Natural Language Processing for Food Security Policymaking: a Scoping Review

Marieke Meeske[1,2*], Frans Cruijssen[1,3], Chris van der Lee[2], and Emiel Krahmer[2]

[1]Zero Hunger Lab, Tilburg University, North Brabant, The Netherlands
[2]Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg University, North Brabant, The Netherlands
[3]Tilburg School of Economics and Management, Department of Econometrics and Operations Research, Tilburg University, North Brabant, The Netherlands
[*]*Address correspondence to: m.meeske@tilburguniversity.edu; PO Box 90153, 5037 AB Tilburg, The Netherlands*

13 November 2024

Abstract

Natural Language Processing (NLP) offers significant opportunities to support the Sustainable Development Goals (SDGs), including Zero Hunger. While many NLP applications have been documented for SDGs such as healthcare and education, its application to food security remains largely unexplored. This paper addresses this knowledge gap through a comprehensive scoping review focused on NLP for food security policymaking. Six key application areas were identified: 1) Early warning systems for food insecurity, 2) Understanding public discourse on food related issues, 3) Knowledge generation and management from food policy and program documents, 4) Understanding dietary habits, 5) Food item classification, and 6) Addressing data gaps in food security statistics and crisis response. However, limited deployment hinders real-world impact. Establishing authentic partnerships from the outset will be essential for successful and sustained implementation of NLP projects that advance progress towards ending hunger and achieving food security and improved nutrition for all.

Keywords: Food security, Natural Language Processing, Policymaking, Sustainable development.

1 Introduction

In 2015, the United Nations (UN) adopted the 2030 Agenda for Sustainable Development and its 17 Sustainable Development Goals (SDGs). Part of this Agenda is SDG 2 ("Zero Hunger"), aiming to "end hunger, achieve food security and improved nutrition and promote sustainable agriculture (UN, 2015, p. 14)". Despite this global commitment, hunger and food insecurity have risen since 2015, exacerbated by, e.g., the Covid-19 pandemic, conflict, climate change and growing inequalities (UN DESA, 2023). Consequently, in 2023, Zero Hunger was the SDG with the largest deterioration in progress compared to baseline (UN DESA,

Abbreviations: AI, Artificial Intelligence; AI4SG, Artificial Intelligence for Social Good; BERT, Bidirectional Encoder Representations from Transformers; FAO, Food and Agricultural Organization of the United Nations; GPT, Generative Pre-trained Transformer; HLPE-FSN, High Level Panel of Experts on Food Security and Nutrition; LDA, Latent Dirichlet Allocation; LLM, Large Language Model; NLP, Natural Language Processing; NLP4SG, Natural Language Processing for Social Good; PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analyses; SDG, Sustainable Development Goal; TRL, Technology Readiness Level; UN, United Nations.

2024a). Specifically, three out of eight targets have deteriorated: undernourishment and food security (2.1), investment in agriculture (2.a), and food price anomalies (2.c). The UN estimated that 9.1% of the world population (733 million people) faced chronic hunger in 2023, rising to 28.9% when considering moderate or severe food insecurity (UN DESA, 2024a). This means that about 2.33 billion people lacked access to adequate food in 2023, and without increased efforts the goal of Zero Hunger will not be achieved by 2030.

These statistics call for urgent coordinated policymaking and actions "to address entrenched inequalities, transform food systems, invest in sustainable agricultural practices, and reduce and mitigate the impact of conflict and the pandemic on global nutrition and food security" (UN DESA, 2023, p. 14). Given the diverse challenges that food systems face, shifts in food and nutrition security policies are needed, along with a stronger enabling environment and a broader understanding of food security and food systems thinking (HLPE-FSN, 2020). Effective food security policymaking requires coordinated action across various administrative levels, with governments and international organizations (e.g., the Food and Agricultural Organization of the UN (FAO) and the World Food Programme) acting as policymakers. In-depth interviews with government actors also highlighted the role of global organizations in strengthening cross-sectoral nutrition policy coordination at the national level (Thow, 2024). Access to high-quality, timely, and relevant data and analysis is crucial for informing these efforts (HLPE-FSN, 2022).

In the last decades, there has been a significant increase in the volume of unstructured data generated from diverse sources, including social media, research publications, and news articles. Due to its complexity and volume, traditional methods of data analysis are often insufficient to extract actionable insights for evidence-based policymaking. An opportunity to efficiently collect and analyze this unstructured data is the use of Natural Language Processing (NLP), a field concerned with employing "computational techniques for the purpose of learning, understanding, and producing human language" (Hirschberg & Manning, 2015, p. 1). Key application areas of NLP for policymaking include data analysis for evidence-based policymaking, interpretation of political decisions, policy communication, and policy evaluation (Jin & Mihalcea, 2023). For example, (Funk et al., 2024), demonstrate how NLP can efficiently analyze vast scientific publications on the SDGs, providing policymakers with insights that go beyond standard indicators. NLP thus holds potential to help policymakers navigate complex information landscapes, enabling more effective food security policymaking.

Recent advancements in NLP, particularly the emergence of Large Language Models (LLMs) with their high performance across diverse NLP tasks (Zhao et al., 2024), have led to the growing initiative of NLP for Social Good (NLP4SG) (Adauto et al., 2023), and Artificial Intelligence for Social Good (AI4SG) more broadly. While many NLP4SG and AI4SG applications have been documented for SDGs such as healthcare and education, its application to food security remains largely unexplored (Adauto et al., 2023; Cowls et al., 2021; ITU, 2024; Nasir et al., 2023; Shi et al., 2020; Singh et al., 2024). Consequently, there remains the need for a concerted effort to identify how NLP can be used for food security. This paper aims to address this knowledge gap through a comprehensive scoping review focused on NLP for food security policymaking. By doing so, it provides an overview of how NLP can be utilized in this domain, including the identification of potential gaps for future research and application areas, to advance the progress towards ending hunger and achieving food security and improved nutrition for all.

This paper is structured as follows: Section 2 defines the main concepts of this scoping review; Section 3 describes the methodology; Section 4 presents the results; Section 5 discusses the implications, including ethical reflections and directions for future research; and Section 6 concludes.

2 Conceptual framework

2.1 Natural Language Processing

As stressed in Section 1, NLP is concerned with employing “computational techniques for the purpose of learning, understanding, and producing human language” (Hirschberg & Manning, 2015, p. 1). NLP combines the fields of computer science, linguistics, and cognitive science, and is considered a subfield of Artificial Intelligence (AI). The field developed from early rule-based approaches in the 1950s to the latest advances in deep learning. Especially with the introduction of Transformers (Vaswani et al., 2017) and pre-trained language models in the 2010s (e.g., Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) and Generative Pre-trained Transformer (GPT) (Radford et al., 2018)), which led to the development of LLMs, the interest in NLP increased significantly (Schopf et al., 2023).

The field of NLP encompasses a wide range of tasks and techniques. (Schopf et al., 2023) proposed a taxonomy of NLP subfields, illustrated in Figure 1, that distinguishes twelve higher-level subfields, each further divided into specific tasks or techniques. Few subfields are particularly relevant for NLP for food security policymaking, including *semantic text processing*, which focuses on deriving meaning and semantic understanding from text. *Language models*, one of its key lower-level subfields, include LLMs which, due to their significant size, have demonstrated strong capabilities in addressing various NLP tasks and have substantially raised the performance bar across the field (Zhao et al., 2024). Another subfield, *information extraction & text mining*, aims to extract structured knowledge from unstructured text, revealing patterns and correlations. It encompasses various tasks, including *classification*, *topic modeling*, and *named entity recognition*. Additionally, *low-resource NLP*, part of *responsible & trustworthy NLP*, is concerned with NLP tasks in data-scarce environments and has recently attracted increased interest from NLP researchers.

Natural Language Processing (NLP)

Higher-level subfields	Lower-level subfields
Multimodality	Visual Data in NLP; Structured Data in NLP; Speech & Audio in NLP; Programming Languages in NLP
Natural Language Interfaces	Dialogue Systems & Conversational Agents; Question Answering
Semantic Text Processing	Discourse & Pragmatics; Knowledge Representation; Text Complexity; Semantic Search; Representation Learning; Word Sense Disambiguation; Semantic Parsing; Language Models; Semantic Similarity
Sentiment Analysis	Opinion Mining; Stylistic Analysis; Intent Recognition; Emotion Analysis; Aspect-based Sentiment Analysis; Polarity Analysis
Syntactic Text Processing	Tagging; Morphology; Chunking; Phonology; Text Error Correction; Text Segmentation; Typology; Syntactic Parsing; Phonetics; Text Normalization
Linguistics & Cognitive NLP	Linguistic Theories; Cognitive Modeling; Psycholinguistics
Responsible & Trustworthy NLP	Ethical NLP; Low-Resource NLP; Robustness in NKP; Green & Sustainable NLP; Explainability & Interpretability in NLP
Reasoning	Textual Inference; Commonsense Reasoning; Numerical Reasoning; Knowledge Graph Reasoning; Machine Reading Comprehension; Fact & Claim Verification; Argument Mining
Multilinguality	Cross-Lingual Transfer; Machine Translation; Code-Switching; Typology
Information Retrieval	Indexing; Document Retrieval; Semantic Search; Text Classification; Passage Retrieval
Information Extraction & Text Mining	Coreference Resolution; Text Clustering; Named Entity Recognition; Event Extraction; Open Information Extraction; Term Extraction; Text Classification; Topic Modeling; Summarization; Relation Extraction
Text Generation	Data-to-Text Generation; Summarization; Question Generation; Dialogue Response Generation; Captioning; Paraphrasing; Machine Translation; Speech Recognition; Text Style Transfer; Code Generation

Figure 1: Taxonomy of subfields in NLP. Figure based on Schopf et al. (2023).

2.2 Food security

The concept of food security has evolved substantially over the past 50 years (HLPE-FSN, 2020). Initially focused primarily on the availability of sufficient food supplies, it was later expanded to include access, nutrition, and cultural dimensions. By 2001, FAO's updated definition distinguished four dimensions of food security: availability, access, utilization, stability. In 2020, the HLPE-FSN called for an expanded definition that explicitly includes "agency" and "sustainability" as additional dimensions (HLPE-FSN, 2020). Table 1 describes the six interconnected food security dimensions. Building on these, food security exists when "All people (*agency*), at all times (*stability, sustainability*), have physical, social and economic access (*access*) to sufficient (*availability*), safe and nutritious food that meets their dietary needs (*utilization*) and food preferences (*agency*) for an active and healthy life (HLPE-FSN, 2020, p. 10)".

Table 1: The six dimensions of food security.

Dimension	Definition
Availability	Having a quantity and quality of food sufficient to satisfy the dietary needs of individuals, free from adverse substances and acceptable within a given culture, supplied through domestic production or imports.
Access	Having personal or household financial means to acquire food for an adequate diet at a level to ensure that satisfaction of other basic needs are not threatened or compromised; and that adequate food is accessible to everyone, including vulnerable individuals and groups.
Utilization	Having an adequate diet, clean water, sanitation and health care to reach a state of nutritional well-being where all physiological needs are met.
Stability	Having the ability to ensure food security in the event of sudden shocks (e.g., an economic, health, conflict or climatic crisis) or cyclical events (e.g., seasonal food insecurity).
Agency	Individuals or groups having the capacity to act independently to make choices about what they eat, the foods they produce, how that food is produced, processed, and distributed, and to engage in policy processes that shape food systems. The protection of agency requires socio-political systems that uphold governance structures that enable the achievement of food security and nutrition for all.
Sustainability	Food system practices that contribute to long-term regeneration of natural, social and economic systems, ensuring the food needs of the present generations are met without compromising the food needs of future generations.

Source: HLPE-FSN (2020).

2.3 Policymaking

The third key concept in this scoping review is policymaking, which is operationalized through the policy cycle framework outlined by (Jann & Wegrich, 2017). This framework describes the policy cycle according to four stages: 1) Agenda setting; 2) Policy formulation & decision-making; 3) Implementation; 4) Evaluation. In the *agenda setting* stage, a problem, such as the severity of food insecurity in a particular region, is recognized and prioritized for action. A proper understanding of the determinants and magnitude of a problem is necessary for making an informed prioritization. Then, during the *policy formulation & decision-making* stage, identified problems are transformed into policies and/or programs, e.g., the development of a food aid program. Next, the *implementation* stage involves executing or enforcing the formulated policies or programs. In the fourth and final stage, *evaluation*, a policy or program is assessed in relation to its intended and unintended outcomes. Evaluation helps identifying successes and areas for improvement, guiding future policy adjustments and adaptations based on the observed outcomes and challenges.

2.4 Technology Readiness

To achieve real-world impact, NLP projects for food security policymaking must be successfully and sustainably implemented. Accordingly, this study evaluates the degree to which identified NLP research projects are implemented in practice. The concept of implementation is operationalized through the Technology Readiness Levels (TRL) scale (Kimmel et al., 2020), which measures the maturity of a technology. Using nine levels, technological maturity is categorized into research (TRL 1-3), development (TRL 4-6), and deployment (TRL 7-9) (see Figure 2). In the *research* phase, the fundamental principles of a technology are observed and documented, followed by identifying initial practical applications and conducting applied research to develop a proof of concept. In the *development* phase, a prototype is designed, developed, and validated both in the lab and in a relevant environment. In the final stage, *deployment*, the prototype evolves into an actual system, demonstrated in an operational environment, and prepared for commercialisation and full-scale deployment.

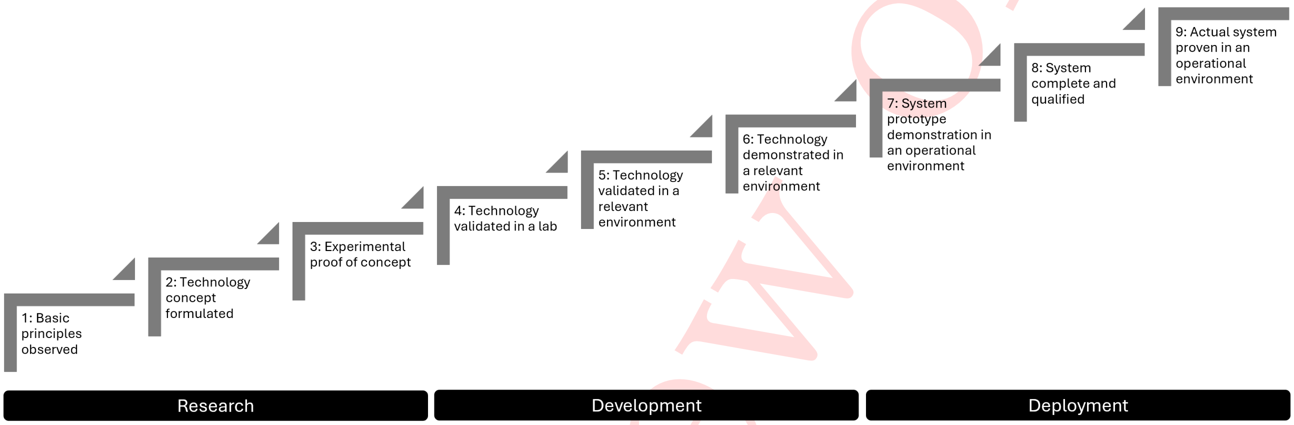

Figure 2: Technology Readiness Levels. Figure based on EC (2024) and Kimmel et al. (2020).

