

Transforming Clinical Trials with AI-Driven Protocol Optimization and Next-Gen Statistical Programming

Phuse US Connect 2025

18th March 2025



Ravindra Shukla

Managing Director,
Data Science &
Analytics

- A seasoned IT leader with 20+ years of experience in global delivery in big data and advanced analytics with expertise in mathematical modeling, service-oriented architecture and predictive analysis and data science
- Current areas of work – Generative AI, LLMs, NLP, computer vision, time series and anomaly detection
- Experience with multiple start-ups (advanced analytics, AI)
- Published and presented papers related to data science, advanced analytics, artificial intelligence
- Published author of classic fiction, A Maverick Heart Between Love and Life and member of the film writer's association, and film direction
- Active in building US and India think-tank and various international social organizations
- Enjoys outdoor sports (tennis, cricket, trekking, travelling)



Aman Bahl

Associate Director,
Statistical
Programming

- 20+ years of clinical industry experience in the areas of leadership, team management, statistical programming, and other statistical services
- Focused on providing innovative solutions and services based on customer needs
- Enables agility and drives change within teams to constantly rethink, revive, respond, and reinvent based on customer needs
- Dedicated to implementing cutting-edge methodologies and technologies to optimize data analysis and drive business growth
- Published and presented over 20 papers related to healthcare, trends and technology, AI/ML, SAS programming, CDISC standards, and R programming at various conferences
- Enjoys outdoor sports, trekking, travelling and volunteering

AGENDA

- 01 Introduction and Current Challenges in programming
- 02 Enhancing Statistical Programming and AI Applications
- 03 AI-Driven Innovations and Use Cases in Clinical Trials
- 04 Technology and Future Pathways
- 05 Results Evaluation and Conclusion

