Survey: Applying Explainable AI (XAI) Techniques On The APP-350 Corpus

Thanks for taking this survey! This survey aims
to assess the effectiveness of several machine learning explanations and your sentiments & opinions about decisions made by artificial intelligence.

I aim to survey law and non-law NUS students on how
interpretable these explanations are. I will ask you whether you find these explanations 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 because of these explanations.

* Required

Part 1

This section captures some demographic information as well as your beliefs and views of artificial intelligence (AI).

1. What is your major? *

Mark only one oval.

- ◯ Law
- ◯ MCS
- ◯ DDP
- ◯ Arts & Humanities
- ◯ Anthropology
- ◯ Economics
- ◯ Environmental Studies
- ◯ Global Affairs
- ◯ History
- ◯ Literature
- ◯ Life Sciences
- ◯ Physical Sciences
- ◯ Philosophy
- ◯ Psychology
- ◯ PPE
- ◯ Urban Studies
- ◯ Other: _____

2. What is your age? (e.g. 21) *

_____

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

Mark only one oval.

None at all

1

2

3

4

5

Very experienced

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

Mark only one oval.

None at all

1

2

3

4

5

Very experienced

5. How much are you concerned about your data privacy? *

Mark only one oval.

Not concerned at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very concerned

6. 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

7. How far do you think decisions that are made by AI can be useful to society? *

Mark only one oval.

Not useful at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very useful

8. How far do you think decisions made by AI are fair? *

Mark only one oval.

Very unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Perfectly fair

9. How far do you think decisions made by AI can be a risk to society? *

Mark only one oval.

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

Part 2

Short primer on how AI works

This capstone involves explaining the predictions of machine learning models (“models”).

These models are used to predict whether sentences from an app 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 or our Advertising Service Providers, like Google Analytics (more on this below) 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 ("Interest-Based Advertisements").”

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. Cookies are files created by websites you visit. They make your online experience easier by saving browsing information. With cookies, sites can keep you signed in, remember your site preferences, and give you locally relevant content. It is also classified as “1stParty” as the data is only collected by the app itself, and not shared with other organisations.

However, as machine learning models make predictions by generalising from examples given to the model, not all the model’s predictions will be correct.

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", cookies or such similar technology are still being used because the app is able to track items in the user's cart in between visits, meaning that tracking technology is still being used even though "cookies" are 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".

A brief description of the model used in this section

This section involves predictions made by a model trained on sentences with the data practice “Identifier_Cookie_or_similar_Tech_1stParty”.
The following sentences were some of the sentences that were predicted correctly by this model (i.e. the predicted practice was correctly predicted to be "Identifier_Cookie_or_similar_Tech_1stParty”):

- You can choose to have your computer warn you each time a cookie is being sent, or you can choose to turn off all cookies.
- If you disable or opt-out of these cookies or other technologies, it may prevent you from using certain parts of our websites and applications, and it may reduce the support or information that we can provide you.
- 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.
- You can see a list of the cookies we use in our Cookie Policy.
- Other technologies, such as Silverlight storage, may be cleared from within the application.

However, the following sentences were some 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”):

- 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.
- In connection with these advertising services, we or our Advertising Service Providers, like Google Analytics (more on this below) 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 ("Interest-Based Advertisements").
- Cookies must be enabled for you to use your Verizon e-mail account.
- 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 correctly classified 311 sentences, and wrongly classified 114 sentences. The accuracy of the model is 73%.

10. 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, you will be presented with three contexts that relate to the use of the abovementioned model in analysing data privacy policies. Answer the questions that follow each context.

Context 1: Imagine that you are an app developer and have been alleged by the Personal Data Privacy Commission (PDPC) of violating the Personal Data Protection Act (PDPA) because your app uses cookies to track user activities without their consent. The PDPC has relied entirely on the model's prediction to prove this data breach. If found true, you could face a fine of up to $10,000.

