Combining Deductive and Statistical Explanations in the FRANK Query Answering System

Combining Deductive and Statistical Explanations in the FRANK Query Answering System.

Alan Bundy
School of Informatics
University of Edinburgh,
Edinburgh, EH8 9AB, UK
Email: A.Bundy@ed.ac.uk

Kwabana Nuamah
School of Informatics,
University of Edinburgh,
Edinburgh, EH8 9AB, UK
Email: K.Nuamah@ed.ac.uk

***Abstract*—Both symbolic and sub-symbolic AI have their limitations, but their combination can be more than the sum of their parts. For instance, statistical machine learning has been hugely successful at classification and decision-making tasks, but not so good at deliberative systematic reasoning nor at explanation. We argue that by combining symbolic and sub-symbolic reasoning into *hybrid* systems, the whole will be more than the sum of its parts.**

To illustrate the potential of hybrid AI system, we describe the FRANK query answering system. FRANK infers new knowledge from the diverse and immense knowledge sources on the Web, using a combination of both deductive and statistical reasoning. This enables it to make predictions. For instance, to answer the question "Which country in Europe will have the highest GDP growth rate by 2032?", it (i) decomposes Europe into its constituent countries, (ii) then for each country uses regression over their previous GDP growth rates to extrapolate each to 2032 and (iii) then returns the country which is predicted to then have the maximum value. The decompositions are explained deductively and the regressions by a prediction model that can be rendered graphically. This explanation of FRANK's reasoning merges deduction and statistics.

In this paper, we highlight recent work on FRANK that focus on leveraging hybrid AI to tackle question answering with emphasis on explainability of the inference process and its inferred answers. We aim for whole system reasoning; that is, we are automating the choices of knowledge sources and the planning that constructs the inference process from the facts found in these knowledge sources. We intend that these 'engineering' choices are also explained to the user.

I. Introduction

We argue for a hybrid and compositional approach to Artificial Intelligence that amalgamates symbolic and sub-symbolic AI [1]. Such hybrid systems can combine the strengths of symbolic and sub-symbolic approaches and curtail their weaknesses. We illustrate this argument using the FRANK (Functional Reasoner for Acquiring New Knowledge) query-answering system. We think that it is a good vehicle for exploring the potential of hybrid reasoning. It is one of the longest standing applications of AI and many other AI problems can be framed as QA. Our aim in combining hybrid and compositional AI is to achieve the following properties discussed in [1]:

- **Interpretability:** a system's inner workings are inspectable by a human user.
- **Generalisability:** a system can deal with a wide range of tasks, including engineering aspects of the tasks that are usually done manually.
- **Robustness:** a system can deal with noise, incomplete knowledge, uncertainty and changes of both the environment and the task.

In this paper we describe our work-plan for achieving them.

II. Limitations of Existing Approaches

The strengths and shortcomings of both statistical machine learning (ML) and symbolic AI paradigms are well documented.

- Spurred by the European Union's General Data Protection Regulation (GDPR), there has been a great deal of research into extracting explanations from ML systems (XAI), which are notoriously opaque. Progress, so far, has been limited with most explanations being geared to system developers rather than lay people. Explanations often deal just with special cases and often apply only to image recognition.
- Although ML has produced some startling successes, they tend to have very narrow scope, e.g., image classification, natural language processing, game playing, etc. Generalising and combining these successes has been limited.
- ML requires large sets of training data or a long sequence of rewards and punishments. Where these do not yet exist, progress is not possible. Any biases in training process are then built into the learnt system.
- Some of the pioneers and advocates of ML, discuss ML's current inability to perform deliberate systematic reasoning and planning [2], as described by Kahneman's *'System 2'* reasoning [3].
- Symbolic AI systems are often brittle and limited in their scope. They lack the ability of ML systems to take advantage of huge amounts of information and to cope with uncertainty and noise.
- Attempts to model commonsense reasoning symbolically have made slow progress and are very limited in scope [4].

Both approaches to AI are heavily dependent on designs and inputs from humans. For instance,

- Designing the ML models to be used to tackle a classification or prediction task.
- Choosing and/or collecting the training data set.
- Manual construction of a logical representation of the environment.

