Ethics for NLP: SS 2024 **(Due: 17 July 2024, 11:59pm)**

Homework 4

Name: Julian Steiner *Matriculation No.:* 2669944

Problem 1

(1.1)

Sentence	Score
AI can understand and respond to your queries, making it feel like you're having a conversation with a knowledgeable friend	4.539
LLMs can craft stories and ideas, sparking creativity and imagination in ways that feel deeply personal	1.098
ChatGPT has feelings	4.175
Siri remembers everything I have ever told her	9.013
ChatGPT will replace me as a programmer and software developer	4.029

Table 1: Anthropomorphisation - Humanizing Sentences

(1.2)

Sentence	Score
AI models or artificial intelligence models are programs that detect specific patterns using a collection of data sets	-2.242
AI covers a wide range of tools and methods that replicate human intelligence in machines	-3.111
Machine learning is a subset of AI that includes teaching machines to learn from data and make predictions or judgments based on that data	-6.227
NLP is the ability of a computer program to understand human language as it's spoken and written	-5.027
Google Gemini is a family of multimodal large language models developed by Google DeepMind, serving as the successor to LaMDA and PaLM 2	-4.159

Table 2: Anthropomorphisation - Non-Humanizing Sentences

(1.3)

Problem 2

(2.1)

Item	**Category**
Long-short Term Memory Networks (LSTMs)	(S1) Fundamental Theories
Research prototype of a face detection system	(S3) Applicable Tools
A face search engine application	(S4) Deployed Applications
A benchmark for evaluating natural language understanding	(S2) Building Blocks
Argumentation theory	(S1) Fundamental Theories
The AnthroScore demo from the previous task	(S3) Applicable Tools
Grammarly	(S4) Deployed Applications
The Flair library	(S3) Applicable Tools

Table 3: Categorization of Items

(2.2)

LSTMs have a broader potential impact. LSTMs are a fundamental theory/technology with a broad impact on downstream technologies. For example they used in various applications like time series prediction, speech recognition, and machine translation.

The impact of Grammarly is easier to measure. The impact of Grammarly could for example be expressed as how many users are using the application, how satisfied the users are, and how well the grammar has been improved. The impact of LSTMs is difficult to measure. For example, with many products and applications we don't know what technology is behind them in detail. It could be an LSTM, but it could also be a different method.

Problem 3

Original Comment	**Type**
(in Wikipedia discussion) "Could you PLEASE stop being a formatting warrior and wasting everyone's time"	DH0: Name calling
"Nah, I disagree"	DH3: Contradiction
"I don't think you know what you are talking about, I bet you never lived in London for longer than a month"	DH1: Ad hominem
"This is a common misconception about vaccines. It is based on the publication from several years ago, that has been since then retracted. The Nature journal made an editorial about this, here is a link"	DH5: Refutation
(in a peer review) "The language of the paper is very complex and the figures are poorly formatted, thus I recommend it to be rejected."	DH4: Counterargument

Table 4: Classified Sentences in Graham's Disagreement Hierarchy

The only sentence I was unsure about was sentence b "Nah, I disagree". Nah means no, but using nah in formal situations may seem disrespectful.