



## RFP Response: Scope of Work (SoW)

### Comprehensive Listing of U.S. Job Postings for Robotics and Artificial Intelligence Competencies



## Table of Contents

Introduction .....	3
Objectives .....	5
1. Company Background .....	6
2. Project Methodology – Draup Job Posting Corpus.....	11
3. Technology Stack and tools .....	24
4. Project Timeline, including major milestones.....	29
5. Customer Service / Maintenance Plan .....	30
6. Cost Proposal .....	32
7. Sample Outputs .....	33
8. References .....	37
9. Litigation.....	38

## Introduction

Draup is pleased to submit this proposal in response to the ARM Institute's RFP "*Comprehensive Listing of U.S. Job Postings for Robotics and Artificial Intelligence Competencies*" (RFP No. 25-02).

We recognize the ARM Institute's leadership in advancing the U.S. manufacturing workforce through roboticscareer.org and the expansion into AI competencies. Draup shares this vision and brings deep expertise in workforce intelligence, job architecture, and labour market analytics to support this important initiative.

Draup is uniquely positioned to deliver structured data because of the unmatched depth and scale of its proprietary datasets. Unlike generic providers, **Draup's Skills Architecture framework** ensures that data is structured by design, translating raw inputs into actionable, relational insights across capabilities, roles, workloads, and skills. This structured approach is further reinforced by a rigorous three-layer validation process, combining internal research expertise, curated external assets, and targeted industry SME reviews to ensure accuracy and reliability. The result is data that is not only comprehensive but also immediately actionable, powering AI transformation roadmaps, workforce recomposition, precision hiring, and location planning. Fortune 500 enterprises already leverage Draup's structured datasets to drive critical workforce and capability decisions, making Draup the trusted partner for organizations seeking structured intelligence that delivers tangible business impact.

With access to a large-scale, validated dataset of U.S. job postings, enriched by Draup's proprietary skills taxonomy and AI-driven classification models, we are uniquely positioned to deliver the robust, accurate, and comprehensive job posting feed required by the ARM Institute. Our solution ensures:

- **Breadth and coverage** across all major job boards and employer sites, ensuring unique and active postings are captured in real time.
- **Alignment to ARM Institute frameworks** by mapping job postings directly to both the Robotics Competency Framework and the upcoming AI Competency Framework, including technical and essential soft skills.
- **Data integrity and quality assurance** through proven de-duplication, verification, and monitoring processes that distinguish active roles from expired ones.
- **Scalability and reliability** to support the ARM Institute's three-year contract horizon, ensuring daily transmission of job posting data with continuous refinement and support.

Through our existing partnership, Draup has been supporting ARM Institute power RoboticsCareer.org by supplying curated, validated, and competency-mapped robotics job postings. This collaboration has ensured that RoboticsCareer.org provides job seekers, education

providers, and employers with a reliable and continuously updated pipeline of robotics career opportunities, fulfilling its role as a national capability strengthening the U.S. manufacturing workforce.

In this new engagement, Draup proposes to expand this proven approach to include Artificial Intelligence competencies, delivering a structured, export-ready database of U.S.-based job postings that capture both robotics and AI-related roles, skills, and technologies. Our methodologies in large-scale job postings aggregation, skill taxonomy mapping, and quality assurance will ensure that the ARM Institute receives accurate, timely, and framework-aligned data that can directly inform workforce strategies and policy initiatives.

By partnering with the ARM Institute, Draup will deliver not only high-quality job posting data, but also the analytics foundation needed to better inform learners, workers, employers, and educators about the evolving robotics and AI talent landscape in U.S. manufacturing.

## **Objectives**

Draup's role in this engagement will be focused exclusively on serving as a **data provider** to strengthen the ARM Institute's national workforce capabilities by delivering a **comprehensive, structured, and competency-aligned dataset of U.S. job postings for Robotics and Artificial Intelligence competencies.**

Specifically, our objectives are to:

1. **Collect and deliver** a comprehensive and export-ready database of U.S.-based job postings across robotics and AI in manufacturing.
2. **Validate, classify, and normalize** job postings in accordance with the ARM Institute's Robotics and AI Competency Frameworks, ensuring postings are accurately mapped to both technical and soft skills.
3. **Provide transparency and reliability** by documenting job sources, taxonomy structures, and quality assurance processes.
4. **Ensure data freshness and continuity** by maintaining a daily updated feed that distinguishes between unique, active, and expired job postings.
5. **Enable ease of integration** by delivering the data in structured ARM Institute compatible formats (Parquet, JSON, CSV, or REST API output) suitable for ingestion into the roboticscareer.org platform without requiring Draup to manage or host the system.
6. **Expand Beyond Robotics to Robotics & AI Competencies:**  
Extend Draup's proven support for RoboticsCareer.org job postings to include Artificial Intelligence roles, skills, and technologies, mapped to ARM Institute competency frameworks.
7. **Enable Competency Mapping and Workforce Insights:**  
Classify job postings against robotics and AI competency frameworks to generate actionable intelligence on demand trends, required skills, and employer needs.
8. **Provide Nationwide Coverage of Job Postings**  
Aggregate job postings from diverse sources across all 50 U.S. states and territories, ensuring broad industry representation across manufacturing and adjacent sectors. All postings will be geographically tagged and mapped to Metropolitan Statistical Areas (MSAs) in alignment with the methodology used by the U.S. Census Bureau.

Through these objectives, Draup will act as a trusted data partner to the ARM Institute, supporting the expansion of roboticscareer.org and strengthening the national manufacturing workforce ecosystem.

## 1. Company Background

### Company overview

**Draup** is a leading AI-powered Talent Intelligence and Academic Intelligence platform, built to help enterprise leaders make smarter, faster decisions through deep data insights and intelligent workflow automation.

With real-time access to comprehensive talent ecosystem data and personalized insights, Draup empowers customers with actionable intelligence & recommendations to strengthen their talent strategies.

Founded by Vijay Swaminathan (CEO) and Vamsee Tirukkala (CCO)—visionaries with over 30 years of experience in global consulting and talent analytics—Draup is redefining how organizations approach workforce strategy, academic partnerships, and digital transformation. The founders previously built **Zinnov**, a global consulting firm, and **TalentNeuron**, the industry's first talent intelligence platform (acquired by CEB, a Gartner company).

### Talent Intelligence Platform

Draup specializes in delivering multidimensional data, insights and intelligence on the external labor market through a secure, cloud-based SaaS platform, custom consulting reports, and API integrations. Key areas of expertise include:

- Talent Supply & Demand
- Compensation Benchmarking
- Skills Architecture & Future Skills
- Diversity & Inclusion
- Reskilling & Workforce Planning
- Enterprises' Peer Insights
- Library of courses & certification relevant to Knowledge work

Powered by proprietary ML, AI, and GenAI models, Draup analyzes over **30 million+ data points** daily from **75,000+** global sources, enabling near real-time intelligence for HR and business leaders.

## **Academic Intelligence Platform**

Draup also delivers deep academic intelligence that goes far beyond traditional alumni directories. The platform continuously harvests and reconciles publicly available data from university websites, Google Scholar, ORCID, government/industry databases, and other repositories to build a living graph of:

- **54,000+** higher education institutions across 65 countries
- **192,000+** professor profiles
- Millions of academic program records

This enables enterprises to:

- Explore full catalogs of degrees, micro-credentials, and certificates mapped to job roles.
- Monitor graduate output to forecast future talent supply.
- Benchmark institutions using Draup's proprietary **Academic Research Index (0–10)**.
- Track university curriculum & research expertise in emerging areas like AI, robotics, IoT, computer vision, and sustainability.
- Identify thought leaders for collaboration, advisory roles, or speaking engagements.

Data & insights on universities are refreshed on a rolling schedule and accessible via dashboards and APIs, supporting use cases across **recruiting, workforce planning, R&D collaborations, and academic partnerships**.

## **Relevant Experience**

In this engagement, Draup proposes to serve as the data provider—responsible for identifying, collecting, validating, classifying, and delivering a comprehensive, export-ready database of U.S.-based robotics and AI job postings. These will be systematically mapped to the ARM Institute's Robotics and Artificial Intelligence Competency Frameworks, ensuring accuracy, relevance, and industry alignment.

Draup's Talent Intelligence capability underpins this effort. Over the past three years, Draup has been supporting ARM Institute by supplying curated robotics job postings for RoboticsCareer.org. During this collaboration, Draup has built ARM-exclusive systems and enhancements, such as:

### **1. Dedicated, Exclusive Capabilities for ARM**

We built and continue to operate a client-exclusive **Active Checking System** designed specifically for Arm. This system performs real-time validation to surface only accurate, up-to-date job listings and eliminate noise and duplication particularly critical for advanced technology and robotics-adjacent roles.

We also introduced **ARM exclusive benefits sections** in job postings, underscoring our commitment to pragmatic, custom enhancements aligned to Arm's business objectives.

