

FoodClimNER: A Manually Annotated Corpus for Food Practice Entities Related to Climate Change

Mathieu Lai-king^{1*}, Louise Deléger², Arnaud Ferré², Yann Le Bodo³, Estelle Chaix^{1*}

¹DER, ANSES, Maisons-Alfort, France

²MaIAGE, INRAE, Université Paris-Saclay, Jouy-en-Josas, France

³DiSSES, ANSES, Maisons-Alfort, France

* Corresponding authors : {mathieu.laiking, estelle.chaix}@anses.fr

Abstract

Climate change is reshaping food systems and consumer practices, creating new challenges for food safety. Understanding how people modify their eating habits in response to climate disruptions and sustainability challenges is critical for assessing emerging microbiological risks. We present FoodClimNER¹, a new named entity recognition (NER) corpus containing extracts from open-access scientific articles on food science in the context of climate change, with manual annotations made by an expert. We define annotation guidelines with expert definitions of food practice entities and climate change entities. We evaluate baselines and a custom prompting method with recent LLMs on this task in a zero-shot setting. Results indicate that expert-level entity types and the presence of nested and discontinuous mentions make this a challenging testbed, with significant room for improving zero-shot methods. We also provide a detailed analysis of the importance of different parts of our custom prompt.

1 Introduction

Climate change is recognized as a major driver of changes in food safety challenges through its multiple impacts on food systems (Intergovernmental Panel on Climate Change (IPCC), 2023). It affects every stage of the food chain, from production and transportation to storage and food availability, while also modifying consumer food practices (Vermeulen et al., 2012). These interconnected changes are expected to increase the burden of food-borne diseases while reshaping food systems and consumption patterns worldwide (Dietrich et al., 2023). However, our understanding of how climate change affects consumer eating habits, the

¹We release our corpus publicly : <https://huggingface.co/datasets/inrae-bibliome/foodclimner>, as well as the code for our experiments : <https://github.com/mathieu-laiking/foodclimner>

Malliaroudaki, 2022

Indeed, dietary GHG emissions of vegan diets are about half that of meat-eaters' diets (Scarborough et al., 2014).

Malliaroudaki, 2022 (also)

In view of the fast-growing world population, the higher per capita income growth and “westernising” diet trends in the East, a significant rise in demand for dairy products sets the long-term sustainability of the sector into question (OECD-FAO, 2020; Pingali, 2007).

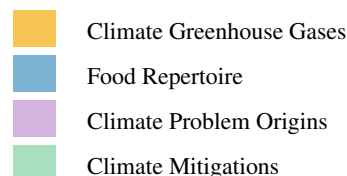


Figure 1: Examples of annotated sentences from FoodClimNER

final link in the food chain, remains limited. Consumers are adapting their eating habits by adopting more sustainable dietary practices (e.g., reduced meat consumption, increased consumption of local, organic or plant-based foods) (Innocent et al., 2024) or in response to climate-related disruptions such as reduced availability of certain products or supply chain disturbances (Hadley et al., 2023). These shifts may introduce or amplify microbiological hazards, as new food practices emerge and foods with limited safety data enter consumers' diets (Duchenne-Moutien and Neetoo, 2021).

Consumer food practices are documented across diverse text sources (scientific literature, grey literature, press articles, and social media), making natural language processing methods a promising choice for extracting structured information from them. Being able to extract information on food practices linked to climate change could help discover and understand emerging food practices that could pose a risk in the upcoming years due to

climate change. For example, the transition from single-use plastic bags to reusable shopping bags and food containers, promoted to reduce environmental impacts, may unintentionally introduce new microbiological hazards. If these reusable items are inadequately cleaned or sanitized, they can act as vehicles for cross-contamination, particularly when used for high-risk foods such as meat, cheese or ready-to-eat products (Kasza et al., 2022).

However, annotating sufficient training data for such tasks remains both time-consuming and costly. The rapid advances in large language models (LLMs) over recent years (Brown et al., 2020; Ouyang et al., 2022; Touvron et al., 2023; Guo et al., 2025) offer a compelling alternative: these models have demonstrated strong zero-shot capabilities through a variety of prompting strategies (Kojima et al., 2022; Wei et al., 2023), thus reducing the need for large annotated corpora. These zero-shot capabilities also transfer to information extraction tasks such as named entity recognition (NER) (Sainz et al., 2024; Xie et al., 2023).

Here we define the term zero-shot as found in the literature, meaning the model is not trained (weight optimization) on the specific task directly. The model instead generalizes to the new task using the knowledge acquired during its training (pre or post-training) or its context (in the case of LLMs). For example, Sainz et al. (2024) include additional task information in their prompt (including entity types descriptions, examples, or annotation guidelines) and still consider this zero-shot. This use of zero-shot can be debatable (especially when using entity examples), but we do not show exact examples

Therefore, we present a new *zero-shot* evaluation corpus for food practices related to climate change. Our contributions include **FoodClimNER**, an expert-annotated NER corpus built from 10 full-text open-access scientific articles, with food practices and climate change entities (see examples in Figure 1), along with expert-defined annotation guidelines, and zero-shot experiments (2 baselines and a custom prompting method with recent LLMs) on this corpus.

2 Related Work

2.1 NER in the climate and food domain

There already exist several works on information extraction in the food domain. However, they are usually recipe datasets annotated with food entities and attributes such as quantities, units, and cook-

ing processes (Popovski et al., 2019; Wróblewska et al., 2022) and do not include food practices or target other types of documents. In the climate domain, Bhattacharjee et al. (2024) introduce Climate Change NER as part of a broader suite of domain-specialized scientific language models, while Pan et al. (2025) propose ClimateIE, a corpus of climate science publications annotated for named entities, relations, and entity linking through an LLM-assisted annotation pipeline combined with expert validation.

Beyond NER, recent work has also explored domain-specialized LLMs for food and nutrition tasks, such as FoodyLLM (Gjorgjevikj et al., 2026), fine-tuned for nutrient estimation and ontology-based entity linking, alongside broader surveys on AI applications in food safety (Lopez et al., 2026).

As the majority of the existing NLP research in the food domain focuses on recipes, the originality of our approach is to annotate research articles in the food safety domain with food practices entities and link this to climate change entities.

