University of Stuttgart
Institute of Literary Studies
Department Digital Humanities
Supervisors: Prof. Dr. Gabriel Viehhauser, Dr. Gabriella Lapesa, (PhD student) Neele Falk

University of Stuttgart
Germany

Master thesis on the subject:

Why (not) vegan? An Investigation of Moral Sentiment and Storytelling in the Vegan Discourse with Natural Language Processing

Submitted by:
Rebecca Pichler
Roßbergstraße 34
70188 Stuttgart

M.A. Digital Humanities
5th semester
st189733@stud.uni-stuttgart.de
Student number: 3453609

Submission date: 14.02.2022

TABLE OF CONTENTS

FIGURES

TABLES

1 About the Vegan Do-gooder

Veganism, according to its official definition by the Vegan Society from 1980, is a way of life that avoids animal cruelty in all aspects of life (The Vegan Society, 2021a). In addition to avoiding harm to animals, one goes vegan because of the environment or their health (Forgrieve, 2018), as this diet is claimed to be good for both aspects (*The EAT-Lancet Commission Summary Report*, 2019). Since 2005, the use of the word “vegan” has grown intensely (*Google Books Ngram Viewer*, 2022) and with “Veganuary”, a one-month challenge to try the vegan lifestyle (*About Us*, 2022), more and more people are considering eating vegan. In Germany alone, the number of vegans has grown from 850’000 in 2015 to 1,13 million in 2020 within just five years (IfD Allensbach, 2021). With a growing community, veganism is more present in society and frequently debated. In these debates, vegans are often criticized as do-gooders because they ground their beliefs and behavior in moral values. Vegans choose a “moral way of life” (Pickett, 2021, p. 18), which automatically suggests that anyone who acts or thinks differently lives immorally (Sandberg, 2019). In the context of the similar concept of vegetarianism, Minson and Monin call this effect “do-gooder derogation”, and define it as “the putting down of morally motivated others” (Minson & Monin, 2012, p. 200).

Portrayal of Veganism

To explore veganism as a phenomenon, in this work, it is represented by online discussions on articles about veganism. Of course, an online source is often easily accessible and can be rich in data. However, this does not necessarily account for a quality discourse. So why was it chosen for this thesis? In contrast to spontaneous face-to-face debates, an online discussion allows for more thorough statements, since online, time is not relevant, and people are more likely to act and react accordingly in discussions. In addition, people deliberately choose to participate in these debates – from reading an article and evaluating it to stepping into the comment section to read what others think and type down their own opinion. This alone helps to establish a high-quality contribution. Moreover, regarding these discussions, it is easy to include information sources in written statements, which helps to build more argumentative confidence for a contribution. Also, particularly concerning veganism, the supposedly moral scolding of non-vegans by their opposite minority, the “do-gooder derogation” (Minson & Monin, 2012, p. 200), likely escalates a debate. This is because this sort of scolding is a moral matter (Aquino & Reed, 2002; Blasi, 2004; Dunning, 2005; Monin & Jordan, 2009; in Minson & Monin, 2012), and because it happens face-to-face where people tend to take things more personally. This can make the entire discussion ineffective or end abruptly. Online, though, the users’ anonymity supports an open exchange where individuals speak more freely. Consequently, online discussions can be more fruitful. Furthermore, many individual viewpoints from several users are represented, which is different from if articles or opinion pieces had been chosen as a source for this work. This concurs with Qian et al.’s argument that online texts provide great coverage linguistically but also culturally (2021). Thus, the data represents highly faceted interaction(s) with higher quality rather than a one-sided monologue(s), providing dynamics crucial for the two aspects of interest in this work. Instead of using this discourse to qualitatively analyze how vegans and their motivation for their way of life are displayed (Christopher et al., 2018), for instance, or checking for vegan-specific persuasive features within the discourse (Sandberg, 2019), for example, the focus lies on moral sentiment and storytelling as well as their interaction within veganism. Both aspects are eligible in this context because they complement the "do-gooder" factor of veganism, hence working as tools to further define the phenomenon and the aspects themselves.

Dimensions of Analysis

Moral sentiment as one of the aspects incorporated in this thesis is based on the Moral Foundations Theory (Graham et al., 2013; Haidt & Graham, 2007), which works as a framework to explore morality in text. Essentially, the theory frames morality as five foundations of morality that are innate and universal to all and become more individual per person depending on their experiences and cultural upbringing. Among others, the researchers define the foundation Care/Harm, for instance, which revolves around the nurturing of others and empathy towards others' pain. The Theory has been researched a lot recently regarding several contexts. Roy and Goldwasser recently investigated whether moral beliefs play a part in politicians' stance towards certain entities (2021). Furthermore, crucial for this work was the corpus created by Hoover et al., whose research focused on the measurement of moral sentiment on the document-level (2020) instead of observing certain personality traits (Iliev et al., 2015) or conditions, like depression (Eichstaedt et al., 2018). Also, moral sentiment was observed diachronically regarding its occurrence in specific contexts like slavery (Xie et al., 2019) and as the phenomenon and its technical detection itself (Garten et al., 2016). Kobbe et al. (2020), in addition, look at moral sentiment within argument mining. A common factor in these studies is that a person's moral belief system is built upon their culture and specific life experiences (Graham et al., 2013). Despite this individuality, moral sentiment detection in texts created by users can encourage a better perspective of conflicts and support a prototypical understanding of social phenomena (Lin et al., 2018). This is because particularly when one's attitude or an ideological opinion is observed, morality forms an important factor (Lin et al., 2018). Moreover, morality, in general, is said to be closely connected to emotions (M. Nussbaum & Craig, 1998), or, according to David Hume, at least not derived from reason but sentiment (Cohon, 2018). Thus, as Graham et al. argued, an engagement with the Moral Foundations Theory can become essential in creating empathy in (non-)researchers for "those who live in a different moral matrix" (2013, p. 43). In summary, these aspects provide interesting leverage for the investigation of the Vegan Discourse. However, moral sentiment is generally difficult to investigate, since morality in text is implied rather than explicitly uttered (Kobbe et al., 2020).

To overcome this obstacle and to further explore the phenomenon of the vegan do-gooder, this work draws on the second aspect mentioned before: storytelling. Telling a narrative instead of bare facts to explain or reason something, helps to effectively process as well as retrieve complex information (Weick & Browning, 1986). Hence, with storytelling, information is schematically structured and explained (Rehm et al., 2021), guaranteeing a better understanding of one's statement in the audience. In addition, in an argumentative context of a discussion, the personal and emotional rather than rational character of storytelling is crucial. Naturally, this feature involves more people and creates awareness for different perspectives through empathy (Vecchi et al., 2021). Also, it allows interpretation and articulates what cannot be expressed rationally (Vecchi et al., 2021). It seems thus logical, that when discussing socially controversial topics, individuals express their identities and moral concerns through stories (Edgell et al., 2016). Furthermore, a study by Polletta and Lee (2006) proved that storytelling enhances people's ability to discuss efficiently by enabling them to represent and appreciate other viewpoints than their own, for example. From a more practical viewpoint, as a study object in the area of energy and climate change, researchers define stories "as data sources, as modes of inquiry, and as creative paths toward social engagement" (Moezzi et al., 2017, p. 1) that help to make people understand, to communicate, and to influence others. For instance, storytelling is researched as a learning tool for disaster prevention in the work environment of engineers (Maslen & Hayes, 2020). Furthermore, Philip N. Meyer argues that lawyers are storytellers and that a story's effectiveness determines the likely life-changing outcome of a trial (2014), proving a story's essential influence. Overall, storytelling can help to lift the veil of moral sentiment in the Vegan Discourse by interlocking with the emotional character of morality and structuring what is not observable or expressible at once.

Research Question

The synergy of the phenomenon of veganism with the two aspects moral sentiment and storytelling hence enables me to evaluate the following research question and hypotheses of this work: first and foremost, how moral is the Vegan Discourse? Assumably, there are strong indications of morality in this discourse because of the moral values underlying veganism, meaning there will be many instances with high morality scores in the data. Consequently, this gives an insight into how much morality becomes perceivable in online debates about veganism. In this context, one needs to examine which specific moral values are expressed in the Vegan Discourse and how strong or weak they are. Most likely, the foundation pairs "Care/Harm" and "Fairness/Cheating" are most prominent, as both involve a sense for the treating of others, similar to the central idea of veganism regarding animals (The Vegan Society, 2021a). Moreover, it is interesting to observe whether there is a relationship to storytelling. Does it enhance the expression of morality in text? Indeed, I expect there to be strong indications of storytelling where there are utterances with detected morality since their interplay is likely to enhance their appearance.

OPERATIONALIZATION

To find concrete answers to these research questions and to be able to draw more general assumptions, this work pursues an empirical approach. First, a corpus was created with roughly 10’000 comments on articles about veganism that were published by the New York Times. Methods from the field of natural language processing (NLP) were used to categorize the data to be able to draw conclusions from it. It was a state-of-the-art BERT multilabel model that was used for the classification of moral sentiment in the corpus. In addition, the instances in the corpus were classified as (non-)storytelling based on a model from the study “Reports of personal experiences and stories in argumentation: datasets and analysis” (2021). The BERT model is trained on the Moral Foundations Twitter Corpus (Hoover et al., 2020), a corpus consisting of about 35’000 tweets labeled for moral sentiment. As labels, one can find the five moral foundations “Care-Harm”, “Fairness-Cheating”, “Loyalty-Betrayal”, “Authority-Subversion”, and “Purity-Degradation”, as well as “Non-moral” for instances with no indications for morality. A gold standard for both domains of interest was created to validate the employed tools, which have been developed for other data sources and whose adaptation falls out of the scope of this thesis. Although these are small concerning the number of labeled instances due to the limitation of this work and a lack of resources, both datasets create room for assumptions on the models’ performances further elaborated in chapter 4.

CONTRIBUTIONS

Thus, this thesis provides future researchers with the Vegan Discourse Corpus, an automatically annotated corpus of about 10’000 comments collected from articles discussing veganism that were published by the New York Times. For moral sentiment, a fully trained BERT model is available for further research. Additionally, the two small gold standard sets for both annotation domains can be used to compare the performances of different models. In essence, the results of this investigation offer researchers a fundamental grasp on moral sentiment in a new discourse topic. Moreover, the investigation of the relationship between storytelling and moral sentiment promises a new approach to the perception of moral values in user-generated text. In the following chapters, I will deep dive into the matter of this research, starting with the background of the theoretical frameworks in this thesis. Further, I will elaborate on the corpus born from this work and on its annotation process. The core of this work with an analysis and interpretation of the automatically classified corpus follows, finished off with an additional chapter on more observations, a conclusion with final thoughts, and an outlook with indications for future work.

2 BACKGROUND

The following chapters discuss the two aspects of interest in this work regarding the phenomenon of veganism in more detail: storytelling and moral sentiment. Both aspects will be defined and further explained for the reader to fully understand the individual aspects and their interaction crucial for this thesis. In addition, there is an overview of findings by other researchers that are interesting in the context of this thesis. Overall, these chapters aim at providing full engagement in the functioning and principles of the two constructs.

2.1 Storytelling

For a deeper understanding of the concept of storytelling, a look at the semantic meaning of the word is necessary. According to the online Cambridge Dictionary (2014), storytelling is defined as “the activity of writing, telling or reading stories” or as “the art of telling stories”. This suggests that storytelling is a multimodal activity, whose reference to the arts alludes to the possession of certain skills, for instance. Furthermore, it is interesting to know about the meaning of the word story: “a description, either true or imagined, of a connected series of events” (“Story,” 2014). Hence, someone who has the skills to read, write or tell a story, is someone who can reason coherently on events either fictitious or real. This can be argued to be contributing to discussions, where it is crucial to structure one’s opinion to make others understand.

CHARACTERISTICS OF A STORY

What does a typical story look like? Polletta and Lee (2006) describe it as an expression of what the speaker wishes for, desires or feels by describing first- or second-hand experiences from their own perspective. This way, a speaker argues for what they deem personally and socially good, thus appealing to a normed version of society’s values (Polletta & Lee, 2006). This understanding connects the dots with what Weick and Browning (1986) claim in their early work, which is that by telling a narrative as an explanation or reasoning, one’s audience can understand and collect the information of what is being said more easily. This is because information is structured and explained schematically through storytelling (Rehm et al., 2021), meaning that the importance and underlying associations of separate events and actions are put together coherently (Polkinghorne & Polkinghorne, 1988; Schank & Abelson, 1995; in Maslen & Hayes, 2020). In general, it is argued that by telling narratives, individuals express what they experienced, feel and value (Polletta & Lee, 2006). Vecchi et al. (2021) also point out the personal and emotional nature of storytelling, but in an argumentative context, and claim that in not pursuing pure rationality when arguing, people become more empathetic, which makes them more aware of different viewpoints. In addition, the researchers state that arguing emotionally allows for more people to be involved and for more room for interpretation. While Schank and Abelson (1995) believe that certain aspects of a story are responsible for its sense-inducing character, a more recent approach suggests people are more susceptible to images than words when processing information in stories (Wyer et al., 2002).

THE STORY'S PURPOSE

In essence, storytelling helps to articulate what cannot be expressed rationally (Vecchi et al., 2021). In this context, it is particularly noteworthy to mention that stories are less likely to be shared when the discussion revolves around politics or technology, for instance, and more likely to be exchanged when people discuss values, as was researched on the grounds of an online consultation forum as database (Polletta & Lee, 2006). In addition, Polletta and Lee claim that when telling a story, one can discuss more efficiently because they can better represent their own viewpoints and understand those of others – they can "grasp [a story's] moral implications" (Polletta & Lee, 2006, p. 703). Hence, they report a greater likelihood for storytelling to trigger engagement of users with other users than for non-storytelling. Furthermore, Moezzi and colleagues (2017) researched storytelling in the context of energy and climate change and are convinced that people struggle less in understanding, communicating, and influencing others in a discussion when using storytelling – they define storytelling as a creative way of engaging socially, among others. In the field of engineering, for instance, storytelling is used for disaster prevention by elaborating on past events (Maslen & Hayes, 2020). Moreover, a key aspect for this work found by Edgell et al (2016) by analyzing debates on socially controversial topics is that individuals portray their identities and moral values through stories. Interestingly, lawyers are said to use stories in their daily work to defend their clients or justify accusations (Meyer, 2014), meaning they rely on the dynamics of stories to determine the outcome of a possibly life-changing trial. Assumably, they create specific identities for their clients or opponents through their stories, make them more relatable than by just naming facts, and thus, appeal to the values of the people present in the courtroom.

STORIES BEING TOLD

To sum up the concept of storytelling, imagine this simple example: because chocolate is known to be popular candy, a person who is very fond of chocolate could find it hard to accept that someone does not like chocolate. If someone argued "I don't like chocolate because it has so much sugar", they would likely be involved in a discussion around the details of sugar in chocolate. However, if they had said "I don't like chocolate because one time I ate too much and was sick for days. I could not get out of bed and the smell alone suffices to make me sick", they might encounter more empathy or at least less controversy than with the fact-based version of the argument. This example can also be transferred to the context of veganism: "I don't want to eat meat because it causes animals to suffer" is, according to the principles of storytelling, less convincing and argumentative. A more successful, in the sense of empathy-triggering, contribution would be: "I don't want to eat meat because it causes animals to suffer. Growing up as a daughter of a veterinarian, I helped to take care of all kinds of animals. I bottle-fed calves and lambs, helped during numerous dog and cat births and got to build a close relationship with animals in general by being in touch with them daily. I cannot imagine hurting one, since I learned that they are very sentient beings, just as humans are". This person draws on their personal experiences to explain why they think animals should not be hurt and illustrates it with concrete examples, which makes their contribution more comprehensible, transparent, and relatable.

Recently, the vegan advocate “Earthling Ed” who is very popular in the vegan community for his activism and campaigning for veganism, took part in a live TV debate about the question of whether veganism is a moral imperative (*LIVE TV DEBATE: Is Veganism a Moral Imperative?*, 2022). Intriguingly, when being asked how he became vegan, he mentions two aspects, both framed as stories of personal experiences. First, he explains that he became vegetarian after hearing of an accident of a truck full of chickens for meat production after which some of these chickens died, and others were hurt. At the same time, he felt sorry for these animals, he looked into his fridge and became aware that he was part of the reason why these animals were in the truck in the first place and being transported to a place where they would die anyway, which was contradictory to him if he sided against animal-cruelty. Moreover, after watching a documentary about the meat industry, he reports that his perception of his hamster “Rupert” changed significantly from a mere pet to an independent being with its own personality. This mindset spread to all animals involved in food production and made him go vegan since he became convinced that it is morally unjust to support the exploitation of animals in any way. Interestingly, the other participants in the debate also tell stories when justifying themselves for not being vegan, for instance: a participant cooked a vegan Christmas dinner for their family since one person attending this dinner was a vegan. Or another participant tells of their regular stops at a pie shop and their love of particular dishes with meat that made them not consider being vegan. Over the course of the interview, many stories are being told to justify one’s ethical choice, and it is interesting to observe this in the context of veganism. Specifically, in the end of the debate, one participant seems to become rather impatient by often interrupting the vegan activist, and the debate comes to an abrupt end. Although this is also due to a limitation in time on live television, it shows that a face-to-face discussion with clashing views on the moral topic of veganism can lead to an escalation in a debate. Overall, this live debate, already in the first three and a half minutes, proves that when talking about moral values or opinions, people like to use storytelling to position themselves and to make them more relatable.

2.2 Moral Sentiment

It is necessary to first define what is understood as being "moral" or what one knows to be a person's "morality". Again, a first look will be shed on definitions in the online Cambridge Dictionary. For the adjective "moral", one can find two definitions: 1) "relating to the standards of good or bad behavior, fairness, honesty, etc. that each person believes in, rather than to laws" ("Moral," 2014) and 2) "behaving in ways considered by most people to be correct and honest" ("Moral," 2014). Noteworthy in the first definition is that it differentiates between one's individual belief in certain standards and what is lawfully right or wrong. Consequently, something that is morally right, fair, honest, etc. according to one person is not necessarily so regarding the law. The second definition further defines the standards that people believe in as something that is shared by a majority of people. In addition, this definition describes moral behavior and, like the first definition, mentions honesty and correctness instead of good or bad for the latter. For the noun "morality", again there are two definitions published by the online Cambridge dictionary: 1) "a set of personal or social standards for good or bad behavior and character" ("Morality," 2014) and 2) "the quality of being right, honest or acceptable" ("Morality," 2014). Here, it becomes evident that the standards related to morality can be either personal or social, relate to one's behavior and character and that the term itself refers to the evaluation of the rightness, honesty, or acceptability of something or someone. In summary, according to the Cambridge dictionary, being moral and morality refer to a) the belief in personal or social standards, b) behavior that is considered good or bad, right or wrong, (dis-)honest or (un-)fair and c) the qualitative evaluation of such behavior.

A Historical Look at Morality

Additionally, according to the authors Gert and Gert in the Stanford Encyclopedia of Philosophy (2020), morality as a term can be used either descriptively or normatively. The former refers to specific standards that are shared by a society or community, like a religious one, or it is interpreted as an individual set of standards for a specific society's behavior. When most rational people follow a specific agreement under certain circumstances, the normative understanding of morality would be valid. Although both definitions are semantically close to what was found in the Cambridge Dictionary, they put forward a clearer boundary between what is personally or socially acceptable, and what is rationally shared among individuals according to the norm – "between what is and what ought to be" (Gert & Gert, 2020). In a society, moral agreements need to be specified to be perceived distinctly from the law, for instance. For a normative code of conduct to be moral, it does not suffice that all moral agents accept it, but it needs to be impartial to some extent or, for instance, enable in-group living for people. Interestingly, accepting a specific content of a descriptive moral code of conduct does not have an impact on the individual if they are not relevant to the code or not a part of the code's society. Interestingly, in the case of an individual justifying themselves or their actions to others, one relies on the most fundamental norms, making one's justification a suitable incorporation of normative morality

(Scanlon, 1982; Scanlon et al., 1998). Stephen Darwall (2009) argues similarly but emphasizes another person's authority to reason with one's justification or accountability. Both approaches to defining morality in the normative sense allude to its social character (Gert & Gert, 2020).

Jonathan Haidt (2008) enumerates many historical events and findings that show that morality is not a new topic in the area of psychology. According to him, morality is an ancient topic. He argues the first evidence of writing found in Mesopotamia to consist of moral writings, since they document debts within merchants' affairs and thus, guide towards the adherence of specific norms among merchants. Among referring to many other historical texts like the Egyptian Instructions of Amenemope or the Hindu Vedas within the context of morality, Haidt explains that the Bible's creation story of Adam and Eve, in fact, describes a story of morality. Also, he names the big philosophers Plato, Aristotle, Hume, Kant or J. S. Mill all to be involved in today's thought process of defining morality. Importantly, Haidt, just as Darwall and Scanlon argues that insights into morality can be changed or lost due to the change of societies. Haidt acknowledges many kinds of theories from renowned researchers like Durkheim's now called "hedonic treadmill" (1971), to Piaget's cognitive-development approach (1932/1965) elaborating on how to reach an understanding of how children learn rules and thus morality, to the strict father view from Lakoff (1996) and to Kohlberg's scheme to measure the development of morality in the brain (1969) among Freud's model of the mind (1900/1976) and Albert Bandura's social-cognitive theory (1991). While in 1975, the sociobiologist E. O. Wilson claimed that through evolution, a human's brain is being shaped in the way that an individual can experience moral emotions, making morality rather intuitive and suggesting that any claims about the definition of morality are mere justifications for one's intuitions, the rise of academic interest in emotions as triggers for (im-)moral behavior from 1990 onwards made room for new and different approaches to defining morality. When giving an outlook on how to define morality, Jonathan Haidt comes back to an older definition of morality as "prescriptive judgments of justice, rights, and welfare pertaining to how people ought to relate to each other" (Turiel, 1983, p. 3; in Haidt, 2008). This approach is, according to Turiel (2006), based on traditional liberal political theory. Haidt, however, came up with a definition that seeks to include all different disciplines involved in the definition process and to exclude any biases based on Western traditions or specific researchers' intellectual interests and movements:

> *"Moral systems are interlocking sets of values, practices, institutions, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible" (Haidt, 2008, p. 70).*

MORAL FOUNDATIONS THEORY

Now that the concept of morality has been discussed in relation to the world's most renowned theorists, the question arises of what these moral systems are. For this, this work draws on Haidt and colleagues' Moral Foundations Theory (MFT) (Graham et al., 2013; Haidt & Graham, 2007). According to Graham and colleagues (2013), the concept of the Moral Foundations is based on several different theories. One of them is the theory of nativism, which states that one's mind does not get shaped initially by certain experiences, but that one is born with a first draft of a structured mind that is malleable by experience. Consequently, one is born with morality, which, from time to time, is shaped by one's environment and experiences of recurrent social challenges. Moreover, these experiences are culture-specific, as the second theory of cultural learning claims on which the MFT draws. Thus, depending on their culture, a child acquires different foundations than another, although both have the same innate morality in their pre-experience structured mind. Reasonably, this aspect is why Graham and colleagues chose to name it moral "foundations". The researchers illustrate this process with an example from the Hindu culture and in reference to the Authority/Subversion foundation, where children are taught to bow in front of elders or guests or in deference to the gods. Subsequently, as adults, these children will automatically bow when being introduced to a person of authority, like for instance a well-known politician, because of their cultural-specific knowledge. A child growing up in a family without this cultural context and particular experiences will not acquire this behavioral pattern, which is certain knowledge about authority or hierarchy to "successfully navigate the moral 'matrix' he or she actually experiences" (Graham et al., 2013, p. 10). Since the foundations are not fixed but ever-developing according to one's culture-specific experiences, one cannot know about the nature and number of foundations in all societies without collecting information about these first. Hence, the MFT attempts to define what is universally available to each and every one concerning morality, and how this gets adapted depending on their culture (Graham et al., 2013). Furthermore, the theory of intuitionism plays an important role in the MFT regarding the phenomenon of moral intuition. This is, according to Haidt's Social Intuitionist Model (2001), the non-deliberative, and thus automatic evaluation of an individual's character or their actions; it is a feeling of the gut rather than a complete evaluation process justifiable by certain evidence. However, there is on-purpose moral reasoning in situations where one is socially required to reason their moral intuition, which is the essential thought of the Social Intuitionist Model. Consequently, reasoning one's moral intuition is guided by intuition and processes of affection that favor a specific outcome. The MFT aims at helping to categorize moral intuition and investigate how and why these can be different across cultures. Furthermore, it is noteworthy to state that the MFT is based on pluralism, arguing that the variety of social challenges account for an equally large number of moral foundations. Graham et al list five challenges that humans had to face and adapt to for millions of years, which thus paved the way for individuals that could solve these challenges more effectively. These challenges are thus reflected in Graham and colleagues' five foundations which can be seen in figure 1 (Graham et al., 2013, p. 58).

Foundation:	Care/ harm	Fairness/ cheating	Loyalty/ betrayal	Authority/ subversion	Sanctity/ degradation
Adaptive challenge	Protect and care for children	Reap benefits of two-way partnerships	Form cohesive coalitions	Forge beneficial relationships within hierarchies	Avoid communicable diseases
Original triggers	Suffering, distress, or neediness expressed by one's child	Cheating, cooperation, deception	Threat or challenge to group	Signs of high and low rank	Waste products, diseased people
Current triggers	Baby seals, cute cartoon characters	Marital fidelity, broken vending machines	Sports teams, nations	Bosses, respected professionals	Immigration, deviant sexuality
Character-istic emotions	Compassion for victim; anger at perpetrator	anger, gratitude, guilt	Group pride, rage at traitors	Respect, fear	Disgust
Relevant virtues	Caring, kindness	Fairness, justice, trustworthiness	Loyalty, patriotism, self-sacrifice	Obedience, deference	Temperance, chastity, piety, cleanliness

Figure 1 The original five moral foundations as defined by and taken from Graham et al. (2013, p. 58)

The authors differentiate between the five foundations Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation. According to the current version of the MFT (Dehghani et al., 2021) that elaborated further on the original one from 2013, the foundations are defined as follows:

1) Care/Harm:
   "This foundation is related to our long evolution as mammals with attachment systems and an ability to feel (and dislike) the pain of others. It underlies virtues of kindness, gentleness, and nurturance." (Dehghani et al., 2021)

2) Fairness/Cheating:
   "This foundation is related to the evolutionary process of reciprocal altruism. It generates ideas of justice, rights, and autonomy. [Note: In our original conception, Fairness included concerns about equality, which are more strongly endorsed by political liberals. However, as we reformulated the theory in 2011 based on new data, we emphasize proportionality, which is endorsed by everyone, but is more strongly endorsed by conservatives]." (Dehghani et al., 2021)

3) Loyalty/Betrayal:
"This foundation is related to our long history as tribal creatures able to form shifting coalitions. It underlies virtues of patriotism and self-sacrifice for the group. It is active anytime people feel that it's 'one for all, and all for one.'" (Dehghani et al., 2021)

4) Authority/Subversion:
"This foundation was shaped by our long primate history of hierarchical social interactions. It underlies virtues of leadership and followership, including deference to legitimate authority and respect for traditions." (Dehghani et al., 2021)

5) Sanctity/Degradation (Dehghani et al., 2021) or Purity/Degradation (Hoover et al., 2020):
"This foundation was shaped by the psychology of disgust and contamination. It underlies religious notions of striving to live in an elevated, less carnal, more noble way. It underlies the widespread idea that the body is a temple which can be desecrated by immoral activities and contaminants (an idea not unique to religious traditions)." (Dehghani et al., 2021)

EMPIRICAL APPLICATION

Graham et al sum up the MFT as "a nativist, cultural-developmentalist, intuitionist, and pluralist approach to the study of morality" (Graham et al., 2013, p. 14). According to the researchers, the theory is tested empirically by using self-report surveys, implicit measures, psychophysiological and neuroscience methods and text analysis. The latter, for instance, is closely related to this thesis in its application by using methods from computer science together with the Moral Foundations Dictionary (Graham et al., 2009). This Dictionary contains words that are labeled for specific vices and virtues of the MFT, like, for instance: "shield" as "HarmVirtue", "kill" or "war" as "HarmVice", or "wholesome" as "DegradationVirtue" and "trashy" as "DegradationVice". Interestingly, some words were captured with a reference to more than one foundation, such as "preserve" with regard to "HarmVirtue", "AuthorityVirtue" and "PurityVirtue" or "apostate" as a referant to "IngroupVice" (meaning "Betrayal"), "AuthorityVice" and "PurityVice". Moreover, some terms were labeled as indicators for "MoralityGeneral", such as "righteous", "lesson", "character" or "goodness". The Dictionary was among others used in an analysis of blogs (Dehghani et al., 2011) or, interestingly, by a digital humanist to investigate texts from the 18th century regarding the appearance of metaphors of mind (Pasanek, 2009; in Graham et al., 2013). More recent computer-scientific studies that based their work on the MFT aimed at finding out whether moral beliefs influence politicians' attitude towards certain entities (Roy & Goldwasser, 2021). Key-giving for this work was the work by Hoover and colleagues (2020), who built an entire corpus consisting of online Twitter tweets annotated for the moral foundations referred to as moral sentiment in their work (and also adapted for this work). These tweets look like this, for instance:

a) *“One of the characters of true leadership is to lead with compassion and not manipulation“* (Hoover et al., 2020, subcorpus “2016 Presidential Election”)
b) *“In order to appreciate and respect diversity we must first recognize and embrace our humanity. #AllLivesmatter #WeAreOne #BeyondFerguson“* (Hoover et al., 2020, subcorpus “All Lives Matter”)
c) *“Solidarity with #BlackLivesMatter #NationalDayOfMourning #NoThanksNoGiving pic.twitter.com/fXSbiNqmKc\ “* (Hoover et al., 2020, subcorpus “Black Lives Matter”)

In example a), the annotators decided on the moral foundations Care/Harm and Authority/Subversion, and in b), they agreed on Care/Harm and Fairness/Cheating. In addition, in example c), the annotators deemed the Loyalty/Betrayal foundation to be reflected. Overall, this annotated corpus represents well how moral sentiment is represented in online discussions. Other researchers went into the direction of investigating certain personality traits within the concept of morality (Iliev et al., 2015), or explored depression in this context (Eichstaedt et al., 2018). Xie et al (2019) performed a diachronic study regarding the appearance of moral sentiment regarding topics like slavery, while others focused more on the moral foundations and their detection with computer science methods (Garten et al., 2016). Within the field of argument mining that is related to discourse analysis and thus interesting for this work, Jonathan Kobbe and colleagues (2020) explored the appearance of moral sentiment, meaning the expression of a sort of moral attitude or moral opinion towards something.

THE THEORY’S CONTRIBUTION

However, the question remains: what benefit comes from the detection of moral sentiment in texts? Although one’s moral values are very individual mainly because of cultural learning (Graham et al., 2013), their detection can help improve the understanding of conflicts and the building of prototypes for social phenomena to better understand these (Lin et al., 2018). Lin et al (2018) also claim that moral values are enlightening to the observation of an individual’s attitude or ideology. Moreover, one’s emotions are closely connected to their moral beliefs, as Nussbaum (n.d.) argues, which is concise with David Hume’s theory that morality comes from sentiment instead of reason (Cohon, 2018). Consequently, as Graham and colleagues claim (Graham et al., 2013), the MFT can be an essential tool for the creation of empathy in (non-)academics for those with a different set of moral beliefs. A challenge, however, for the detection of morality in text is that morality is mostly not expressed explicitly but implicitly (Kobbe et al., 2020). Ming Qian and her team of researchers (2021), for instance, lately explored the concept of reading between the lines by generating a kind of hidden context synthetically with a language model. This model was trained on a large corpus of online text, which is why Qian et al argue for great coverage linguistically but also culturally.

