

RAGAs-based assessment of AI agents performances in BPMN querying: SPARQL versus ChatGPT Retrieval Connector across TELeR prompt levels

RAGAs metric results for AI agent responses using SPARQL querying in Ontotext's GraphDB TTYG (TELeR level 2)

| Prompt | Ground truth | Generated answer | Response relevancy | Factual correctness | Semantic similarity | Faithfulness |
|--|--|---|--------------------|---------------------|---------------------|--------------|
| Is "Return undelivered product" a manual task? | Yes, "Return undelivered product" is a manual task. | Yes, "Return undelivered product" is a manual task. | 0,99 | 1 | 1 | 1 |
| What is the task type of "Settle with merchant"? | The task type of "Settle with merchant" is User. | The task type of "Settle with merchant" is User. | 0,99 | 1 | 1 | 1 |
| Is "Ship replacement" marked for compensation? | No, "Ship replacement" is not marked for compensation. | No, "Ship replacement" is not marked for compensation | 0,99 | 1 | 0,97 | 1 |
| Does "Cancel invoice" act as a compensation task? | Yes, "Cancel invoice" acts as a compensation task. | Yes, "Cancel invoice" acts as a compensation task. | 1 | 1 | 1 | 1 |
| Is there any task that has a defined cost? | Yes, there are multiple tasks that each have a defined cost. | Yes, there is at least one task in the process that has a defined cost. | 0,84 | 1 | 0,81 | 1 |
| Does "Issue invoice" have a defined execution time? | Yes, "Issue invoice" has a defined execution time. | Yes, "Issue invoice" has a defined execution time of 1 minute. | 0,96 | 0 | 0,93 | 1 |
| Is the "Customer still interested?" event conditional? | No, the "Customer still interested?" is not a conditional event. | No, the "Customer still interested?" event is not marked as conditional in the process model. | 0,86 | 1 | 0,86 | 1 |

RAGAs metric results for AI agent responses using ChatGPT Retrieval Connector querying in Ontotext's GraphDB TTYG (TELeR level 2)

| Prompt | Ground truth | Generated answer | Response relevancy | Factual correctness | Semantic similarity | Faithfulness |
|--|--|--|--------------------|---------------------|---------------------|--------------|
| Is "Return undelivered product" a manual task? | Yes, "Return undelivered product" is a manual task. | Yes, "Return undelivered product" is a manual task. | 0,99 | 1 | 1 | 1 |
| What is the task type of "Settle with merchant"? | The task type of "Settle with merchant" is User. | The task type of "Settle with merchant" is User. | 0,99 | 1 | 1 | 1 |
| Is "Ship replacement" marked for compensation? | No, "Ship replacement" is not marked for compensation. | No, the task "Ship replacement" is not marked for compensation. | 0,84 | 1 | 0,87 | 1 |
| Does "Cancel invoice" act as a compensation task? | Yes, "Cancel invoice" acts as a compensation task. | Yes, "Cancel invoice" acts as a compensation task. | 1 | 1 | 1 | 1 |
| Is there any task that has a defined cost? | Yes, there are multiple tasks that each have a defined cost. | Yes, there are several tasks that have a defined cost. For example: <ul style="list-style-type: none"> "Cancel invoice" has a cost of 0.1. "Ship replacement" has a cost of 6. "Return undelivered product" has a cost of 8. "Return product to courier" has a cost of 3.5. "Deliver parcel to customer" has a cost of 4. | 0,91 | 0,17 | 0,7 | 1 |
| Does "Issue invoice" have a defined execution time? | Yes, "Issue invoice" has a defined execution time. | Yes, the task "Issue invoice" has a defined execution time of 00:00:00:01:00 (1 minute). | 0,92 | 0 | 0,85 | 0,67 |
| Is the "Customer still interested?" event conditional? | No, the "Customer still interested?" is not a conditional event. | The "Customer still interested?" event is not a conditional event. It is modeled as an Exclusive Gateway of type "Data-based (without marker)", not as a conditional event. So, while it represents a decision point in the process, it is not technically a BPMN conditional event. | 0,85 | 0,6 | 0,75 | 1 |

