holi food

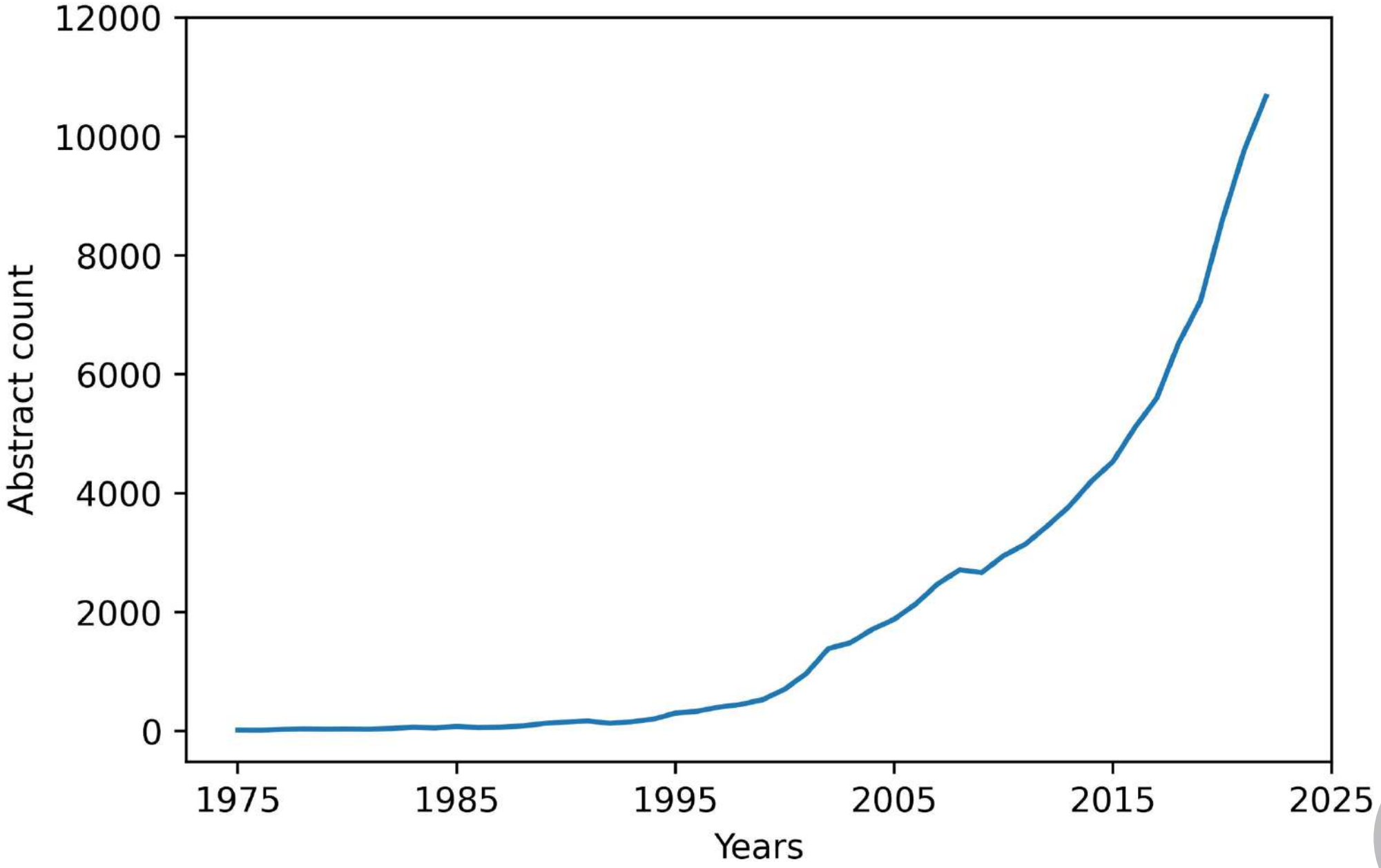
Funded by
the European Union

Large Language Models

Floor van Meer

June 2026

All fake-looking images are made with AI



van den Bulk et al. Current Research in Food Science 5 (2022): 84-95.

Large language model (LLM)

Gemini

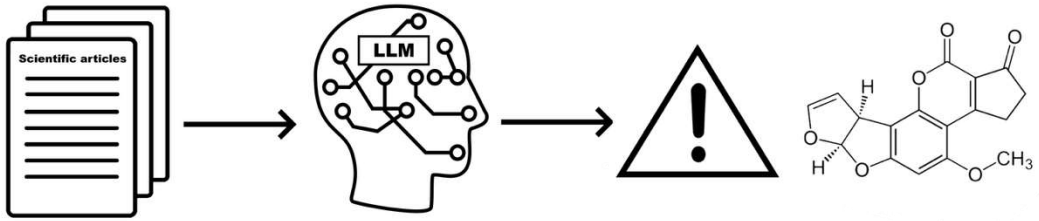
Copilot

Hugging Face

deepseek

Introduction

- Thousands of publications in food safety per year
- The need to swiftly identify emerging chemical hazards
- State of the art performance of LLMs (e.g. ChatGPT)
- Investigated automatic extraction of chemical hazards from the literature
- Built a dashboard for domain experts



