



Highlighting Case Studies in LLM Literature Review of Interdisciplinary System Science

Lachlan McGinness

Australian National University
CSIRO Data61

Peter Baumgartner

CSIRO Data61
Ausralian National University

Esther Onyango

CSIRO Agriculture and Food

Zelalem Lema

CSIRO Agriculture and Food

In a Nutshell

Literature review can be an arduous task

- (1) Find potentially relevant papers
- (2) Skim and reject
- (3) Read (understand) and reject
- (4) Summarize and report
- (5) Goto 1

≈ **SLR**

Systematic
Literature
Review



Increase of scientific literature over the years in many fields

Literature reviews are even more challenging and time consuming

Large Language Models (LLMs) excel at natural language tasks

Leverage the power of LLMs?

Existing approaches

(a) could have better explainability; (b) show hallucinations; (c) do not measure performance

Our contribution

LLMs and symbolic reasoning methods for (2) and (3)

Improved explainability by highlighting (a)

Check against hallucinations (b)

Performance evaluation of different LLM prompt strategies; four case studies (c)

SLR Case Study 1 - Healthy and Sustainable Transitions in Agri-Food Systems

Example *Interdisciplinary System Research Questions*

- (1) *What are the drivers of human health risks in food systems transitions?*
- (2) ...
- (3) ...



Systematic Literature Review - Case Study 1

Retrieve information on ...

- *Global Context*
- ***Associated Health Focus***
- *Transition Pathway*
- ***Agri-food System Aspect***
- *Public Health Risk*
- *Synergies*
- *Constraints*
- Integrated Solutions*

GPT3
GPT4
Researcher

- **Evidence from paper**
- **Information from evidence**

... from paper(s) ...

SLR Case Study 1 - Healthy and Sustainable Transitions in Agri-Food Systems

Example *Interdisciplinary System Research Questions*

- (1) *What are the drivers of human health risks in food systems transitions?*
- (2) ...
- (3) ...



Systematic Literature Review - Case Study 1

Retrieve information on ...

- *Global Context*
- ***Associated Health Focus***
- *Transition Pathway*
- ***Agri-food System Aspect***
- *Public Health Risk*
- *Synergies*
- *Constraints*
- Integrated Solutions*

GPT3
GPT4
Researcher

- **Evidence from paper**
- **Information from evidence**

... from paper(s) ...

“Associated Health Focus = Key threat to human health explored in the paper”

-> Retrieval Example

Information Retrieval Example

Query

Associated Health Focus

Researcher Answer:

Novel pathogens (COVID-19)

GPT3 Answer:

*pandemics, specifically
the COVID-19 pandemic.*

GPT4 Answer:

*The associated health focus of
this paper is COVID-19, a zoonotic infectious
disease caused by the SARS-CoV-2 virus.*

Paper



The two-way relationship between food systems and the COVID19 pandemic: causes and consequences

Marta G. Rivera-Ferre^{a,*}, Feliu López-i-Gelats^a, Federica Ravera^{b,c}, Elisa Oteros-Rozas^{a,c}, Marina di Masso^a, Rosa Binimelis^a, Hamid El Bilali^d

^a Chair Agroecology and Food Systems, University of Vic-Central University of Catalonia, Spain

^b Department of Geography, University of Girona, Spain

^c FRACTAL Collective, Spain

^d International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM-Bari), Valenzano, Bari, Italy

GPT4 Retrieved Evidence

"COVID-19 is the biggest pandemics in the last few decades and definitely the one with the highest socio-economic impacts worldwide and the situation is still characterised by a high degree of uncertainty (HLPE, 2020)."

"COVID-19 is a zoonosis, that is, a disease or infection that is transmitted from other vertebrate animals to humans (Platto et al., 2020)."

"Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19, also known as coronavirus disease, has created a completely novel situation worldwide."

Information Retrieval Example

Query

Associated Health Focus

Researcher Answer:

Novel pathogens (COVID-19)

GPT3 Answer:

*pandemics, specifically
the COVID-19 pandemic.*

GPT4 Answer:

*The associated health focus of
this paper is COVID-19, a zoonotic infectious
disease caused by the SARS-CoV-2 virus.*

Evidence for “what” (Covid-19) (?)

No evidence for “how” (Food <-> Covid-19)”

Paper



The two-way relationship between food systems and the COVID19 pandemic: causes and consequences

Marta G. Rivera-Ferre^{a,*}, Feliu López-i-Gelats^a, Federica Ravera^{b,c}, Elisa Oteros-Rozas^{a,c}, Marina di Masso^a, Rosa Binimelis^a, Hamid El Bilali^d

^a Chair Agroecology and Food Systems, University of Vic-Central University of Catalonia, Spain

^b Department of Geography, University of Girona, Spain

^c FRACTAL Collective, Spain

^d International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM-Bari), Valenzano, Bari, Italy

GPT4 Retrieved Evidence

"COVID-19 is the biggest pandemics in the last few decades and definitely the one with the highest socio-economic impacts worldwide and the situation is still characterised by a high degree of uncertainty (HLPE, 2020)."

"COVID-19 is a zoonosis, that is, a disease or infection that is transmitted from other vertebrate animals to humans (Platto et al., 2020)."

"Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19, also known as coronavirus disease, has created a completely novel situation worldwide."

Information Retrieval Example

Query

Agri-food System Aspect

Researcher Answer:

Supply chain: supplier connections, food processing and food service.

GPT3 Answer:

the impact of COVID-19 on food supply chains.

GPT4 Answer:

the impact of industrial food systems on emergent infectious diseases through land use changes and habitat fragmentation, the vulnerabilities of long food supply chains during crises like COVID-19, agricultural intensification and industrialization, the globalization of food markets, and the shift towards diets rich in ultra-processed foods and animal protein.

Paper



The two-way relationship between food systems and the COVID19 pandemic: causes and consequences

Marta G. Rivera-Ferre^{a,*}, Feliu López-i-Gelats^a, Federica Ravera^{b,c}, Elisa Oteros-Rozas^{a,c}, Marina di Masso^a, Rosa Binimelis^a, Hamid El Bilali^d

^a Chair Agroecology and Food Systems, University of Vic-Central University of Catalonia, Spain

^b Department of Geography, University of Girona, Spain

^c FRACTAL Collective, Spain

^d International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM-Bari), Valenzano, Bari, Italy

GPT4 Retrieved Evidence

"Long food supply chains have been particularly affected by COVID-19 crisis, however, it is important to avoid universalization of impacts and responses as agri-food systems are characterised by a huge diversity and heterogeneity."

"Significant changes have occurred in food systems in the last decades that have contributed to widen such 'holes' in the barriers from phase to phase: agricultural intensification and industrialization (Matson et al., 1997) causing major environmental deterioration, the increasing distance travelled by food in global markets (Nayak and Waterson, 2019), and the nutrition transition towards diets rich in ultra-processed food and animal protein (Popkin et al., 2012) are the three cornerstones of such changes.

“ ””
...

Information Retrieval Example

Query

Agri-food System Aspect

Researcher Answer:

Supply chain: supplier connections, food processing and food service.

GPT3 Answer:

the impact of COVID-19 on food supply chains.

GPT4 Answer:

the impact of industrial food systems on emergent infectious diseases through land use changes and habitat fragmentation, the vulnerabilities of long food supply chains during crises like COVID-19, agricultural intensification and industrialization, the globalization of food markets, and the shift towards diets rich in ultra-processed foods and animal protein.

Paper



The two-way relationship between food systems and the COVID19 pandemic: causes and consequences

Marta G. Rivera-Ferre^{a,*}, Feliu López-i-Gelats^a, Federica Ravera^{b,c}, Elisa Oteros-Rozas^{a,c}, Marina di Masso^a, Rosa Binimelis^a, Hamid El Bilali^d

^a Chair Agroecology and Food Systems, University of Vic-Central University of Catalonia, Spain

^b Department of Geography, University of Girona, Spain

^c FRACTAL Collective, Spain

^d International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM-Bari), Valenzano, Bari, Italy

GPT4 Retrieved Evidence

"Long food supply chains have been particularly affected by COVID-19 crisis, however, it is important to avoid universalization of impacts and responses as agri-food systems are characterised by a huge diversity and heterogeneity."

"Significant changes have occurred in food systems in the last decades that have contributed to widen such 'holes' in the barriers from phase to phase: agricultural intensification and industrialization (Matson et al., 1997) causing major environmental deterioration, the increasing distance travelled by food in global markets (Nayak and Waterson, 2019), and the nutrition transition towards diets rich in ultra-processed food and animal protein (Popkin et al., 2012) are the three cornerstones of such changes.

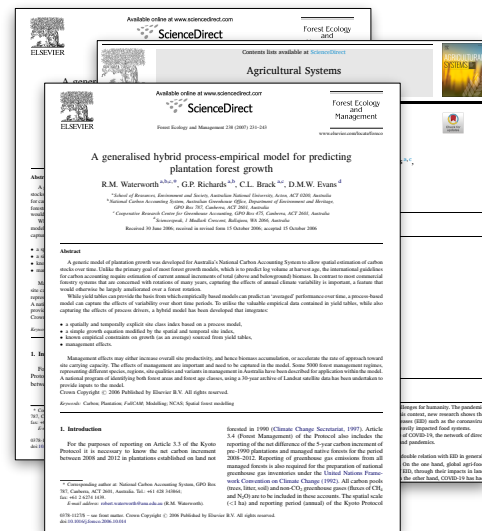
“ ... ”

-> Retrieval Workflow

Complete Information Retrieval Workflow

Complete Information Retrieval Workflow

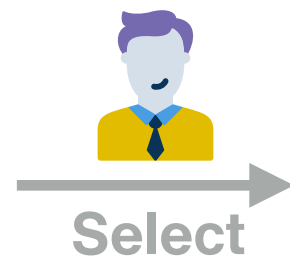
(50) Papers



Complete Information Retrieval Workflow

(50) Papers

The image shows a stack of ScienceDirect journal covers. The top cover is from 'Forest Ecology and Management' and features the article 'A generalised hybrid process-empirical model for predicting plantation forest growth' by R.M. Waterworth, G.P. Richards, C.L. Brack, and D.M.W. Evans. Below it, another cover from 'Agricultural Systems' is visible.

