

Samsung Exynos Edge AI

(Seed AI Challenger)

2023. 10. 27

Agenda

- 개회사
- 프로그램 개요
- ENN SDK 소개
- Exynos Eco System 포털 소개 / 리소스 및 지원 방안
- Best Practices Sample 소개
- 프로그램 주요 일정
- Q&A

프로그램 개요

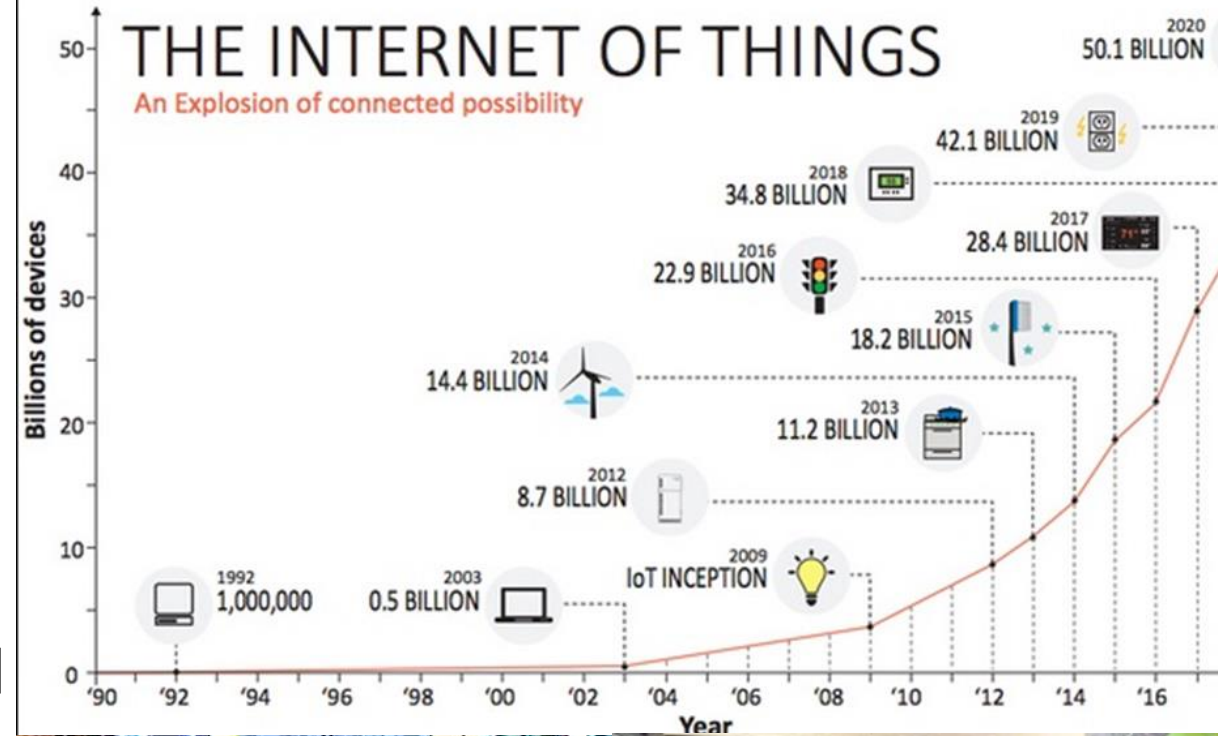
SeedAI (Samsung Exynos Edge AI) 배경

● IoT 기기의 급격한 성장

- 50 billion 이상의 기기 연결 예상
- 다양한 응용 프로그램에 필수적인 방대한 데이터 생성
- 계산 능력의 제약

● Machine Learning과 Edge Computing의 등장

- 데이터를 분석하고 실시간 응답 필요성
- 계산 제약으로 인한 데이터의 클라우드 오프로드의 필요성
- 클라우드 전송의 문제점



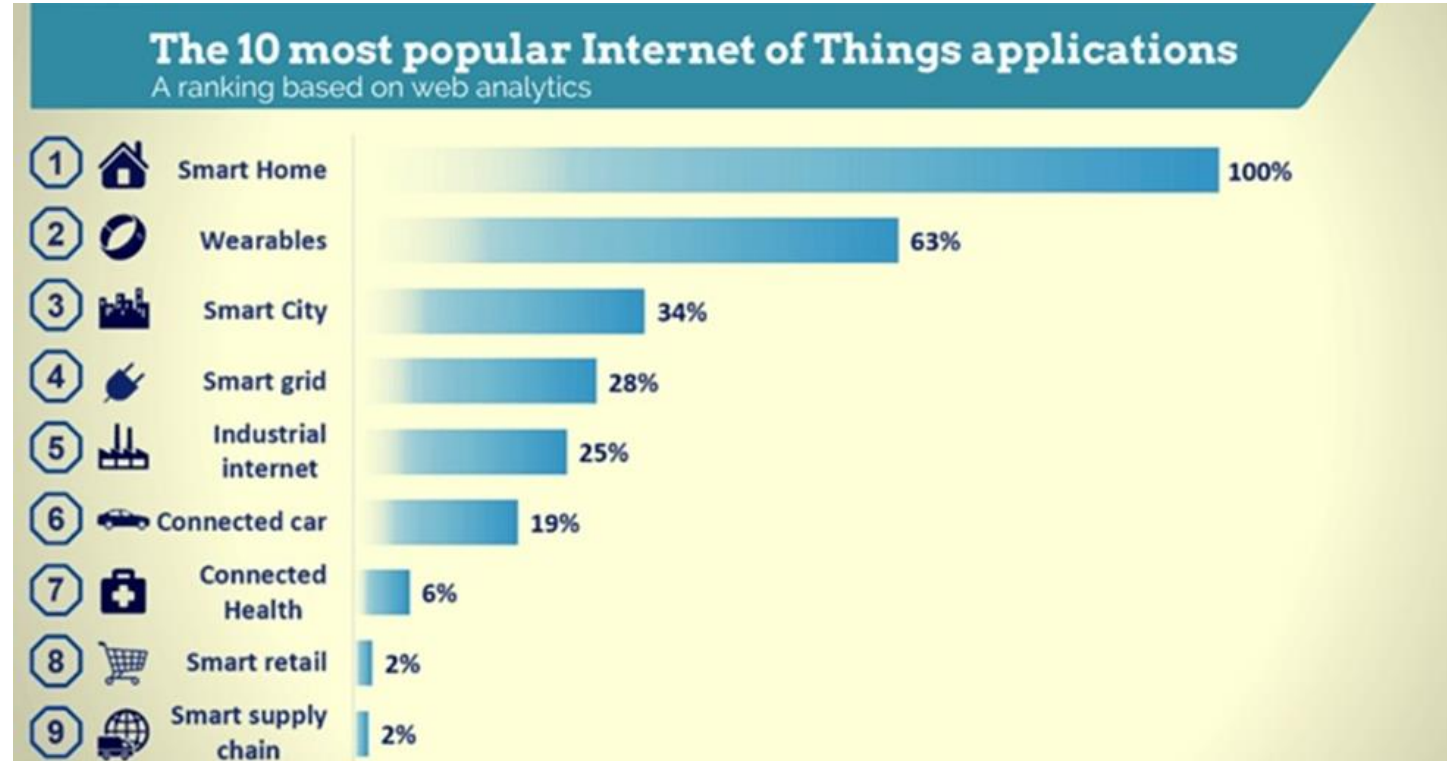
Edge Computing Solution

● Edge Computing 필요성

- 데이터 출처와 가까운 곳에서의 계산 처리
- 클라우드로의 원시 데이터 전송 최소화
- 대기 시간 감소, 사용자 개인정보 보호 등의 이점 제공

● Edge에서의 Machine Learning

- 데이터 관리 및 처리 비용 감소
- 데이터 전송 볼륨 최소화
- 분산 인공지능 제공 및 데이터 볼륨, 실시간 데이터 처리 개선
- 실시간 컴퓨터 비전 작업에서 증강 현실 경험까지 다양한 응용 프로그램 탐구



출처(<https://arxiv.org/pdf/1908.00080.pdf>)

Edge AI Computing 활용

● AI in Consumer Edge Devices

- 엣지 디바이스에서의 AI 계산의 장점 (데이터 보안, 속도 향상, 연결성 감소 등)
- 소비자 기기에서의 엣지 AI 사용 사례 (스마트폰, 웨어러블 등)

