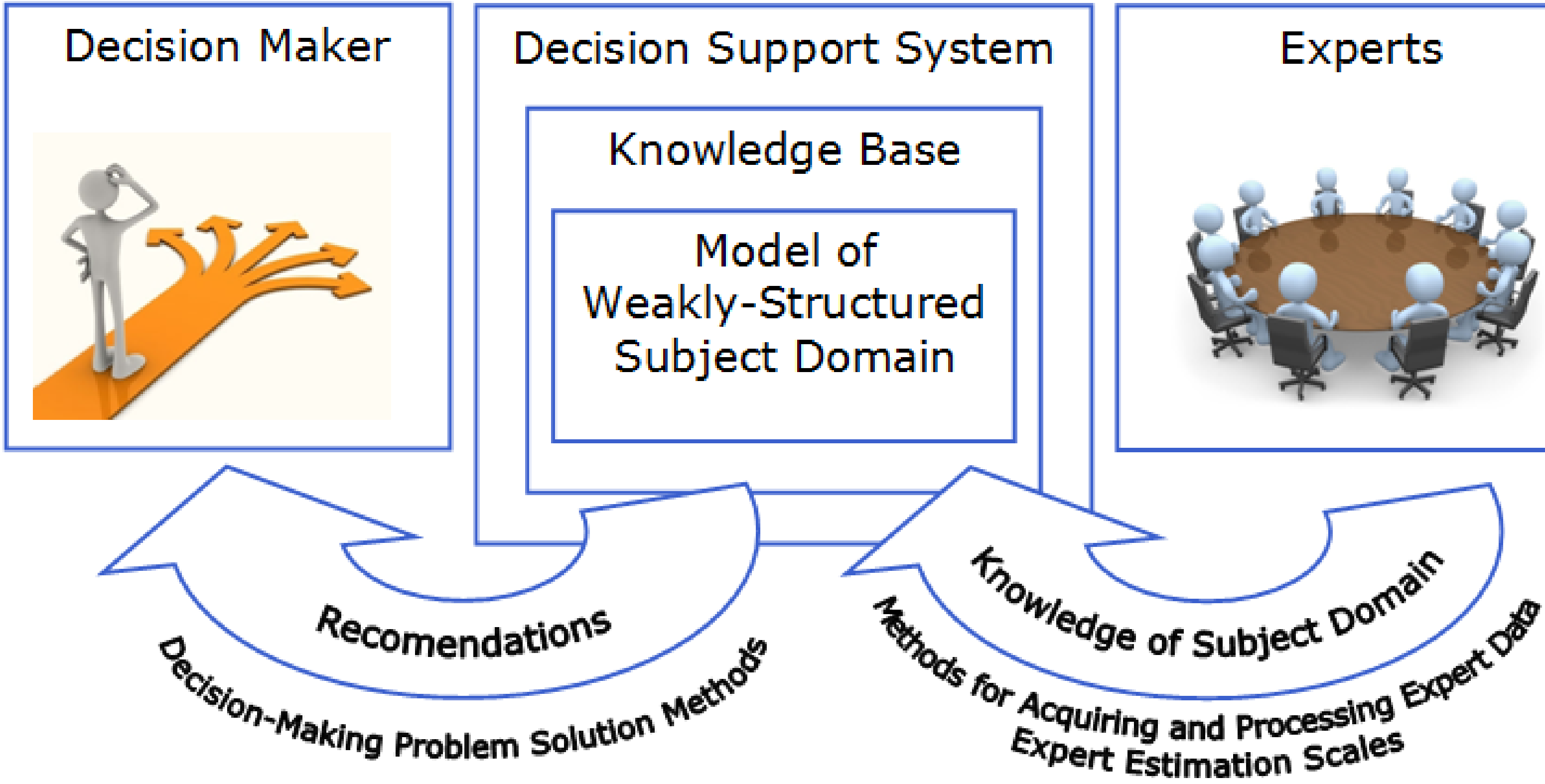
# Usage Artificial Intelligence Toolkit for Improving Translations to English