How far do you:

11. 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

12. 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

13. 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

14. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust completely

Context 2: Imagine that you are a committee member part of the PDPC and have relied entirely on the model's prediction to assess whether an app developer has violated the PDPA by using cookies to track user activities without their consent. The committee is considering whether to formally find the app developer liable of the violation of the PDPA based on this evidence.

How far would you:

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

Mark only one oval.

	Not effective at all					Very effective
	1	2	3	4	5	

16. 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

17. 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

18. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust completely

Context 3: Imagine that you are an user of an app. You decide to analyse the data privacy policy of the app using this model and the model predicts that the app has used cookies to track your online activity without your consent. If this is prediction is true, you could claim damages from the app developer of up to $10,000.

How far would you:

19. 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

20. 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

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

Mark only one oval.

Not risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very risky

22. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust completely

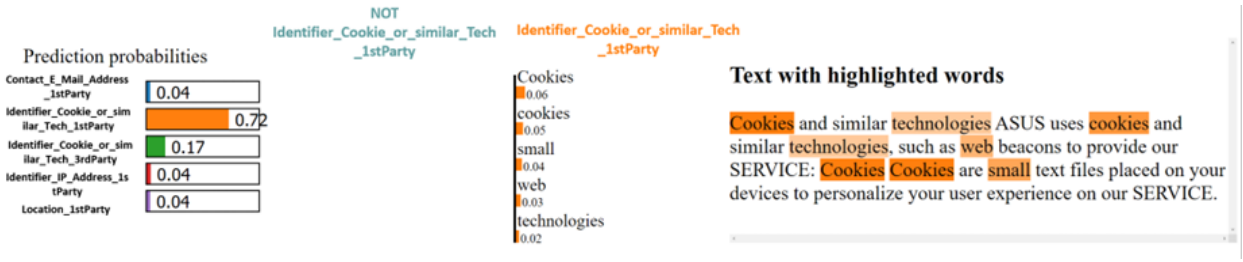
Part 4a

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 given the predicted practice by the model and actual practice.

Visualisation 1

Predicted practice: Identifier_Cookie_or_Similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty



23. How far 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

24. Based on the given visualisation, could you state why the model made this prediction? (1 or 2 sentences would suffice) *

25. How far did you find the visualisation intuitive to understand? *

Mark only one oval.

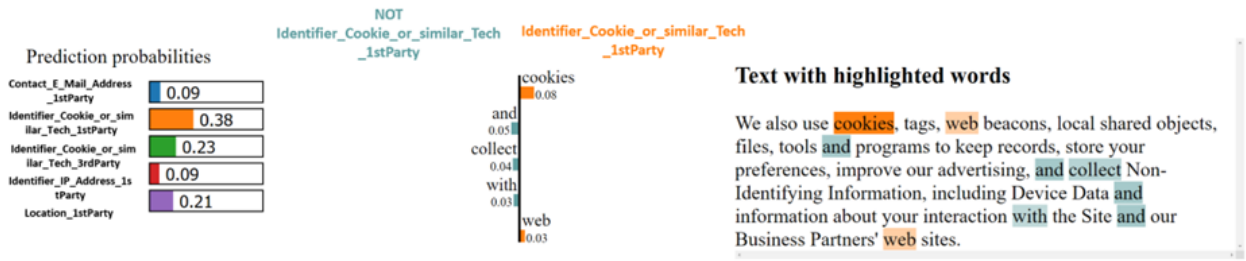
Not understandable at all

1
2
3
4
5

Fully understandable

Visualisation 2

Predicted practice: Identifier_Cookie_or_Similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty



26. How far 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

27. Based on the given visualisation, could you state why the model made this prediction? (1 or 2 sentences would suffice) *

28. How far did you find the visualisation intuitive to understand? *

Mark only one oval.

	Not understandable at all
1	◯
2	◯
3	◯
4	◯
5	◯
	Fully understandable

29. 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."

Mark only one oval.

