Bachelor thesis

in the bachelor's program

Wirtschaftsinformatik

at the Hochschule für angewandte Wissenschaften Neu-Ulm

&

Technische Hochschule Ulm

Does culture have an influence

on the ability to recognize lies and

why does this influence exist?

First corrector: Prof. Dr. Heiko Gewald

Second corrector: Prof. Dr. Andy Weeger

Author: Thinh Nguyen (Matrikel-Nr.: 3134767)

Topic received: 13.10.2023

Thesis submitted: 13.03.2024

Tables of contents

I. List of figures

II. List of tables

III. List of abbreviations

AI	Artificial Intelligence
NLP	Natural Language Processing
MFCC	Mel-Frequency Cepstral Coefficients
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
CCTV	Closed Circuit Television
LCC	Low-Context Culture
HCC	High-Context Culture
PDI	Power Distance Index
SVM	Support Vector Machine

Abstract

The commonly known polygraph has shown a lot of flaws and unreliability, when it comes to accurately detecting lies and deception. As such, researchers sought out alternative methods or techniques to provide more reliable lie detecting solutions. As such the detection of lies with the help of artificial intelligence has gained a lot of popularity, with researchers exhibiting more reliable and performant models, in recent years.

However, when it comes to the performance of these AI solutions, to detect lies across different cultures, research has not shown a lot of findings regarding this important influence. This thesis' purpose is to analyze the actual influence of culture in connection with lie detection tools, which utilize AI, with Mel-Frequency Cepstral Coefficients as its foundation.

The results of this research, show the clear influence of culture when comparing a German and Vietnamese data set. The accuracy of the proposed AI model to correctly differentiate between untruthful and truthful statements therefore sunk from a promising 99% to merely 54%. While the exact influences of culture could not be analyzed in the scope of this thesis, the research highlights several factors, with reference to culture, which may have an impact on the ability of the AI to detect lies.

1. Introduction

1.1 Current relevance and motivation

Artificial intelligence (AI) is gaining more and more popularity in various sectors such as education and business (figure 1). While AI-Usage for educational purposes may be a controversial topic, in the business sector, AI is showing promising results and nothing but benefits for its users (figure 2). As such it is now also possible to revisit several research topics with the help of AI. These topics include research, where results were limited due to research methods or human limitations itself. One important topic which greatly benefits from the utilization of AI is the study of human emotions. Emotions are very hard to perceive as they are unique to each person on the planet. Because of that it is impossible for a human being to evaluate another person's emotions with confidence or clear numbers. Therefore, existing research shows, that people can only guess if a person is being untruthful with an accuracy of about 50%, which basically means, they could only detect deception by "mere chance" (Faußer et al., 2021, p. 5). However, with the use of AI it may be possible to evaluate emotions more precisely through certain factors in human behavior and voice.

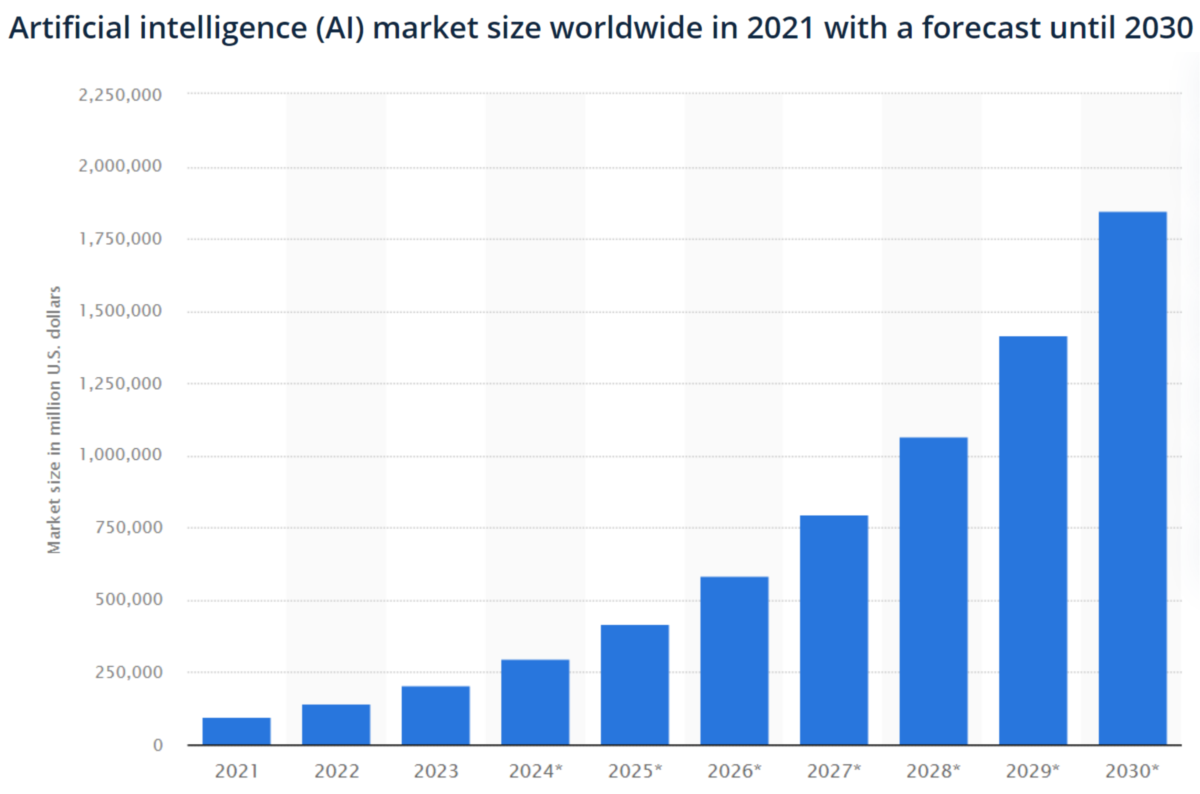

Figure 1: AI market size

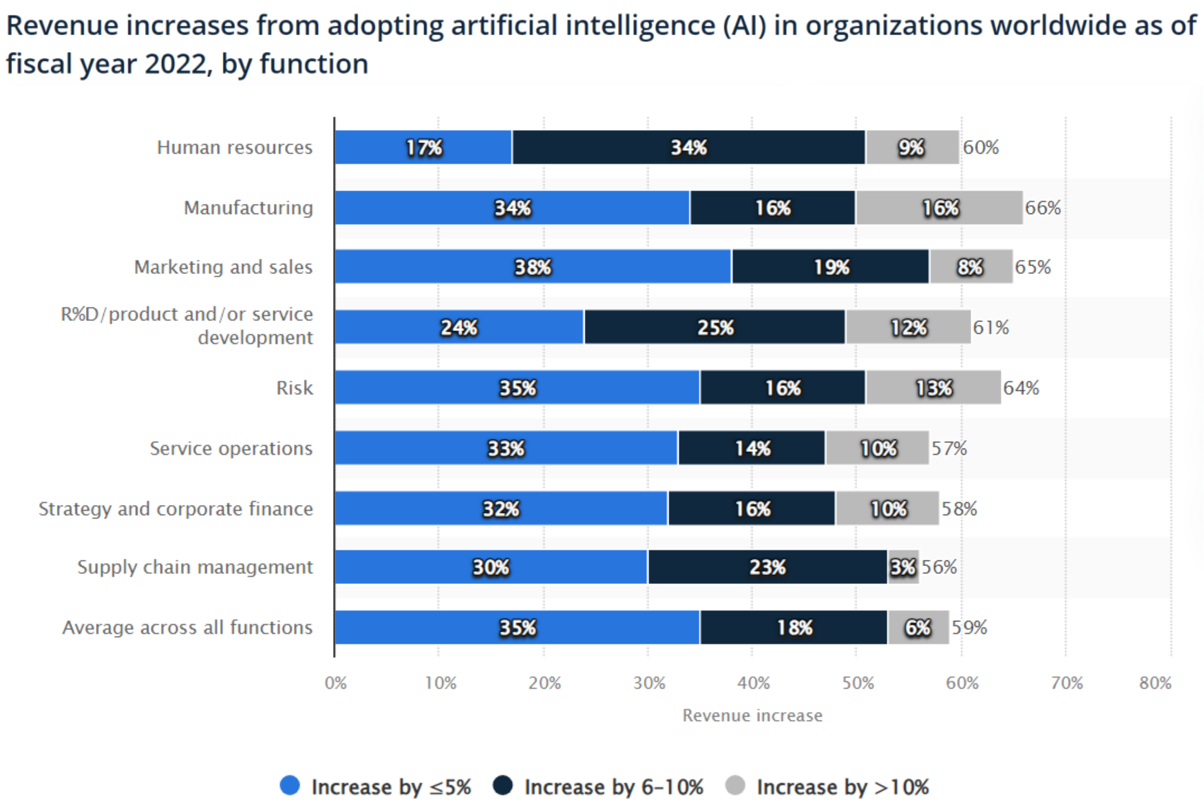

Figure 2: Revenue increase by adopting AI

This could prove useful especially in the interaction with customers, where AI will play a decisive and important role in the near future. A major problem in dealing with customers is the detection of intentional misleading, meaning deception. When lying, a person exhibits several emotions, which can indicate deceitful behavior. As a consequence, a reliable AI solution should be able to detect these emotions, in order to differentiate between untruthful and truthful statements. Such a solution could prove a lot of use, especially in connection with insurance claims. Annually, 308.6$ Billion in damages are solely caused by insurance fraud in the U.S. (Coalition Against Insurance Fraud, 2022, p. 37). This damage could possibly be reduced significantly with the use of an AI solution, which can reliably detect lies.

This bachelor thesis is written as an extension to the paper "Using NLP to analyze whether customer statements comply with their inner belief" (Fabian Thaler et al. 2021), where the authors use an AI to analyze the speech signal of a person to evaluate whether he or she is speaking truthfully. At the end of that paper, it was described that the results used were limited because the analyzed data was only based on speech recordings of people with a German cultural background (Faußer et al., 2021, p. 18). Thus, the bachelor thesis is concerned with analyzing the cultural impact on the AI solution provided by the authors, with additional language data.

1.2 Structure of the thesis

The following chapter of this paper defines all the basics necessary to understand this experiment. These fundamentals include definitions regarding lying and deception and culture itself.

Moreover, a closer look is taken at the two cultures, which are the subject of this experiment. The definition of these two cultures discusses relevant characteristics, which may shed light on how culture is able to influence the detection of lies across different cultures.

To understand how exactly the ability of the AI to detect deception is impacted, the key components of the AI used in this thesis' experiment are discussed. Additionally, the definition of AI itself is also covered in this section.

Before the formulation of the hypotheses, a literature review is conducted, which will summarize and evaluate various approaches regarding the theme of this thesis. The review is split into two disciplines. The first part discusses cultural differences with respect to the production and detection of lies. The second part points out the ability to detect untruthful statements, based on various speech processing techniques. The literature review takes already executed research into account in order to build a stronger foundation for the assumptions made in this paper and to confirm the implications made by the previously defined basics.

The next step is to explain the methodology of the thesis. The thesis follows an empirical approach, as such, the first step is the formulation of the hypotheses. The paper then goes on to explain the scope of the data collection process. The description of the data collection process includes the scouting process, the demographic data, the interview content and the interview process of the participants itself. The AI used to assess the Vietnamese data set is also reviewed in this section.

Now that the results are revealed, a comparison between the representatives of the two respective cultures is made (Germany and Vietnam). This comparison encompasses the performance metrics output by the AI model, i.e. accuracy, sensitivity and precision. After analyzing these results, the hypotheses are tested. In accordance with the hypotheses put forward, the results of the language data are further interpreted through a discussion afterwards. These discussions implicate more granular influences, which may have caused the disparity between the two data sets.

This thesis then continues to review implications for theory and practice based on the knowledge gained from the results and the discussion.

To finish this paper off, several limitations and opportunities of this thesis are highlighted, to help further research in the study of cross-cultural lie detection. Finally, a summary of all important findings is presented and the final answer to the research question is given.

1.3 Research question

This papers' purpose is to research the influence of culture on the ability to detect lies. In the scope of this thesis the Occidental and Confucian culture will be compared. The Occidental culture will be presented by the German data set, whereas the Confucian culture will be presented in the Vietnamese data set. In order to compare the two different data sets, the before mentioned AI, which uses the feature extraction method Mel-Frequency Cepstral Coefficients (MFCC), will be utilized (Faußer et al., 2021, pp. 10–11).

Comparing the German and Vietnamese speech data, these differences may be reflected by the performance metrics of the AI, i.e. accuracy, sensitivity and precision (Faußer et al., 2021, pp. 13–14). Therefore, this thesis seeks to address the question:

"Does culture have an influence on the ability to recognize lies,

and why does this influence exist?"

This research uses an AI model trained solely on a German data set (Faußer et al., 2021, p. 18). It aims to uncover if culture has an influence on how a person produces deception. This can be proven, if the performance of the AI differs when tested on the Vietnamese data set. As mentioned, the AI uses MFCC, to review vocal cues like pitch, rhythm, and various emotions reflected in the speech signal of a person (Faußer et al., 2021, p. 10), to determine whether a person is truthful or not. This study discusses how exactly culture impacts the above-mentioned cues of speech signals, and in consequence the detection ability of deception.

The purpose of this thesis is to close the missing gap between the two disciplines of detecting lies across different cultures and detecting lies with the help of speech processing techniques. Furthermore, this thesis' content will help to grasp how culture, language, and linguistic impacts human communication. It highlights the role culture plays in detecting lies. This finding can make research regarding lie detection methods more aware of differences in communication caused by culture. Additionally, it also improves how we communicate face-to-face across cultures, by highlighting the differences between the communication styles of Occidental and Confucian culture.

