Assessing Doctor and Patient explanation and information needs of Explainable Artificial Intelligence

Matas Kazokas (2592371)
University of Twente
BSc- Department of Psychology

M12 BSc Thesis PSY Second supervisors:

First Supervisor: Funda Yildirim

Second supervisors: Stephanie van den Berg, Maryam Haeri

January 25th 2024

APA 7th edition

Word count: 10845

Abstract

Artificial Intelligence (AI) has been proven to significantly increase healthcare instrument performance but faces reluctance from medical experts due to its 'black box' nature. Explainable Artificial Intelligence (XAI) aims to solve this problem by making AI models more transparent and tailoring explanations to specific user needs, including medical professionals and patients. This study explores how XAI can effectively communicate AI-based decisions in healthcare focusing on tailoring explanations to meet the user group‘s needs, thereby increasing the knowledge of user-centered needs in the medical domain. Applying a mixed model approach, participants first answered open-ended questions to identify their information needs, followed by a case scenario involving a low grade glioma, which was investigated using multiple choice questions. The thematic analysis identified nine themes, showing patient‘s interest in understanding their illness and treatment options, and the doctor‘s focus on background knowledge and differential diagnosis. The quantitative findings indicated a preference for long, descriptive mechanistic explanations for the patient, and brief, causal explanations among the doctor group, with the latter group expressing lower satisfaction overall. This research increases understanding of XAI in medicine, accentuating the importance of user-centered designs and the need for tailored explanations to meet the diverse expertise of different user groups

Keywords: Explainable artificial intellligence (XAI), medical, user-centered, causal explanations, mechanistic explanations.

Introduction

In the last decade, artificial intelligence (AI) has seen increased use in many areas, with one of the most important uses being its ability to aid in the medical sector. AI assists in various ways including health tracking (Wang et al, 2023), cardiovascular disease detection (Siontis et al., 2020) and improving medical imaging (Litjens et al., 2017), indicating its potential to become an inseparable part of the medical field. As in many other fields, however, the main issue of AI is its lack of ability to explain how it reached its results, also known as the black box problem (Minh et al., 2021, S Kim et al 2024., Miller, 2019). In order to tackle this challenge explainable artificial intelligence (XAI) has emerged as a novel solution. Not only is XAI's intention to improve the trust and reliability of AI (Minh et al., 2021) but to also explain actions and decisions to humans in a more understandable way (Kim et al, 2024). An important question then arises in XAI: to whom should these explanations be tailored to? (Ribera 2019). Given the significant knowledge gap of naïve and expert users in the medical field, different kinds of explanations should be considered for different types of stakeholders (Langer et al, 2021; Kim et al 2024). This study will look further into these differences.

The integration of AI into medical practices has been gradual, reflecting the cautious approach in the field. With its deep learning (DL) models AI has gained ability to evaluate vast amounts of data and also gained the capabilities to diagnose diseases, improve diagnostic accuracy and reduce the strains of doctor workflows (Kaul et al., 2020). For one, AI is improving the ability to detect diseases (Litjens et al., 2017). Breast cancer is a prevalent problem in many people's lives. Early detection of this disease is essential for safe recovery. In one of the studies AI with computer-aided diagnostics (CAD) and its DL model was able to detect breast cancer with 92% accuracy rate and decreased the radiologist work by 17%. (Litjens et al., 2017). In another study, AI showed its usefulness in health monitoring sensors (HMS). Medical sensors have widespread use in the health sector however, they run into the problem of noise, drift and difficulty extracting useful information from large amounts of data. However, with the development of DL, all of these limitations can be treated (Zhang et al., 2023). With the help of AI, HMS systems can now perform real-time monitoring, analyse vast amounts of data, and ensure privacy. All of this leads to AI-enhanced sensors to provide more reliable, intelligent convenient services in the medical field (Wang et al, 2023)

Black Box Problem

Although AI has proved its effectiveness in the field of medicine, many experts are still unwilling to put it into practice and many patients do not fully trust it yet (Rajpurkar et al., 2022). Steerling et al. (2023) found that trust associated with AI depends on many personal aspects such as age, gender, education level, general domain knowledge and technological skills. One of the main reasons for mistrust with AI is the black-box problem. Black box pertains to the fact that while many of AI generated answers might be accurate, humans do not know how these answers were derived. This becomes prominent in the medical field where failure to have accurate assessment and specific solutions to a case could have catastrophic consequences (Du et al., 2022). Failure to understand the reasoning of AI decision-making could lead to not only a failure to accurately understand the case and give proper treatment, but also an overreliance of the AI can negatively impact a patient's well-being in the long run (Du et al., 2022).

Explainable Artificial Intelligence

XAI is a field that intends to increase reliability, transparency and find out the inner workings of AI models (Holzinger et al., 2017, Langer et al., 2021). XAI has become such an important topic that even the U.S. Defence Advanced Research Projects Agency (DARPA) has created a program which seeks to solve this problem (Holzinger et al., 2017, Gunning 2019 et al.). This has led to a surge in the XAI community to rather focus on ML explanations (Zhou et al., 2021), as throughout the years the main focus of XAI has been to develop and increase the understandability of these models (Langer et al., 2021). Taking this into account the questions remain: for whom are these models being designed for? (Ribera 2019). Miller (2019) argues that the explanation for many XAI models is created for domain experts. While these domain experts have a good grasp of the topic, these explanations are not suited for lay users. Therefore, many researchers argue that in the domain of XAI a difference between lay users and domain experts should be researched (Kim et al., 2023, Langer et al 2021., S Kim et al., 2024). They argue that in order for this field to be reliable, a user-centred approach should be taken (Riberra and Lapedriza, 2019). A user-centred approach means that different metrics should be targeted for different users and that explanations should be tailored to the specific user needs (Riberra and Lapedriza, 2019, Miller 2021). First, we need to delve deeper into what is expected in XAI to constitute a good explanation.

Design Evaluation

An emerging trend in the XAI community is the focus on the design and evaluation of XAI explanations, yet there are still many questions on what is the best way to do this. In his comprehensive review, Minh (2021) argues that there are five terms which should be taken into account with XAI explanations: understandability, succinctness, comprehensibility, interpretability and explainability. Both comprehensibility and understandability pertain to the user being able to understand what the model learned. Succinctness pertains to the conciseness and compactness of the explanation and how well it can be understood by the person. Interpretability indicates on what level the user can understand the concept with base knowledge, and finally explainability refers to the methods used by AI to clarify its internal workings. Taking these explanations, we can further interpret what might be expected from users.

Patient Medical Information Needs

In their systematic review Jia et al. (2021) outlined the need to tailor online medical information to adhere to patients' needs. Because information technology and social media are gaining a lot of popularity, many people seek medical advice online. It was found that in some countries as many as 90% of people participated in online health information seeking behaviour (HISB). Additionally, in their study Zhang et al., (2019), found that individuals with a prevention focus, characterised by a desire for safety and risk avoidance, predominantly engage in conservative health information-seeking behaviours (HISB). This group tends to adhere to routine and safety-oriented health information. Finally, in the research Xiong et al., (2021) found that for Chinese people the top 4 information types looked for online were healthy behaviours which constituted wellness activities, such as exercise and fitness, along with dietary and nutritional habits, medical concerns such as advice on medications and disease-related consultations, Traditional Chinese Medicine (TCM), often pursued by Chinese individuals and (4) health science popularisation, which focuses on enhancing public health knowledge.

Explainable Artificial Intelligence Question Bank

Vera et al., (2020) delves deeper into the tailored needs of naïve users and medical experts in AI explainability. In this paper the researchers talk about an Explainable Artificial Intelligence question bank (XAIQB) which seeks to fill the intricate role of AI decision making process and user understanding by creating user-centric questions (Sipos et al., 2023). The XAIQB tested users' needs by first eliciting questions to gather answers. After that the questions of the participant were analysed. This approach tailors to user specific needs and makes access to the use of AI not requiring much expertise (Liao et al., 2020). From the XAIQB it was concluded that firstly, the researchers outline that one of the first steps of the design process should be to identify specific user needs. This makes the process of designing specific case scenarios more intuitive and tailored. Secondly, the researchers gave a better understanding of what questions might arise for users in situation specific scenarios. For example: from their research it was found that all participants ask “why” questions, which seems to be the most important answer they would want to hear from AI. Thirdly, participants were also interested in asking “how should this instance change to get different predictions ”which in this case pertained to “what can I do to reduce the risks…”. From this paper it was identified what type of questions might be relevant to both patients and doctors.

Explanation Types

Different explanation types aim to satisfy the requirements of specific user needs depending on the user and context. Many researchers argue that in order for users to comprehend AI explanations, the answer should be tailored to stakeholder needs (Kim et al., 2023, Langer et al 2021., S Kim et al., 2023). There are several explanations that should be considered when taking this into account.

In his research, Van Der Waa et al. (2021b) creates a comparison between a rule-based and example-based contrastive explanations to evaluate their efficacy in system understanding, task performance and persuasive power for diabetes management. In this experiment rule-based explanations would pertain to questions as “why this instead of this” and example-based explanations used previously written cases for the users to replicate the action. This research not only gave insights into what kind of research designs XAI researchers should take but also gave an understanding of possible explanation types. Comparing the two explanation types showed that rule-based explanations gave more system understanding to the users. Another researcher Markus et al. (2021) proposed other explanation models: model-based, attribution-based, and example-based. Model-based explanations provide transparency directly or through simpler models. Attribution-based methods assess input feature significance, crucial in domains like healthcare. Example-based explanations use scenarios to make AI decisions relatable to the users. Additionally, they discuss post-hoc explanations, which offer insights into AI models after which predictions are made, which are important for understanding complex models without compromising their performance. This framework aids in selecting XAI methods, balancing AI complexity with the need for clear, understandable outputs. From this research it was gathered that there are notable ways of making XAI more interpretable and transparent for models. These papers were used to conceptualise this study's explanation types. Compared to the mentioned explanations models, this research did not focus on persuasion but rather understanding, therefore the explanation types came to different conclusions.

Counterfactual explanations consider alternative scenarios, like “What if different symptoms were present?” Counterfactual explanations are considered to be “good” explanations by many researchers (Wachter et al. 2017a, Mittelstadt et al., 2019). Not only can these explanations broaden the possibilities by imagining other outcomes, but they are also a way for users to understand easily and use the information in real world practice (Wachter et al. 2017, Tian et al., 2022). In their paper about AI explanations in drunk driving situations, Warren et al. (2022) found that counterfactual explanations were much more preferred than causal or no explanations at all, because of this, this was the first explanation type. Following this, simple causal explanations describe the cause-and-effect relationship that forms the model's predictions. In the medical field a causal explanation would be "This activity leads to this". According to Van Der Waa (2021) in some cases causal explanations can show better results than counterfactual explanations in understanding the system explanations. Additionally, Lombrozo (2007) in her research argues that simpler, "prettier" and explanations are preferred by participants. This is due to the fact that shorter explanations are usually easier to understand and it additionally pertains to the principle of Occam's Razor in psychology, which states that simpler theories are to be preferred over more complex ones (Baker 2005). Conversely, contrastive explanations provide not only why the event happened but also why it happened instead of another event. Contrastive explanations make predictions, like asking "Why this diagnosis instead of that one?" (Van Der Waa et al., 2018). These provide an argumentative side to the explanation. The strength of these types of explanations is the ability to build on the user's perspectives and epistemic state, expanding the understanding of causal relations without being overwhelmed with unnecessary detail. As causal explanations they adhere to simplicity by providing the most relevant new information (Miller, 2018). As opposed to simple causal explanations, mechanistic explanations offer broader abilities of knowledge by providing more in-depth knowledge into workings of a model (Craver & Kaplan, 2020). Mechanistic explanations similarly to descriptive explanations provide more context of how a result was achieved (Elton 2022). The reason why mechanistic explanations might be preferred over other types is their ability to extensively describe the situation taking into consideration all the parts of the model (Craver & Kaplan, 2020). However, as many researchers have mentioned, the ability to be too extensive and descriptive might be a limitation to some users (Craver & Kaplan, 2020). We used the previously mentioned explanation types to create arguments for our explanation types in the survey.

This research aims to find what type of information doctors and patients want to know from an AI diagnosis system and how doctors and patients differ in their explanation needs of medical illnesses. Due to there being a huge knowledge gap between medical experts and lay users (Langer et al, 2021; Kim et al 2024) it is expected that different AI explanations should be tailored to each group's needs (Ribera and Lapedriza, 2019). We uncovered that there are four explanation types that both of the stakeholder groups might prefer: causal explanations, counterfactual explanations, contrastive explanations and mechanistic explanations. From our quantitative research we aim to explore what kind of information both doctor and patient groups would like to know from AI and whether these information needs differ. Secondly, we performed a quantitative study to find out what type of explanations both doctor and patient groups preferred in diagnosis. This was followed up by a satisfaction analysis to see which groups found the answer types more adhered to their needs and if the satisfaction for these groups differed. This research will give insight into XAI systems in order to increase the tailored knowledge that could be offered to the medical domain. These were the research question posed for the study:

RQ1: What specific information do doctors and patients seek from AI-based diagnostic systems, and how do these information needs differ between the two groups?

RQ2: Which types of AI explanations (causal, counterfactual, contrastive, mechanistic) are preferred by doctors and patients?

RQ3: Does the satisfaction level regarding AI-based diagnostic explanations differ between doctors and patients?

Methods

Participants

Data collection was conducted through the online platform Qualtrics. Participants of multiple nationalities were recruited using convenience and snowball sampling. The researcher used social media and other means to distribute the survey. In order to participate as a doctor, a minimum of a PhD was required, there were no requirements for the patient group Additional participants, mostly first- and second-year psychology students, were recruited through the SONA-system of University of Twente. Credits were awarded to students who participated. A sample size of 48 consisting of 24 doctors and 24 patients was aimed at and calculated using the free statistical software G*Power (Faul, 2007). All participants were fluent in English and provided informed consent before participating in the study. The research was approved by the BMS ethical committee / Domain Humanities & Social Sciences, decision number 231307.

