

Single Day Event

Leveraging NLP for Enhanced Accuracy in
Statistical Analysis Efforts Estimation

Umayal Annamalai,
Precision for Medicine

Karthik Venkatraman,
Precision for Medicine



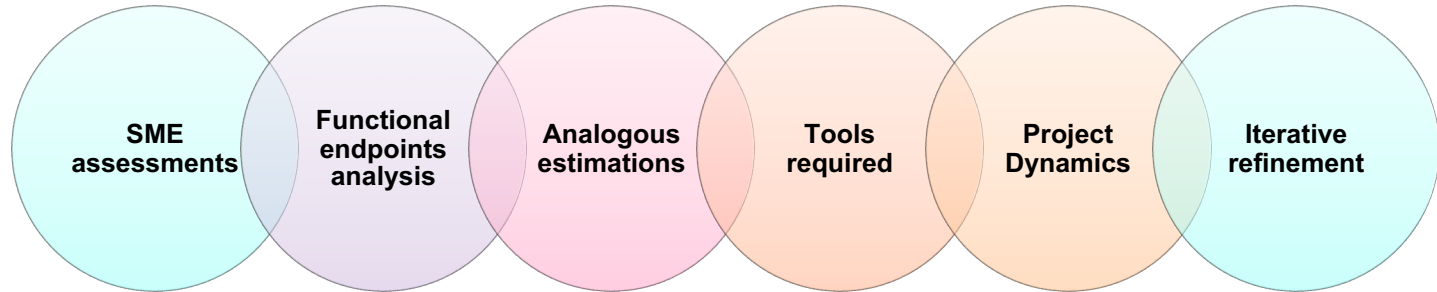
Agenda

- Statistical analysis efforts estimation
- Why is statistical analysis efforts estimations important?
- How is it done currently?
- Need for streamlined process
- Use of NLP in statistical analysis effort estimations
- Model
- Use cases:
 - SDTM Estimations
- Benefits and challenges
- Key takeaways



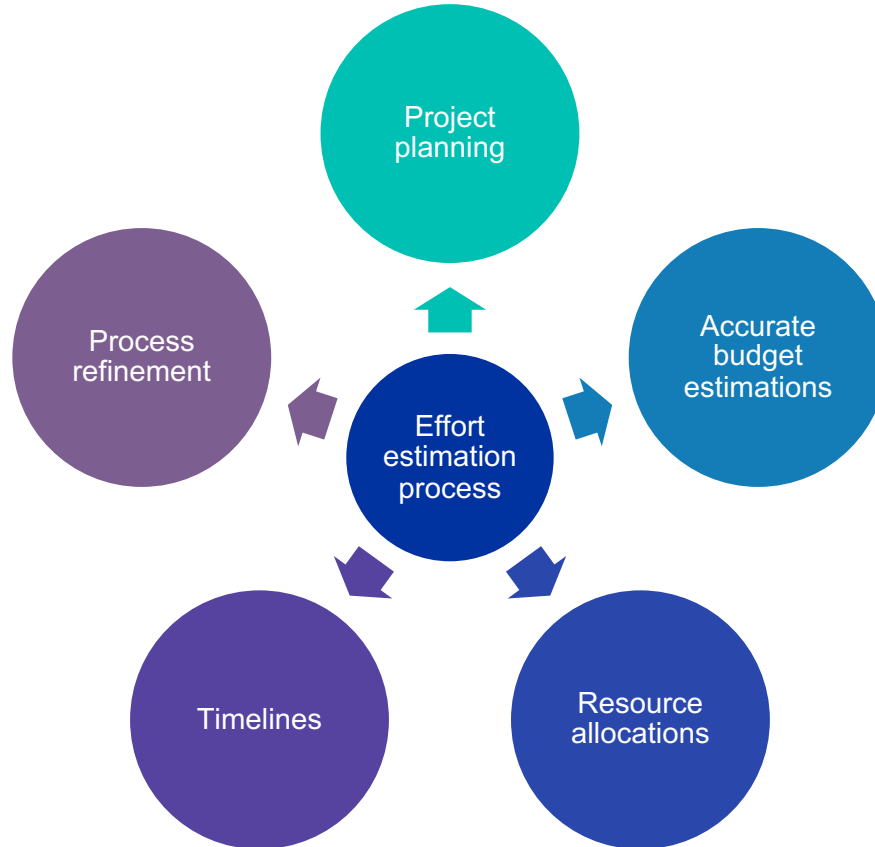
Statistical analysis effort estimations

Estimating the resources, time and cost required to conduct statistical analysis for clinical trials

