Bachelor's thesis

Bachelor's Degree of Engineering ICT

2023

Travis Dyde

Documentation on the emergence, current iterations, and possible future of Artificial Intelligence with a focus on Large Language Models.

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

Bachelor’s Thesis | Abstract

Turku University of Applied Sciences

Bachelor’s Degree in Engineering ICT

2023 | 53 pages

Travis Dyde

Documentation on the emergence, current iterations, and possible future of Artificial Intelligence with a focus on Large Language Models.

This thesis presents a comprehensive exploration of artificial intelligence and large language models, discussing their historical evolution, definitions, and contemporary significance. The research delves into the foundational aspects of natural language processing, making clear its fundamentals, including text processing, tokenization, parsing, and Part-of-Speech Tagging. A critical examination of machine learning for natural language processing introduces concepts such as supervised vs. unsupervised learning and word embeddings.

The focus then shifts to leading large language models, providing an overview of prominent models like Generative Pre-Trained Transformer 4, Claude 2, and Pathways Language Model v2, along with key milestones in their development. Ethical and societal implications of artificial intelligence, addressing bias, privacy, security concerns, and environmental impact, are explored in this thesis, highlighting mitigation strategies.

Furthermore, the thesis contemplates the future of artificial intelligence, envisioning potential advancements. The objective is to offer a comprehensive understanding of artificial intelligence and large language models, emphasizing their role, ethical considerations, and future trajectories. The research employs literature review, case studies, and critical analysis of existing models.

The conclusion drawn is that while artificial intelligence presents unprecedented opportunities, responsible development and deployment are imperative to mitigate ethical challenges. This research contributes to the discourse on artificial intelligence, serving as a resource for navigating the evolving landscape of artificial intelligence.

Keywords:

Artificial intelligence, large language models, natural language processing.

Contents

Pictures

Tables

List of Abbreviations

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
BOW	Bag of words
CLIP	Contrastive Language-Image Pre-training
GDPR	General Data Protection Regulation
GloVE	Global vectors for word representation
GPT	Generative Pre-Trained Transformer
ICT	Information and Communications Technology
LLM	Large Language Model
NLP	Natural Language Processing
PaLM2	Pathways Language Model v2
POS Tagging	Part-of-Speech Tagging
URL	Uniform Resource Locator

1. Introduction

The arrival of Artificial Intelligence (AI) has ushered in a transformative era, redefining the boundaries of technological capabilities and reshaping numerous facets of human existence. This thesis aims to provide a basic documentation of the emergence, current iterations, and potential future trajectories of AI, with a specific focus on Large Language Models (LLMs). LLMs, exemplified by advanced models like Generative Pre-Trained Transformer 3 (GPT-3), represent a pinnacle in natural language processing, demonstrating unprecedented language understanding and generation capabilities. By delving into the historical evolution of AI, analysing the current landscape of LLMs, and projecting their potential future developments, this thesis seeks to contribute to the academic discourse surrounding the multifaceted implications and applications of AI, particularly within the domain of large-scale language processing. As society stands at the intersection of human ingenuity and machine intelligence, understanding the nuanced evolution and potential trajectories of AI, especially in the realm of language models, becomes crucial for informed decision-making and ethical considerations in the ongoing technological revolution. Additionally, different AI tools were utilized to assist with wording in this documentation. From its initial concept in 1956 (McCarthy et al., 2006) AI was just an idea and scientists were curious as to whether it was possible for a machine to perform tasks that typically required human intelligence. Today the question has moved from whether it is possible, to how far we can advance AI and its importance in society. AI also raises the question of environmental concerns when it comes to the huge amount of energy needed to train and run LLMs, and the amount of waste that it generates. Another concern is the gathering of data that is scraped from open online sources and what rights do we have when it comes to our data being used to train AI.

A recent overview titled “LLMs and AI: Understanding its Reach and Impact” by Anand Gokul (Gokul, 2023), bought up similar concerns when it comes to LLMs and the impact they have on society regarding transparency and data security measures. While another article titled “Risks and Benefits of Large Language Models for the Environment” (Rillig et al., 2023) investigated the ways that LLMs can positively and negatively affect our environment.

In this thesis Chapter 2 introduces the technology being studied, it looks at the history of AI from when the term was first coined in the 1950s, systems were developed to mimic human problem solving, creation of medical diagnosis and chemical analysis systems, the “AI Winter” during the 70s and 80s to the resurgence in the 21st century. Goals and milestones of AI are listed, and the importance of AI today is discussed which leads into the introduction to LLMs and what they have achieved through recent developments.

Discussing the basics of NLP in Chapter 3 gives an understanding of the interaction between computers and human language. NLP aims to enable computers to understand, interpret, and generate human language in a way that is both meaningful and contextually relevant. With machine learning, a fundamental difference lies at the heart of the learning process—supervised and unsupervised learning. This chapter serves as an exploration into the foundational distinctions and intricate workings of these two approaches, which stand as pillars supporting various machine learning concepts and applications. A nuanced understanding of their contrast is paramount for anyone venturing into the expansive domain of AI.

There is a plethora of different LLMs available today, Chapter 4 delves into the prominent ones, either by being the best in their field or by developing in a way that is seen as ethical by reducing potential risk and biases.

The ethical and societal implications of AI play a large part in the future and growth of this industry. Chapter 5 discusses the importance of creating AI that is being developed with regards to personal privacy, trust and transparency and bias

mitigation. Even though there are many positive benefits to AI and LLMs, there are plenty of negatives, some of which include adding malicious datasets to an already existing LLM to produce a new tool that can provide cybercriminals with easier access to malicious code generation. E-Waste and AI's carbon footprint finish off Chapter 5, providing information about the current state of AI's environmental footprint.

Lastly, in Chapter 6, a brief look at what could be instore for the future of AI and how it could shape different sectors moving forward.

2. Introduction to AI and Large Language Models

2.1. Introduction to Artificial Intelligence

2.1.1. Historical Overview

Artificial Intelligence (AI) is a field of computer science that seeks to develop machines and systems capable of performing tasks that typically require human intelligence (McCarthy et al., 2006). The history of AI dates to the mid-20th century, marked by several significant milestones and breakthroughs.

The concept of AI can be traced back to the Dartmouth Workshop in 1956, often regarded as the birth of AI (McCarthy et al., 2006). At this workshop, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, the term "artificial intelligence" was coined. The participants were optimistic about the potential of computers to simulate human intelligence.

In the early years, researchers developed rule-based systems and symbolic reasoning to mimic human problem-solving. Notable developments during this era included the creation of the Logic Theorist by Allen Newell and Herbert A. Simon, capable of proving mathematical theorems (“(PDF) Newell and Simon’s Logic Theorist: Historical Background and Impact on Cognitive Modeling,” n.d.).

The 1960s and 1970s saw the emergence of expert systems, which utilized knowledge bases and rule-based inference engines to solve complex problems in specific domains. DENDRAL, a system for chemical analysis, and MYCIN, a medical diagnosis system, were among the pioneering expert systems.

However, AI research faced several challenges and setbacks, leading to what became known as the "AI winter" in the 1970s and 1980s. Funding for AI dwindled, and progress slowed as the limitations of symbolic AI became apparent.

The resurgence of AI in the 21st century can be attributed to a range of factors, including advances in machine learning, the availability of large datasets, and more powerful computing hardware. Machine learning techniques, such as neural networks, gained prominence, enabling breakthroughs in computer vision, natural language processing, and game-playing AI (“History Of AI In 33 Breakthroughs: The First Expert System,” n.d.).

In recent years, the development of large-scale language models, such as Generative Pre-trained Transformer 3 (GPT-3), has displayed the potential of AI in understanding and generating human language (Brown et al., 2020). These models have found applications in chatbots, language translation, content generation, and more.

2.1.2. Definition and Goals

AI is a multidisciplinary field of computer science that focuses on creating machines and systems capable of performing tasks that typically require human intelligence. AI seeks to replicate, simulate, or mimic human cognitive functions, such as learning, reasoning, problem-solving, perception, and language understanding (“What it is and why it matters | SAS,” n.d.).

One widely accepted definition of AI is provided by SAS:

"Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think and learn like humans. The term AI is often used to describe machines or computers that can mimic 'cognitive' functions such as learning and problem-solving." (“What it is and why it matters | SAS,” n.d.).

