Commentary

Kepler: The Pioneer of Data Science and AI

Ronaldo Mota[1]

1. Chair in Artificial Intelligence, Universidade Federal do Rio de Janeiro, Brazil

This article examines Johannes Kepler's pioneering contributions to data-driven scientific discovery and draws parallels with modern advancements in Artificial Intelligence (AI). Kepler's rigorous analysis of Tycho Brahe's astronomical data led to the formulation of the fundamental laws of planetary motion, exemplifying early principles of data science. Contemporary AI techniques—such as symbolic regression, advanced reasoning neural networks, and explainable AI—can now rediscover physical laws from large datasets, mirroring Kepler's methodology but at an unprecedented scale. Despite technological progress, challenges persist, including data quality, interpretability, and validation. The synergy between human intuition and machine intelligence holds promise for accelerating scientific breakthroughs across disciplines, extending Kepler's legacy into the era of big data and AI-driven discovery.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

Kepler as a Data Scientist

In 1600, Johannes Kepler (1571-1630) began working as an assistant to the Danish astronomer Tycho Brahe (1546-1601) in Prague. Brahe had compiled one of the most precise astronomical datasets of the pre-telescopic era through meticulous naked-eye observations. While Brahe excelled in data collection, Kepler's genius lay in his analytical rigor. After inheriting Brahe's observations in 1601, Kepler spent over a decade identifying mathematical patterns, systematically testing hypotheses until he derived his revolutionary laws of planetary motion.

Initially, Kepler, an adherent of Copernicus's heliocentric model, sought evidence that planets orbited the Sun in perfect circular paths—a reflection of the prevailing belief in cosmic harmony. However, Brahe's

 doi.org/10.32388/OEEG86.2

data revealed a groundbreaking truth: planetary orbits were elliptical, with the Sun at one focus. This discovery became Kepler's First Law of Planetary Motion, heralding a paradigm shift in astronomy.

Kepler's three laws of planetary motion are as follows: 1. "The Elliptical Orbit Law": Planets orbit the Sun in elliptical paths, with the Sun at one focus; 2. "The Equal Area Law": A planet's radius vector (the line connecting it to the Sun) sweeps out equal areas in equal time intervals; and 3. "The Harmonic Law": The square of a planet's orbital period is proportional to the cube of its semi-major axis.[1]

These laws not only described celestial mechanics with unprecedented mathematical precision but also established a new scientific paradigm: the universe could be understood through empirical data and quantitative relationships. Unlike Aristotle's reliance on qualitative deduction, Kepler exemplified the empirical approach that would later define the scientific method.

The Role of AI in Scientific Discovery

A compelling question arises in the modern era: Could an AI system, given Brahe's data, replicate Kepler's discoveries? The answer is affirmative. A core objective of science is deriving mathematical models that accurately describe empirical phenomena. Traditionally, scientists manually construct such models using domain expertise, fitting them to observational data. Today, however, machine learning enables automated model discovery from vast datasets.

Cornelio et al.[2] introduced a novel approach combining logical reasoning with symbolic regression to derive scientific principles from both axiomatic knowledge and experimental data. Their method successfully rediscovered Kepler's Third Law and other physical laws, demonstrating AI's ability to identify governing equations even with limited data by evaluating candidate formulas for empirical accuracy.

Similarly, Li et al.[3] proposed an Explainable AI (XAI)-driven paradigm for scientific discovery, wherein AI assists in hypothesis generation, data interpretation, and insight extraction. Their work illustrates how AI can autonomously derive Kepler's Laws from Brahe's astronomical data, bridging computational and experimental methodologies.

Modern machine learning techniques—including neural networks, symbolic regression (SR), and genetic programming—can uncover complex patterns and even derive physical laws from data. Notable examples include: "Eureqa" (Wolfram Alpha)[4], which rediscovered the Law of Conservation of Energy from pendulum motion data; "AI Feynman" (MIT)[5], which reconstructed equations of relativity and

doi.org/10.32388/OEEG86.2

other physical laws; and “CERN’s AI applications”[6], which assist in detecting new particles in high-energy collisions.

AI Building the Future: Advancing Discovery and Reasoning

AI is unlocking powerful new ways to uncover the fundamental laws of nature from data. Among these, symbolic regression stands out as a compelling machine learning approach that extracts interpretable mathematical expressions directly from datasets [7]. In general, driven by genetic programming, recent advances have introduced deep learning techniques as a dynamic, data-driven tool for model discovery. This shift has led to remarkable progress across scientific and engineering domains, from theoretical research to real-world applications.

A particularly promising development is the integration of Kolmogorov-Arnold Networks (KANs) with symbolic regression [8]. Rooted in the Kolmogorov-Arnold representation theorem, KANs break down complex multivariate functions into sums and compositions of simpler univariate functions. This approach offers enhanced interpretability, high approximation accuracy, and broad modelling flexibility, making it a powerful framework for revealing hidden mathematical relationships in data. By combining KANs with symbolic regression, researchers can derive meaningful, transparent models that accelerate scientific discovery.

