

의료관련감염 감시에 생성형 인공지능 활용하기

Using Generative Artificial Intelligence for Healthcare-Associated Infection Surveillance

Clin Infect Dis. 2026

원문 PDF: https://static.potados.com/papers-ko/2026-06-04/214904-morgan-2026-genai-hai-surveillance_original.pdf

번역 생성일: 2026-06-04

※ 기계 보조 번역(자동 검수 1회 포함)입니다. 인용·임상 판단 시 반드시 원문을 확인하십시오. 참고문헌은 원문 그대로 두었습니다.

관점 논문 (Viewpoints)

(Alzyood의 사설 논평은 480-3쪽 참조.)

초록

생성형 인공지능(generative artificial intelligence, GenAI)은 사회 전반으로 빠르게 확산되고 있으나 의학 분야에 미친 영향은 아직 제한적입니다. GenAI는 방대한 문서를 신속하게 소화할 수 있어, 주로 임상 기록에 기반하는 의료관련감염(healthcare-associated infection, HAI) 감시에 특히 적합합니다. 본 논문에서는 HAI 감시에 대한 GenAI 적용 근거를 검토하고, 진료 현장의 변화를 위해 필요한 단계를 강조합니다. 모의 환자를 대상으로 한 소규모 연구와 환자 기록에 대한 후향적 연구들은 GenAI가 HAI의 개별 판정 기준을 분리해 낼 수 있고, 사람 전문가에 준하는 정확도로 HAI 판정을 내릴 수 있음을 확인했습니다. HAI 탐지를 보조하기 위한 GenAI 도구가 개발되고 있습니다. HAI 검토에 GenAI를 활용하면 시설 간 변동성을 줄이고 HAI 검토의 효율을 높이며, 직접적인 예방 활동에 더 많은 시간을 투입할 수 있게 됩니다. 광범위한 도입을 위해서는, GenAI 기반 HAI 탐지가 전향적 임상연구에서 안전하고 효과적임이 입증되어야 합니다.

핵심어. 인공지능; 감염; 감시.

서론

생성형 인공지능(GenAI)은 콘텐츠를 생성하고 설득력 있는 분석을 제공하는 능력으로 빠르게 주목받아 왔습니다. GenAI는 방대한 양의 텍스트와 기타 자료를 단 몇 초 만에 요약할 수 있습니다. GenAI는 컴퓨터 코딩과 그래픽 디자인 같은 여러 산업 분야에 진출했으나, 의료 분야 침투는 상대적으로 제한적이었습니다. 현재 그 용도는 주로 행정적이며 임상치의 부담을 줄이는 데 초점이 맞춰져 있는데, 여기에는 의무기록 작성(scribe) 보조, 환자 포털 응답 처리, 사전승인(prior authorization) 서한 초안 작성 등이 포함됩니다 [1]. GenAI에 대한 대부분의 탐색은, 의사 면허 시험(medical board examination) 완료와 같이 옳고 그른 답이 존재하는 깨끗하고 정제된 데이터셋을 분석하는 데 집중되어 왔습니다. 보다 최근에는 GenAI를 활용해 복잡한 임상 사례를 분석하거나 진단 추론의 근거를 제공하려는 시도가 이루어졌습니다. 그러나 GenAI를 실시간 임상 의사결정에 구현하는 것은 컴퓨팅 인프라, 실시간 데이터 접근성, 규제 요건 등에 의해 제약을 받아 왔습니다 [2, 3].

감염병 분야에서 GenAI에 대한 소수의 탐색은 대부분 임상 비네트(vignette)나 후향적 데이터를 활용한 이론적 사용에 집중되어 왔습니다. 예를 들어, 초기의 비추론(nonreasoning) 모델을 사용한 거대언어모델(large language model, LLM)은 감염병 자문이나 혈류감염(bloodstream infection) 환자 관리에서 사람 임상의 수준에 미치지 못하는 성능을 보였습니다 [4, 5]. 그러나 LLM은 언어 처리에서 고유한 강점을 지니고 있어 의료관련감염(HAI) 감시에 적합할 수 있습니다 [6-9].

본 관점 논문에서는 HAI 감시에 대한 GenAI 연구를, 이론적 접근에서부터 구현을 위한 실무적 고려사항에 이르기까지 검토합니다. GenAI가 현재의 감시 한계를 어떻게 극복할 수 있는지 분석하고, 임상 활용으로의 책임 있는 통합을 위한 경로를 제안합니다. 주요 용어는 용어집(Glossary)에 제시되어 있습니다.

용어집 (Glossary)

- **응용 프로그래밍 인터페이스(application programming interface, API)**: 프로그램들이 서로 인터페이스할 수 있도록 하는 일련의 규칙. 거대언어모델을 이용한 대규모 일괄 처리(batch-processing) 작업을 수행하는 데 흔히 사용됩니다.
- **챗봇(chatbot)**: 사람이 대화형 인터페이스를 통해 거대언어모델(LLM)과 상호작용할 수 있도록 하는 프로그램.
- **생성형 인공지능(generative artificial intelligence, GenAI)**: 생성 모델(generative model)을 사용해 텍스트와 기타 데이터를 생성하는 인공지능(artificial intelligence, AI).
- **GPT-4o**: 흔히 챗봇을 통해 접근하는, 널리 사용되는 거대언어모델.
- **의료관련감염(healthcare-associated infection, HAI)**: 의료의 합병증으로, 특히 중심정맥관 관련 혈류감염(central line-associated bloodstream infection), 도뇨관 관련 요로감염(catheter-associated urinary tract infection), 수술부위감염(surgical site infection) 등 병원 성과 지표로 보고되는 것들.
- **거대언어모델(large language model, LLM)**: 자유 텍스트 입력을 받아들이고 사람과 유사한 텍스트를 생성하며 프로세스를 제어할 수 있는 GenAI 모델.
- **LLaMA**: 널리 사용되는 오픈소스 LLM.
- **기계학습(machine learning)**: 컴퓨터가 명시적 프로그래밍 없이 데이터로부터 패턴이나 예측을 학습하는 AI의 일반적 범주. 전통적으로는 예측적이며, 구조화된 데이터를 사용해 정량적 추정치(예: 확률)를 출력합니다.

