**1. General Description of the AI-Powered Medical Expenditure System**

This technical documentation offers a comprehensive description of the AI-based Medical Expenditure System, developed by our company. The system is designed to predict annual healthcare expenditure for individuals by leveraging demographic data, socio-economic factors, and self-reported medical conditions. Its primary customers are healthcare “payers” like insurance companies, government agencies, employers, and individuals who are financially responsible for medical costs.

User Instructions: Detailed instructions on how to use the system are available in a separate User Guide. The guide can be found in the ‘Help’ section of the application or can be downloaded from our official website. It caters to the needs of multiple user groups, explaining the functionalities tailored to each.

The system’s primary function is the prediction of medical expenditure, but it can also be utilized to prioritize emergency first response services. The system operates under stringent guidelines to ensure transparency, fairness, robustness, and explainability, catering to the needs of both internal and external stakeholders.

The Medical Expenditure System serves multiple user groups, including healthcare data scientists, care managers, and medical staff. Each group has distinct requirements, and the system provides tailored explanation methods for machine learning predictions.

For healthcare data scientists, the system offers interpretable models that facilitate comprehensive model evaluation. These models include directly interpretable forms, allowing for a holistic understanding of the model’s behavior. They not only identify the most predictive features but also reveal the nature of the dependencies, such as their direction and shape. These insights can be cross-referenced with domain knowledge from healthcare experts or reviewed by managers, bolstering confidence in the system’s expected performance. Furthermore, the insights can inform interventions to mitigate costs, contributing to effective care management. Although these models are not causal, they can be interpreted causally with the assistance of expert knowledge or suggest causal hypotheses for future studies.

Care managers play a vital role in making final decisions based on the system’s predictions. To facilitate their understanding of machine learning outputs, we employ the ProtoDash method (class ProtodashExplainer). This method identifies prototypes or representatives in the training data that are similar to a given individual and receive the same class label. By showcasing examples, we enhance the explainability of the system for care managers.

Medical staff, including doctors and nurses, are provided with explanations for the recommendations generated by the system. These explanations are crucial in the care management process as nurses can evaluate the recommendations for quality and correctness, providing valuable feedback. The explanations are generated using the Contrastive Explanations Method (CEM), which highlights features present in the input instance that influence the model’s classification. Additionally, CEM identifies features that are minimally absent but would have altered the classification if present. This level of transparency enables medical staff to understand the reasons behind the system’s recommendations.

The Medical Expenditure System leverages the power of the XGBoost machine learning algorithm, known for its robustness and accuracy. The XGBoost algorithm has been extensively trained on healthcare data to provide reliable predictions and insights.

Our Medical Expenditure System is integrated into a larger Care Management System, offering various services. It allows for prioritization of cases and provides insights to authorized members of the medical staff through the “Medical Staff” interface.

The system is developed on the IBM AIX360 platform and utilizes the released version aix360 v0.2.1. It relies on a range of dependencies, including ‘joblib’, ‘scikit-learn’, ‘torch’, ‘tensorflow’, ‘keras’, ‘matplotlib’, ‘numpy’, ‘pandas’, ‘scipy’, ‘xgboost’, and others. The model has undergone extensive training to ensure optimal performance.

The system can be deployed in various forms:

• On-Premises: For organizations wishing to host the system on their servers, meeting the system requirements outlined below is necessary.

• Cloud-Based: Available as a SaaS (Software as a Service) offering for easier access and lower upfront costs.

• API Integration: For seamless integration into existing systems, the Medical Expenditure System can also be accessed via a secure API.

To deploy the AI-based Medical Expenditure System, the following system requirements should be met:

• Operating System: Compatibility with Windows, Linux, and macOS.

• Processor: Minimum requirement of an Intel Core i5 processor.

• RAM: At least 8 GB for optimal performance.

• Storage: Minimum of 100 GB of free disk space.

• Graphics Card: A dedicated Nvidia graphics card.

• Internet Connection: Stable connection for seamless API access.

• Additional Software: Python 3.6 or later

Installation Instructions: To install the AI-based Medical Expenditure System, please refer to the “Installation Guide” which can be accessed through our official website or within the software package. The guide provides step-by-step installation procedures for On-Premises, Cloud-Based, and API Integration deployment options.