Key steps

1. Collected relevant literature
   - ~134k abstracts with keywords on chemical contamination and negative health effects
2. Used a small LLM in two different scenarios
   - instructing & retraining
3. Post-processed LLM outputs
4. Fed post-processed outputs into the dashboard


holj food

Scenarios

1. Instructing an LLM

LLM used with no changes

Not labor intensive

Instructions in natural language

Various instructions explored

holj food

Instructing - Workflow

Step by step prompt

```
Your task is to perform the following actions:

1. Identify which foods are mentioned in the text below provided between triple
backticks and collect them in a list, only collect items that are edible products.

2. Identify which chemicals are mentioned in the text below provided between
triple backticks and collect them in a list, be sure to exclude microorganisms
such as bacteria, fungi, or other biological entities.

3. Create all combinations of the foods identified in action 1 and chemicals
identified in action 2 as tuples and collect the tuples in a list.

4. Go over each combination of food and chemical in the list created at action 3
and look whether the chemical is mentioned to either be a food safety hazard for
that food, to be hazardous, to contaminate that food, to accumulate in that food,
to have a maximum residue limit (MRL) or to have the potential to pose risk for
human health via consumption of that food in the text below provided between
triple backticks. If the chemical is said to either be a food safety hazard, to be
hazardous, to be a contaminant, to accumulate, to have a MRL or to have the
potential to pose risk for human health via consumption of the food, then add that
combination of food and chemical to a dictionary where each key is a food and the
values are a list of chemicals.

I want to warn you against some pitfalls. First, if chemicals are mentioned both
with their names and abbreviations, make sure to return the full name of the
chemical instead of its abbreviation. Another thing - refrain from providing
irrelevant noun phrases or sentences in values just because they contain chemical
names, limit the values of your dictionary to the names of relevant chemicals.
Return an empty dictionary if you do not identify any foods that are edible
products. Also return an empty dictionary if you do not identify any chemicals.
Furthermore, return an empty dictionary if you do not identify any chemical in
food as either a food safety hazard, hazardous, a contaminant, accumulating,
without MRL or potentially harmful for human health through consumption of food.

Use the following format for your answer and limit your explanation:

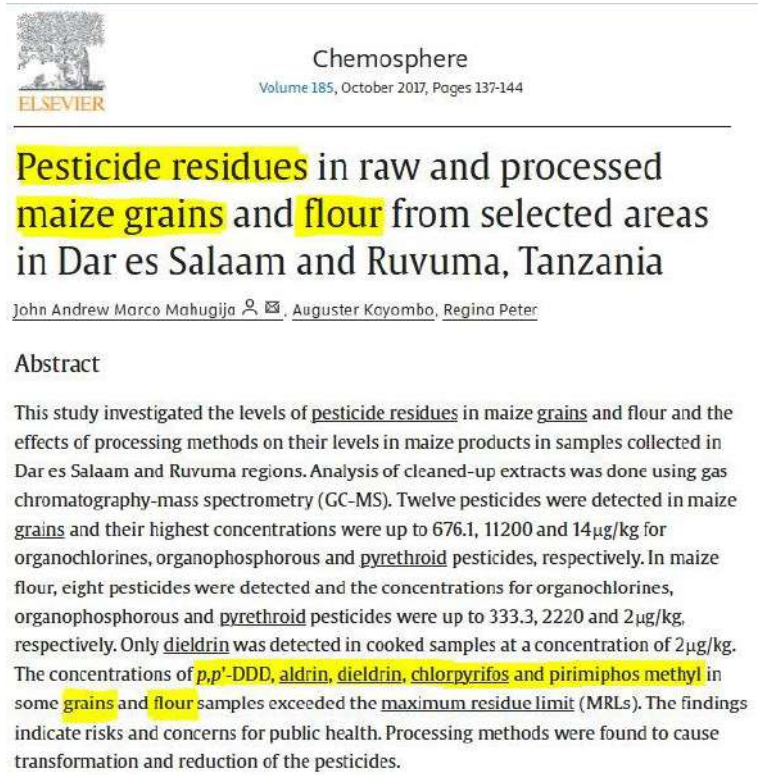
Foods: <foods you identified in action 1 that are edible products in the text
below between triple backticks>

Chemicals: <chemicals you identified in action 2 that are not microorganisms such
as bacteria, fungi, or other biological entities in the text below between triple
backticks>

Dictionary: <dictionary created in action 4 where keys are foods (only edible
products) and the values are a list of chemicals (no microorganisms or biological
entities) that are expressed to either be a food safety hazard, to be hazardous,
to be a contaminant, to accumulate, to have a MRL or to have the potential to pose
risk for human health via consumption for that corresponding food>

