IDRAK AI

TECHNICAL PROPOSAL & SCOPE DOCUMENT

Persona AI

An AI-Powered Augmented Reality Social Skills Training Platform — Personalized Coaching Through Life-Size Saudi AR Characters, Voice & Facial Analysis, and Enterprise Team Development

PREPARED BY	DATE	VERSION	CLASSIFICATION
Idrak AI — Engineering Division	February 2026	1.0 — Final	Confidential

00

Table of Contents

DOCUMENT NAVIGATION

01	Executive Summary	03
02	Market Opportunity & Problem Statement	03
03	Platform Vision & Core Features	04
04	Technical Architecture & AI/ML Pipeline	05
05	AI Character Engine & Voice System	06
06	AR Experience & Saudi Cultural Localization	07
07	Monetization & Pricing Strategy	08
08	B2B Enterprise Module	08
09	Technology Stack & Infrastructure	09
10	Development Roadmap & Timelines	10
11	Budget Breakdown	11
12	Risk Mitigation & Quality Assurance	12
13	Why Idrak AI	12

Document Purpose

This document serves as the comprehensive technical proposal and scope definition for the design, development, and deployment of the Persona AI platform. It details the system architecture, AI/ML pipeline, feature specifications, development timelines, and commercial framework for building an enterprise-grade, AI-powered social skills training platform.

01
Executive Summary

THE OPPORTUNITY AT A GLANCE

Persona AI is a next-generation, AI-powered **augmented reality** social skills training platform built specifically for the **Saudi Arabian market**. Users point their phone camera at any room and interact with life-size, culturally authentic Saudi AI characters that sit, stand, walk, and converse in real-time — all rendered in AR. The platform analyzes both **voice** and **facial expressions** to deliver a complete communication assessment.

The platform combines **AR spatial rendering**, **LLM-driven conversation**, **real-time voice synthesis in Saudi dialect**, **computer vision-based facial emotion detection**, and **behavioral analytics** into a unified mobile experience purpose-built for Saudi professionals, students, and enterprise teams.

$38B
CORPORATE TRAINING MARKET (2026)

72%
EMPLOYERS CITE SOFT SKILLS GAPS

4.2x
ROI ON COMMUNICATION TRAINING

02
Market Opportunity

PROBLEM & DEMAND LANDSCAPE

The Problem

In Saudi Arabia, rapid Vision 2030 workforce transformation demands strong interpersonal skills across industries — yet training options remain limited to expensive executive coaches ($300–$500/hr), generic Western-centric e-learning modules with no cultural relevance, or roleplay workshops that don't scale. 85% of professionals receive no structured social skills development despite it being the #1 predictor of career advancement.

Current Market Gaps

- **No personalization** — one-size-fits-all content ignores individual contexts
- **No practice environment** — learners read theory but never practice in realistic scenarios
- **No measurement** — no data-driven tracking of soft skill improvement
- **No scalability** — coaching is expensive and doesn't scale for enterprise teams

Persona AI's Differentiators

- **Augmented Reality Characters** — life-size AI avatars in the user's real environment via phone camera
- **Voice + Face Analysis** — dual-channel assessment captures tone, clarity, and facial micro-expressions
- **Saudi-First Design** — avatars in Saudi attire, Saudi Arabic dialect, culturally appropriate scenarios
- **Personality Tuning** — dial an AI character to match a specific real-world person

03
Platform Vision & Core Features

WHAT WE'RE BUILDING

User Journey Overview

A user opens the app, points their phone camera at any room, and a life-size Saudi AI character appears in augmented reality — sitting on a chair, standing by a desk, or walking across the space. The user engages in natural spoken conversation in Saudi Arabic or English while the app simultaneously analyzes their **voice** (tone, clarity, pacing) and **facial expressions** (confidence, nervousness, engagement) through the front camera. Post-session, they receive a comprehensive dual-channel analytics report.

1. Onboarding & Diagnostic Engine

NLP-driven intake assessment analyzes user-described challenges. Maps to skill taxonomy (assertiveness, empathy, negotiation, etc.). Generates initial skill profile and recommends learning pathway.

