**Focused Checklist**

**Part 0 — Setup**

* Python env ready (Py3.10+), GPU enabled (Colab Runtime → GPU)
* Install: transformers, torch, accelerate, pysbd, spacy, en\_core\_web\_sm
* Repo structure: src/, data/, lib/, out/

**Part 1 — Data & EDA**

* Story ingestion & sentence segmentation
* Basic EDA: sentence lengths, vocab, NER counts (characters/locations)
* Save plots/screenshots for report

**Part 2 — Summarization (Wiki-style)**

* BART/PEGASUS summaries (chapter gist)
* NER list of main characters/locations
* RAG/wikipedia definitions for key terms
* Output a “Wiki page” JSON/Markdown

**Part 3 — Emotion Detection**

* GoEmotions RoBERTa integrated
* **Margin-based intensity** implemented
* **Keyword boosters** for storywords
* **Majority vote** smoothing (window 2–3)
* (Optional) Zero-shot scene tags (BART-MNLI)
* Eval on labeled set (Accuracy/F1 + confusion matrix)
* Export prosody per sentence

**Part 4 — Dialogue & Speaker Attribution**

* Dialogue detection (quotes)
* **Speaker attribution**: spaCy NER heuristic
* (Optional but ideal) BookNLP or fastcoref for stronger attribution across paragraphs
* **Character Registry** → stable voice presets (speaker\_id, base pitch/speed)

**Part 5 — TTS & Character Styling**

* Choose engine:
  + **Coqui XTTS v2** (local) *or*
  + **Cloud SSML** (Azure/Google/Polly)
* Map **emotion → prosody/style** (SSML or XTTS params)
* Render per-sentence WAVs using speaker\_profile and prosody
* Normalize voice loudness (target ~–16 LUFS)
* MOS mini-survey (3–5 listeners)

**Part 6 — BGM (Context-aware)**

* Build **mood** = fuse(emotion majority, scene tag)
* Pick track from folder by mood (or generate via MusicGen)
* **Auto-ducking** under speech (–8 to –12 dB), **crossfades**
* Normalize music (–26 to –24 LUFS under speech)
* Short interludes at chapter breaks

**Part 7 — Sync & Read-along**

* Word/line timestamps (from TTS if available, or heuristics)
* JSON for UI highlighting
* Gradio UI: play audio, highlight text, show wiki summary

**Part 8 — Evaluation & Reporting**

* ROUGE (summaries), F1 (NER & emotion), MOS (TTS), user pref (BGM)
* Ablations: with/without boosters; window 2 vs 3; emotion-only vs emotion+scene for BGM
* Roles table, timelines, and results slides