Pedagogical Insight in Digital Education from Traditional Big Data Analysis to AI Driven Solution: A Case Study in Junior Science lesson

Fung CHEUNG 1, Hin Nang YUE 1, Sze Lok CHOW1
[1]New Life Schools Incorporation Lui Kwok Pat Fong College, Hong Kong SAR, China
* cf@lkpfc.edu.hk
* yhn@lkpfc.edu.hk
* cslo@lkpfc.edu.hk

Abstract: *This paper evaluates traditional data analysis methods within the Hong Kong Junior Secondary (KS3) science curriculum and proposes a pedagogical shift from manual data collection toward AI-human collaboration. With a pilot scheme, the study suggests pedagogy and curriculum to help teachers, integrating AI, prioritizing intuitive judgment and deliberate practice, equipping students to navigate complex world of science with critical sensitivity.*
Keywords: Digital Education, Science Education, Innovation Technology Education, Artificial Intelligence, Big Data analysis, Anomaly detection

1. Challenges in Data Analysis

Currently, data analysis is included in Hong Kong science education. Students are expected to collect and analyze data before making conclusions in science experiment. In Key Stage 3 (KS3) science curriculum, there are experiments designed for students to learn and experience the process of finding the relationship between two variables, for example, collecting pH values in neutralization and interpreting distance time graphs. What if the way that we process the teaching of data analysis, can be improved with the help of AI?

Research shows that scientific data has unique and difficult properties, as follow, that make it much harder to study than everyday information (Ramalli & Pernici, 2023).

1.1. High Costs and Limited Samples

Collecting scientific data is time consuming and expensive. This is because of the expensive materials and equipment in the on-field measurement. Because of these high costs in time and financial resources, experiments are hard to repeat. Besides experimental data, predictive models and simulations are other types of scientific data involved. While building models is complex and simulations are expensive, these factors lead to small quantities of these data available. Consequently, analyses of scientific data become limited too.

1.2. Problem in anomaly detection

Anomaly detection refers to detecting patterns or behaviors that are different from the norm of other data. The process involves data entry, sorting and manual labeling of the outlier. Manual data analysis therefore can result in human error in handling and interpreting the data. On the other hand, labeling limited amounts of data is easy but when it comes to larger datasets, labeling all the data, be it normal or abnormal, becomes challenging. This stems from inconsistent judgement from different people handling the data, human error in repeating the task. Besides, there is a cognitive burden in understanding and synthesizing data from different sources that usually causes people to overlook relationships and delay in identifying important signals. (Hassani & Silva, 2025)

2. Our Pilot Scheme

Teachers design lesson to teach S.2 students basic data analysis tools, with the help of AI, in the topic of Force and Motion. There will be students learning about the concept of “best-fit line” and its use in the data analysis. The power of

AI, in handling big data, will be introduced to the students. Central to this module is the use of Tracker, an AI-augmented video analysis tool, which enables students to model motion and construct distance-time graphs through automated object tracking. By analyzing video footage of a moving ball, students can directly observe the correlation between physical displacement and real-time data plotting. This study adopts post-test design to evaluate the efficacy of AI-integrated tools in physics education. The participants are divided into an experimental group (n>30), which will utilize the AI-augmented 'Tracker' software and a control group (n>30) consisting of parallel S.2 classes, which will receive instruction through traditional pedagogical methods. Data collections involve summative assessment by the end of the term, to measure the impact of the intervention. The analysis of the data focuses on students' understanding of distance-time graph.

Moreover, basic anomaly detection in big data will be discussed. Students will learn about the importance of anomaly detection as extended learning in data analysis. Furthermore, teachable machine of anomaly detection will be taught, and supervised learning as well as unsupervised learning will be discussed. The pilot scheme acknowledged the supports from the Innovation Technology Education Division of EDB and Physics Department of the Chinese University of Hong Kong.

3. How AI Helps with data analysis

AI acts as a powerful assistant that can handle science data analysis. Some AI models are even described as "unicorn data scientists" (Hassani et al., 2021).

3.1. Handling Big data

There will be a cognitive burden for humans to process and synthesize a lot of information but using AI can help. Take an existing example, when AI is integrated in handling clinical data, research has shown more than 6 times improvement in data processing while accuracy is doubled. While integration of AI does not mean replacing human judgement, the potential of AI and human collaboration in complex situations is great. Besides, AI allows researchers to spend more time thinking about new and more important ideas (Hassani & Silva, 2023).

