1. 첫번째로 장비의 성능의 한계에 있다. 기존 딥러닝 모델은 GPU나 높은 연산 자원을 기반으로 설계되었으나, IoT 디바이스는 **CPU, 메모리, 배터리 용량이 제한적**임. 또한 대개 IoT장비들은 저전력 구동에 무게를 두고 설계되었기에 배터리의 한계도 있다. 두번째는 보안의 문제가 있는데 자체 로컬에서 모델을 구현하는 것에 한계가 있었기에 주로 클라우드에서 딥러닝, 머신러닝이 구동되었는데 이러면 중간에 모든 데이터를 주고 받으면서 보안에 문제가 생길 수 있다는 단점이 있습니다. 마지막으론 앞서 말했듯이 통신으로 연결을 하면 필연적으로 생기는 통신 지연이 있고 더욱이 저전력 통신모델(예를 들어 zigbee, Bluetooth 등)을 쓰다 보니 일반적인 모바일 디바이스 같은 것들 보다 지연이 더 심하고 네트워크에 연결이 되어 있지 않으면 아무것도 할 수 가 없는 문제도 있다.
2. 이에 대한 해결책엔 첫번째로 모델 경량화가 있다. 기존 고성능 컴퓨터에 맞춰진 딥러닝, 머신러닝 모델에서 불필요한 파라미터를 제거하고 데이터의 수를 줄여 모델의 연산량을 줄이는 모델 압축을 할 수 있다. 이러면 비록 정확도는 컴퓨터에서 구동하는 것에 비해 정확도는 떨어질 수 있으나 IoT장비 로컬에서도 돌아가게끔 만들 수 있다. 두번째로는 기존의 컴퓨터에 구축되어 있는 시스템을 쓰는 것이 아닌 IoT장비를 기반으로 경량화 된 알고리즘을 설계하는 방법이 있다. 이는 작고 빠른 신경망 구조를 사용하여 높은 효율을 내도록 만들 수 있다. 최근에 발표된 딥시크와 챗지피티의 사례를 보면 더 낮은 하드웨어 성능을 갖더라도 크게 뒤지지 않는 모델을 만들 수 있음을 볼 수 있다. 마지막으로 위에서 계속 말 했던 로컬 구동이 있는데 앞서 예시를 든 딥시크처럼 클라우드가 아닌 로컬에서 돌릴 수 있게끔 설정을 하면 통신과정이 생략되기에 레이턴시 문제도 없을 뿐 아니라 네트워크 연결과 관계없이 사용이 가능하고 사용자의 행적이 밖으로 새나가지도 않기에 보안적인 측면도 해결 할 수 있다.