- ◯ Yes
- ◯ No

Visualisation 3

Predicted practice: Contact_E_Mail_Address_1stParty
Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty

30. How far 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

31. Based on the given visualisation, could you state why the model made this prediction? (1 or 2 sentences would suffice) *

32. How far did you find the visualisation intuitive to understand? *

Mark only one oval.

	Not understandable at all
1	◯
2	◯
3	◯
4	◯
5	◯
	Fully understandable

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

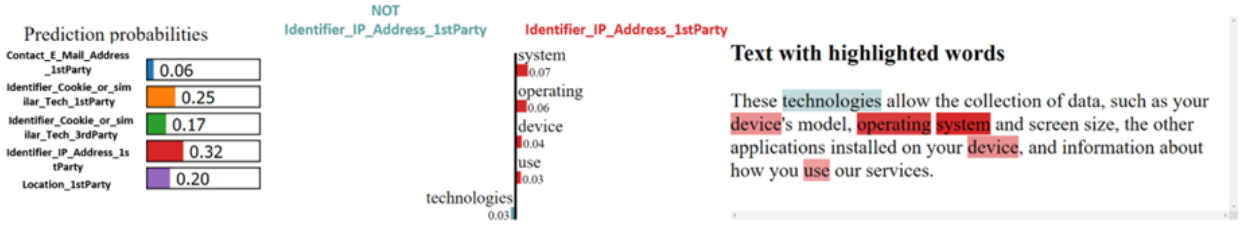
"As explained above, you may either volunteer to us certain information (such as your phone number), or we may automatically collect certain information, such as through the use of your mobile device system's permissions, or through the use of cookies or similar tracking technologies."

Mark only one oval.

- ◯ Yes
- ◯ No

Visualisation 4

Predicted practice: Identifier_IP_Address_1stParty
Actual practice: Identifier_Cookie_or_Similar_Tech_1stParty



34. How far 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

35. Based on the given visualisation, could you state why the model made this prediction? (1 or 2 sentences would suffice) *

36. How far did you find the visualisation intuitive to understand? *

Mark only one oval.

Not understandable at all

1
2
3
4
5

Fully understandable

37. 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."

Mark only one oval.

Yes
No

Part 4b

In
this 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. There is no need to overly scrutinise each explanation.

Visualisation 4.1(i)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

Visualisation 4.1(ii)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty

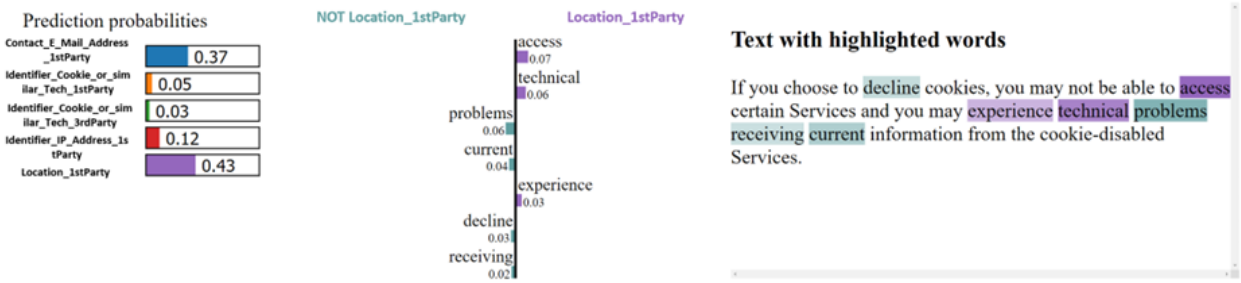
38. Which explanation seems more intuitive to you? *

Mark only one oval.