The AI system is developed by our dedicated team of data scientists and AI experts and was last updated in June 2023. The current version of the system is 1.4.0.

**Opting Out, Overriding, or Reversing Decisions**

**User Discretion:**

While the AI-based Medical Expenditure System provides valuable predictions and insights, it is important to note that these outputs are advisory in nature. The system is not intended to replace human expertise, judgment, or the practice of medicine. Therefore, all users—healthcare data scientists, care managers, and medical staff—have the discretion to opt-out, override, or reverse decisions generated by the system.

**For Individual Consumers:**

Individual consumers who are financially responsible for their medical costs also have the option to disregard the system’s predictions. The outputs should be seen as tools for informed decision-making rather than absolute directives.

**For Healthcare Providers:**

**Care Managers:**

Care managers who play a vital role in making final decisions based on the system’s predictions can exercise discretion in whether to follow or disregard the system’s output. They can also consult with other healthcare professionals and utilize additional tests or diagnostic methods to either validate or challenge the system’s predictions.

**Medical Staff:**

Medical staff, such as doctors and nurses, can also override or reverse the system’s output based on their clinical judgment, patient preferences, or any new information that may not have been included in the data initially fed into the system.

**For Healthcare Payers:**

Insurance companies, government agencies, and employers who might use this system for actuarial or administrative purposes can also exercise discretion. It is essential that these stakeholders validate the AI’s predictions with additional assessments or human oversight, and they have the full ability to opt-out of using the system if it does not meet their specific needs.

**Safety Mechanisms:**

Our system also includes built-in safety mechanisms that allow for easy reversal of automated actions. For instance, if the system is integrated into a larger Care Management System that automates some processes based on the expenditure predictions, there will be a ‘review and confirm’ step where authorized personnel can validate or reverse these automated actions before they are finalized.

**Accountability and Feedback Loop:**

We encourage a feedback loop with all user groups to understand the scenarios where the AI system’s output was overridden or disregarded. This feedback is invaluable for system improvements and future updates.

**Legal and Ethical Guidelines:**

It is imperative that all actions taken based on the system’s predictions adhere to the prevailing medical, ethical, and legal standards. Users should consult their organization’s policies and guidelines regarding the use of AI-based systems in decision-making processes.

**2. Detailed Description of the AI System’s Elements and Development Process**

**Design and Architecture**

The AI system for the medical expenditure prediction is designed with a focus on fairness, interpretability, and accuracy. It utilizes the Linear Rule Regression (LinRR) and Boolean Rule Column Generation (BRCG) algorithms to achieve these goals. The system follows a modular architecture, with distinct components for data processing, model training, fairness assessment, bias mitigation, and explainability.

**Data Source and Pre-processing**

The medical expenditure system uses a dataset containing anonymized patient data, including health status, demographics, and other relevant factors. The data collection process adheres to privacy laws and regulations, ensuring the removal or obfuscation of personally identifiable information (PII) to protect patient privacy.

Prior to training the AI models for medical expenditure prediction, a pre-processing phase is employed to ensure data quality, fairness, and non-discrimination. The following steps are undertaken:

**Data Cleaning**: Inconsistencies, errors, and duplicates in the medical expenditure dataset are meticulously removed to ensure data quality and integrity.

**Feature Selection**: Relevant features that contribute to medical expenditure prediction are identified using domain knowledge and feature importance techniques. This step ensures that only the most relevant features are used in the model training process.

**Data Normalization**: The data is normalized to achieve consistent scaling across all features. Normalization enables the models to train effectively without bias towards specific features that may have a larger scale.

**Bias Detection and Alleviation**: The AI system incorporates various metrics, including disparate impact, average odds difference, statistical parity difference, equal opportunity difference, and Theil index, to detect and quantify biases in the dataset. To mitigate bias, the system employs techniques such as reweighing, prejudice remover, and disparate impact remover.

The reweighing algorithm is applied during the pre-processing stage to transform the training dataset and reduce bias. The algorithm calculates the probability of a favorable outcome for privileged groups and adjusts the training data accordingly, aiming to prevent bias in the model. After the transformation, the model is re-evaluated to ensure a fair comparison with the previous model.

Computational Resources:

• During the data pre-processing and source code development, it is advisable to use a development environment with at least 16 GB RAM and an Intel Core i7 processor or equivalent. Dual-monitor setups can facilitate code review and debugging.

