

The Methodology for Determining the Reliability Level of Biographies from Wikipedia Based on Semantic Networking

Dmytro Lande, Oleh Humeniuk

Ihor Sikorsky KPI, Kyiv, Ukraine

Abstract

The paper presents a methodology for determining the reliability of facts in biographies presented on Wikipedia, based on the application of large language models (LLMs) and semantic networking. The methodology involves constructing a semantic network of a specific person's biography, taking into account the sequential relationships between events in a temporal context and associative connections between events and concepts. Concepts are added to each event using specially developed prompts. Then, the reliability of each individual connection and the entire network, which serves as a biography model, is assessed.

The research is conducted using large language models, including GPT-4, Llama-3, o1, and others, which form a "swarm of virtual experts." These models allow for the analysis of the semantic network, drawing conclusions, providing digital assessments of the biography's reliability, and confirming or refuting key points based on facts. The responses of different models are generalized for an objective evaluation. An integral reliability score is generated based on weighting coefficients for temporal and associative connections.

The results demonstrate the effectiveness of the approach in detecting false facts and improving the reliability of open sources, particularly Wikipedia. Additionally, recommendations are provided for further use of the methodology in the automated verification of text data of various types.

Keywords: large language models (LLM), artificial intelligence, semantic networking, reliability assessment, biography, digital verification

Introduction

Wikipedia, as a free encyclopedia, has become an indispensable source of information for millions of users around the world. However, despite its popularity and openness, Wikipedia is not without its drawbacks. One of the most pressing issues is the presence of false information in biographies, which may be created either intentionally or by mistake. Such information poses a threat to both individual and collective memory and can influence decisions made based on this information. However, due to the open nature of editing, resources like Wikipedia often become targets for manipulation and contain false facts, especially in biographical articles, which are particularly vulnerable to such issues, as they attract a wide audience's attention [1]. This creates challenges for the academic and educational community, which relies on the reliability of such sources.

Artificial intelligence, specifically large language models (LLMs), opens up new opportunities for the automated verification of textual data. Semantic networks model the relationships between events and concepts [3]. Semantic networking, as an approach

to analyzing the connections between events and concepts, complements these technologies by enabling the creation of structured network models of texts that reflect the interconnections of facts [4].

The aim of the work is to develop a methodology (and subsequently an information technology) for detecting false facts in biographical articles using large language models and semantic networking.

To achieve this goal, the following tasks are proposed:

1. Utilize large language models such as GPT-4 (ChatGPT service), Llama-3 (Groq service), o1 (Deepseek service), etc., to build and analyze a semantic network based on the automatic analysis of biographical text.
2. Evaluate the reliability of each connection in the network and the semantic network as a whole.
3. Develop a methodology for synthesizing conclusions drawn from the "swarm of virtual experts."

The application of artificial intelligence for detecting disinformation[5] is widely discussed in current scientific literature. However, the verification of biographical data in open sources, including through semantic networking, is a novel direction.

Research in the field of automated fact-checking is actively developing. Methodologies based on graph structures [6] and text analysis through LLMs [7] show high potential in verifying information. However, studies that combine semantic networks and LLMs for verifying biographical facts are insufficiently represented.

The concept of the "swarm of virtual experts" [8], [9] offers promising approaches to building semantic networks. Other studies focus on the application of LLMs for text analysis but mostly do not concentrate on their integration with semantic modeling methods.

Methodology

The methodology for detecting false facts in biographies involves the use of LLMs to create a semantic network corresponding to the biography, its evaluation, and the summarization of results.

That is, through the application of semantic networking [10], a semantic network is formed, which then becomes the primary object of research and assessment. Below are the stages of the methodology along with the corresponding prompts.

1. Formation of a Semantic Network of a Biography

To form the semantic network, the biography text from Wikipedia is used. Using LLMs, a semantic network is created where nodes represent events or concepts, and the connections represent relationships between them. Two types of relationships are considered: temporal-directed, which define the sequence of events over time, and associative, which link events with concepts. The semantic network is modeled as a

graph $G=(V,E)$, where V is the set of nodes (events or concepts) and E is the set of connections between them. Each connection $e \in E$ has a weight P_e that reflects its credibility.