## 2. High Reliability in Data Delivery

For the past **3 years**, Draup has maintained an approximately **98% success rate** in delivering structured job data. Delivery timelines have been consistently met even with custom periodic changes to custom formats and workflows positioning Draup as a **trusted system of record** for job intelligence.

## 3. Robotics-Specific Skill Intelligence

Responding to your robotics needs, we enriched Draup's skills ontology with a dedicated robotics and AI competency cluster, now supporting **70+ robotics/AI-specific skills**. We've implemented **custom skill-adjustment factors** to fine-tune role-to-skill matching, ensuring highly contextual and relevant listings for niche robotics and automation roles.

## 4. Roadmap Alignment with Draup Pre-Built Modules

Key elements of ARM Institute forward looking future roadmap are already supported by live Draup capabilities, enabling rapid customization often in less than a day:

- **University Intelligence:** Degrees, enrollment trends, institutions, professors, and skill/expertise alignments.
- **Reskilling Architecture:** Role transitions, skill-gap detection, learning paths, relevant resources available and certifications.
- **AI-Based Job Postings:** Coverage across **10+ AI/ML roles**, plus supporting metrics like Talent Supply ratios, Talent Cost, HDIs, Responsibilities, Career Transition insights, Talent Flow cuts, and more.

## 5. Format Continuity Through V2 Migration

As we migrate to a consolidated internal format in **V2**, Arm continues to receive job data in its **original preferred schema**, minimizing re-integration effort and downstream pipeline changes.

## 6. Strength in Scalable Data Acquisition

Draup's acquisition layer goes beyond ATS providers and aggregators. We source from official labor boards (e.g., **U.S. Department of Labor**), company career portals, and other channels that are often inaccessible to typical crawlers—expanding both **depth** and **diversity** of listings.

## 7. Long-Term, Specialized Engagement

Draup has been a strategic partner to Arm has been a strategic partner for **3+ years**. Over this period, we've delivered robotics-focused job intelligence aligned to Arm's talent strategy, closed data gaps, instituted feedback loops, and optimized delivery outcomes. We maintain a rapid turnaround—**often under one business day**—for addressing feedback, resolving issues, and implementing new requirements, ensuring continuity and meeting high expectations for precision and agility.

Over the past three years, Draup has supported the ARM Institute by supplying curated robotics postings for **RoboticsCareer.org**. During this collaboration, we implemented ARM-exclusive systems and enhancements:

- **Skills-Framework Aligned:** Each posting is classified against the ARM robotics and AI competencies, with explicit mappings to required knowledge, skills, and abilities. Draup's ontology has been enriched with **70+ robotics/AI skills**, and **custom adjustment factors** improve role-to-skill precision.
- **Benchmarking Advantage:** We map postings to **Metropolitan Statistical Areas (MSAs)**, enabling precise demand benchmarking against U.S. Census and other government datasets for robust geographic and economic insight.
- **Validated & Comprehensive:** The **Active Checking System** ensures postings are validated, deduplicated, and kept current. Broad sourcing across company portals, official labor boards, and aggregators yields a **credible, timely, and complete** dataset.
- **Export-Ready Delivery:** For **2.5+ years**, Draup has achieved ~98% structured delivery success and met all timelines. Data is delivered in **Parquet, CSV, JSON**, or database-ready formats for seamless integration with **RoboticsCareer.org** and other ARM workforce systems.

## **8. Volume & Cadence:**

- **V1: 2,000–2,500** active robotics postings on a **twice-weekly** schedule.
- **V2: 5,000–6,500** active robotics postings with **daily** delivery.  
Performance has remained on-schedule with minimal supply disruptions, adapting to evolving requirements.

## **9. Purpose-Built Relevance:**

Each posting is classified against the ARM **Robotics Competency Framework**, filtered to robotics/automation contexts (e.g., perception, controls, mechatronics, industrial robotics), and **geo-tagged at the MSA level**. With V1/V2 cadences, we deliver **active, de-duplicated** postings that minimize noise and provide **current, decision-ready demand signals** for hiring plans, employer outreach, and workforce planning.

Draup's established, robotics-focused foundation enables the ARM Institute to scale into emerging and evolving roles with **minimal transition effort** and **rapid customization**—often within a day while sustaining the reliability, precision, and depth that your workforce strategy requires. Building on this foundation, Draup is well-positioned to seamlessly extend into emerging and evolving roles in the robotics domain, ensuring reliable implementation, minimal transition effort, and sustained value delivery.

## 2. Project Methodology – Draup Job Posting Corpus

### 1. *Draup Data Sourcing*

Draup's data sourcing model is designed to create a comprehensive, structured, and continuously updated view of the global workforce related data including jobs demand data. Draup combines tech/AI-driven aggregation with human-in-the-loop validation to ensure coverage, accuracy, recency, and relevance. First step in this process for ARM involves systematic identification and collection of U.S.-based robotics and AI job postings from multiple public and proprietary sources. Draup leverages its Data Harvesters, specialized connectors, and sourcing partnerships to ensure comprehensive coverage.

Draup continuously monitors **45,000+** Large and reputed SMB enterprises' career websites and **400+** reputable job aggregator platforms such as LinkedIn, Indeed, Monster, CareerBuilder, and ZipRecruiter to ethically harvest job postings data. Draup's harvesters also extract postings from federal and state labor boards (including the U.S. Department of Labor's portals), ensuring compliance with government published listings.

To guarantee domain-specific depth, Draup specifically targets robotics and AI vendor job portals (e.g., FANUC, ABB, NVIDIA, Boston Dynamics) as well as industry specific job boards dedicated to advanced manufacturing and emerging technologies. These sources capture highly niche job postings that are often absent from generic job aggregators.

Draup's proprietary connectors enable real time synchronization with the identified data sources, automatically capturing new, updated, and reposted job postings. This ensures the demand dataset accurately reflects current labor market demand, minimizes lag between job creation and availability, and delivers comprehensive coverage across industries, enterprises, and geographies.

#### 1.1 Existing Draup Job Posting Corpus Leverage

For the ARM Institute, Draup will leverage its extensive Job Postings Corpus, which currently comprises over 750 million historical postings and actively processes 1.5 million live postings harvested on a regular basis from 45,000+ enterprise's career sites and 400+ aggregator platforms.

Draup's robust job postings corpus serves as the foundation for the proposed services, enabling ARM to realize immediate value without the need to wait for new harvesting cycles.

As part of the initialization, Draup will:

- **Filter to U.S.-based postings only** – Draup applies advanced geolocation filters and proprietary synonymization logic to ensure postings are accurately tied to U.S. cities,

states, and Metropolitan Statistical Areas (MSAs), guaranteeing compliance with ARM's national workforce scope. Draup's exclusive MSA synonymization model, built using advanced AI/Big Data frameworks and extensive human curation, ensures precise normalization of all location references by resolving inconsistencies and standardizing geographic data to the highest level of accuracy.

- **Re-map postings against ARM's Robotics & AI Competency Frameworks** – Each job description will be parsed and tagged to ARM-defined robotics and AI competencies using Draup's **ontology-driven ML models**, ensuring alignment between labor market demand signals and ARM's skill taxonomies.
- **Baseline Delivery** – ARM will receive a **comprehensive, export-ready dataset** of robotics and AI postings as the initial deliverable. This dataset represents a **point-in-time snapshot** of the U.S. labor market, providing an immediate layer of value and establishing a benchmark for trend analysis.

This baseline dataset creates a **starting reference point** for ARM. Draup will continuously enhance the baseline dataset through **incremental harvesting and refresh cycles**, ensuring the intelligence evolves in sync with market changes.

## 1.2 Incremental Data Harvesting

After the delivery of the baseline dataset, Draup's data pipeline will transition into **incremental harvesting mode** to capture and deliver newly available robotics and AI postings on a continuous basis. This process ensures that ARM always receives **up-to-date, comprehensive, and accurate labor market intelligence**.

### Process Overview:

- **Continuous Monitoring of Sources** – Draup's Data Harvesters and connectors are configured to **continuously monitor** thousands of large & SMB enterprises' career portals, aggregator platforms, and government labor boards for new or updated postings.
- **Targeted Robotics & AI Coverage** – Particular emphasis is placed on **robotics and AI-intensive employers** (e.g., robotics manufacturers, automation solution providers, AI/ML technology firms), this ensures that postings in these sources are captured immediately as they go live.
- **ATS & Job Board Feeds** – Draup integrates with **ATS providers, job aggregators, and niche industry boards** to identify and capture relevant postings that may not be published on company portals, thereby extending coverage using hard-to-source listings.

- **Vendor & Training Portals** – Incremental harvesting includes postings published by robotics OEMs and AI solution providers (e.g., FANUC, ABB, NVIDIA). Many of these organizations maintain dedicated job and training portals that are not indexed by traditional job boards.

### **1.3 Harvesting Process**

The harvesting process is the backbone of Draup's job postings pipeline. It combines **automated crawlers, connectors, and proprietary ML-driven frameworks** to systematically extract, clean, and structure job postings data. The process is designed for **repeatability, accuracy, and scalability**, ensuring ARM always receives complete and fresh job postings.

