Question-answering system for combustion kinetics

Laura Pascazio\textsuperscript{a,}\textsuperscript{,}*, Dan Tran\textsuperscript{a}, Simon Rihm\textsuperscript{b}, Jiaru Bai\textsuperscript{b}, Sebastian Mosbach\textsuperscript{a,b,c}, Jethro Akroyd\textsuperscript{a,b,c}, Markus Kraft\textsuperscript{a,b,c,d,e}

\textsuperscript{a}Cambridge Centre for Advanced Research and Education in Singapore (CARES), CREATE Tower, 1 Create Way, 138602, Singapore
\textsuperscript{b}Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK
\textsuperscript{c}CMCL Innovations, Sheraton House, Castle Park, Castle Street, Cambridge CB3 0AX, UK
\textsuperscript{d}School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, 637459, Singapore
\textsuperscript{e}The Alan Turing Institute, 2QR, John Dodson House, 96 Euston Rd, London NW1 2DB, UK

Abstract

In this paper, we introduce for the first time a natural language question-answering (QA) system specifically designed for the field of combustion kinetics. This system marks a significant step towards achieving the PrIMe vision as outlined by Frenklach in 2007, offering a user-friendly interface that allows researchers and practitioners to easily access and query information about chemical mechanisms. This QA system is a key component of “The World Avatar” (TWA), a dynamic framework built upon semantic web technologies. TWA is characterized by its layered structure, which includes a knowledge graph (KG), software agents, and real-world data integration. These layers collectively create a comprehensive unified system for managing and analyzing complex chemical data from various domains. We detail the enhancements made to TWA's ontologies (OntoSpecies, OntoKin, and OntoCompChem) to meet specific challenges in chemical kinetics and improve their representation accuracy. By focusing on data provenance and interoperability, our approach ensures transparent and reliable data management that adheres to the FAIR principles, which is vital for precise information retrieval and analysis. The role of software agents in populating these ontologies is highlighted, showcasing how they transform raw data into meaningful structured knowledge and generate new insights within the TWA ecosystem. Additionally, the semantic web technologies’ interoperability feature facilitates data integration and exchange across different platforms and tools, making the data machine-actionable. We instantiated in the KG data on four H2/O2 and five CH4/O2 reaction mechanisms taken from the literature and we then demonstrate the QA system’s capabilities in answering questions related to these reaction mechanisms as a proof of concept. Lastly, we discuss the future directions of the TWA framework, which include not only future extensions of the QA system but also the integration of external tool to automate tasks such as generation of kinetic mechanism, further expanding TWA's functionality and application in the field of chemical kinetics.

Keywords: Automated kinetic modelling; Reaction mechanism; Cheminformatics; Artificial intelligence; Knowledge graph
Information for Colloquium Chairs and Cochairs, Editors, and Reviewers

1) Novelty and Significance Statement

The novelty of this research lies in the development of a natural language question-answering system specifically developed for combustion kinetics that marks a significant step towards achieving the PrIMe vision outlined by Frenklach in 2007. The system builds upon a dynamic knowledge graph based on semantic web technologies as part of "The World Avatar”, offering a unified approach to handling complex chemical data. The system emphasizes data provenance and interoperability, ensuring transparent and reliable information retrieval and analysis. This research is significant as it transforms the accessibility and analysis of combustion kinetics data. Our approach, that combines user-friendly access with rigorous data management, represents a significant advancement in making combustion kinetics data not only more accessible but also machine actionable. Answering questions on H2/O2 and CH4/O2 reaction mechanism as a proof of concept, this research demonstrates the QA system’s potential applications.

2) Author Contributions

- L.P.: designed research, performed research, analyzed data, wrote the paper
- D.T.: performed research, wrote the paper
- S.R.: designed research, analyzed data, wrote the paper
- J.B.: designed research, analyzed data, wrote the paper
- S.M.: designed research
- J.A.: designed research
- M.K.: designed research

3) Authors’ Preference and Justification for Mode of Presentation at the Symposium

The authors prefer OPP presentation at the Symposium, for the following reasons:

- Our paper introduces a question-answering (QA) system with wide-ranging implications for the chemical kinetics community, poised to generate significant interest in a room-audience setting.
- An oral presentation will allow for comprehensive discussions and feedback, vital for the refinement and future development of the QA system.
- The oral format is more effective for reaching a diverse audience, crucial for disseminating the innovative aspects of our system across various subfields.
1. Introduction

Combustion is an exothermic process that fundamentally powers our modern world, ranging from internal combustion engines to industrial furnaces [1]. As we progress towards the United Nations’ Sustainable Development Goals (SDGs), the precise prediction and control of the behavior of reacting mixtures during combustion reactions become increasingly important. The key to success is accurate kinetic modeling as it encapsulates our understanding of the underlying chemical reactions and their rates.

