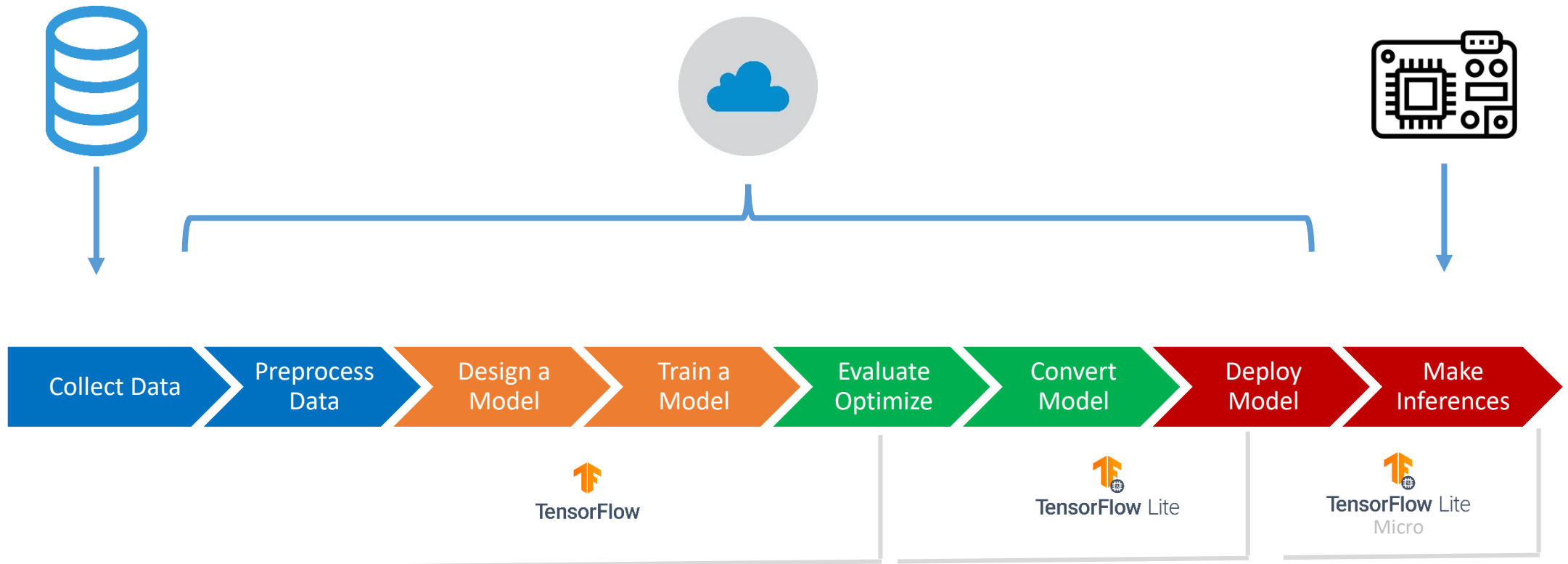


Программирование самообучающихся автономных AI-устройств на микроконтроллерах общего назначения

Область и определения

- Машинное обучение - выявление **эмпирических закономерностей** в данных.
- Искусственные нейросети – взаимодействующие между собой нейроны; реализуют принцип эмпирически формируемого решения с помощью **тренировки обучающей выборкой**.
- Обучение на микроконтроллере (ODL).
- Inference - **исполнение** обученной нейронной сети.
- Специфика исполнения нейросетей на микроконтроллерах:
 - Ограниченные ресурсы вычислителя, памяти.
 - Реальное время.
 - Быстро пополняющееся разнообразие архитектур нейросетей.
- Примеры применения:
 - Распознавание голосовых команд.
 - Детекция состояния оборудования.
 - Классификация объектов на видео камере ...
- Типы входных сигналов - временные ряды:
 - Акселерометр.
 - Микрофон...

