

비정형의 정형화: 데이터 기반의 교육장학금 심사 자동화

1,800건의 비정형 코퍼스(Corpus) 분석을 통한
통제 어휘(Controlled Vocabulary) 구축 및 행정 지식의 디지털 자산화

TECH STACK

Python Playwright Gemini 2.0 Flash

KEY ACHIEVEMENT

1,800건 데이터 마이닝 &
통제 어휘(Codebook) 구축

ROLE

System Architect &
Knowledge Engineer

CURRENT REALITY

인간을 지치게 만드는 '숨은 그림 찾기'

수백 장의 영수증과 성적표,
깨알 같은 숫자들의 미로 속에서
우리의 인내심은 한계를 시험받습니다.

- 반복 업무로 인한 직무 만족도 저하



매 학기 1,800건의 비효율적 검증

연간 약 4,000건의 서류를 수작업으로 검토하는 구조적 한계



극심한 노동 집약성

영수증 승인번호, 금액, 날짜 등 5개 항목을
육안으로 일일이 대조 (단순 반복)



Human Error 리스크

장시간 검토 시 피로 누적으로 인한
오판정 및 필수 항목 누락 발생 불가피



비효율적 인수인계

담당자 변경 시마다 복잡한 검수 규정을
처음부터 재교육해야 하는 비용 발생

50 초
건당 소요 시간

1,800 건+
학기당 처리 (연 4,000건)

High
피로도 레벨

RPA의 손과 AI의 두뇌를 결합한 '디지털 직원'

비정형 데이터를 스스로 수집하고 판단하는 지능형 자동화 파이프라인



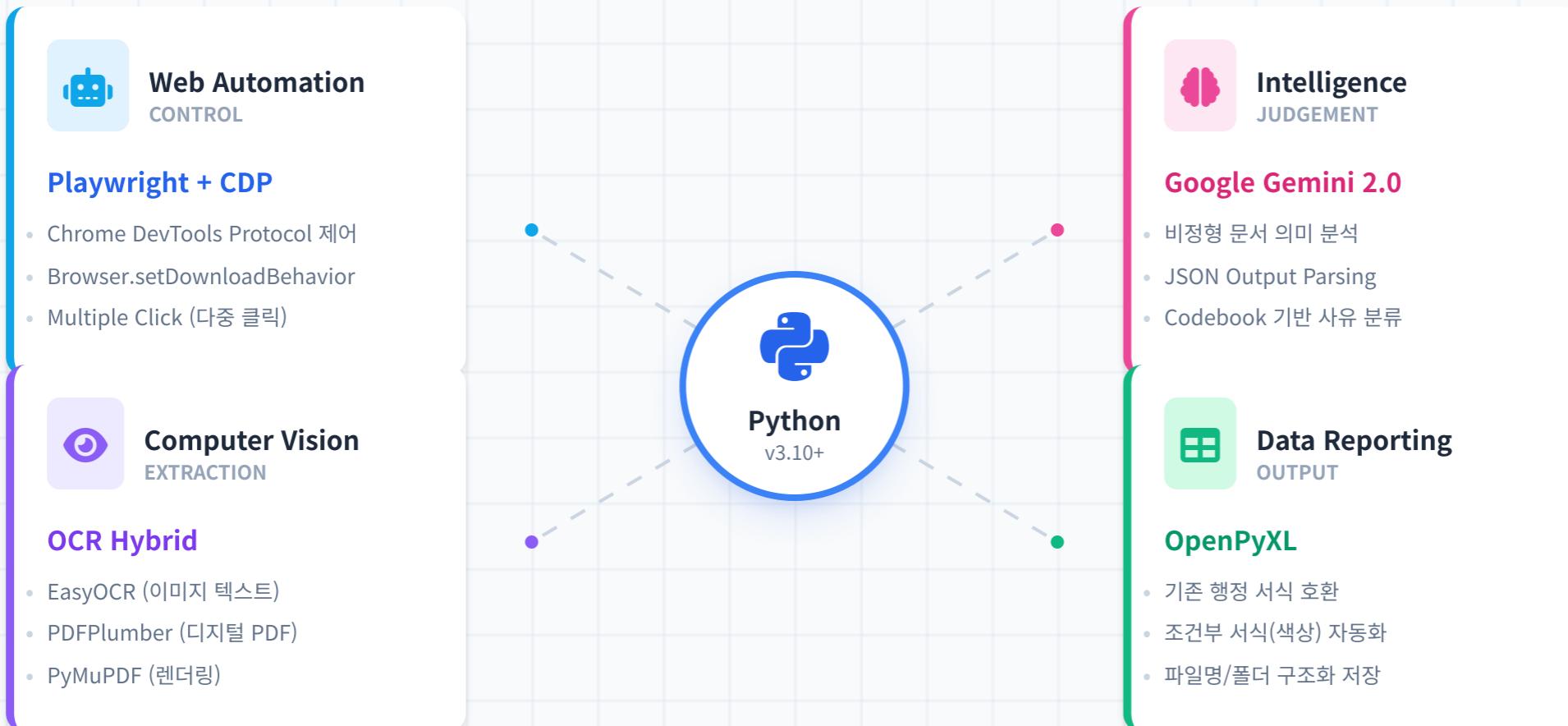
✔ Human-In-The-Loop (최종 승인만 사람 개입)