**Model Training and Validation**

The AI system employs two algorithms for medical expenditure prediction: Linear Rule Regression (LinRR) and Boolean Rule Column Generation (BRCG).

LinRR: LinRR is a version of Generalized Linear Rule Models (GLRM) specifically designed for predicting real-valued outcomes. It produces weighted combinations of AND rules, which offer both accuracy and interpretability.

BRCG: BRCG generates simple OR-of-ANDs classification rules and is used for high-cost patient identification.

Both algorithms undergo training and validation processes to evaluate their performance. The models are trained on the original data, and fairness and accuracy metrics are assessed. Disparate impact and average odds difference are specifically used to identify unfairness favoring privileged groups.

Computational Resources:

• Training Phase: The system’s models are trained on high-performance computing clusters with GPUs. A minimum of 32 GB RAM and Nvidia Tesla V100 GPUs are essential for efficient model training.

• Testing Phase: Testing is carried out on machines with similar specifications to the training environment but can be performed on less powerful hardware, like Nvidia Tesla P100 GPUs and 16 GB RAM.

• Validation Phase: Validation is generally conducted on machines with at least 16 GB RAM and an Intel Core i7 processor or equivalent to ensure the model’s robustness and generalizability.

**Explainability**

To enhance the explainability of the AI models, the **LimeTabularExplainer** method is employed. It generates explanations for the model predictions, providing insights into the label predictions made by the model and illustrating the importance of features and their contributions to the predictions. The explanations help developers and users gain a better understanding of the model’s behavior and decision-making process.

**System Performance and Testing**

The medical expenditure system utilizes a dataset divided into training, validation, and testing sets. The training set is used to train the models, the validation set helps in parameter tuning and preventing overfitting, and the testing set evaluates the model’s performance on unseen data, simulating real-world scenarios.

The performance of the models is assessed using various metrics, including R-squared (R2) scores and accuracy. The system is regularly monitored and updated to detect and address emerging biases. This monitoring process ensures the system’s compliance with regulations and allows it to adapt to changing trends.

**Explainability Features**

A key feature of our AI system is its explainability. We believe it’s essential for data scientists, medical professionals, and system users to understand the reasoning behind the decisions made by the system. The system incorporates built-in human oversight to ensure that the outputs are correct. The models are designed to be transparent, so their workings can be easily understood. They are also monitored during deployment, and adjustments can be made as necessary.

To ensure transparency and explainability in our medical expenditure predictions, we use three different explanatory models: Linear Rule Regression (LinRR), Generalized Linear Rule Models (GLRM), and Boolean Rule Column Generation (BRCG).

**• Linear Rule Regression** (LinRR): LinRR is a version of GLRM for predicting real-valued outcomes. It yields weighted combinations of AND rules that are accurate and have the interpretability of linear models and generalized additive models. LinRR is specifically used for the task of predicting medical expenditures.

**• Generalized Linear Rule Models** (GLRM): GLRM extends the rule-based explanatory model by allowing for more general linear relationships. It produces models that are weighted combinations of rules, which provide insight into medical expenditure predictability. The GLRM algorithm also has the option of combining rules with linear terms.

**• Boolean Rule Column Generation** (BRCG): BRCG generates simple OR-of-ANDs classification rules and is featured for the task of identifying high-cost patients. These rules help identify patterns and factors contributing to high medical costs.

The use of multiple explanatory models allows us to cross-verify the explanations and ensure their reliability. This is particularly important in the context of medical expenditure, where understanding the reasons behind predictions is crucial for decision-making.

In addition to these explainability features, we use GitHub to track and audit the development of our AI system. Every change is logged and timestamped to ensure transparency and accountability. We also prioritize the security of third-party tools and keep them up to date to maintain high standards and comply with relevant regulations.

**Trade-off Considerations**

Designing the system required careful consideration of various trade-offs. For instance, the use of complex models allows us to capture intricate relationships in medical expenditure predictions, but it may reduce explainability. To mitigate this challenge, we have incorporated explainable AI tools and models to balance performance and transparency.

When it comes to data privacy, we prioritize the accuracy of medical expenditure predictions while ensuring the protection of individual privacy. To achieve this, all data is anonymized, and strict data governance protocols are implemented.