2.2 Zero-shot NER methods

Zero-shot named entity recognition has attracted growing attention since the emergence of large language models. A first family of approaches, adapted across different model architectures, shares a common underlying idea: unifying multiple information extraction tasks within a single model by defining a unified output schema and training the model to generate it. This strategy was first explored with encoder-decoder architectures (Lu et al., 2022; Lou et al., 2023), later extended to decoder-only models (Wang et al., 2023), and more recently to encoder-only models (Zaratiana et al., 2025).

A second line of work distills LLMs into smaller, fine-tuned models by using the LLMs as automatic annotators over large text corpora. These approaches rely either solely on entity type labels (Zaratiana et al., 2023; Zhou et al., 2024) or additionally incorporate type descriptions (Aly et al., 2021; Cocchieri et al., 2025), computing similarity scores between sentence-token representations and entity-type (or type + description) representations.

Finally, a third category of approaches directly exploits the emerging capabilities of large language models, using original ways of defining annotation guidelines such as code (Sainz et al., 2024), tool-calling and multi-turn decomposition strategies (Xie et al., 2023), or retrieval-augmented cor-

rection of missing and erroneous entities (Cai et al., 2025).

3 FoodClimNER: A new corpus for food practices related to climate change

FoodClimNER is a new named entity recognition corpus on the subject of consumer food practices linked to climate change. The goal is to provide a testbed for detecting emerging food practices in scientific article documents that could present risks in the future due to climate change. However, this type of corpus is expensive to annotate, as we need expert annotators. Therefore, we aim to create a corpus used for the evaluation of modern zero-shot methods, as we believe our expert-defined entities and annotations will be a challenging task in this setting. Relation extraction was not performed; however, the co-occurrence of food practice and climate-related entities within the same textual context provides valuable information for identifying potential links between emerging food practices and climate change.

3.1 Annotation process

We describe our annotation process from schema creation to document annotation :

1. Two experts (in sociology and food microbiology) first developed the annotation schema for food practices. The annotation framework was designed through an iterative inductive-deductive approach, combining exploration of emerging food practices from the literature with theoretical frameworks describing food behaviors. Food practices were structured according to consumer food pathway, adapted from Desjeux’s consumer itinerary framework (Desjeux, 1998). This structure was chosen to capture food practices at different stages of the consumer pathway and to support subsequent microbiological risk assessment. Climate-related entity types were selected from the schema of an existing climate change-related NER corpus (Bhattacharjee et al., 2024) and adapted to the objectives of this study. We focused on entities describing climate drivers, impacts and mitigation-related concepts so we filtered some relevant entity types from the corpus schema. We provide our annotation schema with detailed entity type definitions for food practices and climate change entity types in section 3.2.
2. Relevant document types included scientific articles, grey literature, press articles and social media content. Given the diversity of available sources, we focused the initial manual annotation corpus on open-access scientific articles to ensure reproducibility and sufficient contextual information. An exploratory literature search was conducted in PubMed and Scopus using a PECO-based framework (EFSA, 2010), combining terms related to climate change, sustainability, food, diets and consumer behaviors. The article selection was further informed by a broader food practice database built independently by three annotators from full-text readings of the screened corpus (Le Bodo et al., 2026). Articles containing at least one instance of a "food preparation" practice were prioritized, as this category was expected to offer a greater variability in surface shape ; two additional articles addressing food practices outside the French context were added among remaining eligible candidates to increase geographic diversity. The resulting ten articles span ten different journals across food science, nutrition, public health, environmental education, and sustainability disciplines.
3. Then, the full-text articles were manually reviewed by one expert, and some sections were discarded because they were judged to contain insufficient entity density or redundant information and also to alleviate the workload of the annotator, avoiding annotating the whole document. For each article the removed part length is very variable, for some documents almost nothing has been removed, and for some other we kept only the abstract or a small part of the complete paper.
4. Annotation was then performed by a single expert annotator, with expertise in food microbiology (PhD), climate change and text annotation (previous work annotating microbes entities in scientific articles). The software used for the annotation is the brat rapid annotation tool (Stenetorp et al., 2012).

The corpus captures two challenging linguistic phenomena: **nested entities**, where spans of different types overlap, and **discontinuous entities**, where a single logical entity spans non-contiguous

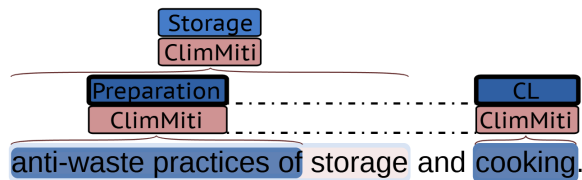


Figure 2: Example of nested + discontinuous annotations. We have one discontinuous mention of the “Food Preparation” entity: [“anti-waste practices of”, “cooking”]; that overlaps with one regular mention of the “Food Storage” entity: “anti-waste practices of storage”. They are also both annotated as climate mitigations, as these practices serve to mitigate climate change.

text segments. We provide an annotated example in Figure 2.

3.2 Annotation Schema

We define two different main concepts of entities which are consumer food practice entities and climate change entities. Both of these concepts contain multiple subtypes of entities that have been manually annotated. We also consider a version where all the subtypes are directly considered as the main concept because we believe that extracting the information of having food practices as well as climate change entities in a given text is already valuable. This simplified version will be noted **FoodClimNER-simple** (results in appendix B).

The first macro-type of entities are consumer food practices (actions directly carried out by an individual or a household as a food consumer). The subtypes of entities are defined according to the food itinerary of the consumer, from planning to waste management:

1. **Food repertoire:** Covers practices and factors related to food choices (in terms of quality or quantity) that may or may not be part of an individual’s diet (e.g., meat consumption, food choices).
2. **Logistics and Travel:** Includes pre-purchase practices (checking existing resources, meal planning, etc.) and travel to the place of purchase (using low-carbon transportation to go grocery shopping)
3. **Purchasing:** Covers purchasing practices (product selection, choice of retail locations, interactions with marketing) as well as non-market procurement strategies (self-production).

4. **Storage:** Covers food storage practices after purchase and before preparation (choice of packaging, adherence to or disregard of expiration dates, preservation, etc.)
5. **Food preparation:** All attitudes and practices related to food preparation and cooking (e.g., raw food diets, water conservation, reducing energy consumption, minimizing waste during preparation, etc.). It also includes skills and actions that we perform when cooking.
6. **Consumption** (table manners): Covers consumption practices related to eating and meals (e.g., portion sizes, choice of utensils).
7. **Waste Management:** Covers practices that take place after a meal, with consideration for the end of a food item’s life cycle (storing leftovers, composting, sorting food waste, expiration date reasoning. . .).