2.3 Approach

This thesis will combine the current knowledge and methods from moral sentiment with those of storytelling in order to overcome the challenge of implicit moral content in texts. This is because the personal and structuring character of storytelling is argued to be likely to enhance the expression of moral values in text. Moreover, as Polletta and Lee (2006) argue, storytelling is a way of arguing for one's individual and social values, which appeal to the norms of society. This ties in with morality, since its concept is related to what is deemed socially good or bad, particularly in the normative or prescriptive understanding as was described in the Cambridge dictionary or by researchers of the Stanford Encyclopedia (Gert & Gert, 2020), for instance. Hence, the use of storytelling can, to some extent, be seen as an expression of one's moral values. Particularly concerning the Vegan Discourse as described in this thesis, which is based on morality and is also a personal or emotional issue, the occurrence of morality and storytelling is expected to be very frequent.

3 The Vegan Discourse

This chapter is dedicated first to the explanation of the concept of veganism by referring to its definition, its motivation and its impact. Secondly, the corpus designed to represent the Vegan Discourse in this work is described regarding its acquisition and descriptive statistics. Overall, these elaborations are meant to induce a fundamental understanding of the phenomenon under investigation in this work and to create transparency regarding its academic relevance and eligibility.

3.1 Being Vegan

What does it mean to be "vegan"? As documented by the Vegan Society in 1988, 44 years after their foundation, the final and current definition of being vegan is as follows:

> *"[…] a philosophy and way of living which seeks to exclude—as far as is possible and practicable—all forms of exploitation of, and cruelty to, animals for food, clothing or any other purpose; and by extension, promotes the development and use of animal-free alternatives for the benefit of humans, animals and the environment. In dietary terms it denotes the practice of dispensing with all products derived wholly or partly from animals."* (The Vegan Society, 2021a)

Thus, the idea is to live an animal-cruelty free life regarding not only one's diet but general consumption and way of living. If you check the use of the words "vegan", "vegetarian" and "plant-based" via the Google Books Ngram Viewer (2022), it becomes evident that the use of the word "vegan" began roughly around 1980, and increased intensely from 2005 onwards. "Vegetarian" as a term was used earlier, from 1843 onwards, and its use started to grow from about 1970. "Plant-based" was becoming popular in 1980 but did not increase as much in its frequency of usage as the other two search components. Solely from this quick analysis, it can already be deducted that the "vegan" topic became more popular over time with a rise from 2005 until today. However, the question remains: what do people eat nowadays?

The Vegan Community

According to the Ipsos MORI Global Advisor Survey in which people from 28 countries participated, the majority of people of these countries (73%) choose an omnivorous diet, meaning a diet consisting of animal and non-animal products on a regular basis (2018). 14% of the attendants are flexitarians, thus they only occasionally eat meat or fish. When it comes to a plant-based diet, the numbers lower to less than a tenth of the global population represented in this survey: 5% are vegetarians and eat all animal products except for meat and 3% are vegans that do not eat any animal products. The same amount of people considers themselves to be pescatarians, which means they eat fish but not meat. Although these numbers seem to be rather small, the vegan lifestyle is becoming more and more prominent, with Veganuary influencing more than 500'000 people in trying the vegan diet during January or longer (Finnerty, 2020). In Germany, for instance, the number of people calling themselves vegan has grown according to a survey performed from 2015 until 2020 with attendants aged 14 years or older (IfD Allensbach, 2021). While in 2015, only about 850'000 people considered themselves vegan, in 2020, this population has grown to about 1,13 million. Interestingly, this is 180'000 people more than in the year before, which shows the biggest growth over the years.

Why Vegan

What are the reasons to go vegan? Mainly, people choose the vegan diet for the animals, the planet or the personal health (Forgrieve, 2018).There is, for instance, research by Chatham House, the Royal Institute of International Affairs, that was supported by UNEP, the United Nations Environment Programme (Benton et al., 2021). They investigated how the food system impacts the loss in biodiversity. More concretely, they explored how to provide more sustainability in the food system and how to improve the state of pressured lands. Among others, the researchers concluded that a plant-based diet should be more encouraged in this context. Moreover, already in 2013, the Food and Agriculture Organization of the United Nations published an article about the essential role meat plays in the planet's greenhouse gas emissions (Food and Agriculture Organization of the United Nations, 2013; Gray, 2020). A more recent survey shows that the production of plant-based foods accounts for only half of the greenhouse gases created by the production of meat, which are, for meat, about 60% of all greenhouse gas emissions from food production (Milman, 2021; Xu et al., 2021). Being vegan for the animals means to not consume any products that contributed to or created animal suffering. As animals have been considered sentient beings since 2012 (Bekoff, 2013; Low, 2012) and have lately been even recognized so by the UK government (Department for Environment, Food & Rural Affairs & The Rt Hon Lord Goldsmith, 2021), for instance, it can be argued morally and ethically wrong to act in any way that causes them to suffer. In addition, article 13 in the treaty on the functioning of the European Union holds that animals are sentient beings and that the member states should ensure their wellbeing with respect to the state's traditions ("Consolidated Version of the Treaty on the Functioning of the European Union," 2012). However, there is enough evidence on how animals are being exploited, the causes ranging from meat or food production to the testing of cosmetics (The Vegan Society, 2021b). Living vegan is thus related to either a strong emotional attachment to animals or the general belief in the right of all sentient beings to life and freedom (The Vegan Society, 2021c). The EAT Lancet Commission Summary Report gives an overview over what needs to be done to improve the health of the planet and the world's population. In its focus is the world's food systems, as "[f]ood is the single strongest lever to optimize health and environmental sustainability on Earth" (*The EAT-Lancet Commission Summary Report*, 2019, p. 5). They claim that a more plant-based diet is healthier than a pure omnivorous diet, which could, if everyone changed their diet to a healthier one, prevent about 11 million deaths (*The EAT-Lancet Commission Summary Report*, 2019, p. 14). This already speaks for a reduction of meat in order to improve one's personal health. Moreover, a purely vegan or vegetarian diet can lower the risk of heart disease, for example, and lead to a better general health (Brown, 2020).

A MORAL MINORITY

Still, being vegan means being part of a minority based on moral accounts. Minson and Monin emphasize that “any group departing from the status quo of moral principle runs the risk of giving [the] impression [of publicly condemning the behaviour of others]” (2012, p. 200). In the context of vegetarianism, whose ideals are similar to those of veganism, the authors reason that this is because moral values are applied universally and inclusively, which complies with the elaboration in chapter 2.2 regarding the five moral foundations. Moreover, Minson and Monin argue that people do not appreciate at all to be criticized on their moral stance, as they value their moral identity a lot (Aquino & Reed, 2002; Blasi, 2004; Dunning, 2005; Monin & Jordan, 2009; in Minson & Monin, 2012). Hence, people that are not vegetarian or vegan might feel deeply offended when their moral beliefs are supposedly questioned by a morally motivated group (Monin, 2007). This key aspect together with the story behind veganism which touches many different issues is what makes the Vegan Discourse as intriguing for the investigation in this thesis.

3.2 Acquisition of the Corpus: a new Test Set for Moral Framing

The Vegan Discourse Corpus (VDC) consists of 10’432 comments on online articles and blogposts of the New York Times, all focusing on Veganism in one way or another. The New York Times is a popular US magazine, with 7.59 million people subscribed to their digital news offers (*Press Release*, 2021). The data was acquired by scraping with python via a free API offered by the New York Times after registration for an online abonnement of the magazine. Interestingly, the New York Times provides other APIs, in order to, for instance, scrape for semantic terms like people, places, or organizations (The New York Times, 2021). Besides the fact that the magazine itself is open to the analysis of their data according to their offer of APIs, it is a magazine popular for its credibility in the US (Morning Consult & The Hollywood Reporter, 2021; *Scholarly vs. Popular Sources*, 2021), making it a suitable candidate for a corpus source. As was elaborated in the introduction, the online discussion character of the data collected for this corpus provides this thesis with diverse and honest user contributions because of the internet’s accessibility and anonymity. In addition, the data was limited to the English language, as the training corpus used for this work (Hoover et al., 2020) consists of English tweets. Currently, there is no other moral foundations corpus available that suits the aim of this thesis regarding size, quality, and research area.

At the beginning of the scraping process, the URLs of several online articles were collected. These were part of an online search result list for the term “vegan” and were chosen based on the total amount and quality of the comments. The latter was checked by looking at samples and deciding whether their content was fitting for this investigation. Hence, with a simple for-loop, the scraping program iterated over the URLs of the 48 articles and stored the content of the comment section in separate files. Together, these files make up the Vegan Discourse Corpus, which distinguishes between comments and their replies to maintain internal discourse structures, although this is not the focus of this thesis. To give more insight into the articles: the article that sparked the highest amounts of comments is called “Stop Mocking Vegans” and led to a discussion with 1’493 comments, with 768 single comments and 725 replies to these single comments. The most recent article is dated 26th May 2021 (“Impossible Dumplings and Beyond Buns: Will China Buy Fake Meat?”) and the earliest one is from 26th January 2010 (“When Chocolate and Chakras Collide”). The time span thus ranges from 2010 to 2021 with a different number of articles per year. To get an insight into the corpus’ content, consider the following examples:

a) *“From the horrific animal cruelty that is the meat, dairy and egg industry, to the fact that our planet is going to hell because of these disgusting violent industries, there's simply no reason for anyone not to go vegan”* (on “Stop Mocking Vegans” (Manjoo, 2019))
b) *“Several vegan manufacturing companies are using biodegradable materials. Cork, bamboo, sugarcane pulp, pineapple leaf fibers, etc. My wife and daughter have very stylish cork purses, for example..... Are there impacts from manufacturing these? Of course, but we have chosen those over the direct cruelty (and major negative environmental impacts) related to the slaughter of billions of animals each year.”* (on “Leather? At Vegan Fashion Week, It's Pineapple Leaf” (Pajer, 2019))
c) *“I am glad to see the vegan and vegetarian diets work so well for many people. I ate a vegan diet for about 7 years and it was not good for my health. I feel so much better with a more varied diet.”* (on “No Meat, No Dairy, No Problem” (Bittman, 2011))
d) *“I’ll pass. I tried the vegan thing and it is most definitely not for me or my family. Best of luck to the restaurant and those they employ.”* (on “The New Menu at Eleven Madison Park Will Be Meatless” (Anderson & Gross, 2021))
e) *“This is a delicious dish, though takes effort to make. My kids born and growing up in USA, loved it and would call it "vegetarian chicken", because it has the texture, consistency of meat, just more delicious.”* (on “The 100-Pound Fruit That Vegans Love” (Rao, 2021b))

While comments a) and b) are assumably written by pro-vegans, comments c) and d) represent opposite opinions on veganism. Although the comments in the VDC can be context-dependent like comment d) about a restaurant going entirely plant-based, they can also be perceived independently like comment a), which states all the reasons to go vegan without referring to a particular context in the respective article. Comment e), on the other hand, represents another kind of comment in the VDC that is more neutral regarding the Vegan Discourse and mostly refers to recipes or particular types of vegan meals. Overall, the collected material consists of all kinds of opinions on veganism within different aspects, and not only surrounding the most popular debated issue of nutrition.

3.3 Descriptive Statistics

To get a more fundamental insight into the corpus, several descriptive analyses were performed. Of all 10'432 comments in the corpus, 5'094 instances are single comments, and 5'338 instances are replies to these. The range of comments per article ranges from 8 to 768. Without performing basic preprocessing like the deletion of stop words or punctuation, most of the comments (7'000) including replies contain up to 500 characters, the total range goes up to about 2'000 characters per instance, meaning per comment. Regarding the number of words, most comments (7'000) fall in the range of up to 80. Around 2'300 comments consist of roughly 80 to 150 words and the minority of the comments are wordier with 150 to 310 words. After pre-processing, thus excluding stop words, for example, most comments (6'500) consist of up to 250 characters and contain up to 40 words. Roughly 2'600 instances range in their length from 40 to 60 words, with the minority of collected comments being longer with 60 to 160 words. Very few comments are 160 to 190 words long. The average word length per comment including stop words is 2 to 9 words, excluding stop words the average word length is 2 to 10 words.

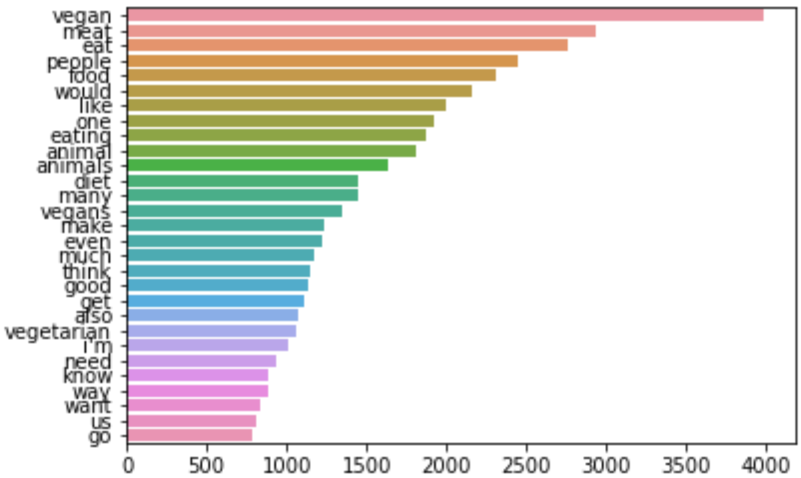

Figure 2 The most common words in the Vegan Discourse Corpus

Non-stop words that occur most frequent (see figure 2) are “vegan”, “meat” and “eat”, with the first occurring about 4'000 times, “meat” roughly 3'000 times and the latter around 2'700 times. The reference to animals is also identifiable with the occurrence of “animal” about 2'000 times and “animals” around 1'700 times. In addition, there are other verbs than “eat” that are frequent in the corpus: “like” (2'000), “make” (1'300), “think” (1'200), “need” (1'000) or “know” (900). An often-present adjective in the corpus apart from “vegan” is “good” with around 1'200 occurrences. Overall, these words can be seen as stereotypical for the Vegan Discourse because of their frequent appearance in the corpus.

Descriptive Analysis

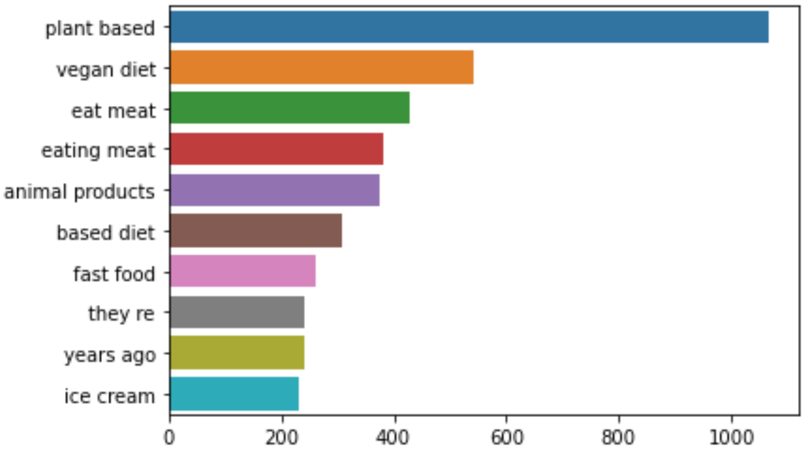

Figure 3 The most common bigrams and trigrams in the Vegan Discourse Corpus (1/2)

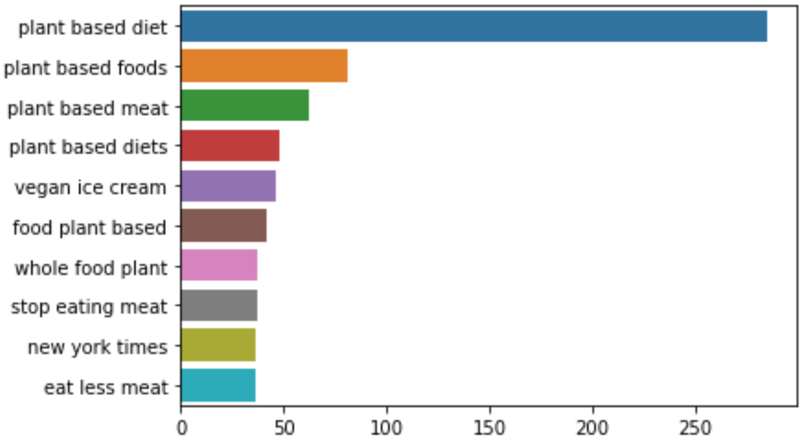

Figure 4 The most common bigrams and trigrams in the Vegan Discourse Corpus (2/2)

Moreover, the most common bigrams (see figure 3 and 4) that can be observed in the VDC are "plant based" with a little more than 1'000 occurrences followed by the term pair "vegan diet" appearing around 600 times. Also, the similar term combinations "eat meat" and "eating meat" occur frequently with around 400 occurrences, closely followed by "animal products". The four most frequent trigrams, interestingly, are all a combination of the term plant-based with another word. "Plant based diet" occurs around 290 times in the corpus, "plant based foods" can be found 80 times, "plant based meat" is 60 times in the corpus and "plant based diets" about 50 times. Other trigrams that can frequently be found are "vegan ice cream" (50), "stop eating meat" (40) or "eat less meat" (40). In general, these term combinations because of their frequent occurrence are defined as typical for the Vegan Discourse in this thesis.

In addition, topic modeling was performed on the corpus. 4 topics were identified, with two of these – in the graph and thus semantically – being close to one another and the other two being more distant. The first topic is the biggest one with 43,7% of all tokens, followed by topic two with 34,1%. Topic 3 consists of 14,2% of all tokens and is close to topic 1. Topic 4 is the smallest one with 8% of all tokens. The first and dominating topic can be described by the three most relevant terms "animal", "vegan" and "meat". It seems as if this topic is the most "general" one, meaning that there is no concrete theme detectable, but rather common terms concerning veganism are used in the comments collected on this topic. Topic 2, on the other hand, is more food-specific. The top three most relevant words in this topic are “vegan”, “food” and “meat”, others are “cheese”, “protein”, “restaurant” or “meal”, hence topic 2 represents the mainly nutrition-based discussion on veganism. The most relevant terms in topic 3 are “meat”, “people” and “one”, but, as in topic 1, no concrete theme can be identified. The same observation can be documented for the last topic, where the terms “milk”, “cow” and “one” are most relevant. Overall, topic modeling showed that if a specific theme is detectable, it is represented by topic 2 and its focus on nutrition and food. The other topics are identified with terms that are considered rather common for the Vegan Discourse and thus, the Vegan Discourse as portrayed in this thesis is not entirely exclusive to certain topics of discussion.

Moreover, the corpus was analyzed regarding sentiment with the annotator of Stanford CoreNLP. Most of the corpus’ instances are neutral, in total 47,9%. In addition, the instances are slightly more negative compared to positive, with 26,6% of all instances being negative and 24,3% being positive. Moreover, the CoreNLP annotator labeled 1,0% of all instances in the VDC as “Very positive” and 0,1% as “Very negative”. Overall, it can be documented that although the corpus is mostly neutral in sentiment, it is also rather negative than positive.

4 Annotation Process

This chapter provides an insight into the process of automatically annotating the Vegan Discourse Corpus for the five moral foundations as well as storytelling. The two models used for this process will be explained briefly in chapter 4.1 and 4.2 to ensure comprehensiveness. In this context, I will also elaborate on the performance of the models based on the commonly used metrics like F-Score, Recall and Precision, and make assumptions about their importance. In addition, two gold standard sets for both annotation domains were created to complete the scientific cycle (see chapter 4.3 and 4.4), although these are rather small due to the limitation of volume in this work and a lack of resources. Indeed, though, both standards make it possible to draw larger generalizations and to fundamentally assess the performance of both models, which is elaborated in chapter 4.5. Their adaptation is out of the scope of this thesis but leaves space for future research.

4.1 Computational Annotation of Moral Foundations in the Vegan Discourse

In the case of this thesis, the training corpus is the Moral Foundations Twitter Corpus (MFTC) created by Hoover and colleagues (2020). It consists of roughly 35’000 English tweets from discussions about different social events or trends. The authors chose to use controversial topic areas and came up with these discourse domains: “All Lives Matter”, “Black Lives Matter”, “Baltimore Protests”, “2016 Presidential Election”, “#MeToo”, “Hurricane Sandy” and a corpus by Davidson and colleagues (2017), consisting of tweets with hate speech and offensive language. The MFTC is labeled for all five moral foundations separately for vice and virtue per foundation. However, in this thesis the labeling approach of vice-virtue pairs by Kobbe and colleagues (2020) was pursued. They aim at keeping the labels as semantically clear as possible and argue that the individual vice and virtue pairs are sometimes hard to distinguish during annotation. Thus, if there was an indication for the Care virtue in an instance, it would be labeled as “Care/Harm” instead of only “Care”. In addition, there is a sixth label called “Non-moral” for instances where the original annotators of the MFTC did not find indications for any kind of morality. It is also important to mention that the labels are not annotated exclusively but that an instance can carry more than one label, except for the Non-moral label, which is mutually exclusive on a conceptual level.

Accordingly, the model trained for this thesis stems from a state-of-the-art BERT multilabel model. BERT stands for Bidirectional Encoder Representations from Transformers and was first published by Google, more concretely from researchers of their AI Language department (Devlin et al., 2018). According to its creators, the pre-trained model works with deep bidirectional representations and its particularity in comparison with other language models is that it can consider a word’s entire context, meaning left and right, jointly. Hence, Devlin and colleagues (2018) argue that one can derive powerful models capable for tasks like language inference or question answering, with less amount of effort in fine-tuning or adaptations for the purpose at hand. For the annotation of the five moral foundations plus the non-moral category, the model was built to deal with multiple labels similar as in the MFTC, meaning it assigns scores for each label instead of just for one per instance. Importantly, the MFTC was pre-processed in a way that links or the mentioning of particular usernames were disregarded by replacing the former with "URL" and the latter with "USER".

EVALUATION ON THE MFTC CORPUS IN KOBBE AND COLLEAGUES (2020)

The evaluation of the model in this work on the original data in the training corpus presented in the following subsections works as a plausibility check of the model's functioning. First, a look at the general distribution of labels in the training corpus will be discussed. Subsequently, evaluative information about the MFTC as a training corpus in the work by Kobbe et al. (2020) will be provided, since this work's approach to classify for moral sentiment with a multi-label BERT model based on the MFTC is strongly oriented on their work. Moreover, evaluation results from Kobbe and colleagues (2020) on their test data will be presented for further insight, and further compared to the results in the gold standard of the VDC (see section *Defining a Threshold*).

To obtain a better grasp of the MFTC and thus any insights on the model's evaluation, a description of Hoover and colleagues' documentation of the amount of vices and virtues annotated per sub-corpora (2020, p. 10, Table 1) will be provided: there are 11'938 instances labeled with a vice or virtue, while 23'170 instances are annotated as Non-moral. In contrast, the Purity/Degradation label is less represented, which can account for lower evaluation scores for this foundation, since a classifier may not have had the chance to learn sufficiently about this foundation. Indeed, besides Subversion, the vice and virtue Purity and Degradation are the two labels annotated the least in the MFTC. Summed up for all sub-corpora that make up the MFTC, Purity was annotated 3'203 times (~ 9,1% of the entire corpus), Degradation 6'127 times (~17,5%), and Subversion 5780 times (~ 16,5%). Hence, one could expect lower evaluation scores for the Subversion foundation, as well. Cheating is very frequent with 8'193 occurrences (~ 0,23%), as well as Harm with 8'185 occurrences (~23,3%) – both are better represented in the MFTC and thus, could correspondingly be assigned more often by a classifier.

Apart from the multi-label BERT model, the Kobbe and colleagues (2020) created two baselines, one random but aware of the distribution in the training data and one based on the Moral Foundations Dictionary (Graham et al., 2009). The Dictionary was also used in combination with three other approaches to differentiate between the sense of specific words. For the first method, they started off by linking entries in the Moral Foundations Dictionary to entries in WordNet (Princeton University, 2010) and created their own lexicon with 61 synsets on average per moral foundation (WN-PPR). Further, they extended the entries in this lexicon with relations such as "similar to" and hypernyms from WordNet. Moreover, Kobbe and colleagues decided to extract features from text in combination with WordNet. Another approach the authors pursued was to generate text that is likely to integrate moral sentiment by using a Sentence-BERT (SBERT)model trained on 317 Wikipedia abstracts manually annotated for the moral foundations. The ready-to-use model was implemented to obtain new texts based on argumentation datasets that are likely to include moral sentiment. Overall, the researchers implemented two baselines and a multi-label classifier built on BERT, the WN-PPR approach together with a k-nearest neighbor classifier, and, in collaboration with a Linear Discriminant Analysis, they implemented an SBERT-Base method, the SBERT-Wiki approach trained on the manually annotated Wikipedia data as well as the combination of SBERT-Wiki and WN-PPR. Note that Kobbe and colleagues differentiate between the five moral foundations and "Moral" or "General Morality" for cases where none of the five moral foundations, but a general moral concept can be detected, for instance with specific terms like "ethic" or "evil". The Binary-F1-scores for the positive class for each moral foundation can be seen in table 1 (Kobbe et al., 2020, p. 35).

Method	Moral	Care	Fairness	Loyalty	Authority	Purity	Average (excl. Moral)
Random baseline	.519	.173	.169	.100	.099	.055	.119
MFD baseline	.630	.332	.213	.166	.231	.141	.217
multi-label BERT	.669	**.510**	**.573**	**.437**	**.377**	**.363**	**.452**
WN-PPR	.628	.334	.379	.311	.210	.088	.264
SBERT-Base	.685	.434	.511	.372	.327	.214	.372
SBERT-Wiki	**.697**	.463	.516	.377	.341	.220	.383
WN-PPR + SBERT-Wiki	.689	.446	.520	.387	.346	.230	.386

Table 1 „Binary F1-scores on the MFTC for individual MFs (F1 for the positive class). The last column shows the average over the F1 scores for the five MFs (excluding Moral)" (Kobbe et al., 2020, p. 35)

Except for "Moral" where the best performer is the SBERT-Wiki approach, the multi-label BERT method yields the best performance results considering the moral foundations. Closest to this with regard to performance are the methods based on SBERT-Wiki and the combination of the two methods WN-PPR and SBERT-Wiki, the latter also being better in performance than the random baseline. To further evaluate on Kobbe and colleagues' approaches, consider table 2 which captures their performance on the chosen argument quality corpus as a test set.

Method	Moral	Care	Fairness	Loyalty	Authority	Purity	Average (excl. Moral)
Random baseline	.658	.257	.179	.096	.134	.105	.154
MFD baseline	**.853**	.056	.237	.043	.200	.086	.124
multi-label BERT	.444	**.517**	**.519**	**.138**	.157	.208	**.308**
WN-PPR	.756	.118	.253	.049	.105	.029	.111
SBERT-Base	.703	.280	.342	.065	.148	.133	.194
SBERT-Wiki	.730	.339	.246	.125	**.233**	**.318**	.252
WN-PPR + SBERT-Wiki	.686	.298	.351	.067	.040	.135	.178

Table 2 „Binary F1-scores on the Dagstuhl ArgQuality Corpus for individual MFs (F1 for the pos. class)" (Kobbe et al., 2020, p. 35)

While the multi-label BERT still performs best on average and in the majority of the labels, the SBERT-Wiki approach yields better results for the foundation Authority and Purity. The baseline based on the Moral Foundations dictionary outperforms regarding the Moral label but this is because of the majority distribution of this label in the training corpus. The approach that combines the WN-PPR method and the SBERT-Wiki one performs partially worse than the random baseline and also worse than the SBERT-Wiki.

<u>EVALUATION ON THE MFTC CORPUS IN THIS WORK</u>

To have a deeper understanding of the model and classifier of this work and to better control its outcome, new code was used to train a classifier. During the process of finding the best performing method, certain parameters were adapted: the training, test and validation set size as well as the number of epochs. Importantly, it was also made sure that there were no duplicates in the test or validation set to not falsify the performance results. In the end, the best model came from a test-train-validation split of 90% for training, and 5% each for validation and testing and an iteration cycle of 3 epochs. Similar to the evaluation results presented by Kobbe et al. (2020), the F1-Scores for the positive class per foundation based on the classification results of the multi-label BERT model in this thesis can be found in table 3. The results for the positive class are interesting to explore, since this class represents the minority in the classification and thus, indicates what is most difficult to predict for the model and thus, gives an insight into the actual classification ability of the classifier. Note that for the calculation of these values, the thresholds which were calculated based on the created gold standard (see chapter 4.3, *Defining a Threshold*) were applied to each foundation respectively. In addition, a majority baseline adapted to the majority class assigned per label in the MFTC as well as a random baseline filled with random assignments per class were created to be able to compare results.

	Care/ Harm	Fairness/ Cheating	Loyalty/ Betrayal	Authority/ Subversion	Purity/ Degradation	Non-moral	average (excl. Non-moral)
multi-label BERT	0,85	0,87	0,45	0,75	0,78	0,36	0,74
Majority baseline	0	0	0	0	0	0	0
Random baseline	0,28	0,27	0,18	0,19	0,13	0,46	0,21

Table 3 Binary F1-scores on the MFTC training set for the positive class for individual moral foundations

Since the majority baseline is always negative (0), the F1-Score for the positive class equals 0 for all labels, since there is no overlap between the model's classifications of the positive class. The F1-Scores of the random baseline show that the model works better than chance – it does not make lucky guesses – , since all evaluation scores for the model are considerably higher than those of the random baseline, except for the Non-moral label. Here, the model works slightly worse than chance; however, one has to consider the high threshold calculated for this label, which makes it to draw conclusions. The average F1-Score for all foundations (excluding the Non-moral label) reaches a value of 0,74 and indicates good performance of the model on the MFTC. Compared to the results by Kobbe and colleagues (2020), it is evident that the scores overall are higher, which is also reflected in the average score. Except for the Loyalty/Betrayal foundation which has a similar F1-score as in Kobbe et al. (2020), all F1-scores calculated based on the model in this work are higher than those in the reference work. Overall, the comparison of the classifier used in this thesis with the random baseline provides sufficient proof that reasonable results can be expected and thus, these results can be involved for a larger generalization. The majority baseline did not reveal any insights for the positive class.

4.2 Computational Annotation of Storytelling in the Vegan Discourse Corpus

For the annotation of storytelling in the Vegan Discourse Corpus, this thesis was provided a ready-to-run model from a recent study exploring the automatic annotation of storytelling in arguments (*Reports of Personal Experiences and Stories in Argumentation: Datasets and Analysis*, 2021). Essentially, the researchers combine the concept's understanding from two domains – Argument Mining and Deliberative Theory – and, with modelling experiments on different data sets from these two domains, prove that annotation approaches from both domains, can successfully be used together. Notably, the researchers use the label "report" for all instances of storytelling in the three data sets, since the two fields differ between storytelling (Deliberative Theory) and testimony (Argument Mining). Moreover, the authors of the study showed that performance varies significantly depending on the chosen classifier and the interaction between training setup and test corpus.