• ⚡ Batch Processing (속도 최적화)

• 🛡 Secure Masking (개인정보 보호)

최적의 도구를 연결하여 완성한 견고한 파이프라인

Architecture & Tech Stack Integration



Batch Inference

Fast-Fail Strategy

Self-Healing Logic

시스템의 한계를 기술적 통찰로 돌파하다

Troubleshooting: Engineering Hurdles & Solutions

🚫 ACCESS DENIED

Browser prevents auto-download popups.

✅ CDP 강제 다운로드

Chrome DevTools Protocol을 직접 주입하여 브라우저 보안 정책을 우회하고 자동 저장 강제.

`Browser.setDownloadBehavior`

⚠ DOM ERROR

DOMException: Element is detached from document.

</> JS Injection 제어

Python 객체 참조 대신, 브라우저 내부에 JavaScript를 직접 실행하여 동적 요소 제어.

`page.evaluate("el => el.click()")`

☒ FILE ERROR

Saved as .jsp or _____.png (Obfuscated)

💡 Magic Number 복원

파일의 바이너리 헤더를 읽어 실제 포맷을 판별하고 확장자를 자동 복구 (.jsp → .pdf).

`Hex: %PDF, 89 50 4E 47`

⌚ TIMEOUT

Click event ignored due to Grid UI lag or overlapping.

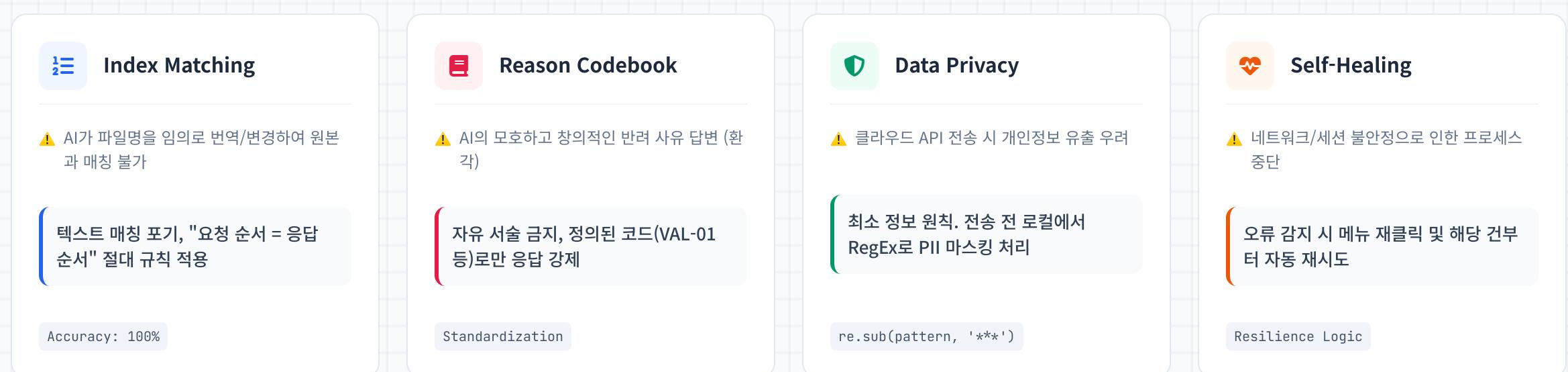
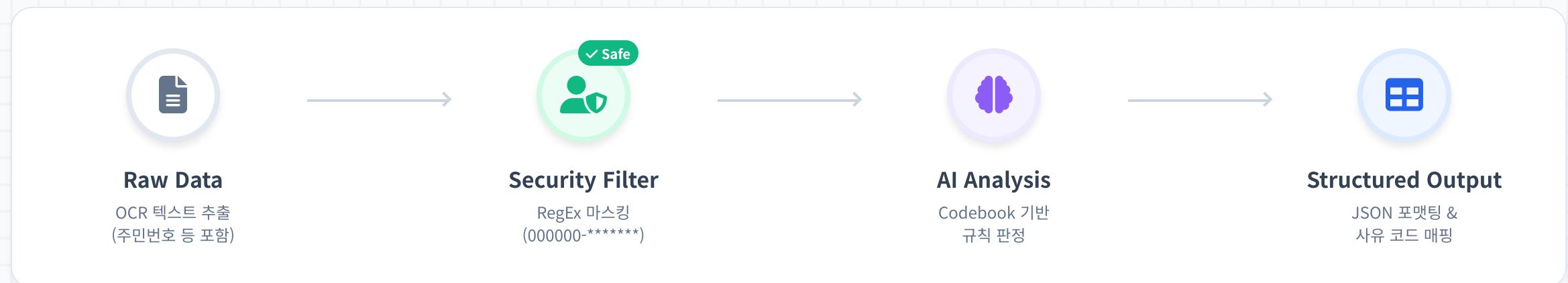
▣ 다중 클릭 전략