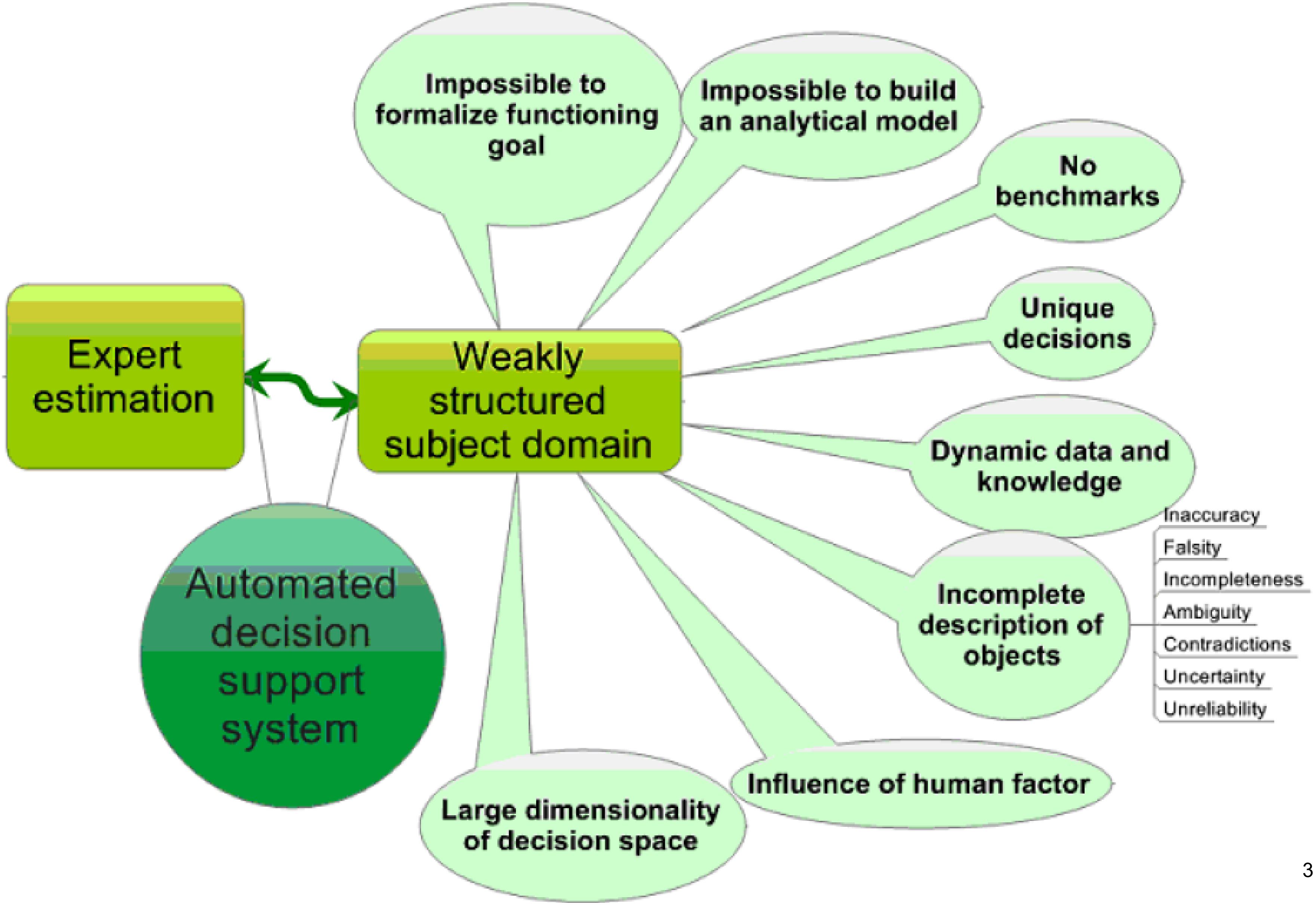
***Oleh Andriichuk***[1,2,3]***, SerhiiKadenko***[1]***, OlhaTsyhanok,***[4]
***OlenaBesklinska***[5] ***and OlenaTrubitsyna***[5]

[1] *Institute for Information Recording of the National Academy of Sciences of Ukraine*
[2] *Taras Shevchenko National University of Kyiv*
[3] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute“*
[4] *State University of Trade and Economics*
[5] *Kyiv National Linguistic University*