3 Methodology

With main concepts defined, the methodology for the study is outlined in this section. The review protocol is available on the GitHub page of the first author, and adhered to PRISMA guidelines (Preferred Reporting Items for Systematic reviews and Meta-Analyses) (Moher et al., 2009; Tricco et al., 2018).

3.1 Database search

A search strategy was employed to retrieve studies from both peer-reviewed (Web of Sciences and Scopus) as well as grey databases (Arxiv, Policy Commons, International Food Policy Research Institute Publications, World Bank Open Knowledge Repository, UN Digital Library System, and FAO Knowledge Repository BETA).

The search query, detailed in Table 2, operationalized the concepts of NLP and food security. The NLP terms were built on several review studies about AI or NLP for social good (Adauto et al., 2023; N. M. Martin, Sedoc, et al., 2022; Meitei et al., 2023; Sarku et al., 2023), while the food security terms were built on the Long-term EU-AU Research and Innovation Partnership for Food and Nutrition Security and Sustainable Agriculture lexicon by (Roche et al., 2022) and expanded building on (HLPE-FSN, 2020). Although policymaking is another key concept for this paper, it was intentionally excluded from the search query to avoid missing studies that, while not explicitly mentioning policymaking, still offer valuable NLP applications for enhancing food security policymaking.

Where specification was possible, the search encompassed the title, abstract, and keywords fields. Advanced search techniques were not feasible for most grey literature databases, requiring the search strategy

Table 2: High-level search query.

Row #	Search query
1	("natural language processing" OR "nlp" OR "language model*" OR "llm" OR "language representation" OR "language anal*" OR "word embed*" OR "text model*" OR "text mining" OR "text analy*" OR "computational linguistic*")
2	("food security" OR "food insecurity" OR "food access" OR "access to food" OR "food aid" OR "food sovereignty" OR "hunger" OR "nutrition security" OR "nutrition insecurity" OR "right to food" OR "self-sufficiency" OR "novel food" OR "resource management" OR "early warning" OR "nutritional quality" OR "malnutrition" OR "undernourish*" OR "socioeconomic sustainability" OR "sustainable intensification" OR "food system*" OR "food availability" OR "available food" OR "food affordability" OR "affordable food" OR "food utility" OR "food stability" OR "food agency" OR "food sustainability" OR "food distribution" OR "food supply" OR "food production" OR "food justice" OR "agrifood" OR "sdg 2" OR "sustainable development goal 2")
3	#2 AND #3
4	Refined by: PUBLICATION YEARS: (2010-2024)

to be customized for each individual database. A detailed overview of the search query and the number of studies retrieved per database is presented in Appendix A.1. The database search was conducted between 26 April - 3 May 2024.

3.2 Study selection

Studies from the various databases were merged and de-duplicated prior to screening. Study selection was then performed in three stages, as detailed in the PRISMA flow chart in Figure 3. In the first stage, one reviewer screened the title and abstract of the candidate studies against the eligibility criteria: the study covers an original and applied study; is written in English; is published after 2009; describes an NLP application; focuses on food and/or nutrition security; and describes a (potential) policy application. A detailed overview of eligibility criteria is presented in Appendix A.2. Only studies that met all the inclusion criteria were included, along with those for which a definitive decision could not be made based solely on the title and abstract. To expedite the screening process, the software ASReview (ASReview LAB developers, 2024) was utilized. This tool utilizes active learning techniques to dynamically present records to the reviewer based on their past decisions, thereby significantly reducing screening time (ASReview, 2024a) (details in Appendix A.3). To minimize risk of bias in the screening process, 12% of candidate studies were double screened by a second reviewer, reaching a 94% agreement rate with the primary reviewer (details in Appendix A.3). Finally, as a validation step, the titles and abstracts of all studies were screened using the GPT-4o model by OpenAI, reaching a 93% agreement rate with the primary reviewer (details in Appendix A.4).

In the second selection stage, the full text of each study that passed the first stage was retrieved and evaluated against the eligibility criteria.

In the third and final stage, the list of selected studies after detailed assessment was inserted into ResearchRabbit, which is an online citation-based literature mapping tool (details in Appendix A.5). Using ResearchRabbit, additional relevant studies were identified based on their similarity to the already identified studies. Additionally, some studies manually identified during the conceptualization of this scoping review but that did not emerge from the database search were also included. The newly added studies underwent a detailed evaluation against the eligibility criteria as in stage two, leading to the final set of included studies.

3.3 Data extraction and analysis

Extracted data items included meta data (e.g., title, authors, publication year, author affiliation, study location), key contributions (e.g., results, limitations, and notes on ethics and/or fairness), and study characteristics (e.g., food security dimension, policy cycle stage, NLP subfield, data source, and TRL) (details in Appendix B.1). Data extraction was performed by one primary reviewer. To minimize the risk of bias, data

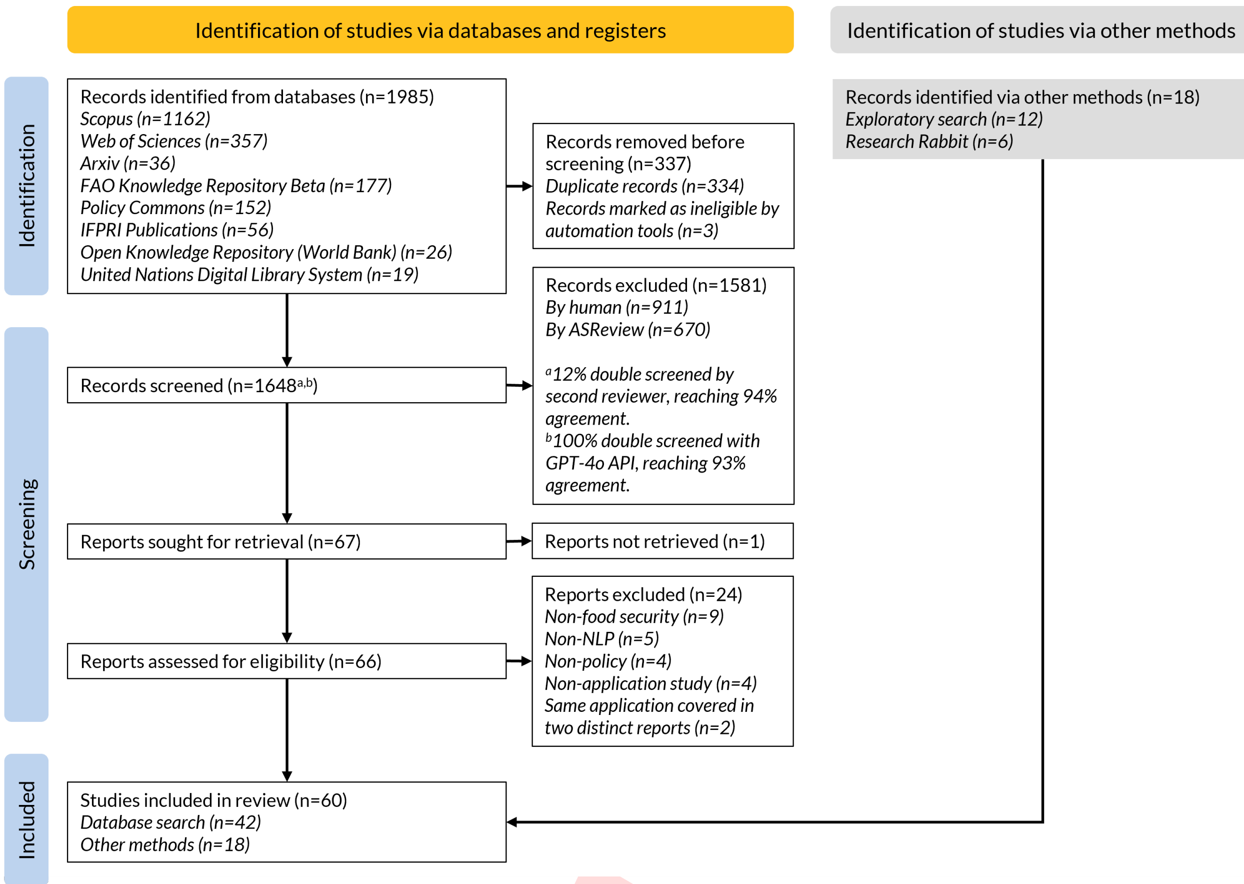

Figure 3: PRISMA flow chart detailing the study identification and selection process.

was also extracted by two additional reviewers for a random sample of 12% of the studies. Additionally, for all included studies, data was extracted using the GPT-4o model (details in Appendix B.2). The extracted data from the three reviewers as well as GPT-4o were subsequently discussed among the three reviewers until consensus was reached.

All studies were manually categorized into distinct NLP application areas for food security policymaking. This process began by reformulating each study's research objective into a high-level objective that considered its potential policymaking application[†]. These reformulated objectives provided a broad overview of the scope of included studies, and subsequently were categorized into application areas. Since the categorization was a manual subjective effort, alternative categorizations could be valid. As a validation step, the GPT-4o model was prompted to suggest broad categories of NLP applications for food security policymaking (details in Appendix B.3), resulting in minor refinements, primarily in the framing of the application areas.

4 Results

Out of 1985 studies retrieved through the database search, 42 were included in this review. An additional 18 studies were selected through the exploratory search and ResearchRabbit, totaling 60 included studies (see Figure 3). Included studies have been published between 2013 and 2024, with about half published from 2021 onwards, reflecting the rising trend of NLP4SG in recent years (Adauto et al., 2023). Of the included

[†]For example, the objective from (Brzustewicz & Singh, 2021), "*Identify the topics users tweet about sustainable consumption and detect the emotion-based sentiments in the tweets.*", was reformulated as "*Understand citizen perception on policies, programs and/or (potential) policy concepts*".

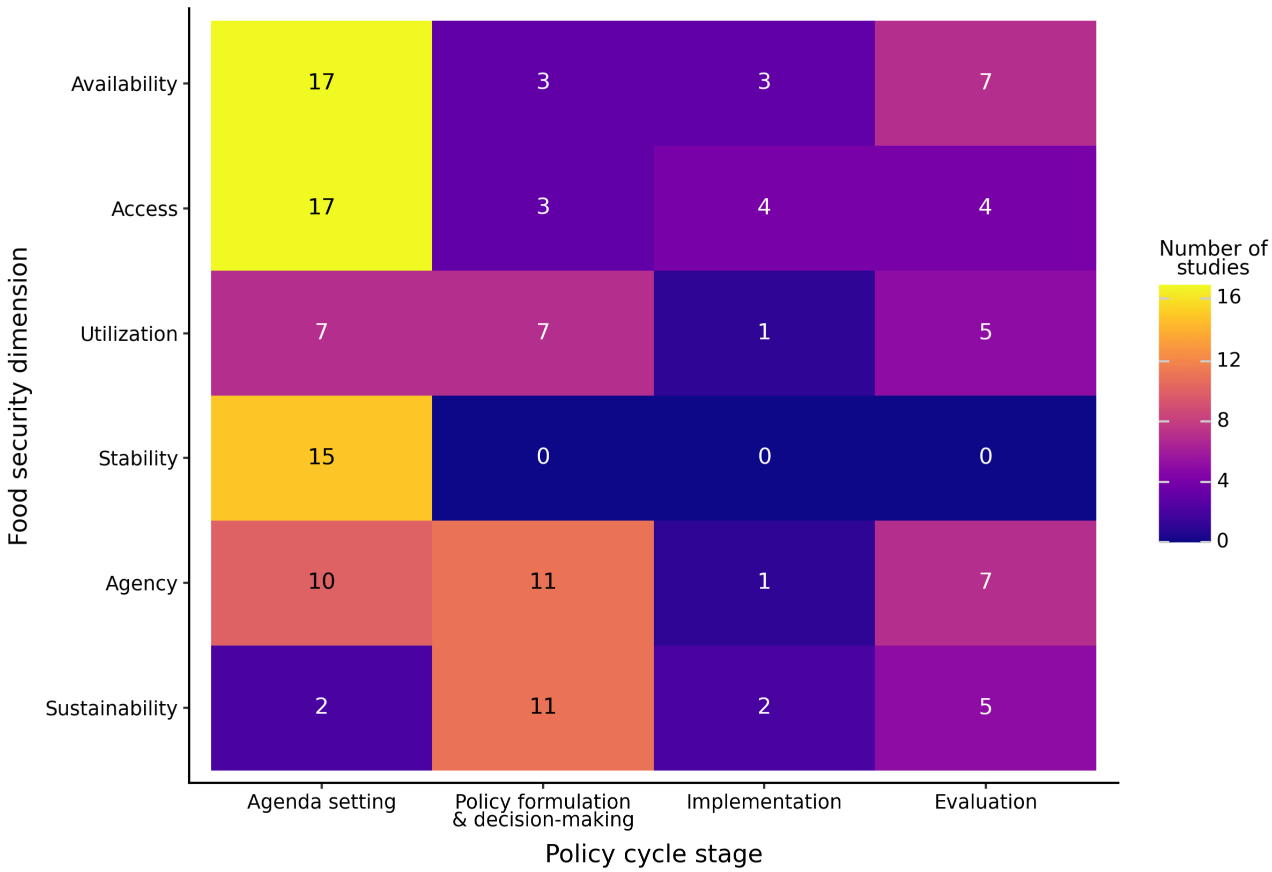

Figure 4: Heat map showing the distribution of NLP for food security policymaking studies across the six food security dimensions and four stages of the policy cycle. Each study could fit into multiple categories along both axes. For clarity, general studies addressing all dimensions of food security (n=10) were excluded from the Figure.

studies, 60% were published in peer-reviewed journals, 28% were conference papers, and the remaining 12% were either working papers or other types of grey literature documents. Studies either had a global or non-country-specific focus (n=21) or they concentrated on individual countries. The latter covered 26 individual countries in total, mostly the United States (n=14). Authors were predominantly affiliated with institutions in the United States or Europe (e.g., United States n=25, Italy n=8, France n=4).

