

AutoProgramming: Agents Are All You Need

Chengxin Li

May 2025

AutoCheng Clinical Data Services LLC

Agenda



Introductions to AI/ML



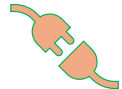
Introductions to Statistical Programming



AI/ML Application Challenges



Solution—All in Agents



Emerging Skills



Discussions

Introduction to AI/ML

- **Revolutionizing Impact:** AI, specifically Deep Learning (DL), is transforming our world, similar to the Internet in the 1990s
- **LLM (Large Language Model):** Pretrained with vast amounts of data (in TBs) and numerous parameters (in billions), these models predict the next token and feature reasoning capabilities
- **AI Agent (or Agentic AI):** Utilize machine learning to automate workflows, planning, reasoning, and acting; freeing professionals from repetitive tasks. 2025 is dubbed the "AI Agent Year"
- **AI Tools:** Lots of LLM Tools available (Gemini, Copilot, ChatGPT, Claude, DeepSeek...) and in affordable services. It is important to clarify that users cannot train these models directly, rather the developers improve the tools based on feedback

Introduction to Statistical Programming

- **Role in Pharma/Biotech:** Supports and implements statistic analyses for clinical studies, involving biggest resources among all functional lines
- **Programming:** All similar across companies, repetitive tasks
 - ✓ Regulated with data integrity and incorporated with domain knowledges (e.g., in CDISC standards and statistic principles)
 - ✓ Well-established process (SOP or WI) with high quality required
 - ✓ Programming scopes not only in coding but lots of documentation as well
 - Specifications (SDTM define-xml, ADaM define-xml, ARM define-xml)
 - Review Guides (cSDRG, ADRG)
 - Annotations (aCRF, Mock Table annotations with ARM)
 - ✓ Documentation is a common pain point
- **Tool:** SAS is dominant, R catching up, and Python in its infancy
- **Automation:** To improve quality and efficiency, programming automations have been explored, and AI/ML potentially plays vital role

Challenges on AI/ML Applications

➤ **Security Concern**

- ✓ Currently the concern is like the data transition to the CLOUD several years ago
- ✓ Gemini publicly embedded in Google Chrome, and all tools also Web accessible
- ✓ Copilot embedded in MS Office365 account

Comment: all hugging AI, security concerns would be addressed quickly

➤ **Business Model**

- ✓ All programming similar across companies, however, those programming codes incorporated with domain knowledge are confidential, company owned IP, not publicly shared as always. Correspondingly:
- ✓ Those AI tools would not have sufficient SAS codes to get trained. Consequently:
- ✓ Not surprisingly, AI generated code is helpful but not production-ready

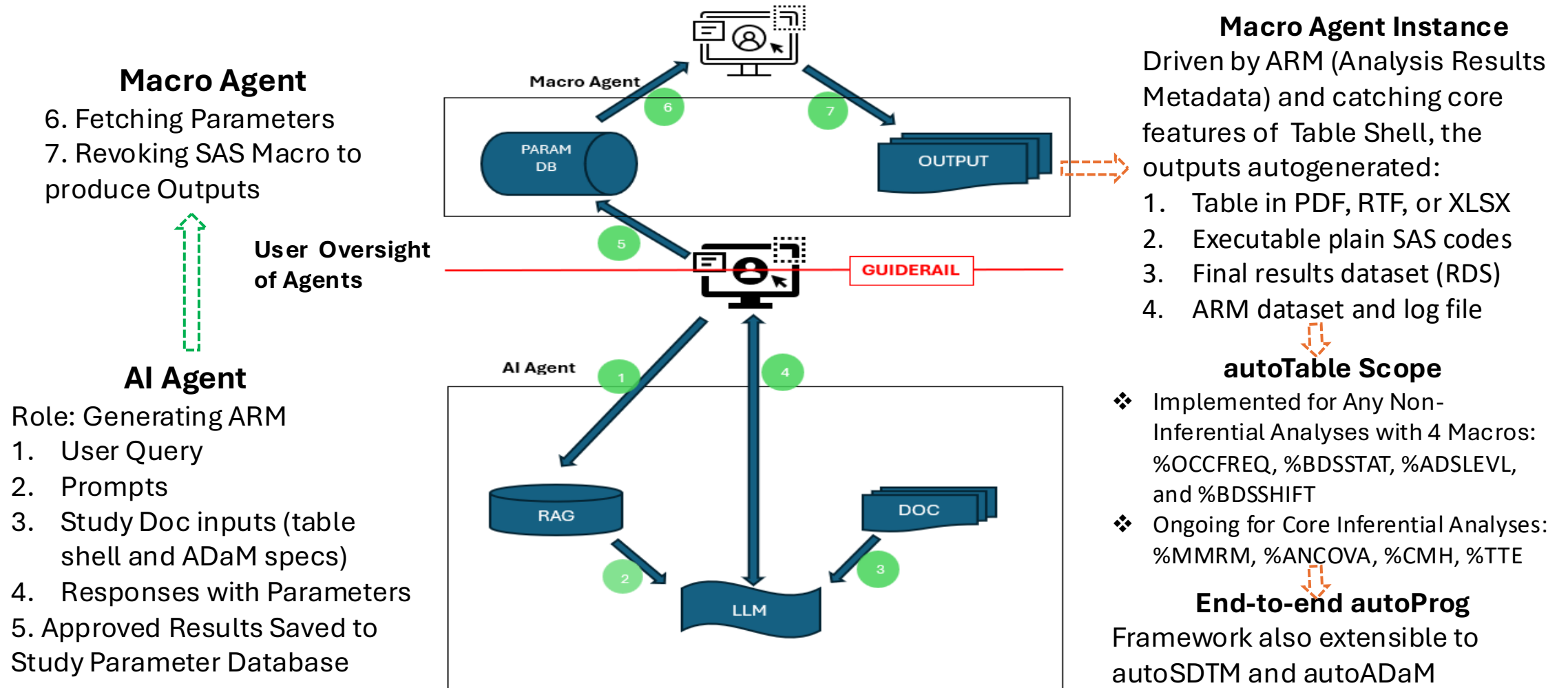
Comment: practical improvements would unlikely occur in the coming few years