Modeling combustion kinetics accurately is not an easy task due to the inherent complexity and multifaceted nature of combustion processes. Even in the combustion of relatively simple hydrocarbons [2], a vast number of chemical species and reactions emerge, resulting in highly non-linear behavior covering a broad range of time and length scales—spanning nanoseconds to milliseconds and from molecular to macroscopic levels. The sensitivity of reaction rates to temperature, described by Arrhenius equations, means that minimal fluctuations in temperature can lead to significantly different outcomes. The real-world interaction between turbulence and chemistry adds an extra layer of complexity, making the task even more challenging.

Compounding these challenges is the variability in the reaction mechanisms developed by different research groups. Depending on focus, methodologies, and available experimental data, diverse mechanisms can emerge for the same fuel [3]. While all aim to capture the combustion phenomenon, inconsistencies may arise when comparing these mechanisms, especially for reduced mechanisms extracted for practical applications [4].

This fragmented landscape necessitates a comprehensive system capable of efficiently managing extensive experimental data and kinetic mechanisms. This is crucial for detecting inconsistencies across mechanisms and predicting reaction rates for missing reactions. In 2007, Frenklach [5] introduced the Process Informatics Model (PrIMe) initiative, a data-centric approach to developing predictive models for complex chemical systems. It envisions an intelligent question-answering (QA) system to assist researchers during this process. One key feature of PrIMe lies in its method of organizing scientific data by merging diverse sources based on both data semantics and provenance. The initiative strives to maintain a complete and up-to-date set of data in combustion chemistry for rapid mechanism benchmarking and analysis. This data curation model holds great potential to foster democratic decision-making and consensus-building within the combustion community.

A significant hurdle in realizing this vision is the isolated nature of existing data, mechanisms, and software tools [6]. This not only impedes interoperability between the outcomes produced from different groups but also results in their unavailability once projects conclude or servers are shut down (e.g.,...
tera [17], face challenges related to interoperability and data integration. These tools often struggle with compatibility of data formats, making it difficult to transfer and use data seamlessly across platforms. Additionally, identifying and eliminating duplicate information in various formats remains a significant issue [18] which highlights the need for improved coherence and reliability in kinetic modeling. Particularly, provenance could ensure better data consistency, eliminate duplicates (e.g., permutations of the same reaction), and help adhere to the FAIR principles – Findable, Accessible, Interoperable, and Reusable [19]. This cannot be achieved without undergoing a process of “transformation” of these disparate data sets into functional models, a process complicated by escalating data volumes.

This situation underscores the importance of evolving databases from mere repositories into dynamic tools for knowledge creation, addressing the meta-data challenges concerning underlying assumptions, parameter origins, uncertainties, and reasons for inclusion/exclusion of reactions [10]. Additionally, as the current landscape in chemical engineering presents itself, many practitioners lack expertise in computing rate coefficients from first principles or using modern kinetic modeling software and machine learning techniques. This calls for a robust and easy-to-use computational framework that accurately determines thermochemical and transport parameters and rate constants for a wide array of chemical species and reactions.

2.2. Available solutions

In response to these challenges, various automated tools like RMG (Reaction Mechanism Generator) [20], CLARA (Class-based Automatic Reaction Alternator) [21], Genesys [22], and AUTO-MECH [23] were developed to generate kinetic mechanisms, employing methods ranging from predefined reaction pathways to unbiased approaches. Tools like RMG, CLARA, and Genesys rely on predefined rules or categories to identify and generate reaction pathways. While effective, this method sometimes fails to select the most relevant species or reactions due to its reliance on predefined templates, necessitating the expansion of the range of reaction templates to improve accuracy [20]. In contrast, AutoMech automates the identification and analysis of reaction pathways without relying on templates, focusing instead on the underlying quantum mechanics of the reactions for potentially more accurate results. However, AutoMech’s reliance on first-principles calculations can be computationally intensive and challenging to integrate seamlessly with other kinetic modeling tools and databases. While these tools are advancing towards the goal of fully predictive kinetic models, gaps remain, particularly in evaluating thermochemistry and rate constants, highlighting the need for increased automation in these areas. A combination of theoretical and experimental methodologies can be used to enhance the predictive capabilities of these mechanisms [24]. This approach, coupled with automated procedures for the investigation of unknown reaction pathways, is critical for discovering new reactive pathways and achieving a comprehensive understanding of combustion kinetics.

