**1) What was the task?**

* Convert my Task 5 season narrative (Syracuse Women’s Lacrosse 2025) into a short, two-voice AI interview (Host ↔ Analyst).
* Emphasize the process over polish: approach, tools explored, failures, decisions, and a reproducible workflow.
* Include a clear ethics disclosure and ensure every spoken claim matches the validated data.

**2) My plan to approach it**

* Start with the validated facts from Task 5 and decide a tight story arc.
* Draft a human-sounding dialogue with brief Host prompts and concise Analyst answers (one stat + one implication).
* Use free, reliable tools to synthesize two distinct voices and assemble one clean MP3.
* Iterate with focused listening passes (cadence, clarity, facts).

**3) How I created the interview script from the narrative**

**Curated facts:** record, closest game, largest win, top scorer, assists leader, narrow losses, and the shots→goals correlation.

**Story arc:** opener → standout players → defining games → narrow losses → strategy (offense vs defense) → “game-changer” → lightning-round stat → recap.

**Dialogue style:** short, conversational Host questions; Analyst answers with one key number + what it means.

**TTS-friendly writing:** numbers in words (“nineteen,” “ten-and-nine,” “zero-point-nine-seven”); split long sentences; add light phonetics for tricky names.

**Ethics first:** disclosure placed as the first spoken line.

**4) Tools used (all free)**

**Google Colab —** hosted environment to run everything consistently.

**edge-tts —** neural text-to-speech with multiple voices, no API key required.

**pydub —** simple audio assembly (concatenate clips, insert pauses, normalize loudness).

**FFmpeg —** backend encoder/decoder used by pydub.

**5) Workflow (end-to-end)**

* Extract 8–12 essential facts from the narrative and verify each one.
* **Draft the dialogue:** disclosure → Q&A sequence → lightning round → recap.
* Select two voices (contrast in tone: warm/curious Host, calm/precise Analyst); test 1–2 lines.
* Synthesize per line (not one long pass) so pacing and re-takes are easy.
* Name clips consistently (e.g., 01\_HOST, 02\_ANALYST, …) to preserve order.
* Stitch clips in sequence; insert ~650–700 ms pauses to mimic turn-taking.
* Normalize loudness so there are no abrupt level jumps between lines/speakers.
* Export a rough MP3 and do three listening passes: cadence, clarity, facts.
* **Targeted fixes:** adjust pause length, rewrite mispronounced words/numbers, re-synthesize only the affected lines.
* Re-export final MP3 and confirm the disclosure and all facts are correct.

**6) Failures and challenges (and how I fixed them)**

**TTS library friction:** initial option didn’t match the runtime / version pinning failed → Switched to edge-tts, which worked reliably and is free.

**Notebook async errors:** “event loop already running” when using the wrong pattern → Used top-level await pattern; avoided asyncio.run() in notebooks.

**Pacing felt rushed:** pauses were too short → Increased inter-line gap in ~100–150 ms steps until it sounded natural.

**Pronunciation issues (names/numbers):** some words sounded odd → Wrote numbers as words, simplified phrasing, and added light phonetic hints.

**7) How I succeeded and generated the audio file**

* Two distinct voices assigned to roles (Host vs Analyst) and synthesized one clip per line.
* Stitched clips in order with consistent pauses; normalized loudness for even levels.
* Iterated quickly by re-synthesizing only lines that needed fixes.
* Exported a clean, human-sounding MP3 that opens with a spoken disclosure.

**8) Quality assurance (focused listening passes)**

**Cadence pass:** adjust pause length to achieve natural back-and-forth.

**Clarity pass:** fix names/numbers via wording; ensure both voices are intelligible.

**Facts pass:** confirm every claim matches the validated data from Task 5.

**9) Ethics and disclosure**

**Spoken disclosure first:** AI-generated, academic use, no real individuals recorded, stats from prior analysis.

**No impersonation:** the Analyst is a generic, fictional voice.

**Data fidelity:** all statements trace back to validated numbers.

**10) Reproducibility**

**Simple rerun recipe:** final script → per-line synthesis (two voices) → stitch with fixed pauses → normalize → export → listen → tweak.

**Recorded choices:** exact voice names used, final pause duration, any wording changes for pronunciation, and the outcomes of each listening pass.

**11) Lessons learned**

* Per-line synthesis is superior for pacing and quick fixes versus one long TTS file.
* Short sentences + numbers as words significantly improve TTS naturalness.
* Small pause changes (±100–150 ms) have a big impact on human feel.
* Light normalization is usually enough; heavy effects aren’t necessary.

**12) Future improvements**

* Add light SSML for emphasis and micro-pauses (used sparingly).
* Apply gentle compression/limiting for a more broadcast-ready finish.
* Audition more voice pairs to optimize warmth, presence, and clarity.
* Explore a video variant (lip-sync avatar + stat cards) and generate subtitles (.srt) for accessibility.