We emphasize the claim in [1], echoed in §I, that a hybrid and compositional approach to AI is needed to tackle these limitations and realise the full potential of AI. We demonstrate how a whole-systems approach to AI, which leverages both symbolic and sub-symbolic methods in a framework, can lead to solutions that are not possible by either one of these paradigms alone. For instance, symbolic reasoning is required to decide which sub-symbolic methods to use and which data sources they should draw on. In this paper, we identify some of the work required to realise such a whole-systems approach.

III. The FRANK Query Answering System

We illustrate this claim via the FRANK (Functional Reasoner for Acquiring New Knowledge) query answering system. FRANK applies inference to knowledge sources on the World Wide Web to answer queries mostly requiring numeric answers. In particular, it can answer queries whose answers are not prestored and reliably assign an uncertainty to the answers. It applies deductive, arithmetic and statistical reasoning to the results of information retrieval.

Although FRANK's main focus is on estimating the values of numeric attributes, it sometimes returns qualitative answers, e.g., the query "Which country will have the largest population in Africa in 2025?" returns the name of the African country with the maximum estimated population.

FRANK combines the following techniques:

- The knowledge required to answer a query is retrieved from a wide variety of different knowledge sources on the Web. FRANK combines knowledge from multiple sources in different formats. It employs APIs for each of the common knowledge formats in order to match the knowledge sought to the knowledge sources from which it retrieves it.
- This knowledge is then dynamically curated into a common format and stored in a query-specific ontology. This enables its inference operations to combine knowledge from diverse sources. Its common format is *alists*, i.e., sets of attribute/value pairs. These pairs are extracted from the particular knowledge item, e.g., the $Subject$, $Predicate$ and $Object$ attributes, but also augmented with attribute values from the source itself, e.g., the $Time$, $Units$ and $Uncertainty$ attributes. Alists provide the flexibility we need to cope with relations of diverse type signatures.
- Queries are represented as conjunctions of alists. Some of their attributes' values are logical variables, whose value is unknown when the query is posed and which it is intended will be instantiated to a concrete value as a side effect of inference. These instantiations are propagated up the inference graph to the root and, thereby, provide the answer to the original query.
- FRANK uses *decomposition rules* to break goals into subgoals. For instance, it can decompose a question about a compound object into questions about its parts, then apply arithmetic functions, such as sum, maximum or minimum, to return the answer for the compound object. Similarly, it can decompose a requested prediction about a future time into questions about past times, apply regression to create a function, then extrapolate[1] this function to the future time to give the required prediction. Decomposition rules are constructed dynamically. For instance, the number of their children will depend the number of parts into which a compound object can be divided or the number of previous times over which a regression is applied.
- FRANK's inference constructs an inference graph with both AND and OR branches. Nodes are labelled either with (sub-)goals represented as alists or with inference rules that enable a parent alist to be inferred from its child alists. The root node is labelled with the original query. Inference is from the root query to the leaf facts. If the search is successful, then the inference graph will contain a proof as a subgraph, which will contain only AND branches. This proof tree will provide just those inference steps required to prove the query. During this proof, the query's variables will be instantiated to provide the required answer.
- The variables associated with the leaf alists of the proof subgraph are instantiated by matching them to facts stored in knowledge sources. The values of variables in parent alists are calculated by applying arithmetic and statistical aggregation operations to some of the variables in their child alists.
- Projected numeric values are assumed to have a Gaussian distribution and are returned as a mean and standard deviation. The mean is regarded as the answer and the standard deviation as an error bar on this answer. Aggregation operations are applied to the child nodes and the values and uncertainties are inherited from leaf to root. Leaf nodes are assigned uncertainty values associated with the knowledge source from which they are taken and these uncertainty values are also inherited up the inference graph as the standard deviations. Knowledge sources are initially assigned default uncertainties, but these uncertainties are incrementally adjusted by a Bayesian process which compares the compatibility of rival sources of the same knowledge. Some inference operations also add additional uncertainty that is inherent in their nature, e.g., regression/extrapolation.
- FRANK's compositional architecture and recursive inference algorithm [5], [6], [7] facilitate the generation of explanations of its answers. Answer alists are translated into English and statistical operations are displayed graphically (see Figure 1).

[1] Or interpolate, to predict the value for a time for which no answer has been pre-stored.

