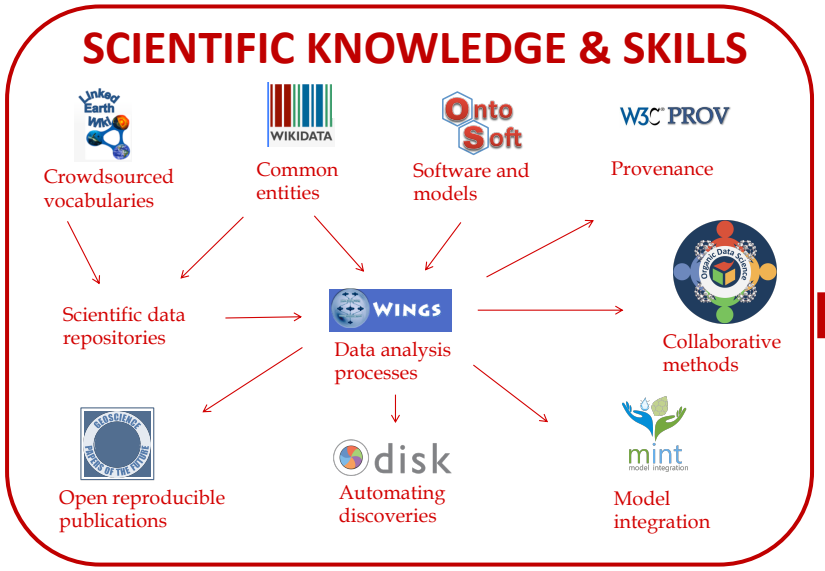
Towards AI Scientists: Critical Partnerships for Future Discoveries

Yolanda Gil
Information Sciences Institute
and Department of Computer Science
Viterbi School of Engineering
University of Southern California

Q52353442

December 1, 2022

Towards AI Scientists



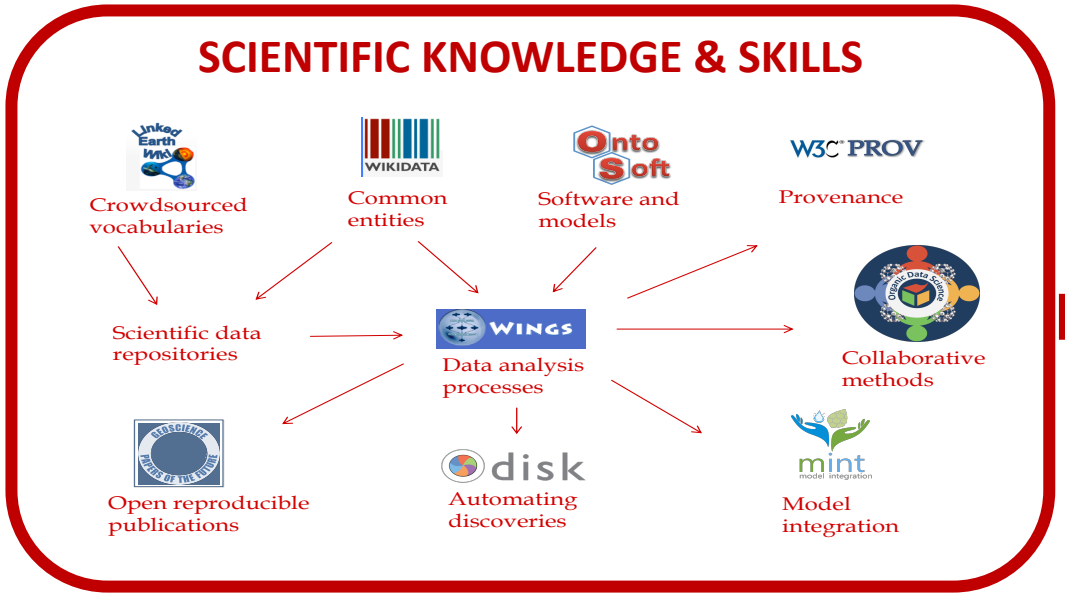
AI reproducing articles → AI as research assistant → AI as co-author

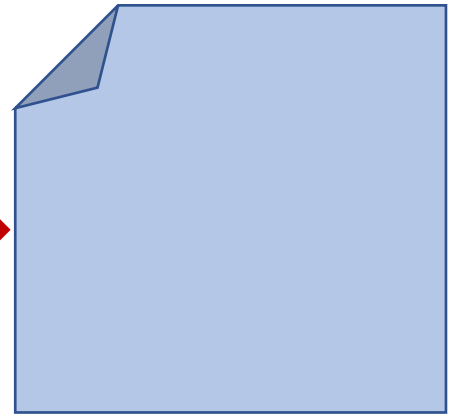
Will AI Write the Scientific Papers of the Future?

