Independent Study Project (ISP) Collection   SIT Study Abroad

Fall 2023

Environmental Loss, Displacement, and Anxiety in Portugal: Analyzing News Articles to Differentiate Manifestations of Environmental Distress

Robin Greene
SIT Study Abroad

Recommended Citation

Environmental Loss, Displacement, and Anxiety in Portugal:
Analyzing News Articles to Differentiate Manifestations of Environmental Distress

Robin Greene

SIT Portugal: Sustainability and Environmental Justice
Fall 2023

Table of Contents

Abstract

Expressions of environmental distress can take different forms, with different symptoms, causes, and treatments. Existing literature generally identifies three primary categories of environmental distress responses: environmental grief, solastalgia, and eco-anxiety. This paper explores these distinctions and identifies words typically associated with each category in Portuguese news articles using Latent Dirichlet Allocation (LDA). LDA is a Natural Language Processing (NLP) technique that groups words into topics and identifies connections between those words based on how often those words appear together in sequences. This paper uses a modified LDA algorithm called GuidedLDA to identify additional keywords within topics defined by a list of "seed" keywords. The coherence scores for two article sets analyzed with these categorizations were -0.87 and -1.05, indicating that these categorizations of environmental distress are clearly delineated in media. Trends identified in the LDA analysis matched findings from qualitative studies, such as the persistence of environmental grief as time-independent and the strong relationship between place attachment and solastalgia. These findings provide quantitative support for previous qualitative metrics and offer a framework for further examinations of environmental sentiments in news media.

Introduction

Climate change, natural disasters, extinctions, and other environmental events can produce a wide range of emotional responses, including distress. The term "environmental distress" is generally understood to mean negative emotional responses to environmental factors and stimuli, but this term is not widely used in environmental or psychological literature (Kim et al, 2020). This paper will examine specific manifestations of environmental distress which are better-researched within the field, using "environmental distress" as an overarching term which encompasses environmental grief, solastalgia and environmental displacement, and eco-anxiety. Some authors describe these distinctions as a continuum, with interrelated expressions between environmental grief, solastalgia, and eco-anxiety (Marchand et al, 2023, p. 132-133). Given this understanding, it is important to note that "environmental distress" may capture emotional responses which fall into multiple or none of those categories.

The aim of this paper is not to examine environmental distress in and of itself, but rather to classify different manifestations of environmental distress and to identify how both the causes and experiences of those emotions differ. This paper identifies three categorizations of environmental distress: environmental grief, which involves a sense of loss and mourning for past environmental states; solastalgia, which represents the sense of displacement and alienation that people feel as ways of life are changed by environmental factors; and eco-anxiety, which describes a sense of negative anticipation for future environmental events or changes. Researchers have differentiated these concepts by noting that eco-anxiety is specifically future-oriented, and solastalgia, which is sometimes characterized as a sub-type of environmental grief, is associated with landscape change (Comtesse et al, 2021). This paper will further distinguish between environmental grief and solastalgia by noting the past-facing nature of

environmental grief and characterizing solastalgia as firmly rooted in the present, consistent with the original definition of solastalgia as tied to "the present state of one's home environment" due to ongoing threats (Albrecht, 2006). Each of these categorizations of environmental distress will be examined in further detail in each subsection below.

Environmental Grief and Loss

As disappearing landscapes induce senses of loss in people, psychologists have identified a sense of environmental grief in response to ecological and climate changes (Cunsolo & Ellis, 2018). Grief and related emotions tend to occur as people witness the disappearance of species, places, and ecosystems with emotional significance (Ojala et al, 2021). Experiential approaches to climate research and activism have incorporated environmental grief as a part of the process, noting that the lived experience of climate change influences responses from both individuals and governments, and so understanding the pathology of environmental grief can help develop pragmatic policy and effective activism (Willox, 2012; Holthaus, 2022).

Environmental grief is experienced differently by various groups of people, as communities differ in their emotional connections to surrounding ecosystems and environmental spaces. Whereas Western perspectives tend to consider the environment as an abstracted "thing" or inanimate surroundings, many Indigenous worldviews conceptualize environmental objects such as mountains, rivers, and forests as living entities with kinship ties to human communities, resulting in a more intimate sense of grief, similar to what one might feel for the loss of a relative (Ritchie & Phillips, 2021; Galway et al, 2022). Indigenous experiences of environmental grief are exacerbated by cultures and traditions which revolve around ecosystems and species that are being lost (Willox, 2012). Older adults, while often ignored in explorations of environmental psychology, experience an acute form of environmental grief due to the firsthand experience of environmental losses (Dennis & Stock, 2023). Environmental grief may co-occur and coincide with other forms of grief, including racial grief (Lennon, 2020). Given the diverse experiences of environmental grief, it is important to understand environmental grief as not a universally identical experience but rather a descriptor of similar emotional phenomena that occur in response to environmental loss.

Experiences of environmental grief are felt at both an individual and communal scale, as places, identities, and human-nature connections are impacted by environmental loss and processes of change (Engstrom, 2019). Feelings of environmental grief are likely to arise and spread through social and communal groups, particularly as activists aim to inspire emotional solidarity with threatened ecosystems as a pathway to action (Holthaus, 2022). Environmental grief can therefore be understood as a reaction to both perceived and persistent loss as well as a response to gained awareness of that loss. The diverse set of mourning experiences described by environmental grief are often not recognized as a form of grief due to the devaluation of ecosystems and nonhuman lives (Kevorkian, 2019, p. 217). Regardless, environmental grief is a clear manifestation of environmental distress, encompassed by mourning ecosystems, species, and the human connections to those ecosystems and species.

Solastalgia and Environmental Displacement

As originally constructed, solastalgia describes a form of environmental distress caused by the disappearance or transformation of one's home by environmental forces such as land degradation, climate change, and epidemics (Albrecht, 2006). While solastalgia is interconnected and often co-occurring with environmental grief, it represents an ongoing form of distress caused by the lived experience of physical, cultural, or social displacement due to environmental factors and the consequent loss of identity or culture associated with dispossession, rather than a mourning for a specific thing which is no longer present (Albrecht, 2020; Cunsolo & Ellis, 2018). The concept of solastalgia is intimately tied to places and landscapes, and therefore expressions of solastalgia are closely linked to the places they occur (Galway et al, 2019). Solastalgia captures a human psychological aspect to ecosystem change which does not fit neatly within the retrospective framework of environmental grief or the anticipatory definition of eco-anxiety, conceptualizing a form of environmental distress caused in response to ongoing processes of environmental harm.

Like environmental grief, experiences of solastalgia differ across social communities. Albrecht notes the unique experiences of Indigenous groups with regards to solastalgia, particularly as the environmental dispossession of Indigenous communities has occurred for centuries longer than the contemporary account of solastalgia in the face of climate change (Albrecht, 2006). Environmental frontline communities, particularly those with high levels of place attachment, are likely to experience acute manifestations of solastalgia, including redefinitions of the human-place and place-community relationship as an adaptation to environmental changes (Phillips & Murphy, 2021). Of particular relevance to the Portuguese context is the relationship between overtourism and solastalgia as urban landscapes are transformed by the process of touristification (Lalicic, 2020). As a place-related form of environmental distress, the nature of solastalgia in a given location is dependent on the relation between a place and its inhabitants, as well as the nature of environmental change occurring there.

The place-based and ongoing nature of solastalgia make it difficult to characterize in the context of other psychological phenomena. Instead, many scholars consider solastalgia to be the psychological aspect of the process of ecosystem change (Albrecht, 2020). Yet like environmental grief and eco-anxiety, solastalgia can occur in response to both short-term and long-term environmental events. For short-term processes of environmental change, such as wildfires, the magnitude of distress is independent of past occurrences or expected future risk, indicating that solastalgia is a primary mediator of psychological responses to these phenomena (Leviston et al, 2023). While solastalgia is relatively underrepresented in the psychology literature compared to other manifestations of environmental distress, its occurrence as a present-facing and ongoing emotional response, particularly in response to short-term natural processes like wildfires, may make it particularly salient in reports of environmental-related news.