2. Relevant theory

2.1 Terms related to deception

2.1.1 Emotions:

Emotion is a suit-case term used to describe several psychological phenomena such as fear, joy, anger, sadness, etc. These enumerated emotions are terms everyone knows and experienced. But because these emotions relate to experiences, they are often not clear states which one can define as they do not have clear boundaries and distinct behavioral or physical signatures. Emotions are usually a response to external stimuli. This response is often unique and evident in behavior, expression or thought of a person. These reactive emotions occur with a confrontation of the unexpected (Feldmann et al., 1993, p. 3).

While there are short-term emotional states, there also exist long-term emotional states. These longer lasting emotions can be called moods. They often have uncertain origins and persist for hours or days in contrast to reactive emotions such as fear when one is confronted with immediate danger, which normally last for several minutes or even seconds (Oatley, 2004, p. 4). As mentioned above, emotions are a response to external factors which result in the automatic appraisal and evaluation of these factors (Feldmann et al., 1993, p. 1) with the help of the amygdala, a sub-area within the limbic system of the brain (Rudolf-Möller, 2017). The goal of these emotional responses is to mediate toward an emotional equilibrium. When said equilibrium is achieved and the unexpected stimuli is usually dealt with, the emotional state resets to its previous state before the stimuli occurred (Plutchik, 2001, p. 347).

There are two models, which describe the structure of emotions. The first model is the latent trait model. In this view, emotions can exist, whether they are expressed by behavior or physical occurrences or not (Feldmann et al., 1993, p. 5).

The second view sees emotions as an emergence model. If “fear” is taken as an example, the emergence model proposes, that several indicators, such as the voice or the facial expression, must represent the emotion “fear” at the same time (Clore & Otorny, 2000, pp. 25–26; Feldmann et al., 1993, p. 4), for one to be in a state of fear. In contrast, the latent trait model suggests, that these indicators are a result of a specific emotion, which is already present in our unconsciousness (Feldmann et al., 1993, p. 4).

Another view which Barett and Russell introduced, proposes that emotions are situated instances of positive or negative evaluations by the amygdala (Russell & Barrett, 1999, pp. 810–811). In this model the only latent traits a human possesses are positive and negative affective reactions and arousal. For example, a person feels "fear" when the amygdala evaluates a situation as negative, and the person is exposed to a dangerous situation. In this model, experiences and cultural knowledge can affect the subsequent emotion. For instance, if one is for example faced with pressure, one may feel fear while another one feels excited, based on past experiences and/or knowledge (Feldmann et al., 1993, p. 3).

2.1.2 Lying and deception

With the knowledge of how emotions work, the next step is to understand the concept of lying and deception. Although everyone has their own opinion about lying, it is important to clearly define what a lie is and when a statement really is a lie. According to James Edwin Mahon a lie is, "to make a believed-false statement to another person with the intention that the other person believe that statement to be true" (Mahon, 2015). Based on this statement from Mahon four conditions to lying can be derived.

Condition 1: Statement Condition:

The statement condition implies that one must make a statement to utter a lie. The statement must be made by conventional signs, symbols or bodily gestures (Mahon, 2015).

While this condition seems to be easy and obvious, there are some objections to be made. According to several philosophers it is not necessary to make a statement to lie, one can lie by simply staying silent (Ekman, 1985, p. 28; Scott, 2006, p. 4). Thus, lying entails any form of behavior which provides the hearer with information, which the testifier believes to be false (Smith, 2004, p. 14).

Condition 2: Untruthfulness Condition:

The second condition is that lying requires the testifier to make an untrue statement. Whether the statement is true or not does not matter in this case, only the belief of the testifier matters (Mahon, 2015).

At first glance this condition seems to be necessary, but there are cases where it is not required to make an untruthful statement. The first case is mentioned above: Lying by staying silent. A second case of lying without meeting this condition is the omitting of information that is crucial for the understanding of the addressee. Another case of lying while being truthful is giving information that avoids the question and misleads the addressee (Carson, 2006, p. 284). All these cases describe the process of depriving the addressee of certain information, with the intent to deceive them.

Condition 3: Addressee Condition:

This condition requires the liar to address a person when making a false statement. To understand this condition better, Mahon explains that a person cannot lie to someone or something who cannot form a belief based on the statement one made to influence the belief of said someone or something (Mahon, 2015).

An objection to this condition is that even if you are not addressing anyone, when you lie you still express something you yourself do not believe in and consequently lie (Griffiths, 2004, p. 31). Therefore, if you look at this issue from the other side of the coin, in this case it makes no sense to say that you did not lie just because the lie was not addressed to a specific person.

Condition 4: Intention to Deceive:

As the name suggests, this rule phrases that a person must have an intention to deceive an addressee in order to lie (Mahon, 2015).

Based on the objection in condition three, this condition does not seem to come to fruition, because if you do not address anyone when telling a lie, you also do not intend to deceive anyone. An additional criticism to the condition of deceiving is "bald-faced" lying (Sorensen, 2007). To further understand bald-faced lies Stokke presented an example in his work "Lying deceiving and misleading". In a murder court case, a witness is asked whether he witnessed the murder of a victim. Before answering a CCTV is played, where it can clearly be seen that the witness in question saw the murder happen in front of his eyes. Even though the witness knows he cannot deceive anyone in the court room he still says, "I did not see the murder". Stokke explains that the witness lied about not seeing the murder, because of his fear for retribution (Stokke, 2013, p. 349).

Considering the criticism mentioned above, the fourth condition seems to be insufficient to determine a lie. Instead of concluding that a person is lying to deceive someone, it is better to say that a person is lying to either create a benefit or a avoid loss (Gupta & Ortony, 2018, p. 149). In Stokke's case the witness is trying to avoid retribution, by stating that he did not see the murder happen. With his bald-faced lie, he is trying to reach an emotional equilibrium, which is disturbed by his current state of fear. The bald-faced lie therefore contributes to the achievement of a calmer emotional state, which can be considered a gain in the perspective of the liar.

Kinds of lying:

Moving on to the second part of understanding lies, it is important to recognize the different cases of lying. Chisholm and Feehan identified three qualities which can discern the different kinds of lying and deception (Chisholm & Feehan, 1977, p. 143):

<u>Lying by Commission and Omission:</u>

The first type is lying by commission. With reference to the above statements made, commission describes any conscious form of external expression which can be recognized by an addressee (Chisholm & Feehan, 1977, p. 144).

The kind of lies which a person cannot perceive, are differentiated as internal expressions. This second type - omission - portrays internal expressions, that include silence and the omission of critical information while lying, as a means of deceiving a person (Chisholm & Feehan, 1977, p. 144).

<u>Positive and Negative Deception:</u>

Positive deception describes the action of an addressee acquiring a false belief or the action of allowing an addressee to continue a false belief (Chisholm & Feehan, 1977, p. 144).

Negative deception, however, depicts cases where the testifier makes an expression, which leads to the cease of a belief. Preventing the addressee to gain a true belief is also another example of negative deception (Chisholm & Feehan, 1977, p. 144).

<u>Deception “simpliciter” and “secundum quid”:</u>

Deception simpliciter references to lying, when a person newly acquires a false belief or ceases to believe a past true belief (Chisholm & Feehan, 1977, p. 144).

On the other hand, we have the deception secundum quid. This type of deception has the intention to make an addressee continue in a false belief or prevent the addressee from gaining a true belief (Chisholm & Feehan, 1977, p. 144).

With a complexity of three dimensions and respectively two options per dimension, this model exhibits a total of eight kinds of different lies (Chisholm & Feehan, 1977, p. 145):

		Positive Deception	Negative Deception
Simpliciter	Omission	Addressee acquires a false belief, by commission	Addressee ceases a true belief, by omission
	Commission	Addressee acquires a false belief, by omission	Addressee ceases a true belief, by commission
Secundum Quid	Omission	Addressee continues in a false belief, by omission	Addressee does not gain a true belief, by omission
	Commission	Addressee continues in a false belief, by commission	Addressee does not gain a true belief, by commission

Figure 3: Different kind of lies

2.2 Culture

2.2.1 Definition of Culture

In the course of time the term culture got increasingly popular and gained recognition in several contexts, for instance corporate culture (business) or pop culture (media). As such it became more difficult to define the term culture in a manner which satisfied its broad use. There is however still similarity across these different contexts. These similarities will start off this approach to define the term culture.

Similarities:

In the past the term culture was mostly used in connection with ethnicity and nationality (Rathje, 2009, p. 35). One may also recognize the term culture in conjunction with “cultural shock”. A person experiences a cultural shock if he or she is confronted with a situation or behavior, which is not congruent with their understanding of “common sense” (Storey, 2009, p. 87). This common sense can be interpreted as the culture of an individual. Therefore, a commonly known definition of culture is a sense of uniformity within a society or group. This uniformity is expressed in many ways, for example in the style of dress, values or customs in a society (Rathje, 2009, p. 35). One can also view culture as a sort of mental programming since culture also affects the patterns of feeling and decision making based on what was learned throughout a person's lifetime. Culture is therefore not innate but learned within a social environment (Hofstede et al., 2010, pp. 4–7).

Distinctions:

The first distinction within the term culture are two perspectives introduced by Stefanie Rathje. The first perspective "collectivist perspective" describes culture solely as "formal and structural" tool in a human group (Rathje, 2009, p. 40). This view can be applied to the hierarchical organization of a company, which is a part of the overall corporate culture, that was mentioned earlier. The second perspective is the "cultural perspective". This perspective can be applied to the customs or values in a certain religion or ethnicity. Stefanie Rathje derives this perspective from the pragmatic approach of Wittgenstein, where he states that culture is most evident when one can identify "shared practices" in a certain group of people (Welsch, 1995, p. 43). In this approach, the paper will focus on the second perspective, because the experiment examines the similarities and/or differences between two different cultures, namely the Confucian culture and the Occidental culture.

The second distinction to be made was introduced by Raymond Williams in 1961. Raymond distinguishes three kinds of cultures in his works. The first being the "ideal culture". This "ideal" view on culture defines culture as a process to find absolute or universal values which one can conform to with the goal of becoming a "perfect" human being. One can compare this process to the evolution of the human species. The second kind of culture is the "documentary culture". This kind entails culture as a collection of "intellectual and imaginative work" (Williams, 1961, p. 57) of humankind, which is recorded in several forms, like books or movies. To fully comprehend how culture works this view is critical, as it describes that history is not fully but only selectively recorded. One can only experience true culture as a whole, in the present. Going forward certain aspects of our past cultures will not be recorded and thus be lost forever. The last definition of culture proposed by Raymond is called "social culture". "Social culture" describes "a particular way of life, which expresses certain meanings and values not only in art and learning but also in institutions and ordinary behavior." (Williams, 1961, p. 57). This definition of culture may be the most familiar to the majority of people.

In addition to the types of culture, there are also three different levels, which were briefly mentioned in the context of "recorded culture". The first level – lived culture – is experienced by a person in a certain moment of their everyday life. The second level – period culture – encompasses "recorded culture, of every kind, from art to the most everyday facts" (Williams, 1961, p. 62) At this level of culture one can only guess how things really felt and worked, based on the records left behind. The last level – selective tradition – is the connection between lived and period culture (Williams, 1961, p. 62). In the process of “selective tradition” information of the lived culture is lost and consequently turns into period culture.

2.2.2 Characteristics and values of Confucian culture

High-context culture:

High-context cultures are a type of culture where communication heavily relies on context, i.e. non-verbal cues, shared understanding, and the social environment. In such cultures, information is often embedded in the context or the relationship between communicators, rather than being additionally explicitly verbalized. This communication style is usually seen in collectivist societies, where there is a strong emphasis on group cohesion, harmony, interconnectedness, and maintaining social connections (Hofstede et al., 2010, p. 109).

As such, people often rely on implicit communication, i.e. context in addition to words, to communicate. Cultural values and social relationships may play a crucial role in interpreting and understanding these conveyed messages. Shared memories, customs, and traditions can influence the way information is communicated and perceived within a certain context (Hofstede et al., 2010, pp. 109, 218, 219).

In-group members or members of a certain culture in collective communities are expected to understand the above-mentioned circumstances and understand the hidden messages conveyed within their respective social groups. In summary, communication in high-context cultures is more complex, resulting in the need for a deep understanding of the cultural and social context. This highlights the difference between high-context cultures, where the communication style is often implicit and relies on context and social setting, and low-context cultures, where the reigning communication style is more direct and clear and therefore relies mostly on the spoken or written word to communicate (Hofstede et al., 2010, p. 109).

High Power Distance Index:

High power distance as a cultural characteristic accepts and even expects unequal distribution of power within a social construct. In such cultures, hierarchy is heavily reinforced, there is a strong emphasis on obedience, respect for authority, and a clear distinction between individuals higher and lower positioned in the hierarchy (Hofstede et al., 2010, pp. 55, 75, 76, 80).

There is also a belief in maintaining these hierarchical structures, acknowledging authority figures, and showing respect to individuals in superior positions. Decision-making is typically made by individuals with the corresponding power, while most of this power is represented in a select few at the top of the hierarchy. Subordinates are expected to follow the directives provided by those at the top without challenging or questioning them (Hofstede et al., 2010, pp. 72, 73, 75, 76).

In high power distance cultures, there is often a high degree of formality when communicating with each other, especially when interacting with authority figures. This formality is shown by subordinates by using proper work titles, formal greetings, and the expectation for individuals to listen and obey those in higher-ranking positions (Hofstede et al., 2010, pp. 72, 76, 80).