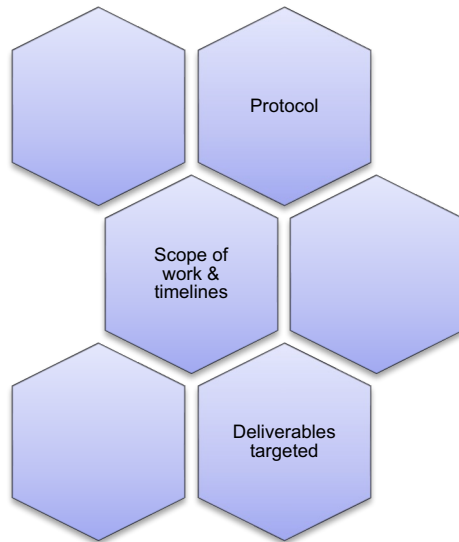




Why are statistical analysis efforts estimation important?



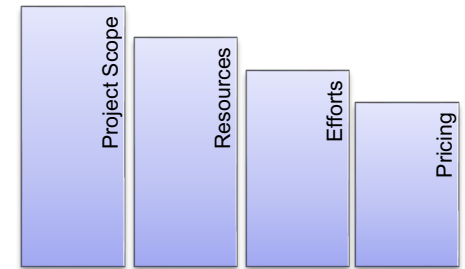
How is it done currently?