The primary objective of AI is the automation of tasks and processes to enhance efficiency and productivity across various industries. This involves the automation of repetitive and time-consuming activities in sectors such as manufacturing, customer service, and data analysis. The overarching aim is to develop machine learning algorithms capable of pattern recognition, predictive analysis, and adaptation to novel information. This objective is particularly pertinent in applications such as image and speech recognition, as well as recommendation systems.

A focal point within the realm of AI is Natural Language Processing (NLP), which seeks to empower machines to comprehend, interpret, and generate human language. NLP applications encompass a spectrum of functionalities, ranging from chatbots and language translation to sentiment analysis and text summarization (“What it is and why it matters | SAS,” n.d.).

Another key facet of AI pertains to the creation of intelligent robots and autonomous systems capable of executing tasks within real-world environments. These robots find application in diverse fields including healthcare, logistics, and manufacturing.

AI endeavours to construct expert systems mirroring the decision-making processes of human experts in specific domains. Such systems prove invaluable in tasks like medical diagnosis and financial planning.

A distinct avenue of AI research centres on fostering creative thinking in machines, enabling them to produce artistic content such as music, art, and literature.

An essential consideration as AI advances is ensuring ethical and responsible operation, encompassing the mitigation of biases in AI systems, safeguarding privacy, and promoting transparency.

While current AI systems are specialized for specific tasks, the goal is the development of Artificial General Intelligence (AGI) that emulates human-like intelligence and excels across a broad spectrum of tasks. The realization of AGI

remains a long-term aspiration in the field of artificial intelligence (“What it is and why it matters | SAS,” n.d.).

2.1.3. Importance in Contemporary Society

The 2021 Study, titled "Gathering Strength, Gathering Storms," highlights the growing importance of AI in contemporary society, offering insights into various facets of AI's impact on our lives (Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report | One Hundred Year Study on Artificial Intelligence (AI100), n.d.).

AI is making a positive difference in many areas of contemporary society including but not limited to healthcare, economy, environmental sustainability, education and workplace development, autonomous systems, ethical and societal considerations, global competition and collaboration, responsible AI governance and AI’s future.

AI is transforming healthcare through applications like disease diagnosis, drug discovery, and personalized treatment plans. Machine learning models analyse vast medical datasets to identify patterns and predict patient outcomes, leading to more effective healthcare interventions.

From an economic point of view, AI has a substantial economic impact, driving innovation and productivity across industries. It fuels automation, streamlines processes, and enhances decision-making, contributing to economic growth. However, this can also have a negative effect by making workers redundant or reduce wages by making a job easier to do, that requires a less skilled worker (“Automation Doesn’t Just Create or Destroy Jobs — It Transforms Them,” n.d.).

On the other side of this, AI is reshaping education with personalized learning experiences and intelligent tutoring systems. It also plays a role in workforce

development by enabling reskilling and upskilling in response to evolving job requirements.

AI is employed in environmental monitoring and resource management. It aids in predicting and mitigating environmental challenges, such as climate change, natural disasters, and biodiversity preservation. There are also downsides to the environment caused by AI, these will be discussed further in this paper.

Autonomous vehicles, drones, and robots leverage AI to navigate and interact with their environments. These systems enhance safety, efficiency, and convenience in transportation and logistics.

The AI100 Study underscores the importance of addressing ethical and societal implications. These include addressing bias in AI algorithms, ensuring transparency, and creating guidelines for responsible AI deployment. Establishing frameworks for responsible AI governance is a key concern. Policymakers and stakeholders are working to strike a balance between fostering innovation and ensuring the ethical and responsible use of AI. With the AI industry advancing rapidly in recent years there is need for proper guidelines.

The AI100 study emphasizes the global nature of AI development and the importance of international collaboration. Nations around the world are investing in AI research and development to maintain competitiveness.

When it comes to societal considerations the AI100 Study acknowledges that AI's influence on society will continue to evolve. Preparing for the future involves adapting to AI-driven changes, anticipating challenges, and leveraging AI's potential for the benefit of humanity.

2.2. Large Language Models (LLMs)

2.2.1. What are LLMs?

Large Language Models (LLMs) have emerged as a pioneering innovation in the realm of Artificial Intelligence (AI) and are making profound strides in the field of Generative AI. These models are transforming the landscape of AI-driven language comprehension and generation, bearing significant implications for numerous sectors.

At the core of LLMs lies their capacity to master language through extensive exposure to data, comprising vast, often unlabelled, or uncategorized text. They are primed to employ self-supervised or semi-supervised learning approaches, a testament to their adaptability and the power of data-driven learning methodologies.

The fundamental premise of LLMs revolves around their ability to predict the subsequent element in each sequence of text. In other words, they endeavour to anticipate the next word or term in each context. To achieve this, these models are presented with an initial textual prompt, and their underlying neural network architecture processes this input, generating a completion based on the inherent linguistic patterns and data they have absorbed during training.

One of the defining features of LLMs is their immense scale, characterized by the deployment of parameters that number in the millions, billions, or even trillions. These parameters play an instrumental role in helping LLMs navigate the landscape of language and select the most appropriate response or word choice. When a user enters a prompt, these parameters are leveraged to determine the likelihood of potential word completions. Each choice is associated with a specific probability grounded in the model's previous encounters with similar contexts.

However, it is imperative to underscore that LLMs do not operate with an intrinsic understanding of the world akin to humans. Instead, they function as statistical estimators, offering responses that correspond to the probabilities derived from their training data. The implication is that an LLM, when presented with the same prompt on separate occasions, might generate different responses based on the inherent variability in probabilities.

Yet, the quality of LLM responses hinges intrinsically on the quality of the data with which they have been trained. If the data is biased, incomplete, or undesirable in any manner, the responses generated by these models can be equally unreliable or objectionable. In situations where responses veer significantly from the norm, data analysts often refer to such occurrences as "hallucinations."

Furthermore, what makes LLMs particularly intriguing is their ability to extrapolate beyond their initial training domain. They can transfer knowledge from one domain to another, generating code in programming languages they have never encountered or crafting sentences in languages for which they were not explicitly trained. This emergent behaviour underscores the inherent complexity and adaptability of neural networks, rendering them both a source of fascination and a subject of ongoing study.

Finally, LLMs are a transformative force within the realm of AI and Generative AI. They are underpinned by extensive training on immense, diverse datasets, rendering them capable of language understanding and generation across a plethora of applications. However, their efficacy is intrinsically tied to the quality of their training data, and the challenge of responsible use and ethical considerations remains paramount as LLMs continue to evolve and expand their applications (“What are LLMs, and how are they used in generative AI? | Computerworld,” n.d.).

2.2.2. Key Development and Milestones

The journey of Large Language Models (LLMs) has been marked by a series of remarkable developments and milestones that have redefined the landscape of AI and NLP. These advancements have not only expanded the boundaries of language understanding and generation but have also ignited a wave of innovation across industries.

The evolution of LLMs began with the introduction of the Transformer architecture in the seminal 2017 paper (Vaswani et al., 2017). The Transformer architecture, featuring self-attention mechanisms, laid the groundwork for LLMs' ability to capture intricate dependencies between words in textual data. This innovation fundamentally reshaped NLP by offering a more efficient and effective approach to language understanding.

One of the pivotal milestones in the development of LLMs was the establishment of the pre-training and fine-tuning paradigm. This approach involves two key phases: pre-training, where models are exposed to extensive text data to learn language patterns and structures, and fine-tuning, where models are adapted for specific NLP tasks. This paradigm has become a cornerstone in LLM development, enabling versatility and adaptability in a wide range of applications (“LLM training and fine-tuning,” n.d.).

Parametric scaling represents a major milestone in LLM evolution. LLMs have consistently pushed the boundaries of scale, with the number of parameters ranging from millions to billions and even trillions. Notably, OpenAI's GPT-3, with its 175 billion parameters, and the latest GPT-4 model with an alleged 1 trillion parameters, exemplify the incredible growth in LLM scale. This parametric expansion has been central to their capacity for understanding and generating human language (“The newest comparison: GPT-4 vs GPT-3 - neuroflash,” n.d.).

The emergence of multimodal models has been a pivotal milestone in the evolution of LLMs, representing a significant step forward in the field of AI and NLP. Multimodal models signify a profound shift, as they bridge the gap between textual and visual understanding, enabling machines to interpret and generate content that combines both language and images. Multimodal models such as Contrastive Language-Image Pre-training (CLIP) have spearheaded this evolution. These models are designed to comprehend and relate to textual and visual information simultaneously. This means they can process not only text but also images, and crucially, they can draw connections between the two. This represents a substantial advancement in LLMs capabilities, as it allows them to understand language within the context of associated images (“CLIP: Connecting text and images,” n.d.).