Процесс разработки AI решения для MCU



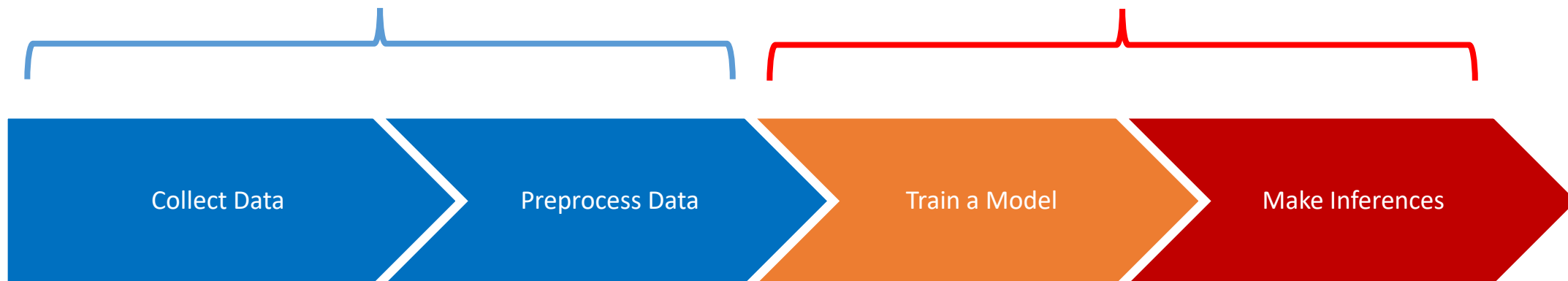
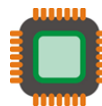
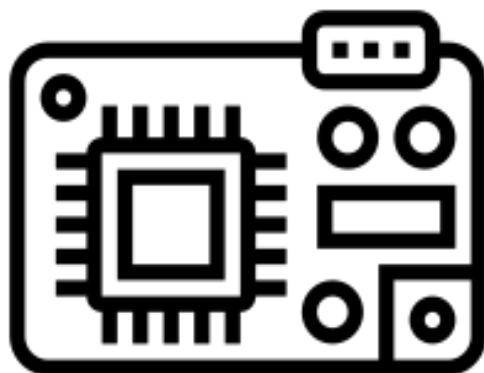
Проблемы