Beyond mathematical modelling, AI is also making strides in advanced reasoning and metacognition. Recent work emphasises the possibility of reinforcement learning (RL) enabling AI systems to self-regulate their cognitive processes, dynamically adjusting their problem-solving strategies for optimal performance [9]. DeepSeek-R1-Zero, an advanced AI model, learned to allocate processing time adaptively, prioritizing complex problems while maintaining coherence in reasoning. Rather than relying on rigid, predefined rules, it developed autonomous reasoning strategies guided by incentive structures, showcasing a shift toward more self-aware AI systems.

As AI continues to evolve, the next generation of model, whether based on symbolic regression, KANs, or advanced reasoning architectures, will push the boundaries of scientific discovery and intelligent automation. These innovations promise to transform fields ranging from fundamental research to industrial applications, paving the way for a future where AI not only assists but also autonomously advances human knowledge.

doi.org/10.32388/0EEG86.2

Despite these successes, AI faces significant challenges: 1. “Data dependency”: Noise or biases in data can lead to erroneous conclusions; 2. “Interpretability”: Some models detect patterns without explaining their underlying mechanisms; and 3. “Validation”: Empirical testing remains essential to confirm AI-generated hypotheses. Thus, while AI enhances discovery, human intuition and experimental verification remain indispensable.[10]

Conclusion

Kepler was a pioneer of data science, demonstrating how empirical observation and mathematical analysis could unveil the fundamental laws of nature. Today, AI amplifies this capability, enabling machines to uncover patterns and formulate hypotheses at an unprecedented scale, unlocking powerful new ways to uncover the fundamental laws of nature from data. Among these AI, symbolic regression and advanced reasoning stand out as compelling machine learning approaches that extract interpretable mathematical expressions directly from datasets

The convergence of AI, big data, and mathematical modeling holds transformative potential across physics, biology, chemistry, and even the social sciences. However, just as Kepler’s creativity and rigor were essential in interpreting Brahe’s data, the collaboration between human insight and machine intelligence will remain crucial in unraveling the universe’s mysteries.

Kepler was not merely an astronomical genius—he was a visionary who foreshadowed the era of data-driven science. His legacy endures, now revitalized by the power of AI.

About the Author

Ronaldo Mota holds the Chair in Artificial Intelligence at the Brazilian School of Advanced Studies at the Federal University of Rio de Janeiro.

Statements and Declarations

Funding

This work was carried out with the support of the Rio de Janeiro State Research Support Foundation (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ), through the “Visiting Researcher

doi.org/10.32388/0EEG86.2

Emeritus Program" associated to the "Artificial Intelligence and Metacognition" Chair at the "Brazilian School of Advanced Studies" of the Federal University of Rio de Janeiro (CBAE/UFRJ).

References

1. ^Kidder SQ (2003). "Kepler Laws." Encyclopaedia of Atmospheric Sciences. https://www.sciencedirect.com/topics/physics-and-astronomy/kepler-laws.
2. ^Cornelio C, Dash S, Austel V, Josephson TR, Goncalves J, Clarkson KL, Megiddo N, El Khadir B, Horesh L (2023). "Logical-Symbolic Discovery of Physical Laws." Nature Communications. **14**(1). doi:10.1038/s41467-023-37236-y.
3. ^Li Z, Ji J, Zhang Y (2021). "From Kepler to Newton: Explainable AI for Science." arXiv. doi:10.48550/arXiv.2111.12210.
4. ^Schmidt M, Lipson H (2009). "Symbolic Regression of Implicit Equations." In: Riolo R, O'Reilly UM, McConaghy T (eds) Genetic Programming Theory and Practice VII. Boston, MA: Springer. 73–85. doi:10.1007/978-1-4419-1626-6_5.
5. ^Udrescu SM, Tegmark M (2020). "AI Feynman: A physics-inspired method for symbolic regression." Science Advances. doi:10.1126/sciadv.aay2631.
6. ^How can AI help physicists search for new particles? CERN. 13 June 2024. https://home.cern/news/news/physics/how-can-ai-help-physicists-search-new-particles; Brandstetter J (2025). "Emmi AI: Data-Driven Industrial Simulations." Emmi AI. https://www.emmi.ai/.
7. ^Makke N, Chawla S (2024). "Interpretable Scientific Discovery with Symbolic Regression: A Review." Artificial Intelligence Review. doi:10.1007/s10462-023-10622-0.
8. ^Harvey TR, et al. (2025). "Symbolic Regression with Multimodal Large Language Models and Kolmogorov-Arnold Networks." arXiv. https://arxiv.org/abs/2505.07956.
9. ^Mota R (2025). "Is DeepSeek a Metacognitive AI?" QEIOS. https://www.qeios.com/read/PJ3POM.2.
10. ^Liang Y, et al. (2025). "When Mathematical Methods Meet Artificial Intelligence and Mobile Edge Computing." Mathematics. **13**(11):1779. https://www.mdpi.com/2227-7390/13/11/1779.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/0EEG86.2