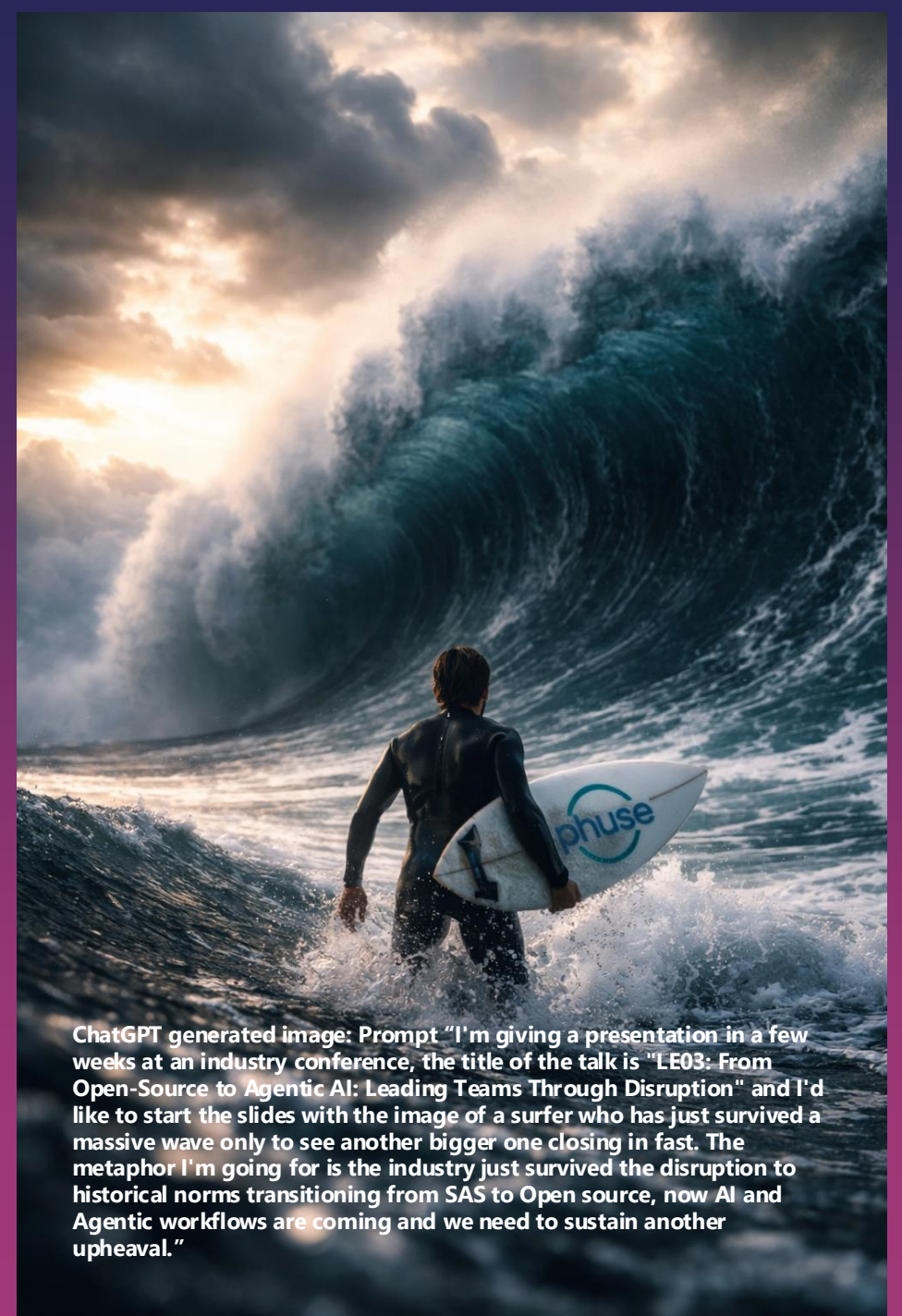


Cytel FSP- AI to Enable the Biometrics Workforce of the Future

Navigating the next big wave...

