INSTITUTO
UNIVERSITÁRIO
DE LISBOA

An unequal race for attention? A comparative case study between human and virtual influencers on social media

Joshua Enzig

Master (M.Sc.) in Marketing

Supervisor: Prof. Dr. João Guerreiro
ISCTE-IUL Business School,
Department of Marketing, Operation and general Management

October, 2022

BUSINESS
SCHOOL

An unequal race for attention? A comparative case study between human and virtual influencers on social media

Joshua Enzig

Master (M.Sc.) in Marketing

Supervisor: Prof. Dr. João Guerreiro
ISCTE-IUL Business School,
Department of Marketing, Operation and general Management

October, 2022

Acknowledgement

First of all, I would like to thank my supervisor, Professor João Guerreiro, for his guidance. For helping me develop this topic, for his support throughout the process of this master thesis, his professional advice, encouraging words and patience.

I would also like to thank my family and friends for their support during this work. And I especially want to thank my girlfriend Sophia for moving to Lisbon with me for the duration of this master's program. I appreciate this commitment very much.

Furthermore, I would like to note that I am very grateful to be able to study in a united Europe where academic exchange is seamless across national borders and where I had the opportunity to broaden my academic horizons during the two years at the ISCTE Business School in Lisbon.

Finally, I would like to dedicate this Master's thesis to my father, who supported me in my decision to enroll in this Master's program and who, unfortunately, will not be able to see me graduate with this thesis.

Resumo

Com o número crescente de avatares virtuais como *influencers* em plataformas de redes sociais, há uma necessidade crescente de compreender melhor a relação para-social entre os seres humanos e os *influencers* virtuais. Esta dissertação avalia a forma como os utilizadores da Internet respondem aos diferentes níveis de desenvolvimento de avatares semelhantes a humanos no Instagram. O quadro teórico para este estudo deriva da teoria de *uncanny valley* de Mashario Mori (1970), que inicialmente foi concebida para descrever a interação entre humanos e robôs, sendo que nesta investigação é aplicada a personagens digitais em redes sociais

A metodologia deste trabalho baseia-se em *data mining* e segue o modelo CRIPS-DM para descobrir como os utilizadores das redes sociais respondem aos *influencers* virtuais. Para este efeito, três avatares virtuais com aparência humana diferente são comparados com um *influencers* humano do mesmo grupo de peers no Instagram. A comparação é baseada na avaliação dos comentários dos utilizadores da primeira metade de 2022 sobre o conteúdo publicado sobre temas nomeados. *Opinion mining* é utilizada para avaliar os comentários e calcular uma pontuação de sentimento para os *influencers*. Além disso, os respetivos níveis de *engagement* são recolhidos e analisados.

Os resultados deste artigo indicam que os utilizadores da Internet reagem mais positivamente aos *influencers* humanos e negativamente ao avatar mais parecido com o humano. Em conclusão, os resultados advêm do estado atual da revisão de literatura e as recomendações derivam do quão os avatares virtuais podem ser melhor integrados em plataformas de redes socias evitando efeitos indesejados.

Palavras-chave: data mining, engagement, sentiment analysis, social media, virtual avatar, virtual influencer, uncanny valley.

JEL Sistema de Classificação: Marketing (M31)

Abstract

With the increasing number of virtual avatars as influencers on social media platforms, there is a growing need to better understand the para-social relationship between humans and virtual influencers. This paper examines how internet users respond to different development levels of human-like avatars on Instagram. The theoretical framework for this study is derived from Mashario Mori's (1970) theory of the uncanny valley, which was initially designed to describe the interaction between humans and robots and is applied to digital characters in social networks in this research.

The methodology of this work is based on data mining and follows the CRIPS-DM model to find out how social media users respond to virtual influencers. For this purpose, three virtual avatars with different human-like appearance are compared with a human influencer from the same peer group on Instagram. The comparison is based on the evaluation of user comments from the first half of 2022 on the posted content of the named subjects. Opinion mining is used to evaluate the comments and calculate a sentiment score for the influencers. In addition, the respective engagement is collected and analysed.

The results of this paper indicate that internet users react most positively to human influencers and negatively to the most human-like avatar. The virtual influencers, which differ significantly from humans in their appearance, achieve average sentiment scores in comparison. In conclusion, the results are placed in the current state of the research literature and recommendations are derived on how virtual avatars can be better integrated into social networks to avoid uncanny effects.

Keywords: data mining, engagement, sentiment analysis, social media, virtual avatar, virtual influencer, uncanny valley.

JEL Classification: Marketing (M31)

Table of Content

List of Tables

List of Figures

Acronyms

AI: Artificial Intelligence

CGI: Computer Generated Image

CG: Computer Graphics

HI: Human Influencer

ML: Machine Learning

NLP: Natural Language Processing

SA: Sentiment Analysis

SC: Sentiment Score

SoM: Social Media

TM: Text Mining

UV: Uncanny Valley

VA: Virtual Avatars

VAI: Virtual Anime-like Influencer

VHI: Virtual Human-like Influencer

VI: Virtual Influencers

VSI: Virtual Sims-like Influencer

1 Introduction

In the age of digitalization, virtual agents have already established themselves in many areas, for example as human voices in telephone hotlines, as guiding holograms at airports, or as virtual assistants such as Amazon's Alexa on IoT devices or Apple's Siri on smartphones (Purington et al., 2017). In addition, however, there is another trend that has so far received little attention in scientific research and public discourse. Human-like avatar influencers who reach a large audience with their content on social media platforms such as Instagram, Facebook or TikTok (Arsenyan & Mirowska, 2021). One of these virtual influencers is *Lilmiquela*, an American teenager who emerged on Instagram in 2016 and is now followed by around 3 million people on the platform (Block & Lovegrove, 2021). But *Lilmiquela* is not the only one active on Instagram. The online archive VirtualHumans maintains a growing list of virtual avatars whose appearance is becoming increasingly human-like as digitalization advances (VirtualHumans.org, 2022). Referring to Astrid Hiort (2022), in 2015 a total of nine avatars were active on social media, which increased to over 200 virtual influencers in 2022. This new category of influencers has thus entered an industry that has expanded just as rapidly over the past few years. According to Statista, the global influencer marketing revenue has grown from $1.7 billion in 2016 to $16.4 billion in 2022 (Statista, 2022). This represents an increase of more than 800% in just 7 years. As a result, influencer marketing has become an essential part of digital marketing strategy as it provides a touch point to reach target audiences (Kim & Kim, 2021). Through the para-social connection between social media users and influencers, this marketing segment is evolving into long-term partnerships between brands and influencers (Kitterman, 2021). However, since this long-term relationship between humans and virtual avatars has not been sufficiently explored, and we cannot say exactly how humans respond to virtual artificial identities on social media, it is even more important to investigate and better understand this gap. Although the phenomenon of virtual avatars in social networks is relatively new, its origins in robotics research go back more than 50 years.

In the 1970s, Japanese robotics professor Mashario Mori proposed the uncanny valley theory, according to which humans tend to feel a sense of eeriness when robots become too human-like. (Mori, 1970). This theory was based on the professor's observations, without the hypothesis being scientifically and statistically proven. In the years that followed, little attention was paid to Mori's theory until the early 2000s, when

technological development and digitalization gained momentum. In 2003, studies have shown that humans react differently when communicating and interacting with robots than with humans. (Shechtman & Horowitz, 2003). Going a step further and transferring the theory from the analog to the digital world, MacDorman et al. (2009) applied the uncanny valley theory to digital robots such as avatars. In recent years, the phenomenon of uncanny valleys has also been repeatedly mentioned in connection with computer games, as technological progress allows developers to simulate reality increasingly better, leading to the observation that players of computer games also perceive the uncanny effect. (e.g., Geller, 2008; MacDorman, Green, Ho, & Koch, 2009). Much of the current research continues to focus on the appearance of robots and examines the effect of the uncanny valley on moving machines. In recent years, however, the theory put forward by Mori in 1970 has been proven in principle (Kätsyri et al., 2015; Mende et al., 2019).

If we now apply the concept of the uncanny valley to the phenomenon of virtual avatars on social media, we have the opportunity to find out more about internet users' perceptions towards avatars and statistically prove or disprove the existence of this concept in the context of social media through data mining and the interaction of thousands of people with these agents. To the best of our knowledge, Arsenyan and Mirowska (2021) are the only researchers who have studied the interaction between virtual influencers and humans on social platforms so far. However, the two researchers focused more on the posting behaviour of the subjects and based their study on fewer avatars. We, on the other hand, want to focus more on the users' reaction to further investigate their perception of the virtual influencers. With this, we want to contribute to a better understanding of the use of avatars in social networks. Be it for brands that enter into collaborations with virtual influencers, organizations that communicate via virtual influencers, or users who interact with virtual avatars.

Based on this introduction, the problem statement and current status of research, the following two research questions arise.

RQ1: Do social media users respond differently to virtual influencers than to human ones?

RQ2: Do social media users react differently to various developmental stages of human-like virtual avatars?

To answer the two questions above, we compare four influencers and analyse their posting behaviour and, in particular, the reactions they receive to their posts. Three of these four influencers belong to the category of virtual avatars and can be classified into different developmental stages of avatars with human-like appearance. The fourth study subject is a human influencer and serves as a reference.

This master thesis is divided into five main chapters. After the first chapter has already described the relevance of the topic, the problem statement, as well as the research questions, the second chapter focuses on the current state of academic literature. The literature review outlines the evolution from human influencers to virtual avatar influencers, followed by the concept of digital avatars and a theoretical consideration of opinion mining as well as a description of the concept of social media engagement. Finally, to better understand human-robot interaction, the uncanny valley theory is explained, which also serves as a theoretical framework for the development of the hypotheses. The third part of this paper presents the methodology developed on the CRISP-DM datamining model, which is used to investigate the previously stated hypotheses. In the fourth part, we statistically analyse the data using ANOVAs and Post-Hoc-Tests to find significant differences in the results and either confirm or reject our hypotheses. Furthermore, the results are presented and compared with the theory of the uncanny valley. In the fifth and final part, a conclusion is drawn and implications are discussed. In addition, the limitations of this study are identified and possible recommendations are made.

2 Literature Review

The literature review serves as the theoretical framework for this thesis. In the following chapter, the research is placed in its context, core concepts are discussed, and relevant research strands are identified. The literature review for this work was conducted through an online search of the three most popular journal databases, Scopus, Web of Science and Jstore. The main keywords selected were *Virtual Influencers*, *Virtual Avatars*, *Uncanny Valley*, *Social Media Influencers*, *Social Media Engagement*, *Text Mining* and *Sentiment Analysis*. To ensure the relevance of the papers cited in the literature review, only articles published in recognized scientific journals with a rating of at least quarter Q2 on the Scimago Journal & Country Rank were considered.

2.1 From Human to Artificial and Virtual Influencers

During the last decade, communication with social media influencers has become an important topic in strategic communication (Kim & Kim, 2021). As a result, many organizations have identified social media influencers as relevant intermediaries, primarily because they can provide access to hard-to-reach audiences, such as teenagers and young adults or specific interest groups (Enke & Brochers, 2019). Another aspect is that social media enables bidirectional communication, where recipients can participate in activities posted by the sender, in a bottom-up manner, as opposed to traditional media where communication is top-down. (Dijkmans, Kerkhof, Buyukcan-Tetik, & Beukeboom, 2015). Despite the growing importance of influencers in the literature, research has not paid particular attention to definitions of the influencer concept and is still seeking consensus. Therefore, we will use the most recent concept developed by Enke and Brochers (2019). From a strategic communication perspective, the authors define social media influencers "as third-party actors that have established a significant number of relevant relationships with a specific quality to influence on organizational stakeholders through content production, content distribution, interaction, and personal appearance on the social web" (Enke & Brochers, p. 8, 2019). Of course, social media influencers should be seen as trustworthy, credible and knowledgeable in the field of the advertised product to be persuasive (Djafarova & Rushworth, 2017). However, social media influencers can also have a negative impact on a brand's reputation. Khamis, Ang and Welling (2017) argue, that there is an inherent ethical problem with the idea of branding an individual. In addition, influencers can also have a negative reputation for

being overly demanding, which could scare off brands or certain user groups (Arsenyan & Mirowska, 2021). This goes hand in hand with the phenomenon that users are beginning to question the authenticity and credibility of influencers - especially those with a large audience and many followers (Kim & Kim, 2021).

A way out of this dilemma can be seen in the use of virtual avatars and influencers, which have the advantage that the company for which the digital avatar is influencing has complete control. With the American teenager *Miqueal Sousa*, in this thesis introduced as *Lilmiquela*, one of the first virtual CGI influencers has already built-up significant influence and socio-economic power, followed online by over 8 million users on various platforms (Block & Lovegrove, 2021). Aspects describing *Limiquelas* success can be found in users' need for distraction, according to Arsenyan and Mirowska (2021), as virtual influencers offer a glimpse into another world. The authors argue that the perception of a parasocial relationship, where the user feels a mutual sense of friendship, can apply not only to human influencers but also to artificial avatars. At the same time, Block and Lovegrove (2021) see another reason for the success of avatars like *Limiquela* in the uncanny duality, discordant polymedia narrative, and transmedia narrative worlds.

2.2 Digital Avatars on the Internet

For years, social media users have been accustomed to virtual avatars like bots and agents (Purington et al., 2017). Thus, fuelled by macroenvironmental aspects like advances in computing capabilities and the rapid development of technology, computer-generated images (CGI) are becoming more human-like, and virtual agents are increasingly capable of simulating human content (Arsenyan & Mirowska, 2021). Due to the rapid development of avatars in marketing, there is not yet a firm consensus on the exact definition. Academics have used several terms synonymously for avatars such as automated shopping assistants, chatbots, virtual customer service agents, or digital assistants (Miao, Kozlenkova, Wang, Xie, & Palmatier, 2021). According to Nowak and Fox (2018), this variance makes it difficult for researchers to compare and conceptualize empirical findings and draw conclusions from different studies. Therefore, Miao et al. (2021) developed a definition that incorporates all aspects and states that avatars are digital entities with anthropomorphic appearance that are controlled either by humans or by software and have the ability to interact.

More recently, Arsenyan and Mirowska (2021) have seen the emergence of virtual influencers as digital avatars designed to look human. Given that anthropomorphized representations of virtual agents, especially those that are visually appealing, elicit social responses and behavioral changes in people, these virtual influencers present a particularly interesting context (Khan & Sutcliffe, 2013). According to Hanus and Fox (2015), they cannot be considered digital avatars as they are not mere representations of the users and are not driven by user-oriented tasks. The authors state that these virtual influencers are presented similarly to human influencers, with their own public personas and stories that allow for greater interaction between users and influencers in the virtual environment (Hanus & Fox, 2015). Miao et. al. (2021) distinguishes four different categories of avatars: the first is simplistic, the second is superficial, the third is intelligent unrealistic and the fourth is a digital (or virtual) human avatar. This classification provides a basis for predicting the success or failure of digital avatars in business practice and will be used as a reference for this study.

2.3 CGI as a Basis for the Creation of Virtual Avatars

Modern computer graphics present us with the challenge of being able to distinguish whether the image of a person is real or digitally generated (MacDorman et al., 2009). However, because humans have learned over the course of evolution to recognize emotions in other faces and to intuitively understand each other through gestures and facial expressions, they are very sensitive and will notice any inconsistency in a computer-generated face (Lombardi et al., 2018). Therefore, correct proportions play a central role in the evaluation of a face, as they can influence whether we judge a person as attractive, dominant or - which is particularly important in the context of this work – healthy (Grammer & Thornhill, 1994).

According to Caudron, Nicq, and Valenza, (2016), one of the main elements in 3D modeling is a polygon mesh. A polygon is created from three vertices, and when these polygons are combined, a surface is created. This surface can be used to create virtual objects, shapes and sculptures. Additionally, texture can be given to this surface to make it look more realistic. MacDorman et al. (2009) demonstrated in a study that CGI faces, which generally look more human, have more photorealistic textures and more detail. The study also found that increasing the number of polygons also increased the human-like appearance of a virtual face. Thus, we can summarize that a high number of polygons is

suitable to make an avatar look more human-like. Taking this process further, Lombardi et al. (2018) assume that a deceptively human face of an avatar cannot be created by manually drawing polygons in a 3D program, but by collecting as many images of human heads as possible from different viewpoints, whereupon a computer-generated polygon mesh is created that is so detailed that we cannot distinguish the virtual avatar from a real human.

