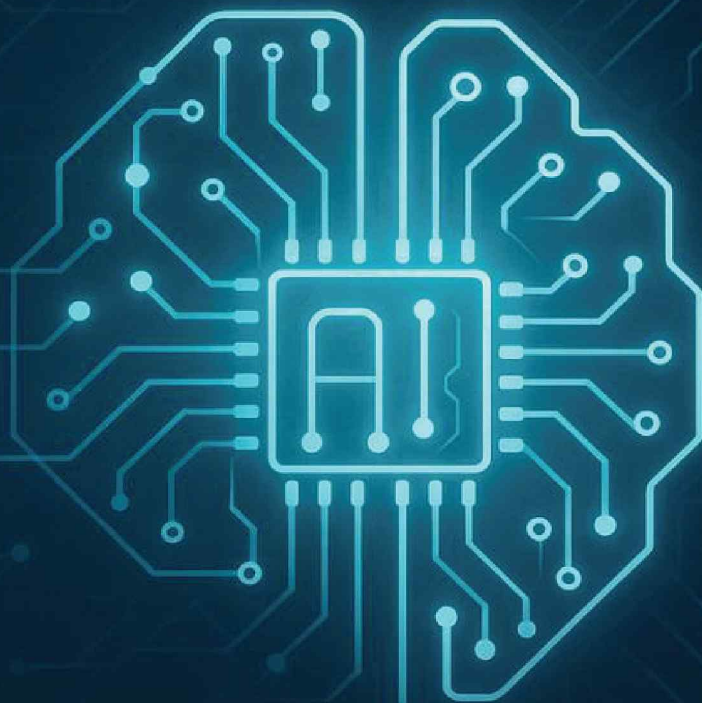


AI NETWORK STRATEGIES

AI 네트워크 전략



입체통신연구소

ETRI 한국전자통신연구원
Electronics and Telecommunications
Research Institute

CONTENTS

AI 네트워크 전략 보고서

01. 개요 1

1.1 AI 네트워크 개념 및 정의	1
1.2 AI 네트워크 연구개발 필요성	3
1.3 AI 네트워크 서비스 요구사항 및 기술의 범위	4
1.3.1 AI 네트워크 기반 서비스 및 활용 시나리오	4
1.3.2 AI 네트워크 요구사항	6
1.3.3 5G 한계와 서비스 제약 요인	8
1.3.4 AI 네트워크 기술의 범위	9

02. 국내외 동향 분석 10

2.1 표준 및 기술 동향	01
2.1.1 글로벌 표준 동향	01
2.1.2 글로벌 제조사 및 통신사 현황	2 1
2.1.3 국내 연구개발 현황	41
2.2 시장 동향 및 전망	51
2.2.1 통신 인프라 및 네트워크 관련 시장 현황 및 전망	5 1
2.2.2 AI 네트워크 시장 전망	7 1
2.2.3 AI 네트워크 기반 Edge AI 활용 서비스 시장 전망	9 1

03. AI 네트워크 핵심 기술 23

3.1 AI 기반 무선 접속망 (AI-RAN) 기술	3	2
3.1.1 AI 기반 무선 링크 성능향상 기술	3	2
3.1.2 AI-RAN 플랫폼 기술	5	2
3.1.3 AI 기반 통신 센싱 융합 기술	9	2
3.2 AI 기반 자율 네트워크 기술	1	3
3.2.1 AI 네트워크를 위한 파운데이션 모델	1	3
3.2.2 AI 기반 지능형 위성통신 기술	5	3
3.2.3 지능형 네트워크 시스템 기술	8	3
3.3 AI 서비스를 위한 네트워크 기술	1	4
3.3.1 AI 데이터센터 네트워크 기술	1	4
3.3.2 AI 서비스 지원 모바일 코어 네트워크 기술	3	4

04. AI 네트워크 기술 로드맵 50

4.1 AI 기반 무선 접속망 기술 로드맵	8	4
4.2 AI 기반 자율네트워크 기술 로드맵	1	5
4.3 AI 서비스를 위한 네트워크 기술 로드맵	6	5

05. 결론 58

약어표	5
참고문헌	6

개요 01

산업계 전반에 걸쳐 혁신의 물결을 몰고 오고 있는 인공지능 기술은 데이터를 분석하고 판단을 보조하는 수준의 인지형 AI를 지나서, 생성형 AI의 등장으로 창작, 코딩, 설계 등 인간의 고유 영역이라 여겨졌던 분야마저 영향권에 들어갔다. 최근 각광받는 에이전틱 AI는 사용자의 명령 없이도 스스로 계획을 수립하고 여러 작업을 연속 수행하는 능력을 갖추게 될 것으로 예상되며, 더 나아가 퍼지컬 AI의 등장은 디지털을 넘어 현실 세계의 조작으로 그 영역을 확장하여 인간의 물리적 한계를 보완하거나 대체하는 기술로, 노동과 생산 방식의 대전환을 이끌 것으로 기대되고 있다.

이러한 인공지능 기술의 영향력은 통신네트워크 산업 또한 예외가 아니다. 통신네트워크의 자동화 및 최적화를 위해서 인공지능 기술을 접목하는 노력은 약 1980년대부터 지속적으로 진행 중이다. 1980~1990년대에는 고정된 조건(rule) 하에서 전문가시스템을 통한 장애진단 등의 영역에서 활용되었으며, 2000년대에는 머신러닝 기반의 트래픽 예측 등에 적용되었다. SDN/NFV 기술의 등장과 함께 2010년대에는 딥러닝 기반의 지능형 네트워크 기술로 진화해 왔으며, 급기야 2020년대에는 AI 기술과 통신네트워크의 융합(Integrated AI and Communication)을 통한 AI-native 기술로 발전하여 2030년대에는 6G 완전 자율네트워크를 실현할 수 있을 것으로 기대하고 있다.

본 보고서에서는 최근 비약적인 발전을 거듭하고 있는 AI 기술을 통신네트워크에 적용하여 지능적인 네트워크를 구축·운영하고 AI가 그 기능을 다할 수 있도록 통신망 구성을 통하여 AI 기반 지능화된 서비스를 제공하기 위한 AI 네트워크 기술 및 서비스를 분석하고, 핵심기술 확보를 위한 추진 전략 및 기술 로드맵에 대해서 설명하고자 한다.

1.1 AI 네트워크 개념 및 정의

AI 네트워크는 AI 기술이 네트워크 운영의 보조적 도구를 넘어서, 네트워크의 구조 전반에 내재화된 형태의 차세대 네트워크의 개념으로서 인공지능이 네트워크의 설계·제어·운영·진화 전 과정에 걸쳐 중심으로 작동하며, 동시에 AI가 요구하는 초대규모 데이터

처리·학습·추론 환경을 네트워크 인프라 자체가 내재적으로 지원하도록 설계된 차세대 네트워크로 정의된다.

기존의 이동통신 도메인에서 진화형 기술(RAN·코어망)에 국한되어 사용되어 오던 6G AI-native 네트워크와 다르게, AI 네트워크는 AI 자체를 중심으로 유무선·전송·엣지·데이터센터·서비스 인프라 전반까지 적용범위를 포함한다는 점에서 차별화된다. <표 1>에서는 AI-native 네트워크와 AI 네트워크의 차별성을 다양한각도에서 정리하여 보여 준다.

표 1 ▶ 6G AI-native 네트워크와 AI 네트워크와의 차별성

구분	6G AI-native 네트워크	AI 네트워크 (본 기획 방향)
목표	RAN/코어 성능·운영 최적화(지능형 RRM, 슬라이싱, 자율운영 레벨 상승)	AI가 네트워크를 자율 운영하고 네트워크가 AI를 내재 지원(양방향)
적용 범위	주로 이동통신 도메인(RAN/코어, fronthaul/backhaul)	유·무선·위성·전송·엣지·데이터센터까지 전 구간
데이터/학습	네트워크 텔레메트리 기반 도메인 한정 학습	멀티도메인·멀티모달 재학습/지속학습, 합성·시물레이션 데이터 포함
제어 범위	페루프 자동화(AN/OAM) 레벨 고도화	완전자율(네트워크 전 계층) + AI 워크로드 지연/비용 최적화 동시 제어
아키텍처	서비스형 RAN지능(xApps/rApps), NWDAF 등 기구 표준 내 컴포넌트	AI 파운데이션 모델·에이전트가 상주하는 네트워크 OS 레이어 + AI-friendly 패브릭
하드웨어	DU/RU/코어의 가속기 부분 탑재	스위치/라우터/엣지·DC의 AI 최적 가속/메모리/스토리지 설계
표준/연계	3GPP/ORAN/ETSI 내 트랙	통신 표준 연계 + AI 모델·데이터 거버넌스·에이전트 상호운용 프레임
사업화 단위	RAN SW, 코어 기능, 자동화 솔루션	AI 네트워크 모델/플랫폼, AI-RAN 기지국 SW/장비, 완전자율 운영 솔루션

즉 다양한 서비스의 복잡한 요구사항에 대응하고 네트워크의 효율성과 성능 향상, 운영비용 절감과 재투자를 통한 사용자 경험 개선과 비즈니스 혁신을 위하여 현재는 네트워크에서 머신러닝과 딥러닝 기술을 보조적으로 활용하고 있는 수준이지만, 향후 데이터의 수집에서 학습, 추론, 검증, 실행, 피드백에 이르는 AI/ML 전주기가 전구간의 통신 네트워크에 내재화되어 상호유기적으로 동작함으로써, 네트워크 운용제어관리의 자동화(AI for Network)와 AI 응용서비스의 최적화(Network for AI)를 지원하는 AI 네트워크로 진화할 것으로 전망된다. [그림 1]에서는 AI 네트워크 개념도를 보여준다.

네트워크 기반 초실감 지능형 AI 서비스

AI 기반 네트워크 자동화 및 최적화 서비스

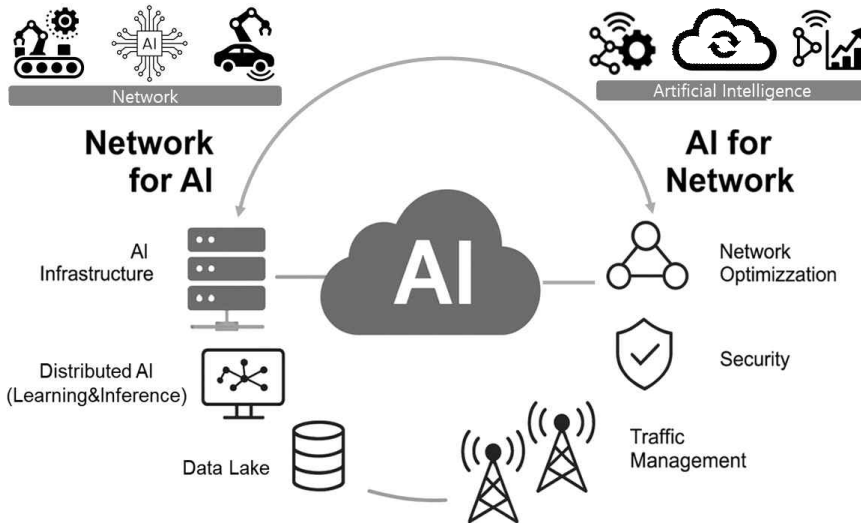


그림 1 ▶ AI 네트워크 개념도

1.2 AI 네트워크 연구개발 필요성

AI 네트워크(AI-Network)의 필요성은 차세대 지능형 통신 환경, 특히 6G와 같은 초연결·초지능 시대의 도래에 따라 점점 더 중요해지고 있으며, 크게 아래와 같은 세가지 측면에서 필요성이 크게 대두되고 있다.

우선적으로 네트워크의 복잡도 증가에 따른 자동화 측면에서 필요성이 증대되고 있다. 6G 시대의 네트워크는 수십억 개의 디바이스, 다양한 서비스 요구사항(eg. 초고속, 초저지연, 초신뢰)을 동시에 처리해야 하며 이를 위하여 기존의 수작업 기반 관리의 한계를 극복하고 트래픽 폭주, 품질 열화와 같은 실시간 예측에 따른 동적 상황 대응을 AI가 스스로 처리하는 자동화 기술이 요구되고 있다.

다음으로는 서비스 다양성 및 사용자 맞춤형 요구 대응 측면에서 AI 네트워크는 자율주행, 원격수술, 산업 IoT와 같은 서비스별 요구사항에 맞는 네트워크 슬라이싱 기반의 네트워크 품질(속도, 신뢰도, 지연 시간 등)을 능동적이고 효율적으로 제공할 수 있어야 한다. 즉 AI가 사용자 트래픽과 행동을 분석하여 최적의 자원 할당을 자동 수행함으로써 사용자 행동 기반 서비스 품질(QoE)의 최적화가 가능하게 된다.

마지막으로 에너지 및 자원 효율화 측면에서도 AI 기술이 네트워크에 적용이 요구된다.

네트워크 인프라의 전력 소비는 지속적으로 증가하고 있으며, 이를 줄이기 위한 그린 네트워크 실현에도 AI 활용이 가능하다. AI가 실시간 트래픽을 예측해 불필요한 기지국이나 네트워크 노드의 전원을 꺼서 전력을 절감하거나, 에너지 사용량 예측 및 스케줄링을 통하여 시간대별 수요에 따라 네트워크를 유연하게 운영할 수 있다.

1.3 AI 네트워크 서비스 요구사항 및 기술의 범위

인공지능과 지능형 네트워크가 결합하면서, 사용자는 단순한 데이터 전송 이상의 맞춤형·상황지능적 서비스를 기대하게 되었고, 서비스 제공자는 이를 가능케 하는 활용 시나리오와 요구사항을 명확히 정의할 필요성이 증대되고 있다.

이러한 배경에서 AI 네트워크 서비스는 초저지연 및 고신뢰 환경에서의 실시간 응용, 다양한 산업 분야와의 융합을 통한 맞춤형 서비스, 지속 가능한 운영과 효율적 자원 관리라는 관점에서 그 중요성이 강조되고 있다. 본 장에서는 먼저 대표적인 활용서비스와 이를 실현하기 위해 충족해야 할 핵심 요구사항과 필요한 기술의 범위를 정의한다.

1.3.1 AI 네트워크 기반 서비스 및 활용 시나리오

가. 인프라 및 자원 최적화

미래 네트워크, 6G 시대는 초연결·초지능 서비스를 뒷받침하기 위해 기존 네트워크의 구조적 진화를 요구한다. 이를 위해 에너지 효율, AI 오프로딩, 컴퓨팅-네트워킹 자원 통합, 다중 접속 동기화, 슬라이싱 기반 AI 인프라 구성 등 다양한 기술이 포함된다. AI 서비스를 위하여 단말 처리 능력에 따라 자원을 분산하거나, AI 연산을 위한 GPU 자원을 효율 분배하고, 사용자의 위치와 상태에 따라 자원을 자동으로 할당하는 구조 및 자원 최적화를 통해서 이동 중에도 끊김 없는 AI 서비스 시나리오가 가능해 진다.

특히, 제한된 단말 능력을 6G/AI 네트워크 인프라가 인지하여 AI 오프로딩을 최적화하고, 슬라이스를 자동 생성하여 필요한 대역폭을 확보하며, AI 모델은 엣지 서버와 실시간 동기화되면서 사용자에게 필요한 서비스를 빠르게 최적으로 제공되어야 한다.

나. 모빌리티 및 지능형 교통

차세대 모빌리티는 자율주행, 협력 주행, UAV 드론 등 다양한 디바이스가 동기화된 위치 인식, 실시간 고신뢰 통신, 센서 융합 기반 환경 인지 등을 통해 상호 작용하는 협력 통신 생태계를 형성한다. AI 네트워크 서비스는 이러한 환경에서 차량, 도로, 공중 로봇

등이 실시간으로 협력하는 상황을 인식하고 초저지연 상태에서 안전 제어와 판단이 이루어지도록 지원한다.

특히, 다중 UAV 협력과 같은 시나리오는 고속 비행체 간 통신, 연합 AI, 대용량 영상 실시간 분석을 필요로 하며, UAV가 사고를 감지하고 협력형 AI로 상황을 분석하며, 지능형 교통 인프라와 연계하여 자율주행 차량의 주행을 조절하는 등 적극적인 대응 시나리오를 제공할 수 있다.

다. 퍼스널 케어 및 헬스

미래에 개인에 대한 퍼스널 케어 서비스는 단순히 생체 정보를 측정하는 단계를 넘어, 지속적인 생체 데이터 수집, AI 분석 기반의 맞춤형 건강 피드백, 실시간 모니터링, 의료진·보호자 연계 기능을 포함한다. 고령자, 유아, 장애인 등 사회적 약자를 위한 AI 보조 기능이 결합되어 사회적 포용성을 증진하고 신뢰 사회를 함께 추구할 수 있게 된다. 또한, 실내·실외의 물리적 제약을 넘어선 모바일/에지 기반 헬스서비스와 진화된 웨어러블/스마트 기기들과의 연동을 고려할 수 있게 된다.

특히, 고령자를 위한 AI 건강 돌봄 서비스는 스마트 워치를 통한 심박, 혈압, 감정 상태 등을 지속적으로 모니터링하고, AI는 건강 데이터를 분석하여 이상 징후를 포착할 수 있으며, 보호자 및 인근 병원에 자동으로 알림을 보내는 서비스를 손쉽게 제공할 수 있고 실내에서는 장애인 보조 AI가 음성과 센서 인식을 통해 낙상 방지, 실내 안내 등의 서비스 시나리오를 고려해 볼 수 있다.

라. AI 협업 및 에이전트 연동

AI 네트워크에서는 다수의 AI 에이전트가 상호 협력하고 동적으로 역할을 분담하는 시스템을 지향한다. 이를 위해 AI 간의 실시간 통신, 지식 공유, 에이전트의 목적 중심 동기화, LLM 협력 등을 기반으로 한다. 이러한 시스템은 목적 중심 네트워크 구성, AI-to-AI 인터페이스 표준화, 다자간 협력 기반 학습 및 판단 기능을 요구하며, AI 윤리와 설명 가능성 또한 중요한 고려 요소로 포함된다.

대표적인 서비스로는 사용자가 회의 중 회의록 정리, 요약, 다음 일정 설정 등을 요청하면 여러 AI 에이전트가 협력 수행하여 하나의 AI는 음성 인식과 요약, 또 다른 AI는 LLM을 기반으로 회의 중 언급된 문서를 검색 및 분석 지원하며, 사용자 피드백을 통해 다음 회의 안건을 추천하는 등 AI 협력형 원격 회의 지원 시나리오를 고려해 볼 수 있다.

마. 로봇틱스 및 생활지원 서비스

6G와 AI 네트워크 서비스 지능형 로봇과 스마트 홈 환경이 사용자의 행동, 감정, 패턴을 이해하고 자율적으로 대응하는 시스템을 가능하게 한다. 홈 로봇, AI 어시스턴트, 가사 보조 AI 등은 인간의 삶을 능동적으로 지원하며, 실내 맵핑, 음성/동작 인식, 스마트홈 연동이 핵심 기술로 작동한다. 이러한 로봇/생활 지원 서비스는 다중 센서 융합, 상황 기반 자동화, 사용자 맞춤형 제어 등에서 고성능 AI와 6G의 초저지연성이 결합되어 실현된다. 또한 스마트홈을 위한 보조 로봇은 단순히 시키는 일만 하는 것이 아니라 모든 기기 및 로봇이 사용자 생활 패턴을 학습하여 능동적으로 작동하는 것을 기대할 수 있다.

바. 멀티미디어/실감형 AI 서비스

6G는 초고속·초저지연·고신뢰 네트워크 특성을 기반으로 실시간 콘텐츠 생성, 몰입형 미디어, 고해상도 영상 인식 등 새로운 차원의 미디어 소비 경험을 제공한다. 분산된 AI 연산, 대용량 스트리밍 처리, 콘텐츠 생성형 AI, 고속 비디오 분석 등은 이 분야의 핵심이며, 이를 위해선 연산 자원의 분산 활용, 모델 동기화, 오탐 방지, 실시간 응답 보장이 필수적이다. 이를 통해서 실시간 스트리밍 방송을 시청하는 시청자의 표정·목소리 등을 AI로 분석해 몰입도를 측정하며, 즉석에서 생성형 콘텐츠를 가공하여 사용자 반응에 최적화된 이야기 흐름으로 전환하는 콘텐츠 제작도 가능해 질 수 있다.

1.3.2 AI 네트워크 요구사항

가. 지능 기반의 네트워크 인프라 요구사항

6G는 초고속, 초저지연, 고신뢰성을 동시에 만족해야 하는 복합 인프라이다. 다양한 서비스 요구에 따라 자원을 유연하게 분배하는 슬라이스 기반 구조가 핵심이며, AI 기반의 자율 제어, 자가복구, 협력 기능이 네트워크 전반에 내재되어야 한다. 클라우드-엣지-단말 간 오프로딩 기술은 지연과 에너지 소비를 줄이며, 단말의 에너지 상태를 고려한 경로 설정과 저전력 연합 연산 구조가 필요하다.

UAV, 웨어러블 등 다양한 환경에서도 안정적인 연결을 위해 동적 라우팅, 네트워크 인지, 간섭 회피 기술이 적용되어야 하며, 실시간 트래픽에 따른 자율적 트래픽 제어, 시간 민감 데이터의 동기화 기술도 필수 요소다.

