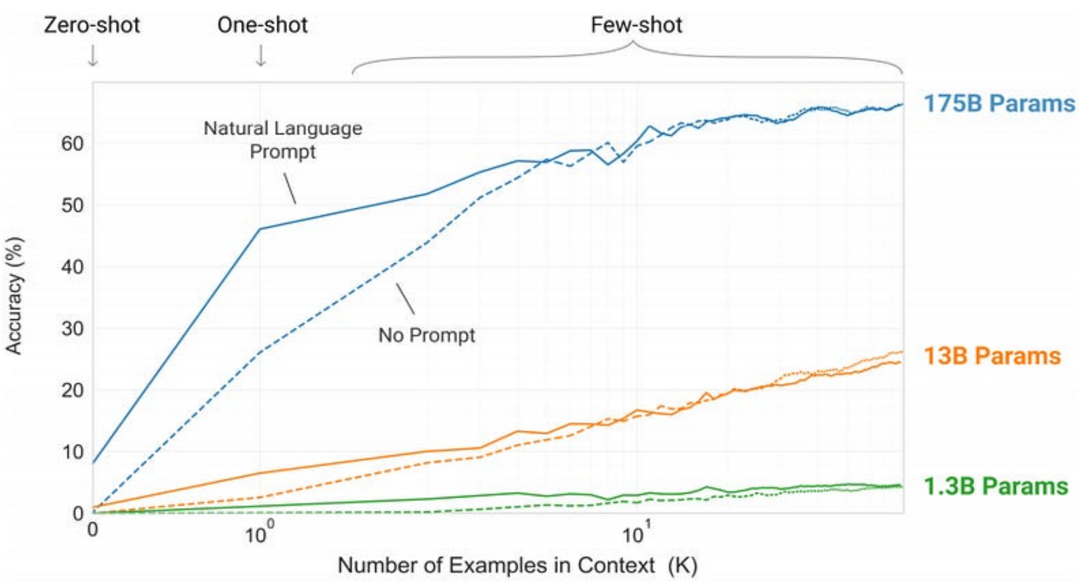


Figure 22:: subset of language models engineered to generate extensive amounts of human-like text.

These models exhibit the remarkable ability to generate text that is not only contextually relevant and coherent but also virtually indistinguishable from text produced by humans. The exponential advancement in performance and the emergence of new capabilities in Large Language Models (LLMs), attributed to their training on vast datasets, has popularized the notion of "scale is all you need" within certain circles of the machine learning community. This phrase, a playful nod to Google's seminal "Attention is all you need" paper, encapsulates an important observation. Essentially, it suggests that as we continue to scale our models with larger datasets and networks, we can anticipate similar performance enhancements seen thus far. However, whether these scaling trends will eventually reach a plateau remains an ongoing research question.

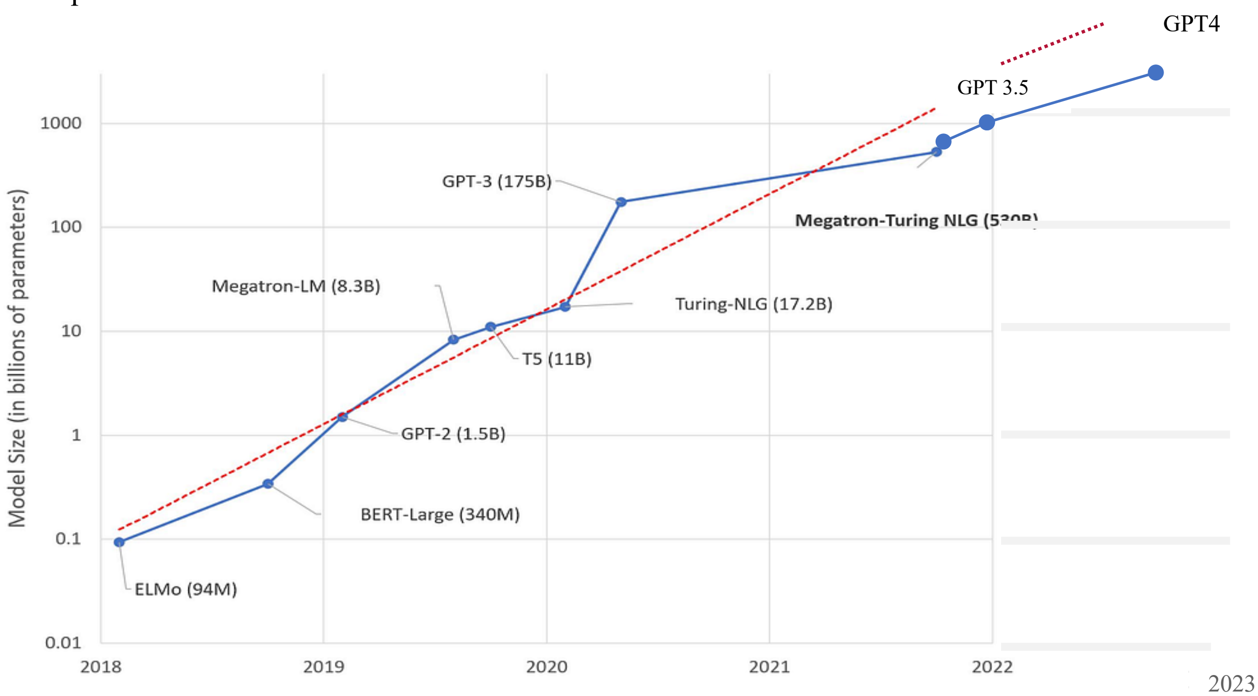


Figure 23:AI tools

3. Large language model (LLM) Architecture [Web 16]

The architecture of a Large Language Model (LLM) can differ based on its particular implementation. Nonetheless, many LLMs leverage a transformer-based architecture, which originated from the groundbreaking "Attention is All You Need" paper in 2017. Transformers have emerged as the predominant architecture for large-scale language models due to their adeptness at managing long-range dependencies and capturing contextual information with greater efficiency. Additionally, transformers offer inherent scalability, as the processing of distinct tokens can be executed in parallel. This feature has proven instrumental for organizations committed to constructing these models and willing to allocate resources to extensive computational power.

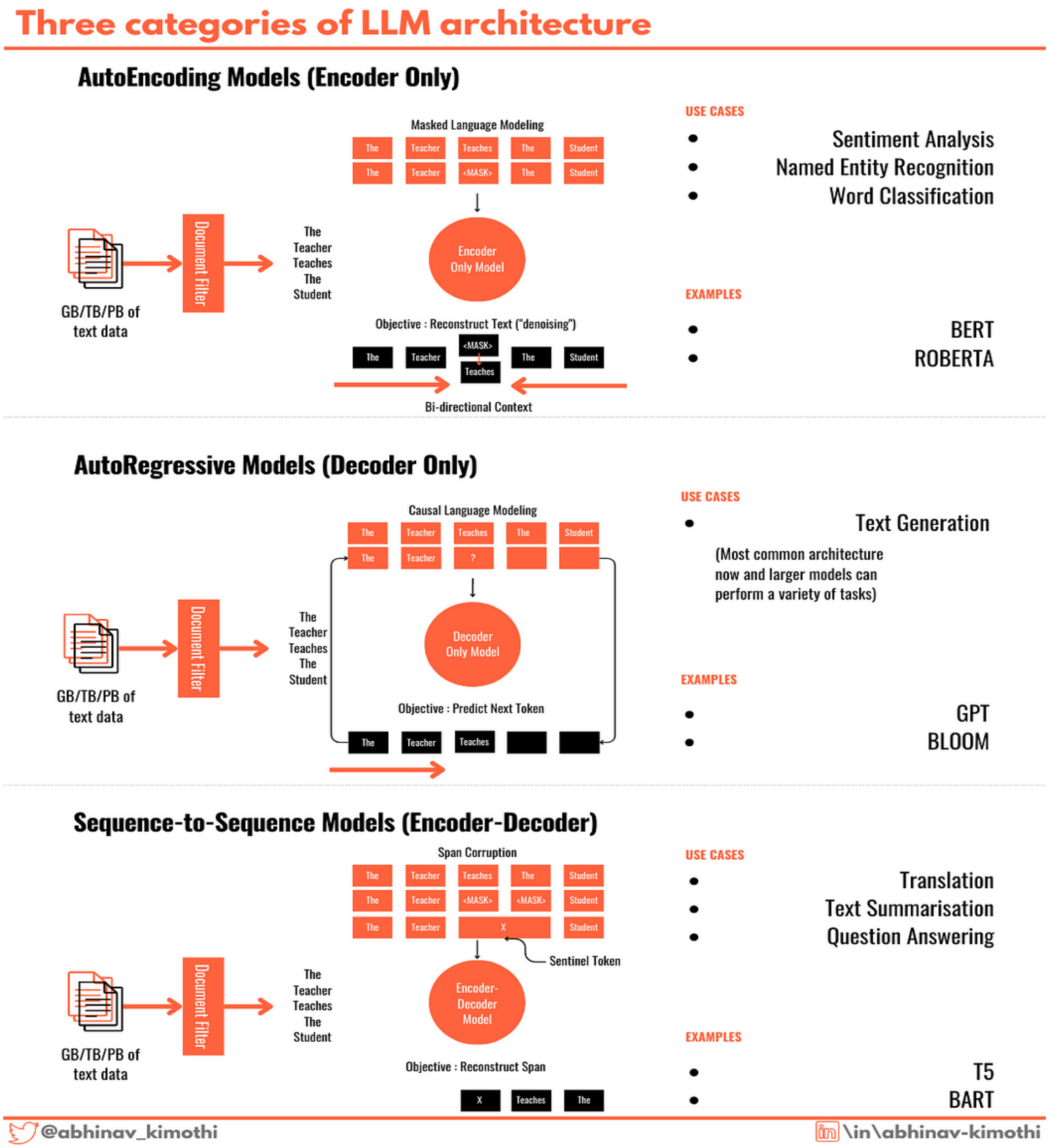


Figure 24:Three categories of LLM architecture

[Web 17]

3.1 Applications of large language model LLM [Web 18]

Although the use of LLMs in production is a relatively new concept, it is becoming clear that LLMs have a wide range of potential applications in NLP and related fields. Some of the most common applications include:

1. Language Generation: LLMs can generate contextually relevant and coherent text in response to a prompt or a question. This has led to the development of language generation applications, such as chatbots and virtual assistants, that can interact with users in natural language.

2. Text Classification: LLMs can classify text into different categories, such as sentiment analysis or topic modeling. This can be useful in applications such as social media monitoring or content moderation.

3. Machine Translation: LLMs can translate text from one language to another, making it easier for people to communicate across different languages.

4. Text Summarization: LLMs are excellent at summarizing large texts into smaller chunks, which can help users in a variety of different communications contexts.

see figure25:

Figure 25: LLM Applications

3.4. Limitations and potential drawbacks [Web 20]

Large language models, despite their remarkable performance across various language tasks, are not without limitations and potential drawbacks:

1. Data Bias: Training on extensive text data may inadvertently incorporate biases reflective of societal norms and values, perpetuating stereotypes and discrimination. Addressing this challenge necessitates meticulous curation of training data and the development of techniques to detect and mitigate biases within language models.

2. Environmental Impact: The computational demands of training large language models result in significant energy consumption, contributing to carbon emissions and environmental concerns. Exploring energy-efficient and sustainable training methods is essential to mitigate these environmental impacts.

3. Limited Interpretability: While proficient in generating coherent text, large language models often lack transparency in their decision-making processes, posing challenges for understanding how they arrive at specific outputs. This lack of interpretability hinders trust, accountability, and regulatory compliance, particularly in sensitive domains like finance and healthcare.

