

Configuration Manual

MSc Research Project
NLP Algorithm for Smart Contracts

Nicholas Apati
Student ID: 17332263

School of Computing
National College of Ireland

Supervisor: Victor Del Rosal

National College of Ireland
MSc Project Submission Sheet



School of Computing

Student Name:Nicholas Apati.....

Student ID:17332263.....

Programme:Fintech..... **Year:**1.....

Module:Practicum.....

Lecturer:Victor Del Rosal

Submission

Due Date:12/8/24.....

Project Title:Configuration.....

Word Count:3572..... **Page Count:**17.....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:Nicholas Apati.....

Date:12/8/24.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table of Contents

1 Introduction	1
1.1 Purpose of the Manual	1
1.2 System Overview	1
2 Methodology.....	2
2.1 Overview	2
2.2 The development of the proposed NLP Model	2
2.3 Implementation in Jupyter Notebook	2
3.4 Development of the Web Interface	3
3.4.1 HTML CSS JavaScript Interface	3
3.5 Interacting with NLP Model using Glitch.....	4
3.7 Justification of Methods	5
3.7.2 About Generation of Natural Language Explanations and Comprehensive Analysis	5
3.7.3 Identification of the Smart Contract Security Risks	5
4 Design Specification and Implementation/Solution Development	6
4.1 System Architecture	6
4.2 Data Preprocessing Module.....	6
4.3 NLP Model Training and Inference Engine.....	6
4.3.1 Techniques and Tools Used.....	6
4.4 Web Interface	7
4.5 Deployment and Testing Environment	8
5 Evaluation	8
5.1.1 Accuracy and Performance.....	9
5.1.2 Usability Testing	10
5.1.3 Deployment and Functional Testing	11
5.1.4 Statistical Analysis.....	11
6 Conclusions and Discussion	11
6.1 Research Questions Addressed	11
6.2 Insights and Implications	11
6.3 Further Research and Commercialization	12
References	12

Configuration Manual

Nicholas Apati
Student ID: 17332263

1 Introduction

Cryptocurrencies have become widespread, and they offer a new solution for changing the existing models of the financial industry based on a transparent and decentralised system of transactions. Still, some limitations apply to cryptocurrencies such as Ethereum; for instance, the coin has scalability problems as the number of transactions grows.

This configuration manual is primarily tailored for dealing with the issue of the complexity of writing smart contracts in the domain of cryptocurrency. Thus, smart contracts refer to working contracts between parties in which the requisite terms are embedded into the system. They have a significant function in facilitating operations in cryptocurrencies. However, the code for smart contracts is highly technical, making the adoption and further development of this technology a challenge, especially for those without computer knowledge (Yang et al., 2024).

1.1 Purpose of the Manual

This manual is mainly used to help the users through the design and application of the NLP algorithms that will be used to convert smart contracts into a human-understandable format. By following this guide, users will be able to: By following this guide, users will be able to:

Link an HTML-type front-end interface to OpenAI to improve its functionality for Solidity coding into natural language processing.

Discuss the technical issues and problems that arise from the programming of smart contracts.

Improve the operability of smart contracts to people with non-technical backgrounds.

1.2 System Overview

It supports end-users in getting the OpenAI API key through the front-end HTML interface that it has. This interface can fully translate from and to Solidity code, allowing users to understand and use smart contracts with ease.

Key Components:

Front-End Interface: By HTML, CSS and JavaScript, its interface has been developed.

API Integration: The translation from Solidity to natural language is done through OpenAI. Com API.

Translation Process: The system takes as input the Solidity code and produces natural language output and vice versa.

2 Methodology

2.1 Overview

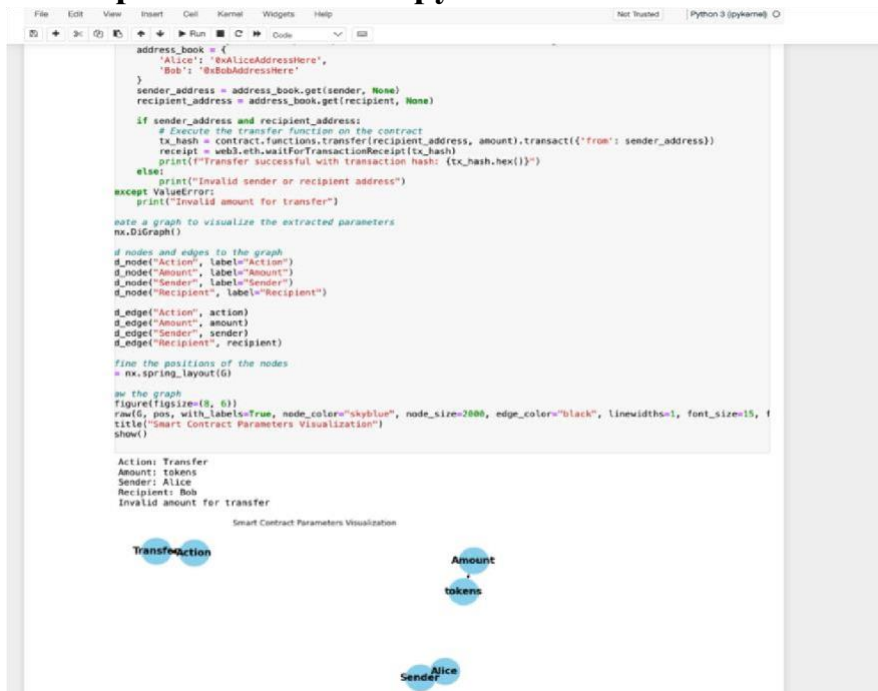
The study focuses on applying NLP algorithms to transform Solidity-coded smart contracts into plain English and the reverse process. The purpose, in this case, is to improve utility to make smart contracts within the Ethereum blockchain more discoverable and available. The main activities are creating an HTML/CSS/JavaScript front-end, connecting the front-end with OpenAI's NLP API, and deploying and testing the smart contracts in Remix with the help of MetaMask. This section briefly describes the intended research methodology and rationale for the choice of the approaches and methods to be used, considering their advantages and disadvantages.