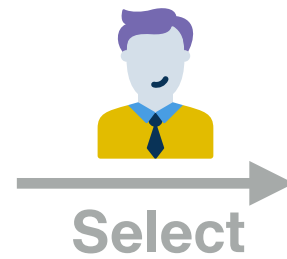
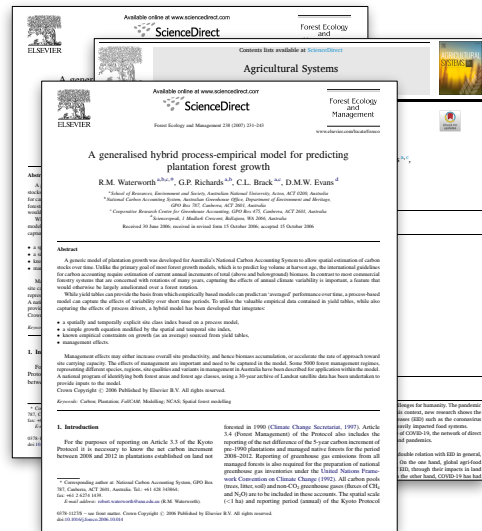


Paper/Excerpt

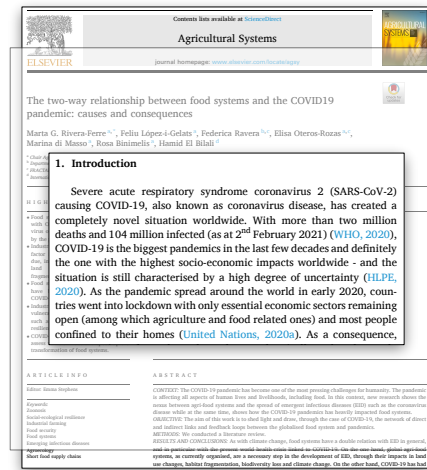
The image shows a detailed view of a paper excerpt. The title is 'The two-way relationship between food systems and the COVID-19 pandemic: causes and consequences'. The authors listed are Marta G. Rivera-Ferre, Felia Lopez-I-Gomez, Federica Ravera, Elias Otero-Rozas, Marina El Manouf, Rosa Rodriguez, and Hamid El Bilali. The excerpt includes an introduction that states: 'Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19, also known as coronavirus disease, has created a completely novel situation worldwide. With more than two million deaths and 104 million infections (as at 2nd February 2021) (WHO, 2020), COVID-19 is the biggest pandemics in the last few decades and definitely the one with the highest socio-economic impacts worldwide - and the situation is still characterised by a high degree of uncertainty (IHLPE, 2020). As the pandemic spread around the world in early 2020, countries went into lockdown with only essential economic sectors remaining open (among which agriculture and food related ones) and most people confined to their homes (United Nations, 2020a). As a consequence, ...'.

Complete Information Retrieval Workflow

(50) Papers



Paper/Excerpt



Queries with definitions and/or examples

Global Context: ...

Synergies: ...

Trade-offs: ...

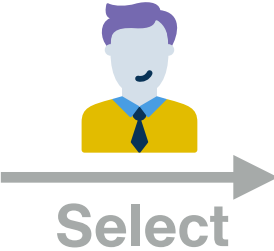
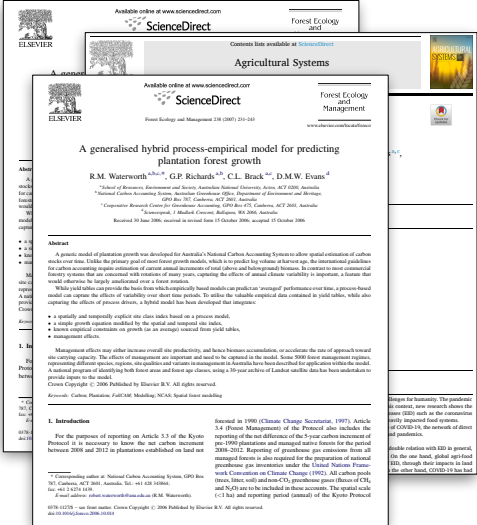
Enablers: ...

Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

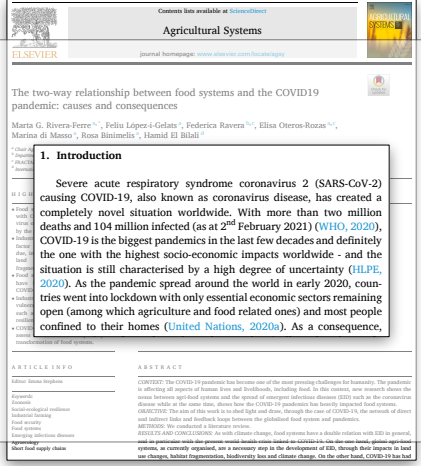


Complete Information Retrieval Workflow

(50) Papers



Paper/Excerpt



Queries with definitions and/or examples

Global Context: ...
 Synergies: ...
 Trade-offs:
 Enablers: ...

Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.



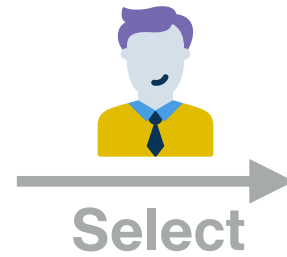
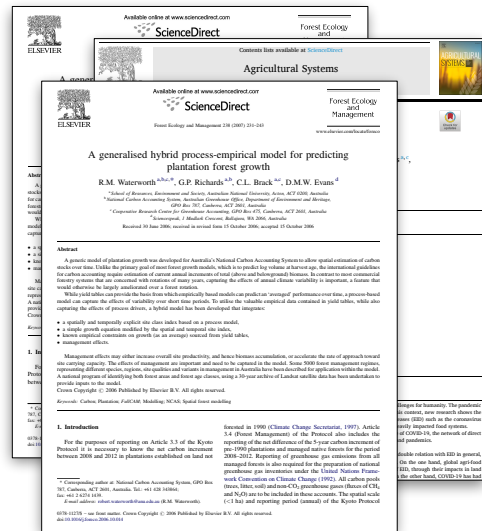
LLM

Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

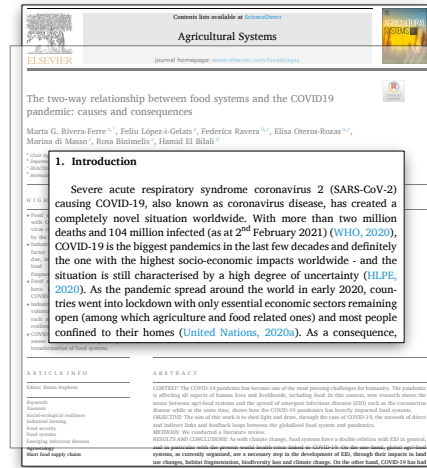
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...

Synergies: ...

Trade-offs:

Enablers: ...

Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes

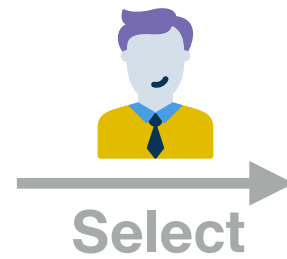
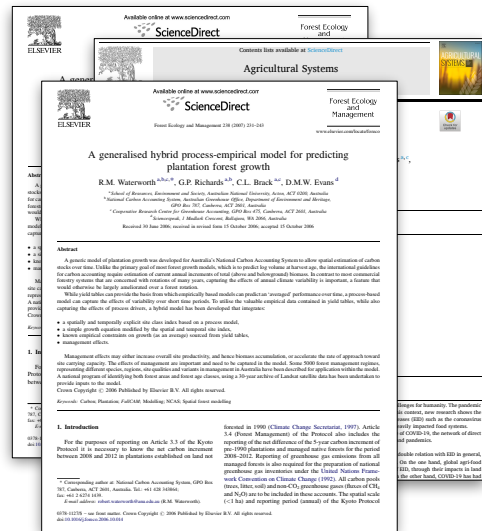
Summarize evidence as <answ>

- Query Answer:** <answ>
- Evidence:** <evidence>

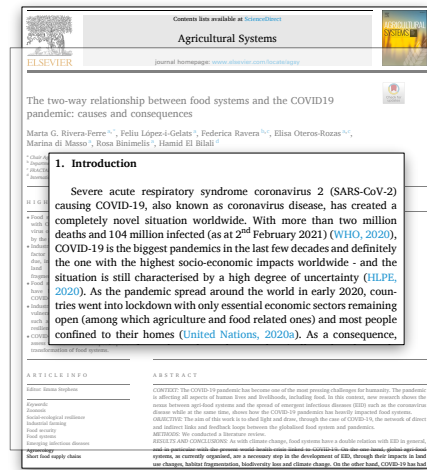
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...
Synergies: ...
Trade-offs:
Enablers: ...



Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

Query Answer: <answ>

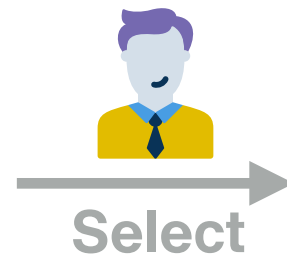
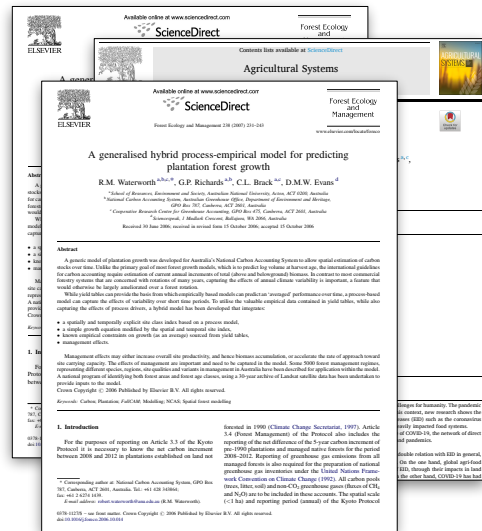
Evidence: <evidence>

Not found

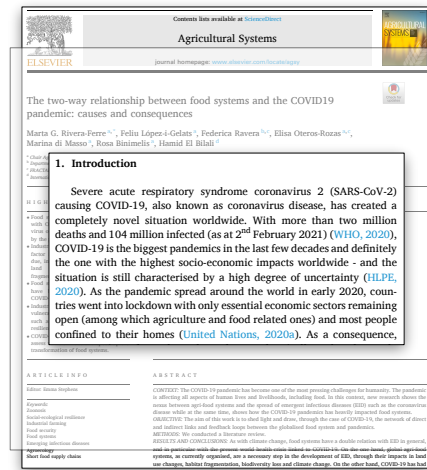


Complete Information Retrieval Workflow

(50) Papers



Paper/Excerpt



Queries with definitions and/or examples

Global Context: ...
Synergies: ...
Trade-offs: ...
Enablers: ...



Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

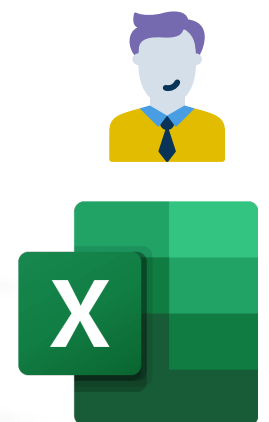
Query Answer: <answ>

Evidence: <evidence>

Highlight

<answ>

<evidence>



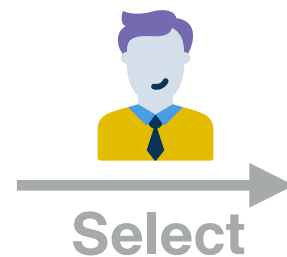
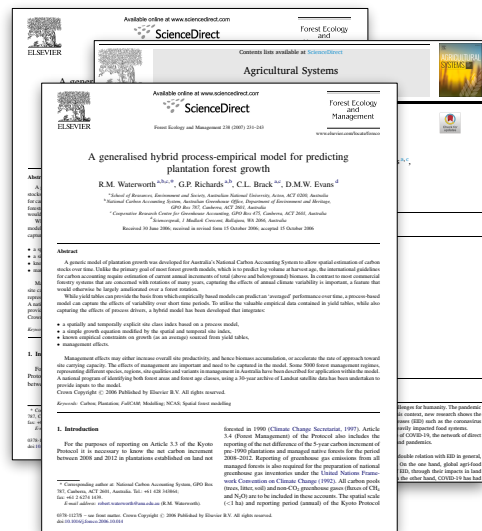
Not found



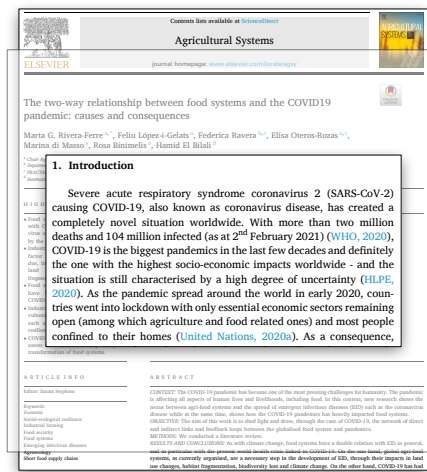
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...
Synergies: ...
Trade-offs: ...
Enablers: ...



Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

Properties

- ① Mitigate selection bias
 - Exclude irrelevant/misleading parts of paper
- ② Compensate lack of domain knowledge
- ③ Check for hallucinations

Query Answer: <answ>

Evidence: <evidence>

Highlight

<answ>

<evidence>



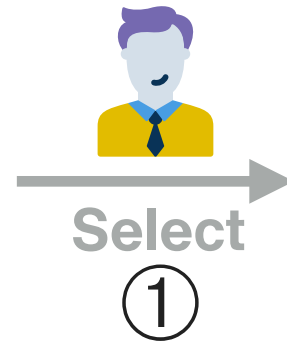
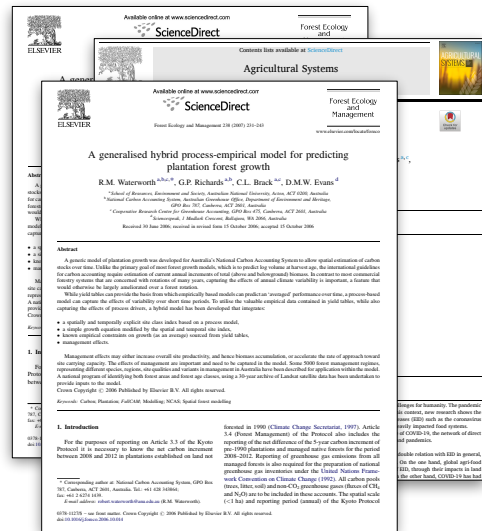
Not found



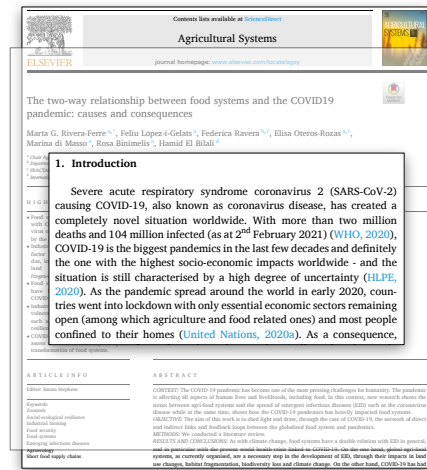
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...
Synergies: ...
Trade-offs: ...
Enablers: ...



Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

Properties

- ① Mitigate selection bias
 - Exclude irrelevant/misleading parts of paper
- ② Compensate lack of domain knowledge
- ③ Check for hallucinations

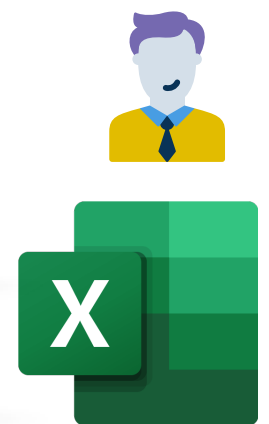
Query Answer: <answ>

Evidence: <evidence>

Highlight

<answ>

<evidence>



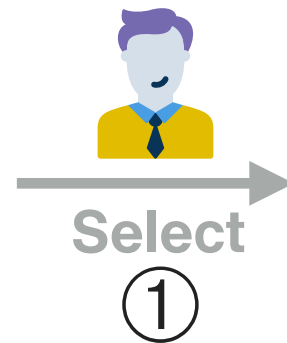
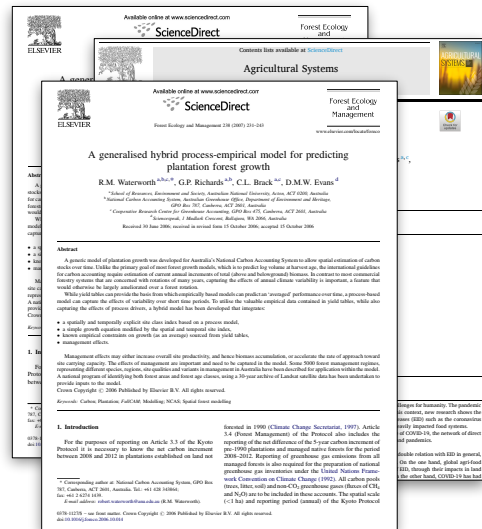
Not found



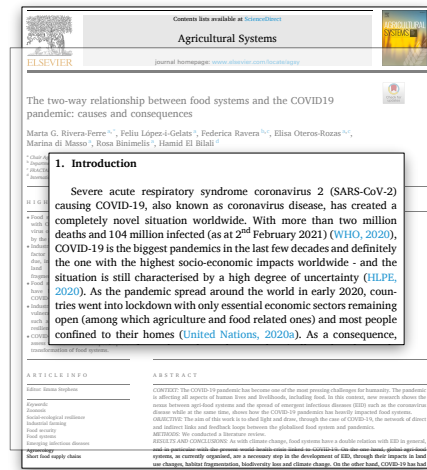
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...
Synergies: ...
Trade-offs:
Enablers: ...



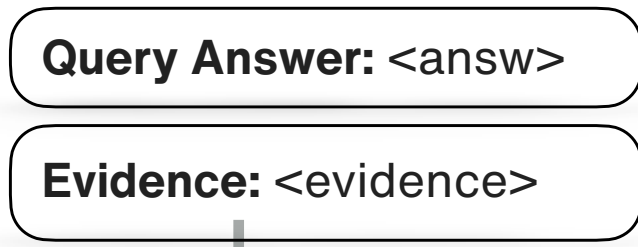
Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.



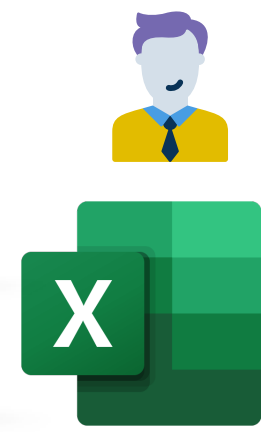
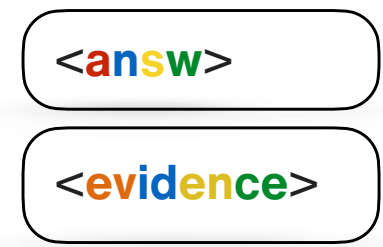
Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

Properties

- ① Mitigate selection bias
 - Exclude irrelevant/misleading parts of paper
- ② Compensate lack of domain knowledge
- ③ Check for hallucinations



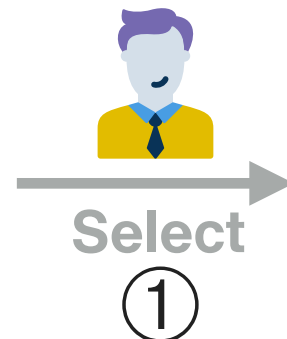
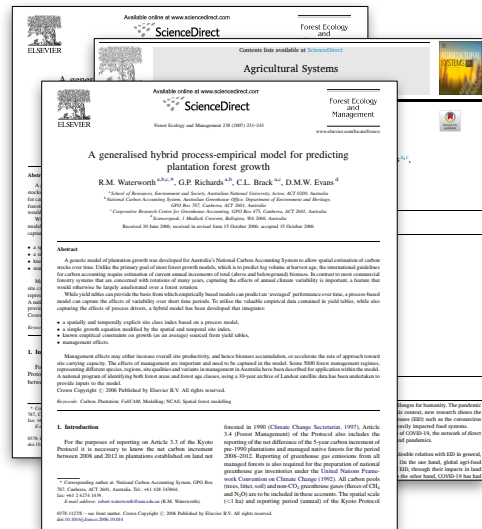
Highlight



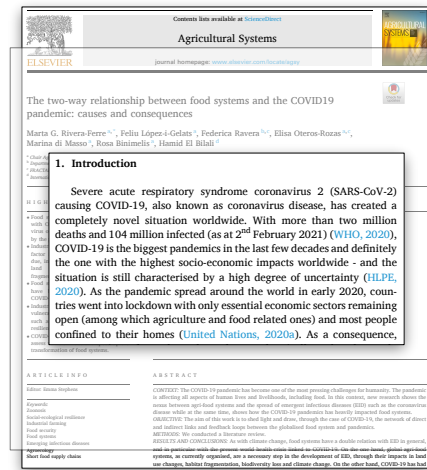
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...
 Synergies: ...
 Trade-offs: ...
 Enablers: ...
Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes
 Summarize evidence as <answ>

Properties

- ① Mitigate selection bias
 Exclude irrelevant/misleading parts of paper
- ② Compensate lack of domain knowledge
- ③ Check for hallucinations

Query Answer: <answ>

Evidence: <evidence>

Highlight

<answ>

<evidence>



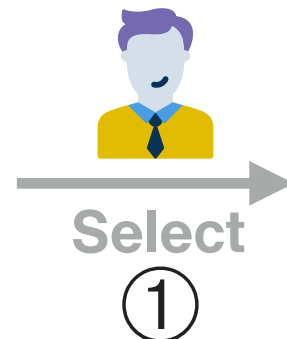
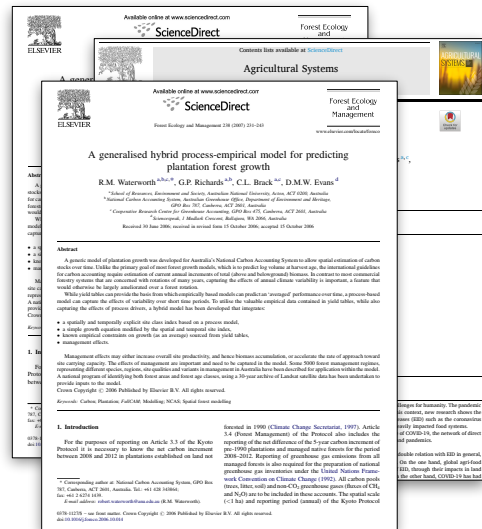
③ Not found



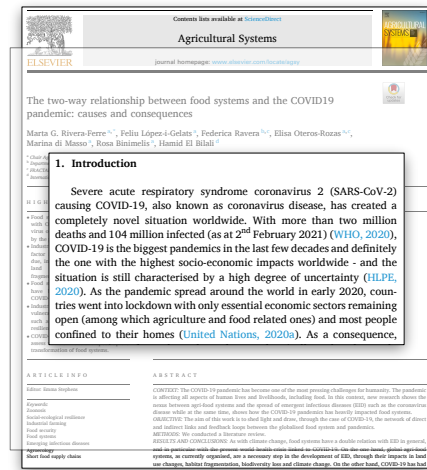
Complete Information Retrieval Workflow

Queries with definitions and/or examples

(50) Papers



Paper/Excerpt



Global Context: ...
Synergies: ...
Trade-offs: ...
Enablers: ...



Constraints: a condition/situation when it was reported to obstruct or impede the achievement of certain key elements.

LLM

For each

Provide evidence for <query>: 3 quotes
Summarize evidence as <answ>

Properties

- ① Mitigate selection bias
Exclude irrelevant/misleading parts of paper
- ② Compensate lack of domain knowledge
- ③ Check for hallucinations

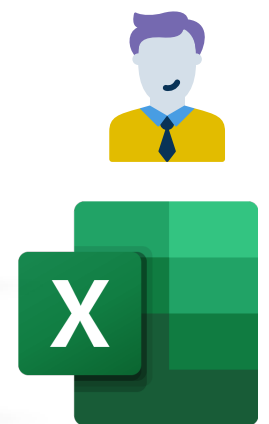
Query Answer: <answ>

Evidence: <evidence>

Highlight

<answ>

<evidence>



③ Not found



-> Highlighting?