Climate change entities encompass the phenomena, factors, impacts, causes, and strategies associated with climate change, as described in the text or recognized as falling within the realm of climate. They include environmental events, greenhouse gases, climate impacts, anthropogenic causes of climate change, and actions aimed at mitigating or responding to its effects. Climate change entities may be identified using domain knowledge. However, any causal or relational link between these entities and food practices must be explicitly supported by the text. The subtypes of entities are defined according to entities selected by the expert annotator from the Climate-change-NER corpus (Bhattacharjee et al., 2024), as well as an additional environmental concern type that was judged relevant by our expert:

- **Climate greenhouse gases:** Gases that cause heating of the atmosphere (e.g., “carbon dioxide”, “CH4”, “GHG emission”)
- **Climate hazards:** Dangers with potential negative impact on climate, such as floods, wildfires, droughts, and heatwaves.
- **Climate impacts:** Effects of hazards, primarily negative effects on humans. We also consider climate impacts on livestock as impacts, as it indirectly affects humans.
- **Climate problem origins:** Problems that describe why the climate is changing. Key examples are fossil fuel and deforestation. We also

mention sectors that can be cited as causes of energy use. For instance, in a text about the energy consumption by the transport sector, the transport sector is annotated as a problem.

- **Climate mitigations:** Activities to reduce climate change or to better deal with the consequences.
- **Environmental concern:** This entity encompasses several aspects, e.g., pollution, biodiversity, waste, water, climate, etc.; it does not specifically refer to climate change; it is an intermediate/residual category within the environmental sphere for which the link to climate is not established or direct.

We also provide additional annotation rules in our annotation guidelines (Appendix A).

4 Experimental Setup

We want to evaluate two existing baseline methods and also a custom prompting method using recent state-of-the-art LLMs (baseline-comparable open-weight models and a large proprietary model) because the progress of LLMs is moving fast and the baselines may no longer be competitive even with task-specific fine-tuning. We also perform a pre-processing of the corpus before running the methods.

4.1 Corpus pre-processing

We ran the experiments on a pre-processed version of the corpus. The pre-processing is described here:

1. We replaced the references with a [REF] tag for single references and a [MREF] tag for multi-references, using regular expressions.
2. We split the documents into sentences using the spaCy sentencizer².

Even if the annotations are performed at the document level, we choose to split them into sentences, as some documents were long, and preliminary experiments showed that having a context too long was hurting performances for LLMs (possibly due to the lost-in-the-middle problem when giving large contexts to LLMs (Liu et al., 2023)). The citation replacement is principally added for easier sentence splitting and because we noticed a lot of multi-citations that did not add any useful context and were taking a lot of space in some sequences.

²<https://spacy.io/api/sentencizer/>

4.2 Existing Baseline Models

We select the following existing methods to have a lightweight encoder baseline and a fine-tuned LLM baseline:

- GLiNER (Zaratiana et al., 2023): It is a suite of encoder-based models where they fine-tuned a DeBERTa model using special [ENT] tokens that precede entity type labels on a large corpus with more than 13k different entity types automatically annotated using ChatGPT (gpt-3.5-turbo-0301). Then the special tokens and entities are concatenated with the sentence in which we want to find entities (using a special [SEP] token), and a dot product is computed between the input sentence token embeddings and the different [ENT] tokens for each entity type. We specifically use version 2.5 and the large variant (459M parameters) of the model (gliner_large-v2.5³).
- GoLLIE (Sainz et al., 2024): GoLLIE is a suite of decoder-based models fine-tuned on multiple information extraction tasks. The approach represents annotation guidelines as code syntax, allowing the model to use in-context learning to identify entity boundaries. The authors fine-tuned different sizes of CodeLLaMA models, transforming annotation guidelines (entity type labels, descriptions, and examples) in Python syntax, defining each entity type as a class. The input text is then appended as a Python string variable in the prompt, and the model generates the entity mentions as a structured list format. We use the 7B parameter (GoLLIE-7B) and 34B parameter (GoLLIE-34B) variants of the model (smaller and bigger one).

4.3 Custom Prompting Method

Because GoLLIE is based on an already outdated model and LLM capabilities have progressed significantly since then, we also use a custom prompting method so that we can evaluate more recent LLMs. We design a prompt that includes all the guideline information. We use one prompt for each document in the corpus, and we handle all entity types at once. Preliminary experiments strongly suggest that it helps disambiguate closely-related entity types (compared to handling one entity type

³https://huggingface.co/gliner-community/gliner_large-v2.5/tree/main

at a time), and it is also faster and cheaper as it divides the number of inferences by the number of entity types. Recent LLMs follow a system/user roles schema, so we design our prompt following this. Here is the composition of this prompt:

- For the system part: contains a basic persona (where we specify the model as expert in NER annotation)
- For the user part:
 1. Task instruction: simple sentence describing the NER task
 2. Entity types, names, and descriptions, and if available, entity mention examples
 3. Rules: Corpus-specific annotation guidelines if available, presented as a mark-down section
 4. Output format: We require the output to follow a specific format, with an illustrative example provided. Here we only tried the JSON format.
 5. Then finally the input text from which we want to extract entities.

The full template is provided in Appendix D.

4.4 Models and configuration

We choose recent LLMs to evaluate this prompting method. We first choose two open-weight models that have a size comparable to the baseline LLM model (GoLLIE). We also add a large model that has state-of-the-art performance on text generation benchmarks, as well as a proprietary model:

- Gemma 4 (Team et al., 2026): Gemma 4 is a suite of vision-language models. We select the dense 31B parameter, instruction-tuned (gemma-4-31B-it) version of this model for our experiments.
- Qwen 3.5 (Qwen Team, 2026): Qwen 3.5 is also a suite of vision-language models. We try two variants of this model: the dense 27B parameter version (Qwen3.5-27B) and the dense 9B parameter version (Qwen3.5-9B). We choose these versions so that their size is close to the models used by the GoLLIE baseline.
- GPT 5.6 Sol (Singh et al., 2026): this is the latest⁴ and (supposedly) largest variant of the

GPT 5.6 model. According to its authors, it achieves state-of-the-art performance on many LLM textual benchmarks⁵.

