

생성형 AI 및 딥러닝 립싱크 기반

「VOD 강의 영상 자동 제작 및 초고속 일괄 통합 솔루션」

① 어떤 상황에서 AI 를 활용했나요? (배경 및 문제 정의)

- **대학원 비대면·VOD 원격 강의의 급증과 제작 부담:** 디지털 교육 환경 확산으로 매 학기 수십~수백 편의 대학원 및 학부 VOD 강의 영상을 제작해야 하는 상황입니다. 강의 영상마다 통일된 규격의 '표지 인트로(강의명, 주차별 주제, 교수 소개 및 오프닝 멘트)'를 삽입하는 것은 교육의 전문성과 신뢰도를 결정짓는 핵심 요소입니다.
- **기존 수작업 제작 방식의 과도한 시간 소모:** 전문 영상 편집 툴(Premiere, After Effects 등)을 사용하여 슬라이드 디자인, 음성 녹음, 싱크 조절, 렌더링, 본 강의 영상과의 결합을 수작업으로 진행할 경우 영상 1 편당 30 분~1 시간 이상이 소요되었습니다.
- **오디오/영상 녹화의 물리적 제약:** 매번 교수가 직접 스튜디오나 마이크 앞에서 오프닝 멘트를 반복 녹음해야 하므로, 주변 소음, 발음 오류, 컨디션 난조에 따른 품질 편차가 크게 발생했습니다.
- **반복적인 대량 일괄 작업의 비효율:** 한 학기 16 주차 분량(수십 개 클립)에 동일한 인트로를 결합할 때마다 발생하는 긴 인코딩 대기 시간 및 반복 파일 관리의 병목이 심각했습니다.

② 어떤 AI 를 어떻게 활용했나요? (기술 파이프라인 및 구현 방식)

- **초고음질 다국어 Neural TTS (Edge TTS 엔진):** 교수의 오프닝 인사말 대본을 입력받아 한국어·영어 등 13 개 언어의 고품질 신경망 성우 음성으로 실시간 변환합니다. 속도(0.8x~1.2x) 미세 조절과 ctypes WinMM 기반 실시간 프리뷰를 구현하여 사전 청취 검증이 가능합니다.
- **생성형 딥러닝 비전 립싱크 엔진 (Wav2Lip & Video-ReTalking):** 단 1 장의 교수 프로필 사진(정적 이미지)을 입력받아, 생성된 오디오 파형에 맞추어 실제 말하는 듯한 자연스러운 입모양과 얼굴 표정의 가상 아바타 비디오를 실시간 생성합니다. SFD 얼굴 인식 및 PyTorch 2.6+ 호환성 패치를 내장했습니다.
- **COM Automation & 4 배 슈퍼샘플링 그래픽 합성:** PowerPoint COM 개체 자동화를 통해 고화질 PPT 템플릿에 세미나 타이틀을 자동 타이포그래피 정렬(60pt Center) 후 1080p FHD 무손실

PNG 를 추출합니다. 아바타 영상의 테두리를 4 배 슈퍼샘플링 안티앨리어싱 둥근 모서리 마스크(alphamerge)로 배경 슬라이드 우하단에 이질감 없이 자연스럽게 합성합니다.

- **초고속 FFmpeg 멀티스레드 인코딩 및 일괄 결합(Batch Concat):** 방송 오프닝과 같은 1.5 초 딜레이 시퀀스를 부여하며, 선택된 수십 개의 본 강의 영상(MP4)들을 재인코딩 손실 없이 해상도(1080p)·프레임레이트(30fps)를 자동 통일하여 초고속 일괄 결합합니다.