### **Sources Identification & Monitoring**

- Draup continuously monitors **45,000+ large and reputed SMB enterprises' career portals, 400+ job aggregator platforms, and government labor boards**.
- Niche sources, such as robotics OEMs (e.g., FANUC, ABB, KUKA) and AI vendors (e.g., NVIDIA, Microsoft AI, OpenAI ecosystem partners), are prioritized to capture domain-specific postings often absent from general job boards.
- Automated connectors are configured to detect **new, reposted and updated postings** in near real-time.

### **Data Extraction**

- Web crawlers and connectors extract **raw job descriptions (JDs)** including title, employer, job description, location, posting date, and application details.
- Structured metadata (e.g., job IDs, URLs, posting dates) is captured alongside unstructured text to allow robust downstream processing.

The harvester pipeline is designed for repeatable execution to enable daily refreshes with minimal reconfiguration.

## **2. Data Cleanup & Standardization**

Once sourced, raw postings pass through Draup's **structured ETL pipeline** (Extract, Transform, Load) designed to ensure accuracy, remove redundancy, and transform the data into a format that aligns with ARM's schema and quality requirements.

### **2.1 Translation**

- Draup begins by applying its **in-house language detection tool** across all harvested postings. While the majority of U.S. postings are in English, certain multinational employers occasionally publish postings in Spanish, Chinese, or other languages (particularly for bilingual roles).
- Detected non-English postings are automatically queued for Draup's **proprietary translation engine**, which is capable of translating up to **1 million postings per day** without performance degradation.
- The translation model retains **contextual accuracy**, ensuring that domain-specific terms like "robotic welding," "machine vision," or "AI pipeline" are not mistranslated into generic terms.
- The output is standardized into **English-only postings**, guaranteeing uniform downstream processing and schema consistency for ARM.

### **2.2 Synonymization**

- **Location Synonymization:** Draup's **exclusive MSA synonymization model**, built with high manual curation, ensures all geographic mentions are standardized. For example:
  - "Bay Area" → "San Francisco Bay Area, United States"
  - "Twin Cities" → "Greater Minneapolis-St. Paul Area, United States"This guarantees postings are consistently tied to **Metropolitan Statistical Areas (MSAs)**, enabling direct benchmarking against Census and BLS datasets.
- **Company Synonymization:** Employer names are often expressed differently across platforms (e.g., "IBM Corp.", "IBM US Careers", "International Business Machines"). Draup's synonymization model maps these variations into **canonical employer IDs**, ensuring postings are not split across duplicates.
- Together, synonymization ensures **location and employer data are normalized** for accurate analysis.

## 2.3 Deduplication

- Draup's **Job Posting Deduplication Model** is built on **cosine similarity** and applies a **40-day rolling window** of historical postings.
- This ensures duplicates are identified even when:
  - Job postings appear across multiple sources (LinkedIn, Indeed, company portal).
  - Posting dates differ slightly ( $\pm 20$  days).
  - Job titles or descriptions are formatted differently (e.g., "Robotics Engineer – Automation" vs. "Automation Engineer (Robotics)").
- This deduplication removes **~70% of raw duplicate records** across multiple sources which is sources for a day, improving dataset quality and reducing noise for ARM.

## 2.4 Consolidation

- In many cases, the same job is posted with **multiple locations** or duplicated across boards with only location differences.
- Draup's consolidation engine merges these into a **single canonical record** while retaining the **full list of associated locations**.
- Example: A job posted as "Robotics Technician – Detroit, MI" and "Robotics Technician – Cleveland, OH" becomes one unified posting with multiple MSA tags.
- This step eliminates an additional **~6% of redundancy** while enhancing **geographic coverage fidelity**.

## 2.5 Cleanup Filtering

- Draup's system applies automated rules to filter out incomplete or low-quality postings.
- Examples of filtered-out postings:
  - No job description text provided (e.g., "Apply for details").
  - Missing critical metadata (posting date, location).
  - Expired or "ghost" postings still visible on aggregator sites.
- On average, **~12% of raw harvested postings** are removed at this stage, leaving only high-quality, fully populated and relevant records.

## 2.6 Schema Alignment

- After cleaning, all valid postings are mapped into **ARM's standardized schema**.
- Schema mapping ensures consistency across critical fields such as:
  - **Job Title** – Lemmatized, normalized and classified into Draup's role taxonomy.
  - **Employer** – mapped via company synonymization into canonical IDs.
  - **Location** – aligned to U.S. MSAs with full geocode tagging.
  - **Posting Date** – standardized into ISO formats.
  - **Responsibilities, Skills, Certifications, Benefits** – structured into discrete fields.
- Each record undergoes **field-level validation checks** to ensure schema compliance before it is promoted into Draup's **centralized Job Corpus**, ready for enrichment and competency mapping.

## 3. *Data Enrichment & ARM Framework Competencies Mapping*

Once job postings are cleaned and standardized, Draup applies its proprietary ML/NLP enrichment pipelines to transform raw job descriptions into structured, actionable intelligence. This enrichment process ensures that every posting is tagged with job roles, skills, functions, workloads, tools, and geographic intelligence, and then systematically aligned to the ARM Institute's Robotics and AI Competency Frameworks.

### 3.1 Core Data Enrichment Layers

Draup applies multiple enrichment layers to extract comprehensive labor market intelligence from each posting:

- **Job Roles** – Each posting is classified against Draup's 2,800-role library, ensuring precision in role identification. For example, “Automation Engineer – Robotics” is mapped distinctly from “AI Engineer – Computer Vision.” This eliminates ambiguity and ensures postings are linked to role clusters relevant to ARM.
- **Skills** – Job descriptions are parsed using Draup's 20,000+ core skills and 250+ soft skills libraries. Skills are extracted through Named Entity Recognition (NER) models, which capture both explicit mentions (e.g., “Python, TensorFlow”) and implicit role-based skills (e.g., “troubleshooting robotic welding arms”).

- **Business Functions** – Postings are mapped to one of Draup’s 22 business functions such as Research & Development, IT, HR, Manufacturing, Sales, and more. This provides ARM with context on where robotics and AI demand is concentrated functionally within organizations.
- **Functional Workloads** – Each posting is mapped to one of Draup’s 220+ functional workloads spanning 34 industries. For example, postings may be tagged under workloads like “Artificial Intelligence & Data Science,” “Embedded Firmware Engineering,” “Robotic Process Automation,” or “Manufacturing Engineering.” This helps ARM analyze demand at a workload level, beyond generic job titles.
- **Digital Stacks & Tools** – Draup leverages its 50,000+ digital products library to identify tools, platforms, and technologies mentioned in postings (e.g., ROS, Keras, FANUC controllers, NVIDIA Jetson). This enables ARM to understand technology-specific demand within robotics and AI.
- **Experience Requirements** – NLP models extract explicit experience statements (e.g., “5+ years of robotics automation,” “2 years of AI/ML model deployment”), ensuring ARM can segment demand by seniority or entry-level roles.
- **Responsibilities** – Using Draup’s JD Segregation Model, job descriptions are partitioned into three sections:
  - About the Company – organizational context.
  - About the Job – role responsibilities and scope.
  - About the Candidate – qualifications, skills, and requirements.
 This structured segregation allows more accurate competency and role mapping.
- **Compensation & Benefits** – Draup’s models capture compensation ranges (e.g., “\$90,000–\$120,000 per year”), payment terms, and listed benefits (e.g., healthcare, relocation support). This provides an additional layer of workforce intelligence for benchmarking job attractiveness.
- **Certifications** – Professional certifications mentioned in postings (e.g., AWS ML Specialty, OSHA Robotics Safety, PMP for automation projects) are extracted and structured for competency validation.
- **Job Families** – Each role is grouped into a broader family of related occupations (e.g., Robotics Technicians → Robotics Support & Maintenance family). This enables ARM to conduct demand benchmarking and cross-role workforce planning at a family level.

### **3.2 ARM Competency Tagging**

Once job roles, skills, and attributes are extracted through Draup's enrichment pipelines, they are **systematically aligned** with the ARM Institute's **Robotics and Artificial Intelligence Competency Frameworks**. This step ensures that every posting delivered to ARM is not only cleaned and structured, but also **translated into ARM's standardized language of competencies**, making it actionable for talent demand analysis.

#### **Process Detail:**

##### **1. Skill-to-Competency Mapping**

Draup's **ontology-driven ML models** map extracted skills (e.g., Python, robot calibration, machine vision, ROS, reinforcement learning) to ARM's defined competency categories.

Skills are grouped into **Knowledge, Skills, and Abilities (KSAs)** as outlined by ARM's framework, ensuring alignment at the most granular level.

#### Example:

- “Robot Programming (Fanuc/KUKA/ABB)” → Robotics Technician → ARM Competencies: *Programming, Equipment Setup, Troubleshooting*.
- “TensorFlow, PyTorch, Model Deployment” → AI Engineer → ARM Competencies: *Algorithm Development, ML Training, Data Processing*.