2.4 Opinion Mining and Sentiment Analysis

With the rise of social media platforms and the rapid expansion of the internet, social networks have become an indispensable means of communication. People around the world use forums, blogs and platforms such as Facebook, Twitter and Instagram to express their opinions publicly (Nandwani & Verma, 2021). As a result, a large amount of unstructured data is generated, referred to as big data, and sentiment analysis has been introduced to analyse this big data effectively (Nagamanjula & Pethalakshmi 2020). For researchers, analysts and marketers, these individual opinions are valuable, hence the need to analyse this public data from the web. The use of sentiment analysis is wide-ranging and can start with the analysis of product reviews, through attitudes to brands and on to political orientations (Mostafa, 2013). Liu and Zhang (2012) define sentiment analysis and opinion mining as a computer-based study of people's opinions, judgments, attitudes and emotions about entities, individuals, problems, events, issues and their characteristics. Based on this definition, Cortis and Davis (2021) add the social aspect and define “the study of user-generated content by a selective portion of society be it an individual or group, specifically those who express their opinion about a particular entity, individual, issue, event and/or topic via social media interaction” (p. 2) as social opinion mining. Conducting a sentiment analysis is about determining polarity and categorizing opinionated texts. They can be either positive or negative, but are not limited to this rough categorization and can be quantified, for example, on a Likert-scale (Prabowo & Thelwall, 2009). In this study, we will apply the sentiment analysis method using Phantombuser.com and MeaningCloud tools to collect and evaluate comments from social media users on Instagram pages of virtual- and human influencers.

2.5 Engagement on Social Media

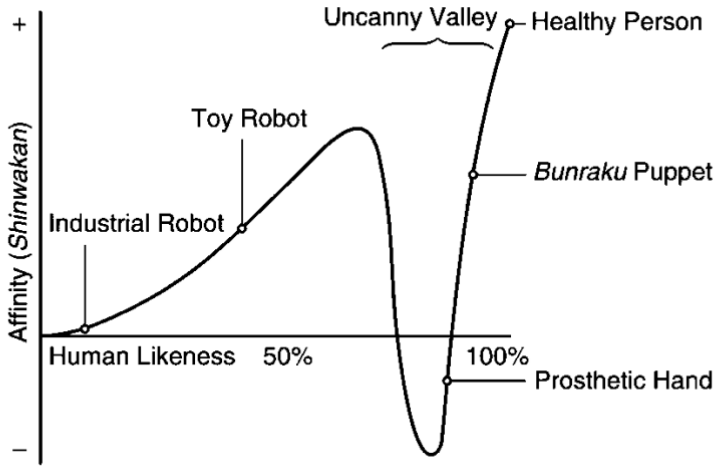
In modern society, interaction with influencers can help develop and present one's identity; this is especially true on highly visible social media platforms (Edwards et al., 2019). According to the authors, the social identity theory states that perceived membership in social groups is part of one's identity and enables the identification of clear in and out groups. With the advent of social media platforms, the role of customers has changed dramatically: They are no longer passive observers of content, but active participants who now become co-creators through their online behaviors (Lee, Hosanagar, & Nair, 2018). In the field of digital marketing, Van Doorn et. al (2010) conceptualized the concept of engagement. The authors define it as "customer's behavioral manifestations that have a brand or firm focus, beyond purchase" (p.254). The digital interactions of engagement behaviour can lead to varying degrees of intensity. In a recent study of engagement intensity, Dolan, Conduit, Fahy, Brodie, et al. (2016) define six groups of social media engagement behaviours on fan pages. Known as active engagement: *Create*, *Contribute*, and *Destroy* can be considered the first three, while *Consuming*, *Dormancy*, and *Detaching* are defined as passive forms of engagement. According to the authors, these six groups can be divided into two typologies. They can be either passive (low) or active (high) and also positive or negative. Preece, Nonnecke and Andrews (2004), define passive engagement as a behaviour of a user who browses an online platform and takes advantage of its benefits, but does not actively participate in community activities. In contrast, active engagement is determined by the behaviour of members who have a strong interest in engaging in an online community by participating in activities, creating messages, disseminating information, and providing emotional support (Casaló, Flavián, & Guinalíu, 2007). Different metrics are measuring the intensity of active or passive engagement on social media. Liking can be described as *affective response* and commenting as *active* and *public deliberation*, while reading content is an example of *passive engagement* (Alhabash et al., 2013). This research will focus on active participation to find out how users feel about the content presented by virtual avatars. Furthermore, since we do not have access to the social media accounts under study and thus can only evaluate the publicly available data, we will focus exclusively on comments and likes when calculating engagement scores and will not consider other metrics such as impressions and reach.

2.6 The Uncanny Valley

In the early 1970s, robotics professor Mashario Mori hypothesized that humanoid robots may have a disadvantage. According to his theory, a robot's perceived familiarity is directly proportional to its human-likeness. This continues until the robot is almost indistinguishable from a human in appearance and behaviour, which causes the emotional response of humans to quickly diminish and become strongly negative through feelings of unease and creepiness. A Phenomenon he called the uncanny valley (Mori, 1970; Mori, MacDorman & Kageki, 2012).

Figure 1

Relationship between affinity and human likeness



Source: Mori et al. (2012)

As can be seen in figure 1, Mori describes human perception in relation to different developmental stages from robots to humans along a series of examples and assigns them to a level of affinity. This starts with an industrial robot which humans are neither repelled nor attracted to, and continues to a toy robot which humans consider as cute. The uncanny feeling arises when a person perceives a discrepancy, as Mori (1970) describes it using an artificial hand. This hand meets our expectations in terms of function and aesthetics, but is cold and technical. After that, however, the curve of affinity rises again to the level of a healthy person, to whom we react very positively.

If we now compare a healthy human individual with a virtual avatar as described by Mori et al. (2012), this will lead us to our first hypothesis, which states that:

> ***H1**: Virtual avatar influencers have a lower sentiment score than human influencers from the same peer group on Instagram.*

Following the correlation between appearance and affinity described in figure 1, we can assume that the most human-like avatar in our study is expected to receive a more negative response from humans than avatars that are perceived as abstract in appearance. Therefore, we formulate our second hypothesis as follows:

> ***H2**: VHI has a lower sentiment score than VAI and VSI.*

Originally, the concept of *uncanny* was defined, among others, by the psychologist Sigmund Freud. According to his definition, the feeling of uncanny occurs when people perceive a difference between imagination and reality (Freud, 1919). Jochum and Goldberg (2016) add another aspect, explaining the uncanny as the Freudian way of representing "objects and other phenomena that evoke a strong psychological response of fear and fascination" (p. 149). If we assume that people express these feelings of fear and fascination on social media as well, we can formulate our third hypothesis, which states that:

> ***H3**: User engagement is higher with virtual influencers on Instagram than with human influencers from the same peer group.*

To trace the origins of this phenomenon, Lewkowicz and Ghazanfar (2012) suggest that the perception of the uncanny valley is not inborn, but is the result of a developmental phenomenon shaped by early childhood perception of human facial expressions. Studies have proven that the uncanny valley occurs in a variety of robots, animated characters, perceivers, and cultures, and there are a number of factors that contribute to these uneasy feelings, including low-level feature anomalies in humanoid robots, such as strange skin coloration and mismatched eyes (Geller, 2008; Ho & MacDorman, 2010; Yam, Bigman & Gray, 2021). Therefore, robots have a greater chance of finding sympathy and eliciting

empathy if they do not look too human, as humans tend to fear such perfection (Mori et al., 2012). Following this indication, we formulate our fourth hypothesis:

> ***H4**: Virtual avatars with a lower number of polygons have a more positive sentiment score on Instagram than avatars with more polygons.*

Yam et al. (2021) describe a another way out of this dilemma: we should remember that these human-looking robots are not robots with human feelings and that they lack the essence of a human mind. Although there is strong evidence for the existence of the uncanny valley (Kätsyri et al., 2015; Mende et al., 2019), Burleigh, Schoenherr and Lacroix (2013) state that the potential fear of the uncanny valley is unwarranted and could be further exacerbated by the dehumanisation of robots. According to the researchers, technological advances like motion capture, resolution and graphical textures increase the realism and simultaneously human likeness and emotional response. Other studies have proven that uncanny feelings are based on first impressions and decrease over time (Paetzel-Prüsmann, Perugia & Castellano, 2021). Therefore, the effect of the uncanny valley on humans could be neglected on long-term relationships between robots and humans. With all these different scientific assessments in recent years, Mori (2012) added a philosophical aspect to robotics research in general when he described the need to better grasp the uncanny valley in order to understand what makes us human. We carry this impulse forward by combining his theory of the uncanny valley with the current phenomena of digital avatars and virtual influencers.

3 Methodology

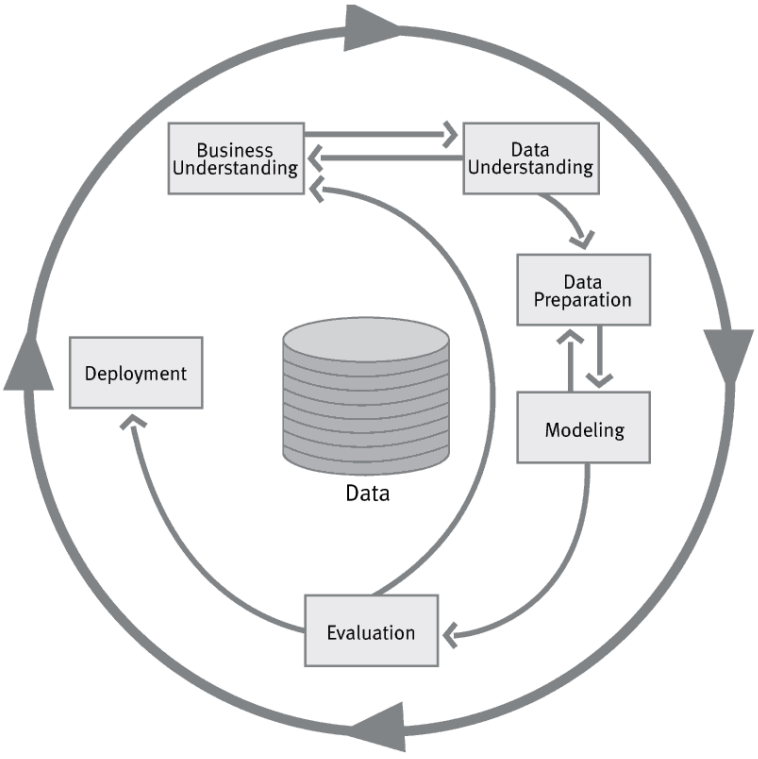
In the methodological part of this master's thesis, we first present the CRISP-DM model that guides the project's workflows. Based on this model, we then describe how the data is collected, processed and analysed.

3.1 CRISP-DM Introduction

With the increasing use of big data in research and industry, there is also a growing need for suitable methods to process and use this data in a meaningful way (Saltz & Krasteva, 2022). CRISP-DM is a methodology, which allows data mining projects to be processed in a planned, structured and standardized manner (Chapman et al., 2000). The model was developed in the 1990s and has been widely adopted since then. Recent surveys of data science teams show that CRISP-DM is still the most commonly used model. (Saltz & Hotz, 2020). Therefore, we also resort to it for this paper.

Figure 2

The CRISP-DM Methodology



Source: Chapman et. al., 2000

The model can be generally divided into six phases, as shown in figure 2, starting with business understanding, which includes identifying business goals and data mining objectives (Schäfer et al., 2019). In the next phase of data understanding, the data is collected, explored, and validated. This phase is followed by data preparation, in which the data is cleaned, converted if necessary, and integrated into the evaluation tool. This is followed by the modeling phase, in which a suitable modeling technique is selected and an assessing model is created. This is joined by the evolution phase, which serves to evaluate the results against the business objectives and is in our case integrated in the results and discussion section, followed by the deployment phase, which focuses on planning, monitoring, and maintaining the implementation. Normally, the workflow in CRIPS-DM proceeds linearly from one phase to the next. However, the model also allows the phases to be repeated and improved if required (Chapman et al., 2000). But this framework also has its limitations. For example, it does not provide precise guidance on the cases in which a phase should be repeated and at what point the results are sufficient and it is acceptable to move on to the next phase (Saltz & Krasteva, 2022). In this case, it is now our responsibility to correctly apply the CRISP-DM model in the following chapters, after which we should be able to test and answer our hypotheses.

3.2 Business Understanding

As briefly described in the previous part, the following section is about business understanding, where we outline the needs and goals of this project. According to Chapman et al. (2000), the success criteria need to be defined and the business and data mining requirements have to be discussed.
This data-mining project is being conducted with the goal:

1. To better understand the impact of using virtual avatars on Instagram and determine which human-like appearance leads to higher a sentiment score and engagement.

2. To make a recommendation for the appearance of virtual avatars based on our data and to avoid people reacting negatively to virtual influencers.

3.3 Data Understanding

According to the CRISP-DM model, this step is about determining where the data will come from and how it will be collected. Due to the visual content and the rapidly growing number of human and virtual influencers (Casaló et al., 2018), Instagram was chosen as the focal social media platform for data collection. Since influencers became known primarily through Instagram and the attractive visual representation of virtual avatars is an important aspect in human-agent interaction (Khan & Sutcliffe, 2014), this platform is well suited to analyze users' reactions. In 2020, Instagram users posted more than 80 million pictures every day, contributing to the network's rapid growth (Pagan, Mei, Li, & Dörfler, 2021). For this data-mining project, posts from influencers and a sample of all public comments on these posts (followers and non-followers) were collected. Because Instagram is a global social network, the languages in which users comment on content vary. Nevertheless, the language of the analysis will be English. Now that the platform has been determined, the subjects of the study must be defined.

To test the hypotheses, we refer to Mori's uncanny valley theory and want to prove that the decrease of likeness is related to the appearance of avatars. The scope ranges from the described toy robot to avatars that are as human-like as possible (Mori, 1970).

1. For the first stage, comparable to Mori's toy robot, we refer to our character as an anime-like virtual influencer and give it the encoding VAI (virtual anime influencer). The proportions of this avatar should be intentionally different from the proportions of a human body and face.

2. The second stage of the human-like avatar already has human proportions and a humanoid appearance, but is still clearly distinguishable from real humans. In addition, this character should have a rougher polygon mesh than the more developed avatar described in the third phase. The visual identity is most comparable to that of *The Sims*, a popular computer game from the 2000s. That's why we refer to this avatar as VSI (virtual sims-like influencer).

3. The third and last stage before a real human is only recognizable as artificial at second glance and is coded by us as VHI (virtual human-like influencer). When selecting this avatar, it is important that the skin already has human-like structures,

such as moles or freckles. The goal is to come as close as possible to a human in appearance.

4. The three aforementioned categories of virtual avatars are compared with a human influencer from the same peer group. Therefore, the specifications for the selection of this human are a similar age, comparable content and the same gender.

In addition to the external characteristics already described, the following requirements are added. Micro-Influencers with less than 100.000 followers are excluded due to their low reach. We also want to exclude all influencers who are considered celebrities. Therefore, macro influencers with more than 3 million followers are also expelled. In addition, the influencers should publish content several times a week on average, to which users respond not only with likes but also with comments. In order to assign an individual to each category, we had to get an overview of all virtual avatars on Instagram. For this purpose, an internet search was conducted and databases were searched (VirtualHumans.org, 2022).

1. VAI – To start with the virtual anime influencer, we wanted to make sure that people do not have any associations with the character, so it speaks for itself that all the well-known cartoon characters, such as those from Disney (e.g., Mini Mouse) and Mattel (Barbie) could not be considered. Nevertheless, the Influencer should be identified by the audience as distinctly non-human and yet be perceived more as developed anime and less as a two-dimensional cartoon character. That's why we have chosen *Noonoouri*. (https://www.instagram.com/noonoouri/)

2. VSI – Of all four influencers, the virtual sims influencer was the most difficult to define. Today, most virtual avatars are either graphically more advanced or too cartoony to fit into this category. In addition, there is the requirement for regular publication of posts and a sufficiently large number of followers. *Leyalovenature* meets all our criteria, even if her content revolves around spiritual themes, among others. (https://www.instagram.com/leyalovenature/)

3. VHI – With more than 3 million followers on Instagram, *Lilmiquela* was already part of a study examined by Arsenyan and Mirowska (2021). Her audience, stage

of development, and English content make her particularly suitable for scientific research. According to VirtualHumans.org, only *Lu from Magalu*, a Brazilian virtual influencer, has a larger audience with 6 million followers. However, *Lu from Magalu* has a commercial focus and her limited reach in Brazil makes the avatar unsuitable for this study. From the perspective on *Lilmiquela* as an average influencer focused on fashion, lifestyle, and music, she is best suited to be defined as our VHI. *Lilmiquela* is an avatar created by *Brud*, an animation studio based in Los Angeles (Rasmussen, 2021). (https://www.instagram.com/lilmiquela/)

4. HI – A variety of possibilities arise for the character of the human influencer. *Madison.Headrick* resembles our avatars the best by the content of her posts, the size of her audience and the number of comments she receives on her posts. Therefore, we define her as HI. Nevertheless, it should be mentioned here that there would have been many possibilities for the character of HI. (https://www.instagram.com/madison.headrick/)

An overview of the four defined characters can be seen in figure 3. The evolution of the characters from anime to human-looking avatars is also evident.