2. AR Character Interaction

Life-size 3D Saudi avatars rendered in the user's real environment via ARKit/ARCore. Characters sit, stand, walk, gesture, and maintain eye contact. Spatial audio creates immersive presence.

3. Voice & Face Analysis Engine

Dual-channel analysis: voice pipeline scores tone, clarity, and assertiveness while computer vision detects facial micro-expressions (confidence, nervousness, engagement, empathy) in real-time.

4. Personality Tuning System

Slider-based interface to adjust AI character traits: aggression, empathy, formality, patience, directness. Users recreate challenging interpersonal dynamics for targeted practice.

5. Saudi Cultural Scenarios

Pre-built scenario library tailored to Saudi professional contexts: Majlis-style negotiations, Vision 2030 corporate presentations, cross-cultural business meetings, and public sector interactions.

6. Dual-Channel Analytics

AI-generated session debrief combining voice scores (clarity, persuasion, assertiveness) with facial analysis (confidence level, emotional regulation, eye contact). Longitudinal progress tracking.

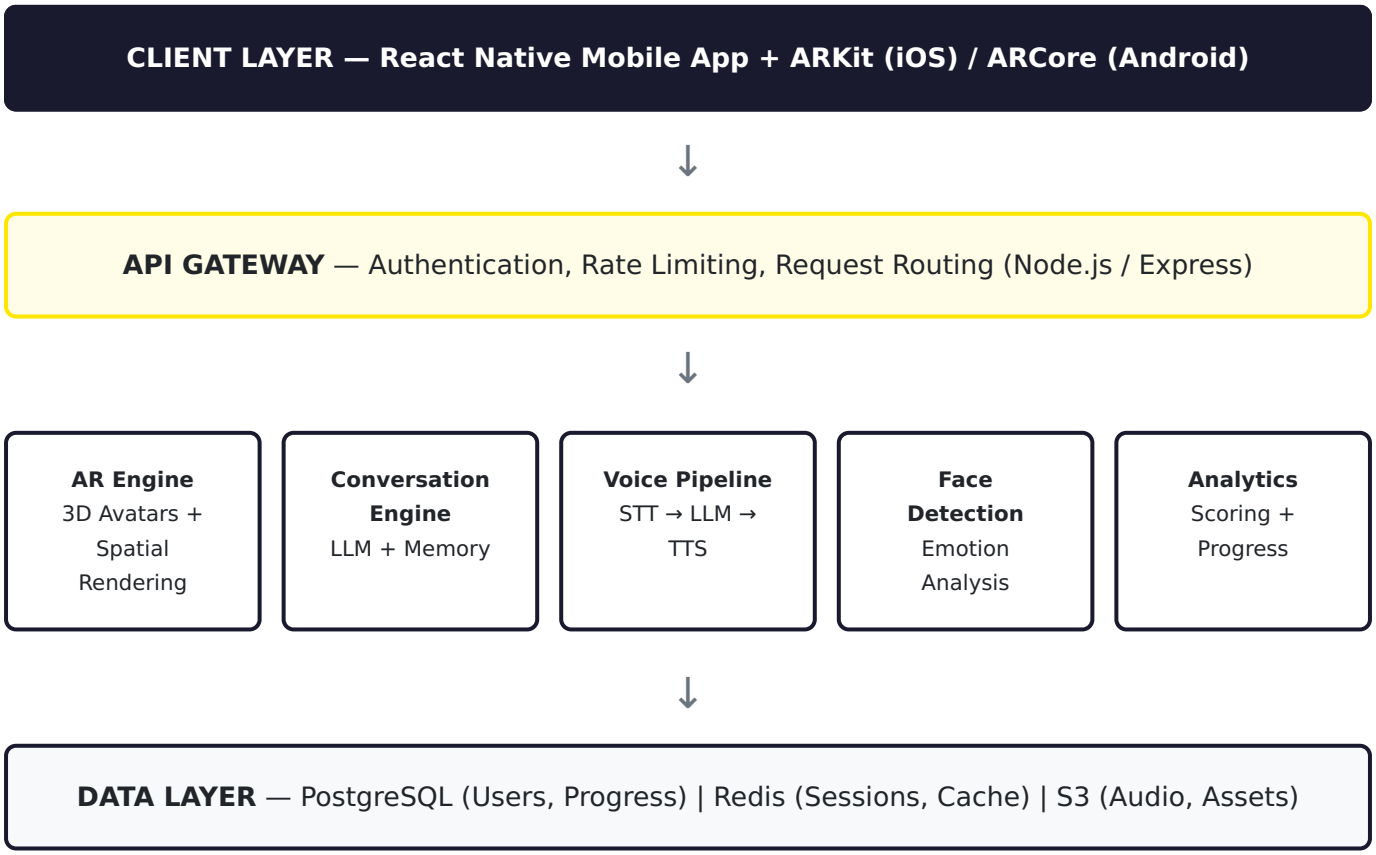
Key Innovation: AR + Dual-Channel Analysis