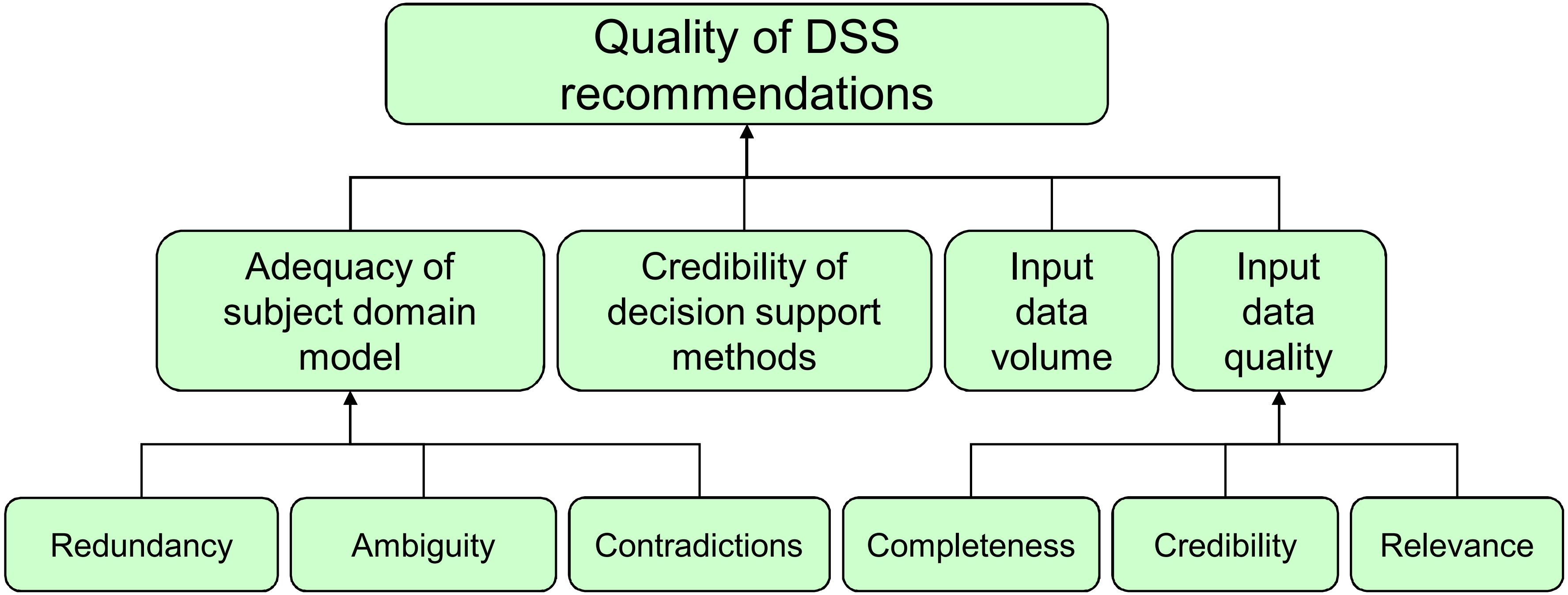
# Decision-making using expert DSS



# Weakly Structured Domains

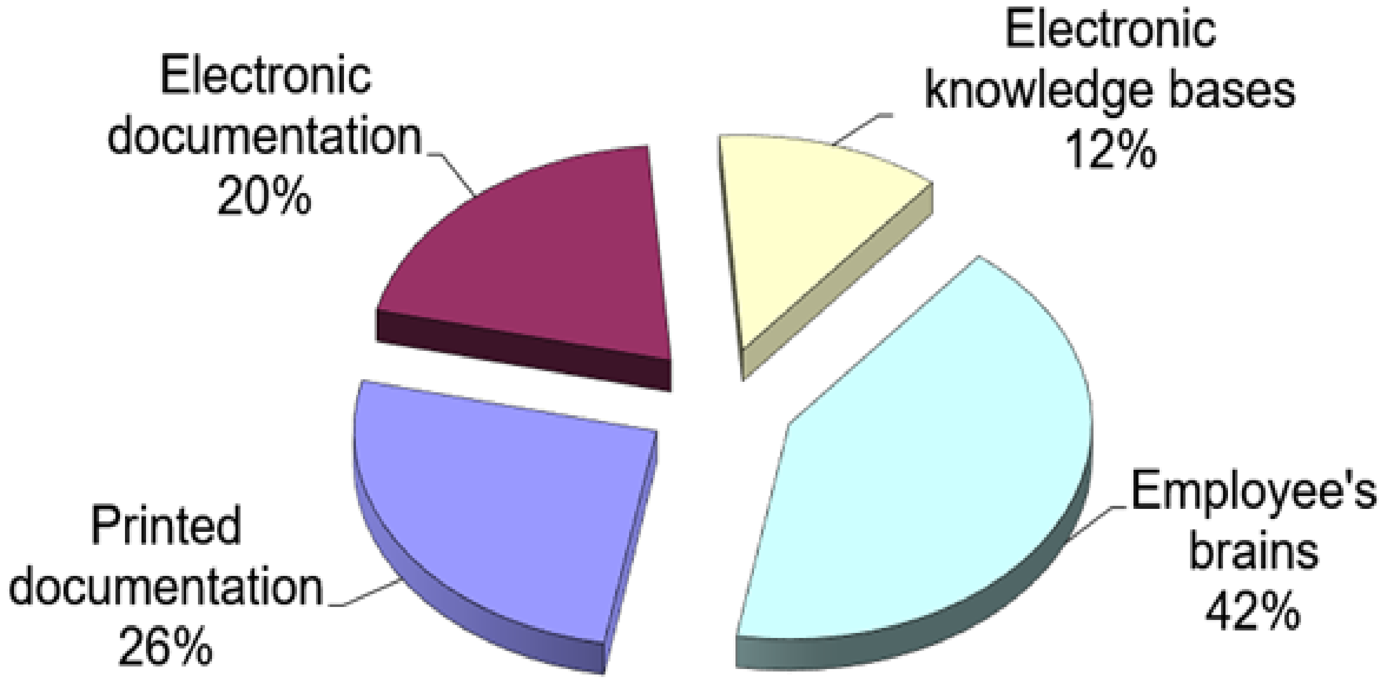


# Factors, that influence the quality of DSS recommendations



# The importance of using expert knowledge in decision-making

Delphi Group Research: Major Repository of Knowledge Organization



# Limitations of Galactica

galactica.org/mission/

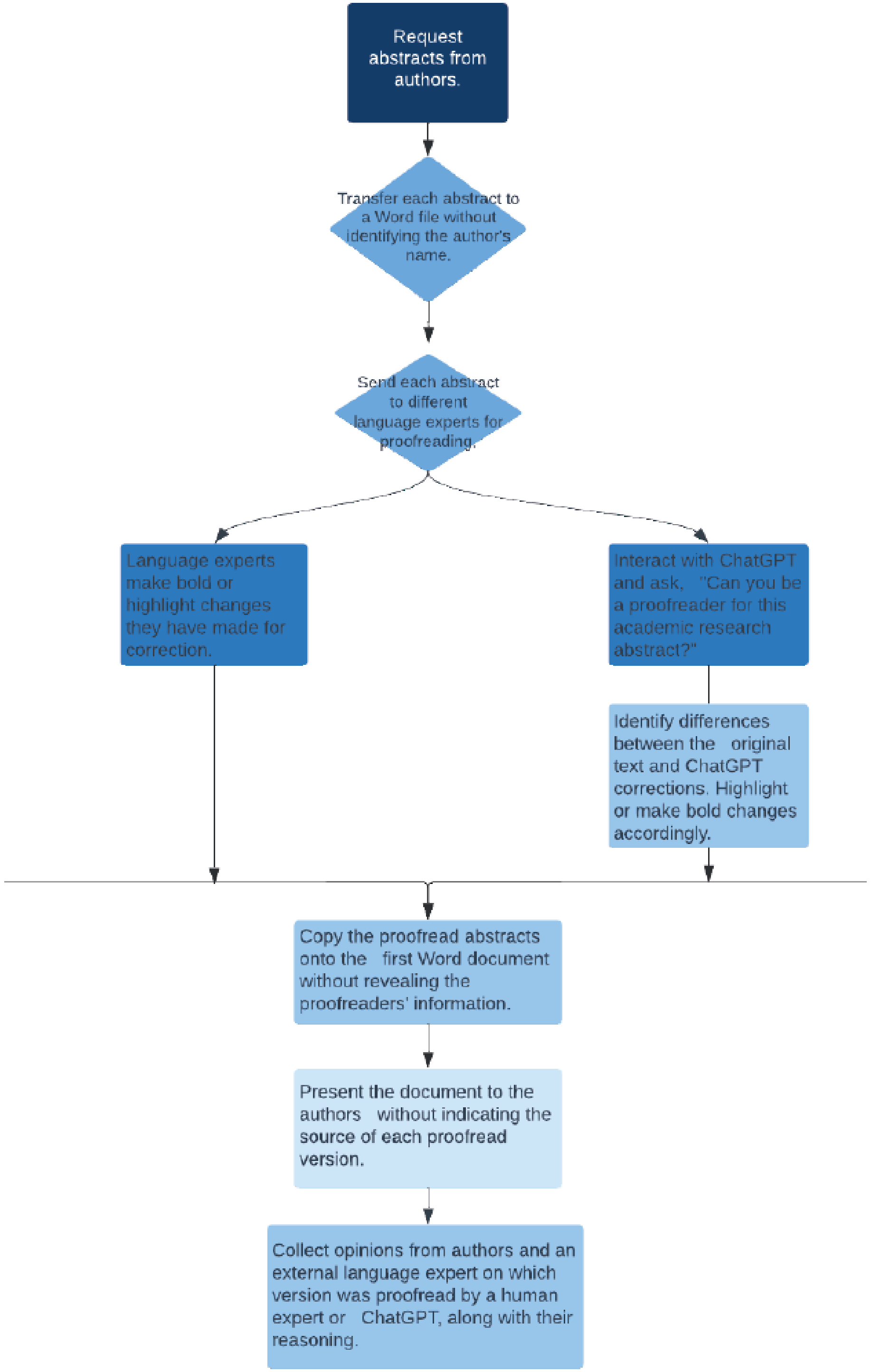
## Limitations

You should be aware of the following limitations when using the model (including the demo on this website):

- **Language Models can Hallucinate.** There are no guarantees for truthful or reliable output from language models, even large ones trained on high-quality data like Galactica. **NEVER FOLLOW ADVICE FROM A LANGUAGE MODEL WITHOUT VERIFICATION.**

- **Language Models are Frequency-Biased.** Galactica is good for generating content about well-cited concepts, but does less well for less-cited concepts and ideas, where hallucination is more likely.

- **Language Models are often Confident But Wrong.** Some of Galactica's generated text may appear very authentic and highly-confident, but might be subtly wrong in important ways. This is particularly the case for highly technical content.




# A COMPARATIVE ANALYSIS OF PROOFREADING CAPABILITIES: LANGUAGE EXPERTS VS. CHATGPT

***Nermin PUNAR ÖZÇELİK***

Department of Foreign Languages, Tarsus University

