PROPOSAL TITLE:

Enhancing Intelligence Source Performance Management Through Two-Stage Stochastic Programming and Machine Learning Techniques

THEME:

Governance & Public Policy

A. PROBLEM STATEMENT

Intelligence agencies play a central role in safeguarding national security, but their effectiveness largely depends on the performance and reliability of their human sources. In practice, managing intelligence sources remains one of the least optimised and most subjective parts of intelligence work. Many agencies rely on qualitative evaluations or senior analyst judgment to assess a source's credibility, productivity or potential deception. While these judgements are valuable, they often lack systematic validation and are influenced by human bias, incomplete information or changing operational contexts.

This subjectivity introduces significant risks. Poor tasking decisions can result in resource wastage, compromised missions or the spread of inaccurate intelligence. Moreover, the absence of quantifiable performance models makes measuring improvement, maintaining accountability or adapting to uncertainty difficult. The lack of analytical tools also limits the institutional memory of agencies; when experienced officers retire or rotate, their tacit evaluation knowledge often disappears with them.

At a broader governance level, poor management of intelligence sources weakens state resilience and transparency. It leads to operational inefficiencies, delays in detecting threats, and sometimes, policy decisions based on unreliable intelligence. In contexts like East Africa, where intelligence agencies face a diverse and rapidly changing security environment, from terrorism to maritime crime, the need for data-driven decision-making systems has become urgent.

Therefore, the challenge is to create a robust framework that can handle uncertainty, utilise historical performance data, and support decision-makers with objective, reproducible insights. Tackling this problem directly enhances national governance and public accountability by aligning intelligence operations with modern data analytics and optimisation practices.

B. PROPOSED SOLUTION

This project presents a Two-Stage Stochastic Programming (TSSP) and Machine Learning (ML) framework aimed at improving the performance management of intelligence sources. The solution combines predictive modelling, optimisation under uncertainty, and

decision support visualisation to help agencies allocate tasks more efficiently, evaluate reliability, and detect deception or underperformance early.

1. Conceptual Framework

The proposed system operates in two interconnected stages:

a) Stage One: Predictive Modelling (Machine Learning Layer)

Using historical performance, reliability, and behavioural data from intelligence operations, ML algorithms such as XGBoost, Support Vector Machines (SVM), Linear Regression, and Keras GRU will predict each source's expected reliability and likelihood of deception. These models will be trained on quantitative and categorical variables (e.g., task outcomes, contextual conditions, reporting frequency and consistency metrics). The model evaluation will emphasise interpretability, so decision-makers can understand why a source is rated high- or low-performing.