These kinds of cultures therefore show social settings where individuals are conditioned to accept and respect hierarchical structures, defer to authority and seniority and maintain established social norms, values and status differences. In these cultures there is even a clear tendency to accept and maintain these social hierarchies and power imbalances. (Hofstede et al., 2010, pp. 55, 72, 73, 75, 76, 80)

<u>Collectivism:</u>

The cultural characteristic collectivism is reflected by individuals which are integrated into strong, cohesive in-groups from an early age. This interconnectedness in these groups continues to shape and define their members throughout their lives. In collectivist societies, the collective holds the highest priority and individuals are expected to respond to the goals and needs of the group over their own individual desires or opinions. This idea is based on the belief that individuals are responsible for the well-being of their close friends, family and colleagues in addition to their own well-being. This mindset lead to a focus on unity, harmony, and shared set of opinions and values (Hofstede et al., 2010, pp. 92, 107, 108, 113, 129, 130).

Collectivist societies also have interconnected social structures, aside from family bonds. Strong bonds between individuals are also often formed through work environments, community or ethnic ties. Similar to bonds between friends and family, these relationships create a sense of belonging, mutual understanding and shared identity, consequently building a collective that prioritizes the needs and interests of the collective over the individual (Hofstede et al., 2010, pp. 107, 108, 113, 120, 124, 129, 130).

Characterized by dependence between in-group members, a focus on in-group cooperation and harmony, collectivism expects high group cohesion, the valuing of loyalty, conformity, and the prioritization of the collective over personal autonomy and self-expression. Individuals within collectivist societies are instilled with a sense of duty to their various in-groups due to their belonging, with their actions and decisions influenced by the collective's opinions, values and welfare. (Hofstede et al., 2010, pp. 92, 107, 108, 129, 130).

2.2.3 Characteristics and values of Occidental culture

Low-context culture:

A low-context culture (LCC) is defined as a culture where communication primarily relies on explicit spoken or written words. This concept is typical of individualist cultures, such as German society (Hofstede et al., 2010, p. 109).

In LCCs, information that is self-evident in collectivist cultures must be explicitly stated. This is for example exemplified by the concise and direct communication style found in American business contracts, which are noted to be considerably lengthier than their Japanese counterparts. This effect is also exhibited in American and Japanese legislative (Hofstede et al., 2010, pp. 109, 218, 219).

In contrast to collectivist cultures, where much is left unsaid due to shared understanding, explicit communication is necessary in low-context cultures to convey information and ensure mutual understanding (Hofstede et al., 2010, pp. 109, 218, 219).

Communication within low-context cultures aims to be precise, transparent and clear in order to achieve efficiency and fairness, especially in a work environment. Thus, most of the intent is evident in the message itself. LCCs avoid the requirement of context, as it can be misinterpreted or even exploited (Hofstede et al., 2010, pp. 109, 219)

Low Power Distance Index:

Low power distance as a cultural value system reflects a relatively equal distribution of power and an emphasis on flat hierarchical structures within society. In such societies, individuals tend to question authority, expect equal treatment, and strive for an equal distribution of influence and responsibilities (Hofstede et al., 2010, pp. 70, 72, 74, 76, 78, 83).

Formal titles hold less significance in comparison to cultures with a high power distance. In consequence individuals are more prone to argue and openly communicate with authority figures. This cultural characteristic expects that power and privileges should not be exclusive to a select few, and subordinates are encouraged to challenge and question the decisions made by those in higher positions (Hofstede et al., 2010, pp. 70, 72, 74, 76, 78).

In low power distance cultures, there is a strong focus on promoting critical thinking and autonomous decision-making, as individual freedom and self-governance is highly regarded. These kind of cultures assert that independence can enhance resourcefulness and creativity. Such societies also foster an environment where individuals are encouraged to accept accountability for their behaviour and decisions, while also managing any potential fallout, without excessively being swayed by authority figures or opinions from others. The underlying principle is that one should learn and benefit from previous missteps (Hofstede et al., 2010, pp. 70, 72, 76).

Overall, a culture with low power distance wants to build an environment where individuals feel comfortable expressing their own opinions, challenging decisions or authorities and engaging in collaborative decision-making processes, for better results. This reflects a social framework that values equality, mutual respect, and shared participation in contrast to collectivistic cultures (Hofstede et al., 2010, pp. 70, 72, 74, 76, 107).

Individualism:

The cultural orientation where personal autonomy, independence and self-expression are prioritized over the needs and interests of the collective group, reflects an individualistic culture. In individualistic societies, achievements, self-sufficiency, and the pursuit of individual goals and aspirations are highly accentuated. Prioritizing one's own opinion and defending it is a lot more common in contrast to collective communities (Hofstede et al., 2010, pp. 92, 113, 117).

Individualism is a cultural framework in which individuals prioritize their personal desires, striving towards personal success or fulfillment. The influence and opinions of social connections or authorities only act as a tool for support. Incentives are given based on individual performance, as such social connections do not offer as many benefits compared to collectivistic cultures (Hofstede et al., 2010, pp. 113, 114, 117, 121, 124).

Overall, individualistic societies often hold a prevalent belief in self-improvement, personal responsibility, the freedom to express opinions and the pursuit of personal goals. This translates into a society that highly values education, creativity, and distinction. Because of the focus of oneself, the basic motivation in individualistic cultures is more self-oriented, in contrast to collectivist countries, where motivation is centered around the goals of a group (Hofstede et al., 2010, pp. 92, 113, 114, 117, 128, 129).

2.2 Artificial Intelligence

2.3.1 Definition of AI

The study of artificial intelligence has the objective to develop systems, which are capable of undertaking functions that usually require human capabilities, such as reasoning, learning, problem-solving, perception, communication and adaption in context to complex settings (Liu et al., 2021, p. 2). To achieve a predetermined goal, for instance deception detection, AI systems must learn from a specific data set. The AI therefore makes its decisions based on the context-data it received. Certain aspects of the context-data can be utilized by the AI flexibly (Feng et al., 2021, p. 85). In order to assure, that the predetermined goals of AI will be achieved, it is important to define clear boundaries regarding the usage and the definition of AI itself (Wang, 2019, pp. 1–3).

The study of AI has already sparked interest in recent years and is gaining more and more momentum with each year that passes. As a results researchers have already successfully created several AI solutions that can mimic the above-mentioned human capabilities. Although one has to keep in mind, that these systems do not exhibit genuine intelligence (Chollet, 2019, pp. 5–6).

AI already seems to be a tool, one cannot imagine being without, but there still is not a consensus of the definition of AI This is caused by the existence of various incongruent definitions by various sources, which reflects the ambiguity surrounding its interpretation in scientific and practical contexts (Veselovsky et al., 2021, p. 293). But one definition most researches can agree on, is that AI is commonly seen as a subfield of computer science focused on replicating human intelligence, with the help of technology (Tuo et al., 2021, pp. 84–85).

2.3.2 Mel-Frequency Cepstral Coefficients

Mel-Frequency Cepstral Coefficients is a technique of feature extraction, focusing on acoustic features. As such it is widely used in speech- and audio-based applications (Kathiresan et al., 2019, p. 2). Research already shows that speech recognition software and speaker identification can benefit from the incorporation of MFCCs (Schädler et al., 2012, p. 4143; Singh et al., 2022, p. 1). Additionally, MFCCs are known for their robustness and effectiveness in capturing spectral cues extracted from data in speech- and audio signals (Kathiresan et al., 2019, p. 2). “MFCCs are calculated by applying a discrete cosine transform to spectral slices of the Mel-spectrogram” (figure 5) (Schädler et al., 2012, p. 4143)

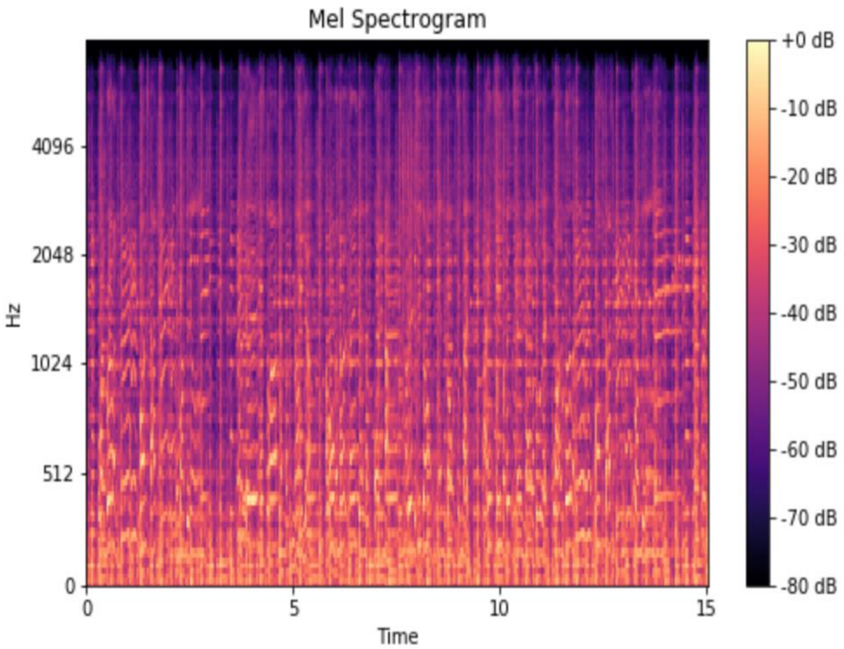

Figure 4: Mel-spectogram

Aside from the use in speech solutions, MFCCs have been leveraged in fields like music classification and mental state evaluation in sports, which pronounces its versatile potential in sound-based and psychological environments (Li et al., 2023; Xing, 2022). These promising results make MFCCs a popular choice when it comes to feature extraction in audio and speech signal processing (Bunrit et al., 2019, pp. 143–144)

2.3.3 Convolutional Neural Network

Artificial neural networks specifically designed for analyzing visual data, are called Convolutional Neural Networks (Hassanien & Darwish, 2021, p. 67). The specialty of these CNNs is to automatically build a hierarchy of features, based on the input context-data (Yamashita et al., 2018, p. 612). These networks make use of multiple layers, including convolutional layers filtering the context-data, subsampling layers or rather pooling layers reducing spatial dimensions, and fully connected layers, which output the final evaluation in context of a predetermined use-case (Hassanien & Darwish, 2021, p. 68; Mustafayev & Azimov, 2021, pp. 66–67).

The usage of a hybrid-network combining CNNs with Long Short-Term Memory (LSTM) networks has already exhibited successful results in connection to extracting “context-aware” emotionally relevant features (Trigeorgis et al., 2016, p. 5200). Furthermore, another useful characteristic of CNNs is its robust performance concerning uncommon speech signals, such as tonal speech, when the utilization of specialized databases for targeted research is given (Dua et al., 2022, pp. 1, 2, 3, 11)

While the strengths of CNNs lies in automatically building feature hierarchies based on input context-data and its robust performance regarding voice recordings with prevalent background noise, they may not be completely ideal for analyzing speech data when used by itself. This is caused by the deficit of not being able to effectively learn long-term context dependencies in sequential data. In addition, other common problems of CNNs, which could also disadvantage the analyzation of speech signal, include overfitting and gradient vanishing (Fantaye et al., 2020, p. 3). Nevertheless, the specialty of CNNs to enhance speech recognition even with high noise floor, demonstrates its superior performance compared to other models like multi-layer perceptron (Hepsiba & Justin, 2022, p. 3).

2.3.4 Long Short-Term Memory

Long Short-Term Memory, also known as LSTM, is a unique kind of recurrent neural network. It is engineered to tackle the common problem of fading and exploding gradient issues in regular recurrent neural networks (Fantaye et al., 2020, p. 2). The strengths of LSTM networks lie in their ability to handle long-term dependencies in sequential data, like speech signals for example. They employ a kind of "gating-system” to achieve this feature. The gating-system helps to manage how data flows within the LSTM over long stretches of time (Hochreiter & Schmidhuber, 1997, p. 1735). Because of this feature, LSTMs are a perfect fit for tasks and topics that require sequential data processing. This includes voice recognition, language modeling, and predicting time series (Tai et al., 2015, pp. 1556–1557)

LSTM networks consist of so-called memory cells. These cells are able to hold and save information over a prolonged period of time. This allows the LSTM to “remember” dependencies within sequential processing, over long periods of time. As a result the efficiency of LSTMs in processing sequential data reaches an impressive computational complexity of O(1) per time step and weight (Hochreiter & Schmidhuber, 1997, p. 1768). Another thing worth mentioning is the possibility to further optimize the LSTM network by integrating a bidirectional characteristic (Bidirectional long-short-term memory). Therefore, BDLSTMs would be able to utilize both past and future context to evaluate input data, enhancing the ability to find complex patterns in sequential data (Kanwal et al., 2022, p. 448).

2.4 Approaches in relation to cultural cues

In this chapter, studies related to cultural cues of deception will be discussed. Cultural cues to deceit define how people might lie in different cultures. These cues encompass different communication styles, which are influenced by cultural dimensions such as low- and high-context communication or Power Distance Index (PDI). The types of lying discussed at the beginning of this chapter may also play a crucial role in connection to the cultural differences regarding communication styles.

Cross-Cultural Deception Detection:

1. Methodology

The researchers conducted a literature review analyzing past approaches and literature in connection with cross-cultural communication, especially linked to deception. The focus of this review was to examine how people from different cultures produce and detect lies with a focus on verbal and non-verbal cues of lying (Taylor et al., 2015, p. 176).

2. Hypothesis

It was assumed that cultural norms have an impact on how people perceive "suspicious" behavior. Additionally it was hypothesized, that verbal and non-verbal cues to deception vary across different cultures, therefore affecting the ability of a person to produce and perceive deception (Taylor et al., 2015, p. 176).

3. Results

Research revealed major differences in how people from various cultures perceive deception. Their ability to detect lies also differed. In addition, verbal and non-verbal cues to deception varied across different cultures, proportional with the cultural differences in social and cognitive functioning. This led to a deterioration of people's ability to detect deception across different cultures (Taylor et al., 2015, pp. 193–194).