Eco-Anxiety

Eco-anxiety is perhaps the best-known manifestation of environmental distress, reflecting fear and negative anticipation surrounding future environmental events (Kurth & Pihkala, 2022). Eco-anxiety is typically framed as an indirect, existential fear response to environmental stress, with parallels to the existential aspects of solastalgia and environmental grief (Pihkala, 2018; Soutar & Wand, 2022). Like environmental grief, eco-anxiety is considered to be an emotional distress response to guilt and loss caused by climate change and may be a psychological motivator of environmental actions (Ojala, 2018; Ágoston et al, 2022). The connections to grief and guilt driven by anticipated disruptions to human-environment connections are an important aspect of eco-anxiety, differentiating it from broader climate anxiety (Pihkala, 2021). A common form of eco-anxiety is climate anxiety, which describes anxiety associated with anthropogenic changes to the global climate (Clayton, 2020). Eco-anxiety covers a wide range of timescales, occurring in response to expected or potential threats which may occur in the immediate or distant future.

Though eco-anxiety is considered to encapsulate a common set of psycho-physical responses, eco-anxiety may be felt differently across different groupings of people. Social analyses of eco-anxiety often emphasize the disproportionate burden of eco-anxiety on young people, who experience the highest eco-anxiety levels of any age group (Usher, 2022). Cultural and socioeconomic factors also have a strong influence on an individual's resilience and susceptibility to eco-anxiety, as connections to nature, strong social links, and the absence of additional trauma factors are likely to reduce maladaptive eco-anxiety responses (Crandon et al, 2022). Eco-anxiety is primarily discussed from a Western and affluent perspective, and more research is needed to examine the specific forms of eco-anxiety that present in vulnerable groups, including Indigenous people and women (Coffey et al, 2021). Eco-anxiety is a widespread phenomenon, as substantial environmental changes are predicted to occur around the world, and populations with high potential climate risk are likely to feel especially severe eco-anxiety.

Due to the highly existential nature of eco-anxiety and the justifiable worry that people, especially young people, fear for their future in a world with a rapidly changing climate, it is helpful to isolate the maladaptive aspects of eco-anxiety from other parameters associated with anxiety induced by climate change. Though eco-anxiety may produce anger responses which inspire effective climate action, it is considered to have a negative overall effect on human wellbeing, with potential impacts on functioning in everyday activities and a reduction in the effectiveness of climate action due to "eco-paralysis" (Stanley et al, 2021; Clayton, 2020). Eco-anxiety contains emotional, psychological, existential, and pathological responses, with strong links between experiences of environmental loss and solastalgia and expressions of eco-anxiety (Pihkala, 2020). The cognitive-emotional aspects of eco-anxiety, such as nervousness, lack of concentration, and inability to perform daily tasks, have been identified as the most potent mechanisms affecting expression of eco-anxiety (Heeren et al, 2023).

Eco-anxiety represents a common form of environmental distress, and its expression is likely to be common in response to predictions of future change or potential future human actions which may cause environmental harm.

Research Question and Objectives

       This paper aims to identify the causes and manifestations of environmental grief, solastalgia, and eco-anxiety in Portugal. In particular, it will ask the question of how different factors and causes produce unique manifestations of environmental distress in Portugal. To answer this question, a natural language processing method called GuidedLDA will be performed on a sample of roughly 2000 news articles from prominent Portuguese news publications, as described in the following section.

Methodology

Obtaining Article Text

       To identify expressions of environmental distress in Portugal, nearly 1000 articles were analyzed from each of two prominent Portuguese news publications: Diario de Noticias (www.dn.pt) and Expresso (www.expresso.pt). Articles on these sites are accessible online and present articles in a convenient text format with no paywall. A third newspaper, Jornal de Noticias, was originally identified as well, but the news article text was coded in JavaScript rather than HTML, which made it difficult to access through an automated program. Analysis of online newspaper articles has recently gained prominence as a useful technique to understand particular sentiments relating to environmental psychology in Portugal (Batel & Pataco, 2020). Collecting articles from multiple sources with editorial independence and limited political bias allows for a sample with balanced perspectives and the ability to explore the range of reporting across sources (Boager & Castro, 2021).

       First, sample article text was collected using a scraper program (Appendix A). This program identified the 1000 most-recent articles tagged "*ambiente*" (environment) on the web pages for each news site. *Ambiente* was chosen as the search term over other potential filters, such as *alterações climáticas* (climate change) as it covers a broader set of issues, so potential causes of environmental distress such as lithium mines or overfishing may have been tagged under *ambiente* but not *alterações climáticas*. To collect article links, the scraper program collected the href objects (hyperlinks) in the HTML code on each site's "results" page for the term *ambiente*, and an automated browser loaded additional articles by either clicking a button or scrolling to the bottom of the page, depending on how each site loaded articles. Once 1000 article links were obtained, the program stopped loading articles and wrote these 1000 links to a csv (comma-separated value) file to obtain the text from each article. A csv file was used as the storage format because it is an easily-accessible, data-efficient standard with spreadsheet-like functionality.

       After loading article links into a csv file, the scraper program sequentially opened each link and attempted to combine the text from each p object (paragraph) with a tag or object class

indicating that the text was a part of the article, rather than text from site navigation, ads, or other extraneous text. This text was then added to a second column of the csv file on the row corresponding to the article link. As some articles, like features or multimedia posts, contained different formatting, this code was modified to account for alternative formats and ensure that all articles had corresponding text. Any articles with minimal or no text, such as video or audio posts, were manually deleted. As a result, the number of articles analyzed per publication was slightly below 1000: the sample included 979 articles from Diario de Noticias and 995 articles from Expresso.

Articles were collected from Diario de Noticias on 10 November 2023 and from Expresso on 11 November 2023. These dates were shortly after the warrants and resignations associated with Operation Influencer, and therefore the sample of articles may focus disproportionately on issues related to Operation Influencer, such as government corruption, lithium mining, and green hydrogen. However, the samples were large enough that articles related to Operation Influencer only represent a small portion of total articles – the revelations of Operation Influencer were reported on 7 November 2023, while the Diario de Noticias sample includes articles since 19 December 2018, and the Expresso sample includes articles since 5 June 2023. Therefore, unavoidable sampling bias due to the time frame of data collection is not expected to have a large impact on the results of the study.

Text Cleaning

After article text was obtained and saved in a csv file, text was prepared or "cleaned" for easier processing. First, all letters were converted to lowercase so that capitalized words were not considered distinct from their uncapitalized forms. After this was done, line breaks, punctuation, and extra spaces were removed, so that all words were separated only by a single space. Some characters, such as hyphens and forward slashes, were replaced by spaces to separate words.

Next, the spaCy lemmatization model, which is optimized for news articles in European Portuguese, was used to reduce all words to their unconjugated form. This was done to ensure that alternative forms of the same word, such as *menino* and *meninas* or *corriu* and *corremos*, were considered identical in the analysis. Lemmatization was chosen over other NLP techniques to obtain root words, like stemming, as lemmatization is best able to identify alternate forms of words which do not follow standard conjugation rules.

In the final cleaning step, "stopwords", or common words that provide minimal information, such as "de", "ser", and other words with a primarily grammatical function, were removed. The Natural Language Toolkit, a Python library, contains a list of stopwords in many languages, including Portuguese. These stopwords were removed from the article texts, along with a list of additional stopwords recommended by other Portuguese-language LDA modelers (Pires, 2017). After all stopwords were removed, the cleaned text was saved to a csv file for analysis.

An example of the text cleaning process is shown below in Figure 1. Sample text is taken from the topline of an article published in Diario de Noticias on 1 November, 2023.

Figure 1. An example of the text cleaning process.

Original text:

As condições climatéricas permitiram, na última semana, que o consumo de energia em todo o país fosse assegurado exclusivamente por fontes renováveis. Os ambientalistas dividem-se entre aplaudir o feito e recear pelo impacto na biodiversidade.

Remove capitals and punctua-tion

as condições climatéricas permitiram na última semana que o consumo de energia em todo o país fosse assegurado exclusivamente por fontes renov áveis os ambientalistas dividem–se entre aplaudir o feito e recear pelo impacto na biodiversidade

Lemma-tization

o condição climatérico permitir em o último semana que o consumo de energia em todo o país ser assegurar exclusivamente por fonte renová vel o ambientalista divider se entre aplaudir o fazer e recear por o impacto em o biodiversidade

Remove stopwords

Final text:

condição climatérico permitir último semana consumo energia todo país assegurar exclusivamente fonte renovável ambientalista divider aplaudir fazer recear impacto biodiversidade

Sample text from Luz, 2023.