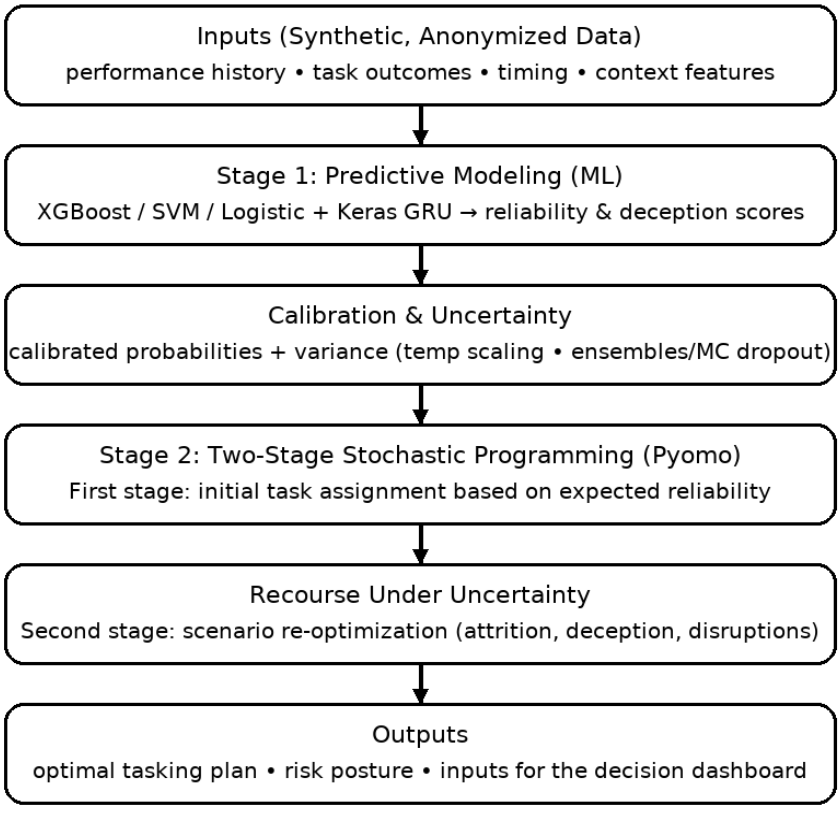

Figure 1: Conceptual Framework for the two-stage system

The first stage employs machine learning to assess source reliability and potential deception using performance and context features, delivering outputs that are easy to interpret. Stage two utilises Two-Stage Stochastic Programming to allocate and reallocate tasks under uncertainty, aiming to reduce operational risk while maximising expected reliability and mission coverage. (The Prototype was validated on synthetic data.)

b) Stage Two: Optimisation (Stochastic Programming Layer)

The predicted performance scores will feed into a Two-Stage Stochastic Programming (TSSP) model that allocates intelligence tasks under uncertainty. The first stage will determine the initial assignment of sources to missions based on predicted reliability. In contrast, the second stage will re-optimise these decisions under simulated uncertainty scenarios (e.g., source withdrawal, false reporting, environmental disruptions). The objective function minimises operational risk while maximising expected reliability and coverage of critical intelligence tasks.

2. AI–Driven Decision Support

The decision-support layer functions as the operational interface that translates model outputs into decisions. It offers an interactive dashboard integrating three components: predicted source reliability, optimal task allocation under uncertainty, and risk alerts.

First, machine learning models assign each source a continuously updated reliability score. This score reflects accuracy, consistency, and past task performance. Managers can observe how a source's reliability evolves instead of relying on static or memory-based assessments.

Second, the system displays tasking recommendations from the two-stage stochastic programming model. It shows which sources should be assigned to which missions and why. Users can also run "What if" scenarios (for example, if a high-value source becomes unavailable or operational priority shifts) and see how the model would reassign coverage to maintain mission reliability.

Third, the dashboard issues early-warning flags when it detects performance anomalies, sudden drops in reliability, or reporting patterns that may indicate deception or fatigue. Each alert is paired with an explanation using interpretable AI methods, such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanation (LIME), so supervisors can see why the system is raising concern instead of treating it as a black box.

All recommendations and decisions are logged with supporting evidence. This creates an auditable record for accountability, after-action review, and policy oversight.

