# Critical AI Theories and Methods for Health Services Research: Enhancing Investigative Tools and Strategies

Philip Payne & Aditi Gupta, WashU Medicine, Institute for Informatics, Data Science, and Biostatistics (I2DB)

## Introduction

Health Services Research (HSR) is essential to guide the design, delivery, and evaluation of healthcare services to optimize patient outcomes, cost-efficiency, and access. As healthcare data becomes increasingly complex and abundant, the role of artificial intelligence (AI) has expanded from an auxiliary tool to a crucial element of modern HSR methodologies. To fully leverage the advantages of AI, HSR researchers must acquaint themselves with both the theoretical foundations and practical applications of different AI methodologies. This knowledge is essential for selecting the most appropriate and effective methods for their specific use cases. This post explores the critical AI theories and methods that HSR professionals should understand and potentially adopt or adapt to elevate their research and achieve more meaningful insights.

## 1. Machine Learning (ML)

Machine learning is at the core of AI and is indispensable for processing and analyzing vast datasets in HSR. ML theories focus on building models that learn from existing data to make predictions or classify information and identify novel patterns with minimal human intervention. For example, ML application can support health-related research through enhanced predictive analytics, resource optimization, cost reduction, and clinical decision support. Investigators should be well-versed in different ML techniques, including:

- **Supervised Learning** involves training models on labeled data to predict outcomes or classify data points. This approach is ideal for projects that require predicting patient outcomes, readmission rates, and disease progression based on features derived from electronic health records (EHRs). Examples of supervised learning methods include linear/logistic regression, decision trees, random forests, and support vector machines (SVM). A common example of the use of supervised learning methods is predicting hospital readmission rates based on patient demographics, medical history, and treatment patterns.

- **Unsupervised Learning** helps discover hidden relationships or groups within unlabeled data. This is useful for exploratory data analysis, where large-scale datasets may not have pre-labeled outcomes. Examples of unsupervised learning methods include K-means clustering, hierarchical clustering, principal component analysis (PCA), and anomaly detection. An example of using unsupervised learning methods is identifying patient subgroups with similar characteristics to inform targeted interventions or discovering trends in healthcare service utilization.

- **Reinforcement Learning (RL)** is an emergent methodology but holds significant potential for modeling dynamic decision-making processes. In RL, algorithms learn patterns and make decisions by interacting with an environment and receiving feedback through rewards or penalties. Examples of RL methods include Q-learning, deep Q-networks (DQN), and policy gradient methods. An example of using RL is simulating policy changes and observing the impact on system- or individual-level health outcomes over time, such as optimizing treatment regimens or hospital resource utilization (e.g., staffing schedules) based on demand and patient flow.

## 2. Deep Learning (DL)

Deep learning (DL) is the branch of ML that uses a specific algorithmic structure called neural networks with multiple layers to solve complex problems with high-dimensional datasets. DL is able to detect patterns and relationships within the data sets that may not be ascertainable using surface-level analyses. HSR investigators seeking to use DL methods should seek to understand the architecture and training of these models, especially for projects involving unstructured data such as images or sequential data like clinical notes derived from EHRs. Due to the nature of DL methods, such models' selection, parameterization, and optimization tend to be highly context-specific. Examples of DL methods include:

- **Convolutional Neural Networks (CNNs)** for image recognition and analysis. CNNs process images and extract relevant features, which can then be utilized to categorize the images into different classes.

- **Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM)** networks analyze sequential data with meaningful temporal relationships, such as time-series data like those extracted longitudinally from EHRs.

- **Autoencoders** for data compression and anomaly detection within large and multi-scale data sets. Autoencoders can simplify complex dataset analysis by reducing their dimensionality, making it easier to process high-dimensional medical data such as genomic or imaging data.

These types of DL methods can be used in HSR projects to evaluate medical imaging data for research on diagnostic accuracy, identify patterns in EHRs for predictive analytics, and automate analysis of thematic or semantic content found in narrative text.

## 3. Natural Language Processing (NLP)

HSR investigators often need to extract valuable insights from unstructured data sources, such as clinical notes, patient feedback, and policy documents. Natural Language Processing (NLP) methods can transform these datasets into structured data for subsequent analysis. Most NLP "pipelines" involve a combination of techniques, including:

- **Named Entity Recognition (NER) and Terminology Standardization** are used to identify specific entities (e.g., drug names and conditions) and map to standard medical vocabularies, ontologies, or representations such as the International Classification of Diseases (ICD), Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), or the Unified Medical Language System (UMLS).

- **Topic Modeling** for discovering high-order thematic or semantic themes found in a collection of documents. For instance, topic modeling can process patient feedback collected through surveys and questionaries to uncover common themes related to patient satisfaction, complaints, and suggestions for improvement.

- **Large Language Models (LLMs) and Transformers** for processing and generating text by leveraging mechanisms that enable models to capture contextual relationships between words in a sequence, allowing for efficient processing and language understanding.

NLP methods can be applied in HSR contexts for a variety of purposes, as alluded to earlier, including analyzing patient surveys and comments to identify issues with access to and use of healthcare services, extracting relevant data from clinical notes for outcomes assessment, and interpreting policy documents for insights into healthcare system changes.

## 4. Hybrid AI Models

Hybrid AI models combine various AI techniques to create more robust and adaptable solutions. For instance, combining ML/DL and NLP can allow HSR investigators to build comprehensive models capable of processing textual data and integrating findings with the results of complementary quantitative data analyses. Examples of such hybrid approaches include:

- **Ensemble Learning:** Models such as Extreme Gradient Boosting combines techniques like 'boosting and bagging' that combine multiple models for more robust predictive performance.

- **Multi-modal Models:** Models incorporating different types of data (e.g., structured data, text data, image data) for a holistic approach to data interpretation.

As an example, using hybrid models, HSR investigators can conduct studies that include both patient-reported outcomes and objective clinical data, producing richer insights and more accurate predictions.

## 5. Theoretical Foundations of AI Ethics and Bias Mitigation

In addition to selecting appropriate AI methods, HSR investigators must also be equipped with an understanding of the ethical implications of their use and how to manage potential sources of bias they may introduce. Theories and frameworks related to fairness, accountability, and transparency are crucial in addressing such issues and can include:

- **Algorithmic Fairness:** Techniques such as fairness through unawareness, disparate impact analysis, and rule-based fairness constraints can ensure AI models do not perpetuate existing biases.

- **Explainable AI (XAI):** Methods like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) assist in making AI model decisions transparent and interpretable by humans.

- **Privacy-Preserving Techniques**: Federated learning, data synthesis, and differential privacy controls can enable model training on sensitive healthcare data without compromising patient confidentiality or requiring the transmission or storage of sensitive data beyond originating data stewards.

Ethical frameworks, bias mitigation strategies, and privacy-conserving techniques are critical when conducting HSR, which may ultimately influence healthcare policy or patient care. Ensuring models are trained and validated on diverse and representative datasets helps maintain the integrity of research findings and recommendations and improves the generalizability of results.

## Conclusion

Given the increasing volume and diversity of data available to HSR investigators, using a range of AI theories and methods can help to accelerate and amplify high-impact projects. Proficiency in ML, DL, NLP, the application of ethical AI frameworks, and hybrid modeling can significantly enhance research quality and outcomes. As AI theories and methods evolve, adopting and adapting these theories and methods will empower HSR investigators to generate insights that drive policy changes, improve healthcare delivery, and foster equitable access to care. Further, by integrating these tools thoughtfully and responsibly, the potential for positive change in the healthcare and research ecosystems is significantly increased.