THESIS

ADVANCING MEDIUM- AND HEAVY-DUTY ELECTRIC VEHICLE ADOPTION MODELS WITH NOVEL NATURAL LANGUAGE PROCESSING METRICS

Submitted by

Fletcher Ouren

Department of Systems Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2024

Master's Committee:

Advisor: Thomas H. Bradley

Timothy Coburn
Bret Windom

Copyright by Fletcher Thomas Ouren 2024

All Rights Reserved

ABSTRACT

ADVANCING MEDIUM- AND HEAVY-DUTY ELECTRIC VEHICLE ADOPTION MODELS WITH NOVEL NATURAL LANGUAGE PROCESSING METRICS

The transportation sector must rapidly decarbonize to meet its emissions reduction targets. Medium- and heavy-duty decarbonization is lagging behind the light-duty industry due to technical and operational challenges and the choices made by medium- and heavy-duty fleet operators. Research investigating the procurement considerations of fleets has relied heavily on interviews and surveys, but many of these studies need higher participation rates and are difficult to generalize. To model fleet operators' decision-making priorities, this thesis applies a robust text analysis approach based on latent Dirichlet allocation and Bi-directional Encoder Representations of Transformers to two broad corpora of fleet adoption literature from academia and industry. Based on a newly developed metric, this thesis finds that the academic corpus emphasizes the importance of *suitability, familiarity, norms,* and *brand image*. These perception rankings are then passed to an agent-based model to determine how differences in perception affect adoption predictions. The results show a forecast of accelerated medium- and heavy-duty electric vehicle adoption when using the findings from the academic corpus versus the industry corpus.

ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Vehicle Technologies Office, award number DE-EE0009664.

In addition to the financial support provided by the Department of Energy, I would also like to thank my thesis committee and David Trinko for their consistent and invaluable support.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iii

1 Introduction 1

2 Background 3

2.1 Text Mining 3

2.2 Agent-Based Modeling 8

3 Methods 14

3.1 LDA Method 15

3.2 BERTopic Method 20

3.3 Score of Topic and Related Terms 21

4 Results 21

4.1 LDA Results 22

4.2 BERTopic Results 24

4.3 START Results 24

4.4 eMDHD Adoption Perception Category Rankings 26

4.5 Results Applied to Adoption Modeling 26

5 Discussion 28

5.1 START Costs and Benefits 28

5.2 Ground Truth Examples from Academia and Industry Text Corpora 29

6 Future Work 31

7 Conclusion 32

Sources 34

Appendix 40

A1 40

A2 41

A3 42

A4 48

List of Abbreviations 49

1 Introduction

Given the aggressive greenhouse gas emission reduction goals set out on every level, from international goals [1] to federal [2] and state [3] legislation, the fleet transportation landscape is poised to enable the adoption of a new generation of electrified vehicles. Battery electric vehicles (BEV) are currently the most used zero-emissions vehicles [4]. While the use of BEVs for personal transportation is rapidly expanding [5], and the number of electric medium- and heavy-duty (eMDHD) fleets is growing [6], the latter are experiencing a slower transition. This delay can be attributed to several considerations undertaken by a fleet operator during purchasing.

To quantify the overall process of fleet adoption, agent-based modeling (ABM) is employed. An ABM simulates the actions of a network of agents based on the data that defines the agents. While technical data, such as vehicle capabilities and costs, are readily available, providing an agent object with data on perceptions is more challenging.

The perceptions that a fleet manager holds during the purchasing process have received little academic attention. So far, researchers have relied on fleet manager interviews and surveys using various techniques and with varying success to address this question. For example, Bae et al. [7] synthesize theories of organizational adoption in the context of interviews conducted with 20 fleet managers to characterize the adoption process framework typically followed by MDHD fleet managers (see Figure 1-1). They identify a list of factors that play an essential role in the adoption decision (the middle box in Figure 1-1) for each of 20 fleet managers interviewed, including relative advantage characteristics (cost reductions, emissions benefits), operational compatibility and complexity, and sources of uncertainty such as fuel price risk.

In addition to the work of Bae et al. [7], other efforts [8–11] attempt to construct a quantifiable model through ranking statements from interviews using area expertise as the basis for these rankings. Each of these studies is based on a small number of fleets that cannot possibly represent the wide vocational, organizational, and geographic diversity of the MDHD transportation sector. Such efforts are also vulnerable to acknowledged methodological errors, including interviewer/interviewee bias, selective nonresponse, and other non-sampling errors.

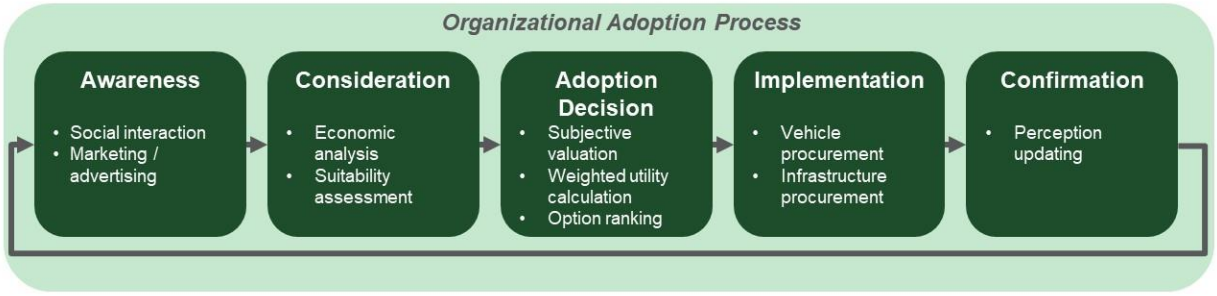

Figure 1-1: Adoption process framework adapted from Bae et al.

This undertaking deviates from the traditional systematic literature review, citing texts from peer-reviewed academic journals and trade journals. Upon introducing a new metric called the "Score of Topic and Related Terms" (START), this study will enumerate the differences in focus between the texts associated with academia and trade journals. Each body of texts, or corpus, fits two complimentary topic models. Results from this process are used to determine the importance of pre-determined categories within each corpus. These categories, shown in Figure 1-2, are analogous to the perceptions of fleet managers during the eMDHD purchasing decision. Discussion and conclusions focus on identifying insights into the differences between topic importance and perceptions from these two bodies of texts but also demonstrate the use of these types of topic models as inputs for ABMs of eMDHD adoption.

The present investigation aims to explain how perceptions impact eMDHD adoption decisions with a data-driven approach. The first task is to quantify the given perceptions using various data sources.

The second task is to demonstrate how perceptions defined from different sources affect eMDHD adoption predictions. This thesis explains how to complete these objectives by referencing publications from the cohort's efforts [12–14].

- *Long-term Cost*
- *Policy/Incentives*
- *Suitability*
- *Reliability*
- *Driver Acceptance*
- *Sustainability*
- *Brand Image*
- *Short-term cost*
- *Knowledge/Familiarity*
- *Norms/Attitudes*

Figure 1-2: Categories of Interest

2 Background

2.1 Text Mining

Both text mining and natural language processing (NLP) have theoretical origins in the 1950s. These ideas remained in the realm of theory due to the lack of available computing power until the 1990s. The onset of the Big Data era, in combination with the rapid development of computing technology, allowed text mining practices to enter the mainstream [15]. They are used in several industrial sectors, such as business analytics, engineering, public health, and the physical and social sciences [15].

Text mining is the process of wrangling and analyzing unstructured text data. Tasks in this field include content extraction, sentiment analysis, pattern recognition, topic modeling, dimensionality reduction, and many more. The connecting theme between these tasks is using term frequencies and semantic relationships in tandem with clustering and classification methods [16]. The models

employed in the present thesis heavily rely on the progression of the text mining field over the past 70 years.

Latent-Dirichlet Allocation

Latent Dirichlet allocation (LDA) [17] is a product of some of the most cited authorities in data science, including David Blei, Andrew Ng, and Michael Jordan, who collaborated to eliminate issues in overfitting experienced when using its predecessor, probabilistic latent semantic analysis (pLSA) [18]. pLSA is one of many alternatives to LDA. Other methods include latent semantic analysis (LSA) [19] and non-negative matrix factorization (NMF) [20]. LDA holds the capability, unique among these options, of assigning multiple topics to a single document. LDA also fulfills the experiment's scalability, accuracy, and explainability requirements. This allows LDA to return a more accurate representation of essential terms in the corpus. NMF shows promising topic modeling results in Kumar and Ng [20], reducing the amount of topic guessing in the training process. Still, LDA was selected over NMF because NMF is usually reserved for shorter texts such as social media posts and web chat captions since an NMF model may generate less interpretable topics [21].

LDA models are applied in various applications, from social media fake news detection [22] and movie reviews [23] to medical journal database organization [24]. The present usage of LDA slightly differs from this common implementation in that it evaluates the global importance of terms instead of limiting the scope to their contribution to generated topics. Although it is well suited for this purpose, similar examples in the literature were not found.

LDA relies on the central assumption that every document is a random combination of words where word order does not matter; it is generally a clustering approach. The set of words ($\boldsymbol{w}$) a is then placed into a set of N topics ($\boldsymbol{z}$) by defining the likelihoods of how many topics a given document will contain (α) and how many words are associated with each topic (β). These hyperparameters, in tandem with

the marginal topic distribution (θ) are used to build the resulting joint distribution of a topic mixture, as specified in (1):

$$p(\theta, \boldsymbol{z}, \boldsymbol{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta)\,. \tag{1}$$

Integrating this distribution and summing over the topics, the marginal distribution of topics in a document is achieved, as specified in (2). Such distributions provide the information on topic probability needed to calculate saliency, the final scoring metric (defined below):

$$p(\boldsymbol{w} \mid \alpha, \beta) = \int p(\theta \mid \alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n \mid \theta)\, p(w_n \mid z_n, \beta) \right) d\theta \tag{2}$$

With the framework of distributions created, they need to be optimized. LDA completes this task by using Gibbs sampling. Depending heavily on Monte Carlo Markov Chains and Bayes' Theorem, this algorithm pulls the word under consideration out of its initial randomly assigned topic. It infers the best fitting (e.g., most closely aligned) topic based on the number of times that the document containing the word uses a topic and the number of times a topic uses the word. One LDA training iteration consists of this process for each word in the corpus. Tomar [25] provides a deeper explanation of the execution of Gibbs sampling.

LDA Performance Metrics

The key output of a typical LDA model is a "coherency score." However, the proposed methods measure "saliency" as the basis for an attribute's ranking in the fleet decision process. In layperson's terms, saliency is simply the extent to which an item is distinct or pronounced.

Saliency is a metric developed initially for TERMite, a visualization package produced by Chuang et al. [26] that shows the overall importance of a word in a corpus. Using the distributions derived in (1) and (2), saliency ranks each word based on its frequency scaled by its distinctiveness as defined in (3):

$$Saliency = P(w) * distinctiveness(w), \tag{3}$$

where w is a word in the corpus and distinctiveness is defined in (4):

$$distinctiveness(w) = \sum_{T} P(T|w) log \frac{P(T|w)}{P(T)}, \tag{4}$$

where T is a topic.

Given a specific word, distinctiveness is the sum of each topic's conditional probability multiplied by the Kullback-Leibler divergence [27] of the topic and word under consideration.

This step is implemented to help eliminate irrelevant terms by looking for the importance of a term within a document and ensuring it is not a common term across the corpus. Several preliminary models were trained to find the terms with the highest saliency to ensure important words were placed into relevant categories. These terms were then placed in the categories listed in Figure 1-2. Topic structure and term saliency of the academic corpus are shown in Figure 2-1.

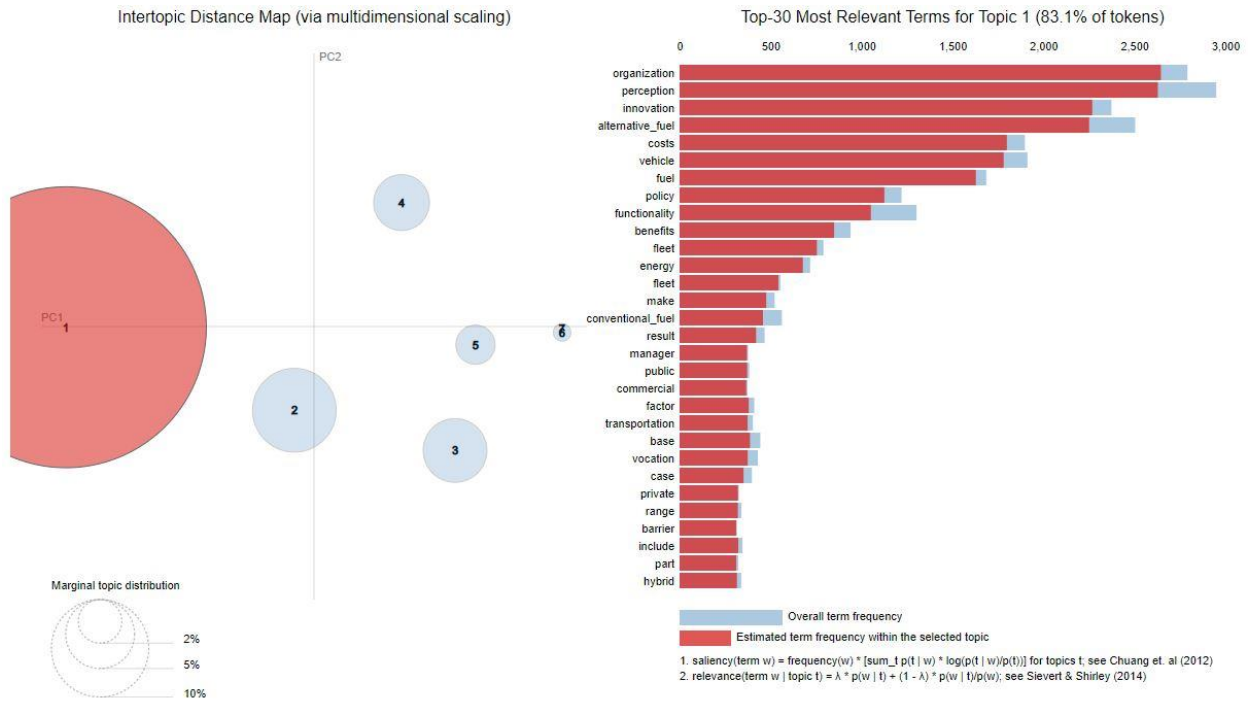

Figure 2-1: TERMite visualization of LDA topic model of academic corpus.