Current Challenges in Statistical Programming



Technical expertise bottlenecks

Lack of experienced statisticians and data analysts



Repetitive tasks

Manual coding and reporting are slow and error-prone



Delays in data processing

Inefficient data processes delay clinical research



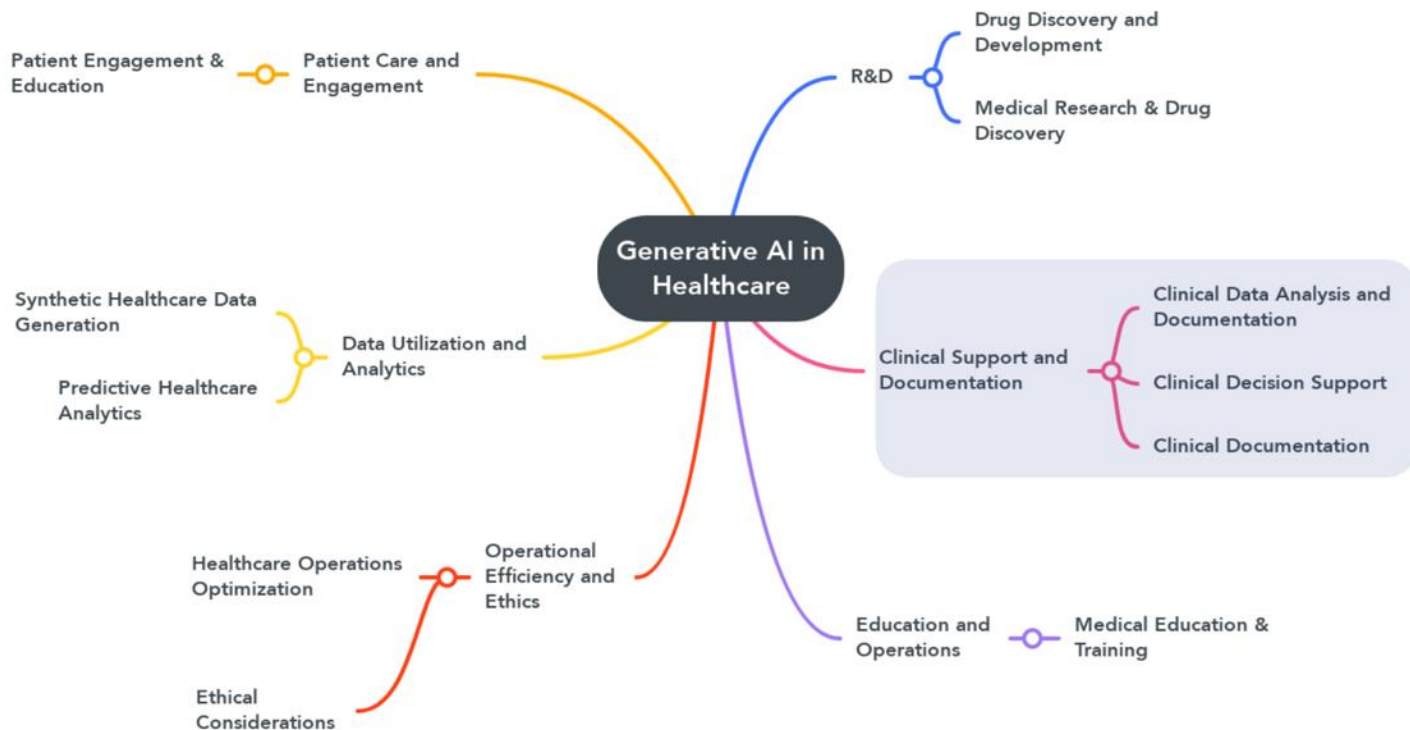
Error-prone processes

Manual coding causes analysis inconsistencies and inaccuracies.

These challenges hinder the efficiency, reliability, and reproducibility of statistical programming in clinical trials

Generative AI in Healthcare

Source: <https://gradientflow.com/generative-ai-in-healthcare/>



Enhancing Statistical Programming and AI Applications



Automated code generation

Generative AI models can automatically generate statistical code based on researcher inputs like data structure and analysis goals, speeding up programming tasks.



Intelligent data preparation

Generative AI can handle data wrangling tasks like imputing missing values, identifying outliers, and standardizing variables, reducing time spent on data preprocessing.



Optimized and reproducible analysis

AI-Generated code can be optimized for efficiency and readability, leading to faster analyses that are more reproducible for future researchers.

By automating repetitive tasks and generating optimized code, generative ai has the potential to significantly enhance the speed, reliability, and accessibility of statistical programming.

Automated Code Generation

Analyze task requirements

The ai system processes user input on analysis type, data structure, and sample input/output to understand the task.

Generate optimized code

The AI generates optimized statistical code based on task requirements.

Accelerate development

AI-Generated code reduces development time, enabling focus on problem-solving and innovation

Enhance debugging and optimization

The AI analyzes the code to detect errors, suggest fixes, optimize performance, and recommend refactoring for a high-quality program.

Intelligent Data Preparation



**Automated Data
Cleaning**



**Time
Savings**



**Improved Data
Quality**



**Anomaly
Detection**



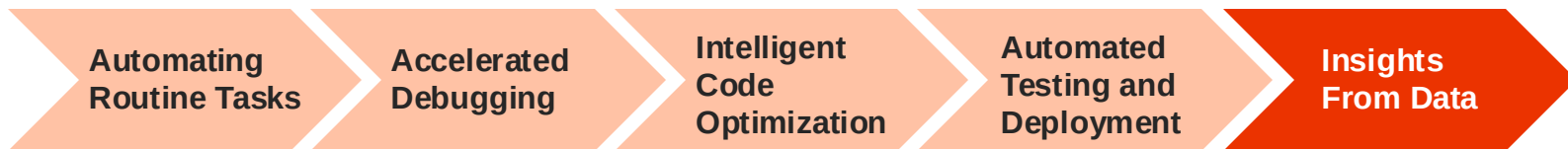
Explainability

Optimized and Reproducible Analysis

Increased Efficiency and Reproducibility

- Time Savings
- Resource Efficiency
- Reduced Syntax Errors
- Improved Reproducibility

Enhanced Productivity



Reduced Errors

Reduction in Common Coding Errors Using AI-Generated Code in Statistical Programming

SYNTAX ERRORS

80-90%

LOGICAL ERRORS

20-40%

MISSING DATA ISSUES

30-50%

INCONSISTENCIES

40-60%

Examples of Common Programming Tasks and ChatGPT Prompts

Code Commenting Made Effortless

Can I send you a SAS program to comment?