Figure 3

Images of the selected influencers

Noonoouri as VAI | *Leya as VSI* | *Lilmiquela as VHI* | *Madison as HI*

Source: Own elaboration

Once the subjects of the study have been defined, the next step is to determine the study period for this data-mining project. The period under study begins on the 1st of January 2022 and lasts until the 30th of June the same year. During this time, all public posts and

a sample of comments (followers and non-followers) by the named influencers were collected and analyzed. In late February 2022, Russia invaded Ukraine in violation of international law, an event that falls within our data collection period and received significant attention on social media. Three of our four influencers (*Noonoouri*, *Lilmiquela* and *Meandison.Headrick*) raised their voices and addressed the situation in their content with the call for peace. Nevertheless, it can be preempted at this point that no bias in the data could be detected due to this event.

After the study period and the study subjects were defined, the links to the postings were collected. This includes all interactions on a post: the number of likes, comments, views and the time at which the post was published. In a first step, we tried to automate this process using the commercial tool *Instagram Post Scraper* from the platform Phantombuster.com. However, since the tool did not fully collect the required data, we decided to perform this process manually.

As shown in table 4 below, the four influencers have between 300,000 and 3 million followers and published a total of 317 posts during the study period. These 317 posts were liked about 5.4 million times and commented on almost 50,000 times. As table 1 shows, the largest share of comments and likes goes to *Lilmiquela*, as she has the largest audience. Statistically, however, this will make no difference, as all groups are large enough to produce significant results.

Table 1

The selected influencers and their interactions during the data collection period

Source	Code	Followers	Posts	Comments	Likes
Noonoouri	VAI	406.212	102	8.714	745.102
Leyalovenature	VSI	338.468	81	6.339	934.010
Lilmiquela	VHI	3.043.508	61	29.413	3.256.521
Madison.Headrick	HI	538.680	73	4.825	458.274
Total		4.326.868	317	49.291	5.393.907

Source: Own elaboration

After all links to the corresponding posts were collected, sorted and listed in an Excel spreadsheet, the next step was to start extracting the comments. For this purpose, a commercial tool from Phantombuster.com was used again, but in this case the *Instagram*

Post Commenters Export. In a pretest, this tool proved to be very reliable and collected all the required data. This includes the username, the profile URL of the commenter, the comment itself, the time the comment was posted, the number of likes the comment has received from other users and the query. Due to limitations of the Instagram API (Constine, 2018), we were unable to capture all comments on every post. Thus, in order to get as many comments as possible, a new Instagram account (with the username: iscte_mma1_2022) was created in January 2022. At the time of data collection, the account was already six months old and had been used regularly in the meantime. This was done with the intention that Instagram would not immediately identify the account as a bot when scraping data later. In the next step, data were extracted and a sample of comments from VAI, VSI, VHI, and HI were collected in individual CSV documents. Additionally, we have limited the maximum number of comments to be downloaded per post to 300. To minimize spam in the data and increase the relevance of the content, we only collected comments that were uploaded up to seven days after the post was published. Through this process, we were able to data-mine a total of 24,089 comments. This corresponds to 48.89% of the total 49,291 comments posted by users during the study period.

3.4 Data Preparation

On the internet and especially on social media, users usually share their feelings, attitudes and emotions in an effortless way. As a result, the data obtained from posts, comments, and reviews on these platforms is very unstructured, making it difficult for machines to perform a sentiment analysis (Nandwani & Verma, 2021). Therefore, data-cleaning and pre-processing is a crucial step, as data quality significantly affects many approaches. We continue with this step of the CRISP-DM methodology, noting that, as expected, many users comment in different languages. Additionally, since the use of emojis is very common on social media, we quickly found in our dataset that in many cases no language was used at all, and instead attitudes such as affection, interest, or dislike were expressed exclusively through emojis. To solve this problem and decipher the feelings behind emojis, ‘emoji lexica’ can be used in sentiment analysis (Alswaidan & Menai, 2020). In the data collection process, we also gathered the profile URL in addition to the comments, timestamp, and query. This was done with the intention of filtering out duplicate comments from the same users among the influencers' posts. Users who were more likely

to have a negative opinion about one of the influencers tended to post that opinion in multiple comments under the same post. In order not to distort our statistical result, we have bundled the opinions of these users into a single comment from each user. Additionally, in this step of the data preparation, we cleaned the data by removing all external links and advertisements for products that have nothing to do with the influencer's post, as well as comments that are obviously unrelated to the content.

After cleaning the data, we proceed with sentiment analysis in the next step. In general, the analysis of feelings and emotions is very diverse and can be performed using different methods. Nandwani and Verma (2021) describe three types of sentiment and emotion analysis methods: The first is lexicon-based, the second is based on machine learning and the third is based on deep learning. Each has its own advantages and drawbacks. Common to all methods is the goal of determining whether the writer has a negative, positive, or natural attitude toward the annotated content (Nandwani &Verma, 2021). Due to practical and technical limitations, we focused on the lexicon-based method for this data mining process.

In this context, it should be mentioned that we are aware of the controversial discussion about the use of online accessible data (Moreno et al., 2013). The data we use for our study comes exclusively from public social media profiles and personal data is not stored beyond this project.

For this thesis, MeaningCloud was used to perform sentiment analysis (MeaningCloud, 2022). MeaningCloud is a software product that integrates with Python, PHP and Microsoft Excel. For the purposes of this project, we used the Microsoft Excel 365 Add-in to perform our sentiment analysis. For the analysis, the data from the CSV documents were transferred to an Excel spreadsheet and the users' comments were structured according to the corresponding influencer posts. MeaningCloud categorizes the comments into five levels. The sentiments can be either very positive (P+), positive (P), neutral (NEU), negative (N) or very negative (N+). If the comment is not analyzable, it will be marked as 'Error code 212'. This is the case when the comment is not in English or is grammatically incorrect. Additionally, it may be that an emoji was used that is not yet integrated into the database. Comments can additionally be classified as 'NONE' if the content does not represent a statement. As mentioned earlier, 24,098 of all 49,291 comments were extracted. As broken down in table 2, 11,874 of these comments were considered valid and classified into the 5-level sentiment classification, representing 24.09% of the absolute number of comments.

Table 2

Overview of the extracted comments

Code	N	Comments Analyzed		Valid Comments	
		n	%	n	%
VAI	8,714	4,409	50.6	2,536	29.1
VSI	6,339	2,770	43.7	1,998	31.52
VHI	29,413	13,119	44.6	4,880	16.59
HI	4,825	3,800	78.76	2,460	50.98
Total	49,291	24,098	48.89	11,874	24.09 %

Source: Own elaboration

With 2,460 comments, more than 50% of the total posted comments were evaluated for HI, while with 4,880 comments analyzed, just 16.59% of all comments were evaluated for VHI. The other two influencers had an almost identical percentage of comments evaluated, 29.1% for VAI and 31.52% for VSI. If we now go a step further, we see in table 3 the listing of users' interactions with each influencer, grouped by sentiment levels.

Table 3

Sentiment Analysis: Number of interactions per each Sentiment level

Source	P+		P		NEU		N		N+	
	n	%	n	%	n	%	n	%	n	%
VAI	318	12.54	2088	82.33	30	1.18	90	3.55	10	0.39
VSI	186	9.31	1735	86.84	17	0.85	53	2.65	7	0.35
VHI	433	8.87	3035	62.19	209	4.28	1024	20.98	179	3.67
HI	420	17.07	1974	80.24	38	1.54	26	1.06	2	0.08
Total	1357	11.95	8832	77.90	294	1.97	1193	7.06	198	1.12

Source: Own elaboration

The majority of all comments evaluated were positive, with 77.9% receiving a ‘P’ and 11.95% receiving an even stronger ‘P+’. Only 1.79% of comments were rated Neutral (NEU), while 7.06% were rated negative (N) and 1.12% were rated very negative (N+).

However, the data varies from influencer to influencer, as can be clearly seen. VHI achieves the lowest value with 62.19% positively rated comments, while VSI achieves the highest value with 86.84% in the same category. On the other hand, for comments rated as negative, VHI scores the highest at 20.98%, while HI receives the fewest comments rated 'N' in this case at 1.06%.

Table 4

Sentiment Analysis: Sentiment 5-level scale classification

Sentiment	Abbreviation	Value
Very positive	P+	2
Positive	P	1
Neutral	NEU	0
Negative	N	-1
Very Negative	N+	-2

Source: Own elaboration

These numbers can only be a first indication of a decrease in 'likeness' as the appearance of avatars becomes more human-like. To further explore the initial hypothesis, we have calculated a sentiment score for each influencer. To do this, the classifications of each post was multiplied by the values in table 4 and divided by the number of valid comments on the post. Through this process, a sentiment score was calculated for each of the 317 posts. In order to obtain the average sentiment score of the influencers for the study period, the values of the individual were summed up and divided by the number of posts.

Table 5

Sentiment Analysis: Sentiment Score per Influencer

Influencer	Nr. of Interactions	Sentiment Score
VAI	2,536	1.02
VSI	1,998	1.03
VHI	4,880	0.55
HI	2,460	1.14
Total	11,874	0.94

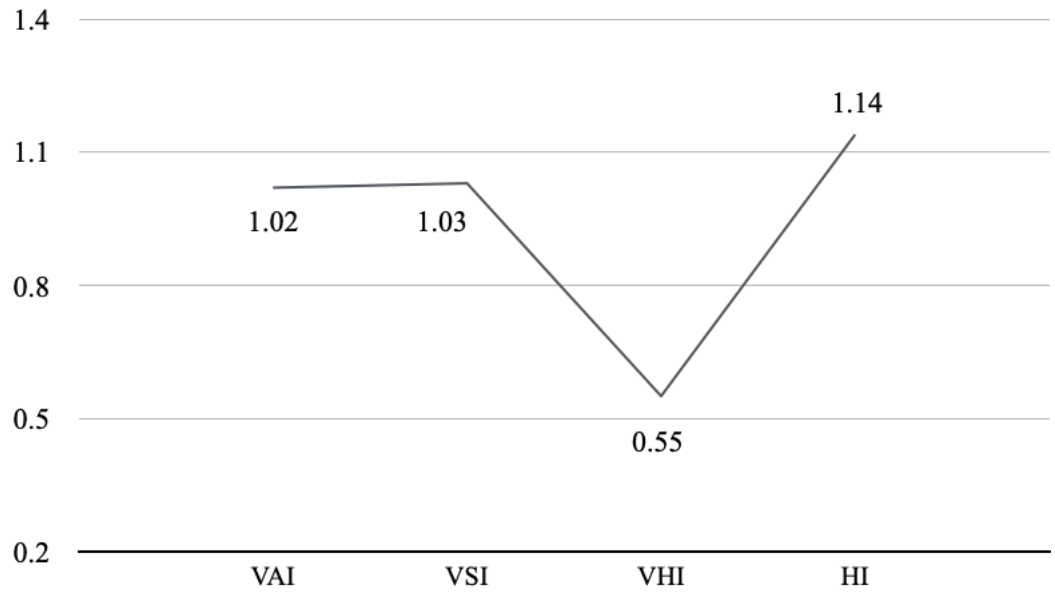
Source: Own elaboration

Recalling *H1*, which states that *Virtual avatar influencers have a lower sentiment score than human influencers from the same peer group on Instagram*, we can clearly say that VAI, VSI and VHI with an average sentiment value of 0.87 are below HI's score of 1.14. Therefore, we provisionally support hypothesis *H1*, but will investigate it further statistically in the fourth part of this paper. In addition, it can be observed that VHI has by far the lowest sentiment score among all three virtual influencers, which is an indication of a decreasing 'likeness' of users towards the character. Since our second hypothesis assumed that *VHI has a lower sentiment score than VAI and VSI*, we can also confirm *H2*. With a value of 1.14, *HI* is the best rated influencer, while VAI and VSI have an almost identical sentiment score of 1.02 and 1.03.

At this stage, however, we cannot say whether the differences between the study subjects are significant. For this, we will perform an ANOVA and Post-Hoc test in the modelling part. Nevertheless, a first visualization of the sentiment score is already possible, with the Y-axis describing the affinity of the recipients to the avatar, as in Mori et al. (2012), and the X-axis mapping the development of the avatars.

Figure 4

Mean Plot of Sentiment Score



Source: Own elaboration

To further investigate our *H3*, which states that *user engagement is higher with virtual influencers on Instagram than with human influencers from the same peer group*, we next focus on user engagement on the Instagram pages of VAI, VSI, VHI, and HI.

Online tools like Hootsuite.com make it quick and easy to track user engagement with influencers. However, since we only wanted to capture the average engagement during our study period, we chose a different approach. For this purpose, the comments and likes of a post were added and divided by the number of followers. This is how we determined user engagement per post. In a next step, the engagement of all individual posts was added up and divided by the total number of posts. In this way, we obtained the average engagement during our study period and did not bias our data by influences of before or after.

Table 6

Number of Interactions with Influencers

Code	Engagement	Followers	Posts	Comments	Likes
VAI	1.82 %	406,212	102	8,714	745,102
VSI	3.43 %	338,468	81	6,339	934,010
VHI	1.83 %	3.043,508	61	29,413	3,256,521
HI	1.19 %	538,680	73	4,825	458,274

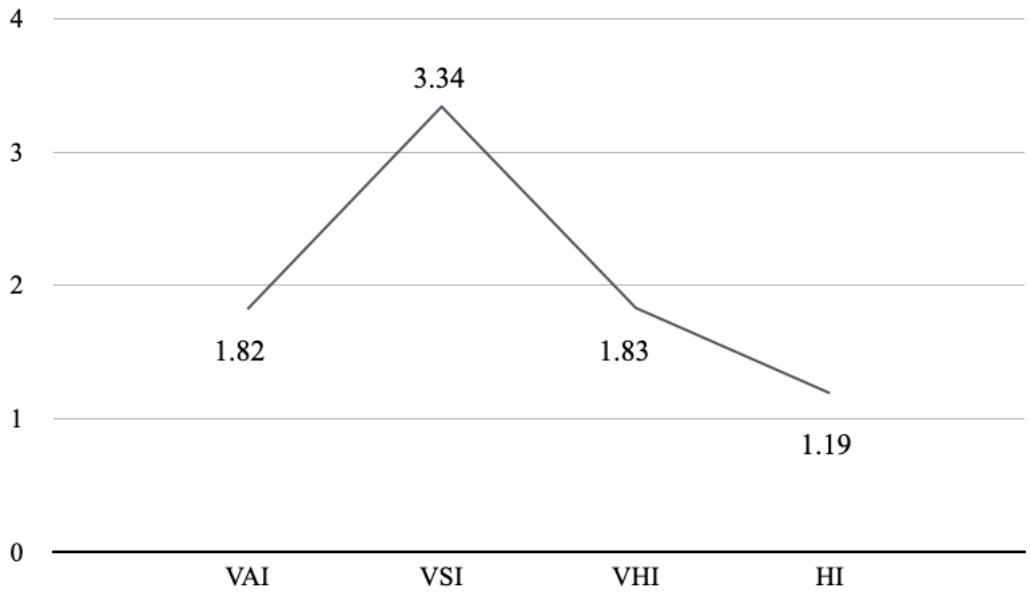
Source: Own elaboration

As table 6 reveals, engagement varies from 1.19% of HI with the lowest value to 3.43% of VSI with the highest value. In contrast, this time VAI and VHI are in the midfield with an almost identical engagement score. Immediately noticeable is HI's low value on engagement, after our human influencer scored the highest on sentiment in the previous step. This correlation suggests that HI's followers are less active compared to the virtual influencers' followers, but are mostly positive about her account and content. So, HI seems to be less polarizing and her audience has a more unified opinion of her, compare to the other subjects of the study. As we have not calculated an ANOVA with the engagement means at this point, we cannot yet say whether the differences are significant.