Example Spreadsheets



Associate Health Focus

LLM Answer - Context	
diet-related noncommunicable diseases (DR-NCDs).	1
Foodborne Disease.	1
The associated health focus of this paper is COVID-19, a zoonotic infectious disease caused by the SARS-CoV-2 virus.	1
Nutrition and diet-related non-communicable diseases (NCDs).	1
the double burden of malnutrition.	1
Non-communicable diseases (NCD).	1
the double burden of malnutrition, which includes undernutrition and obesity, in the context of climate change.	1

Associate Health Focus Evidence

Context Evidence
"Dietary patterns associated with the nutrition transition have contributed to Africa's complex burden of malnutrition – obesity and other diet-related noncommunicable diseases (DR-NCDs) – along with persistent food insecurity and undernutrition." "Unhealthy or obesogenic food environments provide an increased availability of energy-dense, nutrient-poor foods (e.g., refined grains, fast foods) at cheaper prices (14)." "The interaction of the food environment and its influence on food acquisition and consumption is not clearly elucidated in low and middle-income countries (LMICs)."
"WHO estimates for 2010 show that the burden of illness from unsafe food worldwide is comparable to that of malaria and tuberculosis." "For the 2010 base year used, it is estimated that foodborne illnesses (FBI) arising from 31 of the most important hazards lead to 420,000 deaths and 600 million illnesses annually." "Unsafe food is implicated in childhood stunting, demonstrating one of the close links between safety and nutrition."
"COVID-19 is the biggest pandemic in the last few decades and definitely the one with the highest socio-economic impacts worldwide and the situation is still characterised by a high degree of uncertainty (HLPE, 2020)." "COVID-19 is a zoonosis, that is, a disease or infection that is transmitted from other vertebrate animals to humans (Platto et al., 2020)." "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19, also known as coronavirus disease, has created a completely novel situation worldwide."
"The multiple forms of malnutrition are worsening" "Malnutrition encompasses undernutrition, including childhood stunting, wasting, and micronutrient deficiencies; overweight and obesity; and diet-related NCDs." "The Global Burden of Disease identifies the makeup of diets as a significant risk factor for mortality and morbidity, with 11 million deaths and 255 million disability-adjusted life years attributable to dietary risk factors including high intake of sodium and low intake of whole grains and fruits."

Example Spreadsheets



Associate Health Focus

LLM Answer - Context
diet-related noncommunicable diseases (DR-NCDs).
Foodborne Disease.
The associated health focus of this paper is COVID-19, a zoonotic infectious disease caused by the SARS-CoV-2 virus.
Nutrition and diet-related non-communicable diseases (NCDs).
the double burden of malnutrition.
Non-communicable diseases (NCD).
the double burden of malnutrition, which includes undernutrition and obesity, in the context of climate change.



LLM Answer - Context Highlighted
diet-related noncommunicable diseases (DR-NCDs).
Foodborne Disease.
The associated health focus of this paper is COVID-19, a zoonotic infectious disease caused by the SARS-CoV-2 virus .
Nutrition and diet-related non-communicable diseases (NCDs).
the double burden of malnutrition .
Non-communicable diseases (NCD).
the double burden of malnutrition , which includes undernutrition and obesity, in the context of climate change .
Zoonotic diseases and Emerging Infectious Diseases (EIDs).
Non-communicable diseases (NCDs).

Associate Health Focus Evidence

Context Evidence
"Dietary patterns associated with the nutrition transition have contributed to Africa's complex burden of malnutrition – obesity and other diet-related noncommunicable diseases (DR-NCDs) – along with persistent food insecurity and undernutrition." "Unhealthy or obesogenic food environments provide an increased availability of energy-dense, nutrient-poor foods (e.g., refined grains, fast foods) at cheaper prices (14)." "The interaction of the food environment and its influence on food acquisition and consumption is not clearly elucidated in low and middle-income countries (LMICs)."
"WHO estimates for 2010 show that the burden of illness from unsafe food worldwide is comparable to that of malaria and tuberculosis." "For the 2010 base year used, it is estimated that foodborne illnesses (FBI) arising from 31 of the most important hazards lead to 420,000 deaths and 600 million illnesses annually." "Unsafe food is implicated in childhood stunting, demonstrating one of the close links between safety and nutrition."
"COVID-19 is the biggest pandemics in the last few decades and definitely the one with the highest socio-economic impacts worldwide and the situation is still characterised by a high degree of uncertainty (HLPE, 2020)." "COVID-19 is a zoonosis, that is, a disease or infection that is transmitted from other vertebrate animals to humans (Platto et al., 2020)." "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19, also known as coronavirus disease, has created a completely novel situation worldwide."
"The multiple forms of malnutrition are worsening" "Malnutrition encompasses undernutrition, including childhood stunting, wasting, and micronutrient deficiencies; overweight and obesity; and diet-related NCDs." "The Global Burden of Disease identifies the makeup of diets as a significant risk factor for mortality and morbidity, with 11 million deaths and 255 million disability-adjusted life years attributable to dietary risk factors including high intake of sodium and low intake of whole grains and fruits."

Context Evidence Highlighted
Dietary patterns associated with the nutrition transition have contributed to Africa's complex burden of malnutrition – obesity and other diet-related noncommunicable diseases (DR-NCDs) – along with persistent food insecurity and undernutrition. Unhealthy or obesogenic food environments provide an increased availability of energy-dense, nutrient-poor foods (e.g., refined grains, fast foods) at cheaper prices (14). The interaction of the food environment and its influence on food acquisition and consumption is not clearly elucidated in low and middle-income countries (LMICs).
WHO estimates for 2010 show that the burden of illness from unsafe food worldwide is comparable to that of malaria and tuberculosis . For the 2010 base year used , it is estimated that foodborne illnesses (FBI) arising from 31 of the most important hazards lead to 420,000 deaths and 600 million illnesses annually. Unsafe food is implicated in childhood stunting, demonstrating one of the close links between safety and nutrition .
COVID-19 is the biggest pandemics in the last few decades and definitely the one with the highest socio-economic impacts worldwide and the situation is still characterised by a high degree of uncertainty (HLPE, 2020). COVID-19 is a zoonosis , that is, a disease or infection that is transmitted from other vertebrate animals to humans (Platto et al., 2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19, also known as coronavirus disease , has created a completely novel situation worldwide.
The multiple forms of malnutrition are worsening . Malnutrition encompasses undernutrition, including childhood stunting, wasting, and micronutrient deficiencies ; overweight and obesity; and diet-related NCDs . The Global Burden of Disease identifies the makeup of diets as a significant risk factor for mortality and morbidity, with 11 million deaths and 255 million disability-adjusted life years attributable to dietary risk factors including high intake of sodium and low intake of whole grains and fruits .
Improving nutrition is a development priority , particularly in low and middle-income countries (LMICs) in Africa and South Asia, in which there is a persistent burden of undernutrition and increasing obesity . By undernutrition, we mean underweight (low weight for age), stunting (low height for age), wasting (low weight for height and weakness), and micronutrient deficiencies . Drivers of the double burden of malnutrition .

Highlighting Example

Query

Agri-food System Aspect

Keywords

*food.n production.n produce.v
environment.n livestock.n agriculture.n
consumption.n consume.v diet.n farming.n*

GPT4 Answer:

the impact of industrial food systems on emergent infectious diseases through land use changes and habitat fragmentation, the vulnerabilities of long food supply chains during crises like COVID-19, agricultural intensification and industrialization, the globalization of food markets, and the shift towards diets rich in ultra-processed foods and animal protein.

Highlighting



GPT4 Answer Highlighted:

*the impact of **industrial food systems** on **emergent infectious diseases** through **land use changes** and **habitat fragmentation**, the vulnerabilities of **long food supply chains** during crises like COVID-19, **agricultural intensification** and **industrialization**, the globalization of **food markets**, and the shift towards **diets** rich in **ultra-processed foods** and **animal protein**.*

Highlighting Algorithm

Text

Foodborne illnesses significantly influence individuals nutritional status.

Highlighted text

1. It is also likely that climate change will contribute to novel occurrences of disease emergence and transmission.
- ➔ 2. Foodborne illnesses significantly influence individuals nutritional status.
3. Changing lifestyles, mainly due to work commitment, have fuelled the increase in numbers eating out and the need for convenience foods.
4. Significant changes have occurred in food systems in the last decades that have contributed to widen such 'holes' in the barriers from phase to phase: agricultural intensification and industrialization causing major environmental deterioration, the increasing distance traveled by food in global markets, and the nutrition transition towards diets rich in ultra - processed food and animal protein are the three cornerstones of such changes.

Highlighting Algorithm

Text

Foodborne illnesses significantly influence individuals nutritional status.

+

Keywords

Entities: *health, disease, outcome, food, lifestyle*

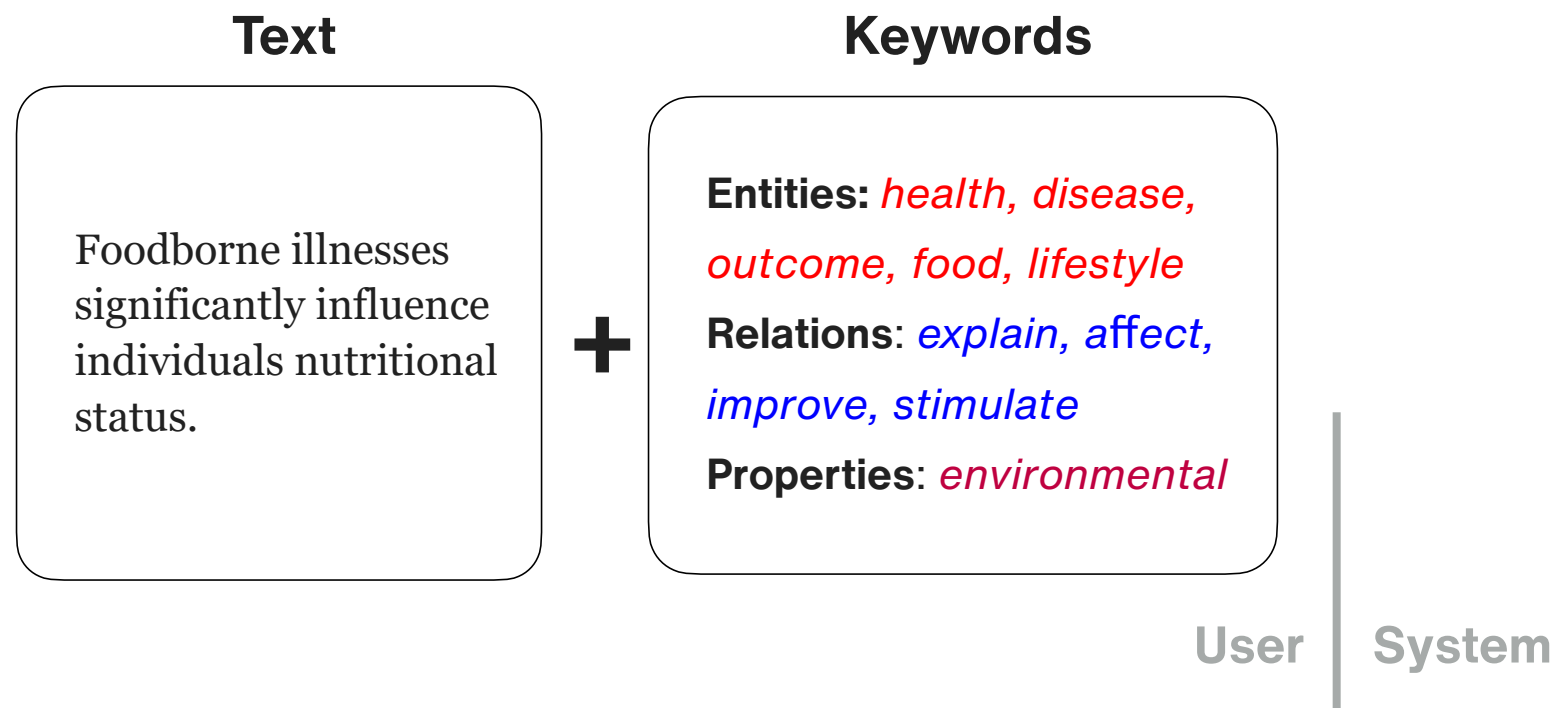
Relations: *explain, affect, improve, stimulate*