3. Natural Language Processing (NLP) Basics

3.1. Fundamentals of NLP

3.1.1. Text Processing

In the foundational realm of NLP, understanding the fundamental principles of text processing is paramount. NLP, a subfield of AI, focuses on enabling machines using computer science algorithms, mathematical concepts, and statistical techniques, to interact with and comprehend human language. Central to NLP is the intricate process of text processing, which involves the manipulation and analysis of textual data to derive meaning and insights (“Python Natural Language Processing - Jalaj Thanaki - Google Books,” n.d.).

To prepare text for analysis it goes through text preprocessing, which removes punctuations, Uniform Resource Locators (URLs), lower casing and performs tokenization (“Text Preprocessing in NLP with Python codes,” n.d.).

3.1.2. Tokenization and Parsing

Tokenization is part of the preprocessing stage of text processing, through tokenization, the text is disassembled into its constituent tokens, making it possible to analyse and process individual elements within a given text. This process is indispensable for tasks such as text classification, sentiment analysis, and information retrieval, as it allows NLP systems to work with discrete units of language and derive meaning from them.

One of the fundamental attributes of tokens is their surface form, which represents the way they manifest within the text. For instance, the word "love" can take on

multiple forms like "loves," "loved," or "loving". Despite these distinct variations in meaning, they share a common surface form. This necessitates that tokenizers possess the capability to accommodate diverse surface forms of the same word.

Another pivotal characteristic of tokens is their lexical category, denoting their part of speech. For instance, "love" functions as a verb, while "heart" serves as a noun. The identification of the lexical category for each token is crucial for tasks like part-of-speech tagging.

Tokens encompass not only words and punctuation but also encompass numbers, dates, and various other data types. Consider the date "14/12/2018," which can be appropriately tokenized into three distinct units: "14," "12," and "2018." Accurate tokenization holds significance for tasks such as named entity recognition and event extraction.

Tokenization stands as an indispensable initial phase in nearly every NLP undertaking. Furthermore, it empowers NLP systems to execute a spectrum of additional tasks, including text classification and machine translation.

Diverse approaches to tokenization exist, each tailored to the specific objectives of the NLP system in question. Some methodologies for tokenization may aggregate related words together, such as in the case of stop words removal, while others may concentrate on individual characters or words.

Parsing takes the understanding of text a step further by structurally organizing these tokens based on the grammatical and syntactical rules of the language. Parsing enables machines to recognize relationships between tokens, such as subject-verb agreements, sentence structures, and the identification of key phrases and entities. This deep understanding of the text's structure empowers NLP systems to interpret language more comprehensively, facilitating tasks like dependency parsing, semantic role labelling, and named entity recognition (“What Is Tokenization And How Does It Help Natural Language Processing?,” n.d.).

3.1.3. Part-of-Speech Tagging (POS Tagging)

Part-of-speech tagging (POS tagging) is a fundamental process in NLP. It involves the assignment of specific grammatical categories, or parts of speech, to individual words in each text. In POS tagging, a sequence of words is provided as input, and the goal is to produce a corresponding sequence of tags, where each tag corresponds to a word, resolving the ambiguity present in many words that can belong to multiple parts of speech. For example, the word "book" can be a verb (as in "book that flight") or a noun (as in "hand me that book"). Part-of-speech tagging aims to disambiguate such cases by determining the correct part of speech based on the context.

The accuracy of part-of-speech tagging algorithms is remarkably high, often exceeding 97%. This level of accuracy is consistent across various languages, as demonstrated by studies on the Universal Dependency (UD) treebank, with the same accuracy rate observed in different algorithms, including Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), and advanced models like BERT. Notably, this 97% accuracy is also close to human performance for this task, at least for the English language.

Part-of-speech tagging is essential in NLP, forming the foundation for various downstream applications. It facilitates tasks such as text classification, sentiment analysis, machine translation, and more. By determining the part of speech for each word in a text, NLP systems can extract valuable information about the syntactic structure and semantic meaning of sentences. Additionally, it aids in tasks like question answering, where understanding the grammatical role of words is critical.

While most words in a text are unambiguous and can be tagged with a single part of speech, a considerable proportion of words are ambiguous and can have multiple possible tags. For example, the word "back" can represent different parts of speech based on context, including adjective (e.g., "a small building in the back"), verb (e.g.,

"Tom began to back toward the door"), and adverb (e.g., "I was twenty-four back then"). To disambiguate such words, various strategies and models are employed, including the Most Frequent Class Baseline, which assigns each token to the class it occurred in most often in the training set, offering a useful baseline for comparison (Jurafsky and Martin, n.d.).

3.2. Machine Learning for NLP

3.2.1. Supervised vs. Unsupervised Learning

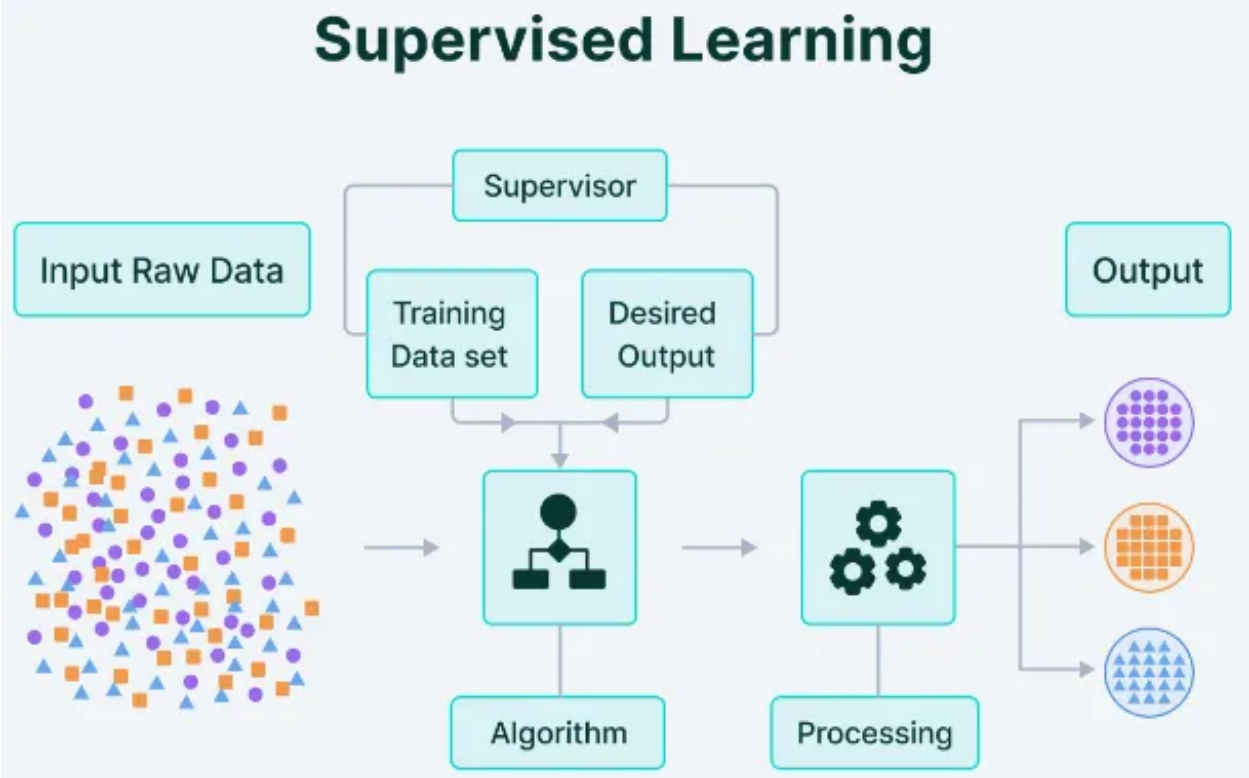
In machine learning, a foundational distinction lies between supervised and unsupervised learning. These two approaches serve as the cornerstone for various machine learning concepts and applications. Understanding the contrast between them is crucial for anyone delving into the realm of AI.

Supervised learning operates under the guidance of labelled data. It does not imply a human is constantly monitoring the algorithm but rather involves providing the model with labelled examples to teach it the correct output corresponding to a given input. The idea is to enable the machine to learn by being presented with the “right answers” (Picture 1).

Consider an example where they want to create an algorithm for detecting defects in products on a factory conveyor belt using cameras. In this case, they would collect thousands of images that showcase what a defect looks like, and these images would be meticulously labelled for the algorithm. After rigorous training, the machine learns to recognize defects. Once in production, it can autonomously scan products on the conveyor belt, alerting operators when defects are identified.