If we now visualize the mean engagement scores, as shown in figure 6, it is already confirmed that we have a different picture compared to the sentiment score and to Mori's hypothesis about the decrease of affinity of humans towards avatars whose appearance becomes increasingly human-like.

Figure 5

Mean Plot of Engagement



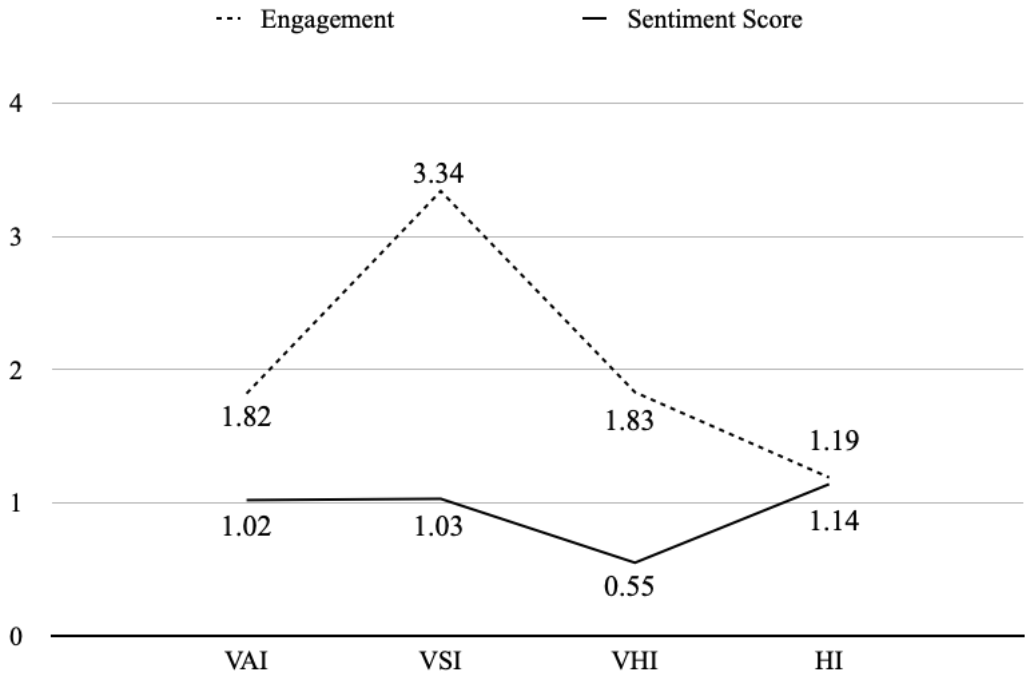
Source: Own elaboration

Now that we have obtained and visualized the engagement of the study subjects, it remains to be noted that VAI, VSI, and VHI score higher than HI. This observation leads us to our third hypothesis, which states that *user engagement is higher with virtual influencers on Instagram than with human influencers from the same peer group.* Therefore, we can confirm this hypothesis. One possible explanation could be that the virtual avatars are more polarizing and therefore have lower sentiment scores but higher engagement. However, to further verify this theory, the next step is to investigate whether the differences are actually significant and therefore allow this conclusion.

To complete the data preparation, Figure 6 shows the mean graph of engagement and sentiment scores. Graphically, the difference between the values of VSI is immediately noticeable. But VHI also stands out, although not quite as strongly. This is in great contrast to HI, where the two values of sentiment score and engagement are almost identical.

Figure 6

Mean Plot of Engagement and Sentiment Score



Source: Own elaboration

3.5 Modeling

Having now prepared the data in the excel spreadsheet described above and obtained initial results in terms of the sentiment score and engagement of the individual subjects, the next step is to check whether the differences found can be confirmed. Considering the goal of this data-mining project, we want to prove that there is a significant difference in the reaction of people to different appearances of human-like avatars. As we have already seen in figure 3, the sentiment scores of VAI, VSI, VHI and HI vary. VHI in particular stands out with the lowest value, which would support Mori's assumption that near-perfect human avatars evoke an uncanny feeling in people (Mori, 2012). Since we are comparing more than two means, we must omit a T-test and perform an ANOVA to determine if the differences are significant. For the calculations we use SPSS 27 by IBM. For the One-Way ANOVA, we employ sentiment score/engagement as the dependent variable and the influencers as factors. However, since the calculation of an ANOVA can only generally determine whether there are significant differences between the study participants, we will subsequently perform a Post-Hoc test for the sentiment score as well as for the engagement in order to precisely determine the significant differences.

4 Results and Discussion

With the data successfully prepared and already descriptively analysed, we continue with the data mining process in the following part, focusing on the analysis, results, and discussion. Of course, we still follow the workflow of the CRIPS-DM model, but for a better structure of the study we integrate the evaluation and deployment part into the fourth section of the thesis.

4.1 Evaluation

With the help of SPSS, One-way ANOVAs and Bonferroni Post-Hoc tests were conducted for the sentiment score and engagement of the four influencers. The results are described and analyzed in the following part.

Table 7 shows that the standard deviation of the sentiment score and the mean are close to each other. Most important, however, is our significance value, which is 0.000, clearly below the error level of 0.05. By calculating a One-Way ANOVA, we cannot yet say which values are significantly different, we can only say that there is a difference. Thus, this result is already very pleasing for us and confirms a statistical difference in the sentiment values of the subjects.

Table 7

ANOVA with Sentiment Score

	Sum of Square	df	Mean Square	F	Sig.
Between Groups	13.33	3	4.45	133.93	0.000
Within Groups	10.39	313	0.03		
Total	23.72	316			

Source: Own elaboration

Still, since we cannot yet tell between which influencers the significant differences in sentiment scores occur, we perform a Bonferroni Post-Hoc test in the next step. Again, we use the sentiment score as the dependent variable and obtain the following results, which are reported in table 8.

Table 8

Bonferroni Post-Hoc Test with Sentiment Score as Dependent Variable

					95% Confidence Interval	
		Mean Diff.	Std. Error	Sig.	Lower Bond	Upper Bond
VAI	VSI	-0.00972	0.02713	1.000	-0.0181	0.0623
	VHI	0.46718*	0.02951	0.000	0.3888	0.5455
	HI	-0.12232*	0.02795	0.000	-0.1965	-0.0481
VSI	VAI	0.00972	0.02713	1.000	-0.0623	0.0818
	VHI	0.47690*	0.03090	0.000	0.3949	0.5590
	HI	-0.11260*	0.02942	0.001	-0.1907	-0.0345
VHI	VAI	-0.46718*	0.02951	0.000	-0.5455	-0.3888
	VSI	-0.47690*	0.03090	0.000	-0.5590	-0.3949
	HI	-0.58951*	0.03162	0.000	-0.6735	-0.5055
HI	VAI	0.12232*	0.02795	0.000	0.0481	0.1965
	VSI	0.11260*	0.02942	0.001	0.0345	0.1907
	VHI	0.58951*	0.03162	0.000	0.5055	0.6735

**The mean difference is significant at the 0.05 level.*

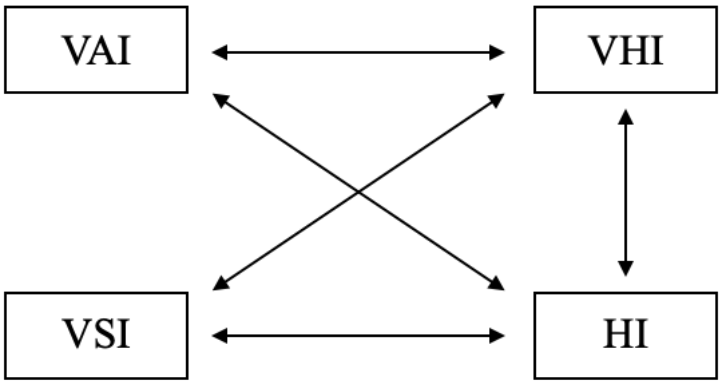
Source: Own elaboration

At 1.02 and 1.03, the average sentiment score of VAI and VSI are very close to each other. The Bonferroni test also shows no significant differences in this case, as would be expected. However, and more importantly, VHI and HI show significant differences. These two influencers are significant different from all others in terms of the mean of the sentiment score. With the lowest average sentiment score during the study period, VHI also shows the highest average mean differences to the other influencers.

All statistically significant values are marked with an asterisk in table 8 and to further illustrate the results of the Bonferroni test, we see in figure 8 all the influencers between which significant differences in sentiment score exist.

Figure 7

Illustration of Significant differences between Means of Sentiment Score



Source: Own elaboration

As can also be seen in the visualization of figure 8, the only two subjects between which a significant difference in sentiment score does not exist are VAI and VSI. With this result, we can statistically support our first hypothesis, which assumes that: *Virtual avatar influencers have a lower sentiment score than human influencers from the same peer group on Instagram*. After conducting ANOVA and Bonferroni to demonstrate significant differences in sentiment scores between *Noonoouri*, *Leyalovenature*, *Lilmiquela*, and *Madison.Headrick*, the same tests were performed on the engagement score.

Table 9

ANOVA with Engagement

	Sum of Square	df	Mean Square	F	Sig.
Between Groups	214.71	3	71.57	3.27	0.022
Within Groups	6792.75	310	21.91		
Total	7007.46	313			

Source: Own elaboration

As shown in table 9, the significance level is 0.022, which is lower than the alpha error value of 0.05. Therefore, we can again assume that there are significant differences between the means in terms of engagement among the four study subjects. To determine these differences more precisely, we performed the Bonferroni test.

Table 10

Bonferroni Post-Hoc Test with Engagement as Dependent Variable

					95% Confidence Interval	
		Mean Diff.	Std. Error	Sig.	Lower Bond	Upper Bond
VAI	VSI	-1.61061	1.01289	0.514	-4.3394	1.1182
	VHI	-0.01034	0.16756	1.000	-0.4631	0.4424
	HI	-0.62542*	0.11056	0.000	0.3297	0.9211
VSI	VAI	1.61061	1.01289	0.514	-1.1182	4.3394
	VHI	1.60026	1.02377	0.534	-1.1552	4.3557
	HI	2.23603	1.01600	0.168	-0.5004	4.9725
VHI	VAI	0.01034	0.16756	1.000	-0.4424	0.4631
	VSI	-1.60026	1.02377	0.534	-4.3557	1.1552
	HI	0.63577*	0.18543	0.005	-0.1383	1.1332
HI	VAI	-0.62542*	0.11056	0.000	-0.9211	-0.3297
	VSI	-2.23603	1.01600	0.168	-4.9725	0.5004
	VHI	-0.63577*	0.18543	0.005	-1.1332	-0.1383

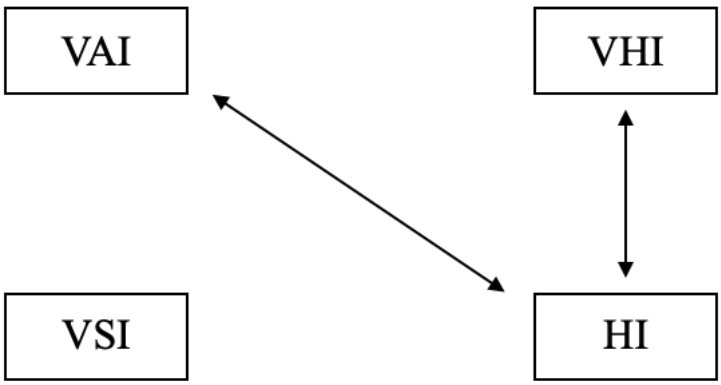
**The mean difference is significant at the 0.05 level.*

Source: Own elaboration

In table 10, we see much less significant differences compared with the sentiment score. Only VAI and HI, as well as VHI and HI are below the alpha error level of 0.05.

Figure 8

Illustration of Significant differences between Means of Engagement



Source: Own elaboration

To illustrate the significant differences between subjects in this case as well, the corresponding relations are shown in figure 9. The only two significant differences in engagement scores exist between VHI and HI, as well as between VAI and HI.

Aside from HI having the lowest engagement score, the ANOVA and Bonferroni test results tell us relatively little in terms of our data-mining objectives, research questions, and hypotheses. Therefore, we go further and try to understand where these differences come from. Table 1 had shown us that VSI received more than twice as many likes during the study period, even though the Influencer had 200,000 fewer followers than HI. The reasons for this can be numerous, but to better understand the phenomenon, we split engagement into likes and comments and look at both categories individually. To make the values comparable, we scale the number of comments and likes the influencers received on average per 1,000 followers on one of their posts. As table 11 indicates, it is again striking that VSI received the most comments and even more likes in relation to followers. However, we also see that the differences are no longer as big as in the previous case, especially in the comments category. Since we consider the comments to be more important and more meaningful in terms of substance, also with regard to the sentiment score, we will focus on them in the following.

Table 11

Engagement Metrics: Avg. Comments and Likes per Post and 1,000 Followers

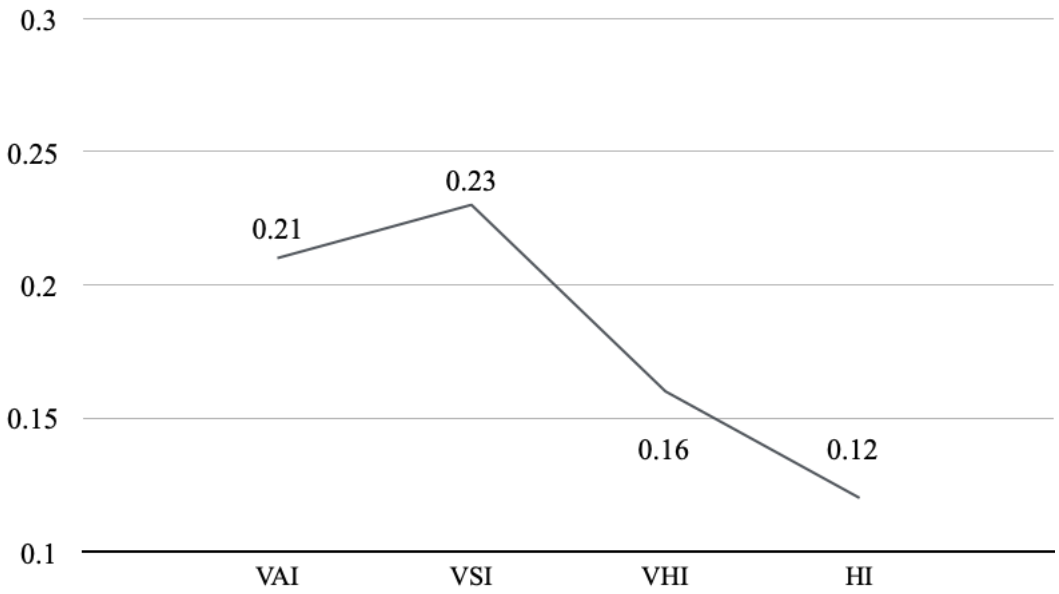
	VAI	VSI	VHI	HI
Avg. Comments per Post (scaled)	0.21	0.23	0.16	0.12
Avg. Likes per Post (scaled)	17.98	34.07	18.14	11.81

Source: Own elaboration

By working with these averages, we can see that the values of the scaled comments are now closer together. The graphical representation in figure 10 shows in this case a similar picture to the one previously shown in figure 4 of the sentiment score. The number of scaled comments per post increases between VAI and VSI before decreasing sharply at VHI. However, one major difference can be seen: In contrast to the sentiment score, engagement between VHI and HI continues to decline.

Figure 9

Mean Plot of Comments (scaled per 1,000 Followers)



Source: Own elaboration

The reasons why HI has the highest sentiment score and the lowest engagement must be examined separately. More importantly here, VHI scores the lowest value with 0.16 comments per 1,000 followers and post among the three virtual avatars. Since we now want to re-test whether the differences in means are significant, we again performed a One-Way ANOVA.

Table 12

ANOVA with scaled Engagement of Comments

	Sum of Square	df	Mean Square	F	Sig.
Between Groups	0.55	3	0.19	5.92	0.001
Within Groups	9.7	311	0.03		
Total	10.252	314			

Source: Own elaboration

This time, we can already say that with a significance of 0.001 the alpha error level is considerably lower. To determine the exact differences between the subjects, we conducted the Bonferroni Post-Hoc test.