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

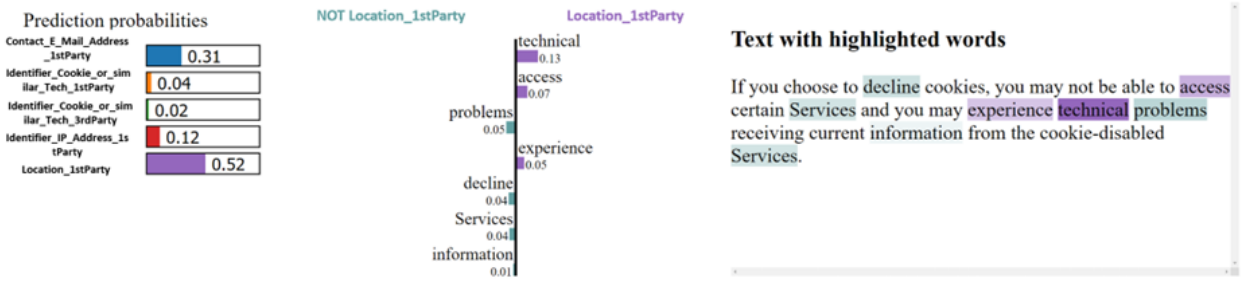
Visualisation 4.2(i)

Predicted practice: Location_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



Visualisation 4.2(ii)

Predicted practice: Location_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



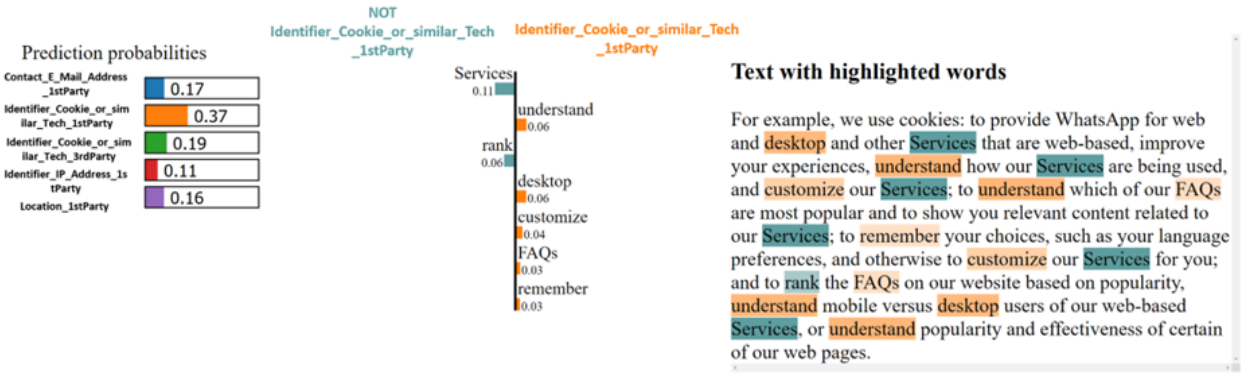
39. Which explanation seems more intuitive to you? *

Mark only one oval.

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

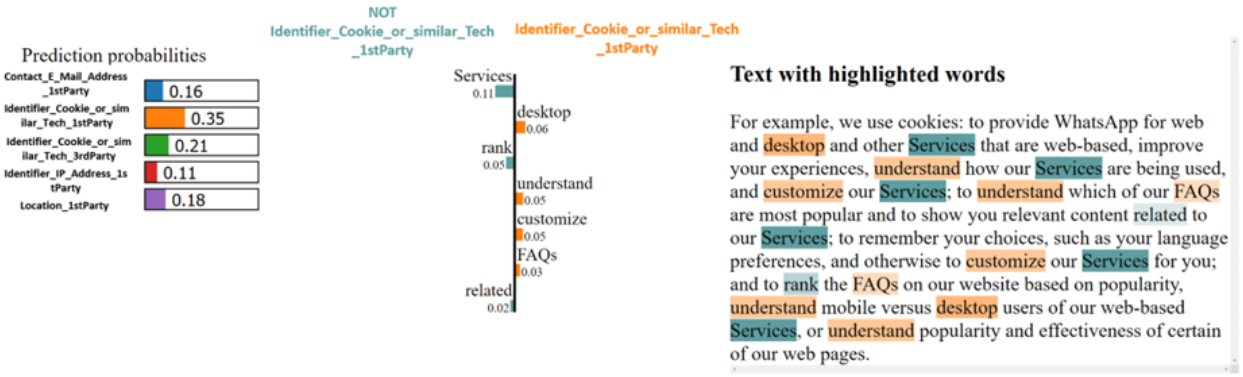
Visualisation 4.3(i)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



Visualisation 4.3(ii)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



40. Which explanation seems more intuitive to you? *

Mark only one oval.

