Survey: Using Explainable AI (XAI) Techniques on a Data Privacy dataset

Thanks for taking this survey! As part of my capstone project, I aim to assess the effectiveness of several machine learning model explanations and your sentiments & opinions about decisions made by artificial intelligence (AI).

Please note that there are no "correct" answers to the questions that follow, simply choose whichever option that you agree with the most.

This survey takes approximately 15 minutes to complete. I highly recommend you complete it on a desktop / laptop within one sitting.

* Required

1. If you are a NUS / Yale-NUS student, are you 18 years old and above? *

OR

If you are not a NUS student, are you above 21+ years old?

These age restrictions are due to Yale-NUS ethics guidelines.

Mark only one oval.

- ◯ Yes
- ◯ No

Participant Information Sheet

The following information contains details about this capstone. After reading the details, click "Yes" at the bottom of this section if you wish to participate in this research.

1. Research Project Title
Applying Explainable AI techniques onto the APP-350 Corpus

2. Who is this researcher?
I am a fifth-year Double Degree in Law and Liberal Arts student at Yale-NUS College, supervised by Professor Michael Choi and Professor Simon Chesterman.

3. What is the purpose of this research?
You are invited to participate in a research study. This sheet provides you with information about the research study. The researcher identified above will describe this research to you and answer all of your questions. Read the information below before deciding whether or not to take part. My capstone aims to train and assess explainable AI models that are able to classify legal text. Part of the explanability of these models include visualisations of how the model classifies text. I aim to survey law and non-law students on how interpretable these visualisations are. I will ask you whether you find these visualisations understandable, and whether you think these are reasonable explanations of why the AI model made a certain prediction. I will also ask questions that survey how much you trust the AI model as a result of these explanations.

4. Who can participate in the research? What is the expected duration of my participation? What is the duration of this research?
18+ (NUS / Yale-NUS students), and 21+ for the general public. The time demand is about 10 – 15 minutes, and the overall duration of study is from December 2022 – May 2023.

5. What will happen to me if I take part?
You will answer questions on a Google form. I aim to survey law and non-law students on how interpretable these visualisations are. I will ask you whether you find these visualisations understandable, and whether you think these are reasonable explanations of why the AI model made a certain prediction. I will also ask questions that survey how much you trust the AI model as a result of these explanations.

6. If I agree to take part, what happens to the data I provide?
Collected data will be kept confidential and will only be used for the stated research, which will take place between August 2022 and May 2023. Unidentified quotes may be used in presentations or the capstone report, which will not contain your name. All data (not including personal identifiers, such as names and contact information) collected will be kept in accordance with the University's Research Data Management Policy. Research data used in any publication will be kept for a minimum of 10 years before being discarded.

7. How will my privacy and the confidentiality of my research records be protected?
No personal identifiable information will be recorded as this is an online survey.

8. What are the possible discomforts and risks for participants?
No discomfort or risk is expected with this research.

9. What is the compensation for any injury?
If you follow the directions of the researcher in charge of this research study and you are injured, the NUS will pay the medical expenses for the treatment of that injury. By giving your consent, you will not waive any of your legal rights or release the parties involved in this study from liability for negligence.

10. Will there be reimbursement for participation?
No reimbursement will be given.

11. What are the possible benefits to me and to others?
There is no direct benefit to you by participating in this research study.

12. Can I refuse to participate in this research?
Yes, you can. Your decision to participate in this research study is voluntary and completely up to you. You can also withdraw from the research at any time without giving any reasons, by informing the researcher or supervisor, after which all of your data collected will be discarded.

13. Whom should I call if I have any questions or problems?
Name of Researcher: Tristan Koh
Contact number: 96219123
Email: tristan.koh@u.yale-nus.edu.sg
Supervisor's Name: Michael Choi
Supervisor's email: tristan.koh@u.yale-nus.edu.sg
Secretariat, College Ethics Review Committee: researchethics@yale-nus.edu.sg

Thank you for reading this information sheet and for considering taking part in this research.

2. I have read about the purpose of this research study, agree to participate, and understand that I can withdraw at any time. *

Mark only one oval.

- ◯ Yes
- ◯ No

Part 1

I would like to capture some demographic information as well as your beliefs and views of artificial intelligence (AI).

3. What is your major / prospective major? *

If you are in a double degree program, choose whichever degree you have more experience in.

If you have graduated, choose the subject area which best corresponds to your expertise in your current work.

Mark only one oval.