Identifying Keywords

In a Guided LDA analysis, a program identifies which words are most often associated with a list (or several lists) of pre-identified keywords. This method allows for identification of relevant words or concepts when keywords associated with specific concepts have already been identified (Wang et al, 2012). Of course, before the Guided LDA can be completed, it is necessary to compile a list of keywords for each topic to be considered by the program. As most existing literature on environmental distress is in English, keyword lists in Portuguese were necessary to examine Portuguese-language media. One survey in Portuguese was identified for terms related to eco-anxiety (Sampaio et al, 2023), but no Portuguese-language surveys or analyses on environmental grief or solastalgia were identified.

Tables 1-3 provide survey questions and categorizations used by previous studies to quantify environmental grief, solastalgia, and eco-anxiety, respectively, alongside Portuguese keywords derived from those questions or categorizations. Only the "base" form of each word is listed in the tables, as the lemmatization step groups together different forms of the same word. For example, *perdido* is a masculine singular adjective, so *perdida* and *perdidos* would also be considered as *perdido*. Some potential keywords were omitted if they have multiple meanings. For example, "límite" was not included as a keyword for eco-anxiety because of potential confusion with emission limits. The LDA algorithm counts hyphens as spaces and pronouns as "stop-words" which are ignored, so reflexive verbs like *preocupar-se* are only counted as the main verb stem (*preocupar*). Keywords repeated in the same table are placed in brackets. Any potential keywords identified in multiple manifestations of environmental distress are marked with strikethrough.

Table 1. Keywords associated with environmental grief.

Pathway of Ecological Grief	Portuguese Keywords
Grief associated with physical ecological losses and attendant ways of life and culture	~~perder~~, ~~faltar~~, ausente
Grief associated with disruptions to environmental knowledge systems and resulting feelings of loss of identity	identidade, tradição, conhecimento, ~~pertencer~~
Grief associated with anticipated future losses of place, land, species, and culture	lugar, cultura, extinto, morto, ~~desaparecer~~

Adapted from Cunsolo & Ellis, 2018.

Table 2. Keywords associated with solastalgia.

English Survey Question	Portuguese Keywords
Sad when look at degraded landscapes and mine voids	triste, degradado
Farming lifestyle depending on good land and water is threatened by change	ameaçar, recursos, ~~trabalhar~~
Worried that valued aspects of place—clean air and water, scenery—are being lost	[recursos], paisagem, ~~perder~~, ~~faltar~~
Unique aspects of nature in this place are being lost	natureza, ~~perder~~, ~~faltar~~
Miss peace and quiet once enjoyed in this place	solidão

Sad that familiar animals and plants are disappearing	~~desaparecer~~
Ashamed of the way this area looks now	vergonha
Thought of my family being forced to leave this place upsets me	sair
Sense of belonging undermined by change	~~pertencer~~

Adapted from Higginbotham et al, 2006.

Table 3. Keywords associated with eco-anxiety.

English Survey Phrasing	Portuguese Survey Phrasing	Portuguese Keyword(s)
Feeling nervous, anxious or on edge	Sentir-se nervoso(a), ansioso(a) ou no limite	nervoso, ansioso
Not being able to stop or control worrying	Não ser capaz de parar ou controlar a preocupação	preocupar
Worrying too much	Preocupar-se em demasia	[preocupar]
Feeling afraid	Sentir-se com medo	medo
Unable to stop thinking about future climate change and other global environmental problems	Ser incapaz de parar de pensar sobre alterações climáticas futuras e outros problemas ambientais globais	refletir
Difficulty sleeping	Ter dificuldade em dormir	dormir
Difficulty enjoying social situations with family and friends	Ter dificuldade em desfrutar de eventos sociais com a família e os amigos	descansar
Difficulty working and/or studying	Ter dificuldade em trabalhar e/ou estudar	foco, ~~trabalhar~~

Adapted from Sampaio et al, 2023.

Table 4 provides a synthesis of Tables 1-3, noting all unique keywords in each category. Nine or ten (9-10) unique keywords were identified for each category. Previous research indicates that the minimum number of labels sufficient to identify a distinct topic in a Guided LDA is between 4 and 8 (Weston et al, 2023). As the number of unique labels for each topic is higher than this range, this keyword list should be suitable for a Guided LDA analysis.

Table 4. Unique keywords to identify manifestations of environmental distress.

Environmental Grief	Solastalgia	Eco-anxiety
ausente	triste	nervoso
identidade	degradado	ansioso
tradição	ameaçar	preocupar
conhecimento	recursos	medo
lugar	paisagem	refletir
cultura	natureza	dormir
extinto	solidão	descansar
morto	vergonha	foco
	sair	

Sentiment Analysis

The sentiment analysis was accomplished using the GuidedLDA model, which establishes topics based on user-set seeds and then identifies the words most closely associated with those seeds and each other. The GuidedLDA Python library, which runs a modified version of the Gensim LDA model, was used to accomplish this task. The guided LDA was run twice, once for the article texts from Diario de Noticias and then for the article texts from Expresso.

The GuidedLDA model works by computing bigrams, which are all sequences of two words within a piece of text. For example, in the sentence "quick brown fox jumped over sleeping dog", the bigrams would be quick-brown, brown-fox, fox-jumped, jumped-over, over-sleeping, and sleeping-dog. From there, topics of ten words are created using an algorithm that maximizes the frequency of bigrams for each pair of two words in the topic, including seed words.

After the GuidedLDA was performed, a coherence score was generated using the U_{mass} metric, which indicates the relative frequency that words within the topic appear together. As U_{mass} is a logarithmic metric, scores range from negative infinity to zero, with a higher score indicating a higher degree of co-occurrence between topic words. The LDA results were visualized using the matplotlib package in Python to create bar graphs showing word weight for each topic word identified (Figure 1 and Figure 2).

Validation of Results

Data were reviewed at each stage of the research process to ensure that data collection and analysis processes are accurate. Given the large volume of data used in this sample, it was not feasible to manually check all results, so a sub-sample of ten articles were used to manually verify results. These ten articles were selected using a random number generator to generate two lists of five numbers between 1 and 1000, and then selecting the corresponding Diario de Noticias articles from the first list of numbers and the corresponding Expresso articles from the second list of numbers.

After obtaining article text, results were validated by ensuring that the article text in the csv file matched the text within the article at the corresponding link for all 10 randomly selected articles. Though all article text was confirmed to be accurate, some articles from Diario de Noticias contained extraneous text from site pop-ups. However, since this text did not include any of the keywords in Table 4, it would not be flagged as part of any of the environmental distress categories, and therefore this irregularity was left in and assumed to not impact the final results.

Following the text cleaning step, results were validated by comparing pre-cleaned text with cleaned text from 10 randomly selected articles. Cleaning appeared to have occurred correctly, though several words which should have been selected as stopwords were not removed. For example, *ha* was not removed as a misspelling of *há*, some English quotes or proper nouns used the English stopword "the", and some other Portuguese terms such as "*até*" or abbreviations of stopwords were not removed. These words were added to a secondary list and removed from article text.

Due to the complexity of the Guided LDA algorithm, it was not possible to verify results directly. The Guided LDA model is intended to process large amounts of text, so running the model on a small sample of articles would produce inaccurate results. Instead, the coherence score was used as a verification metric, as a high coherence score indicates that topic words are often found near seed words. Another form of validation occurred through the inclusion of two separate samples, as the results from Diario de Noticias were compared with the results from Expresso to ensure that the observed trends in keywords were a result of genuine associations with keywords rather than data artifacts based on the reporting styles of each publication.

Ethics

As this study did not directly involve human participants, no ethical concerns related to human subjects and data collection procedures were considered. The work of journalists, particularly journalists who work for national newspapers, is published for public consumption and is therefore expected to be analyzed in a variety of contexts, including scholarly and computer-driven analysis. The use of a text-scraping program is not directly against the Terms of Use of the Diario de Noticias or Expresso sites, though excessive scraping may overload the server, resulting in a denial-of-service, which is against both sites' Terms of Use. To mitigate

this, a delay was implemented between article access attempts to minimize strain on site resources, and since the text scraping was for a legitimate use and any denial-of-service would be unintentional, the author did not interpret this use as a violation of the sites' Terms of Use.