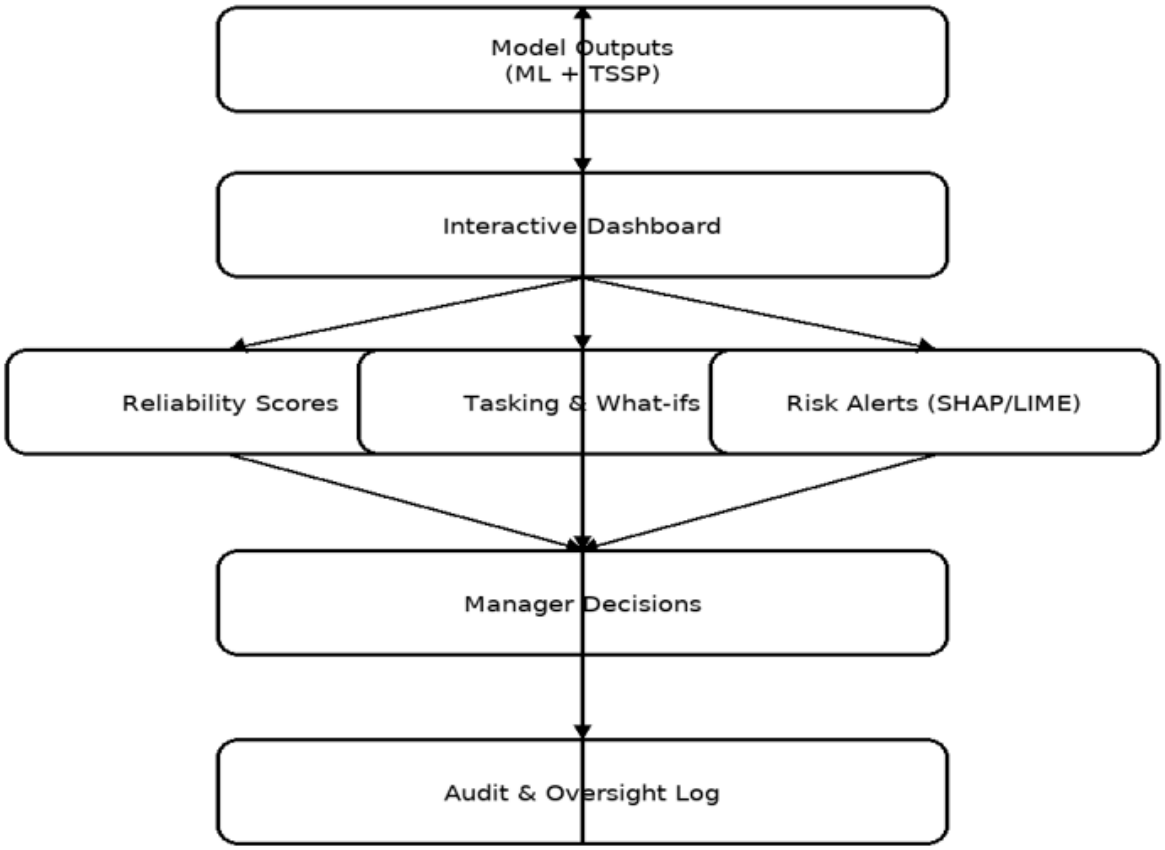

Figure 2: AI-Driven Decision Support Workflow

Model outputs (ML and TSSP) feed into an interactive dashboard with three components: reliability scores, tasking and what-ifs, and risk alerts with SHAP and LIME. Outputs inform manager decisions, and an audit and oversight log provides feedback to update the models. (*The Prototype uses synthetic data only.)*

3. Methodology: CRISP-DM

a) Business Understanding

We seek to improve performance management (tasking, reliability assessment, and risk control) under uncertainty. The success criteria are ML performance (F1, ROC-AUC, precision-recall, calibration) and optimisation quality (mission coverage, expected risk, expected value, or the stochastic solution). All experiments use synthetic, non-identifiable data for methodological validation only.

b) Data understanding

We generate synthetic datasets according to the agreed schema with controlled noise, missingness and drift. We standardise categorical, numeric and time features and log each transformation for reproducibility. From these data, we construct reliability indicators (for example, success ratio and deviation from expected performance) and derive deception or fatigue flags from injected inconsistencies. We use time-based splits to preserve sequence effects and fix random seeds to make results repeatable.

c) Modelling

We use a hybrid stack: transparent tabular baselines plus a compact neural model for short-term dynamics. We train baseline predictive models (Logistic Regression, XGBoost and SVM) in Python using scikit-learn and XGBoost to estimate source reliability and task suitability. We apply k-fold or time series cross-validation and perform probability calibration.

We add a compact TensorFlow (Keras) GRU model to capture non-linear interactions and recent behaviour. Inputs are standardised numeric features, embedding for categorical variables, and a fixed-length event window per source (for example, the last 10 tasks with outcomes and timing deltas). A single GRU (or 1D temporal CNN) encodes the window, which we fuse with static features and pass through two dense layers to output a reliability score between 0 and 1. We train with Adam (lr 1e-3 with cosine decay), early stopping on validation PR-AUC, and apply class weights or focal loss if imbalance persists. Probabilities are calibrated via temperature scaling or isotonic regression. Uncertainty is obtained from deep ensembles (five seeds) or MC dropout at inference.