Bi-directional Encoder Representations of Transformers-Topic

Bi-directional Encoder Representations of Transformers-Topic (BERTopic) [28] is a topic modeling approach that is based on the foundational BERT model [29], but adds two more steps to create a topic model output similar to those from LDA. It is implemented here due to its successful implementation relative to topic modeling benchmark problems [30]. The main job of BERT in BERTopic is to create embeddings from sentences. These embeddings are numerical vectors that capture the meaning of the sentence. The version of BERT implemented in BERTopic by prior researchers was pre-trained on two libraries (Wikipedia and Book Corpus) containing over three billion words [29]. After transforming sentences into embeddings, these numerical vectors are clustered by uniform manifold approximation and projection (UMAP) [31] and hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [32] to form topics. Within each cluster, the class-based term frequency-inverse document frequency (c-TF-IDF) [28] score is calculated for each word in each embedding. As an offspring from the early text mining metric of TF-

IDF [33], c-TF-IDF operates similarly, but instead of looking at the relevance of a word in a document, it looks at the relevance of a word within its cluster. Some of the most common issues with TF-IDF are that it cannot capture the semantic meaning of words and needs to be more efficient with high-dimension datasets. c-TF-IDF is appropriate for use in the present application since BERT captures the semantic meaning of each sentence, and the dimensionality is reduced in the clustering process. The literature on BERTopic [28] explains this concept more deeply.

It is an important nuance of BERTopic that while LDA models assume that each document is a random collection of unordered words, BERTopic only analyzes one sentence at a time and uses the correlation between sequential words to understand the context. This is the primary justification for employing both models to arrive at the results. While other researchers [30,34,35] focus on integrating LDA and BERTopic to maximize coherency [36] or silhouette scores [37] in generated topics, this work focuses on combining the results of both models to gain insight into the importance of a predetermined topic in each corpus. These predetermined topics are considered categories for the remainder of this thesis. The main objectives of this research are to develop a metric that ranks categories mined from topic models and apply this metric to the problem of extracting eMDHD adoption perceptions.

2.2 Agent-Based Modeling

ABM is an individual-level modeling technique in which “agents” are modeled to operate autonomously, influencing outcomes with their decision-making. By representing behavioral and social aspects of decisions and how market phenomena emerge from individual choices, ABM addresses aspects of technology adoption that conventional top-down adoption modeling approaches cannot. Many such elements are of interest, including the effects of targeted policies

and incentives, the roles played by interaction and observation, and the geographic arrangement of agents and factors for decision-making [12].

Adoption models often focus either on demand or supply. Still, the market will be driven by the interaction between supply and demand, which can constrain different market segments. For example, there is potential for a "chicken-or-egg" predicament, where vehicle uptake remains low because of a lack of vehicle and charging infrastructure availability, and, conversely, infrastructure and vehicle supply remain low in response to a lack of vehicle demand and utilization. Thus, eMDHD system growth will likely involve a complex coevolution of supply and demand driven by feedback loops between producers, purchasers, and energy and infrastructure suppliers [8,38].

To address these dynamics, agents in the model are of three types:

1) fleet operators,
2) manufacturers of vehicles, and
3) utility and infrastructure managers.

Policy decisions are treated as inputs to the adoption model. Each of these agent varieties makes decisions in pursuit of individualistic objectives based on a subset of information available and according to individual preferences and risk characteristics. The decisions of agents exert mutual influence in a variety of ways. For example, an electric utility installing a set of public fast charging stations improves the suitability and TCO of PEVs for ride-sharing fleets in the surrounding area, increasing their likelihood to adopt PEVs. The adoption of PEVs, in turn, increases the utilization and payback potential of the charging stations. Furthermore, fleets driving similar duty cycles elsewhere might observe the uptake, increasing their familiarity and confidence in PEVs and potentially provoking them to adopt PEVs. These and many other means of feedback and diffusion can be investigated via ABM [12,39].

Decision-Making Models and Data Sources

The mechanism of decision-making is essential in defining an ABM. In this thesis, it is determined using a combination of theory and empirical study. In contrast to the purchasing choices of individual consumers, which typically involve hard-to-quantify factors such as perceived norms, symbolism, and emotions, business decisions such as those for fleets are heavily influenced by well-understood economic factors like TCO. In some adoption models, business decision-makers are approximated to adhere strictly to utility theory, wherein they invariably purchase with the highest "utility" (e.g., always choose the option with the smallest TCO). However, this can lead to unrealistic dynamics if "utility" is defined too narrowly. Researchers have previously employed the theory of planned behavior (TPB) to model decision-making in fleet settings [8]. This theory enables quantitative modeling of factors such as attitude, familiarity, and perceived operational ease associated with a technology, which have been shown to play a role in the decisions of fleets and other businesses. In the ABM, the theory of decision-making followed by agents will incorporate aspects of both utility theory and TPB [12].

Empirical studies of the decision process followed in fleet settings have been conducted for decades, primarily through surveys, interviews, and focus groups. For example, researchers as far back as 2001 found that "bureaucratic" fleets typical to the public sector are unlikely to respond to incentive-based policies but are responsive to mandates, whereas "hierarchic" fleets typical to larger companies in the private sector exhibit the opposite behavior [40]. Decision-making preferences are also found to differ significantly for strategic, non-routine decisions (such as electrification) and for urgent choices that policy mandates might spur.

Adoption Process Model: Agents are organized in networks, one each for fleets, utilities, and OEMs. Fleet agents own and control vehicles and charging infrastructure, about which they form and share

opinions. Utility agents control the costs of installing and operating electric infrastructure. OEM agents control the attributes and availability of vehicles, including costs. To this ABM population, the study applies a conceptual framework for alternative fuel vehicle adoption by commercial fleets based on interviews with HD fleet managers [7]. The theories underpinning their framework account for 1) how innovations spread across social networks (Diffusion of Innovations theory), 2) factors particular to organizational adoption (Technology-Organization-Environment framework), and 3) the roles played by individuals within the organization, including decision-makers and those directly operating the new technologies. The framework they present consists of an adoption process and, within that process, a characterization of the factors influencing the decision to adopt. The following sections describe how each step in the process is modeled [12].

Awareness

Agents' state of awareness is binary; they are either unaware or aware of each vehicle available to them. They can gain awareness either through the advertising and outreach efforts of OEMs and others or through interactions with other fleets. Advertising is modeled as a stochastic process, where all agents who are unaware of a particular OEM have a probability of discovering it each year. Interactions between agents are modeled in detail [12].

Consideration

Fleets will consider adopting a vehicle if they perceive it to be capable of satisfying their operational requirements. These requirements are represented in terms of range and cargo/passenger capacity. A vehicle will not be considered for adoption if it has less range than the longest distance traveled in a day by existing vehicles or if it has less cargo/passenger capacity than the greatest capacity used

by existing vehicles. Vehicles perceived to be operationally suitable are also subjected to economic analysis. Finally, agents assess the relative utility of all the vehicles they are aware of [12].

Adoption Decision

Having evaluated vehicles for suitability and subjectively assessed them, agents decide whether to replace each existing fleet vehicle and what to replace it with. (Vehicles that were purchased within the previous eight years are never replaced. Every existing vehicle is paired with the highest-utility option suitable for replacing it and entered into equation (5) to find the replacement probability:

$$p_{\text{replace}} = \begin{cases} 1 & u_1 \geq u_0 \\ 10^{k(u_1 - u_0)} & u_1 < u_0 \end{cases} \quad (5)$$

where u_0 is the utility of the existing option, u_1 is the utility of the potential replacement, and k is a constant. Thus, it is always adopted if the potential replacement is perceived as equal or superior to the existing vehicle. Otherwise, it may still be adopted, but with a probability that decreases sharply with its perceived inferiority. This enables fleets to adopt vehicles as pilot projects, to gain information, and to evaluate feasibility (each vehicle option can only be “piloted” once) [12].

Implementation and Confirmation

Finally, fleets procure and operate new vehicles and charging infrastructure. In the implementation step, fleets sell vehicles that are being replaced, procure new vehicles, and design a minimum-cost EV charging system, accounting for existing charging assets, to enable all BEVs to complete their drive cycles. All BEV charging is assumed to be performed at the fleet’s home base. For the

confirmation step, fleet agents record observations of all attributes of all vehicles they own, old and new [12].

In summary, an ABM framework is defined by 1) the types of agents being modeled and 2) the decision theory followed by agents. To put a framework to use, it must be populated using real-world or synthetic data. This entails quantifying and distributing agent attributes, including decision-making preferences and fleet characteristics. Fleet characteristics, including fleet size and vocation, will be initialized from publicly available data and data provided by partners. Decision-making characteristics, such as the relative weights of key metrics, including TCO and up-front cost, are the subject of a significant effort of supporting research and will be initialized based on results of studies in the literature correlating decision preference with fleet characteristics [7].

3 Methods

This section shows how these two methods, LDA and BERTopic, are used to create a comprehensive metric to inform the understanding of how academia and industry might view eMDHDs differently. Figure 3-1 illustrates the activities by which a given category is ranked. This analysis refers to categories as the areas of interest pre-determined by experts with several years of experience in the transportation electrification sector. The right side shows the methods to produce saliency scores

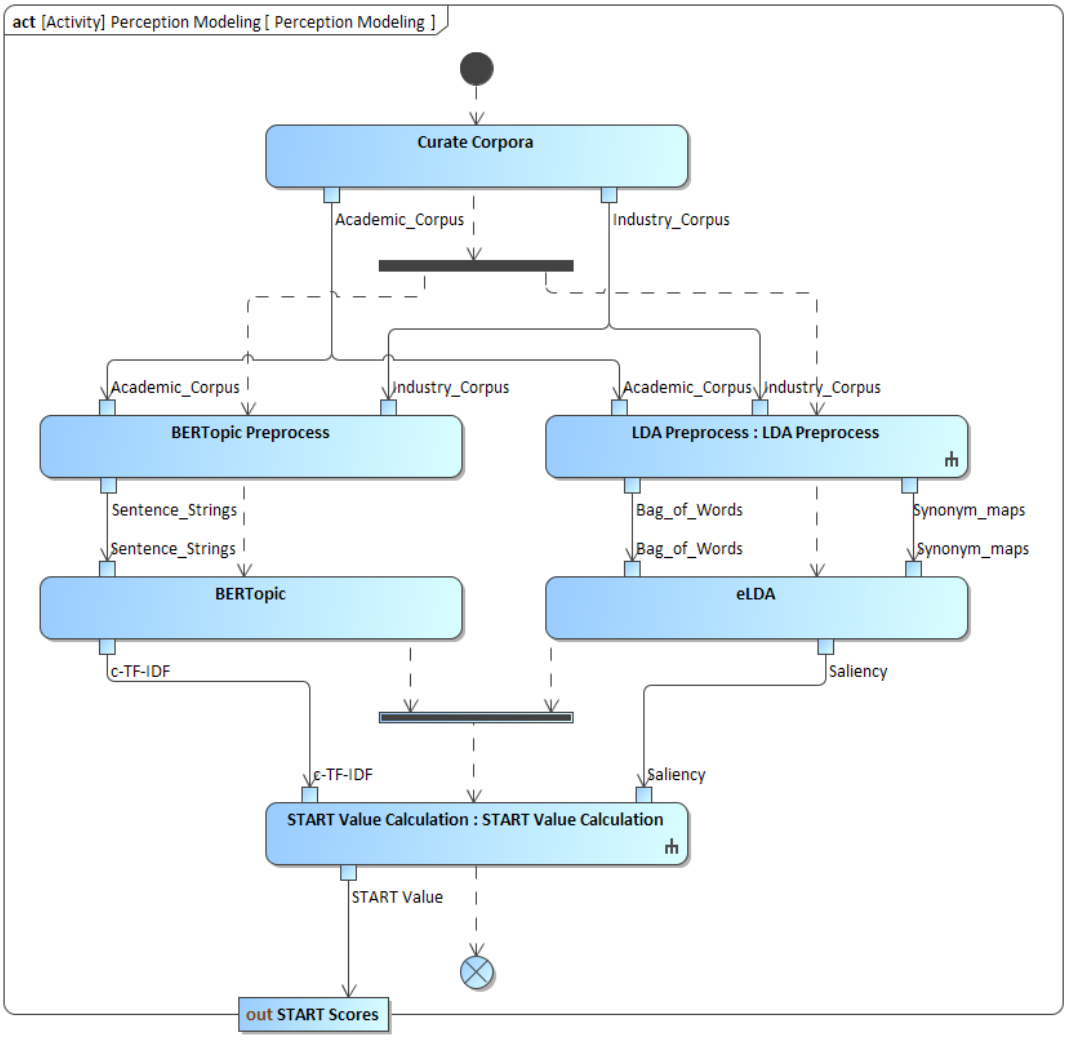

Figure 3-1: Perception modeling activity diagram.

with LDA, while the left shows the methods to produce c-TF-IDF scores with BERTopic. These values are then consolidated into a START score.