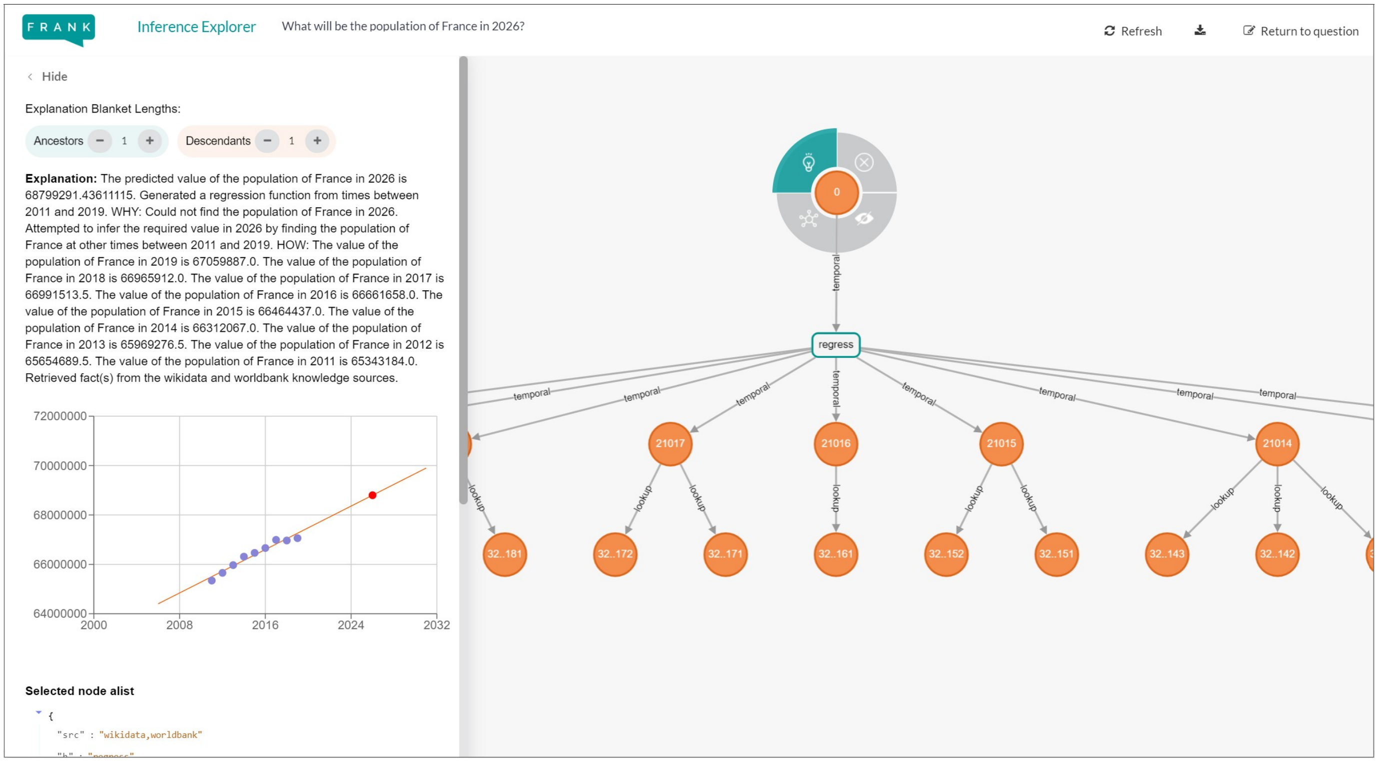

Fig. 1. FRANK's **Inference Explorer** showing the inference graph for the question "What will be the population of France in 2026?". A context menu on the nodes (e.g. on node 0 above) allows users to hide or show branches and to get details of the inference nodes. The "regress" node is selected and its explanation and regression graph are displayed in the sidebar.

IV. Explainable AI (XAI)

A critical aspect of intelligence is the ability to explain one's decision to others. As much as it important to make decisions or predictions in an automated way, it is also necessary to be able to convey the rationale behind the decision to others for several reasons. These include the desire to: (i) verify decisions, (ii) integrate them into larger decision processes, (iii) improve the system, (iv) comply with legislation and (v) allow the subjects of the decisions to appeal them. These social aspects of AI systems are necessary in order for intelligent systems to coexist with humans and provide the needed support to them.