나. AI 연산 및 학습을 위한 구조적 요구사항

6G 환경에서 AI는 사용자의 맥락을 실시간 반영하고 다양한 기기에서도 안정적으로 작동해야 한다. 이를 위해 모델은 경량화되고, 엣지-클라우드 간의 분산 오프로딩과 병렬 추론이 가능해야 하며, LLM 및 생성형 AI는 사용자 피드백을 기반으로 지속적으로 학습하고 진화하는 개인화된 서비스를 제공해야 한다. 추론 파이프라인은 자동화되며, AI 간 협력 연산을 위한 표준 인터페이스와 역할 기반 구조가 요구되며, 자원은 네트워크 상황과 서비스 특성에 따라 유연하게 할당되어야 한다.

다. 신뢰·보안 기반의 운영 요구사항

AI 서비스 확산에 따라 개인정보 보호, 무결성 보장, 설명 가능성과 같은 신뢰 요소가 전제되어야 한다. 오작동에 대비한 정책 기반 대응 체계, 편향 방지를 위한 윤리 설계가 필요하며, 위기나 재난 상황에서도 안정적으로 작동할 수 있도록 회복탄력성이 확보되어야 한다. 이러한 운영을 위해 국제적 수준의 보안·윤리 표준과 인증 체계를 기반으로 한 통합 프레임워크가 요구된다.

라. 사용자 중심 상호작용 지원 요구사항

AI 네트워크 인터페이스는 사용자 상태와 의도를 실시간으로 파악하고, 맞춤형 반응을 제공해야 한다. 음성, 제스처, 표정 등 멀티모달 데이터를 인식하여 몰입감 있는 경험을 제공하고, 사용자 선호도와 행동에 따라 UX가 진화해야 한다.

특히 고령자, 아동, 장애인 등 다양한 계층의 접근성을 고려하고, 협업 환경에서 다중 사용자 간 충돌 없이 상호작용할 수 있는 구조가 필요하며, 지속적인 피드백 반응을 통해 사용자 경험이 점진적으로 개선되어야 한다.

마. 실감형 서비스 및 상황인지 기반 요구사항

6G는 AI와 융합된 VR/AR, XR 콘텐츠를 실시간으로 생성하고 제공해야 하는 동시에, 고대역폭·저지연 특성을 활용해 몰입형 콘텐츠를 끊김 없이 스트리밍할 수 있어야 한다. 또한 로봇, UAV 등은 센서를 통해 환경을 인식하고 자율적으로 동작하며, 다중 기기가 협력하는 군집 지능도 필요하다. 또한 실시간 콘텐츠 생성, 압축, 전송, 사용자 반응 분석이 연결된 상황 인지에 기반한 AI 네트워크 시스템은 원격교육, 진료, 자율이동 등 다양한 분야로 확장될 수 있다.

1.3.3 5G 한계와 서비스 제약 요인

5G 네트워크에서도 RAN, Core 등의 도메인에서 AI 기술을 적용하기 위한 많은 노력이 있었음에도 불구하고 상에서 언급한 AI 네트워크 서비스가 실현되기 어려웠던 이유는 다음과 같다.

우선 데이터의 부족이 원인으로 지적된다. AI 학습에는 고품질의 대규모 네트워크 운용 데이터가 필요하지만, 5G 네트워크는 아직 그만한 수준의 데이터 수집 체계가 미비하였으며, 분산된 네트워크 장비에서 발생하는 데이터가 통합되지 않아 AI 분석에 활용하기 어려웠다.

다음으로는 컴퓨팅 자원의 한계로서 RAN(무선 접속망)이나 코어망에서의 실시간 AI 처리를 위해서는 상당한 컴퓨팅 리소스가 필요하다. 5G 네트워크는 여전히 중앙 집중형 아키텍처 기반이 많아서, 엣지 또는 분산형 AI 처리에 한계가 있었다.

또한 구조적인 면에서 기존 5G 네트워크 아키텍처는 AI 모델을 네이티브하게 탑재할 수 있는 구조가 아니었다. 즉, AI가 네트워크 제어 loop에 직접 참여하지 못하고 트래픽 예측, 장애 예측 등과 같은 보조적인 분석 도구로 사용되는 경우가 많았다.

6G 이후의 AI 네트워크에서 지능형 네트워크가 가능해진 배경은 다음과 같다.

6G 네트워크는 처음부터 AI를 네이티브로 포함하는 구조로 설계되어 AI가 네트워크의 의사결정 루프에 직접 참여하여 실시간 최적화 수행할 수 있다. 예를 들어서 네트워크 슬라이싱 자동 설계, 에너지 최적화, 사용자 QoE 기반 자원 할당 등에 활용이 가능하다.

또한 6G는 통신뿐 아니라 센싱(Sensing)을 포함한 기능을 포함함 네트워크 상태, 단말, 환경에 대한 고정밀 데이터 수집 가능하며, 이런 데이터는 AI 학습에 필수적인 Context-rich 데이터셋 제공이 가능하기 때문에 AI 기술의 적용의 기반이 마련되었다.

6G는 엣지 및 RAN 레벨에서도 AI 추론 및 학습이 가능하게 되어 지연 최소화 및 실시간 대응 가능하다. 또한 Federated Learning, TinyML, On-Device Learning 등의 발전으로 프라이버시 보호하면서 학습 지속성을 동시에 확보할 수 있는 환경이 제공된다.

마지막으로 강력한 컴퓨팅 인프라와 네트워크-컴퓨팅 통합을 통하여 네트워크 인프라 자체에 GPU/NPU 등 AI 연산 기능이 내장되어 실시간 학습 및 추론이 가능하며, 사용자의 의도(Intent)나 목표(QoS/QoE 목표)를 인식하여 스스로 네트워크 구성 및 최적화 가능하기 때문인 것으로 분석된다.

1.3.4 AI 네트워크 기술의 범위

상기의 AI 네트워크 요구사항에 따라서 본 문서에서는 AI 네트워크 기술을 적용되는 도메인과 기술을 통해 이루고자 하는 목적과 용도에 따라서 크게 3개의 범주로 정의하여 기술하고자 한다.

가. AI 기반 무선 접속망(AI-RAN) 기술

AI 기반 무선 접속망 기술(AI-RAN)은 차세대 이동통신 네트워크에서 무선 접속망(RAN)에 인공지능을 적용하여 성능을 극대화하고 운영 효율성을 높이는 기술 영역으로, RAN에 AI를 도입하여 무선 성능을 최적화하고, AI 플랫폼을 RAN에 적용할 수 있도록 데이터를 개방하고, 이를 시험할 수 있는 가상 네트워크 제공을 하는 AI 플랫폼 기술의 도입이 필요하다. 이는 RAN의 성능 최적화, 무선 네트워크 최적화, 힐링과 같은 기술에 AI 도입을 가능하게 한다. 세부 기술은 다음과 같다.

- AI 기반 무선 액세스 지능화 기술
- AI-RAN 플랫폼 기술
- AI 기반 통신 센싱 융합기술

나. AI 기반 자율네트워크 기술

AI 기반 자율네트워크는 기존의 정적이고 수동적인 네트워크 관리 방식을 넘어, 스스로 학습·적응·제어하는 지능형 네트워크를 지향한다. 이는 대규모 연결, 초저지연, 초신뢰성이 요구되는 차세대(6G 포함) 통신 환경에서 필수적인 기술이다. 세부기술은 다음과 같다.

- 네트워크 파운데이션모델
- AI 기반 지능형 위성통신 기술
- 종단간 지능형 네트워크 시스템 기술(완전자율네트워크)

다. AI 서비스를 위한 네트워크 기술

마지막으로 AI 서비스가 고도화될수록 네트워크는 단순 데이터 전달을 넘어 AI를 효과적으로 구동하고 지원하는 기반 인프라로 진화해야 하며, 이를 위해 네트워크는 AI 응용서비스 뿐만 아니라 지능형 센싱 기술과의 밀결합이 요구된다. 또한 Agentic/Physical AI 응용서비스를 지원하기 위한 네트워크 기술 등의 AI 서비스를 위한 네트워크 기술도 하나의 범주로 분류된다. 세부기술은 다음과 같다.

- AI 데이터 센터 인프라 기술
- AI 서비스 지원 모바일 코어 네트워크 기술

2.1 표준 및 기술 동향

2.1.1 글로벌 표준 동향

가. ITU-T

ITU-T는 SG13을 중심으로 통신망에 AI/ML을 통합하기 위한 아키텍처, 기능, 요구사항, 사용 사례를 표준화하고 있으며, Y.3172를 통해 ML 파이프라인 기반 아키텍처 프레임워크를 제시했다^[1]. 이후 Y.3177 등 다양한 권고안을 통해 자동화, 성능 평가, 모델 마켓플레이스 등 구체 기술을 다루고 있다^[2].

2024년 7월에는 AI 네이티브 네트워크를 연구하는 FG-AINN(Focus Group on AI-Native Networks)이 출범했으며, AI가 네트워크 내재화 구조로 통합되는 전환을 모색하고 있다^[3]. 이 그룹은 개념 정의, 아키텍처 설계, PoC 프레임워크 수립 등을 중심으로 활동 중이다.

또한, 생성형 AI의 통신망 활용을 다루는 TR.GenAI-Telecom이 작성되었고, 신뢰성과 효율성 확보를 위한 요구사항과 평가 방법론이 논의되고 있다^[4]. 데이터셋 구조 및 관리에 대한 표준화도 병행 추진 중이다.

나. ETSI

ETSI는 AI를 ICT 전반의 전환 기술로 보고, Operational Co-ordination Group on Artificial Intelligence(OCG AI)를 중심으로 다양한 Technical Committee(TC) 및 Industry Specification Groups(ISG)와 협력하고 있다. 특히 TC Securing Artificial Intelligence(SAI)는 AI 보안 및 신뢰성 확보를 위한 기술 규격을 개발하고 있으며, 사이버 보안 요구사항과 법률 정합성까지 고려한 문서를 발행 중이다^[5].

ISG Experiential Networked Intelligence(ENI)는 경험기반 네트워크 적응, ISG Zero-touch network and Service Management(ZSM)은 제로터치 자동화, ISG Integrated Sensing and Communication(ISAC)은 AI 기반 센싱·통신 통합 기술을 6G

요소로 다룬다[6][7][8]. ISG Multiple Access Techniques(MAT)는 자율성과 신뢰성을 갖춘 AI 모델의 적용 가능성을 연구한다[9].

다. 3GPP

3GPP는 Rel.15부터 AI/ML을 통신망에 적용하기 위한 표준화를 시작하여, Rel.16에서 Network Data Analytics Function(NWDAF)를 통해 네트워크 데이터 기반 인사이트 제공 구조를 구축하였다. 이후 Rel.17~19에서 연합학습, QoS 예측, 데이터 전송 협상 등의 기능이 발전하며 실용화가 가속화되었다. 5G-Advanced에 대한 표준과 6G에 대한 연구가 병행되는 Rel.20에서는 단말(UE) 데이터를 수집해 AI 모델 학습에 활용하는 방안, 사용자 평면(User Plane, UP)의 성능 분석 및 최적화를 위한 5G에서의 AI 고도화 기능이 논의되고 있다[10]. 이와 더불어 AI 에이전트 개념을 포함하여 6G시스템 관점에서의 AI 사용 및 AI/센싱을 위한 데이터 관리 프레임워크 등에 대한 연구가 시작되었다[11].

RAN 측면에서 Rel. 17부터 AI/ML 기반의 지능화 프레임워크를 도입하여, 에너지 절약, 이동성 최적화 등 핵심 기능에 대한 초기 적용을 시작하였다. Rel. 18~19에서는 단말-기지국 간 협력 구조, 모델 수명주기 관리, 빔 관리, 네트워크 슬라이싱 등 다양한 기능이 구체화되며 AI/ML 활용 범위가 확장되었다. 이어서 Rel.20에서는 5G-Advanced 기능으로 기존 NR 아키텍처 내에서의 AI 활용을 통해 네트워크 효율성과 사용자 경험을 향상을 진해하고 있다[12]. 이와 더불어 6G RAN 설계를 위한 구조적 준비가 진행 중이며, AI 기반 기능 확대와 함께 AI 없이도 안정적으로 동작 가능한 시스템 설계에 대한 논의가 병행되고 있다[13].

라. O-RAN alliance

O-RAN Alliance는 2018년 설립된 글로벌 표준화 기구로, 폐쇄적이고 벤더 종속적인 무선 접속망을 개방적이고 지능적인 구조로 전환하기 위해 출범하였다[14]. 특히 RAN Intelligent Controller(RIC)의 도입으로 AI 기반 정책 제어와 네트워크 최적화가 가능해지면서, 무선망 운영은 자율적이고 지능적인 방식으로 진화하였다[14][15]. 여기에 근실시간(near-RT)·비실시간(non-RT) AI 응용과 DU·CU 내부의 실시간 AI 응용이 결합되어, AI는 무선 자원 관리와 제어에 직접 활용된다. 이를 통해 MU-MIMO 스케줄링과 CSI 기반 제어를 최적화하고, 실시간 학습으로 효율적인 자원 배분을 지원한다[16]. 또한 트래픽 부하 예측을 활용한 절전 제어와 AI 기반 이상 탐지 및 위협 대응이 표준 규격에 포함되면서, O-RAN은 성능·안정성·지속가능성을 동시에 추구하고 있다[15][16].

6G 대비를 위해 운영되는 nGRG(Next Generation Research Group)는 6G 활용 사례, 아키텍처, 네이티브·크로스 도메인 AI, 보안을 주요 연구 주제로 삼고 있으며, O-RAN 내부에 AI를 내재화하고 서비스·엣지·코어 전반을 아우르는 협력형 AI 모델 개발을 추진하고 있다[14][15]. 최근에는 생성형 AI의 잠재력이 주목받고 있는데, 이는 네트워크 관리 자동화, 멀티모달 디지털 트윈, 물리계층 신호 처리뿐만 아니라 실시간 상호작용 서비스, 지능형 어시스턴트, 시맨틱 커뮤니케이션 같은 6G 네이티브 서비스로 확장될 수 있다[17]. 이를 통해 O-RAN은 단순한 개방형 구조를 넘어, 생성형 AI를 내재화한 지능형 네트워크 플랫폼으로 진화하고 있음을 보여준다.

마. AI-RAN Alliance

AI-RAN Alliance는 AI 기술을 활용하여 RAN 성능 향상 및 유연성 제고를 목표로 하며, AI-for/and/on-RAN 개념을 통해 무선자원 최적화, AI 워크로드 동시처리, AI 기반 서비스 전달을 추구한다. 최근 Data-for-AI TG와 Test Methodology TG를 신설해, AI 모델 학습용 RAN 데이터 정의와 AI-RAN 시스템의 테스트 프레임워크 개발에 착수하였다[18].

비록 AI-RAN Alliance는 공식 표준 개발 기구(SDO)는 아니지만, 이들이 생산하는 기술 보고서, 평가 프레임워크, 개념 검증(PoC) 결과 등은 3GPP, O-RAN Alliance 등에서의 실질적인 표준화 논의에 중요한 참조 자료로 활용되고 있다. 특히 현실적인 유즈케이스 기반의 사전 표준화 연구를 통해, 기술 적용성과 표준화 간의 간극을 좁히고, 표준이 시장 수요와 기술 진화 속도에 부합할 수 있도록 기여하고 있다.

2.1.2 글로벌 제조사 및 통신사 현황

가. 제조사

Nokia는 완전자율네트워크(Level 5)를 목표로 ANF(AI Network Fabric)를 통해 네트워크 관측·분석·보안·자동화를 통합한 지능형 플랫폼을 개발하고 있으며, Google Cloud 기반 생성형 AI로 트래픽 분석, 이상탐지, 자동복구 기능을 구현 중이다. Microsoft와는 LLM 기반 서비스 자동화를 공동 개발하고 있으며, IOH 등 실증사례를 통해 에너지 절감 성과도 확보했다. AI-RAN Alliance에서도 주요 역할을 맡고 있다.

Ericsson은 AWS와 Gen-AI Lab을 운영하며 OSS/BSS 자동화를 추진하고, 20여개 이상의 AI 앱을 통합한 다중 에이전트 시스템을 통해 운영 효율화를 실현 중이다. NetCloud

Assistant는 LLM 기반 가상 전문가로서, 네트워크 질의응답, 진단 등을 자동화하고 있다. 또한 RAN 에너지 최적화, 신경모방 AI 기반 초저전력 무선처리 기술 등을 개발 중이다.

Huawei는 AI 네이티브 구조 전환을 목표로 AI 코어망을 발표하고, AIBA 기반 Agentic Core 구현을 추진 중이다. 또한 IP/전송망에서도 L4 자율망 솔루션을 상용화하고 있으며, ADN 전략 하에 AI 통합 관리 플랫폼(*iMaster NAIE* 등)을 구축했다. 실운영망에서는 AI 기반 장애 예측 및 복구 성능 향상이 입증되고 있으며, 글로벌 파트너십을 통해 6G 대응을 강화하고 있다.

Samsung은 AI-RAN 기술 선도에 집중하여, NVIDIA와 협력한 AI 가속 기반 vRAN을 PoC하고, MWC 2025에서 AI 기반 신호처리 기술을 시연하였다. O-RAN 기반 상용망에서도 AI 기반 에너지 매니저(*AI-ESM*)를 적용, 실시간 전력 최적화를 구현하고 있다. 자체 개발 NOS, CognitiV 등을 통해 제로터치 자동화와 슬라이싱 최적화도 병행 중이다.

나. 통신사

AT&T는 'Network AI' 전략 하에 자체 데이터레이크 및 분석 플랫폼(*Aether*)을 기반으로, AI/ML을 통해 네트워크 서비스 품질 개선 및 예측 기반 운용을 수행하고 있다. Google Cloud와 협력하여 네트워크 운영 자동화 및 GenAI 기반 고객 지원 시스템을 공동 개발 중이며, AI 기반 트래픽 예측과 장애 사전 대응 시스템을 통해 운영비 절감을 실현하고 있다. 또한, MEC 기반 AI inference 기능을 네트워크 엣지에 통합하려는 시도도 진행 중이다.

China Mobile은 자사 AI 네트워크 전략인 "CUBE-Net 3.0" 하에, 지능형 통신망을 위한 AI 아키텍처를 정립하고 있으며, NWDAF 기반 데이터 분석, 디지털 트윈 기반 네트워크 예측 관리, AI 기반 에너지 최적화를 추진하고 있다. MWC에서 자율 운용 레벨 4 이상의 코어망 및 RAN 자동화를 시연하였으며, 자체 AI 개발 프레임워크(*CM-AI*)와 연계된 모델 마켓플레이스도 운영 중이다.

NTT DOCOMO는 AI-RAN Alliance의 핵심 멤버로서, AI/ML 기술을 기반으로 한 무선 자원 관리 최적화, 네트워크 트래픽 예측, 에너지 절감 등을 위한 다양한 PoC를 진행 중이다. 특히 RIC 기반의 AI 서비스 개발 및 NWDAF 활용을 통해 RAN 지능화를 추진하고 있으며, 오픈 이노베이션을 통한 벤더 협력 및 실증도 병행하고 있다. AI 모델의 학습·추론 위치 분산, 데이터 프라이버시 보호를 고려한 연합 학습 적용 연구도 병행 중이다.

Deutsche Telekom은 'Zero Touch Service and Network Management' 전략

하에, 네트워크 운영 자동화 및 고객 서비스 최적화를 위한 AI 활용을 본격화하고 있다. T-Mobile US와 공동으로 AWS 기반 AI 플랫폼을 구축하여, 실시간 장애 감지 및 복구, 고객 이슈 자동 분류, 자동 자원 할당 등을 실현하고 있다. 또한, OSS/BSS에 LLM 기반 에이전트 도입을 시도하고 있으며, 자체 AI 개발자 포털(AI@DT)을 통해 AI 내재화를 가속 중이다.

T-Mobile US는 Deutsche Telekom과의 협력을 바탕으로, 고객 경험 개선과 운영 효율화를 위한 AI 도입을 확대하고 있다. 특히, 실시간 트래픽 관리와 네트워크 품질 모니터링 자동화, 고객 문의 응답 지원 등에 GenAI와 분석 기술을 접목하고 있으며, 6G 대비를 위한 AI 기반 네트워크 인텔리전스 연구에도 참여하고 있다.

SK텔레콤은 LLM 기반 AI 에이전트를 네트워크 운영에 적용하고 있으며, Telco LLM을 활용한 이상 탐지, 트래픽 예측, 고객 민원 분석 등을 통해 운영 자동화를 추진 중이다. 일부 망 구간에는 AI 기반 자가 복구 기술을 적용하고 있으며, UAM/6G 대비를 위한 위치 예측 및 디지털 트윈 기반 시뮬레이션 환경 연구도 병행하고 있다. 또한 Global Telco AI Alliance(GTAA)에 참여하여 LLM 모델 공동 개발 및 AI 서비스 글로벌화를 모색하고 있다.

KT는 'AI Full Stack' 전략을 기반으로 AI 인프라부터 서비스까지 투자를 확대하고 있으며, 네트워크 운영 자동화와 AI 연계 검증을 진행 중이다. Core/Transport/RAN 각 계층에서 NWDAF 기반 예측·이상 탐지·트래픽 최적화를 검토하고 있으며, MEC 기반 추론 연동과 실시간 제어 구조도 일부 실증하였다.