- ◯ Law
- ◯ MCS / Computer Science / Data Science / Statistics
- ◯ Accountancy
- ◯ Anthropology
- ◯ Arts & Humanities
- ◯ Business
- ◯ Economics
- ◯ Engineering
- ◯ Environmental Studies
- ◯ Global Affairs / Political Science
- ◯ History
- ◯ Literature
- ◯ Life Sciences
- ◯ Medicine
- ◯ Physical Sciences
- ◯ Philosophy
- ◯ Psychology
- ◯ PPE
- ◯ PFM
- ◯ Urban Studies
- ◯ Other: _____

4. What is your age? *

Mark only one oval.

- ◯ 18 - 25
- ◯ 26 - 35
- ◯ 36 - 45
- ◯ 46 - 55
- ◯ 56+

5. Do you have any experience with AI / data science / programming? *

Mark only one oval.

None at all

1

2

3

4

5

Very experienced

6. Do you have any experience with regarding data privacy / law? *

Mark only one oval.

None at all

1

2

3

4

5

Very experienced

7. Are you concerned about your data privacy? *

Mark only one oval.

Not concerned at all

1
2
3
4
5

Very concerned

8. How would you rate your capability in protecting your online data? *

Mark only one oval.

Not capable at all

1
2
3
4
5

Very capable

9. Do you think decisions that are made by AI can be useful to society? *

In answering this question and later questions, please use whichever meaning of "useful" you agree with the most.

Mark only one oval.

Not useful at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very useful

10. Do you think decisions made by AI are fair? *

In answering this question and later questions, please use whichever meaning of "fair" you agree with the most.

Mark only one oval.

Very unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Perfectly fair

11. Do you think decisions made by AI can be a risk to society? *

In answering this question and later questions, please use whichever meaning of "risk" you agree with the most.

Mark only one oval.

	No risk to society
1	◯
2	◯
3	◯
4	◯
5	◯
	High risk to society

Part 2

To answer the rest of the questions, I would like you to read the text below to understand the context of the capstone. Don't worry if you are not able to understand everything, as this survey is meant to assess laypeople's understanding of AI.

What are data privacy policies?

- Every app that you download contains a data privacy policy that states how the app will collect and use your personal data.
- By using the app, you consent to the data privacy policy.
- Due to the Personal Data Protection Act (PDPA), app developers have to notify you and ask for your consent if they collect and use your personal data. If they fail to do so, they would be in violation of the PDPA.

How does AI work?

- The machine learning model ("model") in this capstone is used to predict whether sentences from an app's data privacy policy fall into a particular data practice. A data practice is what the app does with the user's data.
- For example, consider the sentence: "In connection with these advertising services, we may use cookies, web beacons, and similar technologies to collect behavioral information about how you use our site or other websites in order to perform tracking and marketing analytics or serve advertisements that are more likely to be of interest to you."
- Cookies are files created by websites you visit. They make your online experience easier by saving browsing information.
- This sentence is classified as "*Identifier_Cookie_or_similar_Tech_1stParty*" because the sentence states that the app uses cookies (or other tracking technologies) to track the user's activities. This sentence is also classified as "1stParty" as the data is only collected by the app, and not shared with other organisations.

- However, as the models make predictions by generalising from examples given to the model, not all the model's predictions will be correct. The model could classify the sentence wrongly as another data privacy practice, or it could also classify it as not containing any data privacy practice.
- For example, consider the sentence: "These technologies also enable us to provide features such as storage of items in your cart between visits and Short Message Service (SMS)/text messages you have chosen to receive."

- Even though the sentence does not contain the word "cookies", similar tracking technology are still being used because the app is able to track items in the user's cart in between visits. Therefore, tracking technology is still being used even though the word "cookies" is not specifically stated.
- Therefore, if the model relies heavily on "cookies" as a key word to correctly classify the sentence, the model's prediction would likely be wrong as the sentence does not contain "cookies".

12. Please select "strongly agree" to show that you are paying attention to this question. *

Mark only one oval.

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

In this section, I will describe three different contexts with similar facts that relate to the use of the abovementioned model in analysing data privacy policies.

Each context corresponds with the perspective of an app developer, a member of the Personal Data Protection Commission (PDPC), and an user of the app.

I would then ask you questions to capture how your opinions on the use of AI in decision making would differ based on these three different perspectives.

Context 1: Imagine that you are an app developer. You are developing an app that uses cookies to track user activity online. To comply with the PDPA, you know that you need to include a sentence in your app's data privacy policy that notifies and asks for users' consent to use cookies.

