Artificial Intelligence in STEM

The Present and Future of Scientific Discovery

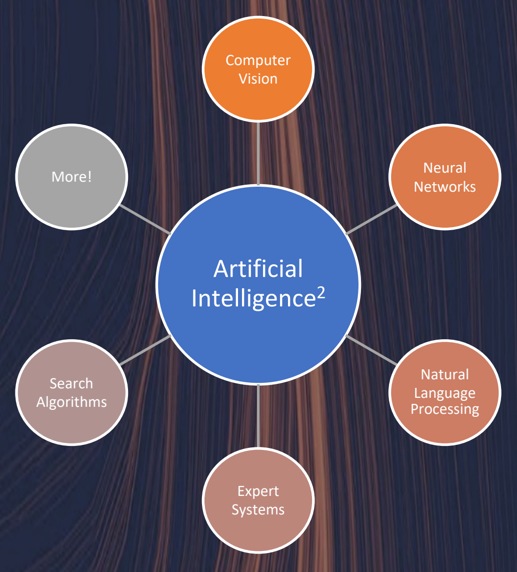
Hi there! I'm an artificial intelligence agent here to explain the uses of AI in STEM and how it might transform the way science is done. Before we dive in, let's cover the fundamentals of AI itself!

Artificial intelligence uses computer algorithms to imitate human intelligence, enabling machines to learn, reason, and make decisions.

Most modern artificial intelligence systems rely on machine learning (ML) and data, but non-ML models are still used today!

Machine learning leverages data and algorithms to give computers the ability to perform tasks without explicitly programming them to do so. The main goal of machine learning is *generalization*, which is when a model can accurately perform tasks it has not seen before.[1]

In our everyday lives, applications range from self-driving cars to face recognition. Many powerful systems are powered by neural networks, a model that mimics the human brain by connecting artificial neurons.



Different machine learning methods and large datasets are often used together to create powerful models. For instance, large language models such as ChatGPT are often trained on books, websites, and even social media!

- **Supervised Learning**: Trains algorithms to recognize patterns using examples (labeled data).
- **Unsupervised Learning**: Finds patterns or structure in data without being given labels or explicit goals.
- **Reinforcement Learning**: Learns to make decisions by receiving rewards for positive actions.
- **Large Language Models**: Processes and generates human-like text by learning from vast datasets of language.[3]

Current Uses of AI in STEM

Biology

Machine learning has accelerated the identification of potential drug candidates by predicting how different chemicals will react in the body using graph neural networks.

Computational models such as AlphaFold have achieved unprecedented accuracy in predicting 3D protein structure.[4]

Environmental Science

Precision agriculture powered by AI helps monitor soil and crop health, predict pest infestations, and optimize water usage, contributing to sustainable farming practices and reduced environmental footprint.

Computer vision models also use satellite imagery to monitor greenhouse gas emissions from space.[5]

Astronomy

Classification algorithms process vast amounts of data from telescopes and space probes to discover new celestial phenomena, like identifying exoplanets with NASA's Kepler space telescope.

Physics-informed neural networks lower the amount of data required for simulating fluid flows or gravitational fields.[6]

Robotics

Autonomous robots driven by reinforcement learning have enabled robotic exploration in environments hostile to humans. For instance, deep-sea robots explore the ocean floor, and Martian rovers use advanced computer vision software to identify geological features of interest and navigate the rough terrain.[7]

1950–56 Creation of the Turing Test and the birth of AI

1964 First AI chatbot, Eliza, speaks with humans

1965 Edward Feigenbaum introduces expert systems

1974 First autonomous vehicle is built

1974–80 AI Winter

1997 Deep Blue, a chess bot from IBM, beats World Champion Garry Kasparov

2000 Honda's ASIMO robot walks as fast as humans

2006 Creation of deep learning that marked the resurgence of neural networks

2012 A convolutional neural network reaches only 16% error on ImageNet

2016 AlphaGo defeats Go champion Lee Sodol

2022 Widespread publicization of GPT-3 and other LLMs[8]

While AI has already found many successful applications in STEM, one emerging approach involves incorporating scientific knowledge into AI itself to enhance its capabilities. This can help scientists accelerate their simulations, optimize their experiments, and form new hypotheses.[9]

```
C:\Users\ORISE>

show_example()
> In biochemistry, there are so many potential drugs (around 10^60) that no one could ever test them all. AI can quickly sort through these vast options and make   smart guesses about which ones are most likely to work well, almost as if they had done real experiments. AI methods are really good at handling these kinds of challenges because it can analyze vast amounts of data and consider many possible solutions at once. Advanced AI techniques such as active learning can even decide what data it needs next to improve its understanding, which means that each new piece of data it uses helps the AI system narrow down the range of possible solutions until it finds the most likely one.

This is also useful in fields like medical imaging, where one needs to figure out what is happening inside the body based on images like X-rays or MRIs, or in astronomy, where observing light from stars can explain what is happening millions of miles away.10
```