The use of a computer program to analyze sentiments in a text dataset also raises some ethical concerns worth noting. First, the results of this study may be construed as a qualitative interpretation of emotions involved in the reporting of environmental news. While this study does aim to identify different sentiments related to environmental events, it is important to note that the LDA model is only capable of quantitative, not qualitative modeling. Essentially, the program identifies correlations between certain words, and this study interprets those correlations as reflecting different sentiments. This interpretation may or may not be valid, and a comparatively simplistic model cannot capture the nuances of emotions and textual connotations in the way that a human mind can. However, qualitative human interpretations of emotions are also subject to bias and inaccuracy, and this paper hopes to provide an alternative method of interpretation to supplement existing human-based characterizations of sentiments related to environmental news.

Additionally, generative AI (ChatGPT) was used in the development of code for this project. While many code edits were the sole work of the author, generative AI wrote the outline of most project code and provided guidance to debug code and to understand some of the process behind the Guided LDA model. The use of generative AI raises several ethical concerns, including inaccuracy and copyright infringement. As generative AI was primarily used to write code, inaccuracy was not a major concern, as any issues would cause an error which would be readily noticed. If AI-generated code was inaccurate but did not produce an error, any inaccurate results would likely be noticed in the validation processes noticed above. Generative AI information on the Guided LDA model was only used to gain a basic level of familiarity with the model, and any information included in this article was obtained from scholarly sources. Several commentators have claimed that AI training datasets infringe on copyright, and litigation on this issue is currently underway in the United States (Zirpoli, 2023). However, due to the difficulty of obtaining proprietary code, code in generative AI datasets is likely open-source or public domain, especially since all code libraries used in this project, such as BeautifulSoup (for web scraping) and Gensim (for LDA analysis), are open-source.

Results and Discussion

Results

The results of the Guided LDA for the Diario de Noticias articles and the Expresso articles are shown below in Figure 1 and Figure 2. The coherence scores for both article sets were relatively high: -0.87 for Diario de Noticias and -1.05 for Expresso. Typical coherence scores for leading LDA models are above -1.1, as both of these scores are, indicating that the three topics have relatively well-defined associations with other sets of words in the texts (Oyshi, 2023, p. 37).

Figure 2. Guided LDA results for Diario de Noticias articles.
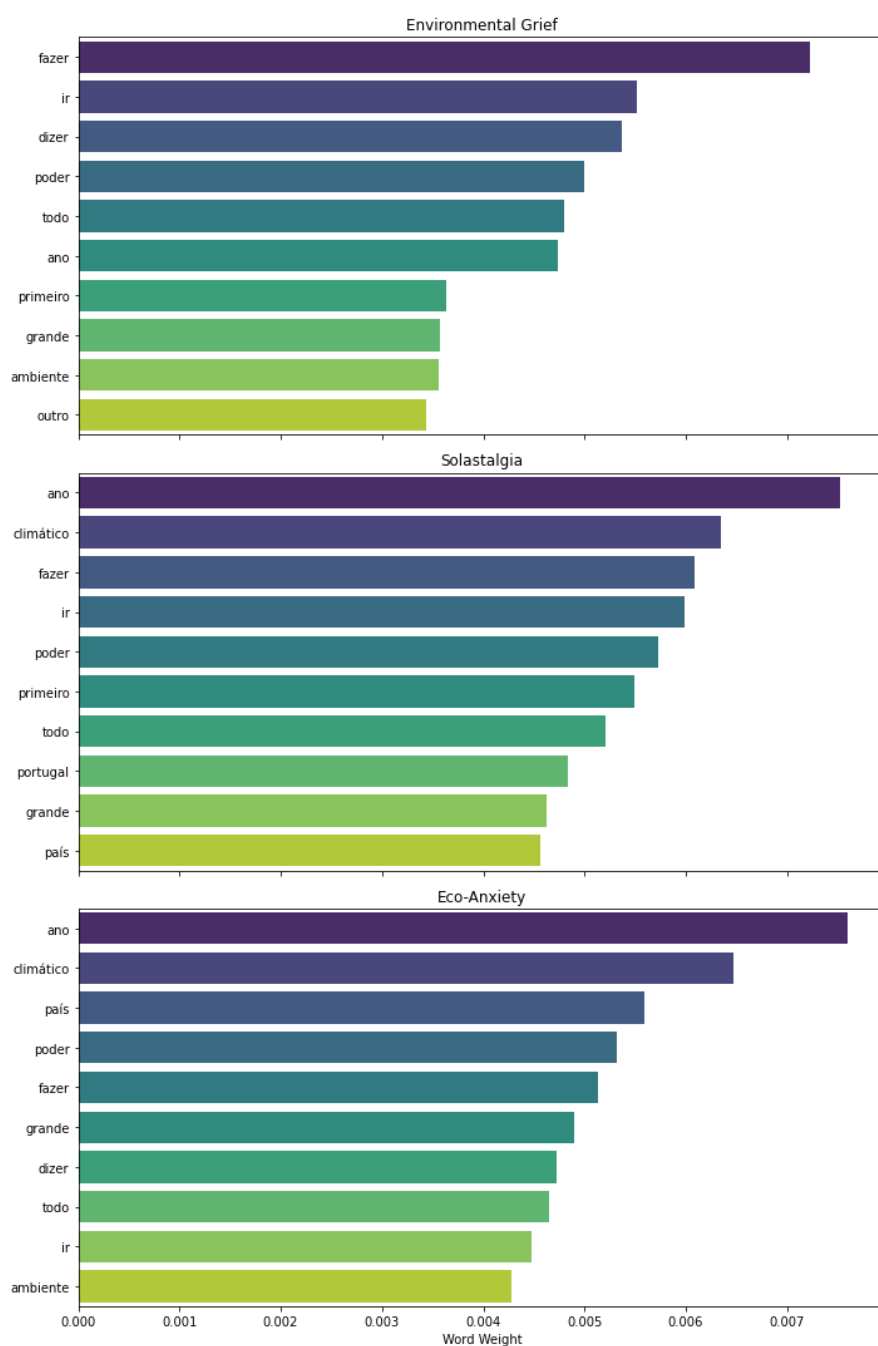
Coherence Score: -0.87

Figure 3. Guided LDA Results for Expresso articles.
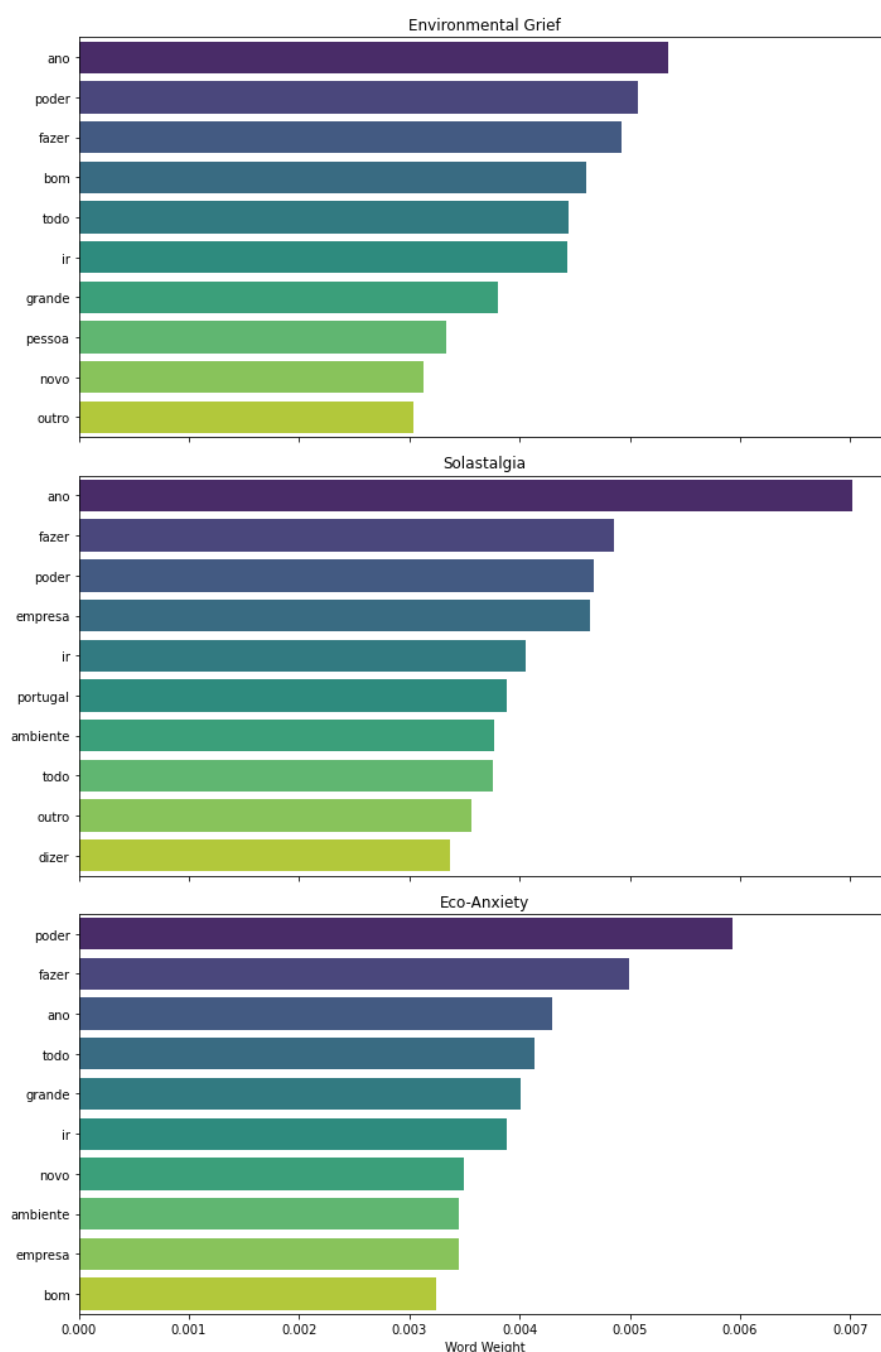
Coherence Score: -1.05