We run the models in greedy decoding, without thinking tokens, on the corpus sentences (after preprocessing the documents) so that the setting stays comparable with the GoLLIE baseline. It also gives a good idea of the performance of the recent models without having to use too much compute by using a multi-pass sampling evaluation and test-time compute.

We ran the open-weight models on single NVIDIA A100 SXM4 80 GB GPUs (except for GLiNER, which we ran on a local GPU NVIDIA RTX A2000 Laptop GPU). The proprietary GPT-5.6 model was ran using the OpenRouter API.

4.5 Evaluation

We perform the evaluation at the entity level and because the corpus has challenging mentions (discontinuous, nested entities), we use two modes of evaluation:

- **Partial:** we count a correct prediction if the predicted mention boundaries at least have an overlap with the true mention and if the predicted type is exact. This removes the discontinuous challenge; if there are multiple spans and the prediction overlaps with one of the spans, it will be counted as correct.
- **Strict:** we count a correct prediction if the boundaries strictly match (multi-spans for discontinuous entities must exactly match, so our methods do not handle this; all discontinuous mentions will be counted as incorrect) and if the predicted type is correct.

We then compute the micro-average for the metrics of precision, recall, and F1-score in both modes.

5 Results

We provide in this section our results on the Food-ClimNER corpus; additional results on the simple variant of this corpus are included in appendix B.

5.1 Resulting corpus

We provide the statistics of our resulting corpus in Tables 1 (for general corpus statistics) and 2 (for the entity type number).

⁴at the time of our last experiments in July 2026

⁵<https://openai.com/index/gpt-5-6/>

Statistics	Value
Number of initial documents	10
Total nb. of words	13599
Number of sentences	527
Sent. avg length (char)	172.7 $\pm$ 88
Sent. avg length (word)	25.8 $\pm$ 13
Number of entity types	12
Number of mentions	1109
Number of unique mentions	727
Mention avg length (char)	26.7 $\pm$ 17
Mention avg length (word)	3.8 $\pm$ 3
Avg mentions per sentence	2.10
Discontinuous mentions	92 (8%)
Nested mentions	478 (43%)
Flat mentions	539 (49%)

Table 1: Corpus statistics for FoodClimNER

Entity Type	Count
Food Repertoire	297 (27%)
Food Preparation	132 (12%)
Food Supply Purchases	101 (9%)
Food Waste Management	72 (6%)
Food Storage	40 (4%)
Food Logistics Travel	36 (3%)
Food Consumption Uses	11 (1%)
Climate Mitigations	222 (20%)
Climate Problem Origins	136 (12%)
Environmental Concern	37 (3%)
Climate Impacts	19 (2%)
Climate Greenhouse Gases	6 (1%)
Climate Hazards	0 (0%)

Table 2: Entity type statistics for FoodClimNER

These statistics include sentence-level statistics, as our model inputs were sentences (split from the initial documents). We observe that annotated mentions are long (almost 4 words per mention in average), and that the corpus have a substantial part of nested entities (43 %, here a nested mention means that it has at least an overlap of one character with another mention). Our annotator did not identify any climate hazard entities in the selected documents, suggesting that the annotated texts primarily focus on other climate-related entities.

5.2 Overall results for both versions of our corpus

We provide the overall results (micro-average metrics of the different methods used) on the FoodClimNER corpus in Table 3.

First, looking at the baseline results, as expected, as we have expert definitions in our corpus, it is a really difficult task for GLiNER, which only supports entity type labels. GoLLIE performs better as it allows adding textual descriptions of each entity type, as well as examples. Similarly, our custom prompt containing information about the task (entity definitions, examples, annotation rules) performs even better and without any fine-tuning. However, the models used are more recent than the backbone model of GoLLIE (CodeLLaMa), so their capabilities have greatly improved since then. Surprisingly, we observe that Gemma 4 obtains the best F1 score overall, even beating GPT-5.6 Sol. The task remains difficult overall (partly due to multiple challenges such as discontinuous entities, which are not handled by the methods we used).

5.3 Entity-type-level results

To better understand where the overall performance gains and losses originate, we report entity-type-level results for the best-performing model on FoodClimNER: Gemma-4-31B-it (Table 4).

For the different food practice entity types, we globally observe reasonable performances given the zero-shot setting and corpus difficulty (between 31 and 51 F1 points), meaning our expert guidelines given to the model are useful, as verified in our ablation experiment (Section 6). For the climate change entity types, we notice greenhouse gases are easy to detect (lexically simple as there are not a lot of different possibilities), while impacts and problem origins obtain low performance (around 6 F1 points) showing that these types are harder to detect (very low recall for both) and might benefit from additional examples and refined definitions.

6 Ablation Study

We perform a basic ablation study on our custom prompt method, showing how each part contributes to the final performance by doing 4 runs, removing each time a different part of the prompt. We perform this ablation only on our best performing model, Gemma 4. The results are presented in Table 5. For the "without persona" experiment, instead of stating a NER-specific persona, we just give a neutral persona.

The ablation results show that every component of the prompt contributes positively to the performance, since removing any single element degrades the F1 relative to the full configuration.

Model	Partial			Strict		
	P	R	F1	P	R	F1
<i>Baselines</i>						
GLiNER	17.83	34.36	23.48 [21.21, 25.89]	6.88	13.26	9.06 [7.30, 10.90]
GoLLIE-7B	50.64	25.07	33.53 [30.65, 36.53]	25.32	12.53	16.77 [14.25, 19.39]
GoLLIE-34B	47.91	26.65	34.25 [30.57, 37.98]	16.98	9.44	12.14 [9.61, 14.79]
<i>Prompting</i>						
Gemma4	62.25	49.70	55.27 [51.97, 58.44]*	42.33	33.79	37.58 [33.89, 41.21]*
Qwen3.5-9B	48.10	32.01	38.44 [35.13, 41.71]	29.54	19.66	23.61 [20.31, 26.85]
Qwen3.5-27B	56.23	46.80	51.08 [47.92, 54.05]	36.51	30.39	33.17 [30.00, 36.34]
GPT5.6-Sol	56.00	40.40	46.94 [43.19, 50.63]	37.62	27.14	31.53 [28.01, 35.04]