LG U+는 네트워크 자동화 및 고객 경험 향상을 위한 AI 적용을 지속 확대하고 있으며, 특히 데이터 기반 AI 장애 감지·예측, 자가 복구 기술, 지능형 품질 분석 시스템 등에 집중하고 있다. 자체 통합 관제 플랫폼에 LLM 기반 분석 기능을 시범 적용하여 운영자 개입 없이 알람 분류 및 원인 진단을 수행하는 구조를 검토하고 있으며, AI 기반 QoE 예측 기술을 통해 사용자 중심의 품질 최적화를 강화하고 있다.

2.1.3 국내 연구개발 현황

한국은 5G 상용화 이후, AI 기반 차세대 네트워크 기술 확보를 위해 전략적 R&D 투자를 본격화하였다.

2021년부터 추진 중인 '6G 핵심기술개발사업'(총 2,000억 원 규모)은 원천기술 확보를 목적으로 하며, ETRI 등 국책연구기관과 주요 대학이 참여하여 AI 기반 무선 및 코어

기술, 지능형 제어 프레임워크, 디지털 트윈 등 기초 기술 수준의 AI 네트워크 요소기술 개발에 집중하고 있다. 또한 2024년부터는 '6G 산업기술개발사업'(총 3,700억 원 규모)이 새롭게 추진되어, 기지국 장비, AI 기반 네트워크 소프트웨어 등 상용화를 위한 실증 기술 개발이 진행되고 있다. 그러나 현재 국내 연구는 대체로 AI를 활용한 전형적인 통신망 자동화 기술에 집중되어 있으며, 생성형 AI기술의 진화와 더불어 글로벌 제조사들이 주도하는 에이전트 기반 자율망(Agentic AI), AI-native 통신망 아키텍처, AI-on-Network와 같은 차세대 패러다임 전환 수준의 기술개발은 아직 미진한 상황이다.

2.2 시장 동향 및 전망

2.2.1 통신 인프라 및 네트워크 관련 시장 현황 및 전망

표 2 ▶ 세계 통신 인프라 및 네트워크 관련 시장 매출액 규모

(단위: 백만 달러)

구분		2023	2025	2027	2029	CAGR (24-29)	
네트워크 장비	유선 가입자 장비	광가입자망	5,602	5,855	6,938	8,268	7.9%
		DSL	3,631	3,145	2,894	2,523	-5.6%
		소계	9,233	9,001	9,833	10,791	3.7%
	데이터센터 장비	15,314	18,424	22,776	27,140	10.5%	
	백본 장비	광전송장비	12,840	12,522	14,752	16,015	5.8%
		스위치라우터	12,598	12,300	14,062	16,705	6.9%
		소계	25,437	24,822	28,813	32,720	6.3%
	모바일 장비	RAN	40,899	34,801	33,055	30,323	-4.0%
		모바일코어	5,976	5,969	6,564	6,717	2.9%
		소계	45,875	40,769	39,618	37,040	-3.0%
기업용(B2B) 셀룰러 장비	2,439	3,650	4,958	6,459	16.7%		
소계	99,298	96,666	105,999	114,150	3.6%		
단말	모바일 폰	473,860	521,412	582,200	634,495	5.1%	
통신 서비스	이동통신서비스	779,520	806,589	865,633	913,366	2.8%	
	저궤도위성통신서비스	3,274	5,962	10,448	17,153	33.5%	
	소계	782,794	812,551	876,081	930,519	3.1%	

출처: ETRI 기술전략연구본부, 2025. 7.

AI 네트워크로 진화하는 통신 인프라 및 네트워크 장비 시장은 유무선 통신사업자 및 기업용(B2B) 셀룰러 장비가 포함된 네트워크 장비 시장, 네트워크와 연계해 사용자와 통신하는 단말 시장, 음성·데이터 서비스를 제공하는 통신 서비스 시장의 세 부분으로 나눌 수 있다. 2025년 기준 네트워크 장비 시장이 967억 달러, 단말 시장이 5,214억 달러, 통신 서비스 시장이 8,126억 달러 규모를 전세계적으로 형성하고 있다^[19].

AI 서비스가 확산되면서 변화하는 통신 인프라 및 네트워크 관련 시장을 전망하면 <표 3>과 같다. 네트워크 장비 시장은 AI for RAN을 넘어 AI on RAN으로 진화하면서, AI 컴퓨팅 장비 시장이 새롭게 포함될 것으로 전망된다. 단말 시장은 스마트폰 중심에서 벗어나, 태블릿-AI와 같은 AI 서비스 지원용 범용 단말 및 VR/AR 디바이스 등으로 확대될 것으로 예상된다. 통신 서비스 시장은 6G AI 네트워크 기반의 초고정밀 IoT 통신 서비스까지 그 영역이 확장될 전망이다^[20].

표 3 ▶ 세계 통신 인프라 및 네트워크 관련 시장 전망

(단위: 백만 달러)

구분		2030	2033	2035	2037	CAGR (30-37)	
네트워크 장비	유선가입자 장비: 광가입자 장비 등	11,201	12,050	12,464	12,739	1.9%	
	데이터센터 장비	28,931	33,971	36,885	39,588	4.6%	
	백본 장비: 스위치, 라우터, 광전송 장비	34,487	39,393	42,607	45,730	4.1%	
	AI 컴퓨팅 장비	1,089	2,921	4,384	5,485	26.0%	
	모바일 장비	RAN	32,749	41,054	42,709	42,793	3.9%
		모바일코어	6,885	7,371	7,677	7,940	2.1%
		소계	39,634	48,425	50,385	50,733	3.6%
	기업용(B2B) 셀룰러 장비	7,234	9,672	11,439	13,280	9.1%	
소계	122,577	146,433	158,163	167,556	4.6%		
단말	모바일 폰	661,684	754,369	822,213	887,595	4.3%	
	태블릿-AI	52,413	60,960	65,618	68,938	4.0%	
	VR/AR 디바이스	21,653	30,012	34,415	39,033	8.8%	
	소계	735,750	845,341	922,245	995,566	4.4%	
통신 서비스	이동통신서비스	938,484	1,021,033	1,079,536	1,139,172	2.8%	
	저궤도위성통신서비스	20,586	30,144	33,867	36,806	8.7%	
	초고정밀IoT통신서비스	71,448	76,942	80,443	83,775	2.3%	
	소계	1,030,518	1,128,118	1,193,845	1,259,753	2.9%	

출처: ETRI 기술전략연구본부, 2025. 7. ^[21]

2.2.2 AI 네트워크 시장 전망

AI 네트워크 시장은 크게 AI 기반 무선 접속 망(AI-RAN 장비)과 AI 자율네트워크 장비로 구분된다. 이 두 영역은 단순한 하드웨어 확장에 그치지 않고, 네트워크의 설계·제어·운영·진화 전 과정에서 지능적으로 작동하는 네트워크용 AI모델과 결합함으로써 자율적이고 지능적인 운용이 가능한 고도화된 네트워크 장비로 발전하고 있다. 이러한 구조는 무선·유선·위성 전 구간에서 AI가 중심으로 작동하는 체계를 구현함으로써, 차세대 통신 인프라의 핵심 축으로 자리매김할 것으로 전망된다[22].



그림 2 ▶ AI 네트워크 시장 구분

가. AI-RAN 장비

AI-RAN은 O-RAN 구조를 기반으로 진화하고 있으며, AI 반도체를 탑재한 인프라 서버가 DU 역할을 수행하면서 동시에 CU 기능을 하는 통합 소프트웨어를 내장하고 있다. RU는 별도의 장비로 구성되어 AI-RAN 인프라 서버와 상호 연결되는 구조를 이룬다. 이 과정에서 AI 모델의 운영과 병렬처리를 지원하기 위해 GPU, 고속 NIC, 스토리지 등 다양한 가속 장치가 추가되며, AI 오케스트레이션과 RIC용 AI 모델 등을 운영하기 위한 소프트웨어 및 라이선스도 필수적으로 포함된다. RU:DU:CU= 70:20:10의 구축 비중과 장비별 비용 증가를 고려할 때, 전체 비용은 기존 RAN 대비 약 27~40% 상승할 것으로 예상된다.

나. AI 자율 네트워크 장비

RAN을 뿐만 아니라 모바일 코어, 전달·백본 망에도 네트워크용 AI 모델을 기반으로 한 "AI 자율 네트워크"로의 혁신이 가속화되고 있다. 이러한 흐름은 공통 AI 플랫폼과의 연계를 통해 더욱 고도화될 전망이다. 이에 따라, AI 자율 네트워크 장비로 진화하는 모바일 코어, 전달·백본 망은 기존 대비 각각 35%와 12.5%의 비용 증가가 예상된다.

다. 전체 네트워크에서 AI 네트워크 장비 비중

기지국(RAN)은 가상화와 개방형 표준을 기반으로 한 오픈랜(O-RAN) 구조로 진화하고 있으며, 이러한 개방성과 유연한 아키텍처는 AI 기능을 네트워크에 통합할 수 있는 토대를 제공한다. 따라서 O-RAN의 도입 비중은 AI 네트워크 확산 수준을 가늠할 수 있는 핵심 지표로 활용될 수 있다. 이에 따라 AI 네트워크 장비 시장의 확산 패턴은 현재 진행되고 있는 O-RAN 도입 패턴을 적용하여 예측한다. 실제로 O-RAN 도입율은 2024년 5.8%에서 시작하여 2025년 9.2%, 2026년 12.4%, 2027년 15.2%, 2028년 17.9%, 2029년 21.3%를 거쳐 2030년에는 24.5% 수준까지 증가할 것으로 전망된다.

라. 시장 전망 결과

AI-RAN과 AI 자율 네트워크 장비로 구성된 세계 AI 네트워크 장비 시장은 1차년도에 53.4억 달러 규모에서 시작하여 8차년도에는 322.7억 달러 규모로 연평균 29.3%의 성장률을 기록하며 시장이 확산될 것으로 전망된다.

표 4 ▶ 세계 AI 네트워크 시장 전망

(단위: 백만 달러)

구분		Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	CAGR (Y1-Y8)
AI-RAN 장비		2,545	4,380	6,435	8,362	10,143	12,190	14,091	15,483	29.4%
AI 자율 네트워크 장비	모바일 코어	539	876	1,207	1,513	1,819	2,207	2,585	2,894	27.1%
	전달·백본 망	2,250	3,741	5,269	6,736	8,258	10,210	12,178	13,890	29.7%
	소계	2,789	4,616	6,476	8,249	10,077	12,417	14,763	16,785	29.2%
합계		5,335	8,997	12,912	16,611	20,219	24,607	28,854	32,267	29.3%

출처: ETRI 기술전략연구본부, 2025. 9. [23]

참고 ▶ 시장 전망 설명자료

□ AI-RAN 장비

- (정의) 기존 RAN 기능을 가상화된 서버 인프라 위에서 구현하면서, 무선 접속망의 물리계층(L1)·데이터링크/네트워크 계층(L2/L3)·오케스트레이션 기능에 AI 기반 최적화 알고리즘을 통합하여, 무선 자원 관리와 트래픽 처리를 지능적으로 수행하는 장비 및 시스템
- (구성) AI-RAN 장비 = Radio Unit + RAN L1 SW + RAN L2/L3 SW + 가상화 플랫폼 + 오케스트레이터 → 기존 RAN 장비 대비 34% 비용 증가

□ AI 자율 네트워크 장비

- (정의) RAN 장비 외에, 모바일 코어, 전달·백본 망 영역에서 AI 기반 자율 제어와 최적화를 위해 추가되는 기능/인프라를 포함하는 장비 및 시스템
- (구성) AI 자율 네트워크 장비 = 모바일 코어(AI-Core) + 전달·백본 망(AI-Transport/Backbone), 각각 35%와 12.5% 비용 증가

□ 모시장 및 시장 전망 (출처: '29년까지 Gartner 전망치, '30년 이후 ETRI 기술전략연구본부 추정)

- AI-RAN 장비 : 이동통신 RAN 시장 × 1.34
- AI 자율 네트워크 중 모바일 코어 장비 : 이동통신 코어 장비 시장 × 1.35
- AI 자율 네트워크 중 전달·백본 장비 : 백본 장비(스위치/라우터, 광 전송장비) 시장 × 1.125
- 확산패턴 : Open RAN 도입 비중 적용 = 5.8%부터 24.5%까지 연차별로 적용

2.2.3 AI 네트워크 기반 Edge AI 활용 서비스 시장 전망

2025년 6월, O-RAN 차세대 네트워크 연구그룹은 AI-RAN 기반 “Edge AI 활용 신규 서비스” 보고서를 발간^[17]하였다. 해당 보고서에서 제시한 3가지 유형, 12가지 신규 시나리오(Use Case)의 핵심 내용을 정리하면 <표 5>와 같다.

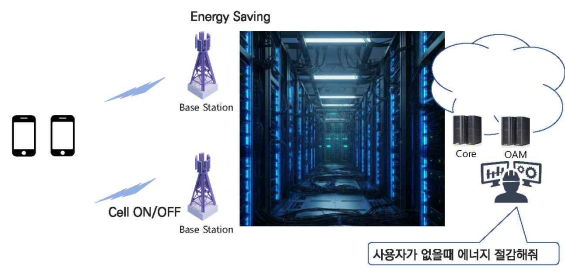
표 5 ▶ AI 네트워크 기반 Edge AI 활용 신규 서비스

분류	신규 서비스(Use Case)	내용 설명
1. AI for Network (AI가 네트워크를 똑똑하게 해줌)	AI Agent 기반 네트워크 자동화	네트워크에 “AI 비서”를 두어 장애 대응·최적화를 자동 처리
	의도 기반 네트워크 관리	사람이 “영상은 끊기지 않게 해줘”라고 말하면, AI가 의도를 이해해 자동 설정
	멀티모달 AI 기반 관리	로그·이미지·센서 등 다양한 데이터를 종합해서 네트워크 상황을 분석·제어
	RAN 디지털 트윈 지원	무선망을 가상으로 복제해 AI가 시뮬레이션·예측
	물리계층 처리 지원	복잡한 신호처리를 AI가 보조해 통신성능 향상
2. Network for AI applications (네트워크가 AI 서비스 구현을 도와줌)	실시간 인터랙티브 비디오 생성	사용자의 입력에 따라 영상·장면을 즉시 바꿔주는 서비스 (게임·XR·교육)
	옛지 기반 GenAI 컴퓨팅 서비스	사용자가 원하는 AI 모델을 가까운 옛지 서버에서 빠르고 안전하게 실행
	분산형 AI 에이전트	여러 장치와 서버에 흩어진 AI들이 서로 협력·거래해 맞춤형 서비스 제공
3. Network-native AI services (AI와 네트워크가 합쳐져 새로운 서비스 탄생)	실시간 자동 통역 통화	한국어로 말하면 상대방 언어로 동시에 변환되어 들리는 실시간 통역 통화
	스마트 어시스턴트 앱	네트워크 속에 내장된 비서가 일상·업무 지원
	의미 기반 통신(시멘틱 통신)	데이터를 모두 보내지 않고 의도/의미만 전송 → 받는 쪽 AI가 재구성
	몰입형 커뮤니케이션	홀로그램·3D 아바타 기반의 초실감 실시간 대화/회의

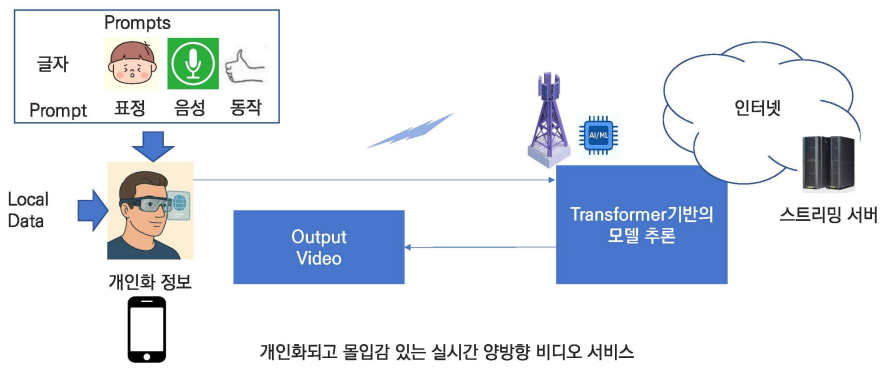
[그림 3]에서는 생성형AI서비스의 제공 관련 O-RAN nGRG 에서 제시한 신규 서비스에 대한 예제를 보여준다[24]. 네트워크를 위한 AI 서비스 예 중 하나는 의도(Intent) 기반의 네트워크 관리로, 완전 자율 네트워크의 대표적 예시가 될 것이다. 네트워크가 AI 서비스를 도와주는 Edge AI 서비스 예 중 하나는 추후 AR/XR등에서 활용될 수 있는 실시간 인터랙티브 비디오 생성과 같은 AI 서비스가 될 수 있으며, AI와 네트워크가 합쳐져 실시간 자동 통역 통화 서비스 등이 새롭게 시장에 등장 할 수 있을 것이다.

1-2. 의도 기반 네트워크 관리

LLM과 RAG를 지원하는 의도 기반 망관리



2-1. 실시간 인터랙티브 비디오 생성



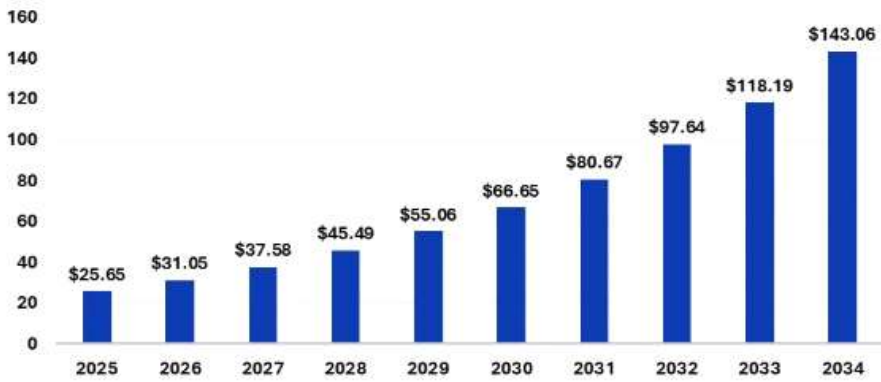
3-1. 실시간 자동 통역 통화



그림 3 ▶ Edge AI 활용 신규 서비스 시나리오 개요도 (예시)

Edge AI 시장은 앞으로 10년 동안 약 7배 이상 성장할 것으로 전망되며, 2034년에는 전체 시장 규모가 약 1,431억 달러에 이를 것으로 예측된다. 이 시장 규모에는 소프트웨어, 하드웨어, 엣지 클라우드 인프라, 그리고 관련 서비스가 모두 포함된다. 특히 셀룰러 B2B 엣지 컴퓨팅 분야에서는 제조, 헬스케어, 교통, 환경 감시, 게임 등 5대 융합산업 분야가 핵심 수요처로 부상할 것으로 기대된다[25].

Edge AI Market Size 2025 to 2034 (USD Billion)



출처: Precedence Research, 2025. 8.

그림 4 ▶ Edge AI 시장 전망

3.1 AI 기반 무선 접속망 (AI-RAN) 기술

AI 기반 무선 접속망(AI-RAN) 기술은 AI를 기반으로 한 기지국(RAN) 기술로 무선접속망의 성능 향상을 위한 AI의 활용을 시작으로 AI와 RAN의 공존, RAN에서의 AI 서비스 제공 방향으로 점차 확대되어 연구되고 있다. 본 절에서는 AI 기반의 무선 접속 망 성능향상 기술과 AI-RAN 플랫폼 구축 기술에 대하여 기술한다. [그림 5]에서는 RAN을 위한 AI와 AI-RAN 플랫폼의 개념도를 보여준다.



그림 5 ▶ AI 기반 무선 접속 망 기술

3.1.1 AI 기반 무선 링크 성능향상 기술

AI 기반 무선 링크 성능 향상 기술은 물리계층(Physical Layer, PHY)과 매체접속제어계층(Medium Access Control, MAC)에 AI를 적용하여, 무선 전송 과정에서 발생하는 복잡한 환경 변화에 정밀하고 유연하게 대응하기 위한 기술이다. 이 기술은 채널 상태 추정, 빔포밍, 신호 복원, 자원 제어 등 전송 성능에 직접 영향을 주는 핵심 기능을 고도화하여, 전체 무선 링크의 효율성과 신뢰성을 향상시키는 것을 목표로 한다.

기존 무선 링크 기술은 수학적 모델 기반 알고리즘에 주로 의존해 왔으며, 단순한 전파

환경에서는 일정 수준의 안정적인 성능을 제공해왔다. 그러나 고주파 대역 사용 확대, 고속 이동성, 대규모 MIMO 환경 등으로 채널 조건이 복잡해지면서, 이러한 변화에 능동적으로 적응할 수 있는 데이터 기반 AI 기법의 적용이 필수적으로 부상하고 있다.

현재 연구 중인 주요 기술 요소는 다음과 같다.

❖ 딥러닝 기반 채널 추정 기술

기존의 LS(Least Squares) 및 MMSE(Minimum Mean Square Error) 방식은 고속 이동 환경이나 주파수 선택적 페이딩 조건에서는 성능 저하가 발생하기 쉽다. 이를 보완하기 위해, CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), Transformer 기반 모델을 이용해 채널 응답(CIR, FDP 등)을 입력으로 받아 채널 상태를 예측하는 기술이 개발되고 있다. 이러한 모델들은 다양한 채널 조건을 학습함으로써 일반화 성능을 확보하며, 실시간 추론이 가능하도록 경량화된 구조로 최적화되고 있다. 결과적으로 고속 이동성이나 다중 경로 환경에서도 안정적인 채널 추정이 가능하다.