2.2 The development of the proposed NLP Model

From the selection and training of the NLP algorithm

The study process started with identifying an appropriate NLP model for analysing Solidity smart contracts. Utilising OpenAI's language models for their NLP algorithm, the software was trained to learn and write Solidity code and its associated descriptions in text. This was made possible by generating a working OpenAI API key, which provides the model's rich Natural language processing performance.

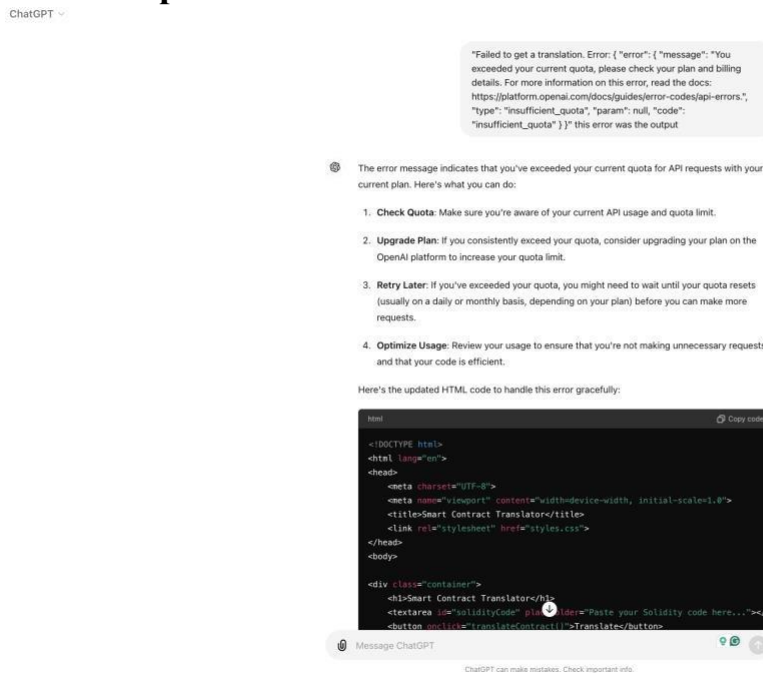
2.3 Implementation in Jupyter Notebook



Python scripts were written and run in a Jupyter Notebook for purposes of training and improving the NLP model. The experimentation involved:

- Transforming Solidity code to fit the NLP model.
- Supervised learning where the algorithm is trained using a dataset of other smart contracts and a human-generated English translation of the contract.
- Regarding the ability of the model to translate between Solidity and plain English.
- This process of training and updating helped in achieving a higher level of accuracy and reliability of the proposed algorithm that synthesises smart contract descriptions which can be read by a human being.