- Evolution from SAS to Open-source presented the largest disruption to our industry in last 20 years
- AI/ ML represents a potential for a major shift in the way we work in the next 3-5 years
- How do we prepare ourselves as leaders and our teams to surf the next big wave?



ChatGPT generated image: Prompt "I'm giving a presentation in a few weeks at an industry conference, the title of the talk is "LE03: From Open-Source to Agentic AI: Leading Teams Through Disruption" and I'd like to start the slides with the image of a surfer who has just survived a massive wave only to see another bigger one closing in fast. The metaphor I'm going for is the industry just survived the disruption to historical norms transitioning from SAS to Open source, now AI and Agentic workflows are coming and we need to sustain another upheaval."



AI Surf Report/ Forecast

What can we expect in the next 3-5 years

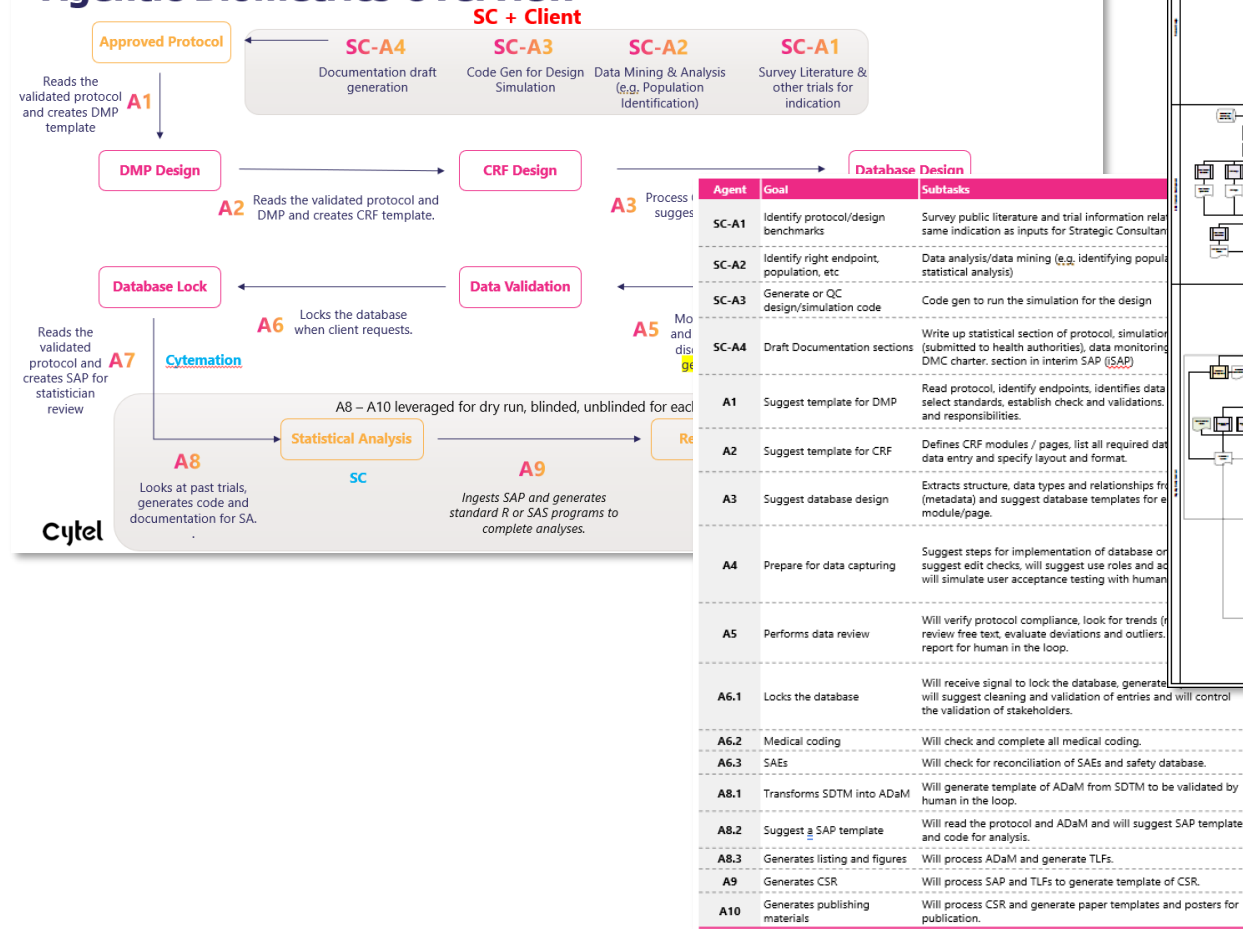


Agentic Solutions Driving Efficiency In Biometrics

Driving Efficiency, Scale & Quality

- +20% efficiency driven by agentic tools in components of the biometrics workflow
- Agentic biometrics is real today - moving from concept to working prototypes that already automate core biostatistics workflows
- Transforming productivity – AI agents accelerating protocol design, CRF set-up, data validation, and analysis, reducing manual cycles and
- Improving quality & compliance – automated checks, simulations, and validations strengthen regulatory alignment and reduce risk of errors
- Scaling with proof points – spanning efficiency gains, trial design & analysis, asset intelligence, and assisted validation, showing immediate tangible ROI

Agentic Biometrics Overview



Future Roles - Agentic Biometrics Specializations

	Clinical AI Model Specialist	AI Workflow & Interaction Designer	Output Quality & Compliance Analyst