Other tools have been developed for accessing and analyzing kinetic mechanisms to overcome the issue of the disorganized state of the collection, evaluation, and selection of chemical reaction models. Frameworks like ReSpecTh [25] and SciExpeM [26] both aim to create structured and reliable repositories freely accessible to the scientific community. ReSpecTh focuses on providing a comprehensive database of reaction mechanisms and experimental data with detailed metadata, adhering to FAIR principles. SciExpeM emphasizes automation in managing and analyzing experimental or modeling data by integrating tools for data validation [27] and analysis [13]. The tool provided by Killingsworth et al. [28] addresses issues in the specification of thermodynamic properties, reactions, and rate coefficients identifying inconsistencies and errors. While these advancements are promising, these tools are still limited by the lack of standardized reporting of experiments and data storage and further efforts are needed to fully integrate these tools into a cohesive, user-friendly platform.

Finally, in order to increase ease of access, QA systems can significantly lower entry barriers. The current wave of large language models illustrates this impressively in many areas. However, general Natural Language Processing (NLP) tools that are available to the public are monoliths in nature, lacking explainability and domain knowledge (e.g., ChatGPT). Going forward, computational chemistry needs a “peer-to-data-driven operation with respect to molecular science to act autonomously on arguments and questions raised in natural language by the operator” [29].

2.3. Semantic Web for representing chemical kinetics

The Semantic Web, developed as an extension of the World Wide Web, allows data to be shared and reused across application, enterprise, and community boundaries using technologies that enable data to be connected and utilized in a meaningful way [30]. Leveraging this framework, the Semantic Web offers an innovative, graph-oriented approach for effectively representing the intricate domain of chemical kinetics. It structures chemical information into a graph, aptly capturing the complex relationships and properties inherent in chemical processes. This approach is particularly well-suited for chemical kinetics, where many tools are graph-based [18] [31], as it naturally aligns with the graph-like nature of chemical interactions, depicted in Fig. 1. In this framework, reaction mechanisms are portrayed as networks of interconnected reactions, with each reaction defined by its reactants and products, represented as individual chemical species within the graph (Fig. 1). A major advantage of this approach is its capacity to manage “multiple properties” of
species and reactions that may originate from various data sources. The unique identification of each species and reaction in the graph permits precise linkage of species and reaction properties to their origin. This is crucial for resolving ambiguities in naming conventions and ensuring data consistency. This structure is also particularly beneficial when designing new chemical kinetics models because it permits the selection of the most suitable thermodynamic models for species or kinetic models for reactions, based on specific reaction conditions. It allows for accurate comparisons and selection among various models, ensuring a more accurate and context-sensitive application of these models, which is vital for understanding and predicting chemical reactions accurately.

Moreover, the Semantic Web's framework facilitates the alignment of common elementary reactions across different mechanisms, offering a unified view of kinetic data. This system replaces traditional data formats, like Chemkin files, with a dynamic graph structure, where each data point becomes a node in the graph. This enhancement not only simplifies navigation and updates within the vast array of chemical kinetics information but also supports advanced querying capabilities.

These features are essential to identify recurring trends, cross-mechanism relationships, and potential knowledge gaps, thus pushing the analysis of chemical kinetics towards a more integrated and insightful approach.

The following subsections detail the specific components of TWA utilized in the context of chemical kinetics, discussing modifications that have been implemented for this work.

3. The World Avatar Infrastructure

The World Avatar (TWA) is a framework that adheres to the FAIR data principles [19]. It is designed to store, process, and analyze chemical data and models. Differently from other frameworks in the field, TWA distinguishes itself with a structured, layered architecture, which includes ontologies for data organization, software agents for data manipulation, and connections to empirical real-world data, as depicted in Fig. 2. This approach embeds dynamicity by default, allowing the system to automatically update and evolve as new data becomes available.