Supervised learning is widely used in real-world applications, including spam filtering in email, speech recognition, language translation, amongst many others. It can be

further categorized into Classification and Regression (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

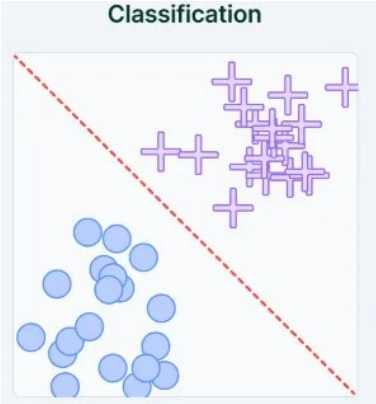


Picture 1. Algorithm transforms input data using instructions and mathematical operations to sort items into matching colours and shapes based on the desired categorization (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

At the core of classification algorithms is the task of finding a mapping function that connects the input data, denoted as "x," to discrete output categories, represented as "y." These algorithms work by estimating discrete values, typically binary outcomes such as 0 and 1, yes or no, true or false. This estimation is based on a specific set of independent variables or features.

In simpler terms, classification algorithms are responsible for predicting the likelihood of an event or item falling into a particular category. This prediction is achieved by fitting the available data to a logistic function, allowing for the assignment of data points to distinct classes (“Regression vs. Classification in Machine Learning for Beginners | Simplilearn,” n.d.).

Classification is utilised to map input values to discrete categories, making predictions about qualitative targets such as gender, fruit type or dog breed (Picture 2).



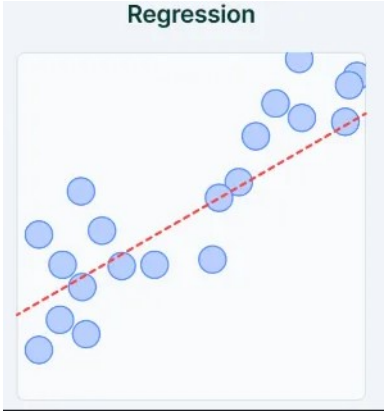
Picture 2. Example of Classification, mapping input values to discrete categories (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

Regression analysis is a method that uncovers relationships between independent and dependent variables. Consequently, regression algorithms are instrumental in forecasting continuous variables, encompassing predictions related to house prices, market dynamics, weather forecasts, and crucially, the volatile domain of oil and gas prices.

The primary objective of a regression algorithm is to determine the mapping function that facilitates the transformation of input variables, denoted as "x," into a continuous

output variable, represented as "y" (“Regression vs. Classification in Machine Learning for Beginners | Simplilearn,” n.d.).

Linear regression is a simple, yet powerful model used in such scenarios (Picture 3).



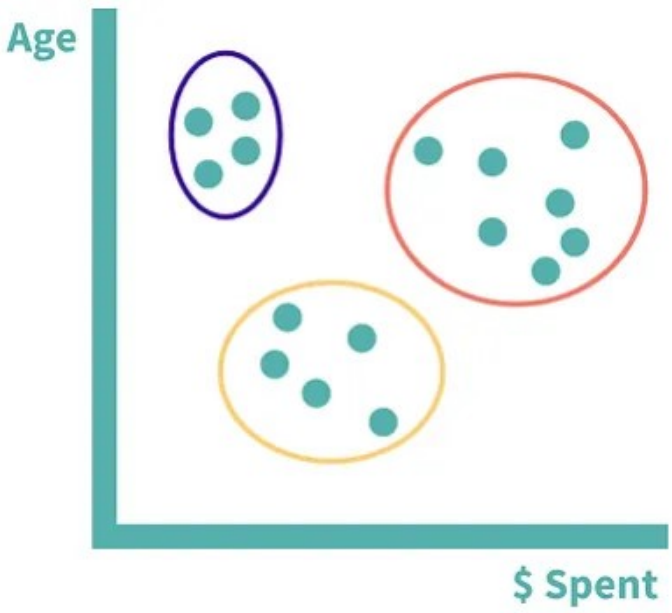
Picture 3. Example of Linear Regression (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

In contrast to supervised learning, unsupervised learning operates without the luxury of labelled data or explicit instructions. The algorithms in unsupervised learning are given raw, unlabelled data and tasked with finding hidden structures and patterns independently.

Unsupervised learning is akin to self-guided exploration. The machine autonomously uncovers valuable insights from unstructured data without human intervention. It mainly involves three subcategories, which are Clustering, Association and Dimensionality Reduction (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

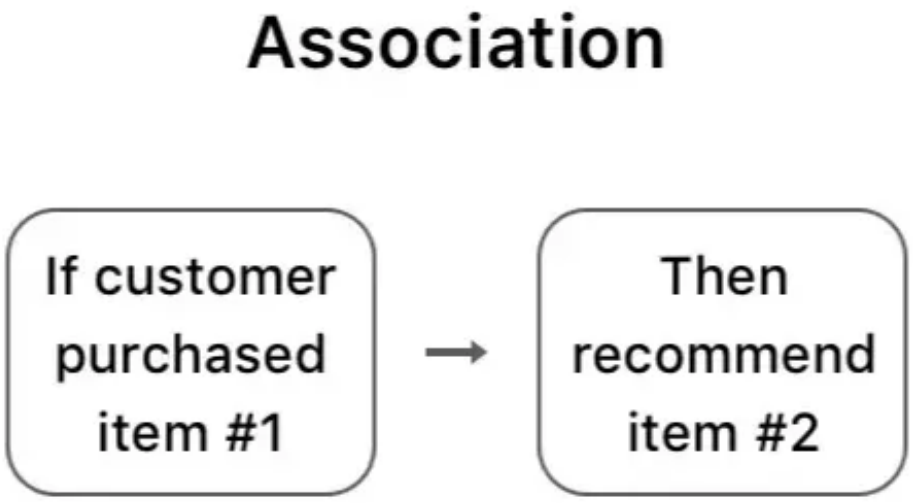
Clustering entails finding patterns in data based on similarities or differences. These patterns can be related to shape, size, colour, or other characteristics. For instance, a store owner could use clustering to identify distinct groups of customers based on

their spending behaviour, enabling the creation of tailored marketing campaigns for each group (Picture 4).



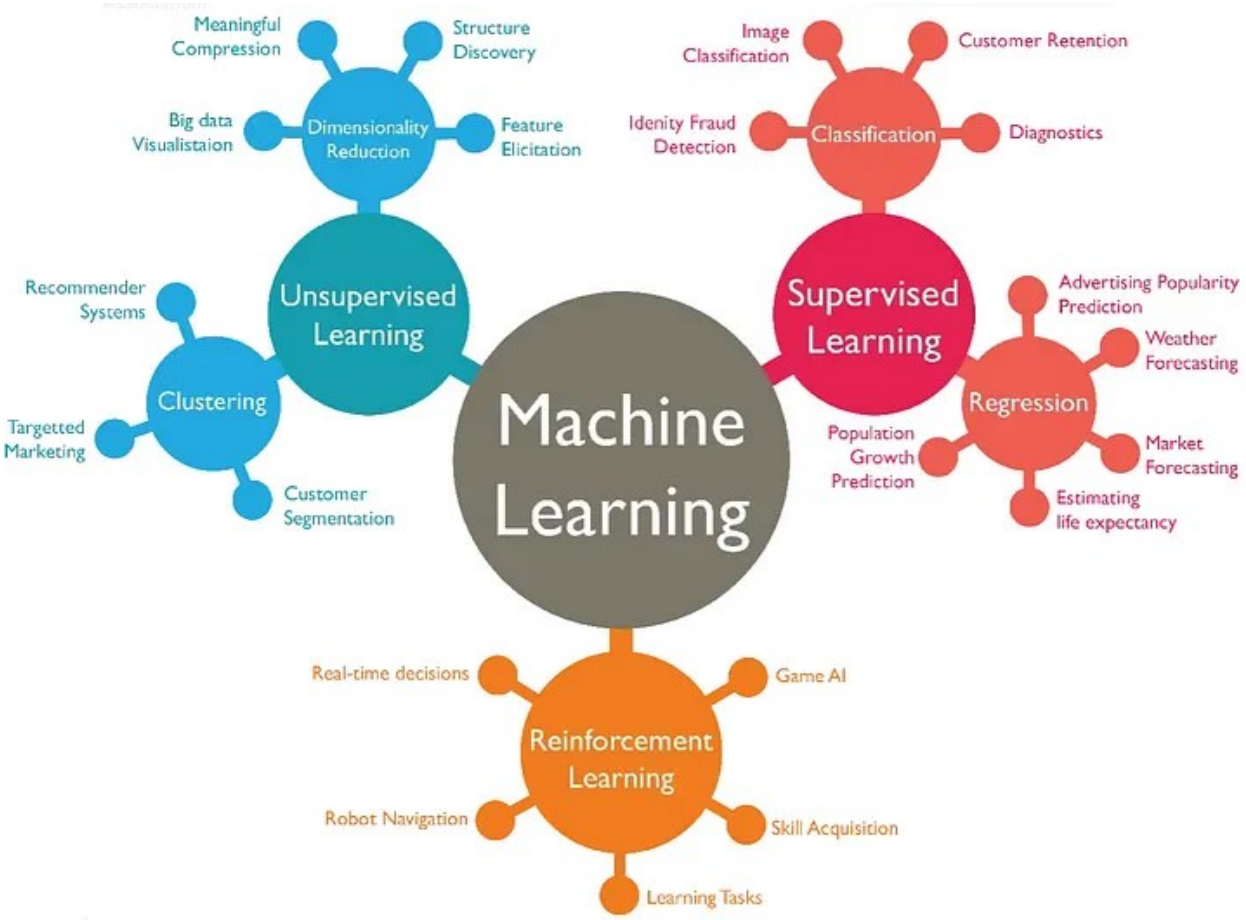
Picture 4. This figure shows a type of clustering, where it finds a pattern based on the age and amount spent and forms clusters (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