The individual words identified under each topic in the LDA analysis do not offer much insight into specific causes or manifestations of environmental distress in Portugal, as the topic words are rather broad. Auxiliary verbs such as *poder* and *fazer* ranked consistently high across all categories. While one might expect these auxiliary verbs to appear consistently in many topics due to their frequent usage, the coherence score for each article set remained high,

indicating that these auxiliary verbs were disproportionately used with keywords identifying environmental distress. The U_{mass} coherence score is calculated as the logarithm of the fraction of word occurrences that are in a bigram with the target word: $U_{mass} = \log \frac{B_d + 1}{d}$, where B_d is the number of bigrams between a topic word and a given keyword, and d is the number of times that keyword appears overall. Therefore, for common words such as auxiliary verbs, a high U_{mass} score, as observed in these results, indicates a high rate of co-appearance with keywords.

Analysis of Results

One potential explanation for the frequent observation of auxiliary verbs near environmental distress keywords is that auxiliary verbs may indicate a time component, as conjugations of *poder* and *ir* are often used to indicate future occurrences, and conjugations of *fazer* are common when discussing past events. This interpretation is supported by other common words in the topic lists with time components, as *ano* (year) was the highest-ranked word in five of six topic lists, and other time-related words such as *novo* and *primeiro* appeared in some topic lists. Some LDA users suggest eliminating these auxiliary verbs as stopwords to improve the usefulness of results (Fatima-Zahrae et al, 2021). When the analysis was re-run with the words *poder*, *ir*, and *fazer* removed as stopwords, the additional topic words identified included other time-related words such as *semana* and *século*, adding further evidence to support this hypothesis.

Several interesting trends also occur across different manifestations of environmental distress. Auxiliary verbs and the word *ano* are less common for keywords associated with environmental grief than for other keywords. This may indicate that environmental grief is less time-dependent than solastalgia or eco-anxiety. Some authors have suggested that environmental grief is a cultivated response with generalized anticipatory or backwards-looking components, which could result in fewer associations with time compared to other forms of environmental grief (Holthaus, 2022). This finding also supports broader psychology literature which finds that complicated forms of grief, including environmental grief, tend to be persistent due to the strong emotional attachments involved (Mancini & Bonanno, 2012). Based on the relatively low prevalence of words indicating a time dimension in the environmental grief topic list, Portuguese news media appears to reflect the persistent nature of environmental grief identified in the literature of environmental psychology.

The word *grande* was more commonly associated with environmental grief and eco-anxiety than with solastalgia, which indicates that environmental grief and eco-anxiety may have a higher intensity of experience. Indeed, researchers have found that while environmental grief and eco-anxiety regularly present as pathologies that produce significant disruption to patients' livelihoods, acute manifestations of solastalgia at the same level of severity are only largely observed in Indigenous communities (Ojala et al, 2021). Interestingly, *portugal* and *país* occurred only in both solastalgia topic lists, which fits with previous research identifying solastalgia as a highly place-based form of environmental distress (Galway et al, 2019). While these data do not identify causes associated with environmental distress, they indicate the

existence of a Portugal-specific place attachment that is evident in news reports and manifests in solastalgia, albeit at a lower level of severity compared to forms of environmental distress. As little literature exists discussing solastalgia in a specifically Portuguese context, this finding signals the need for further research on the causes and impacts of changing place attachment in Portugal.

While these general trends are evident across results from both Diario de Noticias and Expresso, some differences between the two publications exist. Notably, *empresa* occurs relatively frequently in the Expresso topic lists but does not appear in any of the lists from Diario de Noticias. Upon further inspection, it was recognized that many articles on Expresso list the author's email, which uses the email domain @expresso.empresa.pt. Therefore, the frequent occurrence of *empresa* in the topic lists from Expresso is likely a result of this data quirk. The placing of this term in an email makes it relatively isolated from other topic terms, which would explain why the coherence score for the Expresso LDA was lower than the coherence score for the Diario de Noticias LDA. The term *climático* scored highly on the environmental grief and solastalgia topic lists for Diario de Noticias, but was conspicuously absent from topic lists for Expresso. Instead, the term *ambiente* appeared with slightly higher frequency in the Expresso lists. This may have been a result of differing editorial guidelines; while *ambiente* and *climático* are synonyms, *climático* appeared in Expresso article text only 307 times, compared with 2,095 occurrences in Diario de Noticias, which may reflect a publication-wide preference for the term *ambiente* among editors at Expresso. Research indicates that the effectiveness of climate communication, particularly in news media, is highly influenced by the familiarity of terms used, which is in turn audience-dependent, so the choice of different synonymous terms across publications may indicate that Diario de Noticias and Expresso have different audiences with different preferences for certain environmental terms (Nerlich et al, 2010).

Conclusion

This paper explored the existing literature on different forms of environmental distress and then conducted an analysis of Portuguese news media publications to identify terms associated with each form of environmental distress. To perform this analysis, unique Portuguese keywords were identified for each form of environmental distress, which could support further computational or qualitative assessments of environmental psychology in the Lusophone world. The coherence scores obtained using the U_{mass} method were high, indicating a high degree of co-occurrence between the identified keywords for each topic. Several trends in the results were observed, including the relative time-independence of environmental grief and the place-based nature of solastalgia, which supports results identified in previous studies. Future research could expand on these results by applying similar methods to a wider variety of news sources, including news sources with a partisan perspective, local news sources, news sources in other countries, or even non-news sources such as Tweets or academic papers.