● Leveraging AI on Enterprise Edge Devices

- 기업에서 엣지 AI의 사용 사례 (로봇, 센서, 동영상 분석 등)
- 기업 시장의 엣지 AI 도입 이점 및 도전 과제

● AI and Connectivity

- AI와 연결성 간의 상호 작용 소개
- 데이터 전송, 처리 위치, 및 연결성의 중요성
- 하이브리드 AI 모델 (로컬과 클라우드 간의 데이터 전송)

● AI at the Telecom Edge

- 통신 엣지에서의 AI의 역할
- 5G와 엣지 컴퓨팅의 통합과 이점
- 사용 사례: 스마트 시티, 자동차 통신 등

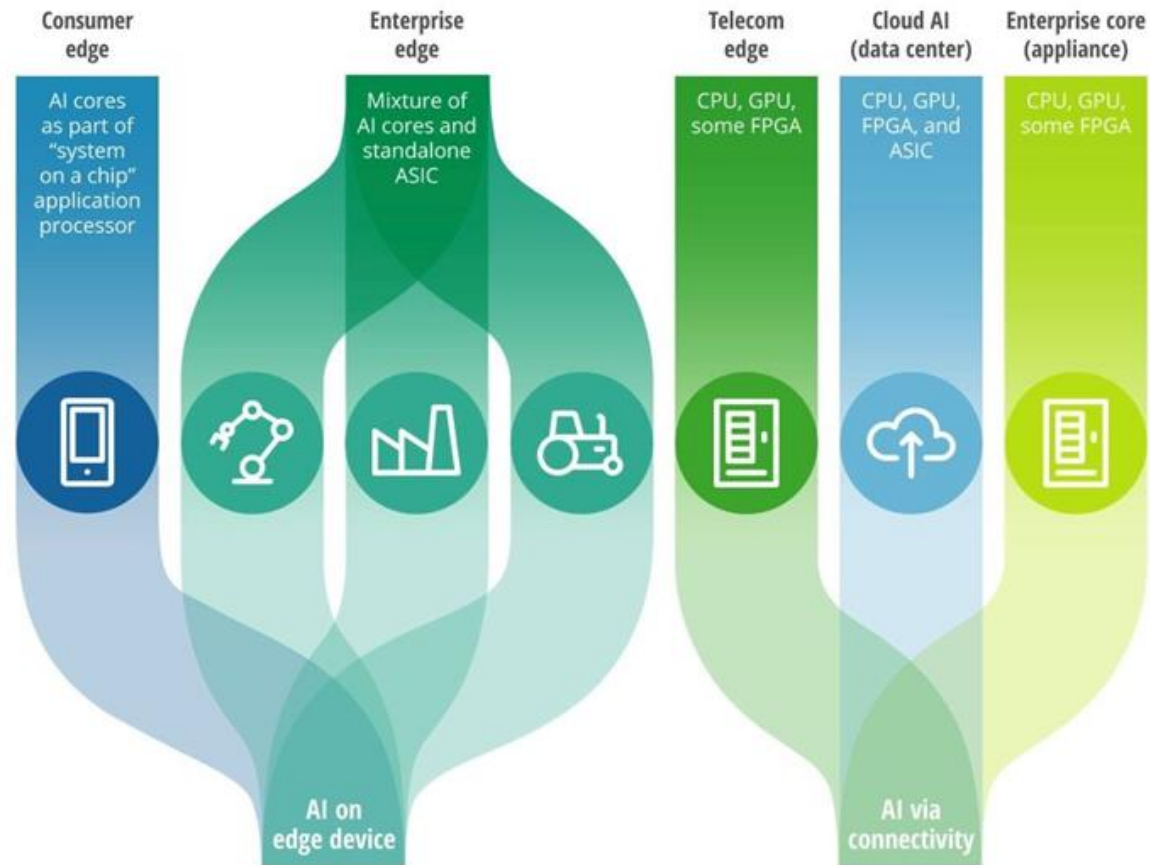
● Integrating Cloud and AI

- 클라우드 AI와 엣지 AI의 시너지
- 클라우드 AI를 이용한 빅 데이터 분석

● AI in the Enterprise Core

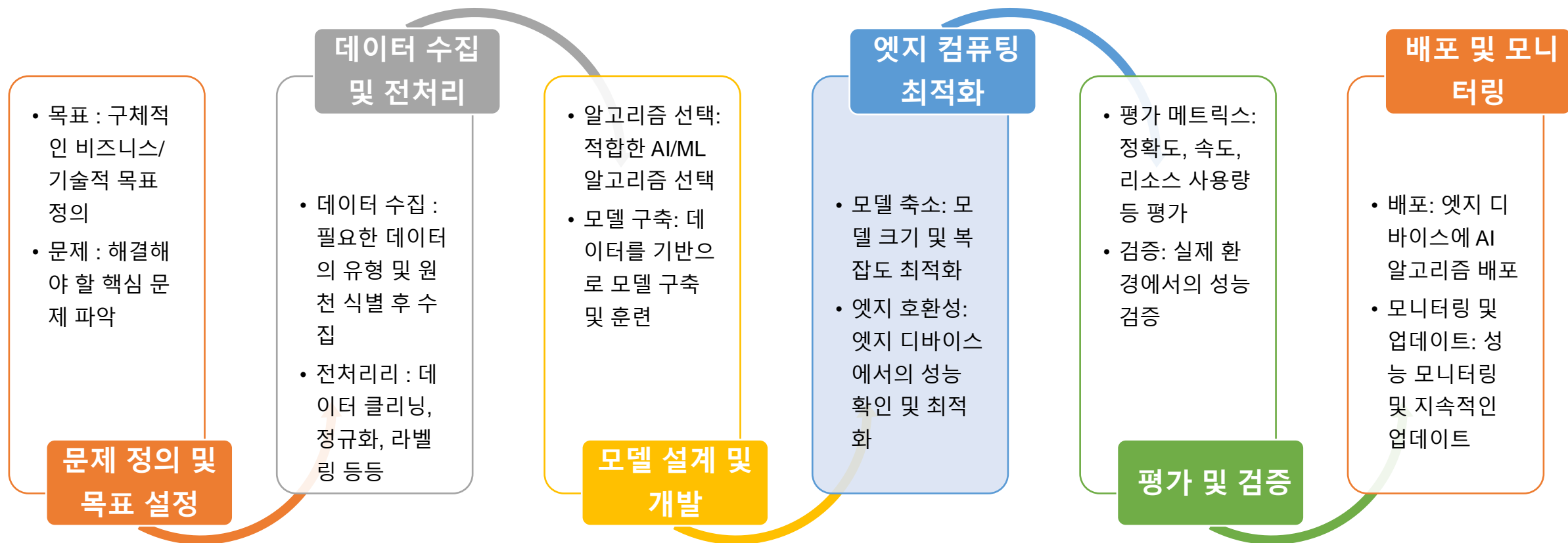
- 엔터프라이즈 핵심에서의 AI 작업(데이터 분석, 예측 모델링 등)
- 엔터프라이즈 데이터 센터의 역할 (사용 사례: 실시간 분석, 비즈니스 인텔리전스 등)

AI computing can occur at different physical locations



출처(<https://www.embedded.com/why-edge-ai-is-a-no-brainer/>)

