

**University of Westminster
School of Computer Science and Engineering
5DATA006C.1 Data Visualisation and Communication**

Module Leader: Mr. Alroy Mascrenghe

Individual Coursework

Name: W.A.Heshan Peiris

UOW No.: w1999712

IIT No.: 20221171

Table of Contents

5DATA006C.1 Data Visualisation and Communication	1
Table of Contents	2
1) Research Question and Data Sourcing	3
1.1) Research Question	3
1.2) Data Source	3
1.3) Relevance of Research Questions	3
2) Data Preparation	4
2.1) Identification of Variables	4
2.2) Observation	4
2.3) Evaluating Data Tidiness	5
3) Exploratory Data Analysis	6
3.1) Univariate analysis	6
3.1.1) AI Tool Name	7
3.1.2) Description	7
3.1.3) Free/Paid/Other	7
3.2) multivariate analysis	13
3.2.1) Charges and Major Categories	13
3.2.2) Review and Free/Paid/Other	14
4) Data storytelling	15

1) Research Question and Data Sourcing

1.1) Research Question

How has the AI tool's performance improved or changed since its launch, and what are the notable updates or advancements introduced in the past year?

1.2) Data Source

Creative Commons – Public Domain

1.3) Relevance of Research Questions

- **Enhanced Accuracy and Fluency:** AI models are now able to respond with greater accuracy and fluency, creating content that is more akin to that produced by humans. The use of bigger and more varied training datasets together with advances in natural language processing (NLP) approaches are credited for this.
- **Improved Task Completion:** These days, artificial intelligence (AI) programs can do more, including creating original text forms and providing insightful, open-ended, difficult, or unusual answers to your inquiries. This shows how machine learning algorithms have advanced and how adaptable AI systems are in new situations.
- **Increased Efficiency and Scalability:** AI models can now analyze bigger volumes of data more quickly and tackle more complicated jobs because they are more scalable and efficient. This is essential for the practical implementation of AI, especially in fields with large data sets.
- **Personalized User Experiences:** AI technologies are becoming more and more customized to offer unique experiences to each user. The ability of AI systems to recognize and accommodate unique preferences is made possible by developments in user profile modeling and recommender systems.

2) Data Preparation

2.1) Identification of Variables

AI Tool Name: The name of the AI tool or software under consideration.

Description: A brief description of the AI tool's features and capabilities that gives a basic idea of how it works.

Free/Paid/Other: Specifies the AI tool's pricing technique, including whether it is free to use, requires a paid membership, or comes under a different price range.

Useable For: Outlines the specific issues or tasks that the AI tool is intended to solve and describes the main use cases or applications for which it is built.

Charges: Describe the price and/or cost of the AI technology, including information on any one-time purchases, subscription fees, and other costs.

Review: Evaluation and ratings submitted by users that provide information about the effective of the AI tool, user contentment, and general reaction among the user community.

Tool Link: The official website or download page for the AI tool may be accessed using this URL or link, giving consumers quick access to more thorough information and resources.

Major Category: Places the AI tool in more general areas or categories, such computer vision, data analytics, natural language processing (NLP), etc. to give background information on its major use case.

2.2) Observation

An AI tool's Name(AI Name),a brief description of its features and capabilities(Description), its pricing model(Free/Paid/Other and Charges), the main use cases or applications for which it is intended(Useable For), user- generated reviews and ratings that reflect the tool's effectiveness and user satisfaction(Review), and a direct link to the AI tool's official website or download page(Tool Link), are all included in a single row of this table.

2.3) Evaluating Data Tidiness

R Script

```
1 # Assuming the data is stored in a CSV file named 'Data.csv'
2 Data <- read.csv('all_ai_tool.csv')
3 # Identify duplicate rows based on a unique identifier
4 column_names <- names(Data)
5 column_names
6
7 duplicate_rows <- duplicated(Data[, c("AI.Tool.Name", "Description", "Free.Paid.Other", "Useable.For", "Charges", "Review", "Tool.Link", "Major.Category")])
8
9 # Example if missing values are represented as empty strings in character columns
10 missing_counts <- sapply(Data, function(x) sum(x == ""))
11 print(missing_counts)
12
13 # Make sure column names are unique and meaningful
14 column_names <- make.names(names(Data))
15 print(column_names)
16
17
18
```

R Environment

Environment History Connections Tutorial

Import Dataset 205 MiB

R Global Environment

Data

Data 4969 obs. of 8 variables

Values

column_names	chr [1:8] "AI.Tool.Name" "Description" "Free.Paid.Other" "Useable.For" "Charges" "Review" "Tool.Li...
duplicate_rows	logi [1:4969] FALSE FALSE FALSE FALSE FALSE FALSE ...
missing_counts	Named int [1:8] 0 0 0 0 2435 3507 0 0

R Console

```
R 4.3.0 · E:/BDA/BDA - 2st Year - 1st Sem/2. Data Visualisation/CW/
> # Assuming the data is stored in a CSV file named 'Data.csv'
> Data <- read.csv('all_ai_tool.csv')
> # Identify duplicate rows based on a unique identifier
> column_names <- names(Data)
> column_names
[1] "AI.Tool.Name" "Description" "Free.Paid.Other" "Useable.For" "Charges"
[6] "Review" "Tool.Link" "Major.Category"
>
> duplicate_rows <- duplicated(Data[, c("AI.Tool.Name", "Description", "Free.Paid.Other", "Useable.For", "Charges", "Review", "Tool.Link", "Major.Category")])
>
> # Example if missing values are represented as empty strings in character columns
> missing_counts <- sapply(Data, function(x) sum(x == ""))
> print(missing_counts)
AI.Tool.Name Description Free.Paid.Other Useable.For Charges Review
0 0 0 0 2435 3507
Tool.Link Major.Category
0 0
>
> # Make sure column names are unique and meaningful
> column_names <- make.names(names(Data))
> print(column_names)
[1] "AI.Tool.Name" "Description" "Free.Paid.Other" "Useable.For" "Charges"
[6] "Review" "Tool.Link" "Major.Category"
>
>
```

My analysis of AI Tools data reveals that it is tidy and well-organized. Furthermore, all column names are unique, meaningful, and non-repeating, and no duplicate rows and missing values were found. This thorough data preparation guarantees consistent and accurate data, ensures the reliability and validity of the subsequent analysis.