HAI 감시 및 보고의 현황

HAI에는 중심정맥관 관련 혈류감염(central line-associated bloodstream infection, CLABSI), 도뇨관 관련 요로감염(catheter-associated urinary tract infection, CAUTI), 수술부위감염(surgical site infection, SSI)이 포함되며, 이는 의료의 합병증으로 그중 상당 부분이 예방 가능합니다 [10]. HAI 발생률을 추적하는 병원 지표는 공개적으로 보고되고 진료비 보상(reimbursement)과 연계된 최초의 지표 중 하나였습니다 [11]. 미국 질병통제예방센터(US Centers for Disease Control and Prevention, CDC)의 국가의료안전네트워크(National Healthcare Safety Network, NHSN)는 정의와 감시 보고 방식을 제공하며, 미국 내 37,000개 이상의 시설에서 사용됩니다 [12]. 유럽과 아시아에서도 유사한 정의와 보고 체계가 사용됩니다 [13, 14].

이러한 지표가 의료기관의 수익과 평판에 미치는 중요성과 가시성이 HAI 추적에 대한 투자를 이끌어 왔으나, HAI 감시에는 여러 어려움이 따릅니다. 이 감시는 일반적으로 의료시스템에 고용된 감염관리 전문가(infection preventionist, IP)가 수행하며, 이들은 복잡한 정의에 기반해 HAI 사례에 대한 전자건강기록(electronic health record, EHR)을 광범위하게 검토합니다. 이 과정은 자원 집약적입니다. 감시 활동에 필요한 인력-시간(person-time)은 CAUTI 한 건당 15-30분, CLABSI 한 건당 최대 75분에 달하며 [6, 7, 15], 이는 IP 전체 시간의 1/3을 차지할 수 있습니다 [16]. HAI 감시는 HAI를 줄이기 위해 병동과 협력하는 등 보다 능동적인 감염 예방 활동에 쓸 시간을 제한합니다. 한 조사에서는 감염관리 전문가의 95%가 HAI 감시가 다른 감염관리 활동을 방해한다고 응답했습니다 [17]. 더욱이 CDC의 정의가 점점 상세해지고 있음에도 불구하고, 보고는 잠재적 편향과 인간 오류에 노출되어 있으며, 더 낮은 HAI 발생률을 보고하려는 지불 구조와 평판상의 압력이 존재합니다 [15].

Lin 등 [15]은 객관적 알고리즘 접근법과 실제 사람 기반 보고를 비교했을 때 병원 간 CLABSI 발생률에서 최대 2배의 변동성을 확인했습니다. 보다 최근의 연구들도 CLABSI 정의에 대한 과소보고(underreporting)와 변동적·비일관적 해석을 계속해서 기록해 왔습니다 [18-21]. HAI 보고의 객관성 문제는 임상가와 행정가들 사이에서 이러한 지표에 대한 신뢰의 침식으로 이어졌습니다 [22]. 이는 전적으로 EHR 데이터에 기반하며 추가적인 수동 검토가 필요 없는 병원 발생 균혈증(hospital-onset bacteremia) 같은 새로운 지표의 개발을 촉진했습니다. 다만 병원 발생 균혈증은 현재 보고가 의무화된 지표는 아닙니다 [23].

전통적 AI를 이용한 HAI 탐지

자동화 및 부분 자동화 감시는 효율을 높일 수 있으나, 역사적으로 이러한 기법은 구현하기 어려웠습니다 [24]. GenAI가 등장하기 전, 자동화된 HAI 감시는 EHR에서 손쉽게 추출할 수 있는 구조화된 데이터 요소(예: 국제질병분류 제10판 [International Classification of Diseases, Tenth Revision] 코드)를 사용하는 기계학습이나 기타 통계 알고리즘에 집중되어 있었습니다 [25]. 이러한 자동화된 HAI 감시 접근법의 근본적 장벽은, 중요한 변수가 비구조화된 임상 기록에만 자주 존재하며, 이러한 기준을 신뢰성 있게 추출하는 것이 어려울 수 있다는 점입니다 [26, 27].

또한 한 시설에서 개발된 자동화 감시 도구는 다른 시설로 쉽게 이전되지 않았습니다. 예를 들어 한 연구는 EHR에서 대장직장(colorectal) SSI를 분류할 수 있는 변수를 식별했으나, 광범위한 구현 전에 모델 매개변수의 지역적 이용 가능성과 검증이 필요했습니다 [28].

HAI를 위한 자연어 처리 (Natural Language Processing for HAIs)

GenAI 이전에 기록의 비구조화 데이터에 접근하는 방법은 자연어 처리(natural language processing)였습니다. 자연어 처리는 LLM에 비해 저렴한 컴퓨팅 비용 같은 일부 장점을 제공하지만, 중요한 방법론적·구현상의 장벽도 가지고 있습니다 [29]. 즉, 모델을 구축하고 시험하기 위해 전문가가 필요하고, 자원 집약적인 전문가 주석(annotation)을 통해 만들어진 기관별 학습 데이터셋에 의존하며, 문서화 관행이 다른 의료시스템 간에 제한적인 이전성(transferability)을 보입니다 [30, 31]. 수백만 건의 이용 가능한 문서로 학습된 GenAI는 임상 정보에 대한 기초 지식을 갖추고 있어 더 일반화 가능하고, 더 적은 전문성으로 더 쉽게 구현할 수 있습니다.