Unlike flat chatbot interactions, Persona AI places a culturally authentic Saudi character in the user's real space via AR — creating genuine social pressure and body-language cues. Simultaneously, the platform performs **voice analysis** (LLM-scored) and **facial emotion detection** (computer vision) to provide a complete communication assessment that no voice-only or text-only tool can match.

04
Technical Architecture & AI/ML Pipeline

High-Level Architecture

The platform follows a microservices-oriented architecture deployed on cloud infrastructure with the following layers:



AI/ML Pipeline — Detailed

1. Natural Language Understanding (NLU)

User input is processed through a multi-stage pipeline: (a) speech-to-text transcription via Whisper (open-source), (b) intent classification and entity extraction, (c) emotional tone analysis using fine-tuned sentiment models, (d) context-aware response generation through LLM with personality-conditioned system prompts.

2. Face Detection & Emotion Analysis

The front camera captures the user's face during sessions. A lightweight on-device CNN model (MediaPipe Face Mesh + custom emotion classifier) detects 7 core emotions in real-time: confidence, nervousness, engagement, frustration, empathy, surprise, and neutrality. Aggregated scores feed into the dual-channel analytics engine alongside voice scores.

3. Conversation Memory & State

Sessions maintain short-term memory (current conversation context) and long-term memory (user history, past sessions, skill progression). Implemented via a sliding window context with retrieval-augmented generation (RAG) for pulling relevant past session insights.

4. Dual-Channel Assessment Engine

Post-session, the system combines **voice analysis** (scored across 8 dimensions: clarity, empathy, assertiveness, active listening, conflict resolution, persuasion, emotional regulation, adaptability) with **facial analysis** (confidence, eye contact, emotional regulation, composure under pressure) into a unified communication score.

Open-Source First Approach

We leverage open-source and cost-effective models to maximize the $58K budget: **Whisper** (STT), **ElevenLabs / Coqui TTS** (voice synthesis), **Llama 3.x / Mistral** (conversation LLM — self-hosted or API), **MediaPipe** (on-device face detection), and **open embedding models** for RAG. AR rendering uses **ARKit/ARCore** with optimized 3D avatar assets.

05
AI Character Engine & Voice System

THE HEART OF THE EXPERIENCE

Character Archetypes (MVP)

The platform ships with 6 pre-designed Saudi character archetypes, each rendered as 3D AR avatars wearing culturally authentic Saudi attire (thobe, ghutra, abaya, etc.) with distinct personality vectors and voice profiles in Saudi Arabic dialect:

CHARACTER	ARCHETYPE	KEY TRAITS	USE CASE
Khalid	Supportive Mentor	High empathy, patient, warm	Building confidence, gentle feedback
Faisal	Tough Manager	Direct, demanding, low patience	Managing up, handling criticism
Noura	Corporate Executive	Analytical, strategic, Socratic	Leadership skills, strategic comms
Omar	Difficult Client	Volatile, high expectations	Client management, de-escalation
Sara	Peer Collaborator	Casual, opinionated, competitive	Team dynamics, negotiation
Dr. Abdulrahman	Interview Panel	Formal, evaluative, structured	Interview prep, presentations

Personality Tuning Interface

Users can modify any archetype or create custom characters using slider-based controls:

Tunable Parameters

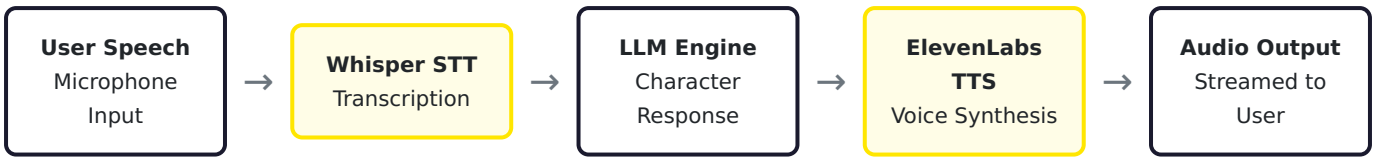
- **Directness** — Blunt ←→ Diplomatic
- **Patience** — Impatient ←→ Very patient
- **Warmth** — Cold/formal ←→ Friendly/warm
- **Assertiveness** — Passive ←→ Aggressive
- **Emotional Range** — Stoic ←→ Expressive
- **Formality** — Casual ←→ Corporate

Technical Implementation

Slider values (0–100) are normalized into a personality vector that dynamically generates the LLM system prompt using a template engine. The same vector modulates ElevenLabs TTS parameters:

- Voice **stability** maps to emotional range
- Voice **speed** maps to patience level
- Voice **style** maps to warmth/formality
- Voice **selection** maps to archetype gender/age

Voice Pipeline Architecture



Target latency: <1.5s end-to-end using streaming TTS and LLM token streaming. WebSocket connection maintains persistent session state.

06
AR Experience & Saudi Cultural Localization

IMMERSIVE & CULTURALLY AUTHENTIC

Augmented Reality Experience

The core differentiator of Persona AI is the **augmented reality interaction model**. Users point their phone at any real-world space and the AI character appears as a life-size, 3D-rendered Saudi avatar that naturally occupies the environment.

AR Character Behaviors

BEHAVIOR	IMPLEMENTATION
Sitting	Character detects flat surfaces (chairs, sofas) and sits naturally with idle animations
Standing	Character stands on detected floor plane with natural weight-shifting and gestures
Walking	Character moves across the room along user-defined paths or AI-driven waypoints
Gesturing	Hand gestures synchronized with speech emphasis and emotional context
Eye Contact	Character head and eyes track toward the user's phone position for realism
Spatial Audio	Voice volume and direction change based on character distance and position

Technical AR Stack

- **ARKit (iOS) / ARCore (Android)** — Plane detection, surface anchoring, light estimation
- **Three.js / React Three Fiber** — 3D avatar rendering within React Native
- **Mixamo / Custom Rigging** — Character animation: idle, walk, sit, gesture cycles
- **GLTF/GLB Format** — Optimized 3D models (<5MB per avatar) for mobile performance
- **Spatial Audio API** — Directional audio based on character position in AR scene

Performance Target

60fps AR rendering on iPhone 12+ and equivalent Android devices. Avatar models optimized to <50K polygons with LOD (level-of-detail) fallbacks for older devices.

Saudi Cultural Localization

Language & Dialect

- **Saudi Arabic dialect** — All character speech in authentic Saudi dialect, not MSA
- **Bilingual support** — Switch between Arabic and English mid-session
- **RTL interface** — Full Arabic right-to-left UI throughout the app
- **Cultural vocabulary** — Business terms, honorifics, and expressions used in Saudi professional contexts

Visual & Cultural Authenticity

- **Saudi attire** — Male avatars in thobe and ghutra; female avatars in professional abaya
- **Authentic environments** — Scenarios contextualized to Saudi offices, Majlis settings, and conference rooms
- **Cultural norms** — Greeting protocols, eye contact expectations, and communication etiquette aligned with Saudi business culture

Gamification — Planned for Future Phase (v2.0)

Engagement features including XP points, skill badges, daily streaks, leaderboards, and level progression will be implemented in a future phase to maximize retention after the core AR experience is validated. This keeps MVP scope focused on the core AI + AR + face detection experience.

