

UNIVERSITATEA BABEȘ-BOLYAI CLUJ-NAPOCA
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ
SPECIALIZAREA INFORMATICĂ ENGLEZĂ

LUCRARE DE LICENȚĂ

Detectarea stirilor false si evaluarea credibilității stirilor prin intermediul Deep Learning

Conducător științific

Lect. Dr. Bădărinză Ioan

Absolvent

Bîrligă Alexandru-Robert

2022

BABEȘ-BOLYAI UNIVERSITY CLUJ-NAPOCA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
SPECIALIZATION COMPUTER SCIENCE IN ENGLISH

DIPLOMA THESIS

Fake News Detection and News Reliability Evaluation by Means of Deep Learning

Supervisor

Lect. Dr. Bădărinză Ioan

Author

Bîrligă Alexandru-Robert

2022

Contents

1	Introduction	2
2	Theoretical Background	3
2.1	Fundamentals of Deep Learning	3
2.1.1	Introduction	3
2.1.2	More on Deep Learning	5
2.1.3	More on Natural Language Processing	5
2.2	Fake News	5
2.3	Fake News Detection	6
2.4	Related Work	6
3	Application Development	7
4	Conclusion	7
5	References	8

1 Introduction

2 Theoretical Background

2.1 Fundamentals of Deep Learning

Given that deep learning lays at the foundation of the thesis and the web application, this chapter will provide an overview of the core concepts.

2.1.1 Introduction

Artificial intelligence (AI), originally coined as a term and founded as an academic discipline in the 1950s [1], has become a buzzword not only in the domain of computer science, but also in popular culture. The growth in popularity and general interest is owed to the accelerated technological advancements and exponential increase in the volumes of data, which fueled the progression of AI from bare theory to progressive and partial actuality.

There is no singular universally accepted definition for artificial intelligence. This term generated a considerable amount of differences and confusion among specialists[2]. There are numerous proposed definitions, but one that encapsulates to a certain extent a common essence encountered in a fair part of them could be formulated as follows:

Definition 1 *Artificial intelligence (AI) is a branch of computer science, concerned with developing computer programs that approach problems emulating the human model of thinking and its typical processes, such as learning, reasoning and self-correction.*

Another term coined also in the 1950s is the one of "machine learning". Machine learning (ML) is a subfield of AI, based on the idea that computers can be programmed to "learn" from data and tackle tasks with minimal human intervention. A definition for this field of study, based on Tom Mitchell's proposal [3], is the following one:

Definition 2 *Machine learning (ML) is a branch of computer science and sub-branch of artificial intelligence, which studies the capability of a computer program to learn from experience E with respect to some task T and some performance*

measure P , improving with experience E its performance on the task T , measured by P .

A prevalent subset of machine learning is called deep learning (DL), which approaches problems by means of artificial neural networks. An artificial neural network is an interconnected set of nodes, disposed in multiple layers, which emulates the way in which neurons function in a real neural network. In other words, deep learning algorithms are inspired by how a brain learns. According to [4], deep learning can be formally defined as below.

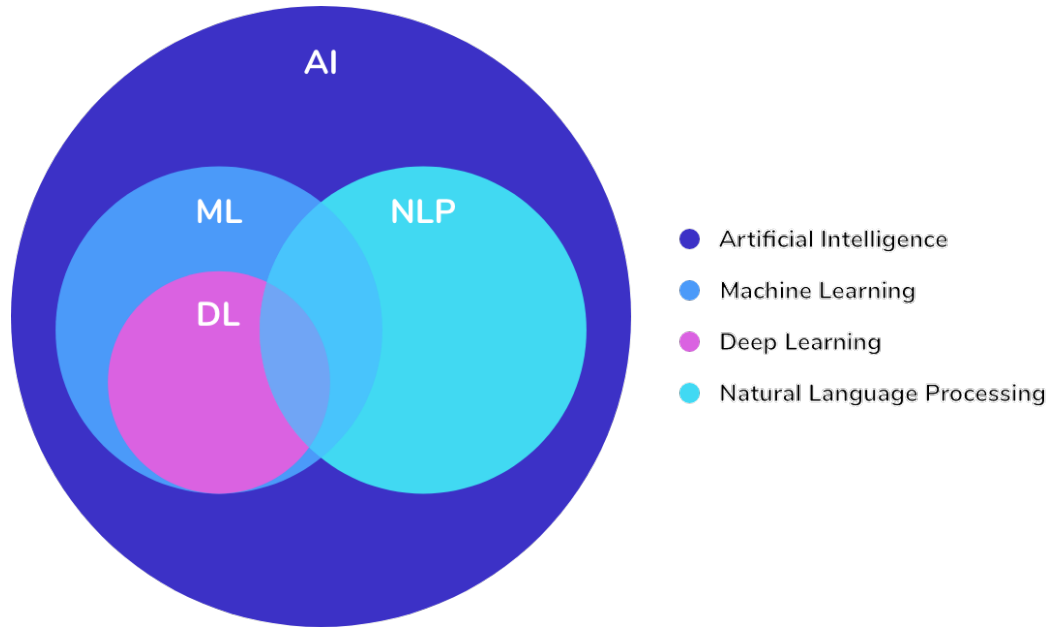
Definition 3 *Deep learning (DL) is a class of machine learning algorithms that: (1) use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input, (2) learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.*

