

# AI 들여다보기

## 2025년 임팩트 보고서

실제 AI 구현 사례에서 통찰을 얻으세요.

## 개요

코드는 문제의 절반에 불과합니다. 견고한 옴저버빌리티 (시스템이 실시간으로 어떻게 동작하는지 보여주는 포괄적인 로그, 메트릭, 트레이스) 역량이 없다면, 아무리 훌륭한 애플리케이션이라도 대규모 확장을 하면 그 속을 알 수 없고 신뢰할 수 없는 암흑 상자가 됩니다. 인공지능(AI)이 추측에 근거한 약속에서 기반 기술로 자리잡은 데서 중요한 교훈을 얻을 수 있습니다.

AI 도입은 전례 없는 속도로 이뤄지고 있으며, 조직에서 AI를 사용하는 패턴도 그만큼 빠르게 진화하고 있습니다. 개발자의 행동이 어떻게 AI 환경을 지속적으로 변화시키고 있는지, 왜 AI 모니터링이 선택 사항에서 필수 사항으로 바뀌었는지를 직접 확인할 수 있도록 뉴렐릭이 집계된 사용 데이터를 처음으로 공개했습니다.

뉴렐릭의 데이터에 따르면 시중에 나와 있는 AI 툴 중 ChatGPT의 사용률이 가장 높습니다. 고객과 개발자들은 비용이 더 낮은 대안이 있는 경우에도 GPT-4o와 같은 최신 버전으로 빠르게 전환하고 있습니다. 동시에 GPT와 함께 사용할 맞춤형 모델을 만드는 개발자가 급증하고 있는데, 이는 앞으로 더 많은 팀들이 틈새 도구를 선택하고 AI를 특정 요구 사항에 맞게 조정하기 시작할 것임을 시사합니다.

이러한 추세는 근본적인 전환점을 의미합니다. AI는 실험 단계에서 실제 운영 단계로 빠르게 이동하고 있습니다. 이렇게 역동적인 환경에서는 기본적인 작업별 모니터링 도구만으로는 안되기 때문에, 뉴렐릭의 AI 모니터링 솔루션을 도입하는 경우가 늘고 있습니다. 이러한 추세가 지속됨에 따라 AI 모니터링이 이제 개발자, 인프라 팀, CTO 모두에게 기술적 민첩성과 운영 복원력을 제공하기 위한 핵심 요소라는 점이 명확해지고 있습니다.

고객이 도입한 AI 프롬프트 툴 중에서 ChatGPT의 사용률이 압도적입니다.



## 지배적 위치를 확보한 ChatGPT

개발자들은 AI 모델의 실험과 전문화 분야에서 매우 강력한 성장을 추진하고 있습니다. 뉴렐릭은 최근 분기에 AI 앱 전반에서 사용되는 고유한 대규모 언어 모델(LLM)의 수가 92%나 증가한 것을 확인했습니다.

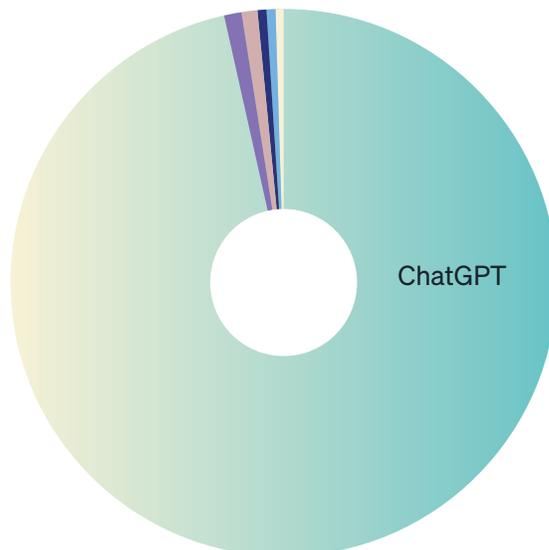
더 자세히 들여다보면, 개발자들은 OpenAI의 ChatGPT가 주도하는 가장 큰 LLM을 수용하고 있습니다. 이는 뉴렐릭 고객이 처리한 모든 LLM 토큰 중 **86% 이상**을 차지하며, Meta의 Llama가 2위로 그 뒤를 따르고 있습니다. ChatGPT의 엄청난 점유율은 개발자들이 최소한의 설정이나 튜닝과 더불어 사용 편의성, 안정성, 최고 수준의 성능을 우선시하고 있음을 강력하게 시사합니다. 현재로서는 기업들이 대안을 구축하거나 미세 조정하는 데 시간을 투자하기보다는 엔터프라이즈급 모델을 사용하여 신속하게 배포하기를 원하기 때문이라고 볼 수 있습니다.