As illustrated in Figure 3-1, the proposed algorithm begins with curating the industry and academic corpora. This study hypothesizes that the saliency of topics that describe the state of the field in eMDHDs will be different between industry and academic corpora. It is commonly suggested that there is a disconnect between industry and academia [41,42]. The present study uses the methods described above to model the difference in focus between academia and industry by analyzing the corpus produced by each group. Academic texts include journal articles and conference proceedings with authors primarily from universities, national laboratories, and government. The industry corpus comprises trade journal articles, white papers, and press releases with authors primarily from industrial trade groups and journalists. The diversity of each corpus is represented in A2, which identifies the publication in each corpus by its source type. This figure's interactive versions can be produced using the “paper_sources” notebook in the GitHub repository provided in A4. The code repository for this project can also be used to gain further insight into the data used.

3.1 LDA Method

Data Wrangling/Preprocessing

The present work independently curates an academic [7,9,10,40,43–84] and an industry corpus for analysis. Industry articles are listed in A3. The reader may notice that there are several more industry articles than academic papers. This is due to the effort to have a similar number of words in each corpus after the preprocessing steps.

A corpus must undergo several pre-processing steps to be suitable for input to a topic modeling engine. These include eliminating stop words, creating n-grams, lemmatization, and synonym grouping. The steps are briefly described below.

"Stop words" are common words such as "a" and "the" that do not contribute to coherent topics. They may appear frequently enough to alter topic distribution results and must be filtered out of the corpus. For removal, 173 application-specific words (e.g., "energy," "factor," and "literature") were added to the default stop words list in the Natural Language Processing Tool Kit (NLTK) [85].

N-grams are combinations of two or more neighboring words, such as "long-term cost" (a 3-gram or trigram). Gensim, the package used to implement LDA [85], supports bigrams, two-word phrases, and trigrams, three-word phrases. Each combination of neighboring words is tested in the corpus, and terms with a score higher than the pre-established threshold of 100 are added to the corpus [86]. The n-gram score threshold was set at 100 to eliminate the addition of irrelevant terms. This threshold score is calculated as in (6):

$$score = \frac{(n\ gram\ count - min\ count) * length}{word_a\ count\ *\ word_b\ count}, \tag{6}$$

where,

- $n\ gram\ count$ = number of times the n-gram is present in corpus
- $min\ count$ = minimum collocation count threshold
- $length$ = size of vocabulary
- $word_a\ count$ = number of times the first word occurs, and,
- $word_b\ count$ = number of times the second word occurs.

Lemmatization allows a single word to represent all words with a similar root. For instance, "truck," "trucks," and "trucking" would all be represented by the word "truck." This investigation chooses to

use lemmatization rather than similar operations such as "stemming" because (1) the root words are guaranteed to be English words, therefore increasing the interpretability, and (2) processing time was not a constraint [87].

A synonym group is a group of highly salient words that all refer to the same idea. Most adoption and perception categories can be referenced using any of several synonyms. For example, "environment" and "green" can be categorized under "sustainability," and "total" and "cost" can be classified under "long-term cost." While manually choosing which words represent certain topics does introduce a certain amount of bias, this process aggregates the saliency of individual terms into a more intuitive single concept. The synonyms that fall in each of these groups are shown in Figure 1-2.

Improvements to baseline LDA Preprocessing

Specific to the eMDHD adoption LDA pipeline, there are a few ways in which the methods proposed in this study advance the state of the art in text analytics and modeling adoption behavior. Previous LDA publications [13] explain that categories consist of a pool of words ("environment," zero-emissions," "decarbonize," and "conservation") that are deemed salient [36] by LDA within the corpus. As part of the preprocessing steps for this study, that category name replaces each word mapped to a given category. After this step, "environment," "zero-emissions," "decarbonize," and "conservation" are all mapped to be represented by, for example, the category of *sustainability* in the corpus. The static nature of these synonym mappings is a weakness in the state of the art since forming these maps is an activity in which error can potentially be injected into the ranking system. To alleviate this problem, a set of S synonyms that map to a category are randomly chosen from the original pool without replacement. The experiment ensures that the size of cap S is less than that of the smallest original synonym pool. This ensures that random selection occurs while determining the representative words of each category. This random selection of synonyms is performed N times,

and the resulting stochastic saliency scores are then analyzed to find the average and sample variance over N synonym map scenarios, thereby reducing error that might be associated with any single synonym mapping. For the example above, only three of the four words chosen to represent *sustainability* in the overall map would be randomly selected to serve as the synonym map for any of the N experiments.

Another improvement relative to the state-of-the-art LDA process in the context of eMDHD adoption is that this proposed process improves the sourcing of synonym pools. For this investigation portion, salient words are extracted from each corpus and added to a collective synonym pool. This was deemed appropriate since there was considerable overlap in salient topics between the two corpora, resulting in a shared synonym pool. Therefore, any differences in topic saliency are attributable to the results of the LDA process, not to differences in the pool of synonyms from which LDA can select. Figure 3-2 defines the random synonym experiments.

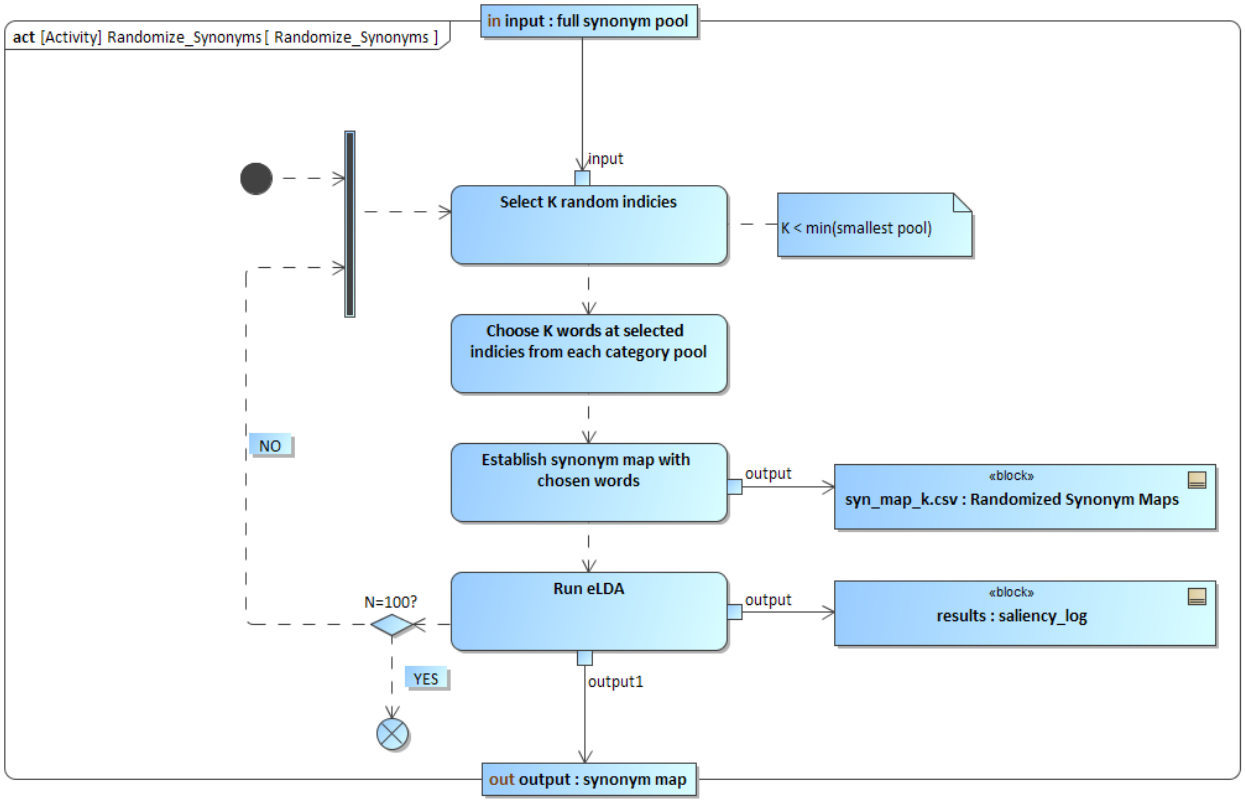

Figure 3-2: Synonym mapping randomization activity diagram.

LDA Implementation

Referencing (1) and (2), the values of α, β, and k significantly impact the model's results. A three-dimensional grid search was initially implemented to optimize parameter selection, as shown in Figure 3-3. After observing inconsistent results between model runs, an ensemble LDA (eLDA) [88] approach was explored as a potential solution. The details of this approach are explained in the following section.

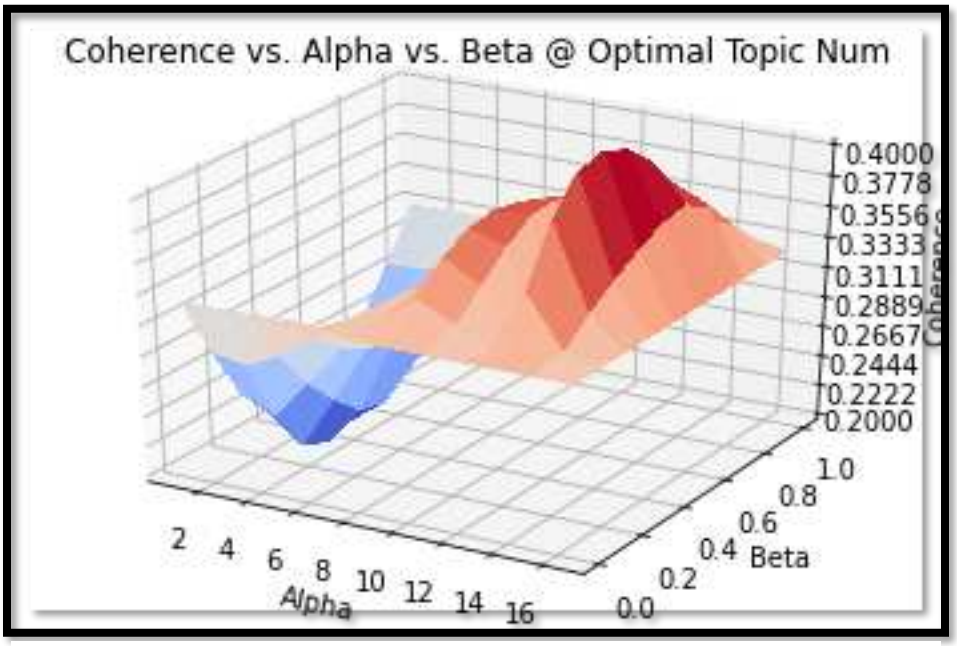

Figure 3-3: Grid search for optimal parameter values

Verification

Typical LDA implementations sample from several random variables during Gibbs sampling (Section 2.1) and therefore suffer from a lack of repeatability. An eLDA routine is used to address this issue, as implemented in Gensim. eLDA uses density-based spatial clustering of applications with noise [89] to categorize the number of topics and beta Dirichlet distributions that create stable results. The process returns a single LDA model with the α, β, and k values shown to produce convergent results.

To ensure consistency, 100 ensembles of eight LDA models each were trained. Section 4.1 shows the results of the validation efforts.

3.2 BERTopic Method

As seen on the left side of Figure 3-1 (page 14), the present process integrates BERTopic results as an evenly weighted contribution to the START metric, as explained in Section 3.3. After splitting each corpus into a list of sentences, the list is used to fit a BERTopic model. The models were then queried for each of the ten categories. A keyword query returns the n topics with the highest similarity to the keyword, and the terms with the highest c-TF-IDF scores within those topics. The BERTopic function *find_topics(search_term)* generates an embedding for the search term and compares this embedding to the previously generated topics to determine similarity [13]. From these top topics (n = 3), the scores of relevant words were summed to arrive at a category c-TF-IDF score. During this step, BERTopic demonstrated a higher recall for relevant words given a search term than LDA. For example, "rebate" scored highly in topics closely related to the "policy/incentives" BERTopic query,

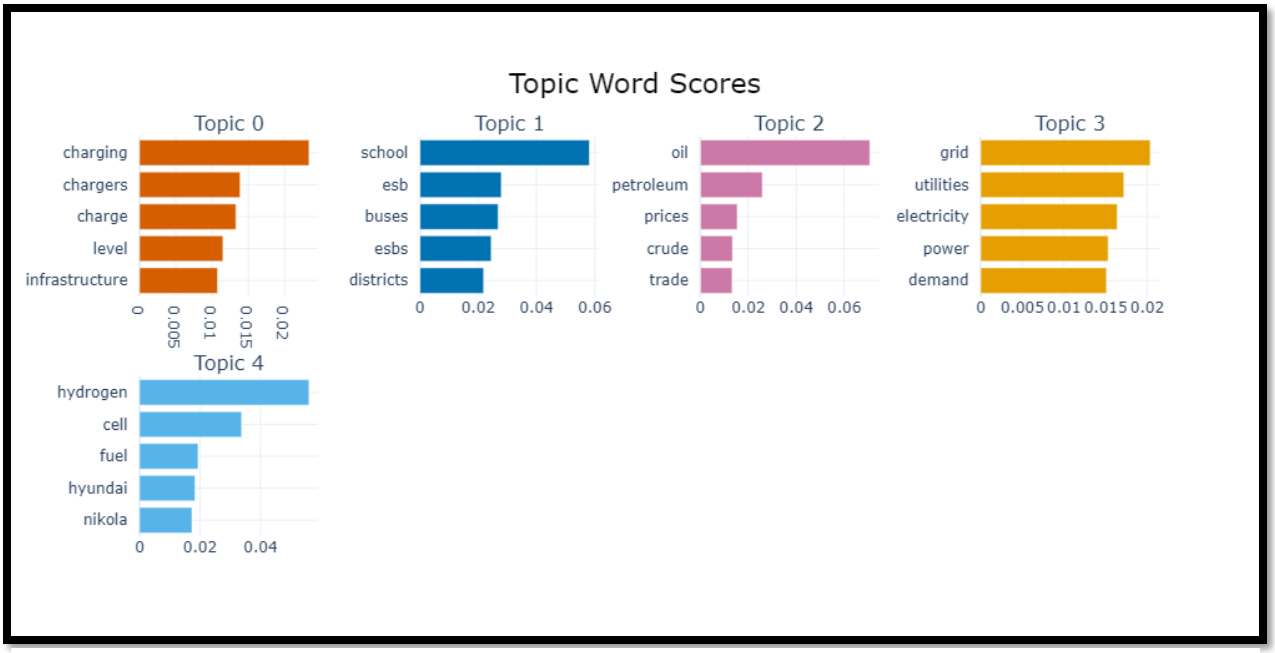

Figure 3-4: c-TF-IDF scores for terms in the top five topics of the industry corpus produced with BERTopic.

whereas “rebate” is not identified as a salient term according to LDA. The top five topics generated from the industry corpus with BERTopic are shown in Figure 3-4.

