

M.A.R.S. 2025

SNUBH 의무기록 생성 데이터톤

Medical **A**uto-documentation with **R**eal-world **S**tructuring:
LLM Clinical Note Generation Challenge

예선 결과 보고서 [MediX]

주최/주관



후원



1. 팀 소개 (MediX)

팀 개요 및 역할 분담 : **의료AI의 혁명을 꿈꾸는 학부 3학년 개인 참가자입니다!**

ABOUT ME

Hi, I'm Seoyoung Oh. I have a strong interest in Healthcare and Medical AI, and I aspire to drive innovation in the medical domain through the power of data and artificial intelligence. Among the many fields of AI and data science, I find healthcare to be the most meaningful and impactful, as it directly relates to people's lives and well-being. My academic focus is centered on exploring how AI and data-driven solutions can address real world challenges in medicine, and I'm actively seeking opportunities to grow through research, interdisciplinary collaboration, and practical applications in this field.

수상 실적

- [2023] 국민대학교 기후변화대응 비즈니스 아이디어 공모전 장려상
- [2024] 제2회 국민대학교 AI빅데이터 분석 경진대회 우수상
- [2025] 국민대학교 경영대학 창업경진대회 최우수상
- [2025] 춘천시 생령형 AI 아이디어 경진대회 우수상
- [2025] 제9회 디지털 헬스케어 MEDICAL HACK 최우수상 (부산대학교총장상)
- [2025] 건양대학교의료원 KHD (Konyang Health Datathon) 최우수상

창업 실적 | TEAM. MEDIX (대표 AI 개발자)

- [2025] 서울 AI 허브 - Seoul AI Young Track (SAY 트랙) | 최종 선정 및 수료
- [2025] 연세대학교 Y-Compass 학생창업팀 | 최종 선정 (A등급 수료)
- [2025] 부산대병원 주최 - 제 9회 디지털 헬스케어 MEDICAL HACK 2025 | 최우수상
- [2025] 학생창업유망팀 U 300+ | 최종 선정 및 활동 중

주요 프로젝트

- [2024] 카카오톡 챗봇 기반 인지증재치료 서비스 제작
- [2025] AI 기반 임신 주수별 개인 맞춤 약물 섭취 금기 모니터링 서비스 제작
- [2025] EMR 기반 의료진 간의 인수인계를 위한 기반 변동사항 요약 시스템 제작 → 창업팀 활동
- [2025] 측면두부규격방사선 분석 기반 부정교합 진단 및 랜드마크 자동 검출 AI 모델 제작 → KHD 수상 프로젝트
- [2025] An Explainable Framework for Diabetic Retinopathy Classification Using Fundus Image Preprocessing and Visual Explanations → 2025 대한의료인공지능학회 정기학술대회 포스터 제출 완료
- [2025] PPG 기반 호흡수 추정의 신뢰도 향상 SQI·SNR 게이팅과 Conformal 예측구간 → 2025 대한의료정보학회 추계학술대회 포스터 제출 완료
- [2025] Surgical Skill Assessment Using Kinematic Features and Deep Learning on the JIGSAWS Dataset → 2025 의료메타버스학회 추계학술대회 포스터 채택
- [2025] Machine Learning Models for Predicting Operative Duration in Assisted Reproductive → JMIR 논문 투고 진행 중

CONTACT

E-mail : 5seo0_oh@kookmin.ac.kr
Linkedin: <https://www.linkedin.com/in/5seo0/>
GitHub : <https://github.com/5seoyoung>
Blog : <https://5seo0.tistory.com/>

학력

- 안양외국어고등학교 졸업 (2020.03 ~ 2023.02)
- 국민대학교 AI빅데이터융합경영학과 재학(2023.03 ~)
- 국민대학교 인공지능학부 부전공(2025.03 ~)

개인 참가이지만, 팀처럼 체계적으로!

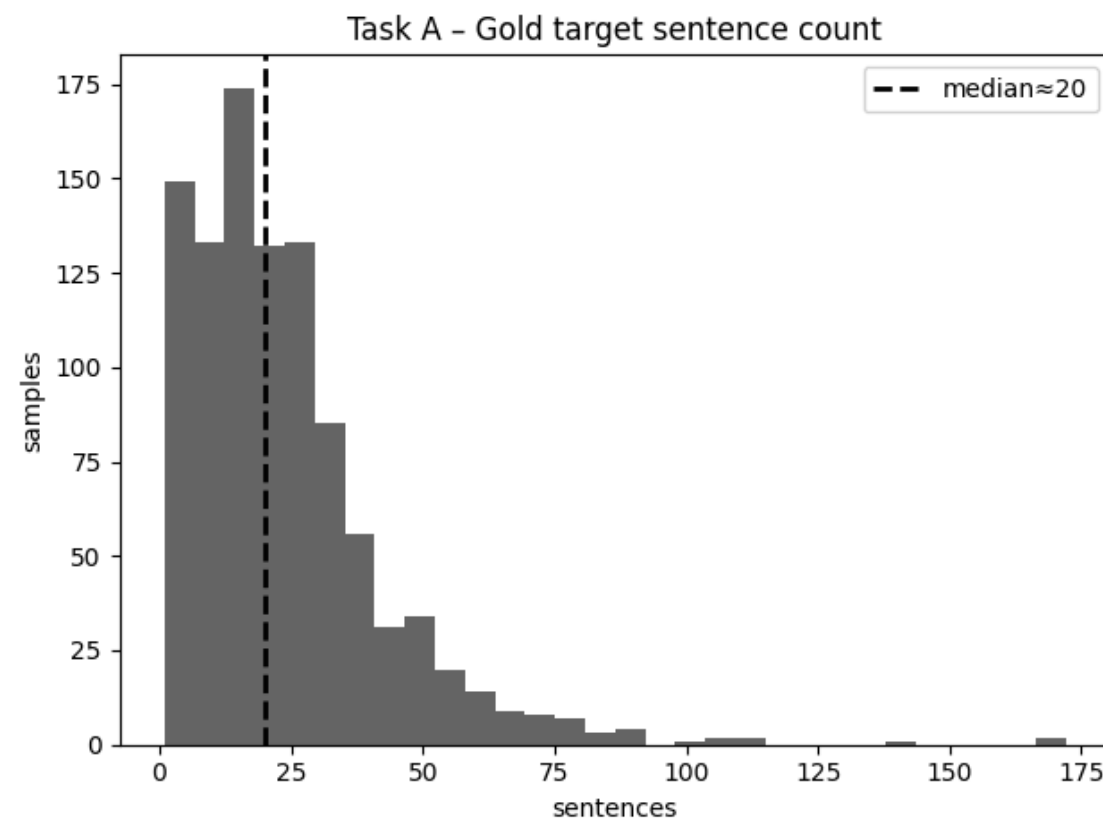
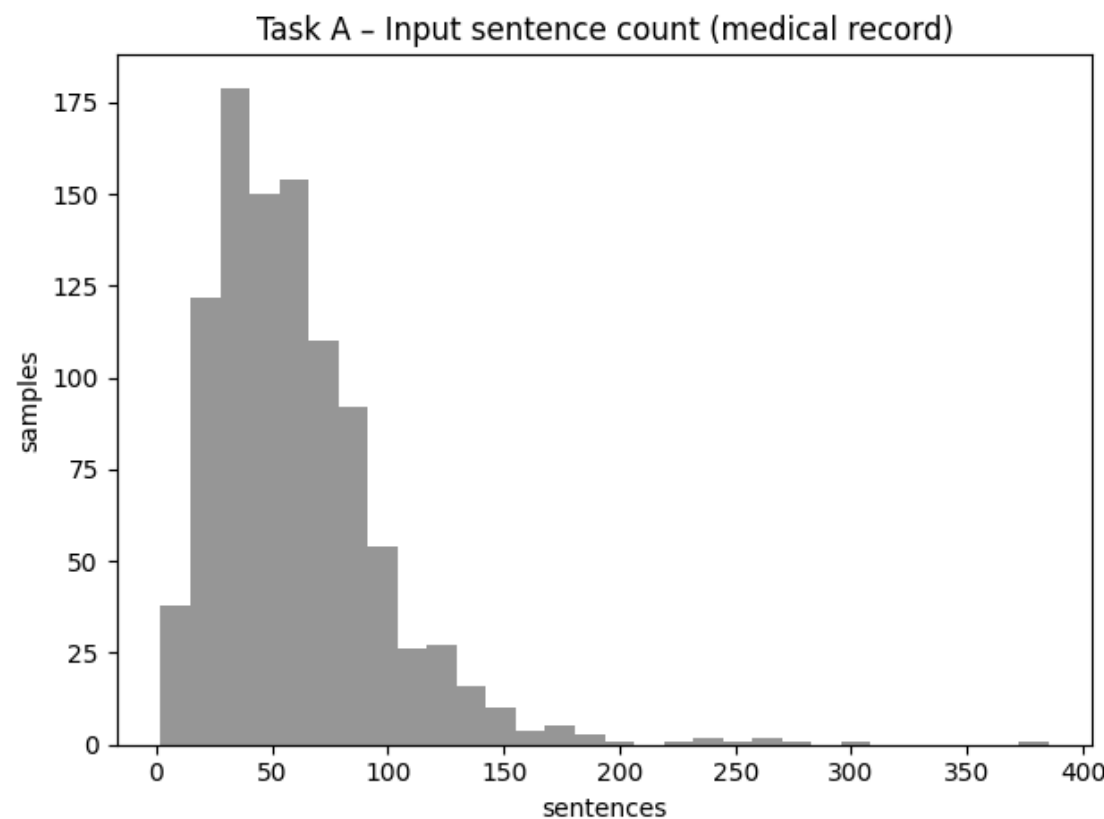
- 프롬프트 설계 → Task별 프롬프트 디자인 및 후처리 규칙 개발
- 데이터 분석 → Train 분포 분석, 시각화 및 근거 제공
- 보고서/발표 → 전략·실험결과를 정리하고 임상적 의미 해석