4. Limited Generalization: Large language models may struggle to generalize to new or unseen data, impacting their performance in dynamic environments with evolving data distributions. Overcoming this limitation is crucial for ensuring robust adaptability in real-world applications.

5. Ethical Concerns: Large language models can be exploited for malicious purposes, such as generating fake news or facilitating cyber attacks, raising ethical concerns around privacy and data security. Addressing these concerns requires comprehensive strategies to promote responsible use and safeguard against misuse of language models.

2. Text generation

Text Generation is an area within Natural Language Processing (NLP) that merges computational linguistics and artificial intelligence to create fresh text. It entails producing synthetic text that is grammatically accurate and semantically meaningful.

Text Generation can be described as the meta capability of Large Language Models (LLMs), enabling the creation of text based on a brief description, with or without example. Generation is a fundamental function present in virtually all Large Language Models (LLMs). It can be extensively utilized through few-shot learning data. The way in which the data is presented, through prompt engineering, determines how the few-shot learning data will be utilized. [Web 21]

2.1. AI content detection

AI content detection involves the identification and categorization of content generated or manipulated by artificial intelligence (AI) systems. As natural language processing (NLP) and generative AI technologies like text generators and deep learning models advance, the necessity to differentiate between human-created and AI-generated content becomes increasingly apparent.

AI content detection encompasses the development of algorithms, techniques, and tools aimed at analyzing and distinguishing between AI-generated content and genuine human-generated content. This process typically entails scrutinizing diverse linguistic, semantic, and contextual aspects of the text to discern patterns, anomalies, or indicators characteristic of machine-generated content.

The main objective of AI content detection is to bolster trust, transparency, and safety within digital environments by empowering users to recognize AI-generated content and grasp its potential implications. This capability holds particular significance in applications like social media, journalism, content moderation, and cybersecurity, where the authenticity and credibility of information are paramount.

- **Common signs of AI generated** [Web 22]

These common signs of AI-generated content serve as valuable indicators for detecting artificially generated text:

1. **Incorrect and outdated information**

Although AI writing can look well-written, it's always important to check how accurate the actual information is. Since most bots are trained on limited data sets (in time, form, or source), they may not have access to the latest and most complete information.

AI-generated content may contain inaccuracies or outdated information, as the model generating the text may not have access to the latest data or may lack the ability to verify the accuracy of the information. This can manifest as factual errors, inconsistencies, or outdated references within the text.

2. **Lack of depth and personality**

Because AI tools don't really write but generate text based on patterns in their training data, they don't "understand" what they're writing about in the same way humans do. This results in very superficial and shallow responses, a lack of critical thinking, and deep topic analysis.

They also don't have a personality, which is why most AI-generated texts lack a personal touch and can sound robotic and emotionless.

In contrast to an AI tool, a journalist or copywriter can have real conversations with subject matter experts in the field they're writing about. These kinds of conversations lead to deeper understandings, interesting stories, and relatable opinions in a way that is hard to replicate with AI.

AI-generated text often lacks the depth, nuance, and personal touch characteristic of human-generated content. While AI models may be proficient at producing coherent and grammatically correct text, they may struggle to imbue the content with genuine emotion, creativity, or unique perspectives. As a result, AI-generated content may come across as flat, generic, or lacking in authenticity.

3. **Repetitive language**

Another common feature of AI is the use of the same words or phrases over and over again. This may be the result of a specific keyword used in the prompt that an AI then repeats word for word. It can also lack context or just have limited and repetitive training information.

AI models are also designed to be cautious and neutral in general, which is why they may rely on more conservative language patterns, which can sometimes look repetitive.

AI models trained on large datasets may inadvertently generate text that exhibits repetitive language patterns. This repetition can manifest as redundant phrases, similar sentence structures, or the reuse of specific vocabulary across different contexts. While some degree of repetition is common in human language, excessively repetitive language in AI-generated content may indicate a lack of originality or creative expression.

4. **Predictable word choice and generic language patterns**

AI-generated texts frequently display predictable word selection and generic language patterns. This tendency arises from AI writing tools drawing upon extensive datasets to produce text, leading to the frequent usage of commonplace phrases and expressions present in the training data. [Web 23]

5. **Sentence conventional structure**

The conventional structure of sentences in AI-generated text often exhibits uniformity in both length and syntax. This tendency emerges from the algorithms utilized in text generation models, which prioritize coherence and grammatical accuracy. Consequently, AI-generated sentences frequently adhere to predictable patterns, giving them a formulaic appearance and diminishing their variety.

A prevalent technique employed in AI text generation models is Next Sentence Prediction, where the model predicts if a given sentence logically follows another. Through training on extensive text datasets, AI models learn to produce sentences that maintain coherence and transition smoothly from one to the next. However, this process can inadvertently lead to the generation of text with repetitive structures and foreseeable language patterns.

While the adherence to conventional sentence structures enhances the readability and grammatical correctness of AI-generated text, it may also contribute to monotony and a lack of diversity. Human writers often utilize a broader spectrum of sentence structures to convey various tones, emotions, and rhetorical effects. In contrast, AI-generated text may struggle to capture the intricacies of language and the creative flair characteristic of human-authored content. [Web 23]

- **AI generated text detection approaches**

Automatic AI-generated text detection methods can be broadly categorized into two main approaches:

Method	Description	Strengths	Strengths
Feature-based methods [16]	These approaches rely on extracting specific linguistic, syntactic, semantic, or stylistic features from the text to identify patterns or anomalies associated with AI-generated content. Feature-based methods may involve analyzing features such as word frequency, sentence structure, grammar usage, sentiment analysis, and coherence. By examining these features, the detection system can flag texts that exhibit characteristics indicative of AI-generated content. (Strengths and limitations)	Feature-based methods are often more interpretable and can provide insights into the linguistic characteristics of AI-generated text. However, they may struggle to capture the complex and nuanced patterns present in large-scale AI-generated content.	- May struggle to capture complex and nuanced patterns in large-scale AI-generated content.
Neural language models [16]	These methods leverage advanced neural network architectures, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based models like GPT (Generative Pre-trained Transformer), to detect AI-generated text. Neural language models are	Neural language models excel at capturing intricate linguistic structures and	- Require significant computational resources for training and inference - May lack

	trained on large corpora of text data and learn to generate coherent and contextually relevant text. Detection systems based on neural language models often involve fine-tuning pre-trained models or training custom models on labeled datasets to recognize the linguistic patterns specific to AI-generated content. (Strengths and limitations)	can adapt to evolving patterns in AI-generated text. However, they may require significant computational resources for training and inference and may lack interpretability compared to feature-based methods.	interpretability compared to feature-based methods
Combining both approaches [16]	Ultimately, the choice between feature-based and neural language model approaches depends on factors such as the nature of the text data, the desired level of accuracy and interpretability, and the available computational resources. Combining both approaches in a hybrid detection system may offer a comprehensive solution for detecting AI-generated text across a wide range of contexts and applications.	- Offers a comprehensive solution for detecting AI-generated text across various contexts and applications	- Requires careful consideration of factors such as the nature of the text data and available computational resources
AI-Generated Text Detection Domains[17]	AI-generated text detection has become a crucial area of research, particularly in specific domains where the impact of misinformation or fake content can be significant. While achieving perfect detection across all possible domains remains a challenge, recent research has made significant strides in developing detection methods tailored to specific areas of concern.	- Significant strides have been made in developing domain-specific detection methods	- Achieving perfect detection across all possible domains remains a challenge
AI-Generated Text Detection Domains [17]	One prominent focus of research is AI-generated text detection in academic and scientific settings. With the proliferation of AI tools capable of generating academic papers or scientific articles, there is a growing need to distinguish between genuine research and AI-generated content. Detection methods in this domain often leverage linguistic analysis, citation	- Leverages linguistic analysis, citation patterns, and domain-specific knowledge	- May face challenges in distinguishing between genuine research and AI-generated content.

	patterns, and domain-specific knowledge to identify anomalies indicative of AI-generated text.		
AI-Generated Text Detection Domains [17]	Another important domain for AI-generated text detection is the detection of fake news, reviews, and misinformation. In an era where online platforms are inundated with deceptive content created by AI algorithms, detecting and combating misinformation has become a pressing concern. Detection approaches in this domain typically involve analyzing textual features, user behavior, and contextual cues to differentiate between genuine and AI-generated content.	- Involves analyzing textual features, user behavior, and contextual cues	- Requires continuous adaptation to evolving strategies used by AI algorithms to create deceptive content.

Table 1: Automatic AI-generated text detection methods

CHAPTER 3:

Conception and Experimentation

1) Introduction

In this chapter we will speak about the creation of dataset and the different stages to creating the texts generated model based on transfer learning.

2) Creating the Dataset

Dataset Selection and Preparation

- **Data Selection**

Source Material: A collection of previous theses in the field of computer science was chosen as the primary source for the dataset.

Scope: The selection included three scientific theses, ensuring a focused and manageable dataset.

- **Data Processing**

Paragraph Segmentation: Each thesis was divided into individual paragraphs to create smaller, more manageable text units.

Rephrasing: Each paragraph was rephrased three times using ChatGPT to generate AI-generated text variations.

Compilation: All human-written and AI-generated texts were compiled into a CSV file.

- **Labeling**

Human-Generated Texts: Labeled with "H" to indicate they were written by humans.

AI-Generated Texts: Labeled with "IC" to denote that they were created using artificial intelligence.

- **Time and Effort**

Duration: This process took a considerable amount of time, reflecting the careful and meticulous approach needed to ensure data quality.

Workload: The work involved thorough reading, segmentation, rephrasing, and labeling, which was diligently carried out to produce a reliable dataset.

The dataset contains 3535 texts where generated by human tickets with H and generated by chatgpt tickets with IC .

The author for 2 categories: 0 refers to humain, and 1 refers to chatgpt, The figure below represents the distribution of the texts in this dataset.

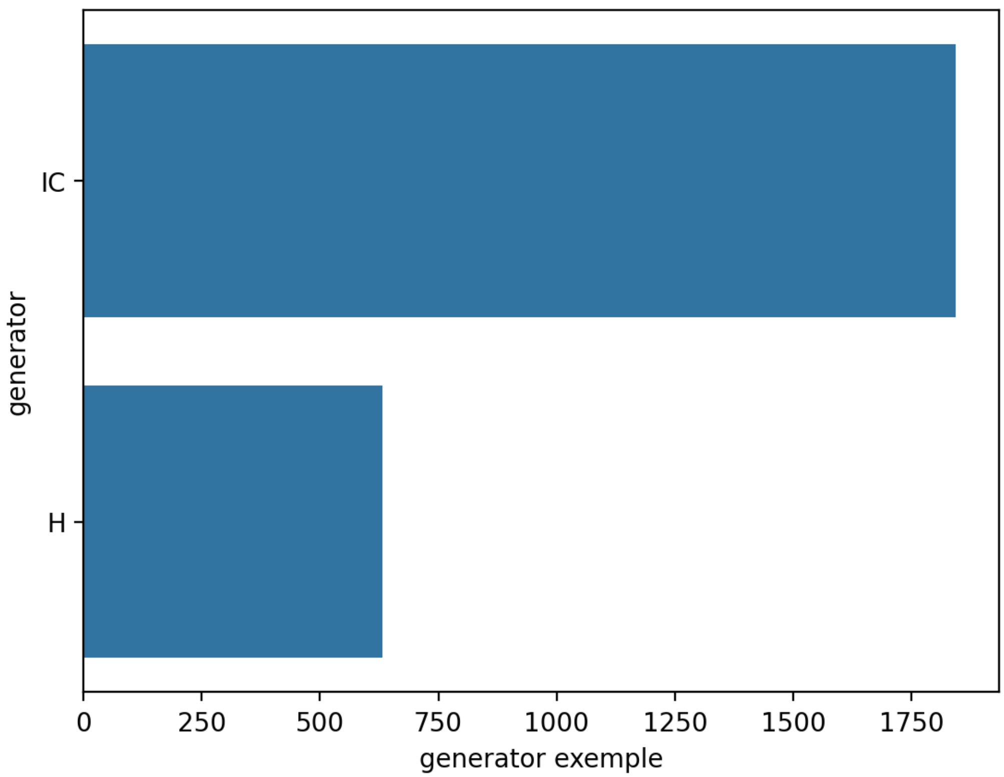


Figure 26:Texts distribution.

