

🚀 [Updated] 화자 식별 및 정체성 규명을 위한 종합 파이프라인

(Goal: 화자별 행동 패턴과 발언 내용을 교차 검증하여, 가장 정확한 이름/역할 부여)

Phase 1. 기반 데이터 구축 (Foundation)

목표: 화자 식별의 기초 자료인 '소리(Who)'와 '텍스트(What)'를 정렬합니다.

1. STT 결과 로더 (STT Loader)

- 팀원이 제공한 STT JSON 로드 및 구조 파싱.

2. 화자 분리 (Diarization with NeMo)

- 오디오에서 화자를 분리하여 SPEAKER_00, 01 등 고유 ID 부여.

3. 정밀 병합 (Precision Merge)

- 단순 시간 매칭을 넘어, 오버랩 구간 등을 고려해 STT 텍스트에 화자 ID를 매핑 (Merged Transcript).

Phase 2. 행동 기반 프로파일링 (Behavioral Profiling)

목표: "말하는 **방식(Form)**을 보니, 이 사람은 이런 역할을 하는 사람이다"라는 1차 단서 확보.

4. 하이브리드 규칙 엔진 (Structural Rule Engine)

- 발화 통계:** 점유율, 턴 빈도, 평균 길이 계산.
- 위치/순서:** 오프닝/클로징 멘트 담당 여부, 타인의 발화 직후 반응 속도 (Latency) 분석.
- 1차 추정:** 통계적 특징만으로 '진행자 추정', '청중 추정' 등의 구조적 태그 생성.

5. 일관성 제약 (Consistency Constraints)

- 물리적으로 불가능한 역할 배제 (예: "한마디도 안 했는데 진행자일 수 없음").

Phase 3. 내용 기반 신원 추적 (Content-Based Identity Tracking)

목표: "말하는 **내용(Context)**을 보니, 이 사람은 특정 전문성을 가진 누구다"라는 결정

적 증거 확보. (여기가 업데이트된 핵심입니다)

6. 화자별 하이브리드 필터링 (Speaker-wise Filtering) [Token Economy]

- 화자 식별에 도움이 안 되는 무의미한 추임새, 단순 호응을 제거하고 **화자의 정체성이 드러나는 '알짜 발화'**만 남깁니다.

7. 벡터화 및 화자 임베딩 (Vectorization & Speaker Embedding)

- 필터링된 문장들을 벡터 공간에 배치합니다.
- **핵심:** 단순히 문장만 찍는 게 아니라, **Speaker 01의 의미적 중심점 (Centroid)**이 어디에 위치하는지 파악합니다.

8. 주제-화자 일치성 분석 (Topic-Speaker Alignment)

- **군집화(Clustering):** 대화 내용을 주제별(예: '예산', '기술', '일정')로 묶습니다.
- **신원 추론:** "Speaker A는 '예산' 군집에, Speaker B는 '기술' 군집에 주로 위치한다"는 사실을 도출합니다.
- 이것이 Phase 2의 결과와 결합되면 **"예산 이야기를 주도하는(Phase 3) 진행자(Phase 2)" = "재무 팀장"**이라는 강력한 추론이 가능해집니다.

9. 동적 입장 변화 추적 (Dynamic Stance Tracking)

- 단일 회의 내에서 화자의 입장이 변했는지(반대 → 찬성)를 추적하여, 단순한 역할명이 아닌 '**설득된 의사결정권자**' 같은 구체적 닉네임 부여의 근거로 삼습니다.

Phase 4. 종합 신원 확정 (Identity Resolution & Tagging)

목표: Phase 2(행동)와 Phase 3(내용)의 단서를 종합하여 최종 이름/별명을 확정합니다.

10. LLM 보조 심층 감별 (LLM Final Judge)

- **Input:** Phase 2의 행동 태그 + Phase 3의 주요 발언 주제.
- **Prompt:** "Speaker A는 회의의 40%를 점유하며 진행자 패턴을 보였고 (Phase 2), 주로 '일정 관리'와 '마감 기한'에 대해 언급했다(Phase 3). 이 사람의 역할/직책을 한 단어로 정의하라."
- **Output:** "PM(프로젝트 매니저)", "팀장" 등 구체적 정체성 도출.

11. 최종 결과 생성 (Final Output Generation)

- 규명된 정체성(이름/역할)이 반영된 최종 JSON 및 자막 생성.

💡 변경된 포인트 확인

이제 ****Phase 3(군집화)****는 단순히 "무슨 이야기가 오갔나"를 요약하는 단계가 아니라, ****"누가 그 이야기를 했는가"**를 통해 그 사람의 정체를 밝혀내는 '수사 과정'으로 재정의되었습니다.

이 흐름이 의도하신 방향과 정확히 일치하나요? 그렇다면 바로 **Phase 1 (기반 구축) 코드 작성**부터 시작해 보겠습니다.