HAI 탐지를 위한 GenAI의 탐색

모의 사례를 대상으로 수행된 전임상 연구들은 챗봇과 LLM 기반 도구가 흔한 HAI를 식별할 잠재력을 입증했습니다 [6]. 실제 환자를 대상으로 한 GenAI 탐색의 대부분은 기록과 데이터를 챗봇 창에 붙여 넣는 방식으로 수행되었습니다. 다만 향후 도구는 EHR에서 데이터를 자동으로 추출해 사람의 노력을 덜 요구하는, 보다 정교하게 개발된 수동적(passive) 감시를 포함할 가능성이 높습니다(표 1).

전임상 개념증명 연구 (Preclinical Proof-of-Concept Studies)

상용 챗봇과 관련된 개인정보보호 우려 때문에, 초기 연구는 합성(synthetic) 또는 모의(simulated) 환자 기록으로 제한되었습니다. 그러나 이러한 예비 조사들은 고무적인 결과를 내놓았습니다. 연구자들은 모의 사례를 사용해 CAUTI를 탐지하는 GPT-4o의 성능을 탐색했고, 허용 가능한 수준의 민감도와 특이도를 확인했습니다 [7]. 마찬가지로 GPT-4o는 CDC 교육 사례에서 CLABSI와 CAUTI를 식별할 수 있었습니다 [8]. 그러나 이러한 모의실험은 비교적 단순했고 복잡한 의무기록의 현실을 반영하지는 못했습니다.

SSI에 대한 GenAI 탐지는 공중보건 기록 발췌자료(public health record abstraction)와 CDC-NHSN 정의를 사용해 탐색되었습니다. GPT-4o는 민감도 100%, 특이도 54%였으며, SSI를 표재성(superficial), 심부(deep), 장기/체강(organ space)으로 분류할 수 있었습니다. 이러한 검토에는 사례당 중앙값 6분이 소요되었으나, 데이터가 의무기록에서 나온 것이 아니라는 점을 강조하는 것이 중요합니다 [9].

환자 기록에 대한 GenAI의 후향적 평가

의료기관들이 안전한 챗봇 환경을 구현하기 시작하면서, 평가는 실제 환자 기록으로 확대되었습니다. CLABSI는 가장 많이 연구된 HAI입니다. 연구자들은 단일 병원에서 20개의 사례군과 20개의 대조군을 대상으로 GPT-4o가 CLABSI 사례의 90%를 식별하고 비-CLABSI 균혈증 사례의 75%를 배제할 수 있음을 확인했습니다 [6]. 이러한 검토는 효율적이어서, 보고된 IP 검토 시간 75분에 비해 중앙값 5분이 소요되었습니다 [6].

미국 보훈부(US Department of Veterans Affairs, VA) 내에서 수행된 보다 대규모의 운영 평가는, 일선 IP와 의사가 사용하는 표준 프롬프트-데이터-평가를 적용해 11개 병원에 걸쳐 CLABSI 판정을 검토했습니다. 그 결과 GPT-4o 기반 AI 보조 검토가 전통적 검토만큼 정확하고, 더 빠르며, 시설 간 변동이 더 적은 것으로 나타났습니다 [32]. 프롬프트에 CDC-NHSN 기준의 상세한 지침을 포함하면 정확도가 향상되었습니다. 더욱이 신중한 프롬프팅을 통해 GenAI는 중심정맥관의 존재 여부나 이차성 혈류감염(secondary bloodstream infection)과 같은 핵심 데이터를 비구조화 기록에서 강조 표시할 수 있어, 보다 효율적인 수동 검토를 가능하게 했습니다. 그러나 GenAI는 데이터 요소를 올바르게 식별한 후에도 일부 CLABSI 또는 기타 HAI를 여전히 오분류했습니다(예: 중심정맥관 삽입일과 균혈증 발생일을 식별했으나, 균혈증이 관 삽입 후 하루 이내에 발생했음에도 CLABSI가 존재한다고 잘못 결론 내림).

한 단일기관 연구에서는 21개의 CAUTI 사례와 21개의 짝지은 대조군을 대상으로 GenAI를 이용한 CAUTI 감시를 수행했으며, 민감도 95%, 특이도 76%(GenAI 출력에 대한 전문가 검토를 거치면 90%로 향상)를 보고했습니다 [33]. GenAI는 또한 시설 검토에서 놓친 대조군 내 진성(true) CAUTI 1건을 식별했습니다.

HAI 감시를 위한 GenAI 구현

이러한 개념증명 연구들이 가능성을 보여주지만, 운영 시스템으로 전환하려면 — 그것이 기록을 복사·붙여넣기 해 검토를 수행하는 안전한 챗봇으로 구현되든, API를 갖춰 HAI 식별을 자동화 또는 부분 자동화하는 도구로 구현되든 — 유사한 기초 단계를 따라야 합니다. 이 단계에는 (1) GenAI 사용의 기술적 측면, (2) 사람의 지원 대비 자동화의 정도에 관한 결정, (3) HAI 탐지를 위한 병원의 GenAI 통합, (4) GenAI를 사람의 업무 흐름에 통합하는 것이 포함됩니다.