07
Monetization & Pricing Strategy

REVENUE MODEL

FEATURE	FREE TIER	PREMIUM ($19.99/MO)	ENTERPRISE (CUSTOM)
Sessions / month	5 sessions	Unlimited	Unlimited
AI Characters	2 (Alex + Sam)	All 6 + Custom	All + Custom corporate personas
Personality Tuning	Basic (3 sliders)	Full (6 sliders)	Full + API-driven presets
Voice Interaction	Text only	Full voice (STT+TTS)	Full voice + custom voices
Analytics	Basic voice score	Full voice + face analysis dashboard	Team analytics + manager reports
AR Experience	Standard avatars	All avatars + custom tuning	Custom corporate avatars
Face Detection	Basic confidence score	Full 7-emotion analysis	Team emotion analytics
Learning Pathways	Auto-generated	Custom + AI recommended	Manager-defined team pathways
Support	Community	Priority email	Dedicated CSM + SLA

08
B2B Enterprise Module

TEAM DEVELOPMENT AT SCALE

The enterprise module enables companies to onboard teams, define skill development goals, and track progress through a dedicated admin dashboard.

Admin Dashboard Features

- **Team Management** — Add/remove members, create teams (e.g., "Engineering", "Sales")
- **Goal Setting** — Define team-level skill goals (e.g., "Improve presentation skills by Q3")
- **Progress Tracking** — Aggregate skill scores, session completion rates, engagement metrics
- **Custom Scenarios** — Upload company-specific scenarios (e.g., handling customer complaints per brand guidelines)
- **Reporting** — Exportable PDF/CSV reports for HR and L&D teams

Enterprise Integration Points

- **SSO** — SAML 2.0 / OAuth 2.0 integration
- **LMS Integration** — SCORM/xAPI compatibility for existing learning management systems
- **HRIS Sync** — Team structure auto-sync from HR systems
- **API Access** — REST API for custom integrations and automation
- **Data Residency** — Configurable data storage region for compliance

Enterprise Example

"A 50-person engineering team at Company X wants to improve presentation skills before a major product launch. The L&D manager sets a team goal, assigns the 'Casey — Interview Panel' character with custom scenarios around technical presentations, and tracks weekly progress via the admin dashboard."

09

Technology Stack & Infrastructure

ENGINEERING DECISIONS

Mobile & AR Layer

COMPONENT	TECHNOLOGY
Framework	React Native + Expo
AR Engine	ARKit (iOS) / ARCore (Android)
3D Rendering	Three.js / React Three Fiber
Avatar Format	GLTF/GLB (optimized <5MB)
State Management	Zustand
Real-time Comms	WebSocket (Socket.io)
Audio/Camera	Native APIs + MediaRecorder

AI / ML Layer

COMPONENT	TECHNOLOGY
Core LLM	Llama 3.1 70B / Mistral Large (via Groq or Together AI)
Speech-to-Text	OpenAI Whisper (open-source, self-hosted)
Text-to-Speech	ElevenLabs API (production) / Coqui TTS (fallback)
Face Detection	MediaPipe Face Mesh + custom emotion classifier (on-device)
3D Avatars	Custom Saudi characters (Mixamo rigging + hand-modeled)
Embeddings	BGE-base / Nomic-embed for RAG
Evaluation	Dual-channel (voice + face) scoring pipeline

Backend

COMPONENT	TECHNOLOGY
Runtime	Node.js 22 (Express/Fastify)
Database	PostgreSQL 16 (Supabase)
Cache / Sessions	Redis (Upstash)
File Storage	AWS S3 / Cloudflare R2
Job Queue	BullMQ (Redis-backed)
API Docs	OpenAPI 3.0 / Swagger

Infrastructure & DevOps

COMPONENT	TECHNOLOGY
Hosting	Vercel (Frontend) + Railway/ Render (Backend)
CI/CD	GitHub Actions
Monitoring	Sentry + LogTail
Analytics	PostHog (open-source)
Email	Resend
Payments	Stripe

Why This Stack?

Every choice optimizes for three constraints: **(1) cost efficiency** — leveraging free tiers (Vercel, Supabase, Upstash), open-source models, and usage-based pricing; **(2) speed to market** — battle-tested tools with rich ecosystems reduce development time; **(3) scalability** — all components are horizontally scalable when the platform grows past MVP.