Properties: *environmental*

Highlighted text

1. It is also likely that **climate change** **will** **contribute** to **novel occurrences** of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly **influence** individuals **nutritional status**.
3. **Changing lifestyles**, mainly due to **work commitment**, **have fuelled** the **increase** in numbers eating out and the need for **convenience foods**.
4. **Significant changes** **have occurred** in **food systems** in the last decades that **have contributed** to widen **such 'holes'** in the barriers from phase to phase: agricultural intensification and industrialization **causing** major environmental deterioration, the **increasing distance traveled** by **food** in **global markets**, and the **nutrition transition** towards **diets** rich in **ultra - processed food and animal protein** are the three cornerstones of **such changes**.

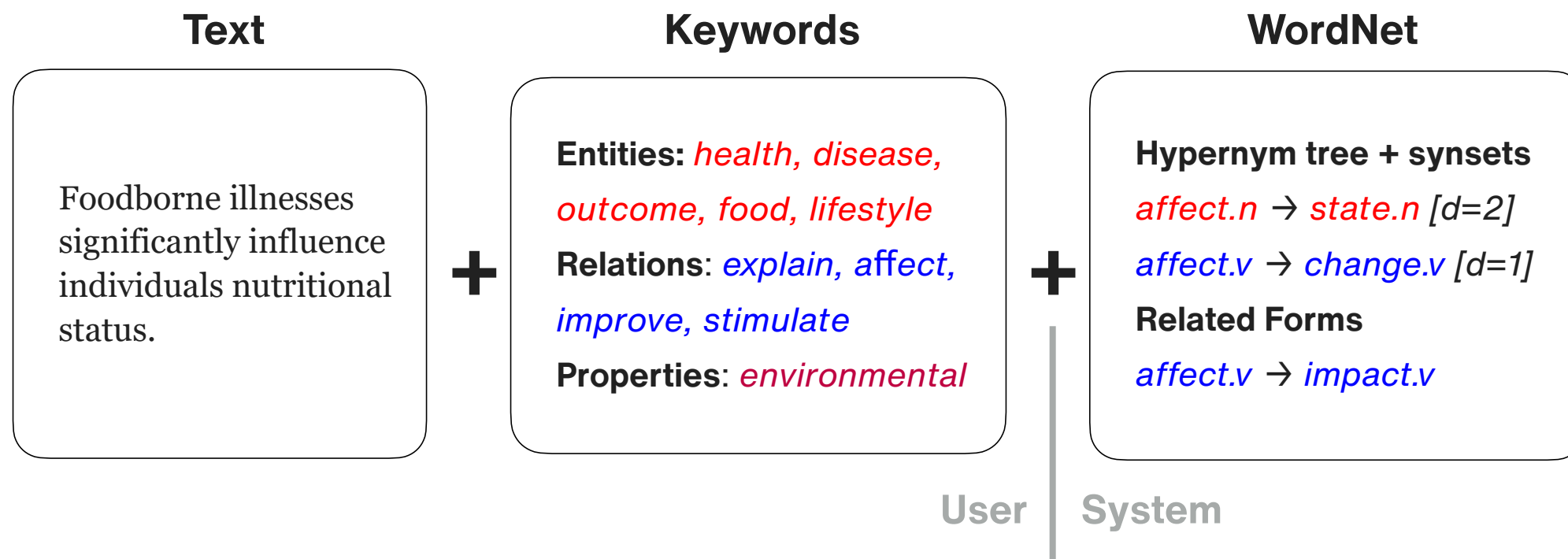
Highlighting Algorithm



Highlighted text

1. It is also likely that **climate change** will contribute to novel occurrences of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly **influence** individuals **nutritional status**.
3. **Changing lifestyles**, mainly due to **work commitment**, **have fuelled** the **increase** in numbers eating out and the need for **convenience foods**.
4. **Significant changes** **have occurred** in **food systems** in the last decades that **have contributed** to widen **such 'holes'** in the barriers from phase to phase: agricultural intensification and industrialization **causing** major environmental deterioration, the **increasing distance traveled** by **food** in **global markets**, and the **nutrition transition** towards **diets** rich in **ultra - processed food and animal protein** are the three cornerstones of **such changes**.

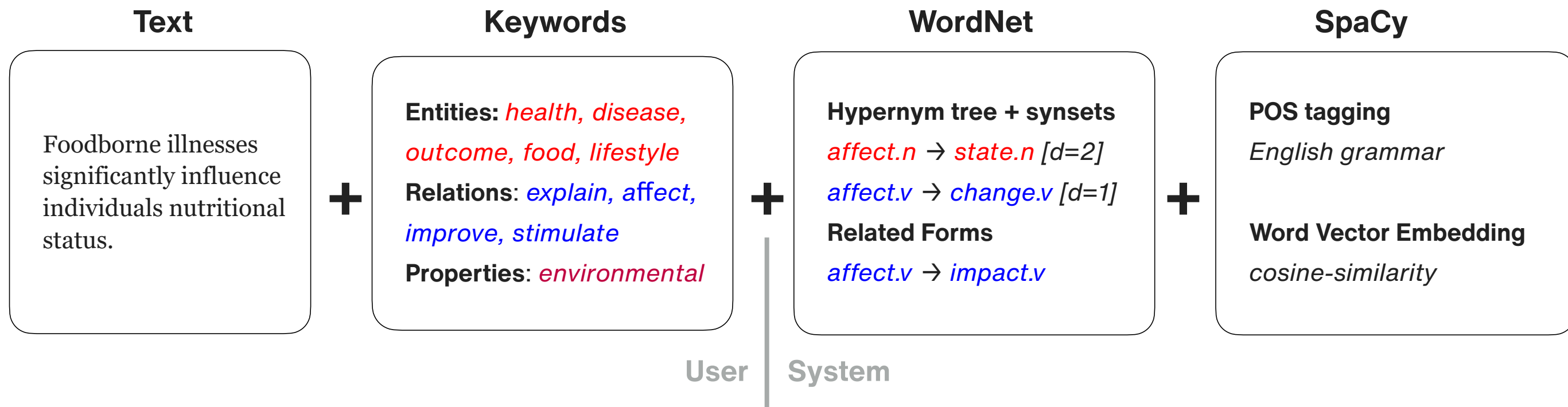
Highlighting Algorithm



Highlighted text

1. It is also likely that **climate change** will **contribute** to **novel occurrences** of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly **influence** individuals **nutritional status**.
3. **Changing lifestyles**, mainly due to **work commitment**, **have fuelled** the **increase** in numbers eating out and the need for **convenience foods**.
4. **Significant changes** **have occurred** in **food systems** in the last decades that **have contributed** to widen **such 'holes'** in the barriers from phase to phase: agricultural intensification and industrialization **causing** major environmental deterioration, the **increasing distance traveled** by **food** in **global markets**, and the **nutrition transition** towards **diets** rich in **ultra - processed food and animal protein** are the three cornerstones of **such changes**.

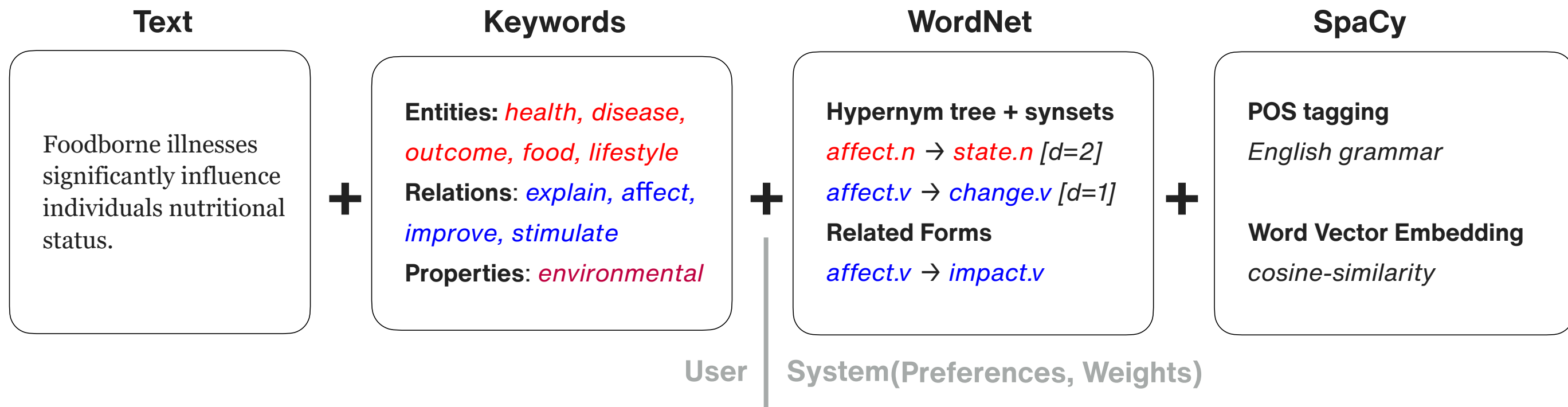
Highlighting Algorithm



Highlighted text

1. It is also likely that **climate change** will contribute to novel occurrences of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly **influence** individuals **nutritional status**.
3. **Changing lifestyles**, mainly due to **work commitment**, **have fuelled** the **increase** in numbers eating out and the need for **convenience foods**.
4. **Significant changes** **have occurred** in **food systems** in the last decades that **have contributed** to widen **such 'holes'** in the barriers from phase to phase: agricultural intensification and industrialization **causing** major environmental deterioration, the **increasing distance traveled** by **food** in **global markets**, and the **nutrition transition** towards **diets** rich in **ultra - processed food and animal protein** are the three cornerstones of **such changes**.

Highlighting Algorithm



Highlighted text

1. It is also likely that **climate change** will contribute to novel occurrences of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly influence individuals nutritional status.
3. **Changing lifestyles**, mainly due to **work commitment**, have fuelled the increase in numbers eating out and the need for **convenience foods**.
4. **Significant changes** have occurred in **food systems** in the last decades that have contributed to widen such 'holes' in the barriers from phase to phase: agricultural intensification and industrialization causing major environmental deterioration, the increasing distance traveled by food in global markets, and the **nutrition transition** towards diets rich in ultra - processed food and animal **protein** are the three cornerstones of such changes.