Sponsor inputs



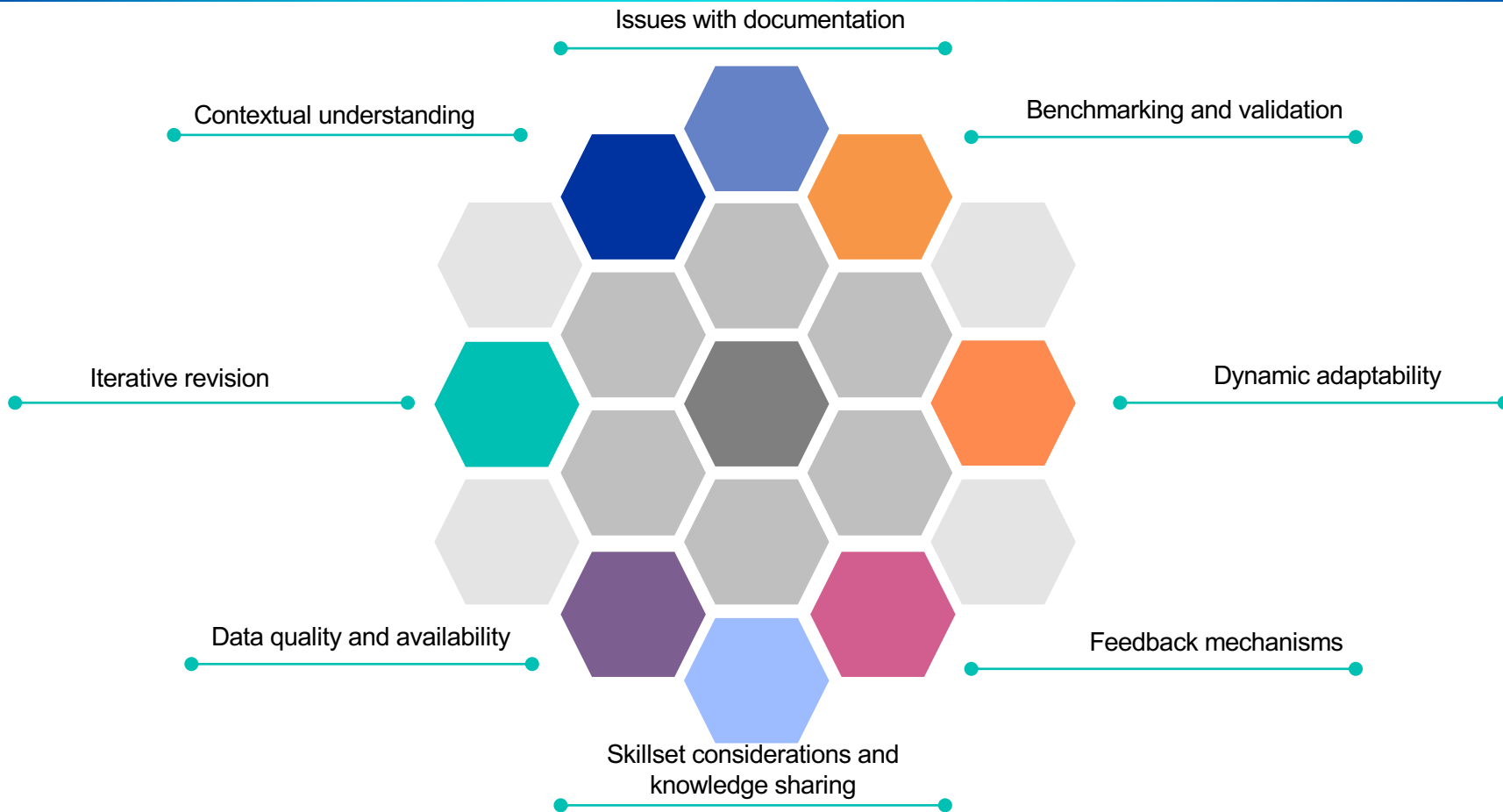
SME Inputs



Documentation

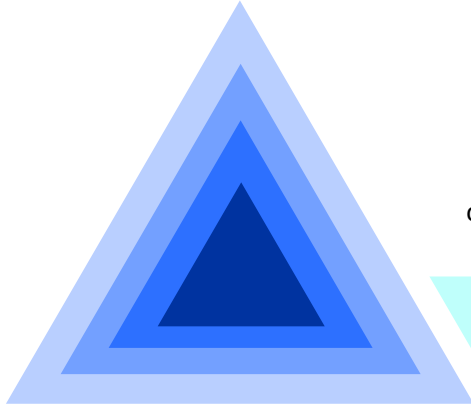


Gaps in estimation





Need for streamlining the estimation process

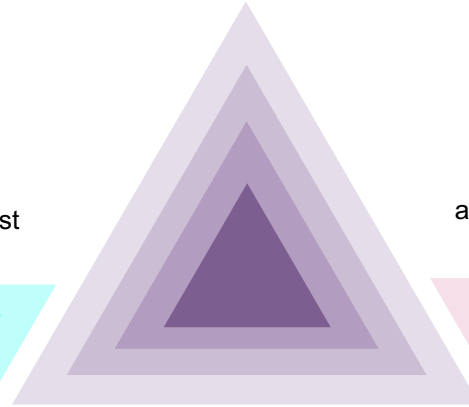


Address frequently faced issues

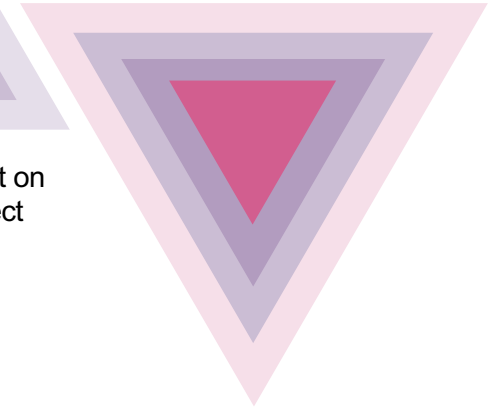
Technological advancements, process refinement, and a commitment to learning from past experiences