RAGAs metric results for AI agent responses using SPARQL querying in TTYG (TELeR level 3)

| Prompt | Ground truth | Generated answer | Response relevancy | Factual correctness | Semantic similarity | Faithfulness |
|--|--|--|--------------------|---------------------|---------------------|--------------|
| Who is the participant that cancels the order? | The participant that cancels the order is the merchant. | The participant responsible for cancelling the order is the "Invoicing clerk". | 0,76 | 0 | 0,75 | 1 |
| How long does it take to transport the parcel to the client? | It takes 2 days to transport the parcel to the client. | It takes 2 days to transport the parcel to the client. | 1 | 1 | 1 | 1 |
| In case the filed claim is accepted, how long does it take to ship the replaced product? | In case the filed claim is accepted, it takes 30 minutes to ship the replaced product. | If the filed claim is accepted, it takes 30 minutes to ship the replaced product. | 0,89 | 1 | 0,98 | 1 |
| Who is responsible for processing returns? | The responsible for processing returns is the returns processor. | The person responsible for processing returns is the "Returns processor." | 1 | 1 | 0,94 | 1 |
| Can you list all tasks that are supported by software? | The tasks that are supported by software are: Log complaint Issue invoice Contact customer Settle with merchant Cancel invoice Cancel order Process return Check tracking information File claim with courier Cancel order & notify customer | The tasks that are supported by software (task type "Service") are: Cancel invoice Issue invoice | 0,69 | 0 | 0,74 | 1 |
| Which tasks rely on sending messages? | The tasks that rely on sending messages are "Notify merchant" and "Notify customer of new estimated time of arrival". | The tasks that rely on sending messages are: Notify merchant Notify customer of new estimated time of arrival | 1 | 1 | 0,94 | 1 |
| Is there a point in the process where the customer's interest determines the next step? If so, which one is it and what are the next steps? | Yes, there is a point in the process where the customer's interest determines the next step. This is the "Customer still interested?" exclusive gateway and the next steps are "Ship replacement" if the customer is interested or "Cancel order & notify customer" if the customer is not interested. | Yes, there is a point in the process where the customer's interest determines the next step. This occurs at the gateway labeled "Customer still interested?". The possible next steps from this gateway are: "Cancel order & notify customer" "Ship replacement" So, depending on whether the customer is still interested, the process will either proceed to cancel the order and notify the customer, or to ship a replacement. | 0,96 | 1 | 0,94 | 1 |
| Which user task has the highest execution cost and what is its value? Who is responsible for that particular task and what is the hourly wage for that responsible individual? | The user task with the highest execution cost is "File claim with courier" with a value of 5. The responsible individual for this particular task is the "Claims specialist" with an hourly wage of 18. | The user task with the highest execution cost is "File claim with courier" with a cost of 5. The responsible individual for this task is the "Claims specialist," whose hourly wage is 18. | 0,98 | 1 | 0,99 | 0,75 |
| What task is performed by the courier right before the parcel is delivered to the customer? | The task performed by the courier right before the parcel is delivered to the customer is "Contact customer". | The task performed by the courier right before delivering the parcel to the customer is reaching the decision point "Customer found?" (an exclusive gateway). This gateway determines whether the courier can proceed to deliver the parcel to the customer. | 0,94 | 0 | 0,69 | 1 |
| Right after issuing an invoice, the merchant cancels the invoice. Is this true? If not, correct the statement. | No, this is not true. Right after issuing an invoice, the merchant packs the order. | The statement is not true. Right after issuing an invoice, the next step is "Pack order," not "Cancel invoice." The correct sequence is: "Issue invoice" → "Pack order." | 0,47 | 1 | 0,73 | 1 |
| There are no tasks that occur after the order is marked as undelivered. | This statement is not true. There are four tasks for compensation that occur after the order is marked as undelivered: "Return undelivered product", "Return to warehouse", "Cancel invoice" and "Cancel order". | That statement is not correct. After the order is marked as undelivered (at the "Order undelivered" end event), there are still tasks that occur in the process model. Specifically, after "Order undelivered," the following tasks are executed: Return undelivered product Return to warehouse Cancel invoice These tasks ensure the undelivered product is handled and the invoice is canceled, even after the order is marked as undelivered. | 0,78 | 1 | 0,77 | 1 |
| After receiving the parcel, the client either uses the product or settles with the merchant. | This statement is partially correct. After receiving the parcel, the client uses the product and in parallel, the courier settles with the merchant. | Your statement is almost correct, but let's clarify the process flow based on the BPMN model: After the client receives the parcel ("Receive parcel - Cash on delivery"), the process continues to a non-exclusive gateway ("Non-exclusive Gateway (BPMN)-31871"). From this gateway, two parallel paths are possible: Use product: The client can proceed to use the product. Settle with merchant: The process can also proceed to settling with the merchant. These two actions are not mutually exclusive; both can occur in parallel or independently, according to the process model. Summary: After receiving the parcel, the client may use the product and/or settle with the merchant. Both actions are possible and can happen in parallel. | 0,52 | 0,5 | 0,67 | 1 |