3) Data Exploration

words used in our corpus.where each word's size is proportional to its importance (more frequent words appear larger).

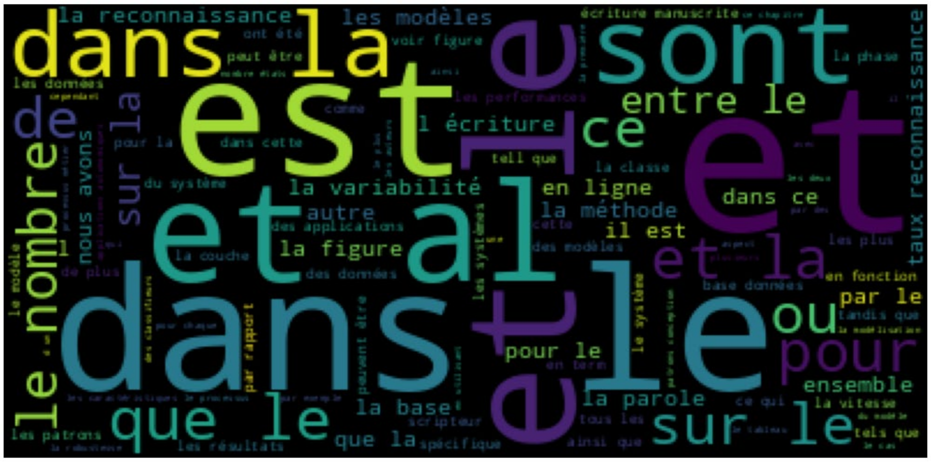


Figure 27: Word cloud of my dataset.

4) Data Transformation

This step consists of converting data to correspond to the fine-tuned camemBERT model (see Figure 3.7). First, we used camemBERT Tokenizer to decompose the sentences into a set of meaningful pieces “tokens” (tokenization), because this tokenizer is the most relevant for French language . Seconde, the special token [CLS] symbol is added at the start of every sentence and the token [SEP] added in the end of each sentence, after that, we get BERT’s format. Third,

we Padding ([PAD]) and truncate all sentences to a single constant length, this Depending on number of tokens in the most comments in our dataset. Then based on the vocabulary of pre-trained camemBERT, each token is mapped to an index. Finally, we create an attention mask; list of 0 and 1 indicating whether the model should consider the tokens or not when learning their contextual representation. We expect [PAD] tokens to have value 0.

5) Experiments on Pre-processing

The pre-processing has a significant impact on building a good model. A wrong pre-processing or not appropriate to the type and structure of the data that we are dealing with will inevitably lead to a decrease in the efficiency of the model, whether in the training or testing. In these experiments, we want to ensure that the pre-processing operations appropriate for our data.

Max length	**Batch size**	**Accuracy in training set %**	**Accuracy in validation set %**
50	16	87,87	90,21
50	32	87,58	90,48
70	16	86,65	90,21
70	32	87,58	90,48
90	16	83,45	92,27
90	**32**	**87,73**	**92,29**
100	16	87,42	91,27

Table 2: Experimental Results.

From the table we note that the best accuracy of the model in the validation set 92,29% was when we set the max_length value to 90 tokens. But when max_length is set to 50 or 70 tokens, the accuracy of the model is reduced somewhat.

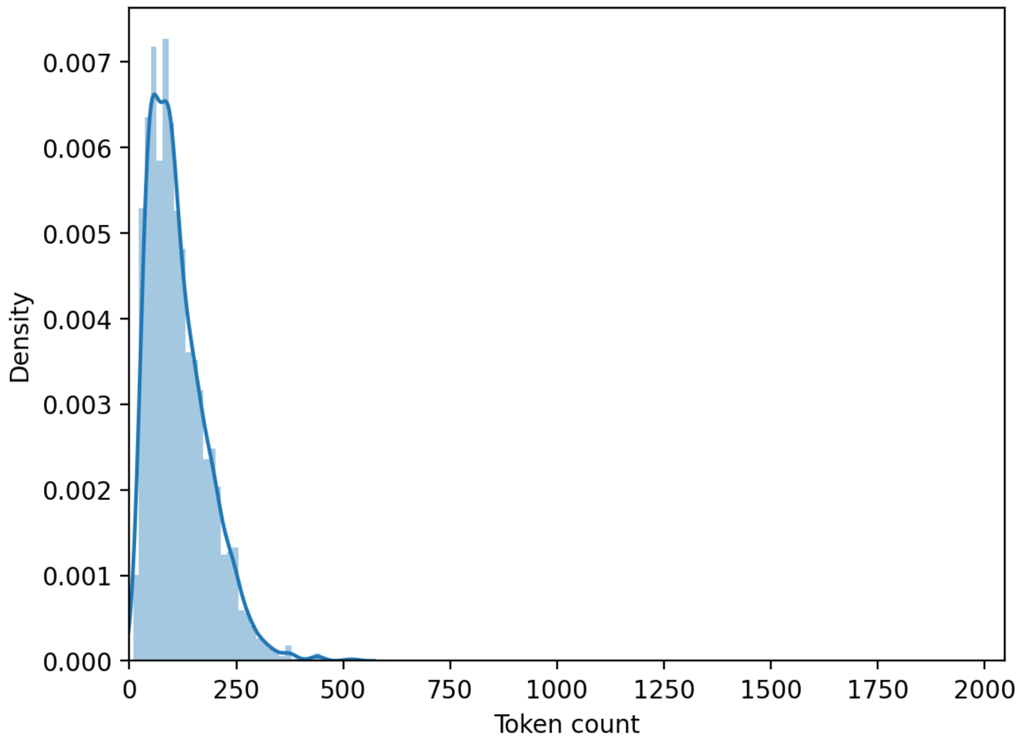


Figure 28: Number of tokens per text.

6) Conclusion

In this chapter, we presented the different stages to creating the texts generated model based on transfer learning. Moreover, a series of experiments were carried out, which allowed us to fix the optimal values of the hyper-parameters, and specify the best pre-processing of our data. Then we present the evaluation of the model and the results obtained, as the model gave an accuracy of 92,29% on a test set, which are considered satisfactory and promising results.

CHAPTER 4:

IMPLEMENTATION

1. Introduction

In this chapter I will speak about the software tools used to develop our model.

2. Software tools

➢ **Google Colaboratory**

Google Colaboratory, often shortened to "Colab" is a cloud service, offered by Google (free or paid), based on Jupyter Notebook and intended for training and research in machine learning. This platform makes it possible to train Machine Learning models directly in the cloud [35]. Without therefore needing to install anything on our computer except a browser. This environment allows us to write and execute Python code in your browser, with no configuration required, free access to GPUs, easy to share.

Figure 29: Google colab logo

➢ **Python Programming language**

Python is robust and user-friendly programming language created by Guido van Rossum and first Realized in 1991. It has simple and effective object-oriented programming techniques and high-level data structures. Python is an ideal language for scripting and rapid application development in many domains and on most platforms due to its easy syntax, dynamic typing, and the fact that it is interpreted.

Python is one of the most popular programming language used by developers today, and it is by far the most used language in the field of Artificial Intelligence, most of the NLP and machine learning libraries are available in Python, and thus it is the best choice for our case.

Figure 30: Python logo

- **Vs Code**

Visual Studio Code (VSCode) is a source code editor and an integrated development environment (IDE) of Microsoft. It is open-source and cross-platform, meaning it runs on Windows, Linux and Mac. It was designed for web developers, but it supports many other programming languages such as C++, C#, Python, Java, etc. It offers many features like syntax highlighting, auto-completion, error highlighting, code navigation, debugging, versioning, integration with Git, and many more. It is also extensible using a wide variety of extensions developed by the community, allowing developers to customize the editor according to their needs.

Figure 31:Vs Code logo

➢ **Django**

Django is a Python framework that makes it easier to create web sites using Python.it takes care of the difficult stuff so that you can concentrate on building your web applications.

Django emphasizes reusability of components, also referred to as DRY (Don't Repeat Yourself), and comes with ready-to-use features like login system, database connection and CRUD operations (Create Read Update Delete).

Figure 32: Django logo

❖ **Used Libraries**

➢ **PyTorch:** It is an open-source Machine learning framework built using python and the torch library. PyTorch uses tensors, which are multidimensional arrays designed for employing the GPU power for computational operations like matrix multiplication.

Figure 33: PyTorch log

➢ **NumPy:** It is a library for Python programming language, intended to manipulate matrices or multidimensional arrays as well as mathematical functions operating on these arrays. More precisely, this free and open source software library provides multiple functions allowing in particular to directly create a table from a file or on the contrary to save a table in a file, and to manipulate vectors, matrices and polynomials. NumPy is the basis of SciPy, a grouping of Python libraries around scientific computing.

Figure 34:NumPy logo

- **Pandas:** Pandas is an open-source Python library that offers powerful tools for data manipulation and analysis, making it a staple in data science workflows. Its primary data structure, the DataFrame, is a two-dimensional, table-like structure with labeled rows and columns. Pandas excels in tasks such as data cleaning, wrangling, exploratory analysis, and visualization. Its extensive range of functions and methods simplifies the handling of structured data, allowing users to load, clean, transform, and analyze their data efficiently.

Figure 35: Pandas logo

- **Transformers:** Transformers is a repository provided by Hugging Face, built with Python, and leveraging PyTorch and TensorFlow 2.0 frameworks. It offers thousands of pretrained models for tasks across various modalities, including text, vision, and audio. Primarily oriented towards Natural Language Understanding (NLU) and Natural Language Generation (NLG) tasks, Transformers supports abstractive summarization. Key features include its simplicity, entry-level code, and helper functions for saving checkpoints and results, facilitating faster model execution and evaluation. This makes it an invaluable tool for researchers and educators to develop and compare custom models efficiently.

Figure 36: Transformers logo

3.Conclusion

In this chapter we speak about the tools used to developed the model.

CONCLUSION GENERAL

Given the impact of artificial intelligence on the creation of scientific texts, we previously discussed how to address this issue and develop a program that allows the detection of AI-generated texts. In the next section, we will address the creative aspect of this project and include the results achieved.

Second part: Commercial aspect

First Axis:

Project submission

First: Summary and Idea of the project:

The project idea revolves around the fact that artificial intelligence has revolutionized many industries, including academic research. With the emergence of AI-generated texts, researchers can now create content that appears high-quality with minimal human effort but actually lacks the foundational elements of scientific research. The use of these texts in theses raises significant questions about their impact on research quality and integrity. This project aims to distinguish and classify instances of scientific texts (articles or theses). The idea of the project is represented in this picture:

Figure 37: 2.1 Project idea

Source: Designed and prepared by the student

1_ Summary of the project

To give an overview of the project, we will discuss the description of the project idea, the reasons for choosing it, and the importance it can play Application through :

Project's name	ThesisGuard AI
the address	An application available on websites
Field of activity	recent ai applications academic research and education
Type of institution	Start-up organization
Project products	Web site for distinguishing AI-generated texts from those written by humans in theses and scientific articles
Labor	01: Aoued Khouloud : Student in Informatique 2 degree master SIAA.