Edge AI 알고리즘 개발



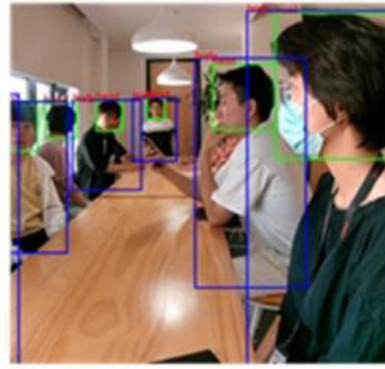
AI Algorithm Example



Face Recognition



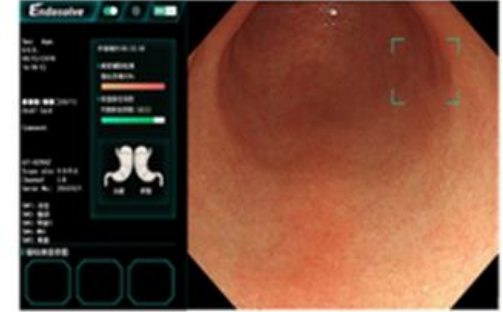
Face Detection



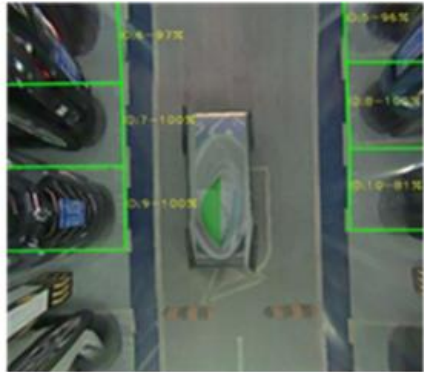
Auto Framing



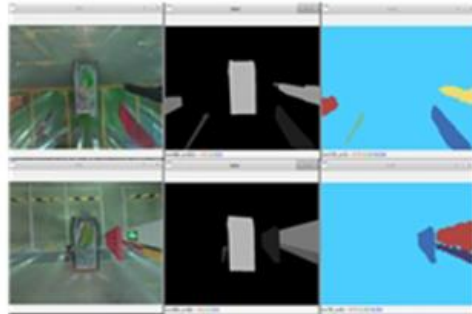
Helmet Detection



Lesion Segmentation



Parking Slot Location



Obstacle Detection



License Plate Recognition



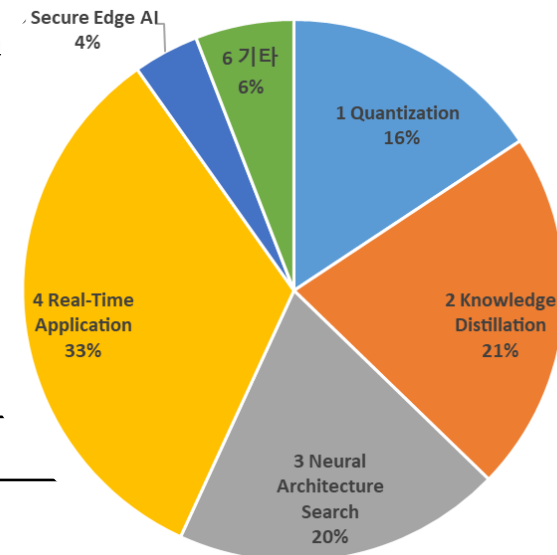
Instance Segmentation



Fall Detection

SeedAI Challenger 관심 분야

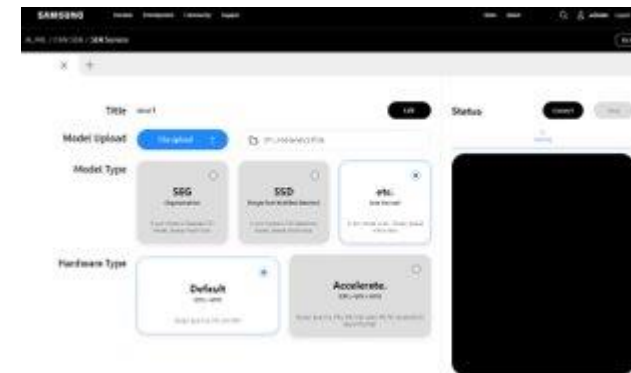
#	Topic	Description	관심 분야/주제
1	Quantization	가중치 값을 정수 또는 최소한의 비트 수를 사용하여 표현하는 방법을 연구	1. Exynos Device에 최적화된 모델 개발 2. 이미지 생성 모델 (Diffusion Model), 초거대 언어모델 (LLM), Structural Re-parameterization based 모델
2	Knowledge Distillation	큰 모델의 지식을 작은 모델로 전달하여 모델의 크기를 줄이는 방법을 연구	1. Vision Transformer model 개발. Self-Knowledge Distillation, Ensemble Distillation 활용
3	Neural Architecture Search	효율적인 모델 아키텍처를 찾기 위한 방법을 연구	1. Edge device 기반, 생체이상 감지 모델 개발 및 경량화 2. Attention 기반의 Transformer 모델 가속기 설계
4	Real-Time Application	모바일 기기에서 실행되는 Edge AI를 활용하여 실시간으로 데이터를 처리하고 분석하는 Application(Computer vision, Voice recognition)을 개발하는 연구	1. AR 앱에서의 localization 및 그래픽스 렌더링 연구 2. Real time application 개발 3. 초저해상도에서 인간의 행동 인식 연구 4. Diffusion 등 이미지 생성모델
5	Secure Edge AI	모바일 기기에서 실행되는 Edge AI가 보안 취약점을 가지고 있을 수 있기 때문에, 보안 기술을 개발하여 Edge AI의 보안성을 향상시키는 연구	1. Secure Edge AI 2. Low-level Optimization
6	기타	상기 분야 외 AI/ML 관련 연구 주제	1. AI/ML 기능을 지닌 서비스 개발



SeedAI Challenger 개요

● 지원 사항

- Samsung Exynos Reference Device 제공
- Exynos Eco System Portal을 통한 기술 지원
- ENN SDK Service 접근 권한
- AI Sample Application 및 자료 제공
- 기술 문서 제공 및 Q&A
- User Project 게시판을 통한 기술 교류의 장 제공



● 기술 습득 및 경험

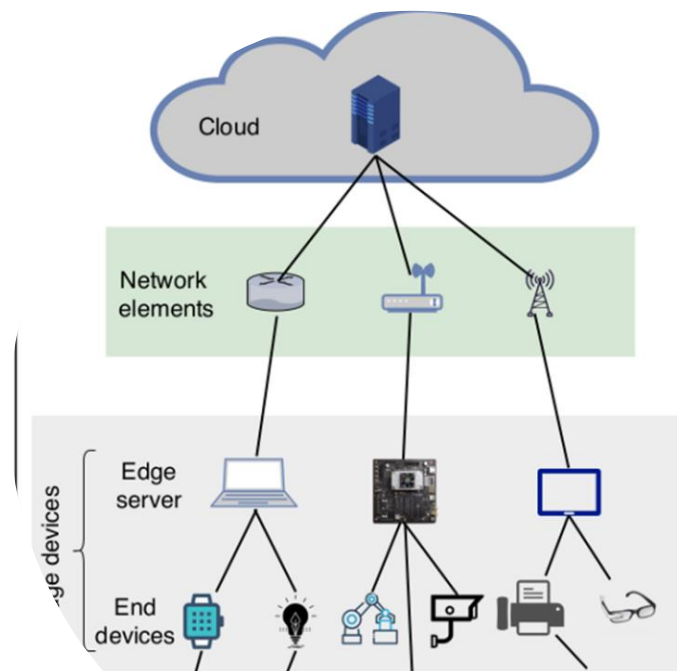
- AI 및 엣지 컴퓨팅 기술 습득: Samsung Exynos Edge AI와 관련된 기술적 지식 및 스킬 습득.
- 실전 경험: 현장에서 사용되는 기술에 직접적으로 접하고 실제 문제를 해결하는 경험.

● 네트워킹 및 멘토링

- 전문가 네트워킹: 삼성 및 기업의 엔지니어, 개발자, 그리고 전문가와의 네트워킹.
- 멘토링: AI 전문가들로부터 기술적인 지도 및 조언을 받을 수 있는 기회.

● 혁신적인 아이디어 실현

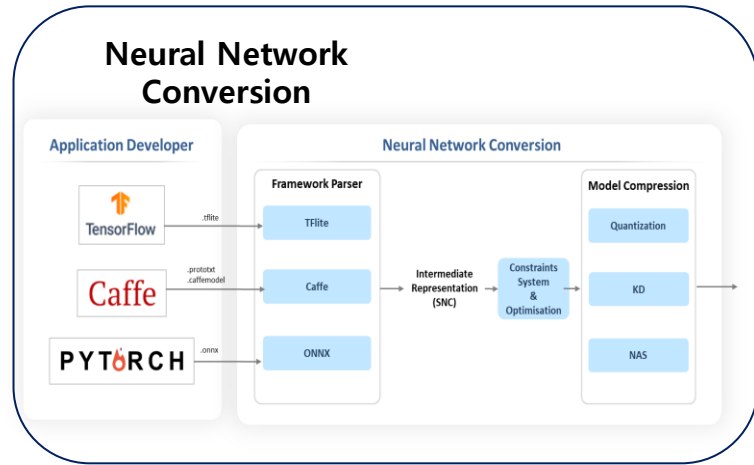
- 아이디어 실현: 창의적이고 혁신적인 아이디어를 실제 프로토타입 또는 솔루션으로 구현.
- 피드백 및 지도: 아이디어 및 프로젝트에 대한 전문가로부터의 피드백.



ENN SDK 소개

Exynos Neural Network Software Development Kit (ENN SDK)

- Samsung develops and provides own SDK for major production products.