The model that was used for this thesis is a BERT model trained on a mixture of all the datasets used in the study, since it was the best model according to the evaluation results. Due to limitation in volume and resources, one can assume from these facts that the model would perform well on the VDC. Still, in chapter 4.4, a small gold standard for storytelling will be discussed to get an indication of how well the model generalizes on the data of the corpus at hand of this thesis.

4.3 Gold Standard – Manual Annotation of Moral Foundations in the Vegan Discourse Corpus

The decision on how to annotate for a gold standard of moral foundations in the Vegan Discourse Corpus was accompanied by a thorough thought process on how to detect the foundations in text. One could either work with explicit signs of morality, like particular words, as described in the Moral Foundations Dictionary (Dehghani et al., 2021). Or it is possible to train annotators in detecting implicitness of morality, meaning, for instance, what the author of a text wanted to say with their statement but did not explicitly verbalize.

GUIDELINE APPROACHES

The researchers Kobbe et al. (2020) work with arguments in this context and focus their annotations for a gold standard on real-world actions or attitudes of morality. Thus, they always check for an agent in the argument and if this agent shows an attitude towards or acts in a way that promotes a moral foundation. This allows them to detect morality even when it was implied. The authors argue that the observation on a word-level, meaning checking for the frequency of words of the Moral Foundations Dictionary in a text, does not suffice to make legitimate claims about the occurrence of a specific moral foundation. Moreover, if, in a document, someone refers to a topic related to equality, for instance, they annotate it as the foundation Fairness/Cheating only if there is an action, attitude or evaluation of this foundation expressed. Hence, they do not annotate a foundation solely based on a reference to a certain topic. The authors and creators of the Moral Foundations Twitter Corpus Hoover et al. (2020), however, focus on explicit signs of moral sentiment for their manual annotation of the corpus, since it is not always clear what the author of a document wants to express. However, they mention that this can lead to the loss of nuances within an individual's language and morality. Therefore, the authors suggest a balanced method between implicit and explicit coding but, at the same time, point out how difficult this is to achieve. Still, they assume that if all annotators were aware of both ways to code, recognizing the implicitness and explicitness of morality, annotator bias could be reduced. Overall, during annotation, they focus on explicitly stated signs of moral sentiment and include inferences of an author's implied intention(s) only when they can be well justified (Hoover et al., 2020).

MANUALLY ANNOTATING THE VEGAN DISCOURSE CORPUS

In this work, the approach by Kobbe and colleagues (2020) to put a larger focus on the implicitness of the moral foundations was chosen to annotate the Vegan Discourse Corpus, as I am convinced it combats well the subtlety of morality in text but also captures when morality is explicitly stated. Kobbe and colleagues define two questions that guide towards the detection of a moral foundation, either explicitly or implicitly articulated, in text: 1) "Who is the agent in the text under consideration (if any)?" and 2) "Does the agent express an attitude / perform an action that promotes the [Moral Foundation] in question?" (2020, p. 4). In the following, these questions together with the following two comments taken from the VDC will be considered:

a) *"While I wouldn't argue that there is a circle of life and that violence is sometimes inevitable, this doesn't give us carte blanche on killing. If you don't have to kill something/somebody to survive, morality would tell us not to."* (on "When Vegans Won't compromise" (Fischer & McWilliams, 2015))
b) *"Vegetarians need to eat healthy too? Duh. The environmental benefits of cutting out or significantly reducing the amount of meat we eat are great."* (on "Good Vegan, Bad Vegan" (Brody, 2021))

In example a), one can clearly observe several explicit expressions related to the Care/Harm foundation, like "killing", "kill" or "violence". In addition, the word "morality" is mentioned. According to Hoover and colleagues, this would suffice to annotate for the Care/Harm foundation. When checking for Kobbe and colleagues' questions, one could argue that there is 1) an I-agent expressed in the text and 2) this agent expresses their attitude towards, in their opinion, unnecessary killing, which relates to the Care/Harm foundation. Thus, Kobbe and colleagues would likely annotate this foundation for this comment. In example b), however, there are no explicit signs in the form of explicit terms of a moral foundation in the text. Consequently, according to Hoover et al.'s annotation guideline, the label "Non-moral" would need to be assigned. According to the guideline by Kobbe et al., however, one would likely decide differently after checking the questions: 1) the comment's author mentions a "we" as a referral to "us humans" and 2) the second part of the statement can be understood as an attitude towards the consumption of meat being harmful to the environment. Hence, one could argue that there is a relation to the Care/Harm foundation. Overall, this brief explanation illustrates that the guidelines by Kobbe and colleagues work well for the detection of moral sentiment, either implied or explicitly stated, which is why they were used for the creation of the moral foundations gold standard in this thesis. Their guidelines describe six labels. Five of them are the moral foundations which are annotated for their meaning as is currently claimed by Deghani and colleagues (2021)., which was also elaborated in chapter 2.2. The sixth label is "NONE", which is assigned in the case when the agent of a statement does not "express an attitude / perform an action that promotes the [Moral Foundation] in question" (Rehbein & Kobbe, 2020, p. 4). In a comment like "I cooked this vegan recipe with my friend and we both liked it", the agent does not promote a certain moral foundation by referring to a certain concept, but merely states a fact. According to the definition of NONE, one could assume that the label equals the label Non-moral in the MFTC. However, the absence of a moral foundation does not necessarily result in the "presence" of non-morality as defined for the Non-moral label. Still, an observation of the behavior of the model with regard to the Non-moral label in the entire corpus and the assignment of the NONE label in the gold standard by the human annotators provided sufficient proof to treat both labels equally in their definition. For the other labels as well, this will be elaborated further with respect to the F1-score for the positive class in the subsection *Defining a Threshold*. Moreover, several examples of the VDC for the Non-moral label are illustrated and discussed in the respective section in chapter 5.1, which facilitates a clearer understanding of the NONE/Non-moral label treated interchangeably.

THE CREATION OF A GOLD STANDARD

Hence, 100 comments of the VDC were randomly, but conditionally, sampled. Per foundation, it was made sure that there are ten comments with a model-assigned probability score greater than 0,5 and the same number of comments with a score lower than this threshold[1]. Since the

[1] The probabilities declared by the model for the comments in the gold standard stem from an older version of the model. The analysis in chapter 4 and 5 is based on the final model and therefore the probabilities assigned by the model for the comments in the gold standard deviate from the ones used during the sampling process.

Authority/Subversion foundation was not classified with a score greater than 0,5 more than two times, this specific foundation was sampled entirely randomly. Three annotators provided with the guidelines by Kobbe and colleagues (2020) worked through these comments, where the majority of annotated labels was assigned as gold. The annotators did not have to choose one of the foundations, but could, just like the model, assign multiple labels for one instance. Besides the five foundations, there was the “NONE” label for instances with no indication of a specific moral foundation. While this does not necessarily account for non-morality in a text, it will be treated as such in this thesis, which is further elaborated in the next chapter.

To reflect the challenge of morality to not always be clearly stated, several approaches for the creation of a gold standard were pursued: to capture the inclusiveness of the foundations, approach a) gold_lax includes a moral foundation as gold if at least one annotator assigned it to a comment. This guarantees any weaker indications of morality detected by the human annotators to also be represented in the gold standard and thus, creates a more “realistic” gold standard. In addition, approach b) gold_strict follows a more classic method of taking a label as gold if two of the three annotators assigned it and else, have no label represented as gold. This produces a gold standard representative of the most dominant labels detected by the human annotators. Overall, since the gold_lax set is representative of all possible indications of morality detected by the human annotators, it consequently somewhat impedes the distinction of the moral foundations per se. Thus, and because of the agreement among annotators on (a) label(s) as gold in approach b) that serve as an indicator of reliability and conciseness of the gold standard, the gold_strict set will serve as a point of orientation for further evaluation in this work. Due to imitation in volume, other approaches like taking a label as gold if all three annotators assigned it to a comment or comparing the model’s predictions with each annotator’s annotations as gold independently cannot be tested – still, these seem eligible in this context and are promising for future work.

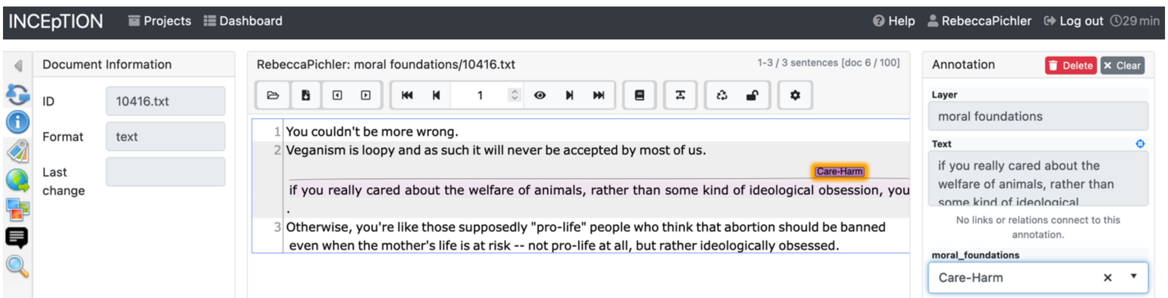

Figure 5 A screenshot of the intelligent annotation platform INCEpTION (Klie et al., 2018) during annotation

The annotation environment (see figure 5) used for the manual annotation of the 100 sampled comments from the VDC is INCEpTION (Klie et al., 2018), a semantic annotation platform that offers intelligent assistance of the manual annotation task and knowledge management. Hence, of all comments in the gold_strict set, 40 received the NONE label and 10 were annotated for Care/Harm. The Care/Harm foundation was co-assigned two times each with

Authority/Subversion and with Fairness/Cheating, three times with the foundation Purity/Degradation and once with the Loyalty/Betrayal foundation. It is the foundation that occurs most and most commonly with co-occurring foundations. The foundations Authority/Subversion, Fairness/Cheating and Purity/Degradation were each assigned five times independently. The Authority/Subversion foundation, in addition, is assigned once together with the Fairness/Cheating foundation. The least occurring foundation is the Loyalty/Betrayal foundation with 2 independent assignments in total. Table 4 shows the averaged Cohen's Kappa score among annotators as well as the Cohen's Kappa score for the model and the two different gold standard sets. Note that in the calculation, the thresholds as defined in the subsection *Defining a Threshold* in this chapter were applied to provide the calculation with a robust basis for the differentiation of negative and positive cases.

	Care/ Harm	Fairness/ Cheating	Loyalty/ Betrayal	Authority/ Subversion	Purity/ Degradation	NONE
Among annotators	0,2821	0,2203	0,1892	0,1912	0,017	0,5484
Model-gold_strict	0,6807	0,8072	0,1685	0,7839	0,3363	0,6352
Model-gold_lax	0,6542	0,4575	0,3112	0,559	0,2597	0,5261

Table 4 Cohen's Kappa Scores among annotators (Average Kappa) and Model-Gold Standard

Given the limitations of this thesis in resources and volume with respect to the gold standard, the scores for the foundations among annotators is reasonable. The best-scoring foundation is Care/Harm, the least-scoring is Purity/Degradation. The scores for Cohen's Kappa for the model and the gold standard sets are reasonably well, too, and indicate an overlap between the model's classification results and the gold standard sets. Except for the foundations Purity/Degradation and Loyalty/Betrayal, the labels received values greater than 0,45 for Cohen's Kappa, which accounts for weak to moderate agreement. Particularly, the Fairness/Cheating foundation with a Kappa score of 0,8072 represents a strong agreement between the model and the gold_strict set. Notably, the gold_lax set together with the model produced similar values for the labels Care/Harm, NONE and Purity/Degradation as the gold_strict set, but for Fairness/Cheating, Loyalty/Betrayal and Authority/Subversion, these differ considerably. To further illustrate the annotation process and the resulting correlation scores, consider the following annotated examples taken from the gold_strict set:

a) *"I started a plant-based way of eating some years ago as an alternative to taking Lipitor for the rest of my life. So far it has done the trick and I plan to continue. I don't feel better than anyone else, and I don't think more people should eat the way I do. I made a personal choice and I'm glad about it."* (on "Stop Mocking Vegans" (Manjoo, 2019))

b) *"I don't think vegans have it rough because of the guilt feelings or defensiveness of others. It's just that it's very difficult to find a restaurant (or a home cooked dinner) that vegans and omnivores will both enjoy. So when a vegan friend tells me that he or she can't join me at a* *restaurant because there will be nothing for him or her to eat there, I don't find it "preachy" but I do find it to be a request that I don't want to honor."* (on "Stop Mocking Vegans" (Manjoo, 2019))

c) *"Sooner or later it will dawn on this country how stupid it really is to respect anyone's tenets of faith in legislation applicable to all. Politics has been mired in this idiocy my whole lifetime."* (on "Satanists, Vegans and Atheists Seek Equal Opportunity in Little Rock (Clines, 2015))

Comment a) was annotated differently per annotator, with annotator 1 assigning NONE, annotator 2 Loyalty/Betrayal and annotator 3 Fairness/Cheating. Moreover, comment b) received the label NONE by one annotator, the labels Loyalty/Betrayal and Care/Harm by the second annotator and the label Fairness/Cheating by the third. Similarly, comment c) represents another case where all annotators assigned differently: annotator 1 found an indication for Authority/Subversion, annotator 2 for Loyalty/Betrayal and annotator 3 for Fairness/Cheating. Although all annotators did not undergo extensive training for annotation which could result in different assignments, these cases can also illustrate how differently the moral foundations are perceived by individuals.

Correlation Between Model Prediction and Annotator Confidence

To further elaborate on the performance of the model in comparison to the gold standard, I evaluate if there is correlation between the agreement of annotators and the model's predictions by calculating the Pearson Correlation Coefficient per foundation and annotator confidence. The idea is that if all three annotators agree on a label, the model predicts a high score for a label. Thus, if all three annotators do not agree on a label and no label can be assigned, the model predicts equally indecisive. Therefore, for each moral foundation, I checked the number of annotators that agreed on it to get a sense of annotator confidence. This is encoded to be represented in the data as follows: a 1 for cases where there is no majority for any label of the moral foundations and thus, no label is assigned. Cases when two annotators agree on a label are encoded with a 2, and in cases where all annotators agree on the same label, a 3 is assigned as label. Thus, for comments with an encoding of a 3 for annotator confidence, I would expect higher probability scores for the gold label of this comment. Comments with an encoding of a 1 or 2 for annotator confidence would concurrently be expected to carry lower probability scores.

Indeed, for three of the six labels, results for an indication for a positive or negative correlation as described can be reported. The Care/Harm foundation correlates positively with the confidence among annotators with a score of 0,23. Hence, one could argue that if there is higher agreement among annotators, the model predicts with a higher probability for this foundation. In addition, the Authority/Subversion foundation as well as the NONE label received negative correlation values: in the former case -0,14 and in the latter -0,15. Arguably, this means that if there is less agreement among annotators, meaning less annotator confidence, for these two labels, the model tends to predict higher probability scores for these two labels.

<u>DEFINING A THRESHOLD</u>

Assumably, the distribution of predicted labels in the VDC is different to the distribution of the labels in the MFTC, although the latter was used to train and evaluate the classifier. This is why the distribution of the predicted labels by the model in the instances of the gold standard (gold_strict) will be examined. Therefore, I explored several thresholds applied on the predictions of the model that mark the definite presence or absence of a particular label within the gold standard set of gold_strict. Hence, per foundation, the best performing threshold with regard to the F1-Score of the positive instances, meaning the comments that received a probability score greater than the set threshold, was calculated based on the manual annotations by the human annotators in gold_strict (see table 5) This lead to six different thresholds for each label: Care/Harm (0,46), Fairness/Cheating (0,44), Loyalty/Betrayal (0,05), Authority/Subversion (0,59), Purity/Degradation (0,27) and NONE (0,98). These thresholds indicate the point where the predictions by the model and the annotations of the gold standard harmonized most and are thus a reasonable tool to evaluate further on the distributions of the labels in the VDC.

	Care/Harm	Fairness/Cheating	Loyalty/Betrayal	Authority/Subversion	Purity/Degradation	NONE
0	0,92	0,98	0,92	0,98	0,93	0,87
1	0,76	0,82	0,21	0,8	0,4	0,77

Table 5 F1-Score (negative and positive class) per foundation and respective threshold for model predictions on the gold_strict set

Notably, the threshold set for the Loyalty/Betrayal foundation is very low, and accordingly, the F1-Score for the positive instances is the lowest among all foundations. This indicates that the model had difficulties detecting the foundation. Since overall, the Loyalty/Betrayal foundation was only captured three times in the gold_strict set, this threshold cannot be seen as considerably representative. Still, for consistency and to be concise with the evaluation results, this threshold will be considered when analyzing the distribution in the VDC. The best F1-Score for positive instances was calculated for the Fairness/Cheating foundation, which shows that at a threshold of 0,44, the model assigned positive cases for the foundation very similar to the gold standard set.

For further evaluation of the model's performance on the VDC, table 6 provides a comparison of the F1-Score of the positive class of the model and the two baselines, majority and random. Conspicuously, on the 100 comments taken from the VDC, the model performs slightly worse than chance concerning the Loyalty/Betrayal foundation. However, this foundation was assigned for only 3% of the gold_strict set, which does not allow for a clear interpretation of this. In all other cases, the classifier performs considerably better than chance – which indicates reasonable performance of the model on the comments in the VDC. This is particularly interesting, since the instances in the MFTC do not resemble those of the VDC significantly. Compared to the model's performance on the MFTC, it can be stated that the positive class of all foundations are similarly well detected by the model, except for the Loyalty/Betrayal foundation and the Purity/Degradation foundation. This could result from the small representation of both foundations – Purity/Degradation is assigned five times independently, three times co-occurring in gold_strict. Compared to Kobbe et al.'s application of their multi-label BERT model on their test corpus, it is noteworthy that the Fairness/Cheating and the Care/Harm foundation received the best F1-Scores, which is similar to the high scores assigned in the case of the model of this work. Interestingly, however, the Authority/Subversion foundation received a value equally low to the Purity/Degradation and Loyalty/Betrayal foundation in Kobbe and colleagues' test set, whereas the foundation is among the best scored ones regarding F1 concerning the model of this work. The two low scored foundations in Kobbe et al.'s case illustrate similar results as in this work. Overall, I argue for a considerably good performance of the model trained on the MFTC and applied to the VDC in this corpus based on the evaluation of the model on the MFTC and the gold_strict set and the comparison to Kobbe et al.'s results. The results allow me to draw larger generalizations, although the two lower scored foundations will be considered as marked cases.

	Care/Harm	Fairness/Cheating	Loyalty/Betrayal	Authority/Subversion	Purity/Degradation	NONE
multi-label BERT	0,76	0,82	0,21	0,8	0,4	0,77
majority baseline	0	0	0	0	0	0
random baseline	0,54	0,37	0,29	0,35	0,51	0,43

Table 6 Binary F1-scores on the gold_strict set for the positive class for individual moral foundations

Defining Non-Morality

In addition to the thresholds per foundation, this thesis explores another approach to define a lower boundary for probability scores, below which there is definite non-morality, to allow a contrastive analysis of weak and high probability scores, meaning moral loadings, within the classified instances. The most assigned label in the gold_strict set is NONE. It is also the only label that works exclusively, meaning it cannot co-occur with a moral foundation – contrastively, foundations are not exclusive and can co-occur with each other. In addition, the NONE label received a considerably well Cohen's kappa score, which indicates moderate agreement among annotators for this label, hence making it more reliable in its importance. These characteristics, particularly the exclusiveness of the label, provide reasonable ground to argue for the NONE label to give a more comprehensive directive to explore the distribution of morality across all labels. Hence, to define a lower boundary below which there is absolute non-morality and above which one can argue for weak morality scores, I will use the NONE label assigned by the annotators in the gold standard (gold_strict) as a point of orientation. Thus, the dataset of the gold standard was filtered for all comments that received the NONE label as gold, and the average of all probability scores assigned by the model for all the NONE labeled instances was calculated per label. This was done to achieve one score per label that represents the model's classification results for all NONE comments as declared by the gold standard. These values are depicted in table 7:

	Care/ Harm	Fairness/ Cheating	Loyalty/ Betrayal	Authority/ Subversion	Purity/ Degradation	Non-moral
positive prediction	0,022767832	0,056860107	0,018721762	0,026185432	0,030850088	0,915812681
negative prediction	0,977232168	0,943139893	0,981278238	0,973814568	0,969149912	0,084187319

Table 7 The average probability scores per label for the NONE comments of the gold standard

It is evident that the average probability scores for the moral foundations are very low compared to those of the Non-moral label, which accounts for the absence of any moral foundations, and the "presence" of non-morality in these instances. This particular observation allows, in this thesis, to argue that the Non-moral and the NONE label both assign non-morality for an instance. In addition, overall, these average scores show that the model classifies reasonably in this regard – where there is non-morality, there should not be morality assigned for an instance. Hence, in the following, NONE and Non-moral can be understood interchangeably. Nevertheless, in order to receive one score that can be used as a boundary, these resulting probability scores were again averaged. This defined the boundary within this approach precisely at 0,178532984, and roughly at 0,18.

This approach assumes that for all instances that received the NONE label in the gold standard, the probability scores for the moral foundations assigned by the model are not meaningful because the human annotators did not detect their expression in the respective instances. Hence, the model's classification is not meaningful with values up to 0,18 regarding the classification of moral foundations, meaning any probability scores up to 0,18 for a moral foundation do not indicate such. Regarding the scores for the Non-moral label in this approach, I will argue that scores above 0,18 are only significant if there is no moral foundation assigned for the same instance, since its assignment simultaneously to a moral foundation is only significant below the boundary, where there is certain non-morality according to the gold standard. According to this approach, if a comment received a probability score of 0,13 for the moral foundation Care/Harm, for instance, this would not count as a meaningful classification by the model and not be interpreted as such. If the score was a 0,19, though, one could effectively argue for the model's classification for weak morality based on the Care/Harm foundation in this comment. This approach for a lower boundary allows for the inclusion of as much data as possible and provides the analysis with more polarity within the scores and diversity in the data itself to contrastively look at the two extremes within the data. Therefore, it enables a qualitative analysis that does not neglect the differences in moral loadings in the individual instances, which potentially helps to further explore the model's functioning and the perplexity of morality with regard to implicitness.

Learnings and Application

Hence, in chapter 5, I will base the qualitative analysis on both approaches: approach 1) where the defined threshold for the model's validity for morality is oriented on the best F1-score for positive instances that will provide a substantial understanding of the respective label under investigation and its appearance and meaning within the Vegan Discourse, whereas approach 2) will focus on the comparison of weak and high probability scores for the labels, allowing for a detailed insight into the issue of implicitness of morality, by applying the lower boundary of absolute non-morality to the data and introducing the concept of moral loadings. Overall, in addition, given the limitations of this thesis in volume, the gold standard together with the metrics calculated reveal sufficiently that the model for the classification of moral foundations works well enough to reason its results for a larger generalization and allows for more fundamental orientation in the qualitative analysis.

4.4 Gold Standard – Manual Annotation of Storytelling in the Vegan Discourse Corpus

For the creation of a gold standard for storytelling, the same three annotators as for the annotation of the moral foundations worked through the data, which also consists of 100 instances from the VDC. In this case, however, there was a 50/50 split: 50 instances were randomly sampled that received a score by the storytelling model between 0 and 0,3 and another 50 samples with a range between 0,7 and 1. The annotators annotated for either evidence of storytelling, or the case where there is no evidence of storytelling. In general, signs of personal experiences formulated as first- or second-hand experiences were treated as signs of storytelling. While the model-based labels consisted of a 50/50 split for (non-)storytelling, the gold standard accounted 54 instances of non-storytelling and, accordingly, 46 instances of storytelling. The average score for Cohen's Kappa among the annotators is 0,45, which shows moderate agreement among annotators. The Cohen's Kappa score between the gold standard and the model reaches a value of 0,72, which represents substantial agreement. Between the agreement of annotators and the model, a correlation score of 0,08 was calculated. This hardly indicates a correlation and refutes the idea that depending on the agreement between annotators (two agree on a label or three agree on a label), the model gives a higher score for higher agreement.

Evaluation on the Corpus

Furthermore, evaluation scores for this model compared to the gold standard were calculated. Interestingly, all values – precision, recall and f1-score – are in the upper range with the lowest being 0,82. The F1-Macro average score reaches a value of 0,86, which indicates rather good values for precision and recall for both storytelling and (non-)storytelling annotations. Indeed, the precision for the detection of non-storytelling instances of the model reaches the highest of all calculated values with 0,90. For the opposite case, it reaches 0,82. The F1-scores for both cases are in the upper range with 0,87 for non-storytelling and 0,85 for storytelling. In total, it can be noted that the non-storytelling instances seem to be detected better by the model compared to the gold standard than the storytelling instances. Still, the storytelling instances are captured well by the model.

In total, the storytelling model was proven to be efficient for this case of labeling instances of the VDC respectively. Although there is the potential to improve certain scores, like the agreement among annotators, all metrics point to a good performance of the model. Hence, its results can be reasoned for further analysis and interpretation in this thesis.

4.5 Assessment

The evaluation results presented, to some extent, legitimize the models' results, which enables me to draw larger generalizations from them. However, two things need to be considered: the evaluation results for the foundation Loyalty/Betrayal were rather poor, so logically, any interpretations and analysis for this foundation are mere assumptions and need to be researched further. In addition, although agreement among annotators for the moral foundations are not unequivocal, the F-Macro average score of 0,55 indicates a moderate performance of the model on the gold standard set and thus accounts for further interpretation.

Descriptive Statistics: Moral Foundations

The following paragraphs serve as a kind of pilot study for the qualitative analysis of this thesis. Features of the instances classified per label in the gold standard will be observed regarding their average comment length, their most frequent words, including the most common bigrams and trigrams within these instances, and the distribution of general sentiment. For this, subcorpora for each label as assigned in the gold_strict set will be created. Overall, the analysis provides a scheme per foundation that can be evaluated with regard to the comments classified by the model as highly likely or less likely for a label and thus, helps to get a fundamental understanding of the two aspects moral sentiment and storytelling within the Vegan Discourse.

<u>CARE/HARM</u>

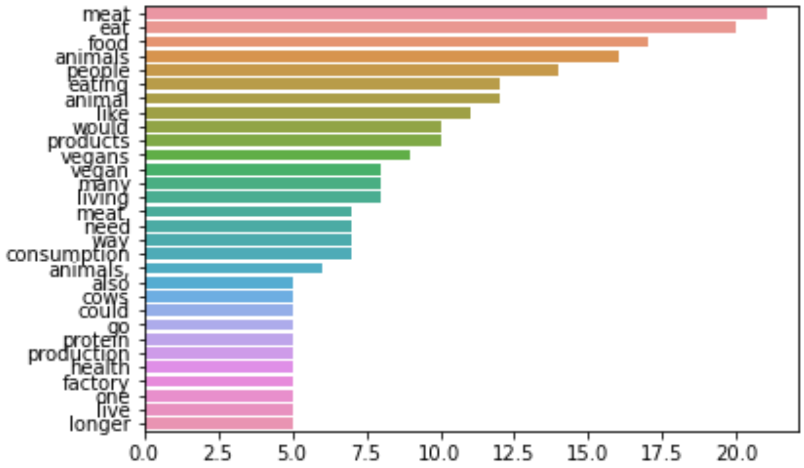

Figure 6 The most common words appearing in the gold Care/Harm instances

The Care/Harm foundation is represented with 25 instances in the gold standard, which were, on average, labeled with a probability score of 0,54 by the model. The comments have an average length of 501 characters without stop words. The most common terms in these instances are “meat”, “eat”, “food”, “animals” and “people” (see figure 6). These seem rather general but do hint towards the importance of animals regarding this foundation. Moreover, other frequently appearing words within these instances are terms related to the subject group of veganism, like “animal” or “cows”, terms that appeal to the meat industry, for instance, such as “products”, “consumption”, “production” and “factory”, and words that can be seen as referrals to the Care aspect of the respective foundation, like “health” and “living”. The most common bigrams in these instances are “eat meat”, “need eat”, “plant based”, “factory farms” and “consumption meat”. Mostly, these term combinations revolve around meat. Also, there is a reference to the meat industry with the last two mentioned bigrams. The term pairings “fruits vegetables make”, “plant based diet” and “new york times” are the most frequent trigrams in the Care/Harm gold instances, which are not necessarily clue-giving for the respective foundation. Trigrams that are interesting for this feature analysis are “non human animals” and “eat animal products”, which can also be seen as referrals to the meat industry. Regarding the general sentiment of the Care/Harm gold comments, it can be observed that mostly, these are labeled as neutral (45,8%). However, they are more negative (29,2%) than positive (25%) in sentiment. Overall, the comments labeled as Care/Harm according to the gold standard are rather long compared to the overall average of all comments in the VDC (250 characters), refer to the central figure of animals and the meat industry on the word-level, and are mostly neutral or negative in sentiment.

Fairness/Cheating

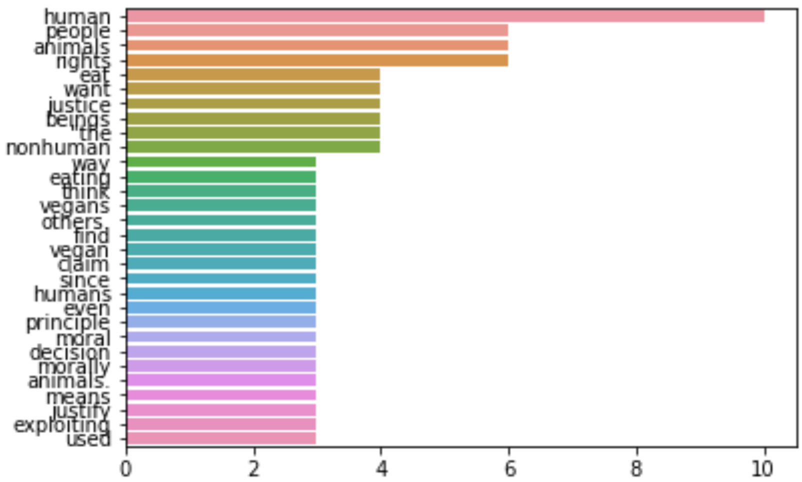

Figure 7 The most common words appearing in the gold Fairness/Cheating instances

The 12 comments labeled as Fairness/Cheating as gold have an average length of 305 characters and are classified by the model with an average probability score of 0,14. The five most frequently occurring stop words (see figure 7) are “human”, “people”, “animals”, “rights” and “eat”, where “rights” specifically alludes to the foundation’s principle within the context of the first three mentioned words. Other interesting words for this foundation that appear often in its gold labeled instances are “justice”, “principle”, “justify” and “exploited”. Interestingly, the terms “moral” and “morally” are frequent in the comments, which can be reasoned as an explicit expression of one’s moral values. Regarding the most common bigrams in the corpus, the foundation’s values are evident, as well: “principle justice”, “beings equal”, “animals rights”, and “justice says” are among the most frequent occurring bigrams. The same observation can be made regarding the most frequent combinations of three terms, with them being represented as “principle justice says”, “justice says sentient”, and “sentient beings equal”. Regarding sentiment, the instances are labeled equally as neutral as negative with a distribution of 0,33% in the Fairness/Cheating gold labeled comments. Comparably, only 25% of the instances are labeled with a positive sentiment along with 8,3% “Very positive” instances. The Fairness/Cheating gold labeled instances are in general a little longer than the average comment in the entire corpus, more negative in sentiment than positive, and show a reasonable representation of explicit words related to the foundation’s principles of justice and equality.