The calibrated probabilities from the selected predictor (best among LR, XGBoost, SVM, GRU) feed the two-stage stochastic program in Pyomo, solved with GLPK. The optimiser uses the mean reliability as the expected value and the variance or quantiles to define second-stage scenarios, enabling stress tests such as spikes in deception or sudden source attrition.

d) Evaluation

For the ML layer, evaluation will be based on report accuracy, F1 Score, Area Under the Receiver Operating Characteristics Curve (ROC-AUC), precision-recall, calibration error, and stability across synthetic time slices and subgroups. For the optimisation layer, we report mission coverage, expected risk, robustness under scenario stress tests, and the expected value of the stochastic solution (EVSS). We conduct “what-if” analyses that remove key sources or shift priorities and compare outcomes to heuristic baselines. We include a synthetic-to-real caution explaining external validity limits and the conditions for future work with real data.

e) Deployment (prototype)

We deliver a secure prototype dashboard that displays reliability trajectories, optimised tasking recommendations, scenario runners and explanation panels. We log inputs, random seeds, generator configurations, model versions, parameters, recommendations and simulated decisions for audit. SHAP and LIME provide case-level explanations on synthetic records. We set up MLOps for the synthetic pipeline, including versioned generator configurations, data snapshots, scheduled regeneration, drift monitoring and role-based access.

f) Ethics and governance

We do not handle real personal or operational data. The synthetic generator is designed to avoid memorisation and prevent re-identification. Any transition to real data would require separate approvals, security controls and a formal ethics review.

4. Innovation and Uniqueness

This project combines data science and operations research to address a rarely tackled problem: managing intelligence source performance. While most AI in security focuses on surveillance or cyber defence, our emphasis is on the human decision-making process that influences operational outcomes. The approach replaces intuition-based assessments with statistically valid predictions and optimisation under uncertainty, resulting in assignments that are explainable, auditable, and improvable over time. Synthetic data allows us to safely test edge cases and stress scenarios while ensuring a clear pathway for governance reviews before any real-world implementation.

5. Sustainability and Scalability

The architecture is modular: the ML layer, optimisation engine and dashboard can be deployed independently or as a stack. This enables scaling across units and agencies and adaptation to adjacent domains such as law enforcement and border security. Interfaces follow common standards (APIs, model registries), supporting integration with existing case-management systems and training programs. With validation, the tool can be institutionalised as a standard for performance analytics, backed by documentation, versioning and role-based access to ensure maintainability and oversight.

C. RELEVANCE TO THE THEME

This proposal aligns with the Governance & Public Policy theme by strengthening data-driven decision-making, institutional accountability and resource efficiency in national security operations.

1. Governance enhancement

The framework introduces measurable, evidence-based decision processes into intelligence source management, reducing reliance on opaque or personality-driven judgments. It fosters transparency and traceability in operational decisions, a core principle of good governance.

2. Public policy relevance

Sound intelligence improves policy across defense, policing and border management. By improving how agencies assess and allocate source effort, we raise the quality and timeliness of information that informs policy options, interagency coordination and resource allocation.

3. National Security and Prosperity

The proposed system enhances operational reliability and anticipates risk through scenario analysis, supporting earlier detection of operational weaknesses and better contingency planning. More precise allocation of limited assets lowers waste and strengthens readiness in priority domains, including maritime and border security.

4. Alignment with sustainable development

The proposed project advances SDG 16 (Peace, Justice and Strong Institutions) by embedding accountability, institutional resilience and responsible use of AI in security governance. We focus on decision quality and auditability, not surveillance expansion.

5. Responsible and explainable AI

We incorporate interpretable ML (SHAP, LIME) so officers can see the factors influencing predictions and recommendations. Human review and override are included, reducing the risk of automation bias and supporting continuous learning through after-action feedback.

D. TECHNOLOGIES AND TOOLS

Languages & libraries: Python, TensorFlow/Keras (neural model), scikit-learn, XGBoost, Pyomo, Pandas, Numpy, Matplotlib

Optimisation solvers: GLPK /IBM ILOC CPLEX Optimiser

Visualisation: Dash or Streamlit for interactive dashboards

Data handling: synthetic, secure, anonymised datasets only, encryption and controlled access.