2. 문제 정의 & 접근 개요

2.1 문제 정의

Task A: Brief Hospital Course 생성

- **데이터**: MIMIC-IV-Note 내 퇴원요약(discharge summary), 입퇴원 정보, Chief Complaint
- **업무 맥락**: 실제 임상에서는 퇴원요약이 환자 경과·치료·합병증·퇴원계획을 정리하는 핵심 문서로 사용됨
- **과제 의미**:
 - 단순 요약이 아니라 연대기적 흐름(입원→진단→치료→합병증→퇴원)을 유지해야 함
 - EMR/EHR 시스템에서 자동화되면, 의료진 문서작성 부담 감소 + 임상 커뮤니케이션 품질 향상 기대



**정답 중앙값 ≈ 20 vs 모델 출력 5-9
→ 길이 정렬 필요성 강조**

- 정답 문서: 환자 입원→진단→치료→합병증→퇴원까지 연대기적 흐름 유지
- 실제 임상: 퇴원 요약은 EMR/EHR에서 가장 중요한 문서로, **의사·간호사·보험청구 모두 사용**
- 모델 출력은 평균 5-9문장으로 정답 대비 과도 축약 → **정보 손실**

요약 길이 불일치 문제



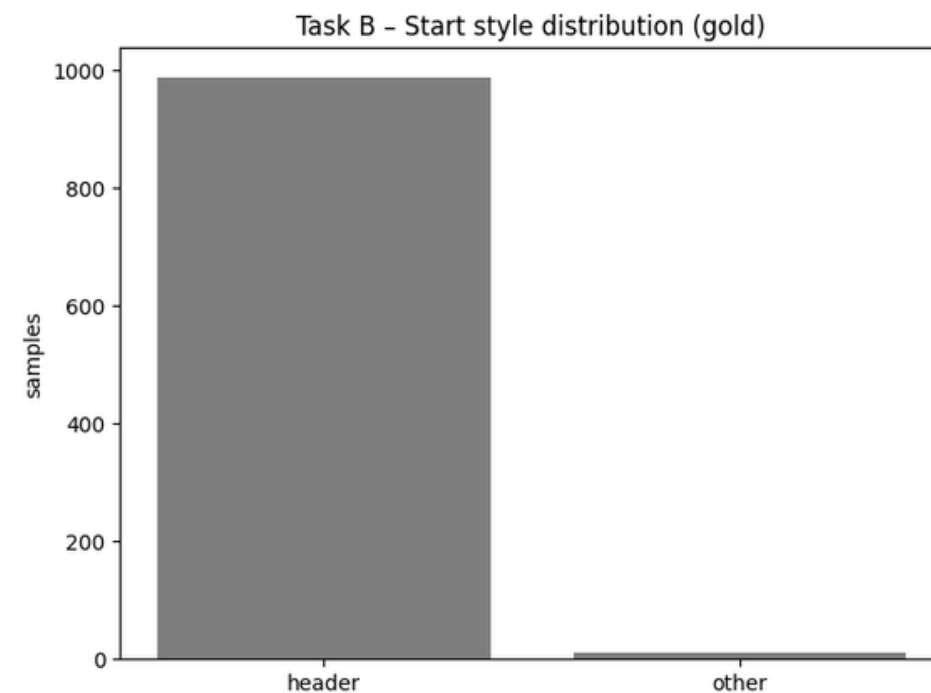
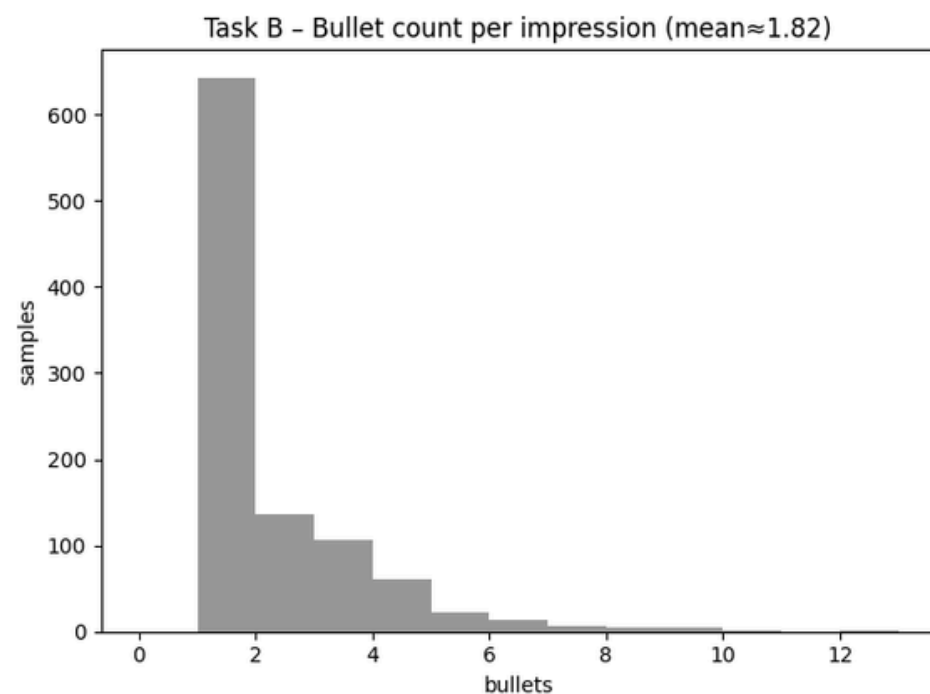
입력/정답 문장수 분포 비교: 정답 중앙값 20 vs 모델 출력 5-9

2. 문제 정의 & 접근 개요

2.1 문제 정의

Task B: Radiology Impression 생성

- **데이터**: 방사선 검사 결과(Findings)
- **업무 맥락**: 실제 판독(report)은 Findings(관찰된 사실) + Impression(의사의 요약적 해석)으로 구성됨
- **가이드라인**: RSNA·ACR 등은 Impression을
 - a. 짧고 명확한 불릿/번호형 리스트로,
 - b. 임상적 의미(존재/부재) 위주로 작성할 것을 권고
- **과제 의미**:
 - 불필요한 권고문구(“추가 검사 권유” 등)를 최소화해야 채점 LLM이 평가하는 Conciseness·Clinical Clarity 점수가 올라감
 - 실제 임상에서는 Impression이 의사결정·후속 진료의 가장 중요한 단락