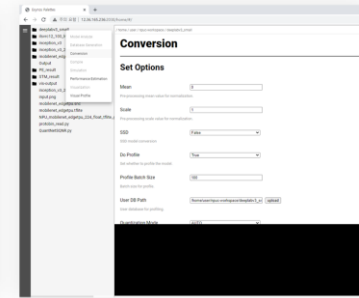
UX/Convenience

```
Enntools
Usage: enntools [OPTIONS] COMMAND [ARGS]...

Options:
  -h, --help Show this message and exit.

Commands:
  analysis      imodel analysis
  compile       imodel compile
  conversion    imodel conversion
  estimation    imodel performance estimator
  generation    idatabase generation
  info          information for workspace
  init          init for workspace
  install-apt   install apt dependent packages
  install-dpdev install dpdev dependent packages
  package-info  package information
  profiling     imodel profiling
  simulation    imodel simulation
  start-gui     start Enntools GUI
```

Enntools-CUI



Enntools-GUI

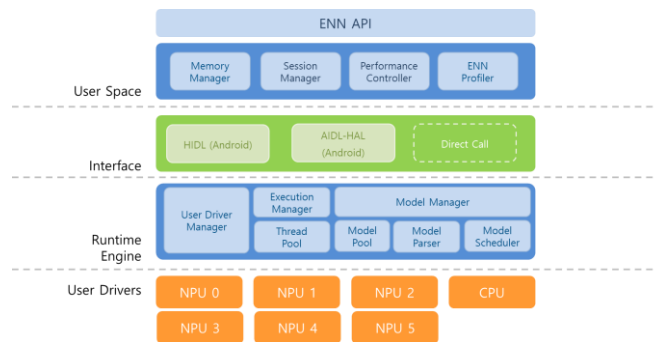
ENN
Tools

ENN
SDK

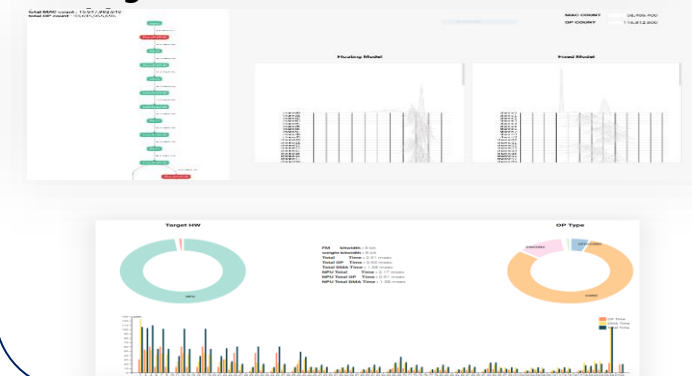
AI
Stack

Test
Suite

Scalable AI core

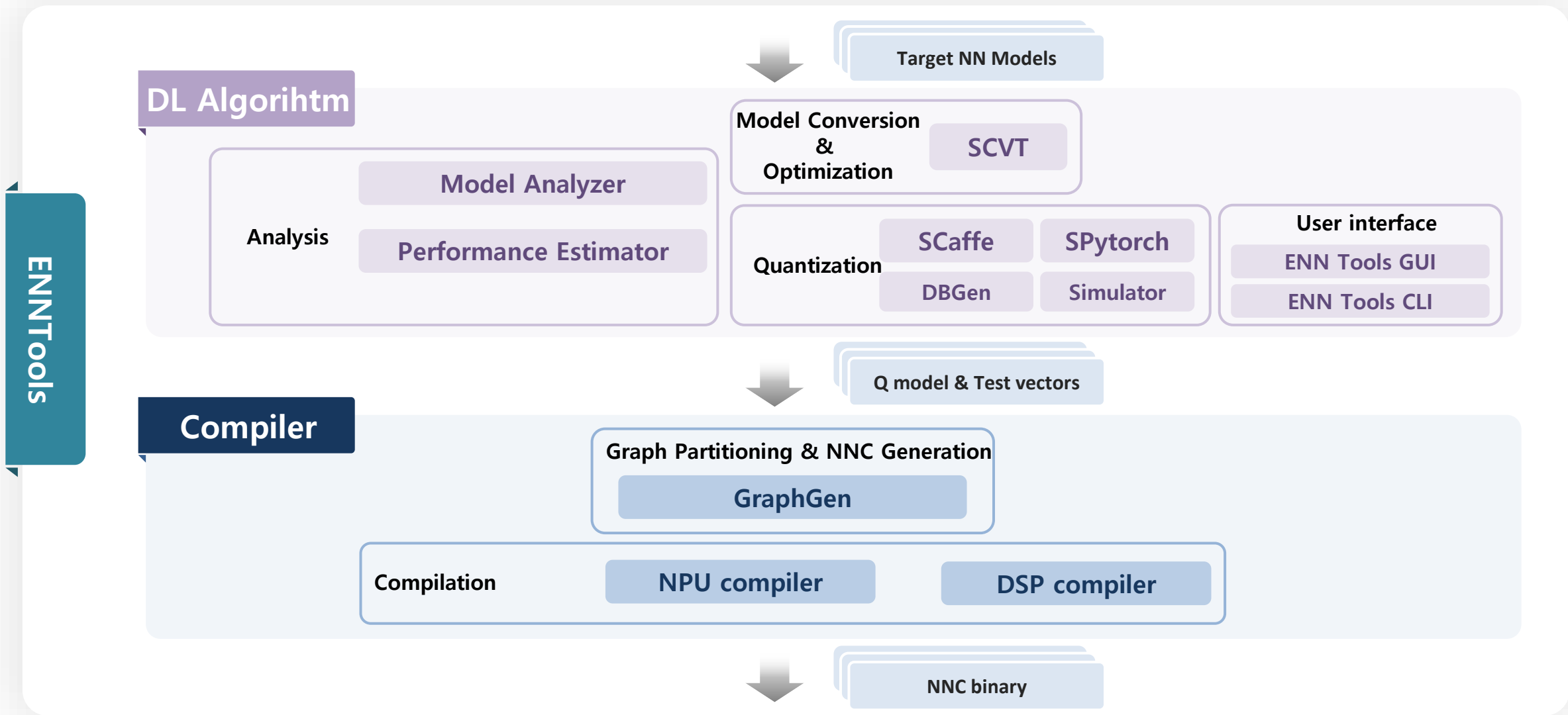


Debug-



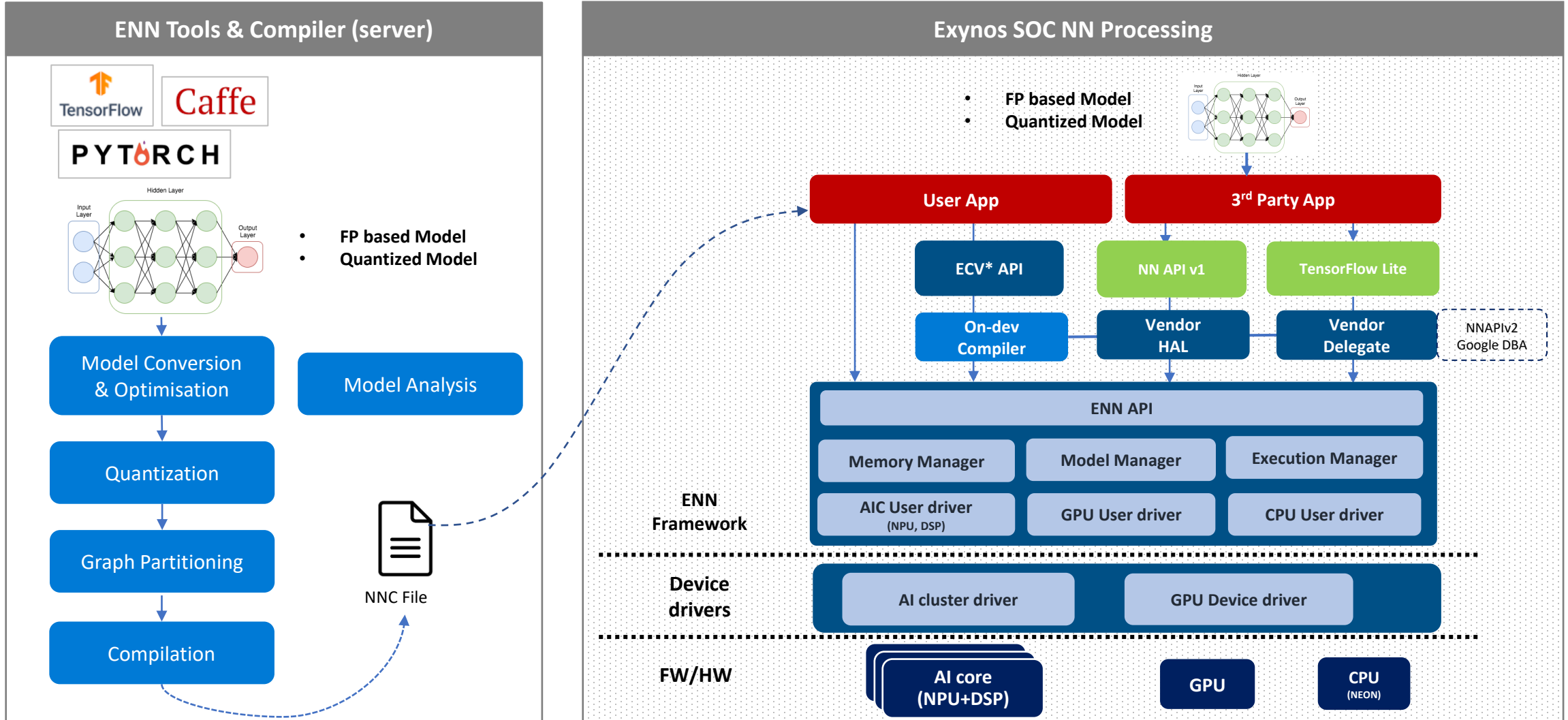
Exynos Neural Network Tools

● Overview & Flow



AI Stack

Scalable AI core Configuration and SW stack support

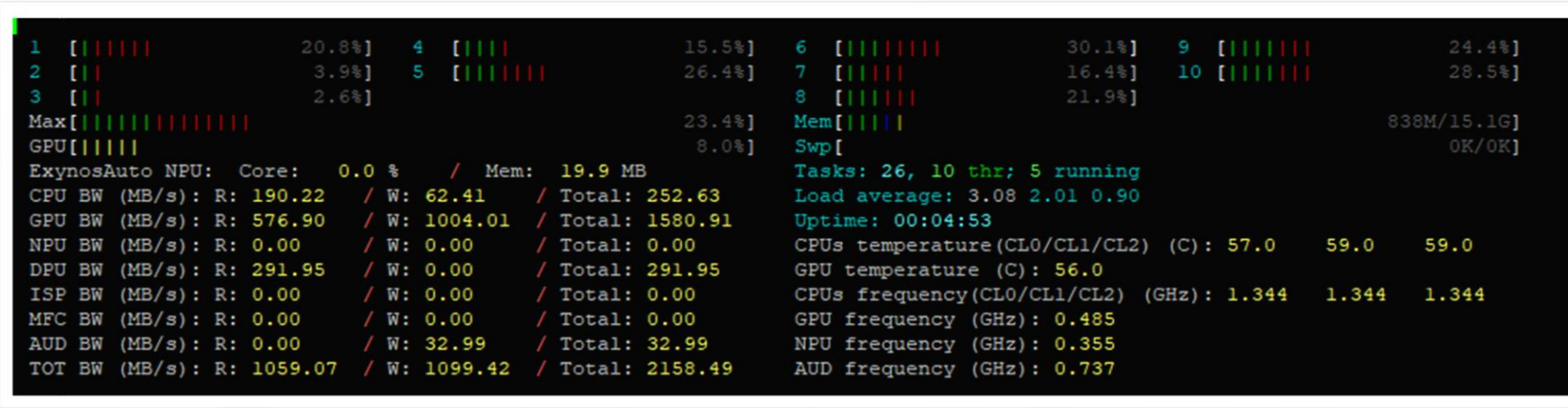


* ECV : Exynos Compute Vision

Test Suite

- Measure the performance of AI stack

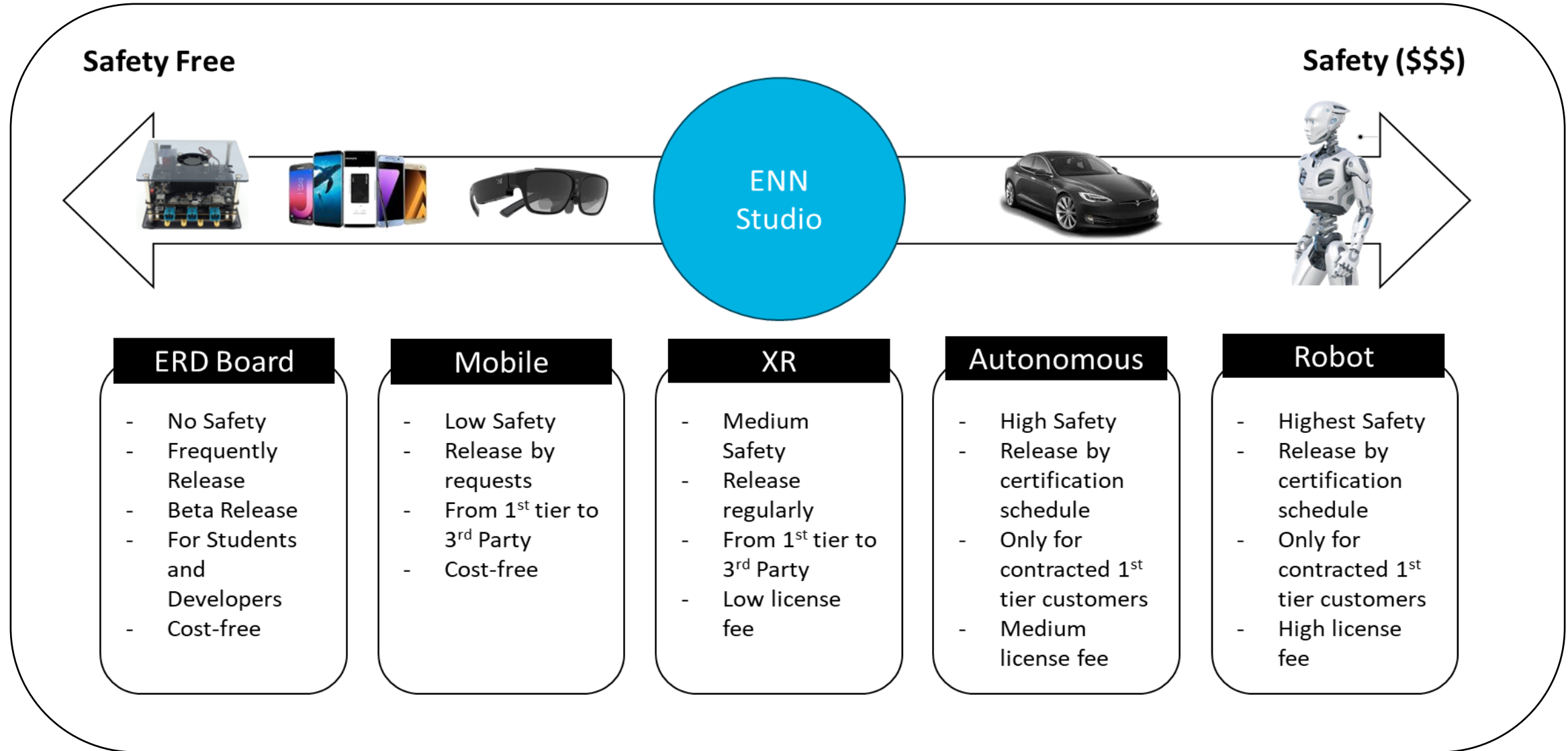
Monitor



Profiler



ENN Studio (On-premise)

