KARYA ILHAM PUBLISHING
CONNECTING MINDS, BRIDGING IDEAS, INSPIRING READERS

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://karyailham.com.my/index.php/arca/index
ISSN: 2462-1927

Development of an Offline Interactive Chemistry Tutor Using Generative AI and Large Language Model

Shafidah Shafian[1,*]

[1] Solar Energy Research Institute, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 4 July 2025
Received in revised form 10 September 2025
Accepted 30 September 2025
Available online 6 October 2025

Keywords:
Offline tutoring; generative AI; large language model (LLM); chemistry education; educational technology

ABSTRACT

This research presents the development of ChemGenAI, an offline, interactive chemistry tutor powered by a large language model (LLM) and cheminformatics tools, designed to enhance digital chemistry education in environments with limited or no internet connectivity. The primary aim is to create a locally deployable platform that supports real-time natural language interaction, molecular structure visualization, and periodic table exploration without relying on cloud-based services. ChemGenAI integrates the pre-trained Mistral-7B-Instruct model executed through Ollama for offline language processing and RDKit for molecular parsing and rendering. The system features a user-friendly graphical interface with three main modules: Chemistry Tutor, Molecule Draw, and Periodic Table. Results show that ChemGenAI accurately responds to chemistry questions, generates and visualizes 2D and 3D molecular structures from SMILES input, and presents detailed elemental data interactively. The interface design follows established usability heuristics to ensure intuitive navigation and engagement. The findings demonstrate that ChemGenAI offers an effective and accessible solution for chemistry instruction, particularly in classrooms where cloud-based systems are not feasible due to limited infrastructure or institutional that restrict external data access. The study concludes that offline AI tools like ChemGenAI can play a significant role in supporting inclusive, adaptable, and technology-enhanced science education.

1. Introduction

Artificial intelligence (AI) has become a transformative tool across various domains, including education [1]. In science education, particularly chemistry, AI is proving valuable in enhancing understanding and engagement [2-4]. Chemistry is often considered a challenging subject due to its abstract concepts, symbolic language, and quantitative problem-solving [5]. Many learners struggle with visualizing molecules, interpreting reactions, and grasping periodic trends. AI technologies are increasingly used to address these difficulties by providing interactive platforms that offer visual aids, simulations, and personalized guidance [6].

Among the most significant advancements in AI are Large Language Models (LLMs), a deep learning system trained on vast and diverse text datasets. Models such as GPT-4 (OpenAI), LLaMA

* *Corresponding author.*
E-mail address: norshafidah@ukm.edu.my

https://doi.org/10.37934/arca.40.1.6375

(Meta), and Gemini (Google) demonstrate the ability to understand and generate human-like text, enabling tasks such as question answering, concept explanation, and language translation [7, 8]. In education, LLMs function as intelligent tutors, capable of adapting to student input, generating problem sets, and explaining complex topics in accessible language. For chemistry, they are increasingly used to help explain reactions, structures, and calculations related to stoichiometry, periodicity, and mechanisms.

However, most LLMs are cloud-based, requiring internet access and remote server support [9]. This dependency poses several challenges, especially for learners in areas with limited connectivity, such as rural schools or underserved regions. In addition, concerns about data privacy and institutional restrictions on cloud-based platforms limit the adoption of such tools, even in well-connected environments. To overcome these barriers, there is a growing need for offline AI systems. The tools that operate locally without relying on the internet. Offline solutions offer key advantages such as they function without connectivity, preserve user privacy, eliminate ongoing cloud service costs, and allow greater customization. In this context, the development of accessible, offline AI tutors becomes crucial for ensuring equitable and scalable support in science education.

This paper introduces ChemGenAI, an offline, interactive chemistry tutor powered by a locally deployed generative AI model. The system utilizes the Mistral-7B-Instruct model, a compact, instruction-tuned LLM executed via Ollama, a lightweight runtime environment that supports quantized models on standard hardware. This setup allows ChemGenAI to function entirely offline, making it suitable for deployment in schools with limited infrastructure or high privacy requirements. ChemGenAI also integrates RDKit, an open-source cheminformatics toolkit, enabling molecular structure visualization and property prediction from SMILES inputs. Users can explore molecules interactively, enhancing spatial reasoning and structural understanding. An additional feature is the interactive periodic table, where users can access detailed information about each element, including atomic properties, classification, and group placement. The graphical user interface (GUI) providing a clean, user-friendly interface that supports text queries, molecular drawing, and periodic table exploration. By combining offline AI capabilities, robust chemistry tools, and an intuitive GUI, ChemGenAI offers a practical, secure, and engaging solution for modern chemistry education.

2. Methodology

The ChemGenAI system was developed using a modular, offline-capable architecture that integrates a LLM, a local execution environment, and a cheminformatics engine. The core components include the pre-trained Mistral-7B-Instruct language model for natural language interaction, the Ollama runtime for local model execution, and RDKit for chemical structure processing and visualization. All components were implemented to function entirely on local hardware without the need for internet connectivity.

2.1 Mistral-7B-Instruct

The Mistral-7B-Instruct model, an open-source, instruction-tuned LLM, was used in its pre-trained form to handle natural language queries related to chemistry. This model was selected for its compact size, strong performance on reasoning and explanation tasks, and suitability for running on consumer-grade CPUs. The model provides domain-agnostic instruction-following capabilities to operate effectively in general chemistry contexts. It supports a broad range of question types, including definitions, reaction explanations, and problem-solving tasks, responding with contextually appropriate answers.

2.2 Ollama

To facilitate offline operation, the Ollama environment was used to execute the Mistral model locally. Ollama is an efficient runtime designed to run quantized LLMs using the GGUF (GPT-generated unified format) standard. The Mistral model was loaded in a 4-bit quantized version, allowing for reduced memory usage and CPU-only inference. This setup enables low-latency, offline interaction with the model, ensuring that all user queries and responses are processed entirely on the host machine. The Ollama runtime handled prompt formatting, model loading, and response generation, and it interfaced directly with the system's front-end without requiring external dependencies.