Acknowledgments

References

Ágoston, C., Csaba, B., Nagy, B., Kőváry, Z., Dúll, A., Rácz, J., & Demetrovics, Z. (2022). Identifying types of eco-anxiety, eco-guilt, eco-grief, and eco-coping in a climate-sensitive population: A qualitative study. International Journal of Environmental Research and Public Health, 19(4), 2461. https://doi.org/10.3390/ijerph19042461

Albrecht, G. (2006). Solastalgia. Alternatives Journal, 32(4/5), 34–36. https://www.proquest.com/docview/218765167

Albrecht, G. (2020). Negating solastalgia: An emotional revolution from the anthropocene to the symbiocene. American Imago, 77(1), 9–30. https://doi.org/10.1353/aim.2020.0001

Batel, S., & Pataco, L. (2020). Portuguese media representations of nuclear facilities in Almaraz, Spain: Beyond borders and risk perception (Representaciones en los medios portugueses de las plantas nucleares en Almaraz, España: más allá de las fronteras y la percepción del riesgo). PsyEcology, 11(1), 104–115. https://doi.org/10.1080/21711976.2019.1644004

Boager, E., & Castro, P. (2022). Lisbon's unsustainable tourism intensification: Contributions from social representations to understanding a depoliticised press discourse and its consequences. Journal of Sustainable Tourism, 30(8), 1956–1971. https://doi.org/10.1080/09669582.2021.1970173

Clayton, S. (2020). Climate anxiety: Psychological responses to climate change. Journal of Anxiety Disorders, 74, 102263. https://doi.org/10.1016/j.janxdis.2020.102263

Coffey, Y., Bhullar, N., Durkin, J., Islam, M. S., & Usher, K. (2021). Understanding eco-anxiety: A systematic scoping review of current literature and identified knowledge gaps. The Journal of Climate Change and Health, 3, 100047. https://doi.org/10.1016/j.joclim.2021.100047

Comtesse, H., Ertl, V., Hengst, S. M. C., Rosner, R., & Smid, G. E. (2021). Ecological grief as a response to environmental change: A mental health risk or functional response? International Journal of Environmental Research and Public Health, 18(2), 734. https://doi.org/10.3390/ijerph18020734

Crandon, T. J., Scott, J. G., Charlson, F. J., & Thomas, H. J. (2022). A social–ecological perspective on climate anxiety in children and adolescents. Nature Climate Change, 12(2), 123–131. https://doi.org/10.1038/s41558-021-01251-y

Cunsolo, A., & Ellis, N. R. (2018). Ecological grief as a mental health response to climate change-related loss. Nature Climate Change, 8(4), 275–281. https://doi.org/10.1038/s41558-018-0092-2

Dennis, M. K., & Stock, P. (2023). Climate grief and older adult environmental activists. Journal of Gerontological Social Work, 1–16. https://doi.org/10.1080/01634372.2023.2259942

Engstrom, S. (2019). Recognising the role eco-grief plays in responding to environmental degradation. Journal of Transdisciplinary Peace Praxis, 1(1), 168–186. https://frontpagepublications.com/product/jtpp-vol-1-no-1-january-2019-article-8/

Fatima-Zahrae, S., Wafae, S., & Amal, E. M. (2021). Application of Latent Dirichlet Allocation (LDA) for clustering financial tweets. E3S Web of Conferences, 297, 1–7. https://doi.org/10.1051/e3sconf/202129701071

Galway, L. P., Beery, T., Jones-Casey, K., & Tasala, K. (2019). Mapping the solastalgia literature: A scoping review study. International Journal of Environmental Research and Public Health, 16(15), 2662. https://doi.org/10.3390/ijerph16152662

Galway, L. P., Esquega, E., & Jones-Casey, K. (2022). "Land is everything, land is us": Exploring the connections between climate change, land, and health in Fort William First Nation. Social Science & Medicine, 294, 114700. https://doi.org/10.1016/j.socscimed.2022.114700

Heeren, A., Mouguiama-Daouda, C., & McNally, R. J. (2023). A network approach to climate change anxiety and its key related features. Journal of Anxiety Disorders, 93, 102625. https://doi.org/10.1016/j.janxdis.2022.102625

Higginbotham, N., Connor, L., Albrecht, G., Freeman, S., & Agho, K. (2006). Validation of an environmental distress scale. EcoHealth, 3(4), 245–254. https://doi.org/10.1007/s10393-006-0069-x

Holthaus, L. (2022). Feelings of (Eco-) grief and sorrow: Climate activists as emotion entrepreneurs. European Journal of International Relations, 29(2), 352–373. https://doi.org/10.1177/13540661221136772

Kevorkian, K. (2019). Non-death loss and grief: Context and clinical implications (D. Harris, Ed.; 1st ed.). Routledge.

Kim, J., Yadav, M., Chaspari, T., & Ahn, C. R. (2020). Environmental distress and physiological signals: Examination of the saliency detection method. Journal of Computing in Civil Engineering, 34(6), 04020046. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000926

Kurth, C., & Pihkala, P. (2022). Eco-anxiety: What it is and why it matters. Frontiers in Psychology, 13, 981814. https://doi.org/10.3389/fpsyg.2022.981814

Lalicic, L. (2020). Solastalgia: An application in the overtourism context. Annals of Tourism Research, 82, 102766. https://doi.org/10.1016/j.annals.2019.102766

Lennon, M. (2020). Postcarbon amnesia: Toward a recognition of racial grief in renewable energy futures. Science, Technology, & Human Values, 45(5), 934–962. https://doi.org/10.1177/0162243919900556

Leviston, Z., Stanley, S. K., Rodney, R. M., Walker, I., Reynolds, J., Christensen, B. K., Monaghan, C., Calear, A. L., Lal, A., Lane, J., & Vardoulakis, S. (2023). Solastalgia mediates between bushfire impact and mental health outcomes: A study of Australia's 2019–2020 bushfire season. Journal of Environmental Psychology, 90, 102071. https://doi.org/10.1016/j.jenvp.2023.102071

Luz, P. S. (2023, November 1). Chuva e vento permitem vários dias de consumo renovável. Vai ser cada vez mais frequente. https://www.dn.pt/sociedade/chuva-e-vento-permitem-varios-dias-de-consumo-renovavel-vai-ser-cada-vez-mais-frequente-17263347.html

Mancini, A. D., & Bonanno, G. A. (2012). The persistence of attachment: Complicated grief, threat, and reaction times to the deceased's name. Journal of Affective Disorders, 139(3), 256–263. https://doi.org/10.1016/j.jad.2012.01.032

Marchand, D., Pol, E., & Weiss, K. (Eds.). (2023). In 100 key concepts in environmental psychology. Routledge, Taylor & Francis Group.

Nerlich, B., Koteyko, N., & Brown, B. (2010). Theory and language of climate change communication. WIREs Climate Change, 1(1), 97–110. https://doi.org/10.1002/wcc.2

Ojala, M. (2018). Eco-anxiety. RSA Journal, 164(4), 10–15. https://www.jstor.org/stable/26798430

Ojala, M., Cunsolo, A., Ogunbode, C. A., & Middleton, J. (2021). Anxiety, worry, and grief in a time of environmental and climate crisis: A narrative review. Annual Review of Environment and Resources, 46(1), 35–58. https://doi.org/10.1146/annurev-environ-012220-022716

Oyshi, U. A. (2023). Topic modeling and prediction of aid data in development studies using LDA and BERT [University of Arkansas at Little Rock]. https://www.proquest.com/openview/07abd752b7f4019b26fb592c701ebeb9

Phillips, C., & Murphy, C. (2021). Solastalgia, place attachment and disruption: Insights from a coastal community on the front line. Regional Environmental Change, 21(2), 46. https://doi.org/10.1007/s10113-021-01778-y

Pihkala, P. (2018). Eco‑anxiety, tragedy, and hope: Psychological and spiritual dimensions of climate change. Zygon, 53(2), 545–569. https://doi.org/10.1111/zygo.12407

Pihkala, P. (2020). Anxiety and the ecological crisis: An analysis of eco-anxiety and climate anxiety. Sustainability, 12(19), 7836. https://doi.org/10.3390/su12197836

Pihkala, P. (2021). Eco-anxiety. In C. P. Krieg & R. Toivanen (Eds.), Situating Sustainability (pp. 119–134). Helsinki University Press. https://www.jstor.org/stable/j.ctv26qjj7d.14

Pires, L. (2017). Modelagem de tópicos em Python utilizando o modelo de Alocação Latente de Dirichlet (LDA). Sauter Digital. https://medium.com/leti-pires/modelagem-de-t%C3%B3picos-em-python-utilizando-o-m odelo-de-aloca%C3%A7%C3%A3o-latente-de-dirichlet-lda-3276a469f421

Ritchie, J., & Phillips, L. G. (2021). Learning with Indigenous wisdom in a time of multiple crises: Embodied and emplaced early childhood pedagogies. Educational Review, 75(1), 54–73. https://doi.org/10.1080/00131911.2021.1978396

Sampaio, F., Costa, T., Teixeira-Santos, L., De Pinho, L. G., Sequeira, C., Luís, S., Loureiro, A., Soro, J. C., Roldán Merino, J., Moreno Poyato, A., Peña Loray, J. S., Rodríguez Quiroga, A., O'Brien, L. V., Hogg, T. L., & Stanley, S. K. (2023). Validating a measure for eco-anxiety in Portuguese young adults and exploring its associations with environmental action. BMC Public Health, 23(1), 1905. https://doi.org/10.1186/s12889-023-16816-z