LOYALTY/BETRAYAL

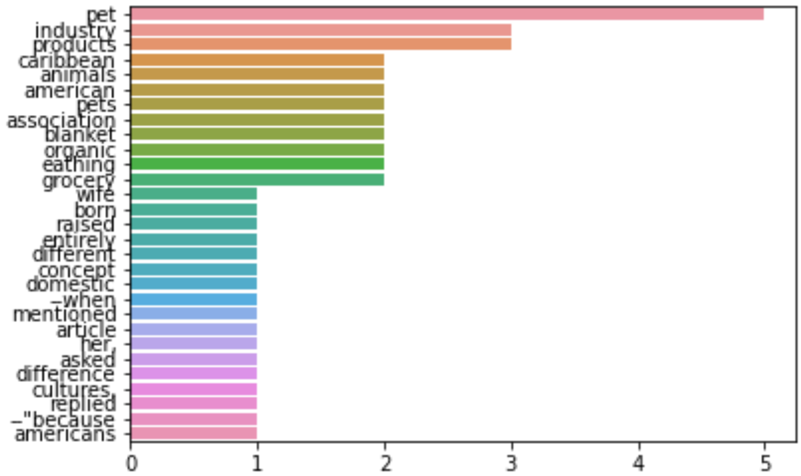

Figure 8 The most common words appearing in the gold Loyalty/Betrayal instances

Four of 100 comments with an average comment length of 378 characters were labeled as instances of the Loyalty/Betrayal foundation in the gold standard. The model detected a probability score of 0,047 on average for these comments, showing it did not identify the foundation's principles in these. The five most-often represented words in these comments (see figure 8) are "pet", "industry", "products", "caribbean" and "animals". Likely, the word "caribbean" is listed here because of a user's specific reference to it, but it does not show a specific reference to the foundation. The other words show an indication for a focus on animals, but also on the meat industry. Other key-giving words regarding the Loyalty/Betrayal foundation that appear often in the four comments are "american", "association" or "cultures", all referring to a sense of group or community. The most common bigrams in these four comments are "pet industry", "organic products", "wife born", "born raised", and "raised caribbean", which are not really indicative of the foundation's principles except for the last two mentioned ones, since they possibly refer to the belonging to a community. Regarding the trigrams, no specific observation can be made, since all of them revolve around the word "caribbean" in similar contexts. Overall, half of the comments are neutral, 25% are positive and another 25% "Very positive". In none of the comments, negative sentiment was identified. Hence, the Loyalty/Betrayal gold instances are longer than the average in the entire corpus, mostly positive or neutral in sentiment and show some reference to the foundation's principles with the frequent occurrence of terms like "american", "association" or "cultures". Still, it is indispensable to recognize that because of the weak representation of this foundation in the gold standard, any features as were detected in this paragraph are not exclusive, meaning they are not necessarily the most characteristic ones for the foundation and need to be further extended.

<u>AUTHORITY/SUBVERSION</u>

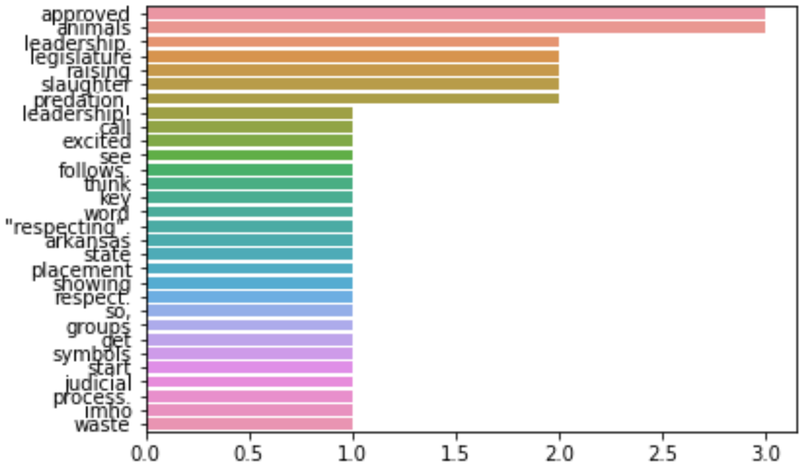

Figure 9 The most common words appearing in the gold Authority/Subversion instances

The Authority/Subversion foundation was detected as gold in seven of 100 instances of the gold standard. These comments have an average length of 137 characters. Interestingly, the model assigned an average probability score of 0,64 for these instances, meaning there is much overlap between the annotators' and the model's assignment. As the five most common words (see figure 9), the comments are represented by "approved", "animals", "leadership", "legislature", and "raising". These indicate a relation to the foundation's principles in the context of animals. Moreover, terms like "slaughter" and "predation" occur frequently, which seem to be more specific for the Vegan Discourse than for the foundation. Bigrams that occur often in the Authority/Subversion gold instances are "legislature approved" and "animals slaughter", where the former confirms the assumption deducted from the most frequently occurring single words. The latter is rather indicative for the Care/Harm instance, which might show an overlap of the two foundations or which is a result of the rather small sample of the gold standard. The most common trigrams are "call leadership excited" or "respecting arkansas state", which do not specifically refer to the Vegan Discourse without knowing the context, but are reasonable with regard to the model's assignment – they include explicit terms in the sense of the Authority/Subversion foundation. 85,7% of these comments are neutral and 14,3% are negative. There are no comments in which positive, "Very positive" or "Very negative" sentiment was detected, making the gold instances for this foundation rather neutral or negative in general. All in all, the comments are shorter than the average comment of the entire corpus, show explicit wording for the foundation and show only neutral or negative sentiment.

PURITY/DEGRADATION

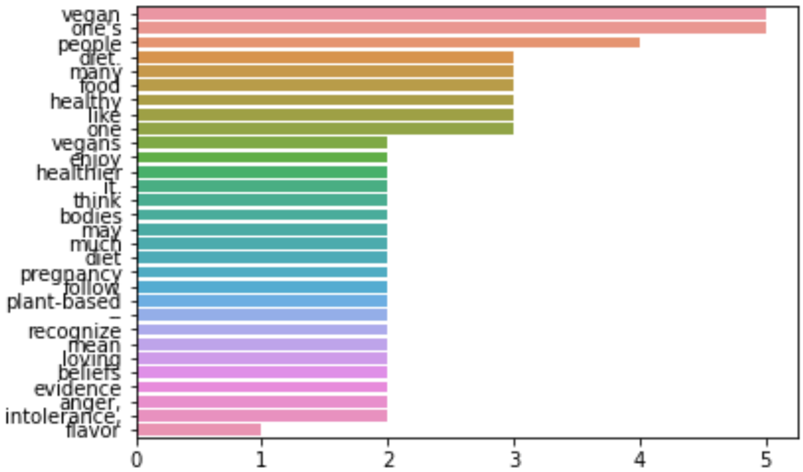

Figure 10 The most common words appearing in the gold Purity/Degradation instances

There are seven comments that were labeled as gold for the Purity/Degradation foundation. They have an average length of 356 characters and were classified by the model with an average probability score of 0,24, showing the model detected a weak indication for this foundation in the comments. The five words that occur most often in these (see figure 10) are “vegan”, “one’s”, “people”, “diet”, and “many” and are hardly allusive for the foundation, except for “diet”, which can be interpreted as a way of living a more elevated life as defined in the foundation’s principles. Other words that are more indicative of the foundation are “healthy”, “healthier”, and “beliefs”. These can be regarded as an expression of relevant words for the foundation within the Vegan Discourse, hence as a referral to the belief of living a healthy life. The term combinations “plant based”, “vegan diet”, “non believers” and “one faith” are the most common bigrams of the Purity/Degradation gold instances. The most frequently occurring trigrams are “anger intolerance hate”, “meat want healthier” and “healthier environmentally friendly”. Both the bigrams and trigrams well represent the foundation’s principle within the Vegan Discourse, since one can interpret that they refer to a healthier lifestyle with a plant based diet, for instance. Regarding general sentiment, most comments are labeled as neutral (57,1%). In addition, they are rather positive (28,6%) than negative (14,3%). In summary, the comments labeled as gold for the Purity/Degradation foundation are longer on average than the average of all comments in the entire corpus, rather positive than negative but overall neutral in sentiment, and are characterized by explicit expressions that can be interpreted in a way that marks the principles of the respective foundation.

NONE

The majority of the comments (45) in the gold standard were labeled with NONE by the annotators, meaning there is no indication of an explicitly or implicitly expressed moral foundation in these instances. The meaning of the NONE label evokes an issue of definition, as the absence of a moral foundation does not necessarily indicate the presence of non-morality, as one could assume. Accordingly, the researchers Kobbe et al. (2020) re-named the Non-moral label to “Moral”, as they reported instances with morality in the Non-moral labeled ones in the original corpus. However, since the average Non-moral probability score assigned by the model for the comments in the gold standard is as high as 0,92, I will argue that the significance of the NONE label equals the meaning of the Non-moral label in this thesis. Still, if an instance was assigned a probability score not only for the Non-moral label but also for a foundation label by the model, the latter will be treated as more significant than the former, to also comply with the observation by Kobbe and colleagues and to acknowledge the importance of the lower boundary set in this thesis.

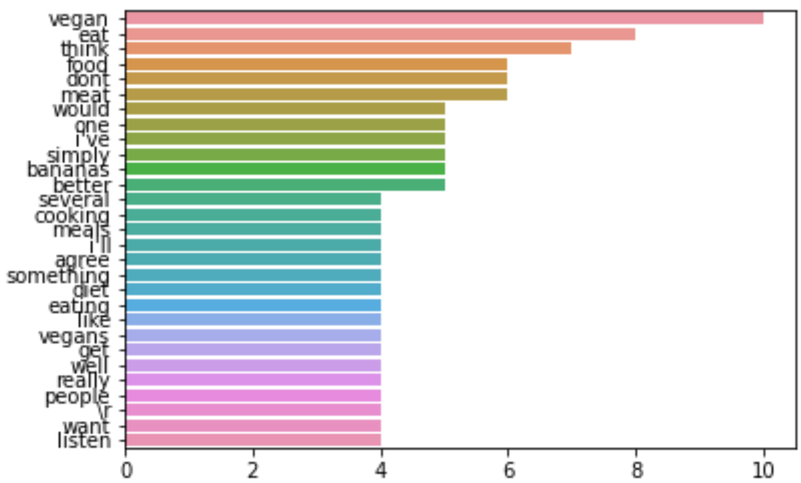

Figure 11 The most common words appearing in the gold NONE instances

Hence, what are the features of the NONE labeled instances of the gold standard? On average, the comments are 183 characters long and are mostly represented by the words “vegan”, “eat”, “think”, “food”, and “don’t” (see figure 11). These are all rather general words with no specific reference to a moral foundation. Words that evoke specific notions or concepts are “meat”, “cooking” or “diet”. Still, these do not indicate a specific reference to the moral foundations, although particularly the first and the last-mentioned term is frequent among the foundations’ comments of the gold standard. However, in these, they co-occur with other words that embed them in a more specific, possibly foundation-related context. The same applies to the bi- and trigrams, where the most frequent appearing ones are “don’t listen”, “listen you”, “ice cream”, or “vegan ice” and “don’t listen you”, “vegan ice cream”, or “great keep up”. Moreover, the comments labeled as NONE in the gold standard are mostly neutral (53,3%) or positive (26,7%) in sentiment. Fewer comments (17,8%) are negative or “Very negative” (2,2%). All in all, the NONE gold instances are shorter than the average comment in the entire corpus and mostly portrayed by general words that can be argued to be stereotypical for the Vegan Discourse, but not for explicit morality. In addition, most of these comments are neutral or positive.

QUALITATIVE INSIGHT: MORAL FOUNDATIONS (GOLD)

The following paragraphs will discuss some interesting cases regarding the model's and the annotators classification results. The presentation of these examples provide further qualitative insight into the classification results and thus, complete the feature analysis of the paragraphs before. First, the following comments shall be examined with regard to their assigned moral foundation by the annotators and the model, as well as the individual comment's statement:

a) "*Thank you for your progressive and responsible leadership. Your new menu is better for the planet, for the animals, and for our health. We look forward to supporting you, and we hope your example encourages many others to make the change*" (on "The New Menu at Eleven Madison Park Will Be Meatless" (Anderson & Gross, 2021))

In example a), the user praises the new, fully meat-free menu at a popular restaurant in New York City. They claim the restaurant's decision to go meat-free to be exemplary of the vegetarian or vegan culture and supportive of the well-being of animals, one's health and the planet alike. For this comment, the model assigned a probability score of 0,50 for the Loyalty/Betrayal foundation, 0,57 for the Care/Harm foundation, and a score of 0,77 for the Authority/Subversion foundation. All scores are above the thresholds defined per foundation. The annotators all assigned the label of the Authority/Subversion foundation making it a gold label for this comment, which is corresponding to the highest score given by the model for this instance. In addition, two of the annotators annotated Care/Harm for this comment and thus, making it a co-occurring gold label. This goes along with the model's classification results, where the Care/Harm foundation was assigned as the second most likely prediction. Moreover, one annotator detected evidence for both the Purity/Degradation as well as the Loyalty/Betrayal foundation. While the model did not detect the former, the latter is indeed represented with a higher score. In this example, the model is very consistent with what the annotators detected. Not only were the gold labels detected as most likely by the model, but all other but one annotation by the annotators for this instance were represented with higher probability scores by the model.

b) *"It's the beginning of the end of atrocities against the innocent defenseless."* (on "How to Cook With Plant-Based Meats" (López-Alt, 2020))

Comment b) perfectly illustrates a nearly unanimous decision by both the model and the annotators. The comment's creator argues that, according to the article's content which is being added in this interpretation, cooking with plant-based meat is a major contribution to end cruelty done to innocent and defenseless creatures. The automatic prediction for this comment was a probability score of 0,98 for the Care/Harm foundation, as well as a 0,18 for the Fairness/Cheating foundation. The latter is below the foundation's calculated threshold, and right at the lower boundary of non-morality, which could indicate a weak moral loading for this foundation. The annotators all agreed on Care/Harm, making it the gold label for this case. In addition, however, one annotator assigned the Loyalty/Betrayal foundation as well, which was not detected by the model.

c) *"I, too, follow a plant-based diet, as much for ethical reasons as for health. And I applaud the work of people like Dr. T. Colin Campbell at Cornell University ("China Study") who advocate plant-based living and back it up with science and with data. I wish The New York Times had looked a little harder for vegans to profile, however. This reads like caricature. "It's my nonalcoholic, meditative, yogic, vegan lifestyle" could -- and perhaps should -- be a line straight out of a Woody Allen film. Still, I guess that the phrase "by any means possible" should apply here. I'll look forward to a follow up piece with people who recognize and appreciate the difference between kicking a substance abuse drug like Oxycontin and giving up dairy. Please."* (on "Vegans Go Glam" (Gordinier, 2015))

This user expresses their dislike about the article's framing of veganism as a celebrities' issue while lacking in proper scientific facts and figures to explain the concept of veganism and its benefits. In this case, the model assigned the highest probability score for the Non-moral category and did not detect any significant indications for the moral foundations. The annotators found consensus in the Purity/Degradation foundation, however, which was represented by the model with a score of only 0,02, which is below the foundation's defined threshold. Hence, comment c) likely displays the challenge of implicitness in the context of morality concerning its automatic detection, since the comment is not necessarily explicit in the expression of foundation related words that could have been picked up. However, this outcome could also have been evoked by a lack of training data for the Purity/Degradation foundation in this specific context, for instance. A misunderstanding of the foundation among annotators is rather unlikely, as all three assigned the same label.

d) *"I differ with your blanket judgment regarding dairy and meat products. Excellent organic products are available. The mere act of breathing air and consuming water exposes one to the "poisons" you mention. Surely breathing and hydrating remain necessities of life. As the mother of pregnant and nursing women and the grandmother of preschoolers, I am familiar with the expense of "going organic" for the sake of the next generation. A Texas grocery chain (with its origins in my town) has a line of remarkably reasonably priced organic products targeting children. I'm thankful for grocery chains sensitive to families trying to reduce vulnerable children's exposure to the chemicals in our environment. But blanket judgmental statements don't help."* (on "Can You Have a Healthy Vegetarian or Vegan Pregnancy?" (Saint Louis, 2017))

In all four cases where the annotators agreed on the Loyalty/Betrayal foundation in the gold standard, the model did not detect any indication for this foundation. For the entire corpus, the model's predictions for this foundation are rather low scored; only one instance receives a probability score of 0,498 (example a)). Interestingly, comment d) is the only instance of the corpus where the annotators reached a majority for this foundation. On the article questioning a healthy vegan or vegetarian pregnancy, this comment is an emphasis on going for organic foods instead of promoting "blanket judgmental statements" about dairy and meat products. The model's predictions for this statement are as follows: 0,76 for Care/Harm and 0,52 for Non-moral. Hence, the Loyalty/Betrayal foundation is not detected by the model, which highlights the challenge of this foundation to be found in the VDC. In this case, however, differently to the Purity/Degradation foundation, one can assume that more training for the annotators would improve results, since out of four instances, only one was a majority-based gold label for this foundation. In addition, the different context of the VDC compared to the MFTC needs to be considered concerning this issue, since it might be the reason why the model did not pick up on any indications for the foundation. The annotators, comparatively, can transfer their knowledge about the foundations on different contexts.

Descriptive Analysis: Storytelling

As for the moral foundations, a pilot study regarding specific features of the comments assigned for a label will be performed for storytelling. The characteristics that will be observed are the comments' average comment length, their most frequent words, including the most common bigrams and trigrams within these instances, and the distribution of general sentiment. The threshold for the storytelling instances is at a value of 0,5.

Storytelling

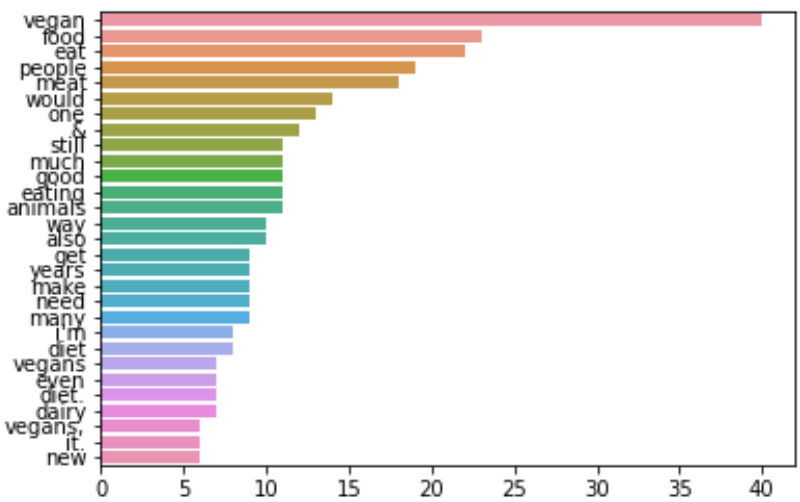

Figure 12 The most common words appearing in the gold storytelling instances

Of 100 comments, 46 are labeled as storytelling in the gold standard. The average length of these comments is 408 characters and the average probability score assigned by the model reaches 0,81. This high probability score confirms that the model confers quite well with the assignment of labels by the annotators. Moreover, the five most frequent words in these comments (see figure 12) are “vegan”, “food”, “eat”, “people” and “meat”, which are rather general and stereotypical for the Vegan Discourse and also, importantly, signs of implicitness in this thesis. This is a valid observation for all other frequently occurring single words within these comments, except that the often-appearing “I’m” could be interpreted as an indication of storytelling, since it likely introduces a personal statement. Bigrams that occur often are “plant based”, “vegan diet” and “years ago”, where the former two relate to the Vegan Discourse in general and the latter can be interpreted as an introduction of a personal utterance similarly as “I’m”. Interesting trigrams that appear often in the storytelling labeled instances are “wider range food”, “run farm sanctuary” and “found good alternative”. The latter two can possibly be observed as a part of personal story, the former does not necessarily indicate an explicit expression of such. Regarding sentiment, the comments are mostly neutral (48,8%) and nearly equally as negative (25,6%) as positive (23,3%). In addition, the sentiment “Very positive” was assigned in 2,3% of the storytelling gold instances. Overall, there are some terms or their combinations that can be seen as allusive to storytelling. In addition, the on average rather long comments compared to the average length of all comments in the entire corpus are mostly neutral and equally as positive as negative.

<u>NON-STORYTELLING</u>

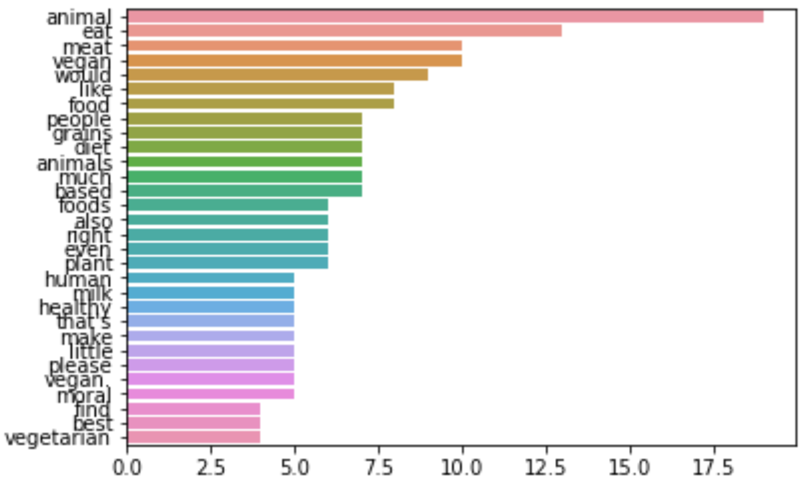

Figure 13 The most common words appearing in the gold non-storytelling instances

The 56 comments labeled as non-storytelling in the gold standard are, on average, 201 characters long and were assigned an average probability score of 0,32 by the model. The latter is compliant with the gold label for these comments, proving that the model is close to the annotators’ assignment. The five words that appear the most in the non-storytelling instances (see figure 13) are “animal”, “eat”, “meat”, “vegan”, and “would” – all are rather general and thus implicit in the context of the Vegan Discourse. Different to the storytelling instances is that the word “animal” occurs the most often in these instances, making the non-storytelling comments’ focus on animals evident. Other words that are frequent in appearance in these instances, such as “diet”, “goods”, “milk” or “healthy”, are all equally as general for the Vegan Discourse. The most common bigrams are “plant based”, “animal products”, and “eat meat”, the most frequent occurring trigrams are “cows bacteria guts”, “plant based diet”, and “pastry chef makes”. Term combinations like “animal products” show a reference to the Care/Harm foundation, for instance, but overall, the most often appearing terms and their combinations do not show an explicit expression of a specific notion or concept.

Their sentiment as predicted by the Stanford CoreNLP classifier, in addition, is mostly neutral (51,9%) and nearly as negative (21,2%) as positive (25%). The sentiment “Very positive” was assigned for 1,9% of all non-storytelling instances of the gold standard. Overall, the non-storytelling instances are, on average, shorter than the storytelling ones and nearly as long as the average in the entire corpus. Furthermore, they are mostly neutral in sentiment and show a focus on animals on the word-level but no specific notions or concepts are expressed; rather general references to the Vegan Discourse.

QUALITATIVE INSIGHT: (NON-)STORYTELLING (GOLD)

The following paragraphs are dedicated to some interesting cases from the automatic classification of storytelling. The aim is to provide an insight into these comments from the VDC regarding its classification as storytelling and, same as in the case of the moral foundations, point out any observations in the process.

e) *“I have two border collies on a vegan diet. Both of them had allergies before the switch to vegan food, and they are definitely doing better. They have been eating V-Dog for a couple of years. They are still super high energy dogs, and their coats are beautiful. I'm a climate scientist, and one of the reasons my dogs and I are on a vegan diet is to reduce our carbon footprint. Meat and dairy production is a large contributor to global greenhouse gas emissions.”* (on “The Vegan Dog” (McDermott, 2017))

Comment e) on the article “The Vegan Dog” is a user’s statement of appreciation of a vegan dog diet that is not only benefitting for their dogs’ health, but also important for the user as a climate scientist with regards to the reduction of carbon emissions. This example illustrates well how the storytelling concept of including personal experiences as a form of justification or illustration works, which is reflected in its annotations – from the model as well as from the annotators. The latter reached a unanimous decision for evidence of storytelling in this instance. The model, too, claimed this comment to contain storytelling with a probability score of 0,89.

f) *“The reason 'most people dont take veganism seriously' is because they are blissfully (or willfully) ignorant of the barbarism of factory farms and slaughter houses. This is partly because the meat and dairy industries have commoditiized animals so that they live in die in deplorable conditions for cheap meat. Additionally, congress has enacted strict ag gag laws which prevent the terrible truth about the daily suffering in factory farms and slaughter* *houses to reach the public. Just look at your dog or cat and then picture him living and dying the way 'food' animals do. 'Most people' would recoil from such barbarism. No gentle sentient creature deserves a lifetime of suffering for your ten minute meal."* (on "When Vegans Won't Compromise" (Fischer & McWilliams, 2015))

A more difficult case regarding its annotation is comment f), which is a statement about the ignorance of people regarding the exploitation of and cruelty towards animals for a "ten minute meal" and the separate view on this matter concerning dogs and cats compared to "food" animals. While the model scored a probability of 0,24 for storytelling, meaning it did not detect substantial evidence for storytelling, two of the three annotators labeled the opposite. Most likely and logically, this is because, compared to example e), there is no actual "I" agent in the comment, which could possibly account for a personal report in an explicit manner. Still, the author illustrates their statement with descriptions on the conditions of "food animals", for instance, and appeals to their potential readers to imagine such a life for their pets. This observation makes the annotators decision reasonable, and it seems as if the model's lower probability score is grounded in the fact that comment f) does not necessarily serve as the prototypical storytelling example with specific features, particularly in comparison with comment e).

g) *"I agree with the hard-core Vegans on this one: if you don't eat meat out of moral, or even health, concerns, then a wedding is the perfect time to educate your guests about how you believe they should eat and why. At the very least, the person throwing the wedding should not have to pay to have animals killed! People in this country eat way too much red meat anyway. Also, it is shocking to me that caterers charge more for a vegan wedding than one with meat, as meat tends to be more expensive than fruits and vegetables. As a final note, raw vegans (a subset that eats only uncooked fruit and veggies) could use the wedding to educate their guests not only on the health benefits, but on the full amazing flavor of a raw diet. And, to invite many more people, they could avoid the caterer and go straight from the farmers' market to the wedding, giving each guest a sharp fruit knife! Enjoy!! And mazal tov to Chelsea and Marc! L'Chaim!!"* (on "At Vegans' Weddings, Beef or Tofu?" (Quenqua, 2010))

A user's opinion about a vegan wedding and the opportunity to teach the concept of veganism to the wedding's guests can be found in comment g). All three annotators agreed that this comment contains storytelling. The model, however, only assigned a probability score of 0,20, declaring it a non-storytelling instance. One can assume the reason for the model's to be the same as in f) – the missing occurrence of the "I" agent in the text together with the description of a personal experience. In the case of comment g), the author tells of some experiences, like the fact that a catering for vegan food is more expensive than for non-vegan food but they do explicitly frame the story itself as a justification, it is rather hidden when compared to comment e), for instance.

h) *“Why does Kenji recommend using a thermometer to gauge when the burger is ready? Is the product not theoretically possible to eat raw? It would help if the recipe gave an approximate time for cooking the burger.”* (on “Vegan Cheeseburgers” (López-Alt, n.d.))

In comment h), its author asks about details of a recipe for a vegan cheeseburger. The model does not detect any significant evidence for storytelling and assigns a probability of 0,19. Accordingly, the annotators unanimously agree on the same state: there is no evidence for storytelling in this comment. Intriguingly, though, the model’s probability score for comment h) is not significantly weaker than the scores assigned for comment f) and g). However, the latter two were, to some extent, identified as storytelling by the annotators, which can be reasoned in a way when exploring the comments. In comment h), one could assume a story to be implied in the last sentence “It would help if the recipe gave an approximate time for cooking the burger”, since the author implies to have tried the recipe and from their experience, this addition to the recipe would provide better cooking results. From this perspective, the similar scores for the three comments could be reasoned. Still, comment f) and g) seem to be more tangible in their storytelling than comment h).

5 Exploring the Vegan Discourse: Analysis & Interpretation

In all cases, meaning the classification results for the moral foundations above the threshold and the two subsets incorporating the poles of moral loading, as well as for the classification results for (non-)storytelling, the following aspects will be examined: first, the general probability distributions per annotation domain are of interest, since they give an overview over how well a specific foundation or storytelling are detected in the entire corpus. In the case of the moral foundations, this will first be investigated with regard to the threshold of 0,52. Moreover, if existent, particularities will be pointed out, such as peaks in frequency or other observations across the results after some cleaning of the data, like the average comment length in characters, the most common n-grams or a general sentiment analysis. The results will be compared with the gold features of the respective label to check on the model's performance, except for the contrastive analysis of the moral foundations oriented on the lower boundary of 0,18, since this aims at documenting the different moral loadings and their importance on the implicitness of morality. Although the gold standard set is rather small, this aspect might provide indications of interest and spark assumptions that can be further explored in future research. Moreover, it is important to provide a qualitative insight into what the actual data looks like by showing samples and discussing their classification result and to make the results more palpable for a deeper understanding. In addition, some specific observations will be shared regarding the foundation's correlation to each other or user patterns regarding their expression of moral values together with storytelling, for instance.

5.1 Moral Foundations

To start off with the investigation of fundamental insights regarding the presence of the moral foundations within the Vegan Discourse, this first section of the chapter will focus on the results obtained by setting the thresholds of moral significance calculated for each foundation on the entire data consisting of 10'432 comments. The resulting distribution is as follows:

- 1'286 Care/Harm instances (12,3% of the corpus at threshold 0,46)
- 168 Fairness/Cheating instances (1,6% of the corpus at threshold 0,44)
- 932 Loyalty/Betrayal instances (8,9% of the corpus at threshold 0,05)
- 20 Authority/Subversion instances (0,2% of the corpus at threshold 0,59)
- 335 Purity/Degradation instances (3,2% of the corpus at threshold 0,27)
- 3'651 Non-moral instances (35% of the corpus at threshold 0,98)

The most dominant moral foundation in the Vegan Discourse Corpus is the Care/Harm foundation, making up about 12,3% of the entire corpus. Second most prominent in the VDC is the Loyalty/Betrayal foundation with around 8,9% of the corpus; however, since the threshold for this foundation is very low and was calculated based on a rather poor representation of this foundation in the gold standard set, this finding cannot be interpreted significantly and needs to be refined further in future work. The VDC is moreover much represented by the Purity/Degradation foundation, which makes up 3,2% of the corpus. It is followed by the Fairness/Cheating foundation with 1,6% of the entire corpus. The foundation assigned least often in the corpus is the Authority/Subversion foundation in about 0,2% of all comments in the corpus. Overall, 35% of the corpus were assigned a probability score higher than the label-specific threshold for Non-moral, which accounts for roughly a third of all comments to contain no morality. Still, based on the lower boundary of non-morality, this needs to be explored further, since probability scores above 0,18 for the Non-moral scores were claimed only to be significant if there is no moral foundation or several ones assigned simultaneously with (a) score(s) above the boundary. Regarding the morality in this corpus, it can be documented that the Care/Harm foundation is most dominating, which is in harmony with the basic principles of veganism to not live a life that causes suffering to other sentient beings. The Loyalty/Betrayal foundation's dominance in the corpus is difficult to interpret in this context, but if one argued for the threshold to be eligible enough, this would indicate that the Vegan Discourse is also strongly characterized by a group-sense and the "one for all, all for one" mentality. Furthermore, the Purity/Degradation is prominent in the corpus, which corresponds to its principles interpreted in a way that vegans want to live a "less carnal" and "more noble" way of life by abstaining the consumption of animal products. In addition, making up a small part of the corpus, the Fairness/Cheating foundation is detected as second least often assigned foundation. Still, its detection in the VDC concurs with the wish of vegans to treat all sentient beings equally and just. This circumstance contradicts one of the hypotheses in this thesis; namely that the Vegan Discourse is mostly characterized by the two moral foundations Care/Harm and Fairness/Cheating, since vegans strongly relate to the principles of these two foundations. However, it is only the former that dominates in the VDC, while the latter is among the least dominant ones. The least represented moral foundation in the corpus is Authority/Subversion in 0,2% of all comments. This can be argued with the fact that the principles of this foundation do not necessarily align with veganism as much as those of the other foundations, and if so, it might not be as obvious in this context as in the other foundations. Moreover, this could be an issue of the different semantic contexts in the training data, since the context of the MFTC is not the same as in the VDC.