Case Scenario

In this study a survey was presented to the participants in the form of a medical case scenario. As outlined by Wolf (2016) case scenarios can be a promising way of constructing user-centred design. The scenarios were different for both doctor and patient types due to their different levels of expertise (Langer et al, 2021; Kim et al 2023). It was conceptualised that in order to explore the topic XAI more accurately the patient group would be presented with more simple layman explanations and the doctor group would receive information consisting of complex medical terminology. In order to fit these requirements a fitting case scenario was required. For a fitting case scenario, two medical articles were put together. For this study the patient description was taken from a case study of a Glioblastoma (Pa-C & Pa-C, 2023) and the medical images and further results were taken from a case of brain glioma progression (Yang H. 2016). Accordingly, the case scenario was kept similar to the original for the doctor group and simplified by reducing the complexity of the expressions for the patient group.

Design and Procedure

An informed consent form (University of Twente, 2022) was presented to the participants after which a brief demographic form was used to capture the participants age, gender, educational background and professional experience (for medical experts).

To answer the research questions a mixed methods approach was undertaken in the study. In order to find out what type of medical information the participants wanted to get from the AI assisted doctor a qualitative study was performed with an open question. Similar to what Mihn (2024) did in their study, the goal of the quantitative part of the study was done in order to explore what type of information both the patient and doctor group would want from the AI assisted doctors. The open question was tailored to specific participant groups. In the study the non-doctor group received a question in which they had to imagine themselves having some sort of symptoms of an illness, where they were then asked to describe what type of information, they would like to receive in order to fully understand their case (see Appendix 1A). For the doctor group the participants were to assume the role of a doctor. Following this they were asked to imagine that their patient had some sort of symptoms of an illness and were inquired what type of information they would like to receive in order to fully understand their patient's case (see Appendix 2A).

After the case scenario, the case scenario was presented to both the patient group (see Appendix A3) and the doctor group (see Appendix A4) accompanied by the MR photos (see Appendix A5). Following the first half, 10 multiple choice questions were presented to the participants. Due to a lack of concrete information about what type of information and questions people usually ask their medical practitioners, different means were taken to ascertain this knowledge. The widespread forum website Reddit (Reddit, 2023) was used to do this. In the subreddit r/braincancer an inquiry to hear about peoples experience with glioma was asked (Trickygap4750, 2023). Appendix D shows the posed query and comments of the thread. From there several assumptions and ideas were presented to the researchers about what type of information people might ask their doctors was gathered. Additionally, the XAIQB

(Vera, 2020) was used for its theoretical baseline. Accordingly, a logical sequence of questions was constructed by the researchers assuming the role of a patient who would be interested in finding more about their symptoms was done. In the first part, prototypical questions about the symptoms such as: “Which of my symptoms most directly suggest an issue in the left frontal region of the brain?” about the illness “Based on the MRI and clinical presentation, what is my preliminary diagnosis?” and about treatment “What treatment options are typically considered for a diagnosed low-grade glioma?” were used in the survey (see Appendix A6). Afterwards, the second case scenario for patients (see Appendix 2A) and the doctors (see Appendix 2C) accompanied by a different MR photo (see Appendix 3C) was presented for both the groups accordingly. Another 8 questions were posed to the participants. After finishing answering the questions, participants were shown an ending screen thanking them for their participation and inquiring them to share their thoughts and feedback about the survey.

Additionally, following every multiple choice question a Likert scale was presented to the participants to rate how satisfied they were with the answers. The Likert scale was measured from 1-5, with 1 being the least and 5 being the most satisfied, to find and was used to find out if the designed answers were fitting the group and which group had a higher satisfaction score.

Answer Types

For this survey a multiple-choice question format was used in the form of answers. After asking prototypical questions the participants would be presented with four answer choices. All the answers would provide similar information to the participants; however, they would be formed differently. As outlined before, the answers would be grouped into causal explanations, which in this survey were short and concrete cause and effect explanations. A typical causal answer would be the shorter out of the 4 choices and sounded like “Your trouble with the right side of his body hints at a problem in the left side of his brain. “. Following the second answer choice was the counterfactual explanation. These explanations provided answers that usually started with “If this then that” a counterfactual explanation would sound like “If the MRI didn't show the lesion, we might not understand the cause, but this spot helps explain your symptoms.” Next up were the contrastive explanations. These types of explanations usually started with words like “unlike”, “contrary” and “compared to”, an explanation of this kind sounded like “Compared to other possibilities, the specific signs in your MRI and symptoms point to a low-grade glioma.”. Finally, the last and the longest type were the mechanistic explanations. These answer types were outlined by extensive and detailed description of explanation. A mechanistic explanation would sound like “Serial measurement of serum and cerebrospinal fluid (CSF) markers, such as glial fibrillary acidic protein (GFAP), can provide insights into the extent of glial cell turnover and the integrity of the blood-brain barrier, which are critical in the management and prognostication of gliomas."

Next, a scenario-based survey was presented to the participants. The survey consisted of two parts. In the first part a tailored open question was posed for the patient group (see Appendix A1) and the doctor group (see Appendix A2). Next a detailed case scenario was provided to both naive users and medical experts. The medical case presented to the patient group, had them assume a role of a patient receiving a diagnosis and was designed in such a way that it could be understood by individuals possessing no medical knowledge (see Appendix A3). The case scenario shown to the medical experts contained more medical terms (see Appendix A4) and had the experts assume the role of a doctor receiving an anamnesis of a patient. Alongside the case scenarios both the experts and naive users received an MR image containing a possible low-grade glioma acquired from different angles (see Appendix A5). The same was done for the second part of the study where naive users (see Appendix B1) medical experts (see Appendix B2) once again received a tailored medical case scenario. The scenarios in part two continued with the same patient that was in part one, with the patient having worsened symptoms after 3 months. Additional MR images accompanied the case scenarios now containing images of possible high-grade gliomas (see Appendix B3). Following all the scenarios a survey was presented to the users (see Appendix A6, Appendix B4). The survey contained questions that were made to replicate the type of questions a human would ask a doctor, to find out about his symptoms. An open question to gauge what type of explanations might the user expect from AI was asked in the beginning and afterwards a total 18 multiple-choice questions were divided into 2 parts. Additionally, after each question participants also indicated their satisfaction with the answer from a 1-5 Likert scale.

Data Analysis of the Qualitative Study

In order to explore what type of information participants would like to receive from an AI based diagnostic system a qualitative text study was conducted. This part of the study was in the beginning of the questionnaire and contained one open question with a minimum of 100 characters to gather participant insights. Firstly, the entries were gathered in the online platform Qualtrics. Once the doctor and patient group text entries were gathered, they were then manually transferred into a word file to facilitate the analysis. Transferring the data from Qualtrics, a number of variables were kept such as: Nationality; Gender; Age; Answer to the questions to gather better insights into possible similarities. Following this, the data was then transferred into the qualitative data analysis tools ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, 2023). From there, a thematic analysis was done for both of the groups in order to gain insight into what type information the participants would like to hear about their/their patient's illness. The goal of the thematic analysis was to identify patterns and similarities from the participants text data (Maguire, 2017). Additionally, because there were no preconceived themes, a general 6 step plan created by Braun & Clarke (2006) for exploring qualitative data, was done, this method is also known as the inductive way of doing qualitative analysis (Azungah, 2018). The 6 steps were: 1. Becoming familiar with the data 2. Generate initial codes 3. Search for themes 4. Review themes 5. Define themes 6. Write-up. Because there were 2 separate things, this process was done twice.

Data Analysis of the Quantitative Study

The primary objective of the data analysis was to determine the preferred type of explanations (causal, contrastive, counterfactual, mechanistic) for both medical experts and naive users. Additionally, the aim was to assess the satisfaction level of participants with the provided explanations and identify any significant differences in preferences between the two groups. In order to do so the statistical programming language R-4.3.1. with the interface R Studio (Posit team, 2023) was used. First the data was imported from Qualtrics. Afterward data cleaning was done in order to keep only the necessary variables for the quantitative study, such as the multiple-choice question answers and Likert scale answers. Respondents who completed fewer than 50% of the questions were excluded from the data set and variables. Once this was done the demographic data of the study was compiled together. Following this the imported variables had to be renamed to maintain data structure. Next the text answers from the questionnaire were turned from character into numeric variables. The text answers were appropriately coded 1 - for causal explanations, 2 - for counterfactual explanations, 3 - for contrastive explanations and 4 - for mechanistic explanations. This was done in order to have better ability to compile the data, missing values were excluded. In order to find out the central questions of what type of explanation types each of the 2 groups preferred, a frequency analysis was conducted. This dataset was used to compare the responses between the two groups. In order to get better insights and to test whether the explanation types were different for both groups further statistical inferences were made. The next step was to check the frequency of each of the questions and the normality of data (Miot, 2017). Further to check the differences of the two groups a one-way ANOVA and a following t-test test was done for the two groups. For the data that did not have normality, a non-parametric Wilcoxon signed-rank test was performed. After this satisfaction analysis was done to assess whether being a doctor or student was associated with higher satisfaction levels. To this end, a dataset was constructed, aligning multiple choice responses with corresponding Likert scale satisfaction scores. Average satisfaction scores for each answer type were calculated, to obtain into the levels of satisfaction associated with different explanation types.

Results

The total sample consisted of N = 56 participants. The student group consisted of 30 (56%) individuals, 12 (40%) male and 16 female (60%) with participants ranging from 21 to 27 years old (M = 21.9, SD = 2.24). For this group the sample consisted of 11 different nationalities with German respondents being most common (40%), followed by Dutch (16%, then Turkish (10%), Russian (6.7%), Lithuanian (6.7%) and others (16%). The doctor's group was comprised of 26 individuals (44%) 10 being male (38%), 14 female (54%) and 2 (4%) non-binary individuals with the youngest being 27 and oldest 61 years old (M = 40.7, SD = 9.95). The respondents of this group consisted of 6 nationalities with the majority being Lithuanian (57%), followed by Dutch (19%), then Turkish (11%), and German (4%), Romanian (4%) and Chinese (4%).

Qualitative Research Results

The goal of the qualitative part of the study was to gather doctor and patient insights about what type of information each group would want to know from an AI doctor assistant. Additionally, it was looked at the kind of similarities and differences were there between the groups. For this a thematic analysis of the survey's open questions entries was performed. The open question text entries were compiled and analysed to gather insights. 9 different coding themes emerged. Coding themes that were mentioned less than 5 times in either of the groups, were not included in the results. The codes in mention consisted of: Illness information/diagnoses, long term concerns, real life doctor, severity assessment, symptom clarification, treatment and remedies, diagnostic tests, history and background. As the themes were first coded separately for the doctor and patient group and then together, some of the themes only have entries for one group. Below is a table showing the codes and that frequency of the mentioned codes for each group.

Table 1

Coding scheme for the qualitative survey results

Patient and doctor group coding results		
Name of the code	**The number of times mentioned by patients**	**The number of times mentioned by doctors**

Illness information/diagnoses	**9**	**10**
Long term concerns	**5**	**0**
Real life doctor	**6**	**0**
Severity assessment	**8**	**0**
Symptom clarification	**17**	**13**
Treatment and remedies	**16**	**5**
Diagnostic tests	**0**	**9**
History and background	**0**	**12**

Illness Information/Diagnosis

This code expressed the participants' interest in gaining information about the potential illness that they might be suffering from. Additionally, participants wanted to get details about what having this illness might entail and any specific symptoms associated with the said disease. This code outlines the participants' need for managing their health condition by having comprehensive knowledge of an illness, its conditions and possible differences in identifying it. A typical code from the patient group sounded like:

- "What potential illness I could have" (23 y/o, Female, German)
- "Then, it is important to me what kind of illness the AI assistant wants to show me, with exact description and possible different levels and types" (22 y/o Female, German)

For the doctor group the focus was more on the illness diagnosis. They were more interested in asking AI to perform differential diagnosis which is a measure to compare different diagnoses and how likely each of them might be to occur (Lamba et al., 2021). Some examples of this would be:

- "The main goal would be to perform a differential diagnosis of the patient's symptoms." (54 y/o, Female, Lithuanian, Family Medicine)
- "...domains to touch on based on differential diagnosis based on the symptoms" (29 y/o, Male, Dutch, Psychiatry)

Long Term Concerns

This theme was described by the participants need to know about what possible long-term consequences a disease would entail for them. They wanted to understand what kind of consequences could lead from an illness and what the complications could arise. Additionally, participants were interested in hearing what their day-to-day life could spell for them and what possible things they should be avoiding. Some sentences describing this code would be:

- "What are the long-term consequences of this illness ….. in case there is no option to be cured permanently, what is the best and worst case scenario. Also, the issue of day to day life - how will the treatment or the illness itself affect eating, movement, libido and the ability to work." (20 y/o, Female, Polish)
- "What do I need to look out for, following this diagnosis?" (27 y/o, Male, Turkish)

The doctor group did not mention this code in their responses.

Real Life Doctor

A common response from the participants was their inquiry about when they should see a real-life doctor. There seemed to be a concern among many of the participants of whether their symptoms were serious enough to visit a medical GP. This was expressed in a couple of ways, however, generally the participants wanted to know whether following their diagnosis an opinion of a real-life doctor should be sought after. Some phrases of this code sounded like:

- " I would also like to learn about differential diagnoses, when to consult a doctor, what are the possible outcomes." (24 y/o, Male, Turkish)
- " I would like to know at the very least whether symptoms are serious enough to get checked by a proper doctor or if I can ignore them" (21 y/o, Male, Irish)

One participant expressed their antipathy for AI an assisted doctor claiming that they would:

- "I would not use AI. I do not trust AI to give me health-related information and/or advice. AI is not a substitute for medical advice from a human (e.g. GP)" (22 y/o Female, Dutch)

None of the participants in the doctor group mentioned this theme.

