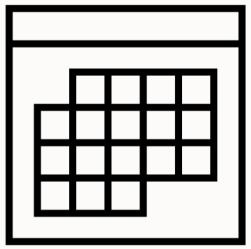
ORACLE
Life Sciences

Leveraging AI to achieve SDTM compliant outputs

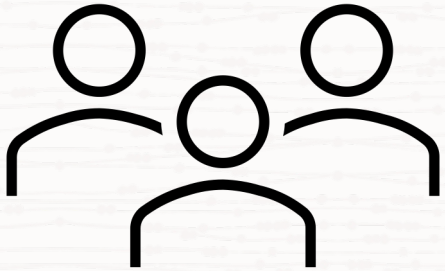
ACDM AI Symposium, 3rd Edition – Basel, 24-25 June

Sarah Jamal
Principal Solutions Consultant
Oracle Centre of Excellence

8-12 weeks

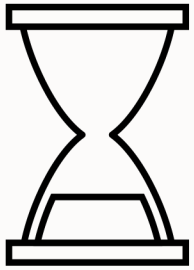


for SDTM datasets generation



Multiple resources

for peer validation or double programming



Delayed start waiting for patient data

Why SDTM and GenAI

SDTM generation today

Required by FDA and PDMA, recommended by EMA

Complex, requiring deep expertise

Time-consuming, causing delays in trial submission

Critical to achieve compliant submissions

Generative AI potential

Performing pattern recognition to **generate content**, based on training data

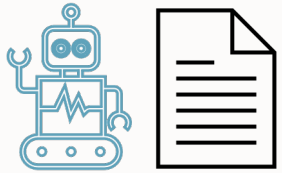
Large language models (LLMs) can process and understand human language, using a **dynamic knowledge base**

Some of the **limitations** of GenAI are misinformation, bias and lack of explainability.

“Let’s start from the beginning”

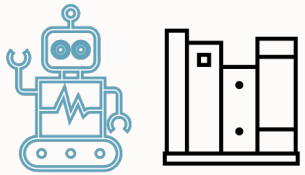
Study protocol

AI-assisted protocol creation



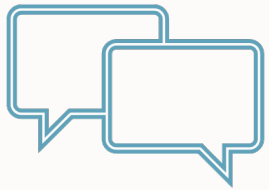
Data Collection

Automated build from standard metadata repositories



SDTM outputs

Generative AI for SDTM mapping



Conversational SDTM generation

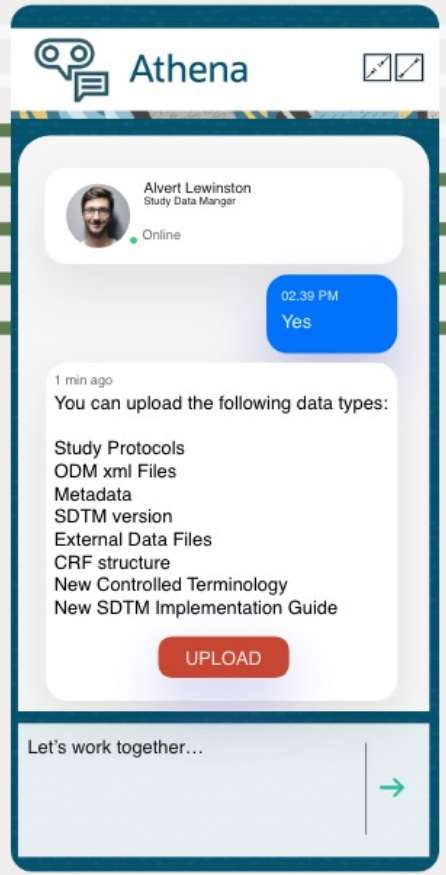
Concept generation

Intelligent Mapper

Search...