3) Exploratory Data Analysis

3.1) Univariate analysis

Variable	Data Type	Count	Missing	Unique	Minimum	Q1	Median (Q2)	Q3	Maximum
AI Tool Name	String	4969	0	4969	-	-	-	-	-
Description	String	4969	0	4969	-	-	-	-	-
Free/Paid/Other	String	4969	0	19	-	-	-	-	-
Useable For	String	4969	0	526	-	-	-	-	-
Charges	Integer	4969	2435	556	0	6.9675	9.9	29	6597
Review	Integer	4969	3507	18	1	1	1	3	42
Tool Link	String	4969	0	4969	-	-	-	-	-
Major Category	String	4969	0	8	-	-	-	-	-

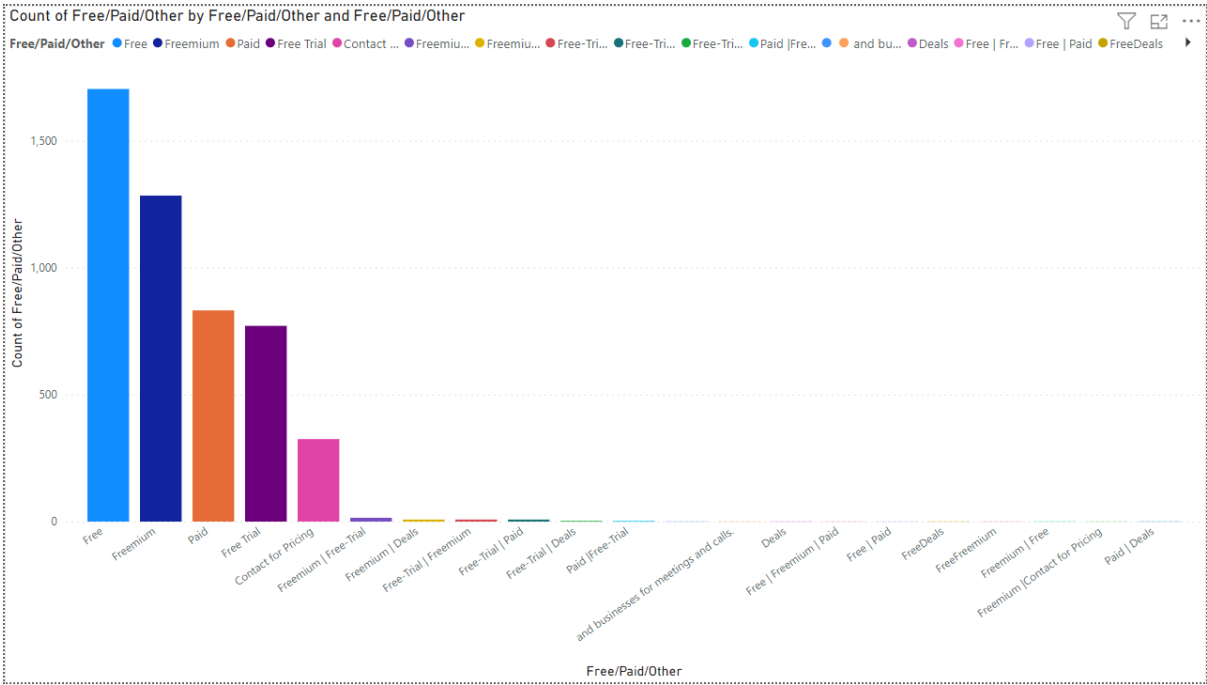
3.1.1) AI Tool Name

The data in the “AI Tool Name” column is either category or nominal, which is why some chart types, such histograms, do not consider it visualizable. Discrete categories or labels without a numerical interpretation or natural order make up categorical data. Histogram and other visualization tools are intended for numerical data with quantitative meaningful values.

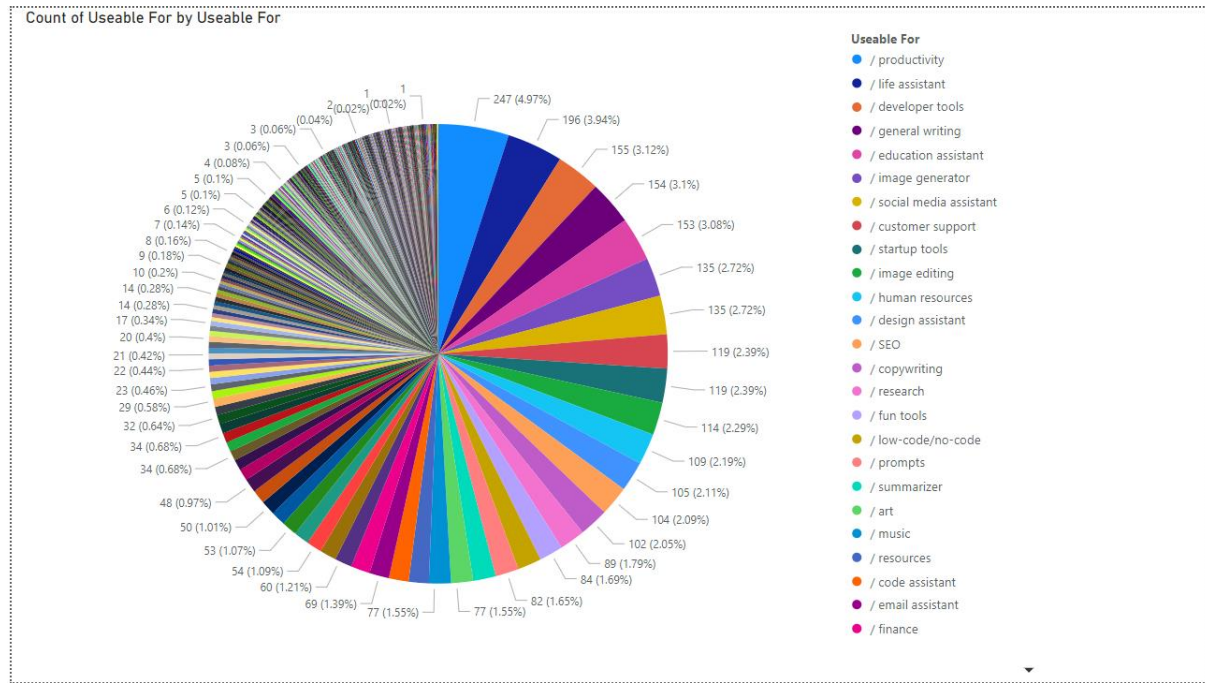
3.1.2) Description

Text contains information that may not be well expressed visually; therefore, it is typically a better format than visualization when discussing the features and capabilities of an AI application. The reason for this frequent preference for textual explanations to explain the functionality of an AI tool is as follows. Text serves as the basis for communicating exact and comprehensive information, while images can improve comprehension, draw attention to important details, or break down difficult structures.

