

```
class Node:
    def __init__(self, value):
        self.value = value
        self.left = None
        self.right = None

class BinaryTreeNode:
    def __init__(self, value):
        self.root = None

    def insert(self, value):
        new_node = Node(value)
        if self.root is None:
            self.root = new_node
            current_node = self.root
        else:
            while True:
                if value < current_node.value:
                    if current_node.left is None:
                        current_node.left = new_node
                        break
                    else:
                        current_node = current_node.left
                else:
                    if current_node.right is None:
                        current_node.right = new_node
                        break
                    else:
                        current_node = current_node.right

    def search(self, value):
        current_node = self.root
        while current_node:
```

AI

풀스택 AI 옵저버빌리티가 요구되는 고급 AI 전략

대규모 투자가 이뤄진 AI는 성능, 위험 및 결과에 대한 강력한 모니터링이 필요합니다.

AI 가시성의 도전과제

모든 업계의 리더들은 고객 서비스 향상, 콘텐츠 생성, 맞춤형 추천, 어시스턴트, 스마트 챗봇, 위험/부정 행위 예측 및 탐지, 프로세스 자동화 등 생성형 AI로부터 이미 다양한 혜택을 누리고 있습니다.

하지만 새로운 역량에는 새로운 기회와 더불어 새로운 위험과 도전과제가 따르기 마련입니다.

AI에 큰 기대를 걸고 있는 기업은 AI 솔루션의 모든 구성 요소를 반드시 이해해야 합니다. AI는 고객 경험, 매출, 비용, 사이버 보안, 브랜드 인지도 등 경영진과 이사회가 관심을 갖는 모든 영역에 잠재적으로 막대한 영향을 미치기 때문입니다. IT 리더와 개발자들 뿐만 아니라 경영진도 AI 모델의 작동 방법, 관련 비용, 잠재적 취약성, 상호 작용 방식을 알고 있어야 합니다.

간단한 일처럼 들릴지 모르지만, AI 기술 스택이 점점 더 복잡해지는 오늘날의 환경에서는 결코 쉽지 않은 과제입니다.

AI 솔루션은 조직에 대규모 언어 모델(LLM)은 물론 고급 데이터 저장소, 데이터 파이프라인, 오케스트레이션 프레임워크 및 머신러닝(ML) 코드 라이브러리 등 새로운 프레임워크와 구성 요소를 도입합니다. 많은 조직에게 이러한 시스템의 문제를 해결하는 것은 그 자체로 완전히 새로운 문제가 될 수 있습니다.

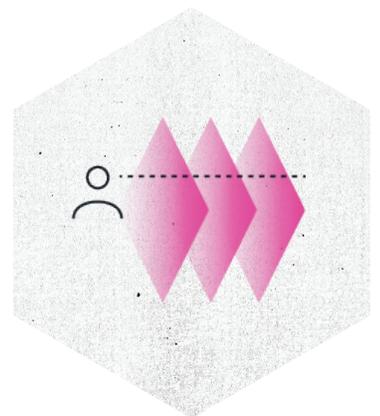
일상적인 행동을 모니터링하는 것조차 까다로울 수 있습니다. AI 사용 사례가 등장하기 전에 만들어진 툴을 사용하는 경우는 더욱 그렇습니다. 또한 각 시스템마다 고유한 API, 고유한 데이터 보고 방식, 고유한 인터페이스가 있으므로 모니터링 툴을 각 모델에 맞게 조정해야 합니다.

“거의 모든 조직이 AI 애플리케이션을 기술 스택에 통합하여 고객 경험과 효율성을 향상하고 수익을 개선하기를 희망합니다. AI는 기술 스택에 복잡성을 야기하기 때문에 보안, 품질, 규정 준수 및 비용을 고려해 책임감 있게 도입해야 합니다.”

스티븐 엘리엇(Stephen Elliot)
IDC 그룹 부사장

이러한 AI 시스템의 성능 및 비즈니스 이점과 관련해 신뢰할 수 있는 데이터가 없으면, 기업은 새로운 AI 투자에 대해 정보에 입각한 결정을 내리거나 이러한 투자를 비즈니스 결과나 가치와 연계시키는 데 어려움을 겪을 것입니다.

AI 성능과 이점에 대한 신뢰할 수 있는 데이터가 없으면, 기업은 정보에 입각해 결정을 내릴 수 없습니다.