METHODOLOGY

For a more detailed insight, the following paragraphs will investigate specific features of the comments per label. With respect to each label and the comments assigned with scores above the label-specific threshold, the comments will be analyzed quantitatively by exploring the average comment length, the most frequent words and n-grams, a brief look at the comments' sentiment distribution via the Stanford CoreNLP annotator and by comparing these observations with the gold standard for each label. Furthermore, to get a better insight into what makes a comment more or less likely for a moral foundation, meaning what made the model decide the way it did and if this is a matter of explicitness vs. implicitness of morality, for all labels in chapter

5.1, I will perform a contrastive analysis per label. Hence, I divide the data into two subsets and look at the same descriptive statistics as before, like the average comment length in characters, the most frequent n-grams and a brief sentiment analysis of the two subsets. Per label, subset 1 will represent the weaker probability scores, which account for a weak moral loading of a comment in this thesis. But when can a probability score for a comment be claimed as indicative for a weak moral loading? What point marks a boundary for non-morality, after which one could argue for weak morality? This is where the lower boundary of 0,18 elaborated in chapter 4.3 becomes crucial. To be able to argue where there is validity in weak probability or not, this boundary defines a probability range of absolute non-morality from 0 up to 0,18, which was calculated by leaning on the gold standard's distribution of the NONE label and its respective classification scores by the model (see chapter 4.3). Everything above this boundary can be claimed as indicative for weak up to high morality. The latter is captured in subset 2 per label and is oriented towards the peaks in data distribution in the higher probability ranges roughly starting at 0,7, which represent higher moral loadings in this work. Comparing the two subsets with each other with regard to certain features enables one to further look into morality's challenge of implicitness and ameliorates the understanding of the model's way of detecting morality despite its implicitness.

CARE/HARM

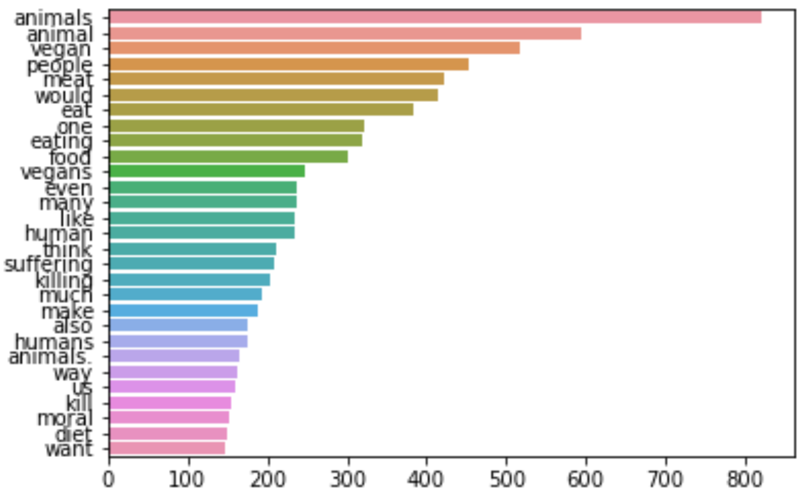

Figure 14 The most frequent occurring words in comments of the Vegan Discourse Corpus labeled for the Care/Harm foundation (threshold 0,46)

Above the threshold, the comments for the Care/Harm foundation peak in frequency in the probability range from 0,8 to 0,85 and 0,89 to 0,93 with around 150 comments per range. The least comments are scored with values of 0,52 to 0,58 or 0,62 to 0,68 with around 80 instances each. The highest score for the Care/Harm foundation is around 0,98 and overall, the comments are rather highly likely for the moral foundation, with accounts for the average probability score of 0,73 for the instances above the threshold. All the comments with a probability score higher than the threshold of 0,46 for the Care/Harm foundation are on average 429 characters long and represented by terms like "animals", "animal", "vegan", "people", and "meat" because of their high frequency (see figure 14). The most frequent bigrams in these comments are “plant based”, “eating meat”, and “animal products” but also term combinations like “killing animals” or “animal agriculture””. Concurrently, trigrams that appear often in these comments are “eat animal products”, “animals feel pain”, and “heart disease diabetes”. Hence, on the word-level, the Care/Harm foundation is represented rather explicitly by the expression of Care/Harm related words like “killing” or “pain”.

In addition, the distribution of general sentiment in these comments is as follows: 41% neutral, 35,8% negative, 22,1% positive, 1,0% very positive and 0,2% Very negative. Hence, the Care/Harm designated comments are mostly neutral but also rather negative than positive. In comparison to the gold standard, it can be noted that the average comment length is a little shorter in the entire corpus and that the most frequent terms and term combinations do not show substantial overlaps with those of the gold standard. Except for words like “people” and “meat”, the explicit wordings closely related to the Care/Harm foundation are not listed. Regarding sentiment, it can be observed that the distribution in numbers is not entirely similar, but the general allocation is the same: mostly neutral or negative, less positive.

CONTRASTIVE ANALYSIS

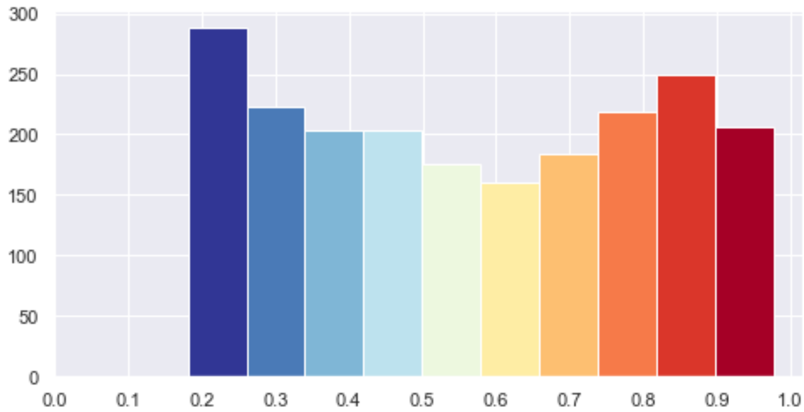

Figure 15 The probability distribution of the Care/Harm foundation in the Vegan Discourse Corpus starting at the lower boundary

After looking at figure 15, it turns out that the frequency peak is in the lowest probability range, from the lower boundary of 0,18 to 0,26, with around 290 instances. The least of the instances received a moderate score between 0,59 and 0,66, in total roughly 160, which is the smallest number of appearances for this foundation in all probability score ranges. In addition, there is one small peak in the probability range between 0,81 and 0,9 with about 250 instances. Overall, the comments received mostly rather low probability scores in the range from 0,18 to 0,26, but they are also represented significantly in the higher range from 0,73 to 0,99.

The first subset consists of all comments with a probability score greater than 0,18 but less than 0,33 to capture the frequency peak in the lower probability range, in total 484 comments. Comments that received a probability score higher than 0,73 are in subset 2, in total 692 comments, and serve to represent the higher probability range including the peak at 0,8 to

0,9. After the deletion of HTML tags and the removal of stop words, the average comment length in subset 1 is 411 characters, for subset 2 it is 421 characters, making the higher scored cases slightly longer on average than their inverse. Furthermore, the most occurring words (excluding stop words) in subset 1 are “vegan”, “meat”, and “people”. In subset 2, the most frequent words are “animals”, “animal”, and “vegan” among words like “suffering”, “kill” and “moral” in the lower frequency range. Intriguingly, this indicates a focus on animals and their suffering in the comments with higher probability scores, while in subset 1, the term “animal” comes at the 9th place in frequency, after “food” and “eating”. Concerning the most common trigrams in the corpus, subset 2 confirms the observation of a greater focus on the suffering of animals with the frequent appearance of the trigram “animals feel pain”, yet among term combinations like “plant based diet” or “non human animals”, which are rather stereotypical for the Vegan Discourse but not necessarily for the foundation. In subset 1, an explicit reference to the Care/Harm foundation via a trigram like in subset 2 is not observed in the n-grams: the most frequent trigram is “plant based diet” (similar in subset 2), among occurrences like “stop eating meat” and “eat animal products”. A similar observation can be made for the most common bigrams in subset 1 and 2. In 2, bigrams like “animal suffering”, “sentient beings” and “killing animals” are frequent but absent in 2. It seems as if the comments in subset 1 were scored with lower probabilities because of their more general view on the topic, meaning less expressive or more implied nature of the Care/Harm foundation related concepts of nurturing or harming others. For the higher scored instances, the inverse is the case: there are more explicit wordings related to the foundation.

Regarding the sentiment in the comments per subset as was performed with the Stanford CoreNLP library, it can be stated that in both, the comments are rather neutral or negative in sentiment than positive. After normalizing the counts through calculating the percentages as will always be done in the following sections, there are more comments in subset 2 that received polarity scores for the negative sentiment (36,2%) than in subset 1 (31,3%). Hence, subset 2 is more negative in sentiment than subset 1. The neutral sentiment distribution among both subsets is nearly similar, with 41,9% in subset 1 and 40,3 in subset 2. With regard to the positive sentiment, it is subset 1 that is more dominant with 25,4% compared to subset 2 with 22,2%. Moreover, subset 1 sparked sentiment classifications in “Very positive” (0,88%) and “Very negative” (0,44%). The former was also observed in subset 2, but with a higher percentage distribution of 1,36%. In total, comments with a higher probability to contain morality of the Care/Harm foundation are, compared to their opposite cases, on average slightly longer and focused on more animal-related issues of veganism. In addition, they are more negative in sentiment than the comments with lower probabilities for the Care/Harm foundation.

<u>QUALITATIVE INSIGHT</u>

In the following, examples for the different probability ranges will be discussed to provide more insight into the foundation related comments as well as the model's classification scores for weak and high moral loading as well as total morality above 0,52.

a) *"@[USER] the animal milk is for animal babies. The babies are taken away from the mothers at birth to prevent them from "stealing" the milk which can be sold profitably for human consumption. Another thing to consider is why human breast milk is not considered acceptable for adult consumption but cow or goat breast milk is acceptable. Why not drink the milk or make cheese using milk from one's own species? Why go to a different species? If the animal is not your mother, then it is not your milk."* (on "Vegan Cheese, but Make It Delicious" (Rao, 2021a))

In comment a), a user is telling about their viewpoint on cow milk – they claim humans should rather use human breastmilk instead of a that of another species and explain that in the process of making cow milk, the cow mother's baby is taken away to gain its milk and sell it. The model assigned a probability score of 0,18 for the Care/Harm foundation. Although this score accounts for a weak moral loading of the comment for this foundation, it seems legitimate in the semantic content of the comment – a baby is taken away from its mother, which goes against the principle of nurturing and gentleness of the foundation. Interestingly, it seems that the implicitness of this image in the user's statement was captured by the model with a weak probability score. Hence, there is a lack of explicit signs of morality, like certain terms, in the comment. Notably, the model also assigned a probability score of 0,74 for the Non-moral category for this instance, which is primarily neglected since the it is below the label's threshold and the foundation's probability score for this comment is right at the lower boundary and thus, valid regarding a weak moral loading. Still, in this comment, one can argue that the high Non-moral score results from the implicitness of the comment, which touches the insights covered in chapter 5.3.

b) *"@[USER] Diet for a Small Planet was definitely not vegan; vegetarian, aggressively so — dairy at every meal to "combine" for max non-"meat" protein. Its central concept is very outdated. Not to mention that dairy cows suffer for years more than animals raised for meat."* (on "At Seasoned Vegan, Soul Food That Never Harmed a Soul" (Mishan, 2018))

A user's claim about the dairy products in a vegetarian diet causing longer-lasting suffering of dairy cows than of animals bred for meat production can be found in comment b). Its foundation probability score for Care/Harm is 0,51, but there is also a probability score of 0,71 for the instance to be Non-moral. The higher score for the foundation, which is above its defined threshold, can be reasoned with the author's explicit expression of the suffering of dairy cows by referring to the term "suffer", for instance, which shows their compassion and care for the animals. The Non-moral score is neglected in this case, since it is below the defined threshold for this label and the foundation's score accounts for a moderate moral loading of the comment. The high Non-moral score, most likely, resulted from the fact that most of the comment does not explicitly mention any foundation related terms as in the last sentence.

c) *"Yes, the animals being slaughtered take precedence over the people who murder them ... are you serious? If it's so traumatic then they shouldn't slaughter them. It's murder, is it supposed to feel good? Do you care more about murderers of people then their victims? Why don't animal lives matter just as much? Good grief. If there's no one left to do the killing then that would be the end of it. At least these "workers" can go home at the end of their day. ..."* (on "Good Vegan, Bad Vegan" (Brody, 2021))

Comment c) is the comment with the highest probability score for the Care/Harm foundation in the entire corpus – a value of 0,98, which is in the range of definite morality. The comment's author complains about the fact that animals' lives do not matter as much to the general population as those of the people involved in the killing of animals. Given the fact that the author alludes to many principles of harm with words like "murder", "killing" or "slaughter" in multiple ways, the model's classification seems reasonable. It is an explicit expression of harm done to others, with these being animals.

d) *"Nobody cares what you eat. They care about the animals you kill and the environment you harm."* (on "Stop Mocking Vegans" (Manjoo, 2019))

A brief and direct statement of a user about the relevance of the harm done to animals and the environment and the irrelevance of the actual food one consumes is captured in comment d). This example was chosen for illustration because it is, unlike the comments before, rather short in length, but still highly likely to contain Care/Harm morality because of the explicit mentioning of harm. It was assigned a score of 0,95 and is thus a portrayal of this foundation in the corpus.

ASSESSMENT: CARE/HARM IN THE VEGAN DISCOURSE

The Vegan Discourse as represented in this paper is significantly characterized by the Care/Harm foundation, since of all instances that were assigned a probability scores above the defined threshold per foundation, the Care/Harm foundation makes up the most. This confirms one of the hypotheses in this work: the Vegan Discourse, if moral, is highly so with regard to the Care/Harm foundation, since the principles of veganism align with those of the foundation, as was detected by the model in the instances. Most of the classified comments of the Care/Harm foundation of the corpus are, if one takes into account moral loadings by applying the lower boundary of 0,18, in the lower range concerning their probability score, which is a sign of a weak moral loading and thus, implicitness, meaning a lack of foundation related explicit expressions. This is confirmed by the fact that the instances in the range of definite morality for this foundation above 0,46 are characterized by explicit wordings like "kill" and "suffering" and thus, received rather high probability scores with an average of 0,73. This explicitness is, however, not portrayed in the gold standard set of this foundation. In addition, there is no substantial overlap in the most frequent words of the comments with a score above the threshold of 0,46 and the gold standard. Both observations can be a result of the limitation in size of the gold standard and need to be explored further in future research. During the contrastive analysis, it became clear that the lower scored comments for this foundation are slightly shorter, less explicit in words regarding the foundation's concepts and more positive in general sentiment than the comments with higher probability scores for this foundation. In comparison with the gold standard, the sentiment distribution is overall similar: mostly neutral or negative and less positive. In summary, comments that received high probability scores for the moral foundation Care/Harm were written by users that explicitly express their moral beliefs, and vice versa, mostly in a neutral or negative sentiment.

FAIRNESS/CHEATING

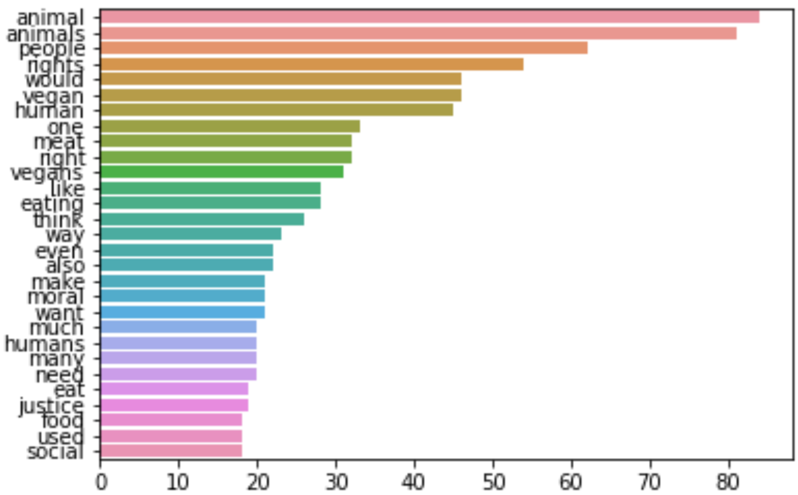

Figure 16 The most frequent occurring words in comments of the Vegan Discourse Corpus labeled for the Fairness/Cheating foundation (threshold 0,44)

Differently to the Care/Harm instances, the comments assigned with a probability score higher than the threshold for the Fairness/Cheating foundation (0,44) peak right at the threshold and decrease in number the higher the probability score becomes. In the range from 0,44 to 0,5, the most comments can be found, in total around 36. After a dip in the data in the range from 0,65 to 0,7 with in total 8 comments, there is another rise in frequency from 0,71 to 0,79 with around 16 instances. This decreases to 6 instances in the highest range from 0,92 to 0,98. Overall, the comments received rather moderate scores mostly from 0,44 to 0,79, with the highest peak around the threshold and an average probability score of 0,64. The 168 comments classified with a probability score higher than the foundation's threshold are on average 363 characters long and characterized by the most frequent words like "animal", "animals", "people", and "rights" (see figure 16). Although the latter already implies a relation to the respective foundation, other frequent words are interesting to mention, such as "moral" or "justice". The comments' relation to the respective foundation can also be observed in the bigrams, with "animal rights" and "social justice” being most common among the comments, among “plant based”. The most occurring trigrams are less direct with this regard in two of three cases: “plant based diet”, “social justice movement”, and “plant based foods”. Still, on the word-level, the foundation Fairness/Cheating is well represented. Furthermore, these Fairness/Cheating assigned comments are mostly neutral (41,6%) or negative (41%) and less positive (17,4%). Compared to the features of the gold standard, the following observations can be made: the average comment length is quite similar, which is also the case for the most common words and bigrams. The trigrams are slightly more significant in the gold standard, but this could be a result of the small sample taken for the gold standard. The sentiment distribution in the gold standard, however, is similar in its general order, with the comments being more neutral and negative than positive. However, the comments of the entire corpus are more neutral and negative than the ones in the gold standard.

CONTRASTIVE ANALYSIS

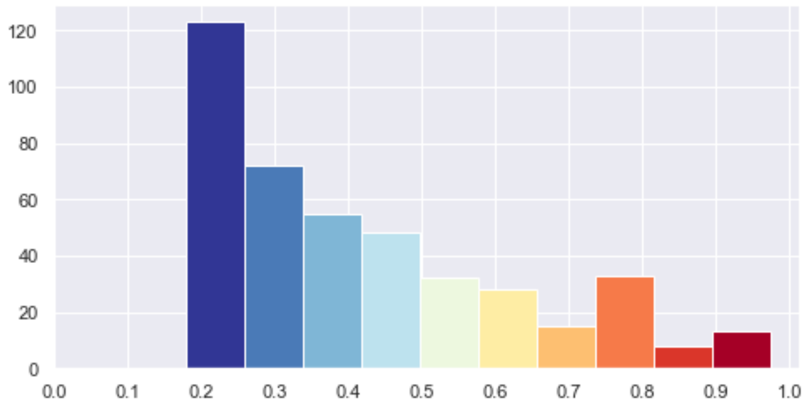

Figure 17 The probability distribution of the Fairness/Cheating foundation in the Vegan Discourse Corpus greater than the threshold

As can be seen in figure 17, most of the instances that received a score above the lower boundary, around 120, received a lower probability score ranging from 0,18 to 0,26, followed by circa 70 comments up to 0,33. While in the highest range from 0,73 to 0,99, fewer instances are recorded, there is a peak in the range from 0,73 to 0,81 with about 35 comments. Overall, the higher the probability score, the less comments were assigned for this foundation. For this foundation as well, a brief descriptive analysis of the features of lower scored and higher scored instances is interesting to observe. Thus subset 1 is representative for comments with probability scores from 0,18 to 0,33, while subset 2 contains of all comments with a probability score higher than 0,73. In this case, the distribution is rather distinct, with 188 comments in subset 1 and only 55 comments in subset 2. Interestingly, subset 1 consists of rather long comments with around 465 characters on average. The comments in the second subset are only about half as long, with roughly 281 characters on average. The most frequent words in the second subset show, similarly to subset 2 of the Care/Harm foundation related comments, a focus on animals and their rights: “animals”, “animal”, “rights”, “human”, “people” and “justice” are the most frequent words in this subset. In subset 1, however, there are, among the most frequent words “animals” and “animal”, more general terms like “vegan”, “meat”, “and “people” that appear most often. Interestingly, the terms “animals” and “animal” are higher in frequency in subset 1 of this foundation than, contrastingly, in subset 2 of the Care/Harm foundation. A look at the most common bigrams in both subsets specifies this observation even more: in subset 1 and subset 2, terms centered around the principles of the Fairness/Cheating foundation occur, like “animal rights”. Nevertheless, subset 2 has a stronger representation of correlated terms, such as “social justice” as one of the most occurring bigrams and “principle justice” or “beings equal”, while in subset 1, there are more general vegan-related term combinations like “plant based” or “eat meat”. Concerning trigrams, subset 1 consists of frequently occurring combinations such as “like us animals”, “turn blind eye” or “feel morally inferior”. Subset 2 is characterized by trigrams like “are humans superior”. In general, subset 2 is more explicit in the expression of Fairness/Cheating related words with regard to bigrams. Regarding three term combinations, subset 1 seems to dominate in explicit wordings, though.

Moreover, regarding a general sentiment analysis, the comments in subset 1 were mostly neutral (36,7%), also nearly equally negative (36,1%) and less positive (26,7%). Furthermore, the label “Very positive” was assigned for 0,56% of all instances in subset 1. Subset 2 is more neutral in sentiment. In addition, the instances in subset 2 are nearly as frequently negative in sentiment as in subset 1 with 35,2% of all instances classified as negative. Particularly, there are less instances with positive sentiment (13%) in subset 2 compared to subset 1. Overall, the longer comments of subset 1 are more polarized in sentiment, meaning they are more negatively and positively classified. Although subset 2 was slightly more explicit in the occurrence of foundation-related terms when looking at the most frequent single words and bigrams, subset 1 was characterized by few, but at least some term combinations that avoided implicitness. Generally, the results for this foundation produced by the model do not show significant differences on the word-level per probability range, framing an interesting case where explicitness and implicitness within the comments do not necessarily account for different probability scores assigned by the model.

QUALITATIVE INSIGHT

A more detailed look at some samples shall provide more insight into the probability distribution of the corpus.

e) *“I live in Boulder County, CO. The number of people with vegan, healthy, active lifestyles in the cancer treatment center is heartbreaking. They feel betrayed since they followed all the “rules” and yet developed cancer. Your comments imply blame must be assigned to the victim of these terrible diseases. If only they’d eaten right they wouldn’t have developed cancer. This superiority attitude hinders progress towards health care for all since getting a disease must be the result of a poor lifestyle choice. While I concede, there are definite unhealthy habits, eating healthy does not guarantee one is disease free. Treating diet as a religion, dogma and all, doesn’t lead to understanding, nor empathy for Those in need of medical care. Empathy seems to be needed* *in the fight for decent medical care for all regardless of whether or not we find our fellow humans living “worthy “ lifestyles or not..”* (on “Good Vegan, Bad Vegan” (Brody, 2021))

Comment e) received a low probability score of 0,18 for the Fairness/Cheating foundation. The comment’s author makes a statement about the fact that a healthy diet does not avoid getting a disease and that instead of treating a diet like “a religion [or] dogma”, people should be more empathetic and less condescending of other people’s lifestyles within the “fight for decent medical care for all”. Particularly the last sentence of this example refers to the fundamental ideas of the Fairness/Cheating foundation: “justice, rights, and autonomy” (Dehghani et al., 2021) by the reference to “decent medical care for all”, which, most likely, sparked the model’s assignment for this foundation. The other parts of the comment are ambiguous regarding its morality. This was also captured by the model with the assignment of several probability scores for the moral foundations. The score of 0,61 for the Care/Harm foundation relies on the comment author’s expression of the caring for others that need medical care. Moreover, the model captured indications of the Loyalty/Betrayal foundation with a probability score of 0,14, which most likely came from the author’s sense of “one for all, all for one” in the achievement of medical care for all – this seems to correlate with the Fairness/Cheating foundation in this example. Furthermore, the Purity/Degradation score was assigned with 0,21 with the reference to the vegan diet as a religion or a worthy lifestyle. Still, it is most important to notice the assignment of the Fairness/Cheating label for this instance, since the comment’s reference to fair treatment of all with regard to medical care is explicitly expressed, but the model did only detect weak morality for the foundation.

f) *“The hypocrisy here is stultifying.”* (on “Good Vegan, Bad Vegan” (Brody, 2021))

The comment with the highest probability score of 0,97 for the Fairness/Cheating foundation can be found in g). In very few words, the user describes the discussion in the comments on the respective article about veganism as a “stultifying hypocrisy”. Since there is no explicit connection to veganism when reading the comment itself without the larger context of the article, it is hard to interpret this example regarding the Vegan Discourse. Concerning its morality, however, hypocrisy is a powerful word correlated to unfairness or cheating. For the performance of the model, in addition, it is interesting to note that this comment is a very short one, which again indicates that not many words are needed for the comment to identify morality.

g) *“More proof that killing animals for food is unnecessary and, therefore, morally unjustifiable. Also, if you don't have the stomach to slaughter a pig or a cow yourself, then hiring some poor person, who has limited choices for work, to do it for you is also cowardly.”* (on “At Seasoned Vegan, Soul Goof That Never Harmed a Soul” (Mishan, 2018))

A comment with a probability score of 0,68 for the Fairness/Cheating foundation can be found in example g), representing the minority of the Fairness/Cheating labeled comments. The model additionally assigned scores for two other foundations: the Authority/Subversion foundation with 0,1 as well as the Purity/Degradation foundation with 0,3. The author of the comment argues that the unnecessity and lacking moral justification for the killing of animals for food was confirmed within the article. Moreover, they condemn people as cowards who cannot kill animals themselves for food but instead, let other people with less job opportunities do it. The reference to veganism is in the defense of the killing of animals, the moral polarization of the statement comes from the allusion to one's morals in justification of the act of killing animals. Although the fact that the reference to the Fairness/Cheating is rather implied, the model picks up a stronger indication for it.

Assessment: Fairness/Cheating in the Vegan Discourse

The comments classified as instances bearing expressions of the Fairness/Cheating foundation mostly received moderate probability scores by the model in the range from 0,44 to 0,79, resulting in an average probability score of 0,64. Although the comments are represented less in number the higher the probability score gets, the contrastive analysis did not show significant differences on the word-level per subset. The comments are mostly characterized by terms like "animals" or "rights", which allude to the principle of the foundation of a sense of justice, equality and fair treatment, and which is particularly confirmed by looking at the bigrams as well. In addition, the foundation's nature analyzed in average comment length, sentiment distribution and in the most frequent words is equally represented in the gold standard as in the comments in the entire corpus, independent of the moral loading. In general, this observation shows that this moral foundation is captured well by the model, since its gold instances are of similar character. The analysis showed overall that when users talk about veganism and they refer to their moral values based on the Fairness/Cheating foundation, they use on average more words than when talking about stereotypical veganism-related topics and explicitly express the relation to morality.

LOYALTY/BETRAYAL

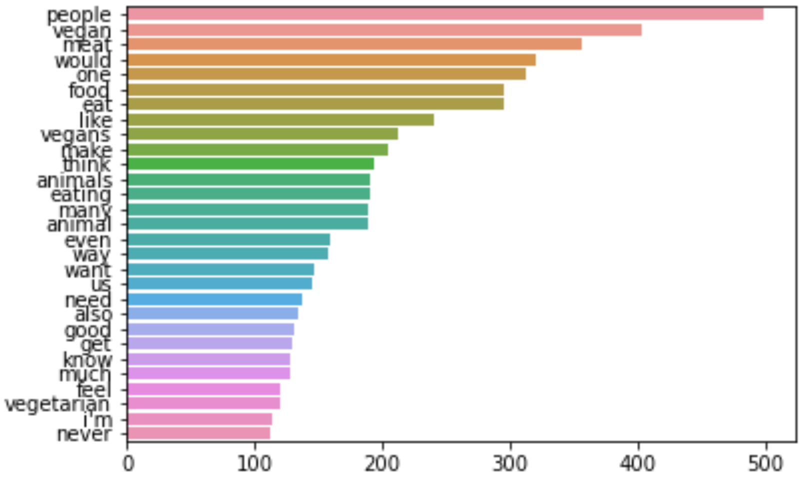

Figure 18 The most frequent occurring words in comments of the Vegan Discourse Corpus labeled for the Loyalty/Betrayal foundation (threshold 0,05)

Similarly to the Fairness/Cheating foundation's distribution, the distribution above the threshold of the comments classified as expressions of the Loyalty/Betrayal foundation decreases as the probability score increases. The peak in frequency for this foundation above its calculated threshold is at 0,05 to 0,11 with around 730 instances. For the higher probability ranges, the number of instances reaches its maximum at 2, its lowest at only 1 from 0,82 to 0,91. Thus, in general, the Loyalty/Betrayal foundation is represented significantly with low probability scores in the corpus, with an average score of 0,12. In general, the Loyalty/Betrayal foundation was assigned second most often within the range of a probability score higher than 0,05, which is primarily owed to this threshold being very low, most likely due to the small gold standard set of this work. However, for consistency, the comments will be analyzed further in the following to get an insight into the foundation's expression in the VDC. The 923 comments have an average length of 444 characters and the terms that can be observed most frequently in these comments are "people", "vegan", "meat", and "would" (see figure 18). The often occurring term "us" is the only assumably explicit term referring to the foundation's principles. The most common bigrams are typical of the VDC, like "plant based" or "eat meat". Intriguingly though, "they are" and "we are" are very common in the Loyalty/Betrayal comments, which clearly correspond to the foundation's group sense principle. Trigrams that occur frequently in the Loyalty/Betrayal assigned comments are "plant based diet", "plant based community", "plant based foods", and "eat plant based". Particularly the second mentioned term combinations are explicit regarding the "one for all, all for one" notion within the Loyalty/Betrayal foundation. In general, on the word-level, the foundation is represented with terms that are not necessarily typical for the Vegan Discourse but rather for the foundation itself, but also with some more context-fitting wordings like "plant based community". Moreover, the comments that were assigned for this foundation are mostly neutral (43,1%) or positive (28,5%). Of all, 26,7% of comments are negative, accompanied by 1,7% of Very positive comments. There is no explicit overlap on the word-level of the comments discussed in this paragraph with those that were labeled for the

Loyalty/Betrayal foundation in the gold standard. Still, the comments are nearly equal in length and have a similar distribution of sentiment, with them mostly being neutral or positive. However, in the gold standard set, there are no negative comments.