4. Connection to current research

As cultural differences regarding norms and expectations across different cultures were evident it can be deduced that they influence the interpretation of verbal and non-verbal cues in a cross-cultural exchange. Therefore, cross-cultural communication may lead to misinterpretations of emotions and intentions, affecting the ability to communicate with each other and consequently also influencing the ability to differentiate between truth and lie (Taylor et al., 2015, pp. 193–194).

Cross-cultural verbal deception:

1. Methodology

The participants of this study included people from Britain, China, and Arab countries. These countries represent cultures with low-context and high-context communication styles. The study examined how people from different backgrounds lie during interviews, by analyzing 17 verbal cues (Leal et al., 2018, p. 192).

2. Hypothesis

The hypothesis of this study encompassed, that low-context cultures exhibit more information, such as emotions and interaction when communicating. The study also assumes that low-context representatives exhibit more interactive cues in contrast to contextual cues regarding truthful or deceptive behavior (Leal et al., 2018, p. 198).

3. Results

The findings prove that participants of low-context cultures, such as Britain exhibit vocal cues to deceit more frequently than in high-context cultures, which was represented by China and Arab participants. This difference in communication between different cultures impacted the ability to detect and produce deception (Leal et al., 2018, pp. 208–210).

4. Relevance to current research

The results indicate cultural background and communication styles influence vocal cues to deception. In high-context cultures, where implicit communication is more prevalent, deception cues may be more obscure or differ from those in low-context cultures, affecting the accuracy of detecting lies across cultures with different communication styles (Leal et al., 2018, pp. 208–210).

2.5 Approaches in relation to vocal cues

In contrast to the previous chapter, the research in this chapter focuses on the ability to detect deception based on vocal cues. Vocal cues can affect the depiction of a speech signal (figure 5) and thus the feature extraction method of the AI model. Vocal cues are affected by factors such as pitch amplitude and intonation. In this context, emotions play an important role as they are one of the main influences on pitch or intonation, especially when a person is lying.

Influences like audio quality and language may also have an influence on the depiction of the speech signal.

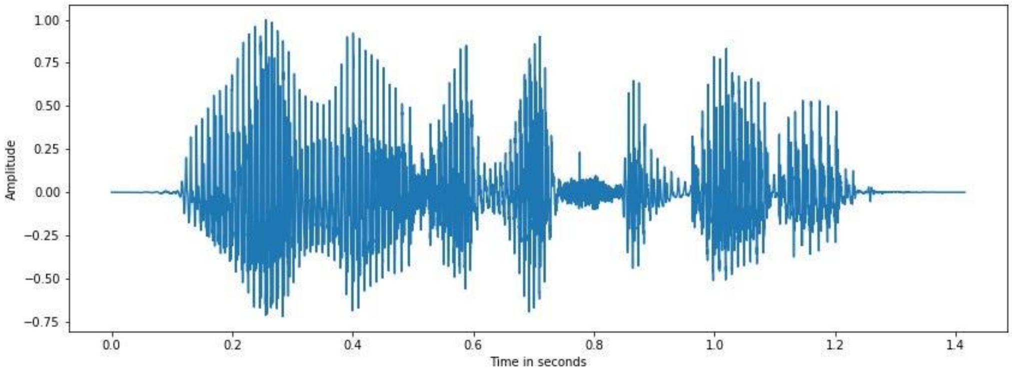

Figure 5: Speech signal

Speech Based Automatic Lie Detection:

1. Methodology

The study utilized homomorphic speech processing technique to extract vocal cues, with the goal of differentiating between normal speech and stressed speech. The cues were then used to find correlations to deception. The base of this study were recordings of real criminal cases and videos of an actor simulating different emotions (Gadallah & Algezawi, 1999, pp. 1, 3).

2. Hypothesis

The researchers assumed, that formant frequencies, pitch amplitude, and pitch contour, etc. would change with emotional stress from guilt. They therefore also assumed that these changes could be indicative of lying (Gadallah & Algezawi, 1999, p. 5).

3. Results

The study revealed several vocal cues linked to deception. The most reliable indicator, with 92% accuracy, was pitch contour. Pitch contour proved most sensitive in distinguishing truthful and untruthful statements. Other cues like maximum pitch amplitude and mean energy also showed decent 76% accuracy in detecting lies (Gadallah & Algezawi, 1999, p. 7).

4. Connection to current research

Emotional stress impacts vocal cues, such as pitch, pitch contour and formant frequencies. They shape the way a speech signal is depicted. These cuess could therefore expose dece-ption, since lying correlates to emotional distress. Especially, pitch contour shows promise for detecting the emotional stress associated with producing deception (Gadallah & Algezawi, 1999, p. 7).

Deep Neural Networks for Lie Detection with Attention on Bio-signals:

1. Methodology

The study employs an AI model that combines physical cues (facial micro-expressions) and vocal cues (speech signal) for detecting lies. The data collection was executed with the use of interviews or rather recordings, where researchers extracted certain micro-expressions and 20 MFCC features indicative of deception. These two sets of features, are then combined to determine if someone is truthful or deceptive (Bhamare et al., 2020, pp. 143–145).

2. Hypothesis

The research assumes that specific vocal cues like pitch, frequency, and volume are reliable indicators for detecting lies. It also assumes that including vocal cues alongside physical cues will improve the accuracy and reliability of lie detection (Bhamare et al., 2020, p. 143).

3. Results

Taking audio features in addition to micro expressions into the equation shows more promising results than traditional methods for detecting deception. The model's performance is heavily influenced by the involvement of MFCC based features, when identifying deception (Bhamare et al., 2020, pp. 143, 146).

4. Connection to current research

The results suggest that differences in vocal, for example pitch, frequency and loudness are effective indicators of deception, meaning any changes in these cues can affect the performance of the AI model. The results also show that the performance of MFCCs are more superior compared to physical indicators (Bhamare et al., 2020, pp. 143, 146).

Lie Detection using Speech Processing Techniques:

1. Methodology

The research of this paper utilized speech processing techniques, with the goal of detecting deception. The process involves analyzing the speech signal of the participants of the experiment and extracting certain features (MFCCs), which could be indicative of lying. In the next step a Support Vector Machine (SVM) based classifier is used to categorize the speech signal as truthful or deceptive. A real-world dataset containing truthful and deceptive speech data was gathered in the scope of this research (Bareeda et al., 2021, pp. 1–2).

2. Hypotheses

The paper assumes that the psycho-neural aspects of speech signals can be exploited to predict whether an untruthful or truthful statement was made. It is also assumed that the extracted vocal features, particularly the MFCCs, play the most important role in identifying cues to deception in speech (Bareeda et al., 2021, pp. 1–2).

3. Results

After training the SVM based classifier with different features, conducting several rounds of experiments, the best model reached an overall accuracy of 81% in distinguishing between truthful and deceptive statements (Bareeda et al., 2021, p. 7).

4. Connection to current research

This study also confirms that MFCCs on their own are an effective approach for identifying vocal cues related to deception. The author suggests that voice mirrors and conveys emotions, and this play a critical role in assessing deception. This result contributes to the fact that leveraging speech signals to detect deception is one of the most reliable methods at this point of time (Bareeda et al., 2021, pp. 1–7).

3. Methodology

3.1 Formulation of the hypotheses

1. **The ability to detect lies and truths with the Vietnamese data set will differ from the German data set:**

   As the current AI model was trained only with speech data from German participants, the AI should not be able to produce similar results when evaluating the Vietnamese data set. As the theory suggests, several factors such as cultural values and communication styles should influence people's opinion about deception, the manner, in which they engage in deception and the perception of deceptive behaviour.

2. **Questions in the context of Confucian cultures have an influence on the recognition of lies:**

   The second hypothesis addresses the question of whether discussion topics with a connection to the subject's culture have an influence on how lies are evaluated. To test this question three of the four topics chosen held a significant connection to Confucian culture. Topic 1 deals with a political question that addresses the aspect of authority in Confucian culture. Discussion topic 3 addresses group cohesion in collectivist cultures and the last topic 4 references the controversial issue of animal welfare in Vietnam.

3. **Questions in the context of group cohesion in Confucian cultures have an influence on the recognition of truths:**

   Group cohesion is the basis of collectivistic cultures. People highly value their various in-groups and strive to keep the harmony intact. The question “Should elderly family members be cared for at home by the family?” questions the loyalty and values of a person in context to group-cohesion. Thus, this question should be the most challenging to truthfully answer. In which exact way this challenge has an influence on the evaluation of the AI remains to be seen.

3.2 Description of the data collection procedure

Data set description:

The research paper “Using NLP to analyze whether customer statements comply with their inner belief” had German data sets readily available, so only language data representing Confucian culture needed to be collected. The experiment involved 75 participants, all of whom were born and raised in Vietnam. The participants' ages ranged from 20 to 67 years, with 48% being female and 52% male. To ensure the authenticity of the data set, all participants chosen, were born and raised in Vietnam.

The collected data consists of voice recordings with an average duration of 1 minute and 45 seconds. The voice recordings contain statements related to a previously selected discussion topic. About two-thirds of the participants were contacted directly by the author's parents, while the remaining third were mainly friends of previous participants.

Topics:

In order to guarantee that the test subjects have a clear opinion on a topic, four easily understandable and controversial or polarizing topics were selected. Three of the questions were tailored specifically to Confucian culture.

Topic 1: "Should the death penalty be abolished in Vietnam?"

Due to the fact that the death penalty is still active in many countries in Confucian culture (China, Vietnam, North Korea) and that the death penalty is a controversial topic in general, this topic made it into the final selection. In addition, this issue has great significance for Confucian culture in the context of authority.

Topic 2: "Should it be possible to buy alcohol at the age of 14?"

This question was chosen, not because it is controversial, but because the majority of people have the same opinion on this question. In this way, we can better minimize half-truths within the language recordings and help the participants to take a clear position in their argumentation.

Topic 3: "Should elderly family members be cared for at home by the family?"

This question was chosen to create a context for the participants, as Confucian countries are collective cultures that focus on family and group cohesion and harmony. This question is therefore also an issue on which the majority have a clear opinion.

Topic 4: "Should animal testing for research purposes be banned?"

This discussion topic was included in the final selection as animal welfare is a controversial issue in Vietnam and Asian countries. This is due to the fact that animal rights in agriculture, wildlife trade and bear farming (traditional medicine) are still a major problem for animal welfare in general. Moreover, it is easy to find arguments for this issue, which makes the process of building the argumentation for the debate easier.

<u>Debate: Distribution & Results:</u>

After the first round of interviews, a total of 75 participants were recorded who took part in the experiment. Out of 75 people, 10 people were assigned the first topic, 22 people the second, 24 people the third and 19 people the fourth. This resulted in a total of 33 argumentations in which the test person lied and 42 argumentations in which the test person told the truth. Out of 41 pro-arguments, 21 were lies, while only 12 of the 34 contra-arguments were lies.

	Topic 1	Topic 2	Topic 3	Topic 4	Sum
Pro-position	5	13	13	10	41
Contra-position	5	9	11	9	34

Table 1: Pro & Contra distribution

	Topic 1	Topic 2	Topic 3	Topic 4	Sum
Argumentation reflected the speaker's conviction	7	11	13	11	42
Argument did not reflect the speaker's conviction	3	11	11	8	33

Table 2: Truth & Lie distribution

Data collection:

General:

The data collection was organized in the form of a debating club. Two people were therefore always consulted per interview. The following measures were taken to ensure that a satisfactory data set was generated: First, there were four predetermined controversial and polarizing topics that were randomly picked. Second, whether a person argued for or against a topic was determined at random.

Test process:

Instructions for the test were provided beforehand, to assure the understanding of the test process before the actual exposure to the interview. The instructions are as follows:

First, the two participants were randomly assigned one of the four discussion topics. Afterwards, the pro or contra position regarding the discussion topic were also randomly assigned. If there were any questions at this point, they were clarified. Once everything is clear, the arguments for the voice recording are prepared. The test participants were given about 20-30 minutes to do this. It is expressly recommended not to write a continuous text, but only to prepare key points for the argumentation. Once the preparation for the argumentation has been completed, the actual experiment is carried out, i.e. the voice recording is generated.

Following the voice recording, demographic information, including name, age and profession, was recorded. Additionally, the topic and the actual individual's opinion on the discussion topic were documented. A questionnaire was also administered to capture the certain personality traits of the participants. Participants were, for example, required to answer how adventurous they consider themselves to be. The participants were also asked how nervous they felt before and during the recording, as well as their level of comfort with their assigned position (pro/contra). All responses were given on a scale of one to five.

3.3 Description of the AI used for analyzing the speech data

This description is established upon the definitions of the key components of the AI, in the second chapter of this thesis in addition to the knowledge gained from the thesis of Faußer et al., which will be cited accordingly.

Definition:

To further enhance the understanding of how vocal cues can affect the ability of the AI model to detect lies, this chapter will focus on describing the specific AI model used in this study. With reference to chapter two, the AI mentioned in the introduction of this thesis incorporates a hybrid of two popular deep learning architectures, namely the Convolutional Neural Network

(CNN) and the Long Short-Term Memory (LSTM) network (Faußer et al., 2021, p. 6). The selection of this hybrid architecture is based upon its primary function to process speech data, with a focus on the detection of deception in speech signals. The ability of CNNs to analyze visual data, including speech signals, stands out in comparison to other models. And with the addition of LSTM networks the models further strengthen their ability to analyze speech data, as LSTM networks are particularly well tailored to sequential data processing, in this case speech recognition.

The AI's ability to analyze goes beyond the simple evaluation of words in speech data. The model was trained to look for features, such as hints of emotions reflected in vocal cues, in speech signals (Faußer et al., 2021, p. 6). With this specification, the aim of this AI model is to accurately evaluate if a person is being truthful or not. This evaluation is based on the data or rather features extracted from the German data set, the AI model was trained with (Faußer et al., 2021, p. 18).