위험 관리와 ROI 증대

투자 수익률(ROI)은 AI 리더들에게 또 다른 중요한 관심사입니다. AI에 대한 투자는 기업 뿐만 아니라 모든 산업을 변화시키고 있습니다. 세콰이어 캐피탈(Sequoia Capital)의 최근 추정치에 따르면, 전 세계 기업들은 2024년에 엔비디아 AI 인프라에 너무 많은 비용을 지출했기 때문에 지출을 정당화하기 위해서는 평생 AI 관련 수익으로 6,000억 달러를 창출해야 합니다. 단 1년 동안의 투자였는데 말입니다.

대부분의 기업들이 아직 AI에 그렇게 많은 투자를 하고 있지는 않지만, ROI를 입증하는 것은 모든 조직에 중요해질 것입니다. 특히 AI에 대한 과도한 기대치가 어느 정도 가라앉고 고위 경영진과 이사회가 AI로부터 실질적인 결과를 요구하기 시작하면 더욱 그러해질 것입니다. 그러나 많은 리더들은 AI 스택에 들어가는 모든 구성 요소와 비용을 알지 못하므로 정확한 계산이 어려울 수 있습니다.

지출을 정당화하려면, AI 모델이 측정 가능한 가치를 제공하고 투자에 버금가는 수익을 창출해야 합니다.



해결책: 풀스택 AI 모니터링

올바른 AI 모니터링 솔루션은 CTO와 CIO가 시기 적절하게 의사 결정을 하고 비용을 관리하며 ROI를 측정 및 극대화하는 데 필요한 데이터를 제공해야 합니다. 올바른 솔루션을 갖추면, 리더들은 서비스 및 인프라와 함께 AI 기술 스택의 모든 구성 요소에서 더 효과적으로 안정성, 품질 및 효율성을 보장할 수 있습니다.

이를 위해서는 AI 스택을 위해 특별히 설계된 옹저버빌리티 솔루션이 필요합니다.

그러나 대부분의 AI 모니터링은 LLM 계층에 중점을 둡니다. AI 스택에 관련된 다른 많은 구성 요소가 있다는 점에 유의해야 합니다.

LLM 계층에만 집중하면 AI 시스템의 성능과 비용에 영향을 줄 수 있는 모든 잠재적인 문제를 이해하지 못할 수 있습니다. 예를 들어, 성능의 측면에서 서버, 애플리케이션 코드 및 벡터 데이터베이스는 모두 문제에 영향을 미칠 수 있기 때문에 총체적으로 고려해야 합니다.

포괄적인 AI 모니터링

리더들은 애플리케이션 계층에서 LLM, AI 계층, 사용자 인터페이스까지, 성능, 오류 및 비용을 포괄적으로 모니터링하는 풀스택 옹저버빌리티 기능이 필요합니다. 전체 디지털 자산에서 실행되는 AI 앱의 수와 전체 비용, 전체 오류율을 확인할 수 있어야 합니다.



LLM을 위한 모든 옹저버빌리티 솔루션에는 몇 가지 기본 메트릭이 포함되어 있으며, 모든 AI 모니터링 솔루션은 기본적으로 이러한 메트릭을 모니터링해야 합니다.

- ✓ 성능
- ✓ 오류
- ✓ 비용

다음은 AI 애플리케이션에서 자주 발생하는 몇 가지 오류입니다.

- 클라우드 인프라에 할당된 메모리가 부족하여 애플리케이션이 충돌할 수 있습니다.
- 디스크 입출력(I/O) 병목 현상으로 인해 앱의 성능이 느려지고 사용자 경험이 저하될 수 있습니다.
- LLM이나 기타 스택 구성 요소의 새 릴리스가 예기치 않은 성능 문제나 설명할 수 없는 오류를 야기할 수 있습니다.

이러한 문제는 인프라와 관련되어 있지만 애플리케이션과 사용자 환경에서 먼저 나타날 수 있습니다. 완전한 모니터링 솔루션은 이러한 유형의 문제를 정확하게 추적하고 진단할 수 있도록 포괄적인 옴저버빌리티를 제공해야 합니다.

이는 서비스 수준 협약(SLC)을 관리하는 기업에게 점점 더 중요해지고 있으며, 다양한 규정과 기타 최종 사용자 약정을 준수하는 데 도움이 될 것입니다.