Table 3: Models performance comparison on FoodClimNER (Partial, Strict). Best per metric in bold, second-best underlined (descriptive). F1 shown as point estimate [95% document-level bootstrap confidence interval (CI)]. * = significantly better than all others (paired bootstrap, B=10000)

Entity Type	Partial			Strict		
	P	R	F1	P	R	F1
Food Consumption Uses	23.81	45.45	31.25	23.81	45.45	31.25
Food Preparation	86.00	65.15	74.14	55.00	41.67	47.41
Food Repertoire	75.00	57.58	65.14	58.77	45.12	51.05
Food Logistics Travel	87.50	58.33	70.00	50.00	33.33	40.00
Food Storage	41.38	60.00	48.98	29.31	42.50	34.69
Food Supply Purchases	52.94	44.55	48.39	36.47	30.69	33.33
Food Waste Management	59.26	66.67	62.75	41.98	47.22	44.44
Climate Greenhouse Gases	71.43	83.33	76.92	57.14	66.67	61.54
Climate Impacts	33.33	21.05	25.81	8.33	5.26	6.45
Climate Mitigations	64.10	45.05	52.91	43.59	30.63	35.98
Climate Problem Origins	63.64	5.15	9.52	45.45	3.68	6.80
Environmental Concern	25.51	67.57	37.04	17.35	45.95	25.19

Table 4: Entity-type performance on FoodClimNER (Partial, Strict) for gemma-4-it-31-B.

Model	P	R	F1
Full prompt	43.47	34.54	38.49
w/o persona	42.26	33.99	37.68
w/o definitions	43.68	30.21	35.71
w/o rules	39.49	32.37	35.58
w/o examples	34.70	28.22	31.13
w/o all	15.71	16.86	16.27

Table 5: Ablation performances of our custom prompt for gemma-4-31B-it on FoodClimNER (strict evaluation only).

The most critical component is the set of examples: removing them causes the largest drop, suggesting that few-shot demonstrations are essential for the model to correctly delimit and classify entities. Removing the rules and definitions each cost around 2.8-2.9 points, indicating that explicit entity definitions and annotation guidelines meaningfully

help disambiguate borderline cases but are less decisive than concrete examples. The persona component has the smallest impact, implying that framing the model’s role contributes only marginally to the extraction quality compared to the more content-bearing parts of the prompt (examples, definitions, rules). Removing all parts goes back to giving only the entity type labels to the model, which drastically reduces its capacity (losing half of its maximum points for F1-score).

7 Conclusion

We introduced FoodClimNER, a manually annotated NER corpus extracted from 10 open-access food science and climate change articles, together with expert-crafted annotation guidelines covering seven food practice subtypes and six climate change subtypes, including nested and discontinuous mentions. We evaluated two existing base-

lines (GLiNER, GoLLIE) and a custom zero-shot prompting method relying on explicit entity type definitions and examples, using several recent open-weight and proprietary LLMs.

Our results show that providing full-text definitions, rather than only entity type labels, substantially helps zero-shot LLMs on this task: prompting-based approaches outperform both baselines despite using no task-specific fine-tuning. However, the task remains far from solved. Even the best model, Gemma 4 31B-it, only reaches a strict F1 of 38.49 on the full schema. Climate problem origins in particular prove considerably harder to recognize than food practice subtypes, independently of their frequency in the corpus. This confirms that our expert-defined entity types constitute a genuinely challenging testbed for current zero-shot NER methods and suggests that gains on more generic NER benchmarks do not necessarily transfer to expert-defined, domain-specific schemas such as ours. Although strict span-level evaluation is challenging, the predictions remain useful for assisting experts in identifying relevant text segments for manual review.

We hope FoodClimNER can serve both as a resource for evaluating future NER methods on expert-defined, fine-grained schemas with nested and discontinuous mentions and as a stepping stone toward downstream applications aiming to detect emerging, climate change-driven food practices that could represent future microbiological risks.

8 Limitations

Annotator agreement As we have only one expert annotating the corpus, we cannot compute an agreement for the different entities, and the annotation is subject to bias. However, our annotator has multiple expertise so having another annotator with a similar expertise was difficult to find in the limited time we had to develop the corpus.

Discontinuous entities The methods we use do not handle the discontinuous entities, but they constitute only a small part of the evaluation corpus (8 %).

Custom Prompt The prompt we developed has all the expert-defined annotation rules, but we did not try to optimize it more than this; it could probably be more detailed. Moreover, we did not try to change the output prompt format, which can also influence the performance of the model.

Size of our corpus The size of the corpus is limited because expert annotation is expensive. So we use it only to evaluate zero-shot methods, as there is not enough data to train a model. However, this is only a first version of the corpus, and we plan to extend it with additional documents, more annotations, and better representation of currently underrepresented or absent entity types, potentially including other text sources such as press articles and social media.

9 Acknowledgments

This work was funded by the French national research agency ANR under the reference ANR-24-CE21-0244. It was also provided with computing and storage resources by GENCI at IDRIS thanks to the grant 2026-AD010317377 on the supercomputer Jean-Zay’s A100 partition .

References

- Rami Aly, Andreas Vlachos, and Ryan McDonald. 2021. [Leveraging Type Descriptions for Zero-shot Named Entity Recognition and Classification](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1516–1528, Online. Association for Computational Linguistics.
- Bishwaranjan Bhattacharjee, Aashka Trivedi, Masayasu Muraoka, Muthukumaran Ramasubramanian, Takuma Udagawa, Iksha Gurung, Nishan Pantha, Rong Zhang, Bharath Dandala, Rahul Ramachandran, Manil Maskey, Kaylin Bugbee, Mike Little, Elizabeth Fancher, Irina Gerasimov, Armin Mehrabian, Lauren Sanders, Sylvain Costes, Sergi Blanco-Cuaresma, and 17 others. 2024. [INDUS: Effective and Efficient Language Models for Scientific Applications](#). Preprint, arXiv:2405.10725.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Ruichu Cai, Junhao Lu, Zhongjie Chen, Boyan Xu, and Zhifeng Hao. 2025. [Handling Missing Entities in Zero-Shot Named Entity Recognition: Integrated Recall and Retrieval Augmentation](#). In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume*