As the main concepts in this review are food security and policymaking, studies were classified on a 6 (dimensions of food security) by 4 (stages of policy cycle) matrix (see Figure 4). The resulting heat map showed that the food security dimensions of availability, access, and agency were most prevalent in the overall set of NLP studies related to food security policymaking (n=27-30). Agenda setting was the most common policy application (n=66), frequently linked to the food security dimensions of availability, access, and stability (n=15-17). These commonly were studies that could contribute to the development of early warning systems for impending food crises. Furthermore, agency and utilization were commonly addressed throughout the policy cycle stages (n=13-17), with the exception of implementation. Examples were studies using NLP to understand dietary habits as input for both policy formulation & decision-making and evaluation. Agency and sustainability were common around the policy formulation & decision-making stage (n=11), and included e.g., studies examining public discourse and sentiment. The evaluation stage of policymaking had a more or less even spread across the food security dimensions (n=4-7), with the exception of stability (n=0). Overall, the implementation stage of policymaking seemed relatively unaddressed (n=11).

Table 3: The six application areas of NLP for food security policymaking.

Application area	Description	Food security dimension	Policy cycle stage	Data source	NLP subfield	TRL	N
Early warning systems for food insecurity	Predicts or forecasts a certain food security status, providing valuable input for the development of early warning systems for impending food crises. Other studies provide contextual explanation for a certain food security status, or predict food price inflation.	Availability; Access; Stability	Agenda setting	News articles; Social media data	Information extraction & text mining; Sentiment analysis	2-5	18
Understanding public discourse on food related issues	Provides insights into public perceptions and behaviors towards food related issues, such as sustainable food practices or governmental food policies or programs.	Agency; Sustainability	Policy formulation & decision-making	Social media data; News articles	Information extraction & text mining	2-3	13
Knowledge generation and management from food policy and program documents	Generates new knowledge and understanding by synthesizing information from a variety of documents, e.g., food policy documents. This application area also encompasses the development of knowledge management systems, which focus on capturing, storing, and disseminating knowledge.	Availability; Access; Stability; Utilization; Agency; Sustainability	Evaluation	Documents	Information extraction & text mining; Text generation	3-5	10
Understanding dietary habits	Provides insights into the nutritional quality of consumed diets. Studies analyze, e.g., food-related social media content, which could inform whether public health interventions, such as strategies for preventing overweight, are needed.	Utilization; Agency	Agenda setting; Evaluation	User-generated content	Information extraction & text mining	2-3	8
Food item classification	Automates the classification of food categories, nutrition quality, and processing levels in food composition datasets, thereby improving their quality. This contributes to enhancing the understanding of nutritional quality of diets as well as energy intake.	Utilization	Policy formulation & decision-making	Food composition datasets	Information extraction & text mining	4	4
Addressing data gaps in food security statistics and crisis response	Addresses data gaps, improves data quality, and utilizes non-conventional data sources for better decision-making in food security policymaking and crisis response.	Availability	Agenda setting	Social media data; Documents	Information extraction & text mining; Sentiment analysis	3	4

Note: Application areas are ordered by size. The listed data items represent the most common ones for each application area and are not exhaustive. Three studies did not fall within any of the identified application areas and are therefore not presented in this section. For a detailed overview of the studies in each application area, including the extracted data items, please refer to Appendix B.1.

Across the 60 reviewed studies, six application areas of NLP for food security policymaking were identified: Early warning systems for food insecurity (Section 4.1); Understanding public discourse (Section 4.2); Knowledge generation and management from food policy and program documents (Section 4.3); Understanding dietary habits (Section 4.4); Food item classification (Section 4.5); and Addressing data gaps in food security statistics and crisis response (Section 4.6). A brief overview of these application areas is presented in Table 3 and they are described in detail in the following sections.

4.1 Early warning systems for food insecurity

The largest application area in the reviewed studies was the use of NLP to predict or forecast a certain food security status, providing valuable input for the development of early warning systems for impending food crises (n=18). Studies that directly focus on this potential included Ahn et al. (2023), Balashankar et al. (2023), de Brito et al. (2020), Dunnmon et al. (2019), Goetz et al. (2023), Kontar et al. (2023), Lukyamuzi et al. (2018, 2020), Molenaar et al. (2024), and Wanrooij et al. (2024). Other studies offered contextual explanation for a certain food security status (Ba et al., 2022; De Simone & Mongeau, 2023; Deleglise et al., 2023; N. M. Martin, Poirier, et al., 2022), predicted food price inflation (J. Kim et al., 2017; Lv et al., 2022; Silva e Silva et al., 2024), or integrated qualitative data into computer simulations of household food security (Paudel & Ligmann-Zielinska, 2023).

Access, stability, and availability were the most common food security dimensions in this application area (n=16, n=15, n=12, respectively). Often this was linked to the agenda setting policy cycle stage (n=17). These studies analyzed a problem situation and its magnitude, informing the urgency of placing it on the political agenda. Food security early warning systems often considered availability of and access to food, while monitoring these dimensions over time (i.e. stability).

News articles and social media data were the most common data types utilized by these studies (both n=9). Social media studies primarily used Twitter data[‡]. The two most common NLP subfields were information extraction & text mining (n=14) and sentiment analysis (n=9). Within the former, topic modeling was a common task (n=8). For example, Ahn et al. (2023) introduced HungerGist, a multi-task deep learning model that predicts food security based on news articles. They employed topic modeling (i.e. Latent Dirichlet Allocation, LDA) to provide contextual insights around the predicted food crisis index. As they used text inference to interpret the "gists", i.e. interpretable signals, in news articles, the study also employed reasoning techniques. Similarly, Wanrooij et al. (2024) utilized BERTopic to extract contextual information from news articles. These extracted topic features, combined with traditional food insecurity features, were used to forecast food security status. An example of sentiment analysis is the study by Deleglise et al. (2023), who used the French version of the Valence Aware Dictionary and Sentiment Reasoner model to assess negativity in news articles about food security, providing insights into how urgently food insecurity is portrayed in the media.

In terms of TRLs, most of the studies were in the research or development phases, ranging from TRL2 to TRL5. Two studies conducted under FAO's Data Lab initiative have reached the deployment stage (TRL9) (De Simone & Mongeau, 2023; Silva e Silva et al., 2024). The first, De Simone and Mongeau (2023), utilized a multilingual dataset of tweets from global newspaper accounts to identify topics and calculate a sentiment index by country, and results are accessible in FAO's online Topic Explorer tool. Silva e Silva et al. (2024) developed a model to estimate food consumer price inflation and a daily food price monitor for key food commodities. They used a combination of food price information from Numbeo (a platform with e.g., crowd-sourced cost of living data), oil prices, exchange ranges, and news articles collected via Twitter. Both tools are available online in FAO's Food Prices tools.

4.2 Understanding public discourse and sentiment on food related issues

The second NLP application area was understanding public discourse and sentiment (n=13), which can be useful for policymakers to understand public priorities, identify concerns, and to assess perceptions of policy interventions. Specifically, Ashtab and Campbell (2021), Brzustewicz and Singh (2021), Cooper et al. (2022), Kil et al. (2023), Krismawati and Panuntun (2023), Matsuoka et al. (2023), Park and Shin (2021), and Singh and Glińska-Neweś (2022) focused on understanding public perceptions and behaviors towards sustainable food practices, such as veganism, organic foods, novel protein sources, local food networks, and

[‡]Twitter was rebranded as "X" in 2023. Since the studies utilizing Twitter/X data were published prior to this rebranding, this paper uses the term "Twitter" to maintain clarity and consistency with pre-rebranding research and references.

smart or local food production. The gained insights can guide the development of policies and programs that align with consumer values and promote sustainable production and consumption, which is essential for fostering regenerative food systems. Additionally, Kumar and Sharma (2020), Lindquist et al. (2021), and Scott et al. (2018) utilized NLP to understand public comments and sentiment regarding government policies or programs. By gathering public feedback, these studies provide valuable input for the creation of new policies or the adjustment of existing ones. Furthermore, Bagheri et al. (2023) analyzed social media posts of the farming community to enhance agricultural knowledge discovery, and Benites-Lazaro et al. (2018) used NLP to understand social policy debates around energy production in Brazil and its relationship with food security and climate change.

The majority of studies focused on the food security dimensions of agency and sustainability (n=11 and n=10, respectively). These studies typically analyzed people's perceptions of various foods, diets, or production methods that have a sustainability component. For example, Matsuoka et al. (2023) examined the cultural acceptance of novel foods, such as cultured meat and insects, by analyzing news articles collected from Google News. The most common policy cycle stage was policy formulation & decision-making (n=10). This links to the argument made earlier that understanding public discourse and sentiment can provide valuable input for drafting new policies or adjusting existing ones. For example, the study by Lindquist et al. (2021) presented how NLP can be used to analyze public comments on the 2020 United States dietary guidelines as part of the policymaking process.

Nine studies used social media data, primarily Twitter. News articles were used as data source for four studies. Additionally, two studies complemented news articles or social media data with other types of data, such as government or company reports (Benites-Lazaro et al., 2018), scientific literature (Kil et al., 2023), or voting records (Scott et al., 2018). The study by Lindquist et al. (2021) used public comments as primary data source.

Similar to the previous section, information extraction & text mining and sentiment analysis were the most common NLP subfields (n=10 and n=8, respectively). Within the former, topic modeling was a commonly used technique (n=7). For example, Benites-Lazaro et al. (2018) used topic modeling in their study on social policy debates in Brazil. Specifically, LDA was applied to text from news articles, reports from non-governmental organizations, government websites, and reports from private sector to identify key themes and keywords. These themes and keywords were categorized by actor and tracked over time, allowing to pinpoint the political stances of the various actors in discussions related to energy, food security, and climate change. As LDA was complemented by discourse analysis to explore underlying themes and contexts, as well as debates and concerns, the study also fitted into the reasoning NLP subfield.

Almost all studies were in the research TRL phase, varying between TRL2 and TRL3. The study by Scott et al. (2018), however, has entered deployment stage (TRL9). This study explored public opinion on the Supplemental Nutrition Assistance Program, a federal food banking program in the United States. The authors analyzed perceptions of the program over time across multiple data sources, including Twitter, news outlets, and elected representatives' voting records on food insecurity-related bills. An online application has been developed that allows the involved food bank to better understand reporting and public opinion on the program, which should help enhance communication about the program.

4.3 Knowledge generation and management from food policy and program documents

Knowledge generation and management from policy and program documents is another NLP for food security policymaking application area (n=10). Specifically, studies have employed customized LLM-based chatbots to summarize and synthesize food policy documents (Benson, 2023; M. Kim et al., 2021). Other studies have utilized information extraction & text mining techniques for multi-country policy analysis (Fabi et al., 2023; Galsurkar et al., 2018; Juventia et al., 2020) or to accelerate thematic knowledge generation from four decades of program documentation (Garbero et al., 2021). Additionally, Aitken et al. (2022), Franzen et al. (2022), and Min et al. (2019) focused on knowledge generation for policy and program evaluation, and P. Martin et al. (2021) have developed a customized knowledge management and knowledge generation pipeline.

All studies in this area relied on documents, such as policy documents or program reports, as their primary data source. Given the broad nature of these documents, most studies encompassed all dimensions of food security (n=7), while the remaining three studies focused specifically on availability, access, sustainability or a combination thereof. When analyzing the policy cycle stages, evaluation emerged as the predominant stage (n=8). For example, Franzen et al. (2022) explored whether AI can improve the efficiency of content analysis to inform the evaluation of a World Bank program on stunted growth and chronic malnutrition, concluding that AI holds promising potential for evaluation synthesis purposes. Additionally, Galsurkar et al. (2018) used NLP to assess country-level national development plans for alignment with the SDGs, including SDG 2 Zero Hunger, which significantly accelerated this type of analysis, known as Rapid Integrated Assessment, compared to manual review alone. Five studies addressed policy formulation & decision-making applications, including the two that utilized LLM-based chatbots (Benson, 2023; M. Kim et al., 2021). These studies explored whether LLMs can accelerate the availability and processing of relevant information and knowledge, thereby improving the efficiency of text synthesis and supporting evidence-based policymaking. While both studies recognized the benefits for rapid policy analysis, they also highlighted notable limitations, such as difficulties in tailoring outputs to specific scenarios and accurately prioritizing policy recommendations by importance.

All studies employed information extraction & text mining techniques to derive and manage knowledge from unstructured textual sources. Additionally, the two studies involving LLM-based chatbots (Benson, 2023; M. Kim et al., 2021) are also in the text generation and natural language interfaces NLP subfields. One of these chatbots used a Retrieval Augmented Generation capability[§], integrating information retrieval methods.

Most studies in this application area were in the research or development TRL phases, ranging from TRL3 to TRL5. Two studies have reached the deployment stage (TRL9). The first is the Knowledge ExtractOr Pipeline System presented by P. Martin et al. (2021), for which preliminary results have been presented to end-users in the Long-term EU-AU Research and Innovation Partnership for Food and Nutrition Security and Sustainable Agriculture project, and which should aid the analysis of project documentation in the related database. Furthermore, Fabi et al. (2023) developed a tool to synthesize country-level "National Pathways" that outline national agri-food system transformation ambitions, online accessible in FAO's Food Systems Summit Analysis dashboard.

4.4 Understanding dietary habits

The fourth application area focused on understanding dietary habits (n=8; (Abbar et al., 2015; De Choudhury et al., 2016; Dondokova et al., 2019; Fried et al., 2014; Hansen & Hershcovich, 2022; Huangfu & Zeng, 2018; Sharma & De Choudhury, 2015; West et al., 2013)). Studying people's food consumption patterns provides insights into the nutritional quality of their diet. This information can help address the utilization aspect of food security and inform public health policies, e.g., those related to issues like obesity or the limited access to healthy and nutritious food in so-called food deserts.

As the majority of studies analyzed food intake represented by food-related social media posts, these studies related to both the utilization and agency dimension of food security (n=7). One study disaggregated food intake for food deserts and non-food deserts (De Choudhury et al., 2016), and hence also touched upon the access and availability dimensions of food security. Another study focuses on sustainable diets specifically (Hansen & Hershcovich, 2022), hence was categorized in the sustainability dimension. Most studies fitted

[§]Retrieval Augmented Generation is the method of generating text based on retrieved documents (Jurafsky & Martin, 2024). These retrieved documents enhance the accuracy and reliability of generated text.

into multiple policy cycle stages. The agenda setting stage was most common (n=7), followed by evaluation (n=4). Studies in the former category, e.g., inform policymakers whether public health interventions related to food consumption, such as strategies for preventing overweight, are needed. NLP applications presented in the evaluation category can serve as a monitoring tool for the development of dietary intake over time.