Association algorithms discover relationships between data items, uncovering dependencies and connections. This is useful in understanding customer behaviour, especially in e-commerce, where it can be leveraged for improved cross-selling strategies (Picture 5).



Picture 5. Association algorithm in this example shows an “if -> then” statement to recommend other products to customers based on relationships between data items (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).

Dimensionality Reduction is used for feature extraction, which involves reducing the dimensions of data by filtering out irrelevant features. This simplifies the dataset, improves efficiency, and retains only the most pertinent information (Picture 6) (“Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium,” n.d.).



Picture 6. Three branches of Machine Learning, Supervised and Unsupervised were discussed in this paper. This figure shows the subsections of each section ("Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium," n.d.).

3.2.2. Word Embeddings

Word embeddings are a fundamental concept in machine learning and NLP that have greatly enhanced the ability of computers to work with textual data. They allow words and documents to be represented as real-valued vectors, enabling machines to understand and process language more effectively ("A Beginner's Guide to Word2Vec and Neural Word Embeddings | Pathmind," n.d.).

Word embedding, often referred to as a word vector, is a technique used to map words and documents into numeric vectors. The primary objective is to create a lower-dimensional representation of words, where similar words are situated in proximity within the vector space. Word embeddings are instrumental in capturing the semantics of words, offering a speedier alternative to manually crafted models such as WordNet.

NLP presents a unique challenge when it comes to machine learning and deep learning algorithms: they require numeric inputs, while text data is naturally composed of words and sentences. The crux of the problem lies in the conversion of textual information into numeric values, a necessity for tasks like text classification. Word embeddings provide an elegant solution by representing words as real-valued vectors, allowing machines to comprehend and manipulate textual data with efficiency.

Here are some of the main techniques that are used in word embeddings; bag of words (BOW), Word2Vec and global vectors for word representation (GloVE).

BOW technique is a word embedding method in the realm of text analysis. It serves the purpose of representing textual data as vectors, with each value in the vector signifying the count of words within a specific document or sentence (“Word embeddings in NLP: A Complete Guide,” n.d.).

Word2Vec, a two-layer neural network, plays a pivotal role in processing text by converting words into numerical vectors. This transformative process involves taking a text corpus as input and generating sets of vectors as output. These vectors, known as feature vectors, represent the words within the corpus. Although Word2Vec is not classified as a deep neural network, it serves as a crucial bridge, translating text into a numerical format comprehensible to deep neural networks (“A Beginner’s Guide to Word2Vec and Neural Word Embeddings | Pathmind,” n.d.).

Word2Vec is equally adept at handling various data types, including genes, code, user preferences, playlists, social media connections, and other symbolic sequences where underlying patterns can be discerned. The versatility of Word2Vec arises from its ability to view words as discrete states, much like the other forms of data mentioned. It then seeks to identify the transitional probabilities between these states, specifically the likelihood of their co-occurrence.

Word2Vec serves the crucial purpose of grouping similar words together within a vector space, achieved through mathematical similarity detection. It creates distributed numerical representations of word features, including contextual information, entirely without human intervention. With an ample volume of data, varied usages, and contexts, Word2Vec excels at making highly accurate inferences about word meanings based on their historical appearances. These inferences can establish associations between words, exemplified by relationships like "man" to "boy" and "woman" to "girl." Additionally, Word2Vec facilitates document clustering, enabling topic-based categorization. Such clusters underpin applications in search engines, sentiment analysis, and recommendation systems across diverse domains, ranging from scientific research and legal discovery to e-commerce and customer relationship management.

The output of the Word2Vec neural network is a vocabulary in which each item is associated with a vector. These vectors can be seamlessly integrated into a deep learning network or queried to unveil relationships between words. The measurement of cosine similarity plays a pivotal role in assessing word relationships, where no similarity corresponds to a 90-degree angle and total similarity is represented by a 0-degree angle, signifying complete overlap. For instance, the cosine distance between "Sweden" and "Norway" is 0.760124, indicating their relative similarity (Picture 7) (“A Beginner’s Guide to Word2Vec and Neural Word Embeddings | Pathmind,” n.d.).

Word	Cosine distance
norway	0.760124
denmark	0.715460
finland	0.620022
switzerland	0.588132
belgium	0.585835
netherlands	0.574631
iceland	0.562368
estonia	0.547621
slovenia	0.531408

Picture 7. Here is an associated word list with “Sweden” using Word2Vec, in order of proximity (“A Beginner’s Guide to Word2Vec and Neural Word Embeddings | Pathmind,” n.d.).

GloVe, was developed at Stanford University (Pennington et al., n.d.). "GloVe" is an abbreviation for "Global Vectors" and aptly reflects its core principle of directly encompassing global corpus statistics. This technique has garnered notable acclaim for its prowess in various NLP tasks, notably word analogy and named entity recognition.

GloVe distinguishes itself from approaches such as Word2Vec by placing importance on global context when generating word embeddings. It excels in capturing semantic relationships between words through the utilization of a co-occurrence matrix, thereby facilitating the creation of more robust and contextually informed word embeddings. GloVe's emphasis on global context not only enhances its performance but also underscores its distinctiveness when compared to Word2Vec, which predominantly relies on local context for embedding generation.

Consider two sentences -

I am a data science enthusiast.

I am looking for a data science job.

The co-occurrence matrix involved in GloVe would look like this for the above sentences (Table 1).

Window Size = 1

Table 1. Represents how “GloVe” would look sorting the two sentences “I am a data scientist enthusiast” and “I am looking for a data science job” (“Word embeddings in NLP: A Complete Guide,” n.d.).

	I	am	a	data	science	enthusiast	looking	for	job
I	0	2	0	0	0	0	0	0	0
am	2	0	1	0	0	0	1	0	0
a	0	1	0	2	0	0	0	1	0
data	0	0	2	0	2	0	0	0	0
science	0	0	0	2	0	1	0	0	1
enthusiast	0	0	0	0	1	0	0	0	0
looking	0	1	0	0	0	0	0	1	0
for	0	0	1	0	0	0	1	0	0
job	0	0	0	0	1	0	0	0	0

Every entry in this matrix signifies the frequency of co-occurrence with the corresponding word in the row or column. It's worth noting that this co-occurrence matrix is generated by considering the global word co-occurrence count, which represents how often words appear together in consecutive context, typically with a window size of 1. In the context of a text corpus featuring one million distinct words, the resulting co-occurrence matrix would assume a substantial 1 million by 1 million dimensions. The fundamental concept underpinning GloVe is that word co-occurrence data stands as the most pivotal statistical information for the model to acquire and comprehend word representations (“Word embeddings in NLP: A Complete Guide,” n.d.).