3.3 Score of Topic and Related Terms

The Score of Topic and Related Terms (START) is a metric that this thesis proposes as a comprehensive ranking for a pre-determined category in a body of text. The START metric weighs a category’s saliency [13] and c-TF-IDF [28] and represents the importance of the category as a single value, as shown in Figure 3-5.

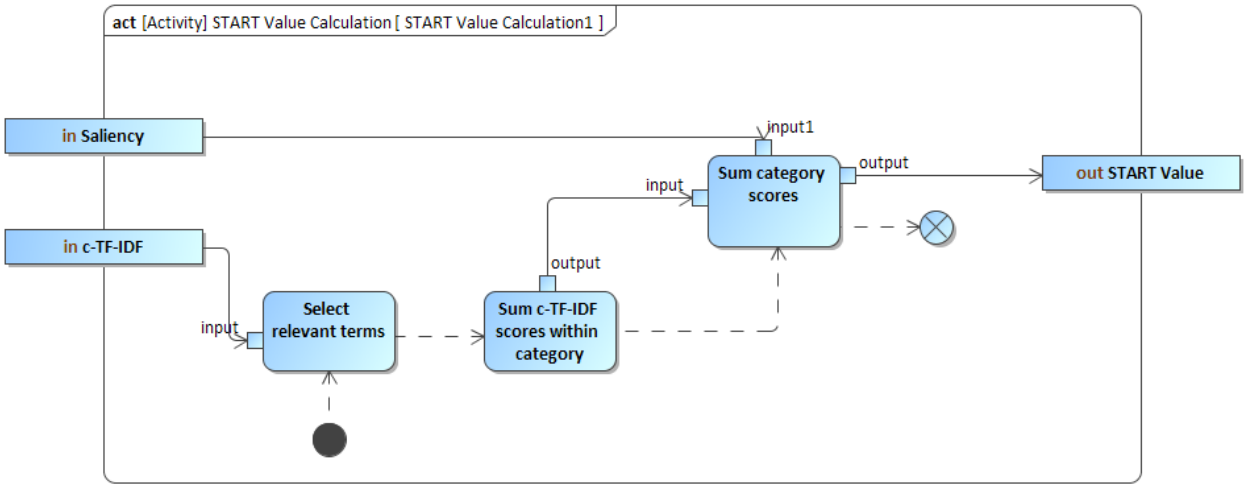

Figure 3-5: START metric calculation for a given category.

4 Results

The proposed methods presented above allow the model user to assess and rank the importance of pre-defined topics within a corpus of text. By applying these methods to both academic and trade literature corpora, one can define the differences between how these bodies of literature understand and address important topics within the field of eMDHD adoption. These results present comparative

results regarding saliency (derived from LDA), c-TF-IDF (derived from BERTopic), and the START metric to inform discussion of these comparisons.

4.1 LDA Results

eLDA Verification Results

Figure 4-1 shows the difference in term saliency as a percent between ensembles. The mean difference in saliency scores between ensembles was 1.49% over 100 independent models, suggesting that the implementation of eLDA is highly repeatable in this project. The norms/attitudes category produced an outlier of just above 4.5%. This leads to the belief that the category's small population of synonyms could make it more susceptible to variation during sampling.

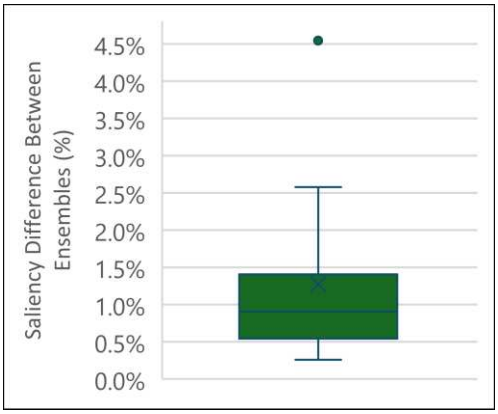

Figure 4-1: Average difference of normalized saliency between ensembles, n = 800

Figure 4-2 shows a comparison of category saliency from both corpora. Each saliency metric is presented with error bars representing the sample variance resulting from the 100 computational experiments described in Section 3.1. As illustrated in Figure 4-2, within the academic corpus, the topics of *long-term cost*, *suitability*, *short-term cost, driver acceptance,* and *brand image* all hold an equal saliency. The industry corpus holds *long-term, suitability, and short-term costs* as highly salient. Still, it reduces reliability and familiarity to intermediate importance before ranking the other categories as a shared bottom-level priority. One can also observe that *familiarity, suitability/compatibility,* and *sustainability* are the only categories that score higher in the industry corpus when considering the displayed error.

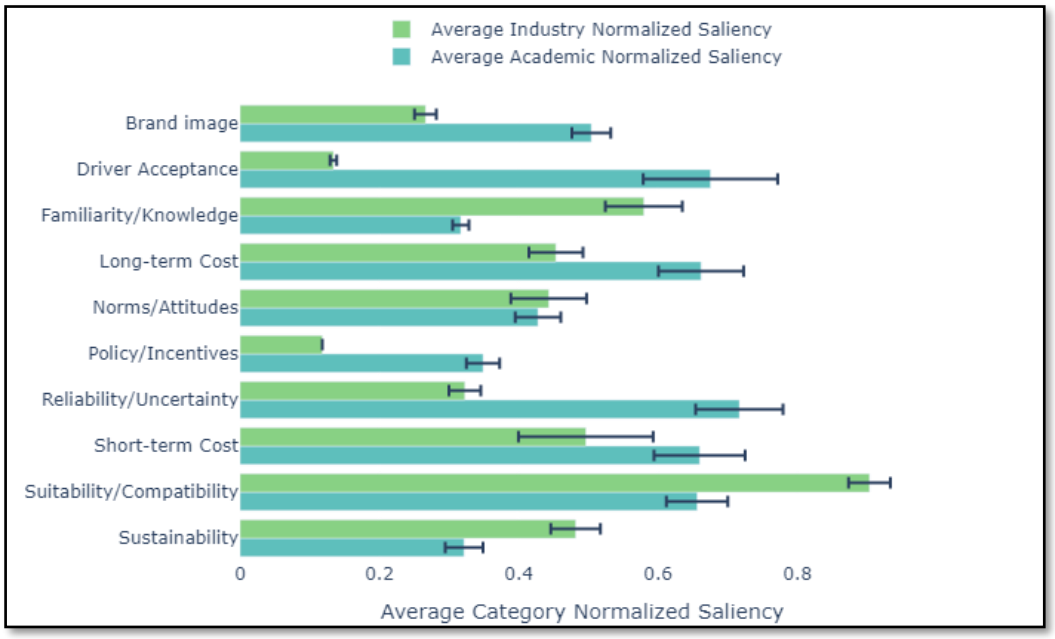

Figure 4-2: Average normalized saliency. Sample variance is displayed with error bars.

4.2 BERTopic Results

Figure 4-3 compares the c-TF-IDF scores from the academic and industry eMDHD adoption corpora. The error bars in this figure are absent because the synonym mapping step does not occur when using BERTopic. This step is not taken due to the model developer's promotion of the model's stability [28]. Using the tokenized sentences from each corpus and referencing them against the original BERTopic training data outlined in Section 2.1, Figure 4-3 shows that both corpora agree on product *familiarity* as the most important category, and *suitability*, the most salient category from LDA, is ranked more evenly with *reliability*.

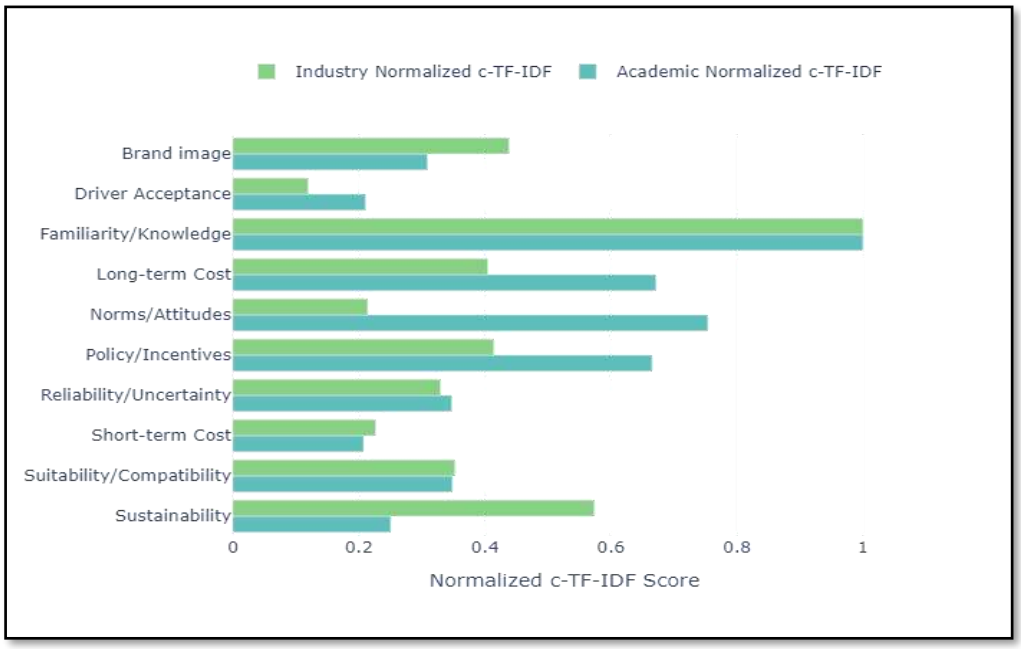

Figure 4-3: Class-Term Frequency-Inverse Document Frequency scores for each corpus

4.3 START Results

As per the procedures outlined in Section 3.3, the mean category saliency was combined with the score resulting from the BERTopic process to arrive at category START scores. A radar plot comparing

the START scores of the two corpora is shown in Figure 4-4. The START scores of the set of categories for the academic corpus are presented in blue, while industry scores are in green.

This visualization shows results that are different from those of LDA alone. *Suitability* commands a high score for both corpora, but concepts such as *short-term cost* show a decline in ranking, justifying the need to employ metrics other than saliency alone. Saliency cannot pick up on contextual references, but the c-TF-IDF can particularly understand context due to its learned relationships between words when placed together or in the same sentence. By combining these metrics into the START metric, both the "commonness" of the topics and their contextual importance within the literature are represented.

Figure 4-4: Radar plot comparison of averaged START values.

4.4 eMDHD Adoption Perception Category Rankings

One objective of this thesis is to clarify that the START scores for ranking perceptions from mining topic models are somewhat volatile and open to different interpretations. This continues to be the case when trying to model something as complex and subjective as the perceptions of human beings. The list in Figure 4-5 provides a firm starting point for understanding the critical factors involved in eMDHD fleet purchasing decisions. To arrive here, the authors mainly adhere to START scores extracted from the industry corpus but consider the saliency experiments described in Section 3.1.

1. *Long-term Cost*
2. *Policy/Incentives*
3. *Suitability*
4. *Reliability*
5. *Driver Acceptance*
6. *Sustainability*
7. *Brand Image*
8. *Short-term cost*
9. *Knowledge/Familiarity*
10. *Norms/Attitudes*

Figure 4-5: Holistic ranking of adoption perceptions

4.5 Results Applied to Adoption Modeling

Another benefit of the START metric of category importance is that it can be used with predefined categories of interest. This characteristic allows START to continuously perform text analytics on an evolving corpus while outputting topic importance weights that can be input into other programs and systems.