2.3 RDKit

For chemical structure handling, the system incorporated RDKit, a widely adopted open-source cheminformatics library. RDKit was used to interpret and visualize molecules based on user-submitted SMILES (Simplified Molecular Input Line Entry System) strings. Upon receiving a valid SMILES input, RDKit parses the molecular structure and generates a two-dimensional graphical representation, which is then rendered within the application's interface. This visual output helps users better understand molecular geometry, bonding, and structural patterns.

2.4 GUI

The interface was organized into three primary modules which are the Chemistry Tutor, Molecule Draw, and Periodic Table. User inputs in the Chemistry Tutor module are routed to the LLM via Ollama, while molecule-related inputs are handled by RDKit. The Periodic Table module uses a locally stored dataset to provide interactive access to elemental properties.

3. Results and Discussion

The ChemGenAI system is designed using a user-centered approach that adheres to Nielsen's 10 usability heuristics to ensure an intuitive, efficient, and educational experience [10, 11]. The GUI, shown in Figure 1, is structured to promote clarity and accessibility. On the left panel, the ChemGenAI logo is prominently displayed, reinforcing system identity and maintaining consistency in design, in alignment with heuristic number four which emphasizes consistency and standards. Below the logo, users can access three main functional modules, namely Chemistry Tutor, Molecule Draw, and Periodic Table. These modules are organized as expandable menu items to support minimalist layout and enhance the visibility of system status, following heuristic number six which promotes recognition over recall and heuristic number eight which emphasizes aesthetic and minimalist design.

A Clear Chat button is placed below the module menu, allowing users to reset the conversation with a single click. This supports user control and freedom, consistent with heuristic number three. At the bottom of the interface, a persistent text input field is provided for users to enter natural language questions or molecular formulas. This feature helps prevent errors by keeping the input mechanism visible at all times and encouraging continuous interaction, as recommended in heuristic number five.

The interface provides immediate and consistent feedback to users, ensuring that the system status is always visible, in accordance with heuristic number one. The use of familiar terminology and chemistry-related language helps align the system with user expectations and mental models, supporting heuristic number two which focuses on matching the system with the real world. Additionally, the design is flexible enough to accommodate both beginners and advanced users, following heuristic number seven which encourages flexibility and efficiency of use.

In Figure 1(a), the system demonstrates its default startup interaction. Upon loading, the Chemistry Tutor module is automatically activated, and the language model responds to the prompt "Who are you?" with a self-introduction. The model identifies itself as ChemGenAI, a professional chemistry tutor developed to support students from high school to university levels. The response is clear, friendly, and aligned with educational expectations, reflecting the model's ability to understand context and generate an appropriate persona. This structured greeting helps set the tone for learning and encourages users to proceed with questions. It also supports the heuristic that users should be able to recognize, diagnose, and recover from any uncertainty during initial interactions.

In Figure 1(b), the user enters a factual query requesting a list of 20 molecular formulas. The system generates an accurate and relevant list of commonly encountered compounds such as water, methane, ethanol, and potassium hydroxide. Each entry is presented in a readable list format with the corresponding chemical formula and name. This output demonstrates the model's competence in factual knowledge retrieval and its capacity to present information in a structured, user-friendly manner. This contributes to overall user satisfaction and aligns with the heuristic that systems should provide accessible and helpful information when needed.

In Figure 1(c), the system is tested for its ability to provide detailed explanations. When prompted with "What is ammonium chloride," the model produces a concise yet informative paragraph that describes the compound's ionic structure, solubility, common applications, and thermal decomposition behavior. A follow-up question, "Can you elaborate more about potassium hydroxide," prompts the model to generate a more comprehensive explanation. The response includes information on the compound's classification as a strong base, its industrial uses, physical properties such as melting and boiling points, and safety considerations during handling. These extended responses illustrate the language model's ability to deliver coherent, accurate, and pedagogically relevant information. The system supports step-by-step learning and allows users to ask iterative questions, which reflects the usability principle of offering flexibility for both novice and experienced learners.

The Molecule Draw module in ChemGenAI provides users with an interactive tool for molecular visualization and basic property inspection, based on user input in the form of molecular formulas or SMILES strings [12, 13]. As shown in Figure 2, the system successfully processes various organic and inorganic compounds and displays their two-dimensional (2D) structures alongside their three-dimensional (3D) conformations. Each rendering is accompanied by essential molecular information including IUPAC name, SMILES notation, molecular formula, and molecular weight.