Many organisations consider explainable decisions by AI systems as a key factor in the success and wide adoption of AI beyond academia and information technology companies. DARPA, an advocate for XAI [8], has challenge problems in data analytics and autonomy where an AI system learns a model to solve a task, generates an *explainable model*, and presents the explanations to the users through an *explanation interface*. John Lauchbury from DARPA has coined the term "The Third Wave of AI" to describe hybrid systems that combine symbolic (first wave) with sub-symbolic (second wave) approaches.

While the field of XAI is now gaining prominence, there is long history of work in developing explanations for intelligent systems. Several of these studies are captured in surveys of the field such as [9] and [10]. These earlier studies underline the fact that explanations are important for users interacting with complex, intelligent systems. Early work such as [11] and [12] highlight the fact that an ability to explain decisions is vital to the acceptance of intelligent systems. Other studies including [13] and [14] stress the importance of explanations in helping users check the accuracy of predictions, thereby increasing their confidence in the outputs of the machine learning system or providing the basis for appeal against them.

There are diverse viewpoints for explanations in intelligent systems. For example, explanation as the search for answers to why, how and what-if questions using causal chains, goal-plan-actions hierarchies and justifications [15] and explanation as a mechanism for discovery [16]. §VII outlines our plans to address what-if and how questions. More detail can be found in [17].

Work in [17] demonstrates how explanations are generated for inferences in FRANK. Explanations in an AI system can be local or global [10]. Local explanation provide insights into specific aspects of the AI system for specific cases. For instance, in a question such as "What will be the population of Europe in 2022?", a local explanation can focus on the prediction component. At the moment, most attempts at explanation in AI systems follow this approach. However, such localized explanations are not sufficient to give an adequate understanding of the answer returned. A global explanation provides a better mental model of the system to the user [18], [19]. Global explanations give insights into how the whole AI system works. In the above example, such an explanation will include details about the data selected, the reasons for decomposing the problem in a particular way, the reason for selecting the prediction algorithm used, etc. This gives users a better mental model of the system and provides them with sufficient insights that they can use to solve similar problems. Such global reasoning requires reasoning *about* the reasoning used to answer the query, i.e., *meta-level* reasoning. It is inherently symbolic. The symbolic and sub-symbolic explanations are complementary, as they deal with different aspects of the required explanations.

A. Interactive Inference Graph

FRANK constructs an inference graph that describes its inference operations. The entire inference graph is presented to the user in real-time via the Inference Explorer view shown in Figure 1. Since this inference graph can get quite large, the user can focus on just a sub-graph. The user specifies a node and the size of the sub-graph both above and below this node. We call this mechanism an *explanation blanket*. Each node has a context menu that allows a user to view the details of the node including its alist as well as an explanation for a specified explanation blanket size. The menu allows a user to choose how much of the inference graph to view by collapsing or expanding sub-graphs. Visually, this makes the entire inference transparent and enables users to decide on which parts of the inference to focus.

B. Graphical Explanations

Graphical explanations can be used to support the textual explanations generated for the alists. It helps users to visually understand parts of the inference process especially when the underlying aggregation operations are dense and difficult to understand textually. A typical example of this is regression. For problems that include regression as an inference step, FRANK returns not only the predicted value, but also the prediction function (e.g. the coefficients in a linear equation). Again, using the explanation blanket, FRANK plots (i) a prediction graph using the regression function calculated in the regress node; (ii) the underlying data points from the child nodes; and (iii) the extrapolated value which is propagated to the parent node. Figure 1 shows an example of the resulting graphical plot. Such prediction graphs provide insights into the intermediate inference steps in FRANK as it infers answer.

V. Compositionality

A key aspect of hybrid AI is compositionality. Below, we highlight some of the main arguments on the relevance of compositionality from work in [1].