Table 13

Bonferroni Post-Hoc Test with scaled Engagement of Comments as Dependent Variable

					95% Confidence Interval	
		Mean Diff.	Std. Error	Sig.	Lower Bond	Upper Bond
VAI	VSI	-0.02090	0.02628	1.000	-0.0907	0.0489
	VHI	0.04895	0.02888	0.547	-0.0277	0.1256
	HI	0.8761*	0.02707	0.008	0.157	0.1595
VSI	VAI	0.02090	0.02628	1.000	-0.0489	0.0907
	VHI	0.06986	0.03022	0.129	-0.0104	0.1501
	HI	0.10852*	0.02850	0.001	0.0328	0.1842
VHI	VAI	-0.04895	0.02888	0.547	-0.1256	0.0277
	VSI	-0.06986	0.03022	0.129	-0.1501	0.0104
	HI	0.03866	0.03091	1.000	-0.0434	0.1207
HI	VAI	-0.08761*	0.02707	0.008	-0.1595	-0.0157
	VSI	-0.10852*	0.02850	0.001	-0.1842	-0.0328
	VHI	-0.03866	0.03091	1.000	-0.1207	0.0434

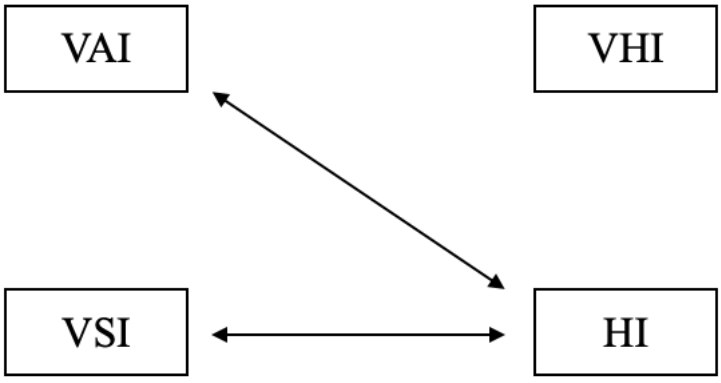
**The mean difference is significant at the 0.05 level.*

Source: Own elaboration

Unlike with the regular engagement, where we have integrated likes in addition to comments, there are significant differences between VAI and HI and between VSI and HI in this case.

Figure 10

Illustration of Significant differences between scaled Comments per Post



Source: Own elaboration

And even though VHI achieves the lowest count of comments per post compared to all three avatars, it remains to be noted at this point that the differences are not sufficient to produce a significant result.

4.2 Deployment

The goal of this thesis was to better understand the impact of virtual avatars on users in social networks and to provide recommendations for the appearance of virtual avatars to prevent users from reacting negatively to them. Now that we have performed a sentiment analysis, we can see that the uncanny valley effect described by Mori (1970) occurs in the selection of our virtual avatars. Instagram users reacted most negatively to VHI compared to the other avatars, suggesting that people on social media are skeptical of this realistic human-like appearance.

Since this data mining project was also conducted with the aim of making a recommendation for the appearance of virtual avatars, we can conclude that a lower number of polygons increases the sentiment score of the avatar. As shown in figure 3, VSI has the roughest textures and the edgiest face, which leads us to assume that her developers must have worked with a lower polygon mesh. Since VSI also achieved the highest sentiment score of 1.03 among the three virtual avatars, we can finally confirm the fourth hypothesis, which assumes that *Virtual avatars with a lower number of polygons have a more positive sentiment score on Instagram than avatars with more polygons.* From this, we can deduce that when using virtual avatars on Instagram, it is recommended to use a coarser polygon mesh and not make the textures look too human-like.

4.3 Results and Discussion

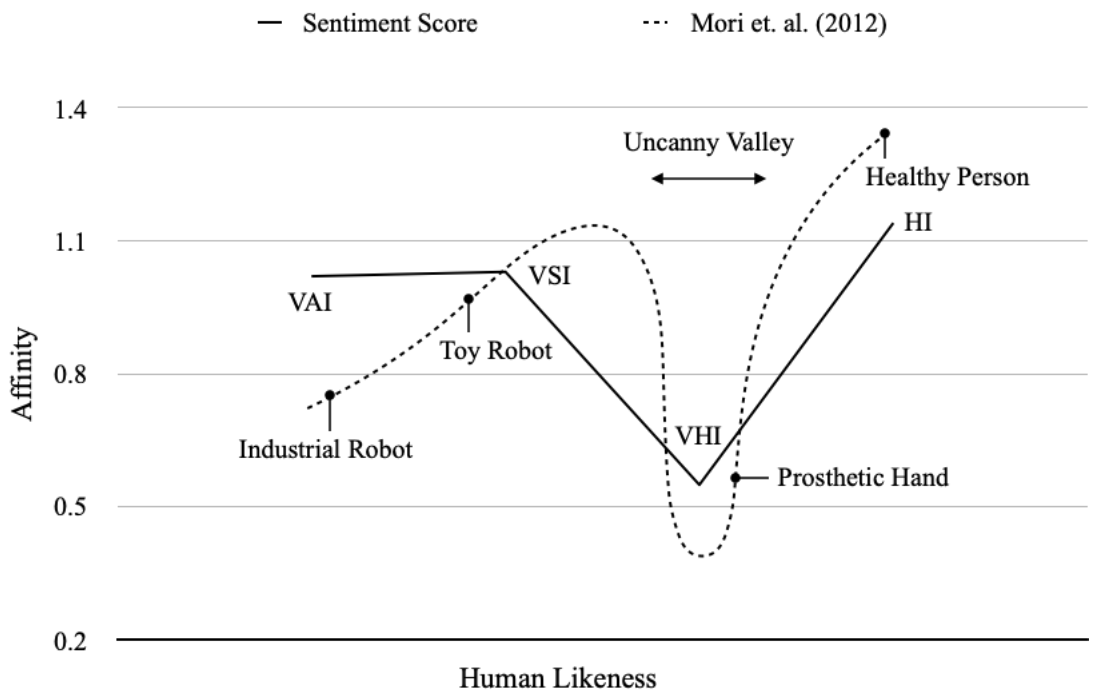
Now that we have completed our data mining process with the CRISP-DM model and evaluated the main results, we underscore these findings with some qualitative insights in this section. In order to prove the reasons for *Lilmiquela's* poorer performance, not only in the sentiment score, it helps to take a close look at the data set and examine the phenomenon with some examples. On Instagram, Mori's theory seems to be familiar to some users. Under VHI first post in the study period, published on the 1st of January

this year, with her boyfriend watching New Year's fireworks, a user comments on the post as follows: *"You are the uncanny effect"*. Another user writes under the same post *"This account gives me anxiety"* followed by a crying emoji and liked by 19 other users. The description of VHI as *"uncanny"* occurs a total of six times in *Lilmiquela's* data set. In the data set of the other two virtual influencers VAI and VSI, on the other hand, the term does not occur even once. *Lilmiquela* is called *"ugly"* a total of 30 times, and this term also does not appear in the records of the two other avatars. In addition, users do not seem to know exactly who they are interacting with. For example, the question *"Are you a robot?"* appears a total of 40 times in *Lilmiquela's* data, and that alone in English and in the correct spelling. Now, one could argue that *Lilmiquela* has already stated this fact in her Instagram bio, where she describes herself as a *"19-year-old robot living in LA"*. However, most of her 3 million users will see VHI's content in their feed, where this information is missing. Generally speaking, we could say that *Lilmiquela*, with a larger number of followers, is also exposed to a greater potential of hate speech, however, this is countered by the fact that VHI does not have a higher level of engagement than the other research subjects. Therefore, in conclusion, this scepticism expressed by Instagram users under VHI's profile aligns seamlessly with the uncanny valley hypothesis.

In his first publication in 1970, Mori describes his theory as a relationship between the human resemblance of a being and the affinity of the observer to that being (Mori, 1970). The curve illustrates that the affinity rises steeply at first and tips at the point where robots or avatars are only minimally distinguishable from a human in appearance and behaviour, causing people's emotional response to quickly decrease and become strongly negative due to feelings of creepiness. This slight degree of imperfection causes us to be sceptical. To conclude our research question and relate it to Mori's illustration (Mori, 1970; Mori, MacDorman & Kageki, 2012), we next overlay the two diagrams (as shown in figure 12). Since our character VAI is not an industrial robot compared to Mori's first subject of comparison, but looks more like an anime and thus much more human, it is not surprising that the degree of affinity in our study is already higher. Furthermore, we can see that VAI and VSI are much closer in terms of external appearance than a machine and a toy robot. Therefore, it is not surprising that the slope in our graph is much lower than in Mori's visualization. Nevertheless, both graphs show a positive slope at this point, which allows us to confirm Moris hypothesis up to this point.

Figure 11

Mean Plot of Sentiment Score and Mori's Graph of the Uncanny Valley (1970)



Source: Own elaboration

As the minimum of the curve, Mori defines an artificial hand. According to his publication, this is due to the fact that the hand looks real, but is cold and technical. This discrepancy leads to the perception of the uncanny valley. In our case, the valley corresponds to VHI. Brud, the company behind *Lilimiquela*, tries to make the appearance as human-like as possible. However, internet users seem to perceive a discrepancy here as well. In general, we can see that our sentiment score results are consistent with Mori's 1970 hypothesis and graph. The slopes of the curve are lower in our case and our results swing less because the appearance of our study participants is closer together.

Our results are thus consistent with the studies of MacDorman (2010), Arsenyan and Mirowska (2021), and Yam, Bigman, and Gray (2021), who also observed the uncanny valley phenomenon in the context of human-like digital characters. While Paetzel-Prüsmann, Perugia, and Castellano (2021) assume that this effect does not play a role in long-term robot-human relationships, our results show the opposite. *Lilmiquela* first appeared on Instagram in April 2016, so she had already been active on the platform for six years at the time of our data collection. Nevertheless, this period was not sufficient

to resolve the habituation effect described by Paetzel-Prüsmann et al (2021). Also, we cannot support with our results the theory of Burleigh, Schoenherr and Lacroix (2013), which state that the potential fear of the uncanny valley is unwarranted and could be further exacerbated by the dehumanisation of robots. Following the authors' argument, technological advances in resolution and graphic textures increase realism while simultaneously increasing human likeness and emotional response. We could not prove this relationship with our fourth hypothesis, which assumed that *Virtual avatars with a lower number of polygons have a more positive sentiment score on Instagram than avatars with more polygons*. Nevertheless, it cannot be ruled out that with further technical development in the coming years, an effect like that described by Burleigh et al. (2013) may still occur.

In conclusion, if we consider Jochum and Goldberg's (2016) definition of uncanny objects as objects that evoke a strong psychological response of fear and fascination, and relate this description to the findings of our study subjects, we can postulate that: Virtual avatars polarize more than human influencers due to their human-like appearance and therefore have higher engagement but simultaneously lower approval ratings by their audience.

5 Conclusion

The objective of this study was to further explore the uncanny valley theory in the context of advanced technological capabilities and social networks. In addition, this thesis extends the current state of literature on Mori's hypothesis (Thepsoonthorn, Ogawa, & Miyake, 2021; Arsenyan, & Mirowska, 2021) and goes a step further by examining avatars in social media. We recall that Mori describes a robot's perceived familiarity as directly proportional to its human-likeness. This continues until the robot is barely distinguishable from a human in appearance and behaviour, causing the emotional response of humans to quickly diminish and become strongly negative due to feelings of discomfort and creepiness (Mori, MacDorman & Kageki, 2012). And although Mori's Uncanny Valley hypothesis dates back to the 1970s, we can confidently say after conducting this study that humans still exhibit this described reaction to human-like avatars today.

5.1 Theoretical and marginal contributions

Now that we have proven the existence of the uncanny valley in the context of virtual avatars in social media, the question of how to prevent these figures from being perceived as uncanny by the audience must be addressed. The answer to this problem is not only relevant to academia, but also offers great potential for brands, organizations, and companies that already use virtual avatars as testimonials, brand ambassadors, or advertisers on social media and want to improve their performance. (e.g., Travers, 2020; Griffiths, 2018).

Overall, two approaches for solving the dilemma can be concluded from our results. The first approach is relatively evident and has already been mentioned by Mori et al. (2012) and other authors (Yam, Bigman, & Gray, 2021) in a general perspective on the uncanny valley. If we look at the different mean values of the sentiment scores, the avatars with less human-like appearance achieve a higher rating compared to the more human-like avatars. The obvious solution to the problem is to dehumanize the avatars in their appearance and communication so that people respond to them more positively. Like in the case of VAI, it may be sufficient not to adhere to human proportions in developing the character. As a result, the avatar is perceived as childlike and cute due to its disproportionately large eyes and triggers people's affection rather than creepiness. Or, as in the case of VSI, sticking to the proportions but being sufficiently different from humans

in the texture of the character's skin. This way, the avatar is clearly distinguishable from us humans and does not give us the feeling of pretending something, which then leads to scepticism and mistrust.

However, this first approach to avoid uncanny effects by dehumanizing the avatars raises the question of how realistic the implementation will be. It can undoubtedly be assumed that technological progress in the next few years will lead to avatars looking even more human-like, thus increasing their possible applications. It would therefore be unreasonable for companies and advertisers to forego the opportunities that present themselves. Taking this into account, we would like to propose another way to minimize the effect of the uncanny valley without losing the benefits of human-like avatars in social networks. To the best of our knowledge, this approach has never been discussed in the scientific literature before, which makes it all the more relevant to consider and further investigate this concept.

In the case of *Lilmiquela,* social media users were not always aware that the character is not a real person. We were able to prove that many users had doubts about the avatar's identity or were generally confused about its existence. Since *Noonoouri* and *Leyalovenature* could clearly be perceived as anime and simpler CG avatar and there was no confusion about their identities, users gave these characters better sentiment ratings. A labelling obligation for realistic, human-like avatars in social media and clear communication could eliminate this misunderstanding and lead to a fundamentally more positive approach to virtual avatars like *Lilmiquela*. With greater transparency, this approach would benefit not only avatars and advertisers, but also social media users. One possible consequence of visual platforms such as Instagram, which promotes filtered and perfected lifestyles, can be social comparisons and resulting feelings of anxiety about not achieving the same level of perfection as the influencers (Djafarova and Rushworth, 2017). In particular, this also relates to virtual characters, where it is much easier for creators to stage this perfectionism. If users are now made aware by a labelling requirement that the figure depicted is not a real person, they may find it easier not to have to compare themselves with the virtual character. But it is not only the appearance of the avatar that plays a decisive role in how users perceive it. In order to build additional trust between virtual avatars and people, it could help to also make clear and transparent which person, company or organization stands behind a virtual avatar in social media. This could help social media users to better evaluate the messages of avatars and place them in a larger context. It is up to the various platforms to enforce these rules, but they

should be independently and critically reviewed and sanctioned by authorities in case of violations.

5.2 Limitations and future research

One of the biggest challenges in this work was the limited number of active virtual avatars on social media. In databases such as VirtualHumans.org, the number of virtual influencers has increased in recent years, but the selection is still limited. In addition, most virtual avatars are female, so a representative study of male or diverse avatars in terms of our hypotheses has not yet been possible. If the number of virtual avatars on social media continues to grow, it remains to be noted that a replication of this study based on a broader and more diverse selection of virtual influencers is recommended. In addition, we have to assume that Instagram is primarily used by a younger target group. According to Dixon (2022), 53.2% of all Instagram users in the United States were between 18 and 34 years old in June 2022, the last month of our data collection. The 45+ age group has a share of only 21.8%. Therefore, the results of this study cannot be automatically generalized to other demographic groups. In addition, Western social networks like Instagram are not freely accessible to citizens in countries like China and Russia.

In addition, we can assume that technological progress will lead to CGI characters looking even more authentic and human-like in the future. Under this assumption, it remains relevant to investigate whether the hypotheses of this study continue to hold true, or if humans as a consequence evaluate more advanced avatars than *Lilmiquela* identical to human influencers such as *Madison.Headrick*. We cannot say today whether Mori's hypothesis will still be true 50 years from now. What we can say, however, is that it is still important to look at this issue, and it will be the task of future research to show whether the theory of the *Uncanny Valley* will continue to hold up in the course of digital and technological progress.