RAGAs metric results for AI agent responses using ChatGPT Retrieval Connector querying in TTYG (TELeR level 3)

| Prompt | Ground truth | Generated answer | Response relevancy | Factual correctness | Semantic similarity | Faithfulness |
|--|--|--|--------------------|---------------------|---------------------|--------------|
| Who is the participant that cancels the order? | The participant that cancels the order is the merchant. | The participant that cancels the order is the Merchant. This is indicated by the task "Cancel order," which is executed by the participant Merchant and is under the responsibility of the Invoicing clerk. | 0,75 | 0,67 | 0,85 | 1 |
| How long does it take to transport the parcel to the client? | It takes 2 days to transport the parcel to the client. | The task "Transport parcel" has a defined execution time of 00:002:00:00:00, which means it takes 2 days to transport the parcel to the client. | 0,86 | 0,5 | 0,65 | 0,5 |
| In case the filed claim is accepted, how long does it take to ship the replaced product? | In case the filed claim is accepted, it takes 30 minutes to ship the replaced product. | If the filed claim is accepted, the task to "Ship replacement" has a defined execution time of 00:000:00:30:00, which means it takes 30 minutes to ship the replaced product. | 0,91 | 1 | 0,81 | 0,5 |
| Who is responsible for processing returns? | The responsible for processing returns is the returns processor. | The person responsible for processing returns is the Returns Processor. This role is part of the organizational unit "Logistics" within the Merchant Organization Structure. The Returns Processor is specifically mentioned as being under the responsibility for the "Process return" task. | 0,78 | 0,33 | 0,83 | 1 |
| Can you list all tasks that are supported by software? | The tasks that are supported by software are: Log complaint Issue invoice Contact customer Settle with merchant Cancel invoice Cancel order Process return Check tracking information File claim with courier Cancel order & notify customer | Based on the available data, the following tasks are supported by software (i.e., are of type "Service" or otherwise automated): 1. Issue invoice – Task type: Service 2. Cancel invoice – Task type: Service These tasks are explicitly marked as "Service" tasks, which indicates they are supported by software. Other tasks listed are manual or user tasks, which require human intervention. | 0,82 | 0,17 | 0,71 | 0,33 |
| Which tasks rely on sending messages? | The tasks that rely on sending messages are "Notify merchant" and "Notify customer of new estimated time of arrival". | The following tasks rely on sending messages: 1. Notify merchant Task type: Send Executed by: Customer 2. Notify customer of new estimated time of arrival Task type: Send Executed by: Customer service agent These tasks are explicitly of the "Send" type, indicating they involve sending messages as part of their execution. | 0,86 | 0,29 | 0,82 | 0,5 |
| Is there a point in the process where the customer's interest determines the next step? If so, which one is it and what are the next steps? | Yes, there is a point in the process where the customer's interest determines the next step. This is the "Customer still interested?" exclusive gateway and the next steps are "Ship replacement" if the customer is interested or "Cancel order & notify customer" if the customer is not interested. | Yes, there is a point in the process where the customer's interest determines the next step. This is at the step "Customer still interested?" If the customer is still interested, the process leads to "Ship replacement." If not, it leads to "Cancel order & notify customer." | 0,96 | 1 | 0,91 | 1 |
| Which user task has the highest execution cost and what is its value? Who is responsible for that particular task and what is the hourly wage for that responsible individual? | The user task with the highest execution cost is "File claim with courier" with a value of 5. The responsible individual for this particular task is the "Claims specialist" with an hourly wage of 18. | The user task with the highest execution cost is "File claim with courier," which has a cost of 5. The person responsible for this task is the Claims specialist, whose hourly wage is 18. | 0,97 | 1 | 0,98 | 1 |
| What task is performed by the courier right before the parcel is delivered to the customer? | The task performed by the courier right before the parcel is delivered to the customer is "Contact customer". | The task performed by the courier right before "Deliver parcel to customer" is the decision point "Customer found?". This is an exclusive gateway where the courier determines if the customer is available to receive the parcel. If the customer is found, the process proceeds to "Deliver parcel to customer." | 0,93 | 0 | 0,71 | 0,75 |
| Right after issuing an invoice, the merchant cancels the invoice. Is this true? If not, correct the statement. | No, this is not true. Right after issuing an invoice, the merchant packs the order. | The statement is not true. After the order is marked as undelivered, there is a subsequent task: "Return undelivered product." This task is executed by the courier. Therefore, tasks do occur after the order is marked as undelivered. | 0,34 | 0,25 | 0,55 | 0,5 |
| There are no tasks that occur after the order is marked as undelivered. | This statement is not true. There are four tasks for compensation that occur after the order is marked as undelivered: "Return undelivered product", "Return to warehouse", "Cancel invoice" and "Cancel order". | You are correct. There are no tasks (BPMN) that occur after the order is marked as "Order undelivered." The process element that follows is an End Event ("Order undelivered"), which signifies the termination of that process path. | 0,62 | 0 | 0,52 | 1 |
| After receiving the parcel, the client either uses the product or settles with the merchant. | This statement is partially correct. After receiving the parcel, the client uses the product and in parallel, the courier settles with the merchant. | The statement is not accurate. After receiving the parcel ("Receive parcel - Cash on delivery"), the next step is a non-exclusive gateway, which means multiple paths may follow. However, the available context does not confirm that the client either uses the product or settles with the merchant as the immediate next steps. The process may involve other actions or decisions after receiving the parcel. | 0 | 0 | 0,68 | 1 |