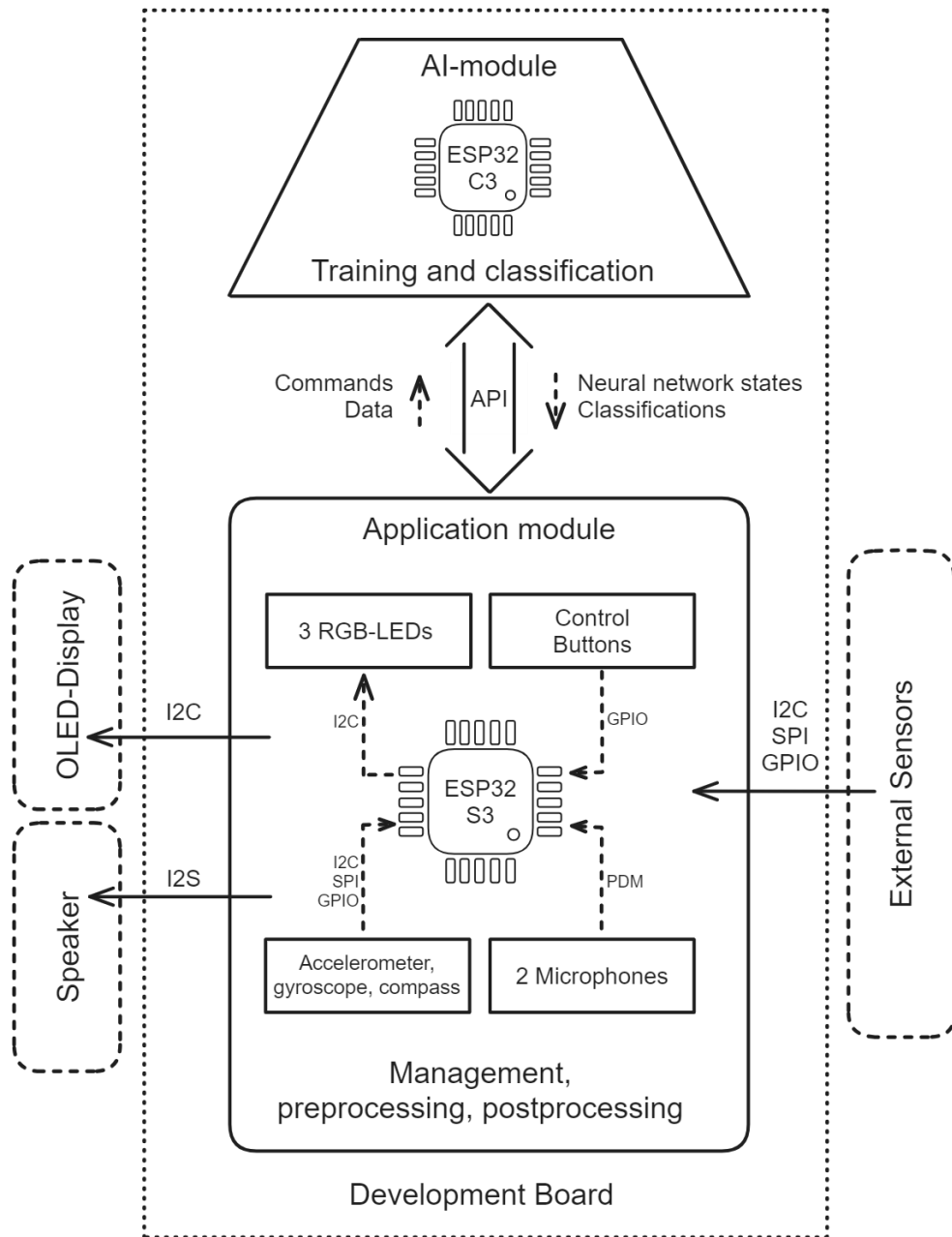
- Затраты времени и ресурсов на:
 - Сбор и подготовка наборов данных
 - Выбор подходящей модели
 - Обучение модели в облаке
 - Оптимизацию и подгонку модели
 - Освоение инструментов
- Существенные расходы на:
 - Специалистов
 - Хранение больших объемов данных
 - Поддержку инфраструктуры для обработки данных
 - Облачные среды разработки

Подход с использованием ML сопроцессора

Host MCU

GRC AI SW Module on Chip





Преимущества GRC AI Module

- Обучение новым классам в режиме реального времени
- Возможность разработки автономных решений благодаря низкому энергопотреблению и отсутствию необходимости передачи данных на сервер
- Гарантирована конфиденциальность и безопасность, т. к. нет необходимости передачи с целевого устройства
- Отличный потенциал разработки гибких решений, настраиваемых под конкретное оборудование