❖ 오토인코더 기반 CSI 피드백 압축 기술

Massive MIMO 환경에서는 단말이 기지국으로 전송해야 하는 CSI(Channel State Information)의 양이 매우 커지며, 이로 인해 피드백 오버헤드가 시스템의 전송 효율을 저해할 수 있다. 이를 해결하기 위해 오토인코더 기반의 압축 기법이 활용되고 있다. 단말에서는 CSI를 저차원 피쳐 벡터로 압축한 후 전송하고, 기지국에서는 이를 복원하여 프리코딩, 링크 어댑테이션 등에 활용한다. 이 방식은 기존 코드북 기반 접근 대비 더 높은 복원 정확도를 제공하며, 피드백 자원의 낭비를 효과적으로 줄일 수 있다.

❖ AI 기반 빔포밍 최적화 기술

밀리미터파 환경에서는 경로 감쇠가 심하여 빔포밍의 정밀도가 통신 성능에 직접적인 영향을 준다. 기존 방식에서는 기지국이 다수의 빔에 대해 참조 신호를 전송하고, 단말이 이를 측정하여 최적 빔을 선택하는 방식으로 수행되며, 이는 높은 자원 소모와 지연을 수반한다. AI 기반 빔포밍 기술은 CNN 또는 MLP(Multi-Layer Perceptron) 모델을 활용하여 일부 빔에 대한 측정값만으로 최적의 빔을 추론한다. 이를 통해 SSB(Synchronization Signal Block) 전송 자원을 절감하고, 탐색 지연을 대폭 감소시킬 수 있으며, 커버리지 유지와 핸드오버 안정성 향상에도 기여한다.

❖ 수신 신호 검출 및 복원 기술

무선 수신 환경에서는 간섭, 잡음, 왜곡 등의 요소가 복조 성능을 저하시킬 수 있으며, 특히 URLLC와 같은 초신뢰 저지연 서비스에서는 신호 품질 확보가 매우 중요하다. AI 기반 복원 기술은 LSTM(Long Short-Term Memory) 또는 Transformer 계열의 시계열 처리 모델을 활용하여 수신 신호의 시간-주파수 특성을 분석하고, 왜곡된 성분을 보정함으로써 신뢰도 높은 신호 복원을 가능하게 한다. 최근에는 기존 신호처리 알고리즘과 AI 모델을 결합한 하이브리드 방식도 제안되고 있으며, 복조 성능과 처리 속도를 동시에 개선하는 방향으로 발전하고 있다.

❖ 스펙트럼 센싱 및 자원 예측 기술

스펙트럼 효율을 높이기 위해서는 혼잡 상태 예측, 간섭 탐지, 동적 자원 할당 등의 기능이 중요하며, AI는 이와 같은 기능 구현에 적합한 접근법으로 평가받고 있다. 강화학습(RL) 기반 모델은 다양한 상황에서 최적의 자원 할당 정책을 학습하며, RNN 또는 LSTM 기반 시계열 모델은 트래픽 부하의 변화나 사용자 수요를 예측하는 데 활용된다. 이를 통해 무선 자원의 활용도를 극대화하고, 슬라이싱 환경에서의 자원 분배, 간섭 회피, QoS 보장 기능을 강화할 수 있다.

이와 같은 AI 기반 기술들은 기존 무선 링크 구조의 한계를 보완하고, 고신뢰·고정밀 통신이 요구되는 5G-Advanced 및 6G 시대의 네트워크 설계에 핵심적으로 기여한다. 특히, 엣지 단말에서의 실시간 추론, 경량화된 모델 구현, 분산 학습 기반 구조 등과의 통합을 통해 무선 링크의 자율적 운용과 지속적인 최적화가 가능해진다.

3.1.2 AI-RAN 플랫폼 기술

AI-RAN 플랫폼 기술은 RAN 최적화를 위한 AI활용 및 RAN에서의 AI 서비스 제공을 위하여 AI와 RAN이 결합되었을 때 이를 이상적으로 지원할 수 있게 하는 기술이다.

이는 무선 접속망 내에서 AI를 위한 데이터 수집, 처리, 학습, 추론, 정책 반영에 이르는 운영 전주기가 설계 가능하도록 하고, RAN 플랫폼에서의 전송 및 제어 기능 전반에 걸쳐 AI 기능이 실시간으로 통합·작동될 수 있도록 구조화된 시스템적 접근을 포함한다. AI-RAN 플랫폼 기술은 AI-RAN을 위한 모델 설계, 학습, 추론 통합 프레임워크, AI 플랫폼에서 수행가능한 프로그래머블 RAN SW, RAN 지능화 제어를 포함하는 오케스트레이션 기술, RAN 디지털 트윈 기반 가상 AI-RAN 플랫폼 기술 및 가상네트워크와 실증망 통합 플랫폼 기술로 이루어 진다.

❖ AI 모델 설계·학습·추론 통합 프레임워크 기술

AI 모델 설계·학습·추론 통합 프레임워크는 RAN 전반에 적용되는 AI 모델의 전주기 생애주기를 지원하는 기반 기술이다. 채널 상태 정보(CSI), 사용자 트래픽 패턴, 셀 단위 KPI(Key Performance Indicator) 등의 대용량 무선 네트워크 데이터를 실시간 수집하여, 전처리를 거쳐 CNN, LSTM, Transformer 등 AI의 딥러닝 모델 학습에 활용한다. 이 과정은 자동화된 데이터 레이블링, 하이퍼파라미터 탐색, 모델 아키텍처 탐색(AutoML), 교차 검증 등의 기술을 통해 최적화된다. 또한 모델의 연산 복잡도와 실시간 추론 가능성을 동시에 고려해 AI 모델은 경량화된다. 훈련된 모델은 온디바이스, 엣지, 클라우드 등 다양한 환경에서 추론이 수행될 수 있으며, 상황에 따라 분산 처리 또는 통합 제어가 가능하다. 특히, 연합 학습(Federated Learning)을 활용해 데이터 프라이버시를 보장하면서도 글로벌 최적화 모델을 확보할 수 있고, 지속 학습(Continual Learning)을 통해 환경 변화에도 적응 가능한 구조를 갖춘다. 이 프레임워크는 AI 기능의 신속한 배포와 업데이트, 검증, 운영 자동화에 필수적이다.

❖ AI-Native 프로그래머블 RAN SW (Programmable DU/CU)

AI-Native 프로그래머블 RAN SW는 무선 접속망의 핵심 구성 요소인 DU(Distributed Unit)와 CU(Centralized Unit)에 AI를 네이티브하게 통합하기 위한 기술로, 기존 정형화된 계층 구조의 한계를 넘어 유연하고 지능적인 무선 전송 처리를 가능하게 한다. AI-Native 프로그래머블 RAN SW는 오픈랜의 O-DU/O-CU/RIC 구조와 유사한 형태의 구조를 가지며, AI DU, AI CU, AI RIC으로 구성된다.

AI-DU/AI-CU가 기존 DU/CU, O-DU/O-CU 등과의 다른 점은 RAN의 Data를 생성하는 모듈이 포함된다는 점과 실시간 AI 제어가 가능한 구조를 제공한다는 점이다. AI-DU는 PHY의 AI 기능 탑재 뿐만 아니라 MAC 및 PHY 계층의 실시간 제어 기능(예를 들어, 스케줄링, MCS(Modulation and Coding Scheme) 결정, 빔포밍, 전력 제어 등)을 AI 알고리즘으로 처리할 수 있도록 구조화되며, 이를 위해 고속 AI 추론이 가능한 연산 모듈을 내장하거나 외부 AI 엔진과 연동 가능한 인터페이스를 포함한다. RAN에서 생성하는 데이터는 무선 채널 정보 (CSI feedback, RSRP, RSRQ), 단말 접속 정보, PRB Usage, 연결 실패 정보, QoS 정보, 이동성 제어 성공률 등 DU와 CU가 생성할 수 있는 통계정보를 포함한다.

AI-RAN 플랫폼을 위하여 AI-RIC(RAN Intelligent Controller)은 RAN 제어 기능을 AI 기반으로 구조화하여, 실시간성과 전략적 최적화 요구를 동시에 충족하는 오케스트레이션 프레임

워크를 제공해야 한다. Near-RT RIC은 수 밀리초에서 수 초 단위의 짧은 시간 지연 요구에 대응하며, xApp 모듈을 통해 DU 및 CU의 자원 할당, 간섭 회피, 빔 선택, 전력 제어와 같은 제어 기능을 AI 기반으로 수행 가능하다. 이는 DU/CU와 개방화된 E2 인터페이스를 통해 연동되어, 초저지연이 요구되는 무선 링크 상황에서도 유연한 대응이 가능하다. AI-RAN을 위한 near-RT RIC은 다양한 AI/ML 프레임워크를 제공하여 AI를 사용하는 xAPP간의 연동이 원활하게 되도록 해야 한다. AI RIC이 기존의 RIC과 다른 점은 실시간, 비실시간 환경의 RIC에 AI/ML 모델을 유연하게 수행가능하게 하며, 다양한 Data를 제공할 수 있다는 점이다. [그림 6]에서는 AI-Native 프로그래머빌 RAN SW의 상위 수준 구조를 보여준다.

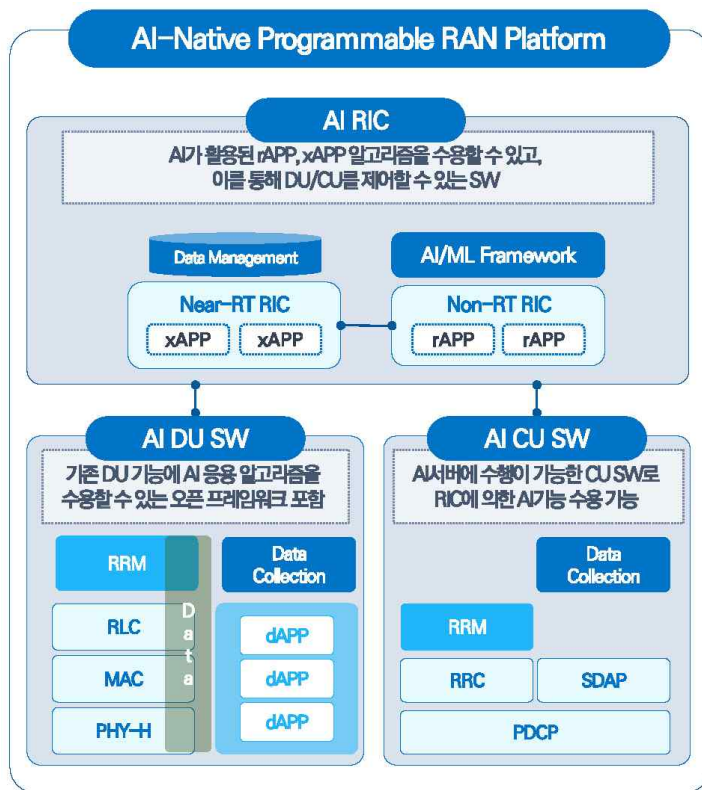


그림 6 ▶ AI-Native 프로그래머빌 RAN SW 상위 수준 구조

AI-Native 프로그래머빌 RAN SW은 5G/5G-A/6G RAN 소프트웨어를 구성함에 있어서 AI 알고리즘 탑재가 가능한 개방화된 구조를 가지며 [그림 6]에서 개념도를 보여준다. RAN에서 추출가능한 데이터를 오픈하여 내부/외부와 공유할 수 있는 구조를

포함하고, 이를 기반으로 RAN 최적화를 위한 AI 알고리즘 수용이 가능한 구조를 가져야 한다.

❖ 디지털 트윈 기반 가상 네트워크 플랫폼 기술

디지털 트윈 기반 가상 네트워크 플랫폼은 실제 네트워크의 동적 동작 특성을 정밀하게 반영하는 고정밀 가상 환경으로, AI 제어 알고리즘과 네트워크 정책을 반복적으로 검증하고 튜닝할 수 있도록 설계되며 [그림 7]과 같은 구조도를 갖는다. 이 플랫폼은 이동성, 채널 페이딩, 경로 손실, 도플러, 다중경로 간섭 등 현실적인 무선 채널 특성을 정밀하게 재현하며, 3D 환경 모델과 실제 측정 기반의 채널 데이터를 결합해 시간과 공간에 따라 변하는 채널 특성을 수치적으로 계산하여, 실제 환경과 유사한 무선 전파 변화를 가상 환경에 적용함으로써 실제 RAN에 근접한 조건을 구현한다. 가상 환경 내에서는 DU, CU, RIC, xApp, rApp이 컨테이너 기반의 마이크로서비스 아키텍처로 구성되며, 다양한 AI 모델을 유연하게 탑재하고 교체할 수 있어 반복 실험이 용이하다. 또한 실증망과 연동되어 페루프 기반의 학습 구조를 구현할 수 있으며, 가상 환경에서 생성된 정책이나 모델을 실증망에 적용하고, 그 결과를 다시 학습에 반영하는 순환 구조를 지원한다. 이를 통해 대규모 테스트베드 없이도 다양한 시나리오에 대한 사전 검증과 최적화가 가능해지며, 궁극적으로 AI 기반 자율 최적화 무선 네트워크 실현을 위한 핵심이 된다. [그림 7]은 AI-RAN 가상네트워크 환경 플랫폼 구조도를 보여준다.



그림 7 ▶ AI-RAN 가상 네트워크 환경 플랫폼 구조도

❖ 실증망 연계 및 라이브 트위닝 기반 운영 구조

실증망 연계 및 라이브 트위닝 기반 운영 구조는 실시간 운영 환경에서 AI 모델의 동작을 안전하고 효과적으로 통제·검증할 수 있도록 설계된다. 실증망에서 수집된 실시간 데이터는 디지털 트윈 가상망의 입력으로 활용되며, 가상망 내에서 추론된 AI 기반 제어 정책은 실증망에 적용된다. 이때 개방형 API 인터페이스를 통해 데이터와 제어 명령이 상호 실시간으로 연계된다. 운영 중인 AI 모델에 대해서는 성능 모니터링, 이상 탐지, 자동 재학습 트리거링, 정책 업데이트 및 롤백 기능이 포함되어, 모델 신뢰성과 제어 안정성을 확보할 수 있다. 특히, 모델의 설명 가능성(Explainability)을 제공하는 XAI(eXplainable AI) 기술이 함께 적용되어, 제어 의사결정의 근거를 분석하고 규제 대응이나 장애 대응에 활용할 수 있다. 이러한 구조는 AI-RAN이 고도로 복잡하고 예측 불가능한 환경에서도 지속적으로 최적화되고 자율적으로 적응하는 지능형 네트워크로 발전하는 데 필수적이다.

3.1.3 AI 기반 통신 센싱 융합 기술

통신 센싱 융합 (Integrated Sensing and Communications, ISAC)은 동일한 시스템에서 무선 통신과 센싱 기능을 결합하여 주파수/하드웨어 자원을 공유하는 기술이다. 기존의 ISAC 기법은 정밀한 수학적 모델에 의존하여 복잡도가 높고 실제 환경에서 성능 저하의 우려가 있었는데, 최근에 AI 기법의 강력한 학습/일반화 능력이 이를 해결할 대안으로 부상하였다^[26]. AI를 통합한 지능형 ISAC은 동적인 환경에서도 빠르게 적응하며, 모델 기반 접근을 데이터 기반으로 전환시키는 핵심 역할을 한다. 실제로 6G 이동통신 비전에서 ISAC은 AI와 더불어 핵심 구성 요소로 인정받고 있으며, 6G의 6대 핵심 활용 사례 중의 하나로 선정되었다^[27]. 이러한 AI 기반의 ISAC 기술은 자율주행 차량, 스마트 공장, 스마트 시티, 환경/기상 센싱, 로봇 등의 다양한 분야에서 6G 시대의 핵심 요소로 연구되고 있다.

AI 결합 ISAC은 자율주행, 스마트 공장, 스마트 시티, 환경/기상 센싱, 로봇 분야 등의 다양한 분야에서 활용될 수 있다. 특히, 자율주행과 스마트 공장 분야에서의 활용은 다음과 같이 요약한다. 먼저, 자율주행 차량에서는 차량 레이더 센싱과 차량통신을 통합해 주행 안전과 통신 효율을 동시에 향상시키는 연구가 활발하다. 특히 차량 간 또는 차량-인프라 간 통신신호를 센싱에 재활용하거나, 차량 레이더 신호에 데이터를 실어 양측 기능을 병행하는 방식 등이 고안되고 있다. 동적인 도로 환경에서는 통신과 센싱 목적 간 자원 할당 트레이드오프가 발생하는데, AI로 이러한 문제를 해결하는 사례들이 보고되었다. AI 기반 파형 및 자원 최적화의 경우 심층 강화학습 기반의 ISAC 파형 최적화에 대한

사례가 보고되었다[28]. 딥러닝 기반 빔포밍 및 위치 예측 관련하여 레이더/센서 데이터를 딥러닝 처리가 가능하다[26]. 통신-센싱 상호보완 및 협력 분야에서는 통신망과 차량 센서 간의 데이터 공유 사례를 보고하였다[29]. 다음으로 스마트 공장에서는 센서와 산업용 통신을 결합하여 공장 자동화와 운영 효율을 높이는 ISAC 기술이 주목받는다. 5G 특화망 (Private Network)가 보급됨에 따라 공장 내 무선 시스템이 생산 설비의 눈과 귀 역할을 할 수 있게 되었고, 6G에서는 ISAC으로 한층 향상된 정밀 모니터링과 제어가 가능해질 것으로 기대된다. 공장은 비교적 환경이 제한적으로 제어가 가능하므로 ISAC 적용에 유리하며, 다양한 AI 기법을 통해 설비 상태 감지, 로봇 경로 최적화, 가상 시뮬레이션 등을 구현하는 연구가 이루어지고 있다. 실시간 객체 인식 및 장애물 회피 관련하여 무선신호로 작업자 제스처나 설비 움직임 인식에 대한 연구가 발표되었다[30]. 디지털 트윈 기반 운영 최적화에 대하여 스마트 공장 디지털 트윈을 통해 로봇/기계 최적화 연구가 보고되었다[29]. 안전 및 보안 모니터링 분야에서는 CCTV 없이 침입자나 이상 움직임을 무선신호 감지할 수 있었다[30].

AI 기반의 ISAC 기술이 가져오는 기술적 이점은 다음과 같이 정리될 수 있다. 첫째, 자원 효율 및 비용 절감이 가능하다. 즉, 하나의 시스템으로 센싱과 통신을 겸하므로 주파수, 하드웨어 자원의 중복 사용이 감소할 수 있다. 둘째는 상호 성능 향상이 가능하다는 점이다. 센싱으로 얻은 환경 정보는 빔포밍 최적화, 간섭회피 등을 통해 통신 신뢰성을 향상시킬 수 있고, 통신망에서 모인 데이터는 AI 분석을 통해서 센싱 정확도 향상에 기여할 수 있다. 셋째, 새로운 서비스와 응용이 가능한데, 예를 들어서 ISAC으로 확보한 정밀 센싱 데이터와 AI 분석으로 디지털 트윈, 자율주행, 스마트 시티 관리 등이다. 넷째, AI 활용에 의한 지능화가 가능해진다. AI 도입으로 ISAC 시스템은 복잡한 최적화 문제를 학습으로 해결하고 동적 환경에 실시간 적용이 가능하다.

위와 같은 장점을 살려 AI 기반 ISAC 기술을 실용화하기 위해서는 해결해야 할 도전 과제도 적지 않다. 첫째, 센싱-통신 간 트레이드오프를 고려해야 한다. 센싱 성능을 높이기 위해 레이더 신호 성분을 늘리면 통신 대역이 감소하고, 반대의 경우에는 센싱 신호가 약해지는 문제가 발생하므로 동작 자원 배분 알고리즘이나 신호 설계 최적화로 균형점을 찾아야 하는 것이다. 둘째, 하드웨어 및 구현 복잡성 문제이다. 하나의 기기에 두 기능을 통합하기 위해서 혁신적인 RF, 안테나, 회로 설계 기법 도입 필요하다. 셋째, 동기화 및 지연 문제이다. 여러 센서 노드에서 모든 데이터를 융합하기 위해서는 마이크로 초 단위의 정밀 동기화, 로봇 분야에서 협업을 위해 낮은 처리 지연 시간 필요하다. 넷째, 보안 및

프라이버시 문제이다. 네트워크로 환경을 감지하는 것은 개인정보 침해나 감시 사회에 대한 우려 발생이 가능하므로 윤리적인 가이드라인 마련이나 프라이버시 보호 기술 필요하다. 마지막으로 표준화 및 규제 관련하여 스펙트럼 공유, 간섭 허용 기준에 대한 규제의 정비, 6G 규격에서의 국제 표준화가 필요하다.

AI 기반의 ISAC 기술은 다양한 활용 분야와 장점을 고려할 때 위에서 언급한 도전 과제들을 해결하면서 6G 시대에 필수 통신 인프라로써 본격적으로 꽃피울 것으로 전망된다.

3.2 AI 기반 자율 네트워크 기술

AI 기반 자율네트워크는 기존의 정적이고 수동적인 네트워크 관리 방식을 넘어, 스스로 학습·적응·제어하는 지능형 네트워크를 지향한다. 이는 대규모 연결, 초저지연, 초신뢰성이 요구되는 차세대(6G 포함) 통신 환경에서 필수적인 기술이다. 세부기술로는 네트워크 파운데이션모델, AI 기반 지능형 위성통신 기술 및 완전자율네트워크를 위한 종단간 지능형 네트워크 시스템 기술이 있다.