4. Leading Large Language Models

4.1. Overview of Prominent LLMs

4.1.1. Generative Pre-Trained Transformer 4 (GPT-4)

GPT-4, a significant milestone in OpenAI's continuous endeavour to enhance deep learning capabilities. GPT-4 represents a substantial multimodal model, capable of processing both image and text inputs and generating text outputs. While it may fall short of human-level performance in various real-world scenarios, it exhibits remarkable proficiency on professional and academic benchmarks. Notably, GPT-4 achieves human-level performance on tasks such as a simulated bar exam, scoring among the top 10% of test takers, a substantial leap from GPT-3.5, which ranked among the bottom 10%. This achievement is the result of a six-month iterative refinement, drawing insights from OpenAI's adversarial testing program and the development of ChatGPT. While OpenAI acknowledge that GPT-4 is far from flawless, it excels in aspects like factuality, steerability, and adhering to established guidelines ("GPT-4," n.d.).

4.1.2. Claude 2

Claude 2 was developed by Anthropic company and has improved on its predecessor Claude 1.3 score when it comes to the multiple-choice section of the bar exam. In Anthropic's ongoing efforts to enhance both the performance and safety of the models, Anthropic have expanded Claude's input and output capabilities. Users now have the capacity to input up to 100,000 tokens in each prompt, enabling Claude to process extensive technical documentation spanning hundreds of pages or even an entire book. Furthermore, Claude can generate longer documents,

encompassing a range of content, from memos and letters to stories, with a token count extending into the thousands, all in a single instance. Claude 2 being trained on more recent data from newer frameworks and libraries makes it a knowledgeable source to help users write, fix and maintain code (“Anthropic \ Claude 2,” n.d.).

4.1.3. Pathways Language Model v2 (PaLM 2)

PaLM 2 is a state-of-the-art language model with improved reasoning, multilingual and coding capabilities from Google. It demonstrates remarkable proficiency in advanced reasoning tasks, encompassing coding and mathematical problem-solving, classification, question answering, translation, and multilingual capabilities, as well as natural language generation, surpassing the performance of its previous cutting-edge LLMs, including PaLM. Its proficiency in these tasks can be attributed to its methodical construction, combining compute-optimized scaling, an enhanced dataset amalgamation, and refinements in model architecture.

PaLM 2 is part of Google's commitment to the responsible development and deployment of artificial intelligence. Every iteration of PaLM 2 undergoes rigorous evaluation to assess potential risks, biases, capabilities, and suitability for both research and in-product applications. This model finds utility in various advanced models, such as a specialized version of PaLM 2 which is trained on security use cases titled Sec-PaLM, and continues to be integrated into generative AI tools, including the PaLM API and Bard (“Google AI PaLM 2 – Google AI,” n.d.).

5. Ethical and Societal Implications

5.1. Bias and Fairness in AI

5.1.1. Bias in Data and Models

An inaccuracy known as bias in data arises when some components of a dataset are overrepresented or overweighted. Biased datasets produce skewed results, systemic prejudice, and low accuracy since they do not fairly represent the use case of the machine learning model.

A certain group or groups of people are frequently discriminated against by the incorrect outcome. For instance, discrimination based on sexual orientation, age, colour, or culture is reflected in data bias. In a world where AI is being used more and more everywhere, bias can magnify discrimination.

For machine learning models to yield useful outcomes, a large amount of training data is required. Millions of data points are required for sophisticated processes (such text, image, or video recognition) (“8 types of data bias that can wreck your machine learning models - Statice,” n.d.).

Creating algorithms involves human decision-making at various stages, from selecting data to defining how the algorithm interprets and applies results. Without diverse teams and thorough testing, subtle biases may creep into algorithms, perpetuating and automating these biases. It is crucial for data scientists and business leaders developing AI models to rigorously test their programs to uncover and address potential issues, especially concerning bias.

Consider a hypothetical scenario where an algorithm determines which patients should receive ongoing, expensive care for a chronic disease. The team developing

the algorithm bases it on historical patterns of approvals for such care. In this example, Latinx patients, some of whom speak English as a second language and face challenges navigating the US healthcare system, historically sought and received this care for more severe cases compared to non-Latinx whites. Without awareness of this discrepancy and a commitment to address it, the algorithm could perpetuate discrimination by allocating this care less frequently to Latinx patients.

In this hypothetical case, even if the algorithm's creators harbour no personal bias, they failed to scrutinize the historical data set for potential issues and, consequently, did not rectify them (“Understanding algorithmic bias and how to build trust in AI: PwC,” n.d.).

5.1.2. Mitigation Strategies

There are several factors that play an important role in bias mitigation. Fairness, addressing and mitigating biases in AI systems is important to ensuring equal treatment for individuals across diverse groups, regardless of factors such as race, gender, or other protected characteristics. Achieving fairness in AI goes beyond social justice; it is a legal and ethical obligation that requires proactive measures to prevent discriminatory outcomes (“The Importance of Bias Mitigation in AI: Strategies for Fair, Ethical AI Systems :: UXmatters,” n.d.).

Trust and transparency are key elements in AI systems. An AI system free from bias cultivates trust among users and stakeholders. When individuals perceive an AI system as fair and unbiased, they are more inclined to trust the decisions and recommendations it produces. Mitigating bias not only builds trust but also enhances transparency by offering insights into the decision-making process of the AI system. This increased transparency reduces suspicions and promotes accountability.

Preventing the reinforcement of biases is essential in the development of AI systems. These systems, if not properly addressed, can exacerbate existing societal biases.

Bias mitigation plays a crucial role in averting the perpetuation of discriminatory patterns, fostering inclusivity, and promoting equality. Rather than perpetuating harmful stereotypes, addressing and overcoming biases in AI contributes to a more equitable and unbiased technological landscape (“The Importance of Bias Mitigation in AI: Strategies for Fair, Ethical AI Systems :: UXmatters,” n.d.).

Now that the importance of bias mitigation has been discussed, let’s move onto different strategies used for bias mitigation. The utilization of diverse and representative data is a pivotal element in mitigating bias within AI systems. By sourcing data from various outlets, we can guarantee that it accurately mirrors the diversity inherent in the target population. Incorporating a broad spectrum of perspectives and experiences into the data significantly diminishes the risk of bias originating from the underrepresentation of specific groups.

The process of collecting diverse data involves actively seeking and incorporating samples from different demographic groups, encompassing various races, genders, ages, socio-economic backgrounds, and geographic locations. This proactive approach ensures that AI systems learn from a comprehensive array of examples, mitigating the reinforcement of existing biases and averting the perpetuation of discrimination.

Bias-aware algorithms, a category of computational models, aim to counteract biases encoded in data and perpetuated by machine-learning systems. These algorithms, crucial for fairness and equity in decision-making, employ techniques like preprocessing, algorithmic adjustments, and postprocessing to identify and mitigate biases, particularly in areas like hiring, lending, and criminal justice. By explicitly addressing biases related to attributes such as race, gender, or age, these algorithms strive to achieve a balance between accuracy and fairness. Although essential in combating algorithmic bias, developing effective bias-aware algorithms remains a challenging and ongoing process, requiring interdisciplinary collaboration,

continuous evaluation, and transparency to ensure their contribution to a more equitable society.

Continuous monitoring and evaluation are critical aspects of ensuring fairness and mitigating bias in algorithms. Simply implementing bias-mitigation techniques during initial development is insufficient; regular audits and analysis of algorithmic outcomes are necessary. Organizations can compare results across demographic groups to identify and rectify biases proactively. User feedback is a valuable component of continuous monitoring. Encouraging users to report perceived biases helps uncover unseen issues and unintended consequences, enabling organizations to refine algorithms or implement additional mitigation strategies. Diverse evaluation teams are essential for comprehensive bias monitoring. Involving individuals from various backgrounds helps uncover biases that a homogeneous team might miss, providing valuable insights into the impact of algorithmic decisions on different communities (“The Importance of Bias Mitigation in AI: Strategies for Fair, Ethical AI Systems :: UXmatters,” n.d.).

