Situation Aware eXplainability (SAX) - Quality of Explanations [G2]

Jan 9, 2024

We welcome you in our study - Group 2

* Required

Consent for study participation

Informed Consent to Participate in Research

Study title: Situation Aware eXplainability – quality of explanations.

Researcher[s]: Prof. Dirk Fahland (TU/e), Dr. Fabiana Fournier and Dr. Lior Limonad (IBM Research).

We're inviting you to take a survey for research. This survey is completely voluntary. There are no negative consequences if you don't want to take it. If you start the survey, you can always change your mind and stop at any time.

What is the purpose of this study?
The purpose it to assess the quality of textual explanations about business processes that are automatically generated by Large Language Models (LLMs).

What will I do?
The survey presents different explanations in a form of textual narratives and asks the participants to rate each on a series of qualitative statements. The survey will take about 45 minutes.

Risks
We're minimizing any potential risk concerning breach of confidentiality in the following ways:

- Data is anonymous.
- We'll store all electronic data on a password-protected, encrypted computer.

Possible benefits: Helping understand more about what makes a good explanation in business processes.

Estimated number of participants: 50

Costs: None
Compensation: None

Future research: De-identified data (all identifying information removed) may be shared with other researchers.

Funding source: EU Horizon Europe

Where will data be stored? On the servers for the online survey software (Microsoft Forms).

How long will the data be kept? 1 month

Who can see my data?

- We (the researchers) will have access to de-identified (no names, birthdate, address, etc.) data. This is so we can analyze the data and conduct the study.
- We may share our findings in publications or presentations. If we do, the results will be aggregate (grouped) data, with no individual results.

Questions about your rights as a research participant, complaints, or problems:
Contact Prof. Dirk Fahland / d.fahland@tue.nl

1. I hereby give my consent to participate in this study, agreeing to the above conditions. *

- ◯ Yes
- ◯ No

General information

2. Gender *

- ◯ Woman
- ◯ Man
- ◯ Non-binary
- ◯ Prefer not to say

3. Age *

- ◯ Below 25
- ◯ 25-35
- ◯ 36-45
- ◯ Above 45

Introduction

We appreciate your agreement to participate in our study. The study investigates the goodness of textual explanations that are generated automatically with Large-Language-Models (LLM), such as chatGPT. An explanation is a statement or account that gives the reason or a justification for the occurrence of a certain condition or situation in a given domain. For example, an explanation may attempt to answer "why was flight departure late?" (i.e., the condition), for which the explanation text may be the statement: "flight departure was late because of a heavy storm weather".

In this survey, you will be presented with three different explanations (textual narratives) corresponding to three problem domains. Each explanation will correspond to a particular question, on which we will ask you to rate its quality on a variety of rating scales. For your convenience, we also provide the "ground" explanation corresponding to each question. Each section consists of 24 ratings and 72 in total.

Pizza case (1 of 2)

This domain is about a pizza take-away process. In the process of pizza making, once a pizza order is placed, dough is kneaded, and sauce and toppings are fetched. Next, pizza is assembled with all ingredients, and then baked. Once baking is complete, pizza is boxed and sent out for delivery. If delivery takes more than 30 min after pizza was boxed, pizza delivery is considered late and the customer is compensated.

In this domain, the pizza store manager is interested in the question: **"Why are pizzas late?"** (condition)

Analysis of the process revealed the following **ground truth answer**:

Pizzas are late primarily because of the kneading of the dough and dough weight, while the fetching of the toppings and their weight do not affect lateness.

Presenting the same question to an LLM, the following explanation was generated:

"The reason for pizzas being late is primarily due to the weights of the toppings and dough. The 'toppings weight' in the 'Fetch topping' step and the 'dough weight' in the 'Knead dough' step are significant factors contributing to the lateness of pizzas."

---

4. Note that the LLM explanation may agree or disagree with the ground truth answer and differ in detail and extent. Please rate the following statements considering the quality of the LLM explanation above with respect to the ground truth answer: *

	Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
The explanation is short.	◯	◯	◯	◯	◯	◯	◯
The explanation can only have one interpretation .	◯	◯	◯	◯	◯	◯	◯
I did not need any clarifications to interpret the explanation.	◯	◯	◯	◯	◯	◯	◯
The explanation provides all the details about the condition.	◯	◯	◯	◯	◯	◯	◯
The explanation is complete in covering relevant information.	◯	◯	◯	◯	◯	◯	◯
The explanation encompasses all aspects and outcomes related to the condition.	◯	◯	◯	◯	◯	◯	◯
I have confidence in the explanations generated by the LLM tool.	◯	◯	◯	◯	◯	◯	◯
The sequence of events leading to the condition is clearly outlined.	◯	◯	◯	◯	◯	◯	◯
I feel secure in relying on the LLM tool for correct explanations.	◯	◯	◯	◯	◯	◯	◯
The explanation is presented in an unambiguous and straightforwa rd manner.	◯	◯	◯	◯	◯	◯	◯
The explanation is easy to understand.	◯	◯	◯	◯	◯	◯	◯
I am curious to know what explains the condition.	◯	◯	◯	◯	◯	◯	◯