Contrastive Analysis

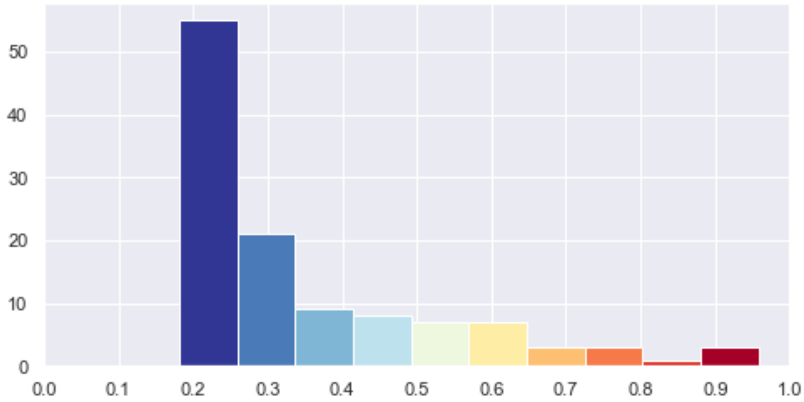

Figure 19 The probability distribution of the Loyalty/Betrayal foundation in the Vegan Discourse Corpus greater than the threshold

Most of the instances higher than the lower boundary, around 55, are classified within a range of 0,18 to 0,26 (see figure 19). The next peak is with around 20 instances at the range 0,26 to 0,33. In the higher ranges, there are less than 10 instances per range, with the highest range being 0,88 to 0,96 with roughly 2 instances. Subset 1, in this case, consists of 75 comments that received a probability score lower than 0,33 but greater than 0,18. Subset 2 contains comments in the probability range from 0,73 and greater, in total 7. These are on average 201 characters long and hence, shorter than the comments in subset 1 with around 419 characters on average. In subset 1, the word "people" occurs most frequently, whereas in subset 2, it comes at 7th place. Rather, the word "would" along with "vegans", "meat", "reduce", and "consumption" are dominant in subset 1. Words of particular interest regarding this foundation which appear often in subset 1 are "respect", "religious" and "Christian". The common bigrams in subset 1 are "eat meat" or "ten commandments", common trigrams are "seeks support position", "could care less", or, particularly interesting, "thanks sharing perspective". Particularly the religion referring combinations do not show considerable indication for the Loyalty/Betrayal foundation's group-oriented principles, since they lack in reference to the "one for all, all for one" principle of this foundation. In subset 2, the most occurring bigrams are "reduce consumption" and "lives matter". Both bigrams seem not entirely explicit regarding a relation to the principles of the Loyalty/Betrayal foundation. In the trigrams of subset 2, there is no explicit reference to foundations principles either with term combinations like "eating meat may" or "may political issue".

Regarding the general sentiment in both subsets, it can be stated that 47,9% of all comments in subset 1 are neutral, whereas about half the comments in subset 2 are neutral (28,6%). In comparison, though, it is interesting to observe that subset 1 is considerably more negative in sentiment (27,4%) than subset 2 (14,3%). Regarding positive sentiment, 27,4% of all comments are marked positive in subset 1, while it is the majority of comments in subset 2 that is positive with 42,9% of all comments. In addition, the label Very positive was assigned for instances in subset 2 (14,3%) and not at all in subset 1. All in all, it is difficult to argue for the reason behind the fact that the model assigned higher scores for some instances and lower scores for others, since both subsets do not necessarily account for very explicit references to the group-oriented foundation. If any, it is subset 1 that has more significant terms that appear to be more explicit regarding this foundation, which is interesting given the fact that it is the subset containing the comments with lower probability scores. Subset 2, contrastively, covers comments with higher probabilities but its features do not show explicit references as one would expect given the higher scores, which, however, could be reasoned with the fact that the second subset is particularly small and thus not necessarily representative.

QUALITATIVE INSIGHT

Nevertheless, some examples and further observations will be presented in the following paragraphs to get a better grasp on the foundation's expression in the Vegan Discourse.

h) *"No single diet works for everyone. I've been an ethical vegetarian for 35 years and would never go back. But I do take certain supplements that are derived from animal products, and my health appears to have improved as a result of taking those supplements - and I make no apologies for that. People who are glam bore me to tears, when they're not making me want to throw up; on the other hand, people who are psychologically deep and personally authentic command my attention and respect. Eat what your body tells you feels right, do it consciously, responsibly - and restrict your preaching to the choir. The rest of us could care less."* (on "Vegans Go Glam" (Gordinier, 2015))

Example h) is a representative for the majority of comments labeled with lower scores for the Loyalty/Betrayal foundation. It was assigned a probability score of 0,18 for this foundation, but also a 0,32 for Care/Harm, 0,48 for Authority/Subversion, 0,39 for Purity/Degradation and 0,25 for Non-moral. Although the user states they have been vegan for a long time, they argue that everyone should listen to their body and eat what feels right in a conscious and responsible way instead of participating in the "do-gooder derogation" (Minson & Monin, 2012). The low score for the Loyalty/Betrayal foundation was sparked by the user's implied sense for a group, a community of people that likely live the way they do ("The rest of us could care less"). The indication for the Care/Harm foundation can be found in the user's reference to the ethics, for Purity/Degradation it is in the description of the user's health and for Authority/Subversion, it is about the user's emphasis on respect and attention for certain people. The Non-moral score is rather neglected since the probability scores for all of the foundations assigned for this comment are above the lower boundary, making it less likely to be non-moral and more likely to have a weaker moral loading.

i) *"Fabulous article, NYT! I had no idea about the history of veganism in the US, especially with regards to the black population. In my country, I feel that vegans are too segregated. I would love to reach out to other vegan cultural groups than my own, and I think most vegans I know would agree. I just don't know how, but I am aware that, on the whole, vegans despise racism. If you abhor one injustice, you are more likely to abhor all injustices. In my mind, black lives matter and all lives matter. That is solidarity, my friends!"* (on "Black Vegans Step Out, for Their Health and Other Causes" (Severson, 2017))

Comment i) is about a user's compliment for the article, about which they elaborate on the black vegan community and its history. In addition, they claim that vegans are likely to be anti-racist, because they fight against the injustice towards animals and are thus likely to do the same with respect to other injustices like racism – which the author calls solidarity, thus sparking the detection of the Loyalty/Betrayal foundation with a probability score of 0,83. The way the author argues reminds of the "one for all, all for one" principle that Dehghani et al. (2021) name as a characteristic for this foundation. In this example, the reference to veganism is expressed explicitly. Hence, this example is an eligible illustration of a Vegan Discourse comment with high morality regarding the Loyalty/Betrayal foundation.

j) *"This is glorious news! I look forward to supporting them."* (on "The New Menu at Eleven Madison park Will Be Meatless" (Anderson & Gross, 2021))

The highest scored instance within the foundation Loyalty/Betrayal can be found in example j). Its probability score reaches a value of 0,96, indicating strong morality. In addition, the model classified a probability score of 0,11 for the foundation Authority/Subversion. The user in this statement is happy that, adding the context from the article, a popular restaurant offers only meatless meals and is excited about supporting it in the future. Although this comment is short, its moral underlying of loyalty in the sense of support becomes clear. However, this is not explicit in the context of the Vegan Discourse because of the lack of context from the article.

<u>ASSESSMENT: LOYALTY/BETRAYAL IN THE VEGAN DISCOURSE</u>

In total, the Loyalty/Betrayal foundation is not represented substantially in the Vegan Discourse. Although there are a few explicit indications for the "one for all, all for one" principle (see comment i), for instance), the on average longer comments with weaker probabilities dominate. Overall, the comments classified with higher probabilities for this foundation above its threshold become less in number the higher the score gets, which is also reflected in the average probability score of 0,12. On the word-level, it must be noticed that the terms that are most indicative for this foundation and occur in the comments frequently are rather specific for the foundation itself and not necessarily for the Vegan Discourse, even though some words such as "plant based community" are suitable for the last case. The comments are mostly neutral or positive, which is similar in the gold standard. The average comment length in characters is slightly longer for the comments in the entire corpus than for those in the gold standard. However, the most frequent words in the comments above the threshold are not found in the gold standard. Regarding the contrastive analysis, it needs to be denoted that the comments in the lower probability score range are characterized by more explicit terms than the comments with higher probability scores. Nonetheless, this could be an issue of representation, since the number of comments in the higher probability ranges is rather small. Overall, the Vegan Discourse is second-most associated with the moral foundation Loyalty/Betrayal, which means that users discussing veganism tend to identify with the values of this foundation, like a particular group-sense, and thus, express those, but they do so not necessarily in the context of veganism based on the word-level, meaning explicitly. However, the representation of this foundation in the Vegan Discourse Corpus needs to be considered with caution, since it is based on a very small representation in the gold standard according to the annotators' estimates. In addition, the evaluation metrics for the Loyalty/Betrayal foundation in the gold standard were poor regarding the agreement among annotators and the model's performance.

AUTHORITY/SUBVERSION

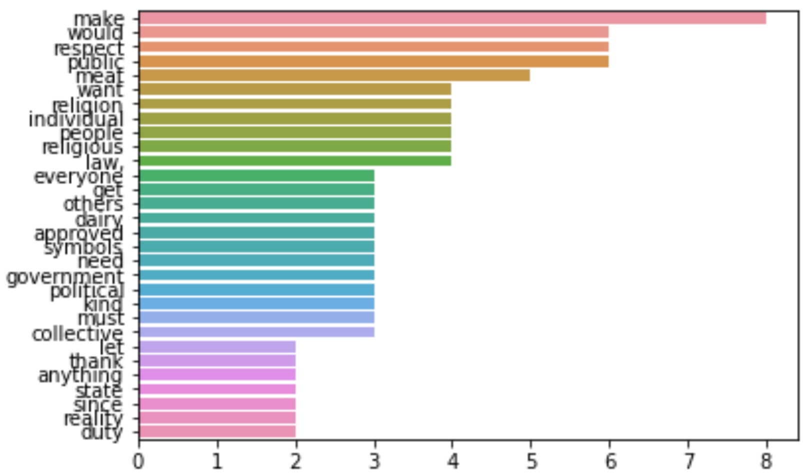

Figure 20 The most frequent occurring words in comments of the Vegan Discourse Corpus labeled for the Authority/Subversion foundation (threshold 0,0,59)

The distribution of the comments that were labeled for the Authority/Subversion foundation is similar to those of the Fairness/Cheating and Loyalty/Betrayal foundation. The most comments are classified with a score in the range from 0,59 to 0,64 with a peak in this range with 5 instances in total. In the higher probability score ranges, the number of instances reaches its maximum at 3 and its minimum at 1. The highest score an instance was assigned regarding this foundation is at 0,83. All things considered, the Authority/Subversion foundation is represented in the comments with more moderate scores than higher ones, which is also reflected in the average probability score of 0,67 for all instances above the foundation-specific threshold of 0,59. In total, 20 comments can be found in the probability ranges above this threshold. The average of these comments is 297 characters long and often consists of the words "make", "would", "respect", and "public" (see figure 20). In addition, words that occur often and are of particular interest for this foundation are “government”, “public”, “religion”, “law”, and “approved”, for instance. Regarding bigrams, the ones that appear most often within the Authority/Subversion comments are “individual choice”, “meat dairy” and “legislature approved” among combinations like “civil disobedience” or “religious symbols”. The most common trigrams within these comments are “respect nutritional choices” and “let thank everyone”. The wording of the Authority/Subversion comments of the entire corpus is in relation to the principles of the foundation, explicitly with the terms “respect” or “disobedience”, for instance. The sentiment of the comments is mostly neutral (50%) and more negative (25%) than positive (15%). In addition, 10% of these comments are labeled as Very positive, which accounts assumably for an equal distribution of positive and negative sentiment in these comments. Interestingly, the comments in the gold standard set that received the gold label for the Authority/Subversion foundation are almost entirely neutral and less negative – there is no positive sentiment among these comments. Although the comments discussed in this paragraph are less similar in sentiment to those of the gold standard and more than twice as long in characters, there are overlaps on the word-level with explicit words like “legislature” and “legislature approved”.

CONTRASTIVE ANALYSIS

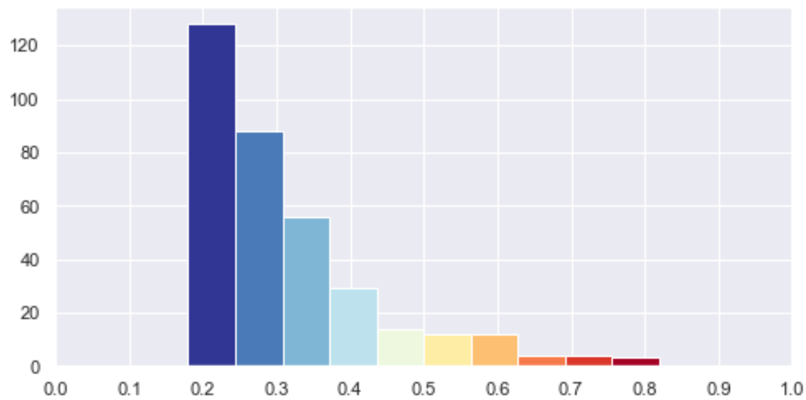

Figure 21 The probability distribution of the Authority/Subversion foundation in the Vegan Discourse Corpus greater than the threshold

The highest probability score for the Authority/Subversion foundation is a 0,82, represented only once in the data (see figure 21). From the highest score to 0,5, there are less than ten instances classified per range. Starting at the range 0,38 to 0,5 with roughly 16 to 25 comments classified for the Authority/Subversion foundation, the number of instances increases. There are around 58 instances with a score in the range from 0,31 to 0,38, about 85 from 0,25 to 0,31 and around 125 from 0,18 to 0,25. Hence, most of the instances with indications for the Authority/Subversion foundation received very low scores by the model. To investigate further what made the model decide on different probability scores, the comments are separated into two subsets. Subset 1 consists of all comments with a probability score between 0,18 and 0,31 to capture both peaks in the lower probability range. There are not many instances that were classified with a high probability score from 0,73 upwards, but still, the threshold for subset 2 will be set consistently. Hence, the comments in subset 2 received probability scores from 0,73 upwards, which accounts for subset 2 to be rather small with 4 instances in comparison to subset 1 with 248 instances. The comments in the latter have an average length of 448 characters, in the former they are on average 286 characters long. Subset 1 is comprised of the words people" or "meat" that occur frequently. The word "vegan" is present in the most common words of subset 1 but not at all in subset 2. Of particular interest is the fact that the words occurring most frequently in subset 2 are less general words, such as "moral", "livestock", "meat", "make", "lands" or "U.S.", among words like "country", "respect", "legislation" or "leadership". Contrastively, in subset 1, the most frequent words seem rather general in the context of the Vegan Discourse, which is also reflected in the most common bi- and trigrams, like "eat meat" or "plant based" or "plant based diet" and "reducing meat consumption". In subset 2, the most frequent bigram is "meat production", but all other bi- or also trigrams do not occur frequently, which is due to the fact that subset 2 is rather small. Still, the word-level observations make subset 2 more semantically distinct from subset 1 and also, more explicit regarding the Authority/Subversion foundation.

Concerning general sentiment, it can be claimed that both data sets are predominantly neutral with 38,3% of all comments in subset 1 and 50% of all comments in subset 2, making subset 2 more neutral than subset 1. Subset 1 is characterized with 32,9% of all comments being negative and 27,9% of all comments being positive. Subset 2 does not contain any negative sentiment but only positive sentiment (25%) and Very positive sentiment (25%), whereas for the latter, subset 1 was only 0,8%. Overall, subset 1 with on average longer sentences seems more general with regard to the Vegan Discourse on the word-level but is more polarized concerning positive and negative sentiment compared to subset 2. In subset 2, there are more specific words related to the Authority/Subversion foundation and the comments are shorter.

QUALITATIVE INSIGHT

In the following, it is interesting to look at some samples to get a clearer picture of the model's classification results for this foundation.

k) *"@[USER], it's likely the FAUX News-viewing parents who feel that veganism has been weaponized as an attack from the liberal left. How dare the liberals interfere with their kids' destinies for obesity, Type 2 diabetes and heart disease!"* (on "When Your Kindergarten Goes Vegan" (Tomky, 2019))

The model labeled comment k) with a probability score of 0,18 for the Authority/Subversion foundation. In addition, it classified the instance to be morally based on the Care/Harm foundation with a score of 0,36 and the Non-moral category with a score of 0,58. The comment's author states that, in the context of vegan kindergartens, "FAUX News-viewing parents" see veganism as a tool of the political liberal left and ironically questions how this community dares to interfere in the children's lives destined to get diseases. The author implies through irony that these diseases could be prevented by eating vegan. Interestingly, there is a strong reference to politics with explicit expressions like “liberals”, which likely caused the model to assign the Authority/Subversion for the instance as a hint to hierarchy based on the training corpus. The Care/Harm foundation was likely sparked in assignment by the reference to diseases. Similarly to the examples of low probability scores for the respective foundations before, the Non-moral score for this instance is rather neglected with regard to the assignment of two moral foundations simultaneously with scores above the defined thresholds.

l) *“Thank you for your progressive and responsible leadership. Your new menu is better for the planet, for the animals, and for our health. We look forward to supporting you, and we hope your example encourages many others to make the change”* (on “The New Menu at Eleven Madison park Will Be Meatless” (Anderson & Gross, 2021))

More importantly, however, is comment l). It was written regarding the same article as comment k) and states, basically, the same: the user is happy about the chef’s “progressive and responsible leadership” and is excited about supporting them in the future. Additionally, though, the author claims that the restaurant contributes to the improvement of the situation of the planet, the animals and one’s health, thus referring to the principles of veganism. The model assigned a score of 0,7 for the Authority/Subversion foundation, that can be reasoned with the explicit reference to the foundation’s principles with terms like “leadership” or “example”. Hence, this comment is an example for the expression of one’s Authority/Subversion based morality in the Vegan Discourse. Notably, the model also assigned probability scores for the Loyalty/Betrayal foundation of 0,63 and the Care/Harm foundation of 0,47, which are both above their defined thresholds an which were likely triggered by the reference to “support” and the benefit of the new menu for the planet, animals and health.

m) *“that is leadership!”* (on “The New Menu at Eleven Madison park Will Be Meatless” (Anderson & Gross, 2021))

Comment m) is the one classified with the highest probability score for the Authority/Subversion foundation with a value of 0,82. With the knowledge about the respective article’s content, one knows that the author of the comment calls the restaurant’s chef that decided to go meatless with their dishes a role model. Specifically by using the term “leadership”, the implied appreciation of the chef’s decision is emphasized, sparking a notion of the Authority/Subversion foundation. There is no explicit reference to the Vegan Discourse in the comment itself, but this example still shows the illustration for a comment with underlying morality for the Authority/Subversion foundation and proves the model’s functioning.

ASSESSMENT: AUTHORITY/SUBVERSION IN THE VEGAN DISCOURSE

In total, the Vegan Discourse is weakly characterized by the foundation's principles of leadership, authority and respect for traditions, since it has not been detected in a significant number of occurrences and if so, with lower than higher probability scores, meaning with weaker moral loadings than their opposite. Above the threshold, it can be stated that the foundation is represented with moderate probability scores resulting in an average score of 0,68. Terms like "disobedience" or "respect" are more prominent in the comments with higher probability scores, which makes the comments with lower probability scores more implicit in their wording. The instances in the gold standard are similar to those of the entire corpus based on the word-level but not concerning sentiment or comment length. Still, the gold instances for this foundation are very few, making any assumptions on this rather difficult. With regard to the Vegan Discourse, though, it can be claimed that when discussing veganism, people do not frequently express moral values based on the Authority/Subversion foundation explicitly, meaning they do not integrate the principles of respect or traditions in this context, and if so, they do not tend to be more negative or positive in sentiment. Still, the observation the weak moral loadings equal implicitness and vice versa can be claimed for the Authority/Subversion foundation.

PURITY/DEGRADATION

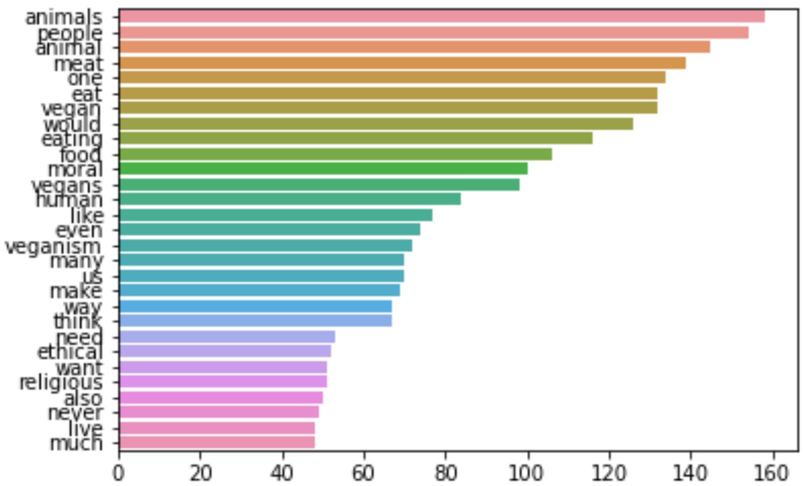

Figure 22 The most frequent occurring words in comments of the Vegan Discourse Corpus labeled for the Purity/Degradation foundation (threshold 0,27)

The comments that received probability scores for the Purity/Degradation foundation above the foundation's threshold at 0,27 decrease in number as the probability score increases. Hence, the highest range from 0,92 to 0,96 is represented in 7 instances, whereas the biggest number of instances is represented in the range from 0,27 to 0,33 with around 100 instances. Overall, the comments that were assigned for the Purity/Degradation foundation received mostly weak to moderate scores, which is confirmed by the average probability score of 0,47 for all instances above the threshold. The average comment length for the 335 comments classified by the model is 510 characters. Words that can be found often in these instances are "animals", "people", "animal", and "meat", but also, more importantly with regard to the foundation, terms like "moral", "live", and "ethical" (see figure 22). Term combinations that are very frequent in these comments are "eating meat", "eat meat", and "animal products". The trigrams that characterize these comments most are "plant based diet", "eat animal products" or "factory farmed animals", which are similar to the bigrams rather stereotypical for the Vegan Discourse and not necessarily explicit regarding the foundation. However, the trigrams "stop eating meat" or "animal rights movement" indicate some explicit reference to morality. On the word level, it is observable that the reference to the principle of living a less carnal and more noble life is being referred to with few explicit words. Moreover, the comments' sentiment is mostly neutral (39,9%), but, in comparison, more negative (32,7%) than positive (27,4%). In the comments of the gold standard, the sentiment is prominently neutral, but twice as positive than negative, which shows a different behavior in sentiment in the entire corpus' Purity/Degradation comments. Furthermore, regarding the most frequent words and their combinations, there are no similarities with explicit foundation-related words, but more with the typical Discourse-referring terms in the gold standard set and the comments of the entire corpus assigned for this foundation.

<u>Contrastive Analysis</u>

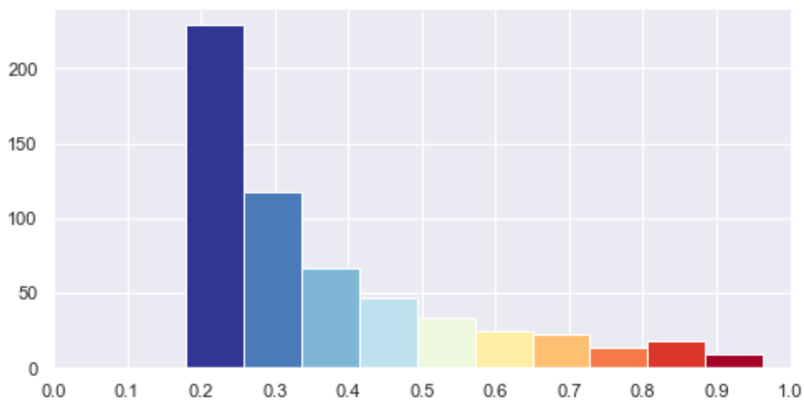

Figure 23 The probability distribution of the Purity/Degradation foundation in the Vegan Discourse Corpus greater than the threshold

Similar to all other foundations except the Care/Harm foundation, the distribution of the Purity/Degradation foundation is high in the lower scores, and low in the higher scores (see figure 23). About 240 instances are classified with a score in the range from 0,18 to 0,26, followed by around 110 instances in the range from 0,26 to 0,33. Around 60 comments were classified within the score range of 0,33 to 0,41. Everything above is represented by less than 50 instances, decreasing in number as the score increases. Still, two subsets were created to obtain an insight into what the features of the two probability ranges lower than 0,33 and greater than 0,73 distinguishes from another, to reason the model's classification results and to get a better understanding of the implicitness or explicitness of morality within the Vegan Discourse. Thus, subset 1 contains 340 comments with lower probability scores and, in contrast, subset 2 consists of only 40 comments with probability scores higher than 0,73. Still, it is interesting to look at the following observations. The comments in both subsets are not equally long in words, with subset 1 having an average comment length of 532 characters and subset 2 an average comment length of 420 characters. The most common words are similar in both subsets, with words like “people”, “animals” or “meat”. However, among these words in subset 2 are also some that are more explicit to trigger a detection for the respective moral foundation, such as “disgusting”, “purity” or “religious”, although the latter is not always indicative for this foundation. Accordingly, in subset 2, there are also more particular bigrams like “animal flesh” or “animal products” than compared to the standard bigrams of the Vegan Discourse like “plant based” or “eating meat” that occur in subset 1. The similar case is depicted in the trigrams, where for subset 1, the most common ones are “plant based diet”, “plant based food” or “stop eating meat”. However, there is also a term combination like “ethical environmental reasons” indicating morality to some extent. In subset 2, most frequent meaningful trigrams with regard to the respective foundation in this section are “would live see” because of its possible implication to a more “elevated live” and “let alone slaughtered” with its reference to carnality. Intriguingly, in subset 2, the reference to the “less carnal way of life” that is sought by people that have high moral values based on the Purity/Degradation foundation becomes evident.

Like all other foundations, the comments of the Purity/Degradation foundation that were part of the two subsets are, concerning general sentiment, mostly neutral (40% subset 1, 45% subset 2). Concerning negative sentiment, subset 2, with 32,5% of all comments, is only a little more prevailing than subset 2 with 31,7% of all comments. Contrastingly, subset 2 consists of 22,5% positively annotated comments and subset 1 of 28% positively annotated comments. Hence, the comments in subsets 1 and 2 are not equal in average comment length, and neither so in the general sentiment that was assigned for them. Regarding the differences observed in the word-level features, it can be claimed that subset 2 contains more explicit words like “disgusting” or “let alone slaughtered” that likely spark the foundation’s classification with higher probability scores.

<u>Qualitative Insight</u>

A more detailed look at some of the classified instances for this foundation will be presented in the following paragraphs to get a more fundamental grasp on its meaning in the Vegan Discourse.

n) *“I agree that people should not distort science and tell lies to make their point. The problem, as seen in this article, is that the ethical argument against hurting animals, which is by far the most compelling reason to stop eating animal products, always gets short shrift. Commentary on dieting generally ignores animal suffering completely or relegates it to an afterthought, just an additional benefit among others, with the main focus being on the individual's health. This article is typical in taking that approach. Those of us who are urgently concerned with ending animal suffering can feel tempted to lean on the health argument, making inflated claims about the health of vegan diets in an effort to persuade people, because persuading people is so important. But it's a misguided approach, and the wrong focus. Abusing sentient animals as slaves and commodities is always wrong, regardless of its impact on our health. That needs to be the* *resounding message of any discussion of the subject.'"* (on "Good Vegan, Bad Vegan" (Brody, 2021))

In comment n), its author argues that the claim that veganism is good for one's health puts the argument to end the cruelty towards animals in the background, although the author finds this significantly more important than the health claim. The model assigned a probability score of 0,18 for the Purity/Degradation foundation, which can be reasoned with the reference to the immoral activity of abusing animals like slaves, as the user frames it. Moreover, the model declared this instance as highly likely for the Care/Harm foundation with a score of 0,72, which makes sense with the author's expressions of the harm done to animals and their suffering. In addition, the instance also received a score of 0,19 for the Authority/Subversion foundation for the slight implication to the persuasion of people. All moral foundations scores are above the lower boundary of non-morality, but the Purity/Degradation foundation score is most interesting in this case, since the comment shows an example of its implied morality.

o) *"@[USER] What vegans care about is the planet and her inhabitants. All of them. This is our sacred duty."* (on "Stop Mocking Vegans" (Manjoo, 2019))

An expression of the striving towards a more "elevated, less carnal, more noble way" (Dehghani et al., 2021) is illustrated in example o). The user calls a vegan's aim of avoiding harm to the environment and its inhabitants everyone's "sacred duty". The model hence assigned a high score for the Purity/Degradation foundation of 0,95, but also noticed some morality founded on Care/Harm with a score of 0,23. This comment shows how veganism can be presented and perceived as something sacred that relies on one's moral foundation of Purity/Degradation.

p) *"You could not be more WRONG! You have it exactly backwards. MEAT-EATING IS THE RELIGION: Veganism is liberation from religion. Not only are all the things you mention taken directly FROM religion, you forgot about these aspects: It is a belief system imposed by society. It considers itself the One True Way. It is based on myth and historical momentum. It depends upon herd mentality. It inculcates and demands rote conformity. It requires blind faith of its adherents. It manipulates human emotions, particularly fear, prejudice and greed. It does not countenance objection. It entwines itself into the very fabric of daily life. It celebrates itself in legend, pomp, ceremony and ritual. It glorifies human superiority, greed and gluttony. It debases life. It cares nothing for "the other." It cares nothing about the Earth. Blood sacrifice by someone or something is required. It is based upon strict hierarchy. It "frees" the adherent from having to think for themselves. It does not hold up to critical scrutiny. It is anti-science. It is anti-spiritual. It pretends to preach virtue, but practices the opposite. It goes to great lengths to hide its own sin. It separates and divides from the rest of world. It fears, denies, rejects, mocks and attacks competing concepts. It justifies violence in order to defend its dogma. It wreaks great damage*

> *upon the individual, human society and the planet. Its true believers perceive none of the above."*
> (on "Good Vegan, Bad Vegan" (Brody, 2021))

Comment p) is a very long statement by a user claiming meat-eating to be a religion and naming all kinds of behaviors and aspects to frame it as such. The author's descriptions triggered several foundations to be detected by the model: Purity/Degradation with a score of 0,84, Care/Harm with 0,55 as well as Authority/Subversion with 0,2. The highest score is assigned to Purity/Degradation, which results from the comment's detailed enumeration of immoralities, like "Blood sacrifice by someone or something is required" or "It glorifies human superiority, greed and gluttony". This comment thus presents well how the moral foundation Purtiy/Degradation can be framed in the Vegan Discourse in an explicit way.

