https://doi.org/10.5281/zenodo.14054137

Comparative Sentiment Analysis and Semantic Meaning in Text using sentiment models from Hugging Face and Power Automate

Aneta Trajkovska
University “St. Kliment Ohridski” – Bitola, Faculty of Information and Communication Technologies,
Bitola, North Macedonia
aneta.trajkovska@uklo.edu.mk

Ilija Jolevski
University “St. Kliment Ohridski” – Bitola, Faculty of Information and Communication Technologies,
Bitola, North Macedonia
ilija.jolevski@uklo.edu.mk

***Abstract*—This paper outlines an innovative approach utilizing Artificial Intelligence models for determining the semantic meaning of sentences. In linguistics, it is crucial to understand how we interpret texts and uncover the underlying messages conveyed by the authors through their sentences. Today, we are witnessing significant advancements in technology and decision-making, driven by previous experiences, diverse databases and knowledge systems. This wealth of data is processed using various robotic systems and artificial intelligence models. Large corporations often receive extensive customer reviews and feedback on their products, but manual reading and analysis can be challenging. In this article we will show how utilization of the Hugging Face models and Power Automate models can do deeply semantic analysis on text and how the punctuation marks can impact the way how the text is written and the whole meaning. Hugging Face can assist by using its advanced natural language processing models to automatically analyze and interpret extensive customer reviews and feedback, streamlining the process. Power Automate can help by automating the collection, organization, and initial analysis of customer reviews and feedback, reducing the manual workload. The synergy between both technologies will be described, as well as their powerful utilization to explore a range of innovative applications to showcase their versatility and practicality.**

Keywords—Hugging Face, Power Automate, AI, Robotic Process Automation, Intelligent systems;

I. INTRODUCTION

The rapid evolution of IT technology and the development of new AI systems are driving numerous scientific innovations and creating abundant opportunities for researchers. Not only are these innovations changing traditional businesses, but they are also opening new areas of research. Researchers can now utilize sophisticated AI models to explore complex datasets, develop intelligent systems, and solve previously inconceivable problems [1]. The integration of AI in various domains is fostering a collaborative environment where interdisciplinary research thrives, leading to groundbreaking discoveries and the continuous expansion of human knowledge [2].

Social media, large corporations, banking systems, educational platforms, universities, and other institutions or industries regularly receive diverse feedback and reviews reflecting user satisfaction levels. These insights are crucial for business growth and determining future strategies. Feedback varies, encompassing customer perspectives and internal organizational assessments. A thorough understanding of feedback can drive substantial growth within organizations. The demand for analyzing feedback, emails, and documents is growing, posing challenges for humans due to potential errors in manual processing. Integration of robotic processes with AI models is the key of how this information’s could be processed and analyzed with different statistical methods and significantly reducing the occurrence of errors.

Hugging Face provides a platform to develop and test custom models or utilize pre-built natural language processing models for language detection and text classification [3]. On the other hand, Power Automate (PA) enables the development of Robotic Process Automation (RPA) workflows that automate repetitive tasks and minimize errors [4],[5]. PA also offers pre-built AI models for language detection and semantic analysis of text, enhancing its automation capabilities.

Through this paper, we will explore the capabilities of AI models in understanding languages and discerning the impact of punctuation marks within sentences. Utilizing models from both Hugging Face and Power Automate (PA), we will conduct comparative analyses across various scenarios.

II. HUGGING FACE AND ITS FEATURES

Hugging Face is an open-source collaborative platform that provides a range of tools that allow anyone to develop, train, and deploy NLP and ML models [6]. It is offering a user-friendly, easily accessible environment for developers, researchers, and enthusiasts to build sophisticated AI applications by utilizing open-source code. The platform supports a wide range of NLP tasks, such as sentiment analysis, translation, and text summarization and offers an extensive library of pre-trained models, making it easier to implement state-of-the-art AI solutions. Hugging Face is a leading AI platform that offers a robust inference API for deploying models as a service, allowing for seamless integration into applications. Ensures scalable and efficient model inference, making it suitable for production environments. Hugging Face provides several powerful libraries and tools for natural language processing (NLP) tasks, including model architectures, pre-trained models, tokenization, training pipelines, etc. There are a wide range of pre-trained models for various NLP tasks, such as text classification, named entity recognition, question answering, and more. We can choose a specific architecture (e.g., BERT, GPT, RoBERTa) and load a pre-trained model using the AutoModel class [7].