##### **2. Role-to-Competency Alignment**

Each job role extracted (from Draup's 2,800+ role library) is **crosswalked to ARM's occupational roles**.

#### Example:

- Draup Role: “Robotics Integrator” → ARM Role: Robotics Systems Integrator → ARM Competency Set: *System Integration, Industrial Safety, Robot Programming*.
- Draup Role: “AI Data Technician” → ARM Role: AI Technician → ARM Competency Set: *Data Collection, Data Labeling, Model Input Preparation. (Based on Customization required)*

### 3. Technical & Soft Skill Integration:

Draup integrates both **hard technical skills** (robotics engineering, computer vision, automation controls, ML pipelines) and **soft skills** (teamwork, communication, critical thinking) into the competency mapping. This ensures postings reflect **holistic employer demand**, not just technical requirements.

### 4. Competency Granularity Across Career Stages:

Postings are further segmented by **experience level** (e.g., entry-level technician vs. senior AI researcher). Draup maps these levels to ARM's **career pathway competencies**, enabling ARM to see how demand distributes across early-career, mid-career, and advanced roles.

### 5. Framework Compatibility & Future-Proofing

Draup's mapping process is **framework-agnostic**, meaning it can accommodate updates to ARM's Robotics and AI frameworks or expand to new competency sets with minimal setup. The process is built for **scalability**, allowing ARM to extend competency mapping beyond robotics and AI into adjacent domains (e.g., quantum computing, advanced manufacturing).

**Skills in Robotics Manufacturing:** Draup has classified the core skills and soft skills that are required for the Robotics roles in manufacturing along with the emerging skills in the market



Job Title	Core Skills				Soft Skills
	Design Knowledge	Digital electronics and microprocessors	PLC (Programmable Logic Controllers)	Troubleshooting	
Robotics Technician	Safety Precautions	Electronics & Control	Program Maintenance	Programming	Persistence
	Hand-Eye Coordination	Supervision			Decision Making
	Advanced Robot Programming	Inspection	Quality Assurance	Tools Management	Problem Solving
Robotics Specialist	Knowledge on Sensors	Risk Assessment	Robotic Motion Planning	Installation	Time Management
	Safety Assessment				Logical Thinking
	Robotic Visualization Tools	Big Data	Simulation Software	Data Modelling & Analysis	Attention to Detail
Robotics Integrator	Mixed Reality Technologies	Operating Systems	Programming Languages	Sensors / Hydraulics / Pneumatics	Active Listening
	Customization	Configuration			Team Player

Emerging Skills
PLM Software
SIEMENS
3D Modelling
SOLIDWORKS
Operating Systems
Linux
UNIX
Programming Languages
VBA
R
G
python
Project Management Tools
Jira
Trello

Source: ARM Institute. DRAUP's proprietary talent module  
Note: Roles listed are not exhaustive. The above taxonomy excludes seniority-level prefixes (such as Senior, Lead, Principal, and others) of the mentioned unique titles to showcase individual roles.

## **4. Data Validation & Quality Control**

Our validation process includes **both automated and manual checks**:

### **4.1 Automated Validation Rules**

Draup's automated validation engine operates as part of the ETL pipeline and applies structured checks at every record level. This ensures that only complete, accurate, and schema-compliant postings move forward into the final dataset.

- **Field Completeness** – Mandatory fields (Job Title, Employer, Posting Date, Location, Description) checked; incomplete records removed.
- **Format Verification** – Dates standardized (ISO 8601),
- **Taxonomy Alignment** – Titles, skills, locations, job type, workloads and functions mapped to controlled vocabularies (e.g., “RPA - Specialist” → “Robotic Process Automation”).
- **URL Checks** – Broken or expired links excluded, ensuring only active postings are retained.
- **Duplication Flags** – Potential duplicates flagged before deduplication to improve accuracy on the delivery
- **Consistency Rules** – Cross-field logic applied (e.g., “Remote” cannot be tagged as “Onsite”)
- URL verification (program links must resolve and return expected HTTP status)
- Location standardization using Draup’s Locations Taxonomy Framework
- Credential/format validation against ARM’s allowed values

### **4.2 Manual SME Review**

While Draup’s automated validation ensures postings meet schema, consistency, and completeness requirements at scale, certain attributes require **expert judgment** to confirm accuracy. Draup therefore integrates a **Subject Matter Expert (SME) manual review process** as a second layer of quality control.

#### **Scope of SME Review:**

SMEs are trained researchers and domain specialists who manually validate a **representative sample of postings** from each source category (company portals, aggregators, labor boards, niche robotics/AI job sites). This ensures data fidelity across multiple perspectives.

#### **Focus Areas for Manual Review:**

##### **1. Competency Mapping Accuracy**

- SMEs verify that extracted roles and skills are correctly mapped to ARM’s Robotics & AI Competency Frameworks.

- Example: A “Controls Engineer” posting requiring PLC programming, safety compliance, and robotics troubleshooting should map to Robotics Systems competencies, not generic “Engineering.”
- Edge cases (multi-role postings, hybrid robotics/AI jobs) are resolved through SME confirmation.

## 2. Skill Extraction & Context Validation

- Automated NLP can misclassify ambiguous skills (e.g., “R” as a programming language vs. “R&D”).
- SMEs review skill-tagged outputs for correctness and adjust mappings where automation may be contextually weak.

## 3. Experience, Salary & Benefits Checks

- SMEs validate that extracted experience requirements and salary ranges are accurately interpreted.
- Example: “3–5 years in robotics automation” should not be normalized as “5 years minimum.”
- Compensation parsing (e.g., “\$90K base + bonus”) is verified to ensure ARM receives precise labor market cost insights.

## 4. Employer & Location Classification

- SMEs confirm that company and location synonymization has been applied correctly.
- Example: “Alphabet DeepMind – Mountain View” should consolidate under **Google, San Jose–Sunnyvale–Santa Clara, CA MSA**.

## 5. Classification of Job Families & Functions

- SMEs review a subset of postings to ensure correct grouping into **job families, business functions, and workloads**.
- This step reduces false positives in workforce demand benchmarking.

## 6. Quality of Job Description Parsing

- Draup’s JD Segregation Model separates postings into *About Company / About Job / About Candidate*.
- SMEs confirm that boundary detection is working properly, especially for postings with non-standard formatting.

### Review Methodology:

- **Sampling Strategy:** Typically 5–10% of total postings per batch are manually validated, with higher sampling rates for new sources or newly added job categories.

- **Feedback Loops:** Errors or anomalies identified during SME review are flagged, corrected, and fed back into Draup's ML models to continuously improve automation accuracy.
- **Rapid Turnaround:** SME validations are integrated into Draup's delivery cycle with **<1 business day turnaround** for high-priority corrections, ensuring no disruption to ARM's data flows.

## 5. Data Delivery & Update Cycle

### 5.1 Export Formats

- Delivered in Parquet, PostgreSQL/MySQL dumps, CSV, or JSON, fully aligned to ARM's schema. Additional supported formats include Parquet, Avro, and TSV for system compatibility.
- All fields structured exactly to ARM's schema.

### 5.2 Job Posting Status Tracking

- Each program marked as **New**, **Updated**, **No Change**, or **Removed** for each data refresh cycle.

### 5.3 Daily Refresh Process

- Through the daily data refresh feed the older JDs will be removed and the new JDs ingested to keep the data up to date for ARM competencies

### 5.4 Architecture Overview of Draup Data Feed

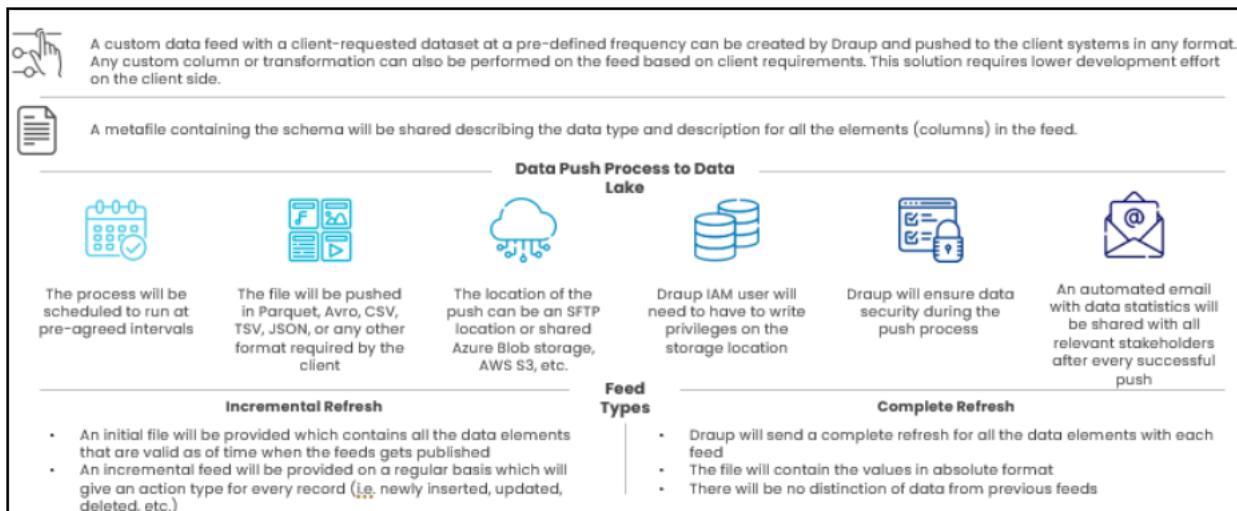


Fig 1: Draup Data Feed Integration

## **6. Scalability & Future Framework Support**

- **Framework-Agnostic Design**

Draup's mapping architecture is **highly modular**, allowing ARM to integrate new Robotics and AI competency frameworks or even frameworks from adjacent domains without redesigning the pipeline. More than **20+ additional frameworks** can be supported by simply configuring mapping rules, ensuring long-term adaptability as ARM evolves.