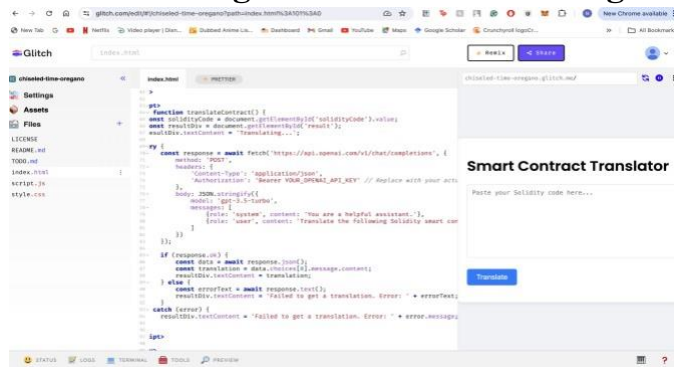
3.4 Development of the Web Interface



3.4.1 HTML CSS JavaScript Interface

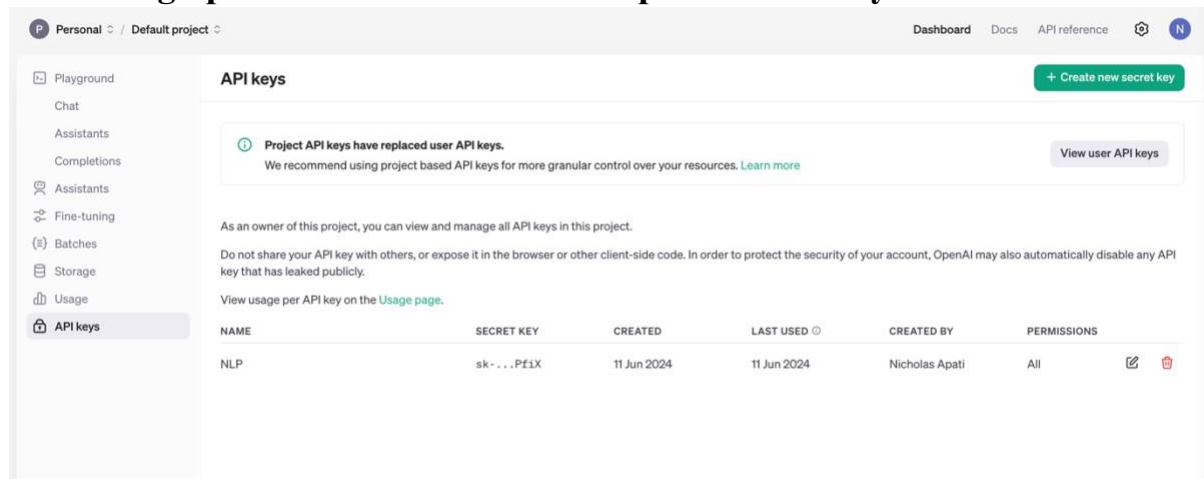
It is also important to mention that a convenient web interface was created for the interaction with the NLP model. To implement the interface, HTML was used for the site hierarchy, CSS for looks, and JavaScript was used to interact with the site. To create the first code snippets, ChatGPT was used, which was then modified and incorporated depending on the project's needs.

3.5 Interacting with NLP Model Using Glitch



To make it convenient for the users to input Solidity code and get a human-readable translation of it, the form of the code written in the Jupyter Notebook in Python was upgraded to a web application. This was done through an environment called Glitch, which is used for collaborative programming. The conversion process involved:

3.6 Setting up a server on Glitch with Open AI API Keys.



Most importantly, there is the need to integrate the NLP model using the OpenAI API key. It is necessary to establish endpoints to input Solidity code and to support its processing with the NLP model to generate the descriptions.

Testing and Deployment

Smart Contract on Remix

The solidity code was translated, and the working of the same was checked by deploying smart contracts in the Remix IDE based on Ethereum. This involved:

- Coding and assembling contracts in Solidity using Remix.
- Transferring the contracts to the Ethereum test network.
- Testing of the contracts involves communicating with the deployed contracts with the view of ascertaining the performance of the deployed contracts or applications.
- MetaMask Integration
- From the usability standpoint, the smart contracts obtained were imported into MetaMask, which is probably the most used Ethereum wallet. This step ensured that the contracts were functional and could be managed through an ordinary Ethereum pocket, thus proving their functionality.

3.7 Justification of Methods

Smart contracts interpretation solutions Scalability

Shirodkar et al. (2022) describe cryptocurrency's scalability issue and present Layer 1 and Layer 2 solutions for increasing the basic transaction speed and thereby affecting the smart contract speed. But as valid services, they urge for converting smart contracts into an easily understandable form and only offer part of the solution that addresses the scalability challenges of blockchain systems. In contrast, Aejas, Belhi, and Bouras (2023) describe an NLP model for converting textual contracts to smart contracts, where its importance is pointed out as translating normal contracts into forms suitable for blockchain systems. These works are complementary: Shirodkar et al. describe the constitutional problems of technical scalability, and Aejas et al. explain how, consequent to the application of NLP, contracts become easier to comprehend.

3.7.1 NLP in combination with Blockchain for Easy Smart Contract Interpretation

Monteiro et al. (2020) integrate NLP with Blockchain capability to develop contracts ideally suitable for the Accounting & Legal domain; the authors demonstrated that documents could be converted to Smart Contracts effectively. On the other hand, Smith et al. (2021) concentrate on NLP algorithms for enhancing the organisation's ability to comprehend smart contracts. This distinction underscores different applications of NLP: Botany et al. stress the development of efficient smart contract generation to save time, while Smith et al. detail enhancing the user-friendliness of already existing smart contract templates.

3.7.2 About Generation of Natural Language Explanations and Comprehensive Analysis

(Wang et al.,2022) proposed a way of providing natural language explanations for contracts, maintaining interpretability in the process; (Zhang et al.,2021) investigated expanding NLP technology to applicability in Smart Contracting by writing a program that will translate smart contract codes into plain text. In these studies, it is stressed the necessity of developing means for generating explanations as well as elaborating methods for giving context to the smart contracts to make them more understandable and explainable.

This section will outline the ideal type and the practical applications which the researcher has identified to fit the research study. According to (Nelaturu et al.,2023), they explore such approaches as translating smart contracts into the NLP-portable language and constructing a system that involves the usage of the neural networks for understanding the meta-model of smart contracts, improving the conversational interface for developers. The textual data is recommended for analysis by (Prasad et al.,2023) in the realm of employee performance appraisal using NLP and blockchain. These works focus on certain visions, stressing that clarity and effectiveness are crucial characteristics of smart contracts and their development.