Hugging Face is widely utilized in scientific and engineering fields for the analysis of complex data sets.

These features make Hugging Face a powerful and versatile platform for anyone working with machine learning and AI [8].

A. *Fundamental Technologies*

Hugging Face utilizes several fundamental technologies to support its suite of tools and libraries. These technologies forms are impacting the progress of their offerings, enabling advanced natural language processing (NLP) and machine learning capabilities. Here are the key fundamental technologies used by Hugging Face:

Python is an effective programing language for utilizing Hugging Face's natural language processing (NLP) features, allowing users to develop, train, and employ sophisticated language models with ease [9].

Transformer Architecture provides APIs and tools to easily download and train state-of-the-art pretrained models [10]. These models support common tasks in different modalities, such as: NLP (text classification, named entity recognition, question answering) computer vision (image classification, object detection, and segmentation), audio (automatic speech recognition and audio classification) and multimodal (table question answering, information extraction from scanned documents). Transformers support framework interoperability between PyTorch, TensorFlow, and JAX. Hugging Face's core technology is built around the transformer architecture, which is the basis for many models like BERT, GPT, T5.

PyTorch and TensorFlow are both popular open-source deep learning frameworks that facilitate the development and deployment of machine learning models, particularly in the domain of neural networks [11],[12].

Tokenization is the process of breaking down a sequence of text into smaller units called tokens shown in Fig.1. The main goal of tokenization is to create a standardized and structured representation of textual data that can be easily processed by machine learning algorithms and natural language processing (NLP) models [13]. Hugging face implements advanced tokenization techniques that convert text into numerical input suitable for models, including Byte-Pair Encoding (BPE), WordPiece, and SentencePiece tokenizers [14].

Data sets is a library and repository that offers a vast array of curated datasets for various NLP tasks. Hugging Face Datasets provides a valuable resource for finding and working with diverse and well-structured datasets.

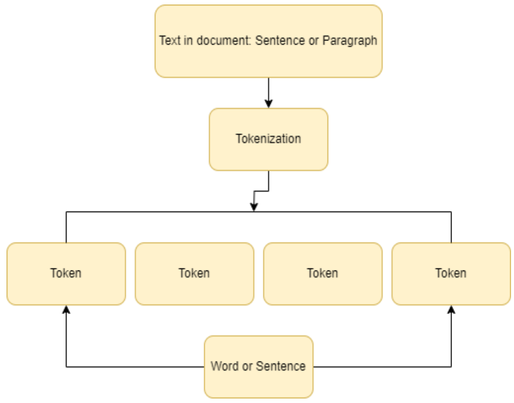

Fig.1. Process of tokenization of a sentence or word

III. Power Automate And Its Featutres

Power Automate is a cloud platform that supports RPA systems. It is a Microsoft product that provides a low-code platform for automating workflows across various applications and services. Power Automate is enabled by default in all Office 365 applications and comes with more than 150 standard connectors that can be utilized and create various automations on the processes. The tool offers an equal number of premium connectors available for purchase to increase automation capabilities. With its user-friendly interface and extensive library of pre-built templates, Power Automate empowers organizations to increase productivity, efficiency, and collaboration by automating routine tasks and processes [15].

A. *Robotic Process Automations in Power Automate and AI connectors*

RPA seamlessly replicates and streamlines business operations, mimicking human actions such as logging into applications, data entry, email correspondence and other repetitive tasks [15]. In PA we can easily develop bots that can execute different tasks. Developing the bots is with configuration of pre-build connectors. There are two options for building bots, desktop bots or cloud flows, additionally there are several ways of real time monitoring them [16].

In order to add intelligence to automated processes, forecast results, and enhance corporate performance, PA includes its own AI Builder feature. With its direct integration with PA, AI Builder directly integrates the power of Microsoft AI through experience, AI Builder enables you to build and deploy AI models effortlessly, empowering you to make data-driven decisions and automate tasks with ease [17].