Purpose	Ensure AI/ML models for dataset generation, anomaly detection, and narrative summarization are trained on accurate, relevant, and regulatory-compliant data	Act as the 'conversation designer' and operator for LLM-based assistants that help automate programming, data queries, and reporting tasks	Serve as the human safeguard who ensures AI outputs are statistically correct, clinically meaningful, and regulatory-compliant
Key Responsibilities	- Curate and label historical clinical trial data to train domain-specific AI models - Fine-tune models for CDISC mapping, data cleaning, and TLF generation - Incorporate regulatory and therapeutic area nuances into AI training data - Collaborate with statisticians and data scientists to define target outputs	- Create and optimize prompts for domain-specific tasks (e.g., FDA submission tables). - Integrate LLM/bot outputs into clinical programming workflows. - Manage multiple AI tools in sync across data pipelines. - Maintain an evolving prompt library for trial tasks	- Compare AI-generated datasets, tables, and narratives against specifications. - Identify and investigate discrepancies or suspicious outputs. - Ensure outputs meet statistical and regulatory requirements. - Maintain audit trails and validation documentation for inspections
Skills Needed	- Understanding of clinical trial data structures (SDTM, ADaM, raw CRF) - Proficiency in Python/R and ML frameworks - Data preprocessing, feature engineering, and bias detection - Regulatory data handling requirements	- Advanced prompt engineering techniques - Knowledge of LLM capabilities/limitations in clinical trials - Familiarity with API integrations for AI tools - Strong communication skills for technical translation	- Deep knowledge of statistical programming and analysis methods - Regulatory compliance expertise (FDA, EMA, PMDA) - Data forensics and debugging skills - Ability to interpret statistical results and clinical implications



How did we ride the Open Source Wave?



Case Study – R adoption

Scenario: Train & validate competency of ~200 associates within a well-established partnership to be able to program tables, figures, and listings using R