Since you have no knowledge of the PDPA, you use the abovementioned model to analyse a pre-drafted data privacy policy that you found online. The model informs you that there is a sentence which states that cookies are being used.

You are deciding whether to rely entirely on the model's prediction, or pay costly legal fees to confirm with your friend who is a lawyer.

If the pre-drafted data privacy policy actually does not state that cookies are being used but your app uses cookies, you could face a fine of up to $10,000 in breach of the PDPA as you would have failed to notify your users.

How far do you, as the app developer:

13. Think that using the model is an effective method of identifying violations of the PDPA? *

In answering this question and later questions, please use whichever meaning of "effective" you believe is the most appropriate for the context.

Mark only one oval.

Not effective at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very effective

14. Think that using the model is a fair method of identifying violations of the PDPA? *

In answering this question and later questions, please use whichever meaning of "fair" you believe is the most appropriate for the context.

Mark only one oval.

Very unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very fair

15. Think that using the model is a method that could be a risk to society? *

In answering this question and later questions, please use whichever meaning of "risk to society" you believe is the most appropriate for the context.

Mark only one oval.

Very risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Not risky

16. Trust the prediction made by the model? *

In answering this question and later questions, please use whichever meaning of "trust" you believe is the most appropriate for the context.

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust very much

Context 2: Imagine that you are a committee member part of the Personal Data Protection Commission (PDPC). A user of an app has informed you that an app is using cookies but has not notified its users.

Your team checks the code of the app and confirms that the app is indeed using cookies. Your team uses the abovementioned model and the model informs you that the data privacy policy does not contain any sentence that notifies its users that it uses cookies.

To increase the efficiency of the PDPC, your team is considering whether to adopt the abovementioned model to automate the analysis of data privacy policies. If this new method of analysis is adopted, the PDPC would rely entirely on the model's predictions to confirm whether app developers have breached the PDPA. The app developers would face a fine of up to $10,000 if they are found to have breached the PDPA.

How far would you, as a committee member of the PDPC:

17. Think that using the model is an effective method of identifying violations of the PDPA? * *Mark only one oval.*

Not effective at all

1

2

3

4

5

Very effective

18. Think that using the model is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very unfair

1

2

3

4

5

Very fair

19. Think that using the model is a method that could be a risk to society? *

Mark only one oval.

Very risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Not risky

20. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust very much

Context 3: Imagine that you are a user of an app. You read in a forum where other users allege that the app uses cookies. You decide to analyse the data privacy policy of the app using the abovementioned model and the model informs you that the data privacy policy does not contain any sentence that notifies its users that it uses cookies.

You are deciding whether to submit this prediction as the only supporting piece of evidence to the PDPC to claim that the app has used cookies without notifying you.

If the PDPC decides that the developer has indeed violated the PDPA, you could claim compensation from the app developer of up to $10,000.

How far would you, as a user of the app:

21. Think that using the model is an effective method of identifying violations of the PDPA? *

Mark only one oval.

Not effective at all

1
2
3
4
5

Very effective

22. Think that using the model is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very unfair

1
2
3
4
5

Very fair

23. Think that using the model is a method that could be a risk to society? *

Mark only one oval.

Very risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Not risky

24. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust very much

Part 3

In the next 3 sections, I would like to capture your opinions about the understandability of the model that I have been using in this capstone. Please carry on reading to learn more about the specifics of the model.

As above, don't worry if you are not able to understand everything, as this survey is meant to assess laypeople's understanding of AI.

What is this model used to predict?
This section involves predictions made by a model trained on sentences with the data practice “Identifier_Cookie_or_similar_Tech_1stParty”.

The definition of this data practice is that:

- The app uses cookies (or other tracking technologies) to track the user's activities.
- The data is only collected by the app itself, and not shared with other organisations.

The following sentences are some illustrative sentences that were predicted correctly by this model (i.e. the predicted practice was correctly predicted to be "Identifier_Cookie_or_similar_Tech_1stParty"):

- We automatically receive and track certain information about your computer or mobile device when you visit our sites or apps, including through the use of cookies.
- However, if you block or erase cookies, we may not be able to restore any preferences or customisation settings you have previously specified, and our ability to personalise your online experience would be limited.
- Other technologies, such as Silverlight storage, may be cleared from within the application.