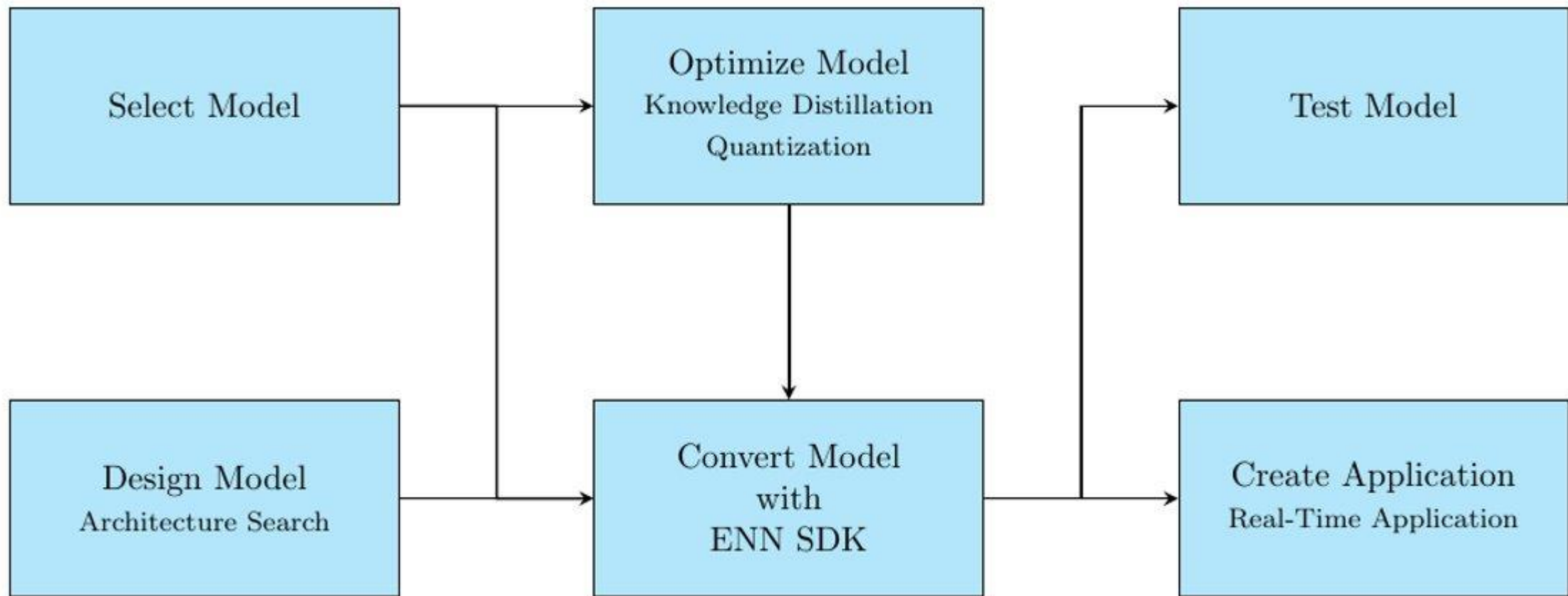


Exynos Eco System 포털 소개 / 리소스 및 지원 방안

Best Practices Sample 소개

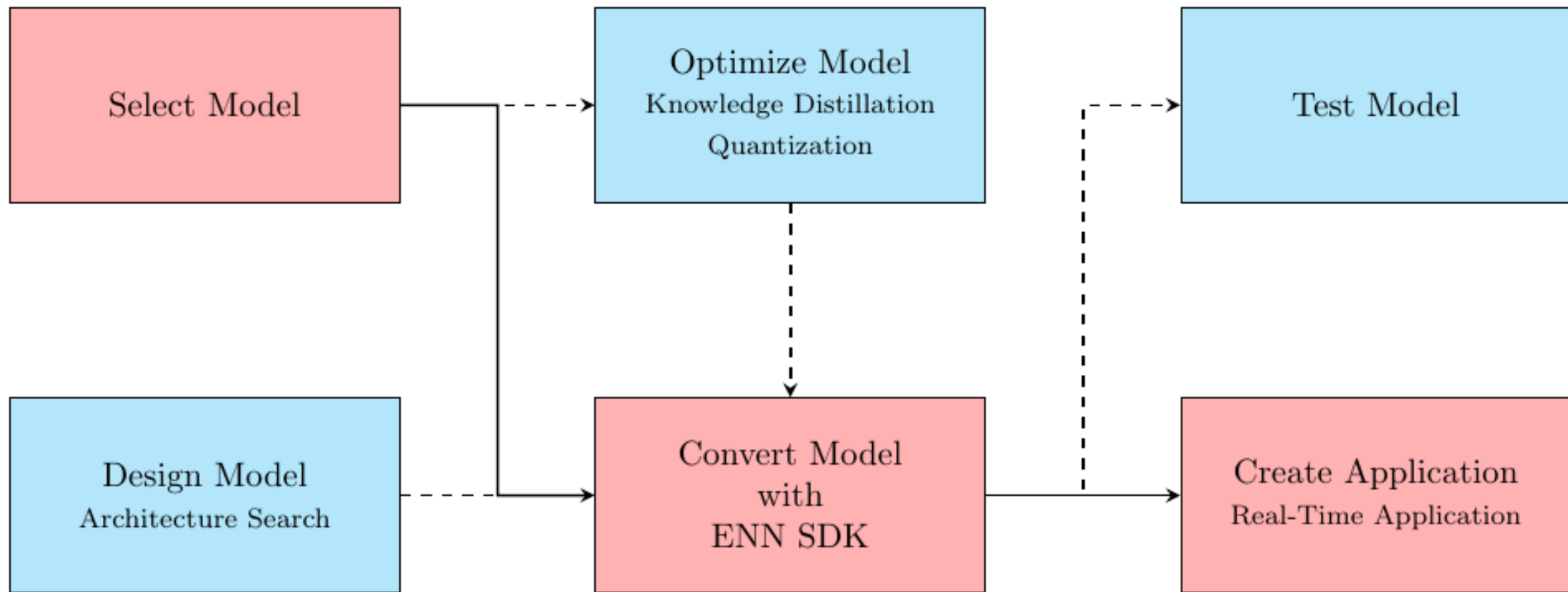
Best Practice Sample 소개 - HelpCrossing

- 예시 프로젝트 진행 과정



Best Practice Sample 소개 - HelpCrossing

- 예시 프로젝트 진행 과정



- Real-Time 건물목 차량감지 어플리케이션 제작

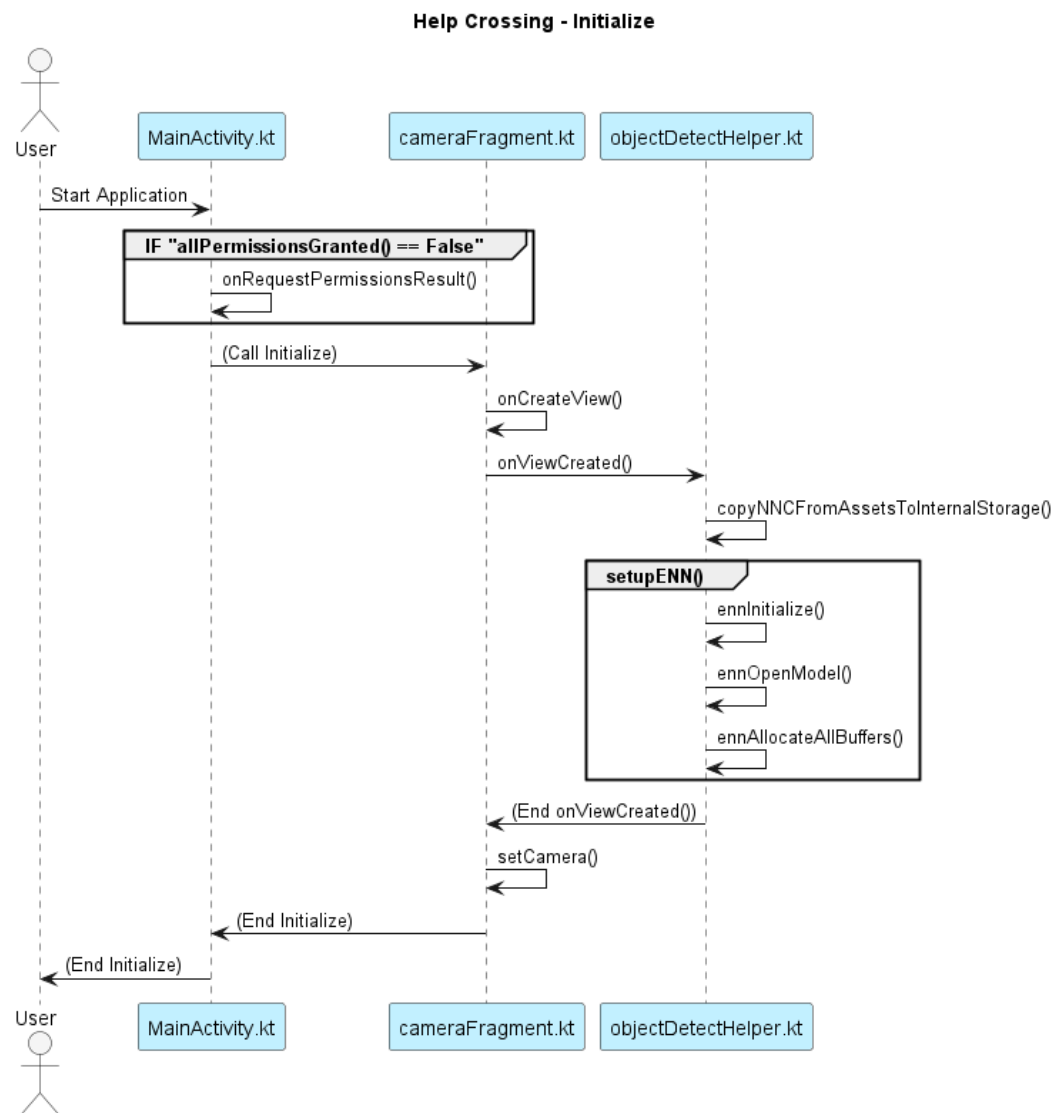
Best Practice Sample 소개 - HelpCrossing

● 모델 선정

- Support Matrix 확인
- Supported Operator 및 Hardware Features에 적합한 모델 선정
- 해당 프로젝트에서는 공개되어있는 ML Commons의 NNC모델 사용
- [mobile_models/v2_0/Samsung/od.nnc at main · mlcommons/mobile_models · GitHub](https://github.com/mlcommons/mobile_models/tree/main/v2_0/Samsung/od.nnc)

