Gathering: Scripted Automation (Python) scrapping for open data hosted on cloud service providers/ vendor platforms. (PDFs docs, HTML pages, JSON manifests, IaaC codes, documentation templates, etc.,)

Processing: Python script for data cleanup tasks - such as: Splitting, Index consolidation, corpus Text extraction, common/ recurrent string removal, tokenization, lemmatization, dictionary creation)

Analysis: Vectorization, bigram, trigram, n gram computation, LDA modelling, cosine similarity calculation.

Visualization/ presentation: PyLDAvis interactive topic model histogram chart, intertopic pca distance quadrant, relevance metric slider. Also, t-SNE cluster visualization using Bokeh for topic grouping of document context.

Preservation: GitRepo, S3 bucket, local HDD, G-Drive.