③ 활용 결과 어떤 변화가 있었나요? (정량적·정성적 성과)

■ 정량적 성과 비교표

구분	AI 솔루션 도입 전 (수작업)	AI 솔루션 도입 후 (본 시스템)	개선 효과
인트로 제작 소요 시간	편당 약 30~45 분	편당 약 30 초~1 분 (원클릭)	97% 이상 시간 단축
16 주차 강의 일괄 결합	반나절 소요 (약 4~5 시간)	백그라운드 자동 일괄 처리 (수 분 내)	업무 피로도 제로화
제작 비용 및 장비	고가 녹음 장비 및 전담 편집자 필요	일반 PC 에서 사진 1 장과 대본만으로 완성	예산 절감 및 상시 제작
강의 콘텐츠 완성도	회차별 음질 편차 및 노이즈 발생	정제된 Neural Voice & 통일된 템플릿	학습 몰입도 극대화

- **정성적 변화 및 교육 집중도 향상:** 강의자는 '영상 편집 및 녹음'이라는 부가적 스트레스에서 완전히 해방되어 순수 강의 연구와 교육 내용 고도화에 전념할 수 있게 되었습니다.
- **글로벌 학습자 접근성 개선:** 다국어(영어/한국어) 음성 인트로를 손쉽게 교체 생성하여 외국인 유학생 및 온라인 글로벌 세미나 대상 강의 접근성을 획기적으로 향상시켰습니다.

④ 나만의 방식 또는 개선 포인트는 무엇인가요? (차별성 및 독창적 노하우)

- **비-개발자도 즉시 쓰는 무결점 원클릭 자동 환경 구축:** 복잡한 PyTorch, FFmpeg, 모델 가중치(Checkpoint), 깃허브 의존 패키지를 사용자가 직접 설치할 필요 없이, 프로그램 실행 시

누락된 가중치(85MB SFD 모델 등)를 자동 감지하여 백그라운드 다운로드 및 패치를 수행하도록 설계했습니다.

- **OpenCV 비-ASCII(한글 경로) 충돌 방지 및 EXIF 회전 안전 파이프라인:** Windows 환경에서 한글 폴더 경로로 인해 OpenCV 및 딥러닝 라이브러리가 크래시되는 고질적인 문제를 ASCII 안전 해시 임시 격리 파이프라인(_sanitize_path_for_opencv)으로 완벽 차단하고, 스마트폰 세로 촬영 사진의 EXIF 메타데이터를 자동 보정하여 얼굴 감지 실패율 제로를 달성했습니다.
- **우아한 Fallback(장애 복구) 아키텍처:** GPU 메모리 부족이나 사진 내 얼굴 인식 실패 시 시스템이 중단되는 대신, 안전하게 정지 사진 마스킹 영상으로 자동 전환(Fallback)되어 작업이 100% 정상 완료되도록 예외 복구 구조를 확립했습니다.
- **커스텀 Tkinter(CustomTkinter) 기반 상용 등급 프리미엄 GUI:** 다크/라이트 테마, 검색 가능한 성우 팝업 필터링(Search & Filter), 실시간 프로그레스 바(0~100%) 지원 등 사용자 경험을 극대화했습니다.

⑤ 다른 사람도 따라 할 수 있나요? (확산 가능성 및 범용성)

- **100% 오픈소스 및 무료 AI 기반 (No Paid API Key):** 유료 구독형 SaaS 를 쓰지 않고, 무료로 제공되는 오픈 모델과 Edge TTS 를 결합하여 개인 교수자, 초·중·고 교사, 공공기관 교육 담당자 누구나 비용 부담 없이 영구 무료로 사용할 수 있습니다.
- **초보자도 클릭 몇 번으로 가능한 직관적 워크플로우:** '[제목 입력] → [사진/템플릿 선택] → [대본 입력] → [제작 버튼 클릭]'의 직관적인 4 단계 프로세스로 구성되어 있어, 컴퓨터 활용 능력이 낮은 고령 교수자나 강사도 5 분 만에 사용법을 숙지할 수 있습니다.
- **타 분야 및 공공 교육으로의 높은 확장성:** 대학 강의뿐만 아니라 지자체 시민 강좌, 공공기관 사내 직무 교육 VOD, 정부 정책 홍보 쇼츠/클립 인트로, 유튜브 지식 채널 등 표준 인트로 영상이 필요한 모든 공공 및 민간 교육 분야에 템플릿만 교체하여 즉시 확산 적용 가능합니다.