GenAI 임상 구현의 기술적 단계

GenAI 임상 구현의 기술적 단계에는 (1) 적절한 데이터 확보, (2) 표준화된 프롬프트 사용, (3) 유용한 형식으로 결과 내보내기가 포함됩니다. 첫째, 임상 기록과 검사 결과 같은 적절한 범주의 데이터 수집이 표준화되고 검증되어야 합니다. 둘째, GenAI 시스템이 동일한 요소를 포착하고 낮은 변동성으로 정확하며 재현 가능한 검토를 수행하도록 유도할 수 있는, 일반화 가능한 모델 프롬프팅 전략이 개발·검증되어야 합니다. 이상적으로는, 검증된 프롬프트 저장소가 CDC와 같은 규제 기관이나 미국의료역학회(Society for Healthcare Epidemiology of America)·감염관리전문가협회(Association for Professionals in Infection Control)와 같은 국가 조직에 의해 공개적으로 개발·유지되어야 합니다. 셋째, GenAI 결과는 현행 HAI 보고 절차나 새로운 감시 체계에 통합될 수 있도록 구조화된 방식으로 내보내져야 합니다.

사람 검토에 대한 GenAI 지원 대 자동화

사람 검토자의 정확도와 일치하거나 그를 능가하는 기술적으로 견고한 GenAI 시스템이라 하더라도, 의료기관과 규제 당국은 중요한 구현 결정에 직면합니다. 감시 구현 옵션은 하나의 연속선상에 존재합니다(그림 1). 한쪽 극단에는 GenAI와 기타

방법을 사용하는 기계가 사례를 독립적으로 검토하고 확정적 판정(예: "CLABSI 여부")을 제공하는 완전 자동화가 있고, 다른 쪽 극단에는 GenAI가 CLABSI 판정을 위한 몇 가지 핵심 요소만 추출하는 데 사용되는 경우가 있습니다.

HAI 탐지를 위한 병원의 GenAI 통합

Epic이나 Theradoc 같은 현행 시스템은 현재 구조화된 데이터 필드로부터 균혈증과 중요한 병원 날짜(hospital dates)를 식별해 주며, 잘 수용되고 있습니다. GenAI 도구의 경우, 대부분의 병원은 자체적인 특정 시스템을 설계하기보다는 대형 EHR 공급업체로부터 이용 가능한 것 중에서 선택하게 될 것입니다. 그러나 개발자, 규제 당국, 병원이 제품을 만들 때 다양한 GenAI 옵션의 범위와, 자동화 대비 사람 감독의 정도를 고려하는 것이 매우 중요합니다(그림 1).

GenAI를 사람의 업무 흐름에 통합하기

기술과 EHR에 관한 결정을 넘어, 사람이 일하는 병원 환경도 고려해야 합니다. 현재 누가 어떤 도구를 사용해 HAI 감시를 수행하는가? GenAI 도구가 그들의 업무 흐름에 어떤 영향을 미칠 것인가? 이 문제는 대부분의 연구에서 다루어지지 않았으며, GenAI 감시가 임상 의사에 의해 어떻게 사용될지와 도구의 수용성·유용성을 다룬 것은 VA의 AI CLABSI 연구뿐이었습니다 [6-9, 32, 33].

거버넌스, 책무성, 신뢰성 (Governance, Accountability, and Reliability)

조직은 국가 조직의 지원을 받아 GenAI 성능을 지속적으로 평가하고 유지하기 위한 계획을 필요로 할 것입니다. 궁극적으로 이는 병원이 수행하도록 의무화되거나, EHR에 접근 권한을 가진 제3자를 통해 이루어져야 합니다. 잠재적으로는 무작위 감사(spot audit)나 보고된 사례에 대한 검토(또는 감사할 신호를 찾기 위해 GenAI가 추천한 사례 대비 실제 보고된 사례의 비율을 검토)가 이루어질 수 있습니다. 끝으로, 기준 표준(reference standard)에 대한 결과의 지속적 평가를 위한 프로토콜이 수립되어야 합니다.

HAI 탐지를 위한 GenAI의 잠재적 이점

GenAI를 활용한 HAI 탐지는 보다 효율적인 기록 검토로 이어지고 IP 전문가가 HAI 감시에 투입하는 시간을 줄일 수 있습니다 [6]. 또한 GenAI 검토는 각 병원의 서로 다른 검토자의 주관성 대신 단일 표준을 적용하므로 변동성이 적고 더 객관적일 수 있습니다(다만 LLM은 성능에 일부 확률적 요소(stochasticity)가 있으므로 이는 연구를 통해 확인되어야 합니다) [32]. 검토의 변동성이 줄어들면 보다 공정한 지표가 되어, 공개 보고가 기대하던 이점, 즉 진정으로 감염을 줄이는 질 향상(quality improvement)을 달성할 수 있습니다. 이러한 잠재적 이점은 실제 현장 연구에서 평가될 필요가 있습니다.

GenAI는 HAI의 잠재적 원인과 HAI를 줄이기 위한 중재(intervention)를 식별하는 데에도 도움이 될 수 있습니다. 예를 들어 CAUTI 환자의 기록 분석을 통해 간호 기록에서 도뇨 백(catheter bag)으로부터 검체를 부적절하게 채취한 사실을 식별할 수 있고, 이는 검체 채취를 개선하기 위한 간호사 교육으로 이어질 수 있습니다.

