Model 1: NER (Named Entity Recognition) model

Model 2: LLM (Large Language Model) model

1. **Phase 1** – APIs, connections, onboarding, cloud, hiring, security clearance, equipment. Depends on how messy their stuff is, and if they have multiple data set systems. **Estimated 4-6 weeks** with FDA cooperation, security and credentialling done in 2 weeks and data loading (need to get with API guy to verify).
2. Past/training data provided to data scientist.
3. APIs and connections set up. Access granted.
4. Credentialing and security clearances established/granted
5. Azure accounts purchased and connected.

**Phase 2** – Training NER Model on past data/ shadowing with current editor staff. **Estimated** **8-12 weeks.**Must start only when all phase one tasks have been completed. *Most Phase 2 tasks can be done simultaneously* by Loader/API expert, 2 Data scientist (one FTE and John PTE) and UI expert. For pricing, the PT position funding for Phase 3 will be used in Phase 2.

1. API Expert PTE 10/hours/week- API pipeline cleaning – Need API timeline from John’s contact. **Estimated 4 weeks.**
2. Data Scientist FTE 40/hours/week- Training Data loaded into model with tuning, cross validation between redacted and unredacted form 3500. – **Estimated 8-12 weeks.**
3. UI expert PTE 10/hours/week - UI refinement and interface updates for greater customer usage and satisfaction. **Estimated 2-4 weeks.**
4. Begin Model 2 Training.
5. Begin cross-training of editors that will be shifted to QC Team, identification of new QC supervisors and preparation of QC Team for 100% QC.

**Phase 3** – Model 1-2 LLM continued refinement. Full quality control on CLARA work. There will be no PTE in this phase, only occasional consulting hours from John as needed. **Estimated 6 weeks total.**

1. Derryck needs to train Quality Control Team on using CLARA in real time. **Estimated 1 week.**
2. Data scientist is working closely with Quality Control Team on catching errors and mistakes. QC Team will report findings on Model 1 and those will be input into Model 2.
3. Percentage accuracy of each model must be established. **Model 1 percent accuracy \_\_ and Model 2 \_\_.**
4. Starting week 2 of Phase 3 we can begin phasing editors into Quality Control at 10 per week. Remaining editors that are not moved to quality control will be moved to other Task Orders or terminated. **Estimated 4 weeks.**

**DEMO**

 Show them a batch import of 50 PDFs. 2 min waiting time. Show hybrid model. Make grammatical changes, adding

UI – Need to be able to edit text, assign B(4) and B(6) labeling. Need to look clean and simple for import and export as a original and redacted files. Side by side comparison or showing original text vs redacted. Need to be user-friendly for

Practice demo 2 time prior to FDA Demo. Have Derryck Demo. Schedule calls.