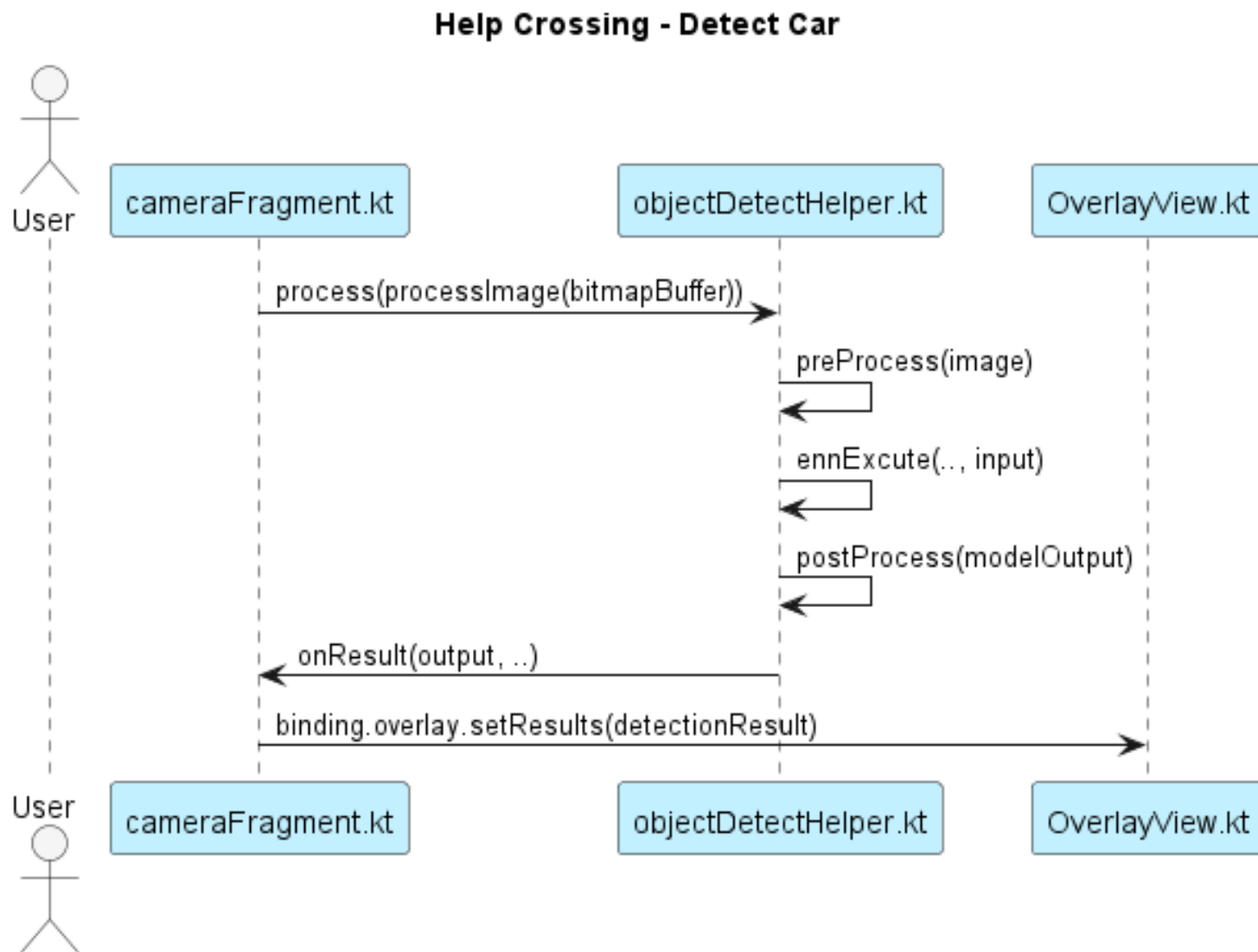
Best Practice Sample 소개 - HelpCrossing

● 어플리케이션의 동작 과정 소개



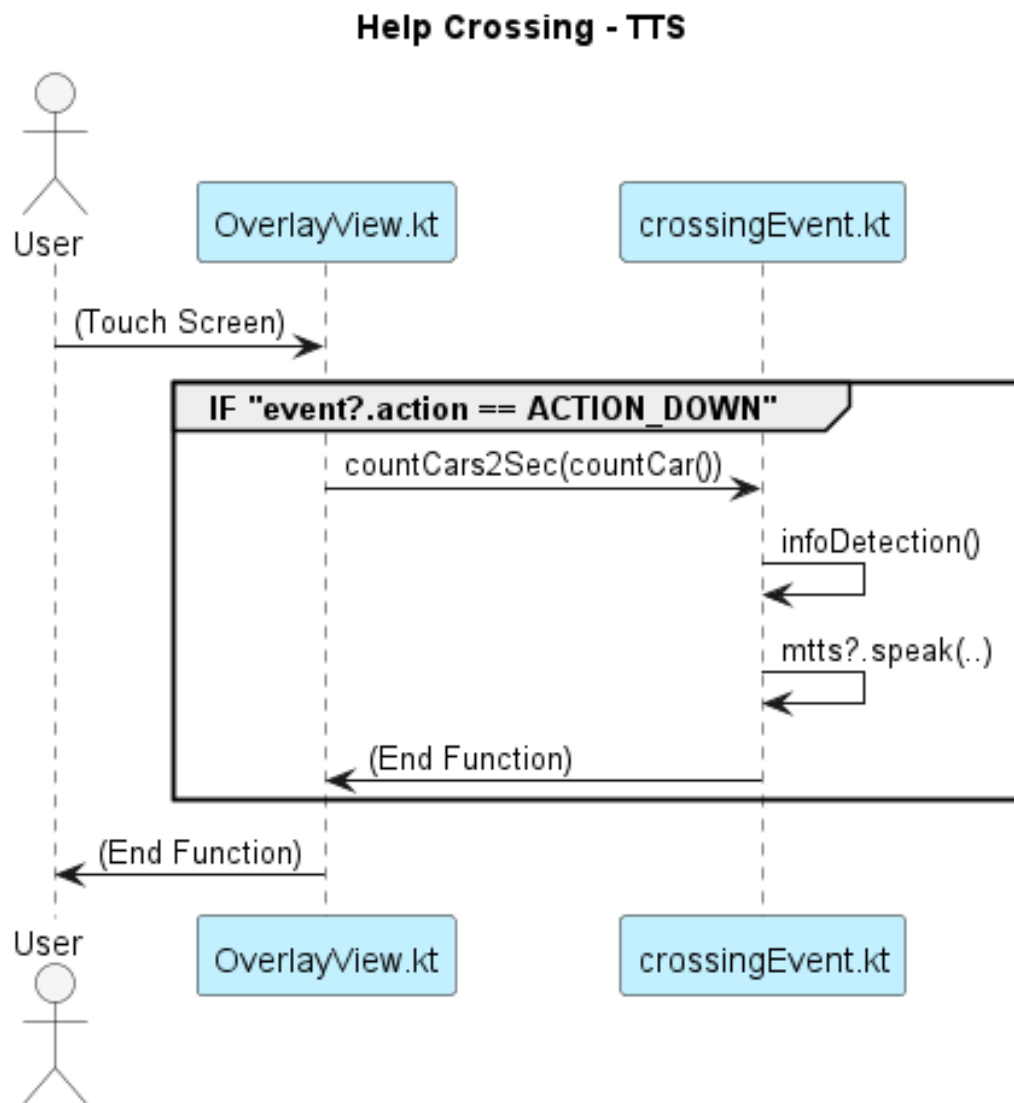
Best Practice Sample 소개 - HelpCrossing

- 어플리케이션의 동작 과정 소개



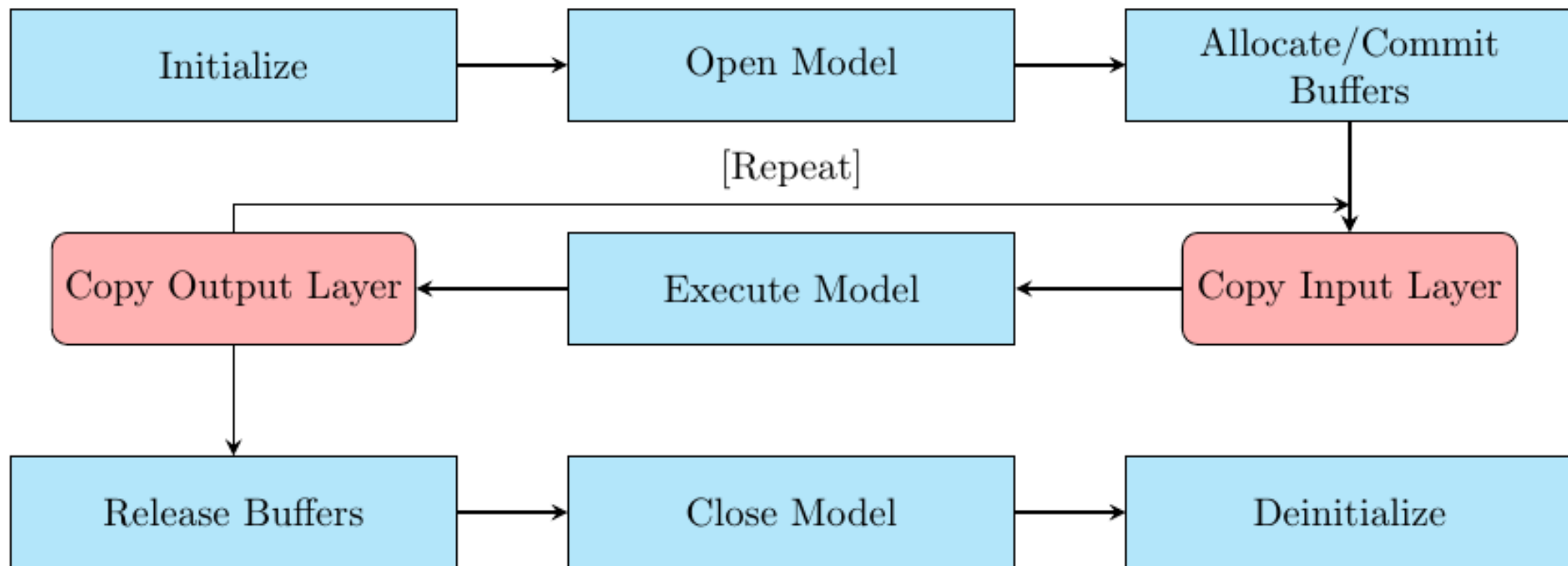
Best Practice Sample 소개 - HelpCrossing

- 어플리케이션의 동작 과정 소개



Best Practice Sample 소개 - HelpCrossing

- ENN Framework

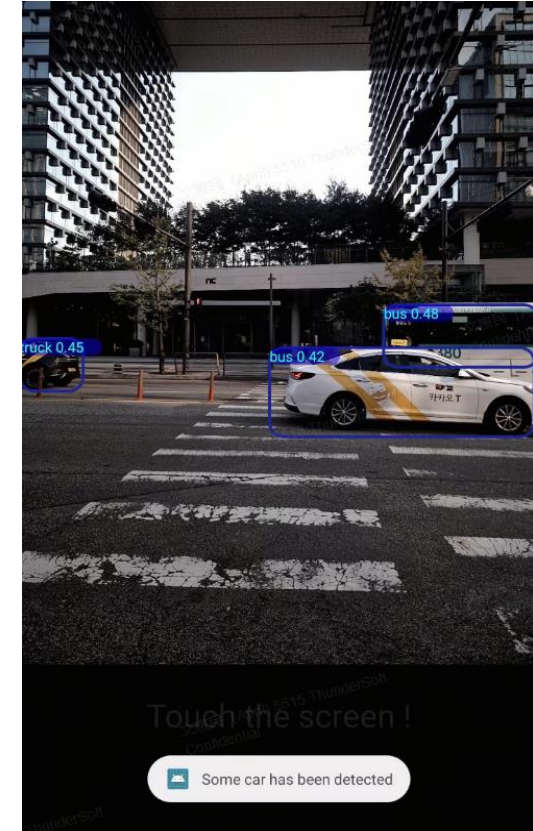


Best Practice Sample 소개 - HelpCrossing

● Application Demo 영상



1) 자동차가 없을 때



2) 자동차가 있을 때

Best Practice Sample 소개 - HelpCrossing

● [User Lab - User Project](#)



Picture



Source Code



Upload

Major ☐

Private ☐

Solved ☐