**Validation and Compliance**

The AI system’s validation involves a rigorous testing process. We evaluate the model’s predictive power and robustness using well-defined metrics such as accuracy, R-squared, and fairness metrics. These metrics help us assess the system’s performance and ensure its reliability. To ensure robustness, the system is tested under different scenarios, including stress tests and edge cases. We simulate various situations to verify the system’s response and ensure its reliability and consistency.

Regarding cybersecurity, the system is designed to comply with industry-standard security protocols and follows best practices to safeguard against potential threats. We employ encryption techniques to protect sensitive data, implement secure authentication mechanisms, and regularly update our system’s defenses against emerging cyber risks.

To ensure compliance with relevant regulations and guidelines, we work closely with legal and regulatory experts. Our AI system is developed and deployed in alignment with applicable laws, including data protection and privacy regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA).

Furthermore, our AI system undergoes regular audits and assessments to maintain compliance with internal policies and external regulations. These audits are performed by independent third-party organizations that assess the system’s adherence to ethical guidelines, data protection standards, and fairness considerations.

We also prioritize ongoing monitoring and evaluation of the AI system’s performance in real-world settings. This includes soliciting feedback from medical professionals, stakeholders, and end-users to identify areas of improvement and address any potential biases or limitations in the system’s predictions.

Additionally, we maintain a comprehensive documentation process to ensure transparency and accountability. This includes documenting the development process, data sources, model architecture, and any updates or modifications made to the system over time. This documentation allows for traceability and enables effective auditing and validation of the AI system’s performance.

Overall, our validation and compliance efforts aim to establish trust and confidence in the AI system’s reliability, security, and adherence to ethical and legal standards. By prioritizing transparency, accountability, and robust evaluation, we strive to ensure the responsible and ethical use of AI in the domain of medical expenditure predictions.

**3. Detailed Information about the Monitoring, Functioning, and Control of the AI System**

The medical expenditure prediction AI system is an advanced, data-driven platform designed with a strong commitment to ethical guidelines, data privacy regulations, and fairness. It employs cutting-edge machine learning models and explainable AI (XAI) algorithms to ensure transparency and accountability in predicting medical expenses. However, like any complex system, it possesses certain capabilities, limitations, and potential risks that require careful understanding and management. This section provides a detailed explanation of these aspects along with the necessary human oversight measures.

**System Capabilities**

The AI system can process significant volumes of anonymized patient data to predict medical expenditure. It utilizes a primary predictive model known as XGBoost, which is a gradient boosting framework that combines multiple decision trees to create a powerful learner. XGBoost is widely recognized for its ability to handle complex relationships, outliers, and large feature sets.

Additionally, the system incorporates three explanatory models: BooleanRuleCG (BRCG), LogisticRuleRegression (LRR), and Generalized Linear Rule Models (GLRM). These models generate interpretable rules that explain the decisions made by the XGBoost model, ensuring transparency in the prediction process.

Moreover, the AI system employs Protodash and CEM, XAI algorithms that provide explanations of medical expenditure predictions to healthcare professionals and patients, respectively. These tools enable users to comprehend the system’s predictions and gain insights into the factors influencing medical expenses.

**System Limitations and Potential Unintended Outcomes**

While the system is robust and highly efficient, it does have certain limitations. Its performance depends on the quality and completeness of the data it processes. Incomplete or biased data may lead to less accurate predictions or unintentional disparities in healthcare outcomes.

Furthermore, similar to any predictive model, the AI system cannot guarantee absolute accuracy. False positives (predicting higher medical expenses for individuals who would have incurred lower costs) and false negatives (underestimating medical expenses for individuals who would have incurred higher costs) are inherent risks. Efforts are made to minimize these risks, but they cannot be entirely eliminated.

There may also be potential unintended outcomes if the system’s predictions are interpreted without considering the contextual information. The rules generated by the explanatory models should be regarded as indications based on historical data rather than absolute truths.

**Sources of Risks**

The primary source of risk in the AI system stems from data bias. Bias in the training data can lead to unfair treatment or disparities in medical expenditure predictions, disadvantaging certain patient groups. To mitigate this, the AI system incorporates a Reweighing algorithm that reduces bias in the training data, promoting fairness and non-discrimination.

Another risk lies in over-reliance on the system’s predictions without human oversight, which may result in unnoticed errors. To address this, the AI system is designed with built-in human oversight mechanisms to ensure the accuracy and appropriateness of the predictions.