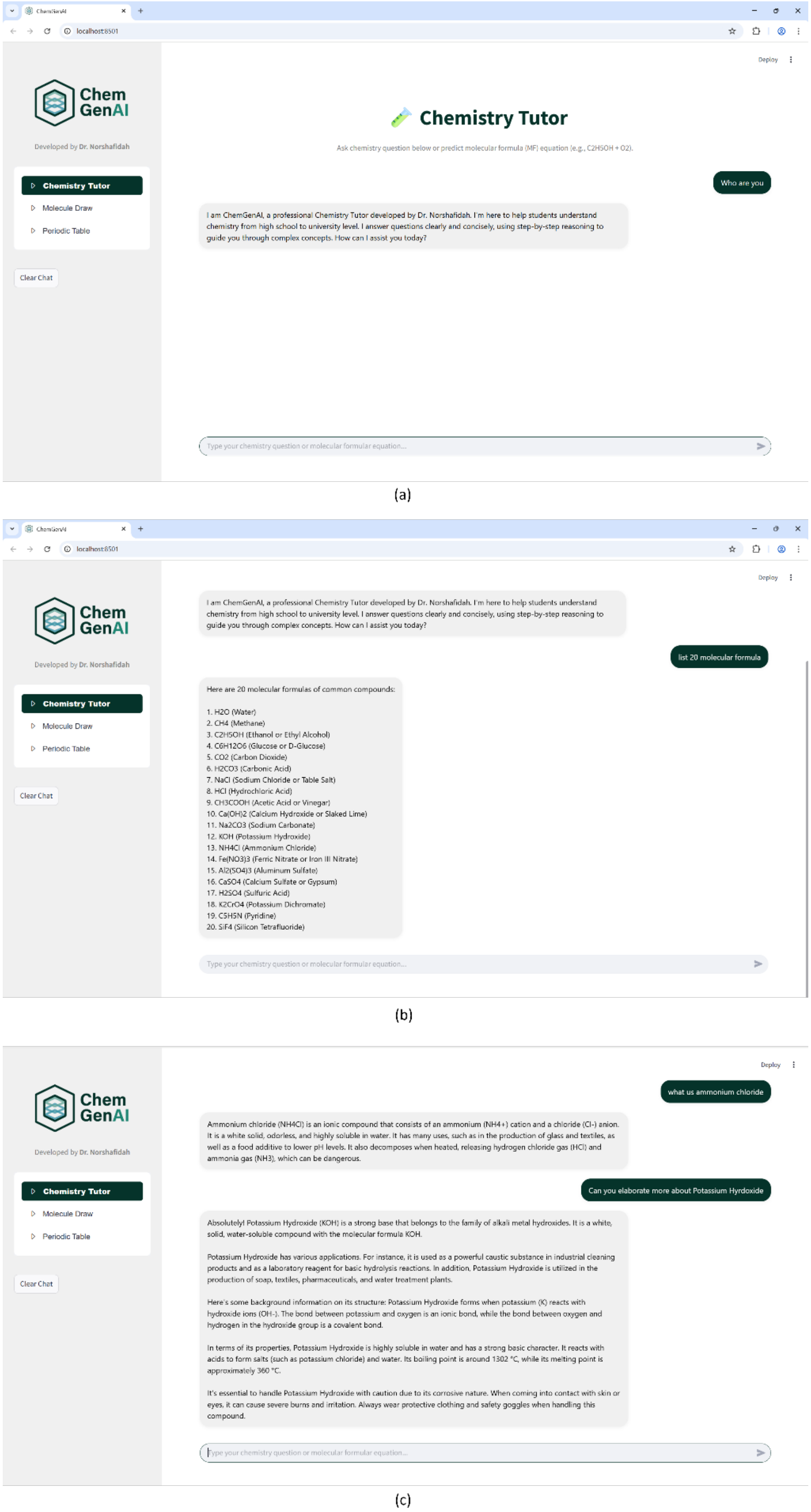

Fig. 1. Screenshots of ChemGenAI's Chemistry Tutor module in use. (a) Initial system response introducing itself to the user. (b) Generated list of 20 molecular formulas of common compounds. (c) Informative explanations on the chemical compounds demonstrating both concise and extended responses

In Figure 2(a), the GUI is divided into three key regions arranged horizontally for ease of comparison and spatial awareness. On the above side, the system displays textual molecular properties, including the IUPAC name, SMILES representation, molecular formula, and molecular weight. This section provides essential descriptive information that supports chemical literacy and allows for quick reference. In the left of the interface, the 2D skeletal structure of the molecule is shown using a standard line-bond representation. This visualization helps users understand atomic connectivity, molecular symmetry, and bond angles in a familiar format. On the right side, the 3D conformation of the molecule is rendered using molecular modeling tools that apply force-field optimization to represent the spatial geometry. For methane (CH_4), the system correctly visualizes a tetrahedral shape, demonstrating its ability to model even basic molecular geometries accurately. This clear separation of properties, 2D, and 3D views within the same window ensures a cohesive, multimodal learning experience that strengthens students' understanding of structural representation and spatial reasoning in chemistry.

Figures 2(b) and 2(c) display water (H_2O) and ethanol (C_2H_6O) respectively. In each case, the correct IUPAC name, molecular formula, and approximate molecular weight are presented. The 3D conformation of ethanol demonstrates its bent geometry and bonding interactions, illustrating the system's ability to represent spatial structure and hybridization, which are important for understanding polarity and hydrogen bonding. Figures 2(d) and 2(e) show carbonic acid (CH_2O_2) and acetic acid ($C_2H_4O_2$). These examples test the module's performance with carboxylic acid functional groups. The outputs clearly show the presence of double bonds and hydroxyl groups in the 2D structures, while the 3D visualizations capture molecular geometry and electron cloud distributions. These visual cues assist users in recognizing acidic behavior and resonance structures in carboxylic acids.

In Figure 2(f), the system processes pyridine (C_5H_5N), a six-membered aromatic heterocycle. The system accurately detects the ring structure and places the nitrogen atom correctly within the ring. The 3D rendering maintains planarity and aromaticity, both key properties of such molecules. This result confirms the module's capacity to handle heterocyclic and aromatic systems, which are crucial in pharmaceutical and biochemical education. Figure 2(g) presents an inorganic molecule which is silicon tetrafluoride (SiF_4). The system generates a tetrahedral geometry in both 2D and 3D. This showcases the module's flexibility in visualizing not only organic but also simple inorganic compounds.