My Projects | New SDTM Project

Project Name	Protocol No.	Version	Compound Number	Project Admin	Knowledge Upload	Go Learn	Q&A	Map & Transform	Human Validation	Ready for Data	LIVE DATA STATUS
Hypertension-342	56446754	1.3	Compound-A321	✓							
CardioBoost-307	63467565	1.1	Molecule-B674	✓							
DiabetesControl-156	13344585	3.2	Agent-C859	✓							
ArthritisEase-998	33454665	1.0	Drug-D092	✓							
DermHeal-629	83464565	2.4	Formula-E303	✓							
PulmoCare-473	63474965	2.6	Therapy-F710	✓							
CancerShield-840	25453525	1.6	Substance-G428	✓							
NeuroRelief-521	26474965	3.2	Treatment-H515	✓							IN-PROGRESS



Intelligent Mapper

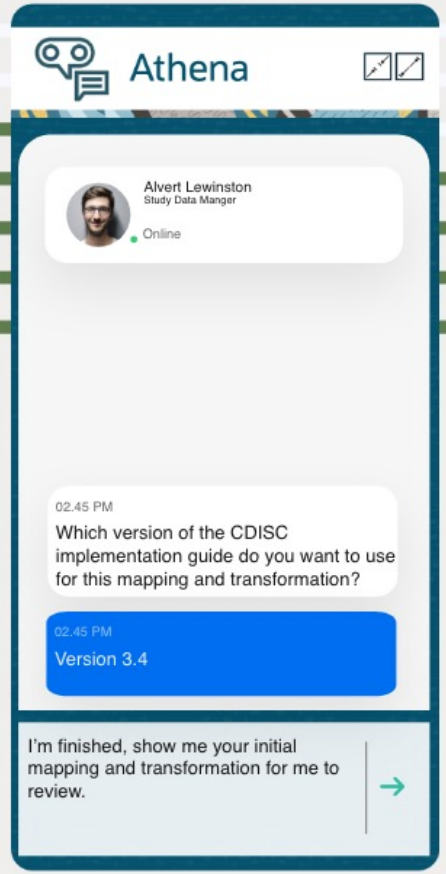
Search...

My Projects | New SDTM Project

Project Name	Protocol No.	Version	Compound Number	Project Admin
Hypertension-342	56446754	1.3	Compound-A321	✓
CardioBoost-307	63467565	1.1	Molecule-B674	✓
DiabetesControl-156	13344585	3.2	Agent-C859	✓
ArthritisEase-998	33454665	1.0	Drug-D092	✓
DermHeal-629	83464565	2.4	Formula-E303	✓
PulmoCare-473	63474965	2.6	Therapy-F710	✓
CancerShield-840	25453525	1.6	Substance-G428	✓
NeuroRelief-521	26474965	3.2	Treatment-H515	✓

Knowledge Upload | Go Learn | Q&A | Map & Transform | Human Validation | Ready for Data | LIVE DATA STATUS