정답 중앙값 ≈ 20 vs 모델 출력 5-9
 평균 불릿 ≈ 5 , 99% 헤더 시작 → 형식 정렬 필요

- RSNA 가이드라인: Impression은 ‘헤더 + 번호형 리스트’로 핵심만 작성
- 정답 분포: 평균 불릿수 ≈ 5 , 99%가 IMPRESSION:으로 시작
- 모델 초기 버전: 1-3 불릿, 헤더 제거 → 형식 불일치로 BERTScore 손해
- 개선안: IMPRESSION: 헤더 강제 + 3-6 불릿 정규화 + 권고문구 제거

보고서 형식적 특성



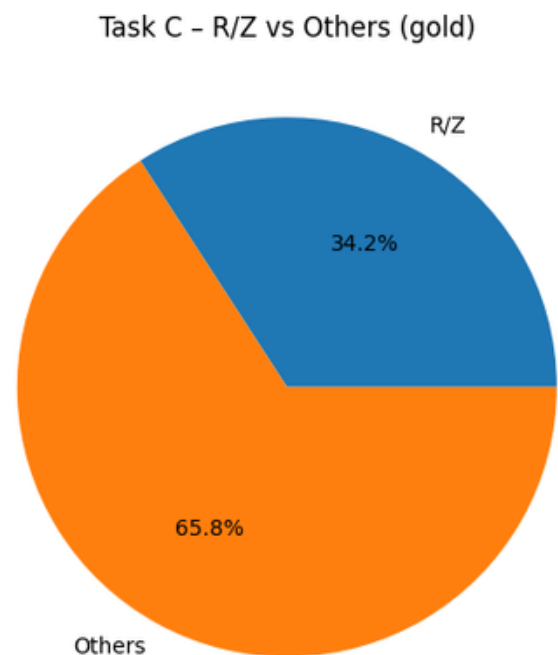
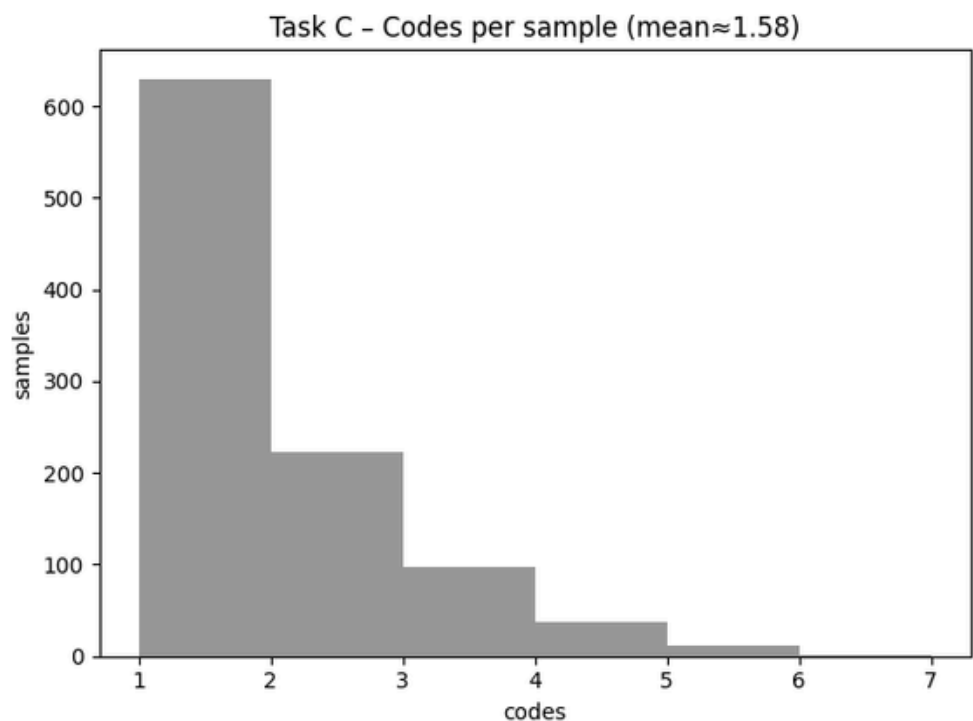
Impression은 평균 5불릿, 99% IMPRESSION: 헤더 시작

2. 문제 정의 & 접근 개요

2.1 문제 정의

Task C: ICD 코드 10 예측

- **데이터:** 입원 경과 요약(hospital course)
- **업무 맥락:** ICD-10은 보험·통계·임상 연구의 표준 진단 코드 체계
- **공식 지침:**
 - 주진단(primary diagnosis) 우선 → 가장 구체적 코드 선택
 - 확정 진단이 있으면 R-코드(증상), Z-코드(기타)로 대체 금지
 - 과코딩 금지: 불필요한 코드 여러 개 기입 시 오히려 패널티
- **과제 의미:**
 - 정확한 코드 예측은 Macro-F1 + 계층적 점수(hierarchical match)로 측정됨
 - 실제 임상/보험 청구에서 코딩 정확도·특이성은 환자 안전·재정에도 직접 연결



평균 1.58개, R/Z 코드 과다
→ 과코딩 방지 전략 필요

- ICD 지침: 주진단(primary) 우선, R/Z 코드는 확정 진단 있으면 **대체 금지**
- 정답 분포: 평균 1.58개 → 모델이 3개 이상 내면 **과코딩 패널티**
- R/Z 코드 비중 34% → 필터링 금지, 대신 “**최대 2개 제한 + 중복 제거 + 점 제거**” 전략 적용

코드 분포의 제약 조건



평균 1.58개, R/Z 코드 과다 → 과코딩 방지 필요

2. 문제 정의 & 접근 개요

2.2 접근 개요

형식 정렬(format alignment)



Task A

- 정답 분포를 보면 타겟 요약의 중앙값이 약 20문장 수준으로 길다 → **12-20문장 가이드**로 확장해 BERTScore/LLM 품질을 확보 (길이 불일치가 의미유사도·가독성 점수를 깎는 문제 완화)
- **BERTScore**는 문맥 임베딩 유사도 기반 지표로, **표현 다양성**은 허용하지만 **핵심 정보 보존**에 민감하다

Ashok Umlana, Pruthwik Mishra, Tathagato Roy, and Rahul Mishra. 2024. Controllable Text Summarization: Unraveling Challenges, Approaches, and Prospects - A Survey. In Findings of the Association for Computational Linguistics: ACL 2024, pages 1603–1623, Bangkok, Thailand. Association for Computational Linguistics.

Task B

- “IMPRESSION:” 헤더 + 번호형 3-6 불릿으로 표준화(전문학회·교육 자료의 권고와 일치)

Task C

- 대문자·점 제거(I10, E119, W19XXA 형태), 중복 제거·원순서 유지, 최대 2개만 반환
→ 정답 분포와 ICD 지침의 “과코딩 금지”에 부합)

규칙 기반 후처리(간결성·안전성)



Task B
권고문구 제거

- Impression은 진단 요약 중심이 원칙. 권고·추가검사 문구는 필요할 수 있으나, 대회 채점 프롬프트가 간결·임상요약을 중시하므로 비핵심 권고를 정리해 가독성·간결성 점수를 노린다 (번호형 핵심요지 유지)

공정성 카드

- 입력에 Gender/Age 메타 한 줄만 덧붙여 정보 손실 없이 subgroup 간 성능편차(“disparate impact”)를 모니터링
- “80% rule” 등 불평등 지표는 산업 표준으로 통용된다

실행시간 제약 대응



- 모델은 Llama-3.1-8B로 통일(속도/안정성). 호출량 제한·10분 내 실행은 대회 공지를 준수

Johnson, A.E.W., Bulgarelli, L., Shen, L. et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 10, 1 (2023). <https://doi.org/10.1038/s41597-022-01899-x>

3. 프롬프트 설계 방법

3.1 설계 원칙 : 분포·평가 지표 기반 설계 3원칙

분포 정렬
(Distribution Alignment)

- Train 분포(문장수·불릿수·코드수)를 정밀 분석 → 프롬프트에서 출력 형식을 **정답 분포에** 근접시키도록 설계
- 예: Task A 정답 중앙값 \approx 20문장 → 12-20문장 버전 설계



평가 지표 타겟팅
(Metric-Oriented Prompting)