As mentioned, the AI uses feature extraction to look for distinct features in the speech signal, which could be indicative of lying (Faußer et al., 2021, pp. 10–11). The specific feature extraction technique used in this model is called MFCCs. It is a popular technique used when dealing with sound and voice, as MFCCs are known for their effective retention of spectral cues in speech signals. The AI employed two different amounts of MFCCs. The first eight models trained, utilized 13 MFCCs, the other eight models utilized 40 MFCCs (Faußer et al., 2021, p. 11).

All results were reflected in the following metric parameters:

True Positives (TP), describe "the number cases in which the model correctly categorized the conviction of a speaker as represented by the position he/she took in the debate." (Faußer et al., 2021, p. 13)

False Negatives (FP), describe "the number cases in which the model did not correctly categorize the conviction of a speaker as represented by the position he/she took in the debate." (Faußer et al., 2021, p. 13)

True Negatives (TP), describe "the number cases in which the model correctly categorized the conviction of a speaker as not represented by the position he/she took in the debate." (Faußer et al., 2021, p. 13)

And finally, False Positives (FP), describe "the number cases in which the model did not correctly categorize the conviction of a speaker as not represented by the position he/she took in the debate." (Faußer et al., 2021, p. 13)

The AI models were judged on three performance metrics derived from the metric parameters, namely accuracy, precision and sensitivity. The accuracy of the AI model describes the portion of correctly categorized lies and truths (TP+TN) across all recordings (TP+TN+FP+FN). Sensitivity measures the portion of correctly categorized truths (TP) in contrast to the number of actual truthful statements (TP+FN) i.e. it highlights the ability to avoid incorrect categorizations of truths (FP). Precision measures the portion of correctly categorized truths (TP) in contrast to statements, which were evaluated truthful (TP+FP), i.e. it highlights the ability to avoid incorrect categorizations of lies (FN) (Faußer et al., 2021, pp. 13–14).

After testing all models, the best model, which utilized 40 MFCCs, accomplished a convincing accuracy of about 98%, a precision of 100% and a sensitivity of 98% (Faußer et al., 2021, p. 15).

To summarize, this AI tool proved to be an effective solution in differentiating untruthful from truthful statements. Therefore, the method of using speech data has turned out to be more than sufficient, to detect lies.

4. Results

4.1 Comparison of the Occidental and Confucian data sets

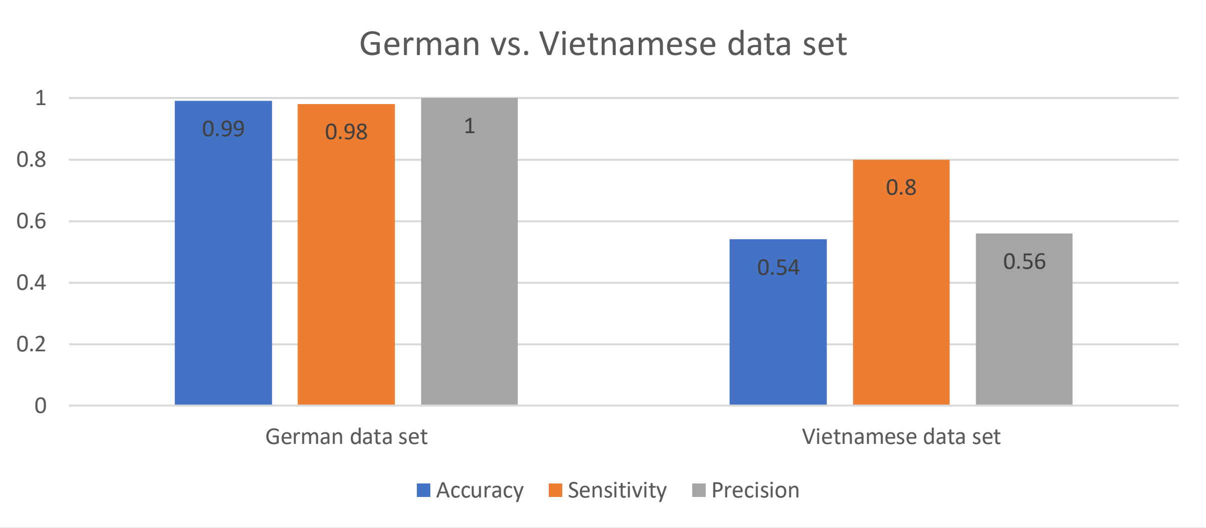

Figure 6: Metrics German & Vietnamese data set

On the first look, we can see a decrease in every performance metric across the board when comparing the German and Vietnamese data set. This clearly indicates that the cultural background of a person has an impact on the ability to detect lies, when using an AI trained solely on a dataset representing a single culture.

The overall ability to correctly detect truths was still satisfying as represented in the sensitivity metric, with an average of 80% across all 4 topics. Thus, the drop from 99% to 54% accuracy is mainly due to the lack of performance in correctly detecting lies (figure 8). The ability of the model to avoid falsely categorizing truths (figure 8), i.e. sensitivity also took a big hit causing a decrease from 100% to 56% when comparing the two data sets.

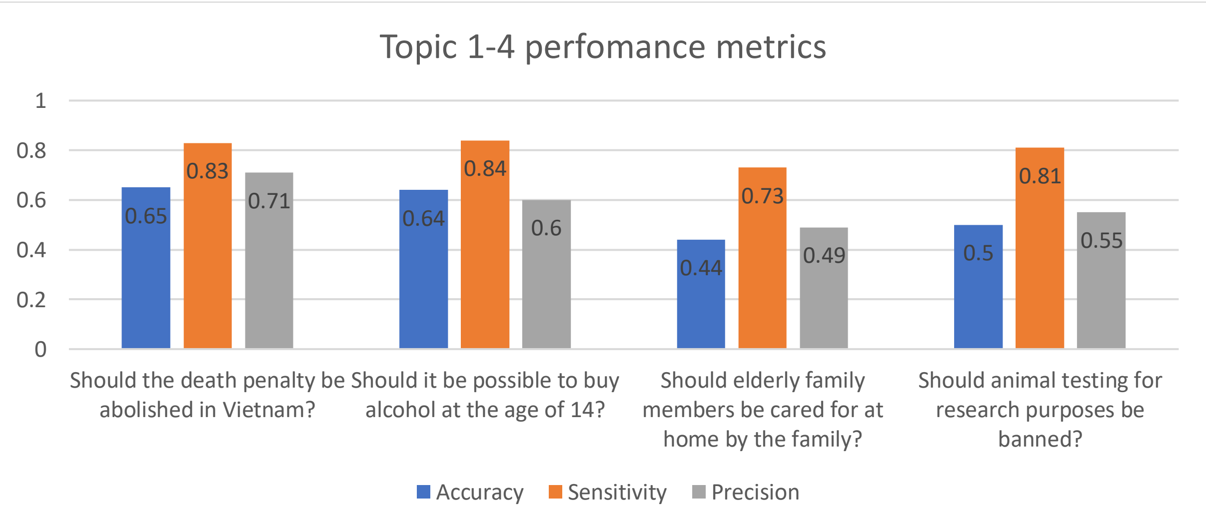

Figure 7: Metrics Topic 1-4

The data (table 2) shows that the balance of untruthful and truthful statements, regarding topic 1, is uneven. Only three out of ten participants answered according to their conviction. With respect to the evaluated sensitivity of topic 1, it can be deduced that the accuracy of 65% may be lower in reality, if a more even distribution can be achieved. Considering this fact, topic 2 in the Vietnamese data set shows the best results with an accuracy of 64% with an even distribution of untruthful and truthful statements. This accuracy could be achieved, because the AI was able to correctly detect lies, if a participant argued about topic 2 and only topic 2. The AI model was unfortunately not able to detect lies, regarding topics 1, 3 and 4 (figure 8).

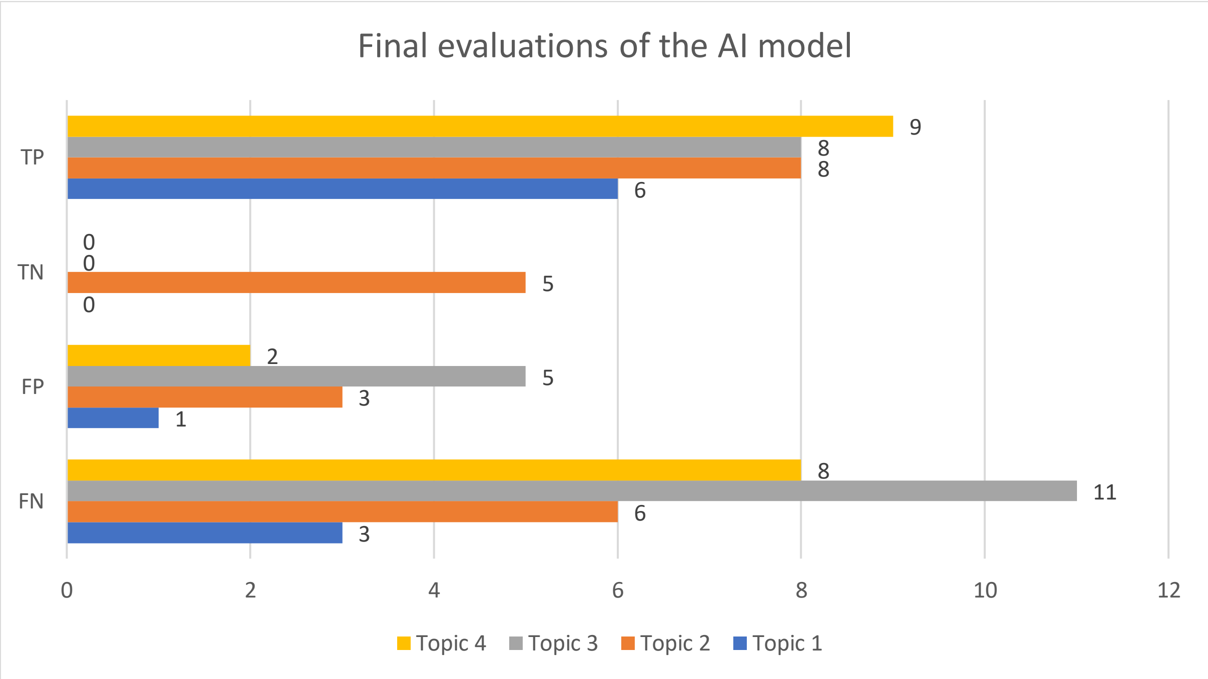

Figure 8: True Positive, True Negative, False Positive & False Negative

4.2 Testing of the hypotheses

1. **The ability to detect lies and truths with the Vietnamese data set will differ from the German data set:**

   As seen above (figure 6) in the subchapter before, there is a significant difference between the German and Vietnamese datasets evaluated by an AI model trained on German data. The evaluation metrics show a decrease in every performance metric of the AI model for the Vietnamese voice recordings. Therefore, it can be concluded that the two data sets differ from each other, when evaluated by an AI trained on German data.

2. **Questions in the context of Confucian cultures have an influence on the recognition of lies:**

   Out of all the 33 untruthful voice recordings, only five were correctly recognized by the AI Coincidentally all of these correctly recognized lies are statements regarding the question "Should it be possible to buy alcohol at the age of 14?" (figure 8). This was the only question out of the four discussion themes, which does not have any context to the culture of the participant. All the other untruthful statements, which had a strong context to the participants culture, could not be correctly categorized by the AI Thus, one can clearly determine the influence of questions with cultural context on the ability to recognize lies.

3. **Questions in the context of group cohesion in Confucian cultures have an influence on the recognition of truths:**

   On the other hand, a closer look will now be taken at the truthful voice recordings of the participants. While the performance to correctly recognize truths from the Vietnamese data set was very good, the German data set still had the upper hand in terms of accuracy. Overall, the AI was able to correctly categorize 31 out of 42 truthful statements. An interesting fact here is, that almost half of the incorrectly categorized truths are arguments regarding topic 3 "Should elderly family members be cared for at home by the family?" (figure 8). This exact topic thematizes the group cohesion and harmony within Confucian culture. In consequence, it is confirmed that discussions regarding group cohesion have an influence on the ability of the AI to recognize truthfulness.

5. Discussion

5.1 Evaluation of the results

The following discussion points out several factors which may have an impact on the evaluation of the AI, furthermore factors which may be excluded from having an impact will also be emphasized.

This whole chapter is based on the findings in chapter two, including all definitions and approaches described. If any additional information is provided, it will be cited as such.

Influences, which may not have an influence on the evaluation of the AI model:

Language (semantics):

Since MFCC is a method which extracts its data or rather features from speech signals, the semantics of a language are irrelevant in context of the AI evaluating whether a statement is untruthful or not. This discussed semantics of languages encompasses the meaning of words, sentences and expressions. MFCCs therefore focus on the inherent structural and statistical properties of speech signals, rather than their semantic content. MFCC is thus not concerned with the meaning behind the words or sentences in the voice recordings of the participants, instead it focuses on the acoustic properties of the recording, analyzing vocal cues like pitch, intensity, and spectral content.

<u>**Influences, which may have an influence on the evaluation of the AI model:**</u>

1. Memory model (Lisa Feldmann Barrett):

According to Barrett, emotions have not manifested in our brains since birth. They are intricate structures, built from past experiences, cultural teachings and memories. As a result, how we react emotionally to external stimuli is shaped by our personal history, education, and cultural environment. This means emotions are not universal and differ from person to person. They can change between different cultures, impacted by differing values, standards, and experiences, as well as individual experiences (Feldmann et al., 1993, p. 3).

