



Article

The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field

Hossein Hassani ^{1,*} and Emmanuel Sirmal Silva ²

¹ The Research Institute of Energy Management and Planning (RIEMP), University of Tehran, Tehran 19395-4697, Iran

² Fashion Business School, London College of Fashion, University of the Arts London, London W1G 0BJ, UK

* Correspondence: hassani.stat@gmail.com

Abstract: ChatGPT, a conversational AI interface that utilizes natural language processing and machine learning algorithms, is taking the world by storm and is the buzzword across many sectors today. Given the likely impact of this model on data science, through this perspective article, we seek to provide an overview of the potential opportunities and challenges associated with using ChatGPT in data science, provide readers with a snapshot of its advantages, and stimulate interest in its use for data science projects. The paper discusses how ChatGPT can assist data scientists in automating various aspects of their workflow, including data cleaning and preprocessing, model training, and result interpretation. It also highlights how ChatGPT has the potential to provide new insights and improve decision-making processes by analyzing unstructured data. We then examine the advantages of ChatGPT's architecture, including its ability to be fine-tuned for a wide range of language-related tasks and generate synthetic data. Limitations and issues are also addressed, particularly around concerns about bias and plagiarism when using ChatGPT. Overall, the paper concludes that the benefits outweigh the costs and ChatGPT has the potential to greatly enhance the productivity and accuracy of data science workflows and is likely to become an increasingly important tool for intelligence augmentation in the field of data science. ChatGPT can assist with a wide range of natural language processing tasks in data science, including language translation, sentiment analysis, and text classification. However, while ChatGPT can save time and resources compared to training a model from scratch, and can be fine-tuned for specific use cases, it may not perform well on certain tasks if it has not been specifically trained for them. Additionally, the output of ChatGPT may be difficult to interpret, which could pose challenges for decision-making in data science applications.

Keywords: ChatGPT; data science; synthetic data; natural language processing; opportunities; challenges; bias; plagiarism; programming; data analysis; ethics



Citation: Hassani, H.; Silva, E.S. The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data Cogn. Comput.* **2023**, *7*, 62. <https://doi.org/10.3390/bdcc7020062>

Academic Editors: Domenico Ursino, Miguel-Angel Sicilia, Nik Bessis and Marcello Trovati

Received: 28 February 2023

Revised: 17 March 2023

Accepted: 21 March 2023

Published: 27 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The continued growth of big data has led to rapid growth in the importance of data science, with data transforming themselves into the most valuable and only inimitable asset for any organization [1,2]. Data science, which involves collecting, analyzing, and interpreting data to extract insights and improve decision-making processes, is defined as “the scientific study of the creation, validation and transformation of data to create meaning [3]. With the exponential growth of big data come significant challenges in terms of managing and processing the vast amount of data. Therefore, data scientists are constantly seeking ways to improve their workflows and streamline their processes to maximize the benefits of data science.

The data science field is rapidly evolving, driven by advancements in technology and an increasing demand for data-driven insights. Recent trends in the field include the growing popularity of machine learning and artificial intelligence, the rise of big data and cloud

computing, and the increasing use of data visualization tools. Data scientists are facing numerous challenges, including data privacy concerns, the need for increased transparency and explainability in machine learning models, and the difficulty of integrating disparate data sources. In addition, the field is facing a shortage of skilled professionals, making it difficult for organizations to find the talent that they need to effectively utilize data science techniques. As the field continues to evolve, it will be important for data scientists to stay up-to-date on the latest trends and developments and to prioritize ethical considerations in their work.

One promising technology that can assist data scientists in their work is OpenAI's Chat Generative Pre-Trained Transformer [4], a conversational AI interface that uses natural language processing to understand and respond to human queries. ChatGPT is based on deep learning algorithms that enable it to generate high-quality responses to a wide range of queries. It achieves this by utilizing natural language processing and machine learning algorithms to generate text that is both fluent and contextually appropriate (see [5] for a concise explanation on the model behind the bot). Whilst it was only launched on 30 November 2022, by March 2023, ChatGPT had succeeded in making it to the TIME magazine cover [6]—giving an indication of its popularity and perceived impact on society. As Thorp (2023) eloquently puts, ChatGPT has become a cultural sensation [7].

It was in 1950 that Alan Turing devised what later became known as the Turing test, which assesses whether a computer could really “think” [7]. Whilst AI systems are yet to be able to think in a way comparable to the human brain, they are easily passing the Turing test today. Reports indicate that Google's chatbot convinced its engineer that it had reached sentience and thus resulted in the engineer being fired and having itself locked behind closed doors until ethicists figured out how to make it safer [7]. In contrast, OpenAI's ChatGPT made it to the public domain, becoming “a very big deal” [8]. This is because, despite its popularity, historically, AI only worked for applications outside of data analysis [8]. By making ChatGPT accessible to the masses, OpenAI has extended the use of AI beyond problems where failure is expensive, thereby opening a new world of applications [8], and with it the potential for abuse as well [9].

This perspective article aims to explore the opportunities and challenges associated with the use of ChatGPT in data science and examines how it can be used to improve various tasks in the field. In addition, the paper is also a quick reference guide for stakeholders interested in learning about the different use cases of ChatGPT in data science, and we hope that this article will prompt more humans to experiment with AI and its capabilities.

That said, ChatGPT presents its own challenges and limitations. Some challenges are associated with its use, mainly around inaccuracies, privacy, bias, and plagiarism [10,11]. One of the greatest concerns is the potential for bias [12,13], as ChatGPT may replicate and reinforce existing biases in the data that it is trained on, thus leading to incorrect and unfair predictions. Additionally, there is a risk of plagiarism if users simply copy and paste information generated by ChatGPT, without proper citations or acknowledgement of its use (see [7,14]). Recently, Check Point Research [15] published information on how ChatGPT is making it easier to engage in cybercrimes too. ChatGPT also presents some limitations in relation to the field of data science. As Marr [16] asserts, at present, these limitations include frequent mistakes, the fact that it is only able to generate text (as opposed to charts and graphs), and information input is limited to text (for example, one cannot upload spreadsheets of data). It is noteworthy that some of the limitations highlighted here are being addressed via its latest release, GPT-4. Nevertheless, it is a black box model that cannot explain how its output was generated [17].