The results collectively demonstrate that the Molecule Draw module performs reliably across a diverse range of molecular inputs. The combination of RDKit for 2D parsing and coordinate generation, along with force-field optimization for 3D rendering, provides a valuable multimodal learning experience [14]. The availability of both symbolic and spatial representations supports a deeper understanding of molecular structure, functional groups, and stereochemistry. This capability is particularly relevant in modern research contexts such as photovoltaics, where redox-active molecules, interfacial engineering, and structural tuning significantly impact device performance [15-22]. It is equally important in medicinal chemistry, where the spatial arrangement of atoms, functional group behavior, and stereoisomerism influence drug activity and receptor binding [23, 24]. Furthermore, in environmental chemistry, molecular understanding is essential for analyzing pollutants, designing greener compounds, and interpreting reaction mechanisms in atmospheric or aqueous systems [25]. By enabling interactive exploration of molecules from multiple domains, ChemGenAI bridges foundational chemical education with real-world applications in sustainable energy, health, and the environment.

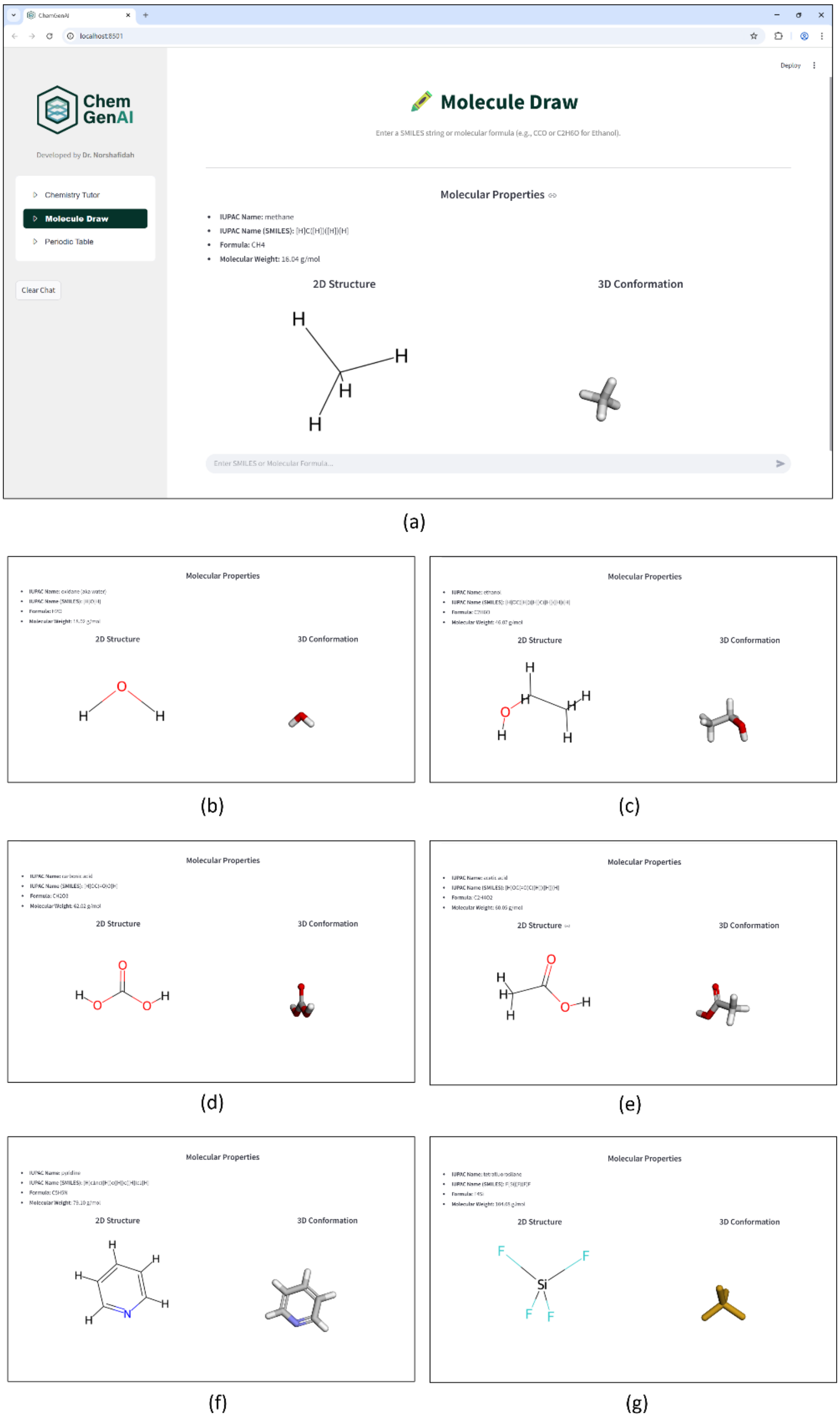

Fig. 2. Screenshots of the Molecule Draw module in ChemGenAI. (a) Methane; (b) Water; (c) Ethanol; (d) Carbonic acid; (e) Acetic acid; (f) Pyridine; (g) Silicon tetrafluoride. Each example shows molecular properties, 2D skeletal structure, and 3D conformation based on user input using SMILES or molecular formula

The Periodic Table module in ChemGenAI is designed to provide users with an interactive visual tool for exploring the elements and their fundamental properties [26]. The implementation of this interactive module provides users with immediate access to elemental data, improving understanding of periodic trends such as atomic size, metallic character, and group behaviors. By presenting data through an interactive and visually organized platform, the system supports both discovery-based learning and quick factual reference.

As shown in Figure 3(a), the periodic table is rendered in full color with appropriate group-based classification to enhance visual recognition and conceptual clarity. Each element block is color-coded according to its chemical category, such as alkali metals, transition metals, nonmetals, noble gases, metalloids, lanthanides, and actinides. This approach allows users to immediately distinguish between element types and recognize periodic trends, supporting chemistry learning through both symbolic and spatial representation.

The layout follows the conventional long-form periodic table format, with lanthanides and actinides placed separately at the bottom. Users can click on any element block to view more detailed information about the selected element, a feature that enhances engagement and supports interactive exploration. This interaction reflects the system's adherence to usability principles, particularly recognition over recall, by making essential data available through direct visual interaction rather than requiring memorization.

