

About Predii

Predii is an Enterprise AI Software Company based in Palo Alto, CA and Pune, India. Our specialized, validated AI platform has been purpose-built to extract predictive and prescriptive insights from unstructured textual, sensor, and procedural data in the automotive aftermarket and service business. Predii's patented AI engine is currently processing 2+ billion historical repair jobs monthly. Our 8+ years NLP and domain expertise enable industry leading companies such as Snap-on, Epicor, Mercedes-Benz, and Valvoline to leverage previously unused data to power predictive solutions, increase aftersales revenue, drive product innovation, and support data-driven decision-making strategies.

Predii has been recognized by Gartner, ABI Research, and the Industrial IoT Solutions World Congress for our focus in Applied AI in repair and maintenance

About this Role: AI Evaluation Engineer

As an **AI Evaluation Engineer** at Predii, you will evaluate and validate AI/ML systems operating on complex automotive data, including component mapping, RAG-based Q&A, and feature extraction from repair orders, catalogs, free text, and technical documents.

This is an **engineering-focused role** involving the design of custom evaluation scripts, datasets, and automated pipelines (e.g., LLM-as-a-judge) to measure quality, detect regressions, and gate releases. Predii provides foundational automotive training, but strong ownership in building domain intuition and high-quality evaluation datasets is essential.

Key Responsibilities

- Evaluate ML and LLM outputs using defined metrics and benchmarks
- Create, curate, and own evaluation datasets and golden test cases
- Analyze results to identify trends, inconsistencies, and regressions
- Execute evaluation-driven smoke and regression tests pre-release
- Track quality metrics and provide go/no-go release signals
- Validate AI services at the API level for correctness and stability
- Monitor performance, latency, and error rates under production load
- Collaborate with ML, backend, and product teams to define and ensure expected AI behavior



Bachelor's Degree/
Master's Degree



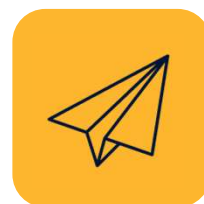
2-3 years experience



Competitive Salary



Pune, India



Apply now.

Basic Qualifications

- Bachelor's/Master's in Computer Science or related fields or PhD in NLP/ML/LLM (one who have submitted their thesis is eligible to apply)
- 2-3+ years of professional experience in production environment. For PhD holders, working experience in a relevant lab setting coupled with top-tier publications (A/A*) would waive the professional experience requirement

- Identify patterns, determine root causes, and implement effective solutions for system challenges
- Strong Python skills for evaluation scripting and data analysis
- Experience working with ML, NLP, or data-driven systems
- Ability to interpret probabilistic outputs and ambiguous correctness
- Experience designing or executing structured evaluations or test cases
- Strong analytical mindset and attention to detail
- Willingness to develop automotive domain intuition through hands-on data work and training

Additional Qualifications

- Experience evaluating LLM or RAG-based systems
- Familiarity with AI evaluation frameworks or custom pipelines
- Experience with API testing (Postman, pytest)
- Experience with performance/load testing tools (JMeter, k6)
- Exposure to automotive, industrial, or equipment data

We're looking forward to receiving your resume and application at jobs@predii.com. For more information about Predii, please visit www.predii.com.