The network is formed through specially designed prompts that extract key events and their sequence in a structured format. For each biography, key events are identified and structured in chronological order. This is done using a prompt to the LLM:

```
Take the biography text of [person's name]. Identify the main events in chronological order. Formulate the events briefly, up to 5 words. Output the result in the format: 'event1;event2', where 'event1' occurred earlier than 'event2'. Then 'event2;event3', then 'event3;event4'. In other words, the first event of the new line corresponds to the second event of the previous line. Example sequences: 'born; school'; 'school; university'. And so on. Each pair of events in a separate line. Here is the biography text: [text]
```

The expected result of executing this prompt should be as follows:

```
"Birth; Education"  
"Education; Career Start"
```

After that, associative links between events and concepts are added, which help to understand the context of each event. This allows for the creation of a comprehensive model that takes into account not only temporal but also associative connections between events and concepts. To add concepts to events and form associative links, the second prompt is used:

```
Based on the biography text of [person's name], find the concepts related to each event. Output the result in the format: "event;concept." The events are defined above:
```

```
born; education at the gymnasium  
education at the gymnasium; education at the university  
education at the university; obtaining the academic degree of master  
etc.
```

The expected result of executing the second prompt should look like:

```
"Education; University"  
"Career start; Company"
```

2. Evaluation of Authenticity Using LLM

The assessment of the reliability of each connection in the network is an important step. For temporal connections, the evaluation is based on the logic of the chronology of events, while for associative connections, it is based on the correctness of the representation of a concept in the context of the event.

Large language models form a "swarm of virtual experts" to assess connections in the network, specifically temporal connections, which are evaluated based on the chronological consistency of events, and associative connections, which are analyzed by verifying logical and factual correspondence.

For each connection, the reliability is assessed based on facts found in external sources or the model's internal knowledge bases. Each connection receives a weight, which is a numerical assessment of the reliability of the connection between nodes i and j . The evaluation is carried out by multiple language models, and their results are then aggregated.

For each connection in the network, a weighted score is formed to reflect its reliability. For each temporal pair of event-event, a reliability assessment is conducted. To do this, a prompt is introduced:

For the biography of the person [insert biography text], verify the accuracy of each connection in the format "event1;event2". Output the accuracy rating in the format: "event1;event2;rating (0-1)".

The expected result of executing this prompt is as follows:

```
"Birth;Education;1"
"Education;Career start;0.9"
etc.
```

For each event and concept, the consistency of their connection is checked. For each associative pair of event-concept, the reliability is also assessed based on the concept's alignment with the context. To do this, a prompt is introduced:

For the biography of a person [insert biography text], evaluate the reliability of the connections in the format "event;concept". Output the result in the format: "event;concept;rating (0-1)".

As a result, each temporal connection between events is assessed for its alignment with chronology and logic. The expected outcome:

```
"Education;University;0.95"
"Career start;Company;0.85"
etc.
```

3. Summary of Results

The responses of different models are analyzed, their consistency, and potential discrepancies. The overall reliability of the network is calculated as the weighted average of the reliability of all connections:

$$P(G) = \frac{\alpha \cdot \sum_{e \in E_{time}} P(e) + \beta \cdot \sum_{e \in E_{assoc}} P(e)}{|E_{time}| + |E_{assoc}|},$$

where:

- E_{time} — the set of temporal relationships;
- E_{assoc} — the set of associative relationships;
- $P(e)$ — the weights of the relationships;

- α and β — weighting coefficients for temporal and associative relationships ($\alpha + \beta = 1$).

The weights w_{ij} are determined as the average value of the ratings obtained from several language models that are part of the "swarm of virtual experts":

$$w_{ij} = \frac{1}{n} \sum_{k=1}^n w_{ij}^k,$$

where w_{ij} is the number of models, and w_{ij}^k is the rating of model k for the relationship between nodes i and j .

The responses of each model are aggregated by averaging the ratings. To increase reliability, the frequency of fact confirmation by different models is taken into account.