- I: Long Papers*), pages 10790–10802, Albuquerque, New Mexico. Association for Computational Linguistics.
- Alessio Cocchieri, Marcos Martínez Galindo, Giacomo Frisoni, Gianluca Moro, Claudio Sartori, and Giuseppe Tagliavini. 2025. [ZeroNER: Fueling Zero-Shot Named Entity Recognition via Entity Type Descriptions](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 15594–15616, Vienna, Austria. Association for Computational Linguistics.
- Dominique Desjeux. 1998. Les échelles d’observation de la consommation. *Comprendre le consommateur*. Paris: Éditions Sciences Humaines.
- Jessica Dietrich, Jens-Andre Hammerl, Annette John, Oliver Kappenstein, Christopher Loeffler, Karsten Nöckler, Bettina Rosner, Astrid Spielmeyer, Istvan Szabo, and Martin H Richter. 2023. Impact of climate change on foodborne infections and intoxications. *Journal of health monitoring*, 8(Suppl 3):78.
- Ramona A. Duchenne-Moutien and Hudaa Neetoo. 2021. [Climate change and emerging food safety issues: A review](#). *Journal of Food Protection*, 84(11):1884–1897.
- EFSA. 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA journal*, 8(6):1637.
- Ana Gjorgjevikj, Matej Martinc, Gjorgjina Cenikj, Riste Stojanov, Jan Drole, Gordana Ispirova, Giulia Menichetti, Nives Ogrinc, Dimitar Trajanov, Sašo Džeroski, Barbara Koroušić Seljak, and Tome Efimov. 2026. [Large language models in food and nutrition science: Opportunities, challenges, and the case of FoodyLLM](#). *Current Research in Food Science*, 12:101351.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shiron Ma, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, and 175 others. 2025. [DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning](#). *Nature*, 645(8081):633–638.
- Kristie Hadley, Stefan Wheat, Heidi Honegger Rogers, Arjun Balakumar, Diana Gonzales-Pacheco, Sarah Shrum Davis, Hanna Linstadt, Tracy Cushing, Lewis H Ziska, Christi Piper, and 1 others. 2023. Mechanisms underlying food insecurity in the aftermath of climate-related shocks: a systematic review. *The Lancet Planetary Health*, 7(3):e242–e250.
- Morgane Innocent, Agnes Francois Lecompte, Samuel Guillemot, and Ronan Divard. 2024. Environmentally sustainable food: an analysis of consumer practices and their diffusion in france. *British Food Journal*, 126(2):548–577.
- Intergovernmental Panel on Climate Change (IPCC). 2023. *Food, Fibre and Other Ecosystem Products*, page 713–906. Cambridge University Press.
- Gyula Kasza, Nina Veflen, Joachim Scholderer, Lars Münter, László Fekete, Eszter Zita Csenki, Annamária Dorkó, Dávid Szakos, and Tekla Izsó. 2022. [Conflicting issues of sustainable consumption and food safety: Risky consumer behaviors in reducing food waste and plastic packaging](#). *Foods*, 11(21).
- Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large Language Models are Zero-Shot Reasoners. *Advances in Neural Information Processing Systems*, 35:22199–22213.
- Yann Le Bodo, Romaric Caillet, and Estelle Chaix. 2026. [Mapping Climate-Related Food Practices: A Data-Driven Framework to Anticipate Microbiological Risks](#). FoodMicro2026. Poster.
- Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. [Lost in the Middle: How Language Models Use Long Contexts](#). *Preprint*, arXiv:2307.03172.
- Marina Arribas Lopez, Yamine Bouzembrak, and Bedir Tekinerdogan. 2026. [Artificial Intelligence in Food Safety: A Tertiary Study](#). *Comprehensive Reviews in Food Science and Food Safety*, 25(2):e70443.
- Jie Lou, Yaojie Lu, Dai Dai, Wei Jia, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2023. [Universal information extraction as unified semantic matching](#). In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, volume 37 of AAAI’23/IAAI’23/EAAI’23, pages 13318–13326. AAAI Press.
- Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2022. [Unified Structure Generation for Universal Information Extraction](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5755–5772, Dublin, Ireland. Association for Computational Linguistics.
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. [Training language models to follow instructions with human feedback](#). *Preprint*, arXiv:2203.02155.
- Huitong Pan, Mustapha Adamu, Qi Zhang, Eduard Dragut, and Longin Jan Latecki. 2025. [ClimateIE: A Dataset for Climate Science Information Extraction](#). In *Proceedings of the 2nd Workshop on Natural Language Processing Meets Climate Change*

- (*ClimateNLP* 2025), pages 76–98, Vienna, Austria. Association for Computational Linguistics.
- Gorjan Popovski, Barbara Koroušić Seljak, and Tome Eftimov. 2019. *FoodBase corpus: A new resource of annotated food entities*. *Database: The Journal of Biological Databases and Curation*, 2019:baz121.
- Qwen Team. 2026. *Qwen3.5: Towards native multi-modal agents*.
- Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, and Eneko Agirre. 2024. *GoLLIE: Annotation Guidelines improve Zero-Shot Information-Extraction*. *Preprint*, arXiv:2310.03668.
- Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, A. J. Ostrow, Akhila Ananthram, Akshay Nathan, Alan Luo, Alec Helyar, Aleksander Madry, Aleksandr Efremov, Aleksandra Spyra, Alex Baker-Whitcomb, Alex Beutel, Alex Karpenko, and 467 others. 2026. *OpenAI GPT-5 System Card*. *Preprint*, arXiv:2601.03267.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: A Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics.
- Gemma Team, Sherif El Abd, Vaibhav Aggarwal, Robin Algayres, Alek Andreev, Olivier Bachem, Ian Ballantyne, Cormac Brick, Victor Cărbune, Michelle Casbon, Mayank Chaturvedi, Victor Cotruta, Alice Coucke, Phil Culliton, Robert Dadashi, Lucas Dixon, Mohamed Elhawaty, Utku Evci, Clément Farabet, and 282 others. 2026. *Gemma 4 Technical Report*. *Preprint*, arXiv:2607.02770.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. *LLaMA: Open and Efficient Foundation Language Models*. *Preprint*, arXiv:2302.13971.
- Sonja J. Vermeulen, Bruce M. Campbell, and John S.I. Ingram. 2012. *Climate change and food systems*. *Annual Review of Environment and Resources*, 37(Volume 37, 2012):195–222.
- Xiao Wang, Weikang Zhou, Can Zu, Han Xia, Tianze Chen, Yuansen Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, Jihua Kang, Jingsheng Yang, Siyuan Li, and Chunsai Du. 2023. *InstructUIE: Multi-task Instruction Tuning for Unified Information Extraction*. *Preprint*, arXiv:2304.08085.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. *Preprint*, arXiv:2201.11903.
- Ania Wróblewska, Agnieszka Kaliska, Maciej Pawłowski, Dawid Wiśniewski, Witold Sosnowski, and Agnieszka Ławrynowicz. 2022. *TASTEset – Recipe Dataset and Food Entities Recognition Benchmark*. *Preprint*, arXiv:2204.07775.
- Tingyu Xie, Qi Li, Jian Zhang, Yan Zhang, Zuo Zhu Liu, and Hongwei Wang. 2023. *Empirical Study of Zero-Shot NER with ChatGPT*. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7935–7956, Singapore. Association for Computational Linguistics.
- Urchade Zaratiana, Gil Pasternak, Oliver Boyd, George Hurn-Maloney, and Ash Lewis. 2025. *GLiNER2: Schema-Driven Multi-Task Learning for Structured Information Extraction*. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 130–140, Suzhou, China. Association for Computational Linguistics.
- Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. 2023. *GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer*. *Preprint*, arXiv:2311.08526.
- Wenxuan Zhou, Sheng Zhang, Yu Gu, Muhao Chen, and Hoifung Poon. 2024. *UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition*. *Preprint*, arXiv:2308.03279.