Технологические преимущества

Уходим от создания
математических и
статистических
моделей

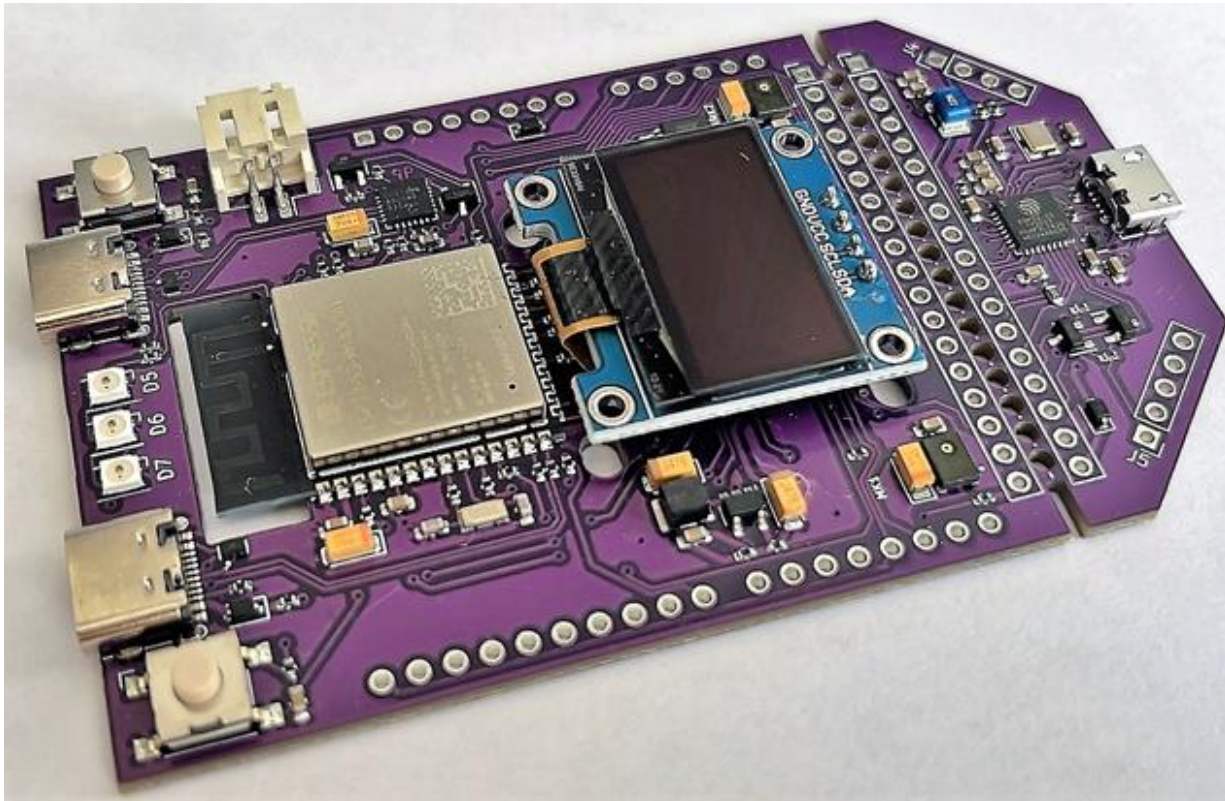
Не нужно
размечать сигнал,
готовить дата-сет

Не нужно обучать
сеть в облаке -
обучаем «на лету»

Минимальное
энергопотребление

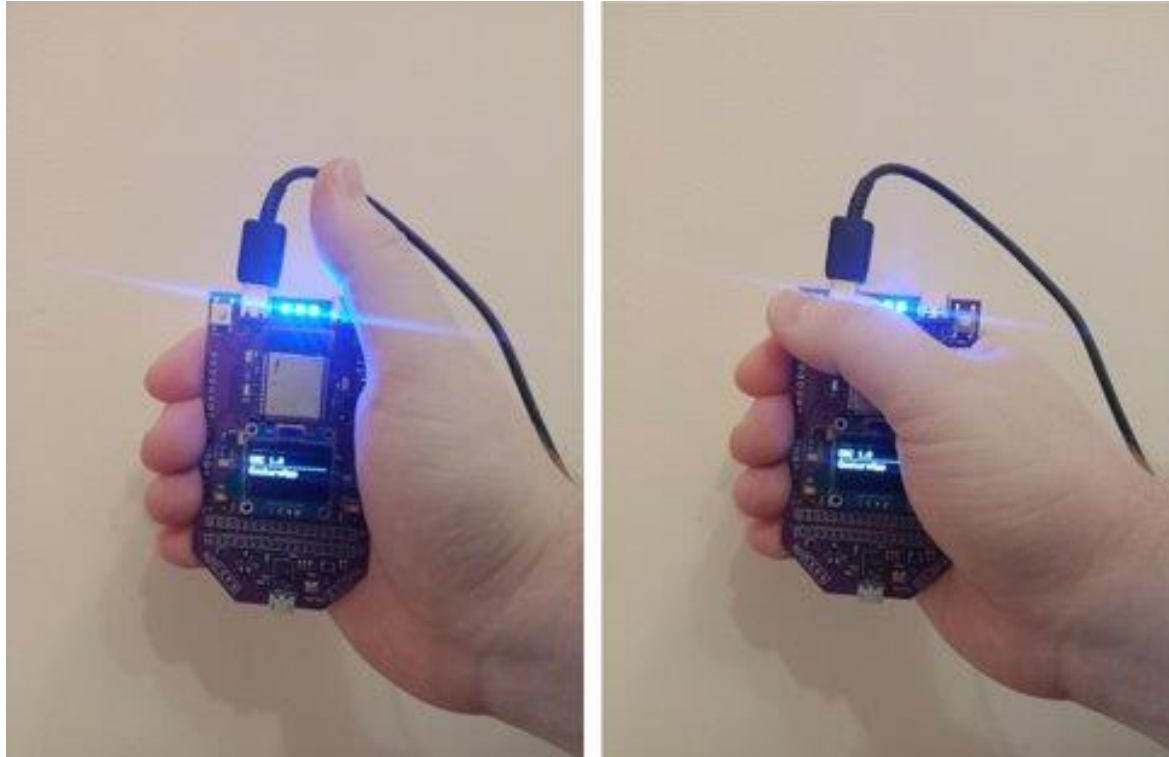
Доступность
вычислителей
(в тч для РФ)