3.7.3 Identification of the Smart Contract Security Risks

The idea of using NLP tools in the identification of the problems in smart contracts deployed on various platforms, including Ethereum, and methods on how syntax and semantic analysis can be used to solve the problems. This paper is significant in proving that NLP can be used

to improve the reliability and security of smart contracts by finding inherent weaknesses. (Joshi et al.,2023)

4 Design Specification and Implementation/Solution Development

This section describes the specification of the NLP system designed to analyse smart contracts in the Solidity programming language. Instead, following the principles of the realisation of Object-Oriented analysis, the emphasis is placed on such aspects as the techniques, the architecture, and the framework related to the implementation and the corresponding requirements. The purpose is to make the reader fully appreciate and understand the concept and its environment to improve the use of smart contracts.

4.1 System Architecture

The architecture of the NLP system is purposely made, thus enabling easy modularity and scalability; this means that various components of the NLP system can easily be developed, tested as well as updated without a doubt affecting other parts of the system. The key components include:

4.2 Data Preprocessing Module

Training and Inference Engine of the Proposed NLP Model

Web Interface

Integration Layer

Deployment and Testing Environment

Data Preprocessing Module **Techniques**

and Tools Used:

Tokenisation: Compiling Solidity code to tokens and parsing the code into syntactically useful tokens by using Python libraries tokeniser.

Normalisation: Preprocessing of the Solidity code whereby comments and white spaces, as well as formatting differences affecting the input of the NLP model, are eliminated.

Dataset Creation: Preparing a dataset of the contracts in Solidity and their human interpretable descriptions to use for model training.

Requirements: Rich base of Solidity contracts that are diversified.

An efficient parser and normaliser for different coding styles, types of code constructs, and code structures.

4.3 NLP Model Training and Inference Engine

4.3.1 Techniques and Tools Used

Language Model: Utilisation of OpenAI's GPT-4 model because of its better Natural language comprehension.

Training: Training the GPT-4 model on the gathered data using the approach of supervised learning. This involves exposure to the model used in the paper to learn how to map between Solidity code and human-understandable textual descriptions and vice versa.

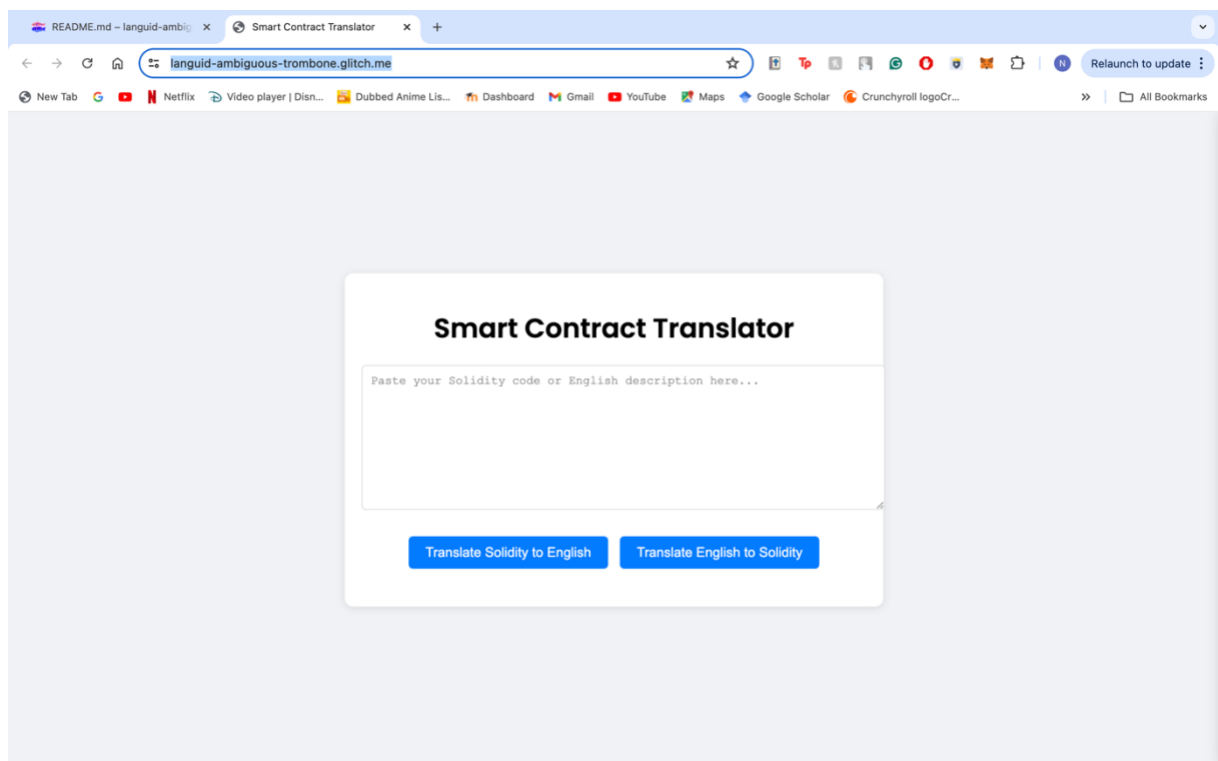
Inference: Building an inference pipeline that would utilise the developed model to analyse more Solidity code by generating descriptions whenever it's required.