3.1.3) Free/Paid/Other



3.1.4) Useable For



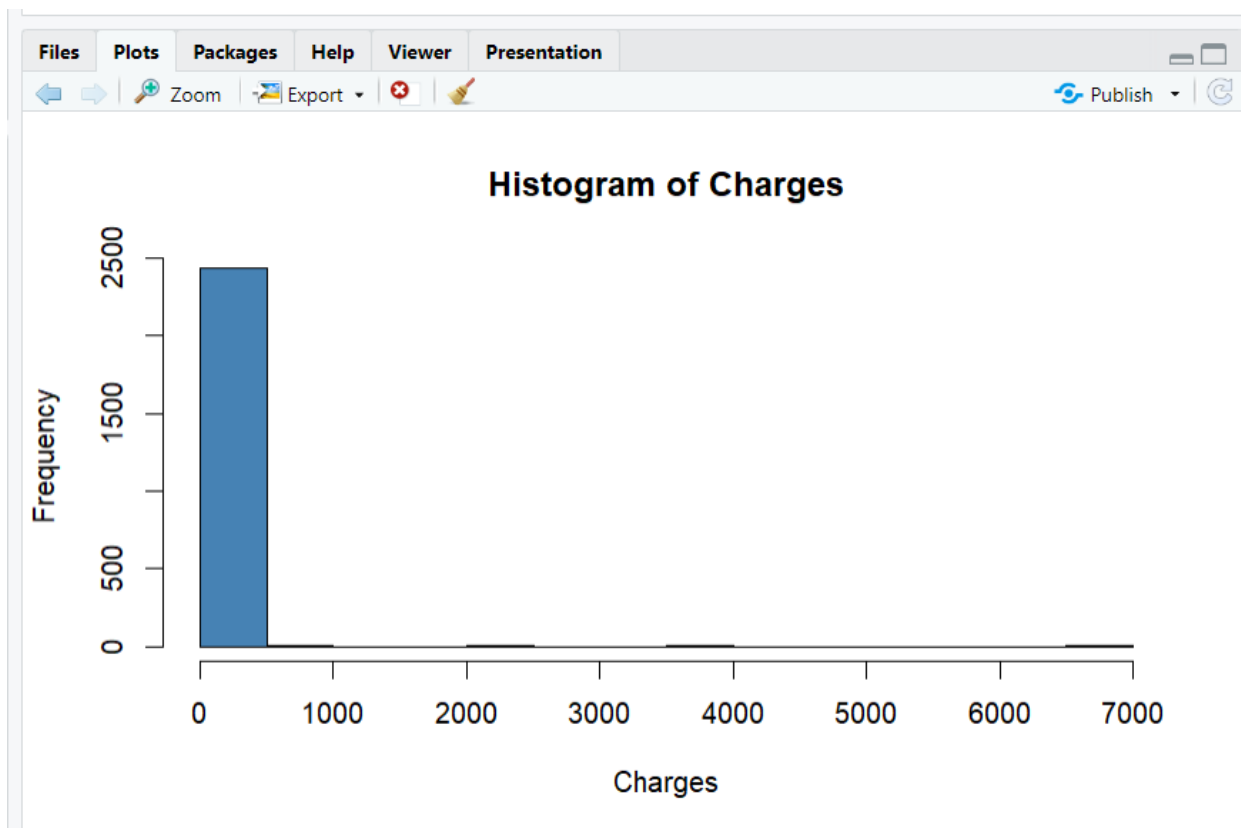
3.1.5) Charges

```
# Check the data type of the 'Charges' column
print(class(Data$Charges))

# Convert 'Charges' column to numeric if it's not already
Data$Charges <- as.numeric(Data$Charges)

# Check the data type again
print(class(Data$Charges))

# Now, create the histogram
hist(Data$Charges, main='Histogram of Charges', xlab='Charges', col='steelblue', breaks=12)
```

```
Console | Terminal x | Background Jobs x
R 4.3.0 · E:/BDA/BDA - 2st Year - 1st Sem/2. Data Visualisation/CW/
> # Check the data type of the 'Charges' column
> print(class(Data$Charges))
[1] "numeric"
>
> # Convert 'Charges' column to numeric if it's not already
> Data$Charges <- as.numeric(Data$Charges)
>
> # Check the data type again
> print(class(Data$Charges))
[1] "numeric"
>
> # Now, create the histogram
> hist(Data$Charges, main='Histogram of Charges', xlab='Charges', col='steelblue', breaks=12)
> plot(x, y, type = "l", main = "Normal Distribution Curve", xlab = "Value", ylab = "Density")
```