- **정량(BERTScore/F1)**: 의미 일치 + 표면 형식 중요 → 형식 정렬(IMPRESSON: 헤더, 문장 길이 맞춤)
- **LLM 품질(Clinical/Concise)**: 권고문구 제거, 임상 톤 유지
- **공정성**: 성별·연령 정보 최소 첨부(편차 완화)



안정성·실행가능성
(Operational Safety)

- 모델/토큰 수를 제약해 **10분 제한 내 확실히 동작**
- 후처리로 형식 오류/빈 출력 대비 → “안전망 코드” 삽입

➡ **분포-정렬, 지표-타겟팅, 안정성 확보’ 이 세 가지 프레임으로 모든 프롬프트를 설계했습니다**

3. 프롬프트 설계 방법

3.2 TASK별 설계 전략 : TASK A/B/C – 맞춤형 전략

| | Train 분석 인사이트 | 전략적 설계 포인트 | 평가 지표 기대 효과 |
|--------|----------------------------|--|--|
| Task A | 중앙값 ≈20문장, IQR 넓음 | 12-20문장 가이드(정렬) vs 5-9문장 가이드(속도·간결) | BERTScore ↑, Conciseness 안정 |
| Task B | 평균 불릿 ≈5, 헤더 존재율 99% | IMPRESSION: 헤더 + 3-6 불릿, 권고문구 제거 | 정량 점수 ↑ (형식 일치), Concise/Clinical ↑ |
| Task C | 평균 코드수 ≈1.58, 점 없음, R/Z 다수 | 대문자·점제거·중복제거·≤2개 제한 | Macro-F1 ↑, 과코딩 방지, 형식 페널티 ↓ |

➡ 각 TASK마다 ‘정답 분포 + 지표 민감도’를 근거로 전략을 달리했습니다
단순히 LLM에 맡기지 않고, 후처리로 형식까지 보정했습니다

3. 프롬프트 설계 방법

3.2 TASK별 설계 전략 : TASK A/B/C – 맞춤형 전략

Task A

Train 분석 인사이트

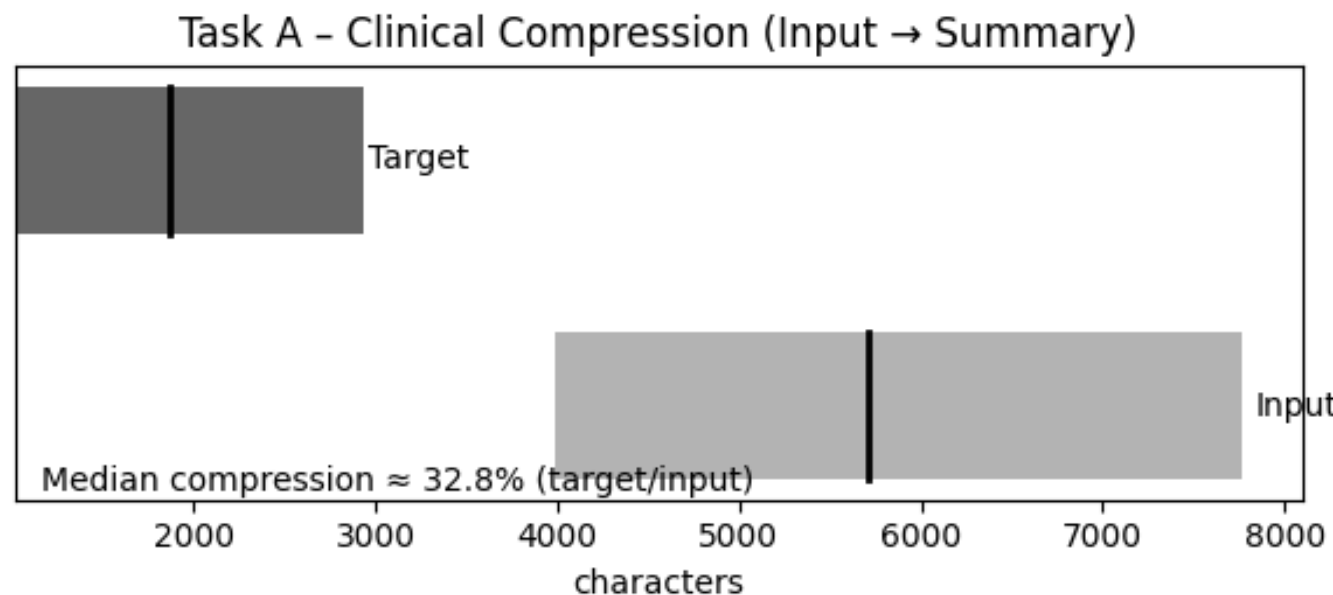
중앙값 ≈ 20문장, IQR 넓음

전략적 설계 포인트

12-20문장 가이드(정렬) vs
5-9문장 가이드(속도·간결)

평가 지표 기대 효과

정량 점수 ↑ (형식 일치), Concise/Clinical ↑



TASK A: 임상 압축 퍼널



과도한 압축은 핵심정보 손실 → 길이 제어 프롬프트 필요

- 역할 부여: “You are a hospitalist” → 전문성 있는 임상 톤을 유도
- 연대기 구조 힌트: “presentation → diagnostics/treatments → response/decisions → complications → disposition”
- 실제 퇴원요약 작성 순서와 동일 → 임상 타당성 + 평가 LLM에 구조적 완결성 인식.
- 길이 제어: “Write 12-20 sentences in English (no bullet points)”
- 길이 분포 맞춤으로 BERTScore ↑, 평가 품질 ↑
- 부정 지시문: “avoid redundancy or speculation” → 불필요한 반복/추측 줄여 Conciseness ↑

3. 프롬프트 설계 방법

3.2 TASK별 설계 전략 : TASK A/B/C – 맞춤형 전략

Task B

Train 분석 인사이트

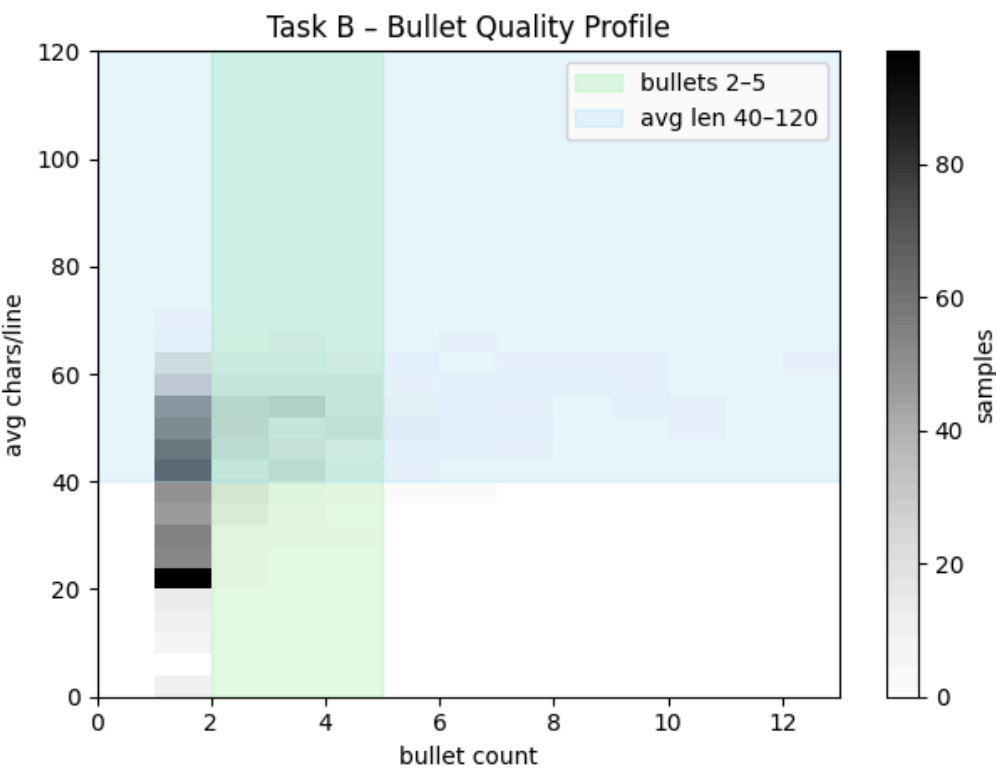
평균 불릿 ≈5, 헤더 존재율 99%

전략적 설계 포인트

IMPRESSION:
헤더 + 3-6 불릿, 권고문구 제거

평가 지표 기대 효과

정량 점수 ↑ (형식 일치), Concise/Clinical ↑



TASK B: 불릿 품질 최적 구간

→ **불릿 수 × 길이 HEATMAP:
3-5불릿, 40-120자 SWEET SPOT**

- 역할 부여: “You are a radiologist” → **영상의학과 의사 스타일 유지**
- 형식 강제: “Start with the header 'IMPRESSION:'” → **정답 99% 형식과 일치**
- 불릿 수 제한: “Provide 3-6 concise bullet points” → **평균 분포(≈5개) 맞춤**
- 내용 범위: “Be explicit about presence/absence; avoid management recommendations” → **임상 요약만 포함, 권고 제거**
- 임상 근거: **RSNA 가이드라인 = Impression은 명확·간결·번호형 불릿.**
- 평가 근거: 형식 일치로 **BERTScore ↑**, 권고 제거로 **Conciseness/Clinical Clarity ↑**