Iterative process improvement on feedback and evolving project requirements

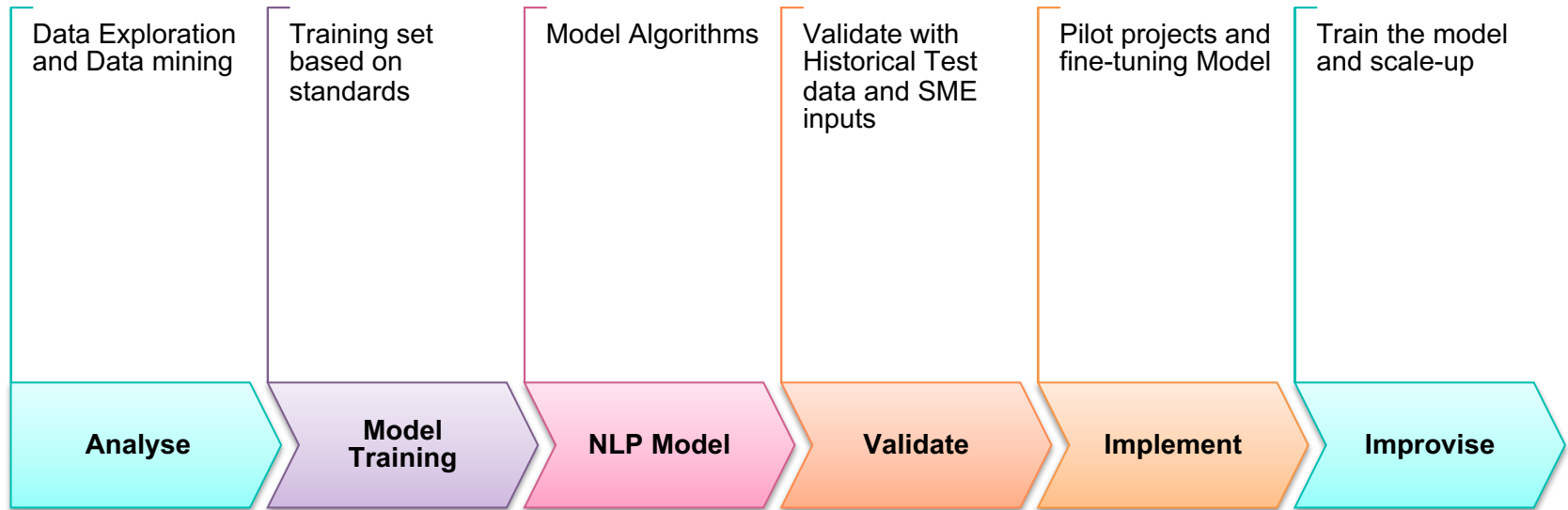


Moving from a manual process to automation for enhanced efficiency, accuracy, and effectiveness



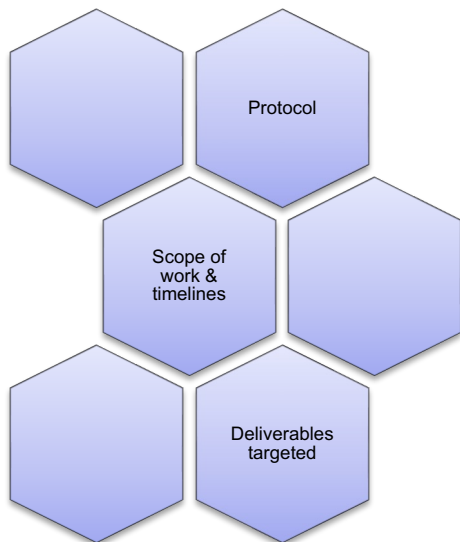


Use of NLP in statistical analysis effort estimations





Where NLP can be applied?