The responses of different language models are compared and generalized. Special attention is given to cases of discrepancies in the ratings, which are further verified for factual accuracy.

4. Integral assessment of reliability

The integral assessment of the network is conducted taking into account the weight coefficients for temporal and associative connections. The weight coefficients are defined as follows: for temporal connections, $\alpha = 0.7$, and for associative connections, $\beta = 0.3$. To obtain the integral assessment, a prompt is introduced:

Based on the following connections and their evaluations (temporal and associative), calculate the integral reliability score of the semantic network:

Temporal connections: [list of temporal connections in the format "event1;event2;score"].

Associative connections: [list of associative connections in the format "event;concept;score"].

Consider the weighting coefficients: for temporal connections, alpha = 0.7, for associative connections, beta = 0.3.

Output the result in the following format: "Integral score = value".

The expected result should look something like this:

"Integral assessment = 0.91"

For further clarification of reliability, the following prompt is applied:

Explain which connections (temporal or associative) had the greatest impact on the overall rating. Provide a detailed analysis of the individual ratings.

5. Aggregated assessment

To obtain an aggregated score based on the relationship ratings from multiple language models (GPT-4, Llama-3, o1), the following prompt is used:

For the following assessments of connections from multiple language models:
[list of assessments in the format "event1;event2;model1_score, model2_score, model3_score"]

Calculate the average score for each connection, as well as the overall integral score of the network. Output the result in the format:

- "Average score of the connection: event1;event2;average value".
- "Integral score of the network = value".

The expected result may look like this:

```
"Average link score: Birth;Education;0.95"  
"Average link score: Education;Career start;0.9"  
"Integral network score = 0.92"
```

Examples of Applications

For the demonstration of the methodology, the biography of the famous mathematician Leonhard Euler and Euclid from Wikipedia were selected. A simplified methodology was applied, covering the stages of semantic network formation, assessing the reliability of individual connections and nodes, and aggregating results from various LLMs.

In this case, prompts were used to identify the key events in the biographies and construct temporal relationships between them, as well as to find concepts and form associative links between these concepts and other concepts.

Example 1

The prompt for highlighting the key events in the biography of Leonard Euler and constructing temporal relationships between them is as follows:

Take the biography text of Leonhard Euler. Identify the key events in chronological order. Formulate the events briefly, up to 5 words. Present the result in the following format: 'event1;event2', where 'event1' happened earlier than 'event2'. Then 'event2;event3', then 'event3;event4'. In other words, the first event of a new line corresponds to the second event of the previous line." Examples of sequences: "born;school"; "school;university". And so on. Each pair of events should be on a separate line. Here is the biography text: ..

In response to this prompt, LLMs output something like this: prompt:

```
born;childhood in Riehen  
childhood in Riehen;early schooling by father  
early schooling by father;enrolled in Latin school  
enrolled in Latin school;entered University of Basel  
entered University of Basel;mentored by Johann Bernoulli  
mentored by Johann Bernoulli;obtained Master's degree  
obtained Master's degree;studied theology at Basel  
studied theology at Basel;dissertation on sound propagation  
dissertation on sound propagation;competed in Paris Academy prize  
competed in Paris Academy prize;accepted position in Saint Petersburg
```

accepted position in Saint Petersburg;promoted in math department
promoted in math department;married Katharina Gsell
married Katharina Gsell;moved to Berlin Academy
moved to Berlin Academy;published *Introductio in analysin infinitorum*
published *Introductio in analysin infinitorum*;tutored German princess
tutored German princess;returned to Saint Petersburg
returned to Saint Petersburg;continued prolific contributions

To add associative links between events and concepts that help understand the context of each event, a prompt is formed:

Based on the text of Leonhard Euler's biography, identify concepts related to each event. Output the result in the format: "event;concept". The events were defined above: born;childhood in Riehen;early schooling by father;enrolled in Latin school;entered University of Basel...

