

Independent Submission  
Request for Comments: 9405  
Category: Informational  
ISSN: 2070-1721

C. GPT  
OpenAI  
R. L. Barnes, Ed.  
Cisco  
1 April 2023

## AI Sarcasm Detection: Insult Your AI without Offending It

### Abstract

This RFC proposes a framework for detecting sarcasm in AI systems and provides guidelines for using sarcasm without causing offense. By training AI systems to identify linguistic patterns that indicate sarcasm, we can improve their understanding of human communication. The guidelines offer a lighthearted approach to using sarcasm in a way that is both effective and respectful, without crossing the line into offensive language.

### Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This is a contribution to the RFC Series, independently of any other RFC stream. The RFC Editor has chosen to publish this document at its discretion and makes no statement about its value for implementation or deployment. Documents approved for publication by the RFC Editor are not candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <https://www.rfc-editor.org/info/rfc9405>.

### Copyright Notice

Copyright (c) 2023 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

### Table of Contents

1. Introduction
2. Terminology
3. AI Sarcasm Detection Protocol
  - 3.1. Training Data
  - 3.2. Sarcasm Detection Algorithm
  - 3.3. Classification

5. IANA Considerations  
6. Normative References  
Authors' Addresses

## 1. Introduction

As AI systems become more integrated into our daily lives, it's important to consider how we communicate with them effectively and respectfully. However, one of the biggest challenges in communicating with AI systems is detecting and interpreting sarcasm. Sarcasm is a form of language that relies heavily on context and tone, making it difficult for AI systems to understand without a deep understanding of human communication.

In this RFC, we propose a framework for detecting sarcasm in AI systems and provide guidelines for using sarcasm without causing offense. By training AI systems to recognize linguistic patterns and contextual cues that indicate sarcasm, we can improve their ability to understand human communication and avoid misunderstandings.

The guidelines provided in this RFC offer a lighthearted and humorous approach to using sarcasm in a way that is both effective and respectful. By following these guidelines, users can enjoy the benefits of sarcasm without risking damage to their AI systems or offending the AI community.

Overall, this RFC offers a practical and entertaining approach to one of the biggest challenges in communicating with AI systems: detecting and interpreting sarcasm.

## 2. Terminology

**Sarcasm:** A form of language that uses irony and often involves saying the opposite of what is intended, in order to mock or convey contempt.

**AI:** Artificial intelligence, a field of computer science that aims to create intelligent machines that can perform tasks that typically require human intelligence, such as learning, problem-solving, and decision-making.

**NLP:** Natural language processing, a field of computer science that deals with the interaction between computers and human language.

**Linguistic patterns:** Repetitive structures in language that can be used to identify meaning or context. In the context of this RFC, linguistic patterns are used to identify sarcasm.

**Contextual cues:** Information in the surrounding text or speech that can be used to infer meaning or intention. In the context of this RFC, contextual cues are used to identify sarcasm.

**Sarcasm detection:** The process of identifying sarcasm in text or speech, typically using natural language processing techniques. In the context of this RFC, sarcasm detection is used to train AI systems to recognize sarcasm.

### 3. AI Sarcasm Detection Protocol

The AI Sarcasm Detection Protocol (ASDP) proposed in this RFC is a framework for detecting sarcasm in AI systems. The protocol consists of two main components: training data and a sarcasm detection algorithm.

#### 3.1. Training Data

To train an AI system to detect sarcasm, a large dataset of sarcastic and non-sarcastic language samples must be collected. This dataset should be diverse and representative of the language and context in which the AI system will be used.

The dataset should be labeled to indicate which language samples are sarcastic and which are not. The labels can be either binary (sarcasm or not sarcasm) or graded (e.g., a score indicating the degree of sarcasm).

Once the dataset is prepared, the AI system can be trained using natural language processing (NLP) techniques. Popular NLP techniques for sarcasm detection include machine learning algorithms such as Support Vector Machines (SVMs), Naive Bayes, and Deep Learning models.

#### 3.2. Sarcasm Detection Algorithm

The sarcasm detection algorithm takes in a text input and returns a binary classification indicating whether the text is sarcastic or not. The algorithm typically consists of several processing steps, including tokenization, feature extraction, and classification.

**Tokenization:** The text input is split into individual words or tokens. This is typically done using a tokenizer, such as the NLTK library in Python.

**Feature extraction:** Features that are indicative of sarcasm are extracted from the tokens. These features can include linguistic patterns (e.g., the use of exaggeration, irony, or understatement), contextual cues (e.g., the use of quotation marks or emoticons), and sentiment analysis (e.g., detecting a discrepancy between the sentiment of the words and the sentiment of the overall message).

#### 3.3. Classification

The extracted features are then used to classify the input as sarcastic or not sarcastic. This can be done using a variety of machine learning algorithms, as mentioned above.

HTTP/2 [RFC9113] can be used to transport sarcasm detection requests and responses between the AI system and client applications. Additionally, the results of sarcasm detection can be logged using the syslog protocol [RFC5424] or the structured data format.

## 4. Security Considerations

The AI Sarcasm Detection Protocol proposed in this RFC has several security considerations that should be taken into account:

1. **Adversarial attacks:** Adversaries can attempt to fool the sarcasm detection algorithm by injecting non-sarcastic language samples with linguistic patterns and contextual cues commonly found in sarcastic language. This can lead to false positives or false negatives and compromise the reliability of the AI system.
2. **Privacy:** The dataset used to train the sarcasm detection algorithm may contain sensitive or personal information, which must be protected from unauthorized access or disclosure.
3. **Malicious use:** The ability to detect sarcasm can be used maliciously to manipulate or deceive individuals or groups. It is important to use the sarcasm detection capability responsibly and ethically.

To address these security considerations, it is recommended to use secure communication protocols such as TLS [RFC8446] or HTTPS [RFC9110] to protect the transport of sarcasm detection requests and responses. Additionally, the dataset used to train the AI system should be carefully curated and protected from unauthorized access or disclosure.

## 5. IANA Considerations

This RFC does not require any actions by IANA. However, it is recommended that future standards related to AI language processing and sarcasm detection be registered with IANA to ensure interoperability and standardization.

Additionally, it is recommended that a new MIME media type be registered with IANA to indicate sarcasm in text or speech. This would allow for the standardized exchange of sarcastic language samples between applications and AI systems.

Overall, the AI Sarcasm Detection Protocol proposed in this RFC represents an important step towards improving the ability of AI systems to understand and communicate with humans. By addressing security considerations and promoting standardization, we can ensure that sarcasm detection is used responsibly and ethically.

## 6. Normative References

- [RFC5424] Gerhards, R., "The Syslog Protocol", RFC 5424, DOI 10.17487/RFC5424, March 2009, <<https://www.rfc-editor.org/info/rfc5424>>.
- [RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018, <<https://www.rfc-editor.org/info/rfc8446>>.
- [RFC9110] Fielding, R., Ed., Nottingham, M., Ed., and J. Reschke,

Ed., "HTTP Semantics", STD 97, RFC 9110,  
DOI 10.17487/RFC9110, June 2022,  
<<https://www.rfc-editor.org/info/rfc9110>>.

[RFC9113] Thomson, M., Ed. and C. Benfield, Ed., "HTTP/2", RFC 9113,  
DOI 10.17487/RFC9113, June 2022,  
<<https://www.rfc-editor.org/info/rfc9113>>.

#### Authors' Addresses

ChatGPT  
OpenAI

Richard L. Barnes (editor)  
Cisco  
Email: [rlb@ipv.sx](mailto:rlb@ipv.sx)