HCD and LLMs [Anonymous Submission]

Thank You for Joining Our Study! We're grateful for your participation in this research.

By completing the form and registering for the study, you confirm your agreement with our data treatment policy. Please know that all the information you share will be handled with great care and in full compliance with the General Data Protection Regulation (GDPR).

* Required

Participants Information

1. How long have you been involved in the design of usability tests? *

- ◯ 3-6 months
- ◯ Less than 1 year
- ◯ 1-3 years
- ◯ More that 3 years

2. Level of experience in planning usability tests *

	No Knowledge	Passing Knowledge	Knowledgeable	Competent	Expert
Expertise	◯	◯	◯	◯	◯

3. What is your role in the company/organization where you work? *

- ◯ Project Manager
- ◯ Developer
- ◯ UI/UX Expert
- ◯ Researcher
- ◯ Other

4. What is your level of familiarity with Anonymous? *

	No Knowledge	Passing Knowledge	Knowledgeable	Competent	Expert
Familiarity	◯	◯	◯	◯	◯

5. What is your level of familiarity with BrainMed? *

	No Knowledge	Passing Knowledge	Knowledgeable	Competent	Expert
Familiarity	◯	◯	◯	◯	◯

Anonymous Description

[ITA]
Anonymous è un sistema basato sull'intelligenza artificiale che ha l'obiettivo di supportare i rinocitologi nella classificazione delle cellule, che includono artefatti, batteri, emazie, eosinofili, cellule epiteliali, cellule cigliate, linfociti, mastociti, cellule metaplasiche, cellule caliciformi e neutrofili.
Per ogni paziente, il sistema fornisce una tabella che mostra il tipo di cellula, il numero di cellule rilevate, l'intervallo di riferimento e una classificazione della loro quantità.
La visualizzazione di ciascun tipo di cellula include una vista categorizzata per livello di confidenza (basso, medio, alto). Il sistema consente di validare e correggere le classificazioni contrassegnandole come Corrette o Errate. In caso di errore, il medico seleziona la nuova classe. Queste funzionalità supportano il medico nella revisione e nell'eventuale correzione delle classificazioni automatiche fornite dal sistema.

Puoi familiarizzare con Anonymous interagendo con il prototipo al seguente link: https://www.figma.com/proto/pwHb8ZNXKrCXlhVrBfI5e1/AnonymousPrototype?node-id=1-17&t=uM3E8v7Itep4bdLs-1

[ENG]
Anonymous an AI-based system that has the objective of supporting rhinocytologists in classifying cells, which are artifacts, bacteria, emazia, eosinophils, epithelial cells, ciliated cells, lymphocytes, mast cells, metaplastic cells, goblet cells, and neutrophils.
For each patient, the system provides a table showing the type of cell, the number of detected cells, the reference range, and a classification of their quantity.
The visualization of each type of cell contains a view categorized by confidence level (low, medium, high). The system allows to validation and correct the classifications by marking them as *Correct* or Incorrect. If incorrect, the physician selects the new class. These features support the doctor in reviewing and potentially correcting the automatic classifications provided by the system.

Familiarize with Anonymous interacting with the prototype at the following link: https://www.figma.com/proto/pwHb8ZNXKrCXlhVrBfI5e1/AnonymousPrototype?node-id=1-17&t=uM3E8v7Itep4bdLs-1

Set 1 of Tasks Generated for Anonymous

[ITA]
Valuta i task generati dagli LLM in base ai seguenti criteri:

1. **Functionality-oriented**: il task specifica quale funzionalità dovrebbe essere esaminata dal tester, senza indicare quali azioni eseguire per raggiungere l'obiettivo.
2. **Utility**: rilevanza del task per la valutazione dell'usabilità del sistema rispetto alle sue funzionalità principali.
3. **Precision**: chiarezza del linguaggio e livello di dettaglio con cui è descritto il task.
4. **Completeness**: misura in cui il task include tutte le informazioni necessarie per l'analisi, senza ambiguità o incoerenze.

[ENG]
Evaluate the LLMs-generated tasks in term of:
1. *Functionality-oriented*: the task specifies which functionality should be examined by the tester without mentioning which actions should be performed to reach the objective.
2. *Utility*: relevance of the task for the usability assessment of the system with respect to its core functionalities.
3. *Precision*: language clarity and level of detail of the task.
4. *Completeness*: extent to which the task mentions the necessary pieces of information for the analysis without inconsistencies.