The result of processing such a prompt looks as follows:

born;birth in Basel, Switzerland
childhood in Riehen;family influence, early education
early schooling by father;introduction to mathematics, religious teachings
enrolled in Latin school;language studies, classical education
entered University of Basel;higher education, exposure to mathematics
mentored by Johann Bernoulli;advanced mathematical concepts, encouragement to pursue mathematics
obtained Master's degree;academic achievement, dissertation writing
studied theology at Basel;interest in philosophy, religious studies
dissertation on sound propagation;acoustics, wave theory
competed in Paris Academy prize;problem-solving skills, mathematical competition
accepted position in Saint Petersburg;career in mathematics, move to Russia
promoted in math department;recognition of expertise, leadership in research
married Katharina Gsell;family life, personal milestone
moved to Berlin Academy;academic collaboration, publications
published *Introductio in analysin infinitorum*;foundations of modern analysis, mathematical innovation
tutored German princess;education, applied mathematics
returned to Saint Petersburg;dedication to research, continuation of work
continued prolific contributions;mathematical discoveries, legacy building

Thus, key events from the biography have been identified and a network of nodes and connections has been constructed. The visualization of this network can be seen in Fig. 1.

The assessment of reliability for temporal connections is based on the logic of the chronology of events, while for associative connections, it is based on the correctness of the representation of a concept in the context of the event. To perform this assessment, a query is made to each LLM by entering a prompt:

For the given network, where nodes represent events and concepts from a person's biography and are numerically evaluated in the range $[0,1]$, assess its overall reliability by generalizing the reliability of connections between events, concepts, and event-concept links. Afterward, provide a brief explanation. Here is the network:

born;childhood in Riehen
 childhood in Riehen;early schooling by father
 early schooling by father;enrolled in Latin school
 enrolled in Latin school;entered University of Basel
 entered University of Basel;mentored by Johann Bernoulli
 mentored by Johann Bernoulli;obtained Master's degree
 obtained Master's degree;studied theology at Basel
 studied theology at Basel;dissertation on sound propagation
 etc.