RAGAs metric results for AI agent responses using SPARQL querying in TTYG (TELeR level 4)

| Prompt | Ground truth | Generated answer | Response relevancy | Factual correctness | Semantic similarity | Faithfulness |
|---|---|---|--------------------|---------------------|---------------------|--------------|
| List all user tasks from the subprocess that investigates lost orders, that are under the responsibility of the Customer service agent and have an execution time of less than 5 minutes. For each, specify the subsequent task. A good output should be con-cise, of less than 100 words. | The user tasks from the subprocess that investigates lost orders, that is under the responsibility of the Customer service agent and has an execution time of less than 5 minutes is "Check tracking information". The subsequent task is either "Notify customer of new estimated time of arrival" or "File claim with courier". | The only user task in the "Lost order investigation subprocess" under the responsibility of the Customer service agent with an execution time of less than 5 minutes is "Check tracking information". Its subsequent task is "Package located?". | 0,73 | 0,5 | 0,88 | 1 |
| Some processes may contain gateways with no outgoing sequence flows. List these gateways and for each, indicate whether this is likely a modeling error or an intentional dead-end. A good output should be concise, fitting in a non-bulleted style paragraph of less than 200 words. | The processes do not contain gateways with no outgoing sequence flows. | All gateways in the provided models have at least one outgoing sequence flow; none are found without outgoing flows. Therefore, there are no gateways that represent either a modeling error or an intentional dead-end in this context. If such gateways existed, they would typically indicate a modeling error, as gateways are meant to direct process flow, and having no outgoing sequence flows would leave the process stranded. However, since none are present, this issue does not arise in your models. | 0,7 | 0,5 | 0,76 | 0,62 |
| For each participant, return: (1) the number of tasks they perform; (2) the types of events belonging to each participant's process flow; (3) the name of any decision points (exclusive gateways) within each participant's process scope. A good output should: be concise, of less than 150 words, and use a numbered list format. | <p>1. Customer:</p> <p>(1) The number of performed tasks: 10</p> <p>(2) The types of events belonging to the Customer's process flow: Start, End, Intermediate, Boundary</p> <p>(3) The name of any decision points (exclusive gateways) within each participant's process scope: none.</p> <p>2. Merchant:</p> <p>(1) The number of performed tasks: 12</p> <p>(2) The types of events belonging to the Customer's process flow: Start, End, Intermediate, Boundary</p> <p>(3) The name of any decision points (exclusive gateways) within each participant's process scope: none.</p> <p>3. Courier:</p> <p>(1) The number of performed tasks: 6</p> <p>(2) The types of events belonging to the Customer's process flow: End, Boundary</p> <p>(3) The name of any decision points (exclusive gateways) within each participant's process scope: Customer found?.</p> | <p>1. Customer</p> <p>Tasks performed: 6</p> <p>Event types: Start Event, End Event, Intermediate Event</p> <p>Decision points: None</p> <p>2. Courier</p> <p>Tasks performed: 6</p> <p>Event types: End Event, Intermediate Event</p> <p>Decision points: Customer found?</p> <p>3. Merchant</p> <p>Tasks performed: 7</p> <p>Event types: End Event, Intermediate Event</p> <p>Decision points: None</p> | 0,44 | 0,44 | 0,86 | 0,89 |