For example, the traditional challenge in developing ABMs for adoption modeling has been assigning weighting and importance to the multiplicity of considerations adopters will consider when making a purchase decision. An ABM model of eMDHD adoption would ask the purchasing agent to simultaneously consider the influence of long-term costs, policy/incentives, the technology's suitability to their application, their perception of the reliability of the technology, and so on [12]. Traditionally, the relative importance of these simultaneous considerations has been assessed via surveys. The START methods presented above now give the analyst data-driven means to weigh the importance of these simultaneous considerations of eMDHD adopters. As a demonstration, the rankings from Figure 4-4 are implemented in the ABM designed by Trinko [12] to model eMDHD adoption, as Figure 4-6 depicts. Figure 4-7 shows the differences between the average number of EVs purchased annually across three fleet managers. The figure shows that using adoption perceptions

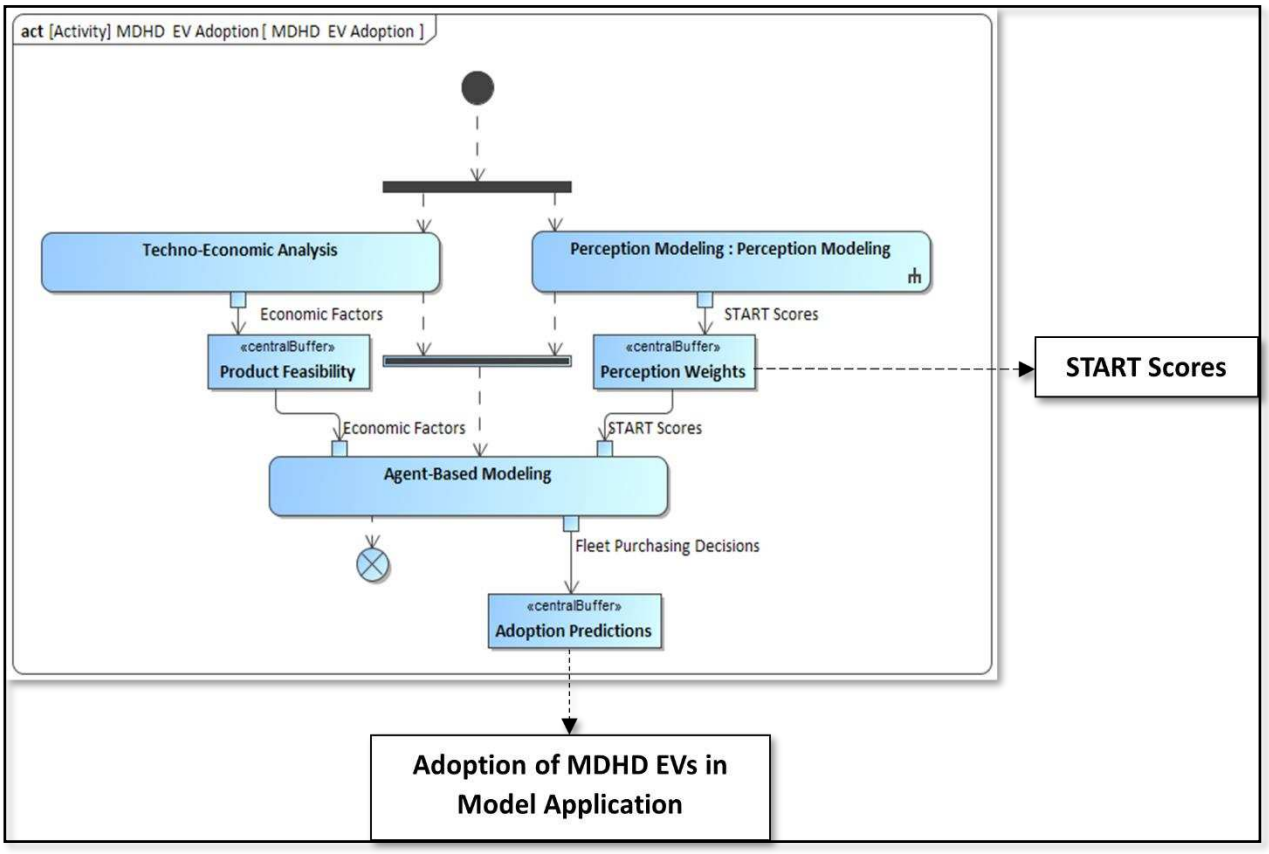

Figure 4-6: Agent-Based Model activity diagram

in academic literature leads to a prediction of much earlier and faster adoption of eMDHDs. In contrast, the use of industry perceptions forecasts a 10-year delay in increased EV purchases by MDHD fleet managers.

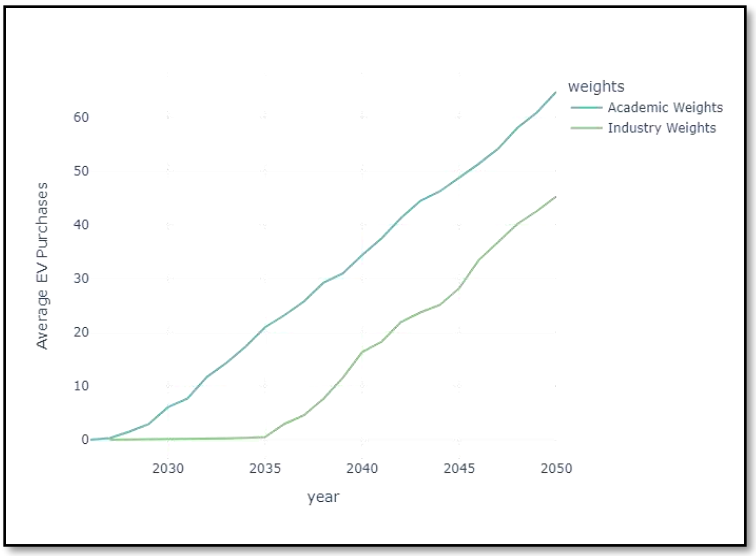

Figure 4-7: Agent-based model results with different perception weights

START implementation opens the door to several opportunities. The possibility now exists to include the data-driven diversity of weightings/opinions in a range set by bounds derived from diverse corpora used in the text mining process. These workflows are unique to the START metric given its ability to distill complex information about various NLP characteristics into a single number.

5 Discussion

5.1 START Costs and Benefits

This thesis introduces START as a more comprehensive text-mining metric of performance that allows for some new meta-algorithmic uses of diverse language models. Using a diversity of methods (such as the LDA and BERTopic methods exemplified here) allows for different components of natural language to be captured and for new uses of NLP to be developed. For example, many clustering

metrics used in NLP (coherence, silhouette score) focus on identifying the most similar topics [90]. These metrics have value in distilling or summarizing a corpus [91], [92], building keyword libraries [93], and assessing sentiment [30]. The benefit of the proposed START methods and metrics is that they allow practitioners to understand how important their pre-determined topics of interest are in a corpus. START gives practitioners a metric to quantify these areas of interest by ranking them from lowest to highest importance.

This study defines topics of interest based on both theories of technology adoption [93,94] and the applied literature on eMDHD vehicle adoption [7,51,75,95–98]. By defining corpora of interest and applying the START metric, analysis can assess the relative importance of these predefined topics experimentally, allowing for direct comparison of the importance of those topics among various corpora (i.e., the academic and industry corpora) and for inputting those of importance into subsequent simulation and analysis (see Section 4.5).

5.2 Ground Truth Examples from Academia and Industry Text Corpora

To validate and build confidence in the results of the START-derived topic scores, one can contemplate examples of the differences between the academia and industry text corpora.

Figure 5-1 highlights differences in wording for sentences regarding the same category. The first sentence is an example from the academic corpus sourced from Bae et al. [7], and the second sentence is an example from the industry corpus sourced from Gurman [99].

Example Academic Corpus Discussion of the topic of "policy/incentives":

- **"Understanding how to navigate this changing marketplace is now a critical challenge for fleets. It is also critical for *policy* makers to understand how fleets will face this challenge in order to establish effective demand-side polies related to AFV adoption in the HDV sector."**

  - Bae et al., 2022 [7]

Example Industry Corpus Discussion of the topic of "policy/incentives"

- ***"Regulations* in the form of sales or purchase mandates are critical to develop ZCEV markets at an accelerated pace."**

  - Gurman, 2021 [99]

Figure 5-1: Excerpts from both corpora discussing the idea of policy and incentives.

While both sentences discuss concepts surrounding policy, only the academic sentence explicitly uses the term, while the trade journal makes a reference to policy through the related idea of regulations while also focusing on capital costs. The LDA results presented in Figure 4-2 show that policy/incentives are a low-saliency topic that does not surpass a score of 0.4 in the best case (academic corpus testing). Figure 5-1 helps to decipher why the ***c-TF-IDF scores*** shown in Figure 4-3 for *policy/incentives* are much higher. This is because BERTopic can pick up on the non-literal references to a category (*policy/incentives*) through its synonyms (*regulations*), such as in the trade journal sentence. While *regulation* is a part of the set of synonyms in the synonym map for *policy/incentives* in the LDA process, the randomization experiment does not always choose it as a

representative term, which demotes its assessed salience when using methods such as LDA. Because START incorporates extracted saliency from LDA and maintains sentence structure importance through BERTopic, START scores can be used to formulate a clear and actionable ranking of the perception categories presented in Figure 1-2.

6 Future Work

Moving forward, eMDHD adoption research holds two promising areas for development. These include agent refinement and increased use of machine learning to provide agents with enhanced information.

Agent refinement refers to improving the system's data foundation for different agent types (stakeholders). In the case of the eMDHD industry, this includes utility providers and original equipment manufacturers. Utility agents should be provided with definitions of local electricity rates. This would allow for an accurate electricity cost input to the TEA of a fleet agent at a localized level. It would also introduce the possibility to test the effects of different electricity subsidy scenarios at a highly granular level. A demonstration of this concept is scheduled to be published later this year.

Another area of improvement lies in using large language models (LLM) to improve perception ranking inputs. LLMs are complex combinations of generative-adversarial networks (GANs) and transformers such as BERTopic. They are considered to be the “next step” in NLP. So far, this investigation has relied on curated papers that provide an explainable result but lacks the ability to simultaneously consider all perspectives supplied on the internet and produce a decision. A potential experiment would be to “interview” an LLM, such as GPT-4. This would entail providing it with parameters including vocation, fleet size, fleet composition, and location, then asking how many vehicles it would choose to replace

with electric ones. The responses for each vocation/fleet scenario would then be passed to the START metric pipeline to produce perception rankings.

The use of LLMs in research introduces some philosophical debate. Results from these models are hard to explain, and the combination of an enormous training corpus paired with a latent space boasting a size that is hard for humans to comprehend means that LLMs can produce incorrect answers. On the other hand, this is an attempt to model the social behavior of human beings. Would it not make sense to introduce this level of error and randomness when trying to model the highly variable and unpredictable nature of one's perceptions?

7 Conclusion

Now having completed the explained research activities we can identify the following primary contributions of this effort as

- the START metric development which allows for modeling nuances of language in a new way,
- the composed comparison of topics embeded within the eMDHD literature of academic and industry corpora, and
- the integration of topic modeling with ABM as a means to provide ABMs with data-derived agent weight schemes.

This thesis helps to clarify the understanding of eMDHD adoption by developing a comprehensive metric, START, that combines results from a traditional statistical model, LDA, and a cutting-edge deep learning model, BERTopic. START allows an individual to provide actionable insights to industry and academics alike. Professionals can use these results to make more informed decisions based on the information distilled from the text that their industry produces, and academic researchers can use the same information to help guide their research in directions relevant to industry needs. While

this study encounters several of the complexities inherent with NLP and cluster analysis, this work contributes a robust ranking of perceptions in the eMDHD adoption arena and objectively shows significant differences in focus between the academic and professional worlds. These results are then utilized to illustrate how the differences in focus between the academic and industry corpora lead to differences in forecasted eMDHD adoption rates.

Sources

[1] Lashof D, Saha D, Barbanell M, Chen J, DeConcini C, Ranganathan J, et al. After Inflation Reduction Act: 5 Ways to Hit the US Emissions Goal 2022.

[2] Paris Agreement. Int Leg Mater 2016;55:743.

[3] CDOT. Planning Initiatives. Colo Dep Transp 2018. https://www.codot.gov/programs/innovativemobility/electrification/planning-initiatives-and-documents (accessed August 30, 2022).

[4] Adnan N, Nordin SM, Rahman I, Vasant PM, Noor A. A comprehensive review on theoretical framework-based electric vehicle consumer adoption research. Int J Energy Res 2017;41:317–35. https://doi.org/10.1002/er.3640.

[5] Nazari F, Rahimi E, Mohammadian A (Kouros). Simultaneous estimation of battery electric vehicle adoption with endogenous willingness to pay. eTransportation 2019;1:100008. https://doi.org/10.1016/j.etran.2019.100008.

[6] Borlaug B, Muratori M, Gilleran M, Woody D, Muston W, Canada T, et al. Heavy-duty truck electrification and the impacts of depot charging on electricity distribution systems. Nat Energy 2021;6:673–82. https://doi.org/10.1038/s41560-021-00855-0.

[7] Bae Y, Mitra SK, Rindt CR, Ritchie SG. Factors influencing alternative fuel adoption decisions in heavy-duty vehicle fleets. Transp Res Part Transp Environ 2022;102:103150. https://doi.org/10.1016/j.trd.2021.103150.

[8] Anderhofstadt B, Spinler S. Factors affecting the purchasing decision and operation of alternative fuel-powered heavy-duty trucks in Germany – A Delphi study. Transp Res Part Transp Environ 2019;73:87–107. https://doi.org/10.1016/j.trd.2019.06.003.

[9] Bennett R, Kottasz R, Shaw S. Factors potentially affecting the successful promotion of electric vehicles. J Soc Mark 2016;6:62–82. https://doi.org/10.1108/JSOCM-08-2015-0059.

[10] Klemick H, Kopits E, Wolverton A, Sargent K. Heavy-duty trucking and the energy efficiency paradox: Evidence from focus groups and interviews. Transp Res Part Policy Pract 2015;77:154–66. https://doi.org/10.1016/j.tra.2015.04.004.

[11] Miller M, Wang Q, Fulton L. Truck Choice Modeling: Understanding California's Transition to Zero-Emission Vehicle Trucks Taking into Account Truck Technologies, Costs, and Fleet Decision Behavior 2017.

[12] Trinko D. Modeling and Simulation to Investigate the Electrification Potential of Medium- and Heavy-Duty Vehicle Fleets. 2023.

[13] Ouren FT, Trinko D, Coburn T, Simske S, Bradley T. Developing a Profile of Medium- and Heavy-Duty Electric Vehicle Fleet Adopters with Text Mining and Machine Learning 2023. https://doi.org/10.2139/ssrn.4373106.

[14] Ouren F, Bradley TH, Trinko D, Coburn T, Miller E. Comparing Perspectives on Medium- and Heavy-Duty Electric Vehicle Adoption with Natural Language Processing. Renew Sustain Energy Rev 2024;RSER-D-23-04514.

[15] Anandarajan M, Hill C, Nolan T. Practical Text Analytics: Maximizing the Value of Text Data. vol. 2. Cham: Springer International Publishing; 2019. https://doi.org/10.1007/978-3-319-95663-3.

[16] Martinez AR. 4 - Data Mining of Text Files. In: Rao CR, Wegman EJ, Solka JL, editors. Handb. Stat., vol. 24, Elsevier; 2005, p. 109–31. https://doi.org/10.1016/S0169-7161(04)24004-4.

[17] Blei DM, Jordan M, Ng A. Latent Dirichlet Allocation 2003:30.

[18] Hofmann T. Probabilistic Latent Semantic Analysis 2013. https://doi.org/10.48550/arXiv.1301.6705.