B. *AI Builder transforming business capabilities*

Below are mentioned some of the capabilities utilization of the AI Builder models, also check the Fig.2. for overview of business automation:

- Low Code/ No code capabilities
- Speed up business processes with AI insights
- Custom Model Training
- User friendly interface
- Seamless integration with over 250 connectors
- Predictive Analysis
- Sentiment Analysis
- Object detection

Fig.2. Extracting text from objects using PA and AI connectors [18]

IV. Differences between Semantic meaning and Sentiment analyse In Text

The semantic meaning of a sentence refers to the interpretation of its content, focusing on the meaning and concepts conveyed by the words and their arrangement. This includes understanding the relationships between words, the context in which they are used, and the overall message or idea being conveyed. Semantic analysis seeks to understand not only the literal meaning of words, punctuation marks, but also implied, inferred, or nuanced meanings, which may be affected by context, tone, and cultural or situational factors.

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique used to determine the emotional tone or attitude expressed in a piece of text. It involves classifying the sentiment of the text into categories such as positive, negative, or neutral. Sentiment analysis is commonly used to analyze customer reviews, social media posts, and other forms of unstructured text data to gauge public opinion or emotional responses.

Sentiment analysis is detecting emotional tone and how is written in message, while semantic meaning aims to understand the actual meaning of the text. Both are crucial for comprehensively analyzing and understanding human language.

V. Utalization of Hugging Face AI Model

Hugging Face is widely used across various industries for tasks such as sentiment analysis, translation services, content moderation, chatbots and virtual assistants, document summarization, voice recognition and image classification. In this paper we will focus more on the sentiment analysis and how changing parameters in text can impact getting different results. For that purpose, in this scenario is utilized the model nlptown/bert-base-multilingual-uncased-sentiment, shown in Fig.3 the implementation of the model.

```
# Load tokenizer and model
model_name = "nlptown/bert-base-multilingual-uncased-sentiment"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)
```

Fig.3 Implementation of the AI model

This model is multilanguage and can recognize different languages in text documents. While detecting the semantic meaning and sentiment analysis in sentences we will use parallel two languages for comparation: English and Macedonian.

A. Semanic meaning and sentiment analize of the sentences in English

1) With punctuation marks

This model was tested by sending a list of sentences that are well formatted and using punctuation marks Fig.4.

```
PS C:\Users\Aneta\Desktop\Hugging Face> python .\test3.py
Sentence: "This book was truly inspiring and I recommend it to everyone!♥"
Sentiment: {'label': '5 stars', 'score': 0.8869526386260986}
```

Fig.4. Results after sending a sentence with using emoji and punctuation marks

The results were also analyzed for the semantic meaning, in the above sentence we can determine that the user experience is highly positive. The model assigned the highest grade, despite the confidence score being around 88%.

Interesting here is that in Fig.4. emoji is added at the end of the sentence. After testing that part and deleting the emoji we got almost the same results from the model, with only difference in the score precent 87%, Fig.5.

```
Sentence: "This book was truly inspiring and I recommend it to everyone!"
Sentiment: {'label': '5 stars', 'score': 0.8756318092346191}

Sentence: "I didn't like the book; the story was too boring and predictable!"
Sentiment: {'label': '2 stars', 'score': 0.5620238184928894}

Sentence: "The book was well-written, but it didn't leave a strong impression on me."
Sentiment: {'label': '3 stars', 'score': 0.5407010912895203}
```

Fig.5. Results for the text classification by the model

With the above results reading the sentences with the marks that are used is understandable, for the human but also for the AI model. It is clearly summarized that the second sentence is with negative meaning and the third sentence is with neutral meaning.

Reading the metrics and the score value the conclusion is that the border between the levels of negative, neutral or positive reaction is based on the characters that are included in the text.

2) Without punctuation marks

Without punctuation marks, sentences can become confusing and their meaning ambiguous, especially for humans. But how do the AI models behave if there are not any punctuation marks in the sentence?

According to the sentences that were used for this testing, it can be notable that the present of how comfortable the model is reducing as shown at the Fig.6.

```
Sentence: "This book was truly inspiring and I recommend it to everyone"
Sentiment: {'label': '5 stars', 'score': 0.8503599762916565}

Sentence: "I didn't like the book the story was too boring and predictable"
Sentiment: {'label': '2 stars', 'score': 0.5230246782302856}

Sentence: "The book was well-written but it didnt leave a strong impression on me"
Sentiment: {'label': '3 stars', 'score': 0.49882516264915466}
```

Fig.6. Representation of the results from the sentences without punctuation marks

If we compare the results now without punctuation marks and before, example for the first sentence: "This book was truly inspiring and I recommend it to everyone", the semantic meaning is still clear, but the model reduce the score from 88% to 85%. Imagine the difference and how much this scope will be reduced if the text is bigger and if the checking is based on many sentences without knowing where the sentence is beginning or ending.