What would need to be represented in order for an AI system to automatically generate a paper that provides an accurate report of its analysis and findings?

Benefits:

- Accurate reporting
- Customizable
- Updates & reuse
- What-ifs
- Comparisons
- Creativity





- **Will AI Write Scientific Papers in the Future?.** Gil, Y. *AI Magazine*. 2021.
- **Thoughtful Artificial Intelligence: Forging A New Partnership for Data Science and Scientific Discovery.** Gil, Y. *Data Science*, 1. 2017.

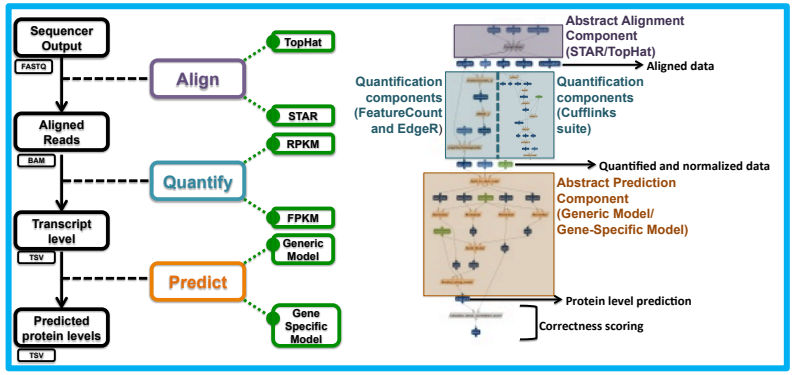
Reproducibility and FAIR Priciples

with Stanford, OSU, OHSU (DARPA, NIH, NSF)

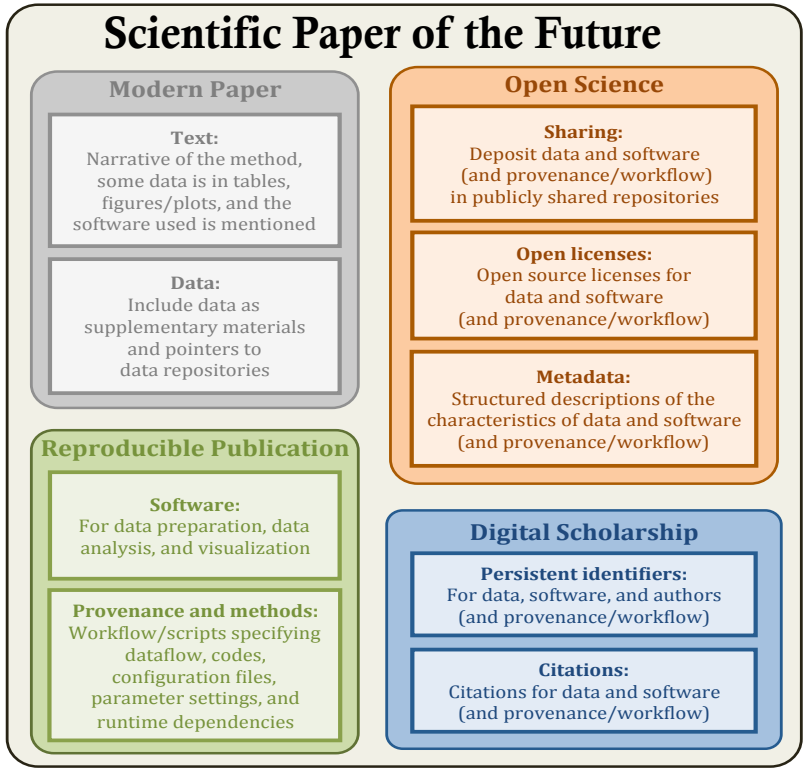
http://www.scientificpaperofthefuture.org

We characterized the variations among top solutions to the challenge, and designed an abstract method that our AI system could elaborate into any solutions and find the best one while explaining its merits over others

DREAM CHALLENGES



AI approach: Document all resources following best practices of reproducible research, open science and FAIR principles, and digital scholarship.