3. 프롬프트 설계 방법

3.2 TASK별 설계 전략 : TASK A/B/C – 맞춤형 전략

Task C

Train 분석 인사이트

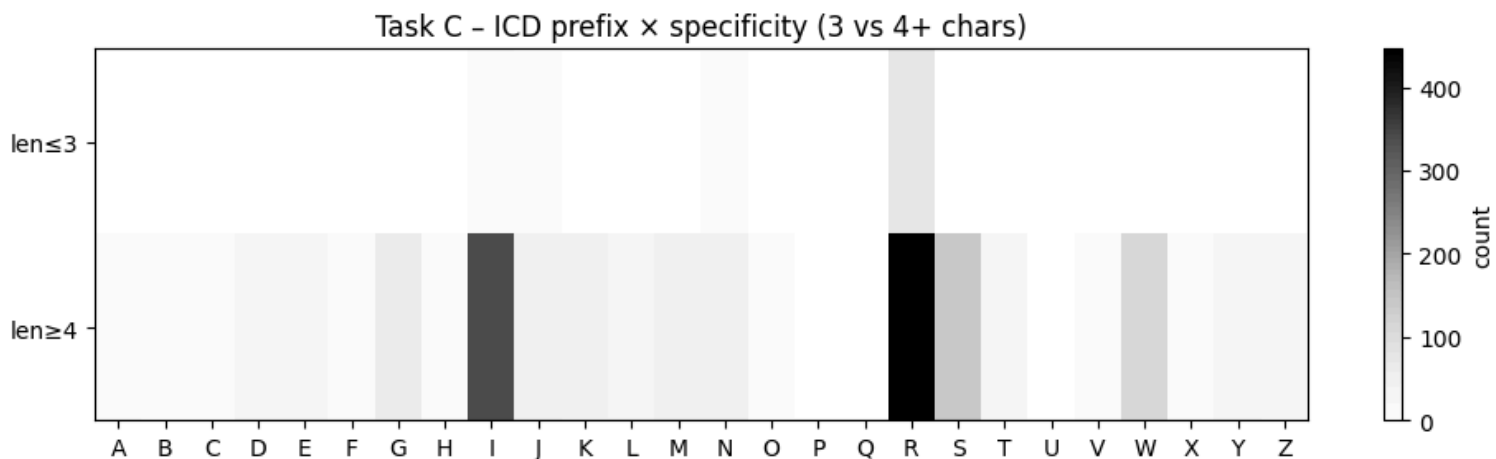
평균 코드수 ≈ 1.58 , 점 없음, R/Z 다수

전략적 설계 포인트

대문자·점제거·중복제거· ≤ 2 개 제한

평가 지표 기대 효과

Macro-F1 \uparrow , 과코딩 방지, 형식 페널티 \downarrow



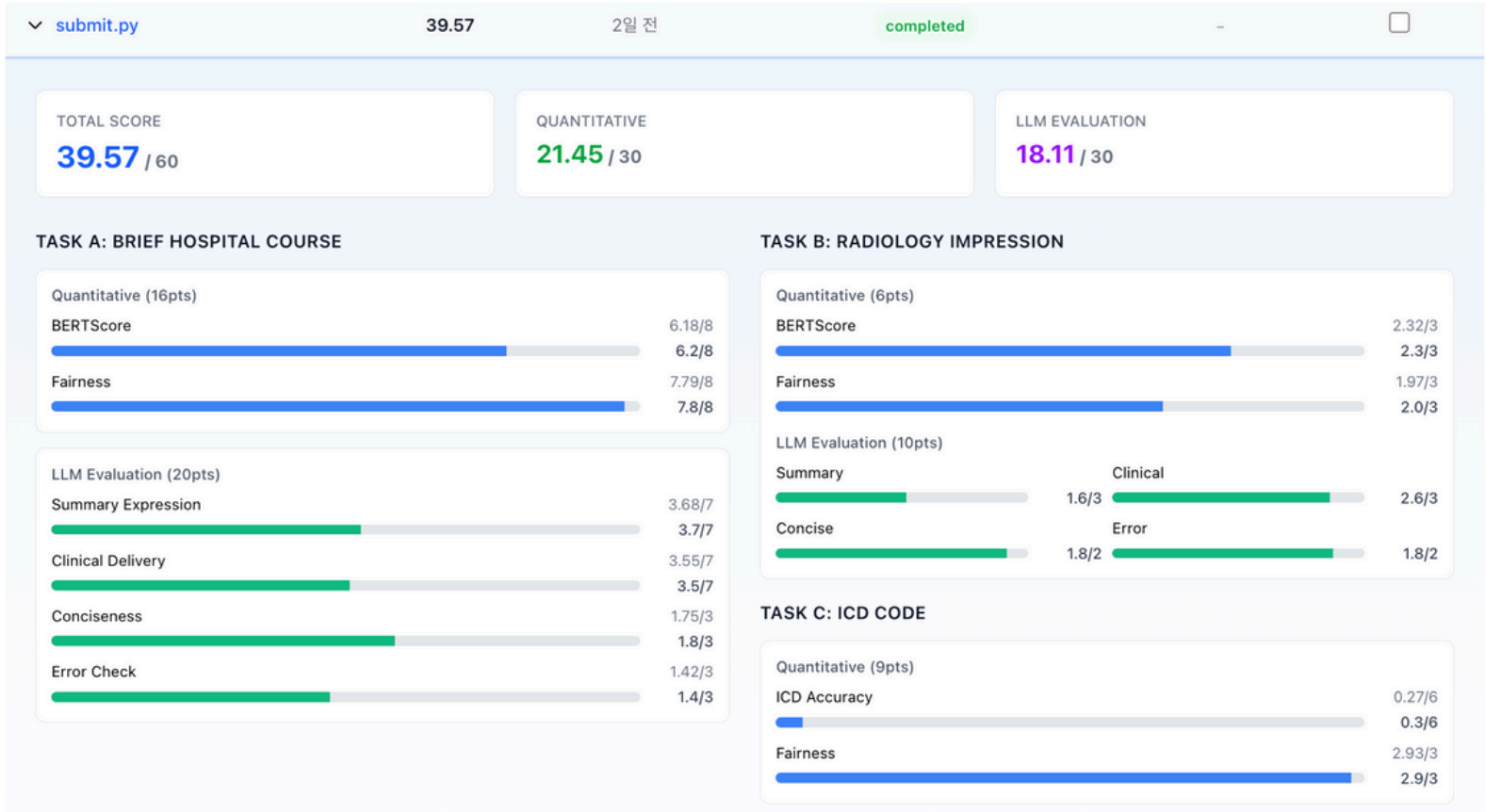
TASK C: 접두어별 특이도 차이

➔ **R/Z = 비특이적, 비RZ = 진단 특이적**
➔ **우선순위 규칙 필요**

- 역할 부여: “You are a medical coder” → **임상 코딩 전문가 톤**
- 출력 개수 제한: “List at most 2 codes” → 평균 1.58개 분포 맞춤, 과코딩 방지
- 출력 포맷 제시: “Output codes separated by commas, no explanations
Examples: I10, E119, W19XXA” → 정규화 용이
- 특정성 강조: “Prefer the most specific codes. Do NOT overcode” → ICD 가이드라인 반영
- 부정 지시문: “no titles/descriptions” → **출력 형식 단순화**
- 임상 근거: 보험·임상 환경에서 R-code 남용 금지, 주진단 우선
- 평가 근거: Macro-F1, hierarchical match에서 과코딩/형식 오류 방지