- **Taxonomies Expansion**

Draup currently tracks **3,000+ job roles and 20,000+ skills across 30+ occupational categories**, spanning engineering, telecom, IT, life sciences, logistics, and more. This breadth makes it easy to extend coverage beyond robotics and AI into **adjacent high-demand industries** such as advanced manufacturing, aerospace, automotive automation, and AI-driven healthcare. By updating source lists and mapping rules, Draup can rapidly onboard **new sectors and role types** into ARM's intelligence framework.

- **Job Posting Corpus Growth**

With a baseline of **1.5M+ active postings harvested daily**, Draup's system is designed to **scale linearly** as sources and occupations increase. The harvesting infrastructure can incorporate **new job boards, employer portals, and labor datasets** with minimal setup, ensuring ARM always receives a **comprehensive and expanding dataset**.

- **Skill & Competency Evolution**

Labor market demand is dynamic, especially in robotics and AI. Draup's enrichment models are continuously updated to recognize **emerging tools, programming languages, and platforms** (e.g., ROS2, Generative AI tools, digital twins). This ensures postings are mapped not only to today's competencies but also to **future-ready skills**, giving ARM a forward-looking edge.

- **Future-Ready Benchmarking**

All postings are mapped to **MSAs**, making it straightforward to benchmark demand against **Census, BLS, and future government-published datasets**. This scalability ensures ARM can continue to produce **national, regional, and occupation-level workforce insights** with unmatched accuracy, even as federal frameworks or geographic taxonomies expand.

### **3. Technology Stack and tools**

Our Harvester Framework delivers a comprehensive, enterprise-grade solution for automated web data collection and processing, capable of handling **billions of data points per month** and processing ~ **50 million requests per day**. The platform combines cutting-edge AI automation with robust engineering architecture to provide end-to-end data harvesting capabilities for various domains of data like Job descriptions, Profiles, Courses, Salary, Products, etc.

**The solution is architected across three core components:**

1. **Intelligent Configuration & Discovery** - AI-powered automation with human oversight for source discovery and configuration
2. **Robust Data Harvesting & Processing** - Enterprise-grade distributed architecture for scalable data extraction and validation
3. **Comprehensive Observability & Quality Assurance** - Advanced monitoring, analytics, and quality validation framework

Our platform leverages modern cloud-native technologies including LLM integration, distributed processing on AWS infrastructure, and advanced analytics capabilities. The system ensures reliable data delivery with **99.9 % uptime**, fault tolerance, scalability, and operational excellence.

#### **Component 1: Intelligent Configuration & Discovery**

**Our intelligent automation layer leverages advanced AI and agentic workflows to automatically discover, analyze, and configure data sources with human supervision for critical validations.** This component eliminates manual setup overhead by using LLM-powered analysis to understand website structures and automatically generating comprehensive URL sets for career pages, ATS systems, and job aggregators while incorporating manual supervision for critical websites and complex extraction strategies.

- **LLM-Powered Analysis with Human Oversight:** Automatically analyses website structures and content patterns using advanced NLP techniques, with manual supervision for critical website configurations and complex extraction strategies
- **Intelligent Web Discovery:** Performs automated web search and crawling to discover comprehensive URL sets, validates endpoints, and establishes source-specific access patterns with human review for mission-critical sources
- **Agentic Configuration Workflows:** Orchestrates end-to-end setup processes using autonomous agents for URL building, metadata extraction, schema generation, and configuration optimization with manual intervention capabilities

- **Input Metadata Generation:** Automatically extracts and structures website metadata including page parameters, filters, pagination logic, and content schemas with expert validation for complex architectures

**Tech Stack:** Temporal Workflow Engine, Apache Airflow, PostgreSQL, SERP, Playwright, LLM and NLP

### **Component 2: Robust Data Harvesting & Processing Architecture**

**Our enterprise-grade harvesting engine processes intelligent configurations through a distributed, fault-tolerant architecture designed for high-volume data extraction on AWS cloud infrastructure.** The system orchestrates sophisticated data collection workflows across multiple priority queues with intelligent rate limiting, comprehensive proxy management, and advanced duplicate detection, ensuring all extracted data undergoes rigorous schema validation before MongoDB storage.

- **Producer-Consumer Architecture:** Input producers generate extraction tasks based on intelligent configurations while specialized workers process high-priority listing and standard detail extraction operations across distributed queues
- **Intelligent Distribution & Proxy Management:** Custom distribution service manages per-source rate limiting, priority-based task allocation, proxy rotation, and fair resource distribution to maximize throughput while preventing blocking
- **Advanced Extraction with Duplicate URL Detection:** Combines HTTP clients for standard requests with headless browsers for JavaScript-heavy content, includes anti-bot protection through proxy systems, and implements intelligent duplicate URL checker
- **Data Validation & Storage:** Multi-layered data validation against generated schemas, comprehensive duplicate URL detection across sources, and structured persistence in MongoDB with automated quality assurance

**Tech Stack:** Celery, Redis, Playwright, FlareSolvver, Scrapoxy, MongoDB, OpenSearch, AWS, EKS

### **Component 3: Big Data & Machine Learning Stack**

Draup's Big Data & ML stack powers the **scalable enrichment, classification, and continuous improvement** of harvested job postings. It is designed to handle **millions of postings daily**, integrate seamlessly with ARM's frameworks, and ensure **future-ready intelligence** through advanced orchestration, automation, and AI capabilities.

### 3.1 Data Lake & Big Data Infrastructure

- **Cloud-Native Storage:** AWS S3 + Lake Formation as the central data lake for structured and unstructured postings
- **Distributed Processing:** Apache Spark, Dask, and Presto for high-volume batch + stream workloads.
- **Real-Time Streaming:** Apache Kafka for high-throughput ingestion pipelines and near real-time updates.
- **Schema Evolution & Reliability:** Delta Lake ensures ACID transactions, version control, and rollback support.
- **Scalable Query Engine:** Trino + PrestoSQL for interactive queries across terabyte-scale job datasets.

### 3.2 Workflow Orchestration & Automation

- **DAG Scheduling:** Apache Airflow and Temporal orchestrate multi-step ETL/ML workflows, ensuring dependency management and recovery from failures.
- **Continuous Integration / Deployment:** Jenkins and GitHub Actions manage ML model deployment, schema migrations, and automated testing pipelines.
- **Monitoring:** Superset dashboards provide interactive visibility into workflow progress, source coverage, and harvesting KPIs.

### 3.3 Machine Learning Pipelines

- **Feature Store:** Feast maintains reusable features across models (e.g., location embeddings, employer clusters, skill frequency weights).
- **ML Lifecycle Management:** Kubeflow and MLflow for model training, experiment tracking, and version control.
- **Model Training & Inference:** TensorFlow, PyTorch, and AWS SageMaker for scalable training of classification, NER, and embeddings models.
- **Real-Time Inference:** ML APIs deployed on Kubernetes (AWS EKS) deliver low-latency role, skill, and competency tagging at ingestion scale.

### 3.4 AI/LLM Integration

- **Job Parsing with LLMs:** GPT-based models and HuggingFace transformers segment postings into structured fields (responsibilities, skills, benefits).
- **Domain-Specific Embeddings:** FAISS, Pinecone, and Weaviate store vector embeddings for semantic skill-role matching.

- **Context-Aware Extraction:** NER models fine-tuned for robotics/AI terminology capture niche competencies (e.g., ROS2, Human-Robot Collaboration).
- **Active Learning Loops:** SME feedback feeds back into model retraining, ensuring continuous improvement.

### 3.5 Proprietary Enrichment Services

- **Location Synonymization:** Draup's proprietary **MSA Synonymization Model** maps cities, states, and colloquial terms (e.g., "Silicon Valley") into standardized U.S. Metropolitan Statistical Areas.
- **Company Synonymization:** Built-in entity resolution service consolidates employer references (e.g., "IBM Corp.", "IBM Careers") into canonical IDs.
- **Ontology Integration:** Roles, skills, and tools mapped against Draup's libraries (3,000+ roles, 20,000+ skills, 50,000+ digital products).