[19] Churchill R, Singh L. The Evolution of Topic Modeling. ACM Comput Surv 2022:3507900. https://doi.org/10.1145/3507900.
[20] Kumar M, Ng J. Using text mining and topic modelling to understand success and growth factors in Global Renewable Energy projects. Renew Energy Focus 2022;42:211–20. https://doi.org/10.1016/j.ref.2022.06.010.
[21] Suri P, Roy NR. Comparison between LDA & NMF for event-detection from large text stream data. 2017 3rd Int. Conf. Comput. Intell. Commun. Technol. CICT, 2017, p. 1–5. https://doi.org/10.1109/CIACT.2017.7977281.
[22] Souza J, Gomes Jr J, Marques F, Julio A, Souza J. A systematic mapping on automatic classification of fake news in social media. Soc Netw Anal Min 2020;10. https://doi.org/10.1007/s13278-020-00659-2.
[23] Liang J, Liu P, Tan J, Bai S. Sentiment Classification Based on AS-LDA Model. Procedia Comput Sci 2014;31:511–6. https://doi.org/10.1016/j.procs.2014.05.296.
[24] Selvi M, Thangaramya K, Saranya MS, Kulothungan K, Ganapathy S, Kannan A. Classification of Medical Dataset Along with Topic Modeling Using LDA. In: Nath V, Mandal JK, editors. Nanoelectron. Circuits Commun. Syst., Singapore: Springer; 2019, p. 1–11. https://doi.org/10.1007/978-981-13-0776-8_1.
[25] Tomar A. Topic modeling using Latent Dirichlet Allocation(LDA) and Gibbs Sampling explained! Anal Vidhya 2019. https://medium.com/analytics-vidhya/topic-modeling-using-lda-and-gibbs-sampling-explained-49d49b3d1045 (accessed October 24, 2022).
[26] Chuang J, Manning CD, Heer J. Termite: visualization techniques for assessing textual topic models. Proc. Int. Work. Conf. Adv. Vis. Interfaces - AVI 12, Capri Island, Italy: ACM Press; 2012, p. 74. https://doi.org/10.1145/2254556.2254572.
[27] Joyce JM. Kullback-Leibler Divergence. In: Lovric M, editor. Int. Encycl. Stat. Sci., Berlin, Heidelberg: Springer; 2011, p. 720–2. https://doi.org/10.1007/978-3-642-04898-2_327.
[28] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure 2022. https://doi.org/10.48550/arXiv.2203.05794.
[29] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding 2019. https://doi.org/10.48550/arXiv.1810.04805.
[30] Egger R, Yu J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Front Sociol 2022;7:886498. https://doi.org/10.3389/fsoc.2022.886498.
[31] McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction 2020. https://doi.org/10.48550/arXiv.1802.03426.
[32] Malzer C, Baum M. A Hybrid Approach To Hierarchical Density-based Cluster Selection. 2020 IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst. MFI, 2020, p. 223–8. https://doi.org/10.1109/MFI49285.2020.9235263.
[33] Qaiser S, Ali R. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. Int J Comput Appl 2018;181. https://doi.org/10.5120/ijca2018917395.
[34] Albanese NC. Topic Modeling with LSA, pLSA, LDA, NMF, BERTopic, Top2Vec: a Comparison. Medium 2022. https://towardsdatascience.com/topic-modeling-with-lsa-plsa-lda-nmf-bertopic-top2vec-a-comparison-5e6ce4b1e4a5 (accessed May 25, 2023).
[35] Robinson D. Is LDA Topic Modeling Dead? Medium 2022. https://towardsdatascience.com/is-lda-topic-modeling-dead-9543c18488fa (accessed May 25, 2023).
[36] Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. Proc. Workshop Interact. Lang. Learn. Vis. Interfaces, Baltimore, Maryland, USA: Association for Computational Linguistics; 2014, p. 63–70. https://doi.org/10.3115/v1/W14-3110.

[37] Joshi T. Silhouette Score. Medium 2021. https://tushar-joshi-89.medium.com/silhouette-score-a9f7d8d78f29 (accessed June 1, 2023).
[38] Forrest K, Mac Kinnon M, Tarroja B, Samuelsen S. Estimating the technical feasibility of fuel cell and battery electric vehicles for the medium and heavy duty sectors in California. Appl Energy 2020;276:115439. https://doi.org/10.1016/j.apenergy.2020.115439.
[39] Ollia S, Ward J, Isaacs R. 2022 Annual Progress Report. DOE VTO; 2022.
[40] Nesbitt K, Sperling D. Fleet purchase behavior: decision processes and implications for new vehicle technologies and fuels. Transp Res Part C Emerg Technol 2001;9:297–318. https://doi.org/10.1016/S0968-090X(00)00035-8.
[41] Gould J. Beyond academia: debunking the industry–academia barrier myth. Nature 2022. https://doi.org/10.1038/d41586-022-00291-4.
[42] Muller G. Industry and Academia: Why Practitioners and Researchers are Disconnected. vol. 15, 2005. https://doi.org/10.1002/j.2334-5837.2005.tb00648.x.
[43] Altenburg M, Anand NR, Balm SH, Ploos van Amstel W. Electric freight vehicles in city logistics: insights into decision making process of frontrunner companies. Eur Battery Hybrid Fuel Cell Electr Veh Congr 2017.
[44] Blynn K, Attanucci J. Accelerating Bus Electrification: A Mixed Methods Analysis of Barriers and Drivers to Scaling Transit Fleet Electrification. Transp Res Rec 2019;2673:577–87. https://doi.org/10.1177/0361198119842117.
[45] Fleet Management and the Adoption of Innovations by Corporate Car Fleets: Exploratory Approach - Virginie Boutueil, 2016 n.d. https://journals.sagepub.com/doi/abs/10.3141/2598-10 (accessed February 14, 2024).
[46] Chocteau V, Drake D, Kleindorfer PR, Orsato RJ, Roset A. Collaborative Innovation for Sustainable Fleet Operations: The Electric Vehicle Adoption Decision 2011. https://doi.org/10.2139/ssrn.1791623.
[47] Christensen L, Klauenberg J, Kveiborg O, Rudolph C. Suitability of commercial transport for a shift to electric mobility with Denmark and Germany as use cases. Res Transp Econ 2017;64:48–60. https://doi.org/10.1016/j.retrec.2017.08.004.
[48] Dziubak T, Wysocki T, Dziubak S. Selection of vehicles for fleet of transport company on the basis of observation of their operational reliability. Eksploat Niezawodn - Maint Reliab 2021;23:184–94. https://doi.org/10.17531/ein.2021.1.19.
[49] Gay D, Rogers T, Shirley R. Small island developing states and their suitability for electric vehicles and vehicle-to-grid services. Util Policy 2018;55:69–78. https://doi.org/10.1016/j.jup.2018.09.006.
[50] Globisch J, Dütschke E, Schleich J. Acceptance of electric passenger cars in commercial fleets. Transp Res Part Policy Pract 2018;116:122–9. https://doi.org/10.1016/j.tra.2018.06.004.
[51] Globisch J, Dütschke E, Wietschel M. Adoption of electric vehicles in commercial fleets: Why do car pool managers campaign for BEV procurement? Transp Res Part Transp Environ 2018;64:122–33. https://doi.org/10.1016/j.trd.2017.10.010.
[52] Golob TF, Torous J, Bradley M, Brownstone D, Crane S, Bunch D. Commercial Fleet Demand for Alternative-Fuel Vehicles in California 1996.
[53] Greene DL. Vehicle Use and Fuel Economy: How Big is the “Rebound” Effect? Energy J 1992;Volume 13:117–44.
[54] Jaller M, Pineda L, Gueldas Y, Alemi F, Otay I. Fostering the Use of Zero and Near Zero Emission Vehicles in Freight Operations 2020. https://doi.org/10.7922/G2QC01SB.
[55] Johns KD, Khovanova KM, Welch EW. Fleet Conversion in Local Government: Determinants of Driver Fuel Choice for Bi-Fuel Vehicles. Environ Behav 2009;41:402–26. https://doi.org/10.1177/0013916507312423.

[56] Kaplan S, Gruber J, Reinthaler M, Klauenberg J. Intentions to introduce electric vehicles in the commercial sector: A model based on the theory of planned behaviour. Res Transp Econ 2016;55:12–9. https://doi.org/10.1016/j.retrec.2016.04.006.
[57] Keller AZ, Goblin MT, Farnworth NR. Reliability analysis of commercial vehicle engines. Reliab Eng 1985;10:15–25. https://doi.org/10.1016/0143-8174(85)90039-3.
[58] Kirk JL, Bristow AL, Zanni AM. Exploring the market for Compressed Natural Gas light commercial vehicles in the United Kingdom. Transp Res Part Transp Environ 2014;29:22–31. https://doi.org/10.1016/j.trd.2014.03.004.
[59] Klauenberg J, Rudolph C, Zajicek J. Potential Users of Electric Mobility in Commercial Transport – Identification and Recommendations. Transp Res Procedia 2016;16:202–16. https://doi.org/10.1016/j.trpro.2016.11.020.
[60] Koetse MJ, Hoen A. Preferences for alternative fuel vehicles of company car drivers. Resour Energy Econ 2014;37:279–301. https://doi.org/10.1016/j.reseneeco.2013.12.006.
[61] Li J, Wang D, Zhang J. Heterogeneous fixed fleet vehicle routing problem based on fuel and carbon emissions. J Clean Prod 2018;201:896–908. https://doi.org/10.1016/j.jclepro.2018.08.075.
[62] Loo BPY, Wong SC, Hau TD. Introducing alternative fuel vehicles in Hong Kong: views from the public light bus industry. Transportation 2006;33:605–19. https://doi.org/10.1007/s11116-006-7947-5.
[63] Mohamed M, Ferguson M, Kanaroglou P. What hinders adoption of the electric bus in Canadian transit? Perspectives of transit providers. Transp Res Part Transp Environ 2018;64:134–49. https://doi.org/10.1016/j.trd.2017.09.019.
[64] Mohammed L, Niesten E, Gagliardi D. Adoption of alternative fuel vehicle fleets – A theoretical framework of barriers and enablers. Transp Res Part Transp Environ 2020;88:102558. https://doi.org/10.1016/j.trd.2020.102558.
[65] Morganti E, Browne M. Technical and operational obstacles to the adoption of electric vans in France and the UK: An operator perspective. Transp Policy 2018;63:90–7. https://doi.org/10.1016/j.tranpol.2017.12.010.
[66] Nesbitt K, Sperling D. Myths regarding alternative fuel vehicle demand by light-duty vehicle fleets. Transp Res Part Transp Environ 1998;3:259–69. https://doi.org/10.1016/S1361-9209(98)00006-6.
[67] Pfoser S, Schauer O, Costa Y. Acceptance of LNG as an alternative fuel: Determinants and policy implications. Energy Policy 2018;120:259–67. https://doi.org/10.1016/j.enpol.2018.05.046.
[68] Pinto R, Lagorio A. Supporting the decision making process in the urban freight fleet composition problem. Int J Prod Res 2021;59:3861–79. https://doi.org/10.1080/00207543.2020.1753896.
[69] Rahm D, Coggburn J. Environmentally Preferable Procurement Greening U.S. State Government Fleets. Public Works Manag Policy 2007;12. https://doi.org/10.1177/1087724X07304302.
[70] ReVelle C, Marianov V. A probabilistic FLEET model with individual vehicle reliability requirements. Eur J Oper Res 1991;53:93–105. https://doi.org/10.1016/0377-2217(91)90095-D.
[71] van Rijnsoever FJ, Hagen P, Willems M. Preferences for alternative fuel vehicles by Dutch local governments. Transp Res Part Transp Environ 2013;20:15–20. https://doi.org/10.1016/j.trd.2013.01.005.
[72] Roemer E, Henseler J. The dynamics of electric vehicle acceptance in corporate fleets: Evidence from Germany. Technol Soc 2022;68:101938. https://doi.org/10.1016/j.techsoc.2022.101938.