All studies used user-generated content as data source. Social media data was the most common source of data (n=7), where five studies analyzed food-related posts on Twitter and two on Instagram. Commonly, the posts analyzed were about what people consumed, when, where and why. Furthermore, West et al. (2013) used web logs of people searching for recipes online, along with the corresponding nutritional information of the recipe, to examine dietary patterns with a specific focus on sodium intake.

Information extraction & text mining was a relevant NLP subfield for almost all studies in this application area (n=7). For example, NLP was used to extract specific food names from social media posts and to link this to caloric or nutritional value of food intake (Abbar et al., 2015; De Choudhury et al., 2016; Sharma & De Choudhury, 2015). Other common tasks within this subfield included classification and topic modeling. For the latter, an example is the study by Fried et al. (2014), who used LDA to identify topics from food tweets, which were then used as features to predict population characteristics like overweight rate, diabetes rate, and political leaning.

All studies in this application area were classified into the research phase (TRL2 - TRL3).

4.5 Food item classification

The use of NLP offers an opportunity to enhance the quality of food composition datasets, particularly in situations where data is scarce or of low quality (Pretorius et al., 2023). Four of the reviewed studies focus on this potential. Specifically, Hu, Ahmed, et al. (2023) presented mechanisms to automate food category classification and nutrition quality score prediction for food composition databases, which usually contain food label information, such as product name, ingredients, and nutritional facts, for food products in a certain country. Similarly, Hu, Flexner, et al. (2023) automated the classification of food processing levels for food composition databases. Eftimov et al. (2017) presented a semi-automated system for classifying and describing food items according to the standardized food classification and description system by the European Food Safety Authority, FoodEx2. This system can be used, e.g., to fill missing data gaps in food composition data. Lastly, Youn et al. (2023) constructed a food composition knowledge base from scientific literature.

High-quality food composition data can be helpful for understanding the nutritional quality of diets as well as energy intake, which contributes to the utilization dimension of food security: all four studies were categorized into this dimension. Furthermore, all studies fell into the policy formulation & decision-making stage of the policy cycle, owing to the role of food composition data in nutrition research, informing nutritional or dietary guidelines, or developing public health policies and strategies.

Information extraction & text mining was the main NLP subfield for all four studies. Tasks used included e.g., text classification and relation extraction. Three studies used BERT-based language models (i.e. BioBERT and Sentence-BERT), thereby leveraging the advanced contextual understanding these models provide.

All studies in this application area were classified as TRL4, marking the first step of the development phase.

4.6 Addressing data gaps in food security statistics and crisis response

The last identified application area focuses on using NLP to address data gaps, improve data quality, and utilize non-conventional data sources for better decision-making in food security policymaking and crisis response. Four studies fitted into this application area, and the exact use case varied per study. For example, Fabi et al. (2022) described how FAO fills data gaps in official statistics. They presented several applications, including the use of web scraping to enhance agricultural production data. In some cases, Optical Character Recognition was employed to digitize text from scanned documents, which was then standardized in terms of commodities and administrative units using information extraction & text mining techniques. The approach helped the FAO fill data gaps and acted as a quality control mechanism by validating reported data. Additionally, Taglioni et al. (2023) described a text-mining tool that searches, downloads, classifies and analyses articles, reports and grey literature documents. Information extracted from these documents included food loss data, in addition to metadata, keywords used in identifying the loss factors for countries and commodities, and a short summary of the document. The studies by Ragini et al. (2018) and Braley et al. (2021) explored the use of Twitter data for timely service delivery in data-scarce environments. Specifically, Ragini et al. (2018) utilized Twitter data to identify the needs of people during disaster situations and analyzed related sentiments. These insights can assist humanitarian actors in delivering more timely and effective support, including food aid. Additionally, the model can be applied during the recovery phase to gather feedback on the public's perception of the emergency response. Similarly, Braley et al. (2021) leveraged Twitter data to gain insights into service delivery priorities.

All studies focused on the availability dimension of food security. Additionally, two studies addressed the access dimension, and one addressed sustainability. The studies could be applied to various stages across the policy cycle, including agenda setting (n=3), policy formulation & decision-making (n=1), implementation (n=2), and evaluation (n=2).

All studies employed information extraction & text mining techniques. Additionally, Ragini et al. (2018) and Braley et al. (2021) used sentiment analysis to interpret the topics identified through topic modeling, and Taglioni et al. (2023) made use of information retrieval techniques.

The Food Loss and Waste Database by Taglioni et al. (2023) is continuously updated, publicly accessible online, and presented in a visual and interactive manner, and is therefore argued to be deployed and classified as TRL9. This is also the case for FAO's approach to filling data gaps in official statistics (Fabi et al., 2022). The other two studies presented an empirical proof of concept (TRL3).

5 Discussion

Main findings

This scoping review aimed to map and identify key application areas of NLP for food security policymaking, advancing the progress towards ending hunger and achieving food security and improved nutrition for all. Six application areas were identified: 1) Early warning systems for food insecurity; 2) Understanding public discourse; 3) Knowledge generation and management from food policy and program documents; 4) Understanding dietary habits; 5) Food item classification; and 6) Addressing data gaps in food security statistics and crisis response.

In addition to mapping existing NLP initiatives, this paper aimed to identify gaps for future research and application areas. A key observation is that several of the identified application areas (specifically, areas 2, 4, and 5) focus predominantly on improving nutrition and promoting sustainable diets, aligning with the utilization, agency, and sustainability dimensions of food security. In contrast, only the first application area focuses primarily on the dimensions of availability, access, and stability, which are critical for food sufficiency in food-insecure regions. This reflects an uneven emphasis on some dimensions of food security, leaving gaps in addressing critical challenges related to others.

This observation links to the critical issue for any NLP application to have access to high quality data. In data scare contexts, NLP opportunities may be limited or of lower performance. There is regional variation in the availability of data on SDG 2 indicators (UN DESA, 2024b), and for many indicators disaggregated data is lacking (UN DESA, 2024c). This may be extended to data availability in general, following from the regional disparities in internet usage (World Bank, 2022). This raises the question of whether insufficient data might cause NLP for food security policymaking, and data-driven innovation more broadly, to inadvertently overlook regions where food insecurity is most severe. Being aware of these disparities is the first step in designing equitable NLP for food security policymaking projects. Furthermore, as discussed in the sixth application area, NLP may at least partly address these data gaps through its ability to process and analyze unstructured data.

Towards successful implementation

For NLP projects to have tangible impact, successful and sustained implementation is essential. However, as demonstrated in the previous section, the majority of NLP for food security policymaking projects have remained within academic research labs and have not been implemented in practice. Specifically, 88% of studies were classified within the research and development TRL phases, rather than deployment, and 73% lacked explicit mentioning of collaboration with potential implementation partners. While this may partly be attributed to the nascent state of NLP for food security policymaking research and the inclusion of academic databases in the search strategy[¶], it also reflects a broader trend in the AI4SG literature, where in most studies the connection to real-world implementation is still limited or non-existent (Shi et al., 2020). Similarly, Sarku et al. (2023) reviewed AI models for food security with an emphasis on local stakeholder involvement. They found that only a small number of studies incorporated stakeholder feedback or facilitated the implementation of findings. The insufficient stakeholder engagement and implementation efforts were attributed to challenges in research practices, such as the emphasis on model outputs for publications, the short time frames of research consortia, and funding issues (Sarku et al., 2023). Furthermore, they found that in most of the reviewed studies, the research was initiated by the researchers rather than the problem owners. This highlights the need to align AI research in food security with the priorities and needs of stakeholders. This alignment, along with contextual intelligence, is essential for ensuring that the resulting projects are both relevant and effective for the communities they are intended to support.

Furthermore, to ensure that technological development aligns with the needs and demands of domain experts, deep partnerships from the outset are crucial for AI4SG projects, including those involving NLP. Tomašev et al. (2020) have developed guidelines for establishing successful collaborations between AI researchers and domain experts working on SDGs, ranging from the overall use of AI technology and its applications, to data handling. Examples of guidelines include AI applications "to be inclusive and accessible, and reviewed at every stage for ethics and human rights compliance", "establishing and maintaining trust", and "improving data readiness" (Tomašev et al., 2020, p. 3).

In addition to the generally limited collaboration, several technical challenges hinder deployment. Baier et al. (2019) identified common challenges in machine learning deployment and operation, which can be extended to NLP deployment in food security policymaking. These challenges range from traditional pre-deployment challenges such as poor data quality, data preprocessing and data management, to deployment challenges like ICT infrastructure set-up and the manual labor involved in ongoing model validation and handling data drifts. Indeed, the limited availability of high-quality, trustworthy data was mentioned by some studies in this review as potentially compromising model outputs and their usability (De Choudhury et al., 2016; Deleglise et al., 2023; Fried et al., 2014; Singh & Glińska-Neweś, 2022; Taglioni et al., 2023). Non-technical challenges, including expectation management, trust, transparency, and explainability are also critical factors potentially hindering deployment (Baier et al., 2019). For example, Silva e Silva et al. (2024) noted that their tool does not directly explain the model outputs, which they consider a potential limitation.

More broadly, enhancing data utilization within food security policymaking institutions is essential for improving data-informed decision-making (HLPE-FSN, 2022). This begins by raising policymakers' awareness of the value that data brings to decision-making. It is also important to harmonize and maximize the sharing of existing data to ensure that all stakeholders have access to comprehensive and up-to-date informa-

[¶]The average TRL was higher for grey studies (6.0) compared to both academic (3.9) and conference studies (3.2). This is not surprising, as 86% of grey studies were conducted in partnership with an implementation partner, compared to only 22% and 6% of academic and conference studies, respectively. Refer to Appendix C for a more elaborate comparison of study characteristics by study type.

tion, and avoid duplicate data collection efforts. Increased and sustained investment in collecting sufficiently disaggregated, granular data is needed to ensure that those most affected by inequalities are properly represented. Additionally, investing in human capital, particularly in data science and statistics skills, is critical for building the institutional capacity to analyze and interpret this data effectively. Lastly, improving data governance and promoting inclusiveness and agency across all relevant data systems are crucial for giving voice to those most affected by food security and nutrition policies, providing an accountability mechanism, and addressing ethical concerns related to power imbalances in data ownership and control. Improving inclusiveness and agency is especially critical given that data-informed decision-making is often influenced by existing power structures and is rarely a neutral process.

Ethical considerations

For NLP projects in food security policymaking to contribute to advancing food security in a socially responsible and sustainable way, they must be fair, inclusive and accessible. This makes it essential to reflect on ethical considerations throughout the development of such projects. However, in this review, only 25% of the studies explicitly discussed ethics, bias, or fairness. The most frequently mentioned concern was potential under-representation in the data (n=8), particularly due to self-selection bias and reliance on platforms like Twitter, which represent only a subset of the population. Additionally, limiting data to English-language content was mentioned as another issue negatively impacting representation (n=4). More specifically, two studies highlighted implications arising from algorithms being developed and trained predominantly in languages like English, which could lead to implementation or performance issues when applied in other contexts, especially when working with low-resourced languages. For example, Braley et al. (2021) discussed the link between data-scarce environments and the presence of underrepresented languages which are not easily translated, emphasizing the difficulties this poses for utilizing regular NLP models. Three studies mentioned ethical considerations related to anonymity and privacy of data. Another ethical concern raised was the lower algorithm performance for different groups of people (n=1), with Elsweiler and Harvey (2015) highlighting that their algorithm generated less effective meal plans for men and for individuals with less diverse dietary preferences. One study mentioned the climate costs associated with running LLMs, and the potential risk of unintended dual usage of model results (Hansen & Hershcovich, 2022). Benson (2023) highlighted the risk of introducing nuanced biases that hinder innovation when deploying LLMs trained on existing data reflecting the status quo. Lastly, Wanrooij et al. (2024) cautioned against the direct integration of model predictions in sensitive decision-making, as they often come with high margins of error and pertain to vulnerable groups. This underscores the need for careful consideration and validation before incorporating such models into high-stakes environments.

Reflecting on ethics is especially important in the context of NLP for food security policymaking, because of the geographical and cultural disparities between where these technologies are developed and where food insecurity is most severe. Specifically, food insecurity is most acute in the African continent, Afghanistan, Haiti, and Yemen (FAO, 2022), while the NLP market size is dominated by countries like the United States and China (Statista, 2024)[‖]. Similarly, the majority of authors in this review were affiliated with institutions in the United States and Europe. This imbalance may have ethical implications, as the dominance of a certain culture and related values within a study field may make it difficult to recognize critical issues when applied to other contexts, or even worse, establish an exclusive and biased environment within that field (Shi et al., 2020). This cultural dominance is exacerbated by the fact that internet access is not evenly distributed. Specifically, language models in NLP are often trained on publicly available data from the internet that underrepresents food-insecure countries (World Bank, 2022). This leads to biased training data, which is problematic because it may perpetuate dominant viewpoints, increase power imbalances, and exacerbate inequalities (Bender et al., 2021). These concerns are even more critical when research is conducted without

[‖]The market size is predicted building on financial reports, funding data, third-party data, and key market indicators such as GDP, number of internet users, number of secure internet serves, and internet penetration.

sufficient stakeholder collaboration. Therefore, this paper advocates for deliberate efforts in any NLP for food security policymaking project to establish authentic partnerships with stakeholders, starting in the initial phase of project development when defining the problem. While doing so, researchers should be cautious about potential power imbalances following from the funding structure involved. Furthermore, employing a Value Sensitive Design approach (Simon et al., 2020) can further enhance the development of responsible and value-aligned NLP systems that consider the diverse values of stakeholders and that mitigate the risk of algorithmic biases. This approach could also deliberately account for values such as privacy (van den Hoven et al., 2020).

A key concern when applying NLP in decision-making settings is the cost of incorrect predictions, which can lead to misguided policies and actions that negatively affect communities or incur substantial costs. To mitigate this risk, Wanrooij et al. (2024), e.g., recommend using their LLM-based prediction model only as a triangulation tool during human-led investigations, rather than integrating it directly into sensitive decision-making. More generally, this reflects human-AI collaboration, where AI supports but does not replace human judgment. Additionally, explainability and transparency are crucial for policymakers to trust NLP models and understand their outputs. These are part of internationally accepted AI governance principles and help mitigate the risks of unregulated AI (Truby, 2020).