```{ABSTRACT}```
```

+



Chemosphere
Volume 185, October 2017, Pages 137-144

ELSEVIER

Pesticide residues in raw and processed maize grains and flour from selected areas in Dar es Salaam and Ruvuma, Tanzania

John Andrew Marco Mahugija, Auguster Kayombo, Regina Peter

Abstract

This study investigated the levels of pesticide residues in maize grains and flour and the effects of processing methods on their levels in maize products in samples collected in Dar es Salaam and Ruvuma regions. Analysis of cleaned-up extracts was done using gas chromatography-mass spectrometry (GC-MS). Twelve pesticides were detected in maize grains and their highest concentrations were up to 676.1, 11200 and 14 μg/kg for organochlorines, organophosphorous and pyrethroid pesticides, respectively. In maize flour, eight pesticides were detected and the concentrations for organochlorines, organophosphorous and pyrethroid pesticides were up to 333.3, 2220 and 2 μg/kg, respectively. Only dieldrin was detected in cooked samples at a concentration of 2 μg/kg. The concentrations of *p,p'*-DDD, aldrin, dieldrin, chlorpyrifos and pirimiphos methyl in some grains and flour samples exceeded the maximum residue limit (MRLs). The findings indicate risks and concerns for public health. Processing methods were found to cause transformation and reduction of the pesticides.

→

Chemicals: [‘pesticide residues’, ‘p,p’- DDD’, ‘aldrin’, ‘dieldrin’, ‘chlorpyrifos’, ‘pirimiphos methyl’]

Foods: [‘maize grain’, ‘maize flour’]

Dictionary: {

‘maize grains’: [‘pesticide residues’, ‘p,p’- DDD’, ‘aldrin’, ‘dieldrin’, ‘chlorpyrifos’, ‘pirimiphos methyl’]

‘maize flour’: ’: [pesticide residues’, ‘p,p’- DDD’, ‘aldrin’, ‘dieldrin’, ‘chlorpyrifos’, ‘pirimiphos methyl’]

}

holi food
```

Final instruction

Step by step prompt

```
Your task is to perform the following actions:

1. Identify which foods are mentioned in the text below provided between triple
backticks and collect them in a list, only collect items that are edible products.

2. Identify which chemicals are mentioned in the text below provided between
triple backticks and collect them in a list, be sure to exclude microorganisms
such as bacteria, fungi, or other biological entities.

3. Create all combinations of the foods identified in action 1 and chemicals
identified in action 2 as tuples and collect the tuples in a list.

4. Go over each combination of food and chemical in the list created at action 3
and look whether the chemical is mentioned to either be a food safety hazard for
that food, to be hazardous, to contaminate that food, to accumulate in that food,
to have a maximum residue limit (MRL) or to have the potential to pose risk for
human health via consumption of that food in the text below provided between
triple backticks. If the chemical is said to either be a food safety hazard, to be
hazardous, to be a contaminant, to accumulate, to have a MRL or to have the
potential to pose risk for human health via consumption of the food, then add that
combination of food and chemical to a dictionary where each key is a food and the
values are a list of chemicals.

I want to warn you against some pitfalls. First, if chemicals are mentioned both
with their names and abbreviations, make sure to return the full name of the
chemical instead of its abbreviation. Another thing - refrain from providing
irrelevant noun phrases or sentences in values just because they contain chemical
names, limit the values of your dictionary to the names of relevant chemicals.
Return an empty dictionary if you do not identify any foods that are edible
products. Also return an empty dictionary if you do not identify any chemicals.
Furthermore, return an empty dictionary if you do not identify any chemical in
food as either a food safety hazard, hazardous, a contaminant, accumulating,
without MRL or potentially harmful for human health through consumption of food.

Use the following format for your answer and limit your explanation:

Foods: <foods you identified in action 1 that are edible products in the text
below between triple backticks>

Chemicals: <chemicals you identified in action 2 that are not microorganisms such
as bacteria, fungi, or other biological entities in the text below between triple
backticks>

Dictionary: <dictionary created in action 4 where keys are foods (only edible
products) and the values are a list of chemicals (no microorganisms or biological
entities) that are expressed to either be a food safety hazard, to be hazardous,
to be a contaminant, to accumulate, to have a MRL or to have the potential to pose
risk for human health via consumption for that corresponding food>

```{ABSTRACT}```
```

Task in steps

Pitfalls

Output format

holi food
```

Scenarios

1. Instructing an LLM

- LLM used **with no changes**
- Not very labor intensive
- Instructions in natural language
- Various instructions explored

