phuse·share
connect·share·advance

Single Day Event

# Accelerating Clinical Data Management, Analysis, and Reporting with Gen AI-powered Code Assistants

Mohit Sehgal, PhD
Hexaware Technologies

phuse.global

# Non-Disclosure Agreement

HEXAWARE

During this presentation, I may disclose confidential information including client names. This information must not be disclosed to third parties in any manner without prior permission from Hexaware Technologies.

For questions or more information about this presentation, please contact mohits12@hexaware.com

# About me

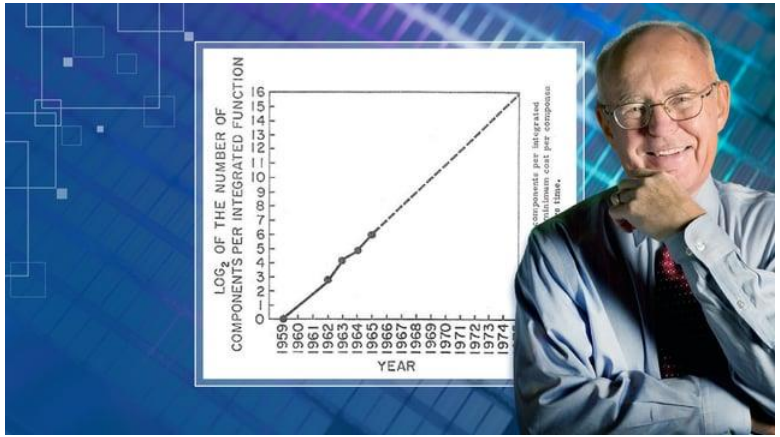
- **PhD in Cell & Molecular Biology (Drexel University, Philadelphia, USA)**
  Studied the role of Dendritic cells in HIV-1 and HCV infection
- **Postdoctoral Research Scientist (Drexel University College of Medicine, Philadelphia, USA)**
  Studied the role of innate immune system in Cancer
- **Postdoctoral Research Scientist (The Wistar Institute, Philadelphia, USA)**
  Studied Myeloid-Derived Suppressor cells and their role in cancer
- **Senior Research Scientist (Cell & Gene Therapy Division, Intas Pharmaceuticals, Ahmedabad, India)**
  CAR-T cell therapy against hematological malignancies and solid tumors
  Hematopoietic Stem Cell Gene-Addition/Editing therapies for β-Thalassemia and Sickle Cell Disease
- **Cell & Gene Therapy SME (Cognizant Technology Solutions)**
- **Cell, Gene & Other Advanced Therapies Practice Lead (Hexaware Technologies)**

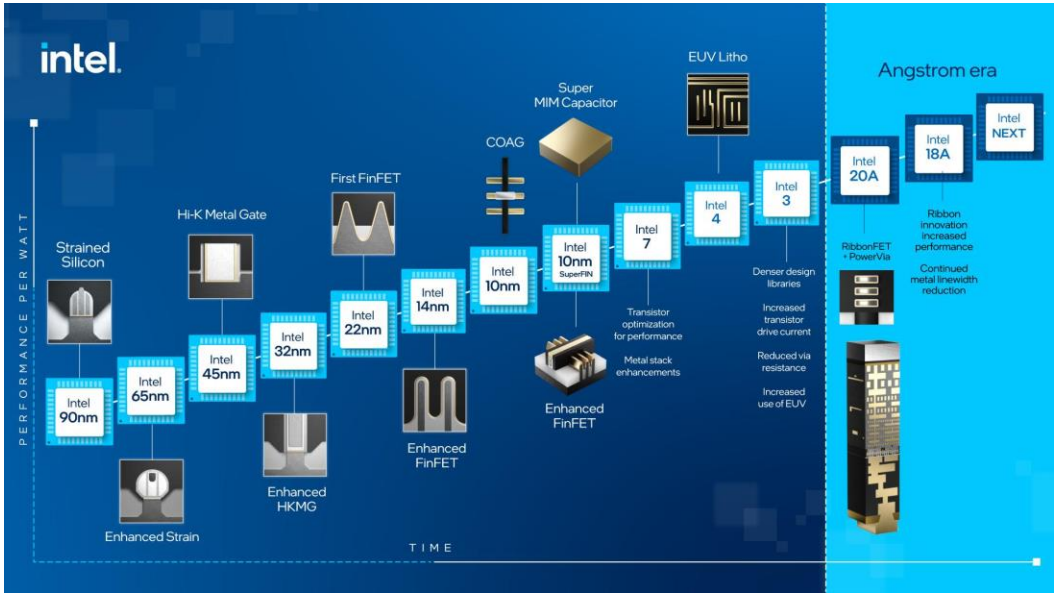
# Agenda

- Rising R&D Costs
- Growing complexity of Clinical Trials
- GenAI-powered code assistants
- Leveraging GenAI-powered code assistants for Clinical Programming
- Q&A