B. Semantic meaning and sentiment analize of the sentences in Macedonian

1) With punctuation marks

Similar to the previous example we did the same analysis, but now utilizing a different language for the same meaning of the sentences. In Fig.7. there is a representation of one sentence with additional emojis.

```
Sentence: „Не ми се допадна книгата приказната беше премногу досадна и предвидлива 😩😩😩"
Sentiment: {'label': '2 stars', 'score': 0.3722081780433655}
```

Fig. 7. Sentence in Macedonian language with using an emojis

The score is around 37% and the feedback of the model is that the semantic meaning of the sentence is with negative understanding.

```
Sentence: „Не ми се допадна книгата, приказната беше премногу досадна и предвидлива."
Sentiment: {'label': '2 stars', 'score': 0.4401301145553589}

Sentence: „Книгата беше добро напишана, но не ме остави со некој посебен впечаток."
Sentiment: {'label': '3 stars', 'score': 0.5224579572677612}
```

Fig.8. Some other sentences on Macedonian

It is notable that comparing the sentences with the English meaning and Macedonian meaning the model is doing good classification and the results are approximately the same Fig.8.

The results of accuracy depend on how well the model is trained to recognize the characters and the meaning of the sentences.

2) Without punctuation marks

Already we mentioned how difficult can be reading of the text without punctuation marks for humans but also for the models. If we are not sending the corresponding data, the models can bring decisions in the wrong way.

```
PS C:\Users\Aneta\Desktop\Hugging Face> python .\test3.py
Sentence: „Оваа книга беше навистина инспиративна и ја препорачувам на секого"
Sentiment: {'label': '5 stars', 'score': 0.255951195955276 5}

Sentence: „Не ми се допадна книгата приказната беше премногу досадна и предвидлива"
Sentiment: {'label': '2 stars', 'score': 0.3717970848083496}

Sentence: „Книгата беше добро напишана но не ме остави со некој посебен впечаток"
Sentiment: {'label': '3 stars', 'score': 0.4831456243991852}
```

Fig.9. Simple sentences in Macedonian language

Due to the brevity of the sentences here we can see that the model behaves as expected and there are small deviations in percentages, compared to when punctuation is used Fig.9.

VI. UTALIZATION OF POWER AUTOMATE AND AI CONNECTORS

In this section we will explore the synergy between the services and how utilization of AI connectors can enhance productivity and reduce human errors. There are many areas of usage as decision making, sentiment analysis of customer feedback or email responses in automated way, document processing for different organizations, integration with other services and workflow customizations.

It is used model for sentiment analysis, Fig.10., with sending the same example set of data as was tested with the Hugging Face model.

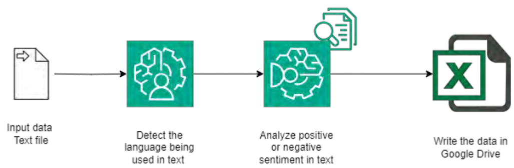

Fig.10. Used AI model in Power Automate flow

Power Automate has many features including AI Builder utilization. With the AI builder there are a lot of capabilities that can be achieved, such as recognition of images, text classification s, automatically filling documents, sending approval messages etc.

In the following part we will focus on the results from the analysis in testing with the builder connector, where we will conclude what are the differences.

A. Semanic meaning and sentiment analize of the sentences in English

1) With punctuation marks

Utilization of the punctuation marks and additional emojis in Power Automate connectors are having results as the human will give while reading the sentences. That we can see on the bellow Fig.11., the model with confidence of approximately

80%, that is high level of confidence of the model – bring decision that the sentiment of the meaning is negative.

Fig.11. Sentiment analysis with included punctuation

2) Without punctuation marks

An interesting comparison can be seen in the analysis of a sentence with and without punctuation marks. When punctuation marks were used, the model determined that the sentiment was negative, with a confidence score of 80%. However, using the same sentence without punctuation marks yielded completely different results: the model classified the sentiment as positive, but with a confidence score of around 49%.