Figure 3(b) demonstrates the behavior of the interface when the element Vanadium (V) is selected. A pop-up overlay appears near the element block, displaying atomic number, atomic mass, group, period, and classification. In this case, Vanadium is identified as a transition metal located in group 5 and period 4, with an atomic number of 23 and an atomic mass of approximately 50.94. This detailed yet concise presentation provides essential data that is useful for both quick reference and classroom instruction. The display remains consistent across all element selections, ensuring uniform feedback and predictability of system behavior.

Figure 3(c) shows a similar interaction when the element Hydrogen (H) is selected. The pop-up correctly displays Hydrogen's atomic number 1, atomic mass 1.01, group 1, period 1, and classifies it as a nonmetal. The classification system is consistent and informative, helping users understand chemical relationships based on position in the periodic table. The inclusion of Hydrogen in the first group but with a nonmetal classification also highlights ChemGenAI's thoughtful handling of edge cases, which are common sources of confusion in chemistry education.

By combining visual interactivity with accurate chemical data, the Periodic Table module not only enhances foundational knowledge but also fosters conceptual understanding transferable to a wide range of real-world chemistry applications. Mastery of periodic trends such as atomic radius, electronegativity, oxidation states, and electronic configurations is essential for rational element selection and predicting chemical behavior across disciplines. In photovoltaic materials research, for instance, choosing metal cations like Pb^{2+}, Sn^{2+}, or Ge^{2+} in perovskite solar cells depends on understanding trends in ionic size and oxidation stability. Halide substitution (Cl^-, Br^-, I^-) affects bandgap tuning, which directly impacts solar cell efficiency [27, 28]. Similarly, in organic photovoltaic interfaces, elements such as molybdenum are chosen for their stable oxidation states and conductivity-enhancing properties. In coordination chemistry, periodic trends help predict ligand field stabilization energy, metal-ligand bonding strength, and preferred geometries. Transition metals across the d-block exhibit characteristic reactivity that is critical in catalyst design, redox tuning, and metallodrug development. For example, platinum and ruthenium complexes are used in cancer therapeutics due to their ability to interact with DNA, guided by their electronic configuration and redox potential [29].

Journal of Advanced Research in Computing and Applications
Volume 40, Issue 1 (2025) 63-75

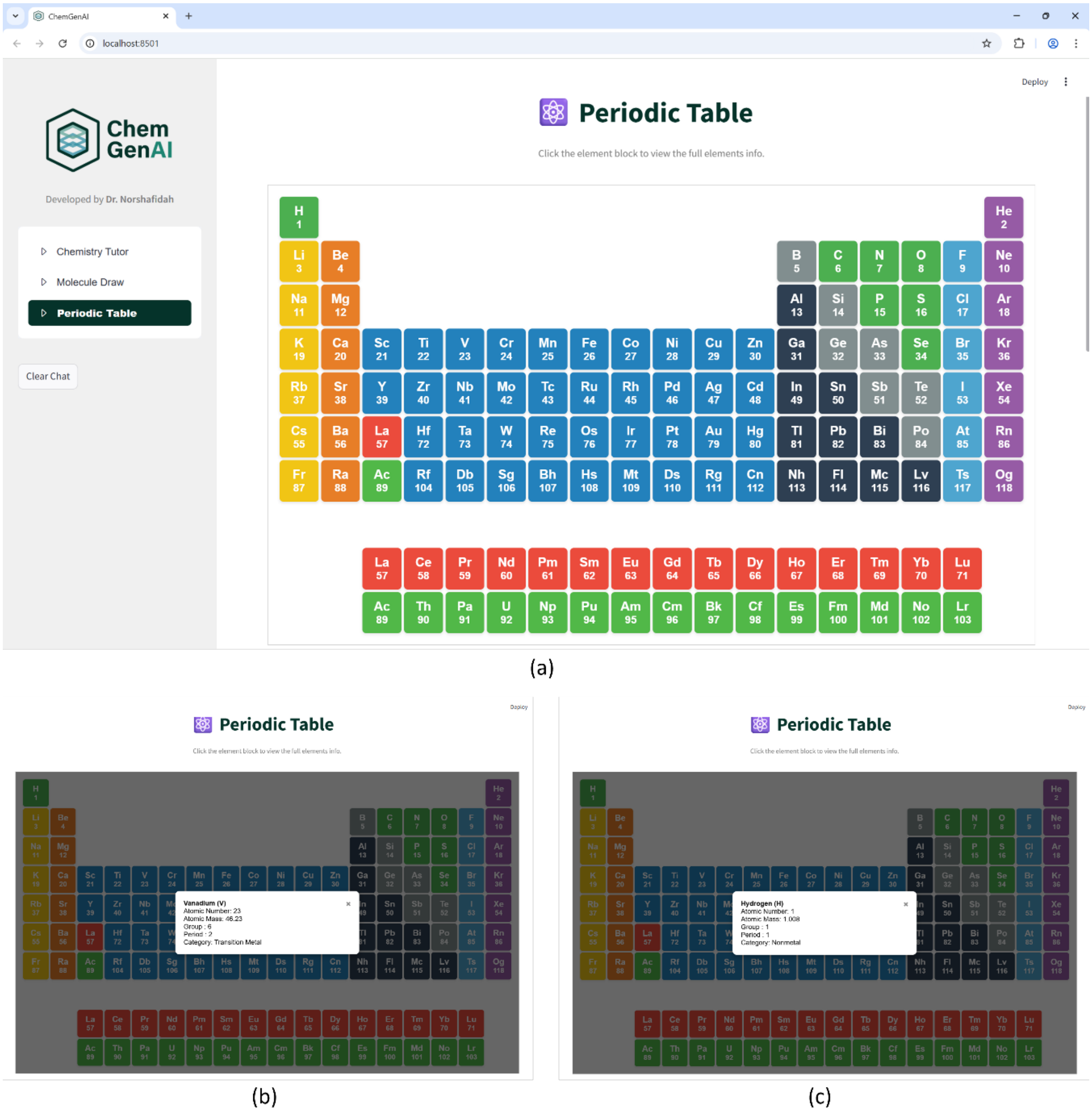

Fig. 3. Screenshots of the interactive Periodic Table module in ChemGenAI. (a) Full periodic table with color-coded element classification for quick visual reference. User interaction with (b) vanadium and (c) hydrogen showing a pop-up with atomic properties and classification