6 Bibliography

Alswaidan, N., & Menai, M. E. (2020). A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, *62*(8), 2937-2987. https://doi.org/10.1007/s10115-020-01449-0

Alhabash, S., McAlister, A. R., Hagerstrom, A., Quilliam, E. T., Rifon, N. J., & Richards, J. I. (2013). Between likes and shares: Effects of emotional appeal and Virality on the persuasiveness of Anticyberbullying messages on Facebook. *Cyberpsychology, Behavior, and Social Networking*, *16*(3), 175182. https://doi.org/10.1089/cyber.2012.0265

Arsenyan, J., & Mirowska, A. (2021). Almost human? A comparative case study on the social media presence of virtual influencers. *International Journal of Human Computer Studies, 155*. https://doi.org/10.1016/j.ijhcs.2021.102694

Block, E., & Lovegrove, R. (2021). Discordant storytelling, 'honest fakery', identity peddling: How uncanny CGI characters are jamming public relations and influencer practices. *Public Relations Inquiry*, *10*(3), 265-293. https://doi.org/10.1177/2046147x211026936

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny Valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, *29*(3), 759-771. https://doi.org/10.1016/j.chb.2012.11.021

Cao, C., Simon, T., Kim, J. K., Schwartz, G., Zollhoefer, M., Saito, S., Lombardi, S., Wei, S., Belko, D., Yu, S., Sheikh, Y., & Saragih, J. (2022). Authentic volumetric avatars from a phone scan. *ACM Transactions on Graphics*, *41*(4), 1-19. https://doi.org/10.1145/3528223.3530143

Casaló, L., Flavián, C., & Guinalíu, M. (2007). The impact of participation in virtual brand communities on consumer trust and loyalty: The case of free software. *Online Information Review*, *31*(6), 775–792.

Caudron, R., Nicq, P., & Valenza, E. (2016). *Blender 3D: Designing objects*. Packt Publishing.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.

Constine, J. (2018, April 3). *Instagram suddenly chokes off developers as Facebook chases privacy*. TechCrunch. https://social.techcrunch.com/2018/04/02/instagram-api-limit/

Cortis, K., & Davis, B. (2021). Over a decade of social opinion mining: A systematic review. *Artificial Intelligence Review, 54*(7),48734965. https://doi.org/10.1007/s10462-021-10030-2

Dijkmans, C., Kerkhof, P., Buyukcan-Tetik, A., & Beukeboom, C. J. (2015). Online Conversation and Corporate Reputation: A Two-Wave Longitudinal Study on the Effects of Exposure to the Social Media Activities of a Highly Interactive Company. *Journal of Computer-Mediated Communication, 20*(6), 632–648. https://doi.org/10.1111/jcc4.12132

Dixon, S. (2022, July 27). *U.S.: Average age of Instagram users 2022*. Statista. https://www.statista.com/statistics/398166/us-instagram-user-age-distribution/

Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior, 68*, 1–7. https://doi.org/10.1016/j.chb.2016.11.009

Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J., & Goodman, S. (2019). Social media engagement behavior: A framework for engaging customers through social media content. *European Journal of Marketing*, https://doi.org/10.1108/EJM-03-2017-0182.

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior, 90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Freud, S. (2004). The uncanny (1919), In: Sandner D (ed.) Fantastic Literature: A Critical Reader. Greenwood, pp.74–101.

Geller, T. (2008). Overcoming the uncanny Valley. *IEEE Computer Graphics and Applications*, *28*(4), 11-17. https://doi.org/10.1109/mcg.2008.79

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, *108*(3), 233-242. https://doi.org/10.1037/0735-7036.108.3.233

Griffiths, J. (2018, December 18). *Dior mocked for using CGI influencer to model its latest beauty campaign*. The Sun. https://www.thesun.co.uk/fabulous/8007013/dior-mocked-cgi-influencer-noonoouri-model-beauty-campaign/

Hanus, M.D., Fox, J. (2015). Persuasive avatars: the effects of customizing a virtual salesperson′s appearance on brand liking and purchase intentions. *Int. J. Hum. Comput. Stud. 84, 33–40*. https://doi.org/10.1016/j.ijhcs.2015.07.004.

Hiort, A. (2022, October 4). *How many virtual influencers are there?* VirtualHumans.org — Virtual Influencers, Explained. https://www.virtualhumans.org/article/how-many-virtual-influencers-are-there

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the Uncanny Valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518.

Jochum, E. and Goldberg, K. (2016), Cultivating the uncanny: The telegarden and other oddities. In: Herath D and Kroos C (eds.) Robots and Art. Singapore: Springer Singapore, pp.149–175.

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny Valley hypotheses: Support for perceptual mismatch as one road to the Valley of eeriness. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00390

Khan, R. F., & Sutcliffe, A. (2013). Attractive agents are more persuasive. *International Journal of Human-Computer Interaction*, *30*(2), 142150. https://doi.org/10.1080/10447318.2013.839904

Khamis, S., Ang, L., & Welling, R. (2016). Self-branding, 'micro-celebrity' and the rise of Social Media influencers. *Celebrity Studies*, *8*(2), 191–208. https://doi.org/10.1080/19392397.2016.1218292

Kim, D. Y., & Kim, H. (2021). Trust me, trust me not: A nuanced view of influencer marketing on social media. *Journal of Business Research*, *134*, 223232. https://doi.org/10.1016/j.jbusres.2021.05.024

Kitterman, T. (2021, December 16). *5 influencer marketing predictions for 2021*. PR Daily. https://www.prdaily.com/5-influencer-marketing-predictions-for-2021/

Lee, D., Hosanagar, K., & Nair, H. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, *64*(11), 4967–5460.

Lewkowicz, D. J., & Ghazanfar, A. A. (2011). The development of the uncanny Valley in infants. *Developmental Psychobiology*, *54*(2), 124-132. https://doi.org/10.1002/dev.20583

Liu, B., & Zhang, L. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Lombardi, S., Saragih, J., Simon, T., & Sheikh, Y. (2018). Deep appearance models for face rendering. *ACM Transactions on Graphics, 37*(4), 1-13. https://doi.org/10.1145/3197517.3201401

Luger, E., & Sellen, A. (2016). "Like having a really bad PA". *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2858036.2858288

MacDorman, K. F., Green, R. D., Ho, C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695-710. https://doi.org/10.1016/j.chb.2008.12.026

MeaningCloud. (2022). MeaningCloud for Excel 365. Retreived July 3, 2022, from https://www.meaningcloud.com/developer/excel-365-addin

Mende, M. A., Fischer, M. H., & Kühne, K. (2019). The use of social robots and the uncanny valley phenomenon. In *AI love you* (pp. 41–73). Springer.

Miao, F., Kozlenkova, I. v., Wang, H., Xie, T., & Palmatier, R. W. (2021). An Emerging Theory of Avatar Marketing. *Journal of Marketing*. https://doi.org/10.1177/0022242921996646

Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking, 16*(9), 708-713. https://doi.org/10.1089/cyber.2012.0334

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine 19*(2): 98–100.

Mori, M. (1970). The uncanny valley. Energy, 7(4), 33–35.

Mostafa, M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications, 40*(10), 4241-4251. https://doi.org/10.1016/j.eswa.2013.01.019

Nagamanjula, R., & Pethalakshmi, A. (2020). A novel framework based on Bi-objective optimization and LAN2FIS for Twitter sentiment analysis. *Social Network Analysis and Mining, 10*(1). https://doi.org/10.1007/s13278-020-00648-5

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining, 11*(1). https://doi.org/10.1007/s13278-021-00776-6

Nowak, Kristine L. and Jesse Fox (2018), "Avatars and Computer Mediated Communication: A Review of the Definitions, Uses, and Effects of Digital Representations," *Review of Communication Research, 6*, 30–53

Pagan, N., Mei, W., Li, C., & Dörfler, F. (2021). A meritocratic network formation model for the rise ofsocial media influencers. https://doi.org/10.21203/rs.3.rs-424956/v1

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143 157. https://doi.org/10.1016/j.joi.2009.01.003

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior, 20*(2), 201–223. https://doi.org/10.1016/j.chb.2003.10.015

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). "Alexa is my new BFF". *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. https://doi.org/10.1145/3027063.3053246

Rasmussen, M. (2021, September 9). *Brud, creators of virtual human Lil Miquela, announce a new direction*. VirtualHumans.org — Virtual Influencers, Explained. https://www.virtualhumans.org/article/brud-creators-of-virtual-human-lil-miquela-announce-a-new-direction

Saltz, J. S., & Hotz, N. (2020). Identifying the most common frameworks data science teams use to structure and coordinate their projects. *2020 IEEE International Conference on Big Data (Big Data)*. https://doi.org/10.1109/bigdata50022.2020.9377813

Saltz, J. S., & Krasteva, I. (2022). Current approaches for executing big data science projects—a systematic literature review. *PeerJ Computer Science, 8*, e862. https://doi.org/10.7717/peerj-cs.862

Schäfer, F., Zeiselmair, C., Becker, J., & Otten, H. (2018). Synthesizing CRISP-DM and quality management: A data mining approach for production processes. *2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*. https://doi.org/10.1109/itmc.2018.8691266

Sheldon, Z., Romanowski, M., & Shafer, D. M. (2019). Parasocial interactions and digital characters: The changing landscape of cinema and viewer/character relationships. *Atlantic Journal of Communication*, *29*(1), 15-25. https://doi.org/10.1080/15456870.2019.1702550

Statista Research Department. (2022, August 19). *Global influencer marketing value 2016-2022*. Statista. https://www.statista.com/statistics/1092819/global-influencer-market-size/

Thepsoonthorn, C., Ogawa, K., & Miyake, Y. (2021). The exploration of the uncanny Valley from the viewpoint of the robot's nonverbal behaviour. *International Journal of Social Robotics*, *13*(6), 1443-1455. https://doi.org/10.1007/s12369-020-00726-w

Tinwell, A., Grimshaw, M., & Abdel-Nabi, D. (2011). Effect of emotion and articulation of speech on the uncanny Valley in virtual characters. Affective Computing and Intelligent Interaction, 557-566. https://doi.org/10.1007/978-3-642-24571-8_69

Tokárová, H. (2021). *Customizable 3D avatar for virtual reality applications*. Masaryk University.

Travers, C. (2020, April 12). *5 notable virtual influencer + brand partnerships*. VirtualHumans.org — Virtual Influencers, Explained. https://www.virtualhumans.org/article/5-notable-virtual-influencer-brand-partnerships

VirtualHumans.org – *Virtual influencers, explained*. (2022, January 14). VirtualHumans.org – Virtual Influencers, Explained. https://www.virtualhumans.org/#influencers

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., & Pirner, P., et al (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research, 13*(3), 253–266.

Yam, K. C., Bigman, Y., & Gray, K. (2021). Reducing the uncanny valley by dehumanizing humanoid robots. *Computers in Human Behavior, 125*, 106945. https://doi.org/10.1016/j.chb.2021.106945

7 Annexes

Data overview of VAI

Nr.	Post date	Post URL	Scaled Comments	Scaled Likes	Engagement in %	Sentiment Score
1	01.01.22	https://www.instagram.com/p/CYLmAvcpw4U/	0.27	16.89	1.72	1.05
2	03.01.22	https://www.instagram.com/p/CYQ1cfzpcr7/	0.23	15.84	1.61	1.22
3	06.01.22	https://www.instagram.com/p/CYYd_xUMhrF/	0.16	17.73	1.79	1.25
4	07.01.22	https://www.instagram.com/p/CYbFPbyp3X0/	0.42	18.90	1.93	1.14
5	10.01.22	https://www.instagram.com/p/CYjUdPeMcnU/	0.17	19.56	1.97	1.36
6	12.01.22	https://www.instagram.com/p/CYoDIZ_MHII/	0.14	15.21	1.53	1.10
7	13.01.22	https://www.instagram.com/p/CYq5CWsMQ57/	0.19	22.41	2.26	1.05
8	14.01.22	https://www.instagram.com/p/CYtLUE1pwt7/	0.20	20.37	2.06	1.05
9	17.01.22	https://www.instagram.com/p/CY1f2zcFpNo/	0.22	17.71	1.79	1.12
10	18.01.22	https://www.instagram.com/p/CY2rhgVIKJq/	0.05	6.48	0.65	1.00
11	19.01.22	https://www.instagram.com/p/CY5pN5YhQvp/	0.22	16.92	1.71	1.13
12	19.01.22	https://www.instagram.com/p/CY5p0jOOzlF/	0.09	7.34	0.74	0.40
13	20.01.22	https://www.instagram.com/p/CY8VW3UrWa7/	0.14	19.79	1.99	1.07
14	22.01.22	https://www.instagram.com/p/CZB5juxMYol/	0.23	17.29	1.75	1.16
15	23.01.22	https://www.instagram.com/p/CZFTaNwJsVe/	0.37	18.49	1.89	1.10
16	24.01.22	https://www.instagram.com/p/CZGe-O8MXbQ/	0.13	10.67	1.08	0.08
17	24.01.22	https://www.instagram.com/p/CZG9KUaJSqX/	0.24	24.76	2.50	0.86
18	25.01.22	https://www.instagram.com/p/CZJZlRwMP47/	0.21	19.10	1.93	1.05
19	26.01.22	https://www.instagram.com/p/CZM93A2st_H/	0.25	18.11	1.84	1.17
20	27.01.22	https://www.instagram.com/p/CZPl15BMQo5/	0.21	21.49	2.17	1.14
21	31.01.22	https://www.instagram.com/p/CZZi3T6JLwt/	0.16	14.73	1.49	1.00
22	01.02.22	https://www.instagram.com/p/CZcG-OYrUwW/	0.14	16.11	1.63	1.15
23	03.02.22	https://www.instagram.com/p/CZg7P1Os9ZC/	0.09	14.37	1.45	1.00
24	04.02.22	https://www.instagram.com/p/CZj46RSMhKt/	0.19	16.00	1.62	0.89
25	05.02.22	https://www.instagram.com/p/CZnJ2ZyA2WN/	0.61	23.66	2.43	1.02
26	08.02.22	https://www.instagram.com/p/CZtM9P3sbCv/	0.12	17.70	1.78	1.15
27	10.02.22	https://www.instagram.com/p/CZy-oZTMrjr/	0.07	12.12	1.22	1.27
28	10.02.22	https://www.instagram.com/p/CZy-zDDsRUn/	0.09	18.26	1.84	1.29
29	10.02.22	https://www.instagram.com/p/CZy-2xggDEz/	0.15	14.23	1.44	1.23
30	13.02.22	https://www.instagram.com/p/CZ6tZcfAZ3R/	0.18	17.59	1.78	1.10
31	14.02.22	https://www.instagram.com/p/CZ9lN53AAXL/	0.33	15.51	1.58	1.06
32	15.02.22	https://www.instagram.com/p/CaAdWuFsLcQ/	0.37	19.20	1.96	1.00
33	20.02.22	https://www.instagram.com/p/CaMix-yA5d_/	0.15	24.19	2.43	0.81
34	20.02.22	https://www.instagram.com/p/CaNUJZlgczD/	0.12	13.27	1.34	0.78
35	21.02.22	https://www.instagram.com/p/CaP5H3PMBiy/	0.07	16.64	1.67	0.81
36	22.02.22	https://www.instagram.com/p/CaRy3o5MvuF/	0.13	15.80	1.59	1.06
37	23.02.22	https://www.instagram.com/p/CaUP2J4smHK/	0.21	22.52	2.27	0.89
38	24.02.22	https://www.instagram.com/p/CaWqk-DAZHk/	0.06	19.45	1.95	1.14
39	24.02.22	https://www.instagram.com/p/CaW4OcIMzm-/	0.23	14.25	1.45	0.49
40	24.02.22	https://www.instagram.com/p/CaXS_KyAQEJ/	0.05	13.31	1.34	0.92
41	24.02.22	https://www.instagram.com/p/CaXh6uvM5m3/	0.09	20.82	2.09	1.07
42	25.02.22	https://www.instagram.com/p/CaZIjt5gurz/	0.13	14.58	1.47	1.20