궁극적으로, GenAI를 사용해 병원 간 표준화된 방식으로 검토를 수행할 수 있게 되면, CDC를 비롯해 HAI를 정의하거나 보고하는 기관들의 HAI 정의가 바뀔 수도 있습니다. 현재 HAI 식별은 복잡하며, 여러 층위의 배제 규칙(exclusionary rule)을 일반적으로 적용하는 훈련된 검토자가 필요합니다. 예를 들어 CLABSI 판정에서는 이차 부위의 배양 없이 발생한 이차성 혈류감염(PNU2나 IAB3 같은 용어로 알려진 것)에 대한 배제 규정이 있습니다. 이러한 배제 규정은 적용하기 어렵고 감사하기는 더욱 어렵습니다. 그러나 GenAI가 병원 간 표준화된 방식으로 적용하고 각 사례에 대해 명확히 표시할 수 있는 정의 변경은, HAI의 임상적 정의와의 투명성 및 일관성을 높일 수 있습니다. 예를 들어 단순화된 CLABSI 정의는 중심정맥관의 존재, 병원체에 의한 균혈증, 그리고 이차 감염에 대한 GenAI의 최선의 설명을 요구하는 형태가 될 수 있습니다.

GenAI HAI 탐지의 과제

정확도 (Accuracy)

첫 번째 과제는 GenAI가 전향적 방식으로 HAI를 얼마나 잘 식별하는가입니다. 각 개별 HAI에 걸쳐 사람 전문가와 비교했을 때 얼마나 정확한가? 어떤 검토 방법이 질 향상으로 해결할 수 있는 예방 가능한 HAI를 가장 잘 식별하는가? 정확도를 극대화하기 위해 GenAI를 임상 시스템에 어떻게 구현할 수 있는가? 일부 연구가 발표되었으나 [6], 보다 대규모의 구현 연구가 필요합니다.

AI 윤리와 편향 (AI Ethics and Bias)

성공적인 GenAI HAI 탐지 도구는 새로운 윤리적 고려사항을 도입할 것입니다. EHR, 청구, 기타 기술에서 보았듯이, GenAI 보조 HAI 감시 도구에 대한 접근성은 — 이러한 도구가 진료비 보상과 다른 고영향(high-impact) 조사·활동에 쓸 IP 시간에 영향을 미친다면 — 병원을 "가진 자(haves)"와 "못 가진 자(have nots)"로 더욱 갈라놓을 수 있습니다. 또한 인종이나 성별 고정관념과 관련된 의식적·무의식적 편향이 LLM의 학습 데이터에 내재될 수 있으며, 이는 LLM의 임상 권고에 영향을 미치는 것으로 입증되었습니다 [34-36]. 이러한 잠재적 편향은 도구 개발 및 평가 과정에서 평가되어야 합니다 [34-36].

끝으로, AI 의무기록 작성기(AI scribe)가 임상 기록을 점점 더 많이 생성하고 있으며, AI가 생성하는 퇴원 요약서(discharge summary)와 같은 다른 문서화 도구도 등장하고 있습니다 [37]. GenAI HAI 도구의 비구조화 입력이 사람이 아닌 LLM에 의해 점점 더 많이 작성됨에 따라, AI 의무기록 작성기가 HAI 발생률을 인위적으로 낮추는 내용과 표현을 포함하도록 학습될 수 있습니다. 예를 들어, 작성기가 CAUTI 기준에 해당하는 증상을 나열하는 것을 의도적으로 회피할 수 있습니다.

회사 또는 모델에 따른 변동 (Variation by Company or Model)

프롬프트가 표준화될 수 있더라도, 작은 변경이나 서로 다른 기반 LLM은 결과를 바꿀 수 있습니다. LLM의 버전 업데이트조차 성능을 변화시킬 수 있습니다. 이는 정기적으로 평가되어야 하며, 기업은 프롬프트, 모델 업데이트, 기능을 투명하게 공개하도록 요구되어야 합니다.

비용 (Costs)

GenAI의 비용은 장벽이 될 가능성이 높습니다. 모델은 실행할 때마다 자원을 필요로 하며, 새로운 도구는 의료시스템에 흔히 비용 부담이 큼니다. 만약 지출이 감염관리 인력에서 특정 기술 도구로 이동한다면, 그 순효과는 감염관리 프로그램 내에서 필요에 대응하는 유연성과 능력의 감소일 수 있습니다.

정책 인센티브 (Policy Incentives)

"어떤 측정치가 목표가 되면, 그것은 더 이상 좋은 측정치가 아니게 된다"는 굿하트의 법칙(Goodhart's law)이 GenAI HAI 검토에도 적용될 수 있습니다. 어떤 절차도 완벽하지 않으며, 이차 감염을 우선적으로 기록하는 것처럼 GenAI 검토에 기반한 새로운 시스템을 "악용(game)"할 방법이 존재할 것입니다.

감염관리 및 환자 안전을 위한 GenAI

감염 감시의 확장 (Expansion of Infection Surveillance)

많은 HAI 감시 프로그램은 규제 요건(예: CLABSI, CAUTI, 그리고 일부 선정된 HAI)에 의해 선택이 좌우되어 몇 가지 유형의 HAI에만 국한되어 있습니다. 그러나 다른 많은 감염도 중요할 수 있으나 드물기 때문에 모니터링이 비현실적입니다 — 예를

들어, 드물지만 치명적인 심장 장치 관련 감염(cardiac device-associated infection)이 그러합니다. 표준화된 정의에 대한 자동 검토가 가능해지면, 효율적으로 HAI 감시를 수행할 수 있는 기술의 수가 극적으로 확대될 수 있습니다. 그 밖의 잠재적 대상으로는 다른 수술 감염, 비인공호흡기 병원관련 폐렴(nonventilator hospital-associated pneumonia), 내시경 합병증, 다양한 수술 후 항생제 예방요법의 합병증 등이 있습니다 [26, 38, 39].