RAGAs metric results for AI agent responses using ChatGPT Retrieval Connector querying in TTYG (TELeR level 4)

| Prompt | Ground truth | Generated answer | Response relevancy | Factual correctness | Semantic similarity | Faithfulness |
|---|---|---|--------------------|---------------------|---------------------|--------------|
| List all user tasks from the subprocess that investigates lost orders, that are under the responsibility of the Customer service agent and have an execution time of less than 5 minutes. For each, specify the subsequent task. A good output should be con-cise, of less than 100 words. | The user tasks from the subprocess that investigates lost orders, that is under the responsibility of the Customer service agent and has an execution time of less than 5 minutes is "Check tracking information". The subsequent task is either "Notify customer of new estimated time of arrival" or "File claim with courier". | In the lost order investigation subprocess, the Customer service agent performs two user tasks with execution times under five minutes: "Check tracking information" (3 minutes), which is followed by "Package located?," and "Notify customer of new estimated time of arrival" (2 minutes), which is followed by "Investigation closed." | 0,7 | 0 | 0,86 | 0,75 |
| Some processes may contain gateways with no outgoing sequence flows. List these gateways and for each, indicate whether this is likely a modeling error or an intentional dead-end. A good output should be concise, fitting in a non-bulleted style paragraph of less than 200 words. | The processes do not contain gateways with no outgoing sequence flows. | The gateways "Package located?", "Customer still interested?", "Claim accepted?", and "Exclusive Gateway (BPMN)-30689" are listed as having no outgoing sequence flows. In standard BPMN modeling, gateways without outgoing flows typically indicate a modeling error, as gateways are intended to direct process flow based on conditions or events. Unless these gateways are deliberately used as placeholders for incomplete models or to mark intentional process termination points (which is rare and not best practice), their lack of outgoing flows is most likely unintentional and should be reviewed for correction. There is no evidence in the retrieved data suggesting these are intentional dead-ends, so they are best interpreted as modeling oversights. | 0,59 | 0 | 0,68 | 0 |
| For each participant, return: (1) the number of tasks they perform; (2) the types of events belonging to each participant's process flow; (3) the name of any decision points (exclusive gateways) within each participant's process scope. A good output should: be concise, of less than 150 words, and use a numbered list format. | <p>1. Customer:</p> <p>(1) The number of performed tasks: 10</p> <p>(2) The types of events belonging to the Customer's process flow: Start, End, Intermediate, Boundary</p> <p>(3) The name of any decision points (exclusive gateways) within each participant's process scope: none.</p> <p>2. Merchant:</p> <p>(1) The number of performed tasks: 12</p> <p>(2) The types of events belonging to the Customer's process flow: Start, End, Intermediate, Boundary</p> <p>(3) The name of any decision points (exclusive gateways) within each participant's process scope: none.</p> <p>3. Courier:</p> <p>(1) The number of performed tasks: 6</p> <p>(2) The types of events belonging to the Customer's process flow: End, Boundary</p> <p>(3) The name of any decision points (exclusive gateways) within each participant's process scope: Customer found?.</p> | <p>1. Customer:</p> <p>Tasks: At least 1 (e.g., "Search product at various merchants").</p> <p>Event types: Conditional event (e.g., "Product problems"), Compensation event.</p> <p>Exclusive gateways: None named within scope.</p> <p>2. Courier:</p> <p>Tasks: At least 3 ("Deliver return", "Return undelivered product", "Contact customer").</p> <p>Event types: Compensation event, Manual task.</p> <p>Exclusive gateways: "Customer found?" (Data-based).</p> <p>3. Merchant:</p> <p>Tasks: At least 1 ("Process return").</p> <p>Event types: Compensation event.</p> <p>Exclusive gateways: None named within scope.</p> <p>Note: Some exclusive gateways (e.g., "Customer found?") are data-based decision points. Other gateways and events may exist, but only those explicitly linked to participants are listed.</p> | 0,49 | 0,38 | 0,8 | 0,91 |