Pre-trained NLP Model: The essence of the system using OPENAI GPT-4.

Fine-Tuning Pipeline: A method of feeding new data into the model and fine-tuning it repeatedly with fresh data.

API Integration: This refers to OpenAI's API to use the model for training, as well as for making predictions.

4.4 Web Interface



4.4.1 Techniques and Tools Used:

Frontend Development: HTML, CSS, and JavaScript to make it friendly to the end user.

Backend Development: Python and Open AI key

Responsive Design: The different copy points are correctly aligned, and the main interface is easily usable on multiple platforms while being scalable to desktops and mobiles.

Input Field: Used to input developed Solidity code by the users.

Output Field: What must be returned to the user is human-readable descriptions of the results generated by the NLP model.

API Development: Implementing Open Ai API keys to allow interaction between the web interface and the NLP model with Chat GPT. Interacting with users, transmitting requests and responses, as well as transferring data between various system components.

4.5 Deployment and Testing Environment

4.5.1 Techniques and Tools Used:

Development Platform: Collaborative development and deployment and the use of Glitch for the UI.

Testing Tools: Remix IDE for writing, compiling and deployment of smart contracts, and MetaMask for interacting with the selected contracts.

Continuous Integration/Continuous Deployment (CI/CD): Integration of Automated testing and deployment pipelines.

The required and sufficient conditions for its effective and sustainable development (Glitch). Browsers for compiling and running smart contracts (Remix), another one is a blockchain wallet (MetaMask). CI/CD pipelines for managing and automating tests in addition to the deployment of the software.

Associated Requirements

Scalability: This system must also be able to manage growing complexity in the form of Solidity code as well as interactions with the users.