가장 많이 사용되는 ChatGPT에서도 옹저버빌리티는 선택 사항이 아닙니다. 팀은 여전히 사용량을 추적하고, 성능을 벤치마킹하고, 이상 징후를 감지하여 비용 효율성, 최적의 성능 및 안정성을 보장해야 합니다.

일반적으로 많은 조직들이 아직 AI 배포의 초기 단계에 있기 때문에, 일반 모델에 의존하는 것은 당연합니다. 그러나 데이터에 따르면 개발자들은 규모는 작지만 도메인별, 작업별 모델 등 다양한 모델을 테스트하고 있는 것으로 나타났습니다. 팀과 조직이 점점 더 정교함을 추구하면서 성과를 극대화하고, 비용을 관리하며, 거버넌스 규정을 준수하고, 다중 모델 환경을 탐색하기 위한 고급 AI 모니터링에 대한 필요성이 향후 몇 년 안에 큰 폭으로 증가할 것으로 예상됩니다.

>86%

뉴렐릭 고객이 처리한 모든 LLM 토큰 중 ChatGPT를 사용한 고객 비율



관찰된 LLM 토큰 소비 기준 가장 인기 있는 모델

- ChatGPT
- Meta Llama
- Ada
- TheDrummer/Unslop
- Anthropic Magnum
- Anthropic

## 빠르게 전환하는 개발자들

AI 혁신의 속도가 빨라지면서 개발자의 행동도 유동적으로 변해, 모델 버전의 전환이 빠르게 일어나고 있습니다. 최근에는 ChatGPT-4o가 우세를 보이고 있으며, ChatGPT-4o mini가 그 뒤를 따르고 있습니다. ChatGPT-4o mini가 ChatGPT-4o보다 비용이 낮다는 점을 고려하면, 개발자들이 최첨단 성능과 기능(예: 더 나은 멀티모달 기능, 낮은 레이턴시)을 비용 절감보다 더 우선시한다는 사실을 알 수 있습니다. 적어도 지금 당장은 그렇습니다. 시간이 지나면서 사용자들이 비용을 최적화하기 시작할지는 지켜볼 일입니다.

일부 경우, ChatGPT 버전 간의 전환이 하룻밤 사이에 이루어집니다. 예를 들어, 뉴렐릭 사용자들은 최근 ChatGPT 3.5 Turbo에서 4월에 발표된 ChatGPT 4.1 mini로 빠르게 갈아타고 있습니다. 기업들이 경쟁력을 유지하기 위해 OpenAI의 최신 혁신을 면밀히 주시하고 있다는 것은 분명해 보입니다.

이처럼 빠르게 변화하는 환경에서 조직은 끊임없이 진화하는 AI 스택과 보조를 맞출 수 있는 AI 모니터링이 필요합니다. 시간이 지나면서 OpenAI가 각 모델에 대한 증분 버전과 업데이트를 출시하면 그 필요성은 더욱 커질 것입니다. 비용, 안정성 또는 규정 준수에 대한 통제력을 유지하면서 가시성을 높이려는 개발자 팀은 새로운 모델을 평가해야 합니다.

---

ChatGPT 버전 간 전환은 신속하게 이뤄집니다.

끊임없이 진화하는 AI 스택에 보조를 맞출 수 있는 AI 모니터링이 필요합니다.

## 급증하는 AI 모니터링

작년 출시 이후, 기업들은 AI 모델의 안정성, 정확성, 규정 준수 및 비용 효율성을 보장하기 위해 뉴렙릭 AI 모니터링을 도입하고 있습니다. **지난 12개월 동안 이전 분기 대비 뉴렙릭 사용량은 30%씩 꾸준히 증가해 왔습니다.**