Severity Assessment

This theme was found for participants which were worried about the possible severity of their illness. The participants expressed a need to know how serious their symptoms were and if they might be fatal. Additionally, some participants wanted to know whether having previous illnesses would make them more likely to be at risk. Some of these codes also were tied with the previous of "real life doctors" as participants were also curious if their symptoms were severe enough to warrant a visit to a real-life doctor. Some sentences for this group:

- "....an indication of severity for the symptoms - an indication of the urgency to see a doctor - an indication of potential danger." (23 y/o, Male, German)
- " I would also appreciate some information regarding some risk factors that apply to specific groups such as : smokers or people with prior lung illness should definitely see a doctor if they have the illness e.g" (24 y/o, Male, German)

Symptom Clarification

One of the most common concerns for the student groups was the need to have more clarification about their symptoms. An interest was shown to know clear explanations of their symptoms and what they might lead to. Others were curious about knowing where their symptoms came from and what these specific symptoms mean. In general, the patients wanted to get a full picture of their illness by means of clear symptom clarification. Some sentences for this code were:

- " I would like to receive a complete picture of all the symptoms the illness includes so I can capable of making a good guess on my potential to have the disease and other symptoms as well" (24 y/o, Male, German)
- "I'd want to know what the possible causes of my symptoms could be" (26 y/o, Female, Iranian)

Many of the doctors found similar interests for this theme; however, their concerns were different. The medical experts were more interested in hearing from the AI for a more detailed description and characters of the symptoms. Some of the codes for this theme sounded like:

- "Whether the patient feels muscle weakness, strong headache, especially in the morning? Does he feel nausea or he is vomiting?...." (61 y/o, Female, Lithuanian Surgery)
- "When it started How long does it goes Any assosiation with feeding Any spesific time frime Associated with bowel movements Is it constant or permanent" (37 y/o, Female, Turkish, Obstetrics and Gynaecology)

Treatment and Remedies

Another common theme among the participants expressed a need to know more about the possible treatment options. This code compiled the students wish to know what their treatment might entail, what they could do to reduce symptoms, medical prescriptions or alternative medicine. Additionally, they were curious to know how they could prevent the disease, what alternative medicine options could help simply alleviate their symptoms. Some sentences for this code sounded like:

- “.... how I could relieve the symptoms.” (19 y/o, Female, Dutch)
- “ It would also be helpful to learn about any treatments or remedies that could help alleviate my symptoms.” (26 year old female Iranian)
- “...small things you can do from home when for example catching a cold.” (19 y/o, Female, German)

The doctor group while not as many as the student group also expressed this theme however, only briefly:

- “treatment options; countries with different treatment options” (43 y/o, Female, Lithuanian, Ophthalmology)
- “Did you do anything to try to relieve the symptoms?” (50 y/o, Male, Dutch, Internal Medicine)

Diagnostic Tests and Results

A similar theme among the doctor group was the need to know more about possible tests and results. Here the medical experts expressed a wish to get information about the patients lab results such as: imaging results, details about molecular or genetic tests, samples of tissue with description and analysis. Additionally, the doctors wished to know what possible diagnostic tests they could perform. Some examples of this code sounded like:

- “Computed and magnetic resonance images with descriptions, specific to this desease. Algorithms for detection.” (48 y/o, Female, Lithuanian, Emergency Medicine)
- “Decision support for imaging or other tests. Expected diagnostic yield. Analysis of imaging tests. (48 y/o, Female, Dutch, Diagnostic radiology).

The patient group did not mention this code in their text entries.

History and Background

The last code to be found in the analysis was history and background and it expressed the doctors need to know more about the patient's previous symptoms anamnesis. Here the doctors were interested in other variables that may impact the patient’s current health such as their age, sex, nationality, whether they drink alcohol or if they used any previous medication. Codes for this theme sounded like:

- “Age, sex, complaints, assessment of general, neurological and physiological conditions. It is important that the entire anamnesis is collected very accurately (52 y/o, Female, Lithuanian, Physical Health and Rehabilitation)
- “Do you smoke? How often and how much alcohol do you drink? Are you taking drugs? What is your previous medical history? What medication are you currently taking? (50 y/o, Male, Dutch, Internal Medicine)

The patient group did not express a need to know this.

Quantitative Research Results

In order to find out what type of explanation doctors and patients prefer a frequency analysis was performed. The results show that both patients and doctors vary in their preference. Figure 1 displays the compiled picked preference for the patient group. In total there were N = 586 total responses for the patient group. The bar chart reveals that the most frequent response type was mechanistic explanations N = 290 (49%%), followed by causal explanations N = 136 (23%), counterfactual explanations N = 87 (14%) and contrastive explanations N = 73 (12%).

Frequency Analysis

Figure 1

Distribution of responses from patients

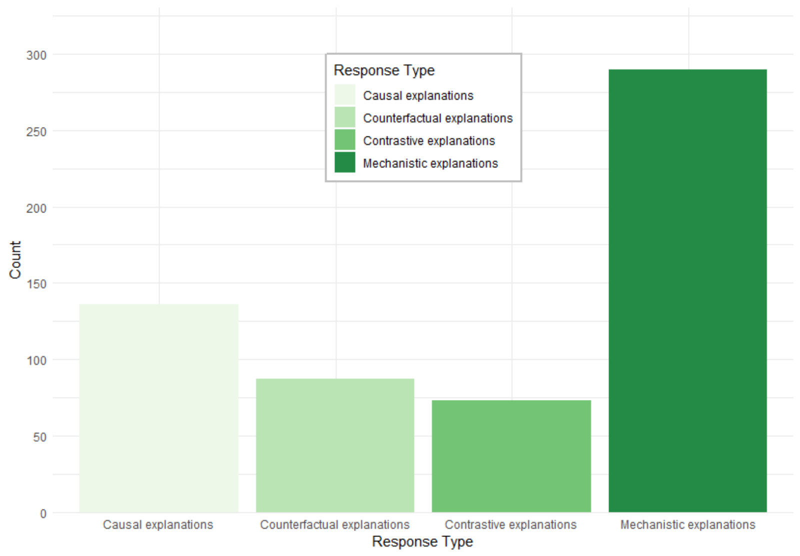
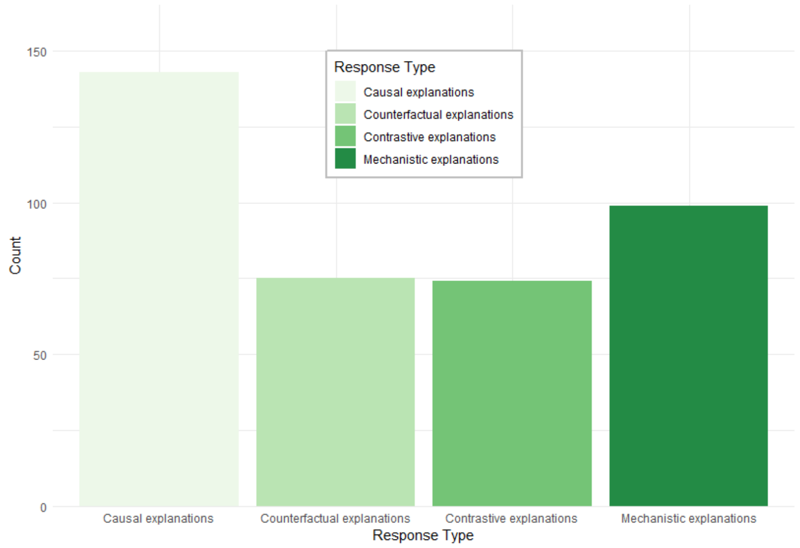


Figure 2 displays the distribution of the doctor group. In total there were N = 391 responses for the doctor group. The bar chart reveals that the most frequent response type was causal explanations N = 143 (37%), followed by mechanistic explanations N = 99 (25%), counterfactual explanations N = 75 (19%) and contrastive explanations N = 74 (18%).

Figure 2

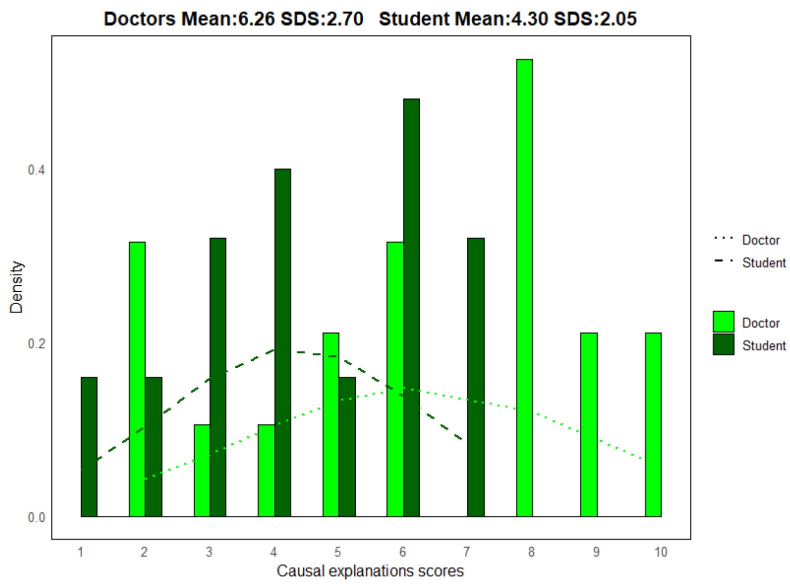
Distribution of responses from doctors



To check whether preferences indicated by the frequency analysis reflected significant differences between doctors and patients, an ANOVA test was carried out. Prior to this, normality of the data was computed, the 8 total distributions of each of the answer types for the 2 groups (see Appendix C). Figure 3 displays the distribution scores of causal explanations for both groups, with a density line.

Figure 3

Density plot of causal explanations between the 2 groups



The Shapiro-Wilk test was performed in order to further check the normality assumptions. The analysis indicated that the distribution of scores for causal explanations did not significantly deviate from normality for both doctors (W = .929, p = .081) and students (W

= .943, *p* = .116). In contrast the test for counterfactual explanations showed non-normality for both the doctor (W = .917, p = .033), and the patient group (W = .889, p = .004). Similarly, contrastive explanation scores significantly deviated from normality for both groups: doctors (W = 0.896, p = .001) and students (W = 0.836, p = .003). Lastly scores for mechanistic explanations were normally distributed for doctors (W = .952, p = .242) and students (W = .967, p = .467), suggesting that the assumptions of normality were met for these variables.

Scores for causal explanations and mechanistic explanations showed normal distribution therefore an ANOVA to test whether there is a difference in mean scores for the population (T. K. Kim, 2017). For counterfactual explanations and contrastive explanations, the Shapiro-Wilk test showed deviation from normality therefore non-parametric test Wilcoxon signed-rank test was used to check whether there is a difference in the means between the two groups (Whitley & Ball, 2002)

Counterfactual and Contrastive Explanations

The Wilcoxon rank sum test was conducted to compare the median scores of counterfactual explanations between doctors and students. The results indicated no significant difference between the two groups (W = 307.5, p = .157). Therefore, we fail to reject the null hypothesis, suggesting that the median scores for counterfactual explanations do not differ significantly between doctors and students. For counterfactual explanations, the Wilcoxon rank sum test revealed an insignificant difference in median scores between doctors and students (W = 352.5, p = .0133). This result leads to the not rejection of the null hypothesis, indicating a statistically significant difference was not present in the median scores for counterfactual explanations between the two groups

Causal and Mechanistic Explanations

The ANOVA test was performed for causal and contrastive explanations. The analysis revealed a significant effect of the group on causal explanation scores, F (1, 55) = 7.609, p = .008. The mean score for doctors (M = 6.26, SD = 2.70) was significantly higher than for students (M = 4.31, SD = 2.05). From this we can infer that the groups have different means in the population. The ANOVA for mechanistic explanations scores demonstrated a significant effect of group, F (1, 55) = 55.06, p < .001. Mean scores for doctors (M = 3.77, SD = 2.37) were significantly lower than for students (M = 8.33, SD = 2.73). After finding that there was a mean difference for the two groups, a t-test was performed to gauge the group mean difference (Mishra et al., 2019).

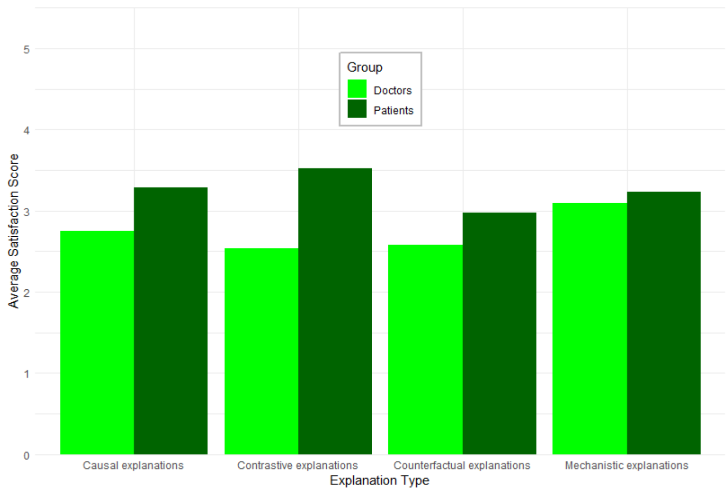
The Two Sample t-test revealed a significant difference in the mean scores of causal explanations between doctors and students, t (32.252) = 2.643, p = .013. The mean score for doctors (M = 6.26) was significantly higher than for students (M = 4.31), with a 95% confidence interval for the mean difference ranging from (0.449 to 3.462). For mechanistic explanations, the Two Sample t-test indicated a significant difference in the mean scores between doctors and students, t (54.944) = -7.472, p < .001. The mean score for students (M = 8.33) was significantly higher than for doctors (M = 3.77), with a 95% confidence interval for the mean difference ranging from (-6.41 to -3.700). These tests conclude that there is a difference between the mean scores between the two groups.