Отладочная плата

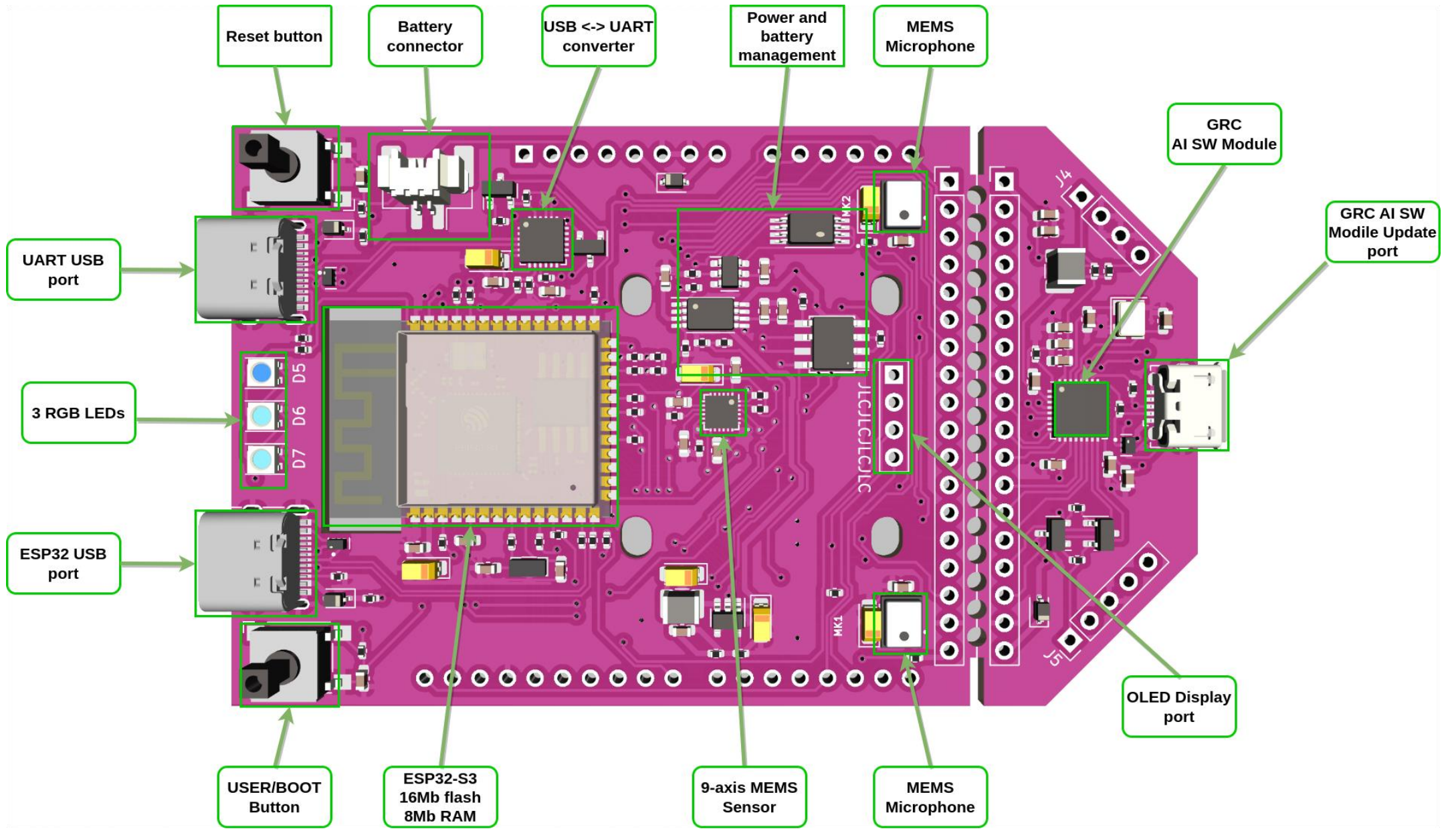


- Возможность опробовать новую парадигму в работе с сигналами
- Попробовать разработать свое решение
- Сделать это:
 - Просто,
 - Быстро
 - Не дорого

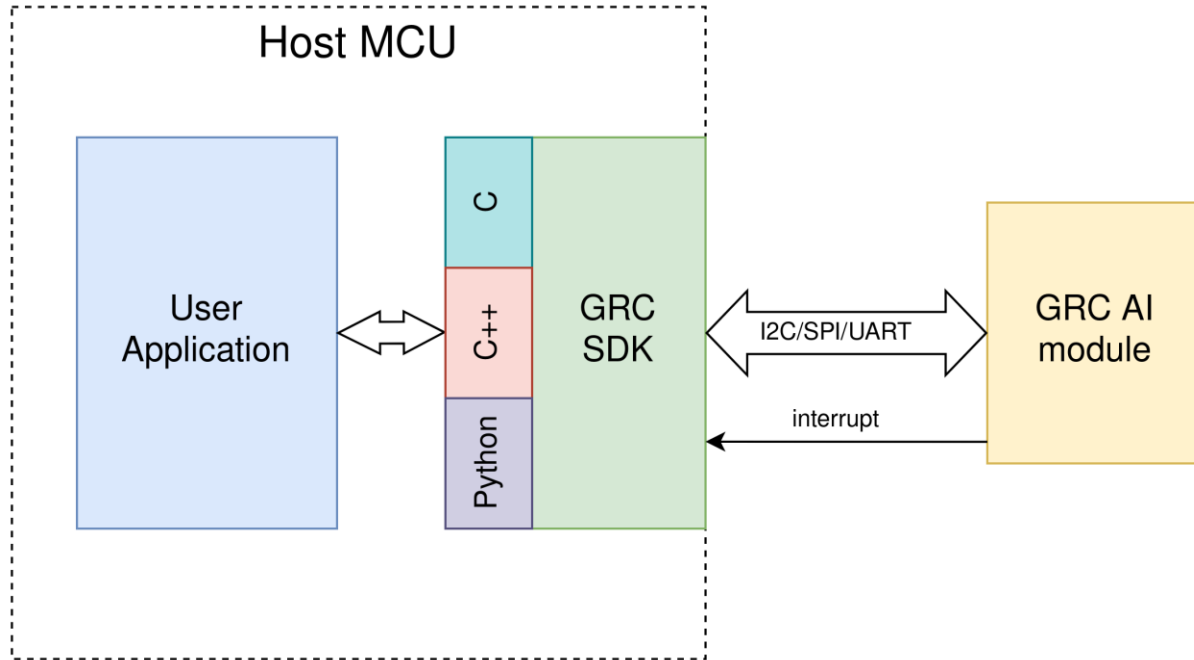
Знакомство



- **Осмотр**
 - Управление платой.
 - Отображение информации.
- **Режим пользователя**
 - Подключение.
 - Режим обучения.
 - Режим классификации.
 - Доступны на GitHub примеры:
 - Классификация жестов.
 - Распознавание ритма.
 - Контроль работы вентилятора.
- **Режим разработчика**
 - Прочитать Developer Guide.
 - Ознакомиться с примерами программирования.
 - Разработка приложения.

