Terms of Reference: Machine Learning Engineer

New Producers for Sustainable Energy (NP4SE) is looking for a Machine Learning Engineer to prepare the data for, and test the accuracy of, two AI tools funded by Energy Corps. The role is part-time: about 65 days over 12 months, at a day rate of USD 300–400, depending on experience.

Background

NP4SE is a charity that supports governments in 25 emerging oil and gas producing countries. With Energy Corps funding, we are building two tools to help small government teams oversee their resources with better information.

The AI Advisory Desk draws on 14 years of NP4SE peer exchanges and expert analysis. It is retrieval-based: it answers from a closed, expert-weighted set of documents and does not train a model on them. Answers on high-risk topics go to a human expert before they reach the user.

The emissions dashboard helps a national regulator check methane and greenhouse gas data reported by operators. It is cloud-hosted and, once tested, run by the government agency.

The Machine Learning Engineer is responsible for the quality of the data going into both tools and for measuring whether their outputs are accurate.

Scope of work

AI Advisory Desk (about 20 days)

- Clean and structure the foundation archive of about 500 documents, including text extraction from scanned files, and tag each with consistent metadata (maturity stage, resource type, discovery phase).
- Run the transcription pipeline for practitioner exchanges recorded as audio or video, and check the transcripts before they enter the document set.
- Apply the expert-weighting scheme so that senior practitioner material takes precedence in retrieval.
- Design and run the accuracy tests: build a set of test questions from historical regulatory scenarios, record results, and review hallucination logs, reporting findings to the Lead Subject Matter Expert.

- Agree written accuracy criteria with the Lead Subject Matter Expert for the escalation protocol.
- Support the pilot with five partner countries, including logging errors and feedback.

Emissions dashboard (about 45 days)

- Build and maintain the uniform data coding maps so that operator submissions follow one reporting structure.
- Process unstructured field logs and paper records into that structure, with the Technical Data Intern.
- Review operator submissions for anomalies and flag them to the regulator's team.
- Compare reported emissions against satellite and other open data sources, and agree the data interfaces with the AI Architect without taking part in the build.
- Document the data pipeline so that the agency's staff can run it after handover.

Across both tools

- Keep data handling consistent with NP4SE's confidentiality commitments: no data from one country is visible to another, and no document leaves the closed environment.
- Test the tools independently of the people who build them, and report results plainly, including failures, to the Lead Subject Matter Expert.

The MLE does not do the strategic impact tracking. That belongs to the MEL Officer, so that performance reporting stays independent of the people building the tools.

Deliverables

Dates are indicative and will be fixed in the consultancy agreement.

Deliverable	Tool	Timing
Cleaned and tagged document set, with a short data note	Advisory Desk	Within the first 2 months
Test question set and accuracy results, with the criteria agreed in writing	Advisory Desk	Before the end of January 2027
Input to the verified pilot evaluation report (Milestone 1)	Advisory Desk	End of January 2027

Deliverable	Tool	Timing
Uniform data coding maps, version 1	Emissions dashboard	Within the first 4 months
Processed field logs and anomaly review notes	Emissions dashboard	Monthly during live testing
Data pipeline documentation for the agency	Emissions dashboard	Before handover
Brief handover note on both tools	Both	End of contract

Profile

We are looking for someone who has built and tested retrieval-based AI systems in a practical setting, ideally for a development organisation, a multilateral body or a government.

Essential:

- Several years of hands-on machine learning or data engineering experience, including retrieval-augmented generation (RAG) systems and ways of testing their accuracy.
- Experience cleaning messy, mixed-format data: scanned documents, spreadsheets, transcripts.
- Python, and comfort with open-source tools such as Dify and Whisper.
- Careful about data protection. Has worked with confidential or government data and understands why it must stay in a closed environment.
- Able to explain results, including errors, in plain language to non-technical colleagues.
- Fluent in English. Able to work remotely with a small team across time zones.

Useful:

- Familiarity with oil, gas, energy or climate data, including emissions reporting.
- Experience working with ministries or regulators in emerging economies.
- French or Portuguese, since the dashboard is meant to be replicated in French- and Portuguese-speaking countries.

Terms

- **Time and fee:** about 65 days between appointment and September 2027 (about 20 for the Advisory Desk and 45 for the dashboard), at a day rate of USD 300-400 depending on experience. Days are called off by email against deliverables.
- **Location:** remote, with occasional travel only if agreed in advance.
- **Supervision:** the Lead Subject Matter Expert directs and reviews the work. The Executive Director manages the contract.
- **Engagement:** consultancy agreement with NP4SE, including confidentiality terms.
- **Ownership of work:** NP4SE will publish its methodology so that others can replicate it. The agreement will say what that means for code and data.

How to apply

Send a CV, a short note on your relevant experience and the names of two referees to contact@newproducersgroup.org by 27 October. We will interview shortlisted candidates and take up references. We aim to appoint by 1 November 2026, as the first milestone is due in early 2027.