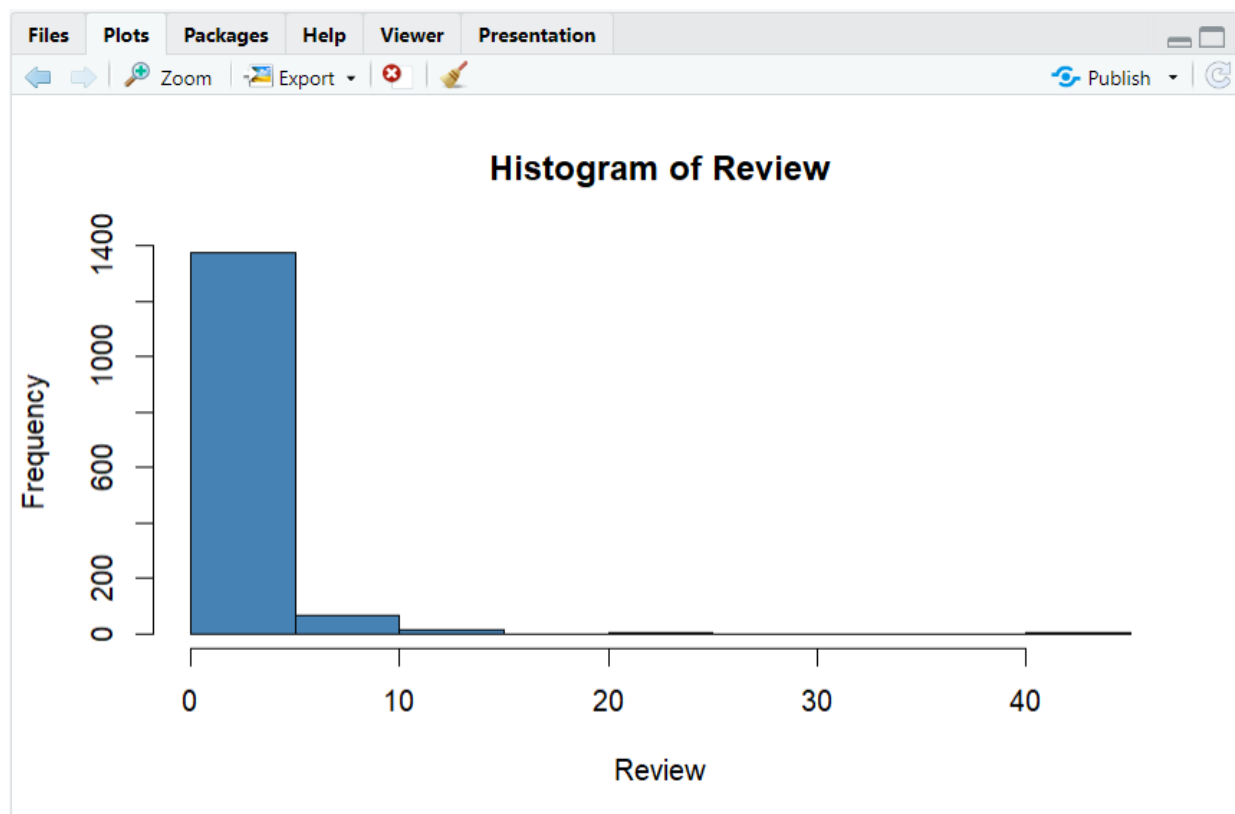
3.1.6) Review

```
# Check the data type of the 'Review' column
print(class(Data$Review))

# Convert 'Review' column to numeric if it's not already
Data$Review <- as.numeric(Data$Review)

# Check the data Review))

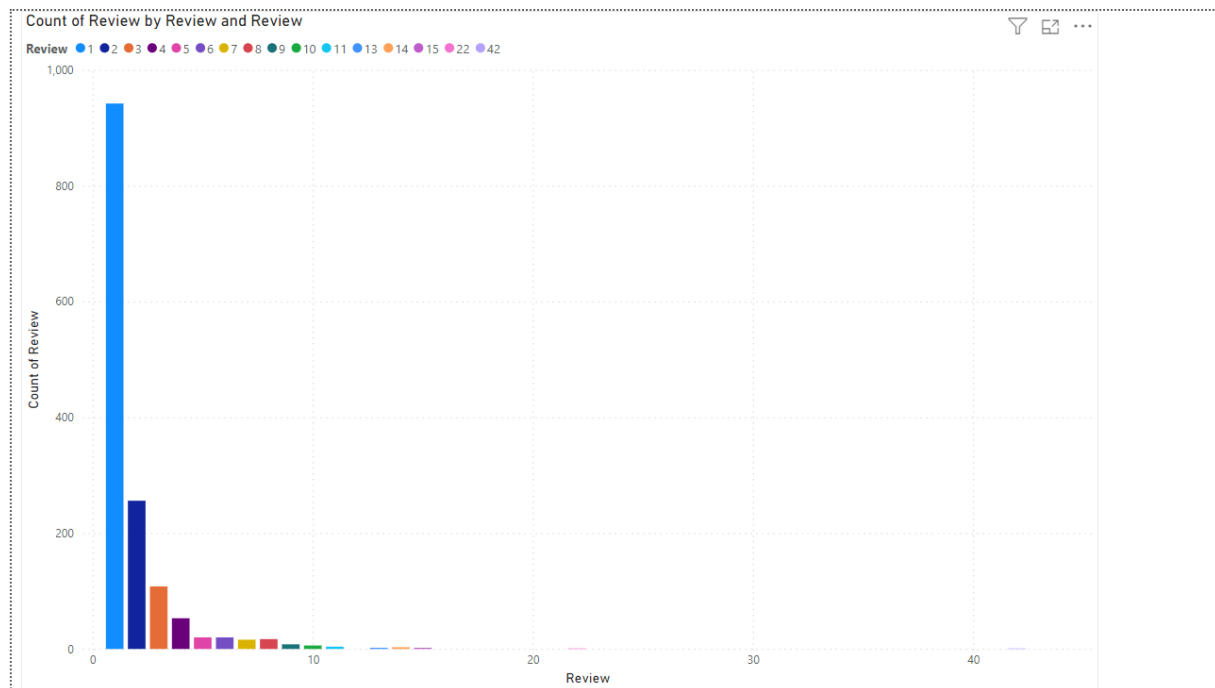
# Now, create the histogram
hist(Data$Review, main='Histogram of Review', xlab='Review', col='steelblue', breaks=12)
```



```

Console Terminal Background Jobs
R 4.3.0 · E:/BDA/BDA - 2st Year - 1st Sem/2. Data Visualisation/CW/
> # Check the data type of the 'Review' column
> print(class(Data$Review))
[1] "numeric"
>
> # Convert 'Review' column to numeric if it's not already
> Data$Review <- as.numeric(Data$Review)
>
> # Check the data Review))
>
> # Now, create the histogram
> hist(Data$Review, main='Histogram of Review', xlab='Review', col='steelblue', breaks=12)
> plot(x, y, type = "l", main = "Normal Distribution Curve", xlab = "Value", ylab = "Density"

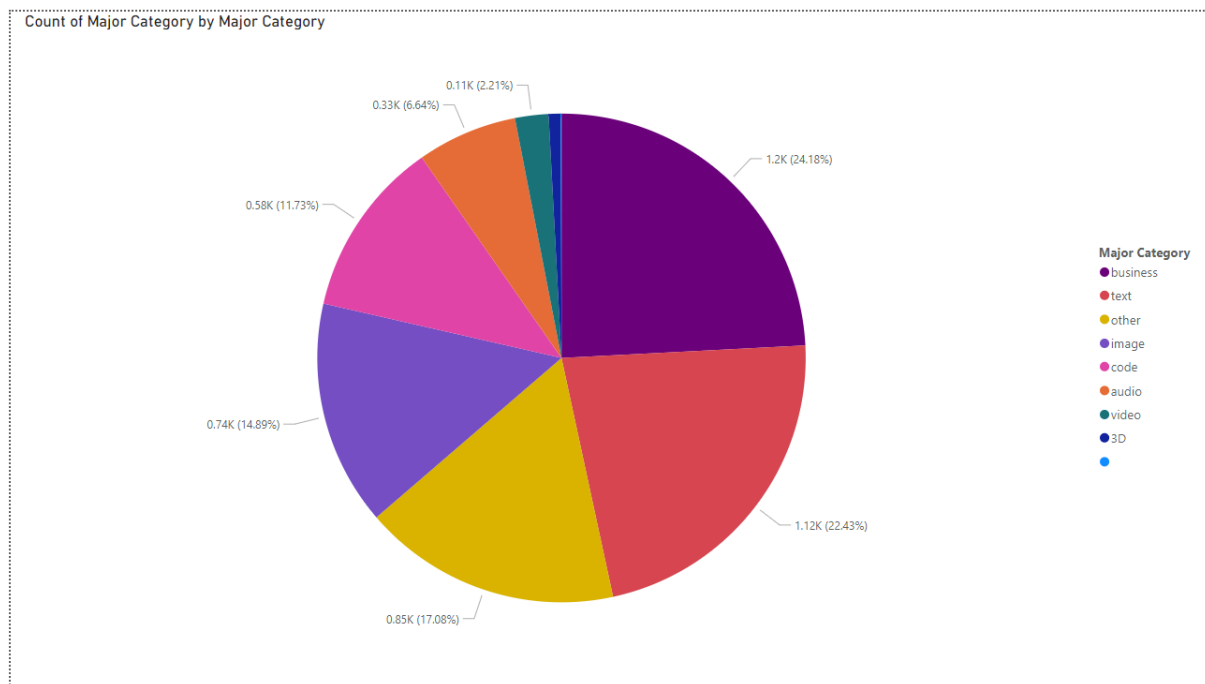
```