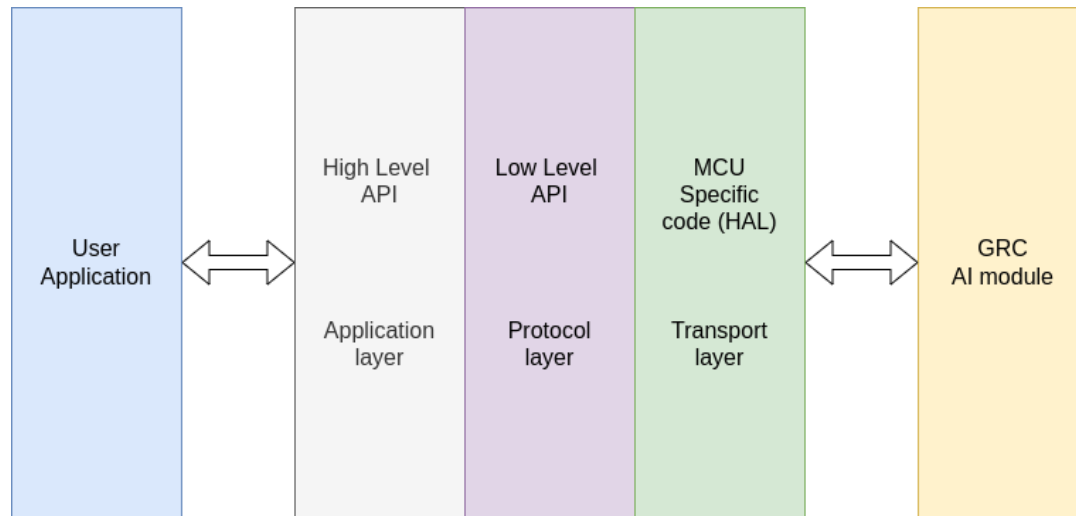


Функционал SDK



- Обучение новым классам в режиме реального времени
- Классификация
- Выбор из нескольких вариантов архитектуры модели
- Конфигурация гипер-параметров для более точной классификации
- Сохранение и загрузка предобученной модели
- Легкое портирование под новые архитектуры

Архитектура SDK



Транспортный уровень

Платформа зависимый интерфейс для работы с I2C/SPI. Предоставляет базовые функции: настройка, чтение и запись

Протокольный уровень

Протокол удаленного вызова функций на GRC, реализованный поверх транспортного уровня

Уровень приложения

Высокоуровневый интерфейс для вызова GRC функций

Производительность GRC на ESP32C3

Длина, сэмпл	Архитектура 1 (1 компонент)		Архитектура 2 (3 компонента)		Архитектура 3 (6 компонент)	
	Обучение на 1 класс, мс	Исполнение на 1 класс, мс	Обучение на 1 класс, мс	Исполнение на 1 класс, мс	Обучение на 1 класс, мс	Исполнение на 1 класс, мс
20	109	26	52	27	130	43
80	351	118	162	89	418	193
160	670	227	314	171	819	379
320	1302	447	621	334	1558	747
640	2547	890	1222	658	6696	1487

Аппаратные платформы для GRC Module

- CPU
 - Большой выбор. Низкая цена
 - ESP32C3
 - Выбираем из: RISC-V, Arm, ARC
- CPU + NNA
 - Большая вычислительная мощность
 - Низкое энергопотребление исполнения нейросетей
 - Не анонсируем
- FPGA
 - Сверхбыстрое исполнение нейросетей
 - Дорого
 - GoWin FPGA – прототип

Спасибо за внимание!

hi@mir.dev

<https://mir.dev>

https://github.com/Grovety/grc_devboard