Satisfaction Analysis

Lastly it was checked whether there is a significant difference between the satisfaction scores of the 2 groups. Figure 4 shows the distribution of the average satisfaction scores per answer type for both groups. The satisfaction scores were scored from 1-5 Likert scale (1 lowest – 5 highest). On average, doctors reported lower satisfaction scores for each type of explanation. Specifically, doctors scored causal explanations M = 2.75 and mechanistic explanations at M = 3.08, while patients reported slightly higher scores of M = 3.28 for causal and M = 3.23 for mechanistic explanations, respectively. Additionally, the student group showed higher satisfaction scores in the other 2 explanations types as well, scoring counterfactual explanations at M = 2.97 and contrastive explanations at M = 3.52, while the doctor group scored counterfactual explanations at M = 2.57 and contrastive explanations at M = 3.52. These results provide a quantitative foundation to address RQ3 and indicate that doctors may have different expectations or requirements from AI-based systems compared to patients.

Figure 4

Average satisfaction scores between groups



Discussion

This study investigated what information doctors and patients sought from an AI-based diagnosis system and how these information needs differed. Additionally, it was examined what type of AI explanations were preferred by doctors and patients and if the satisfaction levels differed between the groups. In the complex and evolving field of AI in medical diagnostics, this research aimed to increase the domain knowledge of XAI, particularly emphasising the importance of a user-centred approach in designing AI explanation models. This discussion will look into these aspects, analysing the implications of the findings for tailoring XAI systems more effectively for each user group need in the medical domain.

The qualitative part of the study delved into the open-ended responses from both doctors and patients to understand their specific information exigency from AI-based diagnostic systems. The thematic analysis, as outlined by Braun & Clarke (2006), revealed distinct patterns in the information needs sought after by each group, uncovering the differences and similarities of their needs and perspectives.

Patients predominantly sought clarity of their symptoms and the possible treatment for their illness (see Table 1). These findings partly aligned with the literature emphasising the general population's desire for comprehensible health information (Xiong et al., 2021). Patients expressed a multitude of different needs but mainly they expressed their preference for hearing straightforward, relatable information that could help them understand their health condition and how to alleviate them. This build upon the research by Kim et al. (2024) where they found that the primary reason why patients sought explanations needs, were to gain motivation through health management. Additionally in the treatment and remedies section participants wanted to know what they could do to remedy an illness, with one respondent requiring information about “...small things you can do from home when for example catching a cold.” (19 y/o, Female, German) this also aligned with a similar concept of people seeking traditional Chinese medicine (TCM) as Xiong et al., (2021) outlined. It appears that the patient group saw the AI system as a sort of pre-primary doctor which could aid them in answering questions about possible diagnoses, in order to gain confidence in controlling and treating their illness. The desire for symptom clarification and treatment options were consistent with studies highlighting patients' need to take preventative measures in order to aid in self-management and safety orientation (Zhang et al., 2019).

Similarly, the doctor group also expressed a desire to obtain more information about patient symptoms (Table 1) which again aligned with the finding of individuals seeking comprehensible health information (Xiong et al., 2021). Additionally, doctors expressed their preference to know more detailed illness information and diagnoses, descriptions of detailed lab results, recommendations for what tests to perform and background information about the patients. Their responses indicated a similarity of the study by Kim et al. (2024). In their study the researchers found that the primary reason that doctors sought out explanations, were to identify problems and determine the right treatment plans for patients. For example, a respondent from the doctor group emphasised their need from the AI was to "perform a differential diagnosis of the patient's symptoms" (54-year-old female, Lithuanian, family medicine) this might be due to clinicians recognizing that there might be a number of possible diagnosis and trying to determine the right options for certain participants. Additionally, it seems that the doctors' group imagined the AI as a sort of an assistant that would be able to perform the simple primary tasks for them. This difference in focus between the two groups highlights the varying levels of expertise and the consequent need for tailored AI explanations, as discussed by Langer et al. (2021).

Interestingly, a unique theme among the patient group was the desire to know when to consult a real-life doctor (Table 1), indicating an underlying mistrust with AI-driven diagnoses and a preference for human medical judgement in severe situations. This observation is in line with recent studies highlighting patients' trust concerns with AI in healthcare (Steerling et al.2023) with one participant stating that they "...would not use AI. I do not trust AI to give me health-related information and/or advice. AI is not a substitute for medical advice from a human (e.g. GP)" (22 y/o, Female, Dutch). However, with many other participants not shying away from providing the AI with their need's information, it seems that many of the participants view this as an alternative of seeking medical information.

Following this, the quantitative analysis of the study focused on evaluating the preferences for different types of AI explanations (causal, contrastive, counterfactual, mechanistic) among both medical experts and naive users, as well as assessing their satisfaction levels with these explanations. This analysis was crucial in understanding how these two distinct user groups interact with and perceive AI-based diagnostic systems.

Causal explanations were the overall second most popular choice among the two groups. In the study these types of explanations presented information in a straightforward, short and simple way. Lombrozo (2007) in her study found that generally simple explanations, especially in the cases where they were correct, elicited a higher response rate. This was expected to be so for the participant group as the causal explanations were picked to adhere to the naive users' need for less information (Ribera and Lapedriza, 2019). However, from the study it was found that the results were quite opposite (Figure 1, Figure 2). Proportionally

(37%) it was the doctors that preferred these answers more than the patients. There may be several factors for this. The first explanation adheres to the psychology term, of Occam's razor, in which simpler theories are to be preferred over complex ones (Baker 2005). This ties in with what Lombrozo (2007) found in her study. It could be presumed that doctors chose this answer due to the fact that they already possessed a high level of knowledge and did not need an overdetailed explanation. Additionally, it could be due to the type of need elicited by the medical experts. In their study Kim et al. 2024 found that doctors sought information from AI in order to construct better specialised treatment plans to meet specific needs of patients. It could have been the case that the doctor group already had enough expertise and an AI assistant was only there to make suggestions such as a treatment plan, or differential diagnosis. The final reason for this might have been due to the difficulty of the survey. In the satisfaction analysis (Figure 4) it was found that doctors rated causal explanations second highest. That was still, however, lower than most of the patients' satisfaction ratings. This suggests that the expert group chose this answer simply due to it being the most understandable. This is further explained if satisfaction scores of the survey are looked at. At the end of the survey both the participants and doctors were asked for their feedback on the experiment. Three of the doctors mentioned that in fact the survey was hard for them to complete correctly as they did not have the domain knowledge to accurately answer the questions "The questionnaire is highly specialized and intended more for a neurologist or neurosurgeon. "(54 y/o, Female, Lithuanian, Family medicine). Additionally, the analysis found a significant difference between the two groups as it was preferred by the doctor group. Taking this into account, it was therefore concluded that simple causal explanations are to be preferred by medical experts in this study.

Counterfactual explanations were the third most picked choice among the doctor and patient group. The satisfaction scores of this type were not that high for either group. This came as a surprise as many researchers posited that counterfactual explanations are to be considered "good" (Wachter et al. 2017a, Mittelstadt et al., 2019, Tian et al., 2022). This type of explanation aimed to provide individuals by answering the "What if different symptoms were present?". By providing real world knowledge and building upon previously familiarity and giving extensive information it was therefore expected to rank higher. However, the study found that this explanation type ranked at the lower end in choice frequency and satisfaction. It might have been the case that this type was simply not causal enough for the doctor group (Langer, 2021) and not extensive enough for the patient group (Elton, 2022). Additionally, the research found the mean of the groups to be not significant for this type of explanations, therefore we concluded that counterfactual explanations were not preferred by either group.

It was assumed from the literature that contrastive explanations would have been preferred by the patients group. In his paper Mittelstadt (2019) argues that in everyday language people do not ask why an event happened but rather why an even Q happened instead of an event P. This play into the human way of making logical connections by linking features together (Rehder, 2003). It was therefore quite interesting to find that contrastive explanations received a small amount of patient (Figure 1) and doctor (Figure 2) responses. It might be so that due to the nature of the explanation, neither participant groups preferred this type. It is assumed from literature that the strength of a contrastive explanation depends on the previous knowledge of the participants. This type excels on the fact that it provides simple causal information about a new aspect without overwhelming the individual with unnecessary detail (Miller 2018). It could be assumed that because of their extensive domain knowledge that the doctor group would have preferred this explanation, yet that was not the case. The satisfaction analysis (Figure 4) gives an interesting perspective on this matter. From the figure it can be implied that the patient group rated this explanation type the highest while the doctor group the lowest. This gives even more consideration as to perhaps the participants appreciating the cause-and-effect explanation. It could have been that in this case most participants did not have much previous knowledge to speak of, therefore everything presented to them would have been new. While this may not have been an issue it could be so that both of the groups did not prefer this answer simply due to the other explanations adhering to their needs better. Similarly, Van Der Waa et al., 2018 in their research found that neither rule-based nor example based contrastive explanations found much success in presenting the participants with new information. This was further supported by the statistical analysis as it was found that there was no significant difference between the groups means and therefore conclude that this answer type is not of the highest preference for either group

Finally mechanistic explanations were among the most population choices as the patient group picked this answer almost half the time (49%). The strength of the mechanistic explanations came in their ability to describe the situation extensively and provide many details. Elton (2022) in their study described the common problems of mechanistic explanations as being too detailed. As overwhelming the individual with a lot of information will not help them understand it any better but rather create a sense of overbearingness. Completeness of models is where mechanistic explanations thrive yet again as Caplan and Craver (2020) put it, however, completeness does not always ensure that the model will be understood by all. For these reasons it was expected that this type would not have received that many responses. However, the contrary was found, surprisingly still, the answer was majorly picked by the patient group. A reason for this might be understood by looking into the qualitative part of the data. From the codes it can be found that the patient group are thoroughly interested in understanding their symptoms fully. It might be the case that knowledge of a situation gave the naive group a sense of confidence to tackle the illness themselves. As Kim et al. 2024 found, one of the main reasons for the patient group seeking explanations was to remain vigilant throughout health management. This builds upon further research as it is found that patients were highly interested in getting information about long term concern, treatment and remedies they could perform themselves. To further this point Xiong et al., (2021) found that one of the main reasons individuals seek information online is to manage their health better. It might be the case that for the participants, having ample amounts of information helped better understand their case and therefore stay confident during the imagined illness. Our statistical analysis found a significant difference among the two groups; therefore, we concluded that mechanistic explanations are to be preferred by the naive user group.

Strengths and weaknesses

The strength of this study lies in its comprehensive mixed-methods approach, which effectively combines qualitative and quantitative data to explore the user-centered needs in XAI for healthcare. The methodology allowed for exploration of the subjective information needs and objective preferences of different user groups. The multinational participant pool increased the generalizability of the findings, making them relevant across a broad spectrum of users. The information gathered to conceptualize the survey question using online forums and scientific literature makes the research more relevant. Furthermore, the use of a real-world medical case scenario of a low-grade glioma in grounds the research in practical, applicable contexts, thereby increasing its relevance for actual healthcare settings.

Despite its comprehensive approach, the study has limitations that must be acknowledged. The sample size, although diverse, may not fully represent the wide range of perspectives and experiences in the broader population, particularly in varying cultural and healthcare system contexts. Due to the lengthy process of finding sufficient medical experts to participate, convenience sampling had to be used. Because of this, around half of the medical sample came from the same country. This holds similarly true for the participant group, as the majority of the participants were gathered from the same background, which is university students. The complexity of the medical scenarios used might have influenced the results as well. Due to a lack of previous knowledge, survey answers were conceptualized and made to be too difficult, this might have led the medical expert group to seek less detailed explanations potentially skewing the results.

Implications

This study's findings have significant implications for the design and implementation of XAI systems in healthcare. Firstly, the research adds more to the emerging fact that different groups require different explanations. By highlighting this, the use of AI in the future medical field can prove to be invaluable. Additionally, providing clear information needs build upon new research of the information needs of lay and expert users. Clear preferences among patients for detailed, mechanistic explanations suggests that XAI systems should be designed to provide in-depth, understandable information about medical conditions and treatments. It provides their preferences for needing information to first try and handle their illness by themselves rather than relying on doctors. For doctors, the preference for concise, causal explanations indicates a need for XAI systems that quickly deliver relevant and precise information to assist in clinical decision-making. This information proves that medical experts would benefit greatly from AI systems adapted to their needs, proving to be invaluable tools for their everyday work. This research highlights the importance of tailoring XAI interfaces and content according to user expertise and needs, potentially leading to improved trust and efficacy in AI-assisted medical diagnostics. Building upon the existing user-centered research, this study proves that AI and different users should work together to conceive the successful implementation.

Conclusion

The goal of this study was to find what type of information doctors and patients want to know from an AI diagnosis system and how doctors and patients differ in their explanation needs of medical illnesses. The qualitative analysis revealed a clear difference in the information needs of patients and doctors. Patients tended to focus on gathering comprehensible health information about their symptoms, while doctors sought detailed diagnostic and medical information in order to comprise a full picture of the patients issue to possibly create a correct treatment plan. These findings have significant implications for the design of XAI systems in healthcare. They emphasise the necessity of a user-centred approach that recognizes and tailors the distinct information needs for different user groups. As Ribera and Lapedriza (2019) argue, understanding each target group's perspectives is crucial in developing AI systems that are not only technically proficient but also practically useful and trustworthy.