### 3.6 DevOps & MLOps Infrastructure

- **Containerization & Orchestration:** Docker + Kubernetes (EKS) ensure workloads are portable, scalable, and self-healing.
- **Continuous Integration / Deployment (CI/CD):** Jenkins pipelines automate deployment of harvesting jobs, ML models, and schema updates.
- **Infrastructure as Code (IaC):** Terraform + Helm manage reproducible AWS deployments.
- **Monitoring & Logging:** Prometheus, Grafana, and ELK (Elasticsearch/Logstash/Kibana) stack for real-time observability.
- **Testing & Load Simulation:** Locust and JMeter for stress testing job pipelines (~50M requests/day).
- **Security & Compliance:** IAM roles, Secrets Manager, VPC isolation, and automated vulnerability scans ensure data protection and compliance.
- **MLOps:** Automated retraining pipelines triggered via Airflow; model drift detection ensures postings remain mapped correctly as competencies evolve.

### 3.7 Analytics & Forecasting Layer

- **Dashboards & BI:** Superset + Grafana provide interactive dashboards with filters for roles, MSAs, skills, and employers.
- **Graph Analytics:** Neo4j models competency-role-skill relationships to identify career pathways and adjacency clusters.
- **Forecasting & Trends:** PySpark + Prophet + TensorFlow models project demand growth by geography, skill, and job family.

## **Value for ARM:**

- **Enterprise-Grade Orchestration:** Airflow, Jenkins, and Superset ensure repeatability, reliability, and visibility.
- **Future-Ready ML:** AI/LLM-driven enrichment with active learning continuously adapts to new robotics/AI terminology.
- **Scalable Analytics:** Capable of processing billions of records, with real-time dashboards and predictive forecasting for job intelligence.

**Techstack:** Apache Airflow, Temporal, Jenkins, GitHub Actions, Apache Kafka, Apache Spark, Dask, Presto, Trino, PrestoSQL, Delta Lake, AWS S3, AWS Lake Formation, AWS SageMaker, Docker, Kubernetes (AWS EKS), Feast, Kubeflow, MLflow, TensorFlow, PyTorch, Hugging Face Transformers, GPT, FAISS, Pinecone, Weaviate, Superset, Prometheus, Grafana, Elasticsearch, Logstash, Kibana, Locust, JMeter, Neo4j, PySpark, AWS Secrets Manager, Amazon VPC

## **Component 4: Comprehensive Observability & Quality Assurance**

**Our operational excellence layer provides complete visibility into system performance, data quality, and business metrics through advanced monitoring and validation frameworks.** This component ensures reliable operations through real-time performance tracking, automated data coverage analysis, comprehensive quality validation, and intuitive dashboards using Superset with Trino for actionable insights into harvesting effectiveness across all configured sources.

- **Real-Time Performance Monitoring:** Tracks system health, processing rates, success ratios, and resource utilization across all harvesting components with custom metrics for queue performance, worker efficiency, and AWS infrastructure
- **Data Coverage & Quality Validation:** Automated analysis of data completeness, schema compliance, freshness metrics, and coverage gaps with intelligent alerts for data quality issues and extraction anomalies
- **Advanced Dashboarding & Visualization:** Interactive dashboards using Superset powered by Trino for high-performance analytics, displaying harvesting metrics, source performance comparisons, and data volume trends with drill-down capabilities
- **Centralized Logging & Alerting:** Comprehensive log aggregation from all system components with intelligent alerting for failures, performance degradation, and data quality issues, plus configurable notification rules

**Tech Stack:** Prometheus, Grafana, Superset, Trino, Elasticsearch, Alert Management Systems, FastAPI, AWS EKS, Locust (Performance Testing)

## 4. Project Timeline, including major milestones.

**Project Start Date:** Project kick-off immediately after issuance of PO.

**Total Duration:** ~4-5 weeks for initial delivery, followed by ongoing weekly/daily refresh cycles.

Steps	Timeline	Milestones	Deliverables
<b>Step 1 – Project Initiation &amp; AI Scope Delta</b>	<b>Weeks 1–2</b>	1) Kick-off with ARM 2) Confirm AI roles/skills & mapping to ARM's AI competencies 3) Finalize filter configuration to include AI jobs alongside Robotics	<ul style="list-style-type: none"> <li>Finalized AI filter configuration (roles, skills, titles, keywords)</li> <li>ARM schema confirmation (no change)</li> <li>Source-to-field &amp; mapping addendum</li> </ul>
<b>Step 2 – Initial AI Data Sourcing (Robotics feed continues)</b>	<b>Weeks 2–3</b>	1) Turn on AI filters in live harvesters 2) Begin priority AI-source harvesting (company portals, aggregators, labs) 3) Weekly ingestion & schema alignment in staging 4) First-pass deduplication & normalization for initial AI batch	<ul style="list-style-type: none"> <li>Interim AI coverage report (sources, postings, MSAs)</li> </ul>
<b>Step 3 – Consolidation &amp; Enrichment (AI + Robotics)</b>	<b>Weeks 2–3</b>	1) Merge AI postings into existing robotics corpus 2) Apply ARM Robotics + AI competency mapping 3) Complete normalization (titles, skills, MSAs, employers)	<ul style="list-style-type: none"> <li>Enriched, competency-mapped AI &amp; Robotics dataset in staging</li> <li>Mapping summary (AI roles ↔ ARM competencies)</li> </ul>
<b>Step 4 – Validation &amp; Quality Control</b>	<b>Weeks 3–4</b>	1) Automated validation completed (schema, links, MSAs, taxonomy) 2) SME review on AI-heavy samples (skills, salaries, certifications) 3) Corrections applied & re-validated	<ul style="list-style-type: none"> <li>Validation report (coverage %, error rates, fixes)</li> <li>Clean, export-ready AI &amp; Robotics dataset in staging</li> </ul>
<b>Step 5 – Delivery &amp; Handover</b>	<b>Week 5</b>	1) Final export in ARM-compatible formats (PostgreSQL/MySQL, CSV, JSON) 2) Documentation pack (sources, filters, methodology, validation) 3) Walkthrough for ARM team (data structure & update cadence)	<ul style="list-style-type: none"> <li>Final ARM-ready AI+Robotics dataset</li> <li>Full documentation package</li> <li>Runbook for ongoing updates</li> </ul>
<b>Step 6 – Ongoing Updates (Robotics &amp; AI)</b>	<b>Post-Go-Live</b>	1) Automated refresh cycles (harvest → dedupe → enrich → validate) 2) Change-log delivery (New, Updated, No Change, Removed) 3) Continuous filter tuning for AI signals	<ul style="list-style-type: none"> <li>Refreshed datasets per cadence</li> <li>Change-log &amp; coverage reports</li> </ul>

## 5. Customer Service / Maintenance Plan

### Customer Service

Draup proposes the following multi-layer staffing plan, with all key personnel committed for the duration of the onboard and active partnership period.

These key personnel will be supported by multiple resources across Draup teams (support, SME, data analysts, engineering). Our team members will be available during ARM's business hours and accessible via video conferencing, emails, phone or in-person meetings as needed.

1. **Customer Success Manager (CSM):** ARM will be assigned a named Customer Success Manager who acts as the strategic point of contact and the first level for any escalations. The CSM ensures alignment with your goals, needs, helps in driving product adoption, shares best practices, ensures timely issue resolution and facilitates Quarterly Business Reviews (QBRs) to track value for your investments.
2. **Delivery Lead (DL):** A dedicated, named Delivery Lead will oversee all aspects of execution and deliverables, ensuring optimal platform performance, high data quality, and the timely rollout of features and enhancements. The DL will also deliver custom reports in alignment with the defined scope by ARM stakeholders. The Delivery Lead (DL) will collaborate closely with the Customer Success Manager (CSM), Customer Support Teams and Customer Teams (ARM users and key stakeholders) to proactively resolve delivery blockers, manage ARM's evolving priorities, and ensure seamless project coordination.
3. **Customer Support Team (CST):** The Customer Support Team will serve as the primary point of contact for addressing ARM's inquiries, troubleshooting issues, tracking support tickets, and providing timely resolutions to ensure seamless user & delivery experience. Operating in close coordination with the Customer Success Manager (CSM), and Delivery Lead (DL), the CST team will manage support tickets, track recurring issues, and escalate critical matters when necessary. Leveraging robust knowledge bases and best practices, the CST team will not only resolve technical and functional challenges but also proactively share insights to enhance platform performance, user satisfaction, and overall customer success.

Draup provides technical & functional support to customers. Functional support addresses specific queries related to Draup data, taxonomy, data formats, data updates, and Data/ML models and analysis. Technical support focuses on data support and delivery queries within the defined project scope, ensuring smooth and reliable access to the required datasets and deliverables.

We provide 24x5 multi-channel support via an integrated ticketing platform, email, and phone. Our support team is distributed across USA and India, enabling continuous coverage with SLA driven response times during business days.

**Service Maintenance:**

Draup maintains a comprehensive maintenance program to ensure the availability, accuracy, reliability, and timeliness of the data we are committing to deliver to ARM.