Lastly, Floridi et al. (2021) offers practical guidance for designing ethical NLP projects for food security policymaking, identifying seven factors and related best practices. These factors align with broader AI ethics principles (i.e. nonmaleficence, autonomy, justice and explicability), and range from falsifiability and incremental deployment to human-friendly semanticisation. Incorporating these factors into NLP for food security policymaking project design ensures that such projects not only contribute to advancing food security but also do so in a socially responsible and sustainable manner.

Strengths and limitations

This scoping review is, to our knowledge, the first to focus specifically on the intersection of NLP and food security policymaking, addressing a gap in both the NLP4SG research field and food security policy domain. A key strength of this review is its inclusion of both academic and grey literature, thereby providing a practical grounding. Furthermore, the review experimented with NLP tools to enhance the literature review process itself, demonstrating the practical utility of NLP in handling large volumes of unstructured data as a complementary outcome of the review.

This review has some limitations, including potential selection bias in the study selection process. For example, to keep search results manageable, the concept of food security was operationalized without explicitly including "agriculture", which may have excluded NLP applications for food availability by design. Readers are referred to Drury and Roche (2019) and Lu et al. (2023) for an overview of text mining and Artificial General Intelligence for agriculture, respectively. Additionally, the search query primarily focused on text-based NLP applications. While this emphasis reflects NLP's core function to process, understand, and generate human language in textual form, it may have excluded relevant studies involving other data types, especially multimodal data. As multimodality is expected to be one of the NLP subfields to characterize the NLP research landscape in the coming years (Schopf et al., 2023), future reviews should consider incorporating a broader range of data types to fully capture the evolving scope of NLP, especially considering the potential of geospatial data in the agricultural sector (Varriale et al., 2024).

Lastly, only English studies were considered for this review, which may have introduced representation bias and exacerbated the geographical imbalance discussed earlier. However, this was a supply- and accessibility-driven decision, as the availability of studies is overwhelmingly higher in English (e.g., for both Scopus and Web of Science 96% of records were English).

Future outlook and recommendations

The findings of this review underscore the promising role NLP can play in advancing food security policymaking, while also revealing several key gaps that present opportunities for future research and application areas. Based on these insights, the following general recommendations for NLP for food security policymaking are proposed:

- **Advance NLP techniques for data scarce regions:** Future efforts should prioritize NLP projects for regions facing the most severe food insecurity but struggling with data scarcity, as NLP holds potential for processing and analyzing unstructured data in these contexts. This includes developing NLP models that are capable of handling low-resource languages.
- **Establish authentic partnerships:** By collaborating closely with stakeholders from as early as project ideation, researchers can foster ownership and alignment of their work with stakeholders' needs and values. This collaborative approach not only enhances relevance, but also facilitates the transition to successful and sustained implementation of NLP projects.
- **Invest in data awareness and technical capacity:** To fully leverage the potential of NLP in food security policymaking, it is essential to enhance data utilization within relevant institutions. This involves raising data awareness, enhancing data quality and accessibility, and fostering a data culture that emphasizes the value of data for informed decision-making and that promotes inclusiveness and agency across all relevant data systems. Additionally, strengthening the technical capacity of these institutions will be crucial for the sustainability and effectiveness of NLP for food security policymaking projects.

Additionally, two more specific suggestions for future research and application areas include:

- **Scale and deploy NLP-enhanced early warning systems for food insecurity:** Developments in NLP-enhanced early warning systems have shown promising potential to address key food security dimensions (i.e., availability, access, and stability), enabling faster and more proactive responses to food insecurity. Yet, to maximize their impact, it is essential to refine and scale these systems and ensure their deployment and implementation in collaboration with practitioners. While doing so, NLP could also help addressing data gaps in early warning systems.
- **Enhance policy analysis and evaluation with NLP:** Given its ability to efficiently analyze large volumes of unstructured data from sources such as reports, policy documents, stakeholder feedback, and citizen science, NLP could significantly contribute to policy analysis and program and policy evaluation, guiding timely and evidence-based adaptive management.

6 Conclusion

This scoping review has mapped key NLP application areas for food security policymaking, identifying both opportunities and gaps for future research and applications. Despite the potential of these applications, limited deployment and implementation poses a challenge to achieving real-world impact. Establishing authentic partnerships from the outset will be essential to promote successful and sustained implementation of NLP projects that could contribute to advancing progress towards ending hunger and achieving food security and improved nutrition for all.

Author contributions

Conceptualization: MM; Data curation: MM (FC second reviewer for study selection; CvdL and EK second and third reviewer for data extraction); Formal analysis: MM; Funding acquisition: FC; Investigation: MM; Methodology: MM, FC, CvdL, EK; Project Administration: FC; Software: MM; Supervision: FC, CvdL, EK; Visualization: MM; Writing - original draft: MM; Writing - review and editing: MM, FC, CvdL, EK.

Acknowledgments

This study was funded by a collaboration between Kickstart AI and the Zero Hunger Lab of Tilburg University. Kickstart AI aims to accelerate the adoption of AI in the Netherlands. Zero Hunger Lab aims to contribute to global food and nutrition security through data science. Additionally, we would like to thank Madelon Meijer, Policy Advisor Agriculture and Food at Oxfam Novib, for the helpful discussion and feedback on the preliminary results and recommendations.

Conflict of interest

The authors and the funding organizations declare to have no conflict of interest.

Supporting information

The supporting information is available on the GitHub page of the first author.

References

Abbar, S., Mejova, Y., & Weber, I. (2015). You tweet what you eat: Studying food consumption through Twitter. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3197–3206. https://doi.org/10.1145/2702123.2702153

Adauto, F., Jin, Z., Schölkopf, B., Hope, T., Sachan, M., & Mihalcea, R. (2023). Beyond good intentions: Reporting the research landscape of NLP for social good. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 415–438. https://doi.org/10.18653/v1/2023.findings-emnlp.31

Ahn, Y., Yan, M., Lin, Y.-R., & Wang, Z. (2023). HungerGist: An interpretable predictive model for food insecurity. *2023 IEEE International Conference on Big Data (BigData)*, 1591–1600. https://doi.org/10.1109/BigData59044.2023.10386346

Aitken, J. A., Rao, D. W., Alaybek, B., Sprenger, A., Mika, G., Hartman, R., & Leets, L. (2022). AI-based text analysis for evaluating food waste policies. *AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges*.

Ashtab, S., & Campbell, R. (2021). Explanatory analysis of factors influencing the support for sustainable food production and distribution systems: Results from a rural Canadian community. *Sustainability*, *13*(9). https://doi.org/10.3390/su13095324

ASReview. (2024a). Active learning explained. https://asreview.nl/blog/active-learning-explained/

ASReview. (2024b). Navigating the maze of models in ASReview. https://asreview.nl/blog/asreview-model-selection-guide/

ASReview LAB developers. (2024). *ASReview LAB - A tool for AI-assisted systematic reviews (v1.6.3rc0)*. Zenodo. https://doi.org/10.5281/zenodo.11185216

Ba, C. T., Choquet, C., Interdonato, R., & Roche, M. (2022). Explaining food security warning signals with youtube transcriptions and local news articles. *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, 315–322. https://doi.org/10.1145/3524458.3547240

Bagheri, A., Taghvaeian, S., & Delen, D. (2023). A text analytics model for agricultural knowledge discovery and sustainable food production: A case study from Oklahoma Panhandle. *Decision Analytics Journal*, *9*, 100350. https://doi.org/10.1016/j.dajour.2023.100350

Baier, L., Jöhren, F., & Seebacher, S. (2019). Challenges in the deployment and operation of machine learning in practice. *Twenty-Seventh European Conference on Information Systems (ECIS2019)*.

Balashankar, A., Subramanian, L., & Fraiberger, S. P. (2023). Predicting food crises using news streams. *Science Advances*, *9*(9), eabm3449. https://doi.org/10.1126/sciadv.abm3449

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Benites-Lazaro, L., Giatti, L., & Giarolla, A. (2018). Topic modeling method for analyzing social actor discourses on climate change, energy and food security. *Energy Research & Social Science*, *45*, 318–330. https://doi.org/10.1016/j.erss.2018.07.031

Benson, T. (2023). *Exploring the potential of customized AI chatbots in food policy research: Capabilities and constraints in comparative perspective* (tech. rep.). CGIAR Research Initiative on Digital Innovation.

Braley, A., Fraiberger, S. P., & Taş, E. O. (2021). *Using Twitter to evaluate the perception of service delivery in data-poor environments* (tech. rep.). World Bank Group. https://doi.org/10.2139/ssrn.4029916

Brzustewicz, P., & Singh, A. (2021). Sustainable consumption in consumer behavior in the time of covid-19: Topic modeling on Twitter data using LDA. *Energies*, *14*(18), 5787. https://doi.org/10.3390/en14185787

Cooper, K., Dedehayir, O., Riverola, C., Harrington, S., & Alpert, E. (2022). Exploring consumer perceptions of the value proposition embedded in vegan food products using text analytics. *Sustainability*, *14*(4), 2075. https://doi.org/10.3390/su14042075

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence*, *3*(2), 111–115. https://doi.org/10.1038/s42256-021-00296-0

De Choudhury, M., Sharma, S., & Kiciman, E. (2016). Characterizing dietary choices, nutrition, and language in food deserts via social media. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1157–1170. https://doi.org/10.1145/2818048.2819956

De Simone, L., & Mongeau, C. (2023). Bridging agricultural data gaps: Innovations in geospatial and non-conventional data sources. *RAF/AFCAS/23-E-105*.

de Brito, M. M., Kuhlicke, C., & Marx, A. (2020). Near-real-time drought impact assessment: A text mining approach on the 2018/19 drought in Germany. *Environmental Research Letters*, *15*(10), 1040a9. https://doi.org/10.1088/1748-9326/aba4ca

Deleglise, H., Begue, A., Interdonato, R., d'Hotel, E. M., Roche, M., & Teisseire, M. (2023). How can text mining improve the explainability of food security situations? *Journal of intelligent information systems*. https://doi.org/10.1007/s10844-023-00832-x

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dondokova, A., Aich, S., Kim, H.-C., & Huh, G. H. (2019). A text mining approach to study individuals' food choices and eating behavior using Twitter feeds. *Frontier Computing*, 520–527. https://doi.org/10.1007/978-981-13-3648-5_60

Drury, B., & Roche, M. (2019). A survey of the applications of text mining for agriculture. *Computers and Electronics in Agriculture*, *163*, 104864. https://doi.org/10.1016/j.compag.2019.104864

Dunnmon, J., Ganguli, S., Hau, D., & Husic, B. (2019). Predicting us state-level agricultural sentiment as a measure of food security with tweets from farming communities. *arXiv preprint arXiv:1902.07087*. https://doi.org/10.48550/arXiv.1902.07087

EC. (2024). Horizon Europe - Work Programme 2023-2025: General Annexes.

Eftimov, T., Ispirova, G., Seljak, B. K., & Korošec, P. (2017). A semi-automatic system for classifying and describing foods according to FoodEx2. *3rd IMEKOFOODS: Metrology promoting Standardization and Harmonization in Food and Nutrition*.

Elsweiler, D., & Harvey, M. (2015). Towards automatic meal plan recommendations for balanced nutrition. *Proceedings of the 9th ACM Conference on Recommender Systems*, 313–316. https://doi.org/10.1145/2792838.2799665

Fabi, C., Mongeau Ospina, C. A., Rosero Moncayo, J., & Silva E Silva, L. G. (2022). The FAO Data Lab on statistical innovation and the use of big data for the production of international statistics [Publisher: IOS Press BV Type: Article]. *Statistical Journal of the IAOS*, *38*(3), 995–1007. https://doi.org/10.3233/SJI-220052

Fabi, C., Ospina, C. A. M., Gerits, H., & Torero, M. (2023). Synthesizing food system summit national pathways: A methodological approach. *Proceedings of the Ninth International Conference on Agricultural Statistics (ICAS IX)*.

FAO. (2022). FAOSTAT: suite of food security indicators (2021-2023). https://www.fao.org/faostat/en/#data/FS

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2021). How to design AI for social good: Seven essential factors. *Ethics, Governance, and Policies in Artificial Intelligence*, 125–151. https://doi.org/10.1007/s11948-020-00213-5

Franzen, S., Quang, C., Schweizer, L., Budzier, A., Hrstich, P., Reissfelder, S., Gold, J., Vellez, M., Ramirez, S., & Raimondo, E. (2022). *Advanced content analysis: Can artificial intelligence accelerate theory-driven complex program evaluation?* (Tech. rep.). World Bank Group. Washington DC, USA, World Bank Group.

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2014). Analyzing the language of food on social media. *2014 IEEE International Conference on Big Data (Big Data)*, 778–783. https://doi.org/10.1109/BigData.2014.7004305

Funk, C., Tönjes, E., Teuber, R., & Breuer, L. (2024). Reading between the lines: The intersection of research attention and sustainable development goals. *Sustainable Development*, *32*, 4545–4566. https://doi.org/10.1002/sd.2906

Galsurkar, J., Singh, M., Wu, L., Vempaty, A., Sushkov, M., Iyer, D., Kapto, S., & Varshney, K. R. (2018). Assessing national development plans for alignment with sustainable development goals via semantic search. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 7753–7758. https://doi.org/10.1609/aaai.v32i1.11424

Garbero, A., Resce, G., & Carneiro, B. (2021). Spatial dynamics across food systems transformation in IFAD investments: A machine learning approach. *Food Security*, *13*(5), 1125–1143. https://doi.org/10.1007/s12571-021-01190-8

Goetz, S. J., Heaton, C., Imran, M., Pan, Y., Tian, Z., Schmidt, C., Qazi, U., Ofli, F., & Mitra, P. (2023). Food insufficiency and Twitter emotions during a pandemic. *Applied Economic Perspectives and Policy*, *45*(2), 1189–1210. https://doi.org/10.1002/aepp.13258

Hansen, M., & Hershcovich, D. (2022). A dataset of sustainable diet arguments on Twitter. *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*, 40–58. https://doi.org/10.18653/v1/2022.nlp4pi-1.5

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266. https://doi.org/10.1126/science.aaa8685

HLPE-FSN. (2020). *Food security and nutrition: Building a global narrative towards 2030* (tech. rep.). High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security [HLPE-FSN]. Rome, Italy.

HLPE-FSN. (2022). *Data collection and analysis tools for food security and nutrition: Towards enhancing effective, inclusive, evidence-informed, decision making* (tech. rep.). High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security [HLPE-FSN]. Rome, Italy, Rome, Italy.