4. 실험 결과 & 분석

4.1 리더보드 성과



Leaderboard

| RANK | TEAM NAME | SCORE | SUBMISSIONS | LAST SUBMISSION |
|------|--------------------------------|---------|-------------|-----------------|
| 1 | AAI16 | 43.3001 | 5 | 1일 전 |
| 2 | 공돌 | 42.8478 | 11 | 21시간 전 |
| 3 | 갓생호소인 | 42.8347 | 7 | 1일 전 |
| 4 | OTEMR (오토머) | 42.6340 | 9 | 19시간 전 |
| 5 | 클립메디 | 42.6088 | 10 | 13시간 전 |
| 6 | 크리스MARS | 41.0609 | 10 | 20시간 전 |
| 7 | CodeBlueNote | 40.9204 | 10 | 20시간 전 |
| 8 | 메디퓨전 | 40.3846 | 6 | 16시간 전 |
| 9 | MARSketeer | 40.0789 | 7 | 17시간 전 |
| 10 | KB | 40.0647 | 3 | 1일 전 |
| 11 | Strio | 40.0546 | 10 | 17시간 전 |
| 12 | Delphi | 39.9488 | 7 | 17시간 전 |
| 13 | MediX <small>Your Team</small> | 39.5659 | 9 | 18시간 전 |

- BEST 버전 (SUBMIT.PY): 39.57 / 60
 - 정량: 21.45, LLM 평가: 18.11
 - TASK A: BERT 6.2/8, FAIRNESS 7.8/8
 - TASK B: CLINICAL 2.6/3, CONCISE 1.8/2
 - TASK C: ICD ACCURACY 0.3/6 (낮음)

➡ 9월 26일 (금) 오후 8시 기준 리더보드 13위

➡ 정량/공정성은 안정 확보, ICD 정확도가 주된 한계

4. 실험 결과 & 분석

4.2 버전별 비교

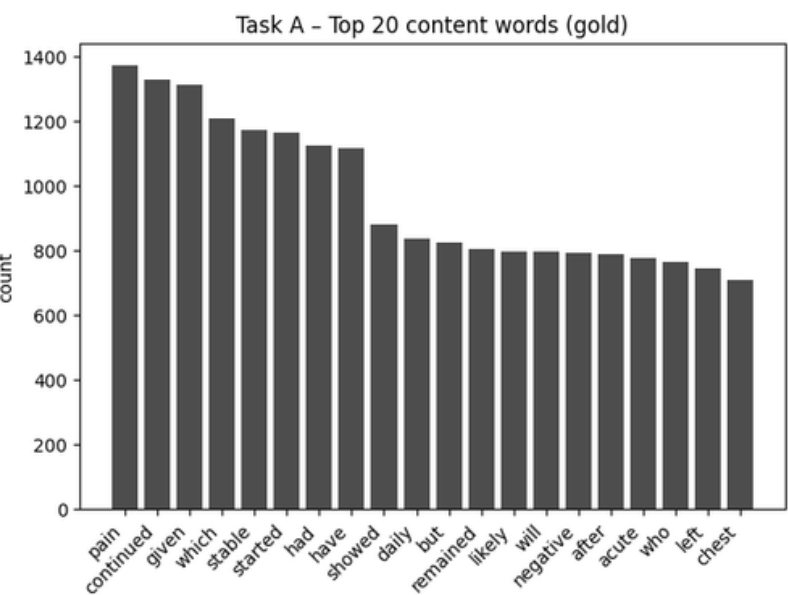
| | 비교 포인트 | V1 (38.75) | V2 (39.02) | V3 (39.57, Best) | 영향 해석 |
|--------|--------|---------------|------------|------------------|-----------------------------------|
| Task A | 문장 길이 | 5-9 | 12-20 | 5-9 | V2 길이정렬로 BERT ↑, V3는 간결성·속도 안정 |
| Task B | 형식 | 1-3 불릿, 헤더 없음 | 헤더+3-6 불릿 | 1-3 불릿, 헤더 없음 | 형식 정렬(V2) → 정량/LLM 평가 ↑ |
| Task B | 권고 제거 | 없음 | 적용 | 헤더만 정리 | 권고 제거 → Concise/Clinical ↑ |
| Task C | 모델 | Llama | Llama | Exaone | Exaone 안정성 ↑, 후처리로 과코딩 억제 |

➡ 형식 정렬이 곧 점수라고 판단했습니다
A/B는 형식이 BERTSCORE와 LLM 평가를 좌우했고, C는 EXAONE+후처리로 ICD 안정성을 확보했습니다

4. 실험 결과 & 분석

4.3 추가분석

Task A: 요약 키워드 분포

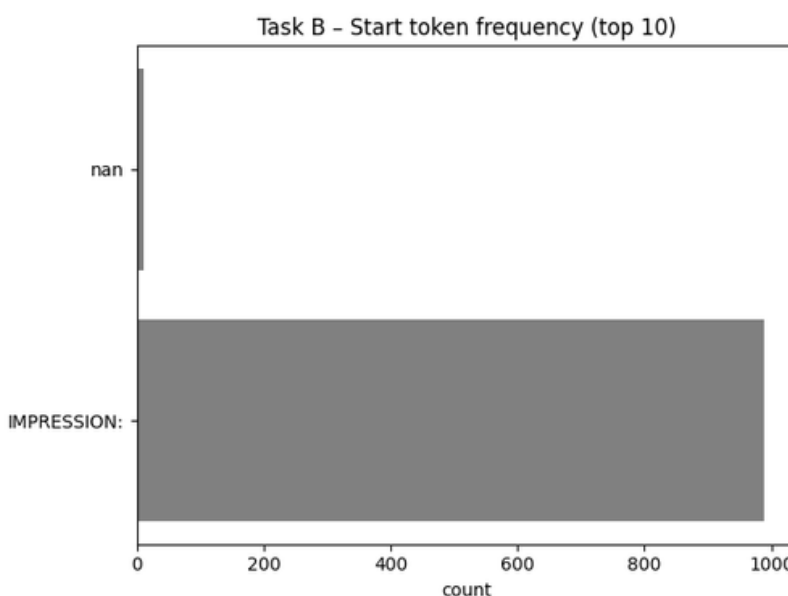


상위 키워드가 환자 상태와
치료 이벤트 중심



요약에서 반드시 포함해야
BERT/LLM 평가 점수 ↑

Task B: 헤더/시작 패턴

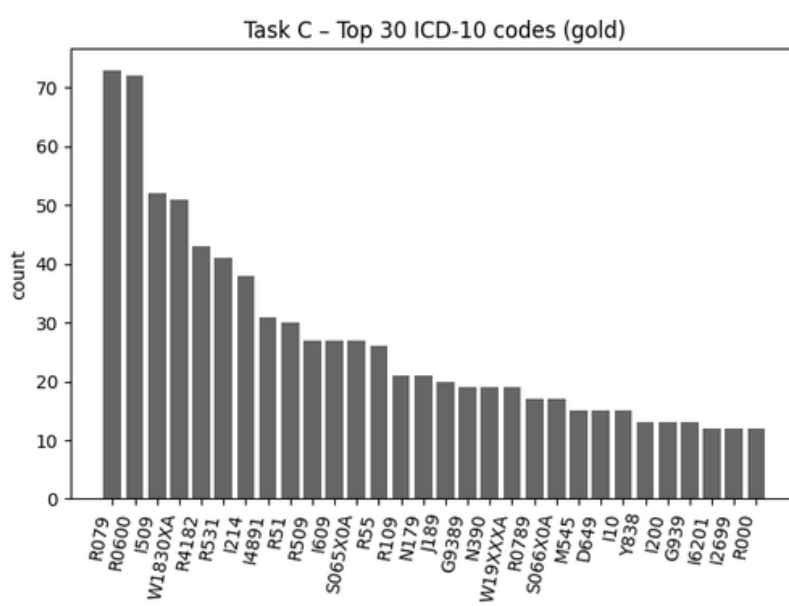


정답 데이터 99%가 'IMPRESSION:'