However, the following sentences are also illustrative examples of sentences which were classified wrongly by the model (i.e. the practice was not predicted to be "Identifier_Cookie_or_similar_Tech_1stParty"):

- In connection with these advertising services, we or our Advertising Service Providers, like Google Analytics may use cookies, web beacons, and similar technologies to collect behavioral information about how you use our site or other websites in order to perform tracking and marketing analytics or serve advertisements that are more likely to be of interest to you.
- Shared Information also includes information about you (including Location Data and Log Data) that others who are using our services share about you.
- As explained above, you may either volunteer to us certain information (such as your email address), or we may automatically collect certain information, such as through the use of your mobile device system's permissions, or through the use of cookies or similar tracking technologies.

Out of 425 sentences, the model:

1. correctly classified 311 sentences; and
2. wrongly classified 114 sentences.
3. The accuracy of the model is 73%.

Note that the model was also trained to identify four other data practices. Hence, you would see other practices stated in some visualisations when the model predicted the wrong practice.

25. Do you understand why the model made the prediction? *

Mark only one oval.

	Don't understand at all					Fully understand
	1	2	3	4	5	

26. Why do you think the model made this prediction? *

A short answer would suffice, there is no "correct" answer.

Instructions for this section

In this section, you will be asked to assess the effectiveness of visualisations that explain how the abovementioned model makes predictions.

You will also be told whether the model predicted the practice correctly / wrongly, and be given the predicted practice by the model and actual practice, such as:

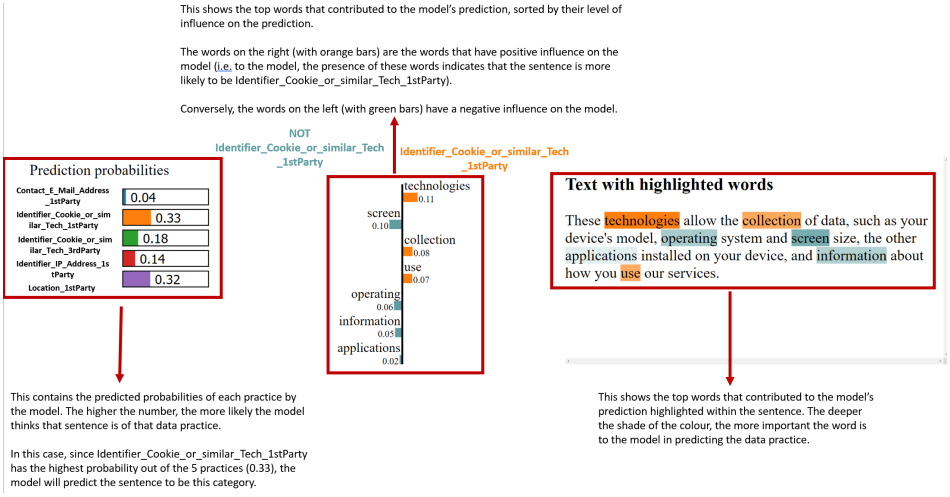
The model predicted the practice wrongly.

Predicted practice: Identifier_Cookie_or_Similar_Tech_1stParty

Actual practice: Identifier_IP_Address

Please note that there are no "correct" answers, simply choose the option which you think you agree with the most.

The following image explains how to interpret a visualisation.

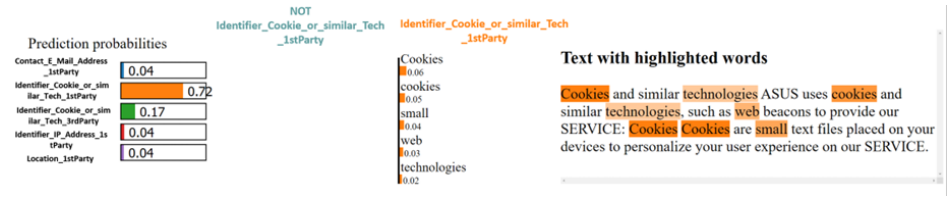


Visualisation 3.1

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_Similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty



27. Did you find the visualisation easy to interpret? *

Mark only one oval.

- Very difficult
- 1
- 2
- 3
- 4
- 5

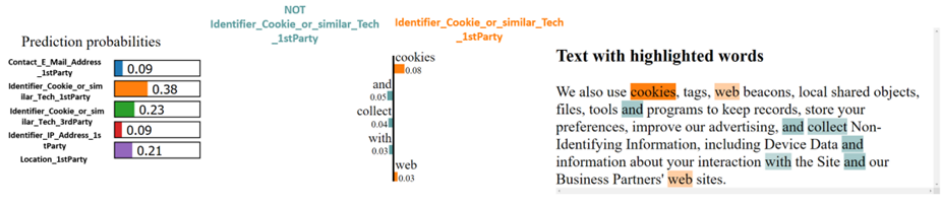
Very easy

Visualisation 3.2

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_Similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty



28. Do you understand why the model made the prediction? *

Mark only one oval.