ADAS AI Automation Classification Compile Conversion DeepLearning Error ML ModelDevelopment NaturalLanguageProcessing NeuraNetworks PredictiveAnalytics Preprocessing Robotics Training
XR

font size ▼



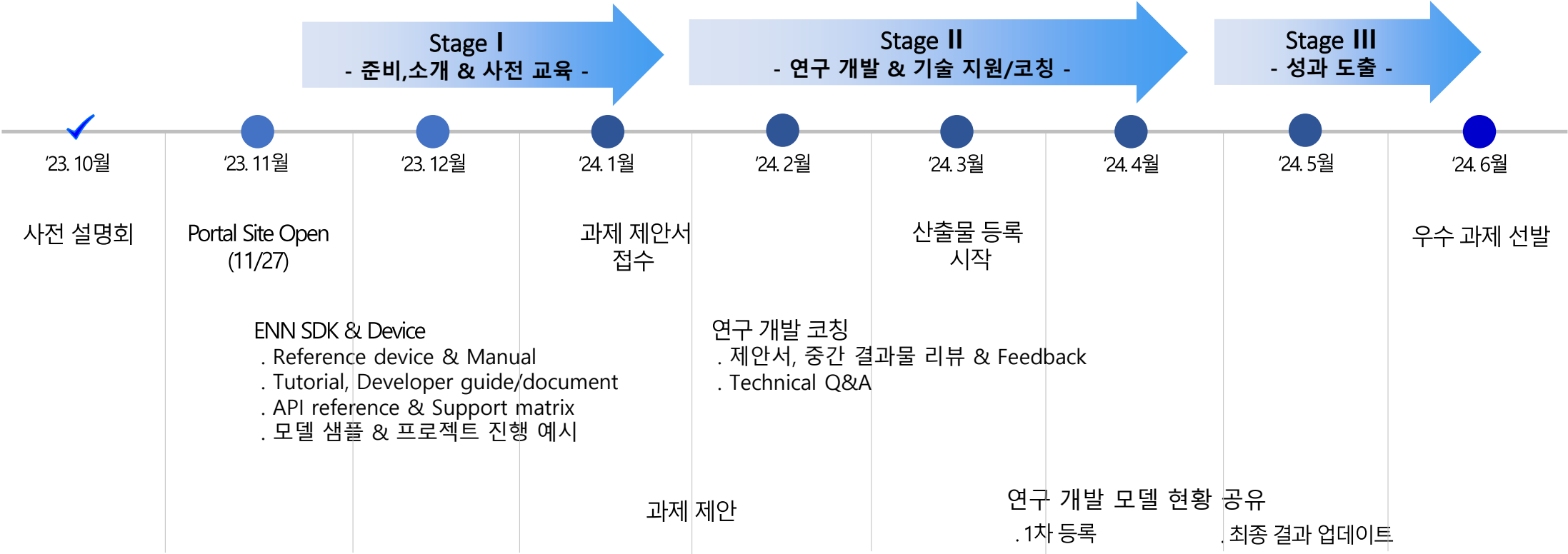
Enhancing Road Safety with HelpCrossing Application via ENN SDK

Description

The core objective of the HelpCrossing application is to augment road safety measures by providing real-time vehicular detection and auditory feedback to individuals intending to cross roads. By harnessing the power of the Exynos Neural Network (ENN) SDK, HelpCrossing aims to show proof-of-concept for minimizing road crossing hazards in urban settings.

프로그램 (AI Challenger) 주요일정

프로그램 주요 일정



주요 이벤트

코칭

Challenger

Q&A

END