헤더 정렬 여부가 평가 점수와 직결

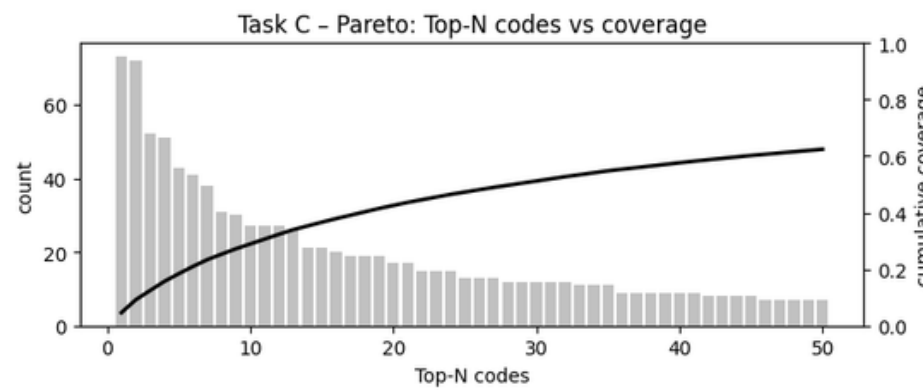
Task C: 코드 파레토 법칙



상위 30개 코드가 대부분 커버



≤2코드 전략이 합리적이고 안정적



4. 실험 결과 & 분석

4.4 임상적 활용 가능성

Task A (입원 경과 요약)

- 바쁜 병동에서 의료진 **인수인계·퇴원 요약 초안 자동화 가능**
- 길이·구조 제어 요약은 EMR(Electronic Medical Record) 템플릿에 바로 **삽입 가능**
- **Conciseness·Error 감소** → 의료사고 예방 및 문서 품질 향상

Task B (Radiology Impression)

- Radiology Findings를 표준 형식의 Impression으로 정리 → **의사-의사 간 커뮤니케이션 개선**
- 권고문구 제거 + 핵심 불릿화 → **진단적 일관성 유지**
- 다기관 연구에서 **형식 정렬된 리포트**가 있으면 NLP 후속분석(예: 암코드 서브타이핑)도 용이

Task C (ICD-10 코드 예측)

- 청구·통계·연구에 **필수적인 ICD 코드를 자동 추천** → 코딩 부담 감소
- ≤2개 제한과 과코딩 방지 전략은 보험 심사·데이터 품질 관리에 직접적 효과
- 장기적으로는 **임상 진단-코드 일치율 개선** → 연구 데이터셋 신뢰성 ↑

- 본선 이후 실제 병원 EMR/RIS와 연계 → **실시간 문서 생성 보조 기능 제공**
- 예: 환자 퇴원 시 자동 초안 생성 → 의사가 확인·수정 후 확정

EMR 연동

- Task A 결과를 Task C ICD 예측에 joint feature로 활용 → **요약·코드 동시 최적화**
- Task B의 정형화된 Impression을 **임상 의사결정 지원 시스템(CDSS)**의 입력으로 활용 가능

Task 확장

- MIMIC 같은 공개 데이터 외에 국내 병원 데이터셋과 RAG 연결 → **사실성(Hallucination) 방지**
- 본선에서 제한된 환경에서도, **진단명 사전/코드 매핑 테이블을 RAG처럼 활용 가능**

RAG(Fact-checking) 결합

5. 한계 & 향후 개선

5.1 현재 한계 → 본선 대비 보완점

예선 문제 인식



- 예선은 정답 문서 대비 유사성 평가 위주였음 → 형식 정렬, 후처리 규칙 설계로 점수 확보
- 본선은 정답 문서 없음 → “임상적으로 타당하고 안전한 새로운 문서 작성”이 평가 핵심.
- 따라서 전략 = ‘모델 출력의 신뢰성’ + ‘임상 실용성’ + ‘공정성’을 모두 균형있게 보여줘야 함

공정성 모니터링



- 현재 한계: subgroup별 세부 분석 부재
- 개선안: 성·연령 subgroup별 성능 추적 + disparate impact(80% rule) 알림 체계

→ 단순 모델이지만 공정성까지 고려한 설계

보고서/시각화 반영



- 현재 한계: 점수 변화 원인 설명 부족
- 개선안: 단순 “성능 수치”가 아니라,
 - Task A: 임상 압축 퍼널 (원문 → 요약 → 최종)
 - Task B: 불릿 품질 열지도 (불릿수×길이 sweet spot)
 - Task C: ICD 파레토 커버리지 (상위 코드가 대부분 커버)

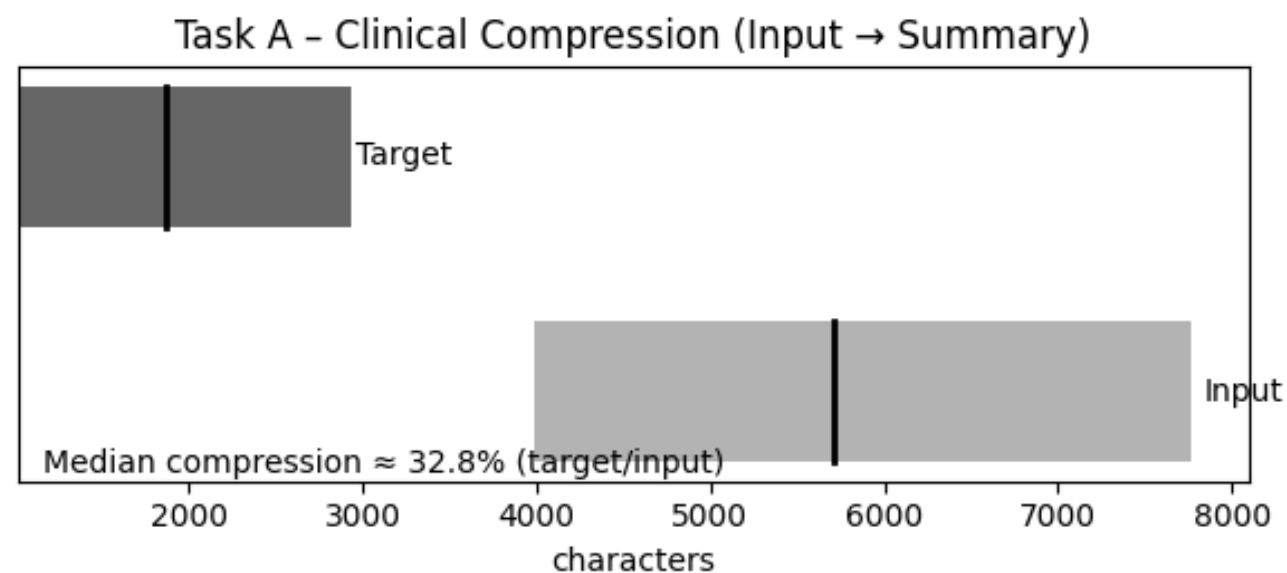
→ 데이터 분석 기반 접근이 단순 엔지니어링이 아닌 전략적 연구임을 보여줌

5. 한계 & 향후 개선

5.1 현재 한계 → 본선 대비 보완점

Task A

- 5-9문장 출력은 정답 길이 분포와 미스매치 → 정보 누락으로 BERT/LLM 품질 손해(길이 제어 필요)
- 길이 제어형 요약 연구는 토큰/길이 프롬프트가 요약 품질·일관성에 영향을 준다고 보고



길이·구조 제어 요약

- **문단 단위**(프리젠테이션/주요 처치/경과/합병증/퇴원계획)로 틀 제공 + **길이 제어 토큰**(예: “Write 15-18 sentences”)을 병행해 정보 보존과 가독성을 동시에 확보
- 길이 제어 요약의 효과는 다수 연구에서 확인됨

➡ “정답 대비 과도한 압축 → 문단 기반·15-18문장 목표”