Table 3: the project idea

2_ Project Idea

❖ Presenting the project idea

This project consists of creating a Tool **ThesisGuard AI ,**The idea for this project arose from recent advancements in natural language processing and machine learning, which have enabled the creation of texts that mimic human writing. This capability has led to increased use of AI-generated texts in academic writing, particularly in scientific theses. However, this practice presents several disadvantages:

❖ Project Description

The project described aims to develop a website that can distinguish between AI-generated and human-written texts. This is a significant endeavor, given the growing impact of AI on the integrity of scientific texts in dissertations and research papers. AI-generated texts can sometimes lack originality, rigor, and adherence to scientific principles.

The proposed website addresses this issue by offering a user-friendly platform for checking the authenticity of scientific texts. Users can register by providing their personal information, then paste the text they want to verify into the website. The system will analyze the text and provide a report indicating whether it is likely AI-generated or human-written.

To ensure accessibility, the website offers a free initial trial, followed by the option for monthly or annual subscription fees. This model allows users to experience the platform's value before committing to a paid plan.

Second: The slogan and vision of the project

- **Project logo**

The logo of **Thesis Guard AI** represents in the picture below:

Figure 38:2.2 ThesisGuard AI logo.

- **Project message**

ThesisGuard AI is your reliable partner to Discerning the Incursion of AI Generated Texts in Dissertations.

- **Project vision**

Research Quality Integrity.

Third: The reasons for choosing the project and his objectives

❖ Reasons for choosing the project

- An inexpensive project.
- It is considered an activity that does not require much physical effort due to the presence of an application.
- Inequality of Opportunities: AI can produce an enormous amount of text, while humans are limited, leading to a prolific intellectual output for researchers relying on AI compared to those working ethically.
- Quality and Accuracy of Texts: AI-generated texts can suffer from issues such as grammatical errors, lack of coherence, and contradictions that are difficult to verify later.
- Piracy(plagiarism): Some AI models may generate texts very similar to existing sources.
- Potential Bias in Algorithms: The algorithms used to generate texts can exhibit biases.
- Threats to Academic Integrity: The use of AI-generated texts poses risks to the integrity of academic research.
- Impact on the Quality of Education: Excessive reliance on AI can negatively affect the quality of education.
- Impact on the individual ability: The excessive reliance on artificial intelligence can negatively impact individuals' abilities.

❖ Project goals

· **Develop and Launch:** Create and deploy a fully functional AI-generated content detection website tailored to the Algerian market.

· **Customer Acquisition:** Onboard at least 50 academic institutions within the first year of operation.

· **User Satisfaction:** Achieve a user satisfaction rate of 95% through excellent customer support and continuous improvement.

· **Profitability:** Reach profitability within two years of operation.

- Distinguishing AI-generated text.
- Promote academic integrity.
- Avoid bias in the algorithms used to generate text.
- Strengthening human thought.
- Avoid direct reliance on artificial intelligence.
- Avoid fabricated sources.
- Combating misinformation.
- Protecting intellectual property.
- Ensuring transparency.

Fourth: Values propositions

❖ **Values propositions for ThesisGuard AI**

➢ **Accurate Detection**

High accuracy in identifying AI-generated content in dissertations.

➢ **Time-Saving**

Quick analysis compared to traditional manual checks.

➢ **Credibility Enhancement**

Enhances the credibility and integrity of academic work.

➢ **User-Friendly Platform**

Easy-to-use interface with comprehensive reporting features.

➢ **Comprehensive Reports**

Detailed analysis and reports for users.

➢ **Customizable Solutions**

Tailored services for different customer needs (students, universities, publishers).

➢ **Scalable Service**

Capable of handling large volumes of data and multiple users simultaneously.

➢ **Cost-Effective**

Competitive pricing models with various subscription plans.

- **24/7 Availability**

  Service accessible at any time, ensuring continuous support.

- **Innovative Technology**

  Utilizes cutting-edge AI technology for text analysis and detection.

ACTIVITY	DESCRIPTION	DURATION
Database Creation		
Data Collection	Gather diverse academic texts and dissertations	6WEEKS
Data Cleaning and Preparation	Organize, clean, and preprocess data	6WEEKS
Model Development		
Model Training	Train the AI model	4 WEEKS
Model Testing	Evaluate the model's performance	4WEEKS
Interface Development		
Front-End Design and Development	Create user-friendly interfaces	3WEEKS
Back-End Development	Implement necessary functionalities	3WEEKS
Integration and Testing		
Integration of Model and Interface	Connect AI model with interfaces	3WEEKS
System Testing	Comprehensive system testing	3WEEKS
User Acceptance Testing (UAT)	Gather feedback and make adjustments	2WEEKS

Table 4:Timetable for project realization

Second Axis:

innovative aspects

1_The nature of innovation

- **Technological innovation:**

ThesisGuard AI stands at the forefront of technological innovation in academic integrity and research quality enhancement. Our platform integrates advanced AI and machine learning technologies to provide robust solutions tailored for educational institutions and individual users.

- **Key Technological Features:**

1. AI-Powered Plagiarism Detection: Our system employs sophisticated algorithms to accurately detect and analyze similarities in academic papers and research documents. This ensures comprehensive plagiarism detection, even across vast databases and diverse sources.

2. Transformers Capabilities: Utilizing Transformers, our platform understands and processes complex language structures, enhancing the accuracy and depth of content analysis. This capability allows for nuanced detection of improper citations, paraphrasing, and originality issues.

3.Scalability and Performance:Designed to handle large volumes of data, our scalable infrastructure ensures consistent performance and reliability even during peak usage periods. Institutions and users can rely on uninterrupted service without compromising on quality or speed.

4. Continuous Improvement and Updates: We are committed to ongoing research and development, continuously enhancing our algorithms and functionalities based on user feedback and advancements in AI technology. This commitment ensures that our users always have access to the latest in academic integrity solutions.

- **Benefits of Technological Innovation:**

  - Enhanced Academic Integrity: By leveraging advanced technologies, ThesisGuard AI helps uphold academic standards and integrity, promoting originality and ethical research practices.

  - Improved Research Quality: Researchers and students benefit from accurate and insightful feedback, improving the overall quality of academic work and publications.

  - User-Friendly Interface: Despite its advanced capabilities, our platform maintains a user-friendly interface, ensuring ease of use for both novice and experienced users in academic settings.

  ThesisGuard AI's technological innovation is not just about advanced algorithms; it's about empowering users with tools that foster integrity, elevate research standards, and support academic excellence. Through continuous innovation and enhancement, we remain dedicated to serving the evolving needs of the academic community worldwide.

2_ Areas of innovation

1. National First

- Innovation: ThesisGuard AI is the first system of its kind to be implemented at the national level.
- Advantage:
  - Pioneering Initiative: Establishes a benchmark for academic integrity within the country.
  - Qualitative Advance: Represents a significant leap forward in the field of research and higher education.
  - Setting Standards: Sets a precedent for other nations to follow, potentially leading to international adoption.

2. Ensuring Integrity and Quality

- Innovation: Provides a robust service focused on maintaining the integrity and quality of scientific theses.

- Advantage:
    - Trustworthy Academic Work: Ensures that academic work is original and credible.
    - Reputation Management: Helps institutions maintain their reputation by preventing academic fraud.
    - Quality Assurance: Acts as a quality control measure, reinforcing the standards of academic research.

3. Collaboration with Key Stakeholders

- Innovation: Actively collaborates with a diverse group of stakeholders including:
    - Universities
    - Educational institutions
    - Individual researchers and students
    - Educational program providers
    - Research and development departments
    - Student organizations
- Advantage:
    - Comprehensive Ecosystem: Creates a supportive and collaborative environment for adoption and integration.
    - Stakeholder Engagement: Involves all relevant parties in the process, ensuring the tool meets the needs of its users.
    - Enhanced Adoption: Facilitates widespread acceptance and utilization through active engagement and collaboration.

4. Commercial Partnerships

- Innovation: Forms strategic commercial partnerships with educational technology companies.

- Advantage:
  - Market Expansion: Broadens market reach by leveraging the networks of established ed-tech companies.
  - Increased Accessibility: Makes ThesisGuard AI available to a larger audience, including smaller institutions and individual researchers.
  - Integrated Solutions: Combines with other educational technologies to offer comprehensive solutions to users.

Third Axis:

strategic analysis of the market

1. **Analysis of the Project's Surrounding Environment:**

In this section, we will present a set of analyses of the project's surrounding environment, including the market, customers, and other factors that affect market activity.

- **Market Study:**

A market study is a fundamental tool for adapting to changes in the market in which the institution operates. This study aims to:

Potential Market and Target Market

1) **Potential Market**

The potential market includes individuals and institutions that require or are likely to require products or services to meet their needs and desires. This encompasses a broad spectrum of entities in the academic and research sectors.

➢ **Key Questions**

- Who will buy our products?

  - Universities and Colleges: Public and private higher education institutions.
  - Research Institutions: Government-funded and independent research centers.
  - Academic Publishers: Publishers of academic journals, books, and conference proceedings.
  - Individual Researchers and Students: Graduate and postgraduate students, academic professionals.
  - Educational Program Providers: Online education platforms, professional training organizations.
  - R&D Departments: Corporate and government research and development departments.

- What motivates them to do so?
  - The need to ensure academic integrity and originality of research.
  - The increasing prevalence of AI-generated content in academic submissions.
  - Institutional reputation and adherence to academic standards.
  - Regulatory requirements and guidelines from educational authorities.
- Where are they located?
  - Major cities with educational hubs such as Algiers, Oran, Constantine, Annaba, and Blida.
  - Regions with significant research activity and educational institutions.
- How many are they?
  - Universities and Colleges: Approximately 50 public and private universities.
  - Research Institutions: Several dozen government-funded and independent research centers.
  - Academic Publishers: Multiple publishers focused on academic and scientific literature.
  - Individual Researchers and Students: Tens of thousands of graduate and postgraduate students and researchers.
  - Educational Program Providers: Growing number of online education platforms and training organizations.
  - R&D Departments: Numerous departments within corporate and government sectors.

2) Target Market

The target market represents the specific group of individuals or institutions to whom ThesisGuard AI offers or presents its products. This segment is chosen based on their high likelihood of needing and benefiting from the service.

- **Target Market Segments**

- Universities and Colleges:
  - Justification: These institutions need to ensure the originality of dissertations and theses to maintain academic integrity.
  - Opportunity: Potential for long-term contracts and integration with existing academic management systems.
- Research Institutions:
  - Justification: High demand for validating the authenticity of research papers and reports.
  - Opportunity: Establishing partnerships to provide regular AI detection services for research outputs.
- Academic Publishers:
  - Justification: Need to verify the originality of manuscripts submitted for publication.
  - Opportunity: Offering subscription services to publishers for routine checks of submitted works.
- Individual Researchers and Students:
  - Justification: Ensuring the credibility of their academic work and meeting submission guidelines.
  - Opportunity: Providing affordable subscription plans for individuals.
- Educational Program Providers:
  - Justification: Verifying the authenticity of coursework and projects to uphold program credibility.
  - Opportunity: Collaborating with online education platforms to offer integrated plagiarism detection solutions.