3.1.7) Tool Link

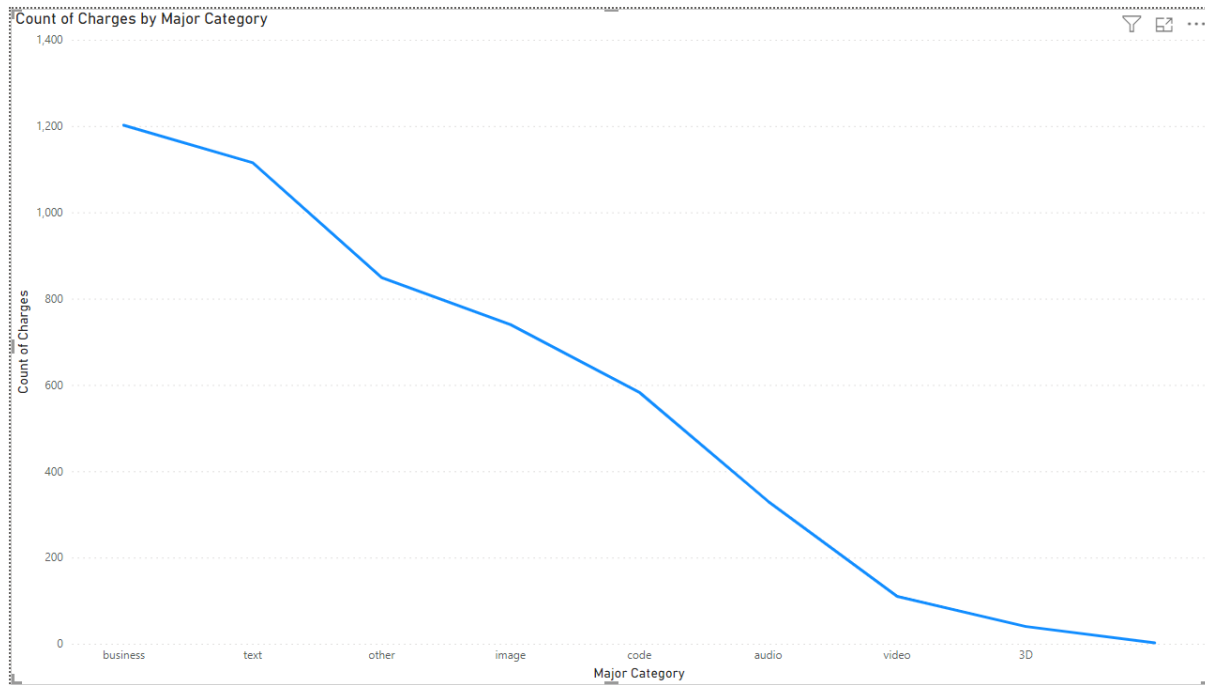
The official website or download page for the AI tool may be accessed by users by clicking on the hyperlink or URL provided in the “Tool Link” column. The goal of this column is to provide users with a straightforward means of accessing detailed information and resources. Therefore, it is not visualizable.

3.1.8) Major Category



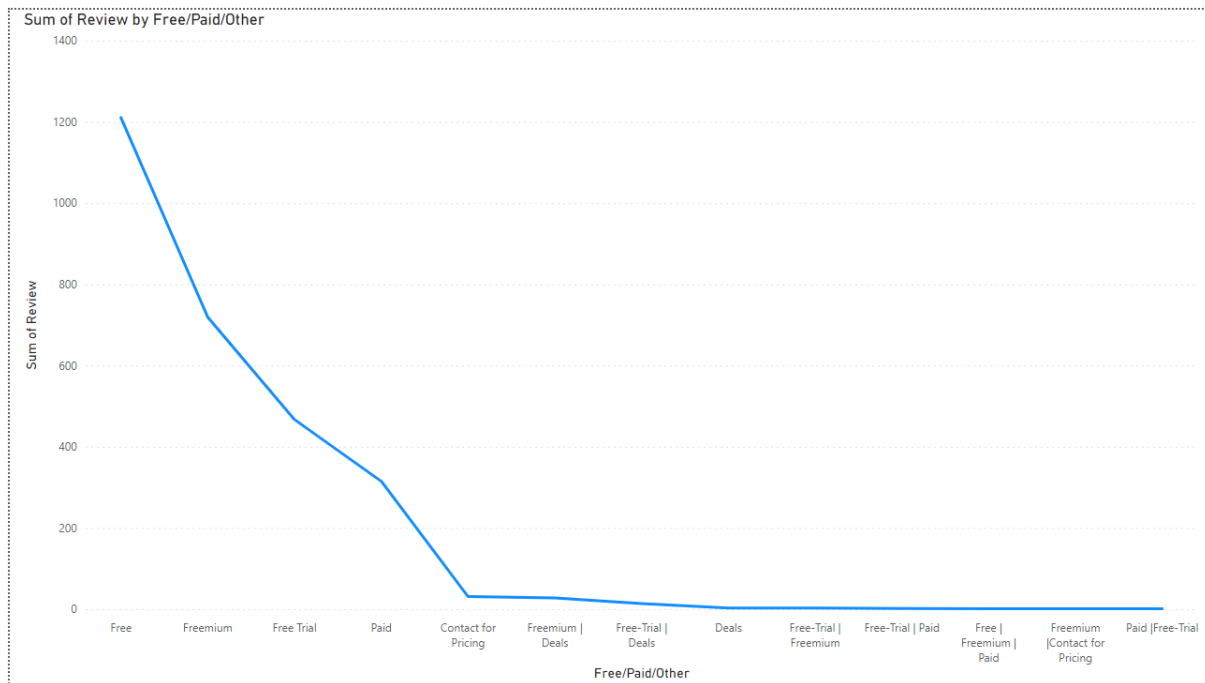
3.2) multivariate analysis

3.2.1) Charges and Major Categories



According to decreasing charge counts, the intermediate categories include The X-axis indicates categories, while the Y-axis shows the total of charges. The line chart shows how charges are distributed throughout the key categories. With the largest count, "Business" is the category that charges the most, while "3D" is the category that charges the least. As the charge counts decrease, the intermediate categories are "Text," "Other," "Image," "Code," "Audio," and "Video." A clear summary of the relative charging frequency across the designated key categories is provided by the graphic depiction, which highlights the dominance of the "Business" category in terms of the number of charges.

3.2.2) Review and Free/Paid/Other



The Y-axis shows the cumulative review counts, while the X-axis represents the various payment forms. The line chart shows the link between the sum of reviews and these payment types. With the greatest number of reviews, "Free" is the most often used payment option, whilst "Paid (Free Trial)" is the least popular and has the fewest reviews overall. According to the collected evaluations, the graphic depiction offers a succinct summary of the user involvement and popularity linked to each payment method, highlighting the significant inclination towards the "Free" payment option and the reduced interest in the "Paid (Free Trial)" category.

4) Data storytelling

Introduction

Innovation in artificial intelligence is rife, and this is especially true in the fast-growing field of general AI tools. What it means for robots to think and behave like humans is quickly changing thanks to these adaptable assistants that can handle a variety of activities and learn on their own.