[73] Saukkonen N, Laine T, Suomala P. How do companies decide? Emotional triggers and drivers of investment in natural gas and biogas vehicles. Energy Res Soc Sci 2017;34:49–61. https://doi.org/10.1016/j.erss.2017.06.005.
[74] Seitz CS, Beuttenmüller O, Terzidis O. Organizational adoption behavior of CO2-saving power train technologies: An empirical study on the German heavy-duty vehicles market. Transp Res Part Policy Pract 2015;80:247–62. https://doi.org/10.1016/j.tra.2015.08.002.
[75] Sierzchula W. Factors influencing fleet manager adoption of electric vehicles. Transp Res Part Transp Environ 2014;31:126–34. https://doi.org/10.1016/j.trd.2014.05.022.
[76] Skippon S, Chappell J. Fleets' motivations for plug-in vehicle adoption and usage: U.K. case studies. Transp Res Part Transp Environ 2019;71:67–84. https://doi.org/10.1016/j.trd.2018.12.009.
[77] Sun L, Zhang Y, Hu X. Economical-traveling-distance-based fleet composition with fuel costs: An application in petrol distribution. Transp Res Part E Logist Transp Rev 2021;147:102223. https://doi.org/10.1016/j.tre.2021.102223.
[78] Teoh T, Kunze O, Teo C-C. Methodology to Evaluate the Operational Suitability of Electromobility Systems for Urban Logistics Operations. Transp Res Procedia 2016;12:288–300. https://doi.org/10.1016/j.trpro.2016.02.066.
[79] Walter S, Ulli-Beer S, Wokaun A. Assessing customer preferences for hydrogen-powered street sweepers: A choice experiment. Int J Hydrog Energy 2012;37:12003–14. https://doi.org/10.1016/j.ijhydene.2012.05.026.
[80] Wikström M, Hansson L, Alvfors P. Investigating barriers for plug-in electric vehicle deployment in fleets. Transp Res Part Transp Environ 2016;49:59–67. https://doi.org/10.1016/j.trd.2016.08.008.
[81] Wikström M, Hansson L, Alvfors P. An End has a Start – Investigating the Usage of Electric Vehicles in Commercial Fleets. Energy Procedia 2015;75:1932–7. https://doi.org/10.1016/j.egypro.2015.07.223.
[82] Wikström M, Hansson L, Alvfors P. Socio-technical experiences from electric vehicle utilisation in commercial fleets. Appl Energy 2014;123:82–93. https://doi.org/10.1016/j.apenergy.2014.02.051.
[83] Wolff S, Madlener R. Driven by change: Commercial drivers' acceptance and efficiency perceptions of light-duty electric vehicle usage in Germany. Transp Res Part C Emerg Technol 2019;105:262–82. https://doi.org/10.1016/j.trc.2019.05.017.
[84] Zhang Y, Jiang Y, Rui W, Thompson RG. Analyzing truck fleets' acceptance of alternative fuel freight vehicles in China. Renew Energy 2019;134:1148–55. https://doi.org/10.1016/j.renene.2018.09.016.
[85] Řehůřek R, Sojka P. Software Framework for Topic Modelling with Large Corpora 2010:45–50.
[86] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality 2013. https://doi.org/10.48550/arXiv.1310.4546.
[87] Manning CD, Raghavan P, Schütz H. Stemming and lemmatization. Introd. Inf. Retr. Online Version, Cambridge University Press; 2008.
[88] Brigl T. Extracting Reliable Topics using Ensemble Latent Dirichlet Allocation. Technische Hochschule Ingolstadt, 2019.
[89] Ester M, Kriegel H-P, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 1996:6.
[90] Basmatkar P, Maurya M. An Overview of Contextual Topic Modeling Using Bidirectional Encoder Representations from Transformers. In: Bindhu V, Tavares JMRS, Du K-L, editors. Proc. Third Int. Conf. Commun. Comput. Electron. Syst., Singapore: Springer; 2022, p. 489–504. https://doi.org/10.1007/978-981-16-8862-1_32.

[91] The challenges of automatic summarization n.d. https://ieeexplore.ieee.org/abstract/document/881692/ (accessed September 26, 2023).
[92] Method Of Text Summarization Using Lsa And Sentence Based Topic Modelling With Bert n.d. https://ieeexplore.ieee.org/abstract/document/9395976/?casa_token=WXzolzIx-mYAAAAA:JeflfztOvTMTc6h2f_crCd8XFRs_BxMtvUNup8vDm7YMiZsHktpMTYQBWmJ4oSJW2lWbPASsOA (accessed September 26, 2023).
[93] Oliveira T, Martins MF. Literature Review of Information Technology Adoption Models at Firm Level. Electron J Inf Syst Eval 2011;14:pp110-121-pp110-121.
[94] Parente SL, Prescott EC. Barriers to Technology Adoption and Development. J Polit Econ 1994;102:298–321. https://doi.org/10.1086/261933.
[95] Liu Y, Klampfl E, Tamor MA. Modified Bass Model with External Factors for Electric Vehicle Adoption. Warrendale, PA: SAE International; 2013. https://doi.org/10.4271/2013-01-0505.
[96] Debnath R, Bardhan R, Reiner DM, Miller JR. Political, economic, social, technological, legal and environmental dimensions of electric vehicle adoption in the United States: A social-media interaction analysis. Renew Sustain Energy Rev 2021;152:111707. https://doi.org/10.1016/j.rser.2021.111707.
[97] Kumar RR, Alok K. Adoption of electric vehicle: A literature review and prospects for sustainability. J Clean Prod 2020;253:119911. https://doi.org/10.1016/j.jclepro.2019.119911.
[98] Anosike A, Loomes H, Udokporo CK, Garza-Reyes JA. Exploring the challenges of electric vehicle adoption in final mile parcel delivery. Int J Logist Res Appl 2021;0:1–25. https://doi.org/10.1080/13675567.2021.1978409.
[99] Gurman RO. TAKING COMMERCIAL FLEET ELECTRIFICATION TO SCALE: CalStart 2021. https://calstart.org/wp-content/uploads/2021/03/Taking-Commercial-Fleet-Electrification-to-Scale-White-Paper.pdf.

Appendix

A1

List of Synonyms for Topic Modeling

Word	Category
cost	long_term_cost
market	long_term_cost
diesel	long_term_cost
value	long_term_cost
infrastructure	long_term_cost
year	long_term_cost
total	long_term_cost
future	long_term_cost
investment	long_term_cost
economic	long_term_cost
expect	long_term_cost
interest	long_term_cost
utility	long_term_cost
economy	long_term_cost
saving	long_term_cost
plan	long_term_cost
purchase	short_term_cost
battery	short_term_cost
price	short_term_cost
cell	short_term_cost
present	short_term_cost
procurement	short_term_cost
pay	short_term_cost
buy	short_term_cost
short	short_term_cost
policy	policy_incentives
public	policy_incentives
state	policy_incentives
government	policy_incentives
incentive	policy_incentives
agency	policy_incentives
city	policy_incentives
program	policy_incentives
regulation	policy_incentives
tax	policy_incentives
national	policy_incentives
country	policy_incentives
fuel	suitability_compatibility
time	suitability_compatibility
charge	suitability_compatibility
transportation	suitability_compatibility
range	suitability_compatibility
operation	suitability_compatibility
need	suitability_compatibility
business	suitability_compatibility
demand	suitability_compatibility
operational	suitability_compatibility
change	suitability_compatibility
performance	suitability_compatibility
available	suitability_compatibility
operate	suitability_compatibility
customer	suitability_compatibility
require	suitability_compatibility
plug	suitability_compatibility
conventional	suitability_compatibility
power	suitability_compatibility
compare	suitability_compatibility
current	suitability_compatibility
requirement	suitability_compatibility
technical	suitability_compatibility
station	suitability_compatibility
mobility	suitability_compatibility
recharge	suitability_compatibility
driving_range	suitability_compatibility
research	familiarity_knowledge
study	familiarity_knowledge
analysis	familiarity_knowledge
make	familiarity_knowledge
information	familiarity_knowledge

potential	familiarity_knowledge
relate	familiarity_knowledge
paper	familiarity_knowledge
report	familiarity_knowledge
theory	familiarity_knowledge
understand	familiarity_knowledge
knowledge	familiarity_knowledge
literature	familiarity_knowledge
estimate	familiarity_knowledge
initial	familiarity_knowledge
conduct	familiarity_knowledge
network	familiarity_knowledge
review	familiarity_knowledge
manufacturer	familiarity_knowledge
know	familiarity_knowledge
fact	familiarity_knowledge
discuss	familiarity_knowledge
similar	familiarity_knowledge
training	familiarity_knowledge
insight	familiarity_knowledge
learn	familiarity_knowledge
introduce	familiarity_knowledge
expert	familiarity_knowledge
decision	norms_attitudes
manager	norms_attitudes
consider	norms_attitudes
private	norms_attitudes
barrier	norms_attitudes
firm	norms_attitudes
perceive	norms_attitudes
see	norms_attitudes
preference	norms_attitudes
management	norms_attitudes
behavior	norms_attitudes
influence	norms_attitudes
attitude	norms_attitudes
inï¬‚uence	norms_attitudes
intention	norms_attitudes
indicate	norms_attitudes
suggest	norms_attitudes
positive	norms_attitudes
social	norms_attitudes
decision_make	norms_attitudes
consideration	norms_attitudes
perspective	norms_attitudes
standard	norms_attitudes
willingness	norms_attitudes
negative	norms_attitudes
perception	norms_attitudes
behaviour	norms_attitudes
seem	norms_attitudes
alternative	brand_image
organization	brand_image
innovation	brand_image
design	brand_image
improve	brand_image
strategy	brand_image
first	brand_image
organizational	brand_image
org	brand_image
early	brand_image
corporate	brand_image
introduction	brand_image
lead	brand_image
early_adopter	brand_image
image	brand_image
promote	brand_image
energy	sustainability
evs	sustainability
environmental	sustainability
afvs	sustainability
hybrid	sustainability
efv	sustainability
afv	sustainability
reduce	sustainability
impact	sustainability
emission	sustainability
green	sustainability
lng	sustainability
reduction	sustainability
bev	sustainability
hydrogen	sustainability
electricity	sustainability
sustainable	sustainability
environment	sustainability
bevs	sustainability
alternative_fuel	sustainability
clean	sustainability
cng	sustainability
gas	sustainability

renewable	sustainability	safety	reliability_uncertainty
variable	reliability_uncertainty	user	driver_acceptance
service	reliability_uncertainty	driver	driver_acceptance
risk	reliability_uncertainty	support	driver_acceptance
work	reliability_uncertainty	employee	driver_acceptance
problem	reliability_uncertainty	acceptance	driver_acceptance
likely	reliability_uncertainty	experience	driver_acceptance
issue	reliability_uncertainty	individual	driver_acceptance
maintenance	reliability_uncertainty	share	driver_acceptance
concern	reliability_uncertainty	consumer	driver_acceptance
availability	reliability_uncertainty	operator	driver_acceptance
lack	reliability_uncertainty	usage	driver_acceptance
access	reliability_uncertainty	personal	driver_acceptance
uncertainty	reliability_uncertainty	ease	driver_acceptance
limited	reliability_uncertainty	implementation	driver_acceptance
reliability	reliability_uncertainty		

A2

Literature Source Classification

The following pie charts are designed to show the diversity of sources for each corpus. Each slice represents a different journal or source type

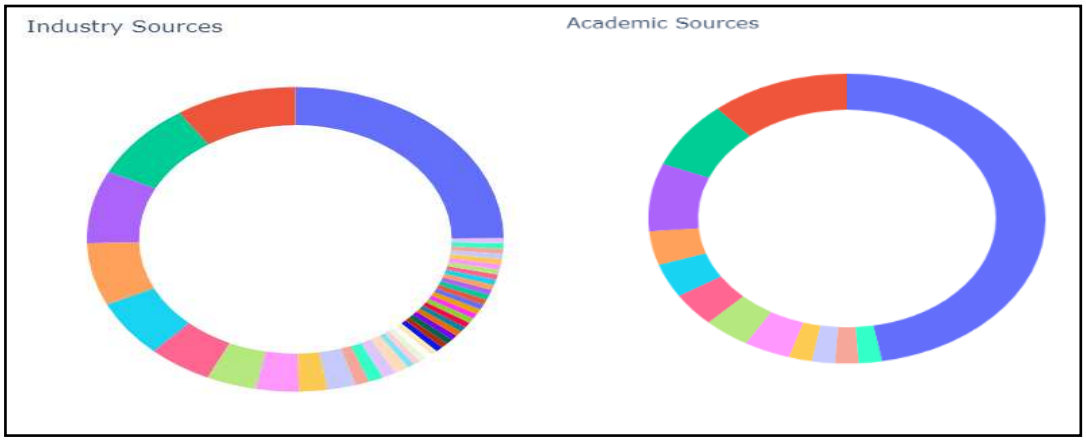

Figure 0-1: Pie charts showing diversity in text sources for each corpus.

A3

List of Industry Articles

Author	Tittle
Wendt	EV Adoption: Take Advantage of Governmental Regulations
Stanton	Flavor of the Decade: What if BEVs Aren't the Answer?
Wendt	Fleet Electrification: A Cost and Benefits Analysis
Shell	Pathways to Help Fleets with Efficiency and Sustainability
Work Truck	Study: Fleet Operators Less Optimistic About EVs, Telematics in Current Economy
Romjue	The Realities Intruding on Mass Electrification
Romjue	SEA Electric Puts Commercial Trucks Onto Cleaner Path
Romjue	Bollinger Motors Pursues Solid, Practical EV Lineup
Staff	Zeem Solutions Launches First EV Transportation-As-A-Service Depot
Romjue	Deep Dive on EV Data Shows Realities of Medium-Duty Fleet Electrification
Tucker	SMUD, Zeus, EAVX: A Triumphant Trifecta for Medium-Duty EVs
Staff	Xos Delivers First Vehicles to UniFirst Corp.
Brown	Designing Truck Bodies for the Coming Wave of Commercial EVs
Romjue	Electric Vehicles Adoption Will Take Longer Than Expected
Leiterman	Schneider previews Freightliner's eCascadia Class 8 electric truck - Company associates get a firsthand look at the future of their fleet
Staff	Xcel Energy is first in the nation to add all-electric bucket trucks to fleet - First truck to arrive in Minneapolis in late June
Ichien	Delivery First/Last-Mile Emissions Estimation Tool (FLEET) Vehicle Analysis
Facanha	GLOBAL ROADMAP FOR REACHING 100% ZERO-EMISSION MEDIUM- AND HEAVY-DUTY VEHICLES BY 2040
Welch	MOVING ZERO-EMISSION FREIGHT TOWARD COMMERCIALIZATION
Gurman	TAKING COMMERCIAL FLEET ELECTRIFICATION TO SCALE: FINANCING BARRIERS AND SOLUTIONS
Al-Alawi	Zeroing in on zero-emission trucks
Staff	Component Costs for Zero-Emission Mediumand Heavy-Duty Commercial Vehicles
Staff	The Beachhead Strategy: A Theory of Change for Medium- and Heavy-Duty Commercial Transportation