It is not always possible to program or train one AI system to solve diverse kinds of problems. In many cases, even the ability of an expert to engineer a system to solve a range of problems, such as that of open-domain QA, is limited by the fact that one cannot anticipate all the possible kinds of questions to answer and how to combine existing AI modules achieve it. *Compositionality* provides a mechanism to compose solutions to problems by automating the combination of existing AI modules to solve new and varied problems. In this work, we use “compositionality” in a loose sense to include the entire spectrum from the high level integration of distinct AI components and systems, through to automatic program composition, all the way to the deeper level integration of knowledge representation, semantics and neural embedding.

In general, analysis of the text of a query is not enough to generate a problem-solving strategy to answer it. This is because the handling of failure of an initial reasoning attempt, e.g., the failure of direct look-up, cannot be anticipated or recovered from by inspection of the query text. For instance, in a question such as “Which country in Europe will have the highest GDP growth rate by 2032”, the kind of data retrieved (or the lack thereof) will determine if retrieval is sufficient, or whether regression on past data for prediction will be needed. Hence, the automatic formulation of new algorithms using existing components requires one to look beyond the initial query text and to work within the constraints of pre- and post-conditions of the underlying symbolic and sub-symbolic modules in order to combine them appropriately. Such meta-reasoning about the proof methods was pioneered by our group’s research on *proof planning* [20]. The plan for the proof of a theorem was constructed by reasoning that the post-conditions of an earlier proof method would enable the pre-conditions of a later one.

The textual explanation in Figure 1 illustrates this proof planning process by showing how the failure of direct look-up of the population of France in 2026 led to the construction of a plan to use regression on previous census data to construct a function and then the extrapolation of this function to 2026 to both predict the answer and estimate the uncertainty of this answer.

VI. Whole System Reasoning

As argued in [1], in addition to the core inference or prediction task, many other tasks, such as the automatic selection of KBs and relevant knowledge, choice of inference algorithms, and how to combine them, are all important to fully automate the QA process. Several of these tasks are scoped out as engineering tasks that experts perform when deploying these AI systems. It is argued that these scoped out tasks should be part of the AI models that are built for QA tasks, as they are key ingredients in the full automation of the QA process. Our work focuses on these tasks as well as the traditional QA problem with the added task of tackling questions that require multiple steps of reasoning to solve. We, therefore, agree with the conclusion that there is the need to refine the scope of problems that AI for QA should solve by incorporating those tasks which, in the real-world applications, look messy and are often tackled by humans experts. Further, “tackling these problems highlights the need for approaches that can appropriately leverage both symbolic and sub-symbolic AI methods and also bring to the fore the need to have AI systems that are compositional in order to adapt seamlessly to different problem types."[1].

Figure 1 shows the beginnings of our attempts to incorporate whole system reasoning into FRANK. Below, and in §VII, we discuss progress in this direction.

A. Automatic Statistical Model Selection

Prediction is a key functionality of FRANK, requiring the use of different prediction models to extrapolate from historical data. Selection of the correct model, e.g. regression, is a non-trivial task given the diverse kinds of data needed to answer questions. As part of our efforts to address this problem, work in [21] develops a lighter, more functional and configurable version of the Automatic Bayesian Covariance Discovery (ABCD) system [22], [23] called *GPy-ABCD* which is based on Gaussian Processes (GPs). Its implementation improves the model-space search as well as generates 1-paragraph descriptions of the underlying GP kernels that the model uses for regression. We use GPy-ABCD to automate the selection of statistical models for prediction in FRANK when generating meaningful shape descriptions of these models contributes to the query's answer. This work has enabled FRANK to answer new kinds of question, e.g. "How does temperature in Switzerland behave over time?". GPy-ABCD now forms part of the SMART system, which includes a wider selection of statistical methods, some of which are non-Gaussian.

B. Automatic Knowledge Source Selection

In many question-answering tasks, data from which to retrieve or infer answers are explicitly provided to the system. As described in [1], this task is often taken out of the scope of what the QA system does and is, instead, handled by human experts. In [24], the emphasis is on allowing FRANK to discover new facts from diverse knowledge sources on the Web from which to infer answers to questions. This is achieved by crawling the web to extract meta-data (based on Schema.org or JSON-LD [2] formats). This enables FRANK to extract useful entities and predicates grounded in knowledge sources such as Wikidata [25]. This knowledge is automatically curated into an ontology which maps entities and predicates to knowledge sources that contain additional relevant information. FRANK uses this ontology to select candidate knowledge sources to search when it needs to instantiate variables at the leaves of its inference graph.