Despite these challenges, ChatGPT is good for the data science community [18]. It can be used to create code to automate processes around data gathering, formatting, or cleansing; define data structures; guide us on what infographics should be produced and what information they should entail; create training material; identify data sources required for particular tasks; create synthetic data; give advice on compliance, regulation, and practical steps that can be taken to ensure that data operations are legal, unbiased,

and ethical; and help to identify analytical processes leading to best practices [16]. As the technology continues to evolve, it is likely that it will become an increasingly important tool in the field of data science.

The remainder of this paper is organized such that Section 2 presents a concise background on what ChatGPT is, its evolution, and how it works. Section 3 expands on how ChatGPT could be used in the context of data science. Section 4 covers its potential to assist programmers, whilst Section 5 seeks to outline the future of ChatGPT in data science. Section 6 concisely refers to its tendency to make errors, whilst Section 7 offers specific suggestions on how universities and data science programs can integrate ChatGPT training in a manner that emphasizes ethics and integrity. The article concludes in Section 8.

2. What Is ChatGPT?

2.1. A Short Overview on Its Structure

ChatGPT is a deep neural network architecture based on the Transformer model, and as a generative model, it can generate new text based on the input it receives. The model is pre-trained on large amounts of text data using a process called unsupervised learning, which allows it to learn the underlying patterns and structures in the language.

The model consists of multiple layers of self-attention and feed-forward neural networks, which enable it to effectively capture the dependencies and relationships between words in a sentence. The self-attention mechanism allows the model to focus on different parts of the input sequence when generating output, which is particularly useful for natural language processing tasks.

During pre-training, the model is trained to predict the next word in a sequence of text based on the previous words. This task is known as language modeling, and it allows the model to learn a high-quality representation of the language. Once the model is pre-trained, it can be fine-tuned on specific natural language processing tasks such as language translation, sentiment analysis, and text classification.

One of the key advantages of ChatGPT is its ability to generate coherent and contextually appropriate responses to text inputs, even for open-ended prompts such as chatbot conversations. This is achieved by using the pre-trained model to generate a probability distribution over the next word in the sequence, and sampling from this distribution to generate the output. By repeatedly generating the next word based on the previous words, the model can produce fluent and coherent text.

In summary, ChatGPT is a deep neural network architecture based on the Transformer model, designed for natural language processing tasks. Its ability to generate coherent and contextually appropriate text has made it a popular tool for applications such as chatbots, language translation, and text generation.

2.2. More on ChatGPT

ChatGPT is a state-of-the-art language model that utilizes deep learning techniques to generate human-like text [19]. It is a product of OpenAI, a research organization dedicated to advancing artificial intelligence and developing cutting-edge technologies that benefit society. The core of ChatGPT's architecture is a Transformer, a neural network architecture that enables the model to analyze sequences of data, such as text. The Transformer was introduced in 2017 and has since revolutionized the field of natural language processing (NLP) [20].

One of the key advantages of ChatGPT is its ability to be fine-tuned for a wide range of language-related tasks. It has been shown that fine-tuned language models can be continual learners, thereby giving us an indication of the future capabilities of innovations such as ChatGPT [21]. By training the model on a specific task with additional data, such as text classification or machine translation, the model can adapt to new domains and perform well on various NLP tasks.

ChatGPT is pre-trained on vast amounts of data (text), primarily sourced from the internet, such as web pages, articles, books, and social media platforms, in several

languages [22]. The pre-training process involves predicting the next word in a sentence, given the preceding words [8]. This allows the model to develop a deep understanding of the structure and meaning of language, which is then used to generate text that is coherent and contextually relevant.

The output from ChatGPT can be used for a wide range of applications, including text summarization, sentiment analysis, language translation, and question answering [23]. Its ability to generate highly accurate and contextually relevant text has made it popular in the fields of chatbots and virtual assistants [24]. With the development of more advanced language models such as ChatGPT, the future of NLP is expected to bring about more innovative and efficient ways to communicate and interact with machines. To date, OpenAI has released several versions of ChatGPT, such as GPT-2, GPT-3, and GPT-4 (as of 14 March 2023), which differ in their size, number of parameters, and the number of languages included in their pre-training data. For instance, GPT-3, which took the public by storm, was trained on a diverse set of texts in 95 languages [22] and with 175 billion parameters [25].

2.3. Some Applications

Here, we provide some examples or case studies that illustrate how ChatGPT has been used in data science to automate various aspects of the workflow and to analyze unstructured data.

Chatbots: ChatGPT has been used to develop chatbots that can interact with customers and provide automated customer support. For example, the National Australia Bank (NAB) developed a chatbot named “Mia” using ChatGPT. Mia is capable of answering customer queries related to banking services, such as the account balance, transaction history, and credit card details.

Data Augmentation: ChatGPT has also been used for data augmentation, which is the process of generating new data samples from existing data to improve the performance of machine learning models. For example, researchers at the University of California, San Diego used ChatGPT to generate synthetic radiology reports, which were then used to augment the training data for a machine learning model for radiology report classification.

Text Generation: ChatGPT has been used for text generation, which involves generating new text based on a given prompt or context. For example, researchers at the University of Washington used ChatGPT to generate news articles from headlines. The generated articles were found to be indistinguishable from articles written by humans.

Text Summarization: ChatGPT has been used for text summarization, which involves generating a shorter summary of a longer piece of text. For example, researchers at IBM used ChatGPT to generate summaries of news articles. The generated summaries were found to be comparable in quality to human-written summaries.

Language Translation: ChatGPT has also been used for language translation, which involves translating text from one language to another. For example, researchers at Google used ChatGPT to develop a machine translation system for the language pair of English to French. The system was found to achieve state-of-the-art performance on several benchmark datasets.

Overall, ChatGPT has been used in a wide range of data science applications to automate various aspects of the workflow and to analyze unstructured data. Its ability to generate coherent and contextually appropriate text has made it a popular tool for applications such as chatbots, text generation, and language translation.

2.4. Generating Synthetic Data Using ChatGPT

Generating synthetic data using ChatGPT involves training the language model on a large corpus of text data, and then using it to generate new synthetic data based on the patterns and structures it has learned from the training data. This can be done by providing a prompt or a seed text to the model, which it then uses to generate new text that is similar in style and content to the original data.

One of the main advantages of using synthetic data is that it can help to address the problem of data scarcity, which is a common issue in many machine learning applications. By generating new data that are similar to the original data, machine learning models can be trained on a larger and more diverse dataset, which can lead to better performance and generalization.

In addition, synthetic data can also be used to augment existing datasets, by adding new examples or variations of existing examples. This can be particularly useful in applications where the dataset is limited or biased, and where adding more data can help to reduce overfitting and improve model accuracy.