단일 클릭 실패를 대비한 3중 복합 클릭 전략과 mtime 기반의 최신 파일 후킹(Hooking).

1. UI Click → 2. JS Click → 3. Tag Click

환각은 줄이고, 신뢰는 높이고

Troubleshooting 2: Logic Accuracy & Security Strategy



데이터로 증명하는 정확도와 판별력

Result Data & Case Study (Accuracy: 99.9%)

검증 결과 리포트 (Excel)

Total: 500 rows

Status	Student	Type	Amount Match (Req vs Doc)	Result Code
✓ PASS	김민수 1234567	TOEIC	52,500 = 52,500	MATCHED
✗ REJECT	홍길동 7654321	인프런	69,300 ≠ 408,100	VAL-01
✗ REJECT	이영희 9876543	YBM 수강	120,000 = 120,000	VAL-03
✓ PASS	박철수	OPIC	78,100 = 78,100	MATCHED
✓ PASS	최지민	TOEIC	52,500 = 52,500	MATCHED

AI Analysis Log

ID: 7654321

VALIDATION LOGIC: AMOUNT CHECK

신청 금액 (DB) 69,300 ≠ 영수증 추출 (OCR) 408,100

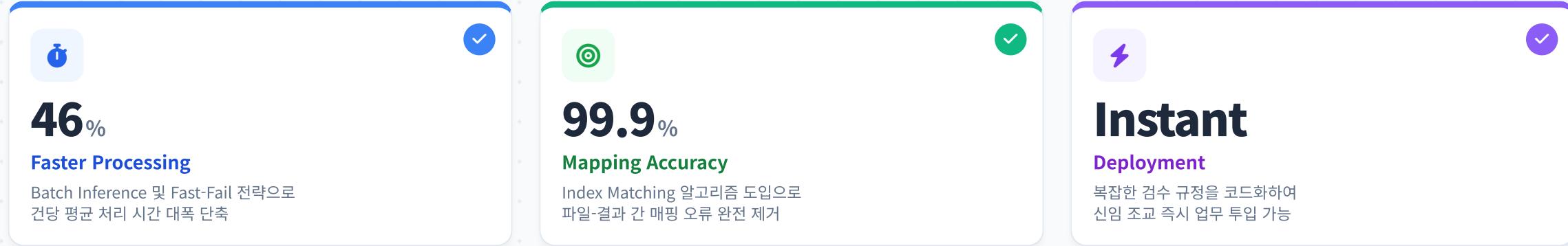
```
{  
    "student_name": "홍길동",  
    "doc_type": "영수증/매출전표",  
    "extracted_amount": 408100,  
    "target_amount": 69300,  
    "is_match": false,  
    "verdict": "Reject",  
    "reason_code": "VAL-01",  
    "comment": "신청 금액과 영수증 결제 금액이 상이함."  
}
```