ASSESSMENT: PURITY/DEGRADATION IN THE VEGAN DISCOURSE

The Vegan Discourse is well represented by numerous instances with regard to the Purity/Degradation foundation. Either the foundation's principle of living a more elevated and noble life in the sense of choosing a moral and less carnal way of life with veganism is expressed in the Vegan Discourse, or the foundation's base related to disgust, contamination or diseases in the context of the killing of animals or the harming of the environment are subject of the discussion. The general probability distribution for the Purity/Degradation assigned comments is characterized by few highly scored and more moderately or weakly scored comments, reaching an average probability score of 0,47. The usage of some explicit words is observable in the comments above the threshold, and the higher scored instances are more expressive of these words with terms like "disgusting" or "let alone slaughtered". The lower scored instances are more implicit and show no overlaps with the higher scored comments in average comment length or sentiment. This is also the case with the gold standard and the comments with a probability score higher than the threshold – there are no overlaps in features, which could result from the small gold standard set. Overall, people participating in the Vegan Discourse fewer times express their values based on the Purity/Degradation foundation explicitly than implicitly. Still, its expression was detected third most dominant in the entire corpus.

NON-MORAL

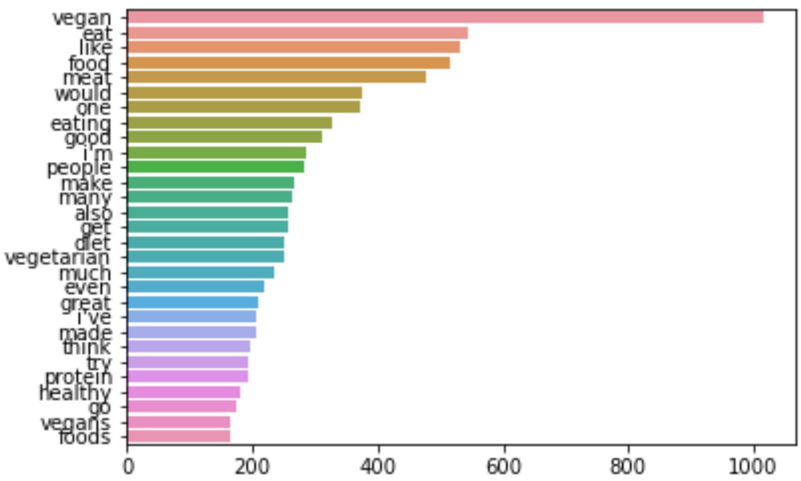

Figure 24 The most frequent occurring words in comments of the Vegan Discourse Corpus labeled for the Non-moral label (threshold 0,98)

For the comments that were assigned a probability score higher than 0,98 as was defined for the Non-moral label, the majority is represented in the highest range from 0,986 to 0,988, in total around 600 comments. In general, the Non-moral comments reach an average probability score of 0,99 and thus, are represented with significantly high probability scores, which is evident considering the high threshold defined for this label. The in total 3'651 comments assigned with the Non-moral label in the corpus with probability scores higher than the threshold consist of 161 characters on average and are mostly represented by words like "vegan", "eat", "like", and "food" (see figure 24). The most common bigrams within these comments are "plant based", "ice cream", "fast food", and "vegan diet", which are all typical for the Vegan Discourse. Similarly, the trigrams that appear most frequently are all combinations with the term "plant based", such as "plant based diet" or "plant based meat", but there are also phrasings like "vegan ice cream" among the most frequent trigrams. Consequently, the comments assigned with the Non-moral label above the threshold are mostly represented with stereotypical words for the Vegan Discourse.

The distribution of sentiment in these comments is as follows: 52,6% neutral, 20,2% negative, 25,8% positive, 1,2% Very positive and 0,1% Very negative. This accounts for a mostly neutral and a less negative than positive data set regarding sentiment. In comparison to the gold standard of the Non-moral comments, it becomes clear that the comments of the entire corpus are on average slightly shorter than those of the gold standard. On the word-level, it can be documented that there are overlaps in explicit wordings like "meat" and "eat" among the most frequent words, as well as "vegan ice cream" as a frequently appearing trigram, for instance. The sentiment distribution is similar in both data sets regarding the neutral sentiment. In addition, the comments of the gold standard with the NONE label are more positive than negative, which is similar in the case of the comments in the entire corpus labeled as Non-moral.

CONTRASTIVE ANALYSIS

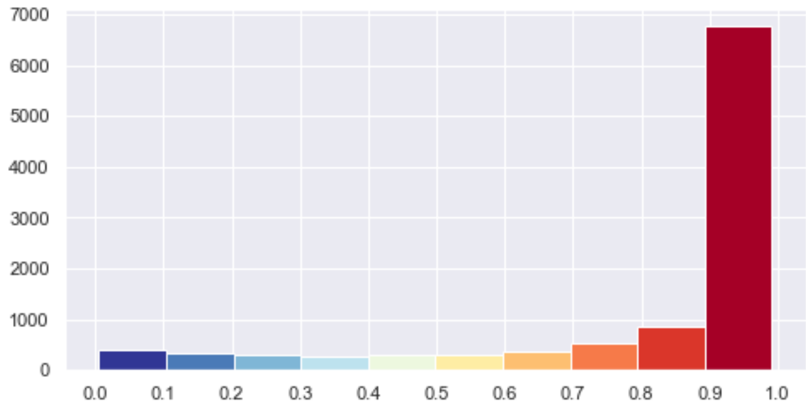

Figure 25 The probability distribution of the Non-moral label in the Vegan Discourse Corpus greater than the threshold

Around 6'500 comments were classified with the Non-moral with a probability score higher than 0,9 (see figure 25). Very few instances received lower probability scores, as can be seen in figure 12. Hoover et al., in their annotation guidelines, advise: "If a document does not have any moral content, it should be coded as […] nonmoral" (2020). Therefore, if one only looks at these classification results, most of the VDC would consist of instances with few to no indications for morality. In the MFTC as training corpus for this analysis, this label dominated across topic domains, too (Hoover et al., 2020), making the meaning of the label itself questionable – if the majority of tweets or comments is labeled with it and the labeling purpose is to identify morality in text, what assumption can be drawn from it, what does it suggest? What difference does its annotation make, what effect does it have? Assumably, the label seems to be lacking more definition, since the model not only captured its occurrence in comments exclusively, but also in combination with a moral foundation. In these cases, as was explained in chapter 4.3, the Non-moral label will not be treated equally as important as the moral foundations one(s), if the score for the moral foundation(s) is greater than the low boundary of 0,18. Hence, if a comment received a high probability for a moral foundation as well as a high probability for the "Non-moral" label, the former is more significant for further analysis and interpretation. To further reason this decision, it is interesting to report that Kobbe and colleagues (2020) renamed this label "Moral" instead of "Non-moral", since they claim to have found results for morality in the Non-moral instances.

For consistency and to get a better understanding of the model's classification results for this label, I will look at two subsets of comments that are different in their assigned probability scores. Subset 1 consists of the lower classified comments regarding probability within the range from 0,18 to 0,33 to be consistent in the procedure. Subset 2 contains comments within the probability range from 0,73 upwards to capture the highest peak but still be consistent. Thus, subset 2 consists of 8'015 comments with an average length of 236 characters, while in subset 1, there are 446 comments with an average length of 470 characters. Among the most frequent words in subset 1 are "animals", "animal", "people" and "meat", thus accounting for animal-focused comments within the context of veganism. In subset 2, the words “vegan”, “eat”, “meat”, and “food” occur the most, which rather represent the stereotypical words in the context of veganism as defined in chapter 3.2. Regarding bigrams, interestingly, the comments with lower probability to be non-moral mostly contain bigrams like “eating meat” or “animal products”, but also, crucially, “animal rights”. This bigram specifically is an explicit expression that could be observed within the Care/Harm or the Fairness/Cheating foundation depending on its context, for instance, and thus concur with a low probability score for the comment of being non-moral. Hence, some sort of distinction from non-moral instances seems observable. This is also confirmed after a look at the most frequent bigrams in subset 2, which are “plant based”, “vegan diet”, “eat meat”, “animal products” or “fast food”. These are all rather general regarding the Vegan Discourse and are thus, because of their non-explicitness regarding a foundation, more likely to spark non-morality. A similar observation can be made for the trigrams. Among the most frequent ones in subset 1 are “harm much possible” but also “plant based diet” as most occurring. While the first is an obvious trigger for a moral foundation and thus explains the low non-moral probability scores, the latter is rather general. This is why this trigram is also most frequent in subset 2, along with combinations like “plant based meat”, “plant based foods”, “plant based meat”, or “vegan ice cream”.

Concerning general sentiment, both subsets are classified as mostly neutral, with subset 2 being more neutral (50%) than subset 1 (38,3%). Subset 1 is dominating concerning positive and negative sentiment, with 35% negatively classified comments compared to 24,3% of such classified comments in subset 2 and 25,1% positively classified instances in subset 1 and 24,5% of those in subset 2. In addition, in subset 1, there are 1,65% of all comments classified as Very positive, while in subset 2, there are 1,1% of those, but, moreover, 0,1% of Very negative classified comments. Overall, the comments that received low non-moral probability scores are reasonably classified as such, since they contain some explicit words that would trigger a classification for certain foundations. The same observation goes for the comments with higher probability scores, although the high frequency of general words of the Vegan Discourse within these comments are likely the reason why several comments received, apart from a higher non-moral score, foundation-related probability scores, since these can sometimes be signs of implicit wordings related to a moral foundation.

QUALITATIVE INSIGHT

A closer look into the comments classified with this label will elaborate more on the label’s meaning in this corpus.

q) *“Sooner or later it will dawn on this country how stupid it really is to respect anyone's tenets of faith in legislation applicable to all. Politics has been mired in this idiocy my whole lifetime.”* (on “Satanists, Vegans and Atheists Seek Equal Opportunity in Little Rock” (Clines, 2015))

In comment q), its author complains about the politics' "idiocy" to include all kinds of faith within the general law that applies to everyone. The model assigned a score of 0,18 for the Non-moral label, meaning there would be less representation of morality in this comment. Interestingly, the word "vegan" or any Discourse-related words are not mentioned at all in this comment. Still, the Authority/Subversion foundation was labeled for this comment with a probability score of 0,76, which seems reasonable with regard to the legislation and politics reference of the author. In general, it makes sense that this comment received a lower score for Non-moral and a higher score for another moral foundation – the labels work complementary.

r) *"This is precisely the kind of "happy exploitation" that Francione is talking about. There is no good reason to kill or exploit animals when there are available alternatives."* (on "When Vegans Won't Compromise" (Fischer & McWilliams, 2015))

A rejection of the concept of "happy exploitation" with regard to the killing of animals based on the availability of alternatives is presented in comment r). While the model assigned a probability score of 0,85 for the Non-moral label, it also categorized it to contain morality of the Care/Harm foundation with a score of 0,42. The latter is likely sparked by the user's referral to the harm done to animals through exploitation. This example is particular interesting, since it seems to be rather explicit regarding the Care/Harm foundation on the word-level.

s) *"I'm a vegetarian but I am disturbed by the odor of pea protein 'meats'. It's difficult to describe beyond 'funky', like overcooked peas. So I've discovered that a healthy dose of spice like cumin, onion and garlic help to mask this."* (on "How To Cook With Plant-Based Meats" (López-Alt, 2020))

Comment s) is a cooking advice from a user to cover the smell of pea protein-based meat substitutes with spices. According to the model, there is no morality in this comment, and it assigned a score of 0,99 for the respective label, which is above its defined threshold. All other foundation scores are close to 0. The model's classification seems logical in this example, given the fact that no other foundations were sparked in the model's classification.

<u>ASSESSMENT: NON-MORAL IN THE VEGAN DISCOURSE</u>

Differently to the moral foundations' probability scores, the number of instances increases as the probability score rises. Hence, the model is very certain when it comes to non-morality in the instances of the Vegan Discourse. Mostly, these comments consist of rather stereotypical words, which could also be observed in the gold instances for this label. The sentiment distribution of the Non-moral comments above the threshold is similar to the distribution in the gold_strict set. The comments that received lower probability scores for the Non-moral label are characterized by explicit wordings that allude to principles of morality. Overall, in the corpus, non-morality is represented by typical or general terms used when referring to veganism, such as “eat” or “meat”.

5.2 Storytelling

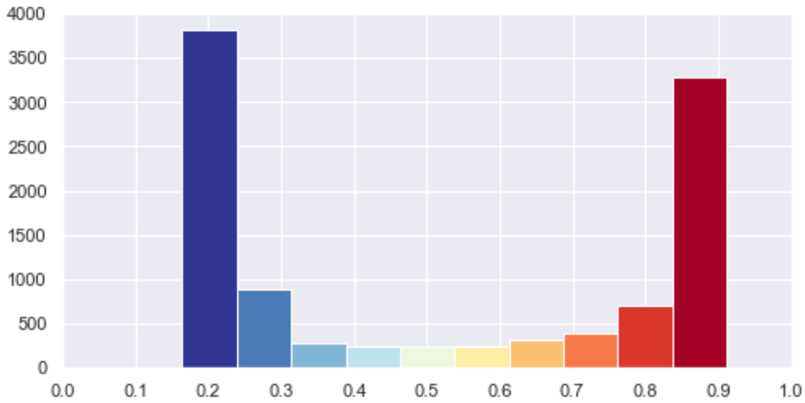

Figure 26 The probability distribution of Storytelling in the Vegan Discourse Corpus

Of all 10’432 instances, 5’588 are classified as non-storytelling and 4’844 instances as storytelling with a threshold of 0,5. Hence, around 53,6% of all instances of the corpus show no storytelling signs and roughly 46,4% of all instances do so, making a little less than half of the corpus storytelling statements. In figure 26, one can observe that most of the comments are either highly probable to be storytelling with a probability of 0,8 or more, or rather unlikely to be storytelling with a probability of 0,3 or less. All probability ranges in between are less represented in the Vegan Discourse Corpus with less than 500 instances. Interestingly, there are no representatives of the VDC in the ranges up to 0,16 or from 0,91 upwards, meaning there are no strong poles regarding probability in this corpus. Roughly 3’750 instances are classified in the lower peak between 0,16 and 0,23 and around 3’200 comments received a probability of 0,84 to 0,91.

The same approach for a feature analysis as in chapter 5.1 regarding the moral foundations will be pursued to get a more fundamental insight into the (non-)storytelling instances in the Vegan Discourse. The average length of all storytelling comments is 362 characters and of all non-storytelling instances, it is 221 characters, making the latter shorter on average than the former. The three most common words in both labels are “vegan”, “eat” and “meat”. In the storytelling instances, other frequently appearing words are “animal” and “animals” and “I’m” and “I’ve” as potential markers for storytelling. In the non-storytelling instances, the latter two do not appear frequently and are not listed within the top 30 of the most frequently appearing words. “Animals” also occurs often in these instances, though, along with “people”, for instance, which is also present in the storytelling instances, but not as much as in the non-storytelling ones. The bigrams are rather general and also similar for both aspects of interest: “plant based”, “vegan diet” or “eating meat” occur very often. In the storytelling comments, however, there are two potential markers hidden, such as “years ago” as a possible reference to a personal experience, or “they are” which could be used in a personal story. The trigrams do not allow significant assumptions concerning both aspects, since they are very similar with term combinations like “plant based diet” or “plant based foods” as most frequent, and very general. Moreover, the general sentiment distribution for the two aspects is as follows: the storytelling instances are more neutral (49,9%) than their opposite (46,2%) and both domains are equally negative (26,8% in storytelling, 26,3% in non-storytelling) but the storytelling comments are slightly more positive (25,7%) than the non-storytelling ones (22,8%). In addition, 1,3% of all comments in subset 1 contain Very positive sentiment and 0,09% contain Very negative sentiment. In subset 2, 0,9% of all comments were assigned the Very positive label and 0,2% the Very negative label. Overall, it can be noted that both the storytelling and the non-storytelling comments are characterized by rather general words. On the word-level, the on average longer storytelling comments show signs of its principles of personal experiences, with markers like “I’ve” or “I’m”. These cannot be observed in the non-storytelling instances.

In comparison to the gold standard, it needs to be documented that the word-level observations are similar concerning the instances of the entire corpus and those of the gold standard. The gold storytelling instances also frequently include the word “I’m” or the bigram “years ago”, which is not the case in the non-storytelling instances. In addition, the non-storytelling instances are shorter than the storytelling ones in the gold standard. Both insights can be observed equally in the classified instances in the entire corpus regarding (non-)storytelling. Furthermore, the sentiment distribution in the gold standard and in all instances of the corpus classified as (non-)storytelling quite similar – only, the storytelling instances are a little more positive in sentiment than the non-storytelling comments. Thus, these overlaps of the features in the entire corpus’ instances and the gold standard ones show that the model classified well. In addition, the rather general vocabulary of the storytelling instances emphasizes that the concept is a bearer of implicitness, which is also a characteristic to morality. This observation promises an interesting insight into the interaction of the two, which will be discussed in chapter 5.3.

QUALITATIVE INSIGHT

A closer look at some samples will provide additional insight into these quantities.

t) *“Thank you so much for this article. In this day and age, anyone making compassionate choices should be applauded rather than ridiculed by others. It is one thing to live with a nonconforming choice but it is another thing to be harassed for it or to be on the constant receiving end of jokes around the office cooler.”* (on “Stop Mocking Vegans” (Manjoo, 2019))

A user expressed their appreciation for the article “Stop Mocking Vegans” and argues that being a vegan as a choice of compassion should, despite its minority character, be applauded rather than ridiculed. The model picked up a probability score of 0,17 for storytelling, although one could argue for it to be higher, since some sort of personal experience is implied with “jokes around the office cooler”. In addition to storytelling, this comment has also been labeled with a

Care/Harm score of 0,58, an Authority/Subversion score of 0,31, and a score for the Non-moral label of 0,17. This indicates that although there are moral values in the comment as detected by the model, storytelling is not as present as it was suspected first in this thesis – in this comment.

u) *“There are far more serious problems than animal rights. Explain being a vegan to a refugee who has nothing to eat and no prospect of finding food in any form. First, figure out that problem and only then can you debate animal ethics. We worry about creatures beneath us in the evolutionary chain and ignore the human suffering around us. Stop the opera - human life is way more precious and oh so fleeting.”* (on “When Vegans Won’t Compromise” (Fischer & McWilliams, 2015))

In u), the comment’s author takes a stance for the prioritization of the poverty and suffering of humans, like in the case of refugees, above animal rights. They argue that animals are “creatures beneath us in the evolutionary chain” and that, consequently, a human’s life is more precious. The storytelling probability score for this instance is moderate with a value of 0,5. Interestingly, a more explicit expression of the concept than in comment t) is not visible at first sight, making it questionable why the model detected a stronger score for u). However, a storyline is visible with the example of the suffering of refugees and the description of their circumstances. The model also assigned scores for the Care/Harm foundation (0,57) and the Fairness/Cheating foundation (0,34). In this example, there are moderate indications for morality as well as for storytelling.

v) *“I am a Type II diabetic on Byetta and Metformin. I am now following a high protein/low carb diet. Protein sources are whey and soy, chicken, eggs, turkey, fish. How can I convert to vegan? I need to have a certain amount of protein per day and I don't know where to start or go for advice. I do want to make changes in my life. I need to lose more weight and get off these horrible diabetic meds.”* (on “Sculptured by Weights and a Strict Vegan Diet” (Pilon, 2012))

The highest probability score for storytelling was assigned for comment v). In this, a person tells about their Diabetic disease and the resulting medication needed, which makes it difficult for them to change their nutritionary habits from eating meat to eating vegan. For this comment, the probability score for storytelling seems reasonable because of the personal experiences described. The model also assigned a probability score of 0,96 for the Non-moral label, indicating that storytelling cannot be observed with explicit mentioning of morality.

Assessment: (Non-)Storytelling in the Vegan Discourse

In general, the Vegan Discourse as represented in this thesis is nearly as much characterized by non-storytelling comments as it is by storytelling ones. If a comment is classified as storytelling, it received either a very high probability score of 0,8 or more, or a very low score of

0,3 or less. On the word-level, the storytelling instances and the non-storytelling instances are not characterized by any specific but rather general words that are typical for the Vegan Discourse, like “eat” or “meat”. Intriguingly, though, the storytelling instances are reported to frequently consist of markers that could trigger the model’s classification results, such as “I’m” or “I’ve”, which is most likely the start to a personal reference or experience. In addition, these terms can also be found in the gold instances for storytelling, where all other features analyzed concur as well. Moreover, the non-explicit words used in the storytelling comments with high probability scores allude to the idea that these are signs of implicitness, which is, after the analysis of the classification results, argued to be characteristic to storytelling. Thus, this reveals itself promising regarding the analysis and detection of moral foundations, since morality is more often implied than expressed explicitly.

5.3 Morality and Storytelling: About moral storytellers

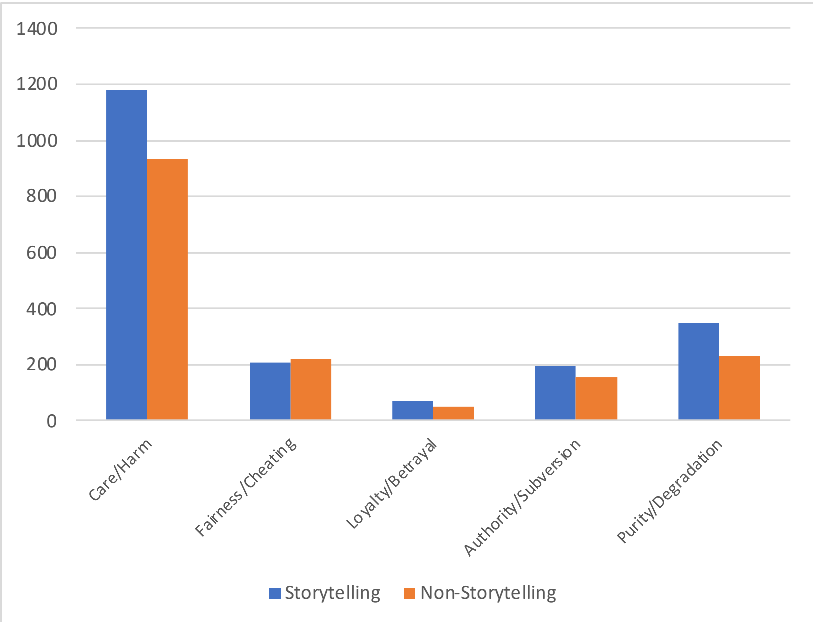

Figure 27 Distribution of (Non-)Storytelling instances across the five moral foundations with probability values above the lower boundary with a value of 0,18

Since the contrastive analysis provided substantial ground to argue for the validity of weak probabilities for the moral foundations based on implicitness in the respective comments, the overview on how many comments per foundation are stories being told or mere facts is measured including the comments from the lower boundary upwards. This is essential regarding the observation made in the chapter before: since storytelling is characterized by more general words, meaning no explicit foundation related terms, which are treated as signs of implicitness in this thesis, it is indispensable to not only cover the strongly loaded comments regarding morality, but also the weaker ones. This is because logically, I expect the weakly morally loaded comments to have high storytelling scores assigned by the model and vice versa. Thus, I argue that *implied* morality is likely to be expressed via storytelling, which can significantly contribute to further research in the detection of morality in texts. In general, in figure 27, it can be observed that across the five moral foundations, the number of (non-)storytelling identified comments varies. Still, one major observation can be documented: for four of the five moral foundations, for which the comments of the Vegan Discourse Corpus received a probability score above the defined lower boundary of absolute non-morality with a value of 0,18, the number of instances with evidence for storytelling as labeled by the model dominates over the number of non-storytelling labeled instances. Only the Fairness/Cheating foundation is characterized by the vice-versa case, with 220 non-storytelling and 207 storytelling labeled comments. Detailed numbers are illustrated in Table 8, not only for the five foundations, but also for the Non-moral category to provide full disclosure of the numbers and figures.

	Storytelling	Non-Storytelling
Care/Harm	1179	932
Fairness/Cheating	207	220
Loyalty/Betrayal	69	48
Authority/Subversion	197	153
Purity/Degradation	350	233
Non-moral	4484	5284

Table 8 Detailed numbers for the (Non-)Storytelling instances across the five moral foundations with probability values above the lower boundary with a value of 0,18

A more Detailed Look at Implied Morality in Stories

In addition, when the loadings in morality and (non-)storytelling are taken into consideration, a similar and crucial observation about the interplay of storytelling and morality can be made. As was pursued in the subsection of the contrastive analysis, the probability range for a weak loading of morality is defined from 0,18 to 0,33. The higher probability range covers scores from 0,73 upwards. For the comments with a weak moral loading as well as for those with a high moral loading, the distribution of the probability scores for storytelling are observed. The Care/Harm assigned instances of the entire corpus with a weak moral loading for this foundation are dominated by high probability scores for storytelling from 0,76 upwards, although there is a peak in frequency in the lower range. In the inverse scenario, meaning in the comments with a high moral loading for this foundation, the probability scores for storytelling behave similarly with two peaks in the very high and very low ranges, but with a smaller delta in between the two, meaning the number of instances with a lower probability score nearly equals the number of comments with higher scores. For the Loyalty/Betrayal foundation, nearly the same can be documented, except that the difference in high and low probability scores for storytelling in the morally weakly loaded instances is quite significant, emphasizing the higher probability range for storytelling in this case. The most representative foundation for this investigation is the Authority/Subversion foundation, where exactly the behavior expected according to the hypothesis can be observed and which is thus documented in the graphs below (figure 28 and 29). In addition, to not neglect comments with intermediate probability scores with respect to this observation, figure 30 displays the probability distribution for storytelling as well as for the Authority/Subversion foundation for all comments that received probability scores greater than 0,33 and lower than ,73 for the Authority/Subversion foundation.

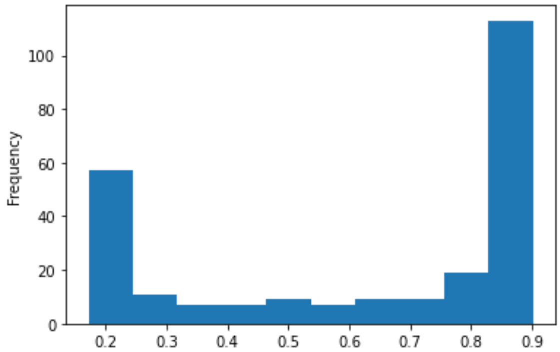

Figure 29 Distribution of Storytelling in the lower scored instances (0,18 – 0,33) labeled for the Authority/Subversion foundation

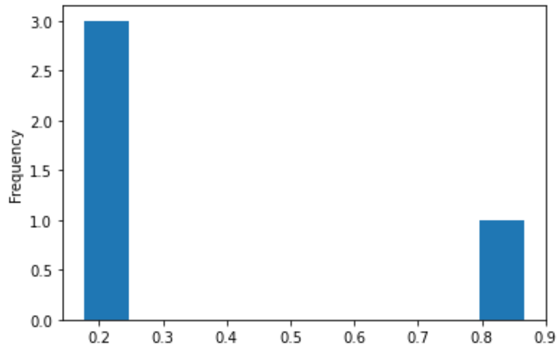

Figure 28 Distribution of Storytelling in the higher scored instances (0,73 – 1) labeled for the Authority/Subversion foundation

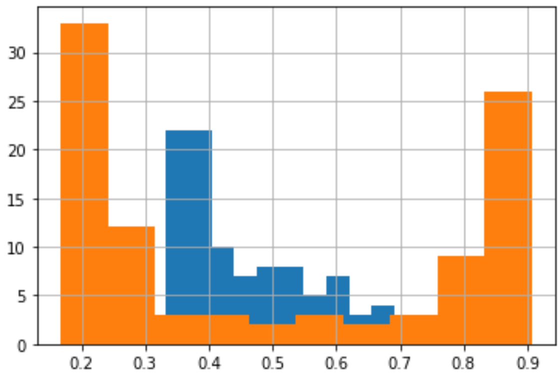

Figure 30 Distribution of Storytelling and the moral foundation scores in the moderately scored instances (>0,33 – <0,73) labeled for the Authority/Subversion foundation

Figure 28 depicts well that for the comments with weaker morality loadings for the Authority/Subversion foundation, the probability scores for storytelling are mostly in the higher range and less in the lower range. Figure 29 shows the opposite perspective: the comments with high morality loadings are mostly represented with low probability scores for storytelling. The same behaviour can be reported for the Fairness/Cheating foundation and the Purity/Degradation foundation. For the comments with scores in between in figure 30, it can be stated that the comments with lower probability scores for Authority/Subversion and storytelling dominate in number compared to the higher probability scores for both categories. Particularly, the storytelling probabilities have a dip in their distribution in intermediate probability scores. Overall, these observations reveal that, in general, comments with weak moral loadings primarily received higher probability scores for storytelling, meaning that mostly, if a person expresses their moral beliefs implicitly, they are likely to use storytelling for this. Also, if a person explicitly argues for their moral values, they do not necessarily tell a story but rather express it explicitly, for instance. Hence, comments with high moral loading predominantly receive lower probability scores for storytelling. Intermediately scored comments with regard to the moral foundations show a similar distribution for storytelling. In total, this confirms the hypothesis that storytelling enhances the detection of morality because both aspects share the nature of implicitness, which is emphasized through storytelling when morality is implied.

FACETED MORALITY

Another observation that was captured during the qualitative analysis of samples of the Vegan Discourse Corpus, but not stated as a hypothesis beforehand, is that if a user expresses one moral foundation, there might be another formulated, too. In more detail, during the qualitative look at some samples of the VCD, it appeared as if some foundations were more likely to occur together than others. To investigate this further, the Pearson Correlation Coefficient was calculated for the individual foundations with each other based on the assigned probability scores by the model to receive an insight into the correlation between the foundations. The detailed numbers are documented in Table 9.

	Care/ Harm	Fairness/ Cheating	Loyalty/ Betrayal	Authority/ Subversion	Purity/ Degradation	Non-moral
Care/Harm	-	0,1425	0,1092	0,2471	0,2937	-0,8569
Fairness/ Cheating	0,1425	-	0,1601	0,2469	0,2320	-0,4286
Loyalty/ Betrayal	0,1092	0,1601	-	0,5000	0,2486	-0,3503
Authority/ Subversion	0,2471	0,2469	0,5000	-	0,3503	-0,5152
Purity/ Degradation	0,2937	0,2320	0,2486	0,3503	-	-0,5792
Non-moral	-0,8569	-0,4286	-0,3503	-0,5152	-0,5792	-

Table 9 Pearson Correlation Coefficient calculated between the individual moral foundations with each other

Notably, the Authority/Subversion foundations has two correlated foundations: Purity/Degradation with a correlation value of 0,35 and Loyalty/Betrayal with a correlation value of 0,5. Both foundations also have a slightly positive correlation as indicated by the value of 0,25. Hence, it is likely that if an individual articulates their statement based on either of the mentioned moral foundations, they do not appear alone but in a combination of each other. In the other cases of the moral foundations, there are mostly also slightly positive correlations, meaning they might be expressed not individually, but combined with another. For the Care/Harm foundation, the most likely foundation for this case would be Purity/Degradation, while for Fairness/Cheating, it would be Authority/Subversion. In total, it can be stated that in the Vegan Discourse, if a person articulates their statement based on a moral foundation, it is more likely that another becomes intertwined in this than the opposite scenario – thus, the Vegan Discourse is rather faceted with regard to the expression of the moral foundations, although there are some that dominate and some that appear less.