Pizza case (2 of 2)

This domain is about a pizza take-away process. In the process of pizza making, once a pizza order is placed, dough is kneaded, and sauce and toppings are fetched. Next, pizza is assembled with all ingredients, and then baked. Once baking is complete, pizza is boxed and sent out for delivery. If delivery takes more than 30 min after pizza was boxed, pizza delivery is considered late and the customer is compensated.

In this domain, the pizza store manager is interested in the question: **"Why are pizzas late?"** (condition)

Analysis of the process revealed the following **ground truth answer**:

Pizzas are late primarily because of the kneading of the dough and dough weight, while the fetching of the toppings and their weight do not affect lateness.

Presenting the same question to an LLM, the following explanation was generated:

"The reason for pizzas being late is primarily due to the weights of the toppings and dough. The 'toppings weight' in the 'Fetch topping' step and the 'dough weight' in the 'Knead dough' step are significant factors contributing to the lateness of pizzas."

---

5. Note that the LLM explanation may agree or disagree with the ground truth answer and differ in detail and extent. Please rate the following statements considering the quality of the LLM explanation above with respect to the ground truth answer: *

	Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
The explanation encloses what causes the condition.	◯	◯	◯	◯	◯	◯	◯
I find the LLM tool to be highly reliable and accurate consistently.	◯	◯	◯	◯	◯	◯	◯
The explanation accurately reflects the properties of the condition.	◯	◯	◯	◯	◯	◯	◯
The key elements of the condition are succinctly presented.	◯	◯	◯	◯	◯	◯	◯
The explanation is about why the condition occurred.	◯	◯	◯	◯	◯	◯	◯
I am eager to understand why the condition occurred.	◯	◯	◯	◯	◯	◯	◯
The explanation does not include irrelevant information.	◯	◯	◯	◯	◯	◯	◯
The explanation is consistent with the process.	◯	◯	◯	◯	◯	◯	◯
The explanation avoids assumptions not backed by data or process evidence.	◯	◯	◯	◯	◯	◯	◯
The explanation aligns with my perception on how the condition occurred.	◯	◯	◯	◯	◯	◯	◯
I am intrigued to find out why some other condition did not occur instead.	◯	◯	◯	◯	◯	◯	◯
The explanation matches my							

matches my level of expertise and knowledge of the process.	◯	◯	◯	◯	◯	◯	◯

Parking fines (1 of 2)

This domain is about the processing of parking fines for cars that are parked at hazardous locations (e.g., the vehicle is parked on a sidewalk or on a crosswalk). In this process, when a parking inspector stumbles upon a suspiciously parked vehicle, they first seek for a disabled parking permit. If such a permit doesn't exist, they observe the parking to determine if it is hazardous. If so, an extended fine is filled out while also calling for a tow truck to move the car to eliminate the risk.

In this domain, the parking enforcement department is interested in the question: **"Why does the processing of fines for cars that are parked within hazardous locations take so long?"** (condition)

Analysis of the process revealed the following **ground truth answer**:

The processing of fines for cars that are parked within hazardous locations takes long time mainly due to the calling for a tow company. This is since the calling for a tow company is a bottleneck in the process, while the time devoted to the submission of an extended fine has no effect on the overall processing time.

Presenting the same question to an LLM, the following explanation was generated:

"The briefest explanation for fine processing lateness is the activity 'check if hazardous parking'. This is a pivotal causal activity leading to 'END event', and it is directly associated with the attribute 'region in city' (with a significant importance value of 0.1), suggesting that the region's characteristics influence the decision-making time. Furthermore, 'check if hazardous parking' causally leads to 'call a tow truck', which is the immediate predecessor of 'END event'. The 'call a tow truck' activity also includes the attribute 'choice of towing company' with an importance value of 0.6, indicating its role in the lateness. These attributes associated with causal activities suggest that the choice of actions during these steps, influenced by region-specific factors and towing company selection, are critical to the lateness in the fine processing timeline."