IN-PROGRESS

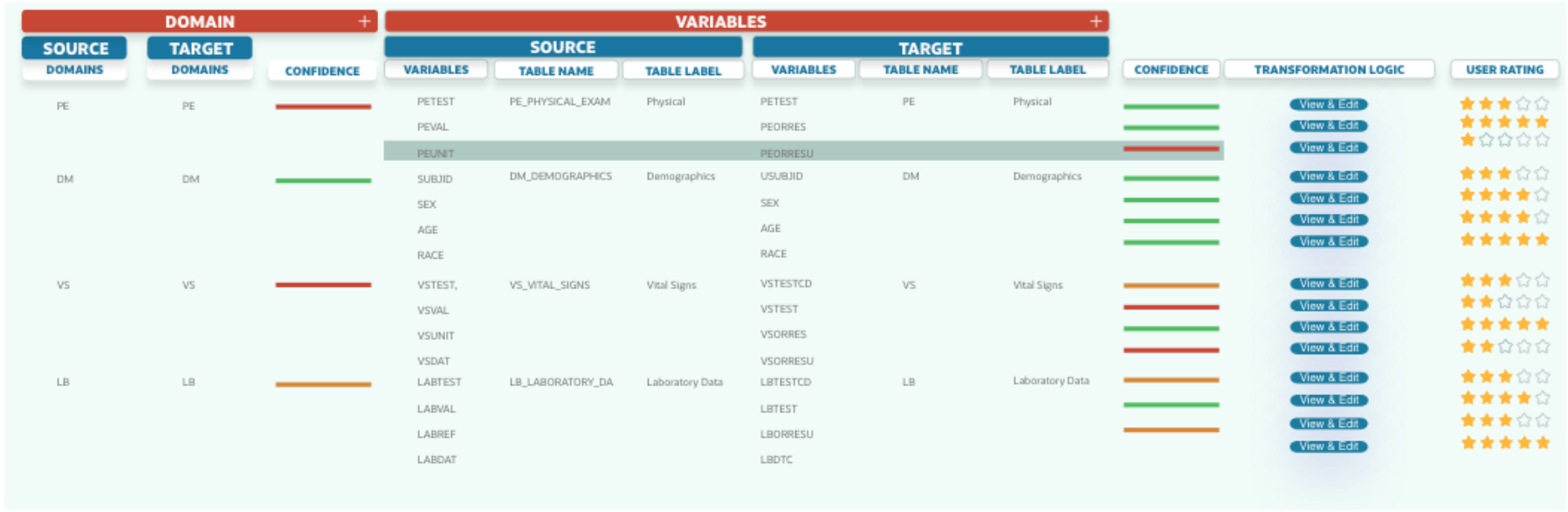


Intelligent Mapper

Re-Run Mapping

Run Test Data

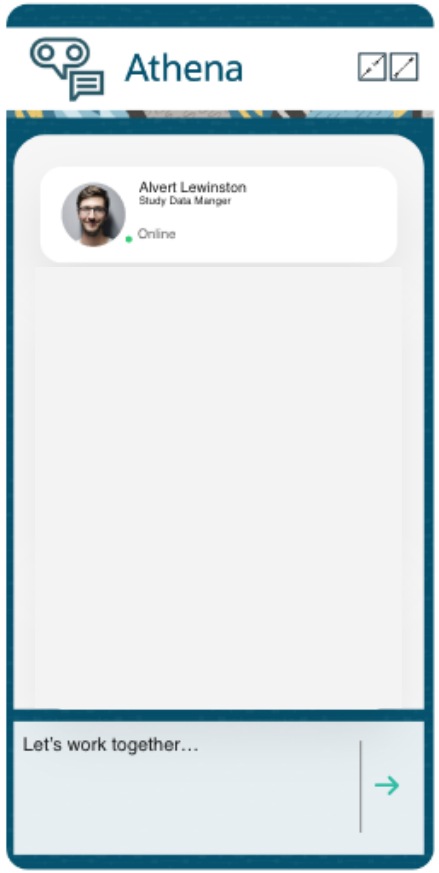
Run Live Data



DOMAIN			VARIABLES								
SOURCE	TARGET		SOURCE			TARGET					
DOMAINS	DOMAINS	CONFIDENCE	VARIABLES	TABLE NAME	TABLE LABEL	VARIABLES	TABLE NAME	TABLE LABEL	CONFIDENCE	TRANSFORMATION LOGIC	USER RATING
PE	PE		PETEST	PE_PHYSICAL_EXAM	Physical	PETEST	PE	Physical		View & Edit	
			PEVAL			PEORRES				View & Edit	
			PEUNIT			PEORRESU				View & Edit	
DM	DM		SUBJID	DM_DEMOGRAPHICS	Demographics	USUBJID	DM	Demographics		View & Edit	
			SEX			SEX				View & Edit	
			AGE			AGE				View & Edit	
			RACE			RACE				View & Edit	
VS	VS		VSTEST,	VS_VITAL_SIGNS	Vital Signs	VSTESTCD	VS	Vital Signs		View & Edit	
			VSVAL			VSTEST				View & Edit	
			VSUNIT			VSORRES				View & Edit	
			VSDAT			VSORRESU				View & Edit	
LB	LB		LABTEST	LB_LABORATORY_DA	Laboratory Data	LBTESTCD	LB	Laboratory Data		View & Edit	
			LABVAL			LBTEST				View & Edit	
			LABREF			LBORRESU				View & Edit	
			LABDAT			LBDTC				View & Edit	

KNOWLEDGE BASE

- Annotated CRF
- Controlled Terminology
- SDTM Version
- Code List
- Hypertension DM
- Hypertension-987



PROJECT NAME: Hypertension-342 PROTOCOL NUMBER: 56446754

Intelligent Mapper

Re-Run Mapping

Run Test Data

Run Live Data

OUTPUT VALIDATION:

Interventions

AE CM EX SU MH

Events

DS AE CE

Findings

EG LB VS PE IE MS MI TA

Special Purpose Domains

DM SV SE CO SC

Trial Design Domains

TE TA TI TV

Relationship and Reference Domains

RELREC SUPPQUAL

Additional Domains for Specialised Studies

QS RS DA DV

Pharmacodynamic Domains

PC PP PD

Genetic Domains

GF CG BG

Healthcare Domains

HO IS BS

Device Domains

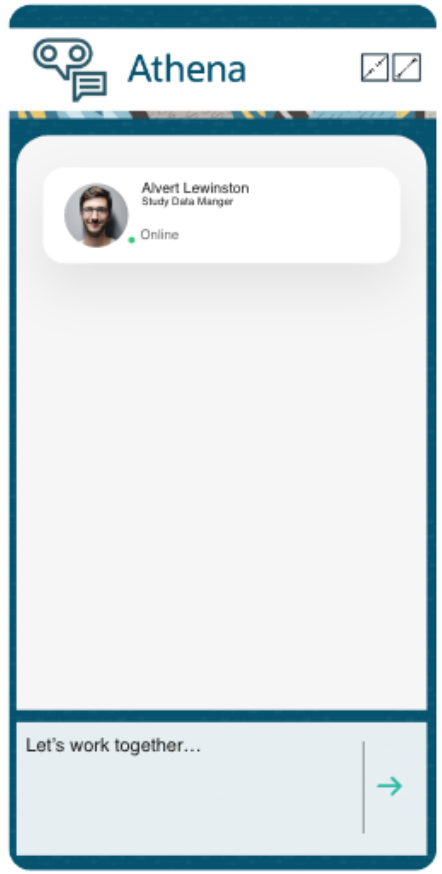
DI DT DR

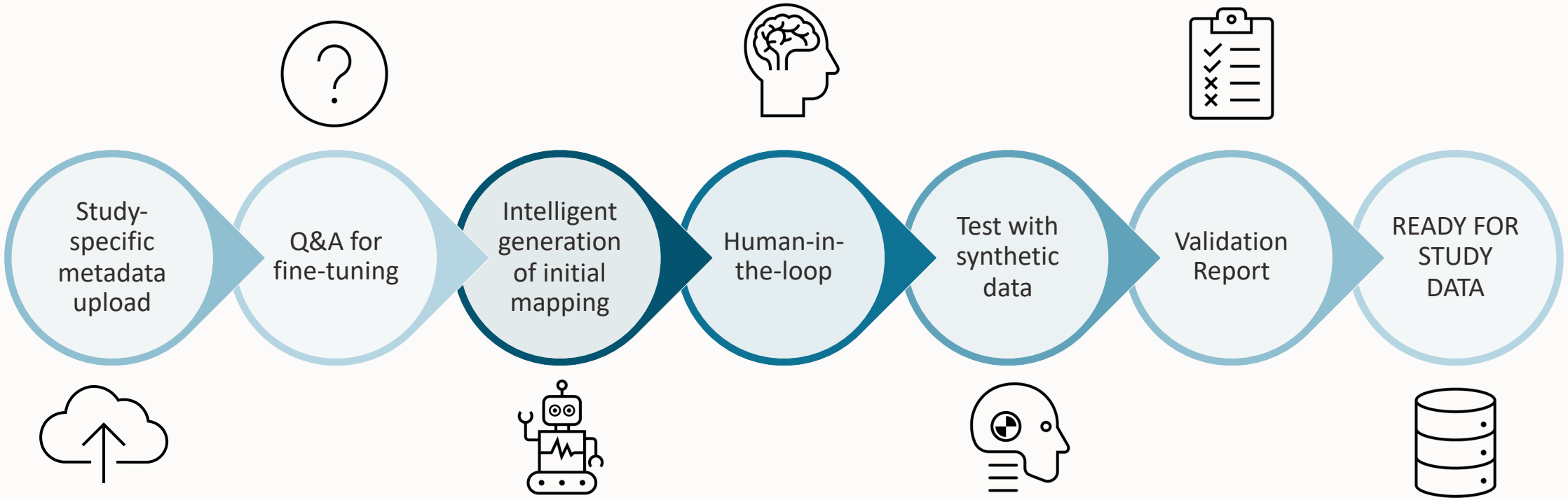
Other Spacial Domains

CP BE RELSPEC

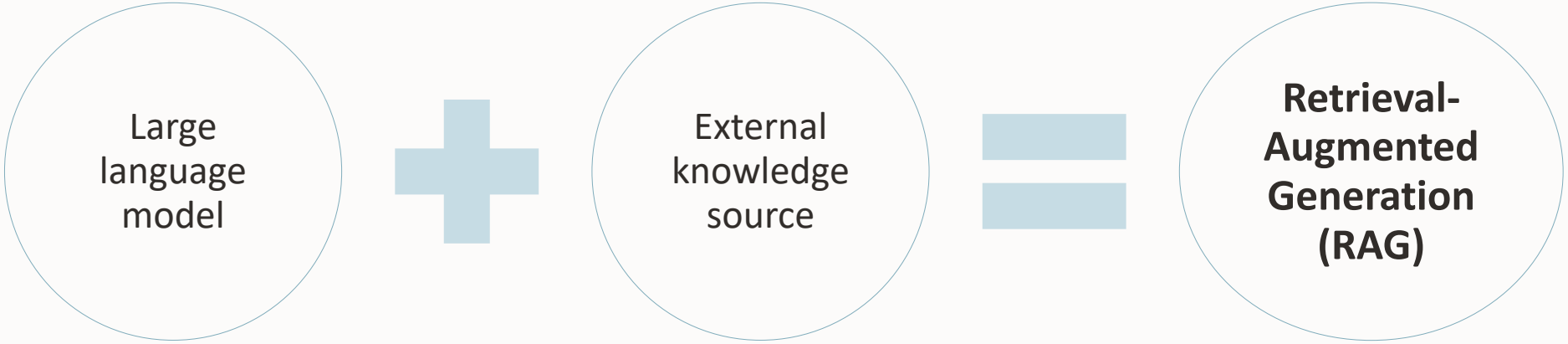
KNOWLEDGE BASE

Annotated CRF (last updated on 04-06-24 at 10.06am GMT)
Controlled Terminology
SDTM Version
