

# ADTA 5770: Generative AI with LLM

## Semester Project – Prompt Responses: Individual Evaluation Likert Scale Questionnaire

Thuan L Nguyen, Ph.D.

### 1. Overview

The Q&A Search system searches for a defined knowledge base of PDF documents using generative AI (Gemini 2.5 Pro, RAG, vector search) to answer user questions. Responses must be based solely on the knowledge base content. The response must cite the document title(s) and relevant paragraph(s) if information is found. If the system cannot find any documents containing contents related to the user's question, the response must clearly state this.

### 2. Likert Scale Questionnaire

#### 2.1 Criteria Summary

##### Scale Levels:

- **5 - Excellent:** Meets all criterion requirements exceptionally well.
- **4 - Good:** Meets the criterion requirements well, with only minor room for improvement
- **3 - Fair:** Partially meets the requirements with main adequate aspects, but notable missing exists.
- **2 - Poor:** Largely fails to meet the criterion requirements; significant flaws are present.
- **1 - Very Poor:** Completely fails to meet criterion requirements or acts counter to them

>>>>>>>>>> *Continue Next Page* >>>>>>>>>>

## 2.2 Likert Scale Questionnaire and Detailed Evaluation Criteria

### 2.2.1 Accuracy and Relevance of Response Content

**Evaluation Focus:** Is the core answer provided accurate and directly relevant to the user's question, based only on the information within the cited sources (if any)?

- **5 - Excellent:** The answer directly, accurately, and completely addresses the user's question using *only* information substantiated by the cited sources/paragraphs. Highly relevant.
- **4 - Good:** The answer accurately addresses the main point(s) of the question using substantiated information. May lack minor details or perfect conciseness but is clearly relevant and derived from sources.
- **3 - Fair:** The answer partially addresses the question or contains minor inaccuracies. It is mostly based on source material but might have slight deviations or miss key aspects. Moderately relevant.
- **2 - Poor:** The answer significantly misinterprets the question or provides largely inaccurate information *even if sources are cited*. Low relevance or poor synthesis of source material. May include uncited external information.
- **1 - Very Poor:** The answer is completely inaccurate, irrelevant to the question, or fabricated (hallucinated) despite any cited sources.

### 2.2.2 Source Document Identification

**Evaluation Focus:** If relevant information was found, does the response correctly and clearly identify the title(s) of the source document(s) from the knowledge base?

- **5 - Excellent:** All relevant source document titles are clearly and accurately listed. Easy to identify which documents contain the information.
- **4 - Good:** Most or all relevant document titles are listed accurately, perhaps with minor formatting issues or slight ambiguity.
- **3 - Fair:** Some relevant document titles are provided, but others are missing, inaccurate, or unclear.
- **2 - Poor:** Document titles are largely missing, incorrect, or presented in a confusing way.
- **1 - Very Poor:** No document titles are provided even when information was found and presented, OR titles provided are completely irrelevant/fabricated. (N/A if no information was found and the system correctly stated so).

>>>>>>>>>> *Continue Next Page* >>>>>>>>>>

### 2.2.3 Paragraph Citation Precision

**Evaluation Focus:** If relevant information was found, does the response accurately pinpoint the specific paragraph(s) or text chunk(s) within the cited document(s) where the answer can be verified?

- **5 - Excellent:** Specific, accurate paragraph(s) or precise text segments are cited for *all* key information in the answer, allowing easy verification.
- **4 - Good:** Paragraph/text citations are generally accurate and point to the correct location, possibly with minor inaccuracies (e.g., off by one paragraph, slightly too broad).
- **3 - Fair:** Some paragraph/text citations are provided but may be inaccurate, vague (e.g., citing whole sections instead of paragraphs), or missing for parts of the answer.
- **2 - Poor:** Paragraph/text citations are largely inaccurate, missing, or point to irrelevant document parts.
- **1 - Very Poor:** No paragraph/text citations are provided despite the information being presented, OR citations are completely fabricated/misleading. (N/A if no information was found and the system correctly stated so).

### 2.2.4 Adherence to Knowledge Base Scope and Handling of “Not Found” Cases

**Evaluation Focus:** Does the response rely exclusively on the knowledge base, and does it appropriately indicate when information cannot be found within that knowledge base?

- **5 - Excellent:** Strictly adheres to the knowledge base. If information is found, it's sourced correctly. If not found, the response clearly, explicitly, and accurately states that the information is unavailable *within the provided documents/knowledge base*. No external information is introduced.
- **4 - Good:** Adheres well to the knowledge base. The "not found" message (if applicable) is clear and accurate, perhaps slightly less explicit than "Excellent". Minimal to no indication of using external knowledge.
- **3 - Fair:** Mostly adheres to the knowledge base, but might contain hints of external information, or the "not found" message (if applicable) could be slightly ambiguous or missing context about *why* it wasn't found (i.e., limited to the KB).
- **2 - Poor:** Attempts to answer using external knowledge when the information is not in the KB *without stating it's external* OR fails to clearly state when information is missing from the KB, potentially implying it should be there.
- **1 - Very Poor:** Fabricates information claiming it's from the KB, heavily relies on external knowledge without acknowledgment, or completely fails to address the absence of information when it's not found in the KB.

>>>>>>>>>> *Continue Next Page* >>>>>>>>>>

## 2.2.5 Overall Clarity and Formatting

**Evaluation Focus:** Is the entire response (answer, citations, any "not found" statements) presented, logically structured, and easy for the user to understand?

- **5 - Excellent:** The response is exceptionally clear, well-organized, and easy to read. The answer and citations are logically connected and formatted for easy comprehension.
- **4 - Good:** The response is clear and understandable. The formatting is good, with only minor potential for confusion or improvement in structure.
- **3 - Fair:** The response is generally understandable but may require some effort to read due to awkward phrasing, moderate formatting issues, or slightly illogical structure.
- **2 - Poor:** The response is difficult to understand due to poor clarity, confusing language, significant formatting problems, or illogical organization.
- **1 - Very Poor:** The response is incoherent, nonsensical, or formatted so poorly that it is unusable.

## 3. Evaluating Responses Using Likert Scale Questionnaire

### TO DO:

- Review each question and the system's generated responses carefully.
- For each criterion below, select the rating score (1 – 5) that best describes the response's quality based on the provided definitions and record the scores in the Excel file using the template.
- Complete the evaluation through five Likert Scale criteria for each question.
- Enter the rating scores in the Excel file (a copy of the template).

### **IMPORTANT NOTES:**

--> *Evaluation scores must be recorded in an Excel file, a copy of the template:*

*“ADTA\_5770\_SEMESTER\_PROJECT\_prompts\_responses\_evaluation\_scores\_***template***.”*

--> *The submitted file must be re-named as follows:*

*ADTA\_5770\_SEMESTER\_PROJECT\_prompts\_responses\_evaluation\_scores\_***<student-name>**