# Moore's Law

In 1975, Intel co-founder Gordon Moore gave Moore's law → Number of transistors on a chip will **double every two years** with a minimal rise in cost

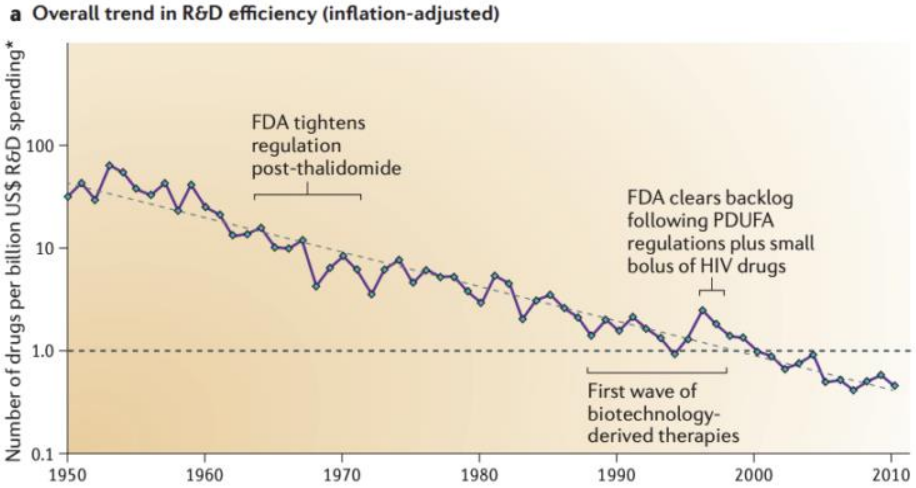



**Source:** Intel Website

# Eroom's Law (Moore spelled backwards)

In 2012, Jack Scannell published a landmark study that highlighted the steady decline in R&D efficiency. He gave **Eroom's law which states that the inflation-adjusted cost of developing a new drug roughly doubles every nine years**




**Source:** Scannell et al; Nature Reviews Drug Discovery (2012)

**According to a 2021 study by Tufts Center for the Study of Drug Development, the average cost of developing a prescription drug is $2.6 billion**

# Clinical Trials account for major portion of the R&D spending

(dollar figures in millions)

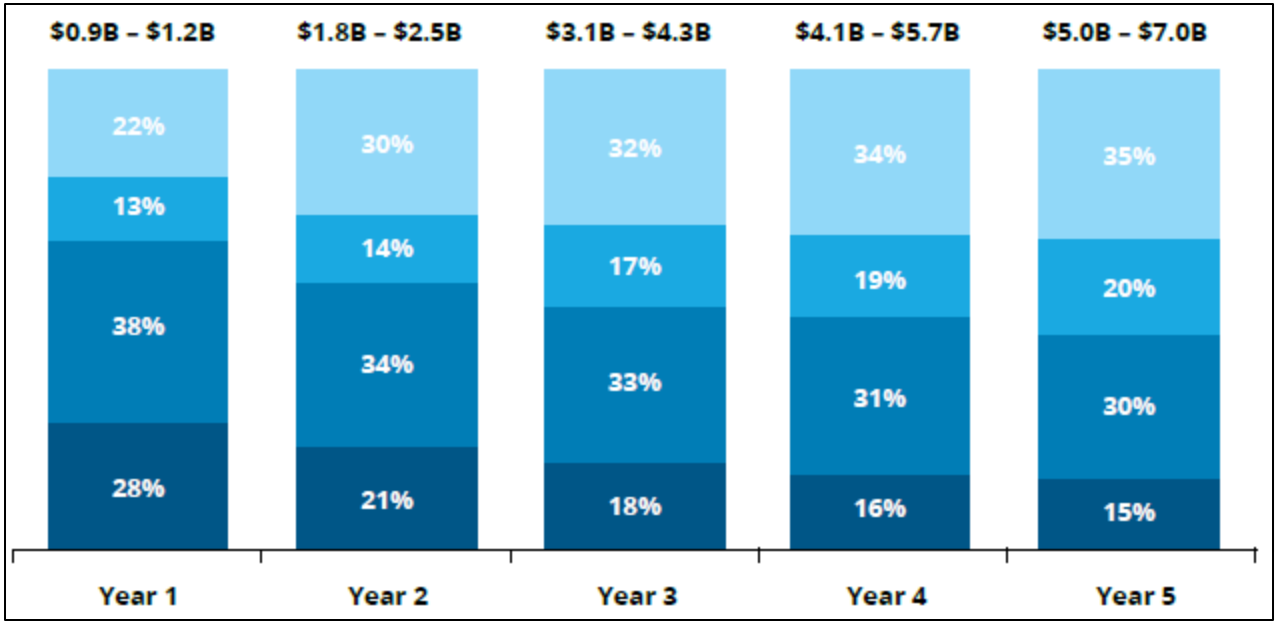
| Function | Dollars | Share |
|---|---|---|
| Pre-Human/Pre-Clinical | $16,030.1 | 15.9% |
| Phase I | $8,531.5 | 8.5% |
| Phase II | $11,197.7 | 11.1% |
| Phase III | $29,050.3 | 28.8% |
| Approval | $4,290.9 | 4.3% |
| Phase IV | $11,597.6 | 11.5% |
| Uncategorized | $20,147.0 | 20.0% |
| TOTAL R&D | $100,845.2 | 100.0% |