Whitespace Enhancement for Readability

Can I send you a SAS program to improve the whitespace?

Cross-Language Code Conversion

Can you help me convert SAS code into R code?

Efficiency Boost Through Code Optimization

Can I send you a SAS program to optimize for efficiency?

Readability Enhancement for Better Understanding

Can I send you SAS code to optimize for readability?

Code Debugging

Can you help me debug my SAS code?

Dynamic Coding Solutions via Conversations

In SAS, I'm trying to concatenate multiple variables in a dataset, separated by a space, into a new variable. The variables all start with "cat," and there are 50 of them.

Dataset Creation Endeavors

Generate a SAS script to create a dataset named `dm` using SAS datasets (`dm1001`, `ds1001`, `ex1001`) located in `'/study_location/xxx'`. Here are the specifications: `xxx`

Learning Aid for Language Proficiency

Can I send you SAS code and have you explained in plain language what it is doing?

Creating Courses

Can you create a beginner's course to transition a SAS programmer to R for clinical programming?

AI Coding Tools and Code Generators

Key Features and Capabilities

- Predict code completions and generate new snippets or entire programs.
- Provide real-time solutions to coding problems.
- Create summaries and comments for better code understanding.
- Optimize, refactor, and translate code across languages.
- Enhance productivity through error detection and security insights

Popular Tools and Generators

- **GitHub Copilot:** Context-aware suggestions for faster coding.
- **Tabnine:** Supports multiple languages with autocompletion and IDE integration.
- **Amazon CodeWhisperer (Q Developer):** Speeds up coding with intelligent suggestions and security features.
- **Replit AI:** Real-time assistance for writing, testing, and debugging.
- **Sourcegraph Cody:** AI-powered search for relevant code examples and context.
- **AskCodi:** Offers documentation and problem-solving suggestions.
- **Cursor AI:** Advanced code completions, refactoring, and editing capabilities.
- **SQLAI:** Specialized in SQL query generation and optimization.
- **Codeium AI:** Autocomplete and refactoring with multi-language support.
- **Snyk (DeepCode AI):** Real-time code quality improvements and security scanning.

Next-Gen statistical programming tools



Emerging
Technologies and
AI Applications



Data Management
and Preparation



Statistical Analysis
and Visualization



Workflow
Optimization and
Collaboration



Regulatory
Submissions and
Compliance



Advanced
Security and
Transparency



Open-Source and
Framework Adoption

AI-Driven innovations in clinical trials



Accelerating Trial Design – Protocol Optimization



AI-Driven Recruitment, Monitoring, and Data Optimization



Synthetic Control Arms



Biomarker Discovery and Precision Medicine



Integration of Real-World Evidence (RWE)