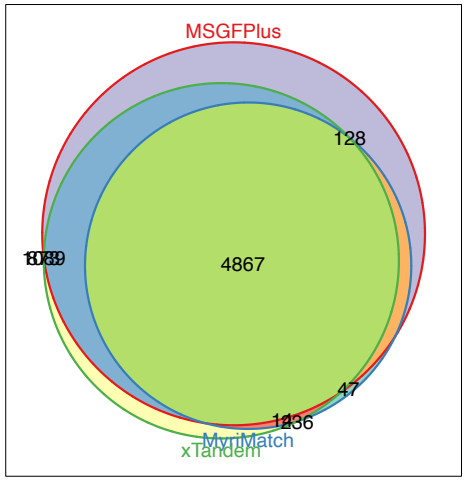


- **Semantic Workflows for Benchmark Challenges: Enhancing Comparability, Reusability and Reproducibility**. Srivastava, A.; Adusumilli, R.; Boyce, H.; Garijo, D.; Ratnakar, V.; Mayani, R.; Yu, T.; Machiraju, R.; Gil, Y.; and Mallick, P. Proceedings of the Pacific Symposium on Biocomputing (PSB), 2019.
- **FAIR Computational Workflows.** Goble, C.; Soiland-Reyes, S.; Garijo, D.; Gil, Y.; Peters, K. *Data Intelligence, Special Issue on FAIR (Findable, Accessible, Interoperable and Reusable) Principles*, 2(1-2). 2020.
- **Use of Semantic Workflows to Enhance Transparency and Reproducibility in Clinical Omics**. Zheng, C. L; Ratnakar, V.; Gil, Y.; and McWeeney, S. K Genome Medicine, 7(73). 2015.

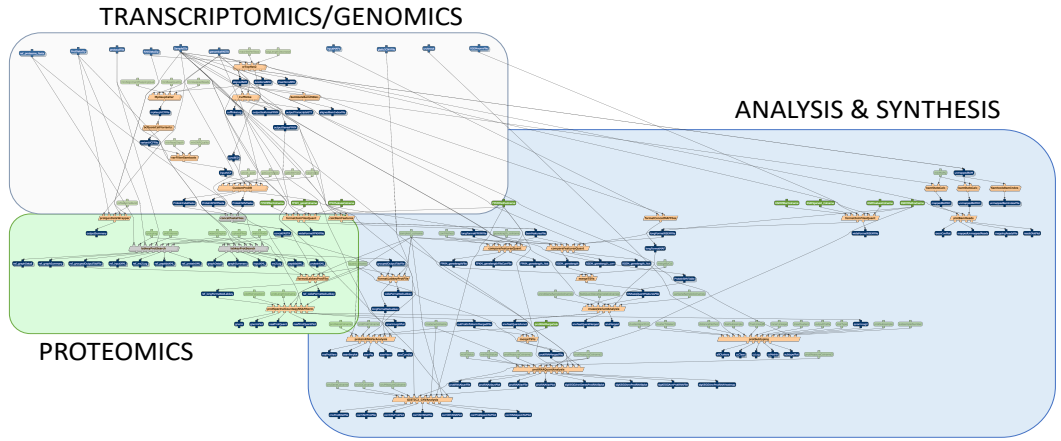
Systematic Continuous Analysis of Data

with Stanford (DARPA, NIH)

We reproduced a seminal cancer study and explored systematic alternative tools/sources, finding that 35% of protein identifications are not robust to changing even just one analysis step. Our AI system can run all methods and do comparisons and ensembles.