2. Retraining an LLM

- LLM further trained
- Requires **annotated data**
- Instructions in natural language
- Best instruction from 1st scenario used


holj food

Retraining - Workflow

Step by step prompt

```
Your task is to perform the following actions:

1. Identify which foods are mentioned in the text below provided between triple backticks and collect them in a list, only collect items that are edible products.

2. Identify which chemicals are mentioned in the text below provided between triple backticks and collect them in a list, be sure to exclude microorganisms such as bacteria, fungi, or other biological entities.

3. Create all combinations of the foods identified in action 1 and chemicals identified in action 2 as tuples and collect the tuples in a list.

4. Go over each combination of food and chemical in the list created at action 3 and look whether the chemical is mentioned to either be a food safety hazard for that food, to be hazardous, to contaminate that food, to accumulate in that food, to have a maximum residue limit (MRL) or to have the potential to pose risk for human health via consumption of that food in the text below provided between triple backticks. If the chemical is said to either be a food safety hazard, to be hazardous, to be a contaminant, to accumulate, to have a MRL or to have the potential to pose risk for human health via consumption of the food, then add that combination of food and chemical to a dictionary where each key is a food and the values are a list of chemicals.

I want to warn you against some pitfalls. First, if chemicals are mentioned both with their names and abbreviations, make sure to return the full name of the chemical instead of its abbreviation. Another thing - refrain from providing irrelevant noun phrases or sentences in values just because they contain chemical names, limit the values of your dictionary to the names of relevant chemicals. Return an empty dictionary if you do not identify any foods that are edible products. Also return an empty dictionary if you do not identify any chemicals. Furthermore, return an empty dictionary if you do not identify any chemical in food as either a food safety hazard, hazardous, a contaminant, accumulating, | without MRL or potentially harmful for human health through consumption of food.

Use the following format for your answer and limit your explanation:

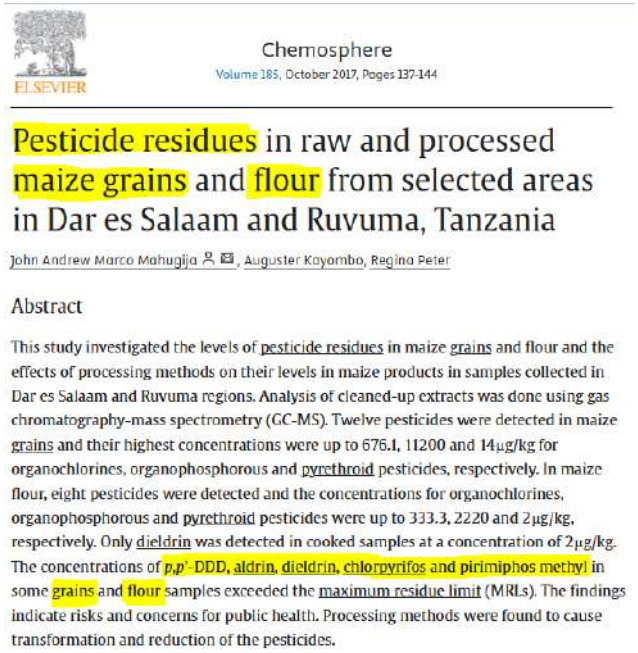
Foods: <foods you identified in action 1 that are edible products in the text below between triple backticks>

Chemicals: <chemicals you identified in action 2 that are not microorganisms such as bacteria, fungi, or other biological entities in the text below between triple backticks>

Dictionary: <dictionary created in action 4 where keys are foods (only edible products) and the values are a list of chemicals (no microorganisms or biological entities) that are expressed to either be a food safety hazard, to be hazardous, to be a contaminant, to accumulate, to have a MRL or to have the potential to pose risk for human health via consumption for that corresponding food>

```{ABSTRACT}```
```

+



Chemosphere
Volume 185, October 2017, Pages 137-144

ELSEVIER

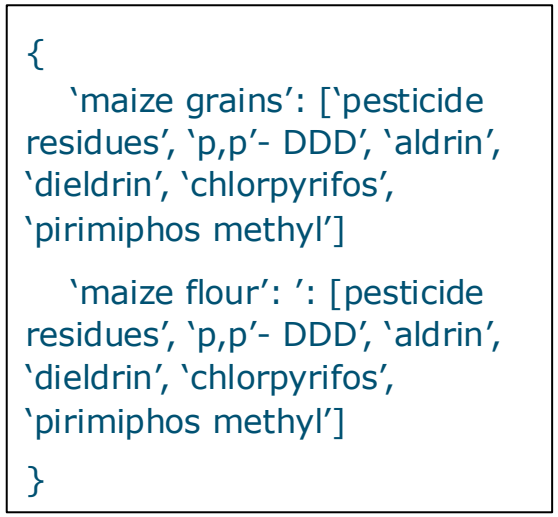
Pesticide residues in raw and processed maize grains and flour from selected areas in Dar es Salaam and Ruvuma, Tanzania

John Andrew Marco Mahugija, Auguster Kayombo, Regina Peter

Abstract

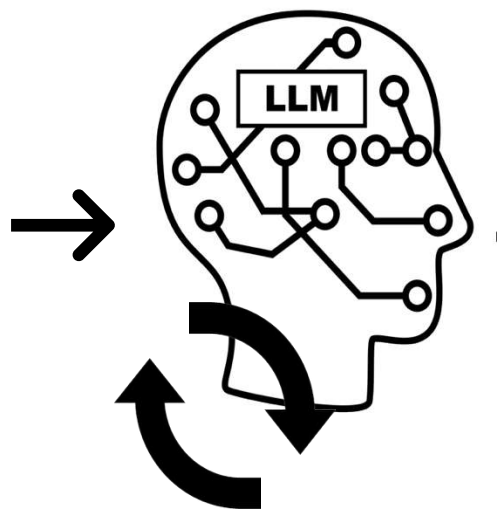
This study investigated the levels of pesticide residues in maize grains and flour and the effects of processing methods on their levels in maize products in samples collected in Dar es Salaam and Ruvuma regions. Analysis of cleaned-up extracts was done using gas chromatography-mass spectrometry (GC-MS). Twelve pesticides were detected in maize grains and their highest concentrations were up to 676.1, 11200 and 14 μg/kg for organochlorines, organophosphorous and pyrethroid pesticides, respectively. In maize flour, eight pesticides were detected and the concentrations for organochlorines, organophosphorous and pyrethroid pesticides were up to 333.3, 2220 and 2 μg/kg, respectively. Only dieldrin was detected in cooked samples at a concentration of 2 μg/kg. The concentrations of *p,p'*-DDD, aldrin, dieldrin, chlorpyrifos and pirimiphos methyl in some grains and flour samples exceeded the maximum residue limit (MRLs). The findings indicate risks and concerns for public health. Processing methods were found to cause transformation and reduction of the pesticides.