At its core, TWA features a middle layer composed of a knowledge graph (KG) based on semantic web technologies that provides a dynamic platform for organizing, querying, and traversing the vast, complex chemical data. The data interaction and connection represented in the graph along with their provenance facilitate data discovery and retrieval of interrelated information and ensure accuracy and traceability of diverse data types, enhancing the understanding of chemical processes. The schema of the KG is defined by ontologies that are split into a Terminology component (TBox) for structural definitions and an Assertion component (ABox) for real-world data instances.

The top layer is characterized by the deployment of advanced software agents. These agents are capable of operating both locally and as web services to provide tailored solutions for varied applications, enhancing TWA’s functionality for specific user needs. The bottom layer establishes a connection with the real world, integrating empirical data for validation, such as sensor measurements in a laboratory setting, and human-machine interfaces like QA systems for chemistry. This ensures the applicability and reliability of TWA-derived insights and models.

Data and software that form the TWA ecosystem are publicly available [8, 32, 33], allowing for community-driven improvements and adaptations.

The following subsections detail the specific components of TWA utilized in the context of chemical kinetics, discussing modifications that have been implemented for this work.

3.1. Knowledge Graphs

In the chemistry domain of TWA KG, several ontologies have been developed to store specific types of information, each addressing a distinct subdomain. This paper concentrates on three subdomains within TWA, each playing a distinct role in enriching the chemical knowledge base as illustrated in Fig. 3.

OntoSpecies, the core ontology in TWA’s chemistry domain (blue boxes in Fig. 3), focuses on species representation and their properties [34]. OntoSpecies allows for precise management of species
identities, including the differentiation of isomers and pseudospecies (see supplemental material), which are essential for accurate chemical reaction modeling and analysis. Moreover, OntoSpecies supports detailed classification, which is particularly important for identifying species that share reactive moieties and functional groups and grouping reactants. The majority of data on species properties in OntoSpecies is experimental, sourced primarily from PubChem [35]. Despite the uniqueness of certain properties like melting points, OntoSpecies acknowledges the variability that arises when data is collected from diverse sources. As in the PrIMe vision, the data is not organized by source but by merging different sources based on scientific meaning, while meticulously tracking the provenance of each data point [34]. In this work, additional properties (filled blue boxes in Fig. 3) that are crucial for chemical kinetics (HeatCapacity, StandardEnthalpy, StandardEntropy) have been added to the OntoSpecies TBox. These properties are defined as sub-classes of ThermoProperty and follow the same schema as defined in our previous work [34].

OntoKin is an ontology that represents information about reaction mechanisms, elementary reactions, and kinetic models as presented in the literature, with each entry enriched with provenance details (red boxes in Fig. 3). Compared to its previous implementation [36], it uniquely identifies reactions across different mechanisms based on reactants, products, and possible third bodies. Although a reaction is uniquely identified, it may be associated with different kinetic models used in different reaction mechanisms. This approach extends to species within OntoKin, where different thermochemical and transport models may be applied across various reaction mechanisms. To take this into account, few modifications have been made to OntoKin TBox (filled red boxes in Fig. 3). The ReactionMechanism is directly connected to the ChemicalReaction by the predicate hasReaction. The ChemicalReaction has a KineticModel that is connected to the ReactionMechanism that uses it by the predicate definedIn. These connections help to avoid duplicate information and follow the idea presented in Fig. 1 so that the same KineticModel can be defined in more than one ReactionMechanism. Equivalently, a Species that takes part in a ChemicalReaction can have more than one ThermoModel and TransportModel that...
are linked to the ReactionMechanism where they are used in.

OntoCompChem is dedicated to information related to quantum mechanical (QM) calculations [36]. This is particularly crucial for deriving some properties of species, such as heat capacity, especially when experimental data are not available. OntoCompChem bridges the gap between theoretical predictions and empirical data, thereby enhancing the framework’s overall accuracy and reliability. The idea of storing calculations and connecting them to the interested species saves time for researchers when the calculation is already available in the ontology. Compared to its previous implementation, species are uniquely identified and connected with their OntoSpecies instance as well as their atoms to avoid any ambiguity.

Lastly, in this work, the addition of unit and provenance subdomains, built upon standard vocabularies and expanded with new units as required, represents an essential step towards standardizing data representation across TWA (orange and green boxes respectively in Fig. 3). Thus, TWA ensures that all data, regardless of its subdomain, adheres to a consistent unit system. This is not only crucial for accurate data representation but also for the interoperability of data between different systems and applications. The standardization of provenance also enhances data interoperability. By embedding this concept directly into the KG, TWA enables the efficient tracking and verification of information sources across its subdomains. This represents a great step forward compared to the widespread use of comment lines in Cantera and Chemkin input files which represented the best practice at the time (see supplemental material).