A Annotation guidelines

In this section, we specify, in addition to section 3.2, the rules created by our annotator to help annotate the defined entity types in some specific and ambiguous cases. First, we define rules for specific types of entities, starting with food practices entities:

- For mentions of "behavior" related to food or diet, it should be annotated as "Food Repertoire" only if the text links these behaviors explicitly to climate change.
- Food practices must explicitly or implicitly involve a consumer. Macro-level phenomena, or those occurring downstream in the food chain (e.g., primary production, industrial processes), are not considered practices. The consumer must be involved directly when a food practice is annotated.
- Do not annotate mentions where we cannot identify a clear practice (e.g., consumer habits); too generic a term is not informative.

- Do not annotate food items (e.g., meat dishes, sustainable food) but take care of similar forms that refer to repertoire and that must be annotated (e.g., meat-free dishes).
- Food categories (e.g., wild foods, edible insects, organic foods) are not annotated as food practices unless they are embedded in a consumption or dietary context (e.g., consumption, diet, intake, eating patterns).
- Cognitive constructs are excluded unless they directly refer to operationalized food preparation activities or competencies immediately applied in food acquisition, preparation, or consumption.
- Include modifiers expressing intensity or direction when a practice is part of the observed behavioral change (e.g., increase/decrease).
- A practice that includes negation is still a practice and should be annotated (e.g., "Do not rinse meat", "Do not pay attention to expiration dates").

Then we also have rules for climate change entities :

- If a climate hazard is named in more detail in a text, the entire term is annotated (e.g., surface water flood or soil liquefaction).
- For the climate impact entity, vector-borne diseases and other diseases explicitly described as climate-sensitive are annotated as impacts. COVID-19 is annotated as an impact only when the text explicitly establishes a link between climate change and the pandemic.
- If an increase or decrease in activity reduces climate change, it should be annotated as a Climate Mitigation (e.g., "food waste has seen a notable decrease in volume").
- Transitions, strategies, policies, or transformations explicitly aimed at the sustainability of a system should be annotated as climate mitigation even when climate change is not explicitly mentioned.
- Environmental concern includes biodiversity, the ecological footprint, etc. Environmental concern includes generic or non-actionable references to sustainability, environmental awareness, or trends in consumption that do not describe explicit consumer actions.

A.1 Global rules

- Discontinuous annotation: If a practice is expressed as a predicate or attribute of diets (e.g., be varied, be seasonal), annotate the full conceptual unit of dietary recommendation using discontinuous spans including the implicit subject (e.g., diets should), and include all coordinated attributes as separate linked spans.
- Nested annotation: a food practice can have the same exact boundary as another entity type. For example, as some practices may have a mitigating effect on climate change, they can also be a climate mitigation entity. In the case of discontinuous mentions, we might even have some spans that have the same types for two different discontinuous mentions (but the other spans composing these mentions will be different).

B Simplified corpus results

As mentioned in section 3.2, we evaluate on two versions of our corpus, FoodClimNER (containing all entity types) and FoodClimNER-simple (aggregating all food practice types into a single type and all climate change types in a single type as well). In this section we provide the results of FoodClimNER-simple in Tables 6 and 7.

Model	P	R	F1
<i>Baselines</i>			
GLiNER	33.69	26.06	29.38
GoLLIE-7B	35.39	15.23	21.30
GoLLIE-34B	38.82	17.10	23.74
<i>Custom Prompt</i>			
Gemma 4 31B-it	57.91	<u>32.29</u>	41.47
Qwen 3.5-9B	54.91	18.99	28.22
Qwen 3.5-27B	<u>55.67</u>	25.69	35.15
GPT-5.6 Sol	51.35	38.44	43.97

Table 6: Strict micro-averaged metrics on the FoodClimNER-simple corpus. Best per metric in bold, second-best underlined.

This time, the results are different from those on the harder version of the corpus. First, the GLiNER model performs better than the GoLLIE models. entity type labels are already sufficient information because they are easier to apprehend than the more fine-grained types.

Globally, the models achieve better performance with these aggregated entity types. We also observe that GPT-5.6 obtains the best results this time.

Our custom prompt with more recent models still beats all the baselines (except for the Qwen 3.5 9B parameter model, which performs slightly below GLiNER).

Similarly to the fine-grained type evaluation, performance is higher for food practice entities than for climate change entities (Table 7).

Entity Type	P	R	F1
Food Practice	56.22	43.98	49.35
Climate Change	41.88	28.93	34.22

Table 7: Strict entity-type performances on FoodClimNER-simple (Partial, Strict) for gpt-5.6-sol.