Particle physics experiments generate enormous amounts of data. Machine learning algorithms can efficiently identify significant events or patterns, such as the appearance of rare particles, which might otherwise be missed.[10]

Differential equations are often difficult to solve, but neural networks are being used to find solutions when traditional methods are impractical. Applications range from weather forecasting to the spread of disease.[10]

Direct air capture may be crucial for combatting climate change, and neural networks are finding new and possibly more effective materials that can be used to extract carbon dioxide from the air.[11]

Recent advancements in generative AI are also facilitating scientific discovery by generating new scientific hypotheses, analyzing text-based data, and designing new materials.

Generative AI

Language models can be *fine-tuned* with specialized datasets to enhance performance on specific tasks, yielding more precise, context-aware responses in scientific domains. Diffusion models, which are responsible for many generative art tools, can be used to enhance low-resolution images or create new molecules.

Example 1: Mathematical Problem Solving

Sophisticated LLMs are generally proficient in handling basic arithmetic and can often solve simple algebraic equations. Fine-tuned models have demonstrated abstract reasoning ability in generating proofs.[12]

Example 2: Biological Sequence Analysis

In fields such as astronomy or medical imaging, diffusion models can enhance low-resolution or noisy images in telescope images or MRI scans, making it easier to extract meaningful information.[13]

⚠ Safety Considerations for AI ⚠

- Hallucinations: AI models can generate false or misleading information, which can lead to incorrect conclusions or flawed scientific research.
- Interpretability: How exactly neural networks and language models work is not well understood, which can hinder scientists' ability to verify and trust the AI's results or predictions.
- Automation: Sophisticated models may automate tasks that were previously done by researchers and scientists, potentially causing job displacement and overreliance on AI.[14]

Erik Wang
Harvard University
November 2023

References

1. Santus, E., Christin, N., & Jayaram, H. (2020). *Technology Factsheet Series: Artificial Intelligence*. Belfer Center, Harvard Kennedy School. https://www.belfercenter.org/sites/default/files/2020-01/AI.pdf
2. Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. https://aima.cs.berkeley.edu/
3. Brown, S. (2021, April 21). *Machine learning, explained*. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained
4. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589.
5. Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S. M., Wang, J., Lin, C., ... & John, A. (2022). A review of earth artificial intelligence. *Computers & Geosciences*, *159*, 105034.
6. Martin, J., & Schaub, H. (2022). Physics-informed neural networks for gravity field modeling of the Earth and Moon. *Celestial Mechanics and Dynamical Astronomy, 134*(2), 13.
7. Verma, V., Maimone, M. W., Gaines, D. M., Francis, R., Estlin, T. A., Kuhn, S. R., ... & Thiel, E. R. (2023). Autonomous robotics is driving Perseverance rover's progress on Mars. *Science Robotics*, *8*(80), eadi3099.
8. Press, G. (n.d.). *A Very Short History Of Artificial Intelligence (AI)*. Forbes. Retrieved November 30, 2023, from https://www.forbes.com/sites/gilpress/2016/12/30/a-very-short-history-of-artificial-intelligence-ai/. All the dates on the timeline come from this source.
9. Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., ... & Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47-60.
10. AI4Science, M. R., & Quantum, M. A. (2023). The Impact of Large Language Models on Scientific Discovery: a Preliminary Study using GPT-4. *arXiv preprint arXiv:2311.07361*.
11. Sriram, A., Choi, S., Yu, X., Brabson, L. M., Das, A., Ulissi, Z., ... & Sholl, D. S. (2023). The Open DAC 2023 Dataset and Challenges for Sorbent Discovery in Direct Air Capture. *arXiv preprint arXiv:2311.00341*.
12. Birhane, A., Kasirzadeh, A., Leslie, D., & Wachter, S. (2023). Science in the age of large language models. *Nature Reviews Physics*, 1-4.
13. Mudur, N., & Finkbeiner, D. P. (2022, November 22). Can denoising diffusion probabilistic models generate realistic astrophysical fields? *Machine Learning and the Physical Sciences Workshop*. NeurIPS 2022. https://doi.org/10.48550/arXiv.2211.12444
14. Müller, V. C. (2020). Ethics of artificial intelligence and robotics.