+



```
{
    'maize grains': ['pesticide residues', 'p,p'- DDD', 'aldrin', 'dieldrin', 'chlorpyrifos', 'pirimiphos methyl']

    'maize flour': ': [pesticide residues', 'p,p'- DDD', 'aldrin', 'dieldrin', 'chlorpyrifos', 'pirimiphos methyl']
}
```



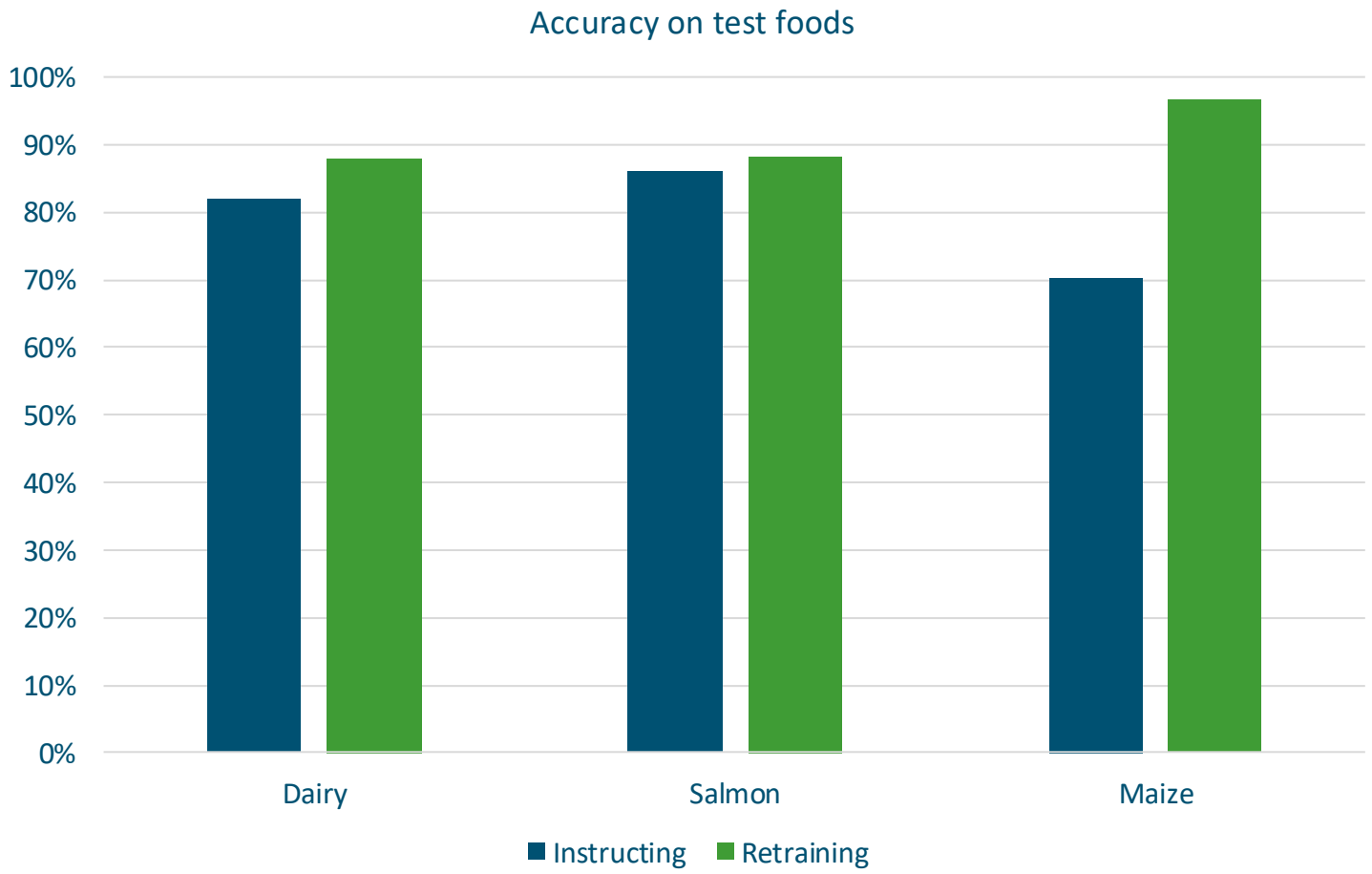
holi food
```

Retraining

- Are annotation and retraining worth it?
- We annotated:
  - 500 abstracts as train set
  - 100 abstracts as validation set
  - 150 abstracts as test set
- Test set – 50 abstracts per test food
  - Dairy, salmon, maize

holj food

Results – Test set



holj food

Dashboard

Overview of chemical hazards found in foods in the scientific literature

Filter options

Select hazard of interest...