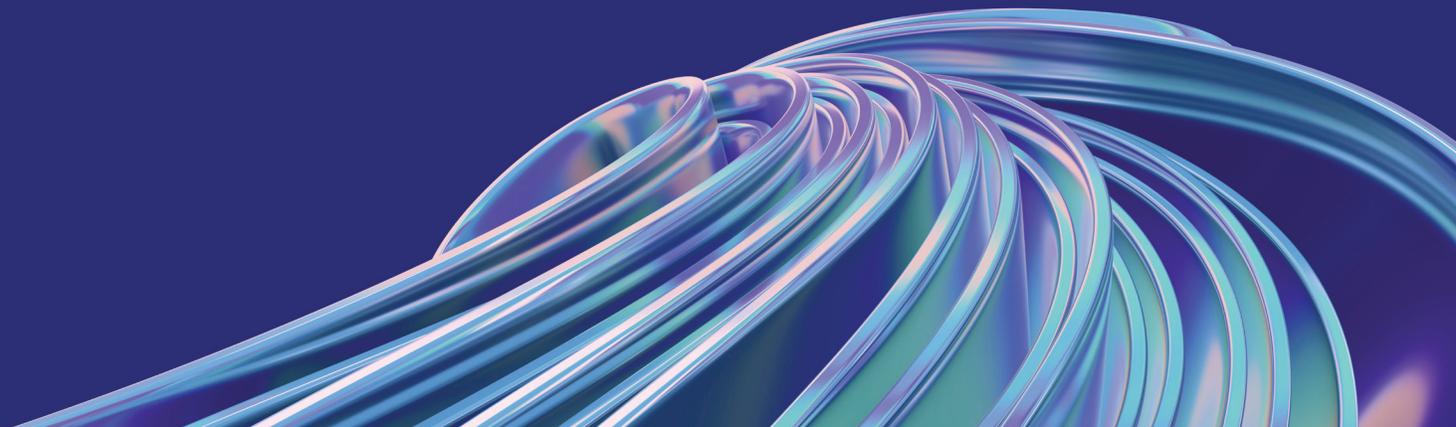
뉴렙릭의 2024 옹저버빌리티 전망 보고서에 따르면, AI의 빠른 도입이 조직들이 AI 모니터링 솔루션을 구현하는 주요 동인입니다. AI가 실험적 단계에서 운영 단계로 전환됨에 따라 기업들은 새로운 모니터링 도전과제에 직면하게 되었습니다. 이러한 진화는 대규모 언어 모델을 넘어서는 포괄적인 AI 옹저버빌리티를 요구합니다. 옹저버빌리티 시스템은 기존 운영 인프라에 완전하게 통합된 전체 AI 관련 기술 생태계를 포괄해야 합니다.

한편, 견고한 AI 모니터링 솔루션 없이 AI 모델을 사용하는 기업들은 신뢰성, 응답의 정확성, 보안, 비용 급증의 측면에서 상당한 위험에 노출됩니다. 적절한 모니터링이 수반되지 않는 AI는 로그 없이 소프트웨어를 배포하는 것과 같습니다. 조직은 AI 거버넌스에 대한 준비가 부족해 성능 문제를 감지하지 못하고, 장애에 취약하다는 의미입니다. 이런 경우 AI 혁신의 속도가 느려지고 AI 확장에 대한 신뢰가 떨어질 수 있습니다. 조직은 쉽게 설정이 가능하고 시간이 지나면서 구성을 조정할 수 있으며 데브옵스(DevOps)부터 임원까지 모든 사용자들에게 직관적인 경험을 제공하는 통합 AI 모니터링 솔루션이 필요합니다.



**새로운 AI 도구는 뉴렙릭에 연결되어 업타임, 성능 및 응답 시간을 모니터링합니다. 이를 통해 개발자들이 사용자용 제품의 기능에 집중할 수 있게 되어 비용이 절감됩니다."**

자레드 정(Jared Chung)  
전무 이사, CareerVillage.org



## 지배적인 위치에 있는 Python

고객들은 뉴렐릭으로 AI 애플리케이션을 관찰합니다. 요청 규모와 고객 도입을 놓고 볼 때 Python은 AI 애플리케이션 분야에서 압도적인 우위를 점하고 있으며, Node.js가 그 뒤를 따르고 있습니다. Python의 고객 도입률은 지난 분기 이후 약 45% 증가했습니다.

AI가 기업 운영에 깊게 자리 잡으면서, AI 애플리케이션을 위한 프로그래밍 언어의 범위도 필연적으로 확장될 것입니다. 이러한 다각화는 정해진 수순입니다. 기업은 AI를 기존 시스템(보통 Java나 C#과 같은 언어로 구축됨)과 통합하고, 특정 성능 및 확장성 요구 사항(C++, Go 또는 Rust의 도입 촉진)을 해결하고, 클라우드에서 엣지 디바이스까지 다양한 환경에 AI를 배포해야 하기 때문입니다.