# Experiment results

| **Abstract Number** | | **1.** | | **2.** | | **3.** | | **4.** | | **5.** | | **6.** | | **7.** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Fields of Abstracts** | | ELT | | Logistics | | Engineering | | Logistics | | Banking | | Finance | | MIS | |
| **Proofread Numbers** | | 1. | 2. | 1. | 2. | 1. | 2. | 1. | 2. | 1. | 2. | 1. | 2. | 1. | 2. |
| The opinions of Writers and a language expert | **W1** | H | **AI** | H | **AI** | AI | H | H | **AI** | AI | H | AI | H | H | **AI** |
| | **W2** | AI | H | H | **AI** | AI | H | AI | H | AI | H | AI | H | H | **AI** |
| | **W3** | H | **AI** | H | **AI** | H | **AI** | AI | H | H | **AI** | AI | H | H | **AI** |
| | **W4** | AI | H | AI | H | H | **AI** | AI | H | H | **AI** | H | **AI** | H | **AI** |
| | **W5** | AI | H | AI | H | AI | H | AI | H | AI | H | AI | H | AI | H |
| | **W6** | H | **AI** | H | **AI** | AI | H | H | **AI** | H | **AI** | AI | H | H | **AI** |
| | **W7** | AI | H | H | **AI** | H | **AI** | H | **AI** | H | **AI** | H | **AI** | AI | H |
| | **E1** | AI | H | H | **AI** | AI | H | H | **AI** | AI | H | AI | H | H | **AI** |
| | | TOTAL: 56 points | | | | | | | | | | | | | |
| | | Correct AI Guesses: 28 points | | | | | | | | | | | | | |

***Abbreviations***:
ELT: English Language Teaching
MIS: Management Information Systems
W:Writer
E: Language Expert
H: Human Language Expert
AI: Artificial Intelligence

# Research Stages

1. Each of the non-native English-speaking respondents provided English-language scholarly texts in the given subject area. Most of these texts were self-translated into English from Ukrainian or possibly other languages. The study did not analyze or control the process of creating or originating the texts. Therefore, it is possible that machine translation tools may have been used, at least, in part. It is important to note that all respondents in this group are competent in the field chosen for this pilot study and the English-language texts they provided also belong to the same field.
2. AI tools (based on the GPT-3.5 architecture) were used to generate recommendations for improving the quality of the resulting English translations. Particularly, we used certain prompts to the AI software system: "I can give you below a new fragment of scientific text. Can you show me a few sentences with poor quality of English (orthography and grammatical correctness) and give step-by-step tips to improve them?"
3. The second group of expert validators assessed the reliability and quality of the original English translated texts and the recommendations for translation improvement provided by the AI tool. This group consisted of experts who were sufficiently competent in the subject area under review. Each expert in the group is proficient in English. The expert group could consult with scholars and professionals in the subject area, if needed.