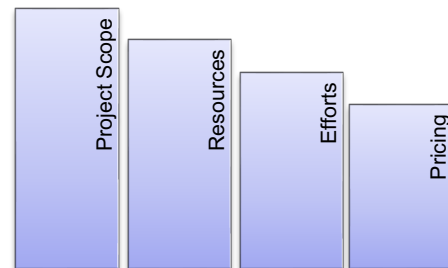
Sponsor inputs



SME Inputs

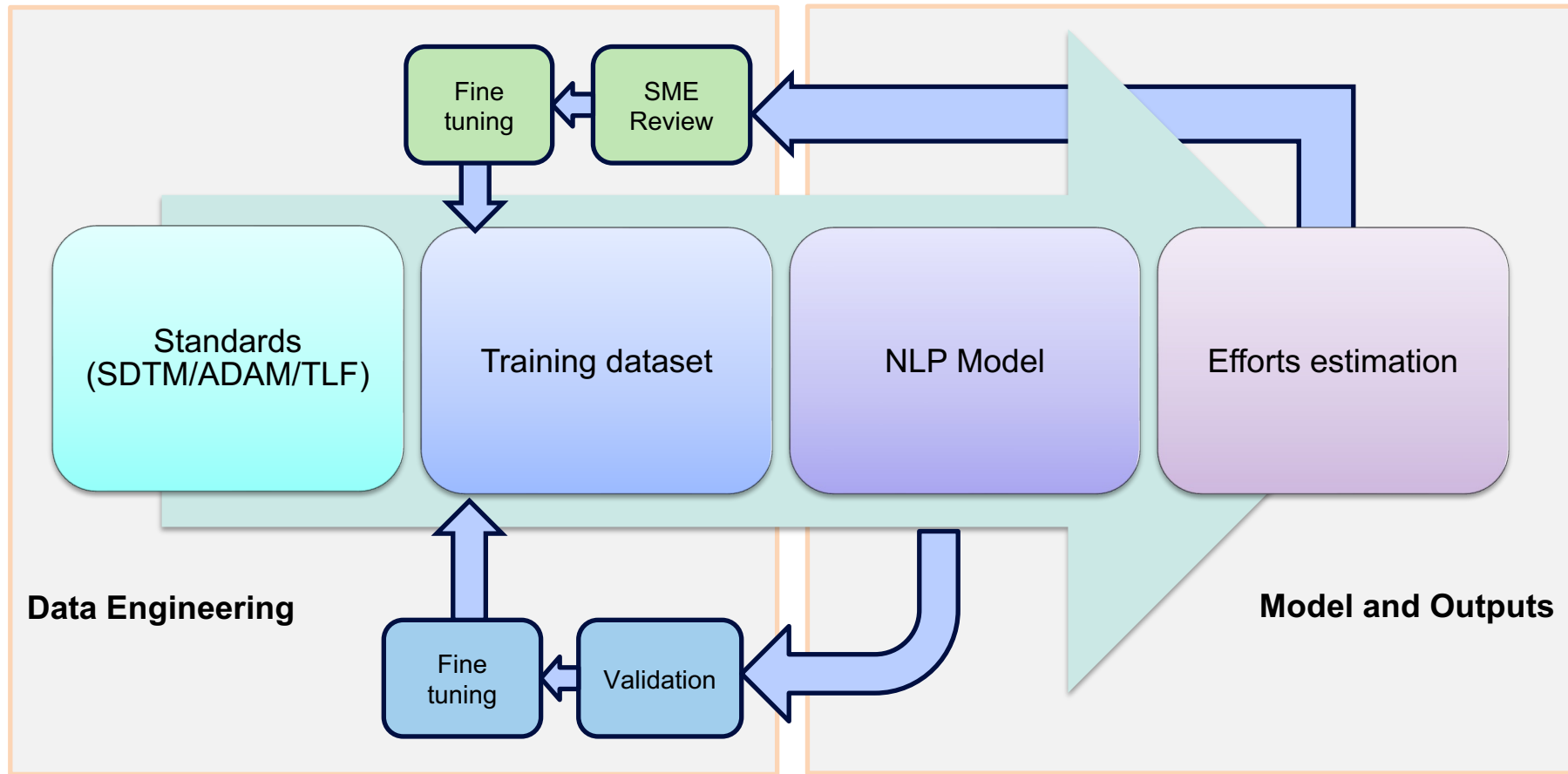
*NLP estimations

**Pre-defined Algorithmic estimations



Documentation

NLP Model

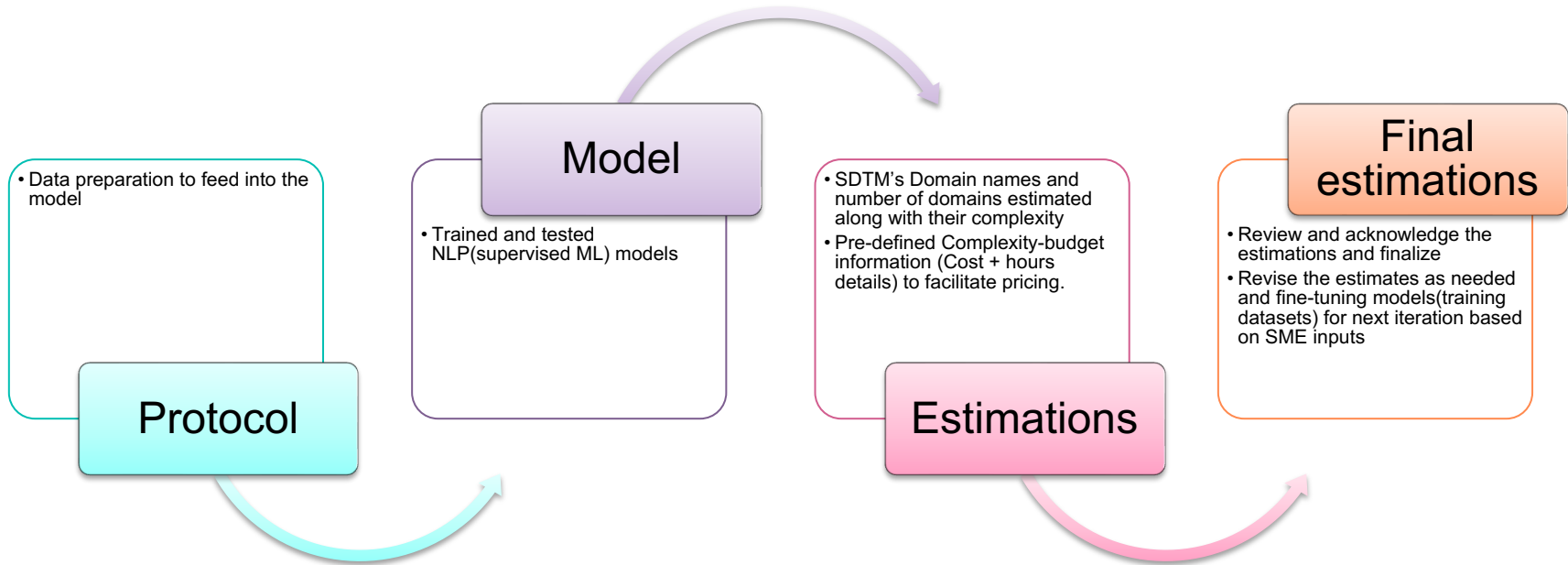


Use Case



Use case- Determining the number of SDTMs with their complexity

Application of NLP methods in determining the number of SDTMs needed for the study and the complexity + efforts estimated





Intermediate outputs from the model

Screenshot 1: Similarity Score

Domain	Cosine_Score_SUM	Cosine_Score_Mean
FA	10.36589838	0.062823627
CV	2.003074436	0.019447325
RP	7.093712373	0.097174142
RE	3.474866562	0.034404619
SR	0.02118832	0.003026903
NV	1.293024112	0.035917336
OE	5.572303815	0.075301403
IS	0.204700007	0.003055224
MB	2.061983812	0.00950223
BS	0.300522391	0.018782649
SS	0.115214724	0.038404908
CP	2.907384549	0.007178727
GF	0.028012351	0.00560247
LB	17.22874321	0.015996976
TS	8.104915611	0.070477527
VS	2.006232746	0.036476959
EG	3.869600184	0.021986365