발생 탐지 및 조사 (Outbreak Detection and Investigation)

유사한 특성이나 노출을 가진 감염의 집단(cluster)이나 유행(outbreak)을 식별하는 것은 흔히 사람이 기록을 검토하는 데 의존합니다. GenAI는 환자 기록이 있을 경우 특정 노출이나 비전형적 감염을 모니터링하는 데 사용될 수 있습니다. 예를 들어, 동일한 군이 식별되면 GenAI의 기록 분석을 통해 공통 노출원을 신속하게 탐색할 수 있습니다. 마찬가지로 GenAI는 조류인플루엔자(avian influenza)에 대한 외래(outpatient) 위험인자를 식별할 수 있습니다 [40].

GenAI로 탐지 가능한 기타 환자 안전 지표

GenAI는 질 향상의 다른 영역, 예를 들어 패혈증 지표(sepsis metric)나 항생제 사용 사유 등에서도 탐색되어 왔습니다 [41]. GenAI는 장기요양시설 노출(long-term care exposure)과 같이 감염 가능성에 영향을 미치는 위험인자에 대한 선별검사를 개선할 수도 있습니다 [42]. HAI를 넘어, 낙상(fall)이나 욕창(pressure injury) 같은 다른 전통적 병원 환자 안전 지표도 GenAI 검토로 식별될 수 있습니다. 이들은 주로 기록에 문서화되며 즉시가 아닌 지연을 두고 보고되어 모든 데이터가 축적될 시간을 허용하기 때문입니다.

표 1. 다양한 유형의 의료관련감염 식별 및 예방에서 거대언어모델의 역할

HAI	GenAI 검토에 필요한 데이터	GenAI 활용 방식	GenAI 활용 근거의 강도	연구의 한계
CLABSI	임상 기록, 장치, 미생물 검사, 활력징후, ADT	CLABSI 식별/배제	일선 의료제공자의 다기관 코호트 [32]	후향적, 미구현
CAUTI	임상 기록, 장치, 미생물 검사, 활력징후, ADT	CAUTI 식별/배제	전임상 결과 [7, 8] 및 단일 코호트	후향적, 1개 병원, 미구현
SSI	임상 기록, 수술 기록, 미생물 검사, 활력징후, ADT, 영상의학	SSI 식별/배제	발체 데이터를 사용한 소규모 연구 [9]	후향적; 환자 기록을 사용하지 않음
HOB	임상 기록, 장치, 수술 기록, 미생물 검사, 활력징후, ADT, 영상의학	HOB의 원인 귀속 및 잠재적 중재 ^a	없음	연구된 바 없음

^a 입원 후 72시간 이상 경과한 혈액 배양으로만 정의되는 지표.

약어: ADT, 입원·퇴원·전실(admission, discharge, transfer); CAUTI, 도뇨관 관련 요로감염(catheter-associated urinary tract infection); CLABSI, 중심정맥관 관련 혈류감염(central line-associated bloodstream infection); GenAI, 생성형 인공지능(generative artificial intelligence); HAI, 의료관련감염(healthcare-associated infection); HOB, 병원 발생 균혈증(hospital-onset bacteremia); SSI, 수술부위감염(surgical site infection).

그림 1. 의료관련감염인 중심정맥관(central line, CL) 관련 혈류감염(CLABSI) 탐지를 위한 구현 접근법의 범위. 전통적 사람 검토에서부터 점차 자동화된 다양한 접근법을 거쳐, 사람 감독 없는 완전 자동화 검토에 이르기까지의 스펙트럼을 보여줍니다. 한 예시가 제시되어 있는데, CLABSI에 대한 잠재적 생성형 인공지능(GenAI) 검토를 시연하며 CLABSI의 구체적 기준을 문서화하면 이후 감염관리 전문가가 최종 판정을 위해 검토할 수 있도록 합니다. 파란색으로 표시된 문서 내용은 전자건강기록(EHR) 내 요소로 연결되는 하이퍼링크를 나타내며, 날짜는 월/

일 형식으로 표기됩니다. 약어: C/D/I, 깨끗하고 건조하며 손상 없음(clean, dry, and intact); CT, 컴퓨터단층촬영(computed tomographic); *E. coli*, *Escherichia coli*; HOB, 병원 발생 균혈증(hospital-onset bacteremia); MBI, 점막 장벽 감염(mucosal barrier infection); PICC, 말초삽입 중심정맥관(peripherally inserted central catheter); *S. aureus*, *Staphylococcus aureus*; *S. capitis*, *Staphylococcus capitis*; UTI, 요로감염(urinary tract infection); WBC, 백혈구(white blood cell, 세포수/마이크로리터). [원문 그림 참조]

결론

GenAI는 HAI 탐지와 기타 환자 안전 지표에 유망합니다. 소수의 연구들은 현재 기술이 감염관리 팀의 사례 검토를 보조할 수 있음을 확인했습니다. 이를 개발하기 위한 계획이 VA와 EHR 기업들에서 진전되고 있습니다. 효과적으로 구현된다면, 이러한 도구는 지표 수집의 효율을 높이고, 잠재적으로 다른 감염으로의 감시 확대를 가능하게 하며, 감염 예방·관리 팀이 고수익(high-yield) 활동에 투입할 시간을 더 많이 확보해 줄 수 있습니다. 이 분야의 우선과제는 (1) GenAI 도구에 대한 대규모 전향적 연구, (2) 도구의 인간 사용 및 발생하는 인지 편향(cognitive bias)에 대한 평가, (3) GenAI가 수행하는 검토에 최적화된 HAI 정의의 재검토입니다.

비고 (Notes)

재정 지원(Financial support). 없음.

잠재적 이해상충(Potential conflicts of interest). 저자들: 보고된 이해상충 없음.