Predictive Modeling and Analytics for Trial Success

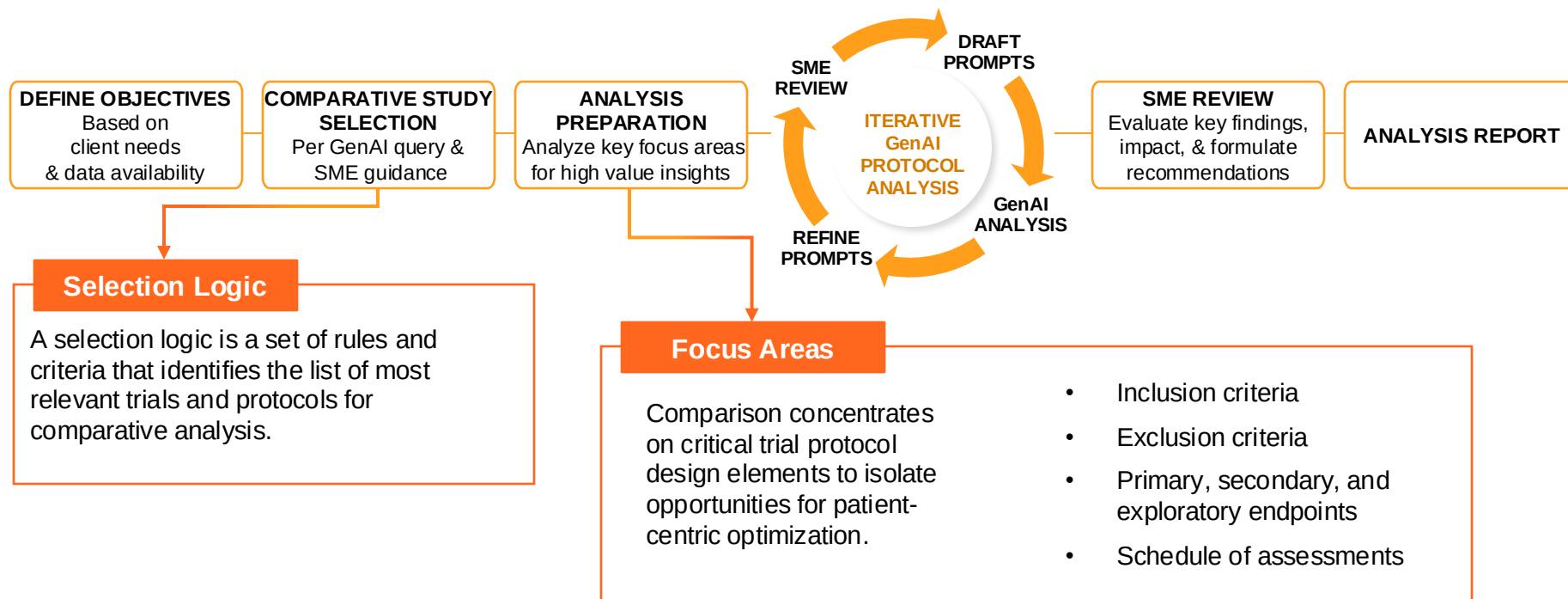
Few Other Key Challenges in Life Science Industry

- High Failure Rate
- Low Patient Enrollment Rates
- Patient Retention
- Regulatory and Ethical Compliance
- Data Quality and Integrity
- Resource Constraints
- Protocol Design and Optimization
- Safety Monitoring and Pharmacovigilance
- Data Management and Integration
- Geographic and Cultural Diversity

Overcoming these challenges demands innovation, collaboration, and AI-driven solutions to optimize trials and enhance safety and efficiency.

Use Case-1: **Trial Optimization**

GenAI analytics to optimize trial selection and protocol analysis



Comparator Protocols: Design Analysis & Performance Correlation

Design Parameters



**Inclusion /
Exclusion Criteria**



**Schedule of
Assessments**



**Primary, Secondary &
Exploratory endpoints**

Key KPIs (performance measures)