Justification for Target Market Selection

- High Demand: All selected segments have a critical need to maintain the originality and integrity of their academic and research outputs.
- Strategic Fit: These segments align with ThesisGuard AI's capabilities and strengths in detecting AI-generated content.
- Market Size: The selected segments represent a significant portion of the academic and research community in Algeria, offering substantial business opportunities.

Potential for Contract Agreements

- Universities and Colleges:
  - Potential for multi-year contracts to provide ongoing detection services for all submitted dissertations and theses.
- Research Institutions:
  - Possibility of annual contracts to verify research papers and reports before publication.
- Academic Publishers:
  - Subscription-based contracts to regularly check submitted manuscripts for AI-generated content.
- Educational Program Providers:
  - Collaborative agreements to integrate ThesisGuard AI into their platforms, ensuring the authenticity of coursework.

- ## Measuring Competition Points

To effectively measure competition in the market for ThesisGuard AI, it's crucial to identify and analyze the most significant competitors. Here's a detailed look at the key competitors and their impact:

Direct and Indirect Competitors for ThesisGuard AI

1. Direct Competitors

Direct competitors are those that offer similar services specifically aimed at detecting AI-generated texts or maintaining academic integrity. And since the project is the first in Algeria, it has no direct competitors.

2. Indirect Competitors

1.1 Turnitin

- **Description:** A well-known plagiarism detection tool used widely in educational institutions.
- **Strengths:**
  - Established reputation and brand recognition.
  - Extensive database for text comparison.
  - Comprehensive suite of academic integrity tools.
- **Weaknesses:**
  - Higher cost, which may be a barrier for some institutions.
  - May not be specifically tailored for detecting AI-generated texts.

1.2 Grammarly

- **Description:** A popular tool for grammar checking, plagiarism detection, and writing enhancement.

- **Strengths:**
  - User-friendly interface and widespread usage.
  - Advanced grammar and writing improvement suggestions.
  - Plagiarism detection integrated into writing tools.
- **Weaknesses:**
  - Primarily focused on grammar and writing, with plagiarism detection as an additional feature.
  - May not specialize in detecting AI-generated content.

1.3 Unicheck

- **Description:** An academic integrity tool that provides plagiarism detection services.
- **Strengths:**
  - Integrates with various Learning Management Systems (LMS).
  - on providing detailed plagiarism reports.
  - Customizable solutions for different educational needs.
- **Weaknesses:**
  - Smaller database compared to Turnitin.
  - Limited focus on AI-generated text detection.

- ## Marketing strategy

1.Marketing Mix (4P)

ThesisGuard AI's marketing mix focuses on the core elements of product, price, place, and promotion to effectively position and promote its AI-generated text detection services for academic integrity in the Algerian market.

1. **Product**

Description: ThesisGuard AI offers advanced AI technology specifically designed to detect AI-generated texts in academic dissertations and theses.

Features

- AI-powered detection algorithms for identifying plagiarism and AI-generated content.
- User-friendly interface tailored for educational institutions, researchers, and students.
- Compatibility with various document formats and integration capabilities with existing educational platforms.

Differentiation

ThesisGuard AI distinguishes itself from traditional plagiarism detection tools by focusing on the unique challenge of AI-generated content detection, ensuring academic integrity and originality.

2. **Price**

Pricing Strategy ThesisGuard AI will adopt a tiered pricing model based on the type and scale of deployment:

Free Trial: Initial free trial period to showcase product efficacy and encourage adoption.

Subscription Plans: Monthly or annual subscriptions tailored for educational institutions and individual users.

Volume Discounts: Discounts offered for bulk purchases by educational institutions or program providers.

Value Proposition: Competitive pricing aligned with the value provided in maintaining academic integrity and enhancing research quality.

3. **Place**

Distribution Channels: ThesisGuard AI will distribute its services primarily through:

Direct Sales: Direct sales team targeting educational institutions, research departments, and program providers.

Online Platform: E-commerce platform for easy access and purchase of subscription plans.

Strategic Partnerships: Collaborations with EdTech companies and educational institutions to integrate ThesisGuard AI into existing platforms and services.

Geographic Focus: Initially targeting the Algerian market with plans for regional expansion based on market demand and regulatory considerations.

4. **Promotion**

Promotional Strategies:

Digital Marketing: Utilize Search engine optimization, content marketing, and paid advertising to increase online visibility and drive traffic to the ThesisGuard AI website.

Social Media: Engage with stakeholders on LinkedIn, Twitter, and Facebook through educational content, customer testimonials, and updates.

Content Marketing: Publish blogs, whitepapers, and case studies focusing on academic integrity, AI technology, and the benefits of using ThesisGuard AI.

Events and Webinars: Host webinars, workshops, and participate in industry conferences to demonstrate product capabilities and engage with potential customers.

Partnership Promotions: Collaborate with partners to co-market ThesisGuard AI's services and leverage their networks for enhanced reach.

Customer Engagement: Implement a robust customer relationship management (CRM) strategy to nurture leads, provide ongoing support, and gather feedback for continuous improvement.

2. SWOT Analysis

Strengths:

Innovative Technology: Utilizes cutting-edge AI algorithms specifically designed to detect AI-generated texts.

First-Mover Advantage: As a national first, ThesisGuard AI has the opportunity to set industry standards.

High Accuracy: Offers a high detection accuracy, ensuring reliable results.

User-Friendly Interface: Intuitive design makes it accessible to users with varying technical expertise.

Strong Partnerships: Collaborations with universities, research institutions, and educational technology companies expand reach and credibility.

Weaknesses:

Development Costs: High initial investment required for developing and maintaining advanced AI technology.

Market Penetration: As a new product, it may take time to gain trust and widespread adoption.

Dependence on Technology: Relies heavily on continuous updates and improvements to stay effective against evolving AI-generated content.

Opportunities:

Growing Concern Over AI Plagiarism: Increasing awareness and concern about AI-generated content in academia create a strong demand for reliable detection tools.

Global Market Expansion: Potential to expand beyond national borders to international markets.

Educational Trends: The rise of online education and digital submissions increases the need for advanced plagiarism detection tools.

Strategic Alliances: Forming alliances with more educational technology companies and academic institutions can boost market presence and user base.

Government Support: Potential to receive support or endorsement from educational authorities and government bodies.

Threats:

Technological Advancements: Rapid advancements in AI could make detection more challenging and require constant updates to the system.

Competition: Emerging competitors with similar technologies could threaten market share.

Privacy Concerns: Handling sensitive academic data requires stringent data protection measures to avoid privacy issues.

Adoption Resistance: Resistance to change from traditional plagiarism detection tools to new AI-based systems among some institutions.

Economic Factors: Economic downturns or budget cuts in education sectors could impact the adoption and investment in new technologies.

Fourth Axis:

Production and organization plan

1_Production

1. Database Creation:

- Objective:

Establish a comprehensive and reliable database for training and testing the AI model.

- Activities:

  - Data Collection:

Gather a diverse set of dissertations and academic texts.

  - Data Cleaning and Preparation:

Organize, clean, and preprocess the data to ensure quality and consistency.

2. Model Development:

- Objective:

Develop a robust AI model capable of discerning AI-generated texts in dissertations.

- Activities:

Developpe and training the AI model.

  - Training the Model:

Split the dataset into training and testing sets.

Train the AI model using the training dataset.

  - Testing the Model:

Evaluate the model's performance using the testing dataset.

3. Interface Development:

- Objective:

Create user-friendly interfaces for interacting with the AI model.

- Activities:

Develop the front-end and back-end interfaces using

- Design and Development:

Design intuitive and accessible user interfaces.

Implement necessary back-end functionalities.

4. Integration and Testing:

- Objective:

Integrate the AI model with the user interfaces and conduct comprehensive testing.

- Activities:

- Integration:

Connect the trained AI model with the front-end and back-end interfaces.

- Testing:

Perform thorough testing to ensure the system works as intended.

Conduct user acceptance testing (UAT) to gather feedback and make necessary adjustments.

- **Steps to deal with the application:**

Step 1: Visit the Website

1. Open Your Browser:

- Open your preferred web browser.
- Go to the ThesisGuard AI website.

Step 2: Registration

2. Sign Up:

- Click the “Sign Up” button on the homepage.
- Fill in the registration form with your name, email, and password.
- Select your Plan .
- Fill the card number and the Amount

Step 3: Login

3. Log In:

- Click on the “Login” button.
- Enter your registered email and password.
- Click “Login” to access your account.

Step 4: Upload Your Dissertation

4. Enter Your Paragraph:

- Navigate to your user dashboard.
- In the designated section, paste or type the paragraph or text you want to check.

5. Check Text:

- Click the “Check Text” button.
- The system will analyze the text and display the results.

Step 5: Analysis Process

6. Automatic Analysis:

- The system will begin analyzing your text automatically.
- This includes checking for AI-generated content and plagiarism.

Step 6: Review Results

8. Result:

- You will receive a result once the analysis is complete.

2 _ Supply

Requirements for the production process:

• Computer: To create the application initially, manage and process orders, respond to users, review financial transfers, and approve displayed services.

• Labor: Currently represented by the computer operator, who manages the various operations related to the application. This individual is also the project owner.

• Internet: Of course, it is the most important factor in managing the application, as a constant connection to the internet must be ensured.

• Payment: It can be through electronic accounts like Baridimob or others.

3 _ Workforce Requirements for ThesisGuard AI

ThesisGuard AI requires a skilled and diverse workforce to effectively develop, deploy, and maintain its AI-based dissertation integrity system. Here are the key aspects of the workforce needed:

The project initially creates 5 job positions, which are as follows:

- 2 Backend Developers.

- 1 Frontend Developer.

- 1 Marketing Manager.

- 1 Thesis Collection and Data Analysis Manager.

4 _ Key Companies

Our project partners include:

- ✓ Universities and Academic Institutions: Collaborations for large-scale implementations.
- ✓ University of August 20, 1955 Incubator in Skikda: They provide us with valuable support in terms of research and development.

Fifth axis:

Financial plan

1. Introduction

In this chapter, we will discuss the financial aspects of a project to develop and implement an application for specific purposes. We will examine the various fees, costs and revenue sources associated with the project, in order to provide information on its financial viability and sustainability over time. Products and Services.

2. Financial Plan

1. Basic Plan: Suitable for individual students (3,000 DZD/month).

2. Institutional Plan: Designed for educational institutions (20,000 DZD/month).

3. Enterprise Plan: For large institutions and publishers (50,000 DZD/month).

4.Collaboration:

Institutional Plan: Designed for educational institutions (15,000 DZD/month).

Enterprise Plan: For large institutions and publishers (38,000 DZD/month).

5.Free plan: 5 tries per day/ 10 days with ads.

	MONTHLY (DZD)	ANNUAL(DZD)
The headquarters rent	70,000.00	840,000
Employees Salaries	200000,00	2400000
Operating expenses (electricity,water, and other consumables and needs)	50,000.00	60 0000
Modifying the headquarters		500,000
CNAS Insurance Fund Charges		612,000
Develop the final script and host it on hostinger		20,000
Marketing and advertising (social media sites, TV channels, logo and banners)		200,000
Office equipment		250,000
Computer equipment		500,000
Programs and website		70,000
TOTAL		**5,992,000**

Table 5 : Cost schedule

3. Business Number

The first year from the start of activity and the year after completion. There are two delay scenarios, the usual optimistic and pessimistic.