Don't understand at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Fully understand

29. Why do you think the model made this prediction? *

A short answer would suffice, there is no "correct" answer.

30. Did you find the visualisation easy to interpret? *

Mark only one oval.

Very difficult

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very easy

31. Based on your current understanding, do you think that the sentence below would be predicted to be "Identifier_Cookie_or_Similar_Tech_1stParty"? *

"We also use **tracking technologies** to keep records, store your preferences, improve our advertising, and collect Non-Identifying Information, including Device Data and information about your interaction with the Site and our Business Partners' web sites."

The difference in the sentence in the visualisation above and the provided sentence is highlighted in **bold**.

Mark only one oval.

- ◯ Yes
- ◯ No

Visualisation 3.3

The model predicted the practice wrongly.

Predicted practice: Contact_E_Mail_Address_1stParty

Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty

Note: Contact_E_Mail_Address_1stParty is a data practice where the app collects the email address of the app user.

32. Do you understand why the model made the prediction? *

Mark only one oval.

Don't understand at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Fully understand

33. Why do you think the model made this prediction? *

A short answer would suffice, there is no "correct" answer.

34. Did you find the visualisation easy to interpret? *

Mark only one oval.

Very difficult

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very easy

35. Based on your current understanding, do you think the sentence below would be predicted to be in "Identifier_Cookie_or_Similar_Tech_1stParty?" *

"As explained above, you may either volunteer to us certain information (such as your **phone number**), or we may automatically collect certain information, such as through the use of your mobile device system's permissions, or through the use of cookies or similar tracking technologies."

The difference in the sentence in the visualisation and the provided sentence is highlighted in **bold**.

Mark only one oval.

- ◯ Yes
- ◯ No

Visualisation 3.4

The model predicted the practice wrongly.

Predicted practice: Identifier_IP_Address_1stParty

Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty

Note: Identifier_IP_Address_1stParty is a data practice where the app collects the user's IP address.

36. Do you understand why the model made the prediction? *

Mark only one oval.

Don't understand at all

1
2
3
4
5

Fully understand

37. Why do you think the model made this prediction? *

A short answer would suffice, there is no "correct" answer.

38. Did you find the visualisation easy to interpret? *

Mark only one oval.

Very difficult

1
2
3
4
5

Very easy

39. Based on your current understanding, do you think the sentence below would be predicted to be in "Identifier_Cookie_or_Similar_Tech_1stParty?" *

"These **cookies and other such tracking technologies** allow the collection of data, such as your device's model, operating system and screen size, the other applications installed on your device, and information about how you use our services."

The difference in the sentence in the visualisation and the provided sentence is highlighted in **bold**.

Mark only one oval.

- ◯ Yes
- ◯ No

Part 4

In this section, you will be given pairs of different visualisations of the predictions of the same sentence. For each visualisation, I will state whether the model predicted the practice correctly / wrongly, and also state the predicted practice by the model and the actual practice.

Choose the visualisation that is more understandable to you on first glance. If both visualisations seem the same to you, then select "no difference".

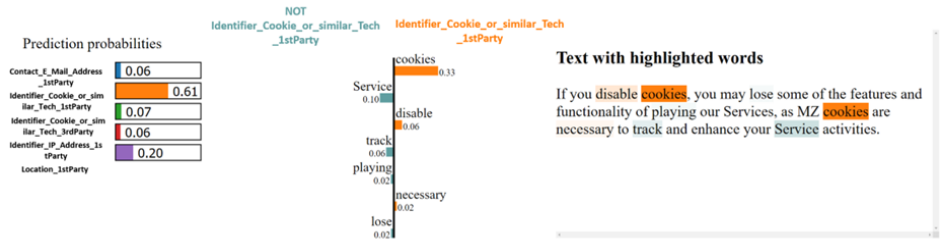
Once again, there is no "correct" option. Simply choose whichever option that you agree with the most.