However, there are also some limitations and challenges associated with using synthetic data in machine learning. One of the main challenges is ensuring that the synthetic data are representative of real-world data, and that they capture the same patterns and structures that are present in the original data. This can be particularly difficult in applications where the data are complex or multifaceted, and where the underlying patterns are not well understood.

Another challenge is ensuring that the synthetic data do not introduce any biases or artifacts that could impact the performance or fairness of the machine learning model. For example, if the synthetic data are generated based on a biased or incomplete dataset, this could lead to a model that is biased or inaccurate in certain contexts.

Despite these challenges, the use of synthetic data is becoming increasingly popular in many machine learning applications, particularly in areas such as computer vision and natural language processing. By combining synthetic data with real-world data, machine learning models can be trained on larger and more diverse datasets, which can help to improve their accuracy and robustness in real-world applications.

2.5. Comparing ChatGPT to Other Similar Applications

Comparing ChatGPT to other similar applications is an important step in evaluating its strengths and weaknesses. When it comes to natural language processing, there are several other popular language models that are commonly used, such as OpenAI's GPT-2 and GPT-3, Google's BERT, and Facebook's RoBERTa.

One key advantage of ChatGPT is its flexibility and ease of use. ChatGPT can be fine-tuned for a wide range of natural language processing tasks, from language generation to text classification and question answering, with relatively few adjustments to the base model architecture. This makes it a versatile tool that can be adapted to a variety of use cases.

In terms of performance, ChatGPT has been shown to be competitive with other state-of-the-art language models, particularly for generating natural language responses. For example, a recent study by Microsoft Research found that ChatGPT outperformed other language models, including GPT-3 and BERT, on a range of language generation tasks, such as summarization and paraphrasing.

However, one potential limitation of ChatGPT is its reliance on large amounts of training data. While the model can be fine-tuned on smaller datasets, its performance may be limited compared to other models that are designed specifically for low-resource settings, such as Google's ALBERT.

Another potential limitation of ChatGPT is its ability to generate coherent and contextually appropriate responses. While the model has shown impressive performance in generating natural language responses, there are still cases where the generated responses may be nonsensical or inappropriate, particularly when dealing with complex or nuanced language.

Table 1 provides a comparison of popular language models used for natural language processing, including ChatGPT, GPT-2, GPT-3, BERT, and RoBERTa. ChatGPT is highlighted for its flexibility, ease of use, and competitive performance, but its potential for nonsensical responses on small datasets is noted. GPT-2 is highlighted for its high-quality language generation, while GPT-3 is noted for its strong performance on a wide range of NLP

tasks but with limited interpretability and potential for bias and ethical concerns. BERT is known for strong performance on text classification and question answering tasks, but has limited performance on language generation tasks. RoBERTa is highlighted for its high-quality language representation and strong performance on various NLP tasks, but also has limited interpretability and potential for bias and ethical concerns. Overall, ChatGPT’s flexibility, ease of use, and competitive performance make it a valuable tool for a range of natural language processing applications. However, as with any language model, it has its limitations and should be evaluated carefully in the context of specific use cases and performance requirements.

Table 1. Comparison of popular language models for natural language processing.

Model	Year	Training Data Size	Strengths and Weaknesses		
			Strengths	Weaknesses	Fine-Tuning Required
ChatGPT	2020	1.5 B	Flexibility, ease of use, competitive performance	Limited performance on small datasets, potential for nonsensical responses	Yes
GPT-2	2019	1.5 B	High-quality language generation, strong performance on text completion tasks	Limited performance on other NLP tasks	Yes
GPT-3	2020	Up to 175 B	High-quality language generation, strong performance on a wide range of NLP tasks	Limited interpretability, potential for bias and ethical concerns	Yes
BERT	2018	340 M–1.1 B	Strong performance on text classification and question answering tasks	Limited performance on language generation tasks	Yes
RoBERTa	2019	125 M–2.6 B	High-quality language representation, strong performance on a range of NLP tasks	Limited interpretability, potential for bias and ethical concerns	Yes

Note that the data presented in the Table 1 are subject to change over time as new models and updates are developed. Additionally, the strengths and weaknesses of each model may vary depending on the specific use case and performance requirements.

3. The Use of ChatGPT in Data Science

Hassani et al. [26,27] when they referred to “unicorn data scientists” as “a rare breed, almost mythical creatures that are experts in multiple specialties, from mathematics to computer science and artificial intelligence (AI)” (p. 1). The emergence of ChatGPT could be bringing us closer to such mythical creatures for several reasons. First, by leveraging the power of ChatGPT, data scientists can automate various aspects of their workflow, such as data cleaning and preprocessing, model training, and result interpretation. Second, ChatGPT is showing its potential for analyzing unstructured data such as customer feedback, social media data, and online reviews to uncover new insights and improve decision-making processes. Third, the automation prospects provided by ChatGPT in relation to the process of data analysis and natural language processing can enable data scientists to focus on more complex tasks, such as developing more accurate predictive models and enhancing data visualization. Fourth, its ability to generate synthetic data also makes it a valuable resource for data scientists who are working with limited or incomplete datasets. Fifth, we no longer need extensive education in mathematics, computer science, or AI to generate a code or program that can be used to solve real-world problems. Therefore, ChatGPT has the potential to revolutionize the field of data science, making it more accessible, efficient, and effective through its influence on data analysis, predictive modeling, and language translation.

As summarized above, the use of ChatGPT in data science has numerous applications [28]. One of the most significant advantages of ChatGPT is its ability to generate synthetic data that can be used for training machine learning models [29,30]. By generating large amounts of synthetic data, data scientists can improve the performance of their models, reduce the time and cost of data collection, and avoid issues related to privacy and security [31–33]. In addition to generating synthetic data, ChatGPT can assist data scientists in various other ways too. For example, ChatGPT can be used to perform text mining and natural language processing tasks, such as sentiment analysis [34,35] and entity recognition [34]. ChatGPT can also be integrated with chatbots and other conversational interfaces to provide more natural and intuitive interactions with users. In a world where time is money, ChatGPT presents the ability to save a significant amount of time associated with data science tasks—time and resources that businesses can now devote to other important tasks [17].

Below, we present a few examples that showcase the statistical capabilities of ChatGPT.