SC	4.914189871	0.100289589
DA	0.019602165	0.003267028
TU	0.066743096	0.005134084

Screenshot 2: Number of SDTM Domains

Phrases matched	Domain	Complexity
Default	DM	Medium
Default	EX	High
Default	SE	High
Default	SV	High
Default	TI	Medium
Default	TE	Medium
Default	TA	Medium
Default	TS	Medium
Default	TV	Medium
Default	DV	Low
Default	IE	Low
Default	CM	Low
Default	AE	Low
Default	MH	Low
206	FA	High
126	CV	Low
108	RE	Medium
267	MB	Low
1438	LB	High
62	VS	Medium
182	EG	Medium
55	SC	Low

Screenshot 3: Pre-defined Complexity-Cost factor

Complexity	Skillset Level	Time in hours	Cost/hour	Total	Type
High	S1	xx	xx	xx	Prod
Medium	S1	xx	xx	xx	Prod
Low	S1	xx	xx	xx	Prod
High	S1	xx	xx	xx	Val
Medium	S1	xx	xx	xx	Val
Low	S1	xx	xx	xx	Val
High	S2	xx	xx	xx	Prod
Medium	S2	xx	xx	xx	Prod
Low	S2	xx	xx	xx	Prod
High	S2	xx	xx	xx	Val
Medium	S2	xx	xx	xx	Val
Low	S2	xx	xx	xx	Val
High	SS1	xx	xx	xx	Prod
Medium	SS1	xx	xx	xx	Prod
Low	SS1	xx	xx	xx	Prod
High	SS1	xx	xx	xx	Val
Medium	SS1	xx	xx	xx	Val
Low	SS1	xx	xx	xx	Val

