

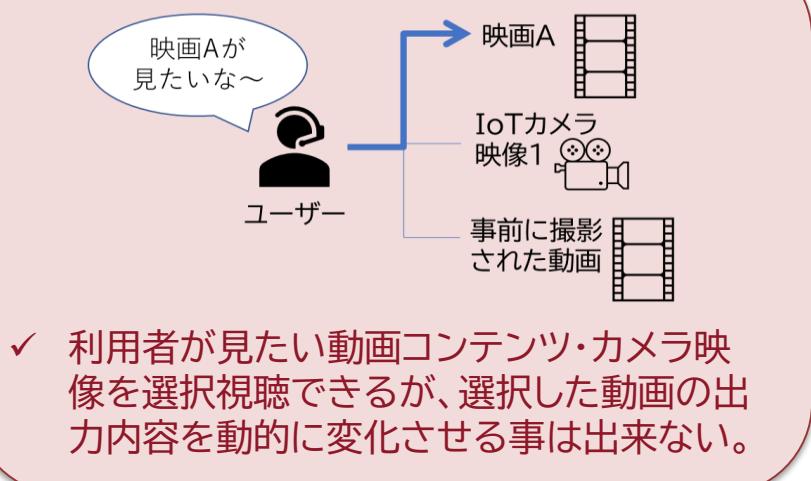
リアルタイム動画サービスを実現する オンデマンドAIのアーキテクチャ設計

アルプスシステムインテグレーション株式会社

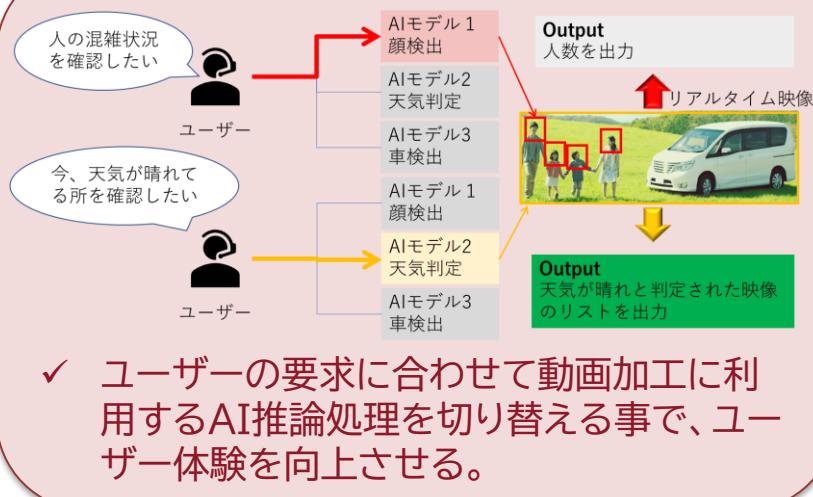
小野村 章伍

shogo.onomura@alsi.co.jp

既存オンデマンドサービスの課題

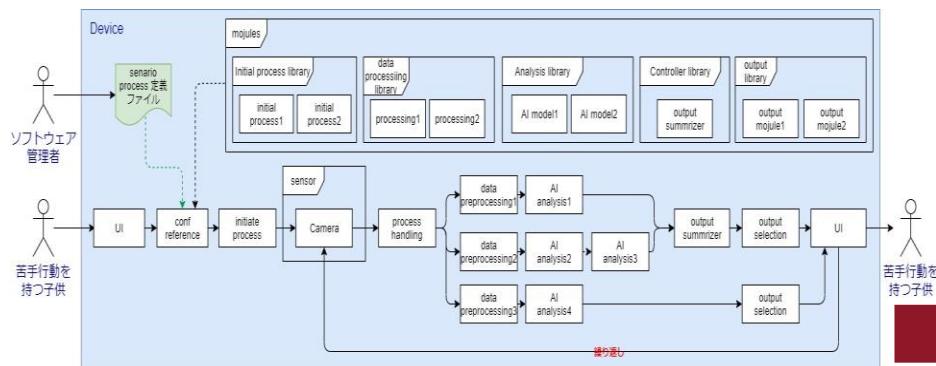


手法・ツールの適用による解決



ADDによるAI推論処理切り替えアーキテクチャの洗練

アーキテクチャ設計(第1イテレーション)