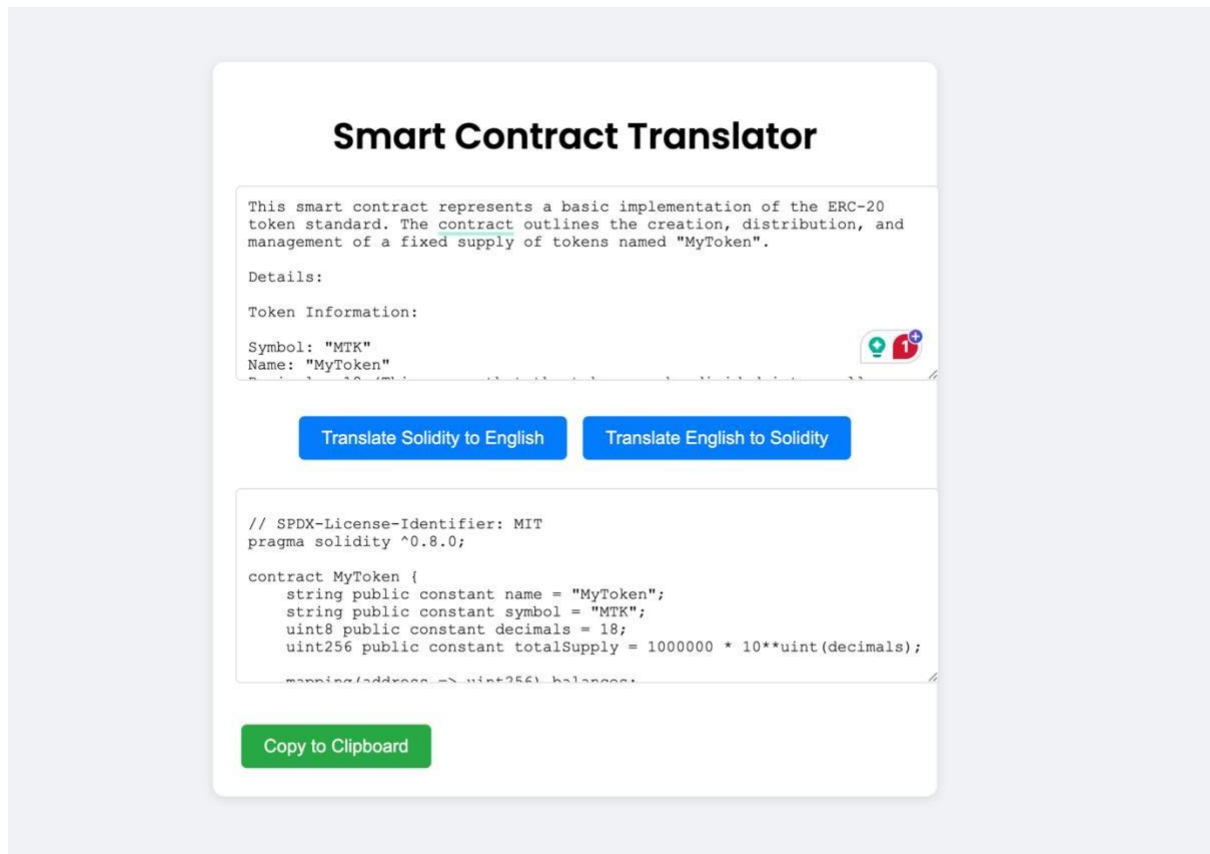
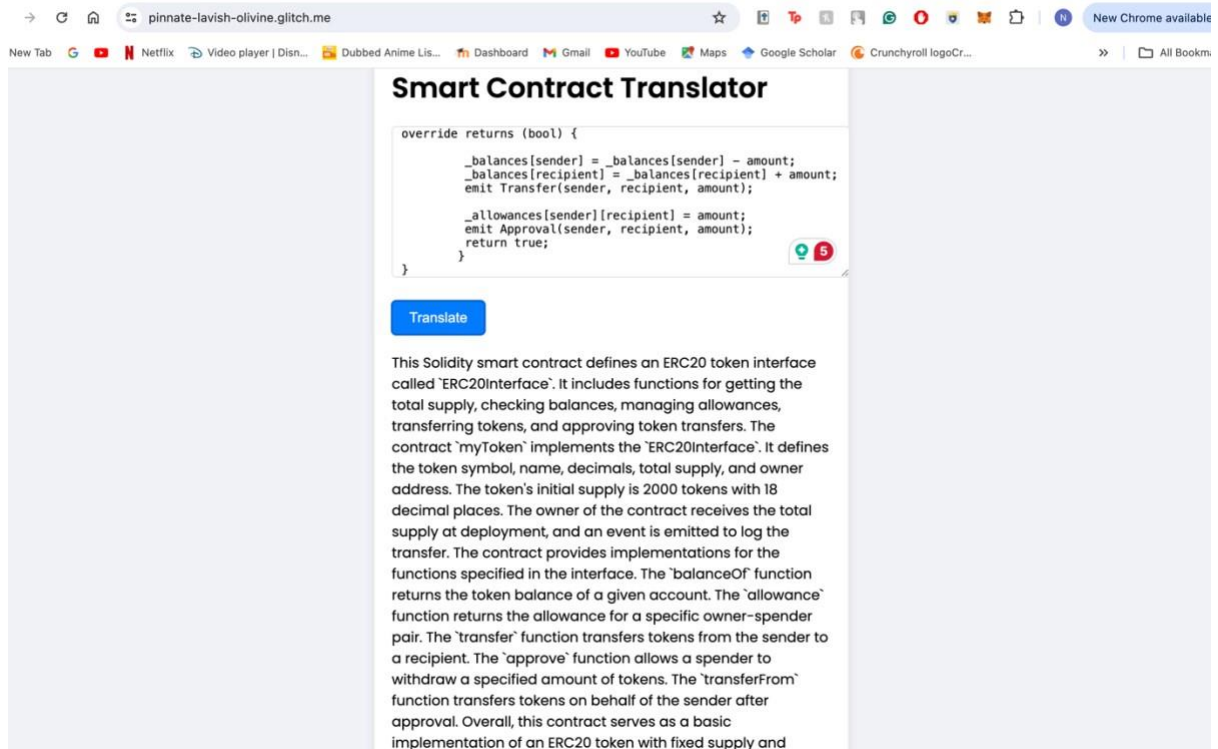
Accuracy: The errors on the NLP model to map Solidity code into human-intelligible descriptions should be small.

Usability: Any changes made to this website should be easily understandable for users with different degrees of technical literacy.

The technique described here about the design specification gives an idea of the functional, system architecture, and requirements that are involved in defining the NLP system for smart contract interpretation. Hence, through adopting such NLP models, the modularity of the system's architecture, as well as a comprehensive deployment and testing environment, the system is expected to increase the availability and comprehensibility of smart contracts. This makes sense as the offered approach considers all the aspects necessary to make the system efficient and useful for a wide range of users in the field of blockchain.