3.2. Anomaly detection

The traditional method of anomaly detection lacks flexibility and scalability since it is hard to detect evolving anomalies and human error is inevitable. (Thanh, 2025) Instead of rule-based detection which uses the criterion set to find anomaly, AI uses deep learning or machine learning in anomaly detection. AI is considered more adaptive because it can handle more complex and dynamic patterns. There are generally 3 ways of anomaly detection: supervised, semi supervised and unsupervised learning. In supervised learning, labeled data are provided and thus the AI is trained to find anomalies using these data. In unsupervised learning, unlabeled data are given for training AI and the patterns that are different from other data are detected as anomalies. In semi supervised learning, small amounts of labeled data and large amounts of unlabeled data are used in the training. The role of semi supervised and unsupervised learning becomes more important especially when labeled data is limited.

3.3. Train the machine learning, train the computer

When there is not enough real data for training the machine, AI can create "synthetic data" based on the structure and pattern learned from the data provided. (Hassani & Silva, 2023) By giving appropriate prompts to the AI, it can then generate data that has similar style to the data provided. This allows scientists to train their computers to find answers even when they only have a few real-life examples to study. This can also help reduce time and cost in the process.

4. Inspiration for Education

Traditionally, students are provided with materials and apparatus to conduct experiments and collect data. These data are usually labeled by students as correct data or unusual data. Here, students find and label the anomalies by themselves, if any, based on the criterion set. As we all know, there is more emphasis on digital education in the education sector, ranging from enhancing learning and teaching from funding, integrating AI into curriculum, buying hardware or software from funding and systematic training provided to teachers. What if the idea of AI unsupervised learning is implemented at the junior level where students communicate and evaluate the accuracy of AI analysis? What if AI semi supervised learning is employed so that students can evaluate the effectiveness of the labeled data provided by themselves? The potential of AI analysis can be explored in the classroom where curiosity is filled.

4.1 The innovation from intuitive judgement

Anima Anandkumar, the Bren Professor of Computing and Mathematical Sciences at Caltech pointed out that it is getting more important to know when to trust, question and pause the recognition of AI(Carter, 2025). It is found that the collaboration of intuitive judgement and computational system shows improvement in performance and efficiency of real-life situation. Therefore, the question comes to how intuitive judgement can be trained so that students can better utilize AI tools.

Traditional pedagogy focuses on teaching students reasoning through reproducing instead of creating by themselves. Analytical skills are valued much more than perceptual training. Also, ambiguity and problem framing are not easily found in the courses because most of the questions are already well defined and students are not required to identify the problem. While the complex system nowadays requires more intuitive sensitivity for better adaptability and precision, it would be beneficial if our student's intuition can be trained at an earlier stage. Chi (2020) and Ericsson and Pool (2016) mentioned the importance of deliberate practice with feedback in evolving intuition which is an area that teachers seldom talk about in the lesson. When the intuition of surgeons can be improved by finding anomaly from AI system, what if this idea can be applied to our education? More can be done to train students intuitive pattern recognition with the aid of AI.

5. Conclusion

The integration of Artificial Intelligence opens a new page for scientific inquiry, moving from the limitations of manual data processing to an era of high-speed, synthetic, and automated analysis. This thesis has demonstrated that while traditional methods struggle with high costs and cognitive burdens, AI tools offer solutions that enhance both accuracy and efficiency.

In the context of Hong Kong's junior science education, this transition gives insight into pedagogical evolution. Rather than viewing AI as a replacement for human effort, it should be embraced as a collaborative partner. By offloading rote technical tasks to AI, the classroom can pivot toward cultivating intuitive judgment and critical pattern recognition. Ultimately, the goal of digital education must be to equip students with the sensitivity to question and guide these complex systems, ensuring they remain informed architects of discovery in an increasingly automated world.

References

Carter, E. (2025, November 4). Essay 2: Intuition in the age of AI. Harvard T.H. Chan School of Public Health. https://hsph.harvard.edu/news/essay-2-intuition-in-the-age-of-ai/

Chi, M. T. H. (2020). Thinking, learning, and problem solving. Routledge.

Curriculum Development Council. (2018). Secondary Education Curriculum Guide.

Ericsson, A., & Pool, R. (2016). Peak: Secrets from the New Science of Expertise. New York: Houghton Mifflin Harcourt.

Hassani, H., & Silva, E. S. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. Big Data and Cognitive Computing, 7(2), 62.

Hassani, H., & Silva, E. S. (2025). The role of LLMs in the evolution of scientific data analysis. arXiv. https://arxiv.org/html/2508.05519v2

Ramalli, E., & Pernici, B. (2023). Challenges of a Data Ecosystem for scientific data. Data & Knowledge Engineering, 148, 102236.

Thanh, D. B. (2025, April 14). AI anomaly detection: Enhancing data security and operational efficiency.

SmartDev. https://smartdev.com/ai-anomaly-detection/#How_AI_Powers_Anomaly_Detection