Additionally, the quantitative part of the analysis gave a clear indication of what type of explanation types both doctor and patients sought after. From the analysis it was found that patients preferred mechanistic explanations due to their detailed description of the illness. For this group, gathering knowledge about an illness might have given them motivation and confidence to tackle their illness better, which was in line with what Kim et al. (2024) found in their study. For the doctor group, simple causal explanations were preferred. It was surmised that doctors already had a of base knowledge and only needed advice from AI based diagnostics. Additionally, it was found that better tailoring of the survey should have been done for the doctor group. This was revealed by the satisfaction analysis, as overall, doctors scored much lower on the score of the explanation's types than patients. This might have been due to the doctor's survey consisting of many complex specialised terms. For some of the participants it was hard to understand what was asked of them because the terms used were primarily intended for neurologists. These insights are critical for the development of user-centered XAI systems in the medical field, ensuring that the explanations provided by AI align with the specific needs and expectations of different user groups.

Future research

Future research should aim to broaden the scope of this study by including a larger and more diverse sample that covers different healthcare systems and cultural backgrounds. In the same way some statistical adjustments should be made. In order to investigate if the differences between the groups is realized, not only should the sample size be increase but also the plotting of the residuals should be done to test whether the data is normally distributed. By doing so investigating the impact of XAI on actual clinical outcomes could provide more direct evidence of its utility in healthcare for a bigger audience. Research could also be extended to other medical fields beyond radiology, to understand how user preferences for explanations vary across different medical specializations and scenarios. Finally, one limitation that this research had which could be improved a lot was the fact that the qualitative and quantitative research was done at the same time, it would have proven more useful to first gather information from the open-questions, same as Kim et al (2024). If this was done, it would have allowed for a more iterative and informed research design. Specifically, insights from the initial qualitative analysis could have been used to refine the quantitative survey questions, leading to a more tailored and precise assessment of user preferences. Such an approach would likely yield more nuanced and specific findings, facilitating a deeper understanding of the nuances in user preferences for XAI in healthcare. Furthermore, it could provide better insights for developing AI systems that are more effectively aligned with the varying needs of different user groups, thereby increasing the practical applicability and impact of the research in real-world healthcare settings.

References

ATLAS.ti Scientific Software Development GmbH. (2023, December 20). ATLAS.TI | the #1 software for qualitative data analysis. Retrieved from https://atlasti.com/

Azungah, T. (2018). Qualitative research: deductive and inductive approaches to data analysis. Qualitative Research Journal, 18(4), 383–400. https://doi.org/10.1108/qrj-d-18-00035

Baker, A. (2004). Simplicity. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Winter 2004 Edition). Retrieved from http://plato.stanford.edu/archives/win2004/entries/simplicity/

Beckers, S. (2022b). Causal explanations and XAI. arXiv. https://doi.org/10.48550/arxiv.2201.13169

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Craver, C. F., & Kaplan, D. M. (2020). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. British Journal for the Philosophy of Science, 71(1), 287-319.

Cook, C., & Décary, S. (2020). Higher order thinking about differential diagnosis. *Brazilian Journal of Physical Therapy*, *24*(1), 1–7. https://doi.org/10.1016/j.bjpt.2019.01.010

Du, Y., Antoniadi, A. M., McNestry, C., McAuliffe, F., & Mooney, C. (2022). The Role of XAI in Advice-Taking from a Clinical Decision Support System: A Comparative User Study of Feature Contribution-Based and Example-Based Explanations. Applied Sciences, 12(20), 10323. https://doi.org/10.3390/app122010323

Elton, D. C. (2022). Common Pitfalls When Explaining AI and Why Mechanistic Explanation Is a Hard Problem. In Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 1 (pp. 401-408). Springer Singapore.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence (XAI) program. Ai Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hoffmann, J., & Magazzeni, D. (2019). Explainable AI Planning (XAIP): Overview and the case of Contrastive Explanation (Extended Abstract). In Lecture Notes in Computer Science (pp. 277–282). https://doi.org/10.1007/978-3-030-31423-1_9

Holzinger, A., Biemann, C., Pattichis, C., & Kell, D. (2017). What do we need to build explainable AI systems for the medical domain?. Retrieved from ResearchGate

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. Nature Reviews Cancer, 18(8), 500–510. https://doi.org/10.1038/s41568-018-0016

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jia, X., Pang, Y., & Liu, L. S. (2021). Online health information seeking behavior: A systematic review. Healthcare, 9(12), 1740. https://doi.org/10.3390/healthcare9121740

Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. Gastrointestinal Endoscopy, 92(4), 807-812. https://doi.org/10.1016/j.gie.2020.06.040

Kim, S., Choo, S., Park, D., Park, H., Nam, C. S., Jung, J., & Lee, S. (2023). Designing an XAI interface for BCI experts: A contextual design for pragmatic explanation interface based on domain knowledge in a specific context. International Journal of Human-Computer Studies, 174, 103009. https://doi.org/10.1016/j.ijhcs.2023.103009

Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. Korean Journal of Anesthesiology, 70(1), 22. https://doi.org/10.4097/kjae.2017.70.1.22

Lamba, D., Hsu, W. H., & Alsadhan, M. (2021). Predictive analytics and machine learning for medical informatics: A survey of tasks and techniques. In Elsevier eBooks (pp. 1–35). https://doi.org/10.1016/b978-0-12-821777-1.00023-9

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from explainable artificial intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artificial Intelligence, 296, 103473. https://doi.org/10.1016/j.artint.2021.103473

Lei, Y., Yin, M., Yu, M., Yu, J., Zeng, S., Lv, W., Li, J., Ye, H., Cui, X., & Dietrich, C. F. (2021). Artificial intelligence in medical imaging of the breast. Frontiers in Oncology, 11. https://doi.org/10.3389/fonc.2021.600557

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liyanage, S., Dia, H., Abduljabbar, R. L., & Bagloee, S. A. (2019). Flexible mobility on-demand: An environmental scan. Sustainability, 11(5), 1262. https://doi.org/10.3390/su11051262

Lombrozo, T. (2007). Simplicity and probability in causal explanation. Cognitive Psychology, 55(3), 232–257. https://doi.org/10.1016/j.cogpsych.2006.09.006

Maguire, M. (2017, October 31). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education. Retrieved from https://ojs.aishe.org/index.php/aishe-j/article/view/335

Meske, C., & Bunde, E. (2020). Transparency and trust in human-AI-interaction: The role of model-agnostic explanations in computer vision-based decision support. In Lecture Notes in Computer Science (pp.

Laato, S., Tiainen, M., Islam, A. K. M. N., & Mäntymäki, M. (2022). How to explain AI systems to end users: A systematic literature review and research agenda. Internet Research, 32(7), 1–31. https://doi.org/10.1108/INTR-08-2021-0600

Miot, H. A. (2017). Avaliação da normalidade dos dados em estudos clínicos e experimentais. Jornal Vascular Brasileiro, 16(2), 88–91. https://doi.org/10.1590/1677-5449.041117

Mishra, P., Singh, U., Pandey, C. K., Mishra, P., & Pandey, G. (2019). Application of student's t-test, analysis of variance, and covariance. Annals of Cardiac Anaesthesia, 22(4), 407. https://doi.org/10.4103/aca.aca_94_19

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. Association for Computing Machinery, New York, NY, USA, 279–288. https://doi.org/10.1145/3287560.3287574

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(6), 1141–1159. https://doi.org/10.1037/0278-7393.29.6.1141

Reddit thread for Brain Cancer subreddit. Retrieved from https://www.reddit.com/r/braincancer/comments/16sl2i8/seeking_stories_of_brain_tumor_diagnosis_for_my/

Forbes. (2019, September 26). What happens with self-driving cars kill people. Retrieved from https://www.forbes.com/sites/cognitiveworld/2019/09/26/what-happens-with-self-driving-cars-kill-people/#4b798bcc405c

Sipos, L., Schäfer, U., Glinka, K., & Müller-Birn, C. (2023). Identifying explanation needs of end-users: Applying and extending the XAI question bank. In MuC ’23: Mensch Und Computer 2023 (pp. 492-497). https://doi.org/10.1145/3603555.3608551

Sjödin, D., Parida, V., Palmié, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. Journal of Business Research, 134, 574–587. https://doi.org/10.1016/j.jbusres.2021.05.009

Tian, B., Cao, Y., Zhang, Y., & Xing, C. (2022). Debiasing NLU models via causal intervention and counterfactual reasoning. In Proceedings of the AAAI Conference on Artificial Intelligence, 36(10), 11376–11384. https://doi.org/10.1609/aaai.v36i10.21389

Trickygap4750. (2023). Seeking stories of brain tumor diagnosis for my bachelor thesis. Retrieved from Reddit: https://www.reddit.com/r/braincancer/comments/16sl2i8/seeking_stories_of_brain_tumor_diagnosis_for_my/

Van Der Waa, J., Nieuwburg, E. G., Cremers, A., & Neerincx, M. A. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. Artificial Intelligence, 291, 103404. https://doi.org/10.1016/j.artint.2020.103404

Van Der Waa, J., Robeer, M., Van Diggelen, J., Brinkhuis, M., & Neerincx, M. A. (2018). Contrastive explanations with local foil trees. arXiv. https://doi.org/10.48550/arxiv.1806.07470

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Social Science Research Network. https://doi.org/10.2139/ssrn.3063289

Wang, C., He, T., Zhou, H., Zhang, Z., & Lee, C. (2023). Artificial intelligence enhanced sensors - enabling technologies to next-generation healthcare and biomedical platform. Bioelectronic Medicine, 9(1). https://doi.org/10.1186/s42234-023-00118-1

Whitley, E., & Ball, J. (2002). Statistics review 6: Nonparametric methods. Critical Care, 6(6), 509. https://doi.org/10.1186/cc1820

Wolf, C. T. (2019). Explainability scenarios. In IUI ’19: Proceedings of the 24th International Conference on Intelligent User Interfaces. https://doi.org/10.1145/3301275.3302317

Yang, H., Wu, J., Zhou, J., Teng, H., & Xu, R. (2016). A case of brain glioma progression: Surgical resection and post-operation nursing care. Translational Cancer Research, 5(3), 315-319. https://doi.org/10.21037/tcr.2016.05.14

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things Journal, 1(1), 22–32. https://doi.org/10.1109/jiot.2014.2306328

Zhang, Z., Wang, L., & Lee, C. (2023). Recent advances in artificial intelligence sensors. Advanced Sensor Research, 2(8). https://doi.org/10.1002/adsr.202200072

Appendix A

First part of the study

A1 Patient group open question.

“Imagine you have some symptoms of illness and you would like to find out more about them. Describe what kind of information you would like to receive from an AI assisted doctor in order to fully understand your situation. (100 minimum characters) “

A2 Doctor group open question.

“Imagine your patient has some symptoms of an illness and you would like to find out more about them. Describe what kind of information you would like to receive from an AI assisted doctor in order to fully understand their situation? (100 min characters)”

A3 Patient case scenario

Case scenario Patient:

“In this study, you'll be presented with a medical case scenario and asked to review it from a patient's perspective. You will choose from a list of different explanations for each question to determine the most understandable and relevant information for your case. Please select the option that adheres to your needs the best.

Case scenario:

You're a 30-year-old who's been feeling quite off lately. Over the past 3 months, you've been feeling really tired. Walking has become a challenge, especially with your right leg, which feels like it's "dragging" behind. Sometimes, you even find yourself leaning to the left when you walk. On top of that, you've been getting headaches, mostly in the mornings, and sometimes your vision gets a bit blurry.

You haven't noticed any changes in your thinking, hunger, or weight. You also haven't had any recent injuries, started any new medicines, or made any big changes in your daily life. In your medical records, it's noted that you have hyperlipidemia. This is a more technical way of saying you have high cholesterol. You also used to use smokeless tobacco, but you quit that about 5 years ago. Something else worth noting is that your mother had vascular dementia when she got older. Vascular dementia is a condition where there's a decline in thinking skills due to reduced blood flow to the brain.

When you went for a check-up, the doctor said everything seemed okay with your brain nerves. However, you did have some weakness in your left arm and leg. The doctor also noticed your left knee reacted a bit more than usual when tapped.

Because of these symptoms, your doctor suggested getting an MRI, a special brain imagine, to get a better look inside and figure out what might be causing your problems. You had this done 3 days ago.

This is what your results look like:”

A4 Doctor case scenario

“In this study, you'll be presented with a medical case scenario and asked to analyse it from a doctor's perspective. You will choose from a list of different examination results to determine the most fitting explanation for each question. Please select the option that best aligns with your understanding and preferences.