Soutar, C., & Wand, A. P. F. (2022). Understanding the spectrum of anxiety responses to climate change: A systematic review of the qualitative literature. International Journal of

Environmental Research and Public Health, 19(2), 990.
https://doi.org/10.3390/ijerph19020990

Stanley, S. K., Hogg, T. L., Leviston, Z., & Walker, I. (2021). From anger to action: Differential
impacts of eco-anxiety, eco-depression, and eco-anger on climate action and wellbeing.
The Journal of Climate Change and Health, 1, 100003.
https://doi.org/10.1016/j.joclim.2021.100003

Usher, C. (2022). Eco-anxiety. Journal of the American Academy of Child & Adolescent
Psychiatry, 61(2), 341–342. https://doi.org/10.1016/j.jaac.2021.11.020

Wang, D., Thint, M., & Al-Rubaie, A. (2012). Semi-Supervised Latent Dirichlet Allocation and
Its Application for Document Classification. IEEE/WIC/ACM International Conferences
on Web Intelligence and Intelligent Agent Technology, 306–310.
https://doi.org/10.1109/WI-IAT.2012.211

Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of
topics in topic modeling: A tutorial. Advances in Methods and Practices in Psychological
Science, 6(2), 251524592311601. https://doi.org/10.1177/25152459231160105

Willox, A. C. (2012). Climate change as the work of mourning. Ethics & the Environment,
17(2), 137–164. https://doi.org/10.2979/ethicsenviro.17.2.137

Zirpoli, C. (2023). Generative Artificial Intelligence and Copyright Law. Congressional Research
Service. https://crsreports.congress.gov/product/pdf/LSB/LSB10922

Appendix A: Project Code

All project code was written in Python 2.7. Before running this code, any included libraries, such as Selenium, BeautifulSoup, pandas, spaCy, and Gensim, should be installed.

Diario de Noticias Scraper

```python
#Get text for articles with different formatting
def get_text(url, home='https://www.dn.pt'):
    # Create a browser instance
    driver = webdriver.Chrome()

    try:
    # Open the URL
    driver.get(home+url)

    # Wait for some time to let the page load (you can adjust this)
    time.sleep(2)

    # Find and extract the article text
    soup = BeautifulSoup(driver.page_source, 'html.parser')
    intro = ''
    try:
        intro = soup.find('div', class_='t-af3-head').find('p').text
    except:
        pass
    paragraphs = []
    try:
        paragraphs = soup.find('div', class_='t-af2-c1-body
js-article-readmore-resize-watch').find_all('p')
    except:
        paragraphs = soup.find('div', class_='t-af3-c1-body').find_all('p')

    article_text = intro
    for paragraph in paragraphs:
        if "cookie" not in paragraph.text:
        article_text += '\n'
        article_text += paragraph.text
    except:
    print("Problem fetching " + url)
    article_text = ''

    finally:
    # Close the browser
    driver.quit()

    return article_text
```

```python
def update_csv_with_empty_text(input_csv):
    # Read the existing CSV content
    with open(input_csv, 'r', encoding='utf-8') as csv_file:
    csv_reader = csv.DictReader(csv_file)
    rows = list(csv_reader)

    # Create a list to store rows with added article text
    updated_rows = []

    for row in rows:
        # Check if the "Article Text" field is an empty string
        if row['Article Text'] == '':
        # Get the article text for the current link
        link = row['Link']
        article_text = get_text(link)

        # Remove extra lines for readability
        article_text = article_text.replace('\n', '')

        # Add the article text to the current row
        row['Article Text'] = article_text

        updated_rows.append(row)

    # Write the updated content to the CSV file
    with open(input_csv, 'w', encoding='utf-8', newline='') as csv_file:
    fieldnames = csv_reader.fieldnames
    csv_writer = csv.DictWriter(csv_file, fieldnames=fieldnames)
    csv_writer.writeheader()
    csv_writer.writerows(updated_rows)

# Replace this as needed
input_csv_filename = "dn_links_with_text.csv"

# Call the function to update the CSV file
update_csv_with_empty_text(input_csv_filename)

# Add text for articles with alternative HTML structures
def get_text(url, home='https://www.dn.pt'):
 # Create a browser instance
 driver = webdriver.Chrome()

 try:
    # Open the URL
    driver.get(home+url)

    # Wait for some time to let the page load (you can adjust this)
```

```python
    time.sleep(2)

    # Find and extract the article text (different structures for some articles
which have weird structures)
    soup = BeautifulSoup(driver.page_source, 'html.parser')
    intro = ''
    try:
        intro = soup.find('div', class_='t-af3-head').find('p').text
    except:
        pass
    paragraphs = []
    try:
        paragraphs = soup.find('div', class_='t-af2-c1-body
js-article-readmore-resize-watch').find_all('p')
    except:
        paragraphs = soup.find('div', class_='t-af3-c1-body').find_all('p')
        article_text = intro
        for paragraph in paragraphs:
            if "cookie" not in paragraph.text:
                article_text += '\n'
                article_text += paragraph.text
    except:
        print("Problem fetching " + url)
        article_text = ''

    finally:
        # Close the browser
        driver.quit()

    return article_text

def update_csv_with_empty_text(input_csv):
    # Read the existing CSV content
    with open(input_csv, 'r', encoding='utf-8') as csv_file:
        csv_reader = csv.DictReader(csv_file)
        rows = list(csv_reader)


    # Create a list to store rows with added article text
        updated_rows = []

        for row in rows:
            # Check if the "Article Text" field is an empty string
            if row['Article Text'] == '':
            # Get the article text for the current link
            link = row['Link']
```

```python
            article_text = get_text(link)

            # Remove extra lines for readability
            article_text = article_text.replace('\n', '')

            # Add the article text to the current row
            row['Article Text'] = article_text

            updated_rows.append(row)

        # Write the updated content to the CSV file
        with open(input_csv, 'w', encoding='utf-8', newline='') as csv_file:
        fieldnames = csv_reader.fieldnames
        csv_writer = csv.DictWriter(csv_file, fieldnames=fieldnames)
        csv_writer.writeheader()
        csv_writer.writerows(updated_rows)

  # Replace this as needed
  input_csv_filename = "dn_links_with_text.csv"

  # Call the function to update the CSV file
  update_csv_with_empty_text(input_csv_filename)
```