C. Interfacing Through Natural Language

FRANK uses the alist data structure as its basic internal representation. Questions are formalized as alists which are then used as inputs for FRANK to process. Earlier implementations of FRANK used a template-based translation of natural language questions to alists and vice-versa. In [26], the translation process was improved by adapting state-of-the-art machine translation and language generation techniques to FRANK. Large pretrained language models (e.g. BART [27]) were fine-tuned using training data constructed from the natural language to SPARQL queries data in the LC-QUAD dataset [28]. This approach to translating natural language questions to alists (and vice-versa) allows for greater flexibility in translation, even when questions are not perfectly grammatical, while also enabling FRANK to tackle a wider variety of questions.

The above instances show how different statistical and AI methods are combined to improve the capabilities of FRANK in a *systems approach* to QA, instead of a single large system whose internal workings are hard to interpret and whose outputs are hard to explain.

VII. Conclusion

We have argued that the future of successful AI lies with the compositional combination of symbolic and sub-symbolic approaches within hybrid systems. These systems will use meta-level reasoning to plan an appropriate combination of AI methods and knowledge sources, and to recover from any failures in this initial plan by re-planning. This meta-level reasoning will automate engineering tasks that are currently done manually. Moreover, these hybrid systems will explain their reasoning to the user and estimate any uncertainty in their conclusions. By combining textual and graphical explanations we can address the challenge of explainable AI.

We further argue that query answering is a good vehicle for the exploration of this approach. As a proof of concept, we are developing the FRANK query answering system. Figure 1 illustrates how FRANK addresses the explainable AI challenge. It shows a combination of textual explanations of deductive reasoning with graphical representations of statistical reasoning. It explains its meta-level reasoning (i) when constructing a problem-solving composition of methods and (ii) when recovering from initial failure.

We are the first to admit that there is a lot more to do. In particular, the recent developments discussed in §VI will facilitate a whole-systems approach. Our next steps in natural language generation will enable FRANK to describe not just the answers to queries but also high-level explanations of how they were derived, including the process of constructing a plan of diverse deductive and statistical reasoning methods and which knowledge sources were chosen and why. We plan to make FRANK's user interface interactive to enable uses to ask what-if and how questions about the inference process. For instance, by changing the results returned at the inference graph's leafs, the user can explore what the answer might have been in different circumstances. By directing that a different method or knowledge source be used to solve a subgoal, e.g., a more accurate but more time-consuming statistical analysis or a rival knowledge source, the user can explore the robustness and uncertainty of an answer.

The range of query types is constantly expanding requiring new types of uncertainty estimate. For instance, while error

[2]https://json-ld.org

bars are well-suited to quantitative queries, probabilities are more appropriate for qualitative ones. We saw in §III that a qualitative query about which African country was predicted to have the largest population required quantitative sub-queries about previous population information. So, these two uncertainty measures will be intermingled during inference, which raises interesting theoretical challenges.

As well as expanding the range of FRANK's statistical methods, we also need new decomposition rules and second-order functions, such as calculus, to apply to functions formed by regression, e.g., to predict when a quantity will reach its maximum or minimum. Some of these new statistical methods are not Gaussian but can be based on Gaussian distributions or Gaussian assumptions about the data and its inherent noise, so will not necessarily return a mean and standard deviation, although one can usually distinguish value and uncertainty components in their results.

These challenges confirm our choice of both query answering as a vehicle for advancing artificial intelligence, and our hybrid and compositional methodology.

Acknowledgment

The authors would like to thank Ruqui Zhu, Thomas Fletcher, Nick Ferguson and Vaishak Belle for feedback on an earlier draft, and Huawei for supporting the research on which this paper was based under grants HO2017050001B8s and YBN2019125018.

References

[1] K. Nuamah, "Deep algorithmic question answering: Towards a compositionally hybrid ai for algorithmic reasoning," in *Workshop on Knowledge Representation for Hybrid and Compositional AI*, 2021.

[2] Y. Bengio, Y. LeCun, and G. Hinton, "Deep learning for AI," *Communications of the ACM*, vol. 64, no. 7, pp. 58–65, Jul. 2021.