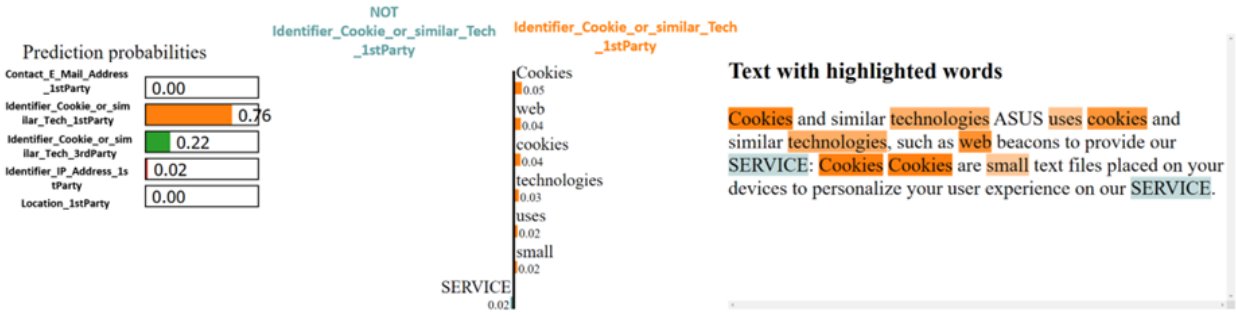
- ◯ 4.3(i)
- ◯ 4.3(ii)
- ◯ No difference

Part 4c

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. There is no need to overly scrutinise each explanation.

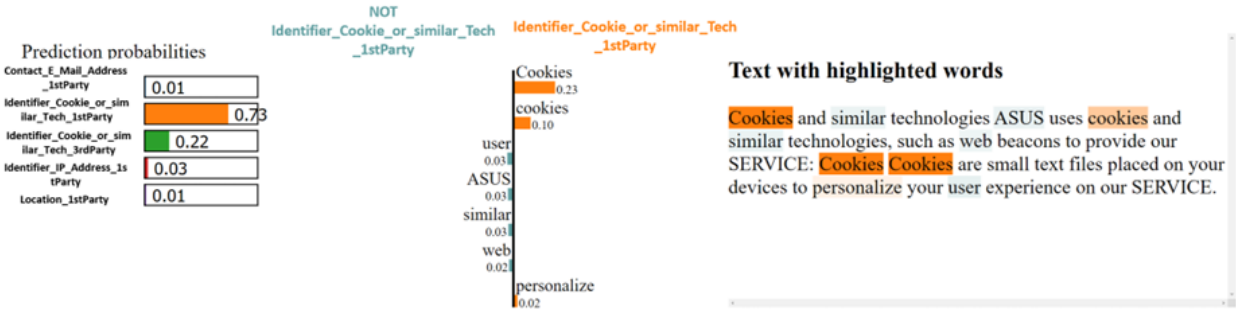
Visualisation 4.4(i)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