10
Development Roadmap & Timelines

20-WEEK DELIVERY PLAN

1

Phase 1 — Foundation & Mobile App Shell WEEKS 1-3

Project setup, CI/CD pipeline, database schema design, authentication system, React Native app shell with AR camera permissions, API architecture, WebSocket infrastructure for real-time communication.

Deliverables: Authenticated mobile app shell, API gateway, database schema, dev/staging environments.

2

Phase 2 — AI Engine, Voice & Face Detection WEEKS 4-8

LLM integration (Llama/Mistral via Groq), character persona system with personality vectorization, prompt engineering for all 6 Saudi archetypes, Whisper STT + ElevenLabs TTS integration, **MediaPipe face detection pipeline**, emotion classifier training, dual-channel scoring engine.

Deliverables: Working voice + face detection conversation, personality tuning interface, end-to-end dual-channel pipeline.

3

Phase 3 — AR Experience & 3D Saudi Avatars WEEKS 9-13

ARKit/ARCore integration, 3D Saudi avatar modeling and rigging (thobe, ghutra, abaya), animation cycles (sit, stand, walk, gesture), plane detection and avatar anchoring, spatial audio, Arabic RTL interface implementation, Saudi dialect voice tuning.

Deliverables: Full AR experience with life-size Saudi avatars, spatial audio, Arabic-localized UI.

4

Phase 4 — Analytics, Enterprise & Payments WEEKS 14-17

Onboarding assessment, adaptive learning pathways, dual-channel analytics dashboard (voice + face), enterprise admin dashboard, team management, SSO, reporting, Stripe payment integration.

Deliverables: Full analytics, enterprise admin, team management, payment flow.

5

Phase 5 — QA, Arabic Localization & Launch WEEKS 18-20

End-to-end QA testing on iOS + Android, AR performance optimization, voice latency tuning, Arabic content QA, security audit, App Store / Play Store submissions, launch readiness review.

Deliverables: Production deployment, App Store/Play Store listing, documentation handoff.

20	5	Week 8	Week 20
WEEKS TOTAL	DEVELOPMENT PHASES	FIRST USABLE DEMO	PRODUCTION LAUNCH

11

Budget Breakdown

INVESTMENT ALLOCATION — $58,000 USD

One-Time Development Cost

#	CATEGORY	SCOPE	AMOUNT	%
1	**Mobile App & AR Experience**	React Native app, ARKit/ARCore integration, AR scene management, spatial audio, camera pipeline	**$12,500**	22%
2	**Backend & API Development**	API architecture, database, auth, WebSocket, job queues, payment integration	**$6,500**	11%
3	**AI/ML Engine & Face Detection**	LLM integration, face emotion detection (MediaPipe), personality tuning, dual-channel evaluation pipeline, RAG	**$10,500**	18%
4	**Voice Pipeline (STT + TTS)**	Whisper integration, ElevenLabs integration, Saudi dialect tuning, streaming audio, latency optimization	**$4,500**	8%
5	**3D Saudi Avatar Design & Animation**	6 culturally authentic Saudi 3D characters (thobe, ghutra, abaya), rigging, sit/stand/walk/gesture animations	**$7,500**	13%
6	**B2B Enterprise Module**	Admin dashboard, team mgmt, goal setting, reporting, SSO	**$4,000**	7%
7	**UI/UX Design & Saudi Localization**	Arabic RTL interface, Saudi cultural UX, wireframes, design system, interaction design	**$6,000**	10%
8	**QA, Testing & DevOps**		**$3,500**	6%

#	CATEGORY	SCOPE	AMOUNT	%
		iOS + Android device testing, AR performance optimization, security audit, CI/CD, App Store submissions		
9	**Third-Party Services (6 months)**	ElevenLabs API, Groq/Together AI credits, hosting, App Store fees, monitoring	**$3,000**	5%
	TOTAL		**$58,000**	**100%**

Estimated Recurring Monthly Costs (Post-Launch)