6 CONCLUSION

This thesis aimed at exploring how moral the Vegan Discourse is, what kind of moral values are expressed most in it, and if there is a relationship between morality and storytelling in this context. In pursuing this, the phenomenon itself together with the three aspects were defined further and explored empirically with methods from Natural Language Processing. Essentially, the Vegan Discourse Corpus consisting of roughly 10'000 comments on articles about veganism published by the New York Times was created and functioned as the core of this thesis. With machine learning models and the creation of gold standards for both aspects of interest, this data was classified for the five moral foundations plus non-morality based on the Moral Foundations Twitter Corpus (Hoover et al., 2020) as well as for storytelling (*Reports of Personal Experiences and Stories in Argumentation: Datasets and Analysis*, 2021). The categorized data was observed independently per moral foundation and per (non-)storytelling, but also in their interaction. The analysis included quantitative and qualitative aspects and conceptual insights, which were interpreted in a larger context based on the empirical nature of this work.

<u>MORALITY IN THE VEGAN DISCOURSE</u>

In the Vegan Discourse Corpus, about a third of the corpus was classified by the model with moral validity for each foundation, which was calculated on the grounds of the gold standard created for this thesis. The majority was labeled for the Care/Harm foundation. Hence, the Vegan Discourse is highly represented by the principles of this foundation, such as the caring for others and the empathy towards pain of others. In particular, the Care/Harm foundation is the only moral foundation that is expressed rather explicitly than implicitly as can be argued on the ground of its probability distribution. On the ground of the contrastive analyses performed per foundation, I confirmed that the usage of explicit words related to a foundation result in high probability scores for a foundation. This claim is also true for the Authority/Subversion foundation and the Purity/Degradation foundation. In the Fairness/Cheating foundation, which is the second least often assigned foundation in the entire corpus, the higher and lower scored instances both show the expression of explicit foundation-related words. The Loyalty/Betrayal foundation is represented with similar behavior but based on poor evaluation results, which does not allow for a clear assumption on this. Overall, the Vegan Discourse is mostly shaped morally by the foundations Care/Harm and Purity/Degradation, and, if valid, the Loyalty/Betrayal foundation – hence, users express vices and virtues related to the caring and empathy towards others and the strive to live a more ethical life, or also, moral values based on a sense of belonging to a group. The latter, however, needs to be researched further and thus, is not considered to be sufficiently arguable for further interpretations. Transposed on the values of veganism, a typical user's statement that touches the two most prominent foundations sounds like this:

> *Veganism is a fair choice for those who wish to make that commitment to the well-being of animals. If someone seriously believes that animals have the right to live just like us humans do, you have the right to be vegan. But vegan should not be a diet, but a lifestyle. Vegans should be promoting their disgust, if not their choice to be vegan is just a moral obligation instead of an animal-rights-movement. Of course, it is okay to see veganism in whatever way you want, a moral obligation or a rights movement. We have the right to speak freely. It is indeed similar to the abolitionists of the slave movement, just not as extreme as slavery. The connection to the man fainting at a show is indeed interesting. Do you let the show go on or do you say something? It is going to be at the expense of other viewers' pleasure, but when the time comes the man's life is more important. Do vegans let the show of killing billions of animals go on, or do they step up and say something?"* (on "When Vegans Won't Compromise" (Fischer & McWilliams, 2015))

This user expresses, according to the model, definite moral values related to the Care/Harm and the Purity/Degradation foundation. They are a stereotypical participant in the Vegan Discourse regarding their moral values expressed in this statement. The case of faceted morality in the sense of the assignment of several moral foundations for one or more comments confers with the finding that the foundations themselves are correlative with each other. This is particularly the case for the Authority/Subversion foundation, which correlates with the Purity/Degradation foundation and the Loyalty/Betrayal foundation. The Vegan Discourse Corpus is also represented by non-morality, since the Non-moral label was assigned to roughly a third of the corpus. These comments are characterized by very general, stereotypical words related to the Vegan Discourse and are mostly represented with high probability scores. However, the lower scored instances of the Non-moral label were shown to consist of explicit foundation-related terms while the higher scores instances were implicit on the word-level, which is in fact argued to be indicative for morality, as well. Hence, the label's dominance in the corpus is rather questionable regarding its actual meaning as well as its assignment for implicit comments, since morality bears the issue of implicitness.

THE ROLE OF STORYTELLING

This becomes particularly interesting when taking into account the storytelling instances in the Vegan Discourse Corpus. They are featured with many general, discourse-related words apart from two storytelling markers like "I'm" or "I've". Overall, there are approximately as many instances in the corpus that are storytelling as there are non-storytelling ones. The comments that were labeled as storytelling make up about half of the corpus. Observed together with the assignment of the labels for the moral foundations, it is noteworthy that there are more comments that are classified as moral and storytelling than the inverse scenario. In addition, the comments in the Vegan Discourse are either highly probable to contain storytelling or highly unlikely. High probability scores for storytelling were documented to more often co-occur with low probability-scores for the moral foundations; thus, explicitness does not necessarily result in a high likelihood of storytelling. This is because the latter was not found to be representative of explicit wordings for any of the moral foundations. Therefore, storytelling does not enhance the expression of morality in text in the way that it uncovers the implicitness around morality and thus makes the implicit more explicit. Rather, as the analysis showed, it puts a special emphasis on implicitness in general as a shared character feature of both morality and storytelling and thus, enables to better detect morality in text. As a result, I argue that the dominant assignment of the Non-moral label results from the fact that it includes implicitness as a label-relevant feature, although it is generally argued that morality is rather implicit, which I confirmed in the cases for weak morality in the contrastive analyses based on the lower boundary evolved from the gold standard.

Methodological learnings

In addition, it should be noted that the evaluation scores for the model trained for the purpose of this thesis indicate good performance, which was also confirmed by comparing its performance with those of the model of other researchers and also, human annotators. Although the gold standard set was rather small, the results provided indications for thresholds of moral validity and a lower boundary of non-morality. This was crucial for the quantitative and qualitative analysis of this work, since it gave a frame to the abstract concept of morality. With regard to the different context of the Vegan Discourse Corpus compared to the one of the Moral Foundations Twitter Corpus, I can report some possible impacts on the classification of the Loyalty/Betrayal foundation, since it was in the lower range regarding the evaluation metrics. Still, this could also be a result of the general nature of the Vegan Discourse, meaning the foundation can merely not be expressed as dominantly in this context. Overall, the methods from Natural Language Processing enabled me to dive deep into the rather abstract concept of morality about which philosophers and other scientists have been theorizing for centuries. Thus, this thesis emphasizes that the work of the Digital Humanities is that of an inspiring field that combines abstract thought processes and complex methods from different academic and scientific disciplines, and through this, allows to introduce new perspectives on seemingly independent and distant concepts.

Summarized findings

I claim the Vegan Discourse to be significantly interesting for an analysis with moral sentiment and storytelling, since the phenomenon itself lends itself well for this with its “do-gooder derogation” (Minson & Monin, 2012, p. 200) character and because the synergy of the two aspects promise a new and fruitful approach to the detection of implied morality in text. The Vegan Discourse itself was proven to be significantly moral in the sense of the five moral foundations given the challenge of implicitness of morality, since more than a tenth of the corpus was classified reliably with morality. Taking into account the moral loadings based on the lower boundary, there are even more comments by users expressing morality in one way or another. In addition, the two most frequently assigned foundations are the Care/Harm and the Purity/Degradation foundation, which does not confirm the hypothesis that the values referring to Fairness/Cheating are expressed frequently among with those of Care/Harm. Furthermore, the second hypothesis that storytelling enhances the expression of morality in text needs to be re-defined: although most instances with morality in the corpus were more characterized by storytelling than by non-storytelling, the comments were not highly likely to be indicative for morality as well as for storytelling, but rather highly likely to be storytelling in cases of weak moral loadings. Since these result from implicitness, as I argue in this thesis, the aspect of storytelling is rather highly significant as a marker of detectable implicitness in these cases. Thus, storytelling enhances the expression of morality in the way that it can serve as a tool to detect implied morality.

So, when asked why (not) vegan, is morality an important factor? And is the typical (vegan) do-gooder likely also a storyteller? This thesis proved that both concepts, moral sentiment as well as storytelling, are significantly observable and intertwined in the Vegan Discourse as represented in this work. I am thus convinced that there are many (non-)vegans that tell their personal experiences and opinions in stories to make their way of life more understandable and palpable for other people. This is particularly the case for those that share a different opinion regarding this topic, since it is an issue of morality, which is innate to everyone, but not equally developed in every individual according to their culture or experiences according to the Moral Foundations Theory. It makes thus sense that one elaborates on personal experiences more often in the Vegan Discourse, since these are actually shaping the way we think and act morally. Overall, a topic like veganism is one that is framed in many ways morally and more often reasoned with the reference to personal experience to make it understandable for others.

7 OUTLOOK

I argue that the implicit kind of expression of morality is likely to be expressed via storytelling, which can significantly contribute to further research in the detection of morality in texts. To ameliorate the model's detection of moral sentiment in text, it is interesting to include an instance's likelihood for storytelling as a feature and make the model learn the interdependencies between morality and storytelling. Thus, an instance that has a high storytelling score but includes no explicit foundation-related terms, for instance, should not be assigned the Non-moral label, as was the case with the model developed for this thesis, but rather checked twice for implied morality.

In addition, it is also noteworthy to mention that the Moral Foundations Theory argues to be universally applicable and shaped culturally per individual – still, it is questionable whether all existing foundations are captured in the Theory, if some went extinct or developed further during time. Hence, it is reasonable to further explore the existence of other moral foundations, which is actively encouraged by the creators of the Theory (Dehghani et al., 2021). In the Vegan Discourse, I report the occurrence of all foundations with more or less dominance, but except for the fact that the comments' authors are readers of the American news magazine The New York Times, no demographic or cultural information is available. Thus, it would be interesting to apply the Moral Foundations Theory on data that includes this information to assess if the foundations are culture-bound.

Moreover, due to a limitation in volume of this thesis, the gold standard created for this thesis is rather small. Still, it helped to draw assumptions from it that framed the entire analysis of the classification results. However, in future research, a larger gold standard would help to establish these assumptions further and to evaluate on the model as well as the two annotation domains themselves. In addition, a bigger and more diverse group of annotators, for instance, that were trained on the annotation guidelines and its challenges, would likely improve the quality of the gold standard in a way that there is less bias and more agreement among annotators. This again can support a better understanding of the principles underlying the moral foundations or storytelling and thus, spark more interesting thoughts in the analysis and interpretation.

Finally, an approach by Lin and colleagues (2018) that introduces the idea to enrich the input text for classification with background knowledge from an independent knowledge base seems promising in the context of this thesis. Since the Moral Foundations Twitter Corpus did not have a significant overlap in the topics of the corpus domains, it might have been helpful to have more context for the model to escape the limitation of the model to specific events that are related to a specific moral value in the training corpus. The authors of the paper name the example of the "Westboro Church", which is known for its racist and anti-LGBTQ+ stance and hate speech, but this knowledge, if not stated explicitly, gets lost when the reference in a sentence is given to the model for classification. A similar example in the context of the Vegan

Discourse would be “Earthling Ed”, who is a popular personality in the vegan community and known for his active engagement in the vegan lifestyle. If the input text was enriched with a brief explanation about this person, it would likely capture some more insights into morality than without this knowledge. It would be interesting to combine this approach with the idea of storytelling as a marker of implicitness and to see whether one could create a scale for implicitness and connect it to morality.

DIGRESSION: MORAL STORYTELLERS

Lastly, it is noteworthy to state that some users draw attention to themselves as was found during the analysis, since their comments received high scores for different moral foundations during classification, indicating a consistent expression of morality. Mostly, a particular foundation is prominent for a user, like the Care/Harm foundation, but there are some cases where multiple foundations are triggered in classification by a user’s statement. 19 users among many more and their comments were handpicked from the classified data based on their pattern particularity and their average probability scores were calculated. Notably, the users’ total average probability score for the Care/Harm foundation is 0,62, with the lowest being a value of 0,36 and the highest a value of 0,8. Moreover, the total average probability score for storytelling is at 0,59, where the lowest average score for a user is a value of 0,33 and the highest average score a value of 0,89. Moreover, ten of the 19 users received an average probability score greater than 0,1 for the Purity/Degradation foundation (apart from their high score for the Care/Harm foundation), making it the second most prominent foundation among the users. In addition, four users expressed the Fairness/Cheating foundation with a probability score greater than 0,1, and three users respectively with the Authority/Subversion foundation. Particularly, for four users the average probability score was greater than 0,1 for three foundations or more. Overall, these users are the perfect portrayal of moral storytellers within the Vegan Discourse and deliver great evidence for this thesis’ purpose of defining the Vegan Discourse in a moral way.

LITERATURE

About us. (2022, February 1). Veganuary. https://veganuary.com/about/about-us/

Anderson, B., & Gross, J. (2021, March 5). The New Menu at Eleven Madison Park Will Be Meatless. *New York Times*. https://www.nytimes.com/2021/05/03/dining/eleven-madison-park-vegan-menu.html

Aquino, K., & Reed, A. 2nd. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037//0022-3514.83.6.1423

Bekoff, M. (2013, June 20). A Universal Declaration on Animal Sentience: No Pretending. *Psychology Today*. https://www.psychologytoday.com/intl/blog/animal-emotions/201306/universal-declaration-animal-sentience-no-pretending

Benton, T. G., Bieg, C., Harwatt, H., Pudasaini, R., & Wellesley, L. (2021, March 2). Food system impacts on biodiversity loss. Three levers for food system transformation in support of nature. *Energy, Environment and Resources Programme*. https://www.chathamhouse.org/2021/02/food-system-impacts-biodiversity-loss/04-recommendations-action-2021

Bittman, M. (2011, December 28). No Meat, No Dairy, No Problem. *New York Times*. https://www.nytimes.com/2012/01/01/magazine/mark-bittman-going-semi-vegan.html

Blasi, A. (2004). Moral Functioning: Moral Understanding and Personality. In *Moral development, self, and identity.* (pp. 335–347). Lawrence Erlbaum Associates Publishers.

Brody, J. E. (2021, February 10). Good Vegan, Bad Vegan. *New York Times*. https://www.nytimes.com/2017/10/02/well/eat/good-vegan-bad-vegan.html

Brown, J. (2020, January 23). Are there health benefits to going vegan? *BBC*. https://www.bbc.com/future/article/20200122-are-there-health-benefits-to-going-vegan

Christopher, A., Bartkowski, J. P., & Haverda, T. (2018). Portraits of Veganism: A Comparative Discourse Analysis of a Second-Order Subculture. *Societies*, *8*(3). https://doi.org/10.3390/soc8030055

Clines, F. X. (2015, February 9). Satanists, Vegans and Atheists Seek Equal Opportunity in Little Rock. *New York Times*. https://takingnote.blogs.nytimes.com/2015/09/02/satanists-vegans-and-atheists-seek-equal-opportunity-in-little-rock/

Cohon, R. (2018). Hume’s Moral Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/hume-moral/

Consolidated Version of the treaty on the functioning of the European Union. (2012). *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012E/TXT&from=en

Darwall, S. (2009). *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press.

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). *Automated Hate Speech Detection and the Problem of Offensive Language*.

Dehghani, M., Atari, M., Koleva, S., Graham, J., Haidt, J., & Ditto, P. (2021, June). *Moral Foundations*. https://moralfoundations.org/

Dehghani, M., Gratch, J., Sachdeva, S., & Sagae, K. (2011, January). *Analyzing Conservative and Liberal Blogs related to the Construction of the 'Ground Zero Mosque.'*

Department for Environment, Food & Rural Affairs, & The Rt Hon Lord Goldsmith. (2021). *Animals to be formally recognised as sentient beings in domestic law* [Press release]. https://www.gov.uk/government/news/animals-to-be-formally-recognised-as-sentient-beings-in-domestic-law

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, *abs/1810.04805*. http://arxiv.org/abs/1810.04805

Dunning, D. (2005). Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself. *Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself*. https://doi.org/10.4324/9780203337998

Edgell, P., Hull, K. E., Green, K., & Winchester, D. (2016). Reasoning Together Through Telling Stories: How People Talk about Social Controversies. *Qualitative Sociology*, *39*(1), 1–26. https://doi.org/10.1007/s11133-015-9321-4

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, *115*(44), 11203–11208.

Finnerty, K. (2020, January 15). Could veganism be the solution to the climate crisis? *Energy & Environment (Ipsos MORI)*. https://www.ipsos.com/ipsos-mori/en-uk/could-veganism-be-solution-climate-crisis

Fischer, B., & McWilliams, J. (2015, August 16). When Vegans Won't compromise. *New York Times*. https://opinionator.blogs.nytimes.com/2015/08/16/when-vegans-wont-compromise/

Food and Agriculture Organization of the United Nations. (2013). *Major cuts of greenhouse gas emissions from livestock within reach*. https://www.fao.org/news/story/en/item/197608/icode

Forgrieve, J. (2018, February 11). The Growing Acceptance of Veganism. *Forbes*. https://www.forbes.com/sites/janetforgrieve/2018/11/02/picturing-a-kindler-gentler-world-vegan-month/?sh=66577d5b2f2b

Garten, J., Boghrati, R., Hoover, J., Johnson, K. M., & Dehghani, M. (2016). *Morality Between the Lines: Detecting Moral Sentiment In Text*.

Gert, B., & Gert, J. (2020). The Definition of Morality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/morality-definition/

Google Books Ngram Viewer. (2022, February 1). https://books.google.com/ngrams/graph?content=vegan%2C+vegetarian%2C+plant-based&year_start=1800&year_end=2019&corpus=26&smoothing=3&direct_url=t1%3B%2Cvegan%3B%2Cc0%3B.t1%3B%2Cvegetarian%3B%2Cc0%3B.t1%3B%2Cplant%20-%20based%3B%2Cc0#t1%3B%2Cvegan%3B%2Cc0%3B.t1%3B%2Cvegetarian%3B

%2Cc0%3B.t1%3B%2Cplant%20-%20based%3B%2Cc0

Gordinier, J. (2015, September 29). Vegans Go Glam. *New York Times*. https://www.nytimes.com/2015/09/30/dining/vegan-diet-lifestyle-recipes.html

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Chapter Two—Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029.

Gray, R. (2020, February 13). Why the vegan diet is not always green. *BBC*. https://www.bbc.com/future/article/20200211-why-the-vegan-diet-is-not-always-green

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2008). Morality. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *3*(1), 65–72. https://doi.org/10.1111/j.1745-6916.2008.00063.x

Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not recognize. *Social Justice Research*, *20*, 98–116. https://doi.org/10.1007/s11211-007-0034-z

Hoover, J., Portillo-Wightman, G., Yeh, L., Havaldar, S., Davani, A. M., Lin, Y., Kennedy, B., Atari, M., Kamel, Z., Mendlen, M., Moreno, G., Park, C., Chang, T. E., Chin, J., Leong, C., Leung, J. Y., Mirinjian, A., & Dehghani, M. (2020). Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. *Social Psychological and Personality Science*, *11*(8), 1057–1071. https://doi.org/10.1177/1948550619876629

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, *7*(2), 265–290. https://doi.org/10.1017/langcog.2014.30

Ipsos MORI. (2018). *An exploration into diets around the world*. https://www.ipsos.com/sites/default/files/ct/publication/documents/2018-08/an_exploration_of_diets_around_the_world.pdf

Klie, J.-C., Bugert, M., Boullosa, B., Eckart de Castilho, R., & Gurevych, I. (2018). The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, 5–9. https://www.aclweb.org/anthology/C18-2002

Kobbe, J., Rehbein, I., Hulpus, I., & Stuckenschmidt, H. (2020). Exploring Morality in Argumentation. *Proceedings of the 7th Workshop on Argument Mining*, 30–40. https://aclanthology.org/2020.argmining-1.4

IfD Allensbach. (2021). *Personen in Deutschland, die sich selbst als Veganer einordnen oder als Leute, die weitgehend auf tierische Produkte verzichten, in den Jahren 2015 bis 2020 [Graph].* Statista. https://de.statista.com/statistik/daten/studie/445155/umfrage/umfrage-in-deutschland-zur-anzahl-der-veganer/

Lin, Y., Hoover, J., Portillo-Wightman, G., Park, C., Dehghani, M., & Ji, H. (2018). Acquiring

Background Knowledge to Improve Moral Value Prediction. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 552–559. https://doi.org/10.1109/ASONAM.2018.8508244

LIVE TV DEBATE: Is Veganism a Moral Imperative? (2022, January 13). [BBC Politics Live]. https://www.youtube.com/watch?v=Hy65IdIr1Rk

López-Alt, J. K. (n.d.). Vegan Cheeseburgers. *New York Times*. Retrieved February 6, 2021, from https://cooking.nytimes.com/recipes/1020867-vegan-cheeseburgers

López-Alt, J. K. (2020, March 3). How to Cook With Plant-Based Meats. *New York Times*. https://www.nytimes.com/2020/03/03/dining/impossible-beyond-meat.html

Low, P. (2012, July 7). *The Cambridge Declaration on Consciousness*. Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals, Cambridge, UK. http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

Manjoo, F. (2019, August 28). Stop Mocking Vegans. *New York Times*. https://www.nytimes.com/2019/08/28/opinion/vegan-food.html

Maslen, S., & Hayes, J. (2020). "This is How we Debate": Engineers' Use of Stories to Reason through Disaster Causation. *Qualitative Sociology*, *43*(2), 191–212. Sociological Abstracts; Sociology Database. https://doi.org/10.1007/s11133-020-09452-1

McDermott, M. T. (2017, June 6). The Vegan Dog. *New York Times*. https://www.nytimes.com/2017/06/06/well/family/the-vegan-dog.html

Meyer, P. (2014). *Storytelling for Lawyers*. Oxford University Press. https://books.google.de/books?id=JEmeAgAAQBAJ

Milman, O. (2021, September 13). Meat accounts for nearly 60% of all greenhouse gases from food production, study finds. *The Guardian*. https://www.theguardian.com/environment/2021/sep/13/meat-greenhouses-gases-food-production-study

Minson, J. A., & Monin, B. (2012). Do-Gooder Derogation: Disparaging Morally Motivated Minorities to Defuse Anticipated Reproach. *Social Psychological and Personality Science*, *3*(2), 200–207. https://doi.org/10.1177/1948550611415695

Mishan, L. (2018, June 9). At Seasoned Vegan, Soul Food That Never Harmed a Soul. *New York Times*. https://www.nytimes.com/2018/09/06/dining/seasoned-vegan-review-harlem.html

Moezzi, M., Janda, K. B., & Rotmann, S. (2017). Using stories, narratives, and storytelling in energy and climate change research. *Energy Research & Social Science*, *31*, 1–10. https://doi.org/10.1016/j.erss.2017.06.034

Monin, B. (2007). Holier than me? Threatening social comparison in the moral domain. *Revue Internationale de Psychologie Sociale*, *20*(1), 53–68.

Monin, B., & Jordan, A. (2009). *The Dynamic Moral Self: A Social Psychological Perspective* (pp. 341–354).

Moral. (2014). In *Cambridge Dictionary*. Cambridge University Press. https://dictionary.cambridge.org/de/worterbuch/englisch/moral

Morality. (2014). In *Cambridge Dictionary*. Cambridge University Press. https://dictionary.cambridge.org/de/worterbuch/englisch/morality

Morning Consult, & The Hollywood Reporter. (2021). *Credibility of the New York Times in the United States as of May 2021* [Graph]. Statista. https://www.statista.com/statistics/239749/credibility-of-the-new-york-times-in-the-united-states/

Nussbaum, M. C. (n.d.). Morality and emotions. *Routledge Encyclopedia of Philosophy*. https://doi.org/10.4324/9780415249126-L064-1

Nussbaum, M., & Craig, E. (1998). Morality and the Emotions. *Routledge Encyclopedia of Philosophy*, *6*, 558.

Pajer, N. (2019, June 2). Leather? At Vegan Fashion Week, It's Pineapple Leaf. *New York Times*. https://www.nytimes.com/2019/02/06/style/vegan-fashion-week.html

Pasanek, B. (2009, September). *Ideology and metaphor*.

Pickett, S. (2021). Veganism, Moral Motivation and False Consciousness. *Journal of Agricultural and Environmental Ethics*, *34*(15). https://doi.org/10.1007/s10806-021-09857-0

Pilon, M. (2012, April 1). Sculptured by Weights and a Strict Vegan Diet. *New York Times*. https://www.nytimes.com/2012/01/05/sports/vegans-muscle-their-way-into-bodybuilding.html

Polkinghorne, D. E., & Polkinghorne, J. C. (1988). *Narrative knowing and the human sciences*. State University of New York Press.

Polletta, F., & Lee, J. (2006). Is Telling Stories Good for Democracy? Rhetoric in Public Deliberation after 9/11. *American Sociological Review*, *71*(5), 699–721. https://doi.org/10.1177/000312240607100501

Princeton University. (2010). *About WordNet*. https://wordnet.princeton.edu/

Qian, M., Laguardia, J., & Qian, D. (2021). Morality Beyond the Lines: Detecting Moral Sentiment Using AI-Generated Synthetic Context. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI - Second International Conference, AI-HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24-29, 2021, Proceedings* (Vol. 12797, pp. 84–94). Springer. https://doi.org/10.1007/978-3-030-77772-2_6

Quenqua, D. (2010, June 8). At Vegans' Weddings, Beef or Tofu? *New York Times*. https://www.nytimes.com/2010/08/08/fashion/08vegan.html

Rao, T. (2021a, April 16). Vegan Cheese, but Make It Delicious. *New York Times*. https://www.nytimes.com/2021/04/16/dining/vegan-cheese.html

Rao, T. (2021b, May 26). The 100-Pound Fruit That Vegans Love. *New York Times*. https://www.nytimes.com/2021/05/26/magazine/jackfruit-sabzi-recipe.html

Rehbein, I., & Kobbe, J. (2020). *Guidelines for the Annotation of Moral Foundations on Arguments*.

Rehm, G., Zaczynska, K., Bourgonje, P., Ostendorff, M., Moreno-Schneider, J., Berger, M., Rauenbusch, J., Schmidt, A., Wild, M., Böttger, J., & others. (2021). Semantic Storytelling. *Computational Analysis of Storylines: Making Sense of Events*, 240.

Reports of personal experiences and stories in argumentation: Datasets and analysis. (2021).

Roy, S., & Goldwasser, D. (2021). Analysis of Nuanced Stances and Sentiment Towards Entities of US Politicians through the Lens of Moral Foundation Theory. *SOCIALNLP*.

Saint Louis, C. (2017, June 30). Can You Have a Healthy Vegetarian or Vegan Pregnancy? *New York Times*. https://www.nytimes.com/2017/06/30/well/family/can-you-have-a-healthy-vegetarian-or-vegan-pregnancy.html

Sandberg, K. (2019). *Persuasion in Discourse: Evaluation Methods of Vegans* [Honors Thesis -- Open Access]. https://scholarworks.wm.edu/honorstheses/1298

Scanlon, T. M. (1982). Contractualism and utilitarianism. In A. Sen & B. Williams (Eds.), *Utilitarianism and Beyond* (pp. 103–128). Cambridge University Press. https://doi.org/10.1017/CBO9780511611964.007

Scanlon, T. M., Scanlon, A. P. N. R. M. P. C. P. T. M., University, H., Press, H. U., & Press, B. (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press. https://books.google.de/books?id=FwuZcwMdtzwC

Schank, R. C., & Abelson, R. P. (1995). Knowledge and memory: The real story. In *Knowledge and memory: The real story.* (pp. 1–85). Lawrence Erlbaum Associates, Inc.

Scholarly vs. Popular Sources. (2021). Yale. Poorvu Center for Teaching and Learning. https://poorvucenter.yale.edu/undergraduates/using-sources/principles-citing-sources/scholarly-vs-popular-sources

Severson, K. (2017, November 28). Black Vegans Step Out, for Their Health and Other Causes. *New York Times*. https://www.nytimes.com/2017/11/28/dining/black-vegan-cooking.html

Story. (2014). In *Cambridge Dictionary*. Cambridge University Press. https://dictionary.cambridge.org/de/worterbuch/englisch/story

Storytelling. (2014). In *Cambridge Dictionary*. Cambridge University Press. https://dictionary.cambridge.org/de/worterbuch/englisch/storytelling

The EAT-Lancet Commission Summary Report. (2019). EAT-Lancet Commission. https://eatforum.org/content/uploads/2019/07/EAT-Lancet_Commission_Summary_Report.pdf

The New York Times. (2021). *Community API*. https://developer.nytimes.com/docs/community-api-product/1/overview

The New York Times Company Reports 2021 Third-Quarter Results. (2021). New York Times. www.nytco.com

The Vegan Society. (2021a, October 12). *History*. The Vegan Society. One World. Many Lives. Our Choice. https://www.vegansociety.com/about-us/history

The Vegan Society. (2021b, October 12). *Treatment of animals*. The Vegan Society. One World. Many Lives. Our Choice. https://www.vegansociety.com/news/media/statistics/treatment-animals

The Vegan Society. (2021c, October 12). *Why go vegan?* The Vegan Society. One World. Many Lives. Our Choice. https://www.vegansociety.com/go-vegan/why-go-vegan

Tomky, N. (2019, February 27). When Your Kindergarten Goes Vegan. *New York Times*. https://www.nytimes.com/2019/02/27/style/vegan-daycare-children.html

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Turiel, E. (2006). Thought, emotions, and social interactional processes in moral development.

In *Handbook of moral development.* (pp. 7–35). Lawrence Erlbaum Associates Publishers.

Vecchi, E. M., Falk, N., Jundi, I., & Lapesa, G. (2021). Towards argument mining for social good: A survey. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1338–1352.

Weick, K. E., & Browning, L. D. (1986). Argument and narration in organizational communication. *Journal of Management*, *12*(2), 243–259.

Wyer, R. S., Adaval, R., & Colcombe, S. J. (2002). *Narrative-based representations of social knowledge: Their construction and use in comprehension, memory, and judgment* (Vol. 34, pp. 131–197). Academic Press. https://doi.org/10.1016/S0065-2601(02)80005-3

Xie, J. Y., Junior, R. F. P., Hirst, G., & Xu, Y. (2019). Text-based inference of moral sentiment change. *EMNLP*.

Xu, X., Sharma, P., Lin, Tzu. S., Ciais, P., Tubiello, F. N., Smith, P., Campbell, N., & Jain, A. K. (2021, September). Global greenhouse gas emissions from animal-based foods are twice those of plant-based foods. *Nature Food*, *2*, 724–732.

Erklärung über die Eigenständigkeit der Masterarbeit

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel

“Why (not) vegan? An Investigation of Moral Sentiment and Storytelling in the Vegan Discourse with Natural Language Processing”

selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

Declaration of Authorship

I hereby certify that the dissertation entitled

“Why (not) vegan? An Investigation of Moral Sentiment and Storytelling in the Vegan Discourse with Natural Language Processing”

is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name/Name: Rebecca Pichler

Matrikelnummer/student number: 3453609

Unterschrift/Signed:

Datum/Date: 14.02.2022