BIDS AND AWARDS COMMITTEE

BID BULLETIN No. 1

Project Title	:	PR No. 22-036: One (1) Lot Supply, Delivery, Setup, Installation, Data Ingestion, Testing, and Implementation of Software for Artificial Intelligence, inclusive of one (1) year Support and Training.
Date	:	17 September 2022

In its meeting on 13 September 2022 and pursuant to the e-mail from prospective bidder dated 14 September 2022, the Bids and Awards Committee resolved to issue this Bid Bulletin to clarify issue in the Bid Documents regarding the purchase of One (1) Lot Supply, Delivery, Setup, Installation, Data Ingestion, Testing, and Implementation of Software for Artificial Intelligence, inclusive of one (1) year Support and Training posted in PhilGEPS website, AMLC website and conspicuous places on 30 August 2022.

ISSUE/S	CLARIFICATIONS / AMENDMENTS
Issue No. 1: **Section V. Special Conditions of Contract** Final User Testing and Acceptance Matrix Item G. Maintenance of the General Requirements No. 1 SW support spans at least **5 years.**	**Amendment No. 1** Final User Testing and Acceptance Matrix Item G. Maintenance of the General Requirements No. 1 SW support spans at least **1 year.**
Issue No. 2: **Section VII. Technical Requirements** **Functional Requirement** *"Must be able to check the completeness of information in the Narrative field, specifically, if it answers the questions Who, What, Where, When, Why, and How"* (found on pages 32, 34, and 40 of the Bidding Document)	**Amendment No. 2** **Functional Requirement** *"Must be able to come up with a completeness index for the Narrative field using natural language processing, preferably based on the 5Ws and 1H".*
Issue No. 3: 1. What is the frequency of the STR/CTR data to be analyzed? Is there further information such as data volume and data growth?	**Clarification No. 1** 1. The STRs and CTRs are filed on a per-transaction basis and received daily by the AMLC. Please see Annex A, Table 1 for the historical data volume.

2. What is the approximate number of STR we may get monthly?

2. Monthly 2021 STR submissions are provided in the table that follows:

Month	No. of STRs
January	258,411
February	203,856
March	215,197
April	616,172
May	244,574
June	758,007
July	268,021
August	456,781
September	298,171
October	401,186
November	230,854
December	276,004
TOTAL	4,227,234

3. What is the structure of STR/CTR data? Please share a comprehensive set if there are different formats

3. Please see Annex B for the data dictionary of SAS tables. The structure is similar to the Oracle database which holds the raw data. The 2021 AMLC Registration and Reporting Guidelines, Part 4 (Reporting Procedures, page 73 onwards)[1] also contains information on the data elements.

4. What is the time-window of available transaction data?

4. The models are required to run on 5 years' worth of transaction data.

In terms of availability of submitted data:
• Oracle DB: 1 day from submission to AMLC. CTR/STR reports submitted within a given day will be processed for uploading to the AMLC DB the following day.
• SAS DB: Ideally, 2 days from submission to AMLC. Input to SAS is from the Oracle

---

[1]
http://www.amlc.gov.ph/images/PDFs/2021%20AMLC%20REGISTRATION%20AND%20REPORTING%20GUIDELINES.pdf

	DB. Reports successfully uploaded to the DB within a given day are ingested by SAS the following day.
5. Kindly provide the data volume to be initially ingested/processed, and the projected data growth.Kindly provide an estimated percentage of annual data growth.	5. Five (5) years' worth of data (approximately 350 GB) will be available. Initial data ingestion will depend on the vendor's model. Please see Annex A, Table 2 table for data on daily data ingestion.
6. Please share a comprehensive set of sample data.	6. Please see below linked file for a sample of redacted CTRs and STRs. Kindly request access, indicating your name and company in the Message field. https://docs.google.com/spreadsheets/d/1SWTKdqhs2hYA41XhgP6462eOxhEt-zdn
7. Can training data be provided for STR? (That covers all range of narrative and reasoning inputs) (STR Reasoning)	7. Yes, training data will be provided for STRs.
8. Based on what data, trend and underlying characteristics will be learned (only transaction or any other external information)?	8. Only CTR and STR data will be used to train the models. Aside from transaction data, STRs contain narratives on the basis of filing institutions' suspicion. More information on the contents of the CTRs and STRs can be found in the 2021 AMLC Registration and Reporting Guidelines.[2]
9. Do you have other external data sources that you connect or get data as reference for CTR/STR verification? In case the client's data is not sufficient, how many outside data sources should be assumed?	9. The minimum requirement is to use the existing CTR and STR data, but we are open to the use of additional data sources. The AMLC has existing access to watchlist databases, name screening