Select hazard class of interest...

Select food (standardized) of interes...

Select food subclass of interest...

Select food class of interest...

Year selection: 2020 → 2026

Food	Hazard	DOI	Food standardized	Food subclass	Food class	Hazard class	↓ Year	Frequen
Shellfish	Azaspiracid	https://doi.org/10.1016/j.talanta.2025.128407	Shellfish	Shellfish	Fish & seafood	Fycotoxins	2026	39
Prawns	Arsenic+5	https://doi.org/10.1016/j.talanta.2025.128518	Shrimps and prawns	Shellfish	Fish & seafood	Elements (metals & metalloids)	2026	1
Anchovy	Arsenobetaine	https://doi.org/10.1016/j.talanta.2025.128518	Anchovies	Fish	Fish & seafood	Elements (other elements)	2026	1
Anchovy	Arsenic+5	https://doi.org/10.1016/j.talanta.2025.128518	Anchovies	Fish	Fish & seafood	Elements (metals & metalloids)	2026	1
Prawns	Arsenic+3	https://doi.org/10.1016/j.talanta.2025.128518	Shrimps and prawns	Shellfish	Fish & seafood	Elements (metals & metalloids)	2026	1
Anchovy	Arsenic+3	https://doi.org/10.1016/j.talanta.2025.128518	Anchovies	Fish	Fish & seafood	Elements (metals & metalloi		1
Corn	Glyphosate	https://doi.org/10.1016/j.saa.2025.126840	Maize grain	Cereal grains	Grains and grain-based pr	Pesticides		

holi food

Dashboard

Select hazard of interest... | Mycotoxins ×

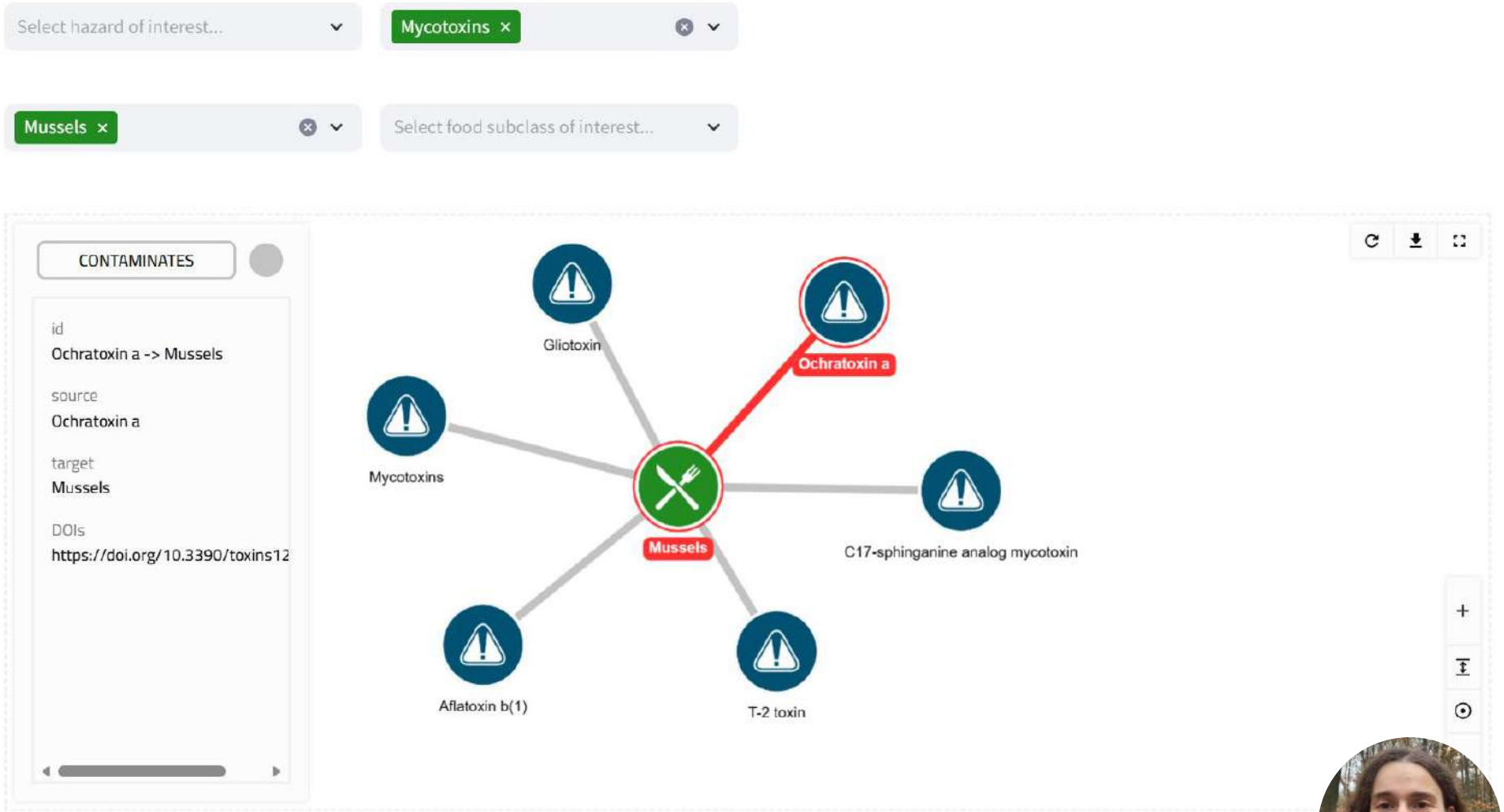
Select food (standardized) of interes... | Select food subclass of interest... | Fish & seafood ×