Together, these ontologies create a synergistic and multifaceted structure within TWA that not only facilitates a comprehensive representation of chemical information but also underscores the depth and complexity inherent in the field of chemical kinetics and computational chemistry. More details on the ontologies’ TBoxes can be found at https://theworldavatar.io/demos/marie.

3.2. Agents

Software agents within TWA infrastructure fall into two primary categories: those that facilitate the collection, organization, and curation of data, and those that enable the processing and analysis of data. For the OntoSpecies domain, a specialized agent is tasked with data acquisition from various scientific databases. It retrieves identifiers, experimental properties, and spectral information from PubMed and utilizes CHEBI for chemical classification and usage data [34]. In this work, the population agent has been expanded to extract thermochemical experimental data from the NIST Chemistry WebBook [37]. This comprehensive approach ensures that the OntoSpecies domain is populated with extensive chemical data.

The OntoKin domain employs a population agent that processes data from Chemkin files. Due to the ambiguity and inconsistency of Chemkin labels for species, an InChI string is provided for each chemical species listed in the Chemkin file to minimize errors and accurately link each species to its IRI in OntoSpecies, ensuring consistency and data integrity across domains. The agent then links every reaction in the mechanism to its unique IRI. Finally, with the help of RMG API, it parses the data reported in the Chemkin file and instantiate it in the graph using a SPARQL update. Currently, the data collated by this agent includes four H2/O2 and five CH4/O2 mechanisms taken from literature [36-43].

The OntoCompChem domain employs a population agent that processes data from Gaussian log files and instantiates the relevant information in the KG.

The second type of agents are used for data augmentation, enhancing the system’s capabilities for analysis and prediction. To cite some, if a vibrational analysis of the species exists in the KG, an agent can calculate thermal properties (enthalpy, heat capacity and entropy) from the calculation outputs [36]. We also demonstrated that agents can do sensitivity analysis and calibration, as demonstrated for combustion experiments [3].

In future expansions, we plan to integrate external tools such as RMG for predicting reactive chemistry or others like Cantera, Chemkin, and kinetics, especially as they evolve and new tools emerge [44]. It is important to stress the flexibility of our approach, that structuring data with clear definitions of concepts, units, and provenance, enables easy conversion of knowledge into formats compatible with various tools, that can be then easily integrated in our framework.

3.3. Natural Language Processing

The integration of advanced NLP techniques for a question-answering (QA) system developed for chemistry within TWA, Marie, marks a significant stride in the realm of chemical kinetics modeling.

The NLP framework in TWA is underpinned by an end-to-end translation approach, leveraging the power of pre-trained text-to-text language models. The system is designed to seamlessly translate questions posed in natural language into SPARQL queries, enabling users to interact with the KG in an intuitive manner. This approach is a departure from traditional methods which relied heavily on hand-crafted templates and were limited in their scalability and flexibility [45]. In its previous implementation, Marie was limited to the OntoSpecies domain [45]. In this paper, we expanded its functionality to work with additional ontologies like OntoKin and OntoCompChem, achieved by performing multi-task prompted training on both the translation and domain classification tasks. This expansion facilitates a broader range of applications in chemistry-related research and industrial scenarios, making the system an even more versatile tool for data retrieval and analysis.
4. Results and Discussion

In the realm of practical application, TWA’s advanced querying capabilities, coupled with its comprehensive representation of chemical information, empowers researchers to navigate and resolve the complexities of chemical kinetics modeling. The integration of the QA system, Marie, into TWA infrastructure represents a significant stride towards achieving the vision Frenklach set out in 2007 [5]. As an advanced tool, Marie helps to access TWA capabilities without the need to know the SPARQL query language. In the context of chemical kinetics, Marie is able to answer questions about specific reactions or species as well as full mechanisms and affiliated kinetic or thermodynamic models based on different criteria. In the context of QM calculation, users can request optimised geometries, energies, and other values computed for specific molecules at varying levels of theory and using different basis sets. Marie also handles species-specific queries, like identifying species with particular characteristics or detailing properties of species that belong to a specific chemical class. Based on its underlying language model, Marie also demonstrates its capabilities in answering more complex questions that require crafting advanced queries across domains and data sets. This can be achieved either by asking for data comparison or chaining simple data retrieval questions, indicating a step closer to having a meaningful conversation with an “AI Scientist”. In this section, we demonstrate the capabilities of Marie in answering questions related to the reaction mechanisms instantiated in the KG as a proof of concept. An illustrative example is presented in Fig. 4, where we compare three different approaches to answering a common research question in chemical kinetics: “Please compare the rate constant parameters of the reaction H2 + OH = H2O + H across all mechanisms it appears in.” Such queries are crucial for researchers because for common combustion processes, many unique mechanisms and kinetic models are published in the literature and sometimes differ only in a few ways that are hard to compare.