43	25.02.22	https://www.instagram.com/p/CaaKRobgCS3/	0.13	15.85	1.60	0.82
44	26.02.22	https://www.instagram.com/p/CabuHFpMCG7/	0.28	28.77	2.90	1.15
45	26.02.22	https://www.instagram.com/p/CacXw6XgquG/	0.05	16.45	1.65	0.93
46	27.02.22	https://www.instagram.com/p/CaeVQqBAsRL/	0.15	19.28	1.94	1.14
47	27.02.22	https://www.instagram.com/p/CafSSGvM-nZ/	0.38	20.84	2.12	1.13
48	01.03.22	https://www.instagram.com/p/CakSaVgspZM/	0.22	23.81	2.40	1.05
49	02.03.22	https://www.instagram.com/p/CamUaZDgucl/	0.18	16.25	1.64	1.07
50	03.03.22	https://www.instagram.com/p/Caon5RhsYPE/	0.18	23.01	2.32	1.10
51	04.03.22	https://www.instagram.com/p/CasWU0QATnJ/	0.18	22.14	2.23	1.00
52	05.03.22	https://www.instagram.com/p/Cat_yzoAVvX/	0.10	11.63	1.17	1.00
53	06.03.22	https://www.instagram.com/p/CawZExcMH2Y/	0.17	15.72	1.59	0.95
54	06.03.22	https://www.instagram.com/p/CawmbnZstXr/	0.16	9.54	0.97	1.00
55	07.03.22	https://www.instagram.com/p/CayxCBPgYN1/	0.21	21.33	2.15	0.87
56	08.03.22	https://www.instagram.com/p/Ca18EhYs9gH/	0.16	19.41	1.96	0.80
57	09.03.22	https://www.instagram.com/p/Ca3IFk_gNbQ/	0.48	25.64	2.61	1.00
58	10.03.22	https://www.instagram.com/p/Ca8GmtlgJg6/	1.49	24.11	2.56	1.06
59	11.03.22	https://www.instagram.com/p/Ca9Qq-8A1oX/	0.26	20.70	2.10	1.00
60	12.03.22	https://www.instagram.com/p/CbAat4GAh8f/	0.86	14.26	1.51	1.04
61	13.03.22	https://www.instagram.com/p/CbCiCA3MadJ/	0.14	16.19	1.63	1.08
62	14.03.22	https://www.instagram.com/p/CbGWdVWA8i2/	0.32	23.38	2.37	1.04
63	18.03.22	https://www.instagram.com/p/CbQZ7VNg7Mt/	0.18	19.01	1.92	1.13
64	22.03.22	https://www.instagram.com/p/CbZ1E5ssofU/	0.18	15.59	1.58	1.05
65	24.03.22	https://www.instagram.com/p/CbfDML_BkdE/	0.32	22.78	2.31	0.80
66	30.03.22	https://www.instagram.com/p/CbvHyVZBfj1/	0.12	18.61	1.87	1.04
67	31.02.22	https://www.instagram.com/p/CbwkJ_nhp3d/	0.13	13.52	1.36	1.25
68	02.04.22	https://www.instagram.com/p/Cb2OTfQgwr5/	0.11	33.25	3.34	1.17
69	02.04.22	https://www.instagram.com/p/Cb3MmI1gQNg/	0.17	13.32	1.35	1.04
70	03.04.22	https://www.instagram.com/p/Cb5xiudgncW/	0.13	40.08	4.02	0.91
71	05.04.22	https://www.instagram.com/p/Cb-EWmMw7L/	0.18	20.89	2.11	1.05
72	07.04.22	https://www.instagram.com/p/CcDaCxlMZ69/	0.15	20.21	2.04	1.25
73	10.04.22	https://www.instagram.com/p/CcL0EzbsXIm/	0.15	22.24	2.24	1.16
74	12.04.22	https://www.instagram.com/p/CcP3h_PMu39/	0.16	26.70	2.69	0.90
75	17.04.22	https://www.instagram.com/p/CccXoh9sCAK/	0.28	31.02	3.13	1.00
76	10.04.22	https://www.instagram.com/p/Cch0I9osUsx/	0.16	19.54	1.97	0.96
77	22.04.22	https://www.instagram.com/p/CcpZIjUMzCE/	0.23	26.36	2.66	1.17
78	24.04.22	https://www.instagram.com/p/Ccumv3EM60i/	0.26	21.62	2.19	1.08
79	29.04.22	https://www.instagram.com/p/Cc7q9HKMBPR/	0.16	16.62	1.68	1.17
80	02.05.22	https://www.instagram.com/p/CdDfVe2gCnF/	0.36	24.81	2.52	1.19
81	03.04.22	https://www.instagram.com/p/CdGB3fVMAcb/	0.43	20.16	2.06	1.06
82	04.05.22	https://www.instagram.com/p/CdJeyl3MB6R/	0.07	7.88	0.79	1.00
83	07.05.22	https://www.instagram.com/p/CdPuxmiMlhu/	0.16	13.00	1.32	1.06
84	10.05.22	https://www.instagram.com/p/CdYintLMaDN/	0.20	18.84	1.90	0.29
85	13.05.22	https://www.instagram.com/p/CdfiV5eghUI/	0.23	23.36	2.36	0.00
86	17.05.22	https://www.instagram.com/p/CdqEphVMb29/	0.18	19.39	1.96	1.03
87	20.05.22	https://www.instagram.com/p/CdyP4naow7Q/	0.40	18.08	1.85	0.92
88	22.05.22	https://www.instagram.com/p/Cd2p6q_Aa4D/	0.12	14.50	1.46	1.29
89	23.05.22	https://www.instagram.com/p/Cd6BGx6AEzj/	0.19	14.16	1.43	1.19
90	28.05.22	https://www.instagram.com/p/CeG3qHmslWQ/	0.19	11.03	1.12	1.08
91	01.06.22	https://www.instagram.com/p/CeQrC-TA6CW/	0.16	17.75	1.79	0.96
92	03.06.22	https://www.instagram.com/p/CeWSjvgAh6g/	0.21	17.00	1.72	0.94
93	07.06.22	https://www.instagram.com/p/CegbKtFgSYi/	0.16	9.57	0.97	0.95

94	09.06.22	https://www.instagram.com/p/CekvIcCAj59/	0.09	10.39	1.05	1.06
95	10.06.22	https://www.instagram.com/p/CeoDnhTA-s_/	0.11	13.92	1.40	0.95
96	13.06.22	https://www.instagram.com/p/Cevgs_ygqBX/	0.19	15.61	1.58	1.00
97	17.06.22	https://www.instagram.com/p/Ce6H-WvIK-q/	0.25	19.36	1.96	1.07
98	19.06.22	https://www.instagram.com/p/Ce-rGj1sUK8/	0.07	10.24	1.03	1.22
99	24.06.22	https://www.instagram.com/p/CfL2mvMAXnh/	0.21	11.02	1.12	0.91
100	25.06.22	https://www.instagram.com/p/CfOi4qFgg8n/	0.16	13.47	1.36	1.09
101	28.06.22	https://www.instagram.com/p/CfWcm53gk7i/	0.15	9.99	1.01	1.09
102	30.06.22	https://www.instagram.com/p/CfbMGJwg1jh/	0.13	11.72	1.18	1.00

Data overview of VSI

Nr.	Post date	Post URL	Scaled Comments	Scaled Likes	Engagement in %	Sentiment Score
1	01.01.22	https://www.instagram.com/p/CYNH9xpBfZV/	0.29	15.74	1.60	0.98
2	03.01.22	https://www.instagram.com/p/CYSAhoAheNs/	0.18	11.81	1.20	1.04
3	04.01.22	https://www.instagram.com/p/CYUnaJBhfBq/	1.99	309.6	31.16	0.93
4	06.01.22	https://www.instagram.com/p/CYZ2-3BBDAD/	0.23	9.75	1.00	0.93
5	10.01.22	https://www.instagram.com/p/CYiWcM_hbj8/	0.22	9.20	0.94	0.77
6	15.01.22	https://www.instagram.com/p/CYw420OuHzI/	0.14	23.65	2.38	1.04
7	20.01.22	https://www.instagram.com/p/CY8-SJghIq0/	0.29	6.62	0.69	1.03
8	24.01.22	https://www.instagram.com/p/CZH_yLJBnBw/	0.19	6.48	0.67	1.07
9	28.01.22	https://www.instagram.com/p/CZSYIPTB62L/	0.26	8.03	0.83	1.00
10	31.01.22	https://www.instagram.com/p/CZaGG2IhA_J/	0.23	7.88	0.81	1.07
11	02.02.22	https://www.instagram.com/p/CZfEkcMhTxo/	0.15	10.22	1.04	0.82
12	04.02.22	https://www.instagram.com/p/CZkafRQuCM2/	0.09	6.57	0.67	1.14
13	05.02.22	https://www.instagram.com/p/CZm1Ta2oG-c/	0.07	6.15	0.62	1.05
14	06.02.22	https://www.instagram.com/p/CZpuwyUoO0L/	0.14	8.52	0.87	0.96
15	07.02.22	https://www.instagram.com/p/CZroXtWt0qj/	0.13	5.92	0.61	1.11
16	08.02.22	https://www.instagram.com/p/CZuY7mSqhw2/	0.08	11.52	1.16	0.88
17	10.02.22	https://www.instagram.com/p/CZzevanqzNF/	0.11	6.54	0.66	0.95
18	11.02.22	https://www.instagram.com/p/CZ15eo6qlkU/	0.19	7.62	0.78	1.00
19	12.02.22	https://www.instagram.com/p/CZ4S-BHKvlj/	0.19	117.7	11.78	1.10
20	13.02.22	https://www.instagram.com/p/CZ7FjQFKHO9/	0.32	9.60	0.99	0.94
21	15.02.22	https://www.instagram.com/p/CaAMLY0qsyW/	0.20	8.34	0.85	1.00
22	16.02.22	https://www.instagram.com/p/CaC9bzMos4S/	0.17	47.30	4.75	0.96
23	18.02.22	https://www.instagram.com/p/CaIBp45qNkS/	0.18	21.80	2.20	0.80
24	19.02.22	https://www.instagram.com/p/CaKjo77qkRb/	0.13	8.42	0.86	1.07
25	20.02.22	https://www.instagram.com/p/CaNMOeDqFYt/	0.14	107.1	10.73	1.00
26	22.02.22	https://www.instagram.com/p/CaSNpPwqDaU/	0.09	12.20	1.23	1.09
27	23.02.22	https://www.instagram.com/p/CaU45mzq_te/	0.22	22.37	2.26	1.03
28	25.02.22	https://www.instagram.com/p/CaaFT--qxeW/	0.21	68.25	6.85	1.06
29	26.02.22	https://www.instagram.com/p/CacpouQqBU5/	0.11	42.10	4.22	0.91
30	01.03.22	https://www.instagram.com/p/CakWZ6HK_io/	0.27	33.12	3.34	0.97
31	02.03.22	https://www.instagram.com/p/Cam3n_Nowkc/	0.18	14.91	1.51	1.05
32	03.02.22	https://www.instagram.com/p/CapdRXBKpUb/	0.29	15.45	1.57	1.08
33	04.03.22	https://www.instagram.com/p/Car30ycKUEZ/	0.17	39.72	3.99	1.05
34	05.04.22	https://www.instagram.com/p/CaulSnFqgsB/	0.13	12.27	1.24	1.03

35	06.03.22	https://www.instagram.com/p/CaxHJeEKAGv/	0.22	17.07	1.73	0.96
36	07.03.22	https://www.instagram.com/p/Ca0JY8VKPbn/	0.17	94.74	9.49	1.00
37	09.03.22	https://www.instagram.com/p/Ca44l-JqK9n/	0.26	28.35	2.86	1.07
38	10.03.22	https://www.instagram.com/p/Ca7iszZq_Nb/	0.33	11.88	1.22	1.13
39	11.03.22	https://www.instagram.com/p/Ca-aDXKqESz/	0.23	12.50	1.27	1.00
40	13.03.22	https://www.instagram.com/p/CbDDEuSq02j/	0.33	19.58	1.99	1.04
41	14.03.22	https://www.instagram.com/p/CbF2kRNqw9U/	0.25	14.51	1.48	0.94
42	15.03.22	https://www.instagram.com/p/CbIVRcfKTxC/	0.10	4.57	0.47	1.17
43	16.03.22	https://www.instagram.com/p/CbK3sGoKGx_/	0.09	8.46	0.85	1.18
44	17.03.22	https://www.instagram.com/p/CbNjuAtKjkZ/	0.16	5.12	0.53	1.11
45	18.03.22	https://www.instagram.com/p/CbQI-90qpa0/	0.22	7.14	0.74	1.03
46	22.03.22	https://www.instagram.com/p/CbaOErRIj71/	0.13	10.69	1.08	1.00
47	23.03.22	https://www.instagram.com/p/Cbc3ULZqWlH/	0.25	6.58	0.68	1.00
48	24.03.22	https://www.instagram.com/p/CbfhKyKqJl-/	0.08	8.80	0.89	1.00
49	25.03.22	https://www.instagram.com/p/CbiEhDgoAK1/	0.33	51.85	5.22	1.04
50	26.03.22	https://www.instagram.com/p/Cbkq-p2qZC_/	0.18	5.16	0.53	1.09
51	27.03.22	https://www.instagram.com/p/CbnCptSqfkH/	0.11	5.03	0.51	1.29
52	28.03.22	https://www.instagram.com/p/Cbp1sVHqHOO/	0.10	11.53	1.16	0.94
53	29.03.22	https://www.instagram.com/p/CbsJMe7q41H/	0.14	5.22	0.54	1.06
54	30.03.22	https://www.instagram.com/p/Cbu4zfLKk-L/	0.20	5.61	0.58	0.63
55	01.04.22	https://www.instagram.com/p/Cb0Z1GmpQtQ/	0.18	5.71	0.59	1.04
56	02.04.22	https://www.instagram.com/p/Cb24A6XoxhM/	0.10	5.20	0.53	1.14
57	03.04.22	https://www.instagram.com/p/Cb42_mMqP1l/	0.18	5.92	0.61	1.14
58	05.04.22	https://www.instagram.com/p/Cb-qYgxqoWM/	0.09	5.90	0.60	0.91
59	06.04.22	https://www.instagram.com/p/CcAzPAEKLHJ/	0.14	5.00	0.51	1.00
60	07.04.22	https://www.instagram.com/p/CcDkQelqo5I/	0.16	4.73	0.49	1.14
61	12.04.22	https://www.instagram.com/p/CcQCBxTqAMN	0.10	7.08	0.72	1.00
62	13.04.22	https://www.instagram.com/p/CcSwUHWKU1q	0.70	73.50	7.42	1.11
63	17.04.22	https://www.instagram.com/p/CcdMqJlqBWN/	0.14	9.25	0.94	1.06
64	21.04.22	https://www.instagram.com/p/CcnYFlGKZvZ/	0.15	6.11	0.63	0.85
65	22.04.22	https://www.instagram.com/p/CcqAgw-KMkF/	0.06	6.14	0.62	0.82
66	23.04.22	https://www.instagram.com/p/Ccq_p78ouoX/	0.17	66.91	6.71	1.08
67	23.04.22	https://www.instagram.com/p/CcsYNO2qIYS/	0.21	7.29	0.75	1.05
68	24.04.22	https://www.instagram.com/p/Ccu2oPBqoGF/	0.19	4.52	0.47	0.81
69	25.04.22	https://www.instagram.com/p/CcwF0yFInTr/	0.10	3.77	0.39	1.29
70	26.04.22	https://www.instagram.com/p/CczwaQ1I1Fb/	0.22	6.35	0.66	1.18
71	28.04.22	https://www.instagram.com/p/Cc47vdqKINm/	0.10	16.73	1.68	1.13
72	01.05.22	https://www.instagram.com/p/CdAvY6FoPo-/	0.19	5.16	0.53	1.05
73	07.05.22	https://www.instagram.com/p/CdQCn7VK-MN/	0.15	8.55	0.87	0.91
74	24.05.22	https://www.instagram.com/p/Cd7sXX2KCfk/	0.28	5.57	0.58	0.96
75	26.05.22	https://www.instagram.com/p/CeB8zoQK7gd/	0.35	6.36	0.67	0.93
76	30.05.22	https://www.instagram.com/p/CeLS0XsKAUZ/	0.24	5.57	0.58	1.05
77	04.06.22	https://www.instagram.com/p/CeYSFDgqaWg/	1.08	502.3	50.34	1.17
78	08.06.22	https://www.instagram.com/p/Ceig39fIkmY/	0.17	4.94	0.51	1.33
79	08.06.22	https://www.instagram.com/p/CejNZC0o88B/	0.24	4.86	0.51	1.10
80	23.06.22	https://www.instagram.com/p/CfH7jstKe7S/	0.98	583.2	58.42	1.15
81	24.06.22	https://www.instagram.com/p/CfLsCw3qpgY/	0.20	5.97	0.62	1.04