C Compute cost and environmental cost

We measure the number of GPU hours spent on our experiments and the estimated environmental cost on our GPU cluster. We spent 135⁶ GPU hours (on NVIDIA A100 80 GB GPUs) running all the open-weight models on the two versions of our corpus (FoodClimNER and FoodClimNER-simple) which translates to 8.5 kg CO2e according to the estimation given by our GPU cluster. We spent \$12 for running the experiment with the OpenRouter API (for the GPT-5.6 Sol model). However, we do not have any way to compute an environmental cost for this model.

D Complete Custom Prompt

In Figure 3, we provide the full template filled with our corpus entity type labels, definitions, examples, and annotation rules.

⁶This is the value given by our cluster in normalized V100 equivalent hours, because they attribute GPU hours budgets depending on the hardware we ask (higher rate for better GPUs). So for A100 GPUs, hours are counted double so it is effectively 67.5 A100 hours.

System prompt

You are a helpful assistant specialized in named entity recognition.

User prompt

Extract all mentions of the declared entities types from the input text in the specified format.

Entity Types

- * climate_greenhouse_gases
 - **Definition:** Gases that cause heating of the atmosphere.
 - **Examples:** carbon dioxide, CH₄, GHG emission
- * climate_hazards
 - **Definition:** Hazards with potential negative impact on climate. Where a hazard is named in more detail in a text, the entire term is annotated.
 - **Examples:** floods, wildfires, droughts, heatwaves, surface water flood, soil liquefaction
- * climate_impacts
 - **Definition:** Effects of hazards, primarily negative effects on humans. Climate impacts on livestock are also considered impacts, as they indirectly affect humans. COVID-19 is annotated as impact only when the text explicitly establishes a link between climate change and the pandemic.
 - **Examples:** Vector-borne diseases, diseases explicitly described as climate-sensitive
- * climate_mitigations
 - **Definition:** Activities to reduce climate change or to better deal with the consequences. If an increase or decrease in activity reduces climate change, it should be annotated. Transitions, strategies, policies, or transformations explicitly aimed at the sustainability of a system should be annotated even when climate change is not explicitly mentioned.
 - **Examples:** food waste reduction, transitions aimed at sustainability, strategies aimed at sustainability, policies aimed at sustainability
- * climate_problem_origins
 - **Definition:** Problems that describe why the climate is changing. Sectors cited as causes of energy use are also annotated as problem origins.
 - **Examples:** fossil fuel, deforestation, transport sector
- * environmental_concern
 - **Definition:** An intermediate/residual category within the environmental sphere for which the link to climate is not established or direct. It does not specifically refer to climate change. Includes generic or non-actionable references to sustainability, environmental awareness, or trends in consumption that do not describe explicit consumer actions.
 - **Examples:** biodiversity, ecological footprint, pollution, water
- * food_consumption_uses
 - **Definition:** Covers consumption practices related to eating and meals.
 - **Examples:** portion sizes, choice of utensils
- * food_culinary_preparation
 - **Definition:** All attitudes and practices related to food preparation and cooking. It also includes skills and actions that we perform when cooking.
 - **Examples:** raw food diets, water conservation during cooking, reducing energy consumption in cooking, minimizing waste during preparation
- * food_repertoire
 - **Definition:** Covers practices and factors related to food choices (in terms of quality or quantity) that may or may not be part of an individual's diet. Mentions of behaviour related to food or diet should be annotated only if the text links these behaviours explicitly to climate change.
 - **Examples:** meat consumption, food choices
- * food_logistics_travel
 - **Definition:** Includes pre-purchase practices and travel to the place of purchase.
 - **Examples:** checking existing resources before shopping, meal planning, using low-carbon transportation to go grocery shopping
- * food_storage
 - **Definition:** Covers food storage practices after purchase and before preparation.
 - **Examples:** choice of packaging, adherence to expiration dates, disregard of expiration dates, preservation techniques
- * food_supply_purchases
 - **Definition:** Covers purchasing practices (product selection, choice of retail locations, interactions with marketing) as well as non-market procurement strategies.
 - **Examples:** product selection, choice of retail locations, interactions with marketing, self-production
- * food_waste_management
 - **Definition:** Covers practices that take place after a meal, with consideration for the end of a food item's life cycle.
 - **Examples:** storing leftovers, composting, sorting food waste, expiration date reasoning

Global Annotation Conventions

- Discontinuous annotation: if a practice is expressed as a predicate or attribute of diets (e.g. be varied, be seasonal), annotate the full conceptual unit of dietary recommendation using discontinuous spans including the implicit subject (e.g. diets should), and include all coordinated attributes as separate linked spans.
- Nested annotation: a food practice can have an exact boundary as another entity type. For example, as some practices may have a mitigating effect on climate change they can also be a climate mitigation entity. In the case of discontinuous mentions, we might have some spans that have the same types for two different discontinuous mentions (but the other spans composing these mentions will be different).

Food Practices General Guidelines

- Food practices must explicitly or implicitly involve a consumer. Macro-level phenomena, or those occurring downstream in the food chain (e.g., primary production, industrial processes), are not considered practices. The consumer must be involved directly when a food practice is annotated.
- Do not annotate mentions where we cannot identify a clear practice (e.g. consumer habits) -- a too generic term is not informative.
- Do not annotate food items (e.g. meat dish, sustainable food) but take care of similar forms that refer to repertoire and that must be annotated (e.g. meat-free dishes).
- Food categories (e.g., wild foods, edible insects, organic foods) are not annotated as food practices unless they are embedded in a consumption or dietary context (e.g. consumption, diet, intake, eating patterns).
- Cognitive constructs are excluded unless they directly refer to operationalized food preparation activities or competencies immediately applied in food acquisition, preparation, or consumption.
- Include modifiers expressing intensity or direction when a practice is part of the observed behavioral change (e.g., increase/decrease).
- A practice that includes negation is still a practice and should be annotated (e.g. "Do not rinse meat", "not pay attention to expiration dates").

Output Format

Return a JSON object where each key is an entity type and the value is a list of all mentions of that entity type found in the text.

Example:

```
““json
```

```
{
  "entity_type1" : [
    "mention_type1_1",
    "mention_type1_2",
    "...",
  ],
  "entity_type2" : [
    "mention_type2_1",
    "...",
  ],
  "...",
}
```

```
““
```

Return only valid JSON, no additional text.

Input Text

```
““txt
{input_text}
““
```

Figure 3: Prompt used for LLM-based named entity recognition on the FoodClimNER corpus.