**Input Data**

The AI system utilizes a dataset containing anonymized patient information relevant to medical expenditures. This data complies with data privacy regulations, including GDPR, to ensure the protection of patient information. Prior to training the machine learning models, all data undergoes thorough cleansing, normalization, and processing to maintain its quality and relevance, essential for accurate predictions.

Additionally, where applicable, the AI system incorporates external datasets and information to enhance the accuracy of its predictions. For example, macroeconomic indicators such as inflation rates, average income levels, and regional healthcare expenditure data can be considered to provide a broader context for predicting medical expenses. However, the inclusion of such external data is subject to availability, relevance, and strict compliance with data privacy laws and regulations.

It is important to note that while the AI system has the capability to process a wide range of data, it is designed to avoid using sensitive personal attributes, such as race or gender, in its prediction process. This ensures that the system’s predictions of medical expenditure remain fair, unbiased, and non-discriminatory.

In conclusion, the medical expenditure prediction AI system is a sophisticated tool that automates and streamlines the prediction process. By leveraging advanced machine learning techniques and explainable AI, it ensures accurate, transparent, and fair predictions. However, the system also has its limitations and potential risks, which are addressed through comprehensive data processing, bias mitigation algorithms, and robust human oversight measures. As the system continues to evolve and improve, it holds the potential to contribute to more efficient healthcare planning, resource allocation, and informed decision-making in the medical field.

**Human Oversight Measures**

Article 14 of the AI Regulation mandates an assessment of human oversight measures to ensure that AI systems are transparent, fair, and in line with regulatory and ethical guidelines. Our AI system for medical expenditure prediction incorporates several layers of human oversight to fulfill these requirements.

**Transparency and Accountability**

**• Version Control**: The use of GitHub for version control enables comprehensive tracking of all changes to the codebase. Each change is logged and timestamped, providing a transparent history of modifications to the AI system.

**• Audit Trails**: The system maintains an audit trail for model predictions, allowing for a review of decisions and actions. This traceability is crucial for assessing the system’s behavior and for future audits.

**• Compliance Audits**: Independent third-party organizations regularly perform audits to ensure the system’s compliance with ethical guidelines, data protection standards, and fairness considerations. These audits validate that human oversight is effectively implemented.

**Monitoring and Adaptability**

**• Continuous Monitoring**: During the system’s deployment, real-time monitoring tools track its performance and fairness metrics. Any deviations trigger alerts for immediate human review.

**• Feedback Loops**: Medical professionals, stakeholders, and end-users can provide feedback on the system’s predictions. This feedback is used to fine-tune the AI models and address any biases or limitations.

**• Update Protocols**: A standardized protocol is in place for updating the system based on monitoring results and feedback. This ensures that the system adapts to new data patterns and regulatory changes effectively.

**Decision Reviews**

**• Explainability Features**: The use of LimeTabularExplainer and other explainability models provides an understandable reasoning behind each medical expenditure prediction. This enables human reviewers to assess the logic behind the predictions.

**• Manual Overrides**: The system is designed with the capability for human operators to manually override predictions. This safeguard is crucial in situations where the AI system’s output may not fully capture the complexities of medical expenditure.

**Ethical and Regulatory Adherence**

**• Fairness Assessment**: The system uses fairness metrics like disparate impact and average odds difference to monitor its fairness. Human reviewers assess these metrics to ensure the system meets ethical and regulatory standards.

**• Data Protection**: Compliance with GDPR and HIPAA is regularly reviewed by legal experts in collaboration with the tech team, ensuring that data protection measures are up to date.

By implementing these layers of human oversight, we aim to ensure that the AI system for medical expenditure prediction aligns with the ethical and regulatory mandates specified in Article 14. This comprehensive approach to human oversight ensures that the system is transparent, adaptable, and ethically responsible.

**4. Detailed Description of the Risk Management System**

In accordance with Article 9 of the EU AI Act, our medical expenditure AI system operates within a robust risk management framework that is continuously updated and thoroughly documented to ensure transparency and compliance.