[3] D. Kahneman, *Thinking, Fast and Slow*, ser. Penguin Psychology. Penguin Books, 2012.

[4] E. Davis and G. Marcus, "Commonsense reasoning and commonsense knowledge in artificial intelligence," *CACM*, vol. 58, no. 9, pp. 92–103, September 2015.

[5] K. Nuamah, A. Bundy, and C. Lucas, "Functional inferences over heterogeneous data," in *International Conference on Web Reasoning and Rule Systems*. Springer, 2016, pp. 159–166.

[6] K. Nuamah, "Functional inferences over heterogeneous data," 2018, unpublished Ph.D. Dissertation, University of Edinburgh.

[7] A. Bundy, K. Nuamah, and C. Lucas, "Automated reasoning in the age of the internet," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11110 LNAI. Springer Verlag, 2018, pp. 3–18.

[8] J. Launchbury, "A DARPA perspective on artificial intelligence," *YouTube video,(February 15, 2017). https://www.youtube.com/watch?v=-O01G3tSYpU*, 2017.

[9] O. Biran and C. Cotton, "Explanation and Justification in Machine Learning: A Survey," *IJCAI Workshop on Explainable AI (XAI)*, pp. 8–14, August 2017.

[10] S. T. Mueller, R. R. Hoffman, W. Clancey, A. Emrey, and G. Klein, "Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications, and Bibliography for Explainable AI," *arXiv*, no. 1902.01876, 2019. [Online]. Available: http://arxiv.org/abs/1902.01876

[11] R. L. Teach and E. H. Shortliffe, "An analysis of physician attitudes regarding computer-based clinical consultation systems," *Computers and Biomedical Research*, vol. 14, no. 6, pp. 542–558, 1981.

[12] L. R. Ye and P. E. Johnson, "The impact of explanation facilities on user acceptance of expert systems advice," *Mis Quarterly*, pp. 157–172, 1995.

[13] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "Moviexplain: a recommender system with explanations." *RecSys*, vol. 9, pp. 317–320, 2009.

[14] O. Biran and K. R. McKeown, "Human-centric justification of machine learning predictions." in *IJCAI*, 2017, pp. 1461–1467.

[15] A. C. Graesser, W. Baggett, and K. Williams, "Question-driven explanatory reasoning," *Applied Cognitive Psychology*, vol. 10, no. 7, pp. 17–31, 1996.

[16] T. Lombrozo, "The instrumental value of explanations," *Philosophy Compass*, vol. 6, no. 8, pp. 539–551, 2011.

[17] K. Nuamah and A. Bundy, "Explainable inference in the FRANK query answering system," in *ECAI 2020*. IOS Press, 2020, pp. 2441–2448.

[18] Z. C. Lipton, "The mythos of model interpretability," *Communications of the ACM*, vol. 61, no. 10, pp. 36–43, 2018.

[19] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 1135–1144.

[20] A. Bundy, *A Science of Reasoning*. MIT Press, 1991, pp. 178–198.

[21] T. Fletcher, A. Bundy, and K. Nuamah, "GPy-ABCD: A configurable automatic Bayesian covariance discovery implementation," in *8th ICML Workshop on Automated Machine Learning*, July 2021.

[22] J. R. Lloyd, D. Duvenaud, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, "Automatic Construction and Natural-Language Description of Nonparametric Regression Models," *AAAI*, p. 10, 2014.

[23] C. Steinruecken, E. Smith, D. Janz, J. Lloyd, and Z. Ghahramani, "The automatic statistician," in *Automated Machine Learning*. Springer, Cham, 2019, pp. 161–173.

[24] G. Brimble, "Automatic knowledge discovery," UG Project Dissertation, School of Informatics, University of Edinburgh, 2021.

[25] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Communications of the ACM*, vol. 57, no. 10, pp. 78–85, 2014.

[26] Y. Li, "Models to translate natural language questions to structured forms and back," MSc Project Dissertation, School of Informatics, University of Edinburgh, 2021.

[27] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[28] P. Trivedi, G. Maheshwari, M. Dubey, and J. Lehmann, "Lc-quad: A corpus for complex question answering over knowledge graphs," in *International Semantic Web Conference*. Springer, 2017, pp. 210–218.