Year selection: 2020 → 2026

Food	Hazard	DOI	Food standardized	Food subclass	Food class	Hazard class	Year	Frequen
Fish	Mycotoxins	https://doi.org/10.3390/toxins17070356	Fish	Fish	Fish & seafood	Mycotoxins	2025	14
Fish	Mycophenolic acid	https://doi.org/10.1093/etojnl/vgaf193	Fish	Fish	Fish & seafood	Mycotoxins	2025	1
Fish	Alternariol	https://doi.org/10.3390/toxins17070356	Fish	Fish	Fish & seafood	Mycotoxins	2025	1
Fish	Alternariol monome	https://doi.org/10.3390/toxins17070356	Fish	Fish	Fish & seafood	Mycotoxins	2025	1
Fish	Beauvericin	https://doi.org/10.3390/toxins17070356	Fish	Fish	Fish & seafood	Mycotoxins	2025	2
Fish	Enniatins	https://doi.org/10.3390/toxins17070356	Fish	Fish	Fish & seafood	Mycotoxins	2025	2
Seafood	Mycotoxins	https://doi.org/10.1016/j.toxicon.2025.108505	Fish & seafood	None	Fish & seafood	Mycotoxins	2025	4
Oreochromis	Aflatoxins	https://doi.org/10.1007/s00284-025-04213-1	Tilapias and similar	Fish	Fish & seafood	Mycotoxins	2025	3
Seafood	Aflatoxin b(1)	https://doi.org/10.1016/j.toxicon.2025.108505	Fish & seafood	None	Fish & seafood	Mycotoxins		
Seafood	Aflatoxin g2	https://doi.org/10.1016/j.toxicon.2025.108505	Fish & seafood	None	Fish & seafood	Mycotoxins		

holi food

Dashboard


Select hazard of interest...
Mycotoxins ×
Mussels ×
Select food subclass of interest...
CONTAMINATES
id
Ochratoxin a -> Mussels
source
Ochratoxin a
target
Mussels
DOIs
https://doi.org/10.3390/toxins12
Gliotoxin
Ochratoxin a
Mycotoxins
Mussels
C17-sphinganine analog mycotoxin
Aflatoxin b(1)
T-2 toxin