Algorithm 1 Similarity

```

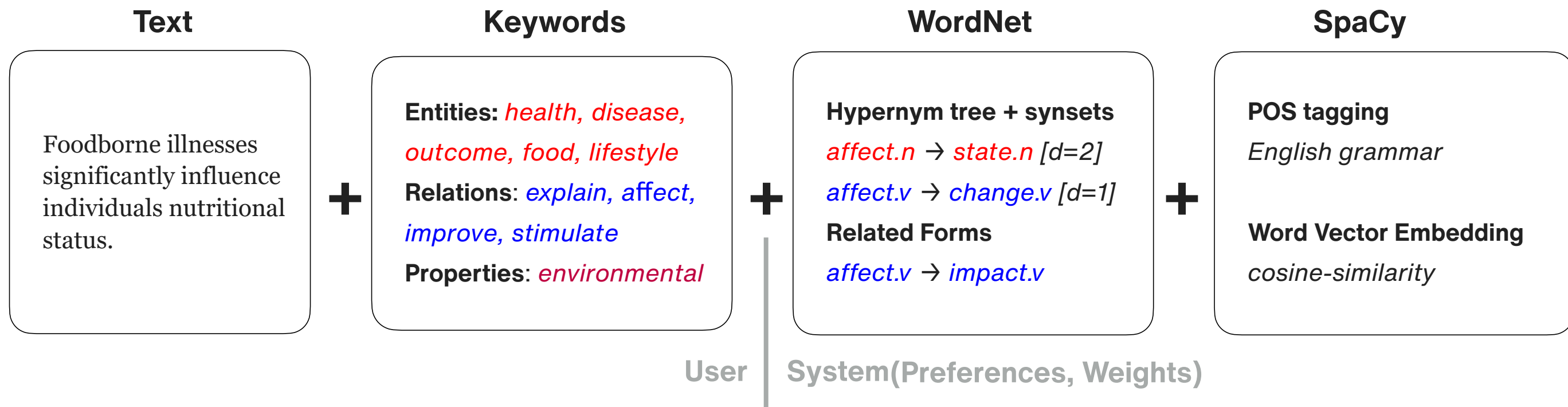
Similarity(w, C, t)
Input: w word, C keywords in canonical form (lemmas), t type of w (noun, verb, adjective ...)
Output: Similarity score for w
P-Weight ← 0.95   RF-Weight ← 0.95   WUP-Threshold ← 0.8   VEC-Threshold ← 0.95
bestwup ← maxc∈C WUP-X(w, c, t) or else 0.0
bestvec ← maxc∈C VEC(w, c) or else 0.0
if bestwup ≥ WUP-Threshold and bestwup ≥ bestvec then return bestwup
elif bestvec ≥ VEC-Threshold then return bestvec
else return 0.0

WUP-X(w, c, t) // Extended Wu-Palmer similarity, considers reachable words from w and c
Sw = Extend(WN-Synsets(w, t)) // WN-Synsets returns synonyms of w
Sc = Extend(WN-Synsets(c, t))
return max(sw→rw, sc→rc)∈Sw×Sc ωw · ωc · WN-WUP(rw, rc, t) // Wu-Palmer from WordNet

VEC(w, c) // Vector similarity
if w = c then return 1.0
elif both w and c have vector embeddings then
    return cosine-similarity of the embeddings of w and of c
else return 0.0

Extend(S, t)
Input: S a set of WordNet synsets
Output: Weighted extension of S by pertainyms and derivationally related forms
R ← ∅ // Result relation
for s ∈ S do
    R ← R ∪ {s  $\xrightarrow{1.0}$  s} // R is reflexive
    for l ∈ WN-Lemmas(s) do
        R ← R ∪ {s  $\xrightarrow{P-Weight}$  WN-SynSet(v) | v ∈ WN-Pertainyms(l)}
        R ← R ∪ {s  $\xrightarrow{RF-Weight}$  WN-SynSet(v) | v ∈ WN-RelatedForms(l)}
return R
    
```

Highlighting Algorithm



Highlighted text

1. It is also likely that **climate change** will contribute to novel occurrences of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly **influence** individuals **nutritional status**.
3. **Changing lifestyles**, mainly due to **work commitment**, **have fuelled** the **increase** in numbers eating out and the need for **convenience foods**.
4. **Significant changes** **have occurred** in **food systems** in the last decades that **have contributed** to widen such 'holes' in the barriers from phase to phase: agricultural intensification and industrialization **causing** major environmental deterioration, the **increasing distance traveled** by **food** in **global markets**, and the **nutrition transition** towards **diets** rich in **ultra - processed food** and **animal protein** are the three cornerstones of **such changes**.

Algorithm 1 Similarity

```

Similarity( $w, C, t$ )
Input:  $w$  word,  $C$  keywords in canonical form (lemmas),  $t$  type of  $w$  (noun, verb, adjective ...)
Output: Similarity score for  $w$ 
P-Weight  $\leftarrow 0.95$    RF-Weight  $\leftarrow 0.95$    WUP-Threshold  $\leftarrow 0.8$    VEC-Threshold  $\leftarrow 0.95$ 
bestwup  $\leftarrow \max_{c \in C} \text{WUP-X}(w, c, t)$  or else 0.0
bestvec  $\leftarrow \max_{c \in C} \text{VEC}(w, c)$  or else 0.0
if bestwup  $\geq$  WUP-Threshold and bestwup  $\geq$  bestvec then return bestwup
elif bestvec  $\geq$  VEC-Threshold then return bestvec
else return 0.0

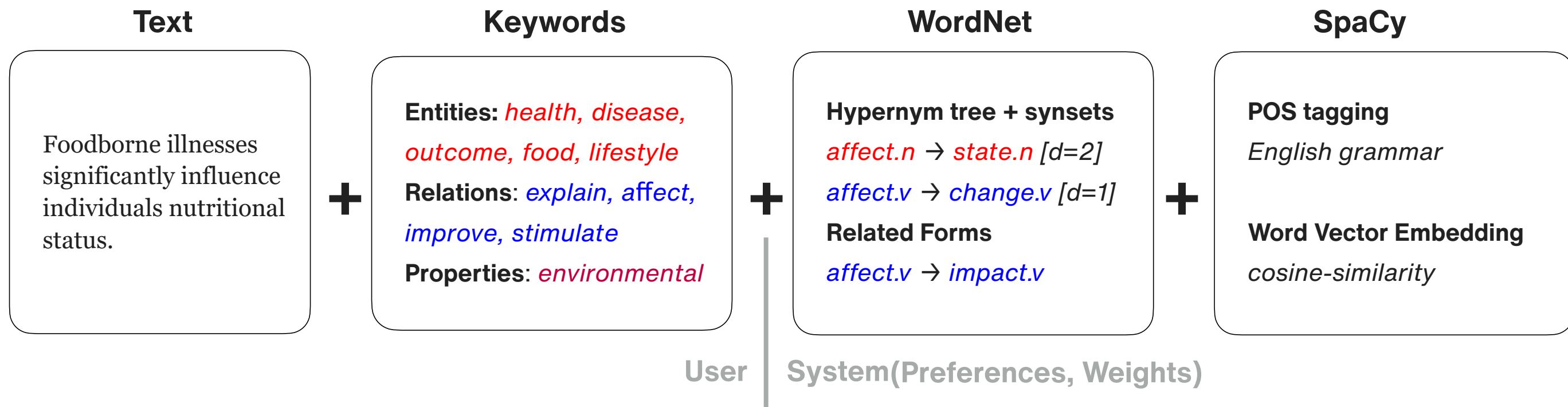
WUP-X( $w, c, t$ ) // Extended Wu-Palmer similarity, considers reachable words from  $w$  and  $c$ 
 $S_w = \text{Extend}(\text{WN-Synsets}(w, t))$  // WN-Synsets returns synonyms of  $w$ 
 $S_c = \text{Extend}(\text{WN-Synsets}(c, t))$ 
return  $\max_{(s_w \xrightarrow{\omega_w} r_w, s_c \xrightarrow{\omega_c} r_c) \in S_w \times S_c} \omega_w \cdot \omega_c \cdot \text{WN-WUP}(r_w, r_c, t)$  // Wu-Palmer from WordNet

VEC( $w, c$ ) // Vector similarity
if  $w = c$  then return 1.0
elif both  $w$  and  $c$  have vector embeddings then
return cosine-similarity of the embeddings of  $w$  and of  $c$ 
else return 0.0

Extend( $S, t$ )
Input:  $S$  a set of WordNet synsets
Output: Weighted extension of  $S$  by pertainyms and derivationally related forms
 $R \leftarrow \emptyset$  // Result relation
for  $s \in S$  do
 $R \leftarrow R \cup \{s \xrightarrow{1.0} s\}$  //  $R$  is reflexive
for  $l \in \text{WN-Lemmas}(s)$  do
 $R \leftarrow R \cup \{s \xrightarrow{\text{P-Weight}} \text{WN-SynSet}(v) \mid v \in \text{WN-Pertainyms}(l)\}$ 
 $R \leftarrow R \cup \{s \xrightarrow{\text{RF-Weight}} \text{WN-SynSet}(v) \mid v \in \text{WN-RelatedForms}(l)\}$ 
return  $R$ 
    
```

← Text/Keywords Similarity

Highlighting Algorithm



Highlighted text

1. It is also likely that **climate change** will contribute to novel occurrences of **disease emergence** and transmission.
- ➔ 2. **Foodborne illnesses** significantly influence individuals nutritional status.
3. **Changing lifestyles**, mainly due to **work commitment**, have fuelled the increase in numbers eating out and the need for **convenience foods**.
4. **Significant changes** have occurred in **food systems** in the last decades that have contributed to widen such 'holes' in the barriers from phase to phase: agricultural intensification and industrialization causing major environmental deterioration, the increasing distance traveled by food in global markets, and the **nutrition transition** towards diets rich in ultra - processed food and animal **protein** are the three cornerstones of such changes.