시장 변화에 대한 대응

AI 기술은 급속하게 발전하고 있기 때문에 보조를 맞추기가 어렵습니다. 연구 개발 예산이 10억 달러에 달하는 기업도 항상 최신 기술을 따라잡는다고 확신할 수 없습니다.

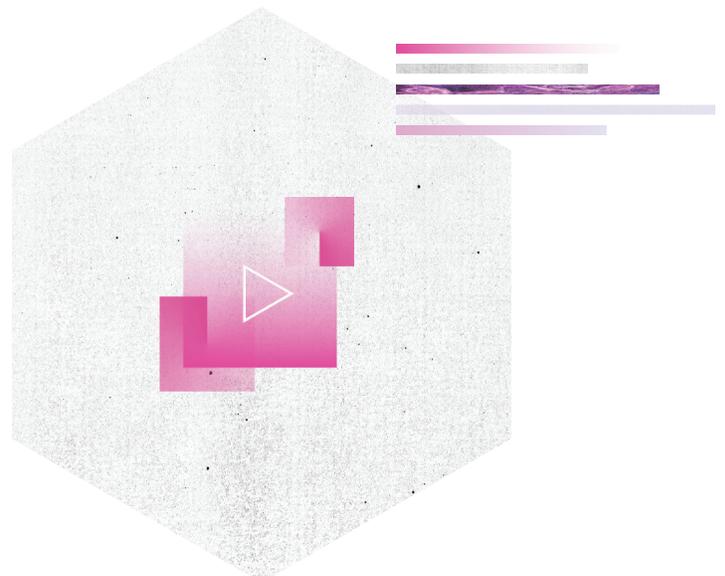
지금 사용하고 있는 AI 스택의 구성 요소가 1년 후나 6개월 후에도 계속 효과적이지는 않을 것이라는 것만은 확실합니다. 확실한 것은 변화가 지속될 것이라는 사실 뿐입니다.

스마트한 기업은 성장하고 변화하는 인프라에 새로운 구성 요소를 통합할 수 있도록 유연하고 확장 가능한 AI 모니터링 솔루션을 찾을 것입니다. 또한 너무나 많은 이종 AI 스택이 존재하기 때문에, OpenAI, Azure OpenAI Service, Amazon Bedrock, Anthropic, LangChain, Nvidia NIM, Groq, Llama를 비롯한 오늘날의 주요 LLM 공급업체와 프레임워크를 모두 포괄하는 솔루션을 찾아야 합니다.

AI 환경에는 많은 옵션이 존재하며, 전체 디지털 자산을 모니터링하는 것이 중요합니다.

예를 들어, 대부분의 사용 사례를 위해 OpenAI 라이선스 계약을 보유하고 있지만 엔지니어가 보다 기밀한 프로젝트를 위해 Bedrock을 사용할 필요가 있다고 판단하는 경우 어떻게 해야 할까요?

고급 모니터링 기능은 기업의 주요 사용 사례 이외에 있는 LLM 및 AI 코드 라이브러리에 대해서도 가시성과 취약성 관리를 제공할 수 있어야 합니다.



품질 모니터링

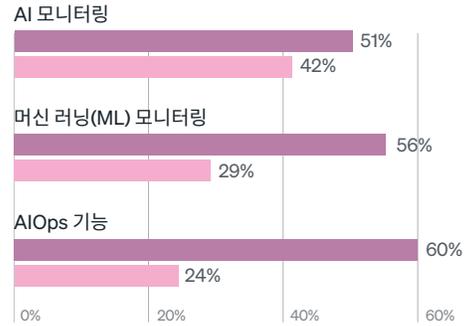
기업은 성능, 비용, ROI 그 이상을 모니터링하고, AI 응답의 품질을 측정할 필요가 있습니다. 이는 AI 모델에서 생성되는 독성, 편향 등 조직에 이러한 톨의 품질과 공정성을 모니터링하고 알리는 데 중요합니다.

실제로, AI는 여러 고유한 위험을 초래합니다. AI 리더들은 LLM이 이야기할 수 있는 편향, 환각, 독성 등의 품질 문제를 선제적으로 관리할 수 있도록 팀원들에게 역량을 제공해야 합니다. 사용자 불만이 제기될 때 어떻게 신속하고 적절하게 대응할 수 있을까요? 더 나아가, 어떻게 AI 모델의 응답이 신뢰할 수 없는 것임을 감지하고 사용자에게 영향을 미치기 전에 대응할 수 있을까요?