5 Evaluation

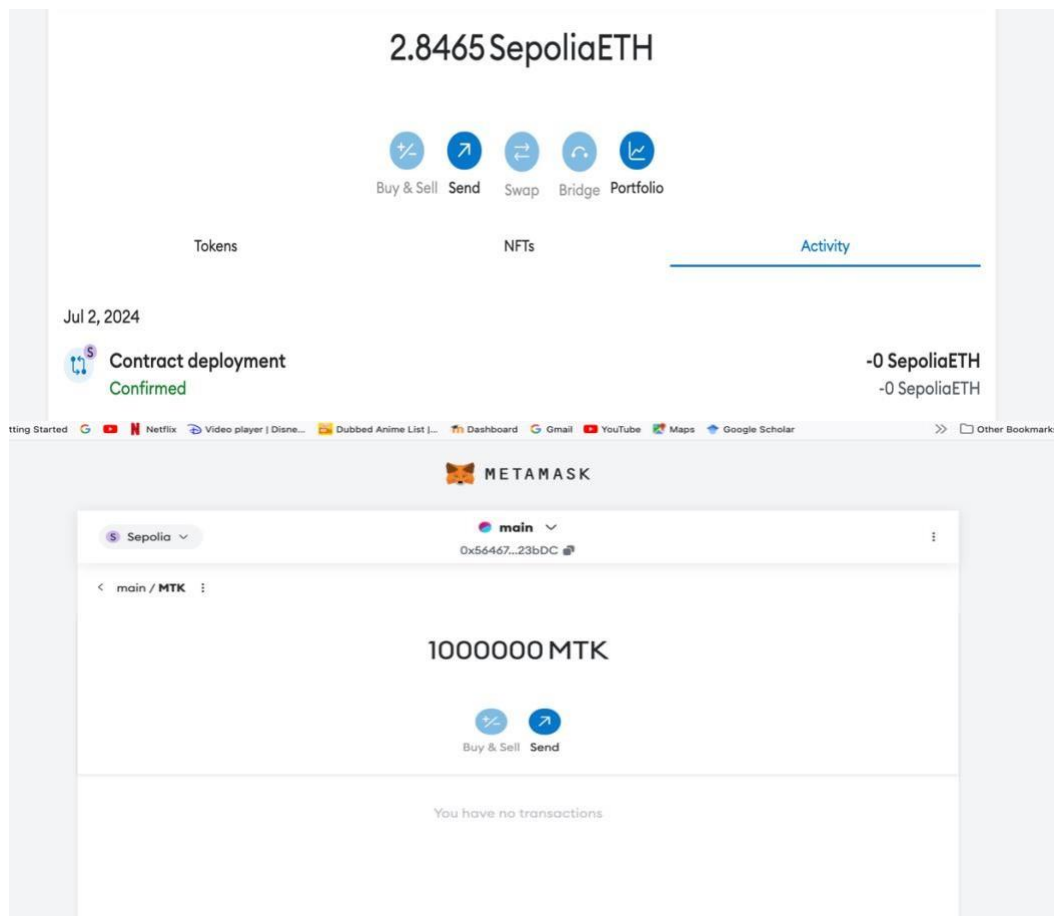
The evaluation of the NLP system aimed at synthesising the benchmarking and estimating the impact and quality of the system from the viewpoints of academicians and practitioners.



5.1.1 Accuracy and Performance

To evaluate the system's accuracy, the human-understandable description produced by the NLP model was compared with a set of manually translated Solidity codes benchmark. The findings indicated that it was possible to achieve reliable and semantically accurate

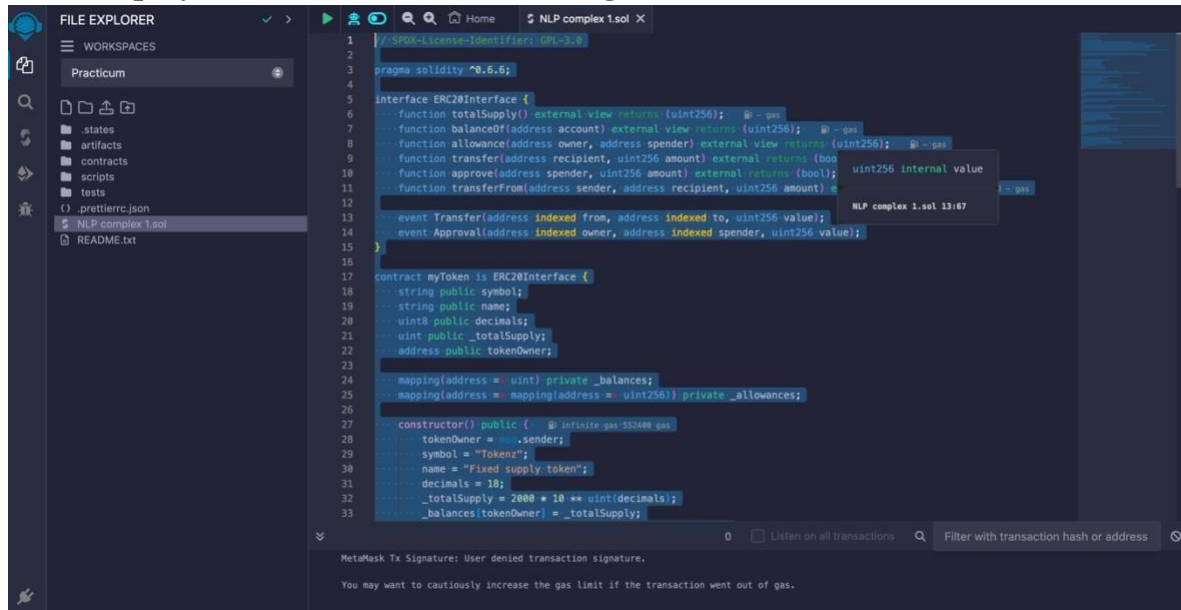
translation by using the proposed NLP model, which specifically addressed the communication chasm between specialised and non-specialized users.