Algorithm 1 Similarity

```

Similarity( $w, C, t$ )
Input:  $w$  word,  $C$  keywords in canonical form (lemmas),  $t$  type of  $w$  (noun, verb, adjective ...)
Output: Similarity score for  $w$ 
P-Weight  $\leftarrow 0.95$    RF-Weight  $\leftarrow 0.95$    WUP-Threshold  $\leftarrow 0.8$    VEC-Threshold  $\leftarrow 0.95$ 
bestwup  $\leftarrow \max_{c \in C} \text{WUP-X}(w, c, t)$  or else 0.0
bestvec  $\leftarrow \max_{c \in C} \text{VEC}(w, c)$  or else 0.0
if bestwup  $\geq$  WUP-Threshold and bestwup  $\geq$  bestvec then return bestwup
elif bestvec  $\geq$  VEC-Threshold then return bestvec
else return 0.0

WUP-X( $w, c, t$ ) // Extended Wu-Palmer similarity, considers reachable words from  $w$  and  $c$ 
 $S_w = \text{Extend}(\text{WN-Synsets}(w, t))$  // WN-Synsets returns synonyms of  $w$ 
 $S_c = \text{Extend}(\text{WN-Synsets}(c, t))$ 
return  $\max_{(s_w, s_c) \in S_w \times S_c} \omega_w \cdot \omega_c \cdot \text{WN-WUP}(r_w, r_c, t)$  // Wu-Palmer from WordNet

VEC( $w, c$ ) // Vector similarity
if  $w = c$  then return 1.0
elif both  $w$  and  $c$  have vector embeddings then
return cosine-similarity of the embeddings of  $w$  and of  $c$ 
else return 0.0

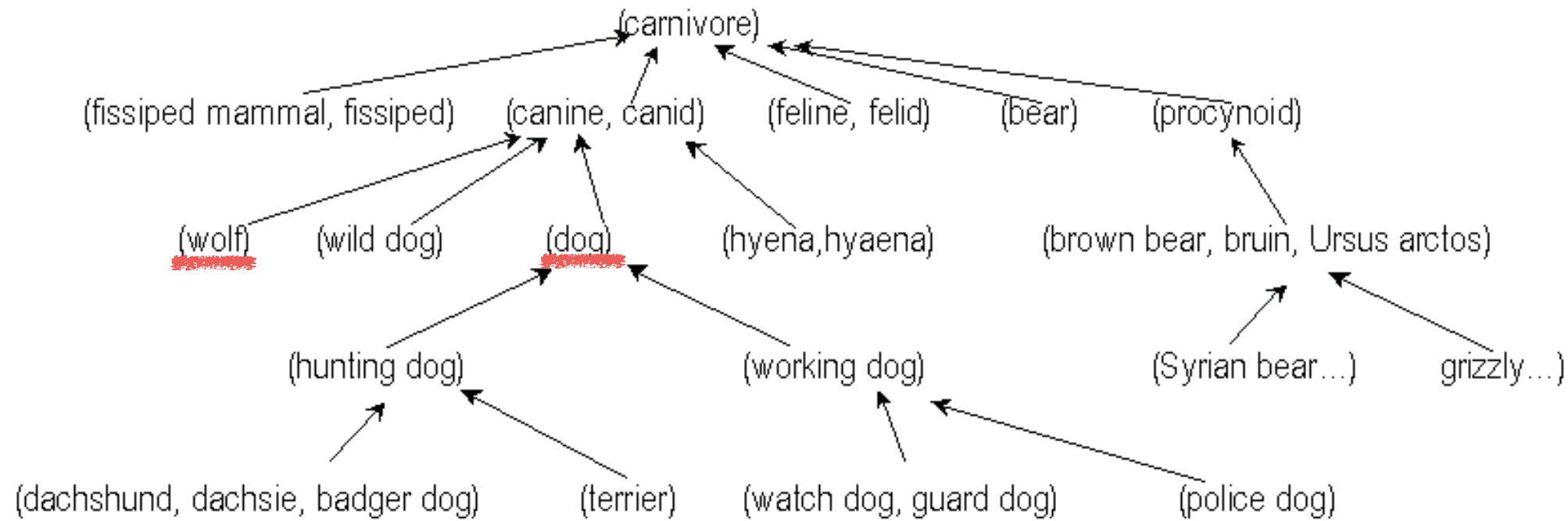
Extend( $S, t$ )
Input:  $S$  a set of WordNet synsets
Output: Weighted extension of  $S$  by pertainyms and derivationally related forms
 $R \leftarrow \emptyset$  // Result relation
for  $s \in S$  do
 $R \leftarrow R \cup \{s \xrightarrow{1.0} s\}$  //  $R$  is reflexive
for  $l \in \text{WN-Lemmas}(s)$  do
 $R \leftarrow R \cup \{s \xrightarrow{\text{P-Weight}} \text{WN-SynSet}(v) \mid v \in \text{WN-Pertainyms}(l)\}$ 
 $R \leftarrow R \cup \{s \xrightarrow{\text{RF-Weight}} \text{WN-SynSet}(v) \mid v \in \text{WN-RelatedForms}(l)\}$ 
return  $R$ 
    
```

← Text/Keywords Similarity

-> WordNet similarity is explainable and accurate

Highlighting Algorithm - WordNet Similarity

WordNet Hypernym Tree



Wu-Palmer("wolf", "dog") = 0.93

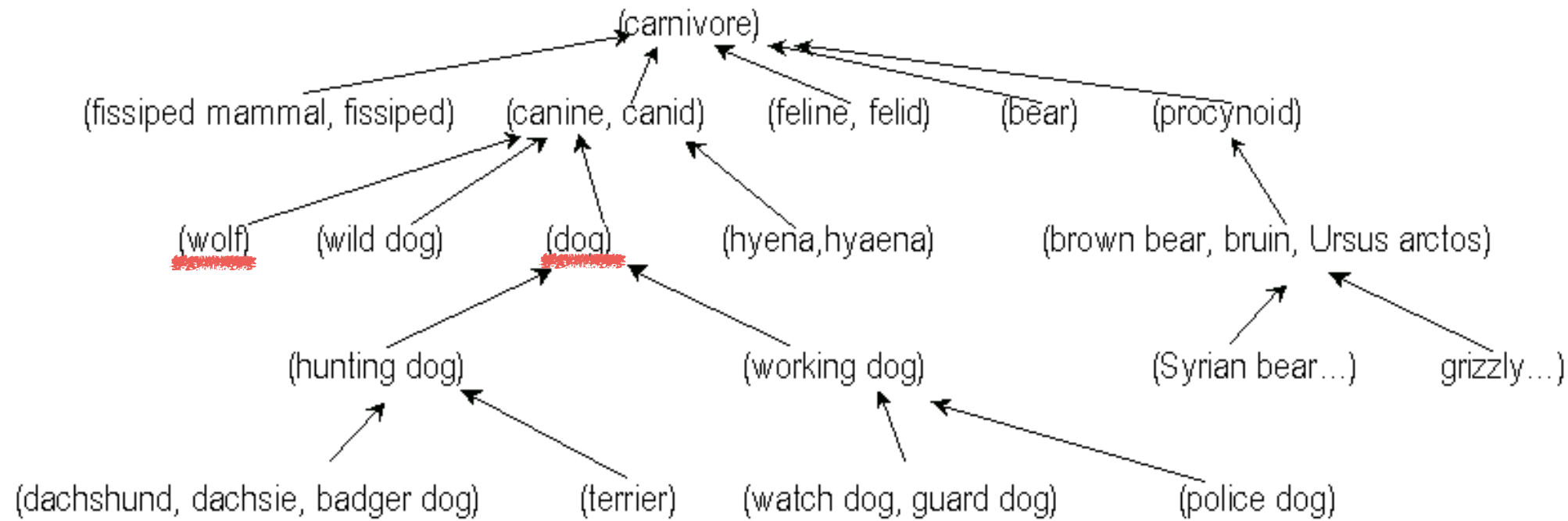
$$\text{Wu - Palmer} = 2 * \frac{\text{depth}(\text{lcs}(s1, s2))}{(\text{depth}(s1) + \text{depth}(s2))}$$

In the Example

2. **Foodborne illnesses** (NCP(Foodborne illnesses, [SimilarTo('disease', 0.95, 'wup']))) significantly **influence** (SimilarTo('affect', 0.84, 'wup')) individuals **nutritional status** (NCP(nutritional status, [SimilarTo('food', 0.91, 'wup']))).

Highlighting Algorithm - WordNet Similarity

WordNet Hypernym Tree



Wu-Palmer("wolf", "dog") = 0.93

$$\text{Wu - Palmer} = 2 * \frac{\text{depth}(\text{lcs}(s1, s2))}{(\text{depth}(s1) + \text{depth}(s2))}$$

In the Example

2. **Foodborne illnesses** (NCP(Foodborne illnesses, [SimilarTo('disease', 0.95, 'wup']))) significantly **influence** (SimilarTo('affect', 0.84, 'wup')) individuals **nutritional status** (NCP(nutritional status, [SimilarTo('food', 0.91, 'wup']))).

- > **Explainable**
- > **Also helps tuning**

SLR Case Study 1 - Results

Query: Associated health focus

Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

SLR Case Study 1 - Results

Query: Associated health focus

How trustworthy/accurate are the LLM answers and evidence?

Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

SLR Case Study 1 - Results

Query: Associated health focus

How trustworthy/accurate are the LLM answers and evidence?

Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

Information	Complexity	Model	Quote Fuzzy Matching Score	Model Average Word Count	Expert Average Word Count	Transformer Similarity	SpaCy Similarity	Model Accuracy
Global Context	Low	GPT-3	97 ± 5	3.5	4.8	0.74 ± 0.25	0.58 ± 0.16	0.9
		GPT-4	98 ± 6	5.2	4.8	0.84 ± 0.06	0.57 ± 0.14	1.0
Associated Health Focus	Low	GPT-3	97 ± 5	5.6	1.6	0.82 ± 0.13	0.57 ± 0.16	0.75
		GPT-4	98 ± 4	7	1.6	0.85 ± 0.05	0.60 ± 0.14	0.94
Transition Pathway	Moderate	GPT-3	95 ± 10	13.4	5.4	0.81 ± 0.04	0.66 ± 0.08	0.65
		GPT-4	97 ± 8	23	5.4	0.85 ± 0.06	0.65 ± 0.21	1.0
Agri-food Boundary	Moderate	GPT-3	97 ± 4	32	17	0.83 ± 0.04	0.77 ± 0.14	0.5
		GPT-4	98 ± 6	50.6	17	0.87 ± 0.03	0.79 ± 0.12	0.85
Public Health Risk	Moderate	GPT-3	99 ± 7	8.8	6.5	0.85 ± 0.05	0.59 ± 0.16	0.7
		GPT-4	97 ± 6	20.5	6.5	0.87 ± 0.06	0.74 ± 0.17	0.95
Synergies	High	GPT-3	97 ± 5	31.3	26.4	0.83 ± 0.03	0.84 ± 0.07	0.25
		GPT-4	98 ± 5	58	26.4	0.81 ± 0.05	0.83 ± 0.07	0.1
Constraints	High	GPT-3	97 ± 5	35	18	0.82 ± 0.02	0.81 ± 0.08	0.44
		GPT-4	98 ± 5	59	18	0.84 ± 0.02	0.83 ± 0.07	1.0
Integrated Solutions	High	GPT-3	97 ± 4	28	30	0.89 ± 0.04	0.90 ± 0.07	0.88
		GPT-4	99 ± 3	50	30	0.89 ± 0.05	0.89 ± 0.07	1.0

SLR Case Study 1 - Results

Query: Associated health focus

How trustworthy/accurate are the LLM answers and evidence?

Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

Information	Complexity	Model	Quote Fuzzy Matching Score	Model Average Word Count	Expert Average Word Count	Transformer Similarity	SpaCy Similarity	Model Accuracy
Global Context	Low	GPT-3	97 ± 5	3.5	4.8	0.74 ± 0.25	0.58 ± 0.16	0.9
		GPT-4	98 ± 6	5.2	4.8	0.84 ± 0.06	0.57 ± 0.14	1.0
Associated Health Focus	Low	GPT-3	97 ± 5	5.6	1.6	0.82 ± 0.13	0.57 ± 0.16	0.75
		GPT-4	98 ± 4	7	1.6	0.85 ± 0.05	0.60 ± 0.14	0.94
Transition Pathway	Moderate	GPT-3	95 ± 10	13.4	5.4	0.81 ± 0.04	0.66 ± 0.08	0.65
		GPT-4	97 ± 8	23	5.4	0.85 ± 0.06	0.65 ± 0.21	1.0
Agri-food Boundary	Moderate	GPT-3	97 ± 4	32	17	0.83 ± 0.04	0.77 ± 0.14	0.5
		GPT-4	98 ± 6	50.6	17	0.87 ± 0.03	0.79 ± 0.12	0.85
Public Health Risk	Moderate	GPT-3	99 ± 7	8.8	6.5	0.85 ± 0.05	0.59 ± 0.16	0.7
		GPT-4	97 ± 6	20.5	6.5	0.87 ± 0.06	0.74 ± 0.17	0.95
Synergies	High	GPT-3	97 ± 5	31.3	26.4	0.83 ± 0.03	0.84 ± 0.07	0.25
		GPT-4	98 ± 5	58	26.4	0.81 ± 0.05	0.83 ± 0.07	0.1
Constraints	High	GPT-3	97 ± 5	35	18	0.82 ± 0.02	0.81 ± 0.08	0.44
		GPT-4	98 ± 5	59	18	0.84 ± 0.02	0.83 ± 0.07	1.0
Integrated Solutions	High	GPT-3	97 ± 4	28	30	0.89 ± 0.04	0.90 ± 0.07	0.88
		GPT-4	99 ± 3	50	30	0.89 ± 0.05	0.89 ± 0.07	1.0

Evidence occurs in text?