1. Language modeling: ChatGPT is trained on massive amounts of text data, which enables it to generate highly accurate and coherent text that mimics human conversation [16].
2. Text classification: ChatGPT can be fine-tuned for text classification tasks such as sentiment analysis, spam detection, and topic classification. The model can learn to classify text based on patterns and associations within the training data (e.g., see [4]).
3. Named entity recognition (NER): ChatGPT can recognize and extract information about named entities in text, such as people, places, and organizations [36]. This is useful for tasks such as information extraction and text summarization.
4. Machine translation: ChatGPT can also be fine-tuned for machine translation tasks, enabling it to translate text from one language to another [37]. While it may not be as accurate as dedicated machine translation models, ChatGPT can provide a useful baseline for machine translation tasks.
5. Question answering: ChatGPT can be fine-tuned for question answering tasks, enabling it to answer questions based on a given context. This is useful for tasks such as customer service and chatbot interactions.
6. Text generation: ChatGPT can generate text that is coherent, contextually relevant, and mimics human language. This capability can be used for tasks such as text completion, story generation, and creative writing.

Overall, ChatGPT's statistical capabilities make it a versatile and powerful tool for NLP tasks, with the potential to transform the way in which we interact with machines and process natural language.

4. Using ChatGPT to Assist in Programming

Doerrfeld [38] states that software experts are in agreement that generative AI is the future of programming. Therefore, ChatGPT can be a valuable resource for programmers looking to improve their skills [39] or for those seeking assistance with coding challenges [40]. In fact, ChatGPT enables those who have never programmed before to create workable code [17]. ChatGPT can be helpful by providing explanations and examples of programming concepts, such as syntax, data structures, algorithms, and more. For example, a programmer might ask ChatGPT a question such as "What is the difference between a linked list and an array?" ChatGPT could provide a clear and concise answer, along with an example code snippet demonstrating how each data structure is used in practice.

Suppose that we have a programmer who is curious about the differences between a linked list and an array. They could ask ChatGPT for an explanation and example code snippet by posting the following question:

Q: What **is** the difference between a linked **list and** an array?

ChatGPT could then provide a clear and concise answer to their question:

A: A linked **list and** an array are both linear data structures, but they differ **in** how they store **and access** data.

An array stores elements **in** contiguous memory locations, making it easy to **access** elements by index but difficult to insert **or** delete elements **in** the middle of the array.

A linked **list**, on the other hand, stores elements as nodes that contain a value **and** a pointer to the **next** node **in** the **list**. This makes it easy to insert **or** delete elements **in** the middle of the **list** but can be slower to **access** elements by index.

Here **is** an example code snippet demonstrating how each data structure **is** used **in** practice:

Array example

```
my_array = [1, 2, 3, 4, 5]
print(my_array[2]) # prints 3
```

Linked **list** example

```
class Node:
def init(self, data):
self.data = data
self.next = None
```

```
class LinkedList:
def init(self):
self.head = None
```

python code

```
def insert(self, data):
new_node = Node(data)
if self.head is None:
self.head = new_node
else:
current = self.head
while current.next is not None:
current = current.next
current.next = new_node
```

```
def delete(self, data):
if self.head is None:
return
```

```
if self.head.data == data:
self.head = self.head.next
return
```

```
current = self.head
while current.next is not None:
if current.next.data == data:
current.next = current.next.next
return
current = current.next
```

```
def print_list(self):
current = self.head
while current is not None:
print(current.data)
current = current.next
create a linked list and print its contents
my_list = LinkedList()
```



```
my_list.insert(1)
my_list.insert(2)
my_list.insert(3)
my_list.print_list() # prints 1 2 3
```

In this way, ChatGPT can provide valuable explanations and examples to help programmers to understand programming concepts and data structures.

ChatGPT can also be used to provide guidance on specific programming languages and tools, such as Python, Java, or Git. For instance, a programmer might ask ChatGPT, “What is the best way to optimize my Python code for performance?” ChatGPT could provide a range of tips and techniques, such as using list comprehensions, avoiding unnecessary function calls, and leveraging built-in libraries and modules.

1. Programmer: *What is the best way to optimize my Python code for performance?*
2. ChatGPT: *There are many techniques for optimizing Python code for performance. Here are some tips:*
 - *Use list comprehensions instead of for loops when possible, as they can be more efficient.*
 - *Avoid unnecessary function calls, especially in loops. For example, instead of calling a function multiple times with the same arguments, call it once and store the result in a variable.*
 - *Leverage built-in libraries and modules, such as NumPy, which provides fast and efficient array operations.*
 - *Use the appropriate data structures for your needs. For example, if you need to perform frequent insertions or deletions, consider using a deque instead of a list.*
 - *Profile your code to identify bottlenecks and areas that can be optimized. Python’s built-in profiler, cProfile, can be helpful for this.*

Furthermore, ChatGPT can assist in solving coding challenges and debugging code [41]. A programmer might ask ChatGPT a question such as, “Why am I getting an ‘IndexError’ when I try to access an element in my array?” ChatGPT could provide insights and suggestions for troubleshooting the issue, such as double-checking the array dimensions, using conditional statements to handle edge cases, or stepping through the code with a debugger.

Suppose that a programmer is working on a Python program and they encounter an error. They could ask ChatGPT for help by providing the error message and any relevant code. For example, if they see the error message “NameError: name ‘x’ is not defined” and their code includes a line such as “print(x)”, they might ask,

```
Q: I'm getting a NameError: name 'x' is not defined when I run my Python
code. Can you help me understand why?
print(x)
```

ChatGPT could then provide insights and suggestions for troubleshooting the issue. For example, it might suggest that they check the scope of the variable “x” to ensure that it is defined before being used in the “print()” statement. In this way, ChatGPT can be a valuable resource for programmers who are stuck or need help with specific coding challenges.

Let us now consider another advantage of ChatGPT for app development using R-Shiny [42]. Shiny is an R package that enables the creation of interactive web applications using R. Shiny makes it easy to turn analyses into interactive web applications that anyone can use, allowing data scientists and analysts to share their work with a wider audience.

Shiny apps are important for data science because they provide a way to communicate data-driven insights to a wider audience. By making it easy to share analyses and visualizations, Shiny can help to democratize access to data and empower decision-makers with the insights that they need to make informed decisions. Additionally, Shiny can help data scientists and analysts to validate their findings and test assumptions in a more interactive and intuitive way. Below, we provide an example that uses ChatGPT to assist with generating a Shiny app.