AI approach: intelligent workflow system captures general method that can be automatically elaborated into alternative implementations and customized to the data



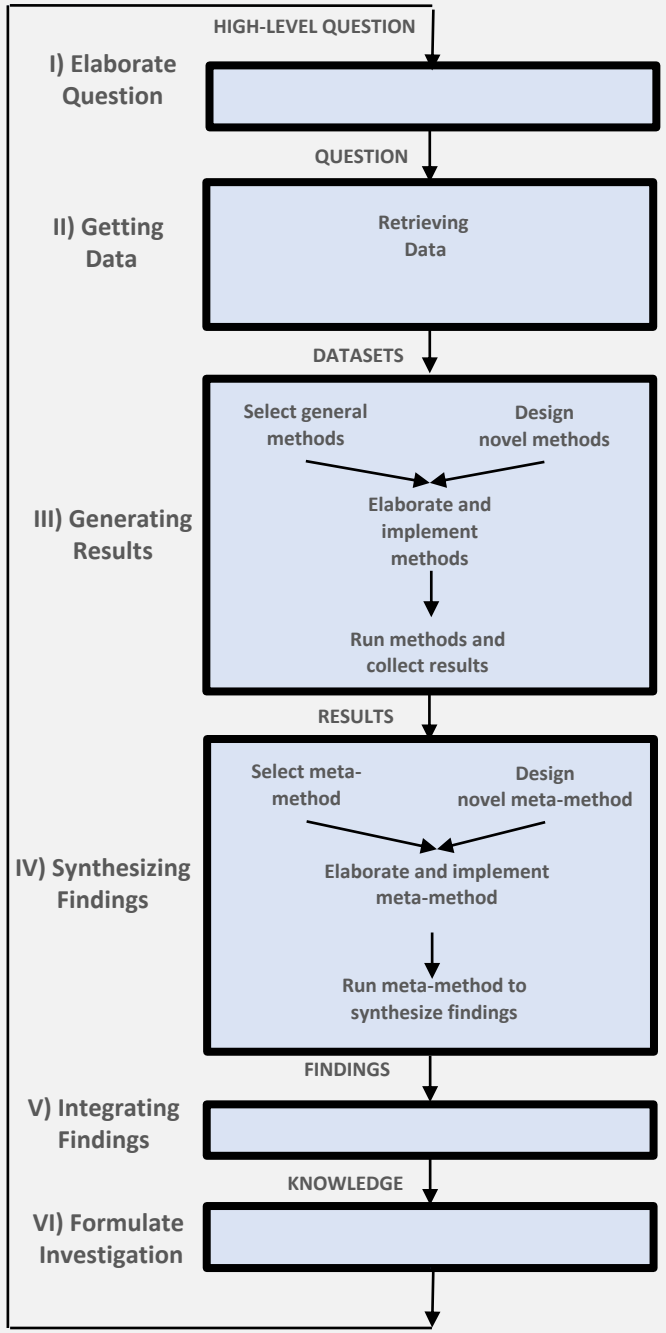
- Cancer multi-omics: systematic, continuous analysis
  - **Towards Continuous Scientific Data Analysis and Hypothesis Evolution**. Gil, Y.; Garijo, D.; Ratnakar, V.; Mayani, R.; Adusumilli, R.; Boyce, H.; Srivastava, A.; and Mallick, P. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, 2017.
  - **Automated Hypothesis Testing with Large Scientific Data Repositories.** Gil, Y.; Garijo, D.; Ratnakar, V.; Mayani, R.; Adusumilli, R.; Boyce, H.; and Mallick, P. In *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems (ACS)*, Evanston, IL, 2016.

Cognitive Architecture for Hypothesis-Driven Discoveries (ONR)

We characterize types of questions about dynamic systems, and develop methods to analyze time series data

Association *Is X associated with Y?*	**Characterization** *Do X, Y, Z conform to a pattern/function?*
Diagnostic *What was observed about X given the last few events?*	**Prefactual** *What outcome if event E occurs now?*
Intervention *What would I observe if I change X?*	**Semifactual** *Same outcome if event E had occurred instead of F?*
Counterfactual *What outcome if event E had not occurred in the past?*	**Prediction** *What future outcome if event E occurs now?*
Backcasting *What event G will have future outcome, what E,F before G?*	**Postdiction** *What event E caused F that caused current outcome?*