VAL-01
부적합

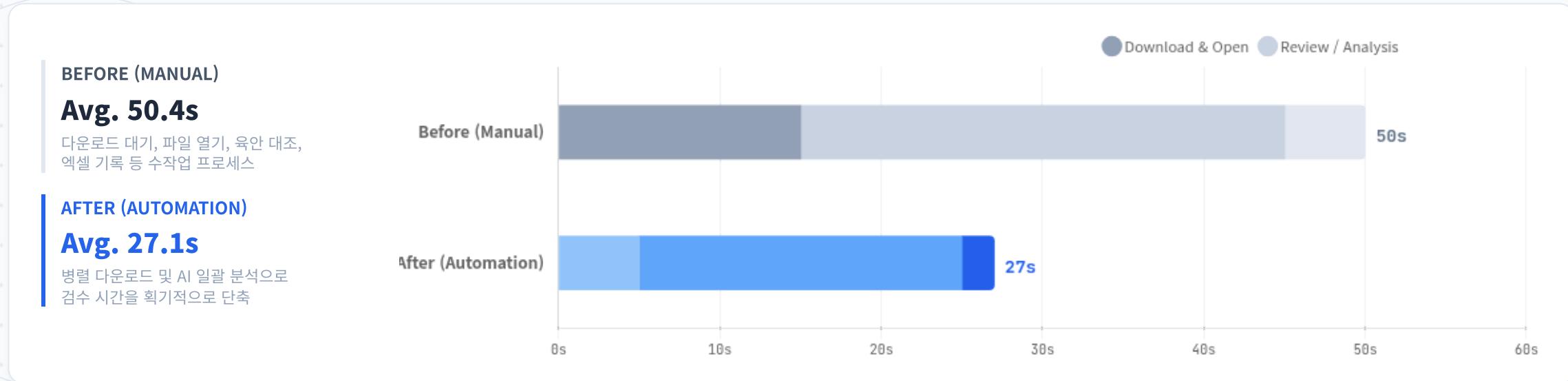
50초 → 27초, 46% 시간 단축의 의미

Performance Metrics & Operational Impact

CORE CAPABILITIES SECURED



PROCESS TIME ANALYSIS



비즈니스 가치와 ROI 분석

Business Value & Current Achievement

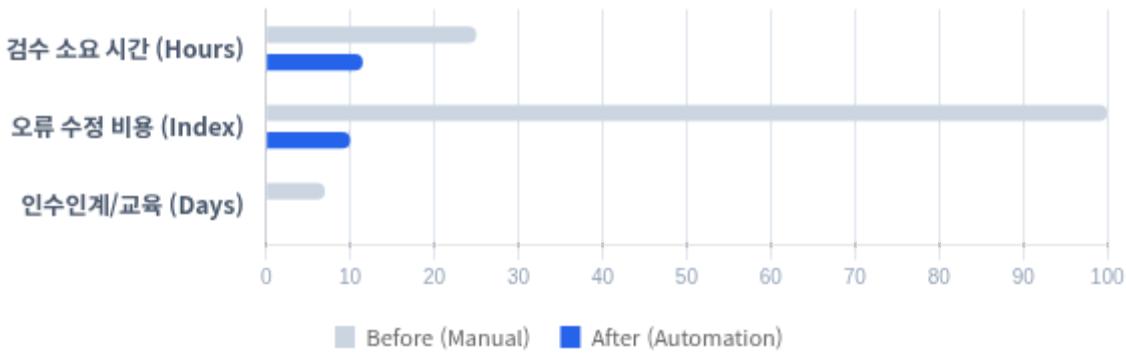
Operational Efficiency (1,800건 기준)

HUMAN RESOURCE EFFICIENCY

↓ 11.5 Hours

학기당 검증 업무 투입 시간 절감 (1,800건 기준)
단순 반복 → 고부가가치(상담/기획) 업무로 전환

업무 효율성 개선 지표 (Efficiency Metrics)



- 민원 감소: 반려 사유 명확화로 학생 문의 30% 감소
- 재검수 비용 0: Human Error로 인한 사후 처리 비용 제거
- 인수인계 비용 0: 업무 매뉴얼 학습 불필요 (System Driven)

Current Milestone

현재 구축 완료된 자동화 시스템 범위

✓ PHASE 1 : COMPLETED

교육장학금 지급 심사 (Scholarship)

신청서, 영수증, 성적표 등 다양한 증빙 서류에 대한
자동 수집(RPA) 및 적합성 검증(AI) 파이프라인 구축 완료.

- 표준 모델 확립: 비정형 문서(영수증) 판독을 위한 OCR+LLM 하이브리드 검증 로직 완성
- 데이터 정합성 확보: 1,800건 데이터 기반의 반려 사유 코드북(Codebook) 적용
- 업무 즉시 투입: 신규 조교 및 담당자가 별도 교육 없이 즉시 운영 가능

CORE CAPABILITIES SECURED

Auto Download

Text Extraction

Logic Check

Excel Reporting

27초의 혁신, 그리고 새로운 가치의 발견

Conclusion & Vision: From Automation to Transformation

27s

AVG. PROCESS TIME

46%

EFFICIENCY GAIN



HANNOVER COST

"단순 반복 업무의 종말, 창의적 업무의 시작"

Scholarship_Automation System has redefined the workflow.

DEVELOPMENT ROADMAP

✓ Completed

⟳ In Progress

Planned



Phase 1: Stabilization

핵심 기능(다운로드, AI 검증) 구현
레거시 시스템 제어 및 안정성 확보



Phase 2: Standardization

통제 어휘(Controlled Vocabulary) 구축
반려 사유 코드화 및 DB 통계 기반 마련



Phase 3: Deployment

실행 파일(.exe) 패키징 및 배포
API 키 관리 UI 제공으로 접근성 강화



Human-AI Collaboration Vision

AI는 **지치지 않는 속도**로 데이터를 처리하고, 인간은 **따뜻한 통찰**로 학생을 지원합니다.
우리가 만든 것은 단순한 봇이 아닌, 행정 업무의 **새로운 동료**입니다.