3.2.1 AI 네트워크를 위한 파운데이션 모델

최근 통신 네트워크 분야에서도 파운데이션 모델(Foundation Model)의 개념이 빠르게 부상하고 있다. 특히 MWC 2024에서 화웨이가 업계 최초의 통신망 파운데이션 모델을 발표하고 이를 자사 자율 주행 네트워크(Autonomous Driving Network, ADN) 솔루션에 적용해 네트워크 운영의 지능화를 달성한 사례를 공개하면서 대규모 AI 모델의 실효성을 현장에서 입증된 바 있다. 이에 따라, 본 절에서는 네트워크 파운데이션 모델(Network Foundation Model, NFM)이 대규모 비지도 학습을 통해 통신망에 특화된 지식을 축적할 수 있게 하는 사전학습(Pre-training) 기법과 이를 통해 일반화 능력과 잠재력을 확보한 NFM을 다양한 과제 해결에 폭넓게 적용하기 위한 최적화(Fine-tuning) 전략에 대해 기술한다.

❖ 대규모 멀티모달 네트워크 데이터 사전학습

가. 네트워크 파운데이션 모델 사전학습

주로 텍스트나 음성, 이미지 등 범용 데이터를 처리하도록 설계된 일반적인 LLM과 달리 NFM은 트래픽 패턴, 장비 로그, 토폴로지, 트래픽 통계, 장애 이력, 네트워크 구성 파라미터, 심지어 현장 장비 이미지나 지리정보 등 네트워크 고유의 멀티모달 데이터를 다루도록

특화된 점이 특징이다. 이러한 NFM을 사전학습 하기 위해서는 먼저 도메인 특화된 토큰화 방법이 필요하다. 예를 들어 패킷 흐름이나 트래픽 매트릭스 같은 다차원 네트워크 데이터를 효과적으로 벡터화하여 모델 입력으로 사용할 수 있어야 한다. IP 패킷의 헤더 필드, 사용자 단말의 위치 변화, 라우팅 토폴로지 등의 정보를 고차원 임베딩으로 변환해주면 모델이 고수준 특성을 추출하기 용이하다.

또한 모델 구조와 학습목표 역시 네트워크 도메인에 맞게 재설계가 필요하다. 다양한 종류의 데이터가 섞인 복합 입력에 대해 수렴하도록 학습 레이어 구성, 매개변수 규모, 손실 함수 및 학습 정책을 최적으로 조정해야 하며, 이를 통해 이질적인 네트워크 데이터들을 함께 학습시켜도 안정적으로 모델이 훈련되도록 해야 한다. 모델은 대규모의 이중 데이터를 자기지도학습 방식으로 학습하여, 데이터 간 숨은 상관관계나 패턴을 내재적으로 익히게 된다. 예를 들어 트랜스포머(Transformer) 기반 모델에 네트워크 타임시리즈 데이터와 로그를 함께 입력하여 다음 발생 이벤트 예측이나 비정상 패턴 검출 등의 목표로 학습시키면, 모델 내부에 네트워크 동향에 대한 일반화된 표현이 축적된다. NFM은 다양한 원천의 대용량 데이터를 장기간 학습함으로써 통신망 전반에 대한 포괄적 지식 기반을 구축하며, 이후 비교적 소량의 추가 학습으로도 여러 전문 업무에 적용될 수 있다. [그림 8]은 국내 산학연 중심의 자립형 네트워크 파운데이션 모델 확보 및 실증을 위한 전략을 하나의 그림으로 도식한 것이다.

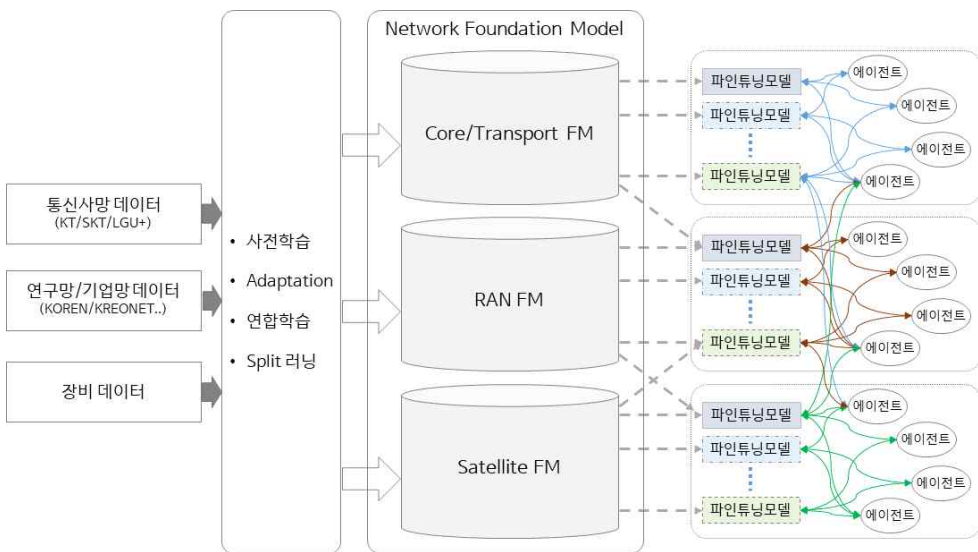


그림 8 ▶ 국내 산학연 중심의 자립형 네트워크 파운데이션 모델 확보 및 실증 전략

나. 네트워크 파운데이션 모델의 종류

통신 분야에 적용되는 파운데이션 모델은 모델의 입력 모달리티와 적용 범위에 따라 여러 유형으로 나눌 수 있다. 우선 대규모 언어 모델 타입은 네트워크 로그나 구성 명령어 같은 텍스트 데이터를 주로 학습하며 질의응답, 문제 원인 분석, 티켓 분류 등 자연어 처리 기반 업무에 활용된다. 다음으로 멀티모달 파운데이션 모델은 여러 형태(Modality)의 데이터를 동시에 입력 받아 통합적으로 이해하고 처리하는 모델을 말한다. 예를 들어 멀티모달 NFM은 기지국 로그(텍스트)와 트래픽 시계열(수치), 커버리지 지도(이미지)를 함께 받아 종합적으로 셀 이상 원인을 추론할 수 있다. 더 나아가 강화학습 기법을 접목해 최적화된 네트워크 제어 결정을 내리는 모델이나, 그래프 신경망(GNN)을 기반으로 토폴로지 구조를 학습하는 모델 등 특수 목적의 파운데이션 모델도 연구되고 있다. 또한 한 통신망 내에서도 코어망, 액세스망, 전송망 등 도메인별로 전문화된 파운데이션 모델을 구축하는 방향도 논의되는데, 궁극적으로는 유무선 위성을 포함하는 전구간의 특화 모델들을 상호 연동하거나 상위에서 조율하는 오케스트레이션 기술을 통해 전체망을 커버하는 지능형 체계를 이루도록 할 수도 있다. 이처럼 NFM은 다양한 형태로 구현될 수 있으며, 상황에 맞게 단일 거대 모델 또는 여러 전문 모델의 조합으로 통신망 AI 생태계를 구성할 수 있다.

❖ 거대 파운데이션 모델 최적화 기술

네트워크 장비들은 계산 자원이 한정적이므로 학습된 대규모 모델을 경량화하고 효율적으로 미세조정하는 기술이 필수적이다. 또한 실시간성 제어를 위한 NFM은 목표 성능을 위해 토큰화부터 아키텍처, 최적화까지 전방위 커스터마이징이 요구된다.

가. 미세조정(Fine-tuning) 학습

미세조정은 사전학습된 거대 모델에 소량의 추가 학습을 시켜 특정 작업이나 도메인에 성능을 최적화하는 기법이다. 이를 통해 모델이 새로운 데이터에 빠르게 적응하여 모델을 처음부터 다시 학습하는 것보다 효율적으로 특정 과제에 고성능을 달성할 수 있다. 특히 통신망 분야에서는 각 통신망의 트래픽 특성, 환경이 다르므로, 기본 모델을 현장 데이터로 미세조정해 맞춤형 모델로 만드는 과정이 중요하다.

최근 거대 모델의 일부 파라미터만 학습하는 경량 미세조정 기법들이 각광받는데, 예를 들면 LoRA(Low-Rank Adaptation)나 어댑터 삽입 같은 방법으로 전체 수백억 개 매개변수 중 극히 일부만 조정하면서도 성능 향상을 얻는다. 이러한 PEFT(Parameters-Efficient Fine-Tuning)

기법은 전체 모델을 재훈련하는 부담 없이도 새로운 네트워크 상황 변화에 모델을 빠르게 업데이트할 수 있기 때문에 통신망처럼 실시간성이 요구되고 잦은 재학습이 필요한 환경에서 특히 유용하다. 더 나아가 미세조정된 여러 버전의 모델을 지식 병합(Ensemble) 하거나 연속 학습(Continual Learning)을 통해, NFM을 지속적으로 현행화하고 성능을 고도화 하는 방향으로 발전시킬 수 있다.

나. 파운데이션 모델 경량화 기술

거대 모델을 실제 네트워크 현장에 적용하려면 추론 속도 향상과 배포 비용 절감을 위한 경량화 노력이 필수적이다. 우선 많이 사용되는 방법은 모델 압축(Model Compression) 기술들로, 예를 들어 중요도가 낮은 가중치를 제거하는 모델 가지치기(Pruning), 높은 정밀도의 파라미터를 저비트로 표현하는 양자화(Quantization) 기법 등이 있다. 이들은 모델 크기와 연산량을 크게 줄여주지만, 과도한 압축은 성능 저하를 유발할 수 있어 정확도와 속도의 균형을 찾는 것이 중요하다. 또 다른 접근으로 지식 증류(knowledge distillation)가 활용되는데, 거대 모델의 지능을 보다 작은 경량 모델에 증류하여 추론 전용으로 사용하는 것이다. 이를 통해 비슷한 출력을 내는 작은 모델을 만들어 엣지 장비 등 제약된 환경에 배포할 수 있다.

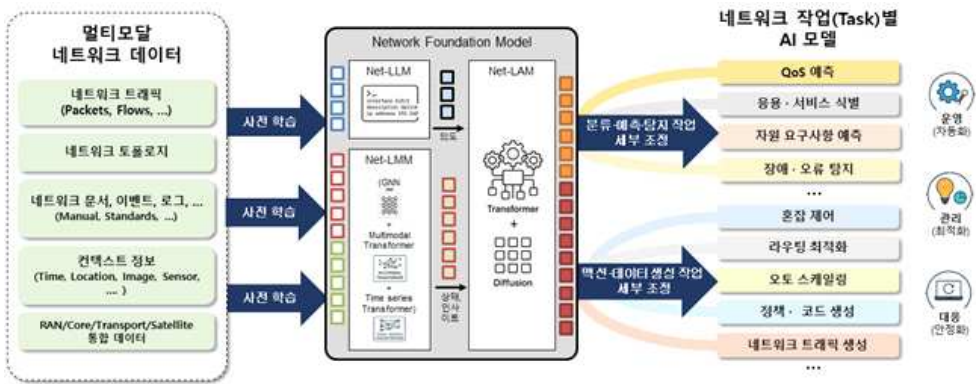


그림 9 ▶ 네트워크 파운데이션 모델 데이터 학습 및 활용 개념도

아울러 분산 추론도 경량화의 한 축이다. 통신망 장비나 모듈에 분산된 형태로 모델을 부분 배치하고, 연합 학습(Federated Learning)이나 분할 학습(Split Learning)으로 각자 계산한 결과를 취합함으로써 중앙 집중형 거대 연산 부담을 줄이는 것이다. 예컨대 일부 레이어는 기지국에서 수행하고 나머지는 중앙 클라우드에서 수행하는 식으로 역할을 분배하면,

실시간 응답이 필요한 연산을 가까운 곳에서 처리하여 지연을 줄일 수 있다. 이런 분산 기법은 통신망 데이터의 프라이버시 보호 측면에서도 이점이 있어 연구가 활발하다. 결국 경량화 기술들의 종합적 목표는 거대 모델을 실시간 운영 가능한 수준으로 최적화하는 데 있다. 아직도 대규모 모델의 밀리초 단위 추론에는 도전과제가 남아 있으나, 압축, 최적화, 하드웨어 가속을 아우르는 노력으로 통신망 현장에서 대규모 AI 모델의 상용 구현 가능성이 점차 현실화되고 있다.

3.2.2 AI 기반 지능형 위성통신 기술

전 세계적으로 인공위성을 활용한 통신 서비스의 수요가 급격히 증가하고 있으며, 특히 저궤도 위성(LEO: Low Earth Orbit)을 기반으로 하는 이동통신 서비스가 주목받고 있다. LEO 위성은 기존 정지궤도(GEO) 위성과 비교해 지연 시간이 짧고 커버리지가 유연하여 향후 6G 초공간 네트워크의 확장에 중요한 역할을 할 것으로 기대된다. 그러나 LEO 위성통신은 위성이 빠르게 움직이며 지속적으로 셀 위치가 바뀌는 특성으로 인해 핸드오버 빈도가 높고, 위성 자원 배분이 동적으로 이루어져야 하며, 도플러 효과와 전파 지연 문제 등도 복잡하게 작용한다. 이러한 환경에서는 기존의 고정 규칙 기반 네트워크 제어 방식이 효과적으로 작동하지 못하고, 실시간 예측 및 자율 제어가 가능한 지능형 접근 방식이 요구된다. 이에 따라 최근 위성통신에 인공지능(AI) 기술의 도입은 단순한 선택이 아닌 필수로 여겨지고 있으며, 위성통신 자율 네트워크 구축, 자원 최적화, 링크 품질 예측, 핸드오버 제어, 장애 탐지 및 복구 등 다양한 분야에서 [그림 10]과 같은 AI 기반 접근법이 활발히 연구되고 있다. 본 보고서에서는 이러한 흐름을 반영하여, AI 기반 지능형 NTN 네트워크의 다양한 기술 개발 분야를 살펴 보고자 한다.

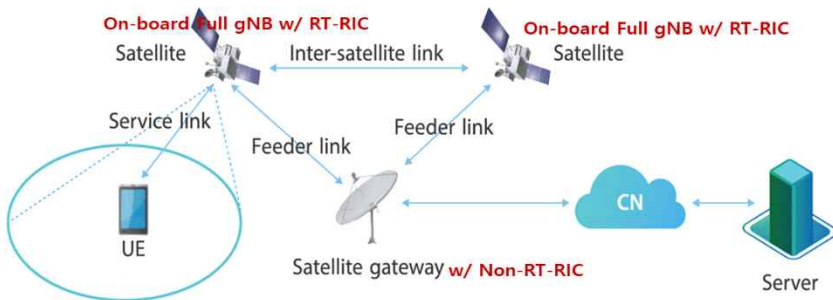


그림 10 ▶ AI 기반 지능형 NTN 개념도 (예시)

❖ AI 기반 지능형 NTN을 위한 개발 동향

AI 기반 지능형 NTN 기술은 위성 통신 시스템의 자율성과 서비스 품질을 향상시키기 위한 핵심 기술로 주목받고 있으며, 특히 통신 탑재체의 OBP(On-Board Processing)는 AI 추론과 제어 기능을 위성 내에서 직접 수행함으로써 네트워크 지연을 최소화하고, 자율성을 극대화하는 기술 요소로 부상하고 있다. 최근 OBP 기술은 RTL(Register Transfer Level) 수준의 하드웨어 고정형 처리 구조에서 벗어나, 유연성과 재구성이 가능한 소프트웨어(SW) 기반 구조로 전환되고 있다. 기존 RTL 기반 시스템은 연산 처리 효율은 높지만 임무별 유연성이나 AI 알고리즘 통합에 제약이 크기 때문에, 최근 위성 통신 및 관측 분야에서는 SW 기반 OBP 아키텍처에 대한 수요가 증가하고 있다. <표 6>은 HW 모델과 SW 모델의 특징을 분석하여 정리된 내용이다.

표 6 ▶ HW모델과 SW모델 특징 분석

항목	HW 모델 (FPGA 구현)	SW 모델 (CPU 구현)
성능	고속 병렬 처리, 낮은 지연 시간, 최적화 가능	직렬 처리, 고속 데이터 처리에 제한적
전력 소비	낮음 (효율적)	상대적으로 높음
유연성	하드웨어에 최적화 되어 제한적	소프트웨어로 구현 가능하여 유연함
개발 시간	개발 기간이 길고 디버깅 과정이 어려움	상대적으로 작업 처리 가능, 범용적 활용 가능
확장성	특정 용도에 최적화되어 확장성 어려움	다양한 작업 처리 가능, 범용적 활용 가능
유지보수	하드웨어 수정 시 재설계 필요	소프트웨어 업데이트만으로 유지보수 가능
우주 환경	우주환경 적용 사례가 다수	고속 동작 CPU의 우주환경 검증 사례 부재

따라서 우주 환경의 특수성을 고려한 가상화 기반 모델 플랫폼 선정 및 기술 개발을 통해 SW기반 6G NTN 저전력, 고신뢰 모델 기술 및 상위 스택 개발이 중요하다.

❖ AI 기반 지능형 NTN 핵심 기술

위성 통신 환경의 특수성을 고려한 통신 품질 향상, 제한된 자원 활용 최적화, 간섭 제어 등과 같은 RRM 분야에 활용될 수 있으며, 크게 다음의 분야들이 포함된다.

가. 자율 위성: 온보드 AI 기반 데이터 처리 및 운영

전통적인 위성은 지상국의 명령을 받아 데이터를 중계하는 '벤트 파이프(bent-pipe)' 역할에 머물렀다. 그러나 AI 기술은 위성을 재생형 기반의 스스로 판단하고 동작하는 '지능형 노드'로 변모시키고 있다. 위성에 탑재된 AI는 실시간으로 데이터를 분석, 필터링, 압축하여 위성-지상 간 통신 링크의 부하를 획기적으로 줄일 수 있다. 이를 위해 고성능 프로세서와

소프트웨어 플랫폼이 위성에 탑재된다. 이를 통해 이미지 분석을 위한 합성곱 신경망(CNN), 시계열 데이터 분석을 위한 순환 신경망(RNN)과 같은 머신러닝 모델을 위성에서 직접 구동할 수 있다. 또한, AI는 위성의 각종 원격 측정(telemetry) 데이터를 지속적으로 모니터링하여 이상 징후를 감지하고 잠재적 고장을 예측한다. 이는 위성의 수명을 연장하고 운영 신뢰도를 높여, 지상국의 개입과 운영 비용을 최소화한다.

나. 지능형 스펙트럼 운용: AI 기반 동적 빔포밍

LEO 위성의 빠른 속도와 메가 컨스텔레이션의 높은 위성 밀도는 빔 관리, 간섭, 채널 안정성 측면에서 전례 없는 어려움을 야기한다.²⁰ AI는 이러한 동적인 환경을 실시간으로 제어하는 데 필수적이다.

- AI 기반 동적 빔포밍(Dynamic Beamforming): 심층 신경망(DNN)과 심층 강화학습(DRL) 같은 AI 기술은 실시간으로 빔의 형태와 방향을 최적화한다. 이를 통해 특정 사용자를 향한 신호 강도를 극대화하고, 인접 빔이나 다른 위성으로의 간섭은 최소화한다. 특히, 계산 자원이 제한된 위성의 한계를 극복하기 위해, 복잡한 AI 모델은 지상의 강력한 서버에서 훈련시킨 후, 가볍고 효율적인 모델만 위성에 탑재하여 실시간으로 실행하는 '지상 훈련, 우주 실행(Train on Ground, Execute in Space)' 모델이 핵심 아키텍처로 자리 잡고 있다. 이 방식은 위성의 채널 상태 정보(CSI)가 지연되어 낡은 정보가 되는 문제를 효과적으로 해결한다.

다. 네트워크 패브릭 최적화: AI 기반 핸드오버, 도플러 보상, ISL 기술

- 도플러 효과 보상(Doppler Effect Compensation): 위성과 지상 사용자 간의 빠른 상대 속도(약 7.8 km/s)는 심각한 도플러 효과를 유발하여 통신을 방해한다. AI 모델은 위성의 궤도 정보를 기반으로 도플러 편이를 높은 정확도로 예측하고, 신호 전송 전에 주파수를 미리 보상함으로써 안정적인 통신 링크를 유지할 수 있게 한다.
- 핸드오버 최적화(Handover Optimization): LEO 위성은 한 지점에서 약 10분 정도만 관측 가능하므로, 위성 간 핸드오버가 빈번하게 발생한다. 비효율적인 핸드오버는 서비스 중단이나 네트워크 과부하(signaling storm)를 유발할 수 있다. 강화학습(DRL)과 그래프 신경망(GNN) 기반의 AI 모델은 최적의 핸드오버 시점과 대상 위성을 예측하여 불필요한 전환을 줄이고 지연 시간을 최소화한다.
- Mega Constellation 저궤도 위성통신 환경에서 ISL(위성 간 링크) 기술은 지상 게이트 웨이에만 의존할 경우 발생할 수 있는 트래픽 집중, 지연 증가, 서비스 불연속성, 그리고 높은 CAPEX/OPEX 문제를 해소할 수 있는 핵심 네트워크 운용 기술이다. 한편 AI

기본 지능형 ISL 네트워크 기술은 위성 간 라우팅 최적화, 배터리·메모리 등 제한된 자원의 효율적 스케줄링, 위성 궤도, 트래픽 패턴, 사용자 수요를 고려한 예측 기반의 네트워크 운영, ISL링크에서 QoS 보장, 이상 탐지 및 자율 복구가 가능하게 된다. 따라서 AI는 ISL 기반 네트워크의 효율성과 지능화를 가속화하는 필수 요소이다.