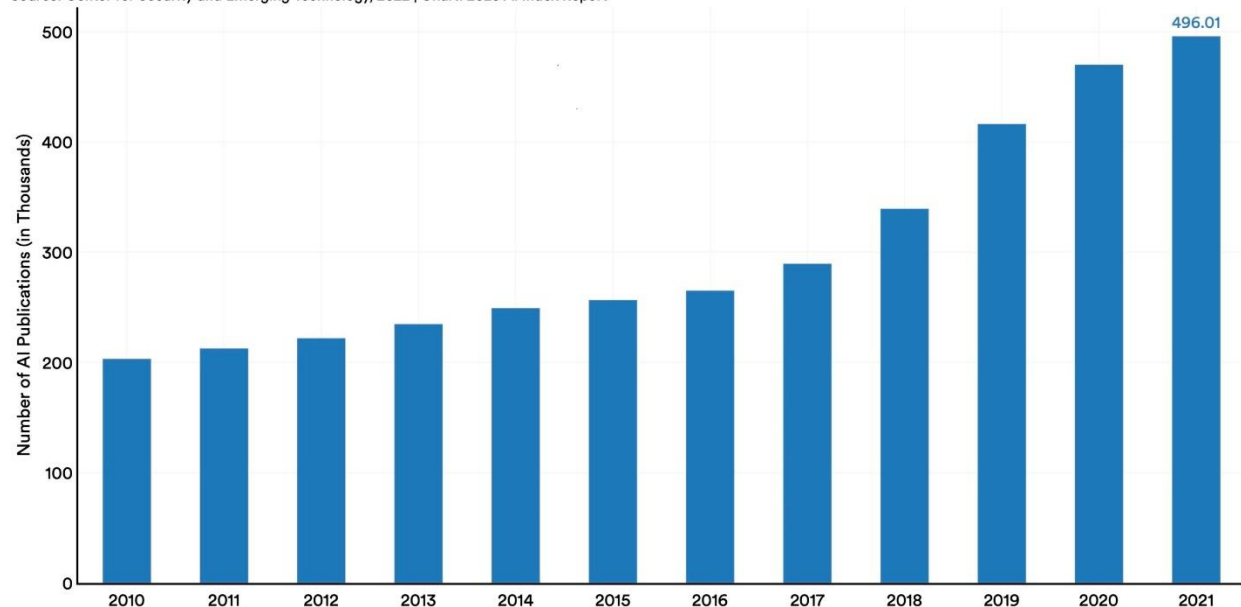
This data narrative looks deeply into the intriguing world of general artificial intelligence (AI) technologies, following their development from early prototypes to the more complex applications we see today. We'll look at their productivity increases, dissect their ground-breaking releases, and investigate the waves they're creating in a variety of fields and sectors.

Get ready to see the emergence of the machines—not as heartless, cold algorithms, but as flexible, learning collaborators that will push the limits of artificial intelligence. We'll analyze the development of General AI tools, illuminating their extraordinary journey and the limitless potential that lies ahead through gripping storytelling and perceptive graphics.

Performance Evaluation

Number of AI Publications in the World, 2010–21

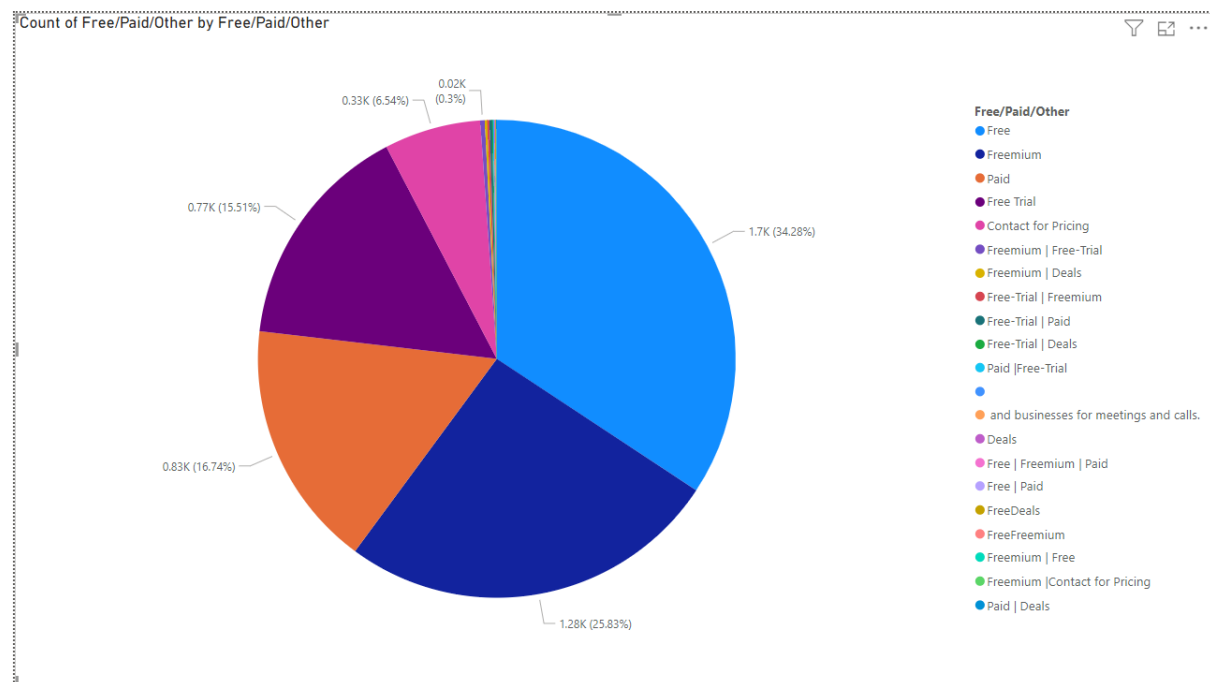
Source: Center for Security and Emerging Technology, 2022 | Chart: 2023 AI Index Report



Over the past ten years, there has been a significant and fast increase in the number of AI publications globally, proving the dynamic evolution and growing significance of artificial intelligence in scientific discourse and academic study. In 2010, there were almost 200,000 AI publications worldwide. The following years saw a sharp increase in the number, reaching 240,000 in 2013, 280,000 in 2016, and an

