

Annex: Current and Future Applications of AI in Science

Task	Example(s)
Technical aids to research	
Literature reviews and knowledge compilation	FutureHouse's paper Q2A generates scientific-literature reviews on demand. ¹
Scoping patent spaces and assisting with grant-proposal applications	HyperWrite's Grant Proposal Generator uses GPT-4 to help turn project ideas and specific goals into detailed grant proposals. ²
Automated coding and code testing	Replit enables scientists to write code automatically, including documentation, and tests for accuracy and safety. ³
Generation of new leads within defined parameters	
Identifying trial populations	National Institutes of Health programme matches potential volunteers to clinical trials. ⁴
Screening tools	2025 NHS trial uses AI to speed up breast-cancer screening in the UK. ⁵
Data management & infrastructure	
Data curation and cleaning	Akkio uses machine learning to spot data-quality issues, perform deduplication and recommend relevant data transformation. ⁶
Research-data management	KNIME Analytics Platform is open-source software that integrates various components for machine learning and data mining through modular data pipelining. ⁷

¹ <https://www.futurehouse.org/research-announcements/wikicrow>

² <https://www.hyperwriteai.com/aitools/grant-proposal-generator>

³ <https://replit.com/>

⁴ https://www.nlm.nih.gov/news/AI_algorithm_volunteers_clinical_trials.html#:~:text=A%20study%20published%20in%20Nature%20Communications%20found%20that,that%20person%20meets%20the%20criteria%20for%20study%20enrollment

⁵ <https://www.bbc.co.uk/news/articles/cly7qx2qx3eo>

⁶ <https://www.akkio.com/>

⁷ <https://www.knime.com/>

Task	Example(s)
Empirical & analytical work	
Optimisation of experimental parameters	CERN's Large Hadron Collider uses machine-learning techniques to optimise particle-collision experiments. AI analyses past data to suggest optimal collision parameters, identify interesting events and enhance simulations. ⁸
Data collection, production and synthesis	AI-driven monitoring in environmental science (for air quality and pollution, for example). AI systems collect, process and analyse data from numerous sensors deployed across environments.
AI-enhanced simulations	Using AI to speed up simulations of novel drugs and materials. ⁹
Embodied physical tasks (robotics, for example)	Lee Cronin's "chemputer" is an automated chemistry lab that converts chemical processes into executable code, automates chemical synthesis and discovery, and enables on-demand production of specialised molecules and materials ¹⁰
Research-ecosystem activities	
Research-funding strategy	(Hypothetical) Impact predictive analysis for funding allocation.
High-level research strategy	(Hypothetical) Alignment of project support with national priorities.
Scientific communication	Using generative AI to simplify the communication of scientific discoveries to the general public. ¹¹
Assembly and curation of large-scale reference databases	Creating the next Protein Data Bank. ¹²
Idea generation & invention	
Data- and literature-driven hypothesis generation	Predictions of multi-factorial correlations/relationships over vast/complex combinatorial spaces – such as relationships between genetic factors, environmental stressors and mental-health outcomes.

⁸ <https://www.sciencedirect.com/science/article/pii/S0168900220310494>

⁹ <https://www.nature.com/articles/s43588-022-00264-7>

¹⁰ <https://medium.com/brain-labs/the-robotic-chemist-5d477af21a96#:~:text=In%202018%2C%20Cronin%20and%20his%20team%20developed%20the,to%20perform%20computer-assisted%20chemical%20synthesis%20with%20high%20precision.>

¹¹ <https://academic.oup.com/pnasnexus/article/3/9/pgae387/7750129?login=false>

¹² <https://www.rcsb.org/>

Task	Example(s)
Theory- and model-guided hypothesis generation	AlphaFold predicts the structure of more than 200 million proteins, ¹³ as well as recommending novel-material designs/systems or alterations to existing ones. ¹⁴
Generation of new mathematical and statistical theories and equations	IBM Watson's AI-Hilbert method integrates background knowledge and experimental data to derive scientific laws expressed as polynomial equalities and inequalities. ¹⁵
Intelligence design of new molecules and materials	MSCA researcher David Baker recently won the Nobel Prize for computational protein design. ¹⁶
AI Scientist	(Hypothetical) An AI system capable of conducting all the steps necessary to complete a given scientific question, including hypothesis generation, conducting experiments, analysing the results and writing up findings.

¹³ <https://www.nature.com/articles/s41592-021-01362-6#:~:text=AlphaFold%20predicts%20the%20structure%20of%20proteins%20with%20novel,prioritizing%20local%20interactions%20over%20recognition%20of%20global%20patterns>

¹⁴ <https://onlinelibrary.wiley.com/doi/full/10.1002/adv.202401401?msocid=1f85e2b0bc5960333d3af61bbd7e6194>

¹⁵ <https://ai-hilbert.github.io/>

¹⁶ <https://marie-sklodowska-curie-actions.ec.europa.eu/news/msca-researchers-2024-nobel-prize>