ICT TECHNOLOGICAL TASK SPECIFICATION

Technical Specification: High-Fidelity Video Face-Swapping Pipeline

Organization: Alan Technologies, LLP
Status: Astana Hub Participant
Contact: Aidos Kuanyshev (+77478124387)

Project Title	Scalable Automated Video Face-Swapping & Restoration Inference Pipeline	**Target Timeline**	2 Months (8 Weeks)
Primary Domain	Artificial Intelligence & Computer Vision (Media Stream Processing)	**Budget Estimate**	2 500 000 KZT

1. EXECUTIVE SUMMARY & OBJECTIVE

The goal of this project is to develop and deploy an automated, modular, high-throughput video face-swapping pipeline utilizing leading open-source deep learning architectures. The system must process high-resolution video streams (1080p, 60fps), preserve temporal consistency across frames, and resolve visual artifacts (flickering, lighting mismatch, and head pose occlusions) without relying on proprietary, expensive SaaS APIs.

2. PIPELINE ARCHITECTURE & MODEL STACK

Pipeline Stage	Recommended Open-Source Stack	Functional Requirements
1. Detection & Alignment	RetinaFace, SCRFD, InsightFace (2D/3D Landmarks)	Robust facial landmark localization under extreme yaw angles (up to ±60°), rapid bounding-box tracking.
2. Identity Transfer	InsightFace (inswapper_128), Ghost, SimSwap, LivePortrait	Latent identity feature extraction, source identity preservation (Cosine ID similarity > 0.92 on ArcFace embeddings).
3. Face Restoration	CodeFormer, GPEN, GFPGAN v1.4	Restoration of skin texture, ocular fidelity, and edge sharpness; elimination of blur and low-resolution artifacting.
4. Temporal Blending	Bi-directional Optical Flow, Exponential Moving Average (EMA)	Suppression of inter-frame jitter and seam inconsistencies; natural skin tone color grading and Poisson blending.

3. KEY PERFORMANCE & QUALITY METRICS (KPIS)

- **Throughput & Latency:** Sustained processing speed of ≥ 24 FPS at 1080p resolution running on a single dedicated GPU (NVIDIA RTX 4090 / A10G 24GB).
- **Identity Retention:** ArcFace face verification retrieval score ≥ 95% between the source reference image and output video frames.
- **Temporal Stability:** Per-frame LPIPS variance ≤ 0.05 across contiguous static facial expressions to ensure zero noticeable frame-to-frame popping.
- **Optimization:** Models converted to ONNX / TensorRT FP16 engines for minimized memory footprint and accelerated GPU inference.

4. PHASED IMPLEMENTATION MILESTONES

Milestone	Deliverable	Acceptance Criteria
Stage 1 (W1–W2)	Architecture Benchmark & Feasibility	Benchmark report comparing SimSwap, Ghost, and InsightFace with CodeFormer; model latency matrix on reference hardware.
Stage 2 (W3–W5)	Core Pipeline & Blending Engine	Integrated Python module with temporal filtering, adaptive occlusion masking, and multi-face tracking support.
Stage 3 (W6–W7)	TensorRT Optimization & Packaging	Dockerized container with NVIDIA runtime, ONNX/TensorRT acceleration, REST API (FastAPI) and gRPC endpoints.
Stage 4 (W8)	Stress Testing & Technical Handover	Complete technical documentation, automated regression test suite, Postman collection, and client acceptance sign-off.

> **Data Compliance & Usage Notice:** The final service must run completely in isolated environments (on-premise or sovereign cloud VPC) without outbound telemetry, meeting all participant data security mandates set forth by Astana Hub.