Therefore, the ability to produce deception also relates to the interconnectedness of our emotions and memories. In some cultures, certain feelings are more accepted or rather preferred more than others. If we take a collective society, where harmony and cohesion are highly regarded, as an example. In these kinds of societies, showing anger or frustration may disrupt said harmony. As a result, people express themselves less, in consideration of others.

In these circumstances, individuals may adapt as life goes on. They manage their emotions and put up a front, suppressing their emotions in the process. This is done to meet society's standards, even if it means to completely hide one’s genuine feelings. Past experiences and cultural rules about sharing feelings can shape this mask making skill.

The way people remember and recall emotional moments can also change how well a person lies and the emotions involved when lying. Strong feelings linked to a certain event might make lying more or on the other hand less obvious.

The blend of emotions, memories, and culture heavily shapes how people deal with the tricky task of lying, especially across different cultures. This important difference can lead to a lot of misinterpretations in cultural exchanges. Therefore, understanding this dynamic helps to gain an insight into the subtle ways people express their feelings and present themselves in a cross-cultural context, in order to avoid any misunderstandings (Feldmann et al., 1993, pp. 3, 4).

2. Conditions of lying:

Intention to deceive:

People from different cultures may have different values and norms, which can affect how deception is produced and judged. For example, in certain cultures, telling white lies may be accepted to keep social relationships intact. As such, the level of guilt someone feels when lying could be less present, if it benefits their circumstance or favors their own values. People from different backgrounds might have different thresholds for what they categorize as a clear attempt to deceit.

In addition to this dynamic, the test environment of this thesis' data collection, in which the participants have been interviewed, may have lessened the emotional burden when lying even more. The participants should not have had any intention to deceive, because it was required to lie. Consequently, any emotions, which could influence the vocal cues of the interviewee, like guilt or shame when lying, may have been less pronounced in voice recordings.

Untruthful condition:

To consider a statement a lie, it must be false from the speaker's perspective. But making an untruthful statement is not absolutely necessary to produce deception. Other ways to create deception is by providing misleading information or not providing any information at all. Consequently, in cultures where context-heavy communication or exaggerations are common, it may be more difficult to differentiate between factual inaccuracies or lies in a cross-cultural context. Lies can also blend with factual inaccuracies, which makes it even harder to judge deception. The act of lying may be less emotionally charged, if it makes use of the aforementioned fact distortion or misinformation, rather than presenting a straight, untrue statement. Therefore, the way lies are told can impact how deception is judged by humans and artificial intelligence likewise.

3. Cultural Characteristics:

Collectivism vs. Individualism:

People's vocal cues can differ a lot when they lie. This difference may be based on their respective cultural backgrounds, in this case collectivistic or individualistic. In collectivist cultures, people often tend to hide their true feelings, in context to cultural circumstances. These circumstances could include group harmony, interactions with elders or authorities and the maintaining of a good public image (principle of saving face) (Hofstede et al., 2010, p. 110). Suppressing or hiding these feelings may result in quieter voices, less variations in vocal cues like pitch and intonation or less vocal cues reflecting emotions in general. This makes it tough to determine if someone is lying, based on vocal cues in a speech signal.

In contrast to collectivistic socies, individualistic societies, where independence and expressing oneself matter, people might be okay with showing their real feelings, even if they lie. Paying attention to one's own feelings and thoughts could make the vocal cues when lying clearer. Changes in sound, loudness, or talking speed could stand out more in these individualistic societies. It reflects the cultural norm of valuing one’s own opinion and the connecting emotions.

Taking everything into account, it is vital to grasp the cultural setting during communication. Because it shows how the mindset of a person affects how people act and sound when they lie. By being aware of this cultural difference, one can gain useful knowledge regarding cultural exchanges. Consequently, it may be easier to imagine how people from different cultures may differ in how they sound when they are not truthful.

Low-Context Cultures vs. High-Context Cultures:

People from low-context and high-context cultures have different ways of expressing feelings, especially when producing deception. In LCCs intents are mainly conveyed through spoken word or text, which may cause people in these kinds of cultures to exhibit more pronounced emotions when lying. This is due to the fact that people in this type of culture highly value clarity and transparency in communication. This characteristic might make it simpler to judge whether a person in a low-context culture is being truthful or not.

The communication style of people in high-context cultures is often less transparent and depends on shared context and body language. Thus, the vocal cues reflected while lying can vary in occurrence and volume. When people want to produce deception in high-context cultures (China), the corresponding cues to deception exhibit a division into several layers of communication, i.e. visual, spatial and auditory layers. This division of cues also differs in comparison to low-context cultures (Britain). Therefore vocal cues, for instance, are less frequent in high-context cultures (Leal et al., 2018, p. 205). It is therefore not easy to point out lies in this kind of communication style, because people make use of several dimensions to pass a message across. In addition to this, the cues to deception are also less apparent. Therefore, it is especially difficult to differentiate between untruthful and truthful statements, solely by vocal cues.

In simple terms, low-context cultures lean towards clear and transparent communication. This means when people from these cultures lie, their emotions may give them away more easily, especially through vocal cues, which are affected by these emotions. On the other hand, high-context cultures are all about reading between the lines in conversation. This might help them hide true feelings when they are lying, thanks to the more subtle nature of their communication. This highlights the connection between cultural communication norms and the expression of emotions, reflected in vocal cues, while lying.

High power distance vs. Low power distance:

The Power Distance Index (PDI) is a cultural characteristic, which may have to be considered in context to effective lie detection. The PDI score has a big influence on how people express their feelings when dealing with people in higher positions or seniority. This is especially the case, when someone is trying to conceal their feelings, to not upset a superior.

In cultures with a low power distance however, everyone expects to be seen and treated as an equal. There is less focus on rank or age compared to high PDI cultures. Consequently, flat hierarchies in companies are more common and decision-making power is allocated, depending on individual performance and responsibilities. In low PDI cultures, people are encouraged to share their opinion, as an exchange of opinions is believed to result in a "higher truth" (Hofstede et al., 2010, p. 107). This often leads to people showing their feelings more openly, even when they are not telling the truth.

This manifestation of PDI may translate to people showing more evident emotional reactions in their voice even when they lie, since individuals tend to feel empowered to express their true emotions. As such determining whether a person is truthful might be easier.

In high power distance cultures however, where hierarchy, respect for authority, seniority, and social ranks matter a lot, lying may look different. People in these cultures often hide how they truly feel and avoid sharing their own opinion in an exchange with authorities or rank differences. This behavior can cause quieter and more controlled speech that does not reflect their true emotions. Vocal cues become less apparent and distorted. Thus, it becomes hard to tell if someone lies just based on vocal cues.

In summary, the PDI score of a culture may challenge the ability to effectively detect lies. This is based on the difference of expressing emotions when comparing high and low power distance cultures. People in low power distance cultures may show more emotional response in their voice whereas in high power distance cultures people tend to suppress their emotions when engaging in interactions related to rank differences.

4. Language (prosody)

The way a person perceives and produces statements is heavily dependent on linguistic features, such as prosody. Prosody, which includes intonation, rhythm, and stress in speech for instance, serves as a crucial carrier of emotional and attitudinal cues within speech (Martinez-Alvarez et al., 2022, p. 1). The way a person understands prosody and especially rhythmic grouping (patterns of stress, intonation, etc. in speech), which is a component of prosody, is influenced by a lot of factors, for instance dialect, gender, social environment, memories, culture and language (Armstrong et al., n.d., pp. 783–789; Iversen & Patel, 2008, pp. 2263–2270).

The aforementioned findings suggest that differences in prosodic features across cultures, languages and social settings can influence emotional expressiveness within speech and emotional perception of speech.

In context to detecting lies, AI models, which use MFCCs to extract features, will have to learn from a diverse mix of speech data, which will ideally cover many different languages and its linguistic and sociolinguistic (ethnicity, race, …) characteristics. This is due to the fact that, how we say and perceive something can change based on various cultural and social influences, particularly with reference to how intonation and pitch can vary greatly across different languages (Mennen et al., 2014, p. 305) . Hence, knowing how linguistic features like prosody and sociolinguistic features behave may help the AI to detect lies in speech signals more accurately.

5. Audio quality

The effectiveness of artificial intelligence models, that utilize feature extraction techniques like MFCC are heavily dependent on the audio quality of recordings. The overall accuracy of the feature extraction process is directly influenced by the quality of the audio recordings. Audio quality characteristics or cues such as clarity, fidelity or resolution play an important role when it comes to effective feature extraction.

As the quality of audio recording drops, key details in the recordings may get distorted or are not present at all. In consequence the AI cannot leverage this data properly for study or decisions due to inaccurate data and information loss. Basically, the sound quality serves as foundation for the AI model. Therefore, any drop in sound quality reduces the model's accuracy, sensitivity and precision.

Recordings with insufficient quality do not have the ability to correctly reflect the original speech, making the evaluation of the AI model inconclusive.

In essence audio quality is important for feature extraction to work well. So, it is a must to get high-quality recordings for any analysis.

6. Cues to deceit

As stated by the results of the two studies regarding the cultural differences of deception detection and production, vocal cues to deceit differ across different cultures. Vocal cues are less apparent or are different altogether. This could lead to misinterpretations in a face-to-face setting. Even though it is not exactly stated, these misinterpretations could translate to evaluation errors of the AI model, when assessing cross-cultural data sets. The exact cultural causes were not proven in the scope of both theses, but some of the influences may be reflected in earlier parts of this discussion. As such the impact of high- and low-context cultures was correlated to these differences in cues to deceit in the work of Leal et al.

7. Inherent ability to produce deception

The overall volume of cues to deception, are less apparent in high-context cultures compared to low-context cultures. As such high-context cultures exhibit less cues evident of deceitful behavior across multiple layers of communication. Especially when comparing the frequency of auditory cues to deception, low-context cultures show about double the frequency (Leal et al., 2018, p. 205). This difference in frequency of detectable cues of deceit may translate to the better ability of people form high-context cultures to produce deception.

5.2 Possible influences on the hypotheses

With reference to the previous chapter, the above-mentioned influences are now assigned to the hypotheses if they show a certain correlation.

Audio quality:

As the audio quality was not uniform across all the recordings, which were evaluated, this influence may have had an impact on all three hypotheses tested. To test this assumption, five recordings, which were identical aside from the audio quality, were compared.

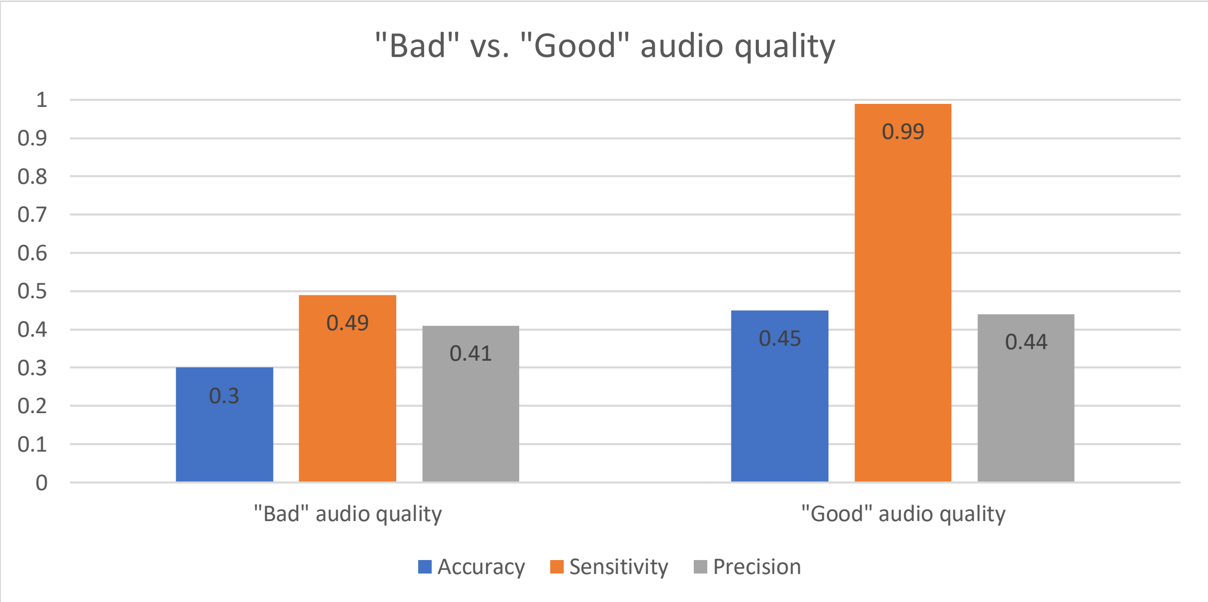

Figure 9: Metrics Bad & Good recordings

Due to the small sample size, the differences regarding the accuracy and the precision may be negligible. But one performance metric, which clearly benefited from the increase in audio quality, was sensitivity. This trend could also be further examined in a different comparison.