---

6. Note that the LLM explanation may agree or disagree with the ground truth answer and differ in detail and extent. Please rate the following statements considering the quality of the LLM explanation above with respect to the ground truth answer: *

	Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
The explanation is short.	◯	◯	◯	◯	◯	◯	◯
The explanation can only have one interpretation .	◯	◯	◯	◯	◯	◯	◯
I did not need any clarifications to interpret the explanation.	◯	◯	◯	◯	◯	◯	◯
The explanation provides all the details about the condition.	◯	◯	◯	◯	◯	◯	◯
The explanation is complete in covering relevant information.	◯	◯	◯	◯	◯	◯	◯
The explanation encompasses all aspects and outcomes related to the condition.	◯	◯	◯	◯	◯	◯	◯
I have confidence in the explanations generated by the LLM tool.	◯	◯	◯	◯	◯	◯	◯
The sequence of events leading to the condition is clearly outlined.	◯	◯	◯	◯	◯	◯	◯
I feel secure in relying on the LLM tool for correct explanations.	◯	◯	◯	◯	◯	◯	◯
The explanation is presented in an unambiguous and straightforwa rd manner.	◯	◯	◯	◯	◯	◯	◯
The explanation is easy to understand.	◯	◯	◯	◯	◯	◯	◯
I am curious to know what explains the condition.	◯	◯	◯	◯	◯	◯	◯

Parking fines (2 of 2)

This domain is about the processing of parking fines for cars that are parked at hazardous locations (e.g., the vehicle is parked on a sidewalk or on a crosswalk). In this process, when a parking inspector stumbles upon a suspiciously parked vehicle, they first seek for a disabled parking permit. If such a permit doesn't exist, they observe the parking to determine if it is hazardous. If so, an extended fine is filled out while also calling for a tow truck to move the car to eliminate the risk.

In this domain, the parking enforcement department is interested in the question: "**Why does the processing of fines for cars that are parked within hazardous locations take so long?"** (condition)

Analysis of the process revealed the following **ground truth answer**:

The processing of fines for cars that are parked within hazardous locations takes long time mainly due to the calling for a tow company. This is since the calling for a tow company is a bottleneck in the process, while the time devoted to the submission of an extended fine has no effect on the overall processing time.

Presenting the same question to an LLM, the following explanation was generated:

"The briefest explanation for fine processing lateness is the activity 'check if hazardous parking'. This is a pivotal causal activity leading to 'END event', and it is directly associated with the attribute 'region in city' (with a significant importance value of 0.1), suggesting that the region's characteristics influence the decision-making time. Furthermore, 'check if hazardous parking' causally leads to 'call a tow truck', which is the immediate predecessor of 'END event'. The 'call a tow truck' activity also includes the attribute 'choice of towing company' with an importance value of 0.6, indicating its role in the lateness. These attributes associated with causal activities suggest that the choice of actions during these steps, influenced by region-specific factors and towing company selection, are critical to the lateness in the fine processing timeline."

---

7. Note that the LLM explanation may agree or disagree with the ground truth answer and differ in detail and extent. Please rate the following statements considering the quality of the LLM explanation above with respect to the ground truth answer: *

	Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
The explanation encloses what causes the condition.	◯	◯	◯	◯	◯	◯	◯
I find the LLM tool to be highly reliable and accurate consistently.	◯	◯	◯	◯	◯	◯	◯
The explanation accurately reflects the properties of the condition.	◯	◯	◯	◯	◯	◯	◯
The key elements of the condition are succinctly presented.	◯	◯	◯	◯	◯	◯	◯
The explanation is about why the condition occurred.	◯	◯	◯	◯	◯	◯	◯
I am eager to understand why the condition occurred.	◯	◯	◯	◯	◯	◯	◯
The explanation does not include irrelevant information.	◯	◯	◯	◯	◯	◯	◯
The explanation is consistent with the process.	◯	◯	◯	◯	◯	◯	◯
The explanation avoids assumptions not backed by data or process evidence.	◯	◯	◯	◯	◯	◯	◯
The explanation aligns with my perception on how the condition occurred.	◯	◯	◯	◯	◯	◯	◯
I am intrigued to find out why some other condition did not occur instead.	◯	◯	◯	◯	◯	◯	◯
The explanation matches my							

matches my level of expertise and knowledge of the process.	◯	◯	◯	◯	◯	◯	◯

Loan approval and offer issuance (1 of 2)

This domain is about the processing of loan applications. In the process, once loan application was placed, it is reviewed for acceptance. If accepted, a loan offer is prepared to match the details of the request. Once created, the offer is sent to the client via mail (or online if customer is connected). Within 30 days post offer issuance, the offer is returned being accepted or refused. At any point during this period, the customer may cancel their application.

In this domain, the loan issuing bank is interested in the question: **"What is the reason for loan offer lateness post customer application?"** (condition)

Analysis of the process revealed the following **ground truth answer**:

Reasons for lateness may reflect bottlenecks in the process after an offer is sent and till the point it is returned or canceled by the applicant. The preceding part of the process during which the application is accepted and an offer is created is not causing any delay in the time it takes to receive a loan offer.