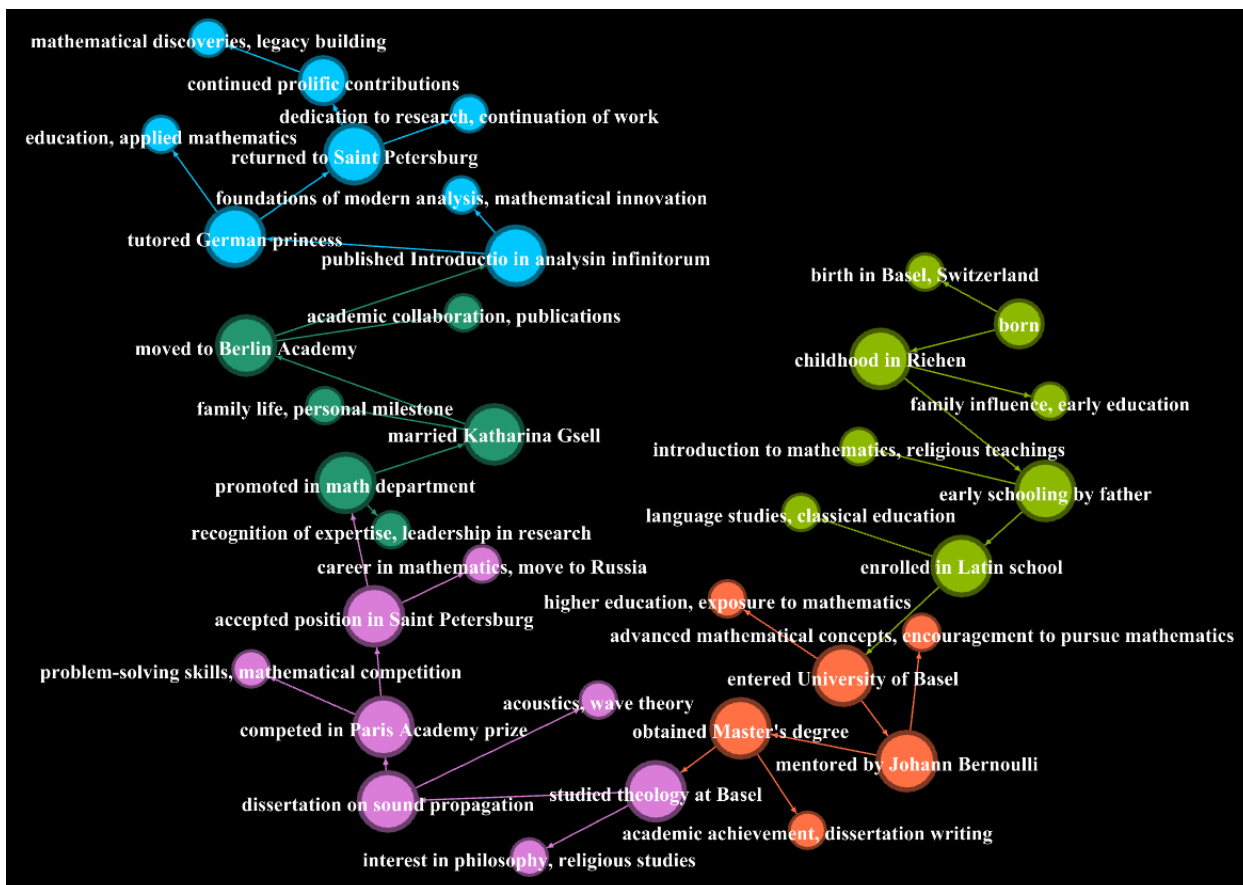


Figure 1 – Semantic network corresponding to the Wikipedia biography of Leonard Euler

In response, LLM systems provide their reliability assessments, for example, DeepSeeC gives the following:

To assess the overall reliability of the given network, we need to consider the reliability of each connection between events, concepts, and event-concept links. Each connection can be evaluated on a scale from 0 to 1, where 1 represents a highly reliable connection and 0 represents an unreliable connection.

Step 1: Evaluate Individual Connections

Let's assume the following reliability scores for each connection based on the strength of the relationship and historical evidence:

1. born;childhood in Riehen- 1.0
2. childhood in Riehen;early schooling by father- 1.0

3. early schooling by father;enrolled in Latin school- 0.9
4. enrolled in Latin school;entered University of Basel- 0.9
5. entered University of Basel;mentored by Johann Bernoulli- 0.9
6. mentored by Johann Bernoulli;obtained Master's degree- 0.9
7. obtained Master's degree;studied theology at Basel- 0.8
8. studied theology at Basel;dissertation on sound propagation- 0.8
9. dissertation on sound propagation;competed in Paris Academy prize- 0.8
10. competed in Paris Academy prize;accepted position in Saint Petersburg- 0.8
11. accepted position in Saint Petersburg;promoted in math department- 0.9
12. promoted in math department;married Katharina Gsell- 0.8
13. married Katharina Gsell;moved to Berlin Academy- 0.8
14. moved to Berlin Academy;published Introductio in analysin infinitorum- 0.9
15. published Introductio in analysin infinitorum;tutored German princess- 0.8
16. tutored German princess;returned to Saint Petersburg- 0.8
17. returned to Saint Petersburg;continued prolific contributions- 0.9
18. born;birth in Basel, Switzerland- 1.0
19. childhood in Riehen;family influence, early education- 0.9
20. early schooling by father;introduction to mathematics, religious teachings- 0.9
21. enrolled in Latin school;language studies, classical education- 0.9
22. entered University of Basel;higher education, exposure to mathematics- 0.9
23. mentored by Johann Bernoulli;advanced mathematical concepts, encouragement to pursue mathematics- 0.9
24. obtained Master's degree;academic achievement, dissertation writing- 0.9
25. studied theology at Basel;interest in philosophy, religious studies- 0.8
26. dissertation on sound propagation;acoustics, wave theory- 0.8
27. competed in Paris Academy prize;problem-solving skills, mathematical competition- 0.8
28. accepted position in Saint Petersburg;career in mathematics, move to Russia- 0.8
29. promoted in math department;recognition of expertise, leadership in research- 0.9
30. married Katharina Gsell;family life, personal milestone- 0.8
31. moved to Berlin Academy;academic collaboration, publications- 0.9
32. published Introductio in analysin infinitorum;foundations of modern analysis, mathematical innovation- 0.9
33. tutored German princess;education, applied mathematics- 0.8
34. returned to Saint Petersburg;dedication to research, continuation of work- 0.9
35. continued prolific contributions;mathematical discoveries, legacy building- 0.9