In battery and electrochemical research, the periodic table guides the choice of anode and cathode materials. Lithium, sodium, and potassium are evaluated based on their standard reduction potentials and atomic mass, influencing energy density and charge mobility [30]. Understanding trends across alkali and alkaline earth metals is crucial in optimizing performance and cost-effectiveness. In nanomaterials and semiconductors, dopant selection such as phosphorus or boron in silicon lattices is governed by periodic trends in valence electrons and atomic size [31]. These small substitutions significantly affect electrical conductivity and are widely used in electronics and sensor technologies. Even in biochemistry, understanding the essential roles of elements like iron (heme cofactors), zinc (enzyme catalysis), or calcium (cell signaling) depends on periodic insights. These elements are often studied for their coordination preferences and charge distribution, which influence biological activity. In the design of optical color filters and multilayer interference coatings, such as those used in semi-transparent solar cells and display technologies, periodic knowledge is essential when selecting materials with appropriate refractive indices, optical bandgaps, and chemical stability. Elements like silver, tin, and indium are commonly chosen for transparent electrodes or reflective layers due to their high conductivity and favorable optical properties. Similarly, dielectric materials such as silicon dioxide (SiO_2), titanium dioxide (TiO_2), and zinc oxide (ZnO) are selected based on trends in polarizability, electronegativity, and thermal stability [32-39]. Understanding these periodic properties enables researchers to fine-tune light transmission, reflection, and absorption for specific wavelengths.

By presenting such chemical data in an accessible and interactive way, ChemGenAI's Periodic Table module helps learners link abstract periodic concepts to concrete scientific applications. This not only enhances learning outcomes but also cultivates chemical intuition applicable across research fields. As interdisciplinary demands grow, tools like ChemGenAI play a vital role in preparing students to think critically about element selection and behavior in both academic and industrial contexts.

4. Conclusion

ChemGenAI demonstrates the feasibility and effectiveness of an offline, AI-driven educational platform for chemistry learning. By integrating a pre-trained large language model with cheminformatics tools such as RDKit, the system enables real-time question answering, molecular visualization, and interactive exploration of the periodic table, all without requiring internet access. The modular interface, designed in accordance with usability heuristics, provides a user-friendly experience that supports both textual and visual engagement. Through its Chemistry Tutor, Molecule Draw, and Periodic Table modules, ChemGenAI caters to a wide range of learning objectives, from conceptual understanding to structural analysis. The system's offline architecture ensures accessibility in low-connectivity environments while maintaining data privacy and operational control. Overall, ChemGenAI serves as a valuable tool for enhancing chemistry education, particularly in settings where traditional cloud-based solutions are not feasible.

Acknowledgement
Author acknowledge Geran Galakan Penyelidik Muda (GGPM), grant number GGPM-2023-048, funded by Universiti Kebangsaan Malaysia. Author also acknowledge Geran Translasi UKM (TR-UKM), grant number UKM-TR2024-09, funded by Universiti Kebangsaan Malaysia.

Conflict of Interest Statement
The author declares that there are no conflicts of interest related to the publication of this manuscript.

Author Contributions Statement
S.S. was responsible for the conceptualization, validation, formal analysis, investigation, resources, visualization, original draft preparation, review and editing of the manuscript, and project administration.

Data Availability Statement
The data supporting the findings of this study are available from the corresponding author upon request.

Ethics Statement

This study did not involve human participants, animals, or sensitive data, and therefore did not require ethical approval.

References

[1] Zhai, Xuesong, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, and Yan Li. "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020." *Complexity* 2021, no. 1 (2021): 8812542.https://doi.org/10.1155/2021/8812542

[2] Shafian, Shafidah, Fitri Norizatie Mohd Salehin, Sojeong Lee, Azlan Ismail, Shuhaida Mohamed Shuhidan, Lin Xie, and Kyungkon Kim. "Development of Organic Semiconductor Materials for Organic Solar Cells via the Integration of Computational Quantum Chemistry and AI-Powered Machine Learning." *ACS Applied Energy Materials* 8, no. 2 (2025): 699-722.https://doi.org/10.1021/acsaem.4c02937

[3] Iyamuremye, Aloys, Francois Niyongabo Niyonzima, Janvier Mukiza, Innocent Twagilimana, Pascasie Nyirahabimana, Theophile Nsengimana, Jean Dieu Habiyaremye, Olivier Habimana, and Ezechiel Nsabayezu. "Utilization of artificial intelligence and machine learning in chemistry education: a critical review." *Discover Education* 3, no. 1 (2024): 95.https://doi.org/10.1007/s44217-024-00197-5

[4] Shafian, Shafidah, Mohd Nizam Husen, Lin Xie, and Kyungkon Kim. "Predicting High-Performance Perovskite Solar Cells Using AI-based Machine Learning Models." *Materials Today Sustainability* 31 (2025): 101176.https://doi.org/10.1016/j.mtsust.2025.101176

[5] Cardellini, Liberato. "Chemistry: Why the subject is difficult?" *Educación química* 23 (2012): 305-310.