- ✓ AI解析を含む動画加工プロセスの直列・並列・拡張・集約配置に対応。各プロセスはYAML形式のファイル設定により定義出来る。

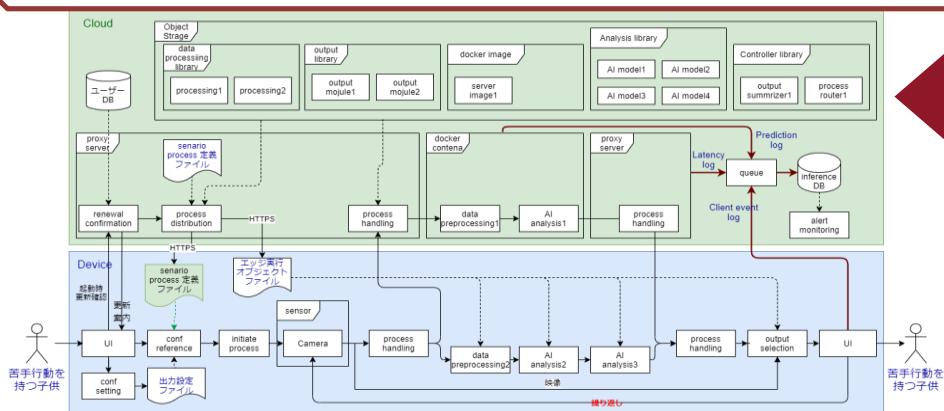
品質シナリオに基づくアーキテクチャ評価

- ✓ 第一イテレーションの設計を踏まえプロトタイプを実装
- ✓ 事前に策定した品質シナリオに基づいて、アーキテクチャの評価を実施。評価結果を次のアーキテクチャ設計へフィードバック

品質シナリオ	評価	理由	対策
撮影している映像が対象となる出力条件を満たし続ける限り、常に対象となる出力に準じた映像加工が継続している。	○	Videoキャプチャ毎に判定処理が働くので、映像加工を継続できる。	-
ユーザーからの指定によって、一時中断を介してAI推論処理を切り替えられる。	○	シナリオを切り替えることにより、推論処理を切り替え可能である。	-
AIモデルサイズの大小に応じてエッジ側とクラウド側にAI推論処理を適用できる事。	×	今回はデバイスのみの設計であるため。	アーキテクチャ設計の管理部分をクラウドに移す。シナリオプロセス定義ファイルに"cloud" or "edge" 選択項目を追加。
AIモデルの想定外が入力されたときに、適切に出力対処出来る事。	○	各モジュールCerr handlingメソッドを必ず定義するように強制する事で対応。 Controllerモジュールにより、エラー時のフロー制御が可能。	-
コンセプトドキュメントやデータリフトを検出し、継続的なモデル評価と改善を実行出来る事。	×	継続的なモデル評価と改善は一切考慮されていない。	推論監視パターンをアーキテクチャに導入する



アーキテクチャ設計(第2イテレーション)



- ✓ 推論結果のログ監視により、データドリフトの検出が可能
- ✓ AIモデルの適性に依じて、クラウド側とエッジ側でそれぞれ推論処理を実行できるように配慮

今後の展望

- ✓ ADD+品質シナリオ評価のプロセスを繰り返すことによって、オンデマンドAI機能を確立
- ✓ DevOps体制を確立
- ✓ アーリーアダプターを発見し、これまで確立したアーキテクチャ基盤に則って迅速なMVP作成を実行



アーリーアダプター