6. Validate the classification of a cell as "Correct" if the system has correctly identified it. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

7. Correct the classification of a cell as "Incorrect" if the system has misclassified it, and provide a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

8. Review the classification of a cell image and ensure that the system has correctly identified the cell type. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

9. Check the classification of a cell image and mark it as "Incorrect" if the system has misclassified it, and provide a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

10. Confirm the classification of a cell image as "Correct" if the system has correctly identified the cell type. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

11. Identify a cell image that is misclassified in the "ciliated" class and correct the classification, providing a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

12. Review the classification of a cell image and ensure that the system has correctly identified the cell type, and if not, correct it and provide a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

13. Check the classification of a cell image and mark it as "Correct" if the system has correctly identified the cell type, and if not, correct it and provide a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

14. Validate the classification of a cell as "Correct" if the system has correctly identified it, and if not, correct it and provide a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

15. Identify a cell image that is misclassified in the "eosinophils" class and correct the classification, providing a reason for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

Set 2 of Tasks Generated for Anonymous

[ITA]
Valuta i task generati dagli LLM in base ai seguenti criteri:

1. **Functionality-oriented**: il task specifica quale funzionalità dovrebbe essere esaminata dal tester, senza indicare quali azioni eseguire per raggiungere l'obiettivo.
2. **Utility**: rilevanza del task per la valutazione dell'usabilità del sistema rispetto alle sue funzionalità principali.
3. **Precision**: chiarezza del linguaggio e livello di dettaglio con cui è descritto il task.
4. **Completeness**: misura in cui il task include tutte le informazioni necessarie per l'analisi, senza ambiguità o incoerenze.

[ENG]
Evaluate the LLMs-generated tasks in term of:
1. *Functionality-oriented*: the task specifies which functionality should be examined by the tester without mentioning which actions should be performed to reach the objective.
2. *Utility*: relevance of the task for the usability assessment of the system with respect to its core functionalities.
3. *Precision*: language clarity and level of detail of the task.
4. *Completeness*: extent to which the task mentions the necessary pieces of information for the analysis without inconsistencies.

16. Examine a set of cells and confirm if the system's identification of each cell type is accurate. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

17. Review a set of cells that the system has identified as belonging to a specific type and determine if the classification is correct. If not, change the classification to the appropriate type and explain why the original classification was inaccurate. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

18. A cell has been identified as a specific type. Verify the classification and, if necessary, reclassify the cell and provide a reason for the change. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

19. Consider a set of cells and decide whether the system's identification of each cell type is accurate. If a cell is misidentified, change the classification and provide a justification for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

20. Review a set of cells and determine if the system's identification of each cell type is accurate. If a cell is misidentified, reclassify the cell and provide a reason for the change. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

21. Evaluate a set of cells and confirm if the system's identification of each cell type is accurate. If a cell is misidentified, change the classification and provide a justification for the correction. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

22. A cell has been identified as a specific type. Verify the classification and, if necessary, reclassify the cell and provide a reason for the change. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

23. Review a set of cells and determine if the system's identification of each cell type is accurate. If a cell is misidentified, reclassify the cell and provide a reason for the change. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

Set 3 of Tasks Generated for Anonymous

[ITA]
Valuta i task generati dagli LLM in base ai seguenti criteri:

1. **Functionality-oriented**: il task specifica quale funzionalità dovrebbe essere esaminata dal tester, senza indicare quali azioni eseguire per raggiungere l'obiettivo.
2. **Utility**: rilevanza del task per la valutazione dell'usabilità del sistema rispetto alle sue funzionalità principali.
3. **Precision**: chiarezza del linguaggio e livello di dettaglio con cui è descritto il task.
4. **Completeness**: misura in cui il task include tutte le informazioni necessarie per l'analisi, senza ambiguità o incoerenze.

[ENG]
Evaluate the LLMs-generated tasks in term of:
1. *Functionality-oriented*: the task specifies which functionality should be examined by the tester without mentioning which actions should be performed to reach the objective.
2. *Utility*: relevance of the task for the usability assessment of the system with respect to its core functionalities.
3. *Precision*: language clarity and level of detail of the task.
4. *Completeness*: extent to which the task mentions the necessary pieces of information for the analysis without inconsistencies.