Expresso Scraper

```python
#This code block collects the first 1000 articles from DN tagged "ambiente" and
scans the article text.
#The automated web page needs to be in active focus on your computer for articles
to load correctly.

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from bs4 import BeautifulSoup
import time
import csv

def write_links_to_csv(links, csv_filename="expresso_links.csv"):
    with open(csv_filename, 'w', newline='', encoding='utf-8') as csv_file:
    csv_writer = csv.writer(csv_file)
    csv_writer.writerow(['Link'])

    for link in links:
        csv_writer.writerow([link])

def scrape_links(url, max_links=1000, wait=2):
```

```python
    all_links = []

    # Create a browser instance
    driver = webdriver.Chrome()

    try:
    # Open the URL
    driver.get(url)

    # Wait for some time to let the page load (you can adjust this)
    time.sleep(wait)

    # Keep track of the number of scrolls
    scrolls = 0

    # Click the "load-more-button" until reaching the desired number of links
    while len(all_links) < max_links:
            # Find all <a> links on the current page
            soup = BeautifulSoup(driver.page_source, 'html.parser')
            a_links = soup.find_all('a')

            # Extract the href attribute from each link
            for link in a_links:
            href = link.get('href')
            if href:
                    if href[0] == '/' and href[0:8] != '/autores':
                            if '1' in href or '2' in href or '3' in href or '4' in
href or '5' in href or '6' in href or '7' in href or '8' in href or '9' in href or
'0' in href:

                                    if href not in all_links and 'podcast' not in href:
                                            all_links.append(href)

                    # Break if we reached the desired number of links
                    if len(all_links) >= max_links:
                            break

            # Scroll down to trigger the loading of more articles
            driver.execute_script("window.scrollTo(0,
document.body.scrollHeight);")

            # Wait for some time to let the new content load (you can adjust this)
            time.sleep(wait)

            # Increment the number of scrolls
            scrolls += 1
            print("Scroll "+str(scrolls)+": "+str(len(all_links))+" articles.")
```

```
            # Wait for some time to let the new content load (you can adjust this)
            time.sleep(wait)

        finally:
        # Close the browser
        driver.quit()

        return all_links

# URL of the page we want to scrape
page_url = "https://expresso.pt/pesquisa?q=ambiente"

# Get the first 1000 links by scrolling down
links = scrape_links(page_url, max_links=1000)

# Write links to the csv file
write_links_to_csv(links)

#scrape article text
def update_csv_with_article_text(input_csv, output_csv):
        with open(input_csv, 'r', encoding='utf-8') as csv_file:
        # Read the input CSV
        csv_reader = csv.reader(csv_file)
        header = next(csv_reader)  # Read the header

        # Find the index of the "Link" column
        link_column_index = header.index("Link")

        # Create a list to store rows with added article text
        updated_rows = []

        for row in csv_reader:
                link = row[link_column_index]

                # Get the article text for the current link
                article_text = get_article_text(link)

                # Remove extra lines for readability
                article_text.replace('\n', '')

                # Add the article text to the current row
                row.append(article_text)
                updated_rows.append(row)

        # Write the updated rows to a new CSV file
        with open(output_csv, 'w', newline='', encoding='utf-8') as output_csv_file:
        csv_writer = csv.writer(output_csv_file)
```

```python
        csv_writer.writerow(header + ["Article Text"])  # Add a new column header

        for updated_row in updated_rows:
            csv_writer.writerow(updated_row)

# Replace these filenames as needed
input_csv_filename = "expresso_links.csv"
output_csv_filename = "expresso_links_with_text.csv"

# Update the CSV with article text
update_csv_with_article_text(input_csv_filename, output_csv_filename)

# Fill in missing articles
def other_article_text(url, home='https://www.expresso.pt'):
    # Create a browser instance
    driver = webdriver.Chrome()

    try:
    # Open the URL
    driver.get(home+url)

    # Wait for some time to let the page load (you can adjust this)
    time.sleep(2)

    # Find and extract the article text (different structures for some articles
which have weird structures)
    soup = BeautifulSoup(driver.page_source, 'html.parser')
    intro = ''
    try:
        intro = soup.find('h2', class_='lead').text
    except:
        try:
            intro = soup.find('div', class_='g-article-lead lead').find('p').text
        except:
            pass
    paragraphs = []
    try:
        paragraphs = soup.find('div', class_='article-content').find_all('p')
    except:
        paragraphs = soup.find('div', class_='t-af3-c1-body').find_all('p')

    article_text = intro
    for paragraph in paragraphs:
            if "cookie" not in paragraph.text:
            article_text += '\n'
            article_text += paragraph.text
    except:
```

```python
        print("Problem fetching " + home + url)
        article_text = ''

    finally:
        # Close the browser
        driver.quit()

    return article_text

def update_csv_with_empty_text(input_csv):
    # Read the existing CSV content
    with open(input_csv, 'r', encoding='utf-8') as csv_file:
    csv_reader = csv.DictReader(csv_file)
    rows = list(csv_reader)

    # Create a list to store rows with added article text
    updated_rows = []

    for row in rows:
        # Check if the "Article Text" field is an empty string
        if row['Article Text'] == '':
        # Get the article text for the current link
        link = row['Link']
        article_text = other_article_text(link)

        # Remove extra lines for readability
        article_text = article_text.replace('\n', '')

        # Add the article text to the current row
        row['Article Text'] = article_text

        updated_rows.append(row)

    # Write the updated content to the CSV file
    with open(input_csv, 'w', encoding='utf-8', newline='') as csv_file:
    fieldnames = csv_reader.fieldnames
    csv_writer = csv.DictWriter(csv_file, fieldnames=fieldnames)
    csv_writer.writeheader()
    csv_writer.writerows(updated_rows)

# Replace filename as needed
input_csv_filename = "expresso_links_with_text.csv"

# Call the function to update the CSV file
update_csv_with_empty_text(input_csv_filename)
```

Text Cleaning

```python
# Importing modules
import pandas as pd

papers = pd.read_csv('test.csv')

# Remove the columns
papers = papers.drop(columns=['Link'], axis=1)

# Load the regular expression library
import re

# Remove punctuation
papers['paper_text_processed'] = papers['Article Text'].apply(lambda x:
re.sub(r'[,\.!?"()]', '', str(x)))

# Convert the text to lowercase
papers['paper_text_processed'] = papers['paper_text_processed'].apply(lambda x:
x.lower())

# Save the processed text as a list
processed = papers['paper_text_processed']

#Lemmatize with spaCy
import spacy

# Load the spaCy Portuguese language model
nlp_pt = spacy.load('pt_core_news_sm')

# Lemmatize function
def lemmatize(l):
    output = []
    for article in l:
        lemmatized_text = [token.lemma_ for token in nlp_pt(article)]
        output.append(lemmatized_text)
        return output

lemmatized = lemmatize(processed)

#Remove stopwords
import gensim
from gensim.utils import import simple_preprocess
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords

stopwords = stopwords.words('portuguese')
```

```
stopwords.append('ja')
stopwords.append('ne')
stopwords.append('ai')
stopwords.append('ta')
stopwords.append('gente')
stopwords.append('nao')
stopwords.append('aqui')
stopwords.append('tambem')
stopwords.append('vc')
stopwords.append('entao')
stopwords.append('até')
stopwords.append('agora')
stopwords.append('ser')
stopwords.append('sempre')
stopwords.append('ter')
stopwords.append('só')
stopwords.append('porque')
stopwords.append('sobre')
stopwords.append('ainda')
stopwords.append('la')
stopwords.append('tudo')
stopwords.append('ninguem')
stopwords.append('de')
#additional stopwords identified in verification process
stopwords.append('the')
stopwords.append('ha')
stopwords.append('xa')


def sent_to_words(sentences):
 for sentence in sentences:
        # deacc=True removes punctuations
        yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))
def remove_stopwords(texts):
 return [[word for word in simple_preprocess(str(doc)) if word not in stopwords]
for doc in texts]
# remove stopwords
data_words = remove_stopwords(lemmatized)
```

LDA and Visualization

```
 #Guided LDA

 from gensim import corpora, models
 from gensim.models.ldamodel import LdaModel
 from gensim.models import CoherenceModel
 import numpy as np
```

```python
# Seed terms for three topics
# grief
d1 = ['ausente', 'ausencia', 'identidade', 'tradição', 'conhecimento', 'lugar',
'cultura', 'extinto', 'morto']
# solastalgia
d2 = ['triste', 'degradado', 'ameaçar', 'recurso', 'paisagem', 'natureza',
'solidão', 'vergonha', 'sair']
# anxiety
d3 = ['nervoso', 'ansioso', 'preocupado', 'preocupar', 'medo', 'refletir',
'dormir', 'descansar', 'foco']

# Create a dictionary and a corpus
dictionary = corpora.Dictionary(data_words)
corpus = [dictionary.doc2bow(doc) for doc in data_words]

# Set up the guided LDA model
guided_topics = [d1, d2, d3]
guided_topics_bow = [dictionary.doc2bow(topic) for topic in guided_topics]
guidedlda_model = models.ldamodel.LdaModel(corpus=corpus, id2word=dictionary,
num_topics=3)

# Update the model with the guided topics
guidedlda_model.update(guided_topics_bow)

# Print the topics and associated terms
for topic_id, topic_terms in guidedlda_model.print_topics(num_topics=3):
        print(f"Topic {topic_id + 1}: {topic_terms}")

# Calculate coherence score
coherence_model = CoherenceModel(model=guidedlda_model, texts=lemmatized,
dictionary=dictionary, coherence='u_mass')
coherence_score = coherence_model.get_coherence()
print(f"Coherence Score: {coherence_score}")

#Visualization

import matplotlib.pyplot as plt
import seaborn as sns

topics = guidedlda_model.show_topics(formatted=False)

# Create a figure and axis
fig, axes = plt.subplots(len(topics), 1, figsize=(10, 5 * len(topics)),
sharex=True)
```

```python
# Plot each topic
for i, (topic_id, topic_words) in enumerate(topics):
    words = [word for word, _ in topic_words]
    weights = [weight for _, weight in topic_words]

    sns.barplot(x=weights, y=words, ax=axes[i], palette='viridis')

axes[0].set_title('Environmental Grief')
axes[1].set_title('Solastalgia')
axes[2].set_title('Eco-Anxiety')

plt.xlabel('Word Weight')
plt.tight_layout()
plt.show()
```