1. Programmer: *I'm new to Shiny and I'm not sure how to get started with building an app. Can ChatGPT provide any guidance?*
2. ChatGPT: *Sure! Here are some tips to help you get started with building a Shiny app:*
 - *Start by sketching out the basic layout and user interface of your app. Consider what types of input and output elements you will need, such as sliders, text boxes, and plots.*
 - *Familiarize yourself with the structure of a Shiny app, which consists of a ui.R file for the user interface and a server.R file for the back-end logic.*
 - *Use the Shiny widgets package to add input and output elements to your app. This package provides a wide range of widgets, such as sliders, text boxes, and select boxes, that can be easily added to your app.*
 - *Leverage the power of reactive programming in Shiny to create dynamic and responsive apps. Reactive expressions and observers allow you to create dependencies between input and output elements, so that changes in one element automatically update others.*
 - *Consider using Shiny modules to modularize your app and make it easier to maintain and extend. Modules allow you to encapsulate parts of your app logic and user interface into separate components that can be reused across multiple apps.*
 - *Finally, use the Shiny app gallery and documentation to find inspiration and examples for your own app. The gallery showcases a wide range of Shiny apps built by the community, and the documentation provides detailed explanations and examples of Shiny functionality.*
3. Programmer: *Thanks for the tips! Do you have any suggestions for specific packages or tools that can help me build my Shiny app?*
4. ChatGPT: *Certainly! Here are some popular packages and tools for building Shiny apps:*
 - *Shinydashboard: provides a framework for building dashboard-style apps with a variety of layout and visualization options.*
 - *Shinythemes: provides a collection of pre-built themes for customizing the appearance of your app.*
 - *Plotly: provides a range of interactive and animated visualization options, including scatter plots, bar charts, and heatmaps.*
 - *DT: provides a flexible and powerful way to display and manipulate data tables in your app, including sorting, filtering, and pagination options.*
 - *ShinyWidgets: provides a collection of custom widgets for creating interactive user interfaces, including color pickers, file uploaders, and progress bars.*
 - *RStudio Connect: provides a platform for deploying and managing your Shiny app, including version control, security, and analytics features.*

Here is an example of a simple Shiny app generated by ChatGPT. This simple Shiny app allows the user to input a number and see its square. The code for the app is as follows:

```

Define UI for application
ui <- fluidPage(
  titlePanel('Square Calculator'),
  sidebarLayout(
    sidebarPanel(
      numericInput(inputId = 'number', label = 'Enter a number:', value
        = 0)
    ),
    mainPanel(
      verbatimTextOutput(outputId = 'result')
    )
  )
)

Define server logic
server <- function(input, output) {

```

```
output$result <- renderPrint({
  input$number^2
})
}
```

```
Run the application
shinyApp(ui = ui, server = server)
```

The UI for the app is defined using the `fluidPage` function, which allows for a responsive layout. The UI includes a title panel and a sidebar layout that includes a numeric input and a main panel that shows the result. The server logic is defined using the `server` function, which takes in the input from the UI and outputs the result. In this case, the output is simply the square of the input number. The `shinyApp` function runs the application, taking in the UI and server logic as arguments.

This simple app demonstrates the ease with which interactive web applications can be created using Shiny, with the guidance of ChatGPT. With a short section of R code, data scientists and analysts can create custom web applications that allow for data exploration and analysis by a wider audience. These examples make it evident that ChatGPT can be a powerful resource for programmers, providing insights, examples, and guidance on a wide range of programming topics. Whether we have a beginner or an experienced developer, ChatGPT can help them to improve their skills, solve coding challenges, and make the most of their programming projects.

5. The Future of ChatGPT in Data Science

As Sam Altman, the creator of ChatGPT, said, the future of AI is both awesome and terrifying and the possibilities of ChatGPT are virtually endless [43]. GPT-4 is undergoing development, with promises of being the most advanced large language model at launch [44].

Against this backdrop, first and foremost, ChatGPT should improve competition and help to curb the lack of specialists in the field of data science given its ability to train laymen to code and program solutions to analytics problems [45]. This should make it more attractive for people who previously found the idea of data science daunting to begin experimenting with data-science-related tasks using ChatGPT. At the same time, some aspects of data science that previously required human interaction could become obsolete [46]. To this end, the concept of intelligence augmentation in the age of AI was recently discussed by Hassani et al. [27], where the authors proposed that AI should be seen and used as a tool for augmenting intelligence and improving human efficiency as opposed to replacing humans. Data scientists will need to differentiate themselves and focus on how they can augment their skills using ChatGPT to remain competitive in the job market.

Secondly, as a model that continuously improves its capacity and learns from the internet by using machine learning, we believe that machine learning is one of the most promising areas for ChatGPT in data science. ChatGPT continues to redefine the limitations that previously governed what a machine can learn [47]. In terms of the processes underlying machine learning, it is well known that machine learning algorithms are data-hungry [48] and rely heavily on large datasets for training. In some cases, data scientists are unable to access the required datasets at scale due to various reasons, and so the ability to generate synthetic data using ChatGPT can significantly reduce the need for "real" large datasets. Additionally, there is evidence that ChatGPT can be used to generate realistic data that simulate complex real-world scenarios [49], which could ultimately lead to better-trained models and more accurate predictions.

Thirdly, having seen the impact of ChatGPT on society, other tech giants such as Google are already working on their own versions of this model (for example, see James (2023) for an account of how Google's Bard is being developed further). As more time and investments flow into related research and development, we could expect models that can perform more sophisticated and complex data-science-related tasks, with ChatGPT likely

to provide significant competition following its partnership with Microsoft. Developments in advanced neural networks (Transformers in particular), reinforcement learning, and unsupervised learning can all contribute towards enhancing ChatGPT's capabilities.

6. ChatGPT Is Not Always Correct

ChatGPT, as with any other language model, is not always correct and can make mistakes. The accuracy and reliability of ChatGPT's responses depend on several factors, such as the quality and diversity of the training data, the complexity and ambiguity of the input text, and the specific task or question being asked.

ChatGPT may struggle to provide accurate answers in certain situations, such as the following.

Ambiguous or unclear questions: If the input text is ambiguous or does not provide enough context for ChatGPT to understand the intended meaning, it may generate inaccurate or irrelevant responses.

Out-of-domain questions: ChatGPT is trained on a large corpus of text from various domains and topics, but it may not have sufficient knowledge or expertise in certain areas, leading to inaccurate responses.

Biased or inaccurate training data: ChatGPT's training data are sourced from the internet and may contain biased or inaccurate information. This can affect the accuracy and reliability of its responses.