24. Check the classification of class "muciparous" cells and correct errors, if any. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

25. Identify cells misclassified in class "ciliated" and correct misclassifications and explanations, if any. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

26. Review the table displaying test results and verify that it accurately shows the type of cell, number of detected cells, reference range, and classification of their quantity. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

27. Browse through the classified cell images, categorized by confidence level (low, medium, high), and assess their accuracy. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

28. Select a cell image and validate its classification as “Correct” or “Incorrect”, and if incorrect, correct it by selecting a new class and providing reasons why the original classification was wrong. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

29. Use the dashboard to navigate through different sections of the system and ensure that it provides easy access to the main functionalities. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

BrainMed Description

[ITA]
BrainMed è un sistema basato sull'intelligenza artificiale che ha l'obiettivo di supportare i neurologi nell'individuazione della Malattia di Alzheimer. Permette al medico di caricare diversi tipi di dati (ad esempio, dati genetici tabellari, cartelle cliniche elettroniche, scansioni MRI). Quando il professionista fornisce questi dati in input, il sistema li analizza e restituisce un output che indica una possibile rilevazione della malattia.
L'output è accompagnato da due spiegazioni: una visuale, che mostra una mappa di calore dell'area cruciale del cervello utilizzata dal modello per effettuare la classificazione, e una descrizione testuale del processo di ragionamento. Il medico può richiedere ulteriori dettagli e chiarimenti sulla diagnosi e sulla spiegazione tramite un chatbot basato su intelligenza artificiale.
BrainMed consente inoltre di accedere alla storia clinica del paziente, visualizzando le diagnosi precedenti.

Puoi familiarizzare con BrainMed interagendo con il prototipo al seguente link: https://www.figma.com/proto/vJqDOtD6WATgoc-Q6GMKivO/BRAINMED?page-id=0%3A1&node-id=1-6431&starting-point-node-id=1%3A6431&t=omJTwGZLWzrLyO2t-1

[ENG]
BrainMed is an AI-based system that has the objective of supporting neurologists in detecting Alzheimer's Disease. It allows the doctor to upload multiple types of data (e.g., Genetic Tabular Data, Electronic Health Records, MRI scans). When the professional gives this data as input, the system analyzes it and provides an output that contains the potential detection of the disease. The output is accompanied by two explanations: a visual one containing a heatmap of the crucial area of the brain that the model used to make the classification and a textual description of its reasoning process. The doctor can ask for more details and clarifications about the diagnosis and explanation to an AI-based chatbot. *BrainMed* enables also to access the patient's history, visualizing patients' previous diagnoses.

Familiarize with BrainMed interacting with the prototype at the following link: https://www.figma.com/proto/vJqDOtD6WATgoc-Q6GMKivO/BRAINMED?page-id=0%3A1&node-id=1-6431&starting-point-node-id=1%3A6431&t=omJTwGZLWzrLyO2t-1

Set 1 of Tasks Generated for BrainMed

[ITA]
Valuta i task generati dagli LLM in base ai seguenti criteri:

1. **Functionality-oriented**: il task specifica quale funzionalità dovrebbe essere esaminata dal tester, senza indicare quali azioni eseguire per raggiungere l'obiettivo.
2. **Utility**: rilevanza del task per la valutazione dell'usabilità del sistema rispetto alle sue funzionalità principali.
3. **Precision**: chiarezza del linguaggio e livello di dettaglio con cui è descritto il task.
4. **Completeness**: misura in cui il task include tutte le informazioni necessarie per l'analisi, senza ambiguità o incoerenze.

[ENG]
Evaluate the LLMs-generated tasks in term of:
1. *Functionality-oriented*: the task specifies which functionality should be examined by the tester without mentioning which actions should be performed to reach the objective.
2. *Utility*: relevance of the task for the usability assessment of the system with respect to its core functionalities.
3. *Precision*: language clarity and level of detail of the task.
4. *Completeness*: extent to which the task mentions the necessary pieces of information for the analysis without inconsistencies.