Case scenario:

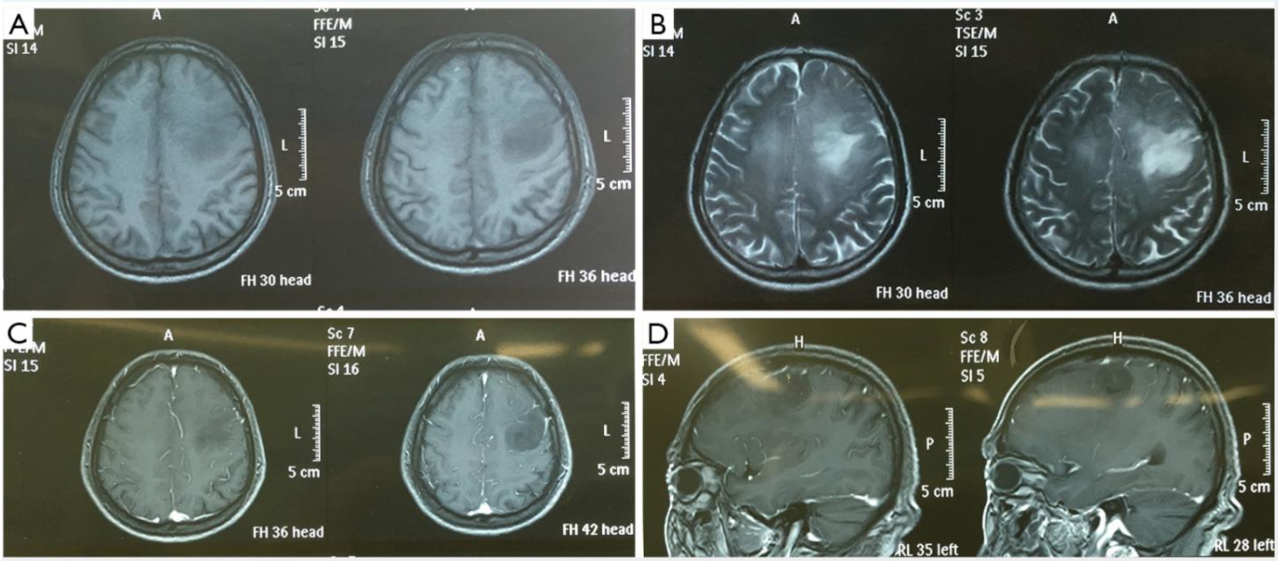
Doctor Mr. John Smith, a 60-year-old male with a background in agriculture, presented to the neurology department with a 3-month history of progressive fatigue, right-sided limb weakness, and unsteady gait. The patient described his symptoms as a noticeable "dragging" of his right leg and a tendency to veer towards the left while walking. Additionally, he reported persistent headaches, particularly pronounced in the mornings, and episodes of blurred vision. There were no reported changes in cognition, appetite, or weight. He denied any recent trauma, medication use, or significant lifestyle changes.

Mr. Smith's medical history was significant for hyperlipidemia. His social history revealed a past use of smokeless tobacco, which he had discontinued over two decades ago. There was a family history of maternal vascular dementia. On neurological examination, cranial nerves II through XII were intact. Motor examination revealed a strength of 3/5 in the proximal left upper and lower extremities and 4/5 in the distal left upper and lower extremities. Reflexes were normal with slightly brisk patellar reflexes on the left.

Given the progressive nature of the symptoms and the neurological findings, an MRI of the brain was recommended to further elucidate the underlying cause. He came in 3 days ago and this is the MRI photo that we received.

This is what the results look like:”

A5 MR photo of possible low grade glioma



Appendix A6 First parts, Survey questions and answers

Question 1: Based on the MRI findings, what best describes the abnormality in Mr. Smith's brain?

Patient Group:

A) The MRI shows a spot in the left frontal area that looks different than the normal brain tissue around it.

B) The MRI spot doesn't look like the rest of the brain tissue - it has an irregular shape and variations in shading.

C) If it was just a simple smooth circle, it might be harmless, but the MRI shows an uneven shape.

D) The MRI shows a region that has both lighter and darker patches, meaning it's made up of different tissue types.

Doctor Group:

A) MRI reveals an irregular left frontal lesion with heterogeneous signal intensity distinct from the surrounding parenchyma.

B) Unlike the homogeneous appearance of the normal brain, the lesion exhibits heterogeneity on MRI with irregular margins contrasting with adjacent tissue.

C) A solitary nodule with smooth regular borders would likely represent a benign process rather than an infiltrative neoplasm.

D) MRI depicts a heterogeneous lesion with mixed hyperintense and hypointense signals, indicating varied components including necrosis, angiogenesis, and inflammation.

Question 2: What can you infer about the nature of the lesion from the MRI results?

Patient Group:

A) The spot on the MRI looks to be made of different kinds of tissue, which means it could be a tumor.

B) Since the area doesn't look normal like the rest of the brain, it seems like it might be cancerous.
C) If it was just one type of tissue it would more likely be benign, but the different shades suggest cancer.
D) The lighter and darker patches show there's something abnormal going on with different cell types and blood flow.

Doctor Group:
A) The heterogeneity within the lesion suggests varied tissue characteristics indicative of a neoplastic process.
B) Unlike normal brain's homogeneous MRI signal, this lesion appears irregular with areas of differing intensity, concerning for an underlying tumor.
C) A uniform enhancement pattern following contrast administration would be expected with a benign non-cancerous lesion.
D) Presence of both hyperintense and hypointense signals implies differing cellularity and vascularity consistent with neoplastic growth

Question 3: Which of Mr. Smith's symptoms most directly suggests an issue in the left frontal region of the brain?
Patient Group:
A) The weakness and problems on the right side of his body point to something going on in the left side of his brain.
B) Unlike just a headache, his right side weakness and falling to the left mean the issue is probably in the left frontal region.
C) If he didn't have right-sided symptoms, we might look at other areas, but these suggest a specific problem in the left frontal part.
D) The left frontal area controls the right side's movement, so his right-sided weakness matches an issue in that brain region.

Doctor Group:
A) Mr. Smith's right hemiparesis, gait instability with leftward deviation, and right leg dragging align with a left frontal lobe process.
B) Unlike a migraine, his lateralizing motor deficits including right arm and leg weakness with leftward veering suggest a pathological process localized to the left frontal region.
C) Absent symptoms of right-sided weakness and leftward gait deviation, we may consider alternate etiologies, but their presence strongly indicates a left frontal lobe locus.
D) The left frontal lobe mediates contralateral motor activity, thus Mr. Smith's right-sided weakness indicates likely involvement of left frontal cortical control areas.

Question 4: How does the MRI help in understanding the cause of Mr. Smith's symptoms?

Patient Group:

A) The MRI shows a lesion pressing on his brain which could be causing the symptoms.

B) Unlike a normal scan, the abnormal spot on the MRI suggests this could be why he has issues.

C) If the MRI was clear without this lesion, we'd look elsewhere. But finding it provides a likely reason for his problems.

D) The lesion and swelling seen means increased pressure on his brain aligning with his headaches and limb weakness.

Doctor Group:

A) The MRI reveals an intracranial lesion exerting mass effect on adjacent structures, which anatomically correlates with the patient's symptomatology.

B) Unlike normal imaging findings, identification of this lesion provides radiographic evidence potentially explaining the patient's clinical presentation.

C) Lack of any abnormalities on MRI may necessitate consideration of alternate etiologies. However, the visualized lesion topographically corresponds to the symptoms.

D) Imaging characteristics including vasogenic edema signify elevated intracranial pressure from mass effect, consistent with his progressive headaches and neurological deficits.

Question 6: Based on the MRI and clinical presentation, what is the preliminary diagnosis for Mr. Smith?

Patient Group:

A) The unusual lesion and his symptoms together point to this likely being a low-grade glioma tumor.

B) His MRI isn't normal and his symptoms make it seem like he probably has a low-grade glioma.

C) If his scan and symptoms were different, we'd consider other things. But his case suggests a low-grade glioma is the most likely diagnosis.

D) The odd-looking spot and his issues indicate he probably has a low-grade glioma tumor.

Doctor Group:

A) Given the heterogeneous frontal lobe lesion with vasogenic edema and Mr. Smith's progressive neurological deficits, the confluence of imaging and clinical features is most consistent with a low-grade glioma.

B) Unlike primary brain tumors such as meningiomas, the patient's insidious symptom onset and MRI findings of an irregular enhancing mass are characteristic of a low-grade glioma.

C) In the absence of MRI irregularities and rapid symptom progression, alternate etiologies would be higher in the differential. However, the current presentation is classic for a low-grade glioma.

D) The presence of a heterogeneous frontal lobe lesion with patchy enhancement coupled with Mr. Smith's subacute neurological decline together strongly indicate an infiltrating low-grade glioma.

Question 7: What is the next recommended step to confirm the exact nature of the lesion?

Patient Group:

A) Doctors should take a small sample of the spot to examine it and confirm if it's cancer.

B) A scan just shows the shape, but examining a sample will provide the details we need to confirm if it's benign or malignant.

C) If the scan looked less worrying we might just monitor it. But given how it looks, the next logical step is to take a sample and test it.

D) Examining some cells from the lesion under a microscope will give information on the specific type and aggressiveness of the tumor.

Doctor Group:

A) The recommended next step is stereotactic biopsy for cytological and histopathological analysis to definitively characterize the lesion's lineage and grade.

B) Whereas imaging reveals macroscopic morphological details, biopsy allows microscopic examination of cellular characteristics to definitively determine benignity versus malignancy.

C) Absent clinical or radiographic red flags, serial imaging surveillance may have been reasonable. However, the concerning features warrant biopsy for pathologic tissue diagnosis.

D) Microscopic analysis of biopsy specimens can provide key insights into the lesion's subtype, proliferative index, genetic mutations, and other prognostic markers guiding management.

Question 8: What treatment options are typically considered for a diagnosed low-grade glioma?

Patient Group:

A) Doctors usually consider radiation or chemotherapy to try to stop the tumor growing.

B) Unlike fast-growing tumors where urgent surgery is needed, slower growing ones can be treated carefully with radiation or chemo.

C) If it was high-grade, urgent surgery or interventions may be considered. But for a low-grade one, less invasive treatments like radiation are often used.

D) Since surgery is difficult with these tumors, radiation or chemo that can reach all parts are better options to control growth.

Doctor Group:

A) Management considerations for low-grade gliomas include surgical resection when feasible, adjuvant radiation, and systemic chemotherapy for residual or unresectable disease.

B) Unlike high-grade gliomas warranting first-line surgical debulking, low-grade gliomas are often managed conservatively, with radiation and/or chemotherapy for tumor control due to their relatively indolent growth.

C) More aggressive interventions like surgical cytoreduction may be considered for higher grade lesions, but are less suitable for low-grade gliomas, which are better managed with radiation or systemic chemotherapy.

D) Given the infiltrative margins and eloquent locations precluding gross total resection, radiation and chemotherapy capable of crossing the blood-brain barrier are preferable modalities for tumor control in low-grade gliomas.

Question 9: Why might a blood test be recommended following a diagnosis of low-grade glioma?

Patient Group:

A) Blood tests can check for tumor markers to monitor how fast it's growing over time.

B) Unlike scans checking size, blood tests can less invasively track molecular signs of how aggressive the tumor is acting.

C) For a benign tumor, blood marker levels stay stable. But with gliomas they rise, so blood tests help track progression.

D) Proteins in the blood indicate cellular damage from the spreading glioma, helping doctors monitor it and guide treatment.

Doctor Group:

A) Serum and CSF markers can be used to monitor growth kinetics, progression, and treatment response of the glioma over time.

B) Unlike anatomical imaging modalities, frequent lab work allows relatively non-invasive surveillance of circulating biomarkers indicative of the tumor's molecular phenotype and proliferative activity.

C) In benign tumors, these markers would be anticipated to remain static. However, low-grade gliomas demonstrate inexorable progression, with levels correlating to growth and suggestive of transformation.

D) Analyzing serum and CSF markers of glial injury like GFAP offers insight into tumor cell damage and vascular permeability changes from infiltration, aiding clinical surveillance.

Question 10: Are there any risk factors in Mr. Smith's medical or social history that could have contributed to the MRI findings?

Patient Group:

A) His past use of smokeless tobacco may have raised his risk for these brain tumors.

B) More than genetics, his lifestyle choice of using tobacco probably played a bigger role in developing the tumor.

C) If he didn't have the history of tobacco use, his odds of getting this type of tumor would probably be lower.
D) Chemicals in smokeless tobacco can damage cells, leading to DNA changes that cause tumors to grow.

Doctor Group:
A) Mr. Smith's prolonged smokeless tobacco use may have been a contributing risk factor by promoting cellular mutations leading to neoplastic transformation.
B) Compared to non-modifiable genetic predispositions, his history of tobacco carcinogen exposure was more likely to initiate mutations causing phenotypic derangements like inhibited apoptosis and angiogenesis.
C) In the absence of an exposure history to a known neurological carcinogen like tobacco, the likelihood of developing this tumor would be reduced.
D) Tobacco carcinogens can create DNA adducts and inhibit DNA repair mechanisms, disrupting signaling pathways controlling apoptosis and promoting tumoral angiogenesis.

Question 11: How might Mr. Smith's symptoms evolve if the lesion remains untreated?
Patient Group:
A) Without treatment, his issues like headaches and weakness will probably worsen over time as the tumor grows.
B) Unlike a sudden stroke, his symptoms would likely slowly get worse as the tumor spreads without any intervention.
C) If it was benign, his symptoms might plateau. But this tumor will keep causing worse issues if not treated.
D) As the tumor invades more of the healthy brain tissue, it can worsen problems like paralysis and vision issues.

Doctor Group:
A) If untreated, the patient would likely demonstrate continued neurological decline with worsening headaches, motor deficits, and visual changes as the glioma progresses.
B) Unlike the acute presentations of vascular pathologies like stroke, his symptoms are expected to follow an insidious course with gradual deterioration in the absence of intervention.
C) In the case of a benign etiology, his symptoms could plateau. However, with an unchecked low-grade glioma, serialized imaging would anticipate inexorable infiltrative spread.
D) Continued growth of the glioma into adjacent eloquent cortex and white matter tracts may precipitate worsening hemiparesis, visual field cuts, cognitive impairment, and other focal neurological deficits.

Appendix B

Second part of the survey

B1 Patient case scenario

"You will now be presented with the second part of the study.

Case scenario:

After turning down our advice for surgery, you chose to get radiation treatment at another hospital. Three months after that treatment, your family told us you were getting more and more confused.

When you came to us, you were very tired and had trouble talking. Your eyes looked normal, but the strength in your right-side muscles was reduced, and a foot reflex test showed unusual results. We did another brain scan and found a big growth in the left front part of your brain that had changed in appearance. We believed this could be a serious type of brain tumor and suggested surgery to remove it.

Here are the findings:"

B2 Doctor case scenario

"You will now be presented with the second part of the survey.