This example clearly demonstrates how changes in sentence structure and the inclusion of additional characters can significantly impact the model's confidence score. The tokenization of sentences and the separation of words, as well as punctuation marks, are crucial factors in the model's final decision-making process.

B. Semantic meaning and sentiment analize of the sentences in Macedonian

Unfortunately, the model in Power Automate didn't support the Macedonian language and we couldn't test the collected data on it. In addition, we are providing which languages are supported by the AI Builder Fig.12.

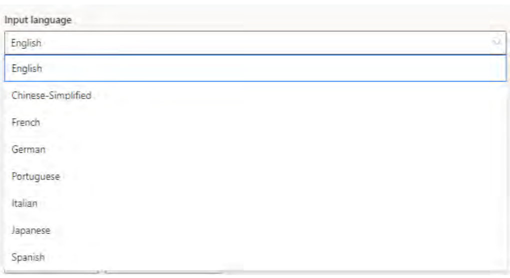

Fig. 12. Supported languages by PA

VII. Comparative Sentiment Analysis From Utilization The Both Models

Based on the results that were accumulated from the separate analyzes on the input text data in Hugging face model and Power Automate model we detected various end parameters. These results will be used to visualize the differences and statistics from the model comparisons.

A. Comparation results between the models for sentences in English

In Power Automate and Hugging face for the same data that we tested we got the following results from Fig.13. for

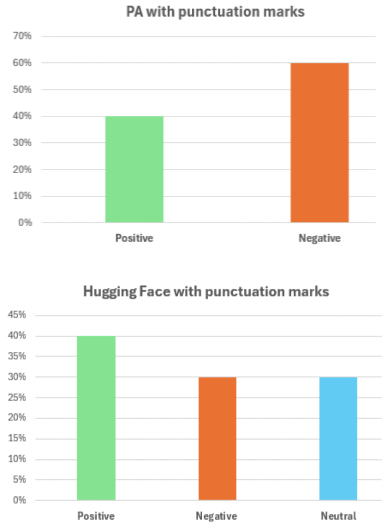

Fig.13. Results with punctuation marks

The data that was selected for testing in both models without punctuation marks give the bellow results Fig.14:

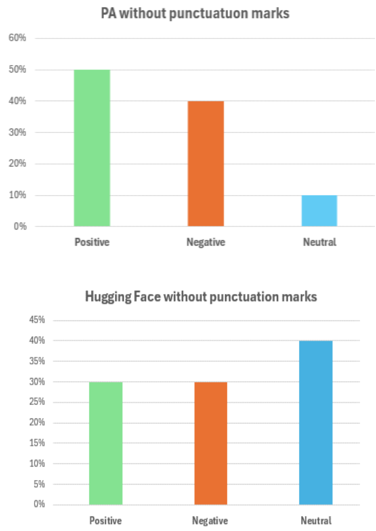

Fig.14. Results without punctuation marks

Statistics can point out the main difference between the fact when some special characters are used or not. The metrics show how the confidence of the models can change if the prompt that is sent is not the same, even if the actual meaning of the sentence is same.

B. Visualisation of the results for the text in Macedonian language from Hugging Face

Below on Fig.16. is shown while utilization of Hugging Face model, the percentage of confident from the model is changing from positive 38% with punctuation marks to 30% without them. Devise for the negative from 31% it is increasing to 40 %.

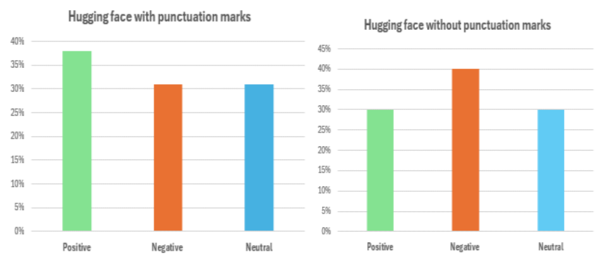

Fig.16. Text with/without punctuation in Macedonian language

VIII. Conclusion

Semantic meaning of the words/sentences is crucial for effective communication, allowing us to convey and understand ideas accurately. In the realm of natural language processing (NLP), it involves tasks like named entity recognition, sentiment analysis, and machine translation, which rely on understanding the underlying meaning of the text to provide accurate and relevant outputs.