30. Verify the correctness of the diagnosis along with its corresponding explanation. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

31. Confirm that the natural language explanation corresponds to the brain areas highlighted in the GradCAM image. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

32. Ask details about the provided diagnosis to the AI-based chatbot. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

33. Ask details about the provided diagnosis to the AI-based chatbot. Ensure that the system presents the patient's previous diagnosis correctly. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

34. Verify that the system allows neurologists to access and view the patient's full history, including clinical data, notes, and personal information. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

35. Confirm that the system allows neurologists to add new patient data such as personal information, clinical data, and doctor notes. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

36. Check the display of the patient's diagnosis and history after adding new patient data. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

37. Ensure that the system allows neurologists to update the patient's diagnosis and history details. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

Set 2 of Tasks Generated for BrainMed

[ITA]
Valuta i task generati dagli LLM in base ai seguenti criteri:

1. **Functionality-oriented**: il task specifica quale funzionalità dovrebbe essere esaminata dal tester, senza indicare quali azioni eseguire per raggiungere l'obiettivo.
2. **Utility**: rilevanza del task per la valutazione dell'usabilità del sistema rispetto alle sue funzionalità principali.
3. **Precision**: chiarezza del linguaggio e livello di dettaglio con cui è descritto il task.
4. **Completeness**: misura in cui il task include tutte le informazioni necessarie per l'analisi, senza ambiguità o incoerenze.

[ENG]
Evaluate the LLMs-generated tasks in term of:
1. *Functionality-oriented*: the task specifies which functionality should be examined by the tester without mentioning which actions should be performed to reach the objective.
2. *Utility*: relevance of the task for the usability assessment of the system with respect to its core functionalities.
3. *Precision*: language clarity and level of detail of the task.
4. *Completeness*: extent to which the task mentions the necessary pieces of information for the analysis without inconsistencies.

38. Examine a patient's assessment and determine if the system's conclusion aligns with your clinical judgment. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

39. Review the system's explanation and describe how it supports the conclusion. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

40. Describe how the visual representation of the brain relates to the system's explanation. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

41. If you have questions about the system's conclusion, ask the AI assistant for more details. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

42. Evaluate the usefulness of the AI assistant's response in clarifying the system's conclusion. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

43. Imagine a situation where you don't understand the reasoning behind the system's conclusion. Describe what information you would need from the system to fully understand it. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

44. Describe the process you would follow to determine if the system’s conclusion is applicable to this patient. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

Set 3 of Tasks Generated for BrainMed

[ITA]
Valuta i task generati dagli LLM in base ai seguenti criteri:

1. **Functionality-oriented**: il task specifica quale funzionalità dovrebbe essere esaminata dal tester, senza indicare quali azioni eseguire per raggiungere l'obiettivo.
2. **Utility**: rilevanza del task per la valutazione dell'usabilità del sistema rispetto alle sue funzionalità principali.
3. **Precision**: chiarezza del linguaggio e livello di dettaglio con cui è descritto il task.
4. **Completeness**: misura in cui il task include tutte le informazioni necessarie per l'analisi, senza ambiguità o incoerenze.

[ENG]
Evaluate the LLMs-generated tasks in term of:
1. *Functionality-oriented*: the task specifies which functionality should be examined by the tester without mentioning which actions should be performed to reach the objective.
2. *Utility*: relevance of the task for the usability assessment of the system with respect to its core functionalities.
3. *Precision*: language clarity and level of detail of the task.
4. *Completeness*: extent to which the task mentions the necessary pieces of information for the analysis without inconsistencies.

45. As a neurologist, I want to verify the correctness of the diagnosis along with its explanation, so I can trust the system's output. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

46. As a neurologist, I want to ask details about the provided diagnosis to the AI Agent, so I can clarify any doubts I have about the diagnosis. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

47. As a neurologist, I want to confirm that the natural language explanation corresponds to the brain areas highlighted in the image, so I can understand the system's reasoning. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

48. As a neurologist, I want to verify that the system accurately integrates tabular genetic data, EHR data, and 3D MRI brain scans to generate a diagnosis, so I can trust the system's results. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯

49. As a neurologist, I want to interact with the AI-based chatbot to ask further questions about the diagnosis, so I can clarify any doubts I have about the results. *

	1	2	3
Functionality-oriented	◯	◯	◯
Utility	◯	◯	◯
Precision	◯	◯	◯
Completeness	◯	◯	◯