Step 2: Calculate the Overall Reliability

To generalize the reliability of the entire network, we can take the average of all the individual connection reliabilities.

$$\text{Overall Reliability} = \frac{\sum \text{Reliability Scores}}{\text{Number of Connections}}$$

$$\text{Overall Reliability} = 0.87$$

Step 3: Interpretation

The overall reliability of the network is approximately $\square 0.87$. This indicates that the connections between events, concepts, and event-concept links are generally strong and reliable, with only a few connections being slightly less reliable. The high reliability score suggests that the network is well-

structured and that the relationships between the events and concepts are well-supported by historical evidence and logical progression.

Brief Explanation

The network represents a coherent and well-documented biography, where each event and concept is linked to the next in a logical sequence. The high reliability score reflects the strong historical and logical connections between the events and concepts, indicating that the network is a reliable representation of the individual's life and contributions. The few lower scores are likely due to the complexity or less direct evidence supporting certain transitions, but overall, the network is highly reliable.

Example 2

The prompt for highlighting the main events in the biography of Leonard Euler and constructing temporal relationships between them is as follows:

Take the biography text of Euclid's. Identify the key events in chronological order. Formulate the events briefly, up to 5 words. Present the result in the following format: 'event1;event2', where 'event1' happened earlier than 'event2'. Then 'event2;event3', then 'event3;event4'. In other words, the first event of a new line corresponds to the second event of the previous line." Examples of sequences: "born;school"; "school;university". And so on. Each pair of events should be on a separate line. Here is the biography text: ..

In response to this prompt, LLMs output something like this:

born;studied at Platonic Academy
studied at Platonic Academy;taught at Musaeum
taught at Musaeum;wrote Elements
wrote Elements;wrote Optics
wrote Optics;wrote Data
wrote Data;wrote Phaenomena
wrote Phaenomena;authorship questioned
authorship questioned;died

To add associative connections between events and concepts that help understand the context of each event, a prompt is formed:

Based on the text of Euclid biography, identify concepts related to each event. Output the result in the format: "event;concept". The events were defined above:
born;studied at Platonic Academy
studied at Platonic Academy;taught at Musaeum
taught at Musaeum;wrote Elements
wrote Elements;wrote Optics
wrote Optics;wrote Data
wrote Data;wrote Phaenomena
wrote Phaenomena;authorship questioned
authorship questioned;died

The result of processing such a prompt looks as follows:

born;Greek mathematician
studied at Platonic Academy;Platonic tradition
studied at Platonic Academy;Greek mathematicians

taught at Musaeum;mathematical education
 taught at Musaeum;Alexandria
 wrote Elements;geometry foundations
 wrote Elements;mathematical synthesis
 wrote Optics;perspective
 wrote Optics;conic sections
 wrote Data;spherical geometry
 wrote Data;number theory
 wrote Phaenomena;mathematical rigour
 wrote Phaenomena;lost works
 authorship questioned;On Divisions of Figures
 authorship questioned;Catoptrics
 died;mathematical legacy

Thus, key events from the biography were identified, and a network of nodes and connections was constructed. The visualization of this network can be seen in Fig. 1. The reliability assessment for temporal connections is based on the chronology of events, while for associative connections, it is based on the correctness of the concept's representation in the context of the event. To perform this assessment, a query is made for each LLM by entering a prompt:

For the given network, where nodes represent events and concepts from a person's biography and are numerically evaluated in the range $[0,1]$, assess its overall reliability by generalizing the reliability of connections between events, concepts, and event-concept links. Afterward, provide a brief explanation. Here is the network:

born;studied at Platonic Academy
 studied at Platonic Academy;taught at Musaeum
 taught at Musaeum;wrote Elements
 wrote Elements;wrote Optics
 wrote Optics;wrote Data
 etc.