모든 저자는 ICMJE 잠재적 이해상충 공개 양식(ICMJE Form for Disclosure of Potential Conflicts of Interest)을 제출했습니다. 편집진이 본 원고의 내용과 관련 있다고 판단한 이해상충은 공개되었습니다.

저자 소속 (Author Affiliations)

¹Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, Maryland, USA; ²Medical Care Center, VA Maryland Health Care System, Baltimore, Maryland, USA; ³Institute for Healthcare Computing, University of Maryland, North Bethesda, Maryland, USA; ⁴Center for the Study of Healthcare Innovation, Implementation, and Policy, Greater Los Angeles VA Healthcare System, Los Angeles, California, USA; ⁵Department of Medicine, UCLA David Geffen School of Medicine, Los Angeles, California, USA; ⁶Infection Control, Mass General Brigham, Boston, Massachusetts, USA; ⁷Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts, USA; ⁸Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; ⁹Department of Medicine, Stanford University, Palo Alto, California, USA; ¹⁰Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA; ¹¹Medical Care, Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah, USA; ¹²Applied Innovations and Medical Informatics (AIMI), Veterans Affairs Long Beach Healthcare System, Long Beach, California, USA; ¹³Office of Research and Development, Veterans Health Administration, Washington, DC, USA; ¹⁴Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA; ¹⁵Institute of Human Virology, University of Maryland, Baltimore, Maryland, USA; ¹⁶Department of Emergency Medicine, Johns Hopkins Medicine, Baltimore, Maryland, USA.

교신저자: D. J. Morgan, Department of Epidemiology and Public Health, University of Maryland School of Medicine, 10 S Pine St, MSTF 334, Baltimore, MD 21201 (dmorgan@som.umaryland.edu); E. Klein, Department of Emergency Medicine, Johns Hopkins Medicine, 5801 Smith Ave., Suite 3220, Baltimore, MD 21209 (eklein@jhu.edu).

접수: 2025년 8월 18일; 채택: 2025년 10월 21일; 온라인 게재: 2025년 11월 12일.

Clinical Infectious Diseases® 2026;82(3):473-9. © The Author(s) 2025. Oxford University Press가 미국감염학회 (Infectious Diseases Society of America)를 대신하여 출판. 모든 권리 보유. <https://doi.org/10.1093/cid/ciaf619>

참고문헌 (원문)

1. Morgan DJ, Rodman A, Goodman KE. How physicians can prepare for generative AI. *JAMA Intern Med*. doi:10.1001/jamainternmed.2025.4914.
2. Rodman A, Buckley TA, Manrai AK, Morgan DJ. Artificial intelligence vs clinician performance in estimating probabilities of diagnoses before and after testing. *JAMA Netw Open* 2023; 6:e2347075.
3. Goodman KE, Morgan DJ, Hoffmann DE. Clinical algorithms, antidiscrimination laws, and medical device regulation. *JAMA* 2023; 329:285-6.
4. Kanjilal S. Flying into the future with large language models. *Clin Infect Dis* 2024; 78:867-9.
5. Schwartz IS, Link KE, Daneshjou R, Cortés-Penfield N. Black box warning: large language models and the future of infectious diseases consultation. *Clin Infect Dis* 2024; 78:860-6.
6. Rodriguez-Nava G, Egoryan G, Goodman KE, Morgan DJ, Salinas JL. Performance of a large language model for identifying central line-associated bloodstream infections (CLABSI) using real clinical notes. *Infect Control Hosp Epidemiol* 2024; 46:305-8.
7. Perret J, Schmid A. Application of OpenAI GPT-4 for the retrospective detection of catheter-associated urinary tract infections in a fictitious and curated patient data set. *Infect Control Hosp Epidemiol* 2024; 45:96-9.
8. Wiemken TL, Carrico RM. Assisting the infection preventionist: use of artificial intelligence for health care-associated infection surveillance. *Am J Infect Control* 2024; 52:625-9.
9. Badia JM, Casanova-Portoles D, Membrilla E, Rubiés C, Pujol M, Sancho J. Evaluation of ChatGPT-4 for the detection of surgical site infections from electronic health records after colorectal surgery: a pilot diagnostic accuracy study. *J Infect Public Health* 2025; 18:102627.
10. Schreiber PW, Sax H, Wolfensberger A, Clack L, Kuster SP, Swissnoso. The preventable proportion of healthcare-associated infections 2005-2016: systematic review and meta-analysis. *Infect Control Hosp Epidemiol* 2018; 39:1277-95.
11. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006; 355:2725-32.
12. Centers for Disease Control and Prevention. National Healthcare Safety Network (NHSN). 2025. Available at: <https://www.cdc.gov/nhsn/index.html>. Accessed 12 August 2025.
13. Hansen S, Sohr D, Geffers C, et al. Concordance between European and US case definitions of healthcare-associated infections. *Antimicrob Resist Infect Control* 2012; 1:28.
14. Asia Pacific Society of Infection Control. 2022. APSIC guide for prevention of catheter associated urinary tract infections (CAUTIs). Available at: <https://apsic-apac.org/wp-content/uploads/2022/06/APSIC-CAUTI-guidelines-2022.pdf>. Accessed 6 October 2025.
15. Lin MY, Hota B, Khan YM, et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA* 2010; 304:2035-41.
16. Mitchell BG, Hall L, Halton K, MacBeth D, Gardner A. Time spent by infection control professionals undertaking healthcare associated infection surveillance: a multi-centred cross sectional study. *Infect Dis Health* 2016; 21:36-40.
17. Schrank GM, Branch-Elliman W, Leekha S, et al. Perceptions of health care-associated infection metrics by infection control experts. *JAMA Netw Open* 2023; 6:e238952.
18. Mayer J, Greene T, Howell J, et al. Agreement in classifying bloodstream infections among multiple reviewers conducting surveillance. *Clin Infect Dis* 2012; 55:364-70.
19. Larsen EN, Gavin N, Marsh N, Rickard CM, Runnegar N, Webster J. A systematic review of central-line-associated bloodstream infection (CLABSI) diagnostic reliability and error. *Infect Control Hosp Epidemiol* 2019; 40:1100-6.