This paper focuses on the reliability of text sentiment models and the accuracy of information processed by bots. The analysis demonstrates that small changes in text, such as adding or removing punctuation marks, can yield different results across various models. This highlights the significant impact of AI and raises several important questions:

- How can we determine if our chosen model is the right one?
- Can we fully trust the analysis provided by AI models?

While AI models have made significant growth, several factors can influence of their analysis: data quality, model testing, context and nuance. The semantic analysis conducted on both models reveals that their results can vary, prompting further inquiry into the reliability. When some organization decides to integrate an AI model into an organization, it is crucial to thoroughly test the model with various scenarios to prevent future issues with production data. Human oversight can catch errors that AI might miss and provide a deeper understanding of complex scenarios.

In conclusion, even if AI models are strong tools, it is wise to proceed carefully when analyzing them and to support their conclusions with thorough testing, continuous assessment, and human expertise.

References

[1] Y. Lierler and M. Truszczynski, “Transition systems for model generators—A unifying approach”, Cambridge University Press, vol. 11, pp. 629-646, July 2011.

[2] R. J. Sternberg, “Human Intelligence”, Cambridge University Press, https://www.book2look.com/book/9781108703864 , pp.20.-36, 2019

[3] B. Lutkevich, “Hugging Face”, https://www.techtarget.com/whatis/definition/Hugging-Face , TechTarget, last accessed 2024/06/23.

[4] Microsoft Azure, “Automate processes with Robotic Process Automation and Power Automate for desktop**”,** https://learn.microsoft.com/en-us/training/paths/work-automation-flow/ , last accessed 2024/06/23.

[5] A. Naqvi and J. M. Munoz: “Handbook of Artificial Intelligence and Robotic Process Automation Policy and Government Applications”, Cambridge University Press, Handbook of Artificial Intelligence and Robotic Process Automation (cambridge.org), January 2022.

[6] J. Ferrer, “What is Hugging Face? The AI Community's Open-Source Oasis”, Data Camp, What is Hugging Face? The AI Community's Open-Source Oasis | DataCamp , last accessed 2024/06/23.

[7] A. Kumar, “Hugging face transformers Hello World: Python Example”, https://vitalflux.com/huggingface-hello-world-python-example-transformers/ , August 2023.

[8] P. Ngwa and I. Ngaruye, “Big Data Analytics for Predictive System Maintenance Using Machine Learning Models” , World Scientific, https://www.worldscientific.com/doi/10.1142/S2424922X23500018 , Vol.15, 2023.

[9] A. Yu, “100 days of Code: the complete Python Pro Bootcamp”, Udemy Course, https://www.udemy.com/course/100-days-of-code/?couponCode=LETSLEARNNOW , last accessed 2024/06/24.

[10] Hugging face, “Transformers”, https://huggingface.co/docs/transformers/en/index , last accessed 2024/06/24.

[11] PyTorch official documentation, https://pytorch.org/foundation, last accessed 2024/06/24.

[12] TensorFlow official documentation, https://www.tensorflow.org/, last accessed 2024/06/24.

[13] A.A. Awan, “What is tokenization”, https://www.datacamp.com/blog/what-is-tokenization , DataCamp, September 2023, last accessed 2024/06/24Hugging Face official documentation, “Tokenizer”, https://huggingface.co/docs/transformers/en/main_classes/tokenizer , last accessed 2024/06/24.

[14] Hugging Face official documentation, “Tokenizer”, https://huggingface.co/docs/transformers/en/main_classes/tokenizer , last accessed 2024/06/24.

[15] Microsoft Azure, “Microsoft Power Automate”, https://www.microsoft.com/en-us/power-platform/products/power-automate , last accessed 2024/06/24.

[16] IWConnect, “RPA Optimization: Advanced Monitoring Solutions for Cloud and Desktop Bots”, https://iwconnect.com/rpa-optimization-advanced-monitoring-solutions-for-cloud-and-desktop-bots/, April 2024, last accessed 2024/06/24.

[17] Microsoft Azure documentation, “Overview of AI Builder”, https://learn.microsoft.com/en-us/ai-builder/overview ,January 2024, last accessed 2024/06/24.

[18] Microsoft Azure documentation, “Extract text from objects using Power Automate and AI Builder”, https://learn.microsoft.com/en-us/azure/architecture/example-scenario/ai/extract-object-text, last accessed 2024/06/24