**Source:** Pharmaceutical Research and Manufacturers of America, PhRMA Annual Membership Survey, 2023

# AI/GenAI can potentially unlock billions of dollars in value for the pharmaceutical industry

A top 10 pharma company with ~$60 billion in revenue could capture between US$5.7 billion of peak value by scaling the use of AI over 5 years

Generative AI is expected to produce $60 billion to $110 billion annual value across the pharmaceutical industry value chain




**Source:** Deloitte Analysis

# Increasing complexity of clinical trials

## Trial Design Complexity

- Complex Innovative Trial Designs such as Adaptive Protocols, Master Protocols
- Increase in Procedures
- Increase in Endpoints

## Data Complexity

- Variety
- Volume
- Velocity

## Operational Complexity

- Managing Global Studies
- Decentralized Clinical Trials

Manual Coding is:

- Expensive
- Labor intensive
- Time-consuming
- Error-prone

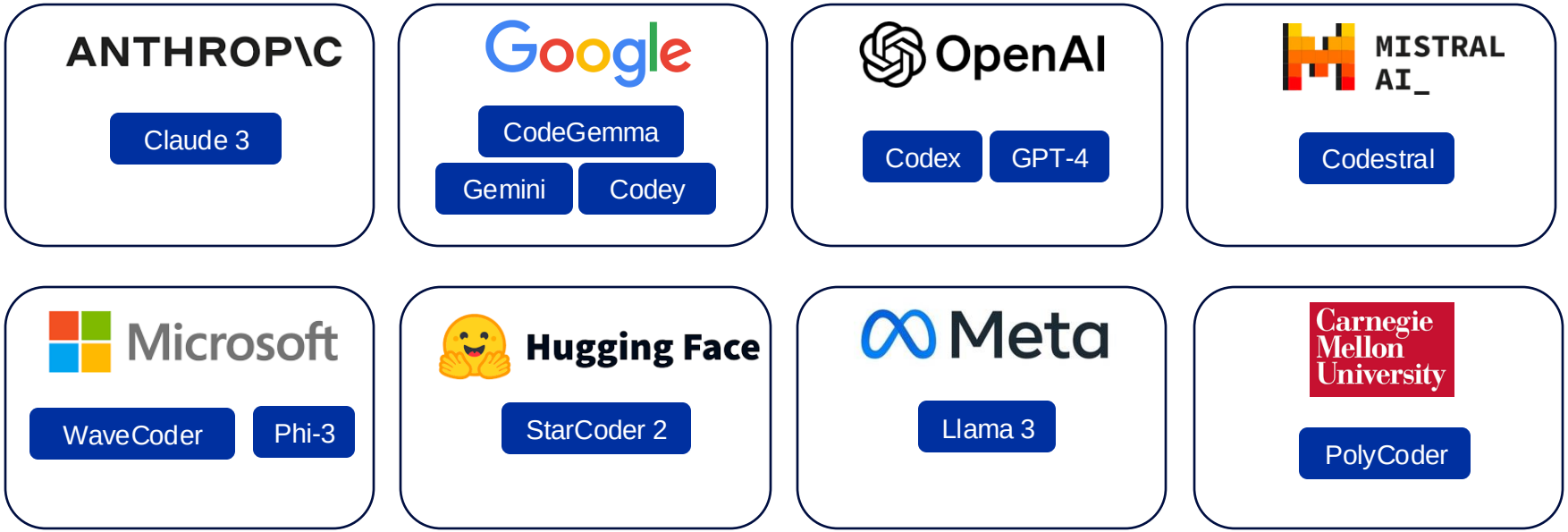
# The future of pharma is open-source

## It's not just about cost-cutting; it's about value-adding

- Pharmaceutical companies are turning to open source at every level to increase the speed and efficiency of clinical trials and bring life-saving treatments to patients faster
- Talent out of university prefers open-source programming languages like R & Python

