General

ID[1]		
Use case name	Explainable artificial intelligence for Genomic Medicine	
Application domain	Healthcare	
Deployment model	Cloud services	
Status	Prototype	
Scope[2]	To explain reason and basis behind AI-generated findings in genomic medicine	
Objective(s)	To improve the efficiency of investigatory work for experts in genomic medicine.	
Narrative	Short description (not more than 150 words)	This technology was deployed to improve the efficiency of investigatory work for experts in genomic medicine, utilizing training data and a knowledge graph that made use of public databases and medical literature databases in the field of bioinformatics. It was then evaluated to validate that it was possible to find and link the basis supporting findings with regard to phenomena whose interrelationships are only partially understood.
	Complete description	Deep Learning is one of the most representative technologies in recent AI and shows high performance in pattern recognition and analysis. However, as it cannot explain the reasons for its judgment, it is called "black box AI." There is a graph-structured data based machine learning technology called "Deep Tensor" that can directly analyze the relations among numerous pieces of real-world data ranging from intercompany transactions to material structures. Additionally, there is also a technology for building a large-scale knowledge base, which is called a "knowledge graph" and consists of vast knowledge existing around the world such as academic papers, by using our unique technology. This technology identifies the factors (partial graphs) that had a significant influence on an inference and coordinates these with partial graphs from a knowledge graph, building a series of pieces of information in the form of connections in the knowledge graph as the basis for the findings. People can combine these two technologies and develop a system that enables AI to explain the reasons and basis (evidence) for its judgment. A use case of applying this explainable AI is genomic medicine (for cancer treatment). The latest genomic medicine helps detect patients' genetic defects that have caused disease (cancer) and uses therapeutic drugs that affect cancer cells produced by such genetic defects. In genomic medicine today, a patient's normal and cancerous cells are analyzed with a next-generation sequencer; then, a medical team uses the obtained genetic data to identify a causal gene and determines the recommended treatment. It takes at least two weeks for the medical team to conduct an examination after

		completing genetic analysis. Unless the cost and time problems are solved, spreading this advantageous genomic medicine far and wide will be difficult. In this use case, the explainable AI trained Deep Tensor using 180,000 pieces of disease mutation data, successfully embedding more than 10 billion pieces of knowledge from 17 million medical articles and other materials into Knowledge Graph. Inputting genetic mutation data into this system enables Deep Tensor to infer disease-causing factors and enables Knowledge Graph to find medical evidence to justify the obtained results. Medical specialists then simply need to review the flow of obtained inference logic, thereby reducing the period between analysis and report submission significantly— from two weeks to a single day.		
Stakeholders[3]	Doctors of genomic medicine, researchers of genomic medicine, patients			
Stakeholders' assets, values[4]	Reducing the determination periods, maintaining the accuracy of predication as well as manual predication			
System's threats and vulnerabilities[5]	Update knowledge graph lately, huge size of knowledge graph			
Key performance indicators (KPIs)	ID	Name	Description	Reference to mentioned use case objectives
	1	Accuracy of predication	Proportion of the true positives and true negatives combined in the disease predication by AI	Improve accuracy
	2	Appropriateness of explanation	Proportion of the appropriate flow of obtained inference logic	Improve efficiency
	3	Determination periods	The periods that a medical team uses the obtained genetic data to identify a causal gene and determines the recommended treatment.	Improve efficiency
AI features	Task(s)	Knowledge processing & discovery, Natural Language Processing, Inference, Prediction		
	Method(s)[6]	Knowledge Graph, Deep Learning (Deep Tensor), Natural Language Processing		
	Hardware[7]			
	Topology[8]			
	Terms and concepts used[9]	Knowledge Graph, Deep Learning, Natural Language Processing, Explainable AI		
Standardization opportunities/ requirements				
Challenges and issues	Challenges: To reduce experts' workloads, shortening determination periods in genomic medicine.			

	Issues: The inability to explain the reason behind inferences from the learning algorithm of black-box AI.	
Societal concerns	Description	1, Accountability for using AI in medical examination 2, Incorrect explanation will cause the determination periods increasing.
	SDGs[10]	Good health and well-being for people

Data (optional

Data characteristics	
Description	Knowledge Graph
Source[11]	Disease mutation data, medical articles and other materials
Type[12]	Graph-structured data in RDF format
Volume (size)	180,000 pieces of disease mutation data, more than 10 billion pieces of knowledge from 17 million medical articles
Velocity (e.g. real time)[13]	Batch
Variety (multiple datasets)[14]	multiple datasets
Variability (rate of change)[15]	Static
Quality[16]	High

Process scenario (optional)

Scenario conditions					
No.	Scenario name	Scenario description	Triggering event	Pre-condition[17]	Post-condition[18]
1	Training	Train a model (deep tensor) with training data set	Disease mutation data for training is ready	To extract disease mutation data from knowledge graph	
2	Evaluation	Evaluate whether the trained model(deep tensor) can be deployed	Completion of training		Meeting accuracy requirement of predication (e.g. accuracy of predication is 90% or more) is the "success" condition
3	Execution	1, Enables Deep Tensor to infer disease-causing factors 2, Enables Knowledge Graph to find medical evidence to justify the obtained results.	The genetic mutation data is ready	To extract mutation data from knowledge graph	

Training (optional)