20. Kim JJ, Kang H, Khan AN, et al. A comparison of National Healthcare Safety Network surveillance definitions to physician review for central-line-associated bloodstream infections. *Infect Control Hosp Epidemiol* 2023; 44:2062-4.
21. Worth LJ, Brett J, Bull AL, McBryde ES, Russo PL, Richards MJ. Impact of revising the National Nosocomial Infection Surveillance System definition for catheter-related bloodstream infection in ICU: reproducibility of the National Healthcare Safety Network case definition in an Australian cohort of infection control professionals. *Am J Infect Control* 2009; 37:643-8.
22. MacEwan SR, Gaughan AA, Beal EW, Hebert C, DeLancey JO, McAlearney AS. Concerns and frustrations about the public reporting of device-related healthcare-associated infections: perspectives of hospital leaders and staff. *Am J Infect Control* 2023; 51:633-7.
23. Howard-Anderson J, Morgan DJ. Moving beyond central line-associated bloodstream infections. *Ann Intern Med* 2024; 177:822-3.
24. van Mourik MSM, Perencevich EN, Gastmeier P, Bonten MJM. Designing surveillance of healthcare-associated infections in the era of automation and reporting mandates. *Clin Infect Dis* 2018; 66:970-6.
25. Streefkerk HRA, Verkooijen RP, Bramer WM, Verbrugh HA. Electronically assisted surveillance systems of healthcare-associated infections: a systematic review. *Euro Surveill* 2020; 25:1900321.
26. Branch-Elliman W, Sundermann AJ, Wiens J, Shenoy ES. Leveraging electronic data to expand infection detection beyond traditional settings and definitions (part II/III). *Antimicrob Steward Healthc Epidemiol* 2023; 3:e27.
27. Shenoy ES, Branch-Elliman W. Automating surveillance for healthcare-associated infections: rationale and current realities (part I/III). *Antimicrob Steward Healthc Epidemiol* 2023; 3:e25.
28. Mulder T, Kluytmans-van den Bergh MFQ, van Mourik MSM, et al. A diagnostic algorithm for the surveillance of deep surgical site infections after colorectal surgery. *Infect Control Hosp Epidemiol* 2019; 40:574-8.
29. Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural language processing for real-time catheter-associated urinary tract infection surveillance: results of a pilot implementation trial. *Infect Control Hosp Epidemiol* 2015; 36:1004-10.
30. Woeltje KF, Lin MY, Klompas M, Wright MO, Zuccotti G, Trick WE. Data requirements for electronic surveillance of healthcare-associated infections. *Infect Control Hosp Epidemiol* 2014; 35:1083-91.
31. Tvardik N, Kergourlay I, Bittar A, Segond F, Darmoni S, Metzger MH. Accuracy of using natural language processing methods for identifying healthcare-associated infections. *Int J Med Inform* 2018; 117:96-102.
32. Morgan D. Using an LLM for frontline providers to identify healthcare-associated infections across 11 hospitals. Presented at: Stanford Artificial Intelligence and Medical Imaging (AIMI) Symposium; 3 June 2025; Palo Alto, CA.
33. Alshanqeeti S, Coffey KC, McDermott K, et al. Comparing generative artificial intelligence vs. experts for detection of catheter-associated urinary tract infection. *Clin Infect Dis* 2026; 82:148-50.
34. Bouguettaya A, Stuart EM, Aboujaoude E. Racial bias in AI-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models. *NPJ Digit Med* 2025; 8:332.
35. Ayoub NF, Balakrishnan K, Ayoub MS, Barrett TF, David AP, Gray ST. Inherent bias in large language models: a random sampling analysis. *Mayo Clin Proc Digit Health* 2024; 2:186-91.
36. Omar M, Soffer S, Agbareia R, et al. Sociodemographic biases in medical decision making by large language models. *Nat Med* 2025; 31:1873-81.
37. Morgan DJ, Branch-Elliman W, Goodman KE. Time to study implementation of AI-generated discharge summaries. *JAMA Intern Med* 2025; 185:825-6.
38. Branch-Elliman W, Sundermann AJ, Wiens J, Shenoy ES. The future of automated infection detection: innovation to transform practice (part III/III). *Antimicrob Steward Healthc Epidemiol* 2023; 3:e26.
39. Branch-Elliman W, Lamkin R, Shin M, et al. Promoting de-implementation of inappropriate antimicrobial use in cardiac device procedures by expanding audit and feedback: protocol for hybrid III type effectiveness/implementation quasi-experimental study. *Implement Sci* 2022; 17:12.
40. Goodman KE, Shams SM, Magder LS, Baghdadi JD, Morgan DJ, Harris AD. Generative AI-based surveillance for avian influenza across a statewide healthcare system. *Clin Infect Dis* 2025; 81:900-3.
41. Yuan K, Yoon CH, Gu Q, et al. Leveraging transformers and large language models with antimicrobial prescribing data to predict sources of infection for electronic health record studies. *medRxiv* 24305966 [Preprint]. April 19, 2024. Available from: <https://doi.org/10.1101/2024.04.17.24305966>.
42. Goodman KE, Robinson ML, Shams SM, et al. Identification of long-term care facility residence from admission notes using large language models. *JAMA Netw Open* 2025; 8:e2512032.