# Examples of recommendations by AI tools

| Experts: | AI tools: |
| --- | --- |
| By utilizing a range of dynamic features, it is feasible to create various biometric vectors and decision-making algorithms, including statistical methods and neural networks. | Dynamic features enable the creation of various biometric vectors and decision-making algorithms, such as statistical methods and neural networks, for improved user authentication based on handwritten signatures. |
| It should be noted that the calculated estimates obtained will be higher when in systems used user`s own signature. | We should note that when users input their own signature into the system, the calculated estimates obtained are typically higher. |
| Processing of complex solutions with big number of connected sub-elements is complicated by the need to analyze a significant number of parameters of these objects contained into potential solutions, and the necessity of unification of their structure. | Processing complex solutions with a large number of connected sub-elements is complicated by the need to analyze a significant number of parameters, as well as the necessity of unifying their structure. |
| In this work, we consider the last subtask where all compared CIOs are based on a single ontology and have a similar structure. | In this work, we focus on the last subtask, which involves comparing CIOs that are based on a single ontology and have a similar structure. |

# Expert Questionnaire

How much better is the quality of the proposed AI recommendation (based on the GPT-3.5 architecture) than the original English-language text?

The answer options included:

- the AI recommendation is unacceptable;
- AI recommendation is acceptable, but equivalent in quality;
- the AI recommendation is acceptable and requires minor adjustments;
- the AI recommendation is acceptable and does not need to be adjusted.

# Statistical credibility of the research

Test experiment series:

| Name of the indicator | Quantity |
|---|---|
| total number of recommendations received | 358 |
| unacceptable recommendations | 139 |
| acceptable recommendations that are equivalent in quality to the original wording | 6 |
| acceptable recommendations that require minor adjustments | 30 |
| acceptable recommendations that do not require adjustment | 183 |

Frequencies: $\{139/358 \approx 0{,}388;\ 6/358 \approx 0{,}017;\ 30/358 \approx 0{,}084;\ 183/358 \approx 0{,}511\}$

# Minimum Number of Instances of the Experiment

$P_\beta = 0.95$

$\beta = 0.05$

$$n \geq \frac{p \cdot (1 - p)}{\beta^2} \left(F^{-1}\left(P_\beta\right)\right)^2$$

$$F^{-1}(0.95) \approx 1.96$$

$$\left(F^{-1}(0.95)\right)^2 \approx 3.84$$

$$n \geq \frac{0.511 \cdot (1 - 0.511)}{(0.05)^2} 3.84 = 383.814$$

# Experiment results

| Name of the indicator | Quantity |
|---|---|
| total number of recommendations received | 797 |
| unacceptable recommendations | 285 |
| acceptable recommendations that are equivalent in quality to the original wording | 10 |
| acceptable recommendations that require minor adjustments | 76 |
| acceptable recommendations that do not require adjustment | 426 |

53.5% of the recommendations do not require adjustments, 63% of recommendations that improve the quality of the source translation, 64.2% – of recommendations that at least do not worsen the translation, 1.3% of recommendations are useless (the quality of the translation remains approximately the same), a 35.8% of recommendations are harmful, as they will lead to a deterioration in the translation quality

# Summary

- As a result of the study, the originally developed methodology was used, which utilizes currently available AI tools to improve the quality of text translations into English. The experimental analysis showed that 53.5% of the AI tool recommendations for improving the translation did not require correction, 63% of them improved the translation, 64.2% did not worsen it, 1.3% of the recommendations were useless, and 35.8% were harmful.

- It is shown that the methodology used in the experimental study can be applied in practice to improve the quality of text translations into English. Furthermore, it is concluded that in the case of practical implementation, this technology is suitable only for automated use with the involvement of a group of experts. It is not advisable to entirely rely on the recommendations of AI tools, specifically the ChatGPT chatbot of the GPT-3.5 language model. Nevertheless, AI systems are useful, powerful, and effective tools for enhancing the quality of English translations.

Thank you for attention!