Manual search in publication: The first approach involves a manual search through all publications we want to compare for the specific reaction (H2 + OH = H2O + H). This is a time-consuming process as notations, order of reactants, etc. can differ as shown in the top left of Fig. 4. Moreover, context on the specific kinetic model as well as parameter units need to be analyzed and tracked independently. Lastly, sources need to be tracked closely as well to assess if the same parameter set given in two publications can be interpreted as validations or just using the same underlying model.

Utilizing ChatGPT for Chemkin file parsing: The second approach (shown in the top right of Fig. 4) uses ChatGPT4 (GPT-4 online chatbot) [46] to parse the uploaded Chemkin files related to the publications. Even after a time-consuming additional step of uploading four files and providing context, ChatGPT is unable to identify the reaction in question in any of the mechanisms. Adjusting the prompt to increase precision and rewording the chemical reaction so that it could be found in the file via classic search function does not help either.

Querying Marie: The third approach, asking Marie to compare kinetic models, results in a precise and comprehensive response. Marie efficiently lists all unique kinetic models, including model types, parameter values, and parameter units. It also correctly identifies only 3 different kinetic models among the four mechanism instantiated in the KG as two of them use the same one [2, 39]. While the example illustrated in Fig. 4 is limited to Arrhenius models, comparisons across different model types are possible. The comparison between Marie and ChatGPT demonstrates the importance of knowledge representation (ontologies vs. text) for the efficacy of language models. Other exemplary use cases can be found in the supplemental material. The ChatGPT prompting strategy designed to provide detailed instructions to the model is also detailed in the supplemental material. Users can interact with Marie by following this link (https://theworldavatar.io/demos/marie). However, the system is still under development, and the accuracy of the results will increase with further refinement of the underlying ontologies.

5. Conclusions and future directions

In this paper, we have presented a proof of concept that showcases the innovative features of TWA in the realm of chemical kinetics modeling and adheres to the PrMe vision. Central to TWA are its NLP capabilities. By allowing users to interact with the system in natural language, TWA significantly simplifies the process of querying and analyzing chemical data. This feature not only enhances the accessibility and usability of the system but also opens up new avenues for research and applications in chemical kinetics. The vision for TWA and its integrated systems extends beyond the current capabilities and aligns with perspectives like Bob Kee’s on Cantera [47] and Bill Green’s on RMG [48], yet seeks to surpass current tool limitations. Key areas for future development include identification of inconsistency, automatic mechanism generation, estimation of rate coefficients, error propagation agents for uncertainty quantification, expanding to other chemistry areas like surface chemistry, integrating with self-driving laboratories and advancing NLP and the reasoning capabilities of Marie. It is crucial to emphasize that the intention is not to reinvent existing tools but to leverage the strengths of established tools, enhancing their utility and effectiveness through seamless integration within the TWA framework, thereby contributing to establishing a global combustion research network as already demonstrated by TWA in other domains [49]. Through these enhancements, TWA and its components aim not only to address the current challenges
Please compare the kinetic models of the reaction \( \text{H}_2 + \text{OH} = \text{H}_2\text{O} + \text{H} \) across all the mechanisms it appears in.

Fig. 4: Comparison of data comparison process between manual search of databases, intelligent use of Chemkin-informed ChatGPT, and Marie.

in the field but also to pave the way for innovative approaches and applications in combustion kinetics and beyond.

Acknowledgments

This research was supported by the National Research Foundation, Prime Minister’s Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. Part of this work was supported by Towards Turing 2.0 under the EPSRC Grant EP/W037211/1 and The Alan Turing Institute.

Supplemental material

The supplemental material includes: additional QA use cases, ChatGPT prompting strategy and more information on TWA ontologies.

References