**Risk Identification and Analysis**

The first step in our risk management process is to identify and analyze known and foreseeable risks associated with the operation of our AI system. This involves a comprehensive examination of the system’s design, the features it utilizes, and the decision-making processes it employs. We pay special attention to potential misuse scenarios and the risks associated with them. Factors such as data quality, potential bias in the training data, overfitting, and the possibility of misinterpreting model explanations are carefully considered during the risk identification phase.

**Risk Estimation and Evaluation**

We estimate and evaluate risks that arise from the intended use of the medical expenditure AI system and under conditions of reasonably foreseeable misuse. Our system utilizes a diverse range of medical expenditure data to model various patient scenarios and test the system’s responses. Risks such as inaccurate predictions, privacy concerns, and miscommunication of medical expenditure estimates are evaluated and quantified.

**Post-Market Monitoring Data Analysis**

In compliance with Article 61, we implement a post-market monitoring system to gather and analyze data. This ongoing process allows us to identify emerging risks, monitor model performance, and detect any unexpected system behavior. By observing the system’s performance and impact in real-world settings, we can ensure that it remains aligned with its intended purpose and does not cause unintended harm.

**Risk Management Measures**

Upon identification and evaluation of risks, we adopt appropriate risk management measures. The system design, including the machine learning models and explanatory algorithms, is continuously reviewed and updated to eliminate or minimize risks. Our risk management measures encompass the application of all system requirements and state-of-the-art practices in the field of AI technology.

We implement mitigation and control measures for risks that cannot be completely eliminated. These measures include regular bias audits, data quality checks, and performance evaluations of the AI system. Furthermore, we provide stakeholders with comprehensive information about the system’s operation and associated risks, ensuring transparency throughout the process.

**Residual Risk Management**

Residual risks associated with each identified hazard and the overall residual risk of the system are considered acceptable as long as the system is used for its intended purpose. We communicate residual risks to users to ensure they have a clear understanding of the system’s limitations and potential issues.

**Testing**

Our high-risk AI system undergoes rigorous testing to determine the most appropriate risk management measures. The system’s performance is evaluated against defined metrics and probabilistic thresholds that are relevant to its intended purpose. Testing procedures are designed to ensure consistent performance and compliance with the requirements outlined in this section. These procedures are suitable to achieve the system’s intended purpose without going beyond what is necessary.

Testing is conducted throughout the development process, prior to deployment, and continues after deployment to ensure ongoing compliance and performance.

**User Considerations**

When developing and updating the risk management system, we consider the technical knowledge, experience, education, and training expected from users. We also take into account the environment in which the system is intended to be used and its potential impact, including considerations related to patients of different age groups.

**Compliance with Applicable Regulations**

As a healthcare organization, we ensure compliance with all applicable regulations and directives relevant to the medical field. The risk management procedures described in this document form part of our institution’s established risk management framework, in accordance with applicable regulations.

**5. A Description of Any Change Made to the System Through Its Lifecycle**

**Version History**

The AI-based Medical Expenditure System has undergone several updates to improve its performance, security, and user experience. Below is a summary of the major changes:

**Version 1.0.0**

• Initial release with XGBoost algorithm for medical expenditure prediction.

• Basic explainability features using ProtoDash and CEM.

**Version 1.1.0**

• Introduced Linear Rule Regression (LinRR) and Boolean Rule Column Generation (BRCG) algorithms for enhanced explainability.

• Improved data pre-processing techniques for better data quality.

**Version 1.2.0**

• Added support for additional operating systems, including Linux and macOS.

• Implemented advanced bias detection and mitigation algorithms.

**Version 1.3.0**

• Upgraded to aix360 v0.2.1.

• Added LimeTabularExplainer for improved model explainability.

• Security patches and bug fixes.

**Version 1.4.0**

• Introduced multi-language support for the user interface.

• Enhanced post-market monitoring features for better compliance with Article 61.

**Rollback Procedures**

In case of critical issues or system failures, rollback procedures are in place to revert the system to its last stable version. All changes are thoroughly tested in a staging environment before deployment to minimize the need for rollbacks.

**6. List of the Harmonized Standards Applied**

In the development and deployment of our medical expenditure AI system, we have integrated several harmonized standards to meet the legal and technical requirements of various jurisdictions, including the European Union. These standards encompass data protection, machine learning, and explainability, ensuring the system operates with integrity and compliance.