| | |
|---|---|
| Make Go/No-Go Decisions Faster | Innovative ways to analyze clinical trial data and make faster decisions at every go/no-go checkpoint |
| Manipulate data more efficiently | Open-source can get you to SDTM/ADaM datasets quicker |
| Submit to the regulatory agencies confidently | Regulators like the FDA are embracing open-source for regulatory submissions |

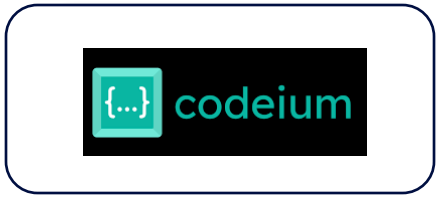
# Rapid proliferation of code generation models



# Most popular GenAI-powered code assistants

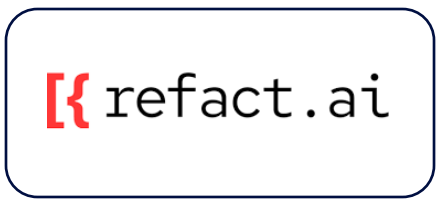















# GenAI-powered code assistants

- ✓ Code completion, code optimization, code debugging, code refactoring in multiple programming languages
- ✓ Natural Language Chat interface
- ✓ Integrate with several IDEs and code editors such as RStudio, Visual Studio Code, Visual Studio, Vim, Neovim, the JetBrains suite of IDEs, and Azure Data Studio
- ✓ Customer-generated code not used for training and improving the performance of the model
- ✓ Option to customize to an enterprise's codebase
- ✓ Deployment: Cloud/Local

# Challenges of leveraging code assistants for clinical programming

## Challenge 1: General Purpose Code Generation Models have not been trained on packages/libraries relevant for Clinical Programming

Solution: Pre-train/fine-tune/Retrieval-Augmented Code Generation using R and Python libraries and web frameworks

### R packages and Web frameworks

| | | | | | |
|---|---|---|---|---|---|
| ROCR | pROC | Shiny | Teal | Tidyverse | Jmpost |
| Readr | Survival | Lme4 | Nlme | Hmisc | Rmarkdown |
| Atable | ClinPK | CRTsize | Blockrand | RandomForest | |
| Knitr | Ggplot2 | Caret | ADCT | Pact | DoseFinding |
| Tidyr | Dplyr | Reshape2 | CompareODM | TidyCDISC | |

### Python Libraries and Web frameworks

| | | | |
|---|---|---|---|
| Pandas | NumPy | SciPy | Scikit-learn |
| TensorFlow | PyTorch | Matplotlib | Seaborn |
| Statsmodels | Plotly | Keras | Streamlit |
| Django | Flask | Dash | |

# Challenges of leveraging code assistants for clinical programming

## Challenge 2: Security & Privacy

Solution: Running LLMs locally (on-prem/local machine)

