

Arabic-language data annotation, benchmarked against industry standards.

In a four-week structured pilot, Karama Data delivered 3,031 annotation tasks across three of the most in-demand task types for AI model training — matching or exceeding published international benchmarks on accuracy and inter-annotator agreement.

3,031

TASKS COMPLETED
IN 4 WEEKS

100%

TASK COMPLETION
RATE

~220–250

TASKS / HOUR
PER ANNOTATOR

18,800+

PROJECTED WEEKLY
CAPACITY AT SCALE

QUALITY RESULTS VS. PUBLISHED BENCHMARKS

TASK TYPE	ACCURACY	INDUSTRY STANDARD	INTER-ANNOTATOR AGREEMENT	RESULT
Preference Ranking	88.55%	83–87% (major AI labs)	κ 0.62 vs. 0.27–0.39 published	Above benchmark
Dialect Identification	71.17%	65–80% (NADI 2024)	κ 0.57 vs. 0.59 (Palestinian Arabic)	On par
Sentiment Tagging	62.53%	60–75% (AraSenTi / ASAD)	κ 0.53 vs. 0.55–0.70 published	Near target

WHAT WE DELIVER

- Native-Arabic annotation across preference ranking (RLHF), dialect identification, and sentiment tagging
- Ground-truth accuracy validation and inter-annotator agreement reporting on every engagement
- Documented, repeatable SOP refined through structured calibration and quality review
- Rapid onboarding — pilot-tested training and guidelines ready to scale to new annotators

SCALING CAPACITY

TEAM SIZE	TASKS / DAY	TASKS / WEEK
2 annotators (pilot)	~3,760	~18,800
5 annotators	~9,400	~47,000
10 annotators	~18,800	~94,000

- Purpose-built impact.** Karama Data trains and employs Palestinian women in Gaza to deliver high-quality Arabic-language annotation for AI companies — pairing rigorous quality standards with meaningful economic opportunity.