[6] Shafian, Norshafidah Binti Abu. "MolEdu: Empowering Chemistry Education with Open-Source Cheminformatics for Molecular Drawing." *Progress in Computers and Learning* 4, no. 1 (2025): 49-58.https://doi.org/10.37934/picl.4.1.4958

[7] Kasneci, Enkelejda, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, and Gjergji Kasneci. "ChatGPT for good? On opportunities and challenges of large language models for education." *Learning and individual differences* 103 (2023): 102274.https://doi.org/10.1016/j.lindif.2023.102274

[8] Buscemi, AlessioDaniele Proverbio. "Chatgpt vs gemini vs llama on multilingual sentiment analysis." *arXiv preprint arXiv:2402.01715* (2024): https://doi.org/10.48550/arXiv.2402.01715

[9] Zhang, Mingxuan, Bo Yuan, Hanzhe Li, and Kangming Xu. "LLM-Cloud Complete: Leveraging cloud computing for efficient large language model-based code completion." *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023* 5, no. 1 (2024): 295-326.https://doi.org/10.60087/jaigs.v5i1.200

[10] Nielsen, Jakob. "Ten usability heuristics." (2005):

[11] Nielsen, JakobRolf Molich. *Heuristic evaluation of user interfaces*. in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1990.

[12] Bento, A Patrícia, Anne Hersey, Eloy Félix, Greg Landrum, Anna Gaulton, Francis Atkinson, Louisa J Bellis, Marleen De Veij, and Andrew R Leach. "An open source chemical structure curation pipeline using RDKit." *Journal of Cheminformatics* 12 (2020): 1-16.https://doi.org/10.1186/s13321-020-00456-1

[13] O'Boyle, Noel M. "Towards a Universal SMILES representation-A standard method to generate canonical SMILES based on the InChI." *Journal of cheminformatics* 4 (2012): 1-14.https://doi.org/10.1186/1758-2946-4-22

[14] Lovrić, Mario, José Manuel Molero, and Roman Kern. "PySpark and RDKit: moving towards big data in cheminformatics." *Molecular informatics* 38, no. 6 (2019): 1800082.https://doi.org/10.1002/minf.201800082

[15] Shin, Solbi, Shafidah Shafian, Ka Yeon Ryu, Young-Kyo Jeon, Won-Suk Kim, and Kyungkon Kim. "Solution-Processed TiO2 Nanoparticles Functionalized with Catechol Derivatives as Electron Transporting Layer Materials for Organic Photovoltaics." *Advanced Materials Interfaces* 9, no. 14 (2022): 2200118.https://doi.org/10.1002/admi.202200118

[16] Kim, Hyunkyoung, Yuchan Heo, Yeji Na, Shafidah Shafian, BongSoo Kim, and Kyungkon Kim. "Cross-Linking-Integrated Sequential Deposition: A Method for Efficient and Reproducible Bulk Heterojunctions in Organic Solar Cells." *ACS Applied Materials & Interfaces* 16, no. 41 (2024): 55873-55880.https://doi.org/10.1021/acsami.4c13237

[17] Kim, Hyunkyoung, Ye-Jin Kong, Won-Suk Kim, Shafidah Shafian, and Kyungkon Kim. "Enhancing Reproducibility in Organic Solar Cell Fabrication via Static Sequential Deposition with Cross-Linked Polymer Donor and Nonfullerene

Acceptor." *ACS Applied Polymer Materials* 6, no. 10 (2024): 5814-5821.https://doi.org/10.1021/acsapm.4c00477

[18] Wang, Peng, Shafidah Shafian, Feng Qiu, Xiao Zhang, Yuping Zhao, Bin Wu, Kyungkon Kim, Yong Hua, and Lin Xie. "Improving redox reactions of Spiro-OMeTAD via p-type molecular scaffold to reduce energy loss at Ag-electrode in perovskite solar cells." *Journal of Energy Chemistry* 102 (2025): 151-160.https://doi.org/10.1016/j.jechem.2024.10.027

[19] Zhang, Xiao, Linqing Wang, Shafidah Shafian, Peng Wang, Yuping Zhao, Pengcheng Wang, Bin Wu, Jiaxue Zhai, Jiayan Chen, and Licheng Sun. "Crosslinking-Driven Chemical Homogeneity Enhances Performance of Pre-Seeded Perovskite Solar Cells." *Small* 21, no. 6 (2025): 2408362.https://doi.org/10.1002/smll.202408362

[20] Ryu, Ka Yeon, Shafidah Shafian, Jongchan Shin, Yu Jin Lee, Minjae Lee, and Kyungkon Kim. "Linear polyurethane ionenes for stable interlayer of organic photovoltaics." *Journal of Power Sources* 542 (2022): 231772.https://doi.org/10.1016/j.jpowsour.2022.231772

[21] Lee, Sojeong, Young Seon Yoon, Shafidah Shafian, Jin Young Kim, and Kyungkon Kim. "Sequential Co-Deposition of Perovskite Film: An Effective Way of Tailoring Bandgap in All Vacuum Processed Perovskite Solar Cells." *Small Methods* (2025): 2500104.https://doi.org/10.1002/smtd.202500104

[22] Hong, Minjeong, Jiyae Youn, Ka Yeon Ryu, Shafidah Shafian, and Kyungkon Kim. "Improving the Stability of Non-fullerene-Based Organic Photovoltaics through Sequential Deposition and Utilization of a Quasi-orthogonal Solvent." *ACS Applied Materials & Interfaces* 15, no. 16 (2023): 20151-20158.https://doi.org/10.1021/acsami.3c02071

[23] Nogrady, ThDonald F Weaver, *Medicinal chemistry: a molecular and biochemical approach*. 2005: Oxford University Press.