**Personal LLMs??**

**Most popular platforms to run LLMs locally**

Ollama

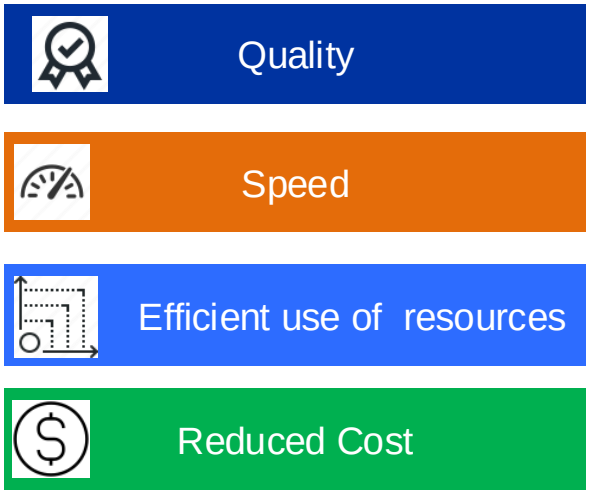
LM Studio

Jan

AnythingLLM

gpt4all

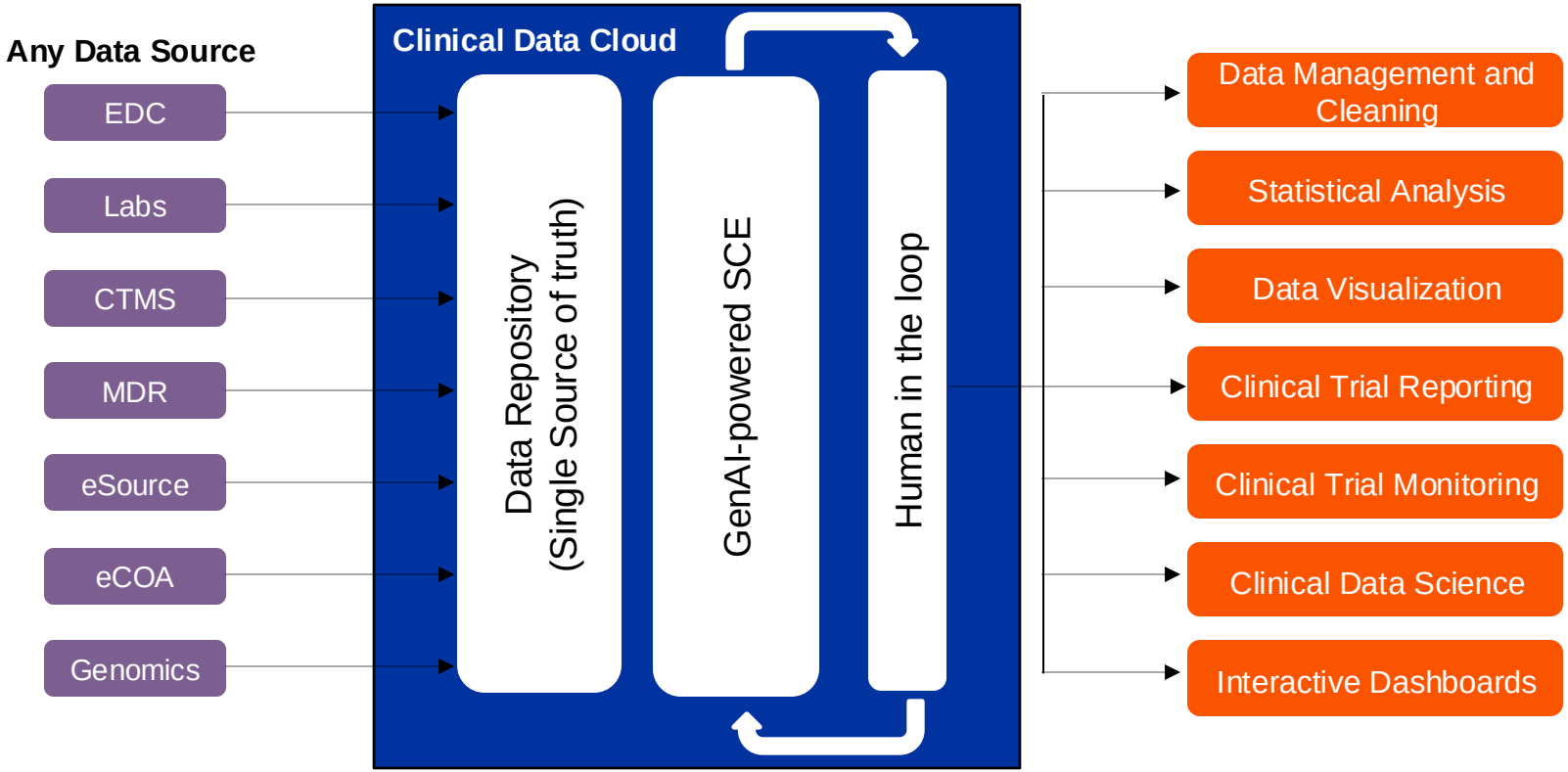
# Advantages of GenAI-powered code assistants




## ‘Trust’ but verify

- It’s just a model that’s been trained on a lot of data
- It’s not always right, appropriate, or optimal
- So, it’s important to verify the output before using it **(Human in the loop)**
- **Pair Programmer?**

# Need of the Hour – GenAI-powered SCE within Clinical Data Cloud



# Thank You!

For any Questions related to this presentation,
please email me at mohits12@hexaware.com


**The Global Healthcare Data Science Community**

**Contact Channels**
UK +44 1843 609600
office@phuse.global
phuse.global

**Social Media**
@phusetwitta
/phusebook
/phusetube
/company/phuse