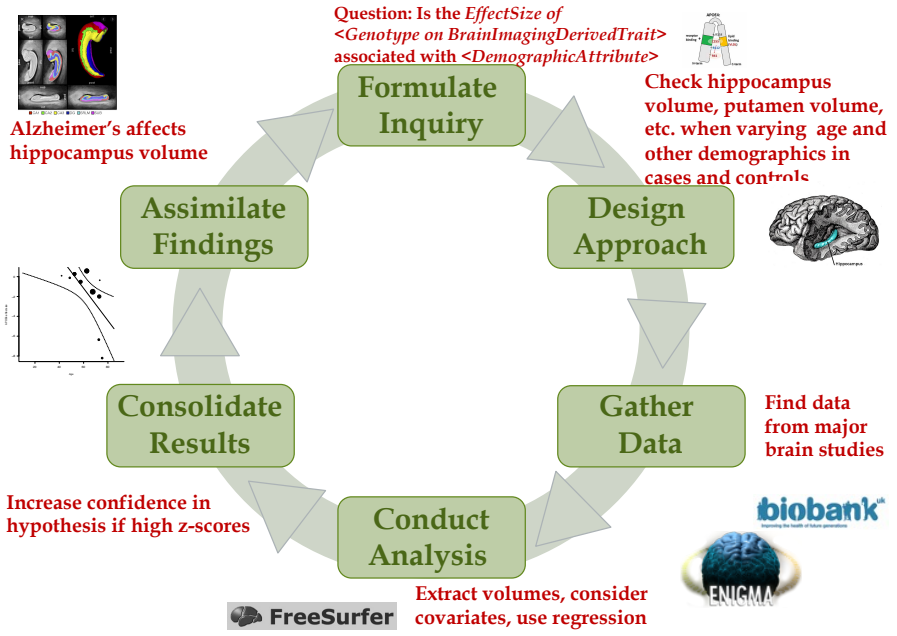
AI approach: Cognitive framework designed to capture how scientists think about questions to set up computational experiments



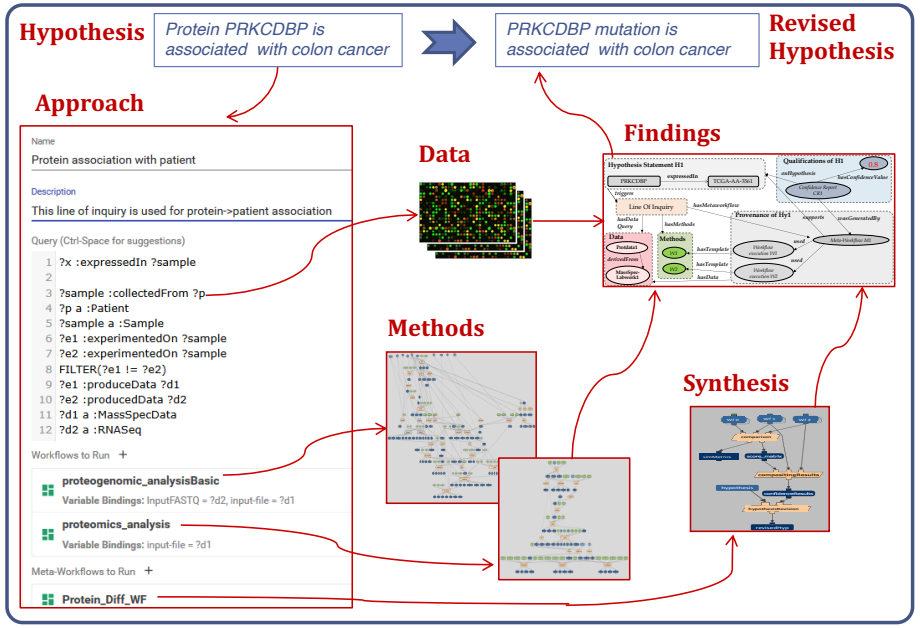
- **Towards Capturing Scientific Reasoning to Automate Data Analysis.** Gil, Y.; Khider, D.; Osorio, M.; Ratnakar, V.; Vargas, H.; Garijo, D.; and Pierce, S. In *Proceedings of the 44th Annual Conference of the Cognitive Science Society (CogSci)*, 2022.
- **Towards Reflection Competencies in Intelligent Systems for Science.** Gil, Y. In "*Artificial Intelligence for Science: A Deep Learning Revolution*," A. Choudhary, G. Fox, T. Hey (Eds). World Scientific, London, UK, 2023.

