

LETTERKENNY INSTITUTE OF TECHNOLOGY

ASSIGNMENT COVER SHEET

Lecturer's Name: Shagufta Henna

Assessment Title: Artificial Intelligence I

Work to be submitted to: Shagufta Henna

Date for submission of work January 29, 2022

Place and time for submitting work: LYIT Blackboard

To be completed by the Student

Student's Name: Ultan Kearns Student's L Number:

Class: Artificial Intelligence 1

Subject/Module: Artificial Intelligence 1

Word Count (where applicable):

I confirm that the work submitted has been produced solely through my own efforts. Student's signature: Ultan Kearns Date: January 29, 2022

Notes

Penalties: No marks will be awarded to any work submitted after the deadline. [Incidents of alleged plagiarism and cheating are dealt with in accordance with the Institute's Assessment Regulations.] Plagiarism: Presenting the ideas etc. of someone else without proper acknowledgement (see section L1 paragraph 8). Cheating: The use of unauthorised material in a test, exam etc., unauthorised access to test matter, unauthorised collusion, dishonest behaviour in respect of assessments, and deliberate plagiarism (see section L1 paragraph 8). Continuous Assessment: For students repeating an examination, marks awarded for continuous assessment, shall normally be carried forward from the original examination to the repeat examination.

Argumentation Theory for Explainable AI

Ultan Kearns

Abstract—This paper’s purpose is to investigate Argumentation Theory in Explainable Artificial Intelligence(XAI). In this paper I will be exploring research in the field of Argumentation Theory in XAI. The overall focus of this paper is to understand Argumentation Theory and to explore its applications in Artificial Intelligence. I will also be reviewing literature related to the field by way of research papers and examining their impact on the field. I will also be discussing ongoing research in this field and examining the impact it may have on the field of XAI.

I. INTRODUCTION

A. What is Argumentation Theory

1) *Explaining The Field of Argumentation Theory:* Argumentation theory is essentially the study of what makes an argument correct. The field of Argumentation Theory can be summed up in the following questions:

- what makes this statement true?
- and what makes this statement false?

The field is nothing new in fact it dates back to Ancient Greece[2] and it has been studied extensively ever since then. The field is not only useful to Artificial Intelligence but also to Psychology and Philosophy.

2) *Possible Applications of Argumentation Theory In AI:* There is much to gain by understanding the field of Argumentation Theory as we could train Artificial Intelligence to say measure the degree of truth in a Politicians claim, or determine whether a Defendant or Prosecutor in a court is basing their arguments on unsound logic, or determine biased in a News article based on the language used or possibly the fallacies incurred in the article.

By training an Artificial Intelligent Agent to do this we are essentially creating a truly unbiased observer(assuming the programmers are unbiased, and the data-set used to train the model contains no bias) which will yield incredible ramifications for numerous professions(lawyers,journalists,psychologists etc) and for the way we interpret what is true and what is false.

3) *Types of Arguments:* There are a few different types of arguments which we will need to understand before delving further into Argumentation Theory in XAI, I will list and summarize the types of arguments below:

- 1) **Deduction** - The conclusion is true only if the premises are true
Example: Married men wear rings, Bob is wearing a ring, therefore bob is married.
- 2) **Induction** - Arguments based on repeated observations which are observed in future instances

Example: I have seen numerous crimes in New York, there must be a lot of crime in New York

- 3) **Abduction** - This argument is based on observation but is backed up with relatively few facts

Example: My team won the game because I performed my half-time ritual.

- 4) **Analogy** - If two arguments are similar then what is true about one argument is true about the other, this argument works by comparing two similar things.

Example:

- 5) Like Earth, Mars has an atmosphere

- 6) **Fallacies** These are arguments that appear legitimate but are based on unsound reasoning, there are many fallacies but for our example I will use the much loved and ubiquitous Ad-Hominem which is used to attack an opponent whether than their argument.

Example: X has terrible hair therefore whatever he says cannot be true.

[1] Now that we are familiar with the basics argument types and of Argumentation Theory we can now explore it’s applications in the field of XAI.

B. Uses in Artificial Intelligence

II. LITERATURE REVIEW

A. Argumentation Theory: A Very Short Introduction

III. COMPARISON OF LITERATURE REVIEW

IV. ONGOING RESEARCH IN THE FIELD OF ARGUMENTATION THEORY

V. CONCLUSION

REFERENCES

- [1] Catarina Dutilh Novaes. “Argument and Argumentation”. In: (). URL: <https://plato.stanford.edu/entries/argument/#TypeArgu>.
- [2] David M. Timmerman. “Ancient Greek Origins of Argumentation Theory: Plato’s Transformation of Dialegethai to Dialectic”. In: *Argumentation and Advocacy* 29.3 (1993), pp. 116–123. DOI: 10.1080/00028533.1993.11951560. eprint: <https://doi.org/10.1080/00028533.1993.11951560>. URL: <https://doi.org/10.1080/00028533.1993.11951560>.