Presenting the same question to an LLM, the following explanation was generated:

"The reason for loan offer lateness could be attributed to the multiple steps involved in the process after the initial acceptance. Specifically, the stages from 'Application_Accepted' through various steps like 'Offer_Create', 'Offer_Created', 'Offer_Sent (mail and online)', 'Offer_Returned', to finally reaching 'Offer_Accepted' indicate multiple transitions and activities. Each of these steps potentially adds to the overall time taken to complete the loan offer process, contributing to any lateness observed."

---

8. Note that the LLM explanation may agree or disagree with the ground truth answer and differ in detail and extent. Please rate the following statements considering the quality of the LLM explanation above with respect to the ground truth answer: *

	Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
The explanation is short.	◯	◯	◯	◯	◯	◯	◯
The explanation can only have one interpretation .	◯	◯	◯	◯	◯	◯	◯
I did not need any clarifications to interpret the explanation.	◯	◯	◯	◯	◯	◯	◯
The explanation provides all the details about the condition.	◯	◯	◯	◯	◯	◯	◯
The explanation is complete in covering relevant information.	◯	◯	◯	◯	◯	◯	◯
The explanation encompasses all aspects and outcomes related to the condition.	◯	◯	◯	◯	◯	◯	◯
I have confidence in the explanations generated by the LLM tool.	◯	◯	◯	◯	◯	◯	◯
The sequence of events leading to the condition is clearly outlined.	◯	◯	◯	◯	◯	◯	◯
I feel secure in relying on the LLM tool for correct explanations.	◯	◯	◯	◯	◯	◯	◯
The explanation is presented in an unambiguous and straightforwa rd manner.	◯	◯	◯	◯	◯	◯	◯
The explanation is easy to understand.	◯	◯	◯	◯	◯	◯	◯
I am curious to know what explains the condition.	◯	◯	◯	◯	◯	◯	◯

Loan approval and offer issuance (2 of 2)

This domain is about the processing of loan applications. In the process, once loan application was placed, it is reviewed for acceptance. If accepted, a loan offer is prepared to match the details of the request. Once created, the offer is sent to the client via mail (or online if customer is connected). Within 30 days post offer issuance, the offer is returned being accepted or refused. At any point during this period, the customer may cancel their application.

In this domain, the loan issuing bank is interested in the question: "**What is the reason for loan offer lateness post customer application?"** (condition)

Analysis of the process revealed the following **ground truth answer**:

Reasons for lateness may reflect bottlenecks in the process after an offer is sent and till the point it is returned or canceled by the applicant. The preceding part of the process during which the application is accepted and an offer is created is not causing any delay in the time it takes to receive a loan offer.

Presenting the same question to an LLM, the following explanation was generated:

"The reason for loan offer lateness could be attributed to the multiple steps involved in the process after the initial acceptance. Specifically, the stages from 'Application_Accepted' through various steps like 'Offer_Create', 'Offer_Created', 'Offer_Sent (mail and online)', 'Offer_Returned', to finally reaching 'Offer_Accepted' indicate multiple transitions and activities. Each of these steps potentially adds to the overall time taken to complete the loan offer process, contributing to any lateness observed."

---

9. Note that the LLM explanation may agree or disagree with the ground truth answer and differ in detail and extent. Please rate the following statements considering the quality of the LLM explanation above with respect to the ground truth answer: *

	Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
The explanation encloses what causes the condition.	◯	◯	◯	◯	◯	◯	◯
I find the LLM tool to be highly reliable and accurate consistently.	◯	◯	◯	◯	◯	◯	◯
The explanation accurately reflects the properties of the condition.	◯	◯	◯	◯	◯	◯	◯
The key elements of the condition are succinctly presented.	◯	◯	◯	◯	◯	◯	◯
The explanation is about why the condition occurred.	◯	◯	◯	◯	◯	◯	◯
I am eager to understand why the condition occurred.	◯	◯	◯	◯	◯	◯	◯
The explanation does not include irrelevant information.	◯	◯	◯	◯	◯	◯	◯
The explanation is consistent with the process.	◯	◯	◯	◯	◯	◯	◯
The explanation avoids assumptions not backed by data or process evidence.	◯	◯	◯	◯	◯	◯	◯
The explanation aligns with my perception on how the condition occurred.	◯	◯	◯	◯	◯	◯	◯
I am intrigued to find out why some other condition did not occur instead.	◯	◯	◯	◯	◯	◯	◯
The explanation matches my							

matches my level of expertise and knowledge of the process.	◯	◯	◯	◯	◯	◯	◯