Visualisation 4.4(ii)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



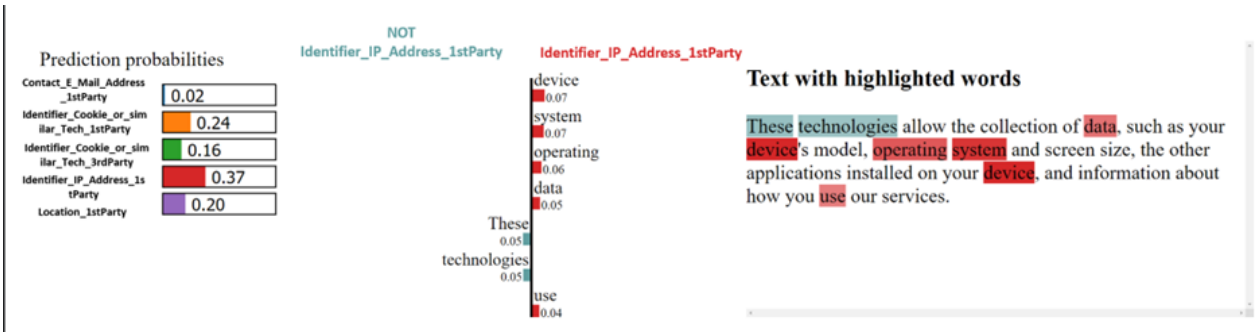
41. Which explanation seems more intuitive to you? *

Mark only one oval.

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

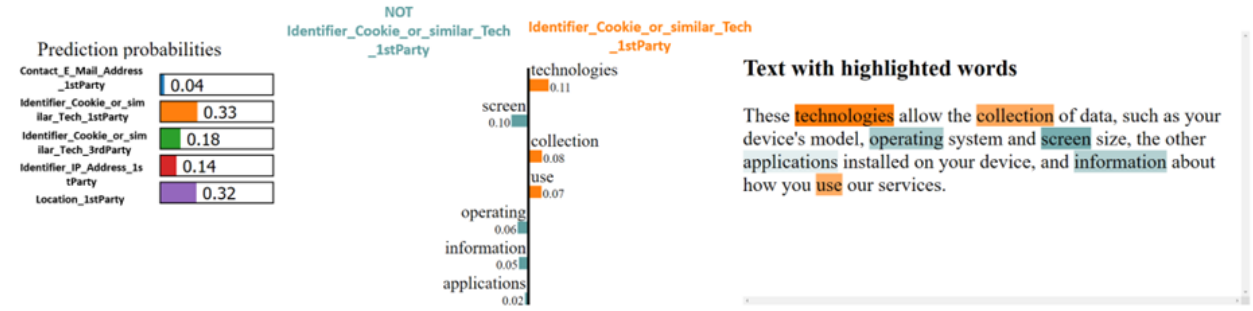
Visualisation 4.5(i)

Predicted practice: Identifier_IP_Address_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



Visualisation 4.5(ii)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



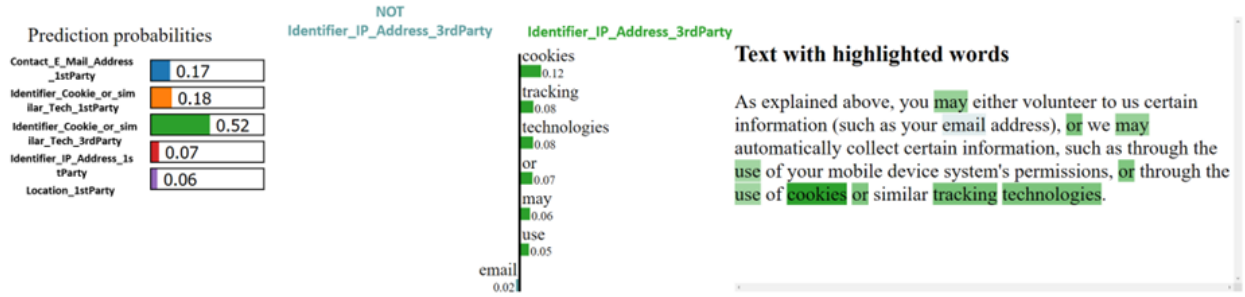
42. Which explanation seems more intuitive to you? *

Mark only one oval.