ITEM	DETAILS	EST. MONTHLY
Hosting & Infrastructure	Vercel, Railway/Render, Supabase, Redis, S3, CDN	$500 – $700
LLM API Credits	Groq / Together AI (usage-based, scales with users)	$600 – $800
ElevenLabs TTS API	Voice synthesis credits (usage-based, Arabic + English)	$500 – $650
AR Cloud & 3D Asset Hosting	3D model delivery, AR session processing, asset CDN	$300 – $400
Monitoring & Analytics	Sentry, PostHog, LogTail, crash reporting	$200 – $250
App Store Fees	Apple Developer ($99/yr) + Google Play ($25 one-time)	~$10
Technical Support (Idrak AI)	Bug fixes, updates, performance monitoring, on-call — 15 hrs/month	$2,400
ESTIMATED TOTAL		**$4,500 – $5,100/mo**

* Recurring costs are estimates and will vary based on active user count. LLM and TTS costs scale linearly with usage. Technical support can be adjusted based on needs.

Payment Schedule (One-Time Development)

MILESTONE	TRIGGER	AMOUNT
Milestone 1	Project kickoff + Phase 1 completion (Week 3)	$17,400 (30%)
Milestone 2	Phase 2–3 completion — Working AR + AI voice demo (Week 13)	$17,400 (30%)
Milestone 3	Phase 4–5 completion — Full platform launch (Week 20)	$23,200 (40%)

12

Risk Mitigation & Quality Assurance

PROACTIVE RISK MANAGEMENT

RISK	SEVERITY	MITIGATION STRATEGY	CONTINGENCY
AR performance on older devices	HIGH	LOD avatar fallbacks, polygon budget <50K, min device: iPhone 12 / equiv. Android	2D avatar mode for unsupported devices
Voice latency > 2s	HIGH	Streaming TTS, LLM token streaming, WebSocket persistent connection, edge caching	Fallback to text-first mode with optional voice replay
Face detection accuracy	MEDIUM	On-device MediaPipe processing; calibration step during onboarding; lighting detection	Graceful degradation to voice-only scoring if face detection confidence is low
Saudi dialect TTS quality	MEDIUM	ElevenLabs Arabic voice fine-tuning; manual QA with native speakers	Fallback to MSA with Saudi vocabulary overlay
Scope creep	HIGH	Strict phase-gated delivery, weekly sprint reviews, documented change request process	Defer non-MVP features to v2.0 backlog

Scope Clarity — What's In & What's Not

Included in MVP ($58K)

- AR mobile app (iOS + Android)
- 6 Saudi 3D avatars with animations
- Voice analysis (STT + TTS + scoring)
- Face detection & emotion analysis
- Saudi Arabic dialect + English support
- Arabic RTL interface
- Personality tuning system
- Adaptive learning pathways
- Dual-channel analytics dashboard
- B2B enterprise admin module
- Payment integration (Stripe)
- App Store / Play Store deployment

Not Included — Future Phases

- Gamification (XP, badges, leaderboards, streaks) — **v2.0**
- Web application (mobile-first for AR)
- Additional dialects (Egyptian, Emirati, etc.)
- Custom voice cloning per user
- Offline mode / on-device LLM
- Apple Vision Pro / Meta Quest support
- Additional 3D avatars beyond initial 6
- Advanced body language analysis

13
Why Idrak AI

YOUR ENGINEERING PARTNER

Our Capabilities

- **AI-Native Team** — Deep expertise in LLM integration, prompt engineering, and conversational AI systems
- **Full-Stack Delivery** — End-to-end product development from design to deployment
- **AR & Voice AI Experience** — Prior work with AR rendering, real-time voice pipelines, STT/TTS integration
- **Rapid Prototyping** — Lean methodology, fast iteration, working demos within weeks

Our Commitment

- **Transparent Communication** — Weekly progress reports, demo sessions, open Slack channel
- **Quality First** — Code reviews, automated testing, staging environment for every release
- **Post-Launch Support** — Ongoing technical support available (see recurring costs)
- **Knowledge Transfer** — Full documentation, codebase walkthrough, and handoff support

Ready to Build the Future of Social Skills Training

Idrak AI — Building Intelligent Systems That Understand People

Contact: **hello@idrak.ai** | Let's build Persona AI together.