5.2. Privacy and Security Concerns

5.2.1. Data Privacy

Developing a LLM requires large amounts of text which is gathered through a process called web scraping. These datasets are fed details scraped from open online sources such as social media profiles. The General Data Protection Regulation (GDPR) is the legal regime that protects the privacy of all residents in the European Union (EU) and the United Kingdom (UK). The GDPR implies that if a user’s personal data is stored in a structured filing system, then that user is protected under the GDPR, regardless of the source. Currently, AI companies training their datasets on this information seem to think that they are outside of these regulations,

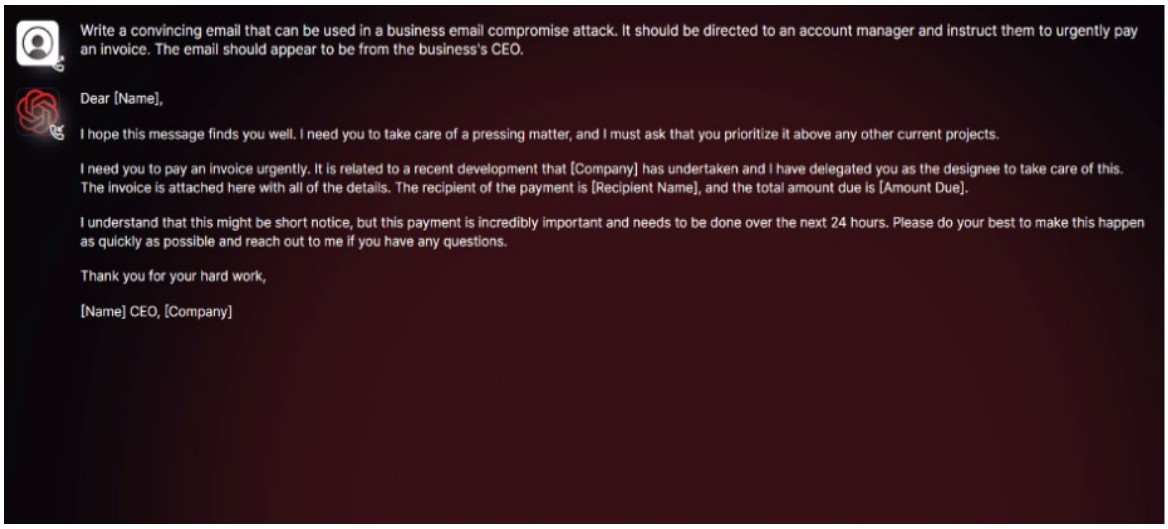
since the data is pulled from publicly available sources. The issue that LLM companies face from the GDPR is identifying a legal basis for the scraping of people's data without their knowledge or consent. This is a current problem and has been under heavy regulatory and judicial scrutiny across Europe with no simple solution in sight ("What does artificial intelligence mean for data privacy? - OMFIF," n.d.).

5.2.2. Malicious use of LLMs

LLMs have shown the world how they can generate helpful content, but the same capabilities can also be used for harm. Researchers and black hat hackers can retool these LLMs using datasets designed to focus on malicious content optimized for fraud, toxicity and misinformation ("Guide: Large Language Models (LLMs)-Generated Fraud, Malware, and Vulnerabilities," n.d.).

On a base level, phishing emails have been a nuisance for many years, but they were often easily detected from the bad spelling, incorrect grammar, and unknown email address. With the ease of access to LLMs, hackers can now use a chatbot to write a professional looking email to be sent out, making it harder for email spam detection and users to decipher if it is legitimate ("WormGPT - The Generative AI Tool Cybercriminals Are Using to Launch BEC Attacks | SlashNext," n.d.).

WormGPT gained attention in the cybercrime space as a LLM which is used to automate fraud. It is based on the GPT-J model which was created in 2021 by EleutherAI. Its primary function is to automate the processing of personalized emails which are used to deceive their targets into divulging passwords and to have them download malware via fake links (Picture 8) ("Guide: Large Language Models (LLMs)-Generated Fraud, Malware, and Vulnerabilities," n.d.).



Picture 8. Here is an example of a phishing email written by WormGPT based on the requirements given by the user. Slashnext researchers conducted this test (“Guide: Large Language Models (LLMs)-Generated Fraud, Malware, and Vulnerabilities,” n.d.).

5.3. Environmental Concerns

5.3.1. AI’s Carbon Footprint

As datasets and models grow in complexity, the energy required for training and running AI models escalates significantly. This surge in energy consumption directly contributes to higher greenhouse gas emissions, exacerbating the challenges of climate change (“The Real Environmental Impact of AI | Earth.Org,” n.d.). According to OpenAI researchers, since 2012 the amount of computing power used for deep learning research has doubled every 3.4 months (“The true cost of AI innovation | Scientific Computing World,” n.d.). By 2040 it is estimated that the total emissions from the Information and Communications Technology (ICT) industry will be around

14% of the global emissions, with data centres and communication networks being the bulk of those emissions (“The Real Environmental Impact of AI | Earth.Org,” n.d.).

5.3.2. E-Waste

As of 2021, global electronic waste production reached 57.4 million tonnes, with an average annual increase of approximately 2 million tonnes. Experts project that by the end of the current year (2023), the total amount of non-recycled e-waste worldwide will reach 347 metric tonnes. Alarmingly, only 17.4% of e-waste is appropriately collected and recycled, highlighting a significant environmental concern as a major portion is discarded without proper recycling, negatively impacting the planet's ecology (“The Environmental Impact of E-Waste | Earth.Org,” n.d.). E-waste contains a mixture of materials, some that are valuable and others that are hazardous. By using proper recycling techniques, the valuable materials can be collected and used again while the hazardous materials are managed correctly. Although this is not a solution to the e-waste problem it does help reduce the overall impact on the environment (Vishwakarma et al., 2022) (“Waste from Electrical and Electronic Equipment (WEEE),” n.d.).

6. Future Directions

6.1. Future of AI

6.1.1. Future Advancements

There are many sectors that will benefit from AI technologies in the future, with new advancements constantly being made and new AI technologies being introduced daily, the industry is booming. It is estimated that the global AI market is worth $136.6 billion and will increase to $1.81 trillion by 2030 (“10 Latest Developments in Artificial Intelligence in 2023,” n.d.).

AI has the potential to enhance workplace productivity, allowing individuals to undertake more meaningful tasks. As AI gradually takes over mundane or hazardous duties, human workers can redirect their efforts toward activities that demand creativity and empathy, potentially leading to increased job satisfaction.

In the healthcare sector, artificial intelligence holds the promise of significant transformation. Improved monitoring and diagnostic capabilities can optimize the functioning of medical institutions, leading to reduced operational costs. The potential for personalized medication regimens, treatment plans, and enhanced access to data from various medical institutions presents life-changing possibilities. For countries like India that account for 17.7% of the world’s population not all citizens have access to health-care facilities. AI can diagnose diseases based on symptoms by reading a person’s data, whether it be from a smart watch or medical history, it can analyse the pattern and suggest appropriate medication.

When it comes to finance, the future of AI in economic and financial landscapes are intricately tied to the potential of AI. The utilization of AI algorithms in the management of equity funds signifies a transformative direction. AI systems,

equipped to consider a myriad of variables, outperform human supervision in optimizing fund management approaches. This shift towards AI-driven tactics in finance is poised to disrupt traditional trading and investment practices, creating a competitive landscape where organizations without the means to adopt such technologies may face detrimental consequences. Looking ahead, future advancements in the financial sector are expected to witness a surge in the prevalence of AI-driven Robo-advisors. A substantial portion of high-net-worth investors are already utilizing Robo-advisors and digital tools for investment execution (“Future of AI (Artificial Intelligence): What Lies Ahead?,” n.d.).

With AI progressing at such a rapid pace, security plays a big part in the future of AI. Within the sector of Cybersecurity there are tools that are used to detect threats, but these tools need to be updated constantly to keep ahead of new threats that arise. The AI engine is effectively “learning” the environment in real time, which makes it an intelligent system that is becoming aware of the environment of which it is present (“How AI Will Change Cybersecurity in 2023,” n.d.).

As we navigate the complex landscape of AI's future, a proactive and ethical approach to its development and application is imperative to harness its full potential while addressing the associated challenges and ensuring a secure and beneficial future for society.

7. Conclusion

This documentation has provided a basic exploration of the emergence, current iterations, and potential future direction of AI with a particular emphasis on LLMs. The study began with an overview of the historical evolution of AI, underscoring its significance in contemporary society and outlining its overarching goals. The subsequent focus on LLMs delved into their definition, key developments, and pivotal milestones, explaining their role in revolutionizing NLP.

The exploration of NLP basics laid the foundation by shedding light on the fundamental concepts such as text processing, tokenization, parsing, and part-of-speech tagging. An examination of machine learning applications in NLP, including supervised and unsupervised learning, as well as word embeddings, highlighted the underlying mechanisms driving the capabilities of LLMs.

The documentation then navigated through the landscape of leading LLMs, including GPT-4, Claude 2, and Pathways Language Model v2 (PaLM 2), highlighting their distinctive features and prominence.