[24] Nicholls, Anthony, Georgia B McGaughey, Robert P Sheridan, Andrew C Good, Gregory Warren, Magali Mathieu, Steven W Muchmore, Scott P Brown, J Andrew Grant, and James A Haigh. "Molecular shape and medicinal chemistry: a perspective." *Journal of medicinal chemistry* 53, no. 10 (2010): 3862-3886.https://doi.org/10.1021/jm900818s

[25] Mamy, Laure, Dominique Patureau, Enrique Barriuso, Carole Bedos, Fabienne Bessac, Xavier Louchart, Fabrice Martin-Laurent, Cecile Miege, and Pierre Benoit. "Prediction of the fate of organic compounds in the environment from their molecular properties: a review." *Critical reviews in environmental science and technology* 45, no. 12 (2015): 1277-1377.https://doi.org/10.1080/10643389.2014.955627

[26] Schwerdtfeger, Peter, Odile R Smits, and Pekka Pyykkö. "The periodic table and the physics that drives it." *Nature reviews chemistry* 4, no. 7 (2020): 359-380.https://doi.org/10.1038/s41570-020-0195-y

[27] Qi, Shuwen, Chenghao Ge, Peng Wang, Bin Wu, Yuping Zhao, Rongjun Zhao, Shafidah Shafian, Yong Hua, and Lin Xie. "Improving Perovskite Solar Cell Performance and Stability via Thermal Imprinting-Assisted Ion Exchange Passivation." *ACS Applied Materials & Interfaces* 16, no. 38 (2024): 51037-51045.https://doi.org/10.1021/acsami.4c08538

[28] Zhai, Jia Xue, Lin Xie, and Shafidah Shafian. "Advancements in perovskite/CIGS tandem solar cells: Material synergies, device configurations, and economic viability for sustainable energy." *Nanotechnology Reviews* 14, no. 1 (2025): 20250196.https://doi.org/10.1515/ntrev-2025-0196

[29] Reed, E. "Platinum-DNA adduct, nucleotide excision repair and platinum based anti-cancer chemotherapy." *Cancer treatment reviews* 24, no. 5 (1998): 331-344.https://doi.org/10.1016/S0305-7372(98)90056-1

[30] Feng, Xuyong, Hong Fang, Nan Wu, Pengcheng Liu, Puru Jena, Jagjit Nanda, and David Mitlin. "Review of modification strategies in emerging inorganic solid-state electrolytes for lithium, sodium, and potassium batteries." *Joule* 6, no. 3 (2022): 543-587.https://doi.org.10.1016/j.joule.2022.01.015

[31] Mehendale, Shruti, Kalliopi Kanaki, Marco Povoli, Andreas Tefre Samnøy, Ganesh Tambave, Angela Kok, Carina Höglund, Susann Schmidt, Saima Sultana Kazi, and Isabel Llamas-Jansa. "Characterization of boron-coated silicon sensors for thermal neutron detection." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 972 (2020): 164124.https://doi.org/10.1016/j.nima.2020.164124

[32] Salehin, Fitri Norizatie Mohd, Puvaneswaran Chelvanathan, Adamu Ahmed Goje, Norasikin Ahmad Ludin, Mohd Adib Ibrahim, and Shafidah Shafian. "Design of blue, green and red colorful semitransparent films using Ag/SnO2/Ag color filter for integrated into solar cells." *Results in Physics* 70 (2025): 108172.https://doi.org/10.1016/j.rinp.2025.108172

[33] Park, Suhyeon, Shafidah Shafian, Juhwan Lee, Seungyun Jo, Seungbae Jeon, Seungjae Lee, Ding Shangxian, Hyungju Ahn, Kyungkon Kim, and Du Yeol Ryu. "High-Efficiency Structural Coloration Enabled by Defect-Free Block

Copolymer Self-Assembly for a Solar Cell Distributed Bragg Reflector." *Advanced Optical Materials* 11, no. 24 (2023): 2301357.https://doi.org/10.1002/adom.202301357

[34] Shafian, ShafidahKyungkon Kim. "Panchromatically responsive organic photodiodes utilizing a noninvasive narrowband color electrode." *ACS Applied Materials & Interfaces* 12, no. 47 (2020): 53012-53020.https://doi.org/10.1021/acsami.0c17183

[35] Kim, Bo Youn, Shafidah Shafian, and Kyungkon Kim. "High-Performance Semitransparent Color Organic Photodiodes Enabled by Integrating Fabry–Perot and Solution-Processed Distributed Bragg Reflectors." *Advanced Materials Interfaces* 10, no. 31 (2023): 2300421.https://doi.org/10.1002/admi.202300421

[36] Shafian, Shafidah. "Optical and Structural Analysis of Layer Thickness Effects on Transmission in Semitransparent Colorful Organic Photovoltaics." *PaperASIA* 41, no. 4b (2025): 1-10.https://doi.org/10.59953/paperasia.v41i4b.569

[37] You, Young-Jun, Muhammad Ahsan Saeed, Shafidah Shafian, Jisoo Kim, Sang Hyeon Kim, Sung Hyeon Kim, Kyungkon Kim, and Jae Won Shim. "Energy recycling under ambient illumination for internet-of-things using metal/oxide/metal-based colorful organic photovoltaics." *Nanotechnology* 32, no. 46 (2021): 465401.https://doi.org/10.1088/1361-6528/ac13e7

[38] Shafian, Shafidah, Jieun Son, Youngji Kim, Jerome K Hyun, and Kyungkon Kim. "Active-material-independent color-tunable semitransparent organic solar cells." *ACS Applied Materials & Interfaces* 11, no. 21 (2019): 18887-18895.https://doi.org/10.1021/acsami.9b03254

[39] Kim, Youngji, Jieun Son, Shafidah Shafian, Kyungkon Kim, and Jerome K Hyun. "Semitransparent blue, green, and red organic solar cells using color filtering electrodes." *Advanced Optical Materials* 6, no. 13 (2018): 1800051.https://doi.org/10.1002/adom.201800051