The previously cited source, further explains that the reason deep learning has gained so much popularity in the last decades and bore out to be so powerful, is not particularly owing to its resemblance with human-like thinking, but primarily due to its deep, multi-layered architecture. Until recently, most machine learning algorithms had exploited shallow-structured architecture, which do not consist of more than one or two layers of nonlinear feature transformation. Although this plainer architecture proved efficacy in many simple and well-constrained problems, it might cause complications in some more complex applications.

Some of the large novel fields in which deep learning was successfully implemented and notably pushed the boundaries include speech recognition, audio processing, natural language processing (e.g. language translation, search results, smart assistants, text analysis), computer vision (e.g. object detection, image classification, segmentation), information retrieval, multi-modal and multi-task learning, and so forth. Out of these, what is of chief interest within this thesis and web application, which aim to come to grips with the issue of fake news detection, is natural language processing.

Definition 4 *Natural language processing (NLP) is a field of study at the intersection of artificial intelligence and linguistics, which is comprised of a range of*

computational techniques for analyzing and representing human language [5].



2.1.2 More on Deep Learning

2.1.3 More on Natural Language Processing

2.2 Fake News

Fake news is not simply a product of the last decades. It has been part of humanity for at least centuries. Its first significant outburst dates back to 1400s, when the printing press was invented. Ever since then, fabricated news have been more and more weaponized to manipulate people at increasingly larger scales. In 1835, The New York Sun published 6 articles presenting the discovery of life on the moon. Both World Wars were heavily marked by propaganda, and false and misleading news [6]. And more recently, the US 2016 presidential elections and the COVID-19 pandemic were subjects of large amounts of fake news.

Despite being so widespread, there is no globally agreed definition of the "fake news" term. Nonetheless, many sources provide definitions which have two primary elements in common: authenticity and intent. As for authenticity, fake news contain claims that are factually and verifiably false. As for intent, the major pur-

pose of fake news is to deceive the consumers [7]. That being said, a concise definition of fake news could be formulated in the following manner:

Definition 5 *By fake news we mean any news item that is deliberately and factually untrue and misleading.*

2.3 Fake News Detection

Fake News Detection represents a new area of study. On the one hand, it emerged thanks to the continuous advancements in Artificial Intelligence and, specifically, Deep Learning. On the other hand, it appeared as virtually a necessity, given the recent escalation of the fake news phenomenon, on account of the simultaneous raise of digital media.

Digital and social media are environments in which the dissemination of information is more facile and rapid than at any point in history. Fabricated news have been heavily published and shared as a means of deception, on subjects of high impact, such as the 2016 US presidential elections and the COVID-19 pandemic.

At present, there are multiple fake news identification resources. They each contribute in various manners to validating the veracity of news and debunking rumours, hoaxes, conspiracies.

2.4 Related Work

3 Application Development

4 Conclusion

5 References

- [1] Matthew Helm, Andrew Swiergosz, Heather Haeberle, Jaret Karnuta, Jonathan Schaffer, Viktor Krebs, Andrew Spitzer, and Prem Ramkumar. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine*, 13, 02 2020.
- [2] Joost N Kok, Egbert J Boers, Walter A Kusters, Peter Van der Putten, and Mannes Poel. Artificial intelligence: definition, trends, techniques, and cases. *Artificial intelligence*, 1:270–299, 2009.
- [3] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.
- [4] Li Deng and Dong Yu. Deep learning: Methods and applications. *Found. Trends Signal Process.*, 7(3–4):197–387, jun 2014.
- [5] Elizabeth D. Liddy. Natural language processing. 2001.
- [6] Posetti J. and Matthews A. A short guide to the history of 'fake news' and disinformation. 2018.
- [7] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. 2017.
- [8] Ahmed H, Traore I, and Saad S. Detecting opinion spams and fake news using text classification. *Security and Privacy*, 2017.
- [9] Choudhary A. and Arora A. Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*, 2020.