**Screen
Failure Rate**



Enrollment Rate



**Discontinuation
Rate**



**Protocol
Amendments**



Time to Approvals



Patient Burden



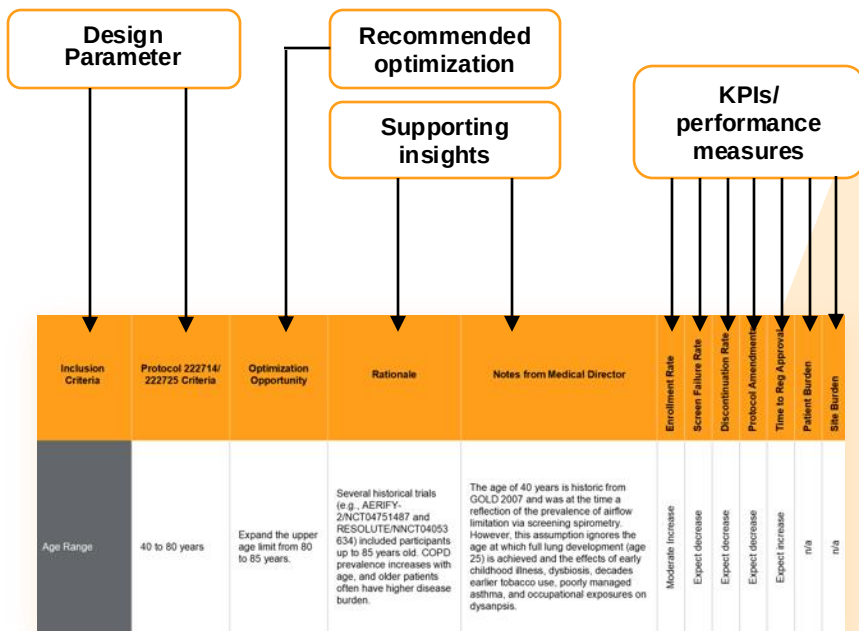
Patient Engagement



Site Burden

Study Design: Comparative Analysis

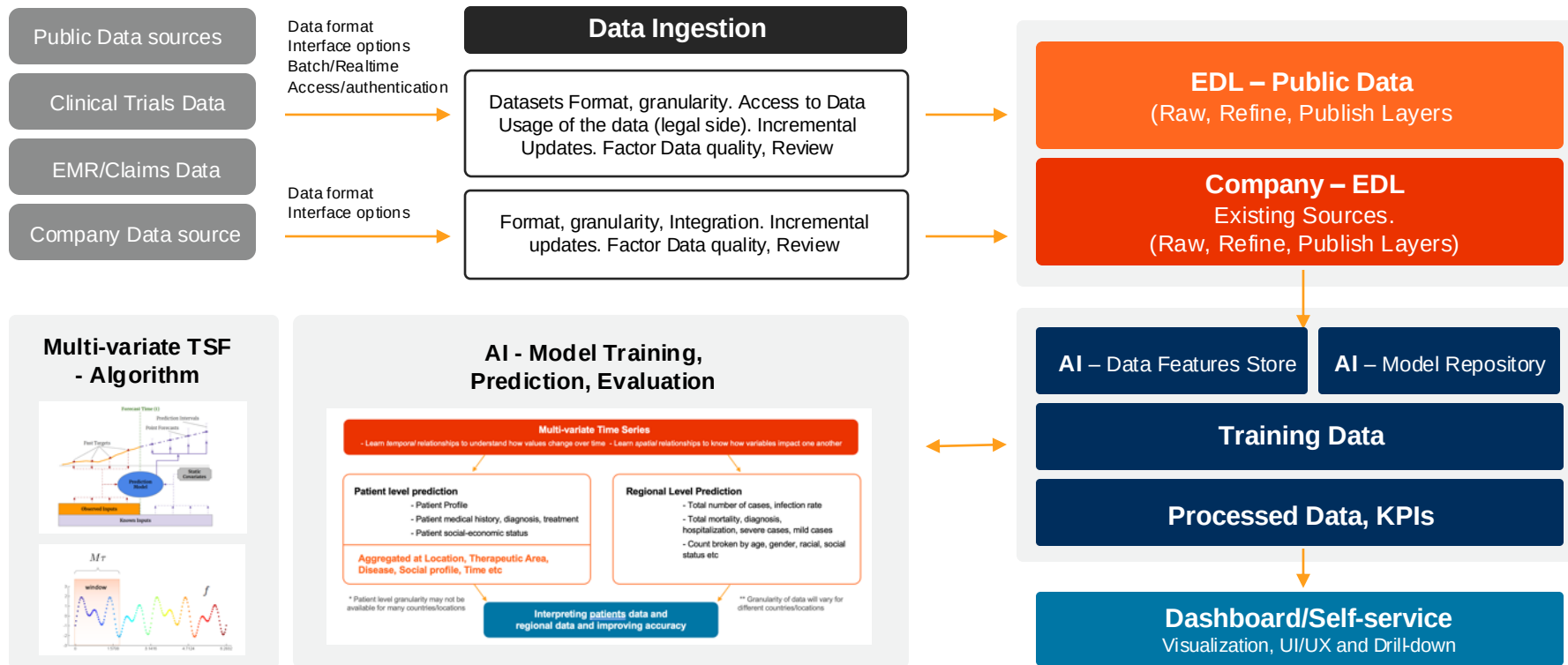
The analysis in this section includes:



	How to interpret KPIs/ Performance Measures
Enrollment Rates	Impact described as • Significant increase (10%+) • Moderate increase (5-10%) • Increase (up to 5%) Calculated by AI tool
Screen Failure Rates	Impact for these measures is estimated by SME. Recommendations that improve enrollment rates are expected to indirectly decrease <i>Screen Failure Rates</i> and <i>Discontinuation Rates</i> . Reducing the complexities driving Patient and Site Burdens are expected to decrease the <i>Number of Protocol Amendments</i> and <i>Time to Regulatory Approval</i> .
Discontinuation Rates	
Number of Protocol Amendments	
Time to Regulatory Approval	Note that there is limited published information on these measures.
Patient Burden	Impact described as • Significant decrease (10%+) • Moderate decrease (5-10%) • Decrease (up to 5%) Calculated by AI tool
Site Burden	

Use Case-2:
Disease Spread Forecasting

Data flow and Prediction Approach



AI based - Prediction Approach - Multivariate Time Series Forecasting

Multi-variate Time Series

- Learn *temporal* relationships to understand how values change over time - Learn *spatial* relationships to know how variables impact one another

Patient level prediction

- Patient Profile
- Patient medical history, diagnosis, treatment
- Patient social-economic status

Aggregated at Location, Therapeutic Area, Disease, Social profile, Time etc

* Patient level granularity may not be available for many countries/locations

Regional Level Prediction

- Total number of cases, infection rate
- Total mortality, diagnosis, hospitalization, severe cases, mild cases
- Count broken by age, gender, racial, social status etc

** Granularity of data will vary for different countries/locations

Interpreting patients data and regional data and improving accuracy

AI/ML approach differentiators

Customized AI Model Training

- Execute combination of training data, algorithms, parameters and evaluate/refine model behavior (including reinforced learning)
- Ability to customized model training based on client specific data
- Augment existing datasets with automated labeling
- Ability to generate synthetic data based on smaller set of data for training purpose

Latest research work factored in AI model development

- Explore open source, AWS, Azure, Google, Open-AI, FB, Stanford latest research and incorporate in data pipe and model development

Innovative, Iterative Approach

- Building Block approach to a digital platform development
- Scalable cloud-based processing pipelines integrated with existing data stores
- Enable machine learning methods while adhering to best practices
- Improve the current process at scale
- Utilize empirical evidence to identify novel patterns for clinicians and analysts

Better Operational Model

- Docker image-based installation to move across different environment and flexible implementation
- Custom scientists work bench
- Factor dynamic learning based on production in model training

Agentic Framework VS RAG Approach

Agentic:

- Tasks/Goals-driven, Agent action-oriented, and Autonomous.
- Best for workflows requiring decision-making and multi-step execution.
- High level instruction
- Users Query broken to multiple steps
 - Mimic the workflow
- Access to Knowledge Base
- Call APIs or various tasks

RAG:

Insight/Data -driven, Fact-grounded, and responsive (Prompt/Users input).

- Best for knowledge retrieval and factual generation
- Searching
- Ranking
- Content Curation
- Open for integrating various key tasks to build application

Future Pathways for Next-Generation Statistical Programming

- Explainable AI (XAI)
- Domain-Specific AI models
- Privacy-Preserving AI
- AI in Decentralized Trials
- Next-Generation Statistical Programming Tools
- AI-Powered Regulatory Submissions
- Human-in-the-Loop Systems



Conclusion

We are at the forefront of a transformative era in statistical programming, driven by AI advancements that are reshaping clinical trials. The challenge lies in balancing AI's precision with human creativity to foster innovation, boost productivity, and deliver greater value. This synergy will enhance trial reliability and enable groundbreaking solutions that redefine digital possibilities.

Generative AI is reducing technical barriers, shifting the focus toward analytical thinking, problem-solving, and design thinking. As tools evolve, their potential for automation, hypothesis generation, and data analysis will revolutionize trial design and analysis.

By addressing inefficiencies, improving data handling, and enhancing compliance, AI integration is accelerating the delivery of life-changing treatments, advancing pharmaceutical research, and improving global health outcomes.

Ravindra Shukla

Managing Director, Data Science & Analytics

Phone: +1-703-679-0117

Email: Ravindra.Shukla@syneoshealth.com

Aman Bahl

Associate Director, Statistical Programming

Phone: +1-905 399 6715

Email: Aman.bahl@syneoshealth.com

Shortening the Distance from Lab to Life®.