라. 지능형 에너지 절감 운용: AI 기반 전력 제어, 자원 스케줄링, 링크 품질 예측

LEO 위성은 빠른 궤도 이동과 수많은 사용자 링크 요구로 인해 에너지 소모가 크며, 제한된 전력 자원 내에서 효율적 운용이 필수적이다. 특히 SW 기반 OBP 구조는 AI 기술을 활용한 에너지 절감 기능의 유연한 구현이 가능하다는 장점을 가진다.

- AI 기반 전력 제어: AI는 통신 상태, 사용자 수요, 채널 품질 정보를 실시간 분석하여, 트래픽이 상황에 맞는 송신 출력을 제어하고, 필요한 자원만 선택적으로 활성화함으로써 에너지 소비를 최적화한다.
- AI 기반 자원 스케줄링: 딥러닝 및 강화학습 기반 알고리즘은 사용자의 위치, QoS 요구사항, 간섭 수준을 종합적으로 고려하여 송수신 타이밍과 주파수 자원을 효율적으로 스케줄링하고, 이를 통해 불필요한 전력 소모를 방지한다.
- 링크 품질 예측 기반 전력 최적화: AI는 과거의 CSI와 궤도 데이터를 학습하여 링크 품질 변화를 예측하고, 이를 바탕으로 송신 전력을 선제적으로 조절함으로써 에너지 절감과 통신 안정성을 동시에 확보한다.

3.2.3 지능형 네트워크 시스템 기술

5G 고도화 및 6G의 등장, 클라우드와 IoT의 확산으로 통신망이 급격히 복잡해지면서, 기존 수동 관리 방식은 한계에 직면하고 있다. 차세대 지능형 네트워크는 관리자의 의도만으로 성능 관리, 장애 대응, 서비스 오케스트레이션을 자동 실행해 실시간 최적화와 Zero-Touch 운용을 구현한다. 결국, 네트워크가 스스로 판단하고 행동하는 자율 지능을 갖추게 되는 것이다. 이를 가능하게 하는 핵심은 여러 전문 AI 에이전트가 협력하고 추론하는 에이전틱(Agentic) AI 기술이다.

❖ 에이전틱 AI

에이전틱 AI는 여러 특화된 AI 에이전트들이 자율적으로 협력하고 소통하며, 복잡한 목표를 달성하기 위해 동적으로 작업을 할당하는 새로운 인공지능 패러다임이다. 기존 통신망 운용의 AI는 주로 인간이 미리 정한 폐쇄형 제어(loop)나 과거 데이터에 기반한 수동적 학습에 의존하여, 새로운 문제 상황에 스스로 해법을 찾아내거나 급변하는 상황에 적응하는

데 한계가 있었다. 반면 에이전틱 AI는 능동적 추론과 행동을 통해 이러한 한계를 극복하고 일반적인 지능과 동적 환경 적응력을 갖춘 네트워크 운영을 지향한다. 예를 들어, 기존 네트워크 AI가 트래픽 최적화나 장애 예측에 단순히 반응하는 수준에 그쳤다면, 에이전틱AI는 목표를 스스로 설정하고 다른 에이전트와 협력 및 자원 협상을 수행하며 실시간으로 문제를 선제적인 해결하는 주체로 동작한다.

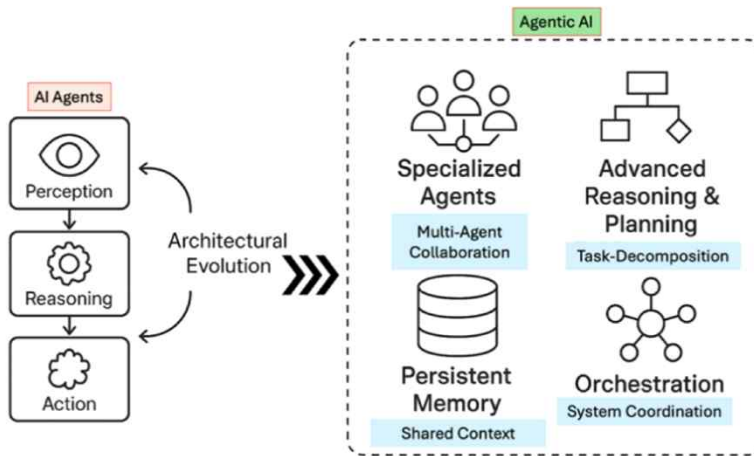


그림 11 ▶ AI 에이전트에서 에이전틱 AI로 진화

에이전틱 AI는 [그림 11]과 같이 여러 특화 에이전트가 유기적으로 협력하기 위한 시스템적 특성을 갖는다. 1) 다중 에이전트 협력; 여러 특화 에이전트가 각자의 역할로 부분 문제를 해결하고 결과를 공유해 복잡한 목표를 달성하며, 시스템 전체 성능을 극대화한다. 2) 시스템 조정; 상위 오케스트레이션 계층이 모든 에이전트의 행동을 통합하고 조율해 자원을 효율적으로 배분하고 충돌을 방지함으로써 일관된 목표 달성을 지원한다. 3) 공유 컨텍스트; 모든 에이전트가 동일한 환경 정보와 목표, 상태를 공유해 각자의 판단과 행동이 전체 시스템 맥락에 부합하도록 보장한다. 4) 작업 분해; 시스템은 복잡한 문제나 목표를 자동으로 하위 작업으로 세분화한 뒤, 각 작업을 가장 적합한 에이전트에 할당해 신속하고 효과적인 문제 해결을 구현한다.

❖ 통신망 특화 추론형 모델(Reasoning)

통신 네트워크의 규모와 복잡성이 급증함에 따라 네트워크의 자율적 운영에 대한 요구가 커지며 이에 부응하는 기술로 에이전틱 AI 개념은 통신 네트워크 영역에서 주목받고 있다.

즉, 통신 네트워크에서 특정 영역이나 작업을 자동화하기 위한 AI 에이전트들이 서로 역할을 분담하고 상호 소통, 중앙 또는 분산 오케스트레이션 하에 집합적으로 동작하여 전체 목표 달성을 위해 협력이 필요하다.

이를 위한 추론형 모델(Reasoning)은 복잡한 문제를 해결하기 위해 여러 변수를 종합적으로 고려해 논리적 해답을 도출하도록 설계된 모델로, AI가 스스로 판단하고 작업을 수행하는 에이전틱 AI 구현의 핵심 기술이다. 추론형 모델은 에이전트의 의사결정과 문제 해결 과정에서 논리적이고 맥락적인 추론을 지원해 복잡한 목표를 자동으로 하위 작업으로 분해하고 최적의 계획을 수립하며, 에이전트 간 협력 및 조정 과정에서 상황을 해석해 적절한 행동을 선택하도록 돕는다.

지능형 네트워크 시스템을 위한 추론형 모델은 범용 LLM이 아니라 통신망의 맥락을 깊이 내재화한 전용 모델이어야 한다. 범용 언어모델은 네트워크 토폴로지나 프로토콜, 장비 구성 정보, SLA 간 인과관계를 충분히 이해하지 못해 잘못된 설정을 제안할 위험이 있다. 게다가 LLM은 대부분 자가회귀(순차의존) 구조로 출력 길이에 비례해 지연이 누적되고, 수조 개의 파라미터에 따른 연산 부담까지 더해져 추론 시간이 길다. 이러한 특성은 실시간 장애 대응이나 초저지연 네트워크 제어에 치명적이다. 특히 RAN 제어와 같이 밀리초(ms) 단위 의사결정이 요구되는 영역에서는 순차 의존성과 모델 규모로 인한 지연을 최소화할 수 있는 경량 및 고속 전용 추론 모델이 필수적이다.

❖ 자율네트워크를 위한 에이전틱 AI 기술

통신망 전용 대규모 AI 모델인 파운데이션 모델과 통합 데이터 레이크를 바탕으로 무선접속망(RAN), 코어망, 전송망, 위성망을 하나의 지능형 패브릭처럼 동작하도록 지원할 수 있다. 이를 통하여 다수의 AI 에이전트가 실시간 KPI와 텔레메트리 데이터를 분석해 이상 징후를 조기 탐지하고, 자동 복구 및 구성 최적화를 실행해 장애 대응 시간을 대폭 단축한다. 아울러 셀 전력 제어와 자원 재배치를 통해 에너지 사용량과 운영비(OPEX)를 절감하는 효과도 얻을 수 있다.

또한 운영지원시스템(OSS)과 서비스 관리 영역에서 통신사가 자체 AI 에이전트를 설계하고 시험, 배포할 수 있는 통합 환경 및 시뮬레이션 기능을 통해 에이전트의 예측 불가능한 행동을 사전에 검증해 신뢰성을 높인다. 이러한 기술을 OSS/BSS 솔루션에 통합해 실시간 데이터 스트리밍과 AI 기반 상관분석이 가능한 텔코용 데이터 레이크 플랫폼을 구축하고 의도 기반 네트워크 운영을 강화할 수 있다.

궁극적으로 에이전틱 AI 개념을 적극적으로 도입하여 자율 네트워크 구현을 선도적으로 추진할 수 있다. Agentic AI 네트워크는 AI 기반 아키텍처를 적용해 종단간 네트워크가 Self-Optimization 및 Self-O&M 기능을 스스로 수행하도록 설계돼 서비스 수요 변동에 따라 네트워크 설정을 실시간으로 조정한다. RAN/Core/Transport/Satellite 영역에서도 여러 AI 에이전트가 셀 간 간섭 관리, 주파수 및 채널 최적화, 트래픽 핫스팟 대응 등 복잡한 작업을 병렬로 수행하도록 조율하여 네트워크 종단간 운용 효율과 사용자 경험을 동시에 높일 수 있으며, 궁극적으로 에이전틱 AI는 사람이 아니라 네트워크가 스스로 운용되는 시대를 앞당기는 촉매이자, 차세대 통신 인프라 경쟁의 결정적 승부수가 되고 있다.

3.3 AI 서비스를 위한 네트워크 기술

“AI 서비스를 위한 네트워크(Network for AI) 기술”은 생성형 AI, 대규모 언어모델(LLM), 연합학습(Federated Learning), 자율지능형 에이전트 등 차세대 AI 서비스가 요구하는 성능·지연·신뢰성·연산 구조를 안정적이고 효율적으로 지원할 수 있도록, 네트워크가 단순한 데이터 전달 경로에서 벗어나, AI 서비스를 위한 실행 플랫폼이자 실시간 오케스트레이션을 가능하게 하는 기반 인프라가 되어야함을 의미한다. 특히 단말 중심 구조에서 벗어나, AI 에이전트를 독립된 통신 주체로 식별하고, 세션, 정책, 자원, 경로를 에이전트 수준에서 제어할 수 있는 Agent-aware 구조로의 진화가 요구된다.

이와 같은 “Network for AI” 아키텍처를 실현하기 위해서는 전송, 제어, 데이터, 엣지, 센싱 인프라 전반에 걸친 구조적 재편이 필요하다. 본 보고서에서는 이 가운데에서도 AI 서비스의 실행 환경과 직접적으로 연결되며, 네트워크 인프라의 경계와 기능 확장을 요구하는 세 가지 핵심 기술 분야를 중심으로 구조적 과제를 살펴보고자 한다.

3.3.1 AI 데이터센터 네트워크 기술

대규모 언어 모델(LLM) 및 생성형 AI 등의 확산으로 데이터센터 내(Intra-DC) 또는 데이터센터 간(Inter-DC) 다수의 AI 컴퓨팅 자원(xPU(GPU, NPU, DPU, ...), Memory 등)을 기술·비용 측면에서 효율적으로 연결하고 운영하기 위한 AI 데이터센터 네트워크 인프라 시장이 급격하게 성장하고 있다.

데이터센터 내 AI 컴퓨팅 서버, 패브릭 가속 카드(DPU/SmartNIC), 고대역 메모리(HBM) 등이

빠르게 발전함에 따라 이들 간 연동을 제공하는 네트워크에서 병목 현상이 발생하고 있다. 이로 인해 다수의 GPU 간 대규모 데이터 통신을 위한 네트워크 부품/장비의 속도/용량 증대와, AI 모델 추론 및 학습 성능 향상을 위해 다수 GPU 간 연동에 최적화된 초(超)저지연 통신이 요구되고 있다. 이러한 AI 데이터센터 네트워크를 제공하기 위해서는 다음 [그림 13]와 같은 주요 핵심 기술 개발이 수반되어야 한다.

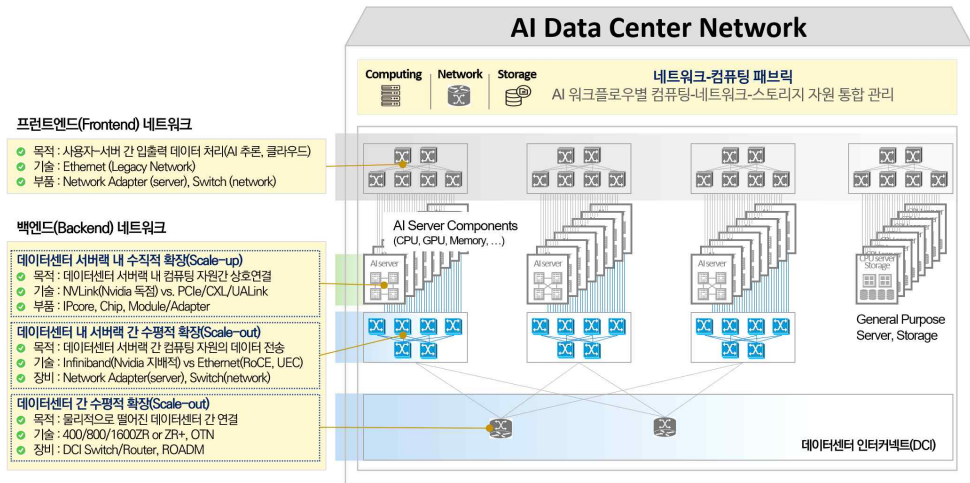


그림 13 ▶ AI 데이터센터 네트워크 핵심기술 구조도

❖ 네트워크-컴퓨팅 패브릭 기술

네트워크-컴퓨팅 패브릭(Network-Computing Fabric) 기술은 폭증하는 생성형 AI 서비스의 원활한 제공을 위해 AI 서비스 요구사항에 따라 네트워크-컴퓨팅 자원을 효율적으로 제공하고, 일관된 서비스 수준 협약(Service Level Agreement)을 제공하기 위한 AI 데이터센터 네트워크-컴퓨팅 패브릭 운영체제(OS) 및 플랫폼 관련 기술이다.

국외의 주요 글로벌 벤더(Nvidia, Broadcom, Cisco, Arista 등)들은 AI 워크로드 특성에 맞춘 AI-Ready 패브릭 장비를 지속적으로 개발하고 있다. Nvidia는 GPU 기반 상호연결 칩/장비 중심, Broadcom은 스위칭 칩 중심의 공급망 생태계 형성, Arista, Cisco, Juniper등은 자동화 및 지능형 운용 관리에 강점을 가지고 이에 대응하고 있다. 또한, Edgecore 등 화이트박스 진영은 개방형 네트워크 OS(SONiC)와 유연한 구성으로 연구 및 테스트베드 환경에 적합한 플랫폼을 제시하고 있다.

❖ 데이터센터 스케일-업 기술

데이터센터 스케일-업(scale-up) 기술은 AI 데이터센터 내 컴퓨팅 자원(CPU, 메모리) 간 고성능, 저지연 네트워킹을 통해 컴퓨팅 자원의 수직적 확장을 제공하는 기술이다.

고성능 AI 데이터센터 플랫폼 구축에서 Nvidia가 개발한 NVLink 기술은 여러 개의 GPU와 CPU들을 연결하여 컴퓨팅 성능과 효율을 높이는 기술로서 AI 플랫폼 시장을 주도하는 주요 기술로 주목받았지만 NVLink 기술은 가속기 중심의 배타적 기술로 멀티 벤더 생태계를 중심으로 하는 개방형 기술(CXL/UALink)이 대안으로 부상하고 있다. 또한, 서로 다른 서버에 분산된 CPU, GPU, 메모리, 스토리지 등의 컴퓨팅 자원들을 초광대역, 초저지연으로 연결하는 광 분산(optical disaggregation) 기술도 고성능, 고효율 AI 데이터센터를 위한 스케일 업 기술로서 연구개발 되고 있다.

❖ 데이터센터 스케일-아웃 기술

데이터센터 스케일-아웃(scale-out) 기술은 AI 데이터센터의 서버랙 간 고성능, 저지연 네트워킹을 통해 컴퓨팅 자원의 수평적 확장을 제공하는 기술이다.

AI 데이터센터 서버랙 간을 연결하는 백엔드 네트워크 분야는 독점적 GPU(Nvidia)를 기반으로 하는 고가의 인피니밴드(Infiniband) 기술이 시장의 상당 부분(80% 이상)을 점유하고 있다. 이러한 특정 기업 중심의 의존적 생태계를 탈피하기 위해, 하이퍼 스케일러(Meta, Microsoft 등), 시스템·칩 벤더(Cisco, Arista, AMD, Broadcom 등)들이 대거 참여한 산업연합체(UEC: Ultra Ethernet Consortium)에서 개발한 개방형 이더넷(Ultra Ethernet)이 대안 기술로 부상하고 있다. 또한, 대규모 지역적으로 분산된 여러 데이터센터들을 초광대역, 초저지연 네트워크로 연결해서 초거대 데이터센터로 확장하는 데이터센터 상호연결(Data Center Interconnect, DCI) 기술도 매우 중요하다.

3.3.2 AI 서비스 지원 모바일 코어 네트워크 기술

AI 서비스의 실시간 처리, 대규모 데이터 교환, 에이전트 기반 자율성이 강조됨에 따라, 모바일 코어 네트워크는 기존 구조만으로는 이를 효과적으로 수용하기 어려운 한계에 직면해 있다. 본 절에서는 이러한 한계를 극복하기 위한 기술 방향으로 ① 컴퓨팅-네트워크 융합, ② Agentic/Physical AI 지원 통신-네트워크 구조, ③ 데이터 프레임워크 중심으로 진화 가능성을 제시한다.

❖ 컴퓨팅-네트워크 융합 기술

6G 시대의 네트워크는 AI 서비스의 본격적인 확산에 따라 기존과는 전혀 다른 특성을 요구받고 있다. 대규모 모델 파라미터 전송, 학습 데이터 교환, 실시간 추론 입력 등은 반복적이면서도 지연에 민감하며, 동시에 다수 노드 간 전송을 필요로 한다. 이러한 특성은 단순한 데이터 전달 효율화를 넘어, 네트워크와 연산 자원을 통합적으로 고려하는 AI-네트워크 인프라 융합을 필요로 한다.

AI-네트워크 인프라 융합의 목표는 단말, 엣지, 클라우드 전 구간에서 연산과 네트워크를 공동 설계하여 분산 추론과 학습을 지원하고, 자원과 경로를 상황에 맞게 동적으로 최적화하며, 일관된 정책과 보안 체계를 유지하는 데 있다. 이는 기존 데이터 전송 최적화나 네트워크 정보 노출을 확장한 개념으로, 연산 배치와 오케스트레이션을 포함해 RAN 및 엣지와와의 긴밀한 결합을 전제로 한다. 이러한 접근은 AI 서비스 실행의 효율성을 극대화하고, XR이나 자율주행, 로봇틱스와 같은 초저지연·고신뢰 서비스의 실현을 가능하게 한다.

이를 뒷받침하기 위해 네트워크는 몇 가지 핵심 역량을 갖추어야 한다. 첫째, 연산이 단말·엣지·클라우드 어디에서 수행되는지에 따라 최적의 경로를 유도할 수 있는 컴퓨팅-네트워크 융합형 트래픽 제어 능력이 요구된다. 둘째, AI 기반의 분석을 통해 혼잡과 지연을 사전에 예측하고 자원과 경로를 선제적으로 조정하는 지능형 오케스트레이션 기능이 필요하다. 셋째, 응용 서비스의 지연·지터·손실 허용도를 QoS와 네트워크 슬라이스 구조에 직접 연계함으로써, 서비스 특화 QoS 보장을 실현해야 한다. 넷째, 연합학습이나 분산 추론이 멀티 클라우드·멀티 사업자 환경에서 이루어지는 만큼, 멀티도메인 최적화와 보안 강화가 필수적이다. 마지막으로, AI 모델이 지속적으로 진화하는 특성을 고려하여, 모델 배포와 업데이트, 성능 검증, 롤백과 같은 과정을 포함하는 AI 모델 수명주기 관리 체계가 요구된다.

❖ Agentic/Physical AI 지원 통신-네트워크 기술

6G 시대에는 네트워크와 상호작용하는 주체가 사람이나 단말에 국한되지 않고, 다양한 형태의 AI 에이전트로 확장될 것으로 예상된다. 이러한 에이전트는 크게 두 가지로 구분할 수 있다. Agentic AI는 대규모 언어모델이나 지능형 알고리즘을 기반으로 네트워크 상에서 동작하며, 디지털 공간에서 다른 에이전트와 협력하거나 사용자를 대신해 자율적으로 서비스를 수행한다. 반면 Physical AI는 자율주행차, 로봇, UAM 등 물리적 환경에서

실시간으로 동작하며, 안전성과 초저지연성을 동시에 요구하는 특징을 가진다.