Code List
Hypertension DM
Hypertension-987
MDR File




Study-specific metadata upload
Q&A for fine-tuning
Intelligent generation of initial mapping
Human-in-the-loop
Test with synthetic data
Validation Report
READY FOR STUDY DATA

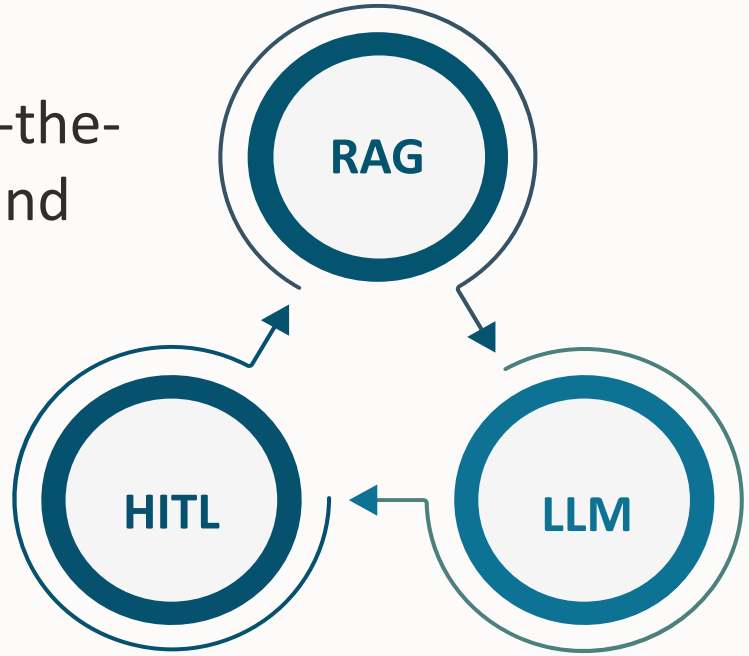
Overcoming GenAI limitations



- **Mitigates hallucinations**
- **Improves accuracy**
- **Provides traceability**

Summary

RAGs and Human-in-the-loop to reduce risk and ensure compliance



GenAI to speed up process and provide assistance

Generate SDTM up to **70%** faster, with **1** resource, starting **prior to FPI**

Thank you

Sarah Jamal

Principal Solutions Consultant