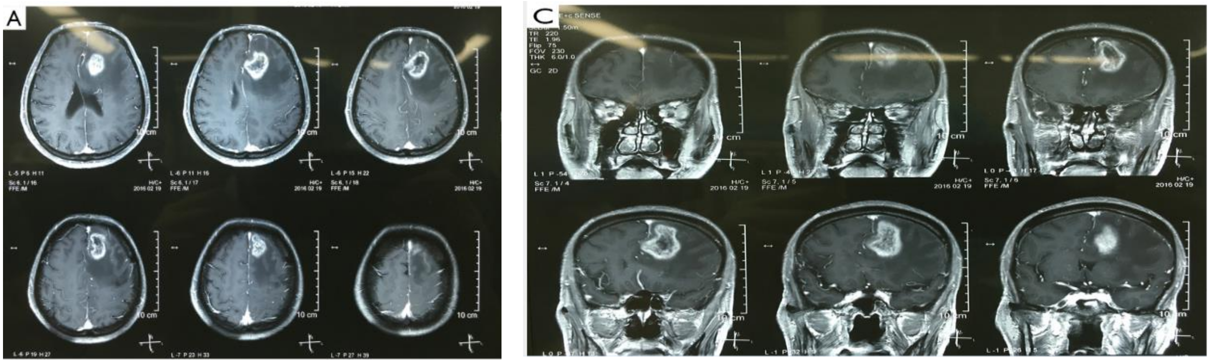
Case scenario:

After the patient declined the recommended surgical intervention, he opted for radiotherapy at an external medical facility. Subsequent to a three-month period post-radiotherapy, the patient's family reported a notable decline in his cognitive state, characterised by increasing confusion.

Upon clinical evaluation, the patient exhibited signs of fatigue and impaired communication abilities. Ophthalmic examination revealed bilaterally symmetrical pupils, each with a diameter of 2.5 mm, and a normal pupillary light reflex. Neurological assessment indicated a Grade III muscle strength on the right side and a positive Babinski sign. A subsequent MRI of the brain revealed a pronounced lesion in the left frontal lobe extending to the basal ganglia, exhibiting irregular contrast enhancement patterns. The radiological findings were highly suggestive of a high-grade glioma, prompting a recommendation for surgical excision.

Here are the MRI results:"

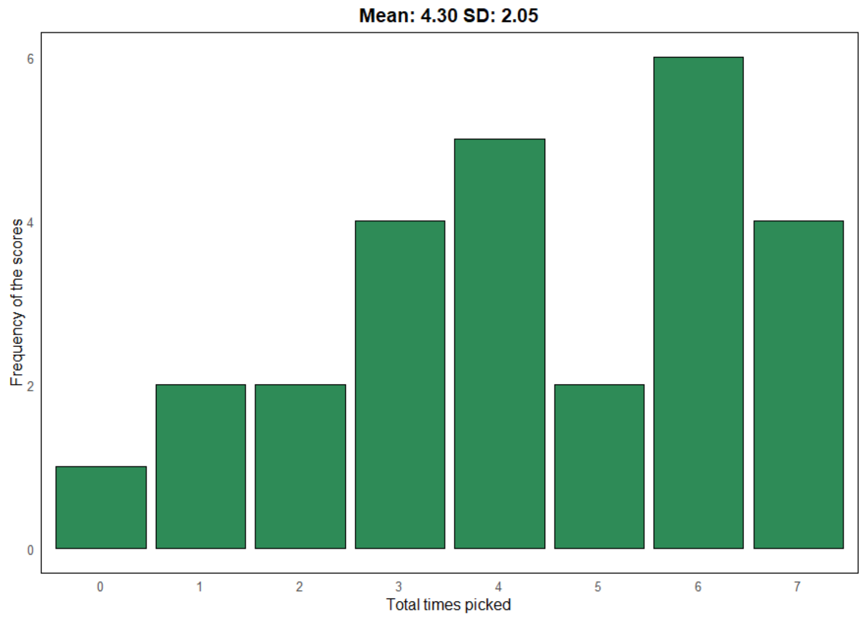
B3 MR photo of a possible glioma



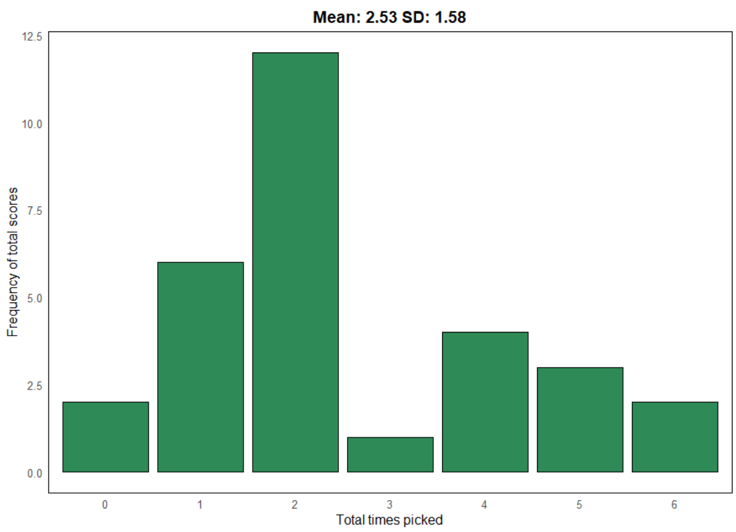
Appendix C

Score frequencies for answer types

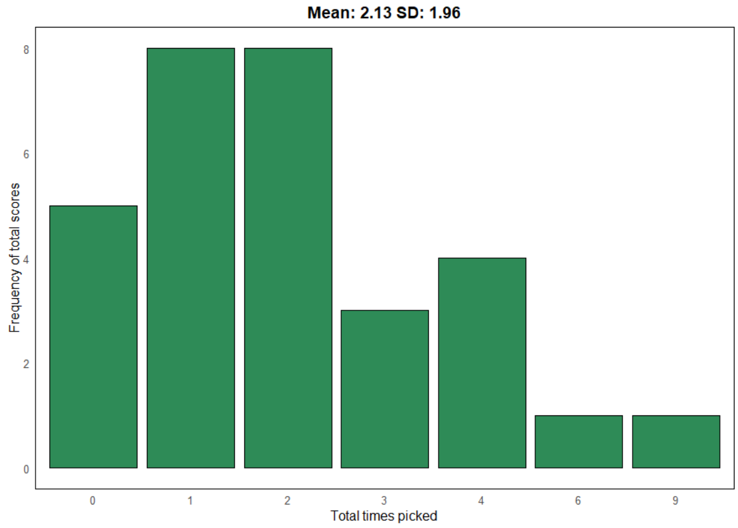
C1.1 Score frequency causal explanations patient group



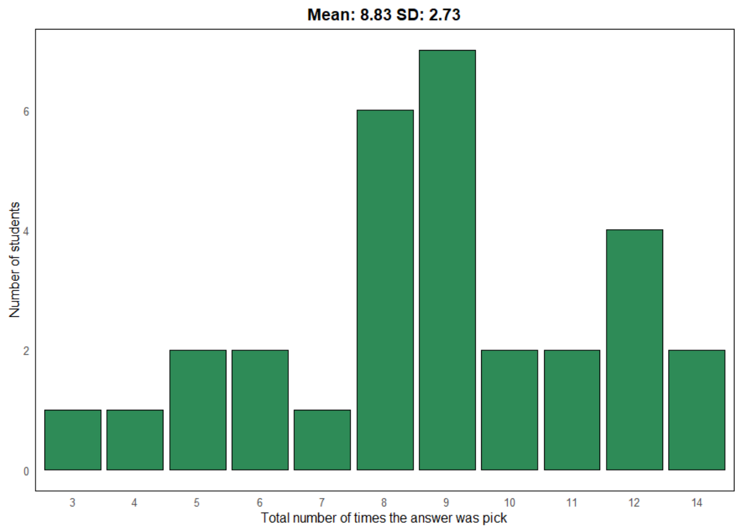
C1.2 Score frequency counterfactual explanations patient group



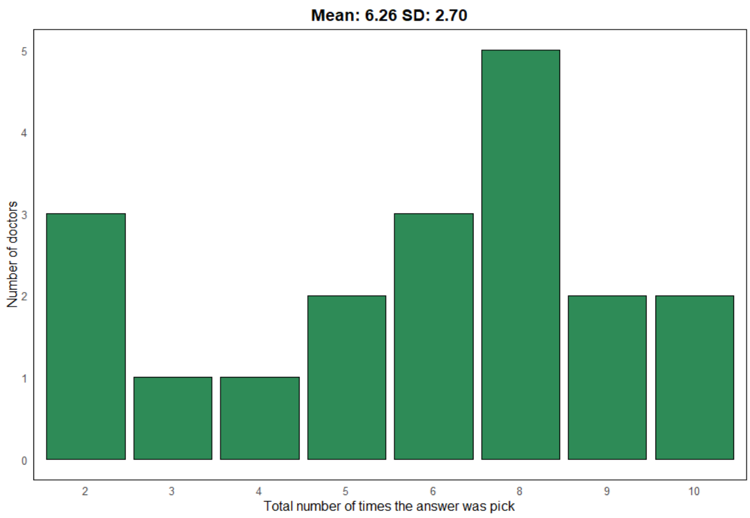
C1.3 Score frequency contrastive explanations patient group



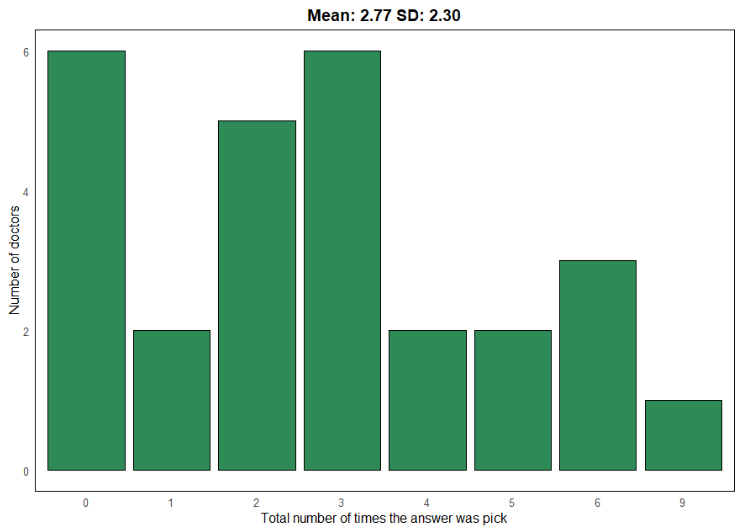
C1.4 Score frequency mechanistic explanations patient group



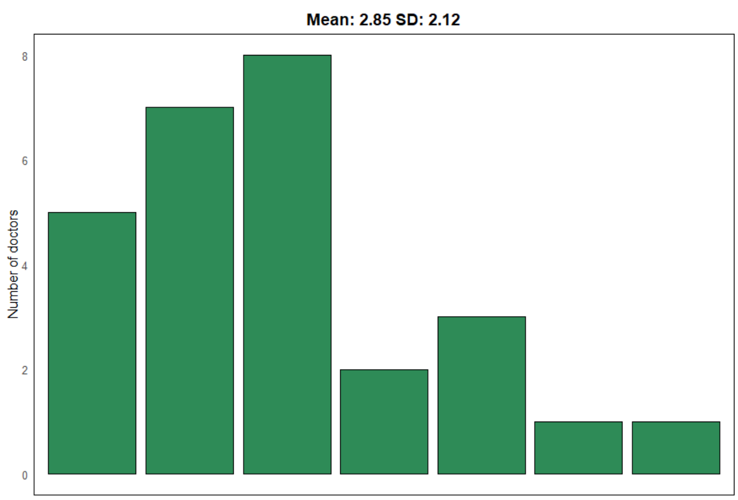
C2.1 Score frequency causal explanations doctor group



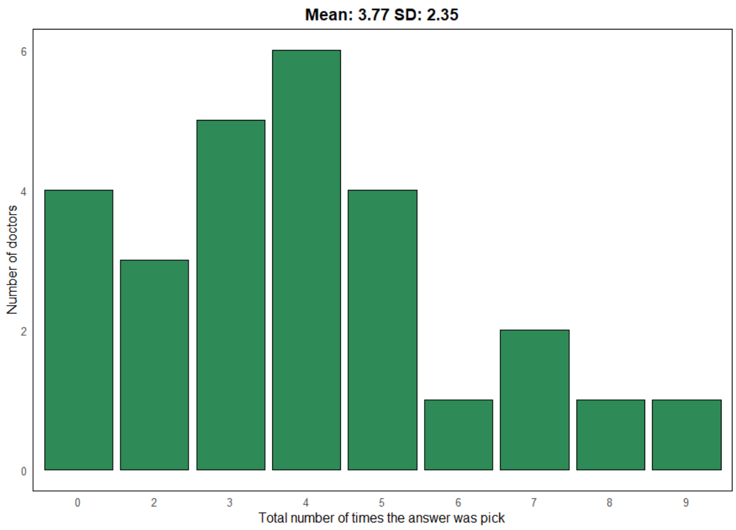
C2.2 Score frequency counterfactual explanations doctor group



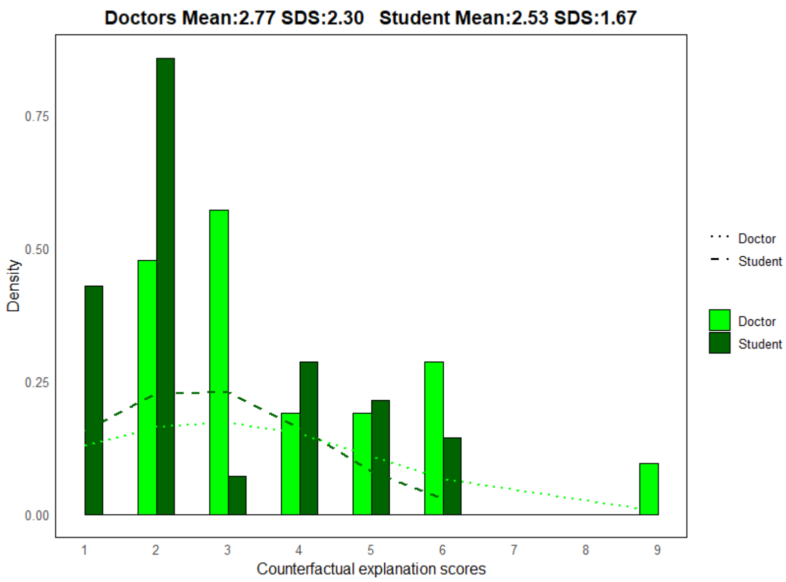
C2.3 Score frequency contrastive explanations doctor group