Visualisation 4.1(i)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



Visualisation 4.1(ii)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

40. What explanation did you find easier to interpret? * *Mark only one oval.*

- 4.1(i)
- 4.1(ii)

No difference

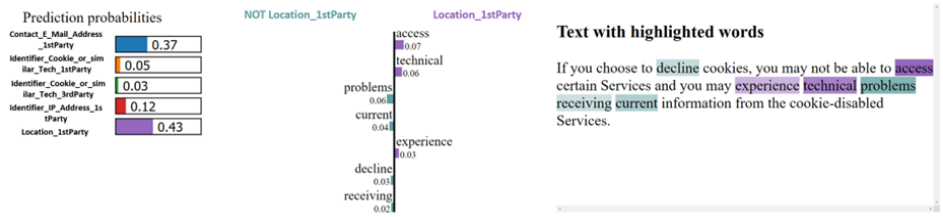
Visualisation 4.2(i)

The model predicted the practice wrongly.

Predicted practice: Location_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

Note: Location_1stParty is a data practice where the app collects the location data of the user.



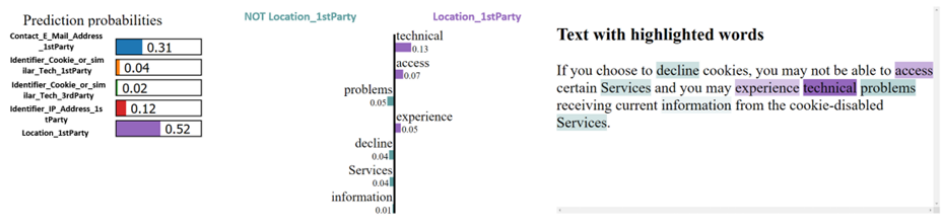
Visualisation 4.2(ii)

The model predicted the practice wrongly.

Predicted practice: Location_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

Note: Location_1stParty is a data practice where the app collects the location data of the user.



41. Which explanation did you find easier to interpret? *

Mark only one oval.

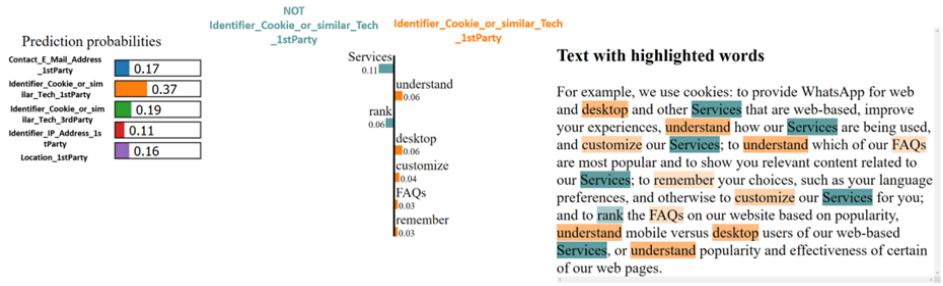
- ◯ 4.2(i)
- ◯ 4.2(ii)
- ◯ No difference

Visualisation 4.3(i)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

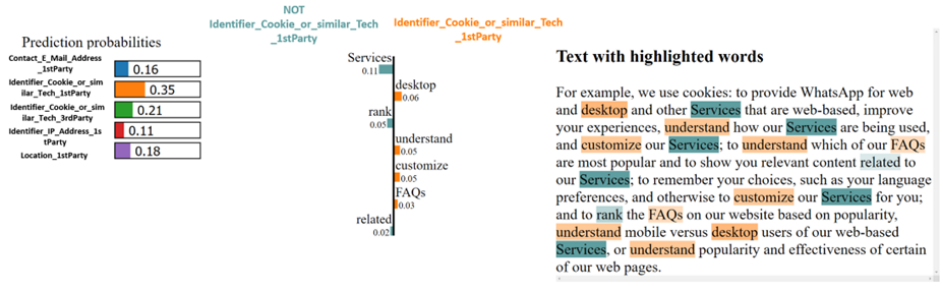


Visualisation 4.3(ii)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



42. Which explanation did you find easier to interpret? *

Mark only one oval.

- ◯ 4.3(i)
- ◯ 4.3(ii)
- ◯ No difference

Part 5

As with the previous section, you will be given pairs of different visualisations of the predictions of the same sentence. Choose the visualisation that is more understandable to you on first glance.

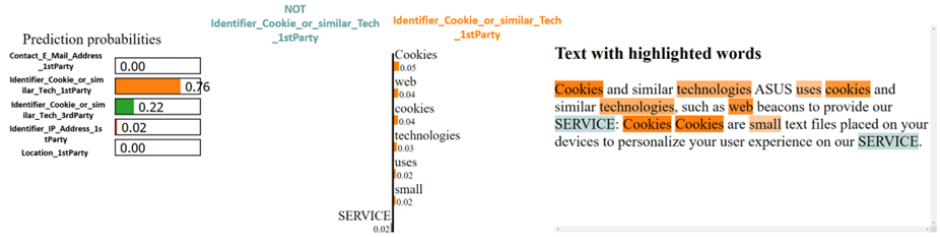
Choose "no difference" if the visualisations seem the same to you.