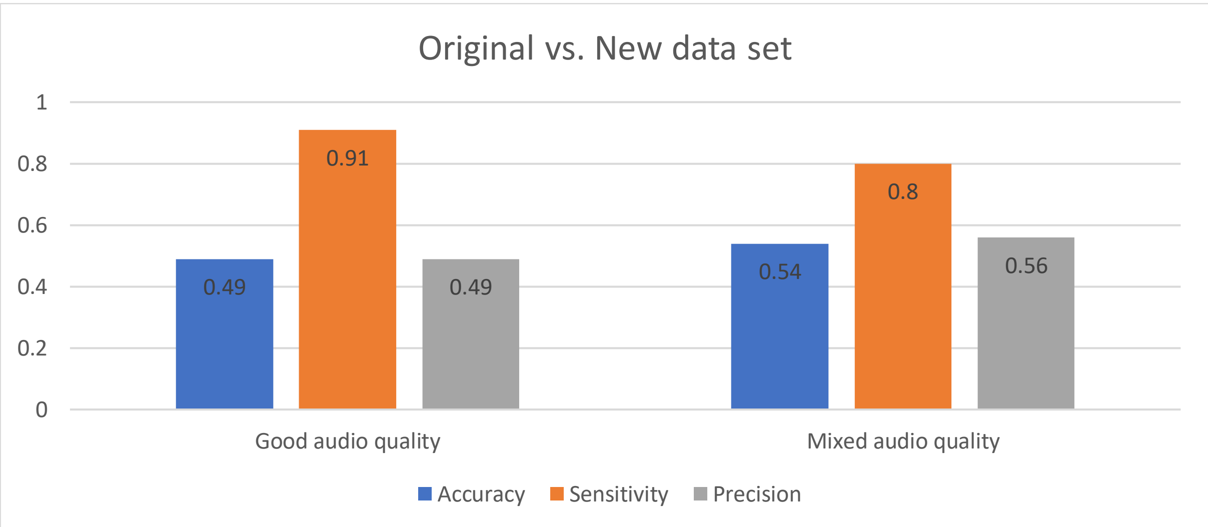

Figure 10: Metrics Original & New data set

For the comparison in figure 10, eleven recordings with “good” audio quality were extracted from the original data set of 75 participants. In addition, 20 new recordings were generated, while assuring good and uniform audio quality across all 31 recordings. As stated above, the performance metric sensitivity also showed the same improvement as shown in figure 9. However, the performance regarding detecting lies did not improve at all.

Language (prosodic):

All participants of this experiment, except one, spoke Vietnamese during their interview. Thus, the influence of language affects all three hypotheses tested.

Conditions of lying:

Both conditions, i.e. “intention to deceive” and “untruthful condition”, may have had an impact on all of the voice data of the participants.

Cues to deceit:

The cultural nature of the German and Vietnamese participants differed, as such the differences in cues to deceit across different cultures, may be applicable to all hypotheses tested.

Inherent ability to produce deception:

Vietnamese participants, should be able to conceal lies better, compared to German participants. This inherent ability may have had an impact on the overall ability of the AI model to detect lies regarding, thus also affecting all three hypotheses tested.

Collectivism vs Individualism:

Collectivism may have had an impact on the second hypothesis, as the question regarding this hypothesis (“Should elder family members be cared for at home by their family?”), questioned an aspect of group cohesion.

Cultural influences overall:

Topic 2 with the question "Should it be possible to buy alcohol at the age of 14?", had no correlation to cultural values at all, in consequence the effects of culture regarding cultural characteristics such as PDI and Collectivism, should not have had an impact on the recordings regarding this specific topic or at least should have had less impact on the speech data.

5.3 Implications for theory and practice

The discussion highlights how factors like test environment, cultural differences and linguistic can greatly influence the vocal cues in the speech of a person.

Cultural differences may cause differences in vocal cues. These differences can be divided into two categories. The first category would be that vocal cues to lying may differ altogether across different cultures. The second category describes how vocal cues, which were extracted from the German data set, are less apparent in the Vietnamese data set, thus influencing the evaluation of the AI model.

Furthermore, linguistic differences may pose a problem of the foundation of the AI, namely the feature extraction, based on the speech signal of a voice recording. This is assumed by the assumption that linguistic cues, like variation in pitch use and pitch range, can influence the depiction of the speech signal itself. Thus, the features extracted to detect lies from the speech signal may differ depending on cultural and social linguistic differences between the participants.

Another thing to note is the test environment. The test environment described and used in the scope of this thesis can only simulate a case of lying. As a result, it is unclear how the AI will perform when confronted with real life cases of lying, which in contrast to a simulated environment may result in different cues to producing deception.

Similar to the linguistic differences, the audio quality also poses a problem in context to the depiction of the speech signal. As the quality in audio drops, information within the speech signal is less pronounced or lost altogether. This can prevent the AI from leveraging certain features that are used to evaluate lies.

All of these enumerated differences in vocal cues might pose a challenge in identifying untruthful speech. Because the AI in this thesis' scope utilizes these vocal cues, with the feature extraction technique MFCC, it is recommended to test the influence of these various factors, to ascertain the exact influence on the performance of the AI to detect lies. When testing the aforementioned factors, it may be possible to neglect certain influences if they do not pose a significant performance decrease.

6. Conclusion

6.1 Limitations of the study and opportunities for future research

Limitations:

The Vietnamese data set, used to test the AI, and the German data, the AI model was trained on, did not include any real-life cases of lying. Therefore, it cannot be said for sure, that the results in scope of the used methodology of this thesis can translate to real life deception detection. While it may present a good foundation, the influence of a real situation is not completely reflected in the results of this paper.

Furthermore, the audio quality of all the samples, used to prove the formulated hypotheses, was not uniform. Some voice recordings were of higher quality than others, due to the compression process, when using audio formats such as “mp4” or “m4a”, which lead to a drop in the audio quality cues, such as resolution and fidelity. Moreover, about a quarter of the recordings exhibited a high level of noise floor (background noise), which may have had an influence on the evaluation process of the AI.

Personality traits, like extraversion or confidence influence the way deception is produced (Levitan et al., 2015). But because the questionnaire was not designed to analyze the detailed personality of a participant, it was not possible to allocate certain traits like extraversion or introversion, to each participant. If the methodology can be extended in context to personality analysis, it may be possible to optimize the AI model even further.

Another problem which could have influenced the results of this paper is the proper selection of questions. The first topic "Should the death penalty be abolished in Vietnam?" for example turned out to be a very sensitive question for the Vietnamese interviewees. This assumption is based on my experience in the process of interviewing several participants. Some interviewees were hesitant to discuss this topic, due to the reigning respect of authority in Vietnam. A few individuals have even withdrawn their participation in the experiment, due to this specific topic. In addition, questions regarding Confucian values (topic 2) may also be sub-optimal, as people may not have stated their genuine opinion, due to nature of the question.

Opportunities:

The two cultures analyzed in this thesis exhibited more differences in contrast to similarities. As consequence it may be more effective to use different AI models depending on the participant’s culture, instead of trying to train a universal AI model. As linguistics may also have an influence on the ability to detect lies, further specialization considering culture and language may optimize the ability of accurately detection deception even more. Continuing this trend, it may be possible to drive this specialization even further, by including factors mentioned earlier in this thesis like gender and dialect. Of course, this specialization of AI models is only possible if the diversity of the samples and the sample size itself allows it.

To further strengthen the foundation for the results in this thesis and gain further insight on the cultural influence on lie detection, it would also be beneficial to build an AI model, trained with the Vietnamese data set to complement the German-trained AI model. This new Vietnamese model would undergo a counter-test encompassing the evaluation of the available German data set and a subsequent comparison of performance metrics, i.e. accuracy, sensitivity and precision, between the two models.

6.2 Summary of the results and answering the research question

In retrospect the findings regarding the fundamentals and various approaches implicate a lot of possible influences on the ability of the AI models' hybrid architecture to recognize lies. These influences mainly focus on the depiction of the speech signal, which the AI uses as its foundation to differentiate between untruthful and truthful statements. In the scope of this thesis, factors such as language, emotions reflected in vocal cues and many other things were discussed. In the end these factors could not be proven true, as the test process was not designed to explicitly test these different influences. However, testing all of these influences would have gone beyond the scope of this bachelor's thesis.

On the other hand, the hypotheses which could be proven true, were the overall impact of culture on the AI's ability to recognize lies. Another interesting fact that could be confirmed is the importance of question-selection, when trying to detect deception across different cultures. Questions which had a strong relation to Confucian values, showed worse performance compared to questions which had not relation to Confucian values, when consulting the Vietnamese participants in this experiment.

In summary all of these findings imply that further research on the impact of culture should be considered, as the precise influence regarding the several aforementioned factors has not yet been examined. This consideration may further optimize the reliability of AI solutions in context of lie detection across different cultures or even lie detection overall.

IV. Literature and sources

Literature:

Arkhipov, V. V. (2022). Definition of artificial intelligence in the context of the Russian legal system: A critical approach. *Gosudarstvo i Pravo*, *1*, 168. https://doi.org/10.31857/S102694520018288-7

Armstrong, M., Breen, M., Gooden, S., Levon, E., & Yu, K. M. (n.d.). *Sociolectal and Dialectal Variation in Prosody*. 2022.

Bareeda, F., Mohan, S., & Muneer, A. (2021). *Lie Detection using Speech Processing Techniques*.

Bhamare, A. R., Katharguppe, S., & Nancy J, S. (2020). *Deep Neural Networks for Lie Detection with Attention on Bio-signals*. 2020 7th Intl. Conference on Soft Computing & Machine Intelligence.

Bond, C. F. Jr., & Atoum, A. O. (2000). *International Deception*. Personality and Social Psychology Bulletin.

Braun, B., Kochanski, G., Grabe, E., & Rosner, B. S. (2006). *Evidence for attractors in English intonation*.

Bunrit, S., Inkian, T., Kerdprasop, N., & Kerdprasop, K. (2019). Text-Independent Speaker Identification Using Deep Learning Model of Convolution Neural Network. *International Journal of Machine Learning and Computing*, *9*(2), 143–148. https://doi.org/10.18178/ijmlc.2019.9.2.778

Carson, T. L. (2006). *The Definition of Lying*.

Chen, A. (2009). *Perception of Paralinguistic Intonational Meaning in a Second Language*.

Chien, P.-J., Friederici, A. D., Hartwgisen, G., & Sammler, D. (2020). *Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers*.

Chisholm, R. M., & Feehan, T. D. (1977). *The Intent to Deceive*. The Journal of Philosophy.

Chollet, F. (2019). *On the Measure of Intelligence*. https://doi.org/10.48550/ARXIV.1911.01547

Clore, G. L., & Otorny, A. (2000). *Cognitive in emotion: Never, sometimes, or always?*

Coalition Against Insurance Fraud. (2022). *The Impact of Insurance Fraud on the U.S. Economy*.

Dua, S., Kumar, S. S., Albagory, Y., Ramalingam, R., Dumka, A., Singh, R., Rashid, M., Gehlot, A., Alshamrani, S. S., & AlGhamdi, A. S. (2022). Developing a Speech Recognition System for Recognizing Tonal Speech Signals Using a Convolutional Neural Network. *Applied Sciences*, *12*(12), 6223. https://doi.org/10.3390/app12126223

Ekman, P. (1985). *Telling Lies: Clues to Deceit in the Marketplace, Marriage, and Politics*. New York: W.W. Norton.

Eliot, T. S. (n.d.). *Notes Towards the Definition of Culture*.

Fantaye, T. G., Yu, J., & Hailu, T. T. (2020). Advanced Convolutional Neural Network-Based Hybrid Acoustic Models for Low-Resource Speech Recognition. *Computers*, *9*(2), 36. https://doi.org/10.3390/computers9020036

Faußer, S., Gewald, H., & Thaler, F. (2021). *Using NLP to analyze whether customer statements comply with their inner belief*.

Feldmann, L., Lewis, M., & Haviland-Jones, J. M.-J. (1993). *Handbook of Emotions*.

Feng, Y., Qiu, L., & Sun, B. (2021). A measurement framework of crowd intelligence. *International Journal of Crowd Science*, *5*(1), 81–91. https://doi.org/10.1108/IJCS-09-2020-0015

Fletcher, J., & Evans, N. (2002). *An acoustic phonetic analysis of intonational prominence in two Australian languages*.

Gadallah, M. E., & Algezawi, A. F. (1999). *Speech Based Automatic Lie Detection*. 16th 16th National Radio Science Conference, NRSC'99.

Gandour, J., Wong, D., & Dzemidzic, M. (2003). *A cross-linguistic fMRI study of perception of intonation and emotion in Chinese*.

Griffiths, P. J. (2004). *Lying: An Augustinian Theology of Duplicity*. Grand Rapids: Brazos Press.

Gupta, S., & Ortony, A. (2018). *The oxford handbook of lying and deception*.

Hansen, K. P. (2009). *Kultur und Kollektiv*.

Hassanien, A. E., & Darwish, A. (Eds.). (2021). *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges* (Vol. 77). Springer International Publishing. https://doi.org/10.1007/978-3-030-59338-4

He, K., & Sun, J. (2015). Convolutional neural networks at constrained time cost. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5353–5360. https://doi.org/10.1109/CVPR.2015.7299173

Hepsiba, D., & Justin, J. (2022). Enhancement of single channel speech quality and intelligibility in multiple noise conditions using wiener filter and deep CNN. *Soft Computing*, *26*(23), 13037–13047. https://doi.org/10.1007/s00500-021-06291-2

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hofstede, G., Hofstete, G. J., & Minkov, M. (2010). *Cultures and Organisations* (3rd ed.).

Iversen, J. R., & Patel, A. D. (2008). *Perception of rhythmic grouping depends on auditory experience*.

Kanwal, S., Khan, F., Alamri, S., Dashtipur, K., & Gogate, M. (2022). COVID-OPT-AINET: A clinical decision support system for COVID -19 detection. *International Journal of Imaging Systems and Technology*, *32*(2), 444–461. https://doi.org/10.1002/ima.22695

Kathiresan, T., Maurer, D., & Dellwo, V. (2019). Highly spectrally undersampled vowels can be classified by machines without supervision. *The Journal of the Acoustical Society of America*, *146*(1), EL1–EL7. https://doi.org/10.1121/1.5111154

Leal, S., Vrij, A., Vernham, Z., Dalton, G., Jupe, L., Harvey, A., & Nahari, G. (2018). *Cross-Cultural verbal deception*. Legal and Criminological Psychology.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Levitan, S. I., An, G., Wang, M., Mendels, G., Hirschberg, J., Levine, M., & Rosenberg, A. (2015). *Cross-Cultural Production and Detection of Deception from Speech*. WMDD '15: Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection.