Cytel

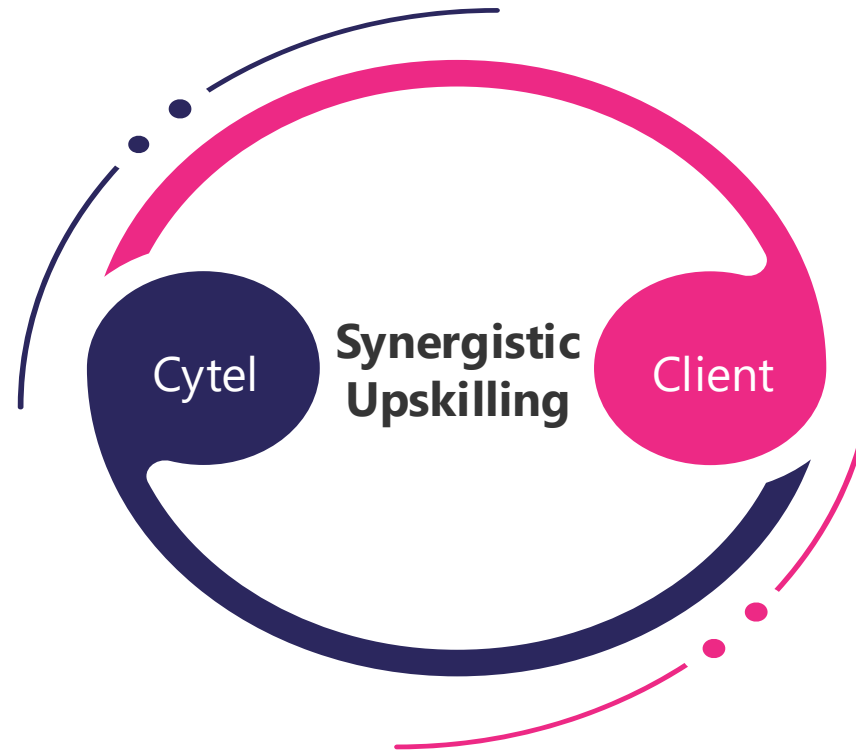
Assigned training curriculum and robust sandbox environment

R SMEs identified and open office hours established

Training exercises created on client reporting package

Regular discussion and feedback sessions centred around Cytel's open source community