네트워크는 이 두 유형의 AI 에이전트를 단말 내부의 단순 부속 기능이 아니라, 독립된 엔티티로 인식하고 제어할 수 있는 구조로 진화해야 한다. 이는 곧, 단일 단말 식별자와 세션 컨텍스트에 종속된 기존 구조를 넘어, 에이전트 개별 단위로 세션을 식별하고, 정책을 적용하며, 서비스 품질을 보장할 수 있는 새로운 제어 체계를 필요로 한다.

우선 Agentic AI 지원을 위해서는 다수의 디지털 에이전트가 동시에 네트워크와 상호작용할 수 있는 세션 관리와 자원 배분 기능이 요구된다. 예를 들어, 다중 에이전트 기반 서비스에서는 각 에이전트가 별도의 QoS나 지연 요구사항을 가질 수 있으므로, 네트워크는 이를 구분하여 관리하고, 필요 시 최적 경로와 정책을 동적으로 제공해야 한다. 또한 Open Gateway API와 NEF 기반의 노출 구조를 활용하여, 에이전트가 네트워크 상태를 조회하거나 정책을 능동적으로 요청·수정할 수 있는 환경을 마련하는 것이 중요하다.

한편, Physical AI 지원을 위해서는 실시간성과 안전성을 보장하는 제어 구조가 필요하다. 이를 위해 기존 단말 기반 NAS 시그널링과 병행하여, 에이전트가 직접 모바일 네트워크와 세션 및 정책을 교환할 수 있는 경량화된 시그널링 혹은 API 기반 연동 방식이 고려된다. 예를 들어, 자율주행차 내의 특정 에이전트가 주행 중 URLLC 슬라이스 전환을 요청하거나, 로봇 제어 에이전트가 이동성 이벤트에 따라 QoS를 조정할 수 있어야 한다.

❖ 데이터 프레임워크 기술

6G에서 논의되는 Data Framework는 단순한 데이터 저장이나 분석을 넘어, AI-Native 네트워크와 차세대 서비스(Agentic AI, Physical AI, ISAC 등)를 지원하기 위한 데이터 수집·관리·공유·활용의 전 주기 체계를 의미한다. AI 모델의 학습과 추론, 네트워크 지능화, 서비스 예측을 위해서는 대규모이면서도 이기종적인 데이터가 실시간으로 처리될 수 있어야 하며, 단말·RAN·엣지·코어·클라우드 등 네트워크 전 구간에서 생성되는 데이터를 일관되게 수집하고 공유·재활용할 수 있는 구조가 요구된다. 특히 데이터의 정확성, 품질, 보안, 프라이버시는 서비스 성능과 SLA 보장에 직접적인 영향을 미친다.

Data Framework는 크게 다음 네 가지가 고려되어야 한다. 첫째, 데이터 수집 단계에서는 단말, RAN, 코어, 엣지 노드, 외부 도메인에서 생성되는 데이터를 이벤트 기반과 스트리밍 기반으로 구조화하여 수집한다. 둘째, 데이터 관리 및 거버넌스 단계에서는 메타데이터 표준화, 접근 제어, 데이터 익명화/가명화, 품질 관리(QoD: Quality of Data) 등을

통해 데이터의 신뢰성과 일관성을 보장한다. 셋째, 데이터 공유 및 노출 단계에서는 NEF와 CAPIF를 확장하여 외부 응용 서비스나 제3자에게 데이터를 안전하게 제공하며, Open Gateway 및 CAMARA API와 연계하여 Data as a Service 모델을 실현한다. 넷째, 데이터 처리 및 활용 단계에서는 분산·연합 학습(Federated Learning), Split Inference, Network Foundation Model 학습 등 AI 서비스와 직접 연계되며, Digital Twin, Predictive QoS, Autonomous Network 제어와 같은 고도화된 응용을 지원한다.

이러한 프레임워크는 5G 대비 뚜렷한 진화를 보여준다. 5G의 NWDAF가 통계적 분석과 정책 최적화에 머물렀던 것과 달리, 6G Data Framework는 예측, 생성, 자율 의사결정까지 포괄하며, 데이터 활용 범위를 RAN과 코어를 넘어 단말·서비스·물리적 환경(센싱, UAM, 로봇 등)으로 확장한다. 또한 데이터 신뢰성과 보안 체계를 강화하여 데이터 경제(Data Economy) 구현의 기반을 마련한다.

현재 표준화 논의는 여러 그룹에서 병행되고 있다. SA2의 FS_6G_ARCH에서는 Data Framework를 6G 공통 아키텍처의 핵심 요소로 정의하고 있으며, SA1의 TR 22.870은 Digital Twin, Predictive QoS 등 AI/ML 기반 서비스 유즈케이스를 통해 데이터 수집 및 활용 요구사항을 다루고 있다. 아울러 SA6은 엣지 컴퓨팅과 연계된 데이터 처리 및 노출 방식을, RAN 및 CT 그룹은 RAN 데이터 노출 및 Edge-native 데이터 처리 구조를 각각 검토 중이다.

AI 네트워크 기술 로드맵 04

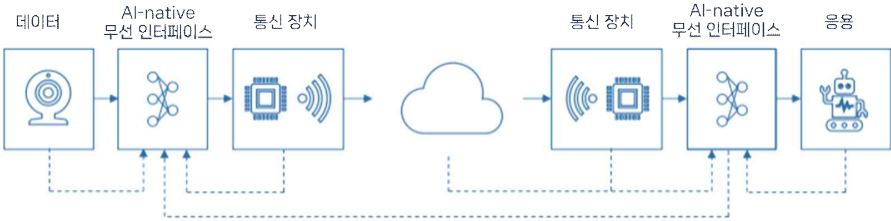
본 장에서는 AI 네트워크 기술을 AI 기반 무선 접속망, AI 기반 자율네트워크, 그리고 AI 서비스를 위한 네트워크 기술로 분류하여 각 기술에 대한 로드맵을 제시한다. [그림 14]는 각 기술에 대한 년도별 로드맵을 정리하여 도시하였다. AI 기반 무선 접속망 기술은 AI-native 무선 인터페이스 개발, AI-RAN 연구플랫폼 개발 및 AI-RAN 가상 실증 통합 구축 및 AI 기반 센싱 통합 6G 핵심 기술 개발이 필요하다. AI 기반 자율 네트워크를 위해서는 네트워크, RAN, 위성 관점에서 네트워크 파운데이션 모델이 각각 개발되어야 비로서 완전 자율네트워크 기술 개발이 완성될 수 있다. 또한 AI 서비스를 위한 네트워크 기술은 AI 데이터 센터 네트워크 기술, Edge AI 네트워크를 포함하는 Physical AI 지원 네트워크 핵심 기술 및 AI-RAN 대규모 실증망 구축을 통하여 이루어 질 것으로 보인다.



그림 14 ▶ AI 네트워크 기술 로드맵

4.1 AI 기반 무선 접속망 기술 로드맵

본 절에서는 AI 기반 무선 접속망 기술은 RAN에 AI가 도입될 때 현재 수행중인 과제를 제외하고 추가적으로 진행되어야 하는 과제에 대하여 기술한다. 향후 새롭게 기획되어야 하는 과제로 AI-native 무선 인터페이스 개발, AI-RAN 연구플랫폼 개발 및 AI-RAN 가상 실증 통합 구축 및 AI 기반 센싱 통합 6G 핵심 기술 개발이 필요하다. 관련된 기술에 대한 과제는 아래와 같이 제안될 수 있다.

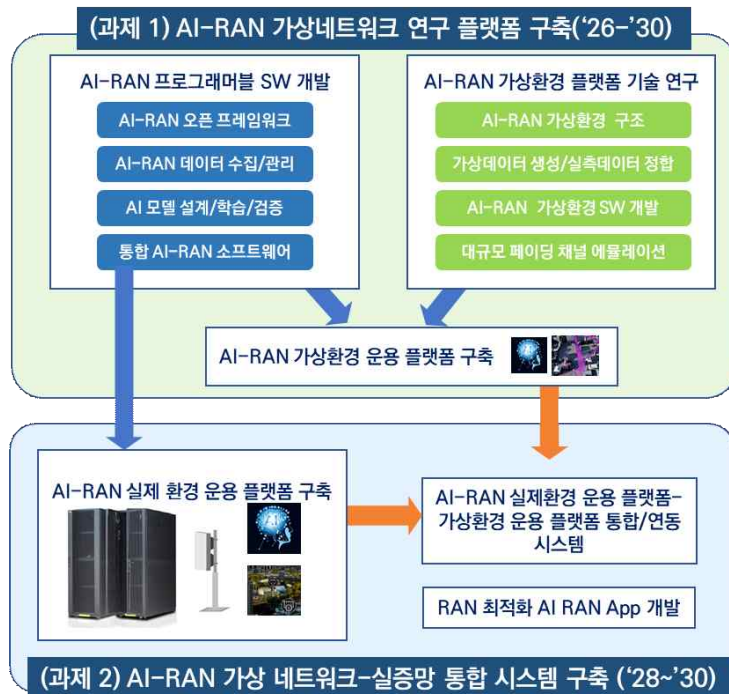
과제명	AI-native 무선 인터페이스 기술 개발
기술분야	AI 기반 무선접속망
<p>과제 개요</p> <ul style="list-style-type: none"> • 끊임없이 변화하는 무선망 환경에서 인공지능이 인간의 개입 없이 알고리즘 성능 평가, 데이터 수집, 학습을 통해 스스로 발전/진화(self-evolving)하는 RAN 제공을 위한 AI-Native 무선 인터페이스 기술 개발 <ul style="list-style-type: none"> - AI-Native 무선 인터페이스 핵심요소기술 개발 - 자기진화형 무선 인터페이스를 위한 분산 기계학습 핵심기술 개발 - AI-Native 무선 인터페이스 핵심기술 표준특허 확보 - AI 가속기 기반 AI 알고리즘 구현 기술 개발 - 통합기술시연(1건): AI-Native 응용서비스 지원 무선 인터페이스 SW 시스템  <p>The diagram illustrates the architecture of an AI-native wireless interface system. It shows a flow from '데이터' (Data) through an 'AI-native 무선 인터페이스' (AI-native wireless interface) to a '통신 장치' (Communication device), which then connects to a cloud. From the cloud, it goes to another '통신 장치' (Communication device), then through another 'AI-native 무선 인터페이스' (AI-native wireless interface) to an '응용' (Application). Dashed arrows indicate a feedback loop from the application back to the AI-native wireless interface and the communication devices.</p> <p>서비스 요구사항, 트래픽 패턴, 환경, 사용자 분포, 주파수, 하드웨어 및 응용 의존도 피드백</p>	

과제명 AI-RAN 연구 플랫폼 개발 및 AI-RAN 가상 실증 통합 구축

기술분야 AI 기반 무선접속망

과제 개요

- AI 중심의 글로벌 네트워크 시장 변화에 선제적으로 대응, AI-RAN 연구플랫폼 구축 및 한·미 협력 기반 AI-RAN 기술력 확보
- (AI-RAN 프로그래머블 SW) RAN기능에 AI 모델을 적용할 수 있는 구조를 수용하며 추론을 포함한 AI Workload의 증가하는 계산 수요에 대응해야 하고 AI와 RAN이 모두 동작가능하도록 구축.
- (AI-RAN 가상환경 플랫폼) 실제 환경과 연계한 무선 자원관리, 효율적인 전력사용등의 측정이 가능하도록 가상의 무선채널을 애플레이션, 가상 AI-RAN 소프트웨어 및 가상데이터와 실제 데이터 정합 기능을 갖는 플랫폼
- (AI-RAN 라이브 트위닝 플랫폼) AI-RAN 가상플랫폼과 실제 플랫폼간의 데이터를 실시간으로 주고 받으며, RAN 에서 AI 서비스를 수용할 수 있는 플랫폼을 구축하여 실증망 시스템과 가상시스템을 실증

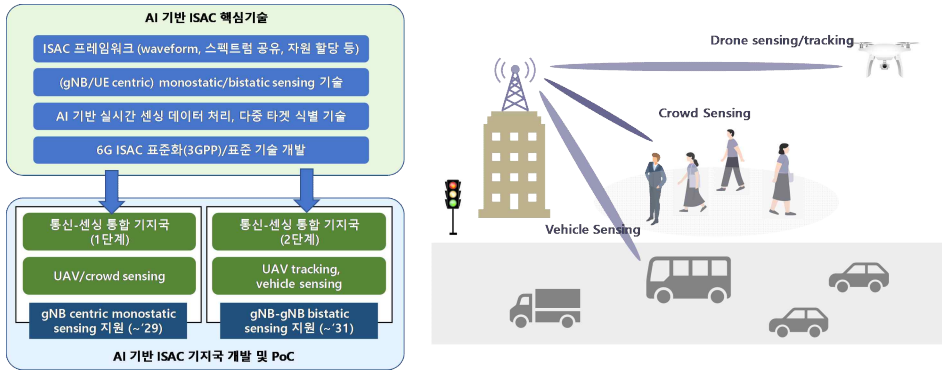


<AI-RAN 연구 플랫폼 사업 구성도>

과제명	AI 기반 센싱 통합 6G 핵심기술 개발 및 실증
기술분야	AI 기반 무선 접속망

과제 개요

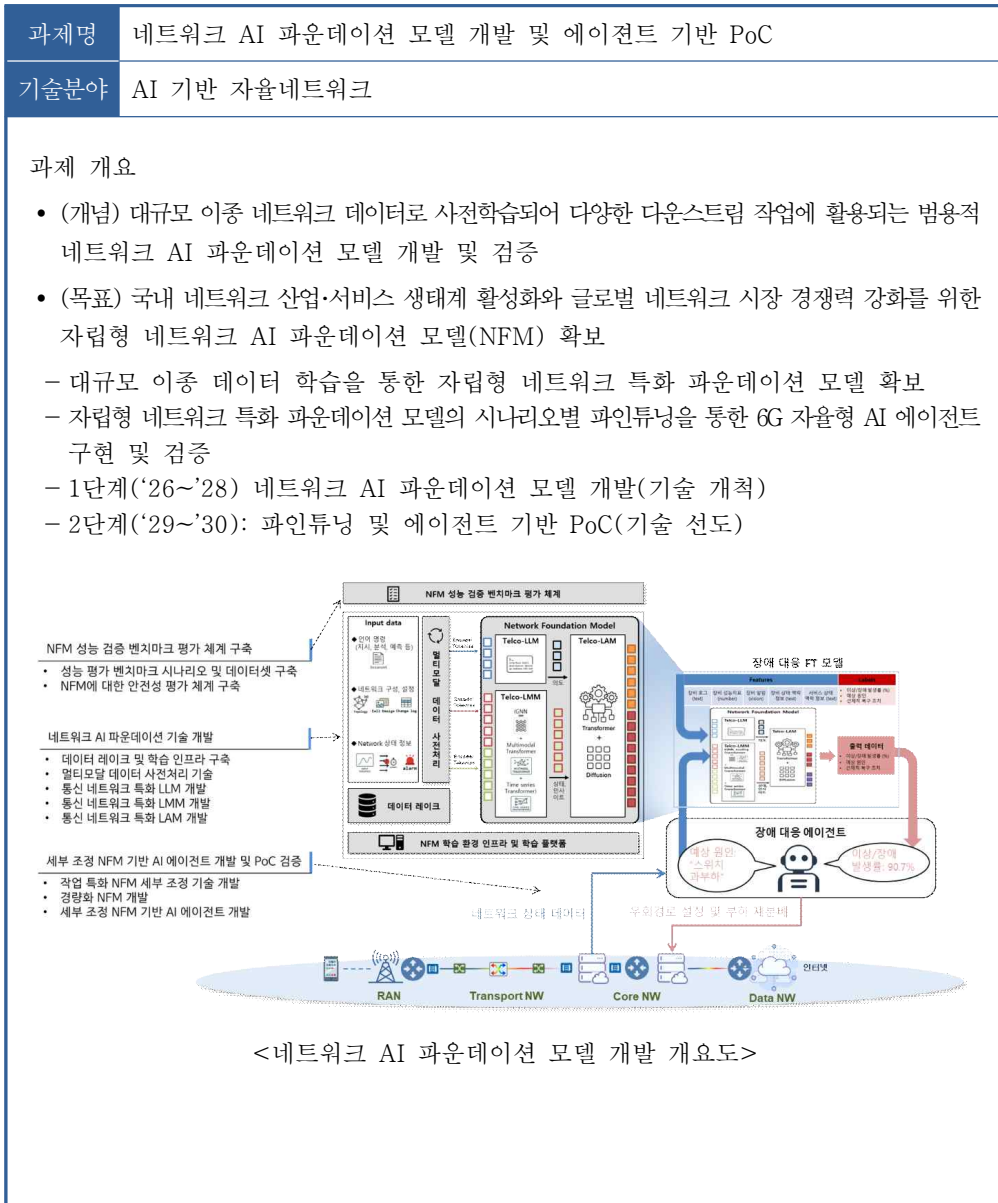
- 6G 이동통신 인프라는 센싱 기술과 결합하여 통신서비스 제공에 그치지 않고 물체 탐지와 정밀 측위와 같은 센싱 기능을 통해 Sensing as a Service (SaaS)를 실현하는 통신-센싱 인프라로 탈바꿈 하게 될 것임
- AI 기반의 통신과 센싱 통합 기지국 개발을 통해 기지국 시스템의 스펙트럼 및 에너지 효율 향상 등의 기술적 효과 기대
- 자율주행의 안정성 향상, 스마트시티의 실시간 환경 모니터링 등 공공 안전과 삶의 질 개선과 같은 사회적인 효과 기대
- 신산업 및 시장 성장을 견인하여 통신장비, 차량용 센서, 스마트시티 등 연관 산업에서 새로운 부가가치와 일자리 창출을 기대
- (‘29) 통신과 센싱 통합에 필요한 핵심기술 개발, AI 기반의 monostatic 센싱 기지국 시스템 개발 및 crowd sensing 기술 개발
- (‘31) Bistatic 센싱 기지국 시스템 개발 및 Sensing as a Service (SaaS) 시연



<AI 기반 센싱 기지국 시스템 개발 과제 개요>

4.2 AI 기반 자율네트워크 기술 로드맵

본 절에서는 AI 기반 자율네트워크와 관련하여 추 후 진행되어야 하는 과제에 대하여 기술한다. 향후 새롭게 기획되어야 하는 과제로 네트워크 AI 파운데이션 모델, RAN, 위성 관점에서 네트워크 파운데이션 모델이 각각 개발되어야 비로서 완전 자율네트워크 기술 개발이 완성될 수 있으며 관련된 기술에 대한 과제는 아래와 같이 제안될 수 있다.