Complex or technical language: ChatGPT may struggle to understand and generate responses to complex or technical language, such as scientific or legal terminology, that is not commonly used in everyday language.

It is important to note that ChatGPT is a tool designed to assist humans in generating responses to various tasks, but it should not be relied upon as the sole source of information or decision-making. Human oversight and critical thinking are essential to ensure the accuracy and integrity of the information generated by ChatGPT.

Let us now provide an example. We asked ChatGPT "what is the sum of sample autocorrelation function for any stationary time series with arbitrary length?". ChatGPT's answer was, "For a stationary time series with arbitrary length, the sum of the sample autocorrelation function (ACF) is not necessarily equal to a specific value. It depends on the specific properties and characteristics of the time series". However, we know that the correct answer is $-1/2$ according to Hassani's $-1/2$ Theorem [50–52].

Theorem 1. *The sum of the Sample Autocorrelation Function (ACF), S_{ACF} , at lag $h \geq 1$ is always $-\frac{1}{2}$ for any stationary time series with arbitrary length $T \geq 2$ [50–52]; that is,*

$$S_{ACF} = \sum_{h=1}^{T-1} \hat{\rho}_h = -\frac{1}{2} \quad (1)$$

In contrast, if we search for such a question on Google, we will find various sources that provide us with the correct answers. In many cases, a quick Google search can provide accurate and reliable information.

It is important to remember that ChatGPT is a machine learning model that is trained on a large corpus of text data, and its responses are based on patterns and associations in these data. It may not have access to the latest or most up-to-date information, and it may not always be able to interpret the context and intent of a question accurately.

In contrast, Google search results are generated by algorithms that take into account various factors, such as the relevance, credibility, and recency of the information. Google's search algorithms are constantly evolving and improving, making it a reliable source of information for many users. Here are some examples of mathematical questions where ChatGPT might provide incorrect answers.

Complex integrals: ChatGPT may struggle with complex integrals that require a deep understanding of calculus and other advanced mathematical concepts. For example, if

asked to solve an integral that requires the use of techniques such as integration by parts, substitution, or partial fractions, ChatGPT may generate an incorrect answer.

Unusual number systems: If asked to perform calculations in an unusual number system, such as base 3 or base 16, ChatGPT may not be able to provide an accurate answer. This is because the model is primarily trained on the decimal system and may not have sufficient exposure to other number systems.

Multivariable calculus: While ChatGPT can handle some basic multivariable calculus questions, it may struggle with more complex questions that involve partial derivatives, gradients, and multiple integrals.

Abstract algebra: ChatGPT may not be able to generate accurate answers to questions related to abstract algebra, such as group theory, ring theory, and field theory. These topics require a deep understanding of advanced mathematical concepts and may be outside the scope of ChatGPT's training data.

7. ChatGPT in University and Data Science Programs: Emphasizing Ethics and Integrity

As data science continues to evolve, universities and data science programs have an important role to play in ensuring that the technology is developed and used in a responsible and ethical manner. Integrating ChatGPT training into data science programs can be an effective way to promote these values, provided that the training is conducted in a manner that emphasizes ethics and integrity. Below, we provide specific suggestions on how this can be achieved.

First, it is important to ensure that ChatGPT training is integrated into data science programs in a way that prioritizes the ethical use of the technology. This could involve incorporating ethics-focused modules into the curriculum, or hosting workshops and events that highlight the importance of ethical considerations in data science. By making ethics a core component of the training, universities can help to ensure that future data scientists understand the potential ethical implications of ChatGPT and other technologies.

Second, universities and data science programs should work to foster a culture of integrity around ChatGPT training. This can involve encouraging students to engage in an open and honest discussion about the ethical considerations associated with the technology, and providing guidance on how to handle situations where ethical concerns arise. By promoting a culture of integrity, universities can help to ensure that data scientists who have been trained in ChatGPT are more likely to use the technology in an ethical manner.

Third, data science programs can also emphasize the importance of transparency when using ChatGPT. This could involve encouraging students to document the data sources and algorithms that they use when training ChatGPT, as well as making the results of ChatGPT analysis more transparent to stakeholders. By promoting transparency, data scientists can help to build trust with stakeholders and ensure that the technology is being used in a responsible manner.

Fourth, universities should also work to ensure that students who are trained in ChatGPT are aware of the potential biases that may be inherent in the technology. This could involve educating students on how to identify and mitigate potential biases, or incorporating bias detection tools into the training curriculum. By understanding the potential biases associated with ChatGPT, data scientists can take steps to ensure that their analysis is unbiased and accurate.

Finally, data science programs should emphasize the importance of ongoing learning and professional development for data scientists who are trained in ChatGPT. This could involve providing access to online courses or workshops that focus on ethical considerations in data science, or hosting networking events where data scientists can learn from one another. By promoting ongoing learning, universities can help to ensure that data scientists who are trained in ChatGPT continue to prioritize ethics and integrity throughout their careers.

In conclusion, ChatGPT training can be a valuable tool for data scientists, but it is important to ensure that the technology is developed and used in a responsible and ethical manner. By integrating ChatGPT training into data science programs in a manner that emphasizes ethics and integrity, universities can help to ensure that future data scientists have the skills and knowledge necessary to use the technology in a responsible and ethical way.

8. Conclusions

In conclusion, this article presents a concise account of how ChatGPT is influencing data science and the likely impact that it will have on this field in the future. The examples shared herein make it clear that it is indeed a powerful tool that can assist data scientists in a variety of ways. ChatGPT's potential to revolutionize the field of data science is unprecedented, and as the technology continues to advance, the use of ChatGPT is likely to become even more widespread, helping data scientists to improve their workflows and achieve better results.

In addition to discussing the many advantages of ChatGPT for data science, we also highlight the challenges and limitations posed by this innovation as it is crucial that future generations of data scientists not only embrace this technology to augment their own skills but also learn to use it ethically, with integrity and full awareness of its benefits and costs.

This perspective article is not without its limitations. Given the novelty of the topic, we have had to rely on several industry publications and opinion pieces in developing our perspective, with access to very limited peer-reviewed academic sources covering the topic of interest. However, we hope that our attempt motivates more academics researching in the field of data science to consider the aspects that we cover in more depth in the future, thereby making significant contributions to knowledge around the impact of ChatGPT on data science.