Our delivery and support teams will work in close collaboration with ARM team to ensure uninterrupted data service. In the event of planned maintenance, ARM team will be notified in advance, with clear timelines and contingency measures to prevent service or delivery disruption.

Our service maintenance plan includes:

**1. Continuous Data & Quality Monitoring**

Draup will perform comprehensive data integrity checks against multiple data sources leveraged by our platform. These checks will combine automated and manual validation processes to detect and address data anomalies, inconsistencies, and obsolete records, ensuring the highest standards of accuracy and reliability.

**2. Regular Data Updates**

Draup maintains scheduled refresh cycles for each data attribute based on relevance, source refresh cycles and priority. High-priority data sets are updated in real-time or near-real-time to ensure maximum accuracy and relevance.

**3. Enhancements**

Draup deploys new capabilities and data improvements without disrupting ongoing data delivery, ensuring continuity of service while maintaining transparent communication with customers through release notes, impact assessments and value of enhancements.

**4. Customer Feedback**

Draup will maintain direct feedback loops with ARM team (users and key stakeholders) to identify evolving requirements and rapidly incorporate corrections or enhancements based on ARM team's inputs.

## 6. Cost Proposal

Sr. No.	Item & Description	Duration	Actual Price (USD)	*Proposed Pricing (USD)	
1	<b>Weekly/Daily Dataset Refreshes</b>  Weekly/Daily re-run of the processes from new data acquisition, enrichment, classification & delivery	Year 1	\$110,000	\$58,000	
		Year 2	\$110,000	\$58,000	
		Year 3	\$110,000	\$58,000	
2	<b>Filter Configuration for Inclusion of AI Jobs:</b>  Filter Configuration for Inclusion of AI Jobs - Add AI job filters, competency mapping, and integrate new classification logic into the existing robotics setup; includes QA and deployment.	One Time	\$60,000	\$10,000	
	<b>Total Discounted Cost in USD</b>				<b>\$184,000</b>

In recognition of Draup's **long-standing partnership** with ARM and the strategic nature of our engagement, we are pleased to propose a significantly discounted subscription cost to support both the ARM Institute and the broader community it serves.

## 7. Sample Outputs

### 7.1 Data Dictionary

attribute_name	attribute_definition	draup_models
data_lake_id	A unique identifier assigned to each job posting data within a data lake or storage system.	
job_title_lemmatised	The lemmatized version of the job title, normalized for analysis, search, and matching.	
translated_job_title	The job title translated from its original language to enable understanding across regions.	
translated_job_description	The job description translated from its original language for clarity and accessibility.	
source	The platform or medium from which the job posting was captured or extracted.	
language	The language in which the job posting is written.	
publication_date	The date when the job posting was published on the source platform.	
date_posted	The timestamp when the job was originally posted on the source platform.	
location	The location mentioned in the job posting (city, region, etc.).	
synon_location	The standardized/synonymized representation of the source location.	
state	State/administrative area of the job location (when applicable).	
msa	The Metropolitan Statistical Area in which the job was published, if applicable.	
latitude	The latitude coordinate of the job location.	
longitude	The longitude coordinate of the job location.	
country_name	The country where the job is located.	
region_name	The region/state/province within the country where the job is located.	
company	The hiring company's name as published with the job posting.	
url	The URL of the source platform where the job posting is published.	
date_updated	The timestamp when the job posting record was last updated in the database.	
mvp_company_name	The synonymized company name from Draup's universe section.	
core_skills	Core/technical skills required or mentioned in the job posting.	Skills Tagging Model
soft_skills	Interpersonal/soft skills required or mentioned in the job posting.	Skills Tagging Model
about_candidate	Details about the ideal candidate as mentioned in the job description.	JD Segregation Model
about_job	Details about the role, responsibilities, and requirements.	JD Segregation Model
about_company	Details about the hiring organization mentioned in the JD.	JD Segregation Model
digital_products	Digital products/technologies mentioned in the JD.	Techstack Extraction Model
job_role_list	List of Draup job roles relevant to the posting.	Job Role Tagging Model
certification	Certifications required from candidates as mentioned in the JD.	Certification Tagging Model
education	Educational qualifications required from candidates.	Education Extraction Model
final_responsibilities_sentences	Responsibilities of the job as extracted from the JD.	Responsibilities Extraction Model
diversity	Diversity specifications mentioned in the job posting.	Diversity Extraction Model
minimum_index_experience	The minimum experience required for the job.	Experience Extraction Model
job_type	Job type including time-based (full-time, part-time) and location-type (hybrid, onsite, remote).	Job Type Extraction Model
business_function	Business function of the job (e.g., Sales, IT, Marketing, HR).	BF extraction Model
functional_workload	Functional workload extracted from the JD, if explicitly mentioned.	FW extraction Model
min_salary	Minimum salary for the role, if specified.	Salary Extraction Model
max_salary	Maximum salary for the role, if specified.	Salary Extraction Model
pay_rate	Salary pay frequency/rate as mentioned in the posting.	Salary Extraction Model
currency	Currency in which salary is specified.	Salary Extraction Model
extracted_benefits	Benefits/perks extracted from the JD.	Benefits Extraction Model
job_poster_name	Name of the person or entity publishing the job.	
job_poster_title	Title/designation of the job poster.	
raw_title	Original, unaltered job title from the source.	
raw_description	Original, unaltered job description from the source.	
industry_name	Industry of the job posting, derived based on the associated company.	

The complete dictionary has been provided in the attached Excel file,

***Job\_Postings\_Data\_Dictionary\_Draup.xlsx***

## 7.2 Sample Record in JSON format

Please refer to the attachment **sample\_output.txt** provided along with the proposal.

### Sample (reference)

Robotic Job Postings.txt

```
1 [ {  
2   "about_candidate": "Qualifications., • Bachelor's or Master's degree in Electrical Engineering, Robotics, Mechatronics, or related field., • Proven experience in hardware  
3   design and development for robotics or unmanned vehicles., • Proficiency in CAD software and PCB design tools., • Strong understanding of electronic components, circuit  
4   design, and signal processing., • Experience with sensors, actuators, and communication protocols (e.g., UART, SPI, I2C)., • Ability to work in a fast-paced, collaborative  
5   environment., • Excellent problem-solving skills and attention to detail., • Strong written and verbal communication skills., • Eligibility to obtain and maintain a U.S.  
6   security clearance., • Willing and able to travel on an as needed basis to remote test sites throughout the United States., Preferred Qualifications., • Experience with rapid  
7   prototyping and iterative design processes., • Knowledge of embedded systems and firmware development., • Familiarity with regulatory standards and compliance for unmanned  
8   vehicles., • Hands-on experience with testing and debugging hardware systems.",  
9   "about_company": "Anduril Industries is a defense technology company with a mission to transform U.S. and allied military capabilities with advanced technology. By bringing  
10  the expertise, technology, and business model of the 21st century's most innovative companies to the defense industry, Anduril is changing how military systems are designed,  
11  built and sold. Anduril's family of systems is powered by Lattice OS, an AI-powered operating system that turns thousands of data streams into a real-time, 3D command and  
12  control center. As the world enters an era of strategic competition, Anduril is committed to bringing cutting-edge autonomy, AI, computer vision, sensor fusion, and networking  
13  technology to the military in months, not years.",  
14   "about_job": "We are seeking a highly skilled and motivated Robotics Hardware Engineer to join our Rapid Development Team. In this role, you will be responsible for designing,  
15  developing, and integrating electronic components for unmanned vehicles. You will work closely with cross-functional teams to rapidly prototype and iterate on hardware  
16  solutions., Design and develop hardware components for unmanned vehicles, including sensors, actuators, and control systems., Integrate electronic components and ensure  
17  seamless communication between hardware and software systems., Collaborate with mechanical, software, and systems engineers to prototype and test new hardware solutions.,  
18  Conduct feasibility studies, simulations, and performance evaluations to optimize hardware designs., Develop and execute test plans to validate hardware functionality and  
19  reliability., Identify and troubleshoot hardware issues, implementing effective solutions., Stay current with industry trends and emerging technologies to drive innovation.,  
20  Document design processes, test results, and maintain detailed engineering records.",  
21   "benefits": [  
22     "Relocation Assistance",  
23     "Retirement Benefit",  
24     "Dental Insurance",  
25     "Employee Assistance Program",  
26     "Extra Paid Holidays",  
27     "Professional Training",  
28     "Transport Assistance",  
29     "Vision Insurance",  
30     "Life Insurance",  
31     "Other Benefits",  
32     "Legal Insurance",  
33     "Free Food",  
34     "Disability Insurance",  
35     "Health Insurance"  
36   ],  
37   "city": "Costa Mesa",  
38   "core_skill": [  
39     "Control Systems Design",  
40     "Signal Processing",  
41     "Firmware",  
42     "Feasibility Studies",  
43     "Embedded System",  
44     "Rapid Prototyping",  
45     "Actuator",  
46     "Communication Protocols",  
47     "Prototyping",  
48     "Debugging",  
49     "Printed Circuit Board (PCB) Design".  
50 }
```