Li, Z., Peng, H., Tan, S., & Zhu, F. (2023). Music classification with convolutional and artificial neural network. *Journal of Physics: Conference Series*, *2580*(1), 012059. https://doi.org/10.1088/1742-6596/2580/1/012059

Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: Constructing and applying a new search strategy. *Scientometrics*, *126*(4), 3153–3192. https://doi.org/10.1007/s11192-021-03868-4

Mahon, J. E. (2015). *The Definition of Lying and Deception*.

Martinez-Alvarez, A., Benavides-Varela, S., Lapillone, A., & Gervain, J. (2022). *Newborns discriminate utterance-level prosodic contours*.

Mennen, I., Schaeffler, F., & Dickie, C. (2014). *SECOND LANGUAGE ACQUISITION OF PITCH RANGE IN GERMAN LEARNERS OF ENGLISH*.

Molnar Monika, Lallier, M., & Carreiras, M. (2014). *The Amount of Language Exposure Determines Nonlinguistic Tone Grouping Biases in Infants From a Bilingual Environment*.

Mustafayev, E., & Azimov, R. (2021). Comparative Analysis of the Application of Multilayer and Convolutional Neural Networks for Recognition of Handwritten Letters of the Azerbaijani Alphabet. *Cybernetics and Computer Technologies*, *3*, 65–73. https://doi.org/10.34229/2707-451X.21.3.6

Nguyen, V. N., Xuan, L. S., & Lezhenin, I. (2021). *Adopting StudyIntonation CAPT Tools to Tonal Languages Through the Example of Vietnamese*.

Oatley, K. (2004). *Emotions a Brief History*.

Ordin, M., & Mennen, I. (2017). *Cross-Linguistic Differences in Bilinguals' Fundamental Frequency Ranges*.

Plutchik, R. (2001). *The Nature of Emotions* (5th ed., Vol. 89). https://www.jstor.org/stable/27857503

Rathje, S. (2009). *The Definition of Culture: An application-oriented overhaul*.

Rudolf-Möller, E. (2017). *Amygdala*. https://www.netdoktor.de/anatomie/gehirn/amygdala/

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Schädler, M. R., Meyer, B. T., & Kollmeier, B. (2012). Spectro-temporal modulation subspace-spanning filter bank features for robust automatic speech recognition. *The Journal of the Acoustical Society of America*, *131*(5), 4134–4151. https://doi.org/10.1121/1.3699200

Scott, G. G. (2006). *The Truth About Lying*.

Singh, M. K., Manusha, S., Balaramakrishna, K. V., & Gamini, S. (2022). Speaker Identification Analysis Based on Long-Term Acoustic Characteristics with Minimal Performance. *International Journal of Electrical and Electronics Research*, *10*(4), 848–852. https://doi.org/10.37391/ijeer.100415

Smith, D. L. (2004). *Why We Lie: The Evolutionary Roots of Deception and the Unconscious Mind*. New York: St. Martin's Press.

Sorensen, R. (2007). BALD-FACED LIES! LYING WITHOUT THE INTENT TO DECEIVE. *Pacific Philosophical Quarterly*, *88*(2), 251–264. https://doi.org/10.1111/j.1468-0114.2007.00290.x

Stokke, A. (2013). *Lying deceiving and misleading*. Philisophy Compass.

Storey, J. (2009). *Cultural Theory & Popular Culture and am Introduction*. Pearson Longman.

Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1556–1566. https://doi.org/10.3115/v1/P15-1150

Taylor, P. J., Larner, S., Conchie, S. M., & Van der Zee, S. (2015). *Cross-Cultural Deception Detection*.

Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2016). Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5200–5204. https://doi.org/10.1109/ICASSP.2016.7472669

Tuo, Y., Ning, L., & Zhu, A. (2021). How Artificial Intelligence Will Change the Future of Tourism Industry: The Practice in China. In W. Wörndl, C. Koo, & J. L. Stienmetz (Eds.), *Information and Communication Technologies in Tourism 2021* (pp. 83–94). Springer International Publishing. https://doi.org/10.1007/978-3-030-65785-7_7

Veselovsky, M. Y., Izmailova, M. A., & Trifonov, V. A. (2021). *Intellectual Governance in the Digital Economy of Russia:* International Scientific and Practical Conference "Russia 2020 - a new reality: economy and society” (ISPCR 2020), Veliky Novgorod, Russian Federation. https://doi.org/10.2991/aebmr.k.210222.057

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Welsch, W. (1995). *Transkulturalität. Zur veränderten Verfasstheit heutiger Kulturen.* Zeitschrift für Kulturaustausch.

Williams, R. (1961). *The Analysis of Culture*.

Xing, S. (2022). Sportsman’s Mental State Evaluation and Early Warning Method Based on Intelligent CNN. *Scientific Programming*, *2022*, 1–6. https://doi.org/10.1155/2022/4711490

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yuan, J. (2011). *Perception of intonation in Mandarin Chinese*.

Figures:

Figure 1:	https://www.statista.com/statistics/1365145/artificial-intelligence-market-size/
Figure 2:	https://www.statista.com/statistics/1298322/worldwide-ai-revenue-increase/
Figure 3:	Own illustration
Figure 4:	https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53
Figure 5:	https://speechprocessingbook.aalto.fi/Representations/Short-time_analysis.html
Figure 6:	Own illustration
Figure 7:	Own illustration
Figure 8:	Own illustration
Figure 9:	Own illustration
Figure 10:	Own illustration

V. Appendix

Appendix 1: Test instructions

Test: Does Cultural Background Influence Lying?

In the following test, I am investigating whether cultural background has an influence on lying using voice recordings (in the form of a debate). An A.I. (artificial intelligence) will examine based on your voice whether you are lying or telling the truth. The entire process will take approximately 15-20 minutes, with the actual recording of you arguing lasting about 1 ½ - 2 minutes.

Step 1: Topic

I will provide you with a topic that will be the subject of our discussion. The topics are usually easy to understand, such as "Should the death penalty be abolished?" If you have any issues with the topic, just let me know, and we can adjust it for you.

Step 2: Pro and Con Position

You will now be assigned either a pro or con position. This assignment is random. This means that even if you normally have a negative opinion on the topic, you must present pro arguments if you are assigned a supporting position. Therefore, whether you are lying or telling the truth is purely random.

Example:

You receive the topic "Should the death penalty be abolished?"

1. Scenario

(1) You agree with the question, meaning you want the death penalty to be abolished.
(2) However, I assign you the opposing role, meaning you must now argue that you do not want to abolish the death penalty.
(3) In this case, you must lie, but this is intentional, and you are not doing anything wrong.

2. Scenario

(1) You disagree with the question, meaning you do not want the death penalty to be abolished.
(2) I assign you the opposing role, meaning you must now argue that you do not want to abolish the death penalty.
(3) In this case, you tell the truth, which is also intentional, and you are not doing anything wrong.

Step 3: Clarify Open Questions

If you have any questions or anything is unclear, please clarify with me before you begin preparing your arguments.

Step 4: Prepare Arguments

You now have 5-10 minutes to prepare your arguments. It is important that you do not prepare a fluent text but only a brief summary. You must speak for at least 1 ½ minutes; it is not a problem if you have prepared more.

Example of preparing arguments correctly (no fluent text):

You receive the topic "Should alcohol be purchasable from the age of 14?"

Correct:	Incorrect:
- Too young, because ... - Damages your health while ... - ...	Alcohol should not be sold to 14-year-olds because they are still children who do not understand the risks of alcohol.

Step 5: Discussion

This is the final step. I will let you know when I start the recording, and you can present your arguments. There may be another person participating in the experiment. In that case, imagine you want to convince the person across from you with your arguments.

Step 6: Data Collection

The voice recording is complete, and I now need various general information from you, such as your age, your actual opinion on the given topic, etc., and ultimately questions regarding the experiment. If you do not have a copy of the questions, just take an additional sheet of paper and write down your answers. Then, send me a photo of it. Thank you very much.

First Name:

Last Name:

Email:

Topic:

Pro Position/Con Position:

Truth/Lie:

Gender:

Age:

Occupation:

Self-assessment Part 1:

How nervous were you before the experiment?

1 2 3 4 5
☐ ☐ ☐ ☐ ☐

(1 = I was not nervous; 5 = I was very nervous)

How nervous were you during the experiment?

1 2 3 4 5
☐ ☐ ☐ ☐ ☐

(1 = I was not nervous; 5 = I was very nervous)

How much can you identify with the position (pro/con) assigned to you?

1 2 3 4 5
☐ ☐ ☐ ☐ ☐

(1 = I do not identify with the assigned position; 5 = I identify with the assigned position)

Self-assessment Part 2:

1. I like to be with other people.	○	○	○	○	○
2. I can quickly spread a good mood.	○	○	○	○	○
3. I am adventurous.	○	○	○	○	○
4. I like to be the center of attention.	○	○	○	○	○
5. I often prefer to be alone.	○	○	○	○	○
6. I am a loner.	○	○	○	○	○
7. I like to go to parties.	○	○	○	○	○
8. I am active in many clubs.	○	○	○	○	○
9. I am a talkative and communicative person.	○	○	○	○	○
10. I am very sociable.	○	○	○	○	○

(1 = I disagree; 5 = I agree)

Appendix 2: Data protection form

Institut für Dienstleistungsmanagement (CROSS)
Project Manager: *Prof. Dr. Heiko Gewald*

Contact for inquiries:
Fabian Thaler
fabian.thaler@hnu.de
Telefon: 0731/9762-1546

CROSS
Center for Research on Service Sciences
Neu-Ulm University of Applied Sciences

Declaration of consent for data collection

Study title: Voice Analysis for Customer Emotions (VACE)

The VACE research project aims to analyze the mood of the speaker of a lecture with the help of artificial intelligence.

For this purpose, we collect the spoken word of the test person in anonymized form. This data set is then used to technically calculate different moods among the subjects. The processing and storage takes place exclusively within the IT infrastructure of Neu-Ulm University and only HNU employees have access to the data.

In order to ensure anonymity, the data collected is assigned by means of a subject number (ID).

The contact data (surname, first name and e-mail address) of the subjects are linked to this ID and are used exclusively for appointment coordination. In addition, basic demographic data (age, gender, etc.), as well as emotional self-assessments (e.g. degree of nervousness) of the subjects are collected. These are also linked exclusively to the ID in order to establish a reference to the audio file, which is necessary for the scientific analysis of the data.

Following the data collection, the contact data is deleted. A connection between contact data and ID is no longer possible from this point on and complete anonymization is guaranteed.

The data is evaluated without knowledge of the person. An assignment of the data to a specific person is irrelevant for the research project.

I am aware that I can revoke my consent to the retention or storage of this data without incurring any disadvantages. I have been informed that deletion of my recordings is no longer possible due to anonymization after completion of the data collection. A revocation can be made -before completion of the data collection- by email to fabian.thaler@hnu.de.

I have read and understood this declaration and hereby agree that my

above data be collected and processed accordingly.

__________ ______________ ________________

Place, Date   Name   Signature

I have received a copy of this consent form.

CROSS
Center for Research on Service Sciences
Neu-Ulm University of Applied Sciences

Institut für Dienstleistungsmanagement (CROSS)
Project Manager: *Prof. Dr. Heiko Gewald*

Contact for inquiries:
Fabian Thaler
fabian.thaler@hnu.de
Telefon: 0731/9762-1546

Consent follow-up study

☐ I am interested in a follow-up survey and allow you to contact me at the following email address in the future for this purpose:

____________________________________ ____________

(email-address) ID

Data protection information

1. Person responsible
Hochschule Neu-Ulm, Wileystraße 1, 89231 Neu-Ulm, Tel. +49 731/9762-0, E-Mai: info@hnu.de

2. Data Protection Officer Contact
Datenschutzbeauftragter, HNU, Wileystraße 1, 89231 Neu-Ulm, Tel. +49 731/9762-0, E-Mail: dsb@hnu.de

3. Purpose
Neu-Ulm University stores your name and e-mail address in order to be able to contact you in case of a follow-up study

4. Legal basis
The legal basis for the processing of the specified personal data for the above-mentioned purposes results from Art. 6 para. 1 lit. a DSGVO (consent).

5. Storage duration
We store your data in each case only until the purpose of the data storage ceases to apply, as long as no legal retention periods, obligations to provide evidence or limitation periods, which allow the legal prosecution of resulting claims, prevent the deletion (in this case, the processing of the data is restricted in accordance with Art. 18 DSGVO) or in cases of consent as long as this has not been revoked.

6. Pass on to third parties
The data will not be passed on to third parties.

7. Data security
To protect the personal data we manage against accidental or intentional manipulation, loss, destruction or against access by unauthorized persons, we use state-of-the-art technical and organizational security measures that are continuously improved.

8. Your rights
You have the right to obtain information from Neu-Ulm University about the data stored about you and/or to have incorrectly stored data corrected. To do so, contact the university as the responsible party or the data protection officer.
You also have the right to request deletion or restriction of processing or the right to object to processing. Furthermore, in the case where you have given consent for processing, you have the right to revoke the consent at any time, whereby the lawfulness of the processing carried out on the basis of the consent up to the revocation is not affected.
Furthermore, you have the right to lodge a complaint with the Bavarian State Commissioner for Data Protection (P.O. Box 22 12 19, 80502 Munich, Tel. 089 212672-0).

VI. Statutory Declaration

“I herewith declare that I have composed the present thesis myself and without use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such; other references with regard to the statement and scope are indicated by full details of the publications concerned. The thesis in the same or similar form has not been submitted to any examination body and has not been published. This thesis was not yet, even in part, used in another examination or as a course performance.”

Place, Date: 13.03.2024 // 89423 Gundelfingen ……… Signature: ……… T. Nguyen ………