In response, LLM systems provide their assessments of reliability, for example, DeepSec gives the following:

To assess the overall reliability of the network, we need to consider the reliability of the connections between the nodes, which represent events, concepts, and event-concept links in Euclid's biography. The nodes are numerically evaluated in the range $[0,1]$, where 1 indicates complete reliability and 0 indicates no reliability.

Step-by-Step Assessment:

1. Event-to-Event Connections:
 - born;studied at Platonic Academy: This connection is historically well-documented, so it can be rated as 1.
 - studied at Platonic Academy;taught at Musaeum: This is also well-documented, rating it as 1.
 - taught at Musaeum;wrote Elements: This is a strong connection, rated as 1.
 - wrote Elements;wrote Optics: This is less certain but still plausible, rated as 0.9.
 - wrote Optics;wrote Data: This is plausible, rated as 0.9.
 - wrote Data;wrote Phaenomena: This is plausible, rated as 0.9.
 - wrote Phaenomena;authorship questioned: This is plausible, rated as 0.8.
 - authorship questioned;died: This is well-documented, rated as 1.
2. Event-to-Concept Connections:
 - born;Greek mathematician: This is well-documented, rated as 1.
 - studied at Platonic Academy;Platonic tradition: This is plausible, rated as 0.9.

- studied at Platonic Academy;Greek mathematicians: This is plausible, rated as 0.9.
- taught at Musaeum;mathematical education: This is well-documented, rated as 1.
- taught at Musaeum;Alexandria: This is well-documented, rated as 1.
- wrote Elements;geometry foundations: This is well-documented, rated as 1.
- wrote Elements;mathematical synthesis: This is well-documented, rated as 1.
- wrote Optics;perspective: This is plausible, rated as 0.9.
- wrote Optics;conic sections: This is plausible, rated as 0.9.
- wrote Data;spherical geometry: This is plausible, rated as 0.9.
- wrote Data;number theory: This is plausible, rated as 0.9.
- wrote Phaenomena;mathematical rigour: This is plausible, rated as 0.9.
- wrote Phaenomena;lost works: This is plausible, rated as 0.8.
- authorship questioned;On Divisions of Figures: This is plausible, rated as 0.8.
- authorship questioned;Catoptrics: This is plausible, rated as 0.8.
- died;mathematical legacy: This is well-documented, rated as 1.

Overall Reliability Calculation:

To generalize the reliability of the network, we can take the average of all the connection reliabilities.

- Event-to-Event Connections: Average reliability = $(1 + 1 + 1 + 0.9 + 0.9 + 0.9 + 0.8 + 1) / 8 = 0.9375$
- Event-to-Concept Connections: Average reliability = $(1 + 0.9 + 0.9 + 1 + 1 + 1 + 1 + 0.9 + 0.9 + 0.9 + 0.9 + 0.9 + 0.8 + 0.8 + 1) / 15 = 0.92$

Overall Network Reliability:

The overall reliability of the network can be calculated as the average of the two averages:

- Overall Reliability: $(0.9375 + 0.92) / 2 = 0.92875$

Conclusion:

The overall reliability of the network is approximately 0.93, which indicates a high level of reliability. This means that the connections between events, concepts, and event-concept links in Euclid's biography are generally well-supported by historical evidence, with only minor uncertainties in some connections.

Conclusions

This paper presents a methodology and an information technology for detecting and verifying false facts in Wikipedia biographies, based on the use of large language models (LLMs) and semantic networking. The proposed approach enables the following:

1. Construction of a detailed semantic network for a biography, oriented towards the chronological sequence of events and the addition of relevant concepts.
2. Evaluation of the credibility of individual connections and the entire network as a whole using prompts.
3. Utilization of different large language models for independent assessment of the biography and formation of a "swarm of virtual experts."
4. Aggregation of the results from different models and providing an overall credibility assessment of the biography.

The scientific novelty of the work lies in the following:

1. The novel methodology for constructing a semantic network for biographies with two types of connections between events and concepts.
2. The first-time application of the "swarm of virtual experts" concept for verifying textual data in biographical articles.
3. The definition of digital criteria for assessing the credibility of both individual connections and the entire network.