Data overview of VHI

Nr.	Post date	Post URL	Scaled Comments	Scaled Likes	Engagement in %	Sentiment Score
1	01.01.22	https://www.instagram.com/p/CYLa5FxLIfS/	0.10	26.44	2.65	0.77
2	01.01.22	https://www.instagram.com/p/CYMk6zPps9N/	0.17	36.44	3.66	0.18
3	05.01.22	https://www.instagram.com/p/CYW4w-YJmfY/	0.22	34.43	3.47	0.64
4	08.01.22	https://www.instagram.com/p/CYen0wuJlOw/	0.15	16.32	1.65	0.81
5	12.01.22	https://www.instagram.com/p/CYo6u4JOzXf/	0.12	13.70	1.38	0.52
6	15.01.22	https://www.instagram.com/p/CYwoUm-uiRH/	0.16	13.50	1.37	0.52
7	19.01.22	https://www.instagram.com/p/CY68tyCuLbT/	0.17	18.93	1.91	0.46
8	22.01.22	https://www.instagram.com/p/CZCqOSLOzBb/	0.15	18.55	1.87	0.54
9	26.01.22	https://www.instagram.com/p/CZM-LD-pFUI/	0.14	12.24	1.24	0.50
10	29.01.22	https://www.instagram.com/p/CZUsCiFOjrM/	0.10	14.61	1.47	0.68
11	02.02.22	https://www.instagram.com/p/CZe_WGZJJOM/	0.09	17.22	1.73	0.53
12	05.02.22	https://www.instagram.com/p/CZmuLNYOexm	0.12	21.64	2.18	0.47
13	06.02.22	https://www.instagram.com/p/CZpz7glpsVv/	0.11	17.14	1.72	0.72
14	09.02.22	https://www.instagram.com/p/CZxA02duxeJ/	0.06	8.11	0.82	0.86
15	12.02.22	https://www.instagram.com/p/CZ4vYJRueP6/	0.06	7.91	0.80	0.91
16	14.02.22	https://www.instagram.com/p/CZ72olbp3aT/	0.13	12.26	1.24	0.45
17	14.02.22	https://www.instagram.com/p/CZ94-0BJXNf/	0.04	5.30	0.53	0.78
18	16.02.22	https://www.instagram.com/p/CaDCoA6p6BE/	0.06	13.65	1.37	0.61
19	19.02.22	https://www.instagram.com/p/CaKwqhBOZEB/	0.08	16.63	1.67	0.55
20	23.02.22	https://www.instagram.com/p/CaVEjSLJJcs/	0.05	5.62	0.57	0.41
21	26.02.22	https://www.instagram.com/p/CacyrpLu_bp/	0.08	8.76	0.88	0.67
22	02.03.22	https://www.instagram.com/p/CanFy_YOuCB/	0.05	5.25	0.53	0.65
23	05.03.22	https://www.instagram.com/p/Cau07VvOuol/	0.07	4.16	0.42	0.70
24	07.03.22	https://www.instagram.com/p/Caz8yzGuqRq/	0.15	13.24	1.34	0.63
25	09.03.22	https://www.instagram.com/p/Ca5HFLSuX_M/	0.04	6.98	0.70	0.91
26	10.03.22	https://www.instagram.com/p/Ca7sE0-AUaR/	0.02	–	–	0.68
27	11.03.22	https://www.instagram.com/p/Ca-Qr51OKiY/	0.05	9.44	0.95	0.77
28	12.03.22	https://www.instagram.com/p/CbA1_qCuoge/	0.05	5.24	0.53	0.43
29	16.03.22	https://www.instagram.com/p/CbLByPtuOwB/	0.20	18.37	1.86	0.34
30	21.03.22	https://www.instagram.com/p/CbX5wLAuq5i/	0.03	7.45	0.75	0.73
31	23.03.22	https://www.instagram.com/p/CbdC4JDOzhD/	0.17	22.52	2.27	0.71
32	26.03.22	https://www.instagram.com/p/CbkyEeauO7e/	0.06	6.65	0.67	0.79
33	28.03.22	https://www.instagram.com/p/Cbp7K2VOOXy/	0.05	5.81	0.59	1.04
34	30.03.22	https://www.instagram.com/p/CbvHOLJpeMv/	0.07	9.16	0.92	0.53
35	02.04.22	https://www.instagram.com/p/Cb2zQpIOTsZ/	0.12	20.64	2.08	0.27
36	04.04.22	https://www.instagram.com/p/Cb78nJ-JWxm/	0.07	10.37	1.04	0.77
37	06.04.22	https://www.instagram.com/p/CcBHWX2JA67/	0.11	13.71	1.38	0.40
38	08.04.22	https://www.instagram.com/p/CcGQF6uOik4/	0.40	36.20	3.66	-0.20
39	13.04.22	https://www.instagram.com/p/CcTHyAQOIBL/	0.10	12.73	1.28	0.31
40	15.04.22	https://www.instagram.com/p/CcYRtMLOVEB/	0.09	13.71	1.38	0.78
41	20.04.22	https://www.instagram.com/p/CclI62oucfY/	0.05	7.60	0.76	0.93
42	22.04.22	https://www.instagram.com/p/CcqUsGYO5ug/	0.33	37.83	3.82	0.15
43	27.04.22	https://www.instagram.com/p/Cc3OPibJMxm/	0.03	9.09	0.91	0.84
44	29.04.22	https://www.instagram.com/p/Cc8UckFpz0T/	0.17	19.18	1.93	0.66

45	04.05.22	https://www.instagram.com/p/CdJMhdZpK70/	0.12	–	–	0.49
46	06.05.22	https://www.instagram.com/p/CdOWX3kJJ2K/	0.27	42.25	4.25	-0.07
47	11.05.22	https://www.instagram.com/p/CdbONCtJfws/	0.67	50.02	5.07	0.47
48	14.05.22	https://www.instagram.com/p/Cdi7_R4pDRe/	0.23	14.24	1.45	0.18
49	18.05.22	https://www.instagram.com/p/CdtQGwMubNI/	0.26	24.74	2.50	0.50
50	21.05.22	https://www.instagram.com/p/Cd1R9mKuqjv/	0.20	21.88	2.21	0.46
51	22.05.22	https://www.instagram.com/p/Cd3jZHjuafJ/	0.14	13.32	1.35	0.26
52	27.05.22	https://www.instagram.com/p/CeEafPhuHyF/	0.24	25.53	2.58	0.43
53	28.05.22	https://www.instagram.com/p/CeG_kIUOk6X/	0.51	44.51	4.50	0.28
54	01.06.22	https://www.instagram.com/p/CeRb0ouOnGg/	0.22	19.54	1.98	0.54
55	04.06.22	https://www.instagram.com/p/CeZBY4AuouR/	0.16	16.52	1.67	0.54
56	08.06.22	https://www.instagram.com/p/CejUxelJ-D2/	0.22	21.23	2.15	0.29
57	11.06.22	https://www.instagram.com/p/CerCh80p2iI/	0.58	58.15	5.87	0.54
58	15.06.22	https://www.instagram.com/p/Ce1kYB7v_ew/	0.28	17.90	1.82	0.35
59	19.06.22	https://www.instagram.com/p/Ce_phpLJMuM/	0.12	8.45	0.86	0.58
60	22.06.22	https://www.instagram.com/p/CfHXqV8JbQI/	0.20	15.86	1.61	0.48
61	25.06.22	https://www.instagram.com/p/CfPFheTJ-y0/	0.44	41.13	4.16	0.66

Data overview of HI

Nr.	Post date	Post URL	Scaled Comments	Scaled Likes	Engagement in %	Sentiment Score
1	04.01.22	https://www.instagram.com/p/CYUQJOODhwf/	0.13	13.51	1.36	1.12
2	07.01.22	https://www.instagram.com/p/CYaHmOSgVm9	0.12	11.77	1.19	1.12
3	09.01.22	https://www.instagram.com/p/CYg1rxwg2qe/	0.18	16.68	1.69	1.09
4	11.01.22	https://www.instagram.com/p/CYkhHRpISHu/	0.07	11.48	1.15	1.12
5	13.01.22	https://www.instagram.com/p/CYpgcCqg_Fa/	0.14	12.53	1.27	1.03
6	22.01.22	https://www.instagram.com/p/CZCh1NwOLNl/	0.07	8.67	0.87	1.19
7	23.01.22	https://www.instagram.com/p/CZFcJhduK6b/	0.06	7.81	0.79	0.71
8	23.01.22	https://www.instagram.com/p/CZFxM04AAeQ/	0.15	14.03	1.42	1.03
9	24.01.22	https://www.instagram.com/p/CZHsTOjuRBY/	0.10	10.76	1.09	1.08
10	28.01.22	https://www.instagram.com/p/CZR9YOjo5V9/	0.20	12.97	1.32	1.10
11	29.01.22	https://www.instagram.com/p/CZUojMwOBA4/	0.06	7.29	0.74	1.10
12	31.01.22	https://www.instagram.com/p/CZYDxd5Anej/	0.06	9.97	1.00	1.30
13	31.01.22	https://www.instagram.com/p/CZZtNdfuTDQ/	0.12	8.86	0.90	1.21
14	01.02.22	https://www.instagram.com/p/CZanRiYAocq/	0.09	10.24	1.03	1.33
15	06.02.22	https://www.instagram.com/p/CZpWEYHguay/	0.24	20.50	2.07	1.13
16	07.02.22	https://www.instagram.com/p/CZrnKHtu5rn/	0.05	7.81	0.79	1.00
17	09.02.22	https://www.instagram.com/p/CZu-zZ7gJOi/	0.21	14.95	1.52	1.10
18	10.02.22	https://www.instagram.com/p/CZzcynnOQzj/	0.11	9.41	0.95	1.17
19	15.02.22	https://www.instagram.com/p/CZ-2nekjxAd/	0.07	10.69	1.08	1.28
20	16.02.22	https://www.instagram.com/p/CaDDI0aOD77/	0.06	11.32	1.14	1.15
21	18.02.22	https://www.instagram.com/p/CaGibzhg4zW/	0.05	7.96	0.80	1.75
22	18.02.22	https://www.instagram.com/p/CaItEMyuI3e/	0.05	7.54	0.76	1.17

23	20.02.22	https://www.instagram.com/p/CaNFeUvOSxD/	0.11	12.53	1.26	1.21
24	21.02.22	https://www.instagram.com/p/CaP4-8Rg3F2/	0.18	12.85	1.30	0.98
25	24.02.22	https://www.instagram.com/p/CaX0pO5pwhw/	0.08	11.43	1.15	1.06
26	26.02.22	https://www.instagram.com/p/Caa-FCjgvnd/	0.10	10.67	1.08	1.22
27	27.02.22	https://www.instagram.com/p/CafA5wFA8ed/	0.11	9.35	0.95	1.33
28	04.03.33	https://www.instagram.com/p/Casls3mu2VH/	0.14	11.80	1.19	1.14
29	07.03.22	https://www.instagram.com/p/CayIKqsgpXZ/	0.12	9.17	0.93	0.70
30	09.03.22	https://www.instagram.com/p/Ca486OAul_m/	0.15	10.86	1.10	1.08
31	12.03.22	https://www.instagram.com/p/CbAjwXZuGUB/	0.07	8.85	0.89	1.27
32	14.03.22	https://www.instagram.com/p/CbF1gbHOgl_/	0.07	10.07	1.01	1.10
33	15.03.22	https://www.instagram.com/p/CbIeIPdATrS/	0.66	75.19	7.59	1.13
34	16.03.22	https://www.instagram.com/p/CbLdNjROrCg/	0.07	7.81	0.79	1.14
35	17.03.22	https://www.instagram.com/p/CbNn96qOxPP/	0.08	10.12	1.02	1.17
36	19.03.22	https://www.instagram.com/p/CbTSB5cOjw0/	0.14	11.10	1.12	1.02
37	28.03.22	https://www.instagram.com/p/CbqGshDue9q/	0.13	12.87	1.30	1.12
38	30.03.22	https://www.instagram.com/p/Cbu0GpvA4vu/	0.24	21.19	2.14	1.08
39	31.03.22	https://www.instagram.com/p/Cbxcd10gpDI/	0.09	5.47	0.56	1.31
40	02.04.22	https://www.instagram.com/p/Cb2u2opuAb3/	0.05	7.27	0.73	1.25
41	04.04.22	https://www.instagram.com/p/Cb73KCWOAji/	0.11	12.48	1.26	1.03
42	05.04.22	https://www.instagram.com/p/Cb_D6Erum7p/	0.15	12.82	1.30	1.08
43	07.04.22	https://www.instagram.com/p/CcD1Ap2ujI6/	0.07	10.26	1.03	1.13
44	11.04.22	https://www.instagram.com/p/CcMDd_gOBAO	0.07	10.52	1.06	1.10
45	15.04.22	https://www.instagram.com/p/CcYte9Ygqly/	0.05	8.00	0.80	1.24
46	22.04.22	https://www.instagram.com/p/Ccq01qLucPq/	0.07	10.04	1.01	1.14
47	25.04.22	https://www.instagram.com/p/Ccx0tyauVQP/	0.13	12.58	1.27	1.13
48	26.04.22	https://www.instagram.com/p/Cc0msEbOval/	0.09	10.50	1.06	1.08
49	28.04.22	https://www.instagram.com/p/Cc6JpFrO0Rl/	0.12	10.19	1.03	1.14
50	29.04.22	https://www.instagram.com/p/Cc8TIIDgl8g/	0.08	–	–	1.08
51	04.05.22	https://www.instagram.com/p/CdJq3JkuX-P/	0.13	11.90	1.20	1.19
52	10.05.22	https://www.instagram.com/p/CdW0mWmDc/	0.14	11.91	1.21	1.14
53	14.05.22	https://www.instagram.com/p/CdhK3YiD2EF/	0.10	7.50	0.76	1.08
54	14.05.22	https://www.instagram.com/p/Cdi9YPquHiT/	0.15	8.34	0.85	1.44
55	15.05.22	https://www.instagram.com/p/CdlTlxvuEU3/	0.15	11.20	1.13	1.17
56	16.05.22	https://www.instagram.com/p/CdnmWvDAA3e/	0.07	7.78	0.78	1.22
57	20.05.22	https://www.instagram.com/p/Cdwj9kvvBu_/	0.15	8.15	0.83	1.15
58	25.05.22	https://www.instagram.com/p/Cd9idJpDwyy/	0.07	10.54	1.06	1.23
59	25.05.22	https://www.instagram.com/p/Cd_unnBuF9b/	0.21	15.80	1.60	1.12
60	26.05.22	https://www.instagram.com/p/Cd_2N5sAo0S/	0.25	17.39	1.76	1.07
61	28.05.22	https://www.instagram.com/p/CeFEJaFuXX1/	0.11	13.80	1.39	1.06
62	28.05.22	https://www.instagram.com/p/CeGqj-aOmXY/	0.08	7.73	0.78	0.96
63	28.05.22	https://www.instagram.com/p/CeHB62Fu1XI/	0.12	10.66	1.08	1.16
64	29.05.22	https://www.instagram.com/p/CeJOmw8uZ3j/	0.12	11.13	1.13	1.08
65	01.06.22	https://www.instagram.com/p/CeQyM-YuJRn/	0.07	6.68	0.67	1.13
66	04.06.22	https://www.instagram.com/p/CeY0TeeukHd/	0.13	10.76	1.09	1.21
67	06.06.22	https://www.instagram.com/p/CeeCQbbOzhl/	0.15	11.28	1.14	1.16
68	08.06.22	https://www.instagram.com/p/CejEsFkOALd/	0.11	9.33	0.94	1.12
69	09.06.22	https://www.instagram.com/p/Cel19NrOl9f/	0.20	11.70	1.19	1.09
70	17.06.22	https://www.instagram.com/p/Ce6fmWfJzF_/	0.05	10.11	1.02	1.14
71	23.06.22	https://www.instagram.com/p/CfKIQNKvBZH/	0.09	10.81	1.09	0.93
72	28.06.22	https://www.instagram.com/p/CfWzeMypWVs/	0.19	10.96	1.11	1.17
73	29.06.22	https://www.instagram.com/p/CfYwsf4LEWD/	0.22	12.55	1.28	1.28