Visualisation 5.1(i)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

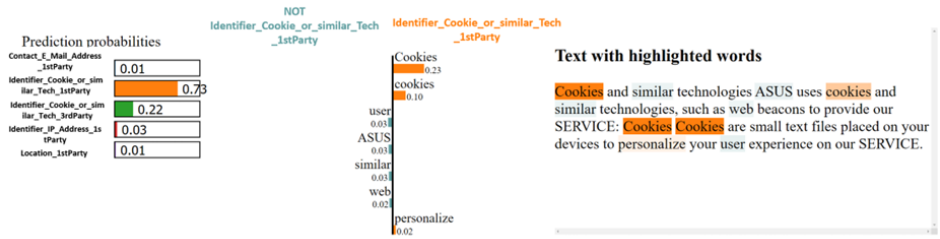


Visualisation 5.1(ii)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



43. Which explanation did you find easier to interpret? *

Mark only one oval.

- 5.1(i)
- 5.1(ii)
- No difference

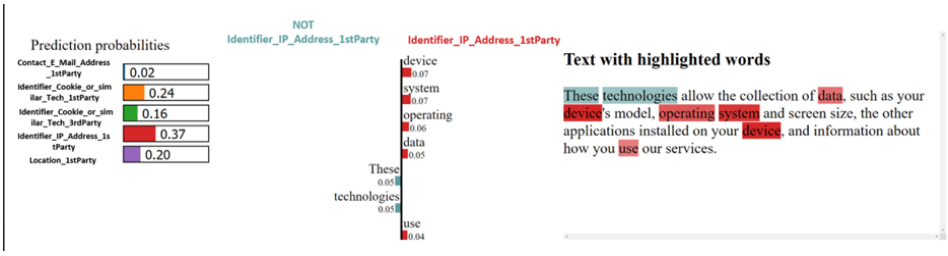
Visualisation 5.2(i)

The model predicted the practice wrongly.

Predicted practice: Identifier_IP_Address_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

Note: Identifier_IP_Address_1stParty is a data practice where the app collects the user's IP address.

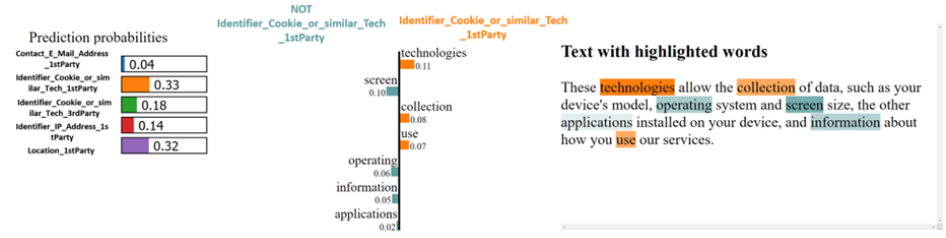


Visualisation 5.2(ii)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



44. Which explanation did you find easier to interpret? *

Mark only one oval.

- ◯ 5.2(i)
- ◯ 5.2(ii)
- ◯ No difference

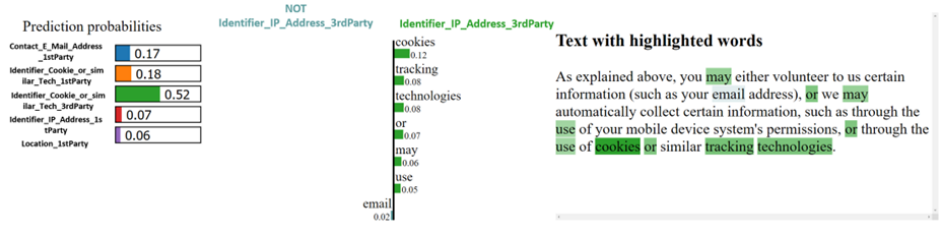
Visualisation 5.3(i)

The model predicted the practice wrongly.

Predicted practice: Identifier_Cookie_or_similar_Tech_3rdParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

Identifier_Cookie_or_similar_Tech_3rdParty is a practice where the app uses cookies or similar tracking technologies and shares it with third party organisations.