Hu, G., Ahmed, M., & L'Abbé, M. R. (2023). Natural language processing and machine learning approaches for food categorization and nutrition quality prediction compared with traditional methods. *The American Journal of Clinical Nutrition*, *117*(3), 553–563. https://doi.org/10.1016/j.ajcnut.2022.11.022

Hu, G., Flexner, N., Tiscornia, M. V., & L'Abbé, M. R. (2023). Accelerating the classification of NOVA food processing levels using a fine-tuned language model: A multi-country study. *Nutrients*, *15*(19), 4167. https://doi.org/10.3390/nu15194167

Huangfu, L., & Zeng, D. (2018). Social media-based overweight prediction using deep learning. *Proceedings of the twenty-fourth Americas Conference on Information Systems.*

ITU. (2024). *AI for good-innovate for impact. an interim report* (tech. rep.). International Telecommunication Union [ITU]. Geneva, Switzerland.

Jann, W., & Wegrich, K. (2017). Theories of the policy cycle. In *Handbook of public policy analysis* (pp. 69–88). Routledge. https://doi.org/10.4324/9781315093192

Jin, Z., & Mihalcea, R. (2023). Natural language processing for policymaking. In *Handbook of computational social science for policy* (pp. 141–162). Springer. https://doi.org/10.1007/978-3-031-16624-2_7

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (3rd).

Juventia, S. D., Jones, S. K., Laporte, M.-A., Remans, R., Villani, C., & Estrada-Carmona, N. (2020). Text mining national commitments towards agrobiodiversity conservation and use. *10.3390/su12020715*, *12*(2), 715.

Kil, S.-H., Park, H.-M., Lee, E., Kim, J.-Y., & Kim, J.-W. (2023). The analysis of research trends and public awareness of smart farms using text mining. *Journal of People, Plants, and Environment*, *26*, 9–21. https://doi.org/10.11628/ksppe.2023.26.1.9

Kim, J., Cha, M., & Lee, J. G. (2017). Nowcasting commodity prices using social media. *PeerJ Computer Science*, *3*, e126. https://doi.org/10.7717/peerj-cs.126

Kim, M., Koo, J., & Jung, Y. (2021). *Can we trust large language models to summarize food policy research papers and generate research briefs?* (Tech. rep.). CGIAR Research Initiative on Digital Innovation.

Kimmel, W. M., Beauchamp, P. M., Frerking, M. A., Kline, T. R., Koorosh Vassigh, K., Willard, D. E., Johnson, M. A., & Trenkle, T. G. (2020). *Technology readiness assessment: Best practices guide* (tech. rep.). National Aeronautics and Space Administration (NASA). Washington DC, USA.

Kontar, N. A. A., Mutalib, S., Hanum, H. F. M., & Abdul-Rahman, S. (2023). Exploratory data analysis: Food security risk among Twitter users. *Journal of Computer Science & Computational Mathematics*, *13*. https://doi.org/10.20967/jcscm.2023.01.003

Krismawati, D., & Panuntun, S. B. (2023). Text analysis study on urban farming news toward food security in Indonesia: Sentiment analysis, named entity recognition, topic modelling, and social network analysis. *Proceedings of The International Conference on Data Science and Official Statistics*, 177–185. https://doi.org/10.34123/icdsos.v2023i1.352

Kumar, A., & Sharma, A. (2020). Socio-sentic framework for sustainable agricultural governance. *Sustainable Computing: Informatics and Systems*, *28*, 100274. https://doi.org/10.1016/j.suscom.2018.08.006

Lampropoulos, G., Garzón, J., Misra, S., & Siakas, K. (2024). The role of artificial intelligence of things in achieving sustainable development goals: State of the art. *Sensors*, *24*(4), 1091. https://doi.org/10.3390/s24041091

Lindquist, J., Thomas, D. M., Turner, D., Blankenship, J., & Kyle, T. K. (2021). Food for thought: A natural language processing analysis of the 2020 dietary guidelines publice comments. *The American Journal of Clinical Nutrition*, *114*(2), 713–720. https://doi.org/10.1093/ajcn/nqab119

Lu, G., Li, S., Mai, G., Sun, J., Zhu, D., Chai, L., Sun, H., Wang, X., Dai, H., Liu, N., et al. (2023). AGI for agriculture. *arXiv preprint arXiv:2304.06136*. https://doi.org/10.48550/arXiv.2304.06136

Lukyamuzi, A., Ngubiri, J., & Okori, W. (2018). Tracking food insecurity from tweets using data mining techniques. *Proceedings of the 2018 International Conference on Software Engineering in Africa*, 27–34. https://doi.org/10.1145/3195528.3195531

Lukyamuzi, A., Ngubiri, J., & Okori, W. (2020). Towards ensemble learning for tracking food insecurity from news articles. *International Journal of System Dynamics Applications (IJSDA)*, *9*(4), 129–142. https://doi.org/10.4018/IJSDA.2020100107

Lv, X., Meng, J., & Wu, Q. (2022). Dynamic influence of network public opinions on price fluctuation of small agricultural products based on NLP-TVP-VAR model—taking garlic as an example. *Sustainability*, *14*(14), 8637. https://doi.org/10.3390/su14148637

Martin, N. M., Poirier, L., Rosenblum, A. J., Reznar, M. M., Gittelsohn, J., & Barnett, D. J. (2022). Enhancing artificial intelligence for Twitter-based public discourse on food security during the covid-19 pandemic. *Disaster medicine and public health preparedness*, 1–25. https://doi.org/10.1017/dmp.2022.207

Martin, N. M., Sedoc, J., Poirier, L., Rosenblum, A. J., Reznar, M. M., Gittelsohn, J., & Barnett, D. J. (2022). Harnessing artificial intelligence to improve food assistance: A scoping review of machine learning tools. *Preprints*. https://doi.org/10.20944/preprints202207.0221.v1

Martin, P., Helmer, T., Rabatel, J., & Roche, M. (2021). KEOPS: Knowledge ExtractOr Pipeline System. *Research Challenges in Information Science*, 561–567.

Matsuoka, H., Uchiyama, Y., Woraitthinan, K., & Kohsaka, R. (2023). Does novel food differ in cultural contexts? a comparative analysis of Japanese and Singaporean cultural acceptance through text analysis of mass media. *Current Research in Food Science*, *6*, 100436. https://doi.org/10.1016/j.crfs.2023.100436

Meitei, A. J., Rai, P., & Rajkishan, S. (2023). Application of AI/ML techniques in achieving SDGs: A bibliometric study. *Environment, Development and Sustainability*, 1–37. https://doi.org/10.1007/s10668-023-03935-1

Min, B., Chan, Y. S., Qiu, H., & Fasching, J. (2019). Towards machine reading for interventions from humanitarian-assistance program literature. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 6444–6448. https://doi.org/10.18653/v1/D19-1680

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International journal of surgery*, *8*(5), 336–341. https://doi.org/10.1371/journal.pmed.1000097

Molenaar, A., Lukose, D., Brennan, L., Jenkins, E. L., & McCaffrey, T. A. (2024). Using natural language processing to explore social media opinions on food security: Sentiment analysis and topic modeling study. *Journal of Medical Internet Research*, *26*, e47826. https://doi.org/10.2196/47826

Nasir, O., Javed, R. T., Gupta, S., Vinuesa, R., & Qadir, J. (2023). Artificial intelligence and sustainable development goals nexus via four vantage points. *Technology in Society*, *72*, 102171. https://doi.org/10.1016/j.techsoc.2022.102171

Park, Y., & Shin, Y.-W. (2021). Trend analysis of grow-your-own using social network analysis: Focusing on hashtags on Instagram. *Journal of People, Plants, and Environment*, *24*(5), 451–460. https://doi.org/10.11628/ksppe.2021.24.5.451

Paudel, R., & Ligmann-Zielinska, A. (2023). A largely unsupervised domain-independent qualitative data extraction approach for empirical agent-based model development. *Algorithms*, *16*(7), 338. https://doi.org/10.3390/a16070338

Pretorius, B., Muka, J. M., Hulshof, P. J. M., & Schönfeldt, H. C. (2023). Current practices, challenges and new advances in the collection and use of food composition data for Africa. *Frontiers in Sustainable Food Systems*, *7*. https://doi.org/10.3389/fsufs.2023.1240734

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, *42*, 13–24. https://doi.org/10.1016/j.ijinfomgt.2018.05.004

Ray, P. P. (2023). Leveraging deep learning and language models in revolutionizing water resource management, research, and policy making: A case for chatgpt. *ACS ES&T Water*, *3*(8), 1984–1986.

Roche, M., Lindsten, A., Lundén, T., & Helmer, T. (2022). LEAP4FNSSA lexicon: Towards a new dataset of keywords dealing with food security. *Data in Brief*, *45*, 108680. https://doi.org/10.1016/j.dib.2022.108680

Sarku, R., Clemen, U. A., & Clemen, T. (2023). The application of artificial intelligence models for food security: A review. *Agriculture*, *13*(10), 2037. https://doi.org/10.3390/agriculture13102037

Schopf, T., Arabi, K., & Matthes, F. (2023). Exploring the landscape of natural language processing research. *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, 1034–1045. https://doi.org/10.26615/978-954-452-092-2_111

Scott, D., Oh, J., Chappelka, M., Walker-Holmes, M., & DiSalvo, C. (2018). Food for thought: Analyzing public opinion on the Supplemental Nutrition Assistance Program. *Journal of Technology in Human Services*, *36*(1), 37–47. https://doi.org/10.1080/15228835.2017.1416514

Sharma, S. S., & De Choudhury, M. (2015). Measuring and characterizing nutritional information of food and ingestion content in Instagram. *Proceedings of the 24th International Conference on World Wide Web*, 115–116. https://doi.org/10.1145/2740908.2742754

Shi, Z. R., Wang, C., & Fang, F. (2020). Artificial intelligence for social good: A survey. *arXiv preprint arXiv:2001.01818*. https://doi.org/10.48550/arXiv.2001.01818

Silva e Silva, L., Mongeau Ospina, C. A., & Fabi, C. (2024). Food price inflation nowcasting and monitoring. *Statistical Journal of the IAOS*, (Preprint), 1–15. https://doi.org/10.3233/SJI-230083

Simon, J., Wong, P. H., & Rieder, G. (2020). Algorithmic bias and the value sensitive design approach. *Internet Policy Review*, *9*. https://doi.org/10.14763/2020.4.1534

Singh, A., Kanaujia, A., Singh, V. K., & Vinuesa, R. (2024). Artificial intelligence for sustainable development goals: Bibliometric patterns and concept evolution trajectories. *Sustainable Development*, *32*. https://doi.org/10.1002/sd.2706

Singh, A., & Glińska-Neweś, A. (2022). Modeling the public attitude towards organic foods: A big data and text mining approach. *Journal of Big Data*, *9*. https://doi.org/10.1186/s40537-021-00551-6

Statista. (2024). Natural language processing - worldwide. market size comparison (2024). https://www.statista.com/outlook/tmo/artificial-intelligence/natural-language-processing/worldwide#analyst-opinion

Taglioni, C., Moncayo, J., & Fabi, C. (2023). *Food loss estimation: SDG 12.3. 1a data and modelling approach* (tech. rep.). Food and Agricultural Organization of the United Nations [FAO]. Rome, Italy.

Thow, A. M. (2024). Enhancing global support to address complex sustainable development policy challenges: Learning from success in cross-sectoral nutrition policy. *Sustainable Development*. https://doi.org/10.1002/sd.3191

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C., Ezer, D., Haert, F. C. v. d., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., de Wever, W., … Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, *11*(1), 2468. https://doi.org/10.1038/s41467-020-15871-z

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of internal medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Truby, J. (2020). Governing artificial intelligence to benefit the un sustainable development goals. *Sustainable Development*, *28*, 946–959. https://doi.org/10.1002/sd.2048

UN. (2015). *Transforming our world: The 2030 agenda for sustainable development.* (tech. rep.). United Nations [UN]. New York, NY, USA.

UN DESA. (2023). *The sustainable development goals report 2023: Special edition* (tech. rep.). United Nations Department of Economic and Social Affairs [UN DESA]. New York, NY, USA.

UN DESA. (2024c). Availability of disaggregated data for Any type of disaggregation for World (total) by SDG regions and by indicators under goal 2 and series (Data for at least one year since 2015). https://unstats.un.org/sdgs/dataportal/analytics/DataAvailability

UN DESA. (2024b). Compare countries accross goal 2 (Data for at least one year since 2015, by country (average across indicators in percent)). https://unstats.un.org/sdgs/dataportal/analytics/DataAvailability

UN DESA. (2024a). *The sustainable development goals report: 2024* (tech. rep.). United Nations Department of Economic and Social Affairs [UN DESA]. New York, NY, USA.

van den Hoven, J., Blaauw, M., Pieters, W., & Warnier, M. (2020). Privacy and Information Technology. In *The Stanford encyclopedia of philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University.

Varriale, V., Camilleri, M. A., Cammarano, A., Michelino, F., Müller, J., & Strazzullo, S. (2024). Unleashing digital transformation to achieve the sustainable development goals across multiple sectors. *Sustainable Development*. https://doi.org/10.1002/sd.3139

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*. https://doi.org/10.48550/arXiv.1706.03762

Wanrooij, C., Cruijssen, F., & Olier, J. S. (2024). Unsupervised news analysis for enhanced high-frequency food insecurity assessment. *Decision Sciences*. https://doi.org/10.1111/deci.12653

West, R., White, R. W., & Horvitz, E. (2013). From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. *Proceedings of the 22nd international conference on World Wide Web*, 1399–1410. https://doi.org/10.1145/2488388.2488510

World Bank. (2022). Individuals using the internet (% of population) (2021). https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2017&start=2015&view=map&year=2021

Youn, J., Li, F., & Tagkopoulos, I. (2023). Semi-automated construction of food composition knowledge base. *arXiv preprint at arXiv:2301.11322*. https://doi.org/10.48550/arXiv.2301.11322

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2024).

A survey of large language models. *arXiv preprint arXiv:2303.18223*. https://doi.org/10.48550/arXiv.2303.18223

A Study identification and selection

A.1 Search history

The high-level search query described in Section 3.1 was customized for each database to fit the specific search engine capabilities. This meant that for most grey literature databases, only NLP-related search strings were used, and studies related to food security were filtered manually. Additionally, to capture relevant NLP studies not explicitly labeled as such, the search was broadened by incorporating general AI and machine learning terms combined with "text" or "language". A detailed overview of the search query per database, along with the number of studies retrieved, is provided in Table A.