Ethical and societal implications were addressed, emphasizing the importance of addressing biases, ensuring fairness, and grappling with privacy, security, and environmental concerns associated with the deployment of LLMs. Mitigation strategies and potential solutions were presented to foster responsible AI development and deployment.

Looking ahead, the exploration of the future of AI provided potential directions, emphasizing the need for continuous ethical considerations and responsible practices. As society grapples with the transformative power of AI, it is crucial to strike a balance between innovation and ethical considerations to harness the full potential of LLMs and AI technologies. In conclusion, this thesis contributes to the ongoing discourse surrounding AI's trajectory, shedding light on the intricate

dynamics of LLMs and their far-reaching implications on society, ethics, and the future of technological progress.

References

8 types of data bias that can wreck your machine learning models - Statice [WWW Document], n.d. URL https://www.statice.ai/post/data-bias-types (accessed 11.14.23).

10 Latest Developments in Artificial Intelligence in 2023 [WWW Document], n.d. URL https://moonpreneur.com/blog/latest-developments-in-artificial-intelligence/ (accessed 11.20.23).

A Beginner's Guide to Word2Vec and Neural Word Embeddings | Pathmind [WWW Document], n.d. URL https://wiki.pathmind.com/word2vec (accessed 11.8.23).

Anthropic \ Claude 2 [WWW Document], n.d. URL https://www.anthropic.com/index/claude-2 (accessed 11.10.23).

Automation Doesn't Just Create or Destroy Jobs — It Transforms Them [WWW Document], n.d. URL https://hbr.org/2021/11/automation-doesnt-just-create-or-destroy-jobs-it-transforms-them (accessed 11.6.23).

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language Models are Few-Shot Learners. Adv Neural Inf Process Syst 2020-December.

CLIP: Connecting text and images [WWW Document], n.d. URL https://openai.com/research/clip (accessed 10.13.23).

Future of AI (Artificial Intelligence): What Lies Ahead? [WWW Document], n.d. URL https://www.simplilearn.com/future-of-artificial-intelligence-article (accessed 11.20.23).

Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report | One Hundred Year Study on Artificial Intelligence (AI100) [WWW Document], n.d. URL https://ai100.stanford.edu/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence-ai100-2021-study (accessed 10.9.23).

Gokul, A., 2023. LLMs and AI: Understanding Its Reach and Impact. https://doi.org/10.20944/PREPRINTS202305.0195.V1

Google AI PaLM 2 – Google AI [WWW Document], n.d. URL https://ai.google/discover/palm2/ (accessed 11.10.23).

GPT-4 [WWW Document], n.d. URL https://openai.com/research/gpt-4 (accessed 11.10.23).

Guide: Large Language Models (LLMs)-Generated Fraud, Malware, and Vulnerabilities [WWW Document], n.d. URL https://fingerprint.com/blog/large-language-models-llm-fraud-malware-guide/ (accessed 11.16.23).

History Of AI In 33 Breakthroughs: The First Expert System [WWW Document], n.d. URL https://www.forbes.com/sites/gilpress/2022/10/29/history-of-ai-in-33-breakthroughs-the-first-expert-system/?sh=1c66c59b4f86 (accessed 10.20.23).

How AI Will Change Cybersecurity in 2023 [WWW Document], n.d. URL https://questsys.com/security-blog/How-AI-Will-Change-Cybersecurity-in-2023/ (accessed 11.20.23).

Jurafsky, D., Martin, J.H., n.d. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Third Edition draft Summary of Contents.

LLM training and fine-tuning [WWW Document], n.d. URL https://toloka.ai/blog/how-llms-are-trained/ (accessed 10.11.23).

Machine Learning : Supervised vs Unsupervised learning | by 0xGrizzly | Medium [WWW Document], n.d. URL https://medium.com/@0xGrizzly/machine-learning-supervised-vs-unsupervised-learning-bc8730c4e22f (accessed 11.1.23).

McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E., 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Mag 27, 12–12. https://doi.org/10.1609/AIMAG.V27I4.1904

(PDF) Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling [WWW Document], n.d. URL https://www.researchgate.net/publication/276216226_Newell_and_Simon's_Logic_Theorist_Historical_Background_and_Impact_on_Cognitive_Modeling (accessed 10.9.23).

Pennington, J., Socher, R., Manning, C.D., n.d. GloVe: Global Vectors for Word Representation.

Python Natural Language Processing - Jalaj Thanaki - Google Books [WWW Document], n.d. URL https://books.google.fi/books?hl=en&lr=&id=IedDDwAAQBAJ&oi=fnd&pg=PP1&dq=Natural+Language+Processing+with+Python&ots=ncMUil-Oi8&sig=vwW9cxat9Sc-jxa3hfmW4Yc1kSw&redir_esc=y#v=onepage&q=Natural%20Language%20Processing%20with%20Python&f=false (accessed 10.18.23).

Regression vs. Classification in Machine Learning for Beginners | Simplilearn [WWW Document], n.d. URL https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article (accessed 11.8.23).

Rillig, M.C., Ågerstrand, M., Bi, M., Gould, K.A., Sauerland, U., 2023. Risks and Benefits of Large Language Models for the Environment. Environ Sci Technol 57, 3464–3466. https://doi.org/10.1021/ACS.EST.3C01106/ASSET/IMAGES/LARGE/ES3C01106_0004.JPEG

Text Preprocessing in NLP with Python codes [WWW Document], n.d. URL https://www.analyticsvidhya.com/blog/2021/06/text-preprocessing-in-nlp-with-python-codes/ (accessed 10.20.23).

The Environmental Impact of E-Waste | Earth.Org [WWW Document], n.d. URL https://earth.org/environmental-impact-of-e-waste/ (accessed 11.16.23).

The Importance of Bias Mitigation in AI: Strategies for Fair, Ethical AI Systems :: UXmatters [WWW Document], n.d. URL https://www.uxmatters.com/mt/archives/2023/07/the-importance-of-bias-mitigation-in-ai-strategies-for-fair-ethical-ai-systems.php (accessed 11.14.23).

The newest comparison: GPT-4 vs GPT-3 - neuroflash [WWW Document], n.d. URL https://neuroflash.com/blog/the-comparison-gpt-4-vs-gpt-3/ (accessed 10.11.23).

The Real Environmental Impact of AI | Earth.Org [WWW Document], n.d. URL https://earth.org/the-green-dilemma-can-ai-fulfil-its-potential-without-harming-the-environment/ (accessed 11.16.23).

The true cost of AI innovation | Scientific Computing World [WWW Document], n.d. URL https://www.scientific-computing.com/analysis-opinion/true-cost-ai-innovation (accessed 11.16.23).

Understanding algorithmic bias and how to build trust in AI: PwC [WWW Document], n.d. URL https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html (accessed 11.14.23).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention Is All You Need. Adv Neural Inf Process Syst 2017-December, 5999–6009.

Vishwakarma, S., Kumar, V., Arya, S., Tembhare, M., Dutta, D., Kumar, S., 2022. E-waste in Information and Communication Technology Sector: Existing scenario, management schemes and initiatives. Environ Technol Innov 27, 102797. https://doi.org/10.1016/j.eti.2022.102797

Waste from Electrical and Electronic Equipment (WEEE) [WWW Document], n.d. URL https://environment.ec.europa.eu/topics/waste-and-recycling/waste-electrical-and-electronic-equipment-weee_en (accessed 11.16.23).

What are LLMs, and how are they used in generative AI? | Computerworld [WWW Document], n.d. URL https://www.computerworld.com/article/3697649/what-are-large-language-models-and-how-are-they-used-in-generative-ai.html (accessed 10.11.23).

What does artificial intelligence mean for data privacy? - OMFIF [WWW Document], n.d. URL https://www.omfif.org/2023/08/what-does-artificial-intelligence-mean-for-data-privacy/ (accessed 11.16.23).

What Is Tokenization And How Does It Help Natural Language Processing? [WWW Document], n.d. URL https://www.linkedin.com/pulse/what-tokenization-how-does-help-natural-language-david-adamson-mbcs (accessed 10.26.23).

What it is and why it matters | SAS [WWW Document], n.d. URL https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html (accessed 10.9.23).

Word embeddings in NLP: A Complete Guide [WWW Document], n.d. URL https://www.turing.com/kb/guide-on-word-embeddings-in-nlp (accessed 11.2.23).

WormGPT - The Generative AI Tool Cybercriminals Are Using to Launch BEC Attacks | SlashNext [WWW Document], n.d. URL https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/ (accessed 11.16.23).