

Subject: Internship Project Proposal

<i>Code Name</i>	PTA_EN_54580_28/06/2026 20.08.38
<i>Date</i>	28/06/2026 20.08.38
<i>Target</i>	Bachelor's Degree - Data Analysis

Host Institution

<i>Host Institution</i>	Kenja
<i>Protocol</i>	54580
<i>Protocol Date</i>	11/04/2025
<i>Country</i>	Japan
<i>City</i>	Tokyo
<i>Address</i>	Moto Akasaka 1-7-10, Minato Ku, Tokyo, Japan, 1-7-10
<i>Website</i>	https://kenja.com/about/
<i>Employees Number</i>	50
<i>Contant Person</i>	Tanisha Gandhi, Head of Process Automation
<i>Phone Number</i>	+1 5083108621/+919711602002
<i>Email</i>	tanisha.gandhi@kenja.com

Project Supervisor

<i>Name and Surname</i>	Tanisha Gandhi
<i>Phone Number</i>	+15083108621/+919711602002
<i>Email</i>	tanisha.gandhi@kenja.com

Internship Project Details

<i>Title</i>	Automated Data Extraction and Validation for Multilingual Document Processing Systems
<i>Detailed Description:</i> Context: The internship takes place at Kenja K.K. within a project developed for Robert Walters Japan. The project focuses on building an AI-driven, automated resume conversion tool that parses unstructured Western-style CVs and structures/translate them into highly standardized Japanese professional formats (Rirekisho and Shokumu Keirekisho). The intern operates within the Quality Assurance (QA) and Data Validation team under direct managerial supervision. Measurable Objectives: Data Accuracy Verification : Validate the precision of the Natural Language Processing (NLP) models in extracting key-value pairs (e.g., dates, job titles, technical skills) from raw text data, aiming to maintain an error rate below a specific project threshold. Structural Consistency Mapping : Test and document the data pipelines to ensure that unstructured inputs are consistently mapped into the rigid, localized database schemas required for Japanese layout generation. Bug Tracking & Reporting : Identify, classify, and log data parsing anomalies and formatting bugs in the tracking system, reducing the cycle time between bug detection and developer resolution. Regression Testing : Execute systematic QA test cases on updated algorithm deployments to ensure model updates do not degrade existing data extraction performance. Tools to be Used: Kenja Enterprise Collaboration Platform (Data workflows and secure environment)	



Bug tracking and project management software.
Data comparison tools, spreadsheet software (Excel/Google Sheets) for error-logging matrices, and basic database query tools.

<i>Topics</i>	Data Validation, Quality Assurance Engineering, and Multilingual Text Parsing Systems.
<i>Reimbursement of Expenses (YES/NO)</i>	NO
<i>Refund Amount</i>	
<i>Availability for Travel (YES/NO)</i>	NO
<i>Kind of employment</i>	Full time
<i>Duration in months (max 12)</i>	2 months
<i>Duration in hours</i>	150
<i>Internship Date Start</i>	03/08/2026
<i>Internship Date End</i>	03/10/2026
<i>Number of Open Position(s)</i>	10

Internship Skills

Required Skills: Understanding of data structures, databases, and structured schemas. Basic concepts of Quality Assurance (QA) methodologies, test-case execution, and regression testing. Familiarity with data tracking, error analysis, and building structured validation reports. Awareness of Natural Language Processing (NLP) data extraction concepts (Named Entity Recognition).

<i>Other Skills</i>	Attention to Detail : Precision in identifying minor discrepancies between raw data inputs and structured outputs. Cross-Cultural & Remote Collaboration : Ability to work effectively within an international, remote team structure (Italy/Japan time zone dynamics). Problem-Solving & Analytical Thinking : Methodical approach to identifying patterns in data failures or software bugs. Clear Technical Communication : Reporting anomalies concisely to development teams and the line manager.