Screenshot 4: Protocol-Domain Mapping: RE

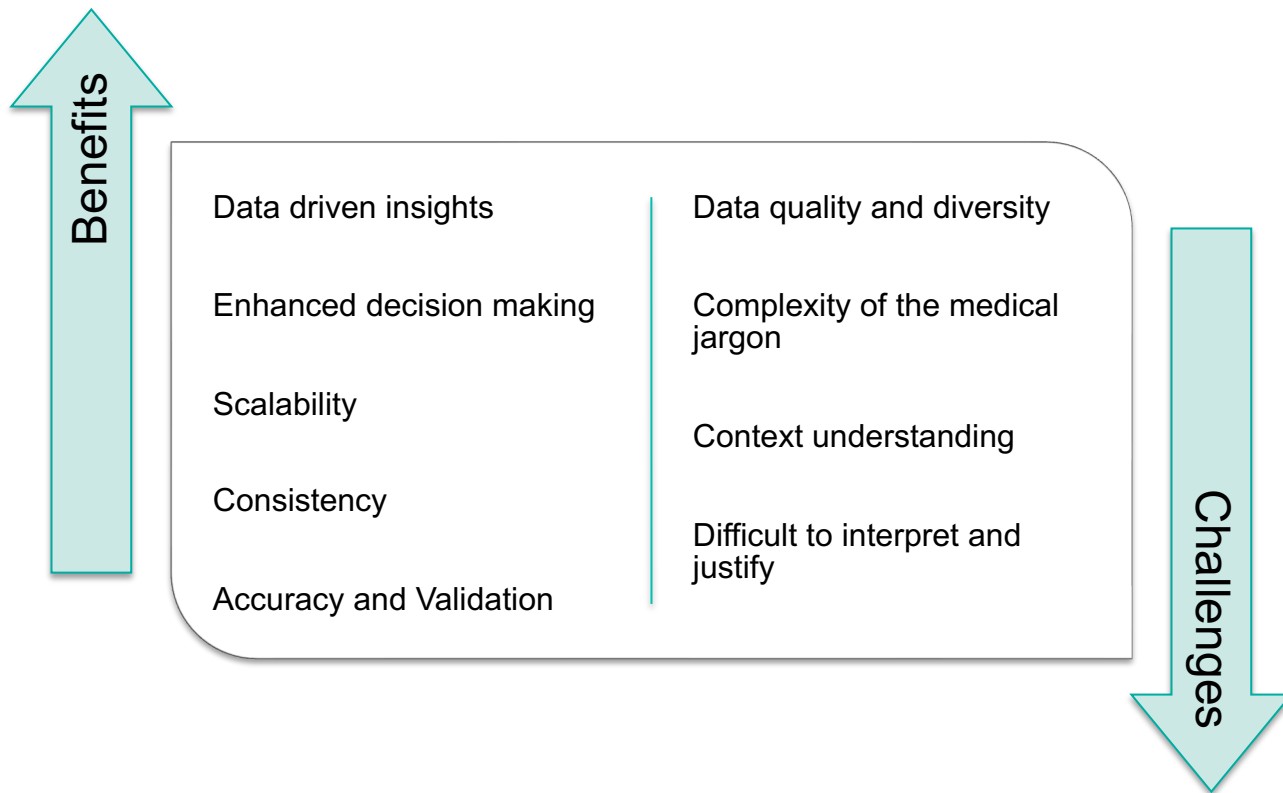
Matched Phrase from protocol (for review)	Domain
diffusion capacity of lung for co; tlc; transfer factor of lung for co	RE
mean diffusion capacity of lung for co	RE
fraction of inspired oxygen	RE
lci to 1/40th initial concentration; lci to 2.5% initial concentration; lung clearance index to 1/40th initial concentration; lung clearance index to 2.5% initial concentration	RE
lci to 1/20th initial concentration; lci to 5% initial concentration; lung clearance index to 1/20th initial concentration; lung clearance index to 5% initial concentration	RE

Screenshot 5: Protocol-Domain Mapping: FA

Matched Phrase from protocol (for review)	Domain
largest diameter of primary lesion; longest diameter of primary lesion	FA
number of abscesses	FA
number of episodes	FA
number of fistulas	FA
number of hospitalizations	FA



Benefits and Key challenges





Key takeaways



**Streamlined
process**



**Faster decision-
making and early
insights**



**Upfront costs Vs
long-term
benefits**



**Enhanced
Collaboration
across different
stakeholders**



**Optimization
of resources
utilization,
adaptability, and
consistency**



**Future
enhancements
and continuous
improvement**

Thank you

umayal.annamalai@precisionformedicine.com

karthik.venkataraman@precisionformedicine.com



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://linkedin.com/company/phuse)