3.1 Revenue Projections for the Optimistic Scenario (First Year)

- **Subscription Plans**

PLAN	Users/Institutions/Enterprises	Monthly Fee (DZD)	Monthly Revenue (DZD)	Yearly Revenue (DZD)
BASIC PLAN	**200**	3,000	600,000	7,200,000
Institutional Plan	**5**	20,000	100,000	1,200,000
Enterprise Plan	**2**	50,000	100,000	1,200,000
Collaboration with institutional	**2**	15,000	30,000	360,000
Collaboration with enterprise	**1**	38,000	38,000	456,000
Total			**868,000**	**10,416,000**

Table 6:Subscription Plans

➢ Costs and Investments (First Year)

- **Initial Investments**

Item	**Amount (DZD)**
Modifying the Headquarters	500,000
Office Equipment	250,000
Computer Equipment	500,000
Programs and Website	70,000
Marketing and Advertising	200,000
Develop the Final Script and Host it on Hostinger	20,000
Total Initial Investments	**1,540,000**

Table 7:Initial Investments

- **Annual Recurring Costs**

• **Item**	**Amount (DZD)**
Headquarters Rent	840,000
Employee Salaries (5 employees)	2,400,000
CNAS Insurance Fund Charges	612,000
Operating Expenses	600,000
Equipment Insurance	37,500
Safety Equipment	15,000
Total Annual Recurring Costs	**4,504,500**

Table 8:Annual Recurring Costs

Grand Total Costs (First Year)

Category	**Amount (DZD)**
Total Initial Investments	1,540,000
Total Annual Recurring Costs	4,504,500
Grand Total Costs	**6,044,500**

Table 9:Grand Total Costs (First Year)

- **Profit Calculation for the First Year**

Item	**Amount (DZD)**
Total Annual Revenue	**10,416,000**
Total Annual Costs	6,044,500
Profit	**4,371,500**

Table 10: Profit Calculation for the First Year

- **Summary Table**

Item	Amount (DZD)
Initial Investments	
- Modifying the Headquarters	500,000
- Office Equipment	250,000
- Computer Equipment	500,000
- Programs and Website	70,000
- Marketing and Advertising	200,000
- Develop the Final Script and Host it on Hostinger	20,000
Total Initial Investments	**1,540,000**
Annual Recurring Costs	
- Headquarters Rent	840,000
- Employee Salaries (5 employees)	2,400,000
- CNAS Insurance Fund Charges	612,000
- Operating Expenses	600,000
- Equipment Insurance	37,500
- Safety Equipment	15,000
Total Annual Recurring Costs	**4,504,500**
Grand Total Costs (First Year)	**6,044,500**
Revenue (First Year)	**10,416,000**
Profit (First Year)	**4,371,500**

Table 11:Summary Table

In the pessimistic scenario for the first year, with detailed revenue and cost breakdowns, the projected profit is **4,371,500 DZD**.

3.2 Revenue Projections for the Pessimistic Scenario (First Year)

- **Subscription Plans Revenue**

PLAN	Users/Institutions/Enterprises	Monthly Fee (DZD)	Monthly Fee (DZD)	Monthly Fee (DZD)
BASIC PLAN	**50**	3,000	150,000	1,800,000
Institutional Plan	**2**	20,000	40,000	480,000
Enterprise Plan	**1**	50,000	50,000	600,000
Collaboration with institutional	**1**	15,000	15,000	180,000
Collaboration with enterprise	**0**	38,000	0	0
Total			**255,000**	**3,060,000**

Table 12:Subscription Plans Revenue

- **Costs and Investments (First Year)**

- **Initial Investments**

Item	**Amount (DZD)**
Modifying the Headquarters	500,000
Office Equipment	250,000
Computer Equipment	500,000
Programs and Website	70,000
Marketing and Advertising	200,000
Develop the Final Script and Host it on Hostinger	20,000
Total Initial Investments	**1,540,000**

Table 13:Initial Investments

- **Annual Recurring Costs**

Item	**Amount (DZD)**
Headquarters Rent	840,000
Employee Salaries (5 employees)	2,400,000
CNAS Insurance Fund Charges	612,000
Operating Expenses	600,000
Equipment Insurance	37,500
Safety Equipment	15,000
Total Annual Recurring Costs	**4,504,500**

Table 14:Annual Recurring Costs

- **Grand Total Costs (First Year)**

Category	**Amount (DZD)**
Total Initial Investments	1,540,000
Total Annual Recurring Costs	4,504,500
Grand Total Costs	**6,044,500**

Table 15:Grand Total Costs (First Year)

- **Profit Calculation for the First Year**

Item	**Amount (DZD)**
Total Annual Revenue	3,060,000
Total Annual Costs	6,044,500
Profit	**8,703,500**

Table 16:Profit Calculation for the First Year

- **Summary Table**

Item	**Amount (DZD)**
Initial Investments	
- Modifying the Headquarters	500,000
- Office Equipment	250,000
- Computer Equipment	500,000
- Programs and Website	70,000
- Marketing and Advertising	200,000
- Develop the Final Script and Host it on Hostinger	20,000
Total Initial Investments	**1,540,000**
Annual Recurring Costs	4,720,500
- Headquarters Rent	840,000
- Employee Salaries (5 employees)	2,400,000
- CNAS Insurance Fund Charges	612,000
- Operating Expenses	600,000
- Equipment Insurance	37,500
- Safety Equipment	15,000
Total Annual Recurring Costs	**4,504,500**
Grand Total Costs (First Year)	**6,044,500**
Revenue (First Year)	3,**060,000**
Profit (First Year)	**2984500**

Table 17:Summary Table

In the pessimistic scenario for the first year, with detailed revenue and cost breakdowns, the projected profit is **2984500 DZD**.

Sixth Axis:

The prototype

1. Introduction

In this chapter I will speak about the interfaces and users functions.

2.Interfaces

2.1. Home page

It’s a simple home page represent the idea of the web site.