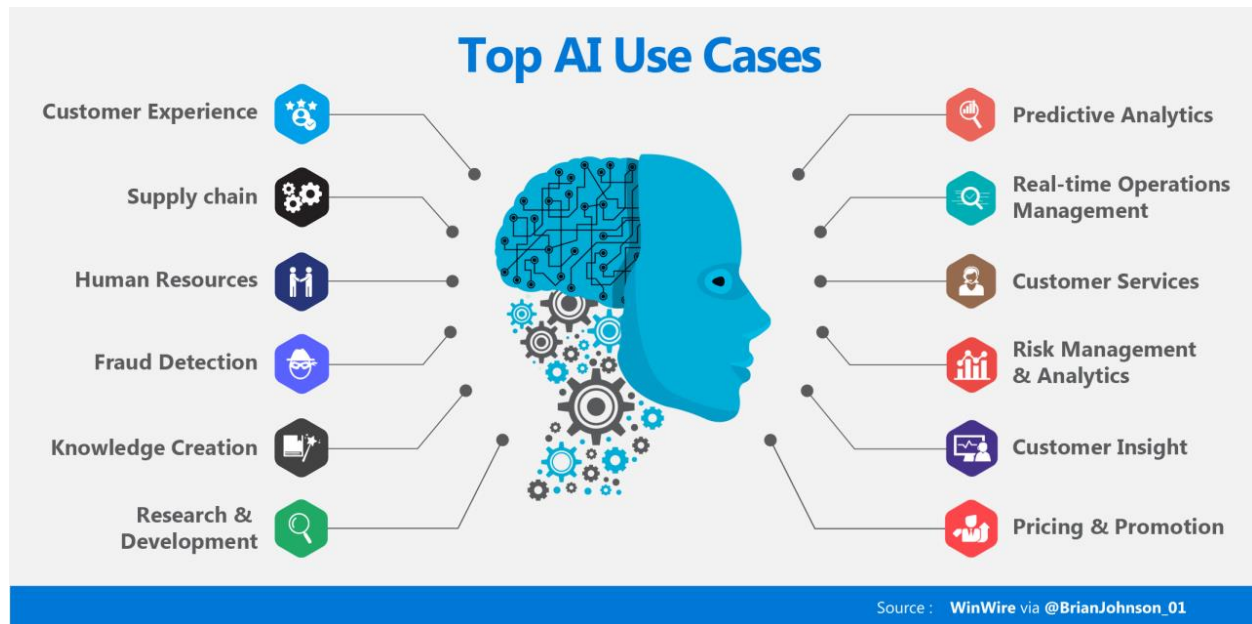
astounding 350,000 in 2018. The trend reached 400,000 publications in 2019, then 450,000 in 2020, and a noteworthy peak of 496,000 in 2021. In 2019, the trend continued its increasing trajectory. This development demonstrates the vast amount of research on artificial intelligence as well as the speed at which it is being added, indicating the growing effect, advancements, and interest in AI in academic and research circles.

Pricing Structure



The pricing structures that have been used display a range of imagination and inclusion. A remarkable 34.28% of the tools take a welcoming stance, kindly providing their services for free, promoting accessibility and stimulating research in the field of artificial intelligence. Closely behind at 25.83%, a freemium business model entices consumers with free access to basic functions, saving premium offers for those choosing to pay for subscriptions. This approach appeals to hesitant users who want to try something out before committing fully. While 15.51% choose free trials so consumers may realize their full potential, 16.74% boldly embrace a paid approach, highlighting their value offer through premium services. 6.54% of tools recommend highly customized solutions and promote direct contact for pricing to provide a more personalized experience. Just 1.1%, a sprinkling of more tools.

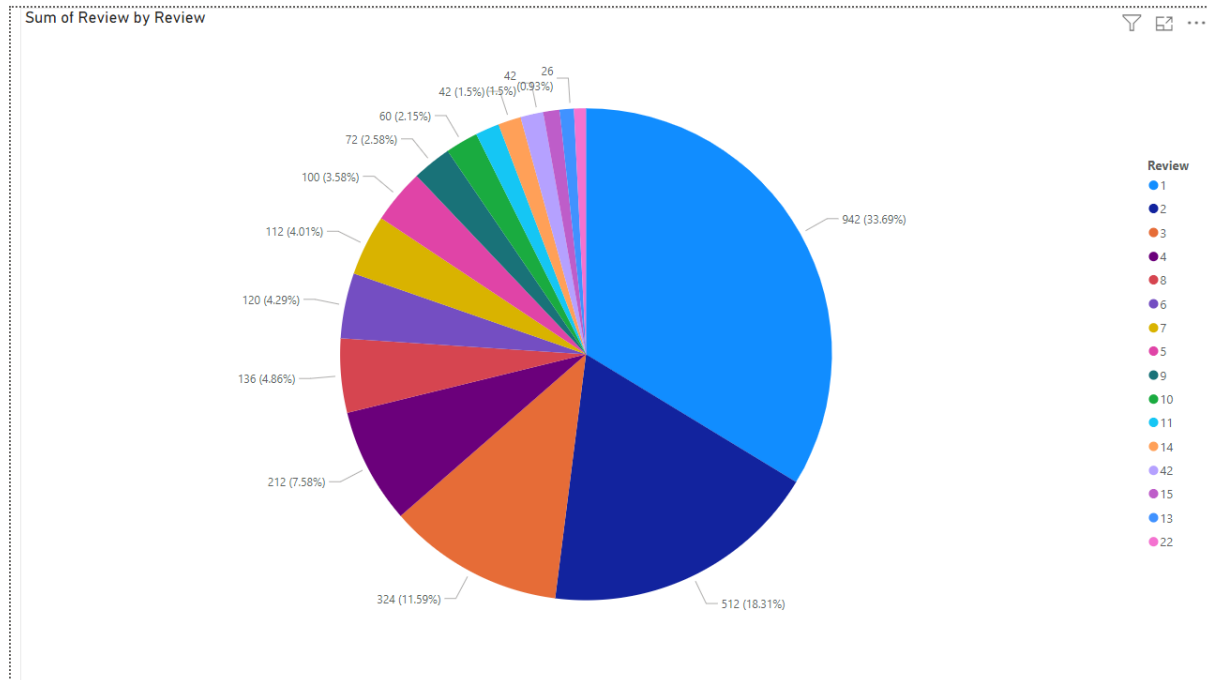
Use Cases



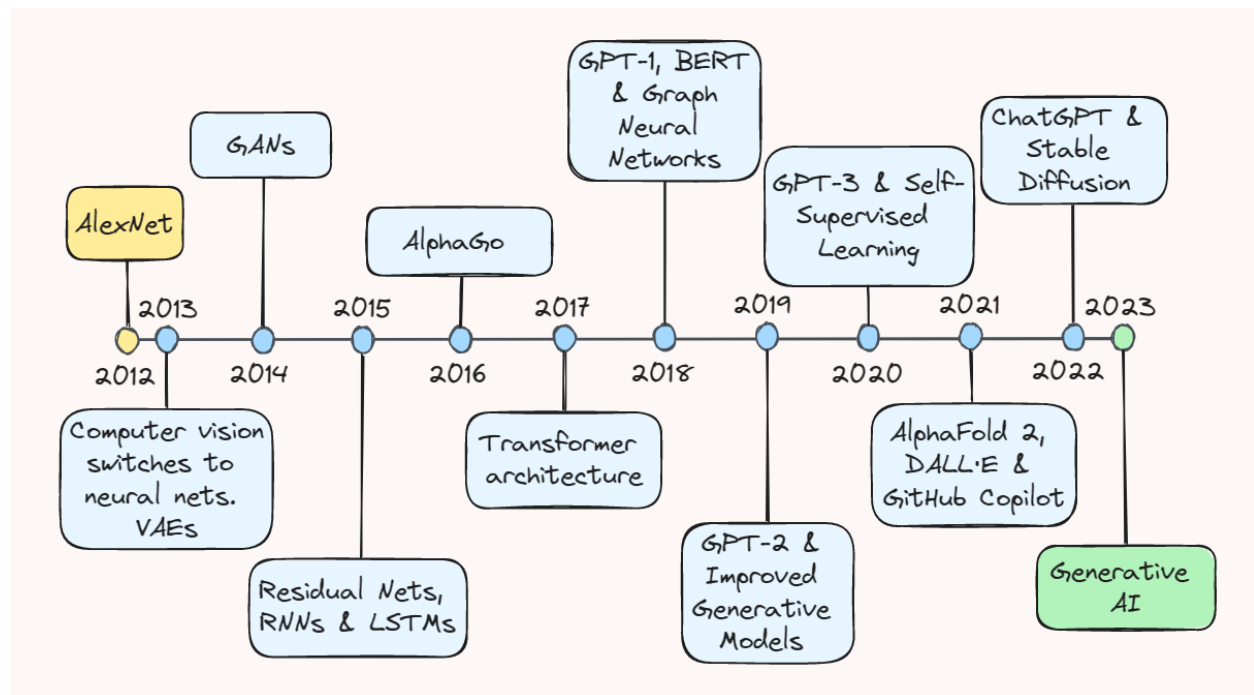
Here are some AI use cases in different industries:

1. **Healthcare:** AI can be used for medical image analysis, drug discovery, disease diagnosis and treatment planning, clinical trial optimization, and personalized medicine.
2. **Finance:** AI can be used for fraud detection, customer service chatbots, predictive analytics, algorithmic trading, risk management, and compliance monitoring.
3. **Manufacturing:** AI can be used for predictive maintenance, quality control, supply chain optimization, product design, and safety monitoring.
4. **Agriculture:** AI can be used for crop yield prediction, soil analysis, disease detection, pest management, and weather forecasting.
5. **Education:** AI can be used for personalized learning, student engagement tracking, plagiarism detection, automated grading, and student support chatbots.
6. **Marketing:** AI can be used for customer segmentation, predictive analytics, chatbots for customer support, sentiment analysis, and marketing campaign optimization.
7. **Media and Entertainment:** AI can be used for content recommendations, personalized advertising, sentiment analysis, automated content generation, and copyright infringement detection.

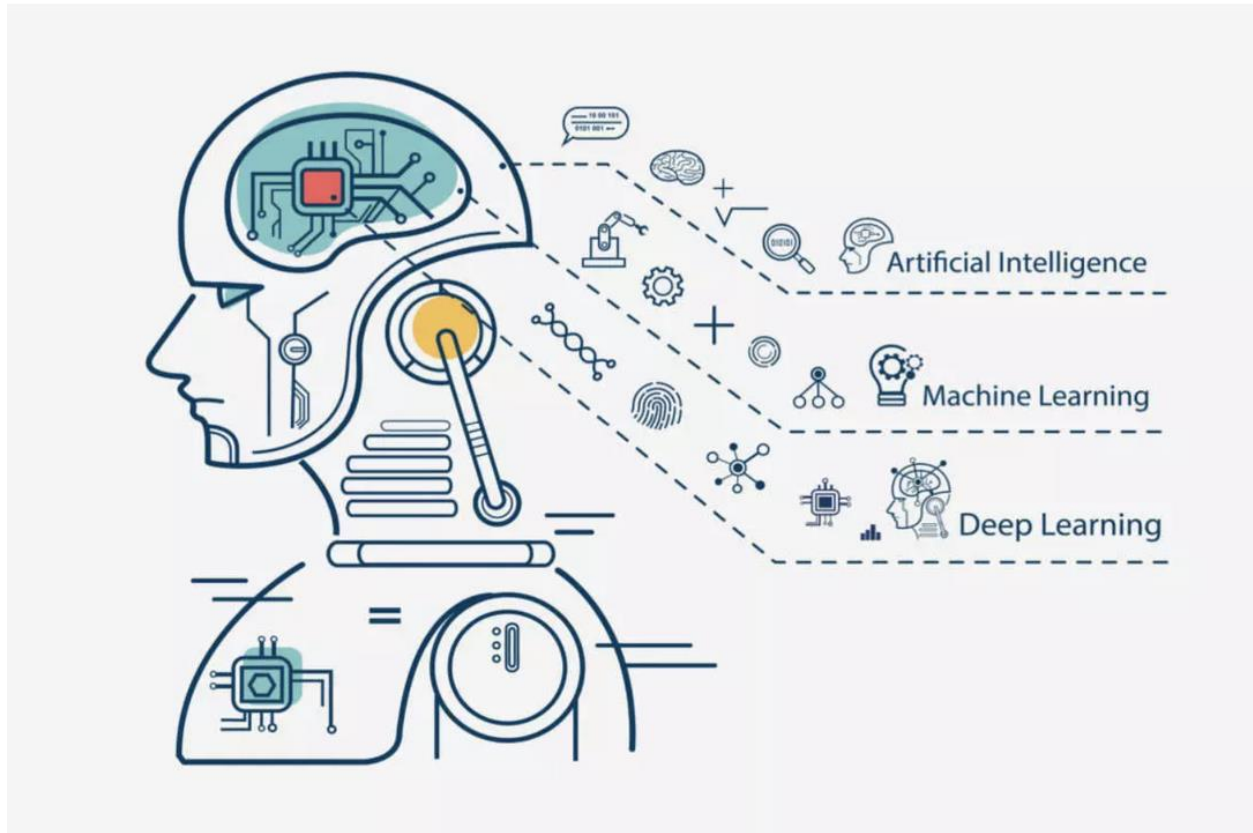
Years of AI in Review



The distribution of reviews for AI tools across different rating categories is shown in the pie chart. 33.69% of the reviews are one star, which makes up the bulk of the total. With 18.31% of the total, 2 stars is the second-highest category; 3 stars is next with 11.59%. Higher rating categories are allocated smaller percentages: 4 stars make up 7.58%, 8 stars 4.86%, and 10 stars have the lowest proportion (2.15%). Overall, the distribution shows a significant concentration of reviews in the lower rating categories, indicating conflicting opinions or worries among AI tool users.



Future Outlook



Significant advancements in deep learning approaches are expected to push advances in pattern recognition, natural language processing, image and audio recognition, and pattern identification. It is anticipated that this machine learning breakthrough would be crucial to the development of autonomous systems like self-driving automobiles and unmanned aerial aircraft. As deep learning technologies advance, additional ethical issues, the necessity for regulatory frameworks, and the cooperative synergy between humans and AI become critical. Deep learning and quantum computing together can solve complicated issues that are outside the scope of traditional computing. Nonetheless, difficulties like cybersecurity flaws highlight how crucial it is to use AI responsibly. The employment landscape is also expected to change, with new occupations emerging and other professions becoming automated.

The AI revolution is here, and its tools are evolving at breakneck speed. It's time to embrace the potential, address the challenges, and shape the future we want to see. Let's unlock the power of AI, responsibly.