Solution—All in Agents

- Although LLM generated code currently has issues with reliability, using AI Agents to perform text-based documentation work provides a big relief to programmers
- Using developed autoProgramming tools such as Coding Agent (Macro Agent) allows for previous developments to remain usable remaining the goal of high quality and high efficiency
- Two types of Agents Applied
 - ✓ AI Agent: LLM Based
 - Generating programming documentations (Specifications, Review Guides, and Annotations)
 - Replacing manual work, generating macro parameters to facilitate autoprogramming
 - ✓ Macro Agent: SAS Macros based to automate programming. For Instance,
 - autoSDTM driven by SDTMIG, autoADaM driven by SAP, autoTLF driven by ARM
- Achieving end-to-end autoprogramming by introducing multi-agent system

Solution—All in Agents

Framework of autoTable Agents



New Skills

➤ **AI/ML Fundamentals**

- ✓ Understanding how AI/ML/DL concepts work
- ✓ Understanding AI Agent and related concepts (RAG, MCP/A2A, Memory, CoT, LangChain, ...)
- ✓ Utilizing AI Agent Development Kits

Some AI Agent Development Kits require little or even zero Python programming skills, just using natural language is enough (e.g., Copilot Agent Studio)

➤ **Prompt Engineering**

- ✓ Developing new communication skills – Prompt Engineering, to improve AI Agent performance in high quality outcomes reliable to production
- ✓ With Prompt Engineering, defining clear roles, specific instructions step by step, RAG access, and how the desired outputs look like with examples

➤ **Process Integration**

- ✓ AI Agent can work within the existing infrastructure and business model without requiring replacements or transformations, thus:
- ✓ Transitioning routine programming practices (usually simple and straightforward workflows) to an AI Agent, letting the AI Agent plan, think, reason, make decisions and act
- ✓ To relieve security concerns, the inputs to external AI tools limited to study documents and metadata, NOT study data

Discussions

➤ **Future Work Direction**

- ✓ AI Agent would be a game changer to statistical programming
- ✓ Reevaluating current working directions including costing digitalization
- ✓ Reevaluating reporting system development framework by introducing multi-agent system
- ✓ Converting current 'LLM+SAS macro' to 'LLM+R/Python package' when R/Python grows up
- ✓ Upgrading Macro Agent to AI based tool once AI tools are capable of producing reliable programs in productions

➤ **Roles of Programmer**

- ✓ High qualified deliverables are the harmonious combination of people + technology (tool) + process. This also remains in the age of AI with programmer as the key player
- ✓ Programmers, freed from routine tedious work with AI Agent, can switch focus to value adding work like concepts, processes, solutions, and managing/overseeing Agents

➤ **Solution Adoption**

- ✓ Keep the solution simple, adopted only if it helps production deliverables meaningfully



Questions

Chengxin Li
AutoCheng Clinical Data Services
Chengxin.Li@autoclindata.com

