# IBM, 인공 지능(AI) 및 양자 컴퓨팅 목표 확장

(2023.12.06., 양자정보연구지원센터)

## □ IBM, 미래에도 인공 지능 및 양자 컴퓨팅 분야에서 혁신 강조

○ 최근 분석가의 날 행사에서 미래에도 **인공 지능(AI) 및 양자 컴퓨팅** 분야에서 혁신을 이어갈 자신이 있음을 강력히 주장함

- 연구와 상용 제품 간의 간극을 줄이고, CEO 명령에 따라 특정 부냐에 중점을 둠으로써 가능해짐
- 이전에는 연구 노력 중 일부가 상용화되지 못한 적도 있지만, 현재는 두 그룹 간 조율이 강화되어 중요한 기술이 신속하게 제품화되고 있음

○ 전략적 이니셔티브가 실제로 미치는 영향으로 **watsonx, AI 도구 패키지**가 빠르게 발전하고 있으며, 다양한 응용 분야에서 중요한 신기술이 상용화되고 있음

- **응용 프로그램** 측면에서 디지털 노동 또는 HR 관련 활동, 고객 관리 또는 고객 지원, 앱 현대화 또는 코드 생성이라는 세가지 주요 범주로 분류됨, 많은 기업들이 watsonx와 genAI 채택하고 있음
- 애플리케이션 외에도 성능 및 규모, 모델 사용자 정의, 거버넌스 및 애플리케이션 활성화에 대한 노력 논의
- **성능 측면**에서 대규모 기반 모델이 효율성 향상을 위해 다양한 새로운 방법 연구 중, 양자화를 통해 모델 크기 축소하고 GPU 분할을 통해 제한된 컴퓨팅 리소스 공유하는 기능을 개선함
- Pytorch 2.0 같은 기술적 노력을 통해 하드웨어 추상화 계층 향상시켜 GenAI 모델의 성능 최적화, 다양한 컴퓨팅 칩 아키텍쳐에서 실행할 수 있는 확장 가능성이 열림

○ **모델 사용자 정의** 측면에서도 많은 노력을 기울이고 있음

- LoRA(Low Rank Adaptation), 매개변수 효율적인 튜닝, 다중 작업

프롬프트 튜닝 등 watsonx 내에서 상용화될 것으로 예상됨, 모델 구축 프로세스에서 교육적 지침 제공할 필요성 설명

- 거버넌스에 대한 IBM의 노력, 모델이 구축되고 진화하는 방법, 모델 생성에 사용된 데이터 등에 대한 세부 정보 추적 및 보고 (차별화 기능)
- 위험 평가 및 예방에 관한 작업, 일반적으로 GenAI 기술의 신뢰와 신뢰성에 대한 기업의 주요 우려 사항과 관련하여 IBM이 시작을 선도하는 위치에 있음

○ 애플리케이션 활성화 영역에서 RAG(Retrieval Augmented Generation) 중심 수행 중인 작업

- RAG는 추론 프로세스 강화, 기업이 자체 데이터 활용에 훨씬 더 쉽고 비용 효율적으로 만드는 비교적 새로운 기술임

○ 2030년까지 확장되는 상세한 기술 로드맵을 선보임

- IBM은 양자 컴퓨팅이 너무 극적이고 미래 지향적인 기술이어서 많은 잠재 고객이 이를 계획할 수 있는 방법을 알아야 할 필요성 강조
- 양자 컴퓨팅 분야에서 미래 기술 로드맵을 자세히 공개하며 신뢰할 수 있는 기업으로 입지를 강화하고 있음
- IBM은 혁신의 의지가 여전히 살아 있으며 미래에 대한 자신감을 높이고 있음

**(원문)**

1. https://seekingalpha.com/article/4651508-ibm-extends-goals-ai-quantum-computing