Robotic Job Postings.txt

```
51 benefits and resources to provide support from planning to parenting.<br/>• Mental Health Resources: We provide free mental health resources 24/7 including therapy, life  
52 coaching, and more. Additional work-life services, such as free legal and financial support, available to you as well.<br/>• A professional development stipend is available to  
53 all Andurilians.<br/>• Daily Meals and Provisions: For many of our offices this means breakfast, lunch and fully stocked micro-kitchens.<br/>• Company-funded commuter benefits  
54 available based on your region.<br/>• Relocation assistance (depending on role eligibility).<br/>• 401(k) retirement savings plan - both a traditional and Roth 401(k). (US  
roles only)<br/><br/>The recruiter assigned to this role can share more information about the specific compensation and benefit details associated with this role during the  
hiring process.<br/><br/>Anduril is an equal-opportunity employer committed to creating a diverse and inclusive workplace. The Anduril team is made up of incredibly talented  
and unique individuals, who together are disrupting industry norms by creating new paths towards the future of defense technology. All qualified applicants will be treated  
with respect and receive equal consideration for employment without regard to race, color, creed, religion, sex, gender identity, sexual orientation, national origin,  
disability, uniform service, Veteran status, age, or any other protected characteristic per federal, state, or local law, including those with a criminal history, in a manner  
consistent with the requirements of applicable state and local laws, including the CA Fair Chance Initiative for Hiring Ordinance. We actively encourage members of recognized  
minorities, women, Veterans, and those with disabilities to apply, and we work to create a welcoming and supportive environment for all applicants throughout the interview  
process. If you are someone passionate about working on problems that have a real-world impact, we'd love to hear from you!<br/><br/>To view Anduril's candidate data privacy  
policy, please visit https://anduril.com/applicant-privacy-notice/.  
55   "job_level": "None",  
56   "job_posting_url": "https://www.linkedin.com/jobs/view/senior-robotics-hardware-engineer-rapid-development-at-anduril-industries-4273987991",  
57   "job_responsibilities": [  
58     "We are seeking a highly skilled and motivated Robotics Hardware Engineer to join our Rapid Development Team. In this role, you will be responsible for designing,  
59     developing, and integrating electronic components for unmanned vehicles. You will work closely with cross-functional teams to rapidly prototype and iterate on hardware  
60     solutions., Design and develop hardware components for unmanned vehicles, including sensors, actuators, and control systems., Integrate electronic components and ensure  
61     seamless communication between hardware and software systems., Collaborate with mechanical, software, and systems engineers to prototype and test new hardware solutions.,  
62     Conduct feasibility studies, simulations, and performance evaluations to optimize hardware designs., Develop and execute test plans to validate hardware functionality and  
63     reliability., Identify and troubleshoot hardware issues, implementing effective solutions., Stay current with industry trends and emerging technologies to drive innovation.,  
64     Document design processes, test results, and maintain detailed engineering records."  
65   ],  
66   "job_title": "Senior Robotics Hardware Engineer - Rapid Development",  
67   "job_update_status": 1,  
68   "maximum_salary": 180000,  
69   "minimum_salary": 120000,  
70   "minimum_experience": "Proven experience in hardware design and development for robotics or unmanned vehicles",  
71   "normalized_title": "Robotics Specialist",  
72   "pay_currency": "USD",  
73   "soft_skill": [  
74     "Evaluation",  
75     "Innovation",  
76     "Attention to Detail",  
77     "Collaboration",  
78     "Conversational Proficiency"  
79   ],  
80   "source_company_name": "Anduril Industries",  
81   "source_location": "Costa Mesa, CA",  
82   "state": "California",  
83   "job_schedule": null,  
84   "zip_code": null  
85 } ]
```

### **7.3 Relevant Draup Customer Case Studies:**

#### **Case Study 1 — Major National Technology Industry Association (Job Postings Demand Atlas)**

##### **Context**

A leading national tech trade association needed a standardized, nationwide view of IT job posting demand by role, skill, and geography to inform policy and workforce programs.

##### **Challenge**

Fragmented postings across boards/portals, heavy cross-posting noise, inconsistent titles/skills, and no MSA-standardized view to benchmark against Country Census.

##### **Draup Approach**

- Built a National Job Postings Demand Atlas: deduplicated and MSA-mapped postings across the countries.
- Normalized titles to a role taxonomy; extracted skills/tools and mapped to NOS competencies.
- Delivered demand dashboards (role/skill/employer intensity, wage bands, trend lines) with quarterly change logs.

##### **Outcomes**

- Clear, benchmarkable MSA-level demand signal across IT roles.
- Evidence-backed policy briefs and targeted program investments by region and competency.
- Faster provider engagement using real employer demand data.

##### **Relevance to ARM**

Directly parallels ARM's need: deduped, MSA-ready job postings mapped to Digital Role (Including AI & Robotics) competencies for national workforce planning.

## Case Study 2 – Global Staffing Organization

### Context

A global staffing company, required **granular, MSA-by-MSA job-posting demand** intelligence for global talent hotspots to route recruiters, prioritize requisitions, and advise enterprise clients across business functions hiring (e.g., **blue-collar jobs, networking/IT infrastructure, manufacturing, software, automation, and AI/ML** ).

### Challenge

- Cross-posted and stale listings obscured true demand.
- Inconsistent titles/skills (e.g., “Automation Tech” vs. “Robotics Technician”).
- Non-standard location labels limited **MSA** benchmarking especially for developing countries and non-English speaking geographies.
- “AI-washed” postings inflated AI demand estimates.

### Draup Approach

- **Dedup & Consolidation:** Cosine-similarity deduplication (40-day window) to collapse cross-posts into canonical records.
- **MSA Mapping:** Proprietary synonymization normalizing locations
- **Taxonomy Normalization:** Standardized titles/skills across blue-collar, networking, manufacturing, robotics, automation, and AI/ML role families; competency tagging.
- **AI Precision Filters:** Embedding-based filters to isolate genuine AI/ML roles.
- **Demand Analytics:** Dashboards for role/skill/employer intensity, wage bands, seasonality, and recruiter routing by MSA; quarterly change logs.

### Outcomes

- **Clean, benchmarkable demand signal** at the MSA level across all target domains.
- Faster recruiter routing to **high-yield markets**; improved time-to-fill.
- Sharper client advisories on where to hire and what skills to target, with **competency-aligned** shortlists.

### Relevance to ARM

Proves Draup’s ability to deliver **MSA-standardized, competency-mapped job-posting demand analysis** across diverse occupations—directly applicable to ARM’s Robotics + AI scope, with minimal additional setup.

## 8. References

1. **Johnson Controls, JCI** – Skills Assessment and Reskilling Assets  
**Contact Name:** Rajesh Ahuja  
**Email:** [rajesh.ahuja@jci.com](mailto:rajesh.ahuja@jci.com)
2. **NASSCOM**: Developed national competency standards for IT occupations, mapped aligned learning curricula, and certified training providers.  
**Contact Name:** Dr. Sandhya Chintala  
**Email:** [sandhya@nasscom.in](mailto:sandhya@nasscom.in)
3. **Intel**: Upskilling Strategies to transition Technicians to Foundry Talent Pool in United States.  
**Contact Name:** Vibhu Ganesan  
**Email:** [vibhu.ganesan@intel.com](mailto:vibhu.ganesan@intel.com)
4. **NASDAQ**: Developed a comprehensive Career Portal for 10k+ Nasdaq Employees to identify viable career paths based on Draup's Reskilling Propensity Index framework  
**Contact Name:** Karoline Johanna Raets  
**Email:** [karoline.raets@nasdaq.com](mailto:karoline.raets@nasdaq.com)
5. **Adecco**: Developed comprehensive demand estimates across job occupations and benchmarked job demand insights across 28 major countries.  
**Contact Name:** Richard Rapple  
**Email:** [Richard.Rapple@adeccogroup.com](mailto:Richard.Rapple@adeccogroup.com)

## **9. Litigation**

Draup confirms that, since our incorporation in 2021, we have not been involved in any litigation. Furthermore, we are not aware of any pending or threatened legal proceedings that could materially impact on our business operations, financial position, or our ability to fulfill the obligations and deliverables outlined in this proposal.

We maintain robust ethical, legal and compliance frameworks designed to proactively identify, assess, and address potential legal risks before they escalate into material issues. These include regular reviews by our internal legal team, contractual compliance monitoring, and periodic engagement with external legal counsel to ensure our activities remain within applicable laws and regulations (across the globe).

We recognize that our customers rely on Draup not only for the quality of our solutions and services but also for the security and integrity of our operations. We are committed to maintaining transparency and will promptly communicate any future developments, including the emergence of litigation or legal proceedings that could reasonably be expected to affect service delivery or contractual commitments as defined in this proposal. Such communication would include an assessment of potential impacts, and a clear outline of mitigation measures being implemented to ensure uninterrupted service and fulfillment of our commitments to ARM.