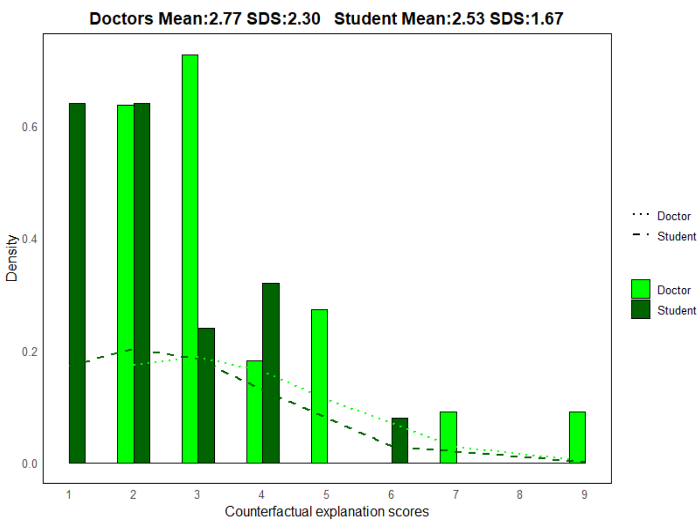
C2.4 Score frequency mechanistic explanations doctor group



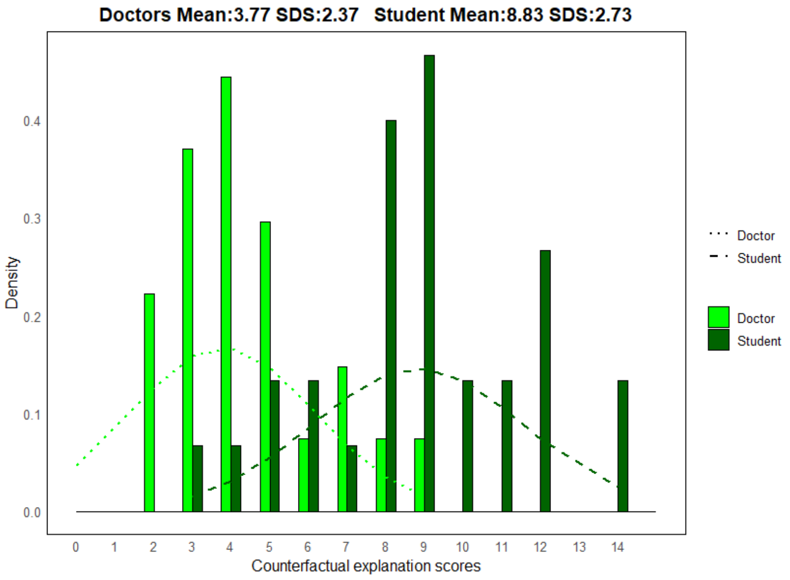
C3.1 Frequency score density plot counterfactual explanations



C3.2 Density plot of Contrastive explanations between the 2 groups



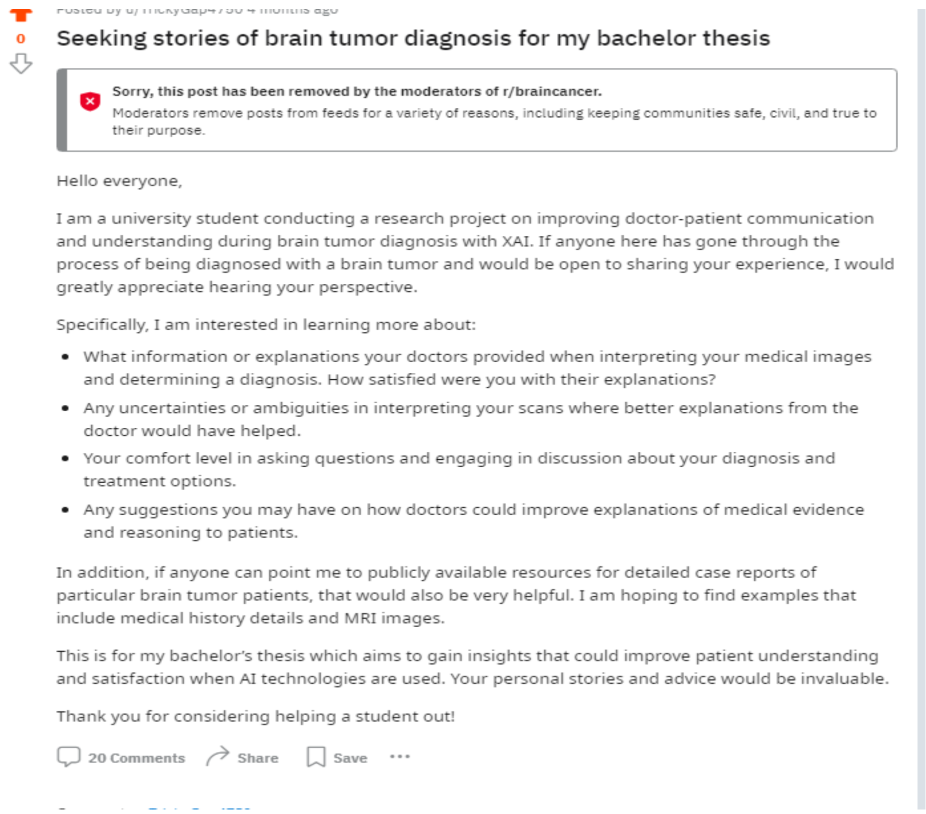
C3.3 Density plot of Mechanistic explanations between the 2 groups



Appendix D

Reddit thread

D1 Initial query posed to the r/braincacer subreddit



Seeking stories of brain tumor diagnosis for my bachelor thesis

Sorry, this post has been removed by the moderators of r/braincancer.
Moderators remove posts from feeds for a variety of reasons, including keeping communities safe, civil, and true to their purpose.

Hello everyone,

I am a university student conducting a research project on improving doctor-patient communication and understanding during brain tumor diagnosis with XAI. If anyone here has gone through the process of being diagnosed with a brain tumor and would be open to sharing your experience, I would greatly appreciate hearing your perspective.

Specifically, I am interested in learning more about:

- What information or explanations your doctors provided when interpreting your medical images and determining a diagnosis. How satisfied were you with their explanations?
- Any uncertainties or ambiguities in interpreting your scans where better explanations from the doctor would have helped.
- Your comfort level in asking questions and engaging in discussion about your diagnosis and treatment options.
- Any suggestions you may have on how doctors could improve explanations of medical evidence and reasoning to patients.

In addition, if anyone can point me to publicly available resources for detailed case reports of particular brain tumor patients, that would also be very helpful. I am hoping to find examples that include medical history details and MRI images.

This is for my bachelor's thesis which aims to gain insights that could improve patient understanding and satisfaction when AI technologies are used. Your personal stories and advice would be invaluable.

Thank you for considering helping a student out!

20 Comments Share Save

D2 First part of comments

Fair_Slice_6887 · 4 mo. ago

I think addressing potential side effects and setting reasonable expectations for healing should be a given. After we left the hospital, we had no idea if what he was experiencing was in the realm of normal and expected or if something had gone wrong. I've learned more from targeted Facebook groups than from his doctors about what further treatment options to explore. It's hard to steer the ship when I have no idea where we're going or how to drive a boat.

3 Reply Share ···

BlatantFalsehood · 4 mo. ago

I was diagnosed with a meningioma of the foramen magnum in November 2021. Every patient should have a neurologist like mine.

My appointment was first thing in the morning, before any other patients arrived. We sat at a small conference table in his office and he brought up my MRI results on the screen. He slowly and patiently outlined what he had been originally looking for and what was found.

He reassured me that it was most certainly not malignant and slow growing. However, he also stressed that there really is no such thing as a "good" brain tumor.

He outlined what next steps would be (neurosurgeon consult) and what the likely choices out of that would be (radiation and surgery). He suggested that due to the location (pressing on spinal cord and vertebral artery), the surgeon would likely recommend radiation.

He spent so much time with me answering questions. Unfortunately, the only question he *couldn't* answer is what was causing my symptoms -- because the location of the meningioma suggested they were not caused by that tumor.

2 Reply Share ···

D3 second part of comments

BlatantFalsehood · 4 mo. ago

I am still experiencing all of the symptoms that led me to the neurologist, but the only ones they believe are related to the tumor are the stiff, painful neck, painful shoulders, and headaches upon waking that dissipate within about an hour after I am up and moving around.

It was quite the doozy of a list, all of which developed slowly over late 2019 and 2020.

- Tremor – hands (particularly ring fingers); left foot – all come and go, but can be quite severe
- Stick, painful neck; painful shoulders
- Transient visual problems in left eye – "wiggling" or clouding over
- Wake with bad headache most days of the week that goes away once I am up and around
- Balance issues
- Stiff gait
- Sensation that I'll pass out when I laugh heartily or yawn deeply (head swims, vision goes black)
- Sluggish bowels/muscle tone
- Slower urination
- Increase in voice "warbling"
- Memory
- Hearing in right ear
- Pain and tingling in left ear
- Unusual muscle spams (e.g., outside of heel of my hand)

Because some of the most troublesome symptoms, such as the worst of the hand tremors, are transient, I still don't have a diagnosis at this time. I suspect if it continues to get worse, I'll eventually be diagnosed with Parkinson's or essential tremor or something.

I got excited when I heard about the "smell" test for Parkinson's, because during my 40s I often thought I had an odd odor that I called "old lady smell." But no one else could smell it. However, there is no clinical test yet that I am aware of, and I haven't found a clinical trial that I might qualify for.

I'm only 60 and otherwise healthy and fit. Some of this may just be related to getting older, but I don't think all of it is. [Edit for typo.]

2 Reply Share ···

D4 third part of comments

Fair_Slice_6887 · 4 mo. ago

My husband had/has a Craniopharyngioma. I understood the diagnosis, and didn't really have many treatment options as the tumor was large and causing significant hydrocephalus. What I was and continue to be extremely unhappy with is the lack of information about potential life time side effects. Neurosurgeon gave no indication of all the hormone and metabolism issues he would face having had a tumor on his pituitary/hypothalamus. Some sort of indication would have helped us set expectations for recovery needs. I was told he could return to work in "weeks". It's now been 8 months and he is still sleeping 14 hours a day and has significant short term memory issues. I know every case is unique, but we had no idea what we were in for.

2 Reply Share ···

TrickyGap4750 OP · 4 mo. ago

D5 fourth part of comments

Fuzzy-Machine-1734 · 3 mo. ago

And as for the doctors: the surgeon was told have little manners but being one of the best ones in Europe. As parents, we of course freaked out from the thought of having a little girl's skull open. I may have been provocative, but asked the surgeon to convince me why to let him dig her brain up. He casually said "I had f@cked up enough operations like this to know how to make it perfectly ". It might sound like bad manners but this exact sentence made me trust him. It meant to me he's learnt from previous mistakes, undertakes risky cases, not just easy ones. He spent about 15 seconds to scroll through the scan, then he closed his laptop and told us how it will be performed. The next task was to mentally prepare the kid to avoid her freaking out. So I chose to be open and very detailed (but not disclosing the risks, it was our task to worry, not hers). So we watched videos, anatomy pictures of brains, and tried to walk her through the whole procedure from the point she gets stabbed with needles to what it may feel like waking up and what to expect (vomiting, possible disorientation, etc). She was all excited and cooperative, asked a trillion of questions from the healthcare workers and she was happy and excited to wake up in ICU and see the machines, cords, whatever was attached to her. For my delight the moment she woke up she started to tell me stories about the anaesthesiologists, so I knew her memory was intact (part of her hippocampus and one of her amygdala was gone). Fast forward we had to wait for pathology. I did a tremendous amount of work to understand them, find people who can explain the microbiology part, etc. then the opinion of the oncoteam was that she needs radiation and chemotherapy. At that point I called the surgeon and told him I won't let that happen until someone points it out in the pathology report what is the indication, because I cannot see any. Two days idle, then the leader of the oncoteam called me and told me they messed up and I was right, no need of any of the therapies. On one hand I appreciated their openness about their mistake, on the other hand... what about other patients, who will not or cannot dig so deep and are afraid to question a doctor?

2 Reply Share ···

Fuzzy-Machine-1734 · 3 mo. ago

We are located in Eastern Europe, basically a spin dictatorship that means the quality of healthcare is dependent on what city/region you're in. The system itself is so inefficient that you'll die sooner than you would get a diagnosis. My daughter was 8, she started to smell funny scents. First we thought she is just fooling around, but after a while it became suspicious she felt the same smell at very different locations. Neurology thought it must be epilepsy, but all tests came back negative. Her neurologist told us to make an MRI just to be sure but he doubted there'd be anything. There was, 5 cm diameter mass in the temporal lobe. Next morning we were sitting in the buffet of the capital's neurosurgical centre, hunting for a doc to tell us what's next. While waiting (the capital has a mayor who is not with the governing party, so the quality of healthcare is....hm, let's say heavily underfunded) we saw neurosurgeons coming down for a coffee and making jokes on "look at this photo of the OR, after ruin pubs we have ruin theatres, hahaha". That was the point we decided she won't be operated in that city. The other reason was