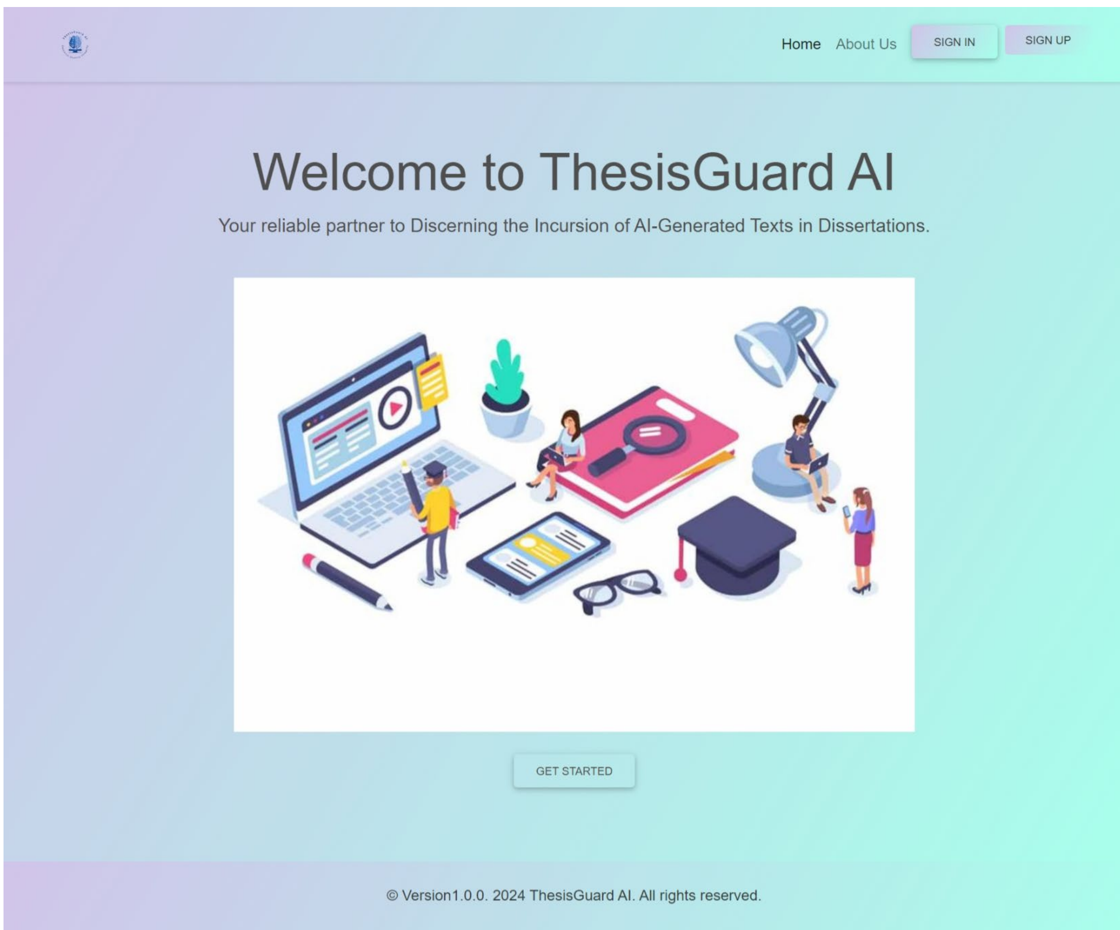


Figure 39:2.3 Home page

2.2 About Us page

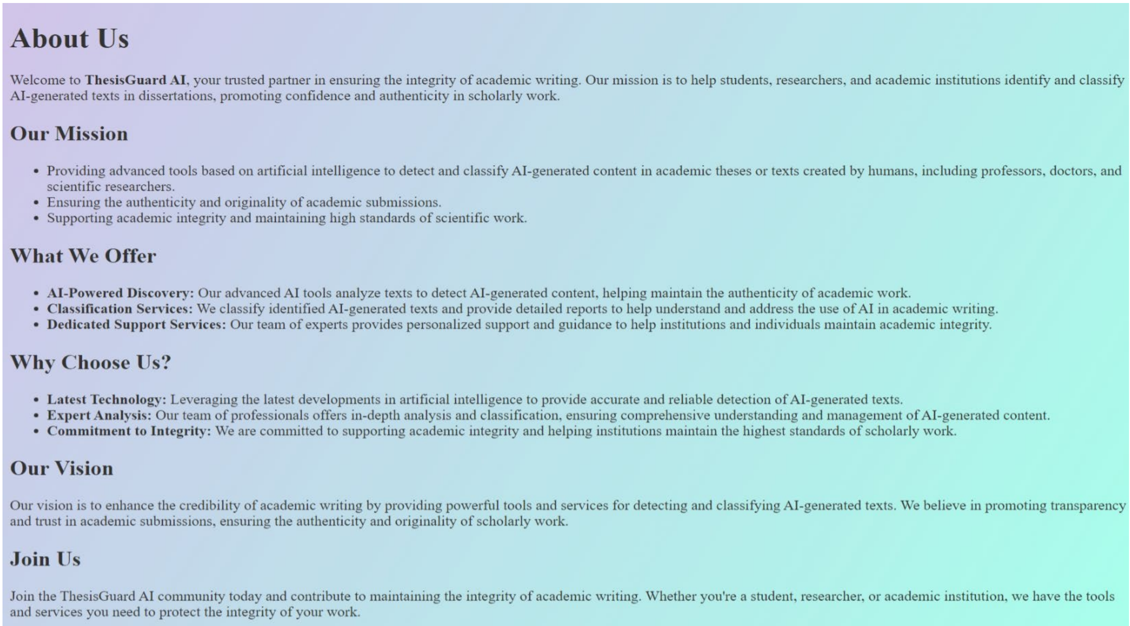


Figure 40: 2.4 About us page

2.3 Login page

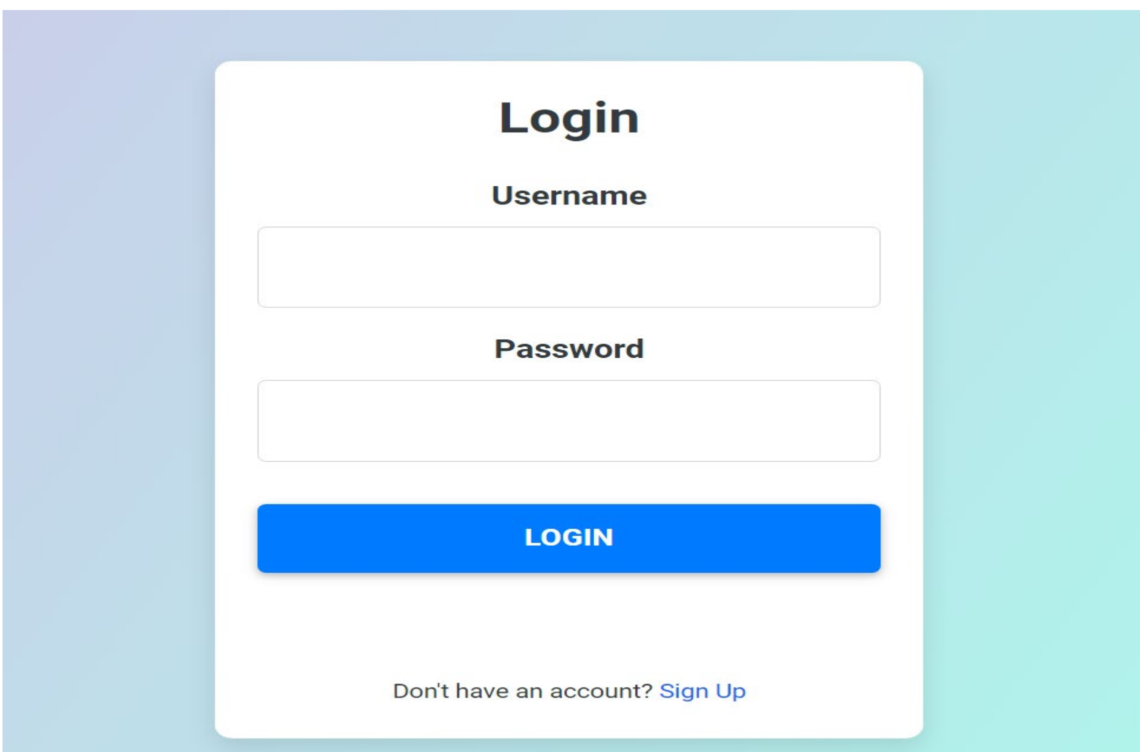


Figure 41: 2.5 Login page

2.4 Signup page

In this page you need to sing with your personnel informations, choosing your plan then add the card number ,and finally the amount.

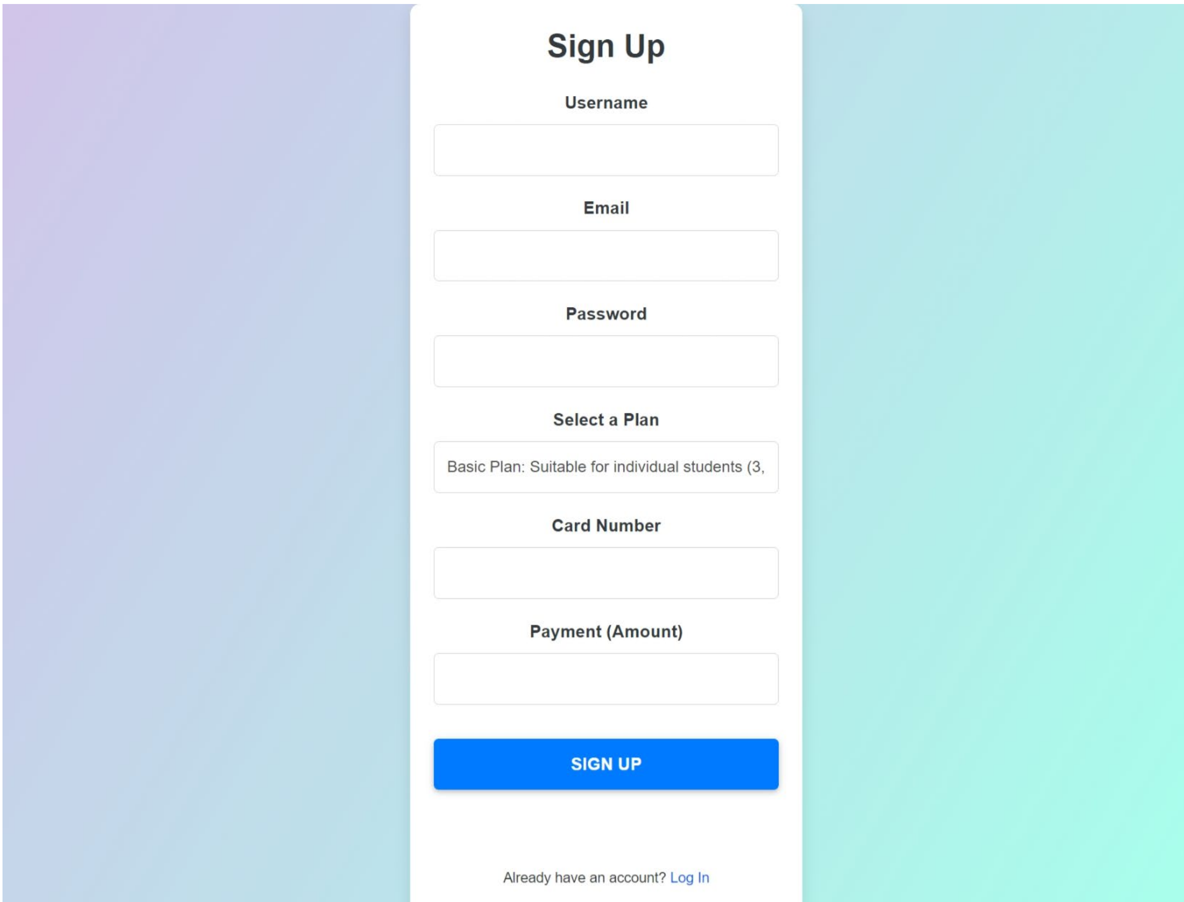


Figure 42: 2.6 Singup page

2.5 Main page

In this page we detect the text if it generated by Human or Chatgpt.

The results represent in figures bellow :

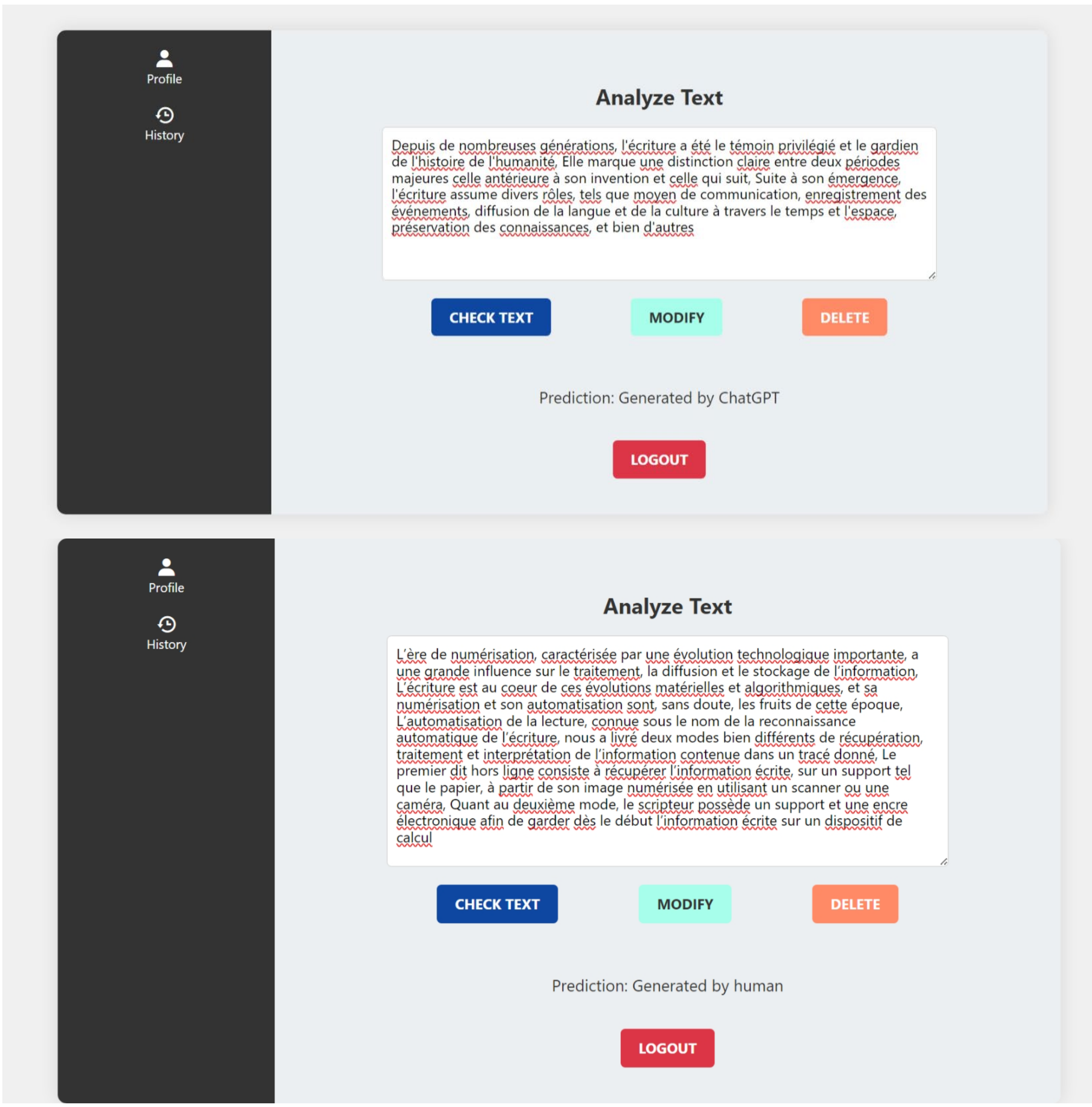


Figure 43: 2.7 Main page

List of appendices

BILANS

ACTIF						
	REALISATION			PREVISION		
En milliers DZD	N-2	N-1	N	N+1	N+2	N+3
Immobilisation Incorporelles						
Logiciels et site web	0	0	20,000.00	20,000.00	20,000.00	20,000.00
Immobilisation Corporelles						
Équipements informatiques	0	0	500,000.00	100,000.00	50,000.00	50,000.00
Équipements de bureau	0	0	250,000.00	10,000.00	10,000.00	10,000.00
TOTAL ACTIF	0	0	770,000.00	130,000.00	80,000.00	80,000.00
PASSIF						
CAPITAUX PROPRES	0	0	6,200,000.00	-	-	-
TOTAL PASSIF	0	0	6,200,000.00	-	-	-