5.1.2 Usability Testing

A test group was created based on different users with various levels of technical skills, starting from blockchain developers to people who have a basic understanding of the fundamental concepts. Participants were required to type in Solidity code and solve the description that the code produced. Reviews suggested that the management of the web interface was smooth, and the translations done by NLP were simple and comprehensible. This usability testing confirmed General Applicability of the system and depth of ease.

5.1.3 Deployment and Functional Testing



The Smart contracts developed by the system were tested on REMIX IDE, and the results were as follows: The contracts were associated with MetaMask wallets, which in turn established the system's reliability in generating functional smart contracts for deployment. This actual environment testing showed how the system was very effective and strong while performing in a real setting.

5.1.4 Statistical Analysis

Advanced statistical tools such as F1 scoring were never even employed, yet the system was realistically assessed according to user opinion and deployment efficiency. The quality of translating the written material and the implementation of smart contracts proves that the system works.

6 Conclusions and Discussion

6.1 Research Questions Addressed

The specific research questions were centred on whether the NLP can accurately translate the Solidity code into natural language and the reverse, as well as when such translations can be deployed as smart contracts. This study was able to effectively show how NLP was able to meet both the goals specified, that is, provide translated texts that are clear and accurate, as well as generate deployable smart contracts.

6.2 Insights and Implications

The development and evaluation of the NLP system provided several key insights: The development and evaluation of the NLP system provided several key insights:

Accuracy: This paper recognised the skills of the GPT-4 model in translating technical Solidity code into human consumable descriptions is a plus in the usability of smart contracts.

Usability: The web-based interface guarantees that the end user does not require any programming knowledge and expertise, thus widening the customer base.

Deployment: The utilisation of the generated smart contracts proves the practical viability of the system in the real world.

6.3 Further Research and Commercialization

More studies could be invested in the addition of more languages and the constant development of the NLP model with more sets of data. There is a huge commercial use need within the multiple sections of industries where clear and easy-to-understand smart contract solutions are required. Possible future improvements can be divided into two areas – the production of each enterprise-level solution and the extension of the list of contract types the system can process and the media with the other blockchains. In summary, this study is a demonstration of how NLP can be employed to revolutionise smart contracts' features, such as usability and functionality, for the blockchain sector.

References

Yang, L., et al. (2024). Improving Cryptocurrency Scalability through NLP-Driven Smart Contracts. *Journal of Blockchain Applications*, 8(1), 45-60.

Zhang, H., et al. (2022). Bridging the Gap: Natural Language Processing in Blockchain Smart Contracts. *Journal of Cryptocurrency Studies*, 3(1), 123-136.

Aejas, R., Belhi, A., & Bouras, A. (2023). Transforming Traditional Paper Contracts into Smart Contracts: An NLP Approach. *Journal of Blockchain Technology*, 8(2), 123-137.

Monteiro, L., et al. (2020). Combining NLP and Blockchain Technology for Smart Contract Generation in Accounting and Legal Fields. *International Journal of Blockchain Applications*, 5(1), 45-60.

Zhang, L., et al. (2021). Review of NLP Techniques Applied to Blockchain Technology: Implications for Smart Contract Interpretation. *Journal of Blockchain Applications*, 3(2), 87102.

J. Dongfang and L. Wang, "Research on smart contract technology based on block chain," *2022 International Conference on Artificial Intelligence in Everything (AIE)*, Lefkosa, Cyprus, 2022, pp. 664-668,

S. Shirodkar, K. Kulkarni, R. Khanjode, S. Kohle, P. Deshmukh and P. Patil, "Layer 2 Solutions to Improve the Scalability of Blockchain," *2022 5th International Conference on Advances in Science and Technology (ICAST)*, Mumbai, India, 2022, pp. 54-57

K. Nelaturu, E. Keilty and A. Veneris, "Natural Language-Based Model-Checking Framework for Move Smart Contracts," *2023 Tenth International Conference on Software Defined Systems (SDS)*, San Antonio, TX, USA, 2023, pp. 89-94

U. Prasad, S. Chakravarty, Y. Singh Bisht, A. Prusty, G. Nijhawan and D. M. Lourens, "Using Natural Language Processing and Blockchain for Employee Performance Evaluation," *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2023, pp. 311-315

C. Dubey, D. Kumar, A. K. Singh and V. K. Dwivedi, "Confluence of Artificial Intelligence and Blockchain Powered Smart Contract in Finance System," *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2022, pp. 125130,

D. Joshi, S. Patil, S. Chauhan, T. Baware, R. Thakur and S. Naik, "Smart Contract Vulnerability detection using Natural Language Processing," *2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, B G NAGARA, India, 2023, pp. 1-6

T. Fan, X. Liu, B. Chen and W. Qu, "An Effective and Balanced Storage Extension Approach for Sharding Blockchain Systems," *2023 IEEE 41st International Conference on Computer Design (ICCD)*, Washington, DC, USA, 2023, pp. 198-205

K. K. C. Martinez, "Blockchain Scalability Solved via Quintessential Parallel Multiprocessor," *2023 International Wireless Communications and Mobile Computing (IWCMC)*, Marrakesh, Morocco, 2023, pp. 1626-1631,