사용자의 기대치에는 LLM이 반환한 텍스트의 독창성과 사실의 정확성에 대한 가정이 포함됩니다. AI 앱이 진실하고 독창적인 응답을 제공할 수 없는 경우, LLM, 데이터 저장소, 사용 중인 ML 라이브러리, 애플리케이션 코드 등 모든 소스에서 문제를 진단하는 데 도움이 되는 모니터링 솔루션이 필요합니다.

장기적으로 AI 리더들은 AI 스택을 지속적으로 개선하여 시간이 지남에 따라 이러한 문제의 위험을 줄이는 방법을 원합니다.

2024년부터 2027년까지 옵저버빌리티 기능의 배포



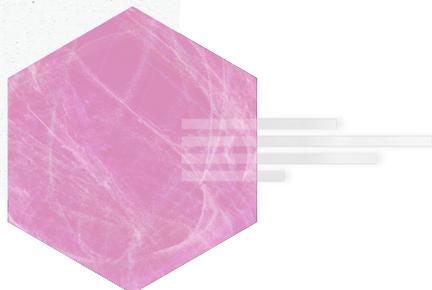
■ 3년 이내 배포 계획
■ 현재 배포됨

2024 옵저버빌리티 전망

AI 모니터링 체크리스트

풀스택 AI 모니터링 솔루션은 LLM 뿐만 아니라 AI 스택의 모든 계층에 대한 포괄적인 옵저버빌리티를 제공해야 합니다.

- ✓ 오케스트레이션 프레임워크
- ✓ LLM
- ✓ ML 라이브러리
- ✓ 모델 제공
- ✓ 벡터 데이터베이스
- ✓ AI 인프라



AI를 어떻게 모니터링할 수 있을까요?

전문적인 AI 옹저버빌리티를 통해 기업은 AI 워크플로우에 대한 포괄적인 가시성을 확보할 수 있습니다. 옹저버빌리티는 다양한 접근 방식과 플랫폼의 문제를 해결하고, 비교 및 최적화하는데 필요한 인사이트를 제공합니다. 이러한 인사이트로 AI 기반 제품을 개선하여, 고객이 대화하고 싶어하는 챗봇이나 진정으로 지능적이고 직관적인 AI 기반 사용자 인터페이스 같은 완전히 새로운 고객 경험을 조성할 수 있습니다.

또한 풀스택 AI 모니터링은 비용 관리, 성능 개선, 결함 감소, 위험 최소화, ROI 향상을 지원합니다.

42%의 기업이 이미 AI 모니터링을 구현했으며, 51%는 3년 이내에 이를 구현할 계획이라는 것은 놀라운 일이 아닙니다. AI 모니터링은 이제 하나의 경쟁력으로 자리를 잡아가고 있습니다.

CTO와 CIO는 전체 AI 스택을 포괄적으로 모니터링할 수 있는 역량을 팀에게 제공해야 합니다. AI 모니터링은 개별 LLM에 대한 심층 분석, 여러 LLM의 성능 비교, 비용 관리, 프롬프트 및 응답 추적, 전체 스택의 성능 최적화, 그리고 편향, 환각 및 독성 같은 품질 문제 관리에 필요한 역량을 제공합니다.

AI 애플리케이션 생태계에 대한 광범위하고 포괄적인 인사이트를 통해, 리더들은 AI 성능을 최적화하고 품질을 높이며 비용을 관리하고 ROI를 달성할 수 있는 최적의 입지를 확보할 수 있습니다.

"AI 애플리케이션에 옹저버빌리티를 적용하는 것은 애플리케이션의 복잡성을 해결하는 스마트하고 효율적인 방법으로, 이를 통해 기업은 혁신을 추진하고 확장할 수 있습니다. 이러한 옹저버빌리티 솔루션은 궁극적으로 조직이 더 나은 제품과 고객 경험을 제공할 수 있도록 지원합니다."

스티븐 엘리엇(Stephen Elliot)
IDC 그룹 부사장

자세히 보기

추가 자료

[AI 및 옹저버빌리티 보고서](#)
[2024년 옹저버빌리티 전망](#)
[뉴렐릭 SNS](#)