SLR Case Study 1 - Results

Query: Associated health focus

How trustworthy/accurate are the LLM answers and evidence?

Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

Information	Complexity	Model	Quote Fuzzy Matching Score	Model Average Word Count	Expert Average Word Count	Transformer Similarity	SpaCy Similarity	Model Accuracy
Global Context	Low	GPT-3	97 ± 5	3.5	4.8	0.74 ± 0.25	0.58 ± 0.16	0.9
		GPT-4	98 ± 6	5.2	4.8	0.84 ± 0.06	0.57 ± 0.14	1.0
Associated Health Focus	Low	GPT-3	97 ± 5	5.6	1.6	0.82 ± 0.13	0.57 ± 0.16	0.75
		GPT-4	98 ± 4	7	1.6	0.85 ± 0.05	0.60 ± 0.14	0.94
Transition Pathway	Moderate	GPT-3	95 ± 10	13.4	5.4	0.81 ± 0.04	0.66 ± 0.08	0.65
		GPT-4	97 ± 8	23	5.4	0.85 ± 0.06	0.65 ± 0.21	1.0
Agri-food Boundary	Moderate	GPT-3	97 ± 4	32	17	0.83 ± 0.04	0.77 ± 0.14	0.5
		GPT-4	98 ± 6	50.6	17	0.87 ± 0.03	0.79 ± 0.12	0.85
Public Health Risk	Moderate	GPT-3	99 ± 7	8.8	6.5	0.85 ± 0.05	0.59 ± 0.16	0.7
		GPT-4	97 ± 6	20.5	6.5	0.87 ± 0.06	0.74 ± 0.17	0.95
Synergies	High	GPT-3	97 ± 5	31.3	26.4	0.83 ± 0.03	0.84 ± 0.07	0.25
		GPT-4	98 ± 5	58	26.4	0.81 ± 0.05	0.83 ± 0.07	0.1
Constraints	High	GPT-3	97 ± 5	35	18	0.82 ± 0.02	0.81 ± 0.08	0.44
		GPT-4	98 ± 5	59	18	0.84 ± 0.02	0.83 ± 0.07	1.0
Integrated Solutions	High	GPT-3	97 ± 4	28	30	0.89 ± 0.04	0.90 ± 0.07	0.88
		GPT-4	99 ± 3	50	30	0.89 ± 0.05	0.89 ± 0.07	1.0

Evidence occurs in text?

LLM / Word Emb. answer similarity

SLR Case Study 1 - Results

Query: Associated health focus

How trustworthy/accurate are the LLM answers and evidence?

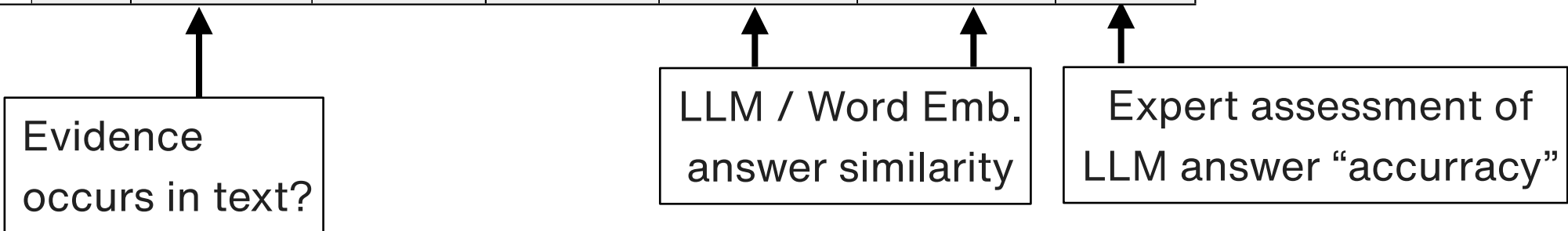
Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

Information	Complexity	Model	Quote Fuzzy Matching Score	Model Average Word Count	Expert Average Word Count	Transformer Similarity	SpaCy Similarity	Model Accuracy
Global Context	Low	GPT-3	97 ± 5	3.5	4.8	0.74 ± 0.25	0.58 ± 0.16	0.9
		GPT-4	98 ± 6	5.2	4.8	0.84 ± 0.06	0.57 ± 0.14	1.0
Associated Health Focus	Low	GPT-3	97 ± 5	5.6	1.6	0.82 ± 0.13	0.57 ± 0.16	0.75
		GPT-4	98 ± 4	7	1.6	0.85 ± 0.05	0.60 ± 0.14	0.94
Transition Pathway	Moderate	GPT-3	95 ± 10	13.4	5.4	0.81 ± 0.04	0.66 ± 0.08	0.65
		GPT-4	97 ± 8	23	5.4	0.85 ± 0.06	0.65 ± 0.21	1.0
Agri-food Boundary	Moderate	GPT-3	97 ± 4	32	17	0.83 ± 0.04	0.77 ± 0.14	0.5
		GPT-4	98 ± 6	50.6	17	0.87 ± 0.03	0.79 ± 0.12	0.85
Public Health Risk	Moderate	GPT-3	99 ± 7	8.8	6.5	0.85 ± 0.05	0.59 ± 0.16	0.7
		GPT-4	97 ± 6	20.5	6.5	0.87 ± 0.06	0.74 ± 0.17	0.95
Synergies	High	GPT-3	97 ± 5	31.3	26.4	0.83 ± 0.03	0.84 ± 0.07	0.25
		GPT-4	98 ± 5	58	26.4	0.81 ± 0.05	0.83 ± 0.07	0.1
Constraints	High	GPT-3	97 ± 5	35	18	0.82 ± 0.02	0.81 ± 0.08	0.44
		GPT-4	98 ± 5	59	18	0.84 ± 0.02	0.83 ± 0.07	1.0
Integrated Solutions	High	GPT-3	97 ± 4	28	30	0.89 ± 0.04	0.90 ± 0.07	0.88
		GPT-4	99 ± 3	50	30	0.89 ± 0.05	0.89 ± 0.07	1.0



SLR Case Study 1 - Results

Query: Associated health focus

How trustworthy/accurate are the LLM answers and evidence?

Expert answer: “**Novel pathogens** (COVID-19)”

word count = 4

GPT4 answer: “The **associated health focus** of this paper is COVID-19, a **zoonotic infectious disease** caused by the **SARS-CoV-2 virus.**”

Word count = 4, Transformer similarity to expert answer = ..., SpaCy word embedding similarity = ...,

Information	Complexity	Model	Quote Fuzzy Matching Score	Model Average Word Count	Expert Average Word Count	Transformer Similarity	SpaCy Similarity	Model Accuracy
Global Context	Low	GPT-3	97 ± 5	3.5	4.8	0.74 ± 0.25	0.58 ± 0.16	0.9
		GPT-4	98 ± 6	5.2	4.8	0.84 ± 0.06	0.57 ± 0.14	1.0
Associated Health Focus	Low	GPT-3	97 ± 5	5.6	1.6	0.82 ± 0.13	0.57 ± 0.16	0.75
		GPT-4	98 ± 4	7	1.6	0.85 ± 0.05	0.60 ± 0.14	0.94
Transition Pathway	Moderate	GPT-3	95 ± 10	13.4	5.4	0.81 ± 0.04	0.66 ± 0.08	0.65
		GPT-4	97 ± 8	23	5.4	0.85 ± 0.06	0.65 ± 0.21	1.0
Agri-food Boundary	Moderate	GPT-3	97 ± 4	32	17	0.83 ± 0.04	0.77 ± 0.14	0.5
		GPT-4	98 ± 6	50.6	17	0.87 ± 0.03	0.79 ± 0.12	0.85
Public Health Risk	Moderate	GPT-3	99 ± 7	8.8	6.5	0.85 ± 0.05	0.59 ± 0.16	0.7
		GPT-4	97 ± 6	20.5	6.5	0.87 ± 0.06	0.74 ± 0.17	0.95
Synergies	High	GPT-3	97 ± 5	31.3	26.4	0.83 ± 0.03	0.84 ± 0.07	0.25
		GPT-4	98 ± 5	58	26.4	0.81 ± 0.05	0.83 ± 0.07	0.1
Constraints	High	GPT-3	97 ± 5	35	18	0.82 ± 0.02	0.81 ± 0.08	0.44
		GPT-4	98 ± 5	59	18	0.84 ± 0.02	0.83 ± 0.07	1.0
Integrated Solutions	High	GPT-3	97 ± 4	28	30	0.89 ± 0.04	0.90 ± 0.07	0.88
		GPT-4	99 ± 3	50	30	0.89 ± 0.05	0.89 ± 0.07	1.0

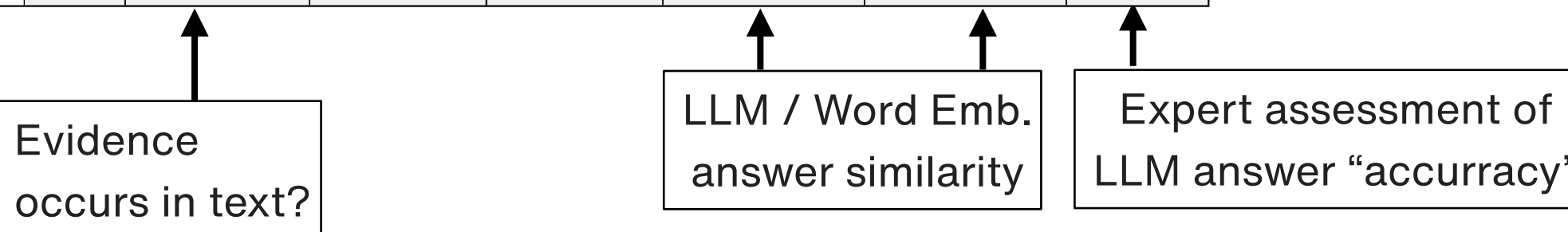
Findings

Transformer similarity better than SpaCy similarity

Correlation between expert answer and GPT3 answer is -0.07 +- 0.08

Correlation between expert answer and GPT4 answer is 0.48 +- 0.09

-> GPT4 “performs well” (on low to moderate complexity?)



More Evaluation and Conclusions

Case Study 2 - Coordinated Response to Crisis Management SLR

Experiment: ask LLM for “evidence” vs “direct” (no evidence to be reported)

-> Retrieving evidence decreases accuracy (overload?)

Case Study 3 - Sustainable Transitions SLR

Experiment: querying LLM “separately” vs “together”, for all research questions

-> “Together” model unusably inaccurate

Case Study 4 - Automatically Marking SLR

20 papers almost all “relevant” / 20 papers almost all “irrelevant”

Experiment: can LLM identify relevant papers only?

-> Very high false positive rate / very low false negative rate

LLM General

GPT4 context window larger than GPT3 -> improves evidence finding

GPT4: Overall 83% accuracy wrt expert query answers (careful with such metrics though)

Performance degrades with difficulty of query (compare “global context” vs “synergy”)

Highlighting

Helps sifting through papers (anecdotal evidence)

Low highlighting rate of paper means: (a) irrelevant paper or (b) expert missed relevant issues

Highlighting rate as similarity measure?