**ISO/IEC 27001: Information Security Management:** Given the sensitive nature of medical data, we adhere to ISO/IEC 27001 principles to establish robust information security management systems. This standard ensures that comprehensive measures are in place to safeguard the confidentiality, integrity, and availability of the medical data we handle.

**ISO/IEC 27701: Privacy Information Management:** To comply with GDPR and other relevant privacy laws, we have implemented ISO/IEC 27701 guidelines. This standard focuses on Privacy Information Management Systems (PIMS), enabling us to effectively manage and mitigate privacy risks associated with handling personal health information.

**ISO/IEC 38505-1: Governance of Data:** Our adherence to ISO/IEC 38505-1 ensures responsible data governance practices, specifically pertaining to decision-making processes. By following this standard, we maintain the integrity and quality of the data used to train our medical expenditure AI system.

**IEEE P7003: Algorithmic Bias Considerations:** In line with our commitment to fairness, we adopt the guidelines of IEEE P7003 to address and mitigate algorithmic bias. We leverage the Reweighing algorithm to reduce bias in the training data, ensuring equitable treatment and non-discrimination in the medical expenditure AI system.

**ISO/IEC TR 24028:2020: Artificial Intelligence (AI) — Overview of trustworthiness in Artificial Intelligence:** Trustworthiness is paramount in our medical expenditure AI system. We follow the principles outlined in ISO/IEC TR 24028:2020 to ensure the system’s robustness, accuracy, privacy, transparency, and explainability. These principles enhance user confidence and contribute to the overall reliability of the system.

In cases where specific harmonized standards are not applicable, we rely on industry best practices within the machine learning and AI domain to fulfill the requirements set out in the medical sector. These include:

**Data Preprocessing and Model Training**: To ensure the accuracy and fairness of our AI system, we employ advanced data preprocessing techniques. These include data cleaning, feature selection based on domain knowledge, data normalization, and bias reduction methodologies. The primary predictive model utilized in our system is XGBoost, renowned for its efficiency and performance in handling medical datasets.

**Explainability and Transparency**: We prioritize the explainability of our medical expenditure AI system to enable healthcare professionals and stakeholders to understand the decision-making process. To achieve this, we employ three explanatory models: BooleanRuleCG, LogisticRuleRegression, and Generalized Linear Rule Models (GLRM). These models provide clear insights and enhance transparency in the medical expenditure decisions made by the system.

**Use of XAI algorithms**: Our system integrates eXplainable Artificial Intelligence (XAI) algorithms, including Protodash and CEM. These algorithms facilitate the explanation of medical expenditure decisions to healthcare professionals and other stakeholders, ensuring clear and understandable insights.

By adhering to these harmonized standards and industry best practices, our medical expenditure AI system operates with integrity, fairness, transparency, and security. These measures align with the principles and requirements set forth in the medical sector, promoting trust and accountability in healthcare decision-making processes.

**7. EU Declaration of Conformity**

1 AI system name and type and any additional unambiguous reference allowing identification and traceability of the AI system:

• AI system name: Medical Expenditure AI Model

• AI system type: Machine learning-based medical expenditure system

• Additional reference: Version 1.2.0

2 Name and address of the provider or, where applicable, their authorized representative:

• Name: Jane Doe

• Address: 221B Baker Street, London

3 A statement that the EU declaration of conformity is issued under the sole responsibility of the provider:

• We hereby declare that this EU declaration of conformity is issued under the sole responsibility of Jane Doe.

4 A statement that the AI system in question is in conformity with this Regulation and, if applicable, with any other relevant Union legislation that provides for the issuing of an EU declaration of conformity:

• We confirm that the Medical Expenditure AI Model is in conformity with the AI Act and, where applicable, with any other relevant Union legislation that provides for the issuing of an EU declaration of conformity.

5 References to any relevant harmonized standards used or any other common specification in relation to which conformity is declared:

• The Medical Expenditure AI Model conforms to the following standards and specifications:

◦ ISO/IEC 27001: Information Security Management

◦ ISO/IEC 27701: Privacy Information Management

◦ ISO/IEC 38505-1: Governance of Data

◦ IEEE P7003: Algorithmic Bias Considerations

◦ ISO/IEC TR 24028:2020: Artificial Intelligence (AI) — Overview of trustworthiness in Artificial Intelligence

6 Where applicable, the name and identification number of the notified body, a description of the conformity assessment procedure performed, and identification of the certificate issued:

• Notified body: English Certification Agency

• Identification number: ECA-6272LON

• Description of conformity assessment procedure: The Medical Expenditure AI Model underwent a comprehensive evaluation by the English Certification Agency. The assessment included a review of the system’s design, development processes, and adherence to the applicable standards and regulations. Additionally, the AI model was subjected to rigorous testing to ensure its performance, reliability, and compliance with the specified requirements.