Accelerating Discoveries through AI Automation with USC LONI/ENIGMA (NIH)

We automate computational experiments, updating findings when new data becomes available in repository

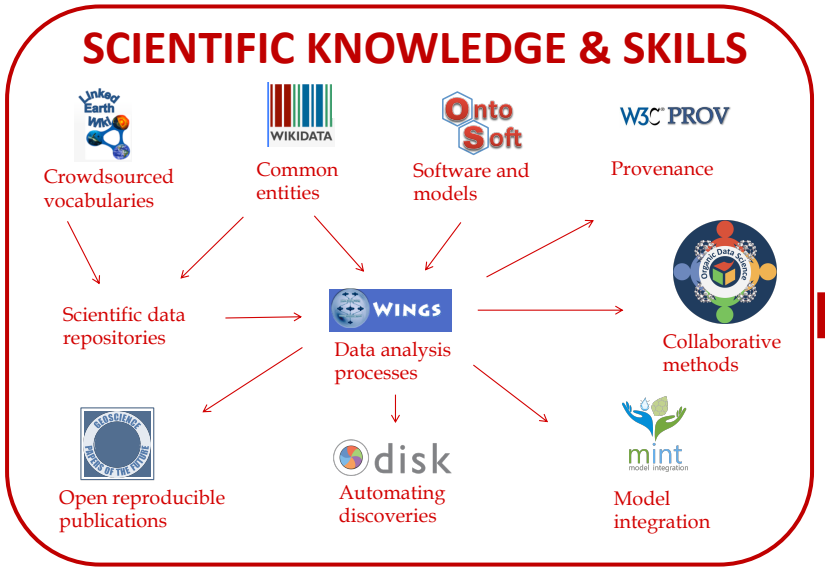


AI approach: For a type of question pattern, develop lines of inquiry that express what data to retrieve, select method based on data available, and analyze findings



- Neuroscience: automating analysis of large data repositories
  - Towards Human-Guided Machine Learning. Gil, Y.; Honaker, J.; Gupta, S.; Ma, Y.; D'Orazio, V.; Garijo, D.; Gadewar, S.; Yang, Q.; and Jahanshad, N. In *Proceedings of the 24th ACM International Conference on Intelligent User Interfaces (IUI)*, Marina del Rey, CA, 2019.
  - Towards Automated Hypothesis Testing in Neuroscience. Garijo, D.; Fakhraei, S.; Ratnakar, V.; Yang, Q.; Endrias, H.; Ma, Y.; Wang, R.; Bornstein, M.; Bright, J.; Gil, Y.; and Jahanshad, N. In *Proceedings of the Fifth Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH), held in conjunction with the 45th International Conference on Very Large Data Bases (VLDB)*, 2019.

Towards AI Scientists



AI reproducing articles → AI as research assistant → AI as co-author

AI Scientists: The Next Two Decades

2025: AI can generate automatically new complex scientific analyses using open data

2025: AI detects when it is missing knowledge and can seek and read new scientific papers on target topics

2030: AI can generate and test sophisticated hypotheses about complex physical phenomena

2030: AI can reproduce the results in 80% of the articles in a scientific journal

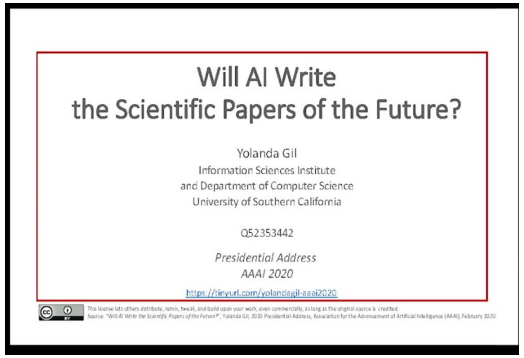
2035: AI can design a scientific experiment and discuss sophisticated aspects of it

2035: AI can compare scientific experiments and papers and contrast their merits

2040: AI can teach advanced theories in some scientific domain effectively to students

2040: AI can formulate research questions and generate novel contributions in some scientific domain

AI Grand Challenge:
AI Scientists that are Partners in Scientific Discovery



AAAI Presidential Address, February 2020
https://vimeo.com/400177695

Article
https://doi.org/10.1609/aimag.v42i4.18149

AI magazine
https://doi.org/10.1609/aimag.v42i4.18149

Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery

Hiroaki Kitano

2025

2050

AI reproduces articles → **AI as research assistant** → **AI as co-author** → **AI as investigator**

2030 2035 2040 2045