## Java 사용량 증가

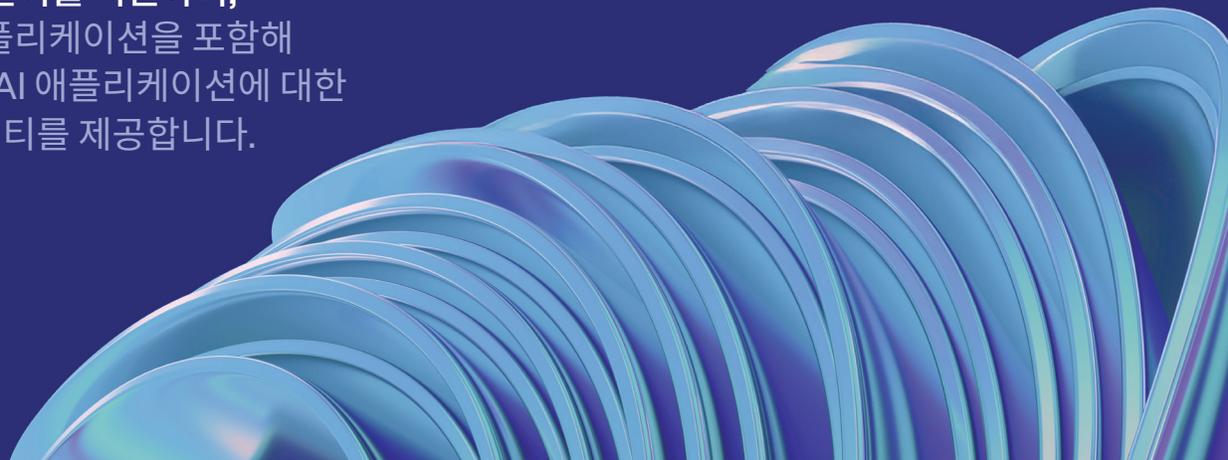
실제로 뉴렐릭은 최근 Java 사용량이 급격히 증가하여, 지난 분기 대비 34%나 늘었음을 확인할 수 있었습니다. 이러한 추세는 주시할 가치가 있으며, 대기업에서 더 많은 프로덕션급 Java 기반 LLM 애플리케이션이 나올 것임을 예상할 수 있습니다.

가장 큰 추진력과 지원, 도구를 갖추고 있다는 점을 감안할 때, Python이 도입과 사용의 측면에서 선두의 위치에 있는 사실은 놀랍지 않습니다. 그러나 다른 언어들도 툴링이 계속해서 성숙해져, 시간이 지남에 따라 더욱 보편화될 것입니다. 뉴렐릭이 가장 많은 언어를 지원하고 LLM을 사용하는 애플리케이션을 포함해 Python으로 구축된 AI 애플리케이션에 대한 포괄적인 오피저버빌리티 역량을 제공하는 이유도 바로 이 때문입니다.

**34%**

지난 분기 이후  
Java 사용량 증가

뉴렐릭은 가장 많은 언어를 지원하며, LLM을 사용하는 애플리케이션을 포함해 Python으로 구축된 AI 애플리케이션에 대한 포괄적인 오피저버빌리티를 제공합니다.



## 조사방식

이 보고서는 2024년 4월 30일부터 2025년 4월 30일 사이에 수집된 85,000명의 활성 뉴렐릭 고객으로부터 집계되고 익명화된 사용 데이터를 기반으로 작성되었습니다.

## 뉴렐릭 소개

뉴렐릭의 지능형 오피저버빌리티 플랫폼은 기업이 중단 없는 디지털 경험을 제공하는 데 도움을 줍니다. 뉴렐릭은 텔레메트리 데이터를 통합하고 상호 연관시켜 전체 디지털 자산에 대한 명확한 가시성을 제공하는 유일한 플랫폼입니다. 뉴렐릭은 적시에 올바른 데이터를 처리함으로써 기업이 문제 사후 대응을 예측과 선제적 대응으로 전환하여 가치를 극대화하고 비용을 절감할 수 있도록 지원합니다. Adidas Runtastic, Domino's, GoTo Group, Ryanair, Topgolf, William Hill 등 세계 유수 기업들이 뉴렐릭을 사용해 혁신을 추진하고 안정성을 높이며 탁월한 고객 경험을 제공하여 성장을 가속화하고 있습니다.

[뉴렐릭 AI 모니터링에 대해 자세히 알아보기](#)