• Certificate issued: Certificate of Conformity (Certificate No. AI-5678)

7 Place and date of issue of the declaration, name and function of the person who signed it, as well as an indication for, and on behalf of whom, that person signed:

• Place of issue: London

• Date of issue: May 15, 2023

• Person who signed: Jane Doe

• Function: AI Lead Engineer and Legal Representative

• Signed on behalf of: The ACME Company

**8. Detailed Description of the System in Place to Evaluate the Medical Expenditure AI System Performance in the Post-market Phase**

After the deployment of the AI system, it is crucial to continuously monitor and evaluate its performance to ensure that it maintains its intended functionality, accuracy, and reliability. In the post-market phase, the AI system’s performance evaluation comprises several critical components, including ongoing monitoring, routine performance evaluations, bias detection and mitigation, and system updates. This section provides a comprehensive description of the system and procedures we have established to effectively evaluate the AI system’s performance in the post-market phase.

**Ongoing Monitoring**

In the post-market phase, the AI system undergoes continuous monitoring to track its performance and identify any deviations in the system’s behavior or the data it processes. This ongoing monitoring process entails collecting and analyzing data on the system’s performance metrics, such as accuracy, reliability, and efficiency. The monitoring process is automated, utilizing real-time alerts to notify the team of any significant changes in these metrics that could indicate issues with the system’s functionality or performance.

The AI system’s performance metrics are compared against predefined thresholds and historical benchmarks to identify any anomalies or unexpected changes. For example, if the system’s accuracy or reliability metrics fall below the established thresholds, an alert is triggered for further investigation. The monitoring process also includes tracking the system’s performance across different patient groups to detect and address any potential bias or discrimination.

**Routine Performance Evaluations**

In addition to ongoing monitoring, we regularly conduct routine performance evaluations of the AI system. These evaluations involve a comprehensive review of the system’s performance metrics and a detailed analysis of the system’s decisions and their impacts.

The routine performance evaluations are conducted on a quarterly basis or as needed based on the ongoing monitoring results. The evaluations include an analysis of the system’s accuracy and reliability metrics, a review of the system’s expenditure decisions and their outcomes, and a comparison of the system’s performance against other benchmark models or industry standards.

**Bias Detection and Mitigation**

Detecting and mitigating bias is a crucial aspect of our post-market evaluation process. We utilize advanced tools and techniques, such as the Fairness Indicators framework, to detect and quantify bias in the AI system’s decisions. This framework allows us to compute fairness metrics and compare the distribution of expenditures for different patient groups.

To mitigate any identified bias, we employ a Bias Correction algorithm that adjusts the decision-making process to ensure equitable expenditure outcomes across different groups. This algorithm reshapes the training data to minimize disparities and promote fairness.

Whenever bias is detected, a thorough investigation is conducted to identify the root cause and implement appropriate mitigation measures. This may involve retraining the AI model, adjusting the model’s parameters, or revising the data pre-processing procedures.

**System Updates**

Based on the results of ongoing monitoring and routine performance evaluations, we may implement updates to the AI system to enhance its performance or functionality. These updates could entail adjusting the AI model’s parameters, incorporating new features or data sources, or upgrading the AI algorithms.

Each update undergoes rigorous validation and testing before deployment to ensure that it improves the system’s performance and does not introduce new risks or biases. All updates are logged and documented to maintain a transparent record of the system’s evolution and improvements.

**Post-market Audit and Compliance**

To ensure transparency and accountability, we conduct regular audits of our post-market evaluation procedures. These audits involve a thorough review of our monitoring and evaluation practices, performance metrics, bias mitigation measures, and system updates.

We also ensure that our post-market evaluation procedures comply with all relevant regulations, including the GDPR and applicable healthcare data protection laws. This involves conducting regular compliance checks and maintaining comprehensive documentation of our evaluation procedures and their outcomes. Any non-compliance issues are promptly addressed, and corrective actions