Arora	Electric School Buses Market Study
Norris	Best Practices on E-Bus and Grid Integration: A Guide for California Transit Fleets
Staff	Bollinger Motors Postpones Electric Pickup and SUV, Looks to Commercial EV Fleet Chassis
Staff	GM, Microvast to Develop Part for Safer, Stronger EV Batteries
Knosp	How To Prepare Electric Vehicles for Winter Months
Romjue	Preparing Your Organization for Fleet Electrification
Staff	Red Flags Emerge on the Way to Electric Vehicle Adoption
Romjue	Why There's No Better Time for Fleet Electrification
Penske	Penske Truck Leasing Continues to Expand Heavy-Duty Electric Vehicle Charging Network
Lewis	Heavy-duty trucks and buses must go electric, says new report
Chambers et al	Fleet Electrification Roadmap
Media Release	Element Fleet Management Issues Q3 Trends Report
Carp	Leading the Charge: 3 Ways to Succeed with Fleet Electrification
fnd	Electric Power Provider in California Aims to Accelerate EV Adoption
fnd	Geotab and Enterprise Fleet Management Study Quantifies Cost and CO2-Saving Opportunities for Fleet Electrification
fnd	GreenPower Hosts Educational Display of All-Electric School Buses and Commercial Vehicles at the U.S. Capitol
fnd	Project opens the door to a truly electrified freight future in which zero-tailpipe emission medium- and heavy-duty trucks are no longer limited to shortmileage
Fisher	EV opportunities are blooming in terminal yards
Jones	Fleets face 'very steep' learning curve on road to EVs
Staff	Lessor Fleet Advantage to order first 200 Class 8 EVs
Commendatore	Voltera execs talk fleet charging infrastructure challenges, opportunities
Roeth	Zero emission isn't only for big fleets
Staff	Florida county introduces electric refuse truck to be powered by city waste
Jones	Fleets 'frustrated' with utilities slow to support electric truck charging
Rohwling	Electric trucks, their tires and wheels
Roeth	Using the Inflation Reduction Act to scale EV adoption
Staff	ZF highlights electric vehicle technology ahead of IAA
Fisher	Small fleets turn to fleet-as-a-service provider to begin electric journey
Commendatore	ZF advances automated, electric technologies for commercial vehicles
Roeth	All medium-duty trucks are electrifiable, but not all will be easy

Staff	Walmart to buy 4,500 electric vehicles for last-mile delivery
Staff	FedEx receives first 150 BrightDrop electric delivery vehicles
Staff	Battle Motors accelerates EV expansion
Fisher	Futurist: Battery-electric trucks are a 'fantasy'
Fisher	More fleets are adding clean vehicles, report finds
Staff	McLane integrates Volvo EVs into Southern California operations
Keith	Kenworth's electric future
Wartgow	Leveraging BEV benefits on urban routes
Hitch	Nikola recalls all 93 Tre battery-electric trucks
Staff	Alabama city buys electric refuse truck
Gaskins	Helping fleets get to a greener future
Clark	The relationship between emissions reduction, technology, and regulations
Fisher	Southeast carrier gathering real-world Class 8 EV data
Fisher	Ruan adds another EV terminal tractor to its fleet
Commendatore	DTNA: Freightliner eCascadia marks shift in product strategy
Staff	Dana, EV maker Lion Electric announce pact for electrified systems
Fisher	Blue Bird enters crowded CV space with medium-duty EV chassis
Staff	Sysco to purchase up to 800 Freightliner eCascadias
Staff	Meritor to acquire Siemens Commercial Vehicles to offer more electric solutions
Staff	Salt River Project receives Kenworth K270E battery-electric truck
Staff	Volvo Trucks opens electric truck training facility
Staff	Mack opens new electric truck training location
Staff	Seattle's county government buys Class 8 BEV to electrify fleet
Staff	CARB showcases zero-emission vehicles
Clark	How to prepare for EV deployment
Roeth	Creating a culture of electrification
Staff	Volvo Trucks to build charging corridor in California for electric vehicles
Staff	WattEV invest $3M in grants for electric trucks
Keith	Industry reacts to proposed national charging station network
Staff	Environmental Defense Fund offers free resource for fleet electrification
Commendatore	NACFE sets sights on electric depots for next Run
Staff	Zeem to buy 100 Nikola Class 8 electrics
Staff	05:20 EST Volvo, DHL Freight partner on heavy duty electric truck project
Staff	Aemetis Enters Electric Vehicle Market Through Stake in Zero Emission Heavy Duty Truck Producer Nevo Motors

Staff	Electrify America and NFI Industries Collaborate on Nation's Largest Heavy-Duty Electric Truck Charging Infrastructure Project - Milestone project provides 34 ultra-fast DC chargers to support the deployment of 60 electric freight trucks that NFI will utilize to serve the Ports of Los Angeles and Long Beach
Staff	GNA Electric Truck & Bus Project Selected by New York Clean Transportation Prizes Program - The innovative project will deploy 38 electric medium- and heavy-duty vehicles and supporting infrastructure in underserved communities
Staff	Technology Innovation Driving the Transformation of the US Electric Commercial Vehicle Market - The medium-duty and heavy-duty truck and bus markets in the US are expected to show strong recovery following the major dip in 2020 due to the dire impacts of the COVID-19 crisis on production and sales.
Staff	Hub Group begins electric truck fleet pilot
Staff	New electric truck fleet to boost mine efficiency at Conuma Coal Resources Limited
Staff	Fleet Advantage Releases Latest White Paper Titled The Future of Heavy-Duty Trucks; Building a Bridge to Alternative Energies for U.S. Long Haul Transportation
Brown	Lordstown Begins Production of Electric Endurance
Clancy	What corporate fleet managers want out of EVs
Lockridge	Attention Shifts Away from Diesel Trucks
Cullen	Challenges of Charging Commercial Trucks
Staff	Fleet Advantage To Order 200 Electric Trucks
Staff	Global News Roundup: Transitioning to Electric Trucks
Staff	MHX Begins Fleet Electrification Project
Staff	North Carolina Adopts Advanced Clean Trucks Rule
Lockridge	The Road to Zero Emissions is Paved with More Than Batteries
Staff	Volvo Trucks Creates BEV Total Cost of Ownership Tool
Lockridge	What Trucking Fleets Need to Know about ESG
Staff	5 Commercial-BEV Questions for Xos
Staff	Global News Roundup: 15 Countries Pledge 100% Zero-Emissions Truck, Bus Sales by 2040
Staff	Xos Unveils Two New Commercial Electric Trucks
Stradling	With trash truck, Cary takes big step to all-electric fleet
Staff	State needs to kick-start funding for electric truck and bus fleet programs
Staff	Lightning eMotors to deploy electric truck, van fleet
The Fleet Podcast	The journey to electric: a deep dive into EV fleets
Darrell	California Voice - Kick-start funding for electric truck, bus fleet programs
Staff	Newmont, Caterpillar form alliance

Moynihan	Colorado Energy Office M/HD ZEV Strategy Colorado Electric Vehicle Coalition Meeting
Staff	A New, Advanced, and Highly Flexible Class 2-5 EV Platform is in the Making
Staff	Autocar, LLC Begins Field Testing All-Electric Cabover Refuse Trucks
Staff	Bollinger Motors Announces Our Next Energy as Battery Supplier for Electric Commercial Trucks
Staff	Challenges When Switching to Electric Vehicles
Staff	Discussing the Fleet Industry's Hottest Topics for 2022 and Beyond
Staff	Going Electric? Learn Tips From Lightning eMotors: Part 1
Staff	Going Electric? Learn Tips From Lightning eMotors: Part 2
Staff	Harbinger and Autel Announce Charging Partnership to Address the Needs of the North American Class 4 to 7 Fleets
Staff	Lightning eMotors Announces Transport Canada Registration, Expanding Commercial EV Presence in Canada
Staff	Lightning eMotors Unveils Next-generation Mobile Charger for Commercial and Consumer Electric Vehicles
Staff	Mack Trucks Introduces Electrify My Refuse Route Program to Incentivize Customers to Purchase the Mack LR Electric
Staff	Red Hook Terminals to Deploy Largest Electric Truck Fleet on US East Coast
Staff	SPARTAN MOTORS' UTILIMASTER BRAND UNVEILS TWO NEW ELECTRIC VEHICLE FLEET CONCEPTS AND LAST MILE DELIVERY GROCERY TRUCK BODY AT 2019 WORK TRUCK SHOW
Staff	The Shyft Group Secures Initial Pre-order for Blue Arc All-electric Delivery Walk-in Vans
Staff	Volvo Trucks Simplifies EV Charger Procurement with Vendor Direct Shipping Program
Staff	Electric Trucks Change the Way Chicago Collects Trash
Staff	MOTIV'S NEXT GENERATION EV TRUCK BATTERY EXCEEDS 150-MILE RANGE IN NEW FIELD TESTING
Staff	Power to the people
NACFE	Electric Trucks Have Arrived: Documenting A Real-World Trucking Demonstration
NACFE	Run On Less – Electric: The Use Case For Vans & Step Vans
NACFE	The Use Case For Heavy-Duty Regional Haul Tractors
NACFE	The Use Case For Medium-Duty Box Trucks
NewsBank	National Grid: N.Y. Clean Truck Conference Focuses on Electric Fleet Conversion

Staff	LION ELECTRIC AND TRANSIT TRUCK BODIES ANNOUNCE LAUNCH OF ELECTRIC VEHICLE FOR THE MEDIUM AND HEAVY-DUTY LAST MILE DELIVERY THAT WAS DEVELOPED UNDER UPFITTER PARTNERSHIP
Saxton	Making the Case for Electric Waste Transfer Vehicles
Staff	Terminal tractors can be ideal starting point for fleet electrification
Watson	The USPS is getting sued for sticking with gas-guzzling trucks
Horaczek	Amazon's futuristic electric delivery vans will hit the road in 2021 - The company plans to deploy 100,000 of the vehicles by 2030.
Staff	First-Of-Its-Kind All-Electric Utility Bucket Truck Makes Clean Energy Breakthrough - Iconic Symbol of Electric Industry Gets Zero-Emission Revamp, Helping to Drive Market for All-Electric Medium- and Heavy-Duty Trucks
Staff	Orange EV Continues to Set Records with Delivery of 500th Fully Electric, Heavy-Duty Truck
Staff	Ideanomics secures third-party financing to accelerate the growth of Solectrac's dealer network
Staff	Terminal Consolidation Leads Industry as First Fleet in State of Missouri to Deploy 100% Electric Class 8 Truck
Staff	Motiv Power Systems Opens New Commercial Electric Vehicle Deployment and Service Center in San Joaquin Valley
Staff	Spartan Motors' Utilimaster Brand Unveils Two New Electric Vehicle Fleet Concepts
Staff	Tropos Technologies Welcomes Electric Truck And Fleet Veteran Jeffrey Esfeld As Director Of National Fleet Sales & Business Development
Staff	Sonepar Selects Penske Truck Leasing to Provide New Electric Light-Duty Fleet of Ford E-Transits in USA
Staff	Orange EV Announces Electric Truck Fleet Milestones
Staff	Gardewine Adds Electric Vehicles to Fleet
Staff	Pride Group Commits to Large eCascadia and eM2 Order
Clare	U.S. Electric Truck Market Research Report
NewsBank	Sierra Club: Postal Service Commits to At Least 40% Electric Truck Fleet in Response to Public Pressure, More Needed
NewsBank	EVgo And MHX, LLC Partner To Power Class 8 Electric Truck Fleet
Staff	AEP to Switch 100% of Cars and Light-Duty Truck Fleet to Electric Vehicles By 2030
Romjue	Hard Facts on Electric Vehicles Favor Fleet Usage
Mehta	Can the US make the switch to electric heavy-duty trucks?
Tucker	A Roadmap for Work Truck Fleet Electrification

Staff	Bollinger B4: A Medium-Duty Electric Truck
Prejean	Bridging the Gap: Electric Trucks
Romjue	Charging Strategies Crucial to Deploying Electric Fleet Vehicles
Fleet Forward Staff	Electric Truck Performance in the Real World
Prejean	Exploring Electrification, Small Vans & Big Wins
Tucker	First Work Truck Advisory Board Meeting Tackles EVs, Management, and More
Fleet Forward Staff	Fleet Electrification: The Straight Dope
Staff	GM Energy Helps Electrify Commercial Fleets
Kaufield	Key Moves to Increase Fleet Sustainability Today
Romjue	Last-Mile Delivery Market Primed for Electrification
Simon	Managing the Transition to an EV Fleet
Fletcher	Past Trends Shape the Future of Work Truck Fleets
Fisher	The Portable, Responsible Approach to Charging EV Truck Fleets
Knight	The Top Costs in Work Truck Fleets
Staff	Don’t Let Medium-Duty EVs Catch Your Fleet by Surprise
Staff	Penske, Cardinal Health Partner to Test Commercial EVs
Fletcher	Issues Created by Driver Perception
Bleizeffer	Wyoming prepares to ‘electrify’ roadways
Staff	Xos Signs Endorsement of Global MOU with CALSTART Advocating for Acceleration of Zero-Emission Medium- and Heavy-Duty Vehicle
Gonzalez et al	Medium- and HeavyDuty Electrification: Weighing the Opportunities and Barriers to ZeroEmission Fleets
Staff	Pegasus To Purchase 25 Zeus All-Electric Cutaway Z-19 Chassis For Type A Electric Bus Market

A4

Link to GitHub Repository

Clone Link: https://github.com/fouren714/MDHD_LDA.git

Online Link: https://github.com/fouren714/MDHD_LDA

List of Abbreviations

Abbreviation	Definition
ABM	Agent-Based Model
BERT	Bi-Directional Encoder Representations of Transformers
BEV	Battery-Electric Vehicle
c-TF-IDF	Class-based Term Frequency Inverse Document Frequency
DOE	Department of Energy
eMDHD	Electric Medium-Duty and Heavy-Duty Vehicles
GPT	Generative pre-trained transformer
HDBSCAN	Hierarchical density-based spatial clustering of applications with noise
LDA	latent Dirichlet Allocation
LLM	Large Language Model
MDHD	Medium-duty/Heavy-duty Vehicles
NLP	Natural Language Processing
NLTK	Natural Language Processing Tool Kit
NMF	Non-Negative Matrix Factorization
OEM	Original Equipment Manufacturer
START	Score of Topic and Related Terms
UMAP	Uniform Manifold Approximation and Projection
VTO	Vehicle Technologies Office
ZEV	Zero-emission Vehicle