A.2 Eligibility criteria

A detailed overview of eligibility criteria is presented in Table B.

A.3 Study selection with ASReview

To expedite the screening process, the software ASReview (ASReview LAB developers, 2024) was utilized, which applies active learning to dynamically present records to the reviewer based on their past decisions, reducing screening time significantly (ASReview, 2024a). As per ASReview's recommendation (ASReview, 2024a, 2024b), different active learning models were utilized: 1) Term Frequency-Inverse Document Frequency with Naive Bayes; 2) Doc2Vec with Random Forest; 3) Sentence-BERT with Neural Network (details below). A heuristic stopping rule was implemented, halting screening in each model iteration after reaching 100 consecutive irrelevant studies (ASReview, 2024a).

Before deploying the first active learning model, studies were randomly screened until 10 relevant studies were identified. This approach provided an estimate of the total number of relevant studies in the dataset. To minimize risk of bias in the screening process, these studies (n=191, 12%) were then double screened by a second reviewer, reaching a 94% agreement rate with the primary reviewer. Disagreements were discussed until consensus was reached, and the double-screened records were subsequently used as training data for the first active learning model iteration. This iteration used the default ASReview setup: Term Frequency-Inverse Document Frequency with Naive Bayes, with the query strategy set to maximum and the balance strategy set to dynamic resampling. This configuration is a powerful and lightweight starting point (ASReview, 2024b). In total, 63 relevant records were retrieved after 689 labeled items. Upon reaching 100 consecutive irrelevant studies, the model was switched to Doc2Vec with Random Forest, again using the maximum query strategy and dynamic resampling. This model, which incorporates contextual and semantic understanding, is considered highly effective for identifying the remaining relevant records (ASReview, 2024b). Labeling 134 additional items yielded 2 more relevant records. After another 100 consecutive irrelevant studies, the model was switched to the final iteration: Sentence-BERT with Neural Network, once again using the maximum query strategy and dynamic resampling. This transformer-based model, the most complex and computationally intensive of the models used, is considered to have the highest performance in textual interpretation, including context and semantics (ASReview, 2024b). One additional relevant record was retrieved after labeling 155 items. Upon reaching 100 consecutive irrelevant studies, the screening process was halted.

Overall, this model switching approach, progressing from simpler to more complex models, facilitated an efficient start while gradually improving contextual and semantic understanding. In total, 59% of the studies in the dataset were screened (n=978), with the remaining 670 studies excluded by ASReview.

Table A: Search history per database.

Database	**Search query**	**Number of records**	**Search date**
Web of Science	(TS=("natural language processing" OR "nlp" OR "language model*" OR "llm" OR "language representation" OR "language anal*" OR "word embed*" OR "text model*" OR "text mining" OR "text analy*" OR "computational linguistic*")) AND (TS=("food security" OR "food insecurity" OR "food access" OR "access to food" OR "food aid" OR "food sovereignty" OR "hunger" OR "nutrition security" OR "nutrition insecurity" OR "right to food" OR "self-sufficiency" OR "novel food" OR "resource management" OR "early warning" OR "nutritional quality" OR "malnutrition" OR "undernourish*" OR "socioeconomic sustainability" OR "sustainable intensification" OR "food system*" OR "food availability" OR "available food" OR "food affordability" OR "affordable food" OR "food utility" OR "food stability" OR "food agency" OR "food sustainability" OR "food distribution" OR "food supply" OR "food production" OR "food justice" OR "agrifood" OR "sdg 2" OR "sustainable development goal 2")) AND Timespan: 2010-01-01 to 2024-05-03 (Publication Date)	357	3 May 2024
Scopus	(TITLE-ABS-KEY ("natural language processing" OR "nlp" OR "language model*" OR "llm" OR "language representation" OR "language anal*" OR "word embed*" OR "text model*" OR "text mining" OR "text analy*" OR "computational linguistic*")) AND (TITLE-ABS-KEY ("food security" OR "food insecurity" OR "food access" OR "access to food" OR "food aid" OR "food sovereignty" OR "hunger" OR "nutrition security" OR "nutrition insecurity" OR "right to food" OR "self-sufficiency" OR "novel food" OR "resource management" OR "early warning" OR "nutritional quality" OR "malnutrition" OR "undernourish*" OR "socioeconomic sustainability" OR "sustainable intensification" OR "food system*" OR "food availability" OR "available food" OR "food affordability" OR "affordable food" OR "food utility" OR "food stability" OR "food agency" OR "food sustainability" OR "food distribution" OR "food supply" OR "food production" OR "food justice" OR "agrifood" OR "sdg 2" OR "sustainable development goal 2")) AND (PUBYEAR >2009)	1.162	3 May 2024
Arxiv	order: -announced_date_first; size: 200; date_range: from 2010-01-01 to 2024-12-31; include_cross_list: True; terms: AND abstract="natural language processing" OR "nlp" OR "language model*" OR "llm" OR "language representation" OR "word embed*" OR "text model*" OR "text mining" OR "text analy*" OR "computational linguistic*"; AND abstract="food security" OR "food insecurity" OR "food access" OR "access to food" OR "food aid" OR "food sovereignty" OR "hunger" OR "nutrition security" OR "nutrition insecurity" OR "right to food" OR "self-sufficiency" OR "novel food" OR "resource management" OR "early warning" OR "nutritional quality" OR "malnutrition" OR "undernourish*" OR "socioeconomic sustainability" OR "sustainable intensification" OR "food system*" OR "food availability" OR "available food" OR "food affordability" OR "affordable food" OR "food utility" OR "food stability" OR "food agency" OR "food sustainability" OR "food distribution" OR "food supply" OR "food production" OR "food justice" OR "agrifood" OR "sdg 2" OR "sustainable development goal 2"	36	26 April 2024
Policy Commons	((summary:("natural language processing" OR "nlp" OR "language model" OR "language modeling" OR "llm" OR "language representation" OR "language analysis" OR "word embedding" OR "text model" OR "text modeling" OR "text mining" OR "text analysis" OR "computational linguistics") AND fulltext:(food OR agriculture OR hunger OR malnutrition OR nutrition OR agrifood OR "early warning" OR "sustainable development goal 2" OR "sdg 2")) OR (summary:("machine learning" OR "machine intelligence" OR "deep learning" OR "neural network" OR "neural model" OR "pre-trained model" OR "pretrained model" OR "supervised learning" OR "unsupervised learning") AND summary:(text OR language) AND fulltext:(food OR agriculture OR hunger OR malnutrition OR nutrition OR agrifood OR "early warning" OR "sustainable development goal 2" OR "sdg 2")) OR (summary:("artificial intelligence") AND summary:(text OR language) AND summary:(food OR agriculture OR hunger OR malnutrition OR nutrition OR agrifood OR "early warning" OR "sustainable development goal 2" OR "sdg 2"))) AND (Years Published: 2010-2024)	152	29 April 2024
International Food Policy Research Institute	("natural language" OR nlp OR "language model" OR "language modeling" OR llm OR "language representation" OR "language analysis" OR "word embedding" OR "text model" OR "text modeling" OR "text mining" OR "text analysis" OR "computational linguistics" OR "artificial intelligence" OR "machine learning" OR "machine	56	29 April 2024

Publications	intelligence” OR “deep learning” OR “neural network” OR “neural model” OR “pre-trained model” OR “pretrained model” OR “supervised learning” OR “unsupervised learning”) AND (Publication date: 2010-2024)		
The World Bank Open Knowledge Repository	((“natural language” OR nlp OR “language model” OR “language modeling” OR llm OR “language representation” OR “language analysis” OR “word embedding” OR “text model” OR “text modeling” OR “text mining” OR “text analysis” OR “computational linguistics”) OR ((“artificial intelligence” OR “machine learning” OR “machine intelligence” OR “deep learning” OR “neural network” OR “neural model” OR “pre-trained model” OR “pretrained model” OR “supervised learning” OR “unsupervised learning”) AND (text OR language))) AND (Date: 2010-2024)	26	29 April 2024
UN Digital Library System	(Abstract and Notes: (“natural language” OR nlp OR “language model” OR “language modeling” OR llm OR “language representation” OR “language analysis” OR “word embedding” OR “text model” OR “text modeling” OR “text mining” OR “text analysis” OR “computational linguistics” OR “artificial intelligence” OR “machine learning” OR “machine intelligence” OR “deep learning” OR “neural network” OR “neural model” OR “pre-trained model” OR “pretrained model” OR “supervised learning” OR “unsupervised learning”)) AND (Specific date period: 2010-2024)	19	29 April 2024
FAO Knowledge Repository BETA	(“natural language” OR nlp OR “language model” OR “language modeling” OR llm OR “language representation” OR “language analysis” OR “word embedding” OR “text model” OR “text modeling” OR “text mining” OR “text analysis” OR “computational linguistics”) AND (Year of Publication: 2010-2024)	177	29 April 2024

Table B: Eligibility criteria.

Database Search date	**Search query**	**Number of records**
Document type	Peer-reviewed articles, publications, technical reports, working papers, policy papers, discussion papers	Meeting documentation, posters, presentations, infographics, newsletters, flyers, blogs
Study type	Original and applied studies	Review studies, purely theoretical studies
Language	English	Non-English
Time frame†	After 2009	Before 2010
Method	NLP	Non-NLP
Thematic domain	Food/nutrition security *Including food security dimensions: availability, access, stability, utilization, agency and sustainability*	Non-food/nutrition security *Including: food safety, food (image) recognition, food fraud, food engineering, food design, food supply chain, customer satisfaction in hospitality*
	(Potential) policy application *Including policy cycle stages: agenda setting, policy formulation & decision-making, implementation, and evaluation*	Non-policy application *Including: private sector, including (agricultural) industry*

† The choice for limiting the time frame to studies from 2010 onwards is based on the occurrence of AI4SG and NLP4SG papers over time (Adauto et al., 2023; Lampropoulos et al., 2024; Shi et al., 2020).

A.4 Study selection with GPT-4o

To validate the screening process, OpenAI’s chat completion API was used to prompt GPT-4o (model “gpt-4o-2024-05-13”) for study selection based on the eligibility criteria, using a one-shot prompting approach. The prompt included both a system message with high-level instructions about the model’s role and its task, and a user message with the specific request. The user message contained the request to the model, an example to guide the model’s response in the desired format, and the study under review. To ensure focused and deterministic output, the temperature was set to 0.2, while top-p sampling was configured at 0.1. The seed parameter was set to 1234 to receive mostly consistent outputs, with a token length of 20. No other hyperparameters were set. This setup, detailed in Table C, was applied to each study.

Table C: GPT-4o prompt for study selection.

System message:
You are a classifier that predicts whether a paper is relevant or irrelevant based on a prompt.

User request:
PROMPT: Is this paper relevant or irrelevant? A paper is RELEVANT if it meets these criteria: 1) Applies natural language processing (NLP); 2) Applied to food security domain (food availability, food access, food stability, food utilization, food agency, food sustainability); 3) Applied to the policy domain (agenda setting, formulation, decision-making, implementation, evaluation), excluding private sector. A paper is IRRELEVANT if it lacks any of these. Provide 'Relevant' or 'Irrelevant' and briefly explain (max 15 words, ';' separated).

EXAMPLE TITLE: Leveraging Deep Learning and Language Models in Revolutionizing Water Resource Management, Research, and Policy Making: A Case for ChatGPT. (Ray, 2023)
EXAMPLE ABSTRACT: The use of artificial intelligence (AI) and advanced language models like ChatGPT is not only innovative but also crucial in handling the multifaceted challenges related to water resource management, policy formulation, and scientific research in the contemporary world. [*Text is truncated to save space. In practice, the prompt contained the full abstract of the article.*] In this context, ChatGPT's ability to generate interactive, scenario-based training can empower communities and stakeholders with the knowledge needed for sustainable water practices. (Ray, 2023)
EXAMPLE RESPONSE: Irrelevant; Applies NLP to water and not food domain.

TITLE: {*Title of an article*}.
ABSTRACT: {*Abstract of an article*}.

GPT-4o screened all 1648 records, with human annotation available for 978. A 93% agreement rate was achieved between GPT-4o and the primary reviewer. Disagreements were reviewed and revealed that GPT-4o tended to be less conservative, often identifying records as relevant that the primary reviewer deemed irrelevant. Notable disagreement cases included studies that did not present an NLP application (n=17), related to food safety which were argued to fall outside the inclusion criteria (n=11), used NLP as a research method in e.g., a literature study rather than an application (n=10), and presented a private sector agricultural production application (n=8). Examples of studies that were deemed irrelevant by GPT-4o but relevant by the primary reviewer include some for which the focus on food security was not explicit (n=6) as well as some studies for which a definitive decision could not be made based solely on the title and abstract (n=5).

Notably, all studies categorized as irrelevant by ASReview (n=670) were also classified as irrelevant by GPT-4o. This consistency further supports the use of ASReview to make the screening process more efficient. No further formal analysis of GPT-4o's responses was conducted. The experimental use demonstrates the potential of automating the study selection process of literature reviews using LLMs. However, the focus remained exploratory, and no definitive conclusions were reached regarding the accuracy or reliability of GPT-4o's outputs in this context.

A.5 Literature searching with ResearchRabbit

The list of selected studies after detailed assessment was uploaded to ResearchRabbit, which is an online citation-based literature mapping tool. Out of 42 studies, only 29 were available in ResearchRabbit. Using these as seed papers, ResearchRabbit identified 1603 studies under "Similar Work", 14 under "Earlier Work", and 4 under "Later Work". The "Similar Work" studies were filtered by relevance, and the titles and abstracts of the first 50 were screened, yielding 5 studies that met the inclusion criteria. From the "Earlier Work" category, only 1 study was included, while none of the "Later Work" studies met the inclusion criteria.

A.6 Included studies

The complete list of included studies, along with their extracted metadata and study characteristics is available on the GitHub page of the first author.

B Data extraction and analysis

B.1 Data items

An overview of extracted data items is included in Table D. Note that multiple responses could apply to each data item in the study characteristics category. For example, a study could address both the availability and access dimensions of food security. Similarly, it could cover both the implementation and evaluation phases of the policy cycle. Allowing multiple responses accounted for the complexity and interrelated nature of the subcategories within each data item. More detailed guidance for data extraction is included in the review protocol, which can be accessed on the GitHub page of the first author. The table with extracted metadata and study characteristics per study can also be accessed there.

Table D: Data items.