Table 18:BILANS

COMPTE DE RUSULTAT PREVISIONNEL

	REALISATION			PREVISION		
En milliers DZD	**N-2**	**N-1**	**N**	**N+1**	**N+2**	**N+3**
Vente et produits annexes	**0**	**0**	**10,416,000**	**11,000,000**	**12,000,000**	**12,500,000**
Production de l'exercice						
Location du local	**0**	**0**	**840,000.00**	**840,000.00**	**840,000.00**	**840,000.00**
Aménagement du local	**0**	**0**	**500,000.00**	**-**	**-**	**-**
Hébergement sur AWS	**0**	**0**	**20,000.00**	**20,000.00**	**20,000.00**	**20,000.00**
Marketing et publicité	**0**	**0**	**200,000.00**	**200,000.00**	**100,000.00**	**50,000.00**
Frais d'exploitation	**0**	**0**	**600,000.00**	**600,000.00**	**600,000.00**	**600,000.00**
Assurance des équipements	**0**	**0**	**37,500.00**	**37,500.00**	**37,500.00**	**37,500.00**
Consommation de l'exercice	**0**	**0**	**2,197,500.00**	**1,697,500.00**	**1,697,500.00**	**1,697,500.00**
Valeur ajoutée d'exploitation						
Charges de personnel	**0**	**0**	**2400,000.00**	**2400,000.00**	**2400,000.00**	**2400,000.00**
Impôts et taxes et versement assimilés	**0**	**0**	**612,000.00**	**612,000.00**	**612,000.00**	**612,000.00**
RESULTAT NET DE L'EXERCICE	**0**	**0**	**3,012,000.00**	**3,012,000.00**	**3,012,000.00**	**3,012,000.00**
Vérificationde l'équilibreActif/Passif	**-220,500.00**					

Table 19:COMPTE DE RUSULTAT PREVISIONNEL

BMC

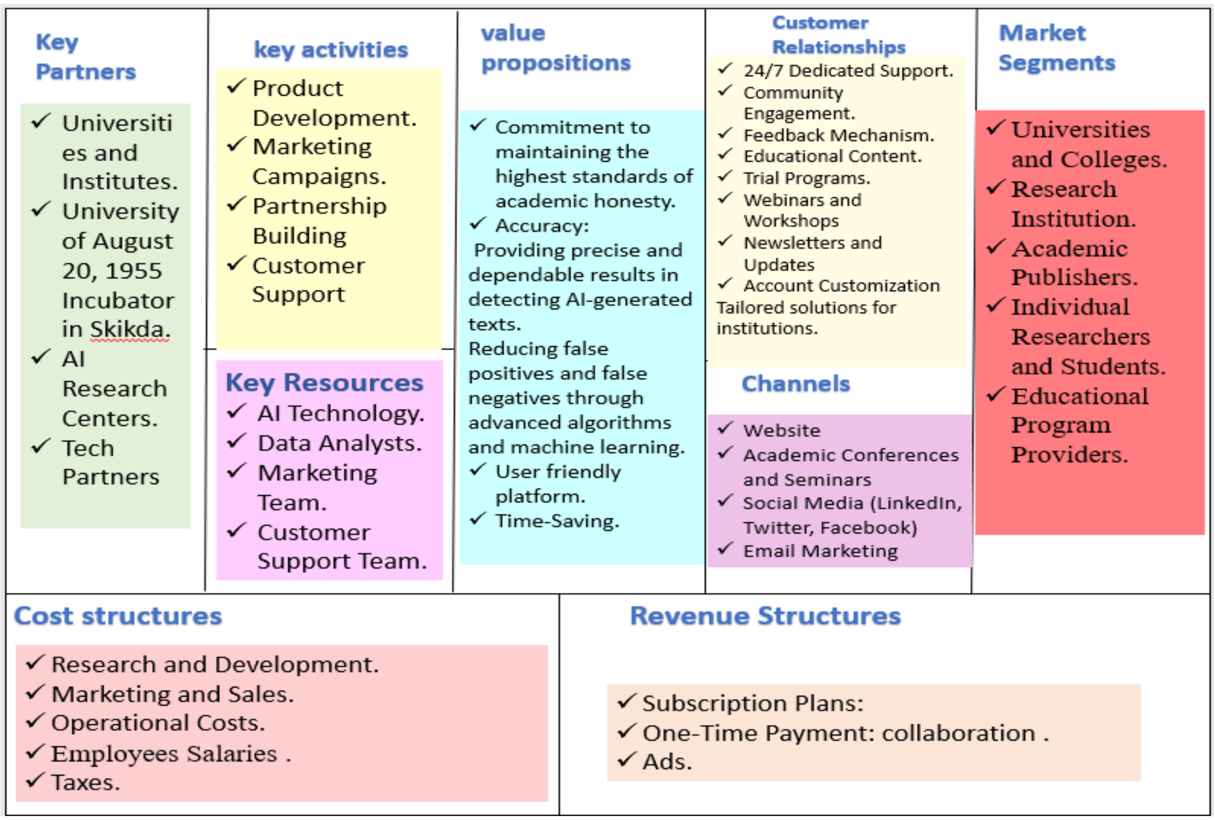


Figure 44: 2.8 Representation of ThesisGuard AI BMC

BIBLIOGRAPHIE:

[1]author = {Mahesh, Batta},year = {2019},title = {Machine Learning Algorithms -A Review}.

[2]author={Sarker, Iqbal H},year={2021}.

[3]author={Busuttil, Steven},title={Support vector machines}.

[4]author={van den Bergh, Don and Clyde, Merlise A and Gupta, Akash R Komarlu Narendra and de Jong, Tim and Gronau, Quentin F and Marsman, Maarten and Ly, Alexander and Wagenmakers, Eric-Jan}, year={2021}.

[5]article{eddy2004hidden,title={What is a hidden Markov model?},

author={Eddy, Sean R}, year={2004}.

[6]author={Topalli, Ndricim},title={Report Title: LHC data analysis for Higgs-Bosons data detection},year={2020}.

[7] Vasilev, I., Slater, D., Spacagna, G., Roelants, P., & Zocca, V. (2019). Python Deep Learning: Exploring deep learning techniques and neural network architectures with Pytorch, Keras, and TensorFlow. Packt Publishing Ltd.

[8]J.W. Boers and Herman Kuiper.

[9](Cour AI mr:BOUGAMOUZA FATEH / 2023).

[10](Author Richa Rao ,Program Authorized to Offer Degree:Electrical and Computer Engineering 2022)

[11]author = {Ghosh, Anirudha and Sufian, A. and Sultana, Farhana and Chakrabarti, Amlan and De, Debashis},

[12]@article{saleem2022comparative,title={Comparative analysis of recent architecture of Convolutional Neural Network},author={Saleem, Muhammad Asif and Senan, Norhalina and Wahid, Fazli and Aamir, Muhammad and Samad, Ali and Khan, Mukhtaj and others}.

[13]@article{saleem2022comparative,title={Comparative analysis of recent architecture of Convolutional Neural Network},author={Saleem, Muhammad Asif and Senan, Norhalina and Wahid, Fazli and Aamir, Muhammad and Samad, Ali and Khan, Mukhtaj and others}.

[14]Lena voita. (2023). NLP Course-Seq2seq and Attention.

[15]@article{han2021pre,title={Pre-trained models: Past, present and future},author={Han, Xu and Zhang, Zhengyan and Ding, Ning and Gu, Yuxian and Liu, Xiao and Huo, Yuqi and Qiu,Jiezhong and Yao, Yuan and Zhang, Ao and Zhang, Liang and others}.

[16]E. Crothers, N. Japkowicz, and H. L. Viktor, "Machine-generated text:A comprehensive survey of threat models and detection methods," IEEEAccess, 2023.

[17]A., & Sung, A. H. (2023, October). How to Detect AI-Generated Texts?. In 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0464-0471). IEEE.

WEBOGRAPHIE

[Web 1]https://www.javatpoint.com/reinforcement-learning

[Web 2]https://semiengineering.com/deep-learning-spreads/

[Web3]https://saurabhnativeblog.medium.com/what-is-the-relation-between-artificial-and-biological-neuron-18b05831036

[Web 4]https://www.javatpoint.com/artificial-neural-network

[Web5]https://www.researchgate.net/figure/Exemple-dune-operation-de-convolution_fig3_337635267

[Web 6]https://www.researchgate.net/figure/Exemple-dune-operation-de-max-

[Web 7]https://www.geeksforgeeks.org/large-language-model-llm/

[Web 8]https://wikidocs.net/167211

[Web9]https://www.google.com/url?sa=i&url=https%3A%2F%2Fgithub.com%2Ftopics%2Fmultihead-attention&psig=AOvVaw1qYXtXurHJQrh3tC-JS8pm&ust=1714336983992000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCLjl14uh44UDFQAAAAAdAAAAABAo

[Web10]https://www.google.com/url?sa=i&url=https%3A%2F%2Fmedium.com%2F%40pennQuin%2Fattention-mechanisms-in-transformers-a726d8f5d5ee&psig=AOvVaw1qYXtXurHJQrh3tC-JS8pm&ust=1714336983992000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCLjl14uh44UDFQAAAAAdAAAAABA1

[Web11]https://medium.com/@reyhaneh.esmailbeigi/bert-gpt-and-bart-a-short-comparison-

5d6a57175fca

[Web12]https://medium.com/@reyhaneh.esmailbeigi/bert-gpt-and-bart-a-short-comparison-5d6a57175fca

[Web 13]https://aclanthology.org/2023.emnlp-main.647.pdf

[Web 14]https://cobusgreyling.medium.com/the-large-language-model-landscape-9da7ee17710b

[Web15]https://medium.com/the-llmops-brief/introduction-to-large-language-models-9ac028d34732

[Web16]https://medium.com/the-llmops-brief/introduction-to-large-language-models-9ac028d34732

[Web 17]https://medium.com/@abhinavkimothi/3-llm-architectures-f527ed781ba9

[Web18]https://medium.com/the-llmops-brief/introduction-to-large-language-models-9ac028d34732

[Web19]https://www.appypie.com/blog/top-10-real-world-applications-of-large-language-models

[Web20]https://medium.com/the-llmops-brief/introduction-to-large-language-models-9ac028d34732

[Web 21]https://cobusgreyling.medium.com/the-large-language-model-landscape-9da7ee17710b

[Web 22]https://www.semrush.com/blog/how-to-detect-ai-written-content-and-plagiarism/

[Web23]https://www.semrush.com/blog/how-to-detect-ai-written-content-and-plagiarism/ " How to Detect AI-written Content and Plagiarism " , Oct 02, 2023, Semrush Team, accessed Apr 24, 2024

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليـــــــــم العالـــــــــــي و البحث العلمـــــــــي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جـامعة 20 أوت 1955- سكيكدة
كــــلية العلـــوم
مكتبـة الكليـة

Université 20 Août 1955 - Skikda
Faculté des Sciences
Bibliothèque de la faculté

سكيكدة في / /2024

بطاقة معلومات خاصة بمذكرة التخرج

اسم و لقب الطالب :
* خلود عواد
*
*
*

رقم التسجيل :
* 36003077
*
*
*

اسم و لقب المشرف على المذكرة : dr. Bougamouza Fateh

عنوان المذكرة : ThesisGuard AI: Discerning the Incursion of AI-Generated Texts in Dissertations

القسم : Informatique

المستوى : Master 2

التخصص : System d'information avanncé et applications

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليـــم العالـــي و البحث العلمـــي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 20 Août 1955- skikda- جـامعة 20 أوت 1955- سكيكدة

Faculté des Sciences كـلية العلـوم

Département d'Informatique قسم الاعلام الآلي

الرقم : ..../ق/ ١/ ١/ ك ع/2024

Autorisation de Dépôt de Mémoire de Master

Je soussigné: ..Bougamouza....Fateh.............................

Certifie que l'étudiant(e) :....Aoued...Khouloud...............

Spécialité : ....Systèmes d'Information Avancés et applications

Ayant soutenu le projet intitulé :...Discerning the Incursion
.........Of AI-Generated Texts in Dissertations...........
....

A apporté les corrections nécessaires sur son manuscrit de Master

Signature de l'encadreur