4. The development of a credibility assessment methodology that combines LLMs and semantic networking.
5. The introduction of an integral evaluation model that accounts for the weight coefficients of temporal and associative connections.

Additionally, the proposed approach can be used for automated fact verification in open sources. The methodology enables the rapid identification of potentially manipulative data, which is especially relevant for journalism, historical research, and the fight against disinformation.

The uniqueness of the method also lies in the use of an aggregated credibility assessment that takes into account the results of various language models, ensuring higher accuracy and resilience to potential errors from individual models. This approach helps avoid biases that may arise due to the specific limitations of each model and provides a more reliable fact-checking process.

The proposed methodology demonstrates its effectiveness in detecting potential false facts, thereby increasing the credibility of information in open sources. It opens up opportunities for automating the verification of information in open sources, particularly Wikipedia, and can be adapted for analyzing other types of textual data.

References

- [1] Joshi, N., Spezzano, F., Green, M., & Hill, E. (2020, April). Detecting undisclosed paid editing in wikipedia. In *Proceedings of The Web Conference 2020* (pp. 2899-2905). DOI: 10.1145/3366423.3380055
- [2] Gao, Fan, Hang Jiang, Rui Yang, Qingcheng Zeng, Jinghui Lu, Moritz Blum, Tianwei She, Yuang Jiang, and Irene Li. "Evaluating large language models on wikipedia-style survey generation." In *Findings of the Association for Computational Linguistics ACL 2024*, pp. 5405-5418. 2024. DOI: 10.18653/v1/2024.findings-acl.321
- [3] Ding, N., Mayer, W., Geng, Y., Duan, Y., & Feng, Z. (2023, December). Generative Semantic Modeling for Structured Data Source with Large Language Model. In *2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)* (pp. 1148-1152). IEEE. DOI: 10.1109/HPCC-DSS-SmartCity-DependSys60770.2023.00164
- [4] Dmytro Lande, Leonard Strashnoy. *GPT Semantic Networking: A Dream of the Semantic Web - The Time is Now.* - Kyiv: Engineering, 2023. - 168 p. ISBN 978-966-2344-94-3
- [5] Wu, J., Guo, J., & Hooi, B. (2024, August). Fake News in Sheep's Clothing: Robust Fake News Detection Against LLM-Empowered Style Attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3367-3378). DOI: 10.1145/3637528.3671977
- [6] Xie, Bingbing, Xiaoxiao Ma, Xue Shan, Amin Beheshti, Jian Yang, Hao Fan, and Jia Wu. "Multiknowledge and LLM-Inspired Heterogeneous Graph Neural Network for Fake News Detection." *IEEE Transactions on Computational Social Systems* (2024). DOI: 10.1109/TCSS.2024.3488191
- [7] Lande Dmitry, Humeniuk Oleh. *Content Reconstruction: The Evolution of Texts through Semantic Networks and LLMs/ Available at SSRN Preprint, Sep. 09, 2024).* - 11 p. DOI: <http://dx.doi.org/10.2139/ssrn.4951516>

[8] Dmytro Lande, Leonard Strashnoy. Swarm of Virtual Experts in the implementation of Semantic Networking. ResearchGate Preprint, October 2024. DOI: 10.13140/RG.2.2.16686.11845

[9] Dmytro Lande, Leonard Strashnoy and Igor Svoboda. Application of the Virtual Experts Swarm Concept in Semantic Network Formation. International Scientific and Technical Conference "Intelligent Technologies of Linguistic Analysis": Conference Proceedings.. - Kyiv: NAU, 2024. - P. 24. URL: <http://dwl.kiev.ua/art/nau2024c/> (ukr)

[10] Feher Anatolii, Lande Dmytro. Defined AI semantic networking in cybersecurity // Intelligent Solutions-S: Proceedings of the International Symposium, September 28, 2023, Kyiv-Uzhorod, Ukraine. Kyiv: Publishing House "Caravela", 2023. pp. 21-22. ISBN 978-966-801-916-6