AI Primer created for Health Research Alliance (HRA) Spring 2025 Meeting
Panel Discussion: The Potential for AI in Discovery Science
Generated by ChatGPT 3/11/2025

Introduction

Artificial intelligence (AI) is reshaping biomedical research by revolutionizing data collection, analysis, and interpretation. From automating complex analyses to generating novel scientific hypotheses, AI-driven tools are driving innovation in discovery science. This primer provides foundational knowledge on key AI-related terms and their relevance to biomedical research funding and grant management.

AI and Machine Learning in Biomedical Research

Artificial Intelligence (AI)

Definition: AI refers to computer systems designed to perform tasks that typically require human intelligence, such as pattern recognition, decision-making, and language processing. It includes machine learning, deep learning, and natural language processing.

Context in Research: AI-powered tools enhance biomedical research by automating image analysis in pathology, optimizing clinical workflows, and personalizing treatment strategies based on patient data. AI-driven algorithms also help in drug discovery by identifying potential therapeutic compounds.

Technical Details: AI systems utilize neural networks, statistical learning methods, and computational frameworks to process complex datasets. Many models improve through iterative learning.

Impact on Discovery Science: AI accelerates hypothesis generation, streamlines experimental workflows, and enables large-scale data analysis and interpretation, fostering new scientific breakthroughs.

Machine Learning (ML)

Definition: A subset of AI, machine learning involves training computer models to recognize patterns in data and make predictions or decisions without explicit programming.

Context in Research: ML models are used in genomics to identify disease-associated genetic variants, in radiology to detect anomalies in medical imaging, and in epidemiology to predict disease outbreaks.

Technical Details: ML techniques include supervised learning (e.g., decision trees,

support vector machines, deep neural networks) and unsupervised learning (e.g., clustering algorithms, principal component analysis) to uncover patterns in data. Training often requires large datasets and substantial computational power.

Impact on Discovery Science: ML enhances research efficiency by automating data analysis and uncovering hidden patterns in complex biological systems, leading to faster scientific discoveries.

Natural Language Processing (NLP)

Definition: NLP is a branch of AI that enables computers to understand, interpret, and generate human language.

Context in Research: NLP is applied in biomedical literature mining, patient record analysis, and automated synthesis of research findings. It can extract meaningful information from vast datasets of published papers and clinical notes.

Technical Details: NLP models use tokenization, named entity recognition, and sentiment analysis. Transformer-based architectures, such as BERT and GPT, improve language comprehension.

Impact on Discovery Science: NLP allows researchers to quickly identify emerging trends, analyze large volumes of text, and enhance knowledge discovery in biomedical research.

Generative AI

Definition: Generative AI refers to models capable of producing new content, such as text, images, or molecular structures, based on learned patterns from training data.

Context in Research: Generative AI is used to design novel proteins, propose potential drug candidates, and generate synthetic datasets for training ML models without real-world data constraints.

Technical Details: These models leverage deep generative frameworks like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) to create new outputs.

Impact on Discovery Science: Generative AI can inspire novel hypotheses, enhance experimental design, and accelerate the creation of new biomedical solutions.

Large Language Models (LLMs)

Definition: LLMs are AI models trained on vast amounts of text data to understand and generate human-like text. While LLMs fall under the broader category of generative AI, not all generative AI models are LLMs—some generate images, music, or other data types instead of text.

Context in Research: LLMs assist in drafting grant proposals, summarizing complex scientific findings, and answering domain-specific queries for researchers and funders.

Technical Details: LLMs analyze large text datasets to learn patterns in language. They rely on deep learning models, particularly transformer architectures, to generate

coherent and contextually relevant text based on input prompts.
Impact on Discovery Science: LLMs streamline literature reviews, enhance data interpretation, and provide researchers with rapid access to synthesized scientific knowledge, improving decision-making in funding and research.

Data Science and Bioinformatics

Data Science

Definition: Data science is an interdisciplinary field that applies statistical and computational techniques to extract insights from large datasets.
Context in Research: It is used to integrate multi-omics data, develop predictive models for disease progression, and optimize research methodologies in clinical trials.
Technical Details: Data science workflows involve data preprocessing, statistical modeling, visualization, and AI/ML techniques for pattern discovery.
Impact on Discovery Science: Advances in data science enable researchers to handle large-scale biomedical datasets more effectively, facilitating new discoveries in systems biology and precision medicine.

Bioinformatics

Definition: Bioinformatics combines biology, computer science, and statistics to analyze and interpret biological data, such as genomic sequences, protein structures, and metabolic pathways.
Context in Research: It is essential for genomics research, drug target discovery, and analyzing high-throughput sequencing data. Bioinformatics supports projects such as the Human Genome Project and cancer mutation profiling.
Technical Details: Bioinformatics relies on sequence alignment algorithms, phylogenetic analysis, and AI-driven functional annotation tools to interpret complex biological datasets.
Impact on Discovery Science: Bioinformatics enables large-scale genomic and proteomic studies, uncovering new therapeutic targets and advancing understanding of complex diseases.

Computational Modeling

Computational Modeling

Definition: Computational modeling involves developing mathematical or algorithmic frameworks to simulate biological processes and predict experimental outcomes.

Context in Research: Models simulate disease mechanisms, predict drug interactions, and optimize clinical treatment plans. They are also used in epidemiology to study the spread of infectious diseases.

Technical Details: Common approaches include differential equation modeling for dynamic systems, agent-based modeling for simulating individual interactions, and molecular docking simulations for drug discovery.

Impact on Discovery Science: Computational models provide a cost-effective and scalable method for testing hypotheses before moving to experimental validation, accelerating the research process and reducing experimental costs.

Conclusion

Understanding AI-related concepts is essential for HRA members involved in funding and reviewing biomedical research proposals. These technologies have the potential to accelerate discovery, improve research efficiency, and open new frontiers in medicine. Familiarity with these terms will help grant managers and program officers make informed funding decisions and support innovative projects at the intersection of AI and biomedical research.