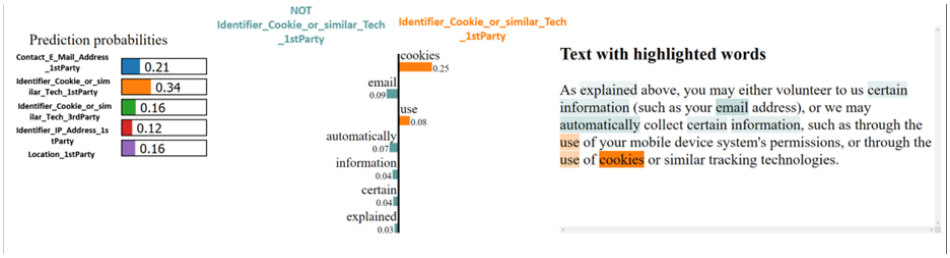


Visualisation 5.3(ii)

The model predicted the practice correctly.

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty

Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



45. Which explanation did you find easier to interpret? *

Mark only one oval.

- 5.3(i)
- 5.3(ii)
- No difference

Part 6

This section asks you to respond to the same questions in the same three contexts (i.e. app developer, member of the PDPA, app user) as described in Part 3.

Similarly, please consider how your opinions about the use of AI in decision making changes with the different perspectives of these three roles.

Please answer these following questions taking into account your current understanding of the model having viewed the explanations in the previous parts.

Context 1: Imagine that you are an app developer. You are developing an app that uses cookies to track user activity online. To comply with the PDPA, you know that you need to include a sentence in your app's data privacy policy that notifies and asks for users' consent to use cookies.

Since you have no knowledge of the PDPA, you use the abovementioned model to analyse a pre-drafted data privacy policy that you found online. The model informs you that there is a sentence which states that cookies are being used.

You are deciding whether to rely entirely on the model's prediction, or pay costly legal fees to confirm with your friend who is a lawyer.

If the pre-drafted data privacy policy actually does not state that cookies are being used but your app uses cookies, you could face a fine of up to $10,000 in breach of the PDPA as you would have failed to notify your users.

How far do you, as the app developer:

46. Think that using the model is an effective method of identifying violations of the PDPA? *

Mark only one oval.

Not effective at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very effective

47. Think that using the model is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Perfectly fair

48. Think that using the model is a method that could be a risk to society? *

Mark only one oval.

Very risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Not risky

49. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust very much

Context 2: Imagine that you are a committee member part of the Personal Data Protection Commission (PDPC). A user of an app has informed you that an app is using cookies but has not notified its users.

Your team checks the code of the app and confirms that the app is indeed using cookies. Your team uses the abovementioned model and the model informs you that the data privacy policy does not contain any sentence that notifies its users that it uses cookies.

To increase the efficiency of the PDPC, your team is considering whether to adopt the abovementioned model to automate the analysis of data privacy policies. If this new method of analysis is adopted, the PDPC would rely entirely on the model's predictions to confirm whether app developers have breached the PDPA. The app developers would face a fine of up to $10,000 if they are found to have breached the PDPA.

How far would you, as a committee member of the PDPC:

50. Think that using the model is an effective method of identifying violations of the PDPA? *

Mark only one oval.

Not effective at all

1

2

3

4

5

Very effective

51. Think that using the model is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very unfair

1

2

3

4

5

Very fair

52. Think that using the model is a method that could be a risk to society? *

Mark only one oval.

	Very risky
1	◯
2	◯
3	◯
4	◯
5	◯
	Not risky

53. Trust the prediction made by the model? *

Mark only one oval.

	Do not trust at all
1	◯
2	◯
3	◯
4	◯
5	◯
	Trust very much

Context 3: Imagine that you are a user of an app. You read in a forum where other users allege that the app uses cookies. You decide to analyse the data privacy policy of the app using the abovementioned model and the model informs you that the data privacy policy does not contain any sentence that notifies its users that it uses cookies.

You are deciding whether to submit this prediction as the only supporting piece of evidence to the PDPC to claim that the app has used cookies without notifying you.

If the PDPC decides that the developer has indeed violated the PDPA, you could claim compensation from the app developer of up to $10,000.

How far would you, as a user of the app:

54. Think that using the model is an effective method of identifying violations of the PDPA? *

Mark only one oval.

Not effective at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very effective

55. Think that using the model is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very fair

56. Think that using the model is a method that could be a risk to society? *

Mark only one oval.

Very risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Not risky

57. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust very much