Scenario name	Training				
Step No.	Event[19]	Name of process/Activity[20]	Primary actor	Description of process/activity	Requirement
1	Disease mutation data for training is ready	Extract training diseases mutation data	Doctors or researchers pf genomic medicine	Extract mutation data from knowledge graph	The software for processing RDF data base has to be provided by the AI solution provider
2	Completion of Step 1	Model training	AI solution provider	Train a model (deep tensor) with the training data set created by Step 1	

Specification of training data[21]	

Evaluation (optional)

Scenario name	Evaluation				
Step No.	Event[22]	Name of process/Activity[23]	Primary actor	Description of process/activity	Requirement
1	Completion of training	Extract evaluating diseases mutation data	Doctors or researchers pf genomic medicine	Extract diseases mutation data from knowledge graph	The software for processing RDF data base has to be provided by the AI solution provider
2	Completion of Step 1	Predication	AI solution provider	Given the mutation data from Step 1, predicate disease / non-disease using deep tensor models that were trained in the scenario of training	
3	Completion of Step 2	Evaluation	Doctors or researchers pf genomic medicine	Compare the result of Step 2 with that of human inspection	

Input of evaluation[24]	
Output of evaluation[25]	

Execution (optional)

Scenario name	Execution				
Step No.	Event[26]	Name of process/Activity[27]	Primary actor	Description of process/activity	Requirement
1	The genetic mutation data is ready	Extract genetic mutation data	Doctors or researchers pf genomic medicine	Extract the target of genetic mutation data from knowledge graph	The software for processing RDF data base has to be provided by the AI solution provider
2	Completion of Step 1	Predication	AI solution provider	Given the mutation data from Step 1, predicate disease / non-disease using deep tensor models that were trained in the scenario of training	
3	Completion of Step 2	Inference	AI solution provider	Enables Deep Tensor to infer disease-causing factors	
4	Completion of Step 3	Explanation	AI solution provider and Doctors or researchers pf genomic medicine	Enables Knowledge Graph to find medical evidence to justify the obtained results	

Input of Execution[28]	
Output of Execution[29]	

Retraining (optional)

Scenario name	Retraining				
Step No.	Event[30]	Name of process/Activity[31]	Primary actor	Description of process/activity	Requirement

Specification of retraining data[32]	

References

References						
No.	Type	Reference	Status	Impact on use case	Originator/organization	Link
1	Brochure				Fujitsu	http://journal.jp.fujitsu.com/en/2018/01/23/02/
2	Brochure				Fujitsu	http://www.fujitsu.com/jp/group/labs/en/business/artificial-intelligence/
3	Press Release				Fujitsu	http://www.fujitsu.com/global/about/resources/news/press-releases/2017/0920-02.html
4	Journal				Nature	http://s3-service-broker-live-19ea8b98-4d41-4cb4-be4c-d68f4963b7dd.s3.amazonaws.com/uploads/ckeditor/attachments/8429/04_UK_Fujistu_AI.PDF

Footnote

[1] Leave this cell blank.

[2] The scope defines the limits of the use case.

[3] Stakeholder involved in the scenario - examples are: type of organization; customers, 3rd parties; end users; humans; environment; negative stakeholders (attackers, criminals, etc).

[4] Assets and values that are valuable to the stakeholders and at the risk of being compromised by the AI system deployment – examples can include competitiveness; reputation or trust; fairness; safety; privacy; stability; etc.

[5] Threats and vulnerabilities can compromise the assets and values above. Examples are: different sources of bias; incorrect AI system use; new security threats; challenges to accountability; new privacy threats (hidden patterns).

[6] AI method(s)/framework(s) used.

[7] Hardware system used.

[8] Topology is the study of geometric forms differentiated by intersection and bifurcation. The term is used for the graphic aspects network architectures.

[9] Terms and concepts listed here can be used to extend the work of WG 1 (AWI 22989 and AWI 23053) as necessary.

[10] The Sustainable Development Goals (SDGs), otherwise known as the Global Goals, are a collection of 17 global goals set by the United Nations General Assembly. SDGs are a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity.
See URL for more details: http://www.undp.org/content/undp/en/home/sustainable-development-goals.html

[11] Origin of data, which could be from instruments, IoT, web, surveys, commercial activity, or from simulations.

[12] Structured/unstructured Images, voices, text, gene sequences, and numerical. Composite: time-series, graph-structured

[13] The rate of flow at which the data is created, stored, analysed, or visualized.

[14] Data from a number of domains and a number of data types. The wider range of data formats, logical models, timescales, and semantics complicates the integration of the variety of data.

[15] Changes in data rate, format/structure, semantics, and/or quality.

[16] Completeness and accuracy of the data with respect to semantic content as well as syntactical of the data (such as presence of missing fields or incorrect values)

[17] Describe which condition(s) should have been met before this scenario happens.

[18] Describe which condition(s) should prevail after this scenario happens. The post-condition may also define "success" or "failure" conditions.

[19] The event that triggers the step. This might be completion of the previous event.

[20] Action verbs should be used when naming activity.

[21] Training data can be further specified.

[22] The event that triggers the step. This might be completion of the previous event.

[23] Action verbs should be used when naming activity.

[24] Specify input of evaluation.

[25] Specify output of evaluation.

[26] The event that triggers the step. This might be completion of the previous event.

[27] Action verbs should be used when naming activity.

[28] Specify input of evaluation.

[29] Specify output of evaluation.

[30] The event that triggers the step. This might be completion of the previous event.

[31] Action verbs should be used when naming activity.

[32] Retraining data can be further specified.