Category	Data items
Meta	Title, Abstract, Keywords, Authors' names, Authors' affiliation, Country of authors' affiliation, Publication year, Item type, Journal name or source, Study location
Key contributions	Main objective or research question, Results, Limitations, Notes on ethics and/or fairness
Study characteristics	Natural Language Processing subfield: Multimodality, Natural Language Interfaces, Semantic Text Processing, Sentiment Analysis, Syntactic Text Processing, Linguistics & Cognitive NLP, Responsible & Trustworthy NLP, Reasoning, Multilinguality, Information Retrieval, Information Extraction & Text Mining, Text Generation (Schopf et al., 2023) (Section 2.1) *The subfield was extracted for the primary NLP task or technique used in the study. Secondary subfields, e.g., associated with data pre-processing steps, were not considered.*
	Food security dimensions: Availability, Access, Stability, Utilization, Agency, Sustainability (HLPE-FSN, 2020) (Section 2.2)
	Policy cycle stages: Agenda setting, Policy formulation & decision-making, Implementation, Evaluation (Jann & Wegrich, 2017) (Section 2.3)
	Data source
	Deployment: TRL1 - Basic principles observed, TRL2 - Technology concept formulated, TRL3 - Experimental proof of concept, TRL4 - Technology validated in a lab, TRL5 - Technology validated in a relevant environment, TRL6 - Technology demonstrated in a relevant environment, TRL7 - System prototype demonstration in an operational environment, TRL8 - System complete and qualified, TRL9 - Actual system proven in an operational environment (EC, 2024; Kimmel et al., 2020) (Section 2.4)
	Implementation partner (if applicable)

B.2 Data extraction with GPT-4o

To explore the potential of enhancing systematic literature reviews with LLMs and to triangulate data extraction, OpenAI's chat completion API was used to prompt GPT-4o (model "gpt-4o-2024-05-13") for the extraction of various data items. The prompt included both a system message with high-level instructions about the model's role and its task, and a user message containing the specific request. The full text of each study was extracted from its PDF format and converted into Markdown using the PyMuPDF4LLM Python package. It was then inputted into the user message, together with a list of questions with one question per to be extracted data item. To ensure focused and deterministic output, the temperature was set to 0.2, while top-p sampling was configured at 0.1. The seed parameter was set to 1234 to receive mostly consistent outputs. No other hyperparameters were set. This setup, detailed in Table E, was applied to each study.

To evaluate GPT-4o's responses regarding the categorical data items (i.e. food security dimension, policy cycle stage, NLP subfield, and TRL), the complete set of responses was analyzed and compared with the

Table E: GPT-4o prompt for data extraction.

System message:
You are a research assistant that helps me with data extraction from articles.

User request:
ARTICLE: {*full text of an article*}.
QUESTIONS: Please answer the following 6 questions about the article just presented to you. Start each answer with “QUESTION 1:”, “QUESTION 2:”, etc. Separate the answers using a comma ’,’. For each answer, provide an explanation of max 15 words in brackets ’()’.

QUESTION 1: Which dimension(s) of food security does this article discuss? You can choose from the following six dimensions:

A) Availability: Having a quantity and quality of food sufficient to satisfy the dietary needs of individuals, free from adverse substances and acceptable within a given culture, supplied through domestic production or imports.

B) Access: Having personal or household financial means to acquire food for an adequate diet at a level to ensure that satisfaction of other basic needs are not threatened or compromised; and that adequate food is accessible to everyone, including vulnerable individuals and groups.

C) Utilization: Having an adequate diet, clean water, sanitation and health care to reach a state of nutritional well-being where all physiological needs are met.

D) Stability: Having the ability to ensure food security in the event of sudden shocks (e.g. an economic, health, conflict or climatic crisis) or cyclical events (e.g. seasonal food insecurity).

E) Agency: Individuals or groups having the capacity to act independently to make choices about what they eat, the foods they produce, how that food is produced, processed, and distributed, and to engage in policy processes that shape food systems. The protection of agency requires socio-political systems that uphold governance structures that enable the achievement of food and nutrition security for all.

F) Sustainability: Food system practices that contribute to long-term regeneration of natural, social and economic systems, ensuring the food needs of the present generations are met without compromising the food needs of future generations.

Multiple dimensions possible. Only mention the dimension name, e.g. Availability or Sustainability. Separate dimensions using ’;’. (max 10 tokens)

QUESTION 2: Which stage(s) of the policy cycle does the application in this article apply to? You can choose from the following four stages:

A) Agenda setting: Problem identification and recognition (incl. determinants and/or magnitude thereof).

B) Policy formulation & decision-making: Identified problem is translated into a policy or program.

C) Implementation: Execution or enforcement of a policy or program.

D) Evaluation: Assessment of a policy or program against (intended) outcomes (incl. monitoring mechanisms & impact assessment).

Multiple stages possible. Only mention the stage name, e.g. Agenda setting or Evaluation. Separate stages using ’;’. (max 10 tokens)

QUESTION 3: Which NLP subfield is applied in this article? You can choose from the following twelve subfields:

A) Mutimodality

B) Natural Language Interfaces

C) Semantic Text Processing

D) Sentiment Analysis

E) Syntactic Text Processing

F) Linguistics & Cognitive NLP

G) Responsible & Trustworthy NLP

H) Reasoning

I) Multilinguality

J) Information Retrieval

K) Information Extraction & Text Mining

L) Text Generation

Multiple subfields possible. Only mention the subfield name, e.g. Multimodality or Text Generation. Separate subfields using ’;’. (max 15 tokens).

QUESTION 4: Which NLP model is used in this article? (max 8 tokens)

QUESTION 5: What is the data source used in this article? (max 5 tokens)

QUESTION 6: What is the technology readiness level of the application discussed in this article? You can choose from the following nine levels:

1. TRL1 Basic Research: Initial scientific research has been conducted. Principles are qualitatively postulated and observed. Focus is on new discovery rather than applications.
2. TRL2 Applied Research: Initial practical applications are identified. Potential of material or process to solve a problem, satisfy a need, or find application is confirmed.
3. TRL3 Critical Function or Proof of Concept Established: Applied research advances and early stage development begins. Studies and laboratory measurements validate analytical predictions of separate elements of the technology.
4. TRL4 Lab Testing/Validation of Alpha Prototype Component/Process: Design, development and lab testing of components/processes. Results provide evidence that performance targets may be attainable based on projected or modeled systems.
5. TRL5 Laboratory Testing of Integrated/Semi-Integrated System: System Component and/or process validation is achieved in a relevant environment.
6. TRL6 Prototype System Verified: System/process prototype demonstration in an operational environment (beta prototype system level).
7. TRL7 Integrated Pilot System Demonstrated: System/process prototype demonstration in an operational environment (integrated pilot system level).
8. TRL8 System Incorporated in Commercial Design: Actual system/process completed and qualified through test and demonstration (pre-commercial demonstration).
9. TRL9 System Proven and Ready for Full Commercial Deployment: Actual system proven through successful operations in operating environment, and ready for full commercial deployment.

Only one level possible. Mention the level name, e.g. TRL8, and provide a short explanation why you choose this TRL. (max 25 tokens)

---

responses of the primary reviewer (n=60). As multiple responses could apply, the focus was on whether GPT-4o and the primary reviewer consistently identified a dimension, without considering combinations with other dimensions. For each of the six food security dimensions, the overall agreement between GPT-4o and the primary reviewer ranged from 68% to 88%, with availability showing the lowest agreement and sustainability the highest. In examining the disagreement cases for availability, both GPT-4o and the primary reviewer typically provided multiple responses (averaging 2.2 and 2.3, respectively). However, the primary reviewer identified availability more frequently than GPT-4o (63% vs 37%). For policy cycle stages, agreement ranged from 37% (for evaluation) to 87% (for implementation). In the majority of disagreement cases for evaluation (74%), GPT-4o provided multiple policy cycle stages with evaluation being one of them, whereas the primary reviewer identified a single stage that differed from evaluation. Agreement between GPT-4o and the primary reviewer regarding NLP subfields was generally high, with the lowest level of agreement at 82% for semantic text processing. In all disagreement cases, GPT-4o listed semantic text processing as one of several NLP subfields, whereas the primary reviewer mostly provided only one NLP subfield that differed from semantic text processing. This discrepancy may have arised from the primary reviewer's focus on the main NLP task or technique in the study, deliberately not considering data pre-processing steps. Lastly, the agreement for TRL was notably low at just 12%. Interestingly, the primary reviewer's average TRL rating was 4.0, whereas for GPT-4o it was considerably higher, equaling 5.3.

To ensure a manageable evaluation of GPT-4o's responses for the open-ended data items (i.e. NLP model and data source), responses were assessed based on the subset of studies manually extracted by three reviewers (12% of studies, n=7). For both data items, all GPT-4o's responses were largely consistent with that of the three reviewers, with only minor deviations occurring in cases where multiple responses could apply.

Beyond incorporating GPT-4o's responses into the data extraction discussion among the three reviewers, no additional formal analysis was conducted on these responses. The primary aim of using GPT-4o in this context was to experiment with and demonstrate the potential of LLMs in enhancing systematic literature reviews. This experimental use highlighted the possibility of automating certain aspects of data extraction, thus offering a glimpse into how LLMs could be integrated for more efficient and scalable literature reviews. However, the focus remained exploratory, with no definitive conclusions drawn on the accuracy or reliability of GPT-4o's outputs in this setting.

Table F: GPT-4o prompt for data analysis.

System message:
You are a research assistant that helps me with the interpretation of literature articles. You will be presented 60 abstracts of articles and I will ask you some questions about them.

User request:
ABSTRACTS: {*list of all the abstracts*}.
QUESTION: What are the broad categories of NLP applications for food security policymaking? I want you to describe the applications from a policy perspective. Please provide a brief description of each category (ca. 5 sentences).

B.3 Data analysis with GPT-4o

To validate the categorization process, OpenAI's chat completion API was used to prompt GPT-4o (model "gpt-4o-2024-05-13") to suggest broad categories of NLP applications for food security policymaking. The prompt included both a system message with high-level instructions about the model's role and its task, and a user message containing the specific request. A list of all 60 abstracts was compiled and inputted into the user message. Various temperature settings (ranging from 0.2 to 2) were tested, with a temperature of 1 producing the most relevant outputs. The seed parameter was set to 1234 to receive mostly consistent outputs. No other hyperparameters were set. This setup resulted in the prompt presented in Table F.

Five application areas were identified: 1) Sentiment analysis and public opinion monitoring; 2) Early warning systems and crisis monitoring; 3) Policy evaluation and impact assessment; 4) Knowledge extraction and decision support; and 5) Trend analysis and behavioral insights. Brief descriptions for each category were provided. These application areas were compared against the manual categorization, resulting in few refinements in the final proposed application areas. This primarily concerned refinements in the framing of the application areas.

C Study characteristics by study type

As described in Section 4, 60% of studies were published in peer-reviewed journals (n=36), 28% were conference papers (n=17), and the remaining 12% were either working papers or other types of grey literature documents (n=7). Table G provides descriptive statistics for the categorical study characteristics by study type. Although the grey literature group was relatively small, a Tukey's honestly significant difference test was conducted at a 95% significance level to compare the means of these characteristics across study types. Statistically significant differences are reported below. Results should be interpreted with caution due to small sample sizes.

Notable differences emerged in the application areas covered by each study type. Grey studies most often addressed knowledge generation and management from food policy and program documents (71%), significantly more than academic studies (8%). Conference studies most often focused on food item classification (41%), more often than both academic (3%) and grey studies (0%).

Differences in food security dimensions were observed for utilization and for studies applying to all dimensions. Specifically, none of the grey studies addressed utilization, while 17% of academic and 47% of conference studies did. In contrast, grey studies more frequently covered all dimensions (57%) compared to academic and conference studies (both 6%). No differences by study type were found for policy cycle stages.

Some differences in NLP subfields were observed across study types. Grey studies more frequently used natural language interface techniques (29%), while none of the academic or conference studies did.

Table G: Study characteristics by study type.

Study type	Application area	Food security dimension	Policy cycle stage	NLP subfield	TRL (average)
Academic studies (60%; n=36)	Early warning 36%, Public discourse 8%, Knowledge generation & management 8%, Dietary habits 3%, Food item classification 3%, Data gaps 33%	Availability 39%, Access 44%, Utilization 17%, Stability 31%, Agency 33%, Sustainability 31%, All 6%	Agenda setting 47%, Policy formulation & decision-making 42%, Implementation 14%, Evaluation 25%	Multimodality 0%, Natural language interfaces 0%, Semantic text processing 3%, Sentiment analysis 47%, Syntactic text processing 0%, Linguistics & cognitive NLP 0%, Responsible & trustworthy NLP 0%, Reasoning 3%, Multilinguality 3%, Information retrieval 3%, Information extraction & text mining 83%, Text generation 0%	3.9
Conference studies (28%; n=17)	Early warning 23%, Public discourse 6%, Knowledge generation & management 12%, Dietary habits 12%, Food item classification 41%, Data gaps 6%	Availability 35%, Access 35%, Utilization 47%, Stability 18%, Agency 47%, Sustainability 12%, All 6%	Agenda setting 59%, Policy formulation & decision-making 29%, Implementation 18%, Evaluation 41%	Multimodality 0%, Natural language interfaces 0%, Semantic text processing 0%, Sentiment analysis 12%, Syntactic text processing 0%, Linguistics & cognitive NLP 0%, Responsible & trustworthy NLP 0%, Reasoning 12%, Multilinguality 0%, Information retrieval 6%, Information extraction & text mining 82%, Text generation 12%	3.2
Grey studies (12%; n=7)	Early warning 14%, Public discourse 0%, Knowledge generation & management 71%, Dietary habits 0%, Food item classification 0%, Data gaps 0%	Availability 43%, Access 14%, Utilization 0%, Stability 14%, Agency 0%, Sustainability 29%, All 57%	Agenda setting 29%, Policy formulation & decision-making 43%, Implementation 0%, Evaluation 43%	Multimodality 0%, Natural language interfaces 29%, Semantic text processing 0%, Sentiment analysis 14%, Syntactic text processing 0%, Linguistics & cognitive NLP 0%, Responsible & trustworthy NLP 0%, Reasoning 0%, Multilinguality 0%, Information retrieval 29%, Information extraction & text mining 100%, Text generation 29%	6.0

Additionally, both text generation and information retrieval were applied to 29% of grey studies, compared to only 0% and 3% of academic studies, respectively. Sentiment analysis was most prevalent in academic studies (47%), compared to 12% in conference studies.

Lastly, the average TRL was significantly higher for grey studies (6.0) compared to both academic (3.9) and conference studies (3.2). This is not surprising, as 86% of grey studies were conducted in partnership with an implementation partner, compared to only 22% and 6% of academic and conference studies, respectively.