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

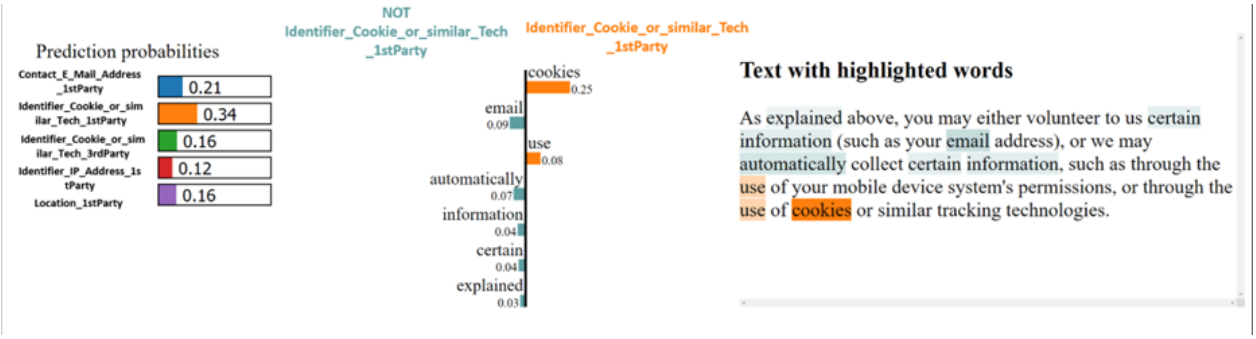
Visualisation 4.6(i)

Predicted practice: Identifier_Cookie_or_similar_Tech_3rdParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



Visualisation 4.6(ii)

Predicted practice: Identifier_Cookie_or_similar_Tech_1stParty
Actual practice: Identifier_Cookie_or_similar_Tech_1stParty



43. Which explanation seems more intuitive to you? *

Mark only one oval.

- 4.6(i)
- 4.6(ii)
- No difference

Part 5

This section asks you to respond to the same questions in the same contexts as described in Part 3. Please answer the following questions taking into account the explanations and your current understanding of the models.

Context 1: Imagine that you are an app developer and have been alleged by the Personal Data Privacy Commission (PDPC) has violated the Personal Data Protection Act (PDPA) because your app uses cookies to track user activities without their consent. The PDPC has relied entirely on the model's prediction to prove this fact. You could face a fine of up to $10,000.

How far do you:

44. Think that this is an effective method of identifying violations of the PDPA? *

Mark only one oval.

	Not effective at all					Very effective
	1	2	3	4	5	

45. Think that this is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very Unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Perfectly Fair

46. Think that this method of making decisions could be a risk to society? *

Mark only one oval.

Not risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very risky

47. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1

2

3

4

5

Trust completely

Context 2: Imagine that you are a committee member part of the PDPC and have relied entirely on the model's prediction to assess whether an app developer has committed a data breach by using cookies to track user activities without their consent. The committee is considering whether to formally find the app developer liable of the breach based on this evidence.

How far would you:

48. Think that this 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

49. Think that this is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very Unfair

1
2
3
4
5

Perfectly Fair

50. Think that this method of making decisions could be a risk to society? *

Mark only one oval.

Not risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very risky

51. Trust the prediction made by the model? *

Mark only one oval.

Do not trust at all

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Trust completely

Context 3: Imagine that you are an user of an app. You decide to analyse the data privacy policy of the app using this model and the model predicts that the app has used cookies to track your online activity without your consent. If this is prediction is true, you could claim damages from the app developer of up to $10,000.

How far would you:

52. Think that this is an effective method of identifying violations of the PDPA? *

Mark only one oval.

	Not effective at all					Very effective
	1	2	3	4	5	

53. Think that this is a fair method of identifying violations of the PDPA? *

Mark only one oval.

Very Unfair

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Perfectly Fair

54. Think that this method of making decisions could be a risk to society? *

Mark only one oval.

Not risky

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

Very risky

55. Trust the prediction made by the model? *

Mark only one oval.

	Do not trust at all
1	
2	
3	
4	
5	
	Trust completely

This content is neither created nor endorsed by Google.

Google Forms