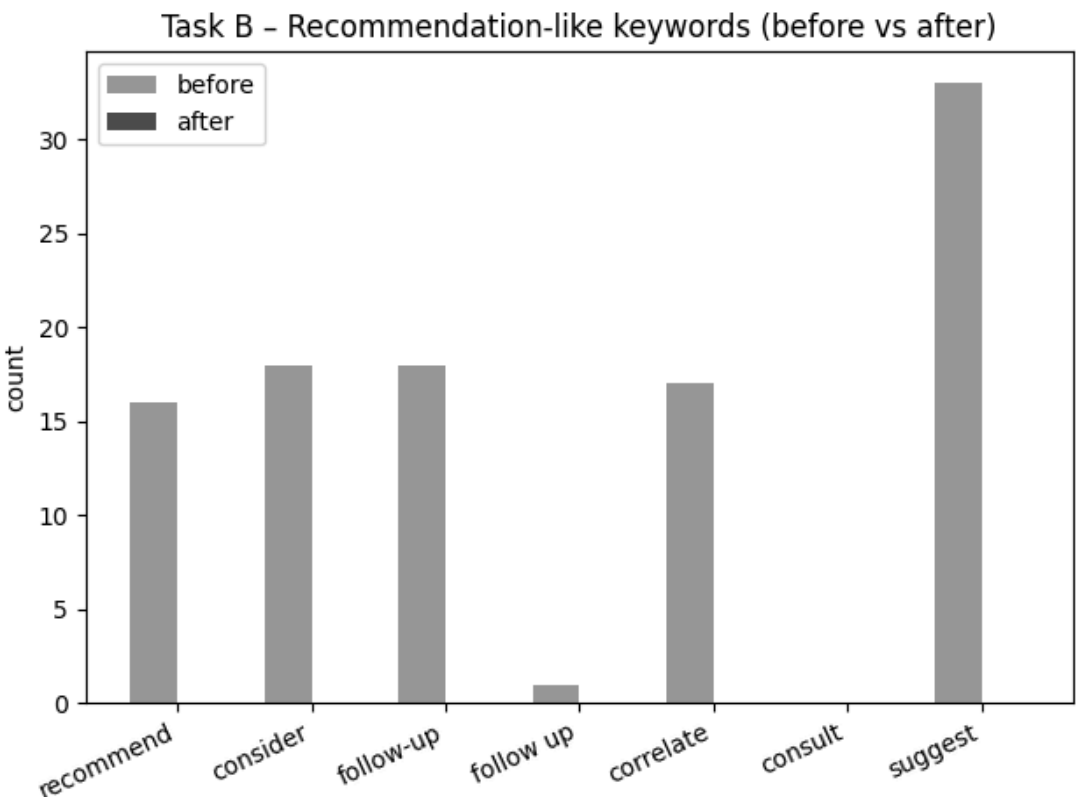
- 프롬프트에 섹션 힌트 제공: Presentation / Diagnostics / Interventions / Clinical Course / Complications / Disposition
- 문장수 15-18개로 고정 → 정보 누락 줄이고 임상 문서로서 자연스러운 길이 확보
- 치료 결과/합병증/퇴원계획까지 반드시 포함 (임상적 의미·안전성 확보)

5. 한계 & 향후 개선

5.1 현재 한계 → 본선 대비 보완점

Task B

- 헤더 미표기·불릿 수 부족은 표준 문서 관습과 불일치 → 평가 모델이 기대하는 형식적 단서(섹션, 나열구조) 부족



표준 형식 & 품질 가이드

- **IMPRESSION** 헤더 고정 + 3-6 numbered bullets
- 중요도 순 정렬(critical first)

→ 표준 템플릿과 교육 자료의 권고와 일치
→ 필요 시 Follow-up/Recommendation은 마지막 한 줄로 분리

→ “권고 제거
→ 본선에서는 **CRITICAL VS FOLLOW-UP** 분리 전략”

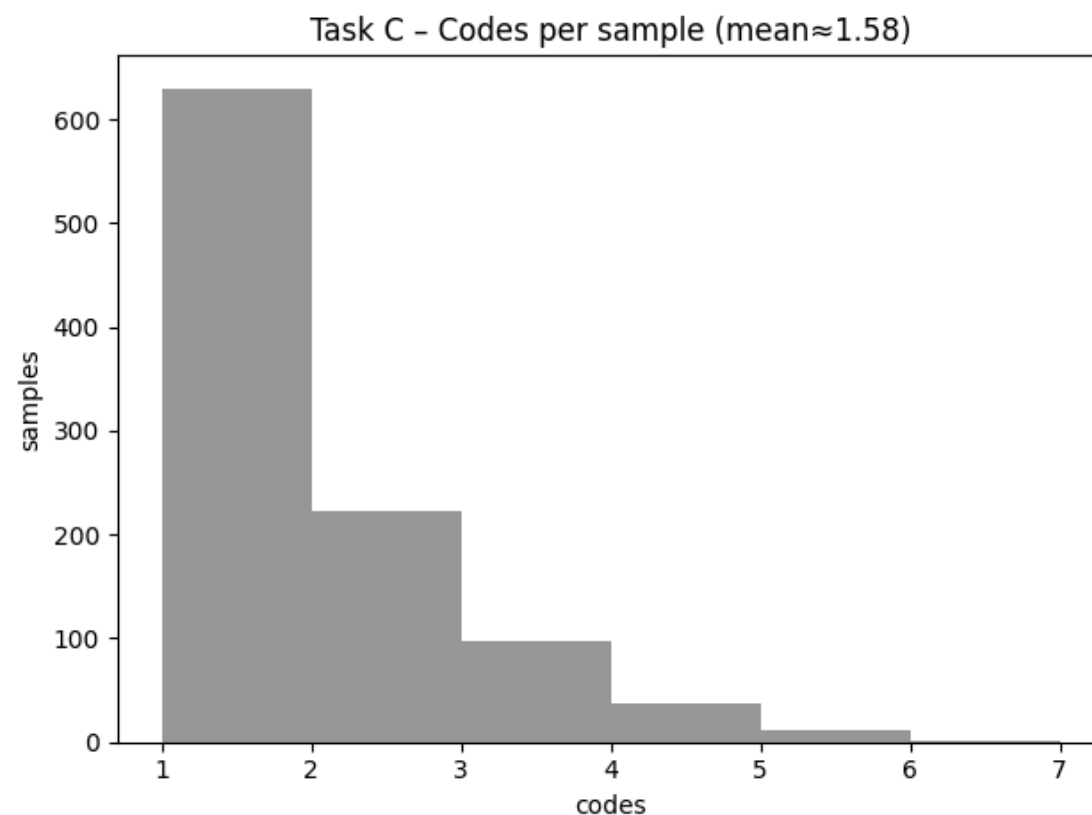
- 헤더 + 번호형 리스트(3-6 bullets) 고정
- 진단적 결론(존재/부재)만 기술, 관리 권고문구(consider/follow-up)는 최종 한 줄로만 별도 분리
- Critical finding 우선순위: 생명을 위협하는 소견은 항상 첫 bullet에 위치

5. 한계 & 향후 개선

5.1 현재 한계 → 본선 대비 보완점

Task C

- 정규식 기반 추출은 오탐/누락에 취약. ICD 지침의 특정성 요구
·증상코드 사용 제한을 세밀히 반영하지 못함



두 단계 예측(후보→선정)

- 후보 생성: 규칙(정규식) + 임상 키워드 매핑으로 상위 N 코드 후보 생성
- 후보 랭킹: (a) 코퍼스 빈도 파레토 기반 priors, (b) 문서-코드 문맥 유사도(간단한 스코어)로 가중 합산

→ 가장 구체적인 1-2개만 반환. R-코드 주진단 금지, 특정성 우선, 증상코드 사용 조건을 규칙으로 강제

→ “≤2코드 제한 + 후보 랭킹 방식으로 정밀도 개선”

- 후보군 생성: 텍스트 내 질환명/증상 키워드 기반 상위 5개 후보 추출
- 후보 랭킹: (a) 빈도 priors (파레토 분포 기반), (b) 문맥 유사도, (c) ICD 지침 룰 (주진단 우선, R/Z 남용 방지)
- 최종 선택: 상위 1-2개만 출력

M.A.R.S. 2025

SNUBH 의무기록 생성 데이터톤

Medical **A**uto-documentation with **R**eal-world **S**tructuring:
LLM Clinical Note Generation Challenge

예선 결과 보고서 [MediX]

- 감사합니다 -