Finally, it is evident that ChatGPT is pushing boundaries and making humans realize the true impact and influence that AI will have on their lives, from education to employment in the field of data science. It also presents opportunities for universities to create new training programs alongside existing data science courses to assist the future generations in using ChatGPT efficiently whilst instilling the values of ethics and integrity. Eventually, as with all technological innovations, it is up to the human race, which has been blessed with creativity and critical thinking and evaluation, to determine how ChatGPT is used for the benefit of society at large—to create a world where AI is used to augment intelligence as opposed to replacing human intelligence.

Author Contributions: Authors contributed equally to this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. de Medeiros, M.M.; Hoppen, N.; Macada, A.C.G. Data science for business: Benefits, challenges and opportunities. *Bottom Line* **2020**, *330*, 149–163. [CrossRef]
2. Nielsen, O.B. A comprehensive review of data governance literature. *Sel. Pap. IRIS* **2017**, *8*, 120–133.
3. Available online: <http://www.datascienceasn.org/code-of-conduct.html> (accessed on 27 February 2023).
4. Available online: <https://chat.openai.com> (accessed on 27 February 2023).
5. Ruby, M. How ChatGPT Works: The Model behind the Bot. Towards Data Science. 2023. Available online: <https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286> (accessed on 17 February 2023).
6. Chow, A.R.; Perrigo, B. The AI Arms Race Is Changing Everything. *TIME Magazine*. Available online: <https://time.com/6255952/ai-impact-chatgpt-microsoft-google/> (accessed on 17 February 2023).
7. Thorp, H.H. ChatGPT is fun, but not an author. *Science* **2023**, *379*, 313. [CrossRef] [PubMed]
8. Pavlik, J.V. Collaborating with ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *J. Mass Commun. Educ.* **2023**, *78*, 10776958221149577. [CrossRef]
9. Romero, A. ChatGPT and the Future (Present) We're Facing. Towards Data Science. 2023. Available online: <https://towardsdatascience.com/chatgpt-and-the-future-present-were-facing-9f2a1cfab0e9> (accessed on 20 February 2023).

10. Lund, B.D.; Wang, T. Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News* 2023, *in press*. [CrossRef]
11. van Dis, A.M.; Bollen, J.; Zuidema, W.; van Rooji, R.; Bockting, C.L. ChatGPT: Five priorities for research. *Nature* 2023. Available online: <https://www.nature.com/articles/d41586-023-00288-7> (accessed on 19 February 2023).
12. Asare, J.G. *The Dark Side Of ChatGPT*; Forbes: Jersey City, NJ, USA, 2023. Available online: <https://www.forbes.com/sites/janicegassam/2023/01/28/the-dark-side-of-chatgpt/?sh=31f2e08a4799> (accessed on 19 February 2023).
13. Getahun, H. *ChatGPT Could Be Used for Good, But Like Many Other AI Models, It's Rife with Racist and Discriminatory Bias*; Insider. 2023. Available online: <https://www.insider.com/chatgpt-is-like-many-other-ai-models-rife-with-bias-2023-1> (accessed on 19 February 2023).
14. King, M.R.; ChatGPT. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cell. Mol. Bioeng.* 2023, 16, 1–2. [CrossRef] [PubMed]
15. Check Point Research. OPWNAI: Cybercriminals Starting to Use chatGPT. Checkpoint.com. 2023. Available online: <https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/> (accessed on 20 February 2023).
16. Marr, B. *Will ChatGPT Put Data Analysts Out Of Work?* Forbes: Jersey City, NJ, USA, 2023. Available online: <https://www.forbes.com/sites/bernardmarr/2023/02/07/will-chatgpt-put-data-analysts-out-of-work/?sh=46433cf64030> (accessed on 19 February 2023).
17. Mollick, E. *ChatGPT Is a Tipping Point for AI*; Harvard Business Review: Brighton, MA, USA, 2022. Available online: <https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai> (accessed on 20 February 2023).
18. Przybyla, M. Why ChatGPT Is Good for the Data Science Community. Medium. 2023. Available online: <https://medium.com/geekculture/why-chatgpt-is-good-for-the-data-science-community-d4dfbebb4ef> (accessed on 19 February 2023).
19. Hutanu, A. How ChatGPT Works and AI, ML & NLP Fundamentals. PentaBlog. 2023. Available online: <https://www.pentalog.com/blog/tech-trends/chatgpt-fundamentals/#:~:text=ChatGPT%20is%20an%20AI%20language,text%20data%20to%20generate%20language> (accessed on 19 February 2023).
20. Heaven, W.D. How ChatGPT works and AI, ML & NLP Fundamentals. *Mit Technol. Rev.* 2023. Available online: <https://www.technologyreview.com/2023/02/08/1068068/chatgpt-is-everywhere-heres-where-it-came-from/> (accessed on 19 February 2023).
21. Scialom, T.; Chakrabarty, T.; Muresan, S. Fine-Tuned Language Models Are Continual Learner. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, DNEC, Abu Dhabi, United Arab Emirates, 7–11 December 2022. Available online: <https://arxiv.org/pdf/2205.12393.pdf> (accessed on 19 February 2023).
22. Christensen, A. How Many Languages Does ChatGPT Support? The Complete ChatGPT Language List. SEO.ai. 2023. Available online: <https://seo.ai/blog/how-many-languages-does-chatgpt-support> (accessed on 19 February 2023).
23. Cyriac, G.P. Understanding ChatGPT Using ChatGPT. Medium. 2023. Available online: <https://medium.com/@ginspcyriac744/understanding-chatgpt-using-chatgpt-7b24e6dde2eb> (accessed on 19 February 2023).
24. Elgan, M. ChatGPT: Finally, an AI Chatbot Worth Talking to. *Computer World.* 2022. Available online: <https://www.computerworld.com/article/3682143/chatgpt-finally-an-ai-chatbot-worth-talking-to.html> (accessed on 19 February 2023).
25. Farseev, A. Is Bigger Better? Why The ChatGPT Vs. GPT-3 Vs. GPT-4 'Battle' Is Just a Family Chat. Forbes.com. 2023. Available online: <https://www.forbes.com/sites/forbestechcouncil/2023/02/17/is-bigger-better-why-the-chatgpt-vs-gpt-3-vs-gpt-4-battle-is-just-a-family-chat/?sh=5055fb2a5b65> (accessed on 20 February 2023).
26. Hassani, H.; Huang, X.; Silva, E.S. The Human Digitalisation Journey: Technology First at the Expense of Humans? *Information* 2021, 12, 267. [CrossRef]
27. Hassani, H.; Beneki, C.; Silva, E.S.; Vandeput, N.; Madsen, D.O. The science of statistics versus data science: What is the future? *Technol. Forecast. Soc. Chang.* 2021, 173, 121111. [CrossRef]
28. Adedeji, A. Harnessing the Power of ChatGPT for Data Science. Analytics Vidhya. 2023. Available online: <https://www.analyticsvidhya.com/blog/2023/02/harnessing-the-power-of-chatgpt-for-data-science/> (accessed on 19 February 2023).
29. Peter, J. Generating Synthetic Datasets with ChatGPT. Medium. 2023. Available online: <https://medium.com/@jan.peter/generating-synthetic-datasets-with-gpt-3-83ad15a32958> (accessed on 19 February 2023).
30. Wiles, J. *Beyond ChatGPT: The Future of Generative AI for Enterprises*; Gartner: Stamford, CT, USA, 2023. Available online: <https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises> (accessed on 19 February 2023).
31. Martineau, K. Five ways IBM is using synthetic data to improve AI models. *IBM Res.* 2022. Available online: <https://research.ibm.com/blog/synthetic-data-explained> (accessed on 19 February 2023).
32. Sarkar, S. Synthetic Data Generation—A Must-Have Skill for New Data Scientists. Medium. 2018. Available online: <https://towardsdatascience.com/synthetic-data-generation-a-must-have-skill-for-new-data-scientists-915896c0c1ae> (accessed on 19 February 2023).
33. Lucini, F. The Real Deal About Synthetic Data. *MIT Sloan Manag. Rev.* 2021, 63, 1–4. Available online: <https://sloanreview.mit.edu/article/the-real-deal-about-synthetic-data/> (accessed on 19 February 2023).
34. Susnjak, T. Applying BERT and ChatGPT for Sentiment Analysis of Lyme Disease in Scientific Literature. *arXiv* 2023, arXiv:2302.06474. Available online: <https://arxiv.org/pdf/2302.06474.pdf> (accessed on 19 February 2023).