과제명	RAN 파운데이션 모델 개발
기술분야	AI 기반 자율네트워크

과제 개요

- 6G 무선통신 물리계층 송수신 성능 증대, Radio Resource control/Mobility control 등 L2/L3 프로토콜 최적화 및 기지국 시스템 저전력화를 위한 RAN 파운데이션 모델 확보
- ('27~'29) RAN 파운데이션 모델 개발
- ('30 ~'31) AI-RAN R&D 플랫폼 이용 성능 검증 및 특화망 등 Real 필드에 적용하여 실증

Input (Real + Synthetic data):
Signal, Text, Metadata

↓

Joint Embedding

↓

Task-Ready Embedding

↓

Outputs for Downstream Tasks

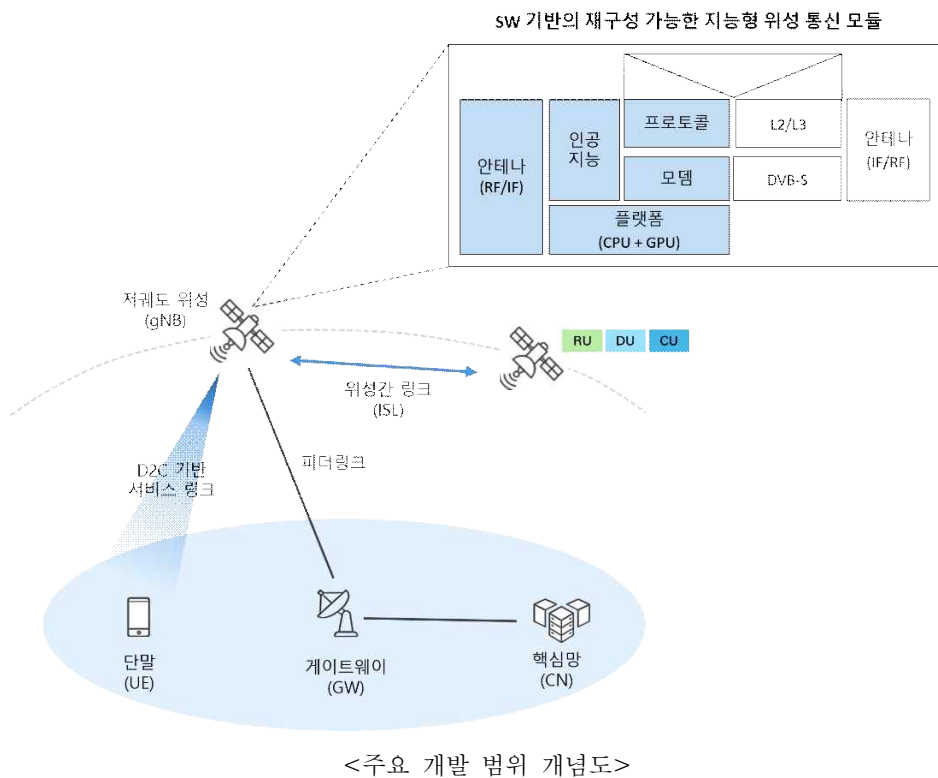
<PHY-MAC-RRC 계층 통합 Wireless Foundation Model 개념도>

과제명 AI 기반 지능형 D2C 위성통신 기술 및 3D ORAN 지능형 네트워크 기술 개발

기술분야 AI 기반 자율네트워크

과제 개요

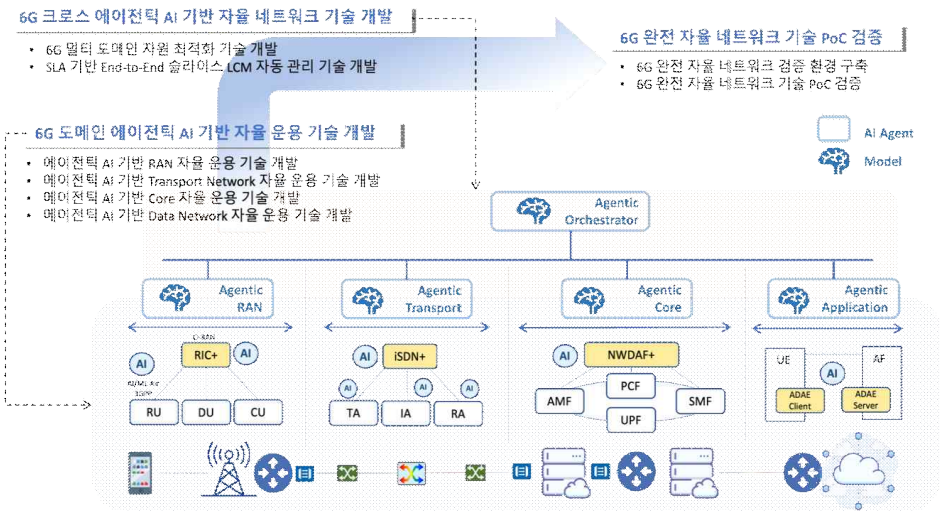
- 위성 직접통신 지원 SW 기반 저궤도 위성통신 핵심기술 확보
- 위성 직접통신 지원 지능형 저궤도 위성통신 핵심기술 확보
- 위성 직접통신 지원 저궤도 위성 지능화를 위한 R&D 환경 제공



과제명	완전 자율 네트워크 기술 개발
기술분야	AI 기반 자율네트워크

과제 개요

- 계층적 구조의 에이전틱 AI 기반 6G 완전 자율 네트워크 기술 개발 및 PoC 검증
 - 1단계('28~'30) 완전 자율 네트워크 기술 개발(기술 추적);
 - 2단계('31~'32): 고도화 개발(기술 선도)



< 완전자율네트워크 개념도 >

과제명 AI 데이터 센터 네트워크 기술 개발

기술분야 AI 서비스 네트워크

과제 개요

- 데이터센터 서버랙 내/간 AI 네트워크, 컴퓨팅 자원을 효율적으로 상호연결 및 운영하기 위한 AI 네트워크 인프라-응용 기술 개발 및 PoC 검증
 - 1단계('26~'28) 국산화 개발(기술 추격)
 - 2단계('29~'30): 고도화 개발(기술 선도)

데이터센터 네트워크 인프라 기술 선도사업

네트워크-컴퓨팅 패브릭 기술 개발

- AI 센트릭 네트워크 운영체제 기술
- AI 센트릭 클라우드 운영체제 기술
- AI 기반 실시간 자원 운영 기술

데이터센터 전송장비 기술 개발

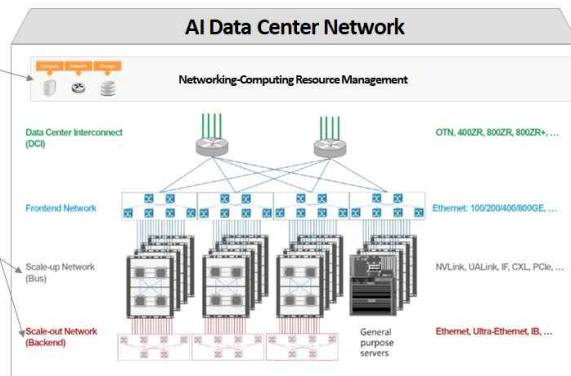
- 이더넷 기반 산업표준 네트워크 인터페이스 기술
- 이더넷/IP 기반 산업표준 전송 프로토콜 기술
- 산업표준 이더넷 네트워크 어댑터 개발 기술

데이터센터 상호연결 기술 개발

- 산업표준 기반 상호연결 인터페이스 기술
- 산업표준 확장 상호연결 프로토콜 기술
- 산업표준 기반 프로토콜 적용 광 연결 모듈 개발 기술

데이터센터 네트워크 장비 검증

- 데이터센터 네트워크 장비 검증 환경 구축 및 연구
- 데이터센터 네트워크 장비 PoC 검증



< 데이터센터 네트워크 인프라 기술 선도사업 구성도 >

4.3 AI 서비스를 위한 네트워크 기술 로드맵

본 절에서는 AI 서비스를 위한 네트워크 기술은 현재 6G 에서도 논의되고 있으며, Edge AI 네트워크를 포함하는 Physical AI 지원 네트워크 핵심 기술 및 AI-RAN대규모 실증망 구축을 통하여 이루어 질 것으로 보인다. 관련된 기술에 대한 과제는 아래와 같이 제안될 수 있다.

과제명	피지컬 AI 지원 통신/네트워크 핵심 기술 및 시스템 개발
기술분야	AI 서비스 네트워크
<p>과제 개요</p> <ul style="list-style-type: none"> • 피지컬AI 에이전트(자율주행차·로봇)를 위한 초저지연·초신뢰 전송, 예측형 RAN 제어, 에이전트 인지 코어 네트워크 및 분산 연합학습 최적화를 지원하는 통신네트워크 핵심기술 및 시스템 기술 확보 <p><주요 연구내용></p> <ul style="list-style-type: none"> • 핵심기술 개발 <ul style="list-style-type: none"> - 피지컬AI 자율주행차·로봇 특화 무선 전송 기술 연구·개발 - 예측형 RAN 품질 제어 기술 연구·개발 - 업링크 편향 트래픽 최적화 네트워크 구조 연구·개발 - 에이전트 인지(Agent-aware) 코어네트워크 제어구조 설계 및 정책 연동 기술개발 - 피지컬AI용 NAS-like 시그널링 구조 설계/개발 - 피지컬AI 특화 Open Gateway API 확장 및 연동 프레임워크 개발 • 시스템 개발 <ul style="list-style-type: none"> - 피지컬 AI를 지원하는 초저지연/초신뢰 특화 통신-네트워크 시스템 (Core-기지국- 단말) 개발 • ('27~'29) Physical AI 지원 통신/네트워크 핵심 기술 및 시스템 기술 개발 • ('30~'31) 기술/시스템 Upgrade 및 Physical AI 응용(로봇 등) 시나리오 기반 실증 	
<p>(Physical AI 지원 초저지연/초신뢰 통신-네트워크 특화망)</p>	

과제명 AI-RAN 대규모 실증 망 구축

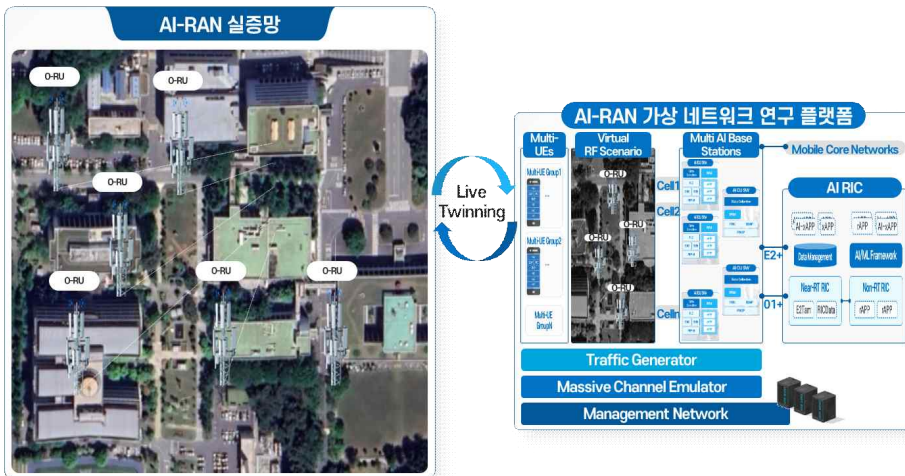
기술분야 AI 서비스 네트워크

과제 개요

- AI-RAN 대규모 실증망을 구축하여 기지국/단말 모뎀, 프로토콜 스택 등을 실제환경에서 검증 또한 트래픽 데이터, 무선 환경 데이터 등을 획득하여 AI-RAN 연구 플랫폼과 디지털 트윈으로 연동하여 각종 성능 분석 및 알고리즘/시스템 최적화 등을 수행

<주요 연구내용>

- 실증망 구축
 - 다수의 기지국 시스템(RU-DU/CU-RIC) 구축 (Core 망 포함)
 - RU-DU/CU-RIC SW 및 파라미터 최적화
 - 다수의 단말을 배치하여 실제 환경에서 시스템 운용
 - 트래픽, 무선 채널 환경 등 각종 데이터 추출
- 디지털 트윈 환경 구축
 - AI-RAN 대규모 실증망과 가상환경 연구 플랫폼을 디지털 트윈으로 연동하여 각종 성능 분석 및 알고리즘/시스템 최적화 수행
 - ('27~'29) AI-RAN 대규모 실증 망 구축
 - ('30~'31) 실증망과 가상망을 디지털 트윈으로 연동하여 각종 성능 분석 및 알고리즘/시스템 최적화



(AI-RAN 대규모 실증 망 및 디지털 트윈 개념도)

AI 기술의 급속한 발전과 초지능·초연결 사회로의 진입은 통신 네트워크의 패러다임을 근본적으로 변화시키고 있다. 데이터 수집과 학습 및 추론을 중심으로 진화하는 인공지능은 더 이상 서비스 단위의 보조 수단이 아니라, 네트워크 설계, 운영, 서비스 제공 전반에 걸쳐 중심적인 역할을 수행하는 핵심 동력으로 자리매김하고 있다. 또한 AI와 네트워크의 융합은 자율화된 지능형 통신망의 실현 뿐만 아니라, AI 응용서비스의 확산과 혁신을 견인하는 기반 인프라로서의 역할을 요구하고 있다.

이러한 시대적 전환점에서, ETRI는 그 동안 축적된 통신네트워크 기술을 토대로 하여 국내 기술 자립을 목표로 현재의 기술 중속 구조를 극복하고, 글로벌 시장에서의 지속가능한 경쟁력 내재화를 위한 핵심 기술을 국내 역량으로 확보하고, 이를 체계적으로 고도화하는 전략이 필수적이다. 특히 AI 기반의 자율 네트워크 기술, AI 서비스 최적화를 위한 지능형 인프라 기술, 그리고 데이터-모델-네트워크 간 상호 작용을 극대화하는 기반 구조에 대한 자체 원천 기술 확보가 선행되어야 한다.

동시에, 기술 개발의 국제 정합성과 시장 확장을 위해 글로벌 표준화와 국제 협력의 주도적 참여가 병행되어야 한다. 3GPP, ITU, O-RAN, ETSI 등 국제 표준 기구에서 AI 네트워크 관련 기술 주도권을 확보하고, E2E AI 네트워크 구조와 AI 협업 모델, AI 연산 자원 최적화 구조 등에 대한 선도형 레퍼런스를 제시함으로써 한국이 AI-통신 융합 시대의 글로벌 기술 리더로 자리 잡을 수 있는 토대를 마련해야 한다.

앞으로 AI와 네트워크의 융합이 만들어낼 새로운 디지털 생태계 속에서, 자율적이고 신뢰 가능한 AI 네트워크 기술 개발을 중심축으로 하여, 국가 경쟁력 강화와 산업 혁신에 기여하는 공공 R&D 기관의 역할을 선도적으로 이끌어 나가기 위해 ETRI는 국내외 산업계와 학계를 포함한 기관과의 협력적 연계를 강화하고, 기술의 실증과 사업화 가능성을 고려한 전략적 R&D를 적극 추진하고자 한다.

❖ 약어표

• 3GPP	3rd Generation Partnership Project
• 6G	Sixth Generation
• AND	Autonomous Driving Network
• AGI	Artificial General Intelligence
• AI	Artificial Intelligence
• AI-RAN	AI-Radio Access Network
• API	Application Programming Interface
• AR	Augmented Reality
• AWS	Amazon Web Services
• B2B	Business to Business
• BSS	Business Support System
• CAGR	Compound Annual Growth Rate
• CAPEX	Capital Expenditure
• CNN	Convolutional Neural Network
• COTS	Commercial Off-The Shelf
• CPU	Central Processing Unit
• CSI	Channel State Information
• CU	Central Unit
• DL	Downlink
• DNN	Deep Neural Network
• DU	Distributed Unit
• ENI	Experiential Networked Intelligence
• ETSI	European Telecommunications Standards Institute
• FG-AINN	Focus Group on AI-Native Networks
• GenAI	Generative Artificial Intelligence
• GEO	Geosynchronous Earth Orbit
• GPU	Graphics Processing Unit
• IEEE	Institute of Electrical and Electronics Engineers

- IP Internet Protocol
- ISAC Integrated Sensing and Communication
- ISG Industry Specification Groups
- ISL Inter-Satellite Link
- ITU-R International Telecommunication Union - Radiocommunication Sector
- ITU-R WP5D ITU-R Working Party 5D
- KPI Key Performance Indicator
- LEO Low Earth Orbit
- LoRA Low-Rank Adaptation
- LLM Large Language Model
- LSTM Long Short-Term Memory
- MAC Medium Access Control
- MCS Modulation and Coding Scheme
- MDAF Management Data Analytics Function
- MIMO Multiple Input Multiple Output
- ML Machine Learning
- MLP Multi-Layer Perception
- MMSE Minimum Mean Square Error
- MAT Multiple Access Techniques
- MU-MIMO Multi-User MIMO
- NFM Network Function Model
- nGRG Next Generation Research Group
- NIC Network Interface Card
- NPU Neural Processing Unit
- NR New Radio
- NTN Non-Terrestrial Network
- NWDAF Network Data Analytics Function
- OAM Operation, Administration, and Maintenance
- OBP On-Board Processor

- OCG AI Operational Co-ordination Group on Artificial Intelligence
- O-CU O-RAN Central Unit
- O-DU O-RAN Distributed Unit
- OPEX Operating Expense
- O-RU O-RAN Radio Unit
- O-RAN Open RAN
- OS Operating System
- OSI Open Systems Interconnection
- OSS Operations Support System
- PEFT Parameters-Efficient Fine-Tune
- PHY PHYsical layer
- PoC Proof of Concept
- PRB Physical Resource Block
- QoE Quality of Experience
- QoS Quality of Service
- RAG Retrieval-Augmented Generation
- RAN Radio Access Network
- RIC RAN Intelligent Controller
- RNN Recurrent Neural Network
- RSRP Reference Signal Received Power
- RSRQ Reference Signal Received Quality
- RT Real Time
- RTL Register Transfer Level
- RU Radio Unit
- SBA Service-Based Architecture
- SLA Service Level Agreement
- SMO Service Management and Orchestration
- SNR Signal-to-Noise Ratio
- SRv6 Segment Routing IPv6

- SSB Synchronization Signal Block
- SW Software
- TN Terrestrial Network
- TR Technical Report
- UAV Unmanned Aerial Vehicle
- UE User Equipment
- UP User Plane
- UPF User Plane Function
- URLLC Ultra-Reliable and Low-Latency Communication
- vRAN virtualized RAN
- VR Virtual Reality
- XR eXtreme Reality
- ZSM Zero-touch network and Service Management

❖ 참고 문헌

- [1] ITU-T, "Architectural framework for machine learning in future networks including IMT-2020," Y.3172, 2019.
- [2] ITU-T, "Architectural framework for artificial intelligence-based network automation for resource and fault management in future networks including IMT-2020," Y.3177, 2021.
- [3] ITU-T, "Focus Group on Artificial Intelligence Native for Telecommunication Networks (FG AINN)," 2024. [Online]. Available: <https://www.itu.int/en/ITU-T/focusgroups/ainn/Pages/default.aspx>.
- [4] ITU-T, "Potential requirements and methodology for deploying and assessing generative AI models in telecom networks," 2025.
- [5] ETSI, "Technical committee (TC) securing artificial intelligence (SAI)," [Online]. Available: <https://www.etsi.org/committee/technical-committee-tc-securing-artificial-intelligence-sai>. [Accessed 23 July 2025].
- [6] ETSI, "Experiential Networked Intelligence (ENI)," [Online]. Available: <https://www.etsi.org/technologies/experiential-networked-intelligence>. [Accessed 23 July 2025].
- [7] ETSI, "Zero touch network & Service Management (ZSM)," [Online]. Available: <https://www.etsi.org/technologies/zero-touch-network-service-management>. [Accessed 23 July 2025].
- [8] ETSI, "Industry specification group (ISG) integrated sensing and communications (ISAC)," [Online]. Available: <https://www.etsi.org/committee/2295-isac>. [Accessed 23 July 2025].
- [9] ETSI, "Industry specification group (ISG) multiple access techniques (MAT)," [Online]. Available: <https://www.etsi.org/committee/2485-mat>. [Accessed 23 July 2025].
- [10] 3GPP, "Study on Core Network Enhanced Support for Artificial Intelligence (AI)/Machine Learning (ML)_Ph2," TR 23.700-04, 2025.
- [11] 3GPP, "Study on Architecture for 6G System," TR 28.801-01, 2025.
- [12] 3GPP, "Study on Artificial Intelligence (AI)/Machine Learning (ML) for NG-RAN Phase 3," RP-251868, 2025.
- [13] 3GPP, "Study on 6G Scenarios and Requirements," TR 38.914, 2025.
- [14] O-R. Alliance, "O-RAN Architecture Description v10.0.," July 2024. [Online]. Available: <https://www.o-ran.org/specifications>.
- [15] A. Maeder, M. Iwamura and e. al., "Towards AI-Native Open RAN for 6G," IEEE Network, vol. 37, no. 4, p. 30-38, 2023.

- [16] O.-R. Alliance, "E2 Service Model for Lower Layer Control (E2SM-LLC), Release 3," [Online]. Available: <https://www.o-ran.org/specifications>.
- [17] O.-R. nGRG, "GenAI in Network, Research Report RR-2025-02,," Feb. 2025.
- [18] "AI-RAN Alliance," [Online]. Available: <https://ai-ran.org>. [Accessed 23 July 2025].
- [19] Gartner, "Forecast: Communications Technology and Services, Worldwide," 2024.04.
- [20] Gartner, "Forecast: PCs, Tablets & Mobile Phones, Worldwide," 2025.06.
- [21] E. 기술전략연구본부, "6G 통신 시장 전망," 2025.07.
- [22] Gartner, "Forecast: Communications Services, Worldwide," 2025.06.
- [23] E. 기술전략연구본부, "AI-RAN 기지국 네트워크 및 Edge AI 활용 서비스 시장 전망," 2025.09.
- [24] O.-R. A. nGRG, "Research Report on Generative AI Use Cases and Requirements on 6G Network," 2025.
- [25] P. Research, "Edge AI Market Size, Share, and Trends 2025 to 2034," 2025.08.
- [26] J. Zhang, W. LU, C. Xing, N. Zhao, N. AL-DHAHIR, G. K. Karagiannidis and X. Yang, "Intelligent integrated sensing and communication: a survey," in Science China, 2025.
- [27] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond," Recommendation ITU-R M.2160-0, 2023.
- [28] N. H. Chu, D. N. Nguyen, D. T. Hoang, Q.-V. Pham, K. T. Phan, W.-J. Hwang 그리고 E. Dutkiewicz, "AI-enabled mm-Waveform Configuration for Autonomous Vehicles with Integrated Communication and Sensing," %1 arxiv, 2022.
- [29] A. Mourad, "Use cases for ISAC in 6G: A Look Ahead," May 2025. [Online]. Available: <https://www.interdigital.com/post/use-cases-for-isac-in-6g-a-look-ahead>.
- [30] A. Mourad, "Exploring How ISAC Paves the Way for Smart Factory Implementation," August 2024. [Online]. Available: <https://www.thefastmode.com/expert-opinion/36780-exploring-how-isac-paves-the-way-for-smart-factory-implementation#:~:text=network%20functions>



AI 네트워크 전략 보고서

작 성 자 :

입체통신연구소	백 용 순
이동통신연구본부	김 일 규
	고 영 조
	나 지 현
	배 정 숙
	정 희 상
네트워크본부	정 태 식
	고 남 석
	김 태 연
위성통신연구본부	이 문 식
	이 준 환
기획팀	이 숙 진
기술전략연구본부	송 영 근

발 행 처 : 한국전자통신연구원 입체통신연구소

발 행 일 : 2025년 9월 22일

ETRI 한국전자통신연구원
입체통신연구소

※ 주의 : 본서의 일부 또는 전부를 무단으로 전재하거나 복사하는 것은
저작권 및 출판권을 침해하게 되오니 유의하시기 바랍니다.