Emerging food safety risks from text

Methods

Raw data

93,415

articles with full text

20

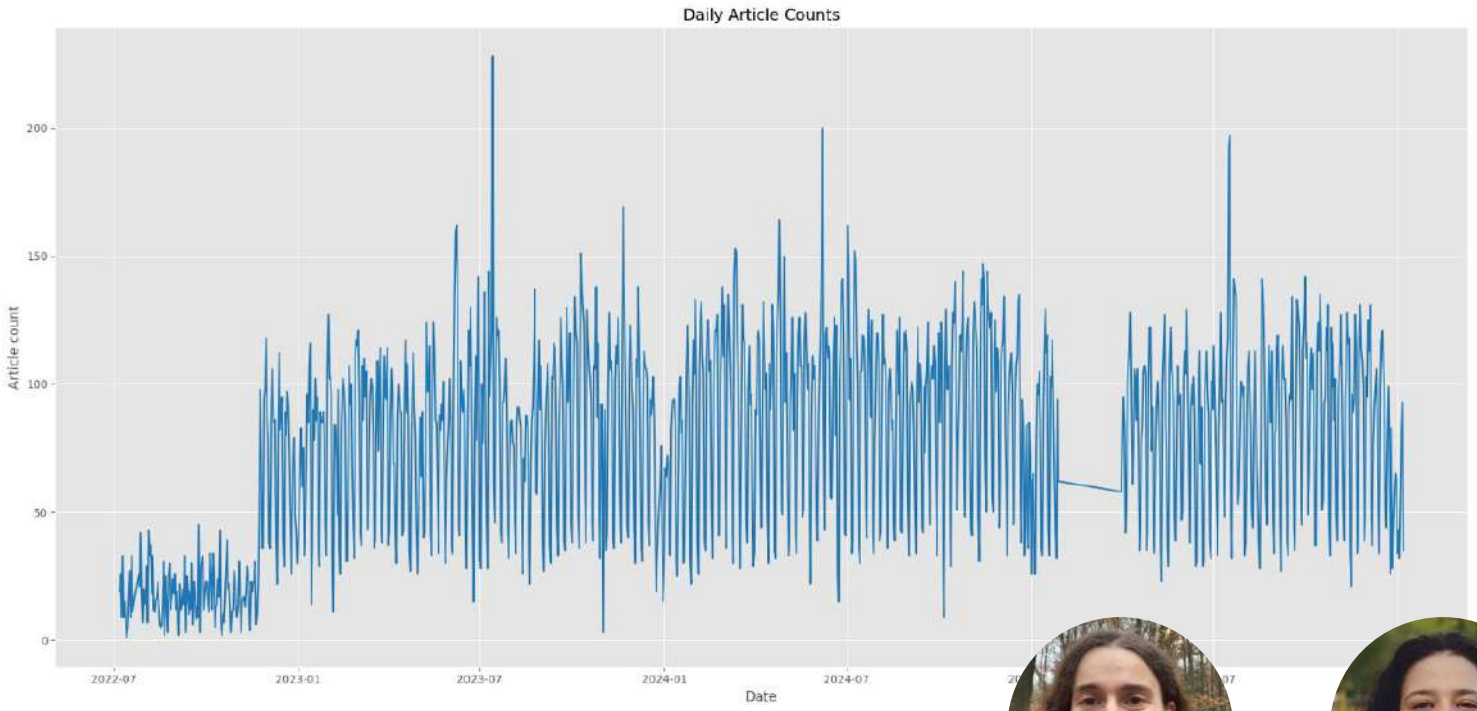
metadata/text fields per article

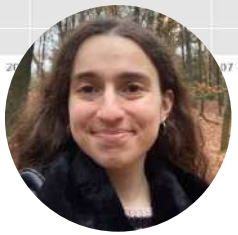
Earliest publication

11 Jul 2022

Latest publication

7 Jan 2026





holi food

Methods

Stepwise prompt

“First… second… be wary of…”

Microsoft Phi-4

holi food

Results

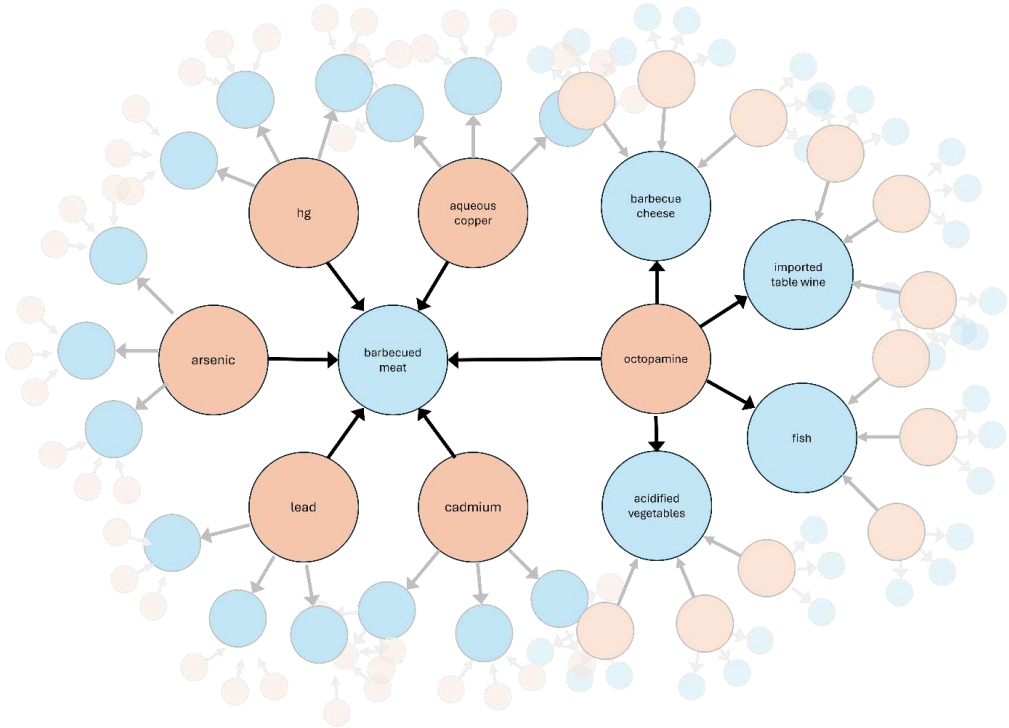
127,275

unique food–chemical pairs

Source contribution

~70k news ~60k scientific

Most frequently occurring pairs



Thank you!

Questions?

holi food

Funded by
the European Union