35. Paialunga, P. Hands-on Sentiment Analysis on Hotels Reviews Using Artificial Intelligence and Open AI's ChatGPT, with Python. Medium. 2022. Available online: <https://towardsdatascience.com/hands-on-sentiment-analysis-on-hotels-reviews-using-artificial-intelligence-and-open-ais-chatgpt-d1939850c79e> (accessed on 19 February 2023).
36. Cupani, M. Advanced NER with GPT-3 and GPT-J. *Towards Data Sci.* 2022. Available online: <https://towardsdatascience.com/advanced-ner-with-gpt-3-and-gpt-j-ce43dc6cdb9c> (accessed on 19 February 2023).
37. Marie, B. Translate with ChatGPT. *Towards Data Sci.* 2023. Available online: <https://towardsdatascience.com/translate-with-chatgpt-f85609996a7f> (accessed on 19 February 2023).
38. Doerrfeld, B. How ChatGPT Raises Software Developers' Agility—And Their Value. Acceleration Economy Network. 2023. Available online: <https://accelerationeconomy.com/ai/how-chatgpt-raises-software-developers-agility-and-their-value/> (accessed on 19 February 2023).
39. Tanvir, F. 7 Ways ChatGPT Can Help You To Become a Better Developer. Geek Culture. 2023. Available online: <https://medium.com/geekculture/7-ways-chatgpt-can-help-you-to-become-a-better-developer-2c490d7a387d> (accessed on 19 February 2023).
40. Tung, L. ChatGPT Can Write Code. Now Researchers Say It's Good at Fixing Bugs, Too. ZDNET. 2023. Available online: <https://www.zdnet.com/article/chatgpt-can-write-code-now-researchers-say-its-good-at-fixing-bugs-too/> (accessed on 19 February 2023).
41. Jain, S. ChatGPT is a new AI chatbot that can find mistakes in your code or write a story for you. *Business Insider India*. 2022. Available online: <https://www.businessinsider.in/tech/news/what-is-chatgpt-and-how-does-it-work/articleshow/95994901.cms> (accessed on 19 February 2023).
42. Available online: <https://shiny.rstudio.com> (accessed on 27 February 2023).
43. Bove, T. Sam Altman, the Maker of ChatGPT, Says the A.I. Future Is Both Awesome and Terrifying. If It Goes Badly: 'It's Lights-Out for All of Us'. Fortune.com. 2023. Available online: <https://fortune.com/2023/01/26/sam-altman-chatgpt-ai-future/> (accessed on 20 February 2023).
44. Awan, A.A. *The Latest On OpenAI, Google AI, and What it Means For Data Science*; Data Camp. 2023. Available online: <https://www.datacamp.com/blog/openai-google-ai-data-science> (accessed on 20 February 2023).
45. Bailey, E. The Future of Tech: Exploring AI/ML and ChatGPT. Cribl. 2023. Available online: <https://cribl.io/blog/the-future-of-tech-exploring-ai-ml-and-chatgpt/> (accessed on 20 February 2023).
46. Iyer, A. ChatGPT Unleashes a Flood of Future Careers You Never Dreamed of. *Analytics India Magazine*. 2023. Available online: <https://analyticsindiamag.com/chatgpt-unleashes-a-flood-of-future-careers-you-never-dreamed-of/> (accessed on 20 February 2023).
47. Rosenbaum, E. The ChatGPT AI Hype Cycle Is Peaking, But Even Tech Skeptics Don't Expect a Bust. CNBC 2023. Available online: <https://www.cnbc.com/2023/02/11/chatgpt-ai-hype-cycle-is-peaking-but-even-tech-skeptics-doubt-a-bust.html> (accessed on 20 February 2023).
48. Adadi, A. A survey on data-efficient algorithms in big data era. *J. Big Data* 2021, 8, 24. [CrossRef]
49. Goodman, M. Can You Generate Realistic Data with GPT-3? We Explore Fake Dating with Fake Data. CodeX. 2022. Available online: <https://medium.com/codex/can-you-generate-realistic-data-with-gpt-3-we-explore-fake-dating-with-fake-data-54e936d29cbb> (accessed on 20 February 2023).
50. Hassani, H. Sum of the sample autocorrelation function. *Random Oper. Stoch. Equ.* 2009, 17, 125–130. [CrossRef]
51. Hassani, H. A note on the sum of the sample autocorrelation function. *Physica A* 2010, 389, 1601–1606. [CrossRef]
52. Hassani, H.; Sirimal Silva, E. A Kolmogorov-Smirnov Based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts. *Econometrics* 2015, 3, 590–609. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.