***General Guidelines (Leaving for other readers)***

*1) Project Proposal:*

*Up to 3 pages write-up + 1 page of references (minimum 6 papers)*

*Guide:*

*– Motivation: Why is this problem important? Why do we care this problem?*

*– Literature Survey: Conduct literature search to understand the state of arts and the gap for solving the problem. Formulate the deep learning problem in details (e.g., classification vs. predictive modeling vs. clustering problem).*

*– Data: Describe the dataset you use, and elaborate on how you would play with the data in your project. Preliminary results are encouraged but not required. It is recommended to try to cover as many aspects as described in project initiation to give you a better navigation in later period of project phase. This is a crucial step, please do it on the first day and never stops until the project.*

*– Approach: Identify the high-level technical approaches for the project (e.g., what algorithms to use or pipelines to use). Identify clearly the success metric that you would like to use (e.g., AUC, accuracy, recall, speedup in running time).*

*– Experimental Setup: Setup the analytic infrastructure for your project (including both hardware and software environment, e.g., AWS or local clusters with Python, PyTorch and all necessary packages).*

*– Timeline: Prepare a timeline and milestones of deliverables with reasonably proposed task distributions for the entire project.*

***Tasks:***

*1) Identify and motivate the problems that you want to address in your project.*

*2) Conduct literature search to understand the state of arts and the gap for solving the problem.*

*3) Formulate the data science problem in details (e.g., classification vs. predictive modeling vs. clustering problem).*

*4) Identify clearly the success metric that you would like to use (e.g., AUC, accuracy, recall, speedup in running time).*

*5) Setup the analytic infrastructure for your project (including both hardware and software environment, e.g., Azure or local clusters with Python, PyTorch and all necessary*

*packages).*

*6) Discover the key data that will be used in your project and make sure an efficient path for obtaining the dataset. This is a crucial step and can be quite time-consuming,*

*so do it on the first day and never stops until the project completion.*

*7) Generate initial statistics over the raw data to make sure the data quality is good enough and the key assumption about the data are met.*

*8) Identify the high-level technical approaches for the project (e.g., what algorithms to use or pipelines to use).*

*9) Prepare a timeline and milestones of deliverables for the entire project.*

***Notes:***

* *Hard requirment of 3 pages + 1 page of References, we cannot exceed or will be rejeced.*
* *Suggests iteration. I think we should consider focusing on a dataset that we can use with the least amount of effort and develop a model. WE could then go back to other datasets and see about inclusion.*
* *Access to ACM(maybe others) will require using IllinoisNet VPN(Seach for it on google). Tunnell All is the setting we want so that the connection to ACM goes through IllinoisNet and we look like we’re in their doman.*