Cytel



Collaboration

Staggered implementation

Adoption rate metrics

Joint training sessions with SMEs from both organisations leading

R reporting package designed to resemble familiar SAS macro suite

Standardisation of package usage across TAs/departments

Cytel – R training curriculum

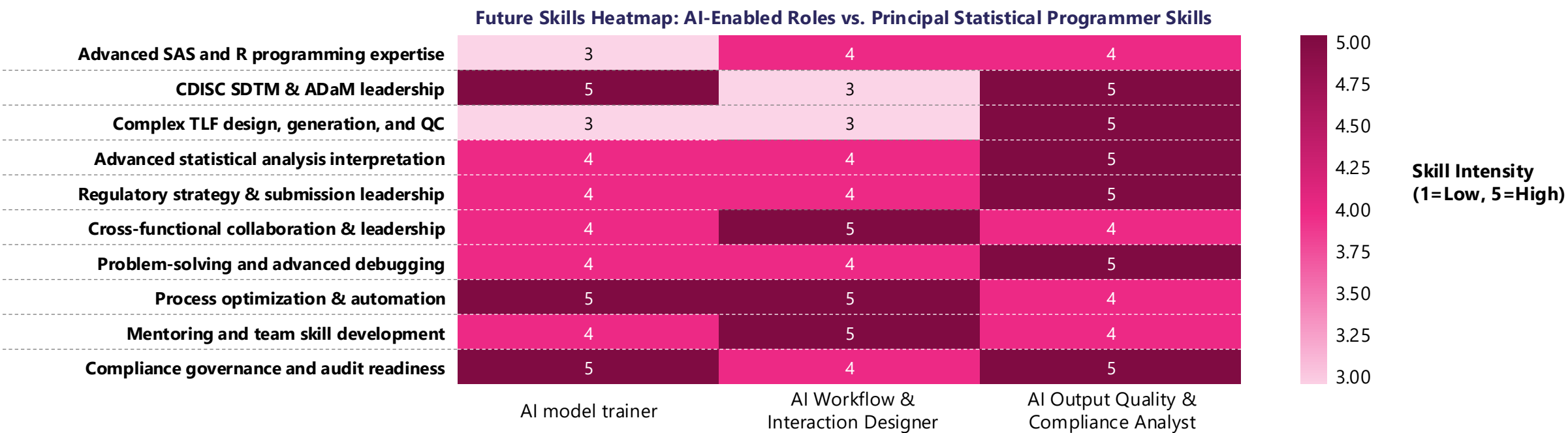
Title	Content	Platform	Level	Duration (hh:mm)
CRUG: Training 1	Rstudio, training resources, arithmetic operations	CRUG	beginner	00:27
CRUG: Training 2	Creating R objects, naming objects, assigning values, comments in the code	CRUG	beginner	00:20
CRUG: Training 3	Functions, content of vectors and manipulating their content, data types, matrices, null data type, handling missing data	CRUG	beginner	00:48
CRUG: Training 4	Installing packages, setwd, reading datasets, looking at data structure	CRUG	beginner	00:09
CRUG: Training 5	Selecting columns and rows from a dataset	CRUG	beginner	00:05
CRUG: Training 6	Converting character to factor, factor to character or numeric, renaming factors	CRUG	beginner	00:07
CRUG: Training 7	Manipulating data using 'dplyr' package: - selecting columns and rows from a data frame - creating new columns - renaming columns - subsetting rows using their positions - calculating summary statistics in the groups.	CRUG	beginner	00:50
	Reshaping a data frame from long to wide format and back using 'tidyr' package			
CRUG: Training 8	Manipulating data, part II - combining two data frames using 'join' functions from 'dplyr' package - selecting first or last row in a group - creating frequency tables - exporting a data frame to a .csv file	CRUG	beginner	00:47
	Visualizing data using ggplot2 (scatter plots, boxplots, time series plots)			
CRUG: Training 9	Visualizing data using ggplot2 (scatter plots, boxplots, time series plots)	CRUG	beginner	00:23
CRUG: Training 10	Introducing package to create tables, styling tables	CRUG	beginner	00:07
Complete Guide to R: Wrangling, Visualizing, and Modeling Data	Importing data, visualizing data with ggplot2, wrangling data, recoding data, exploring data, analyzing data, predicting outcomes	LinkedIn	beginner + intermediate	08:15
Cleaning Bad Data in R	Missing data, duplicate rows and values, converting data, formatting data, working with tidy data, tidying data sets, dealing with suspicious data	LinkedIn	beginner+intermediate	01:45
	Understanding the pipe (%>%) operator, importing .xlsx and .csv files, filtering and summarizing data sets, using tidyr to convert wide and long data sets, non-standard evaluation and programming with the tidyverse			
Learning the R Tidyverse	Understanding the pipe (%>%) operator, importing .xlsx and .csv files, filtering and summarizing data sets, using tidyr to convert wide and long data sets, non-standard evaluation and programming with the tidyverse	LinkedIn	intermediate	03:50
Data wrangling in R	Tidyverse & tibbles, importing data, data transformation, data cleaning and manipulation	LinkedIn	intermediate	02:51
R Programming in Data Science: High Variety Data	Types of big data, considerations used to determine the appropriate R package for Excel & importing workbooks, how to import standard text files using base R and tidyverse, foreign language packages for R, working on SAS files in the foreign language package, extracting data from a PDF into R	LinkedIn	intermediate	01:27
			Total Hours	22:11

Cytel

Waxing our boards



Training for Future needs- Upskilling the workforce



Phase / Month	Prerequisites	Focus Areas	Key Activities / Outcomes
Month 1–2 (Foundations)	SAS/R, CDISC SDTM & ADaM, Regulatory submissions experience	AI/ML Basics, Clinical Trial Applications, Regulatory Context	Python fundamentals, ML basics (bias, supervised/unsupervised), AI in trial case studies, FDA/EMA guidance
Month 3–4 (Applied Skills)	Foundations phase completed	AI Quality & Validation, Explainable AI, Automation & Tools	Validate AI-generated SDTM/ADaM datasets, QC AI TLFs, Apply SHAP/LIME, Build CI/CD validation pipeline, Explore cloud AI platforms
Month 5–6 (Advanced Oversight & Leadership)	Applied Skills phase completed	AI Governance & Ethics, Regulatory Mock Audits, Leadership Skills	AI audit trail documentation, Mock inspections, Lead oversight meeting, Capstone: validate AI outputs & prepare compliance package



The next wave isn't a threat, it's momentum!