---

[2]

http://www.amlc.gov.ph/images/PDFs/2021%20AMLC%20REGISTRATION%20AND%20REPORTING%20GUIDELINES.pdf

	databases, and an internal Foreign Exchange declarations database).
Issue No. 4: **Section VII. Technical Specifications** **Item 2.1 Functional Requirements** 1. How many users are going to use the application?	**Clarification No. 2** 1. 22 concurrent users; 30 named users
2. How many analytics consumers do you have (e.g. reports viewers only)? How many analytics power users / analytics developers do you have?	2. 20 analytics consumers; 5 analytics power users/analytics developers
3. What are the different categories of users?	3. We envision 3 broad roles: • financial analysts, who shall be the end user of software outputs; • machine learning engineers, who shall conduct model assessment and refinement; and • system administrators, who shall oversee software maintenance.
Issue No. 5: **Section VII. Technical Specifications** **Item 2.1.2.A Other information on detection** Is it based on the probability score i.e. high score is maximum, slightly lower is medium, etc.?	**Clarification No. 3** The supplier has the option to implement a scoring mechanism that it deems appropriate, depending on the defined basis of similarity with known cases.
Issue No. 6: What does tactical cases mean, any examples?	**Clarification No. 4** Tactical cases are cases targeting specific persons of interest (POIs) who were found to have likely links to criminal activities. Cases emanate from a variety of sources such as intelligence received from law enforcement agencies (LEAs), disclosures from international financial

	intelligence units (FIUs), and news reports.
Issue No. 7: **Section VII. Technical Specifications** **Item 2.2 Item 1-3 A system that uses trained Machine Learning models to assess the Suspicious Transaction Report (STR) quality and conducts STR prioritization.**	**Clarification No. 5**
Third point is not clear, an example may help.	If the question is referencing the Complementary Function, the users intend to search based on usual query functionalities (through a point and click interface), display the results in tabular format, and retrieve the results in CSV format.
Issue No. 8: **Section II. Instruction to Bidders** **Item 10.2 The Bidder's SLCC as indicated in ITB Clause 5.3 should have been completed within three (3) years prior to the deadline for the submission and receipt of bids.**	**Clarification No. 6**
1. Can we request to extend the SLCC requirement for projects completed in the last 5 years?	1. We maintain the current requirement of three (3) years.
2. Our SLCC, is a contract with Phase 1 & 2 components. Phase 1 is completed, billed, in production and above the required 50% ABC amount. Is this acceptable? **Section V. Special Conditions of Contract** **1.7 Development Team Composition**	2. SLCC pertains to a contract, not just a component of the contract. Hence, completion of just Phase 1 will not meet the requirement.
3. To meet the technical skills requirements, would you be amenable to have resources currently seconded (assigned for a specific duration) to the bidder?	3. Section III (Bid Data Sheet), ITB Clause 7 provides that "subcontracting is not allowed". As mentioned during the Pre-Bid Conference, if a bidder intends to use resources from an affiliate which is 100%

	owned by the bidder, there should be a Joint-Venture Agreement.
Issue No. 9: Since the project will run for 410 days based on the RFP and has a requirement for a one-year support, can we clarify that the software license needed will be for 410days + 365 days or will it be only for 410 days?	**Clarification No. 7** The software license will be for 410 days plus the final user testing period of 10 days.
Issue No. 10: **Section II. Instruction to Bidders** **Item 12.1 Bid Prices** There seems to be hardware related, there is no hardware component in the Bid. Can it be removed?	**Clarification No. 8** The referenced section is a set of general procurement provisions and remains as is. As set out in Section III (Bid Data Sheet) ITB Clause 12 on page 16, the project shall cover software and training.
Issue No. 11: Due to the complexity of the bid, may we request a 2-week bid submission extension from date of release of the final bid bulletin?	**Clarification No. 9** The project timeline does not allow for an extension in bid submission.
Issue No. 12: **Section V. Special Conditions of Contract** **Item 2.2** and **Section VI. Schedule of Requirements** What is the timing of completion? Can the deliverables be completed at a different schedule but within the overall timeline, eg. 14 months?	**Clarification No. 10** Timeline is non-negotiable. Vendor / supplier shall adhere to the schedule of requirements under Section VI, page 37.
Issue No. 13: **Section V. Special Conditions of Contract** **Item 1.5 Implementation and Other Services** What is the expected output and overall expectation?	**Clarification No. 11** The vendor/supplier is expected to follow the lifecycle of data analytics in order to deliver the Technical Specifications under Section VII, pages 38-44.

Issue No. 14: **Section V. Special Conditions of Contract** **Section VII. Technical Requirements** **Item No. 2 Functional Requirements** "Solution providers must develop a machine learning model for at least five (5) high risk predicate crimes." Can we update from "at least five" to "at most five"?	**Clarification No. 12** The minimum requirement is to have **at least** five models.
Issue No. 15: **Section VII. Technical Requirements** 1. What are the 5 high risk predicate crimes? If not yet finalized, can you at least provide the options that you are considering now? 2. We understand AMLC will pre-identify high-risk predicate crimes. Is AMLC expecting one (1) machine learning model to be built for each predicate crime, therefore a total of five (5) machine learning models for five (5) high risk predicate crimes? 3. For these identified predicate crimes, how much true positive data is available for the analysis per high-risk predicate crime? 4. Is AMLC open to non-supervised machine learning models in the event that there are insufficient true positives necessary for building a supervised model?	**Clarification No. 13** 1. The final predicate crimes will be determined after the awarding of the contract and will be chosen from the high-risk predicate crimes listed in Annex C. 2. The project aims to build one model for each predicate crime. The AMLC will provide the five high-risk predicate crimes. 3. Data collection is still in process. AMLC has a lot of true positive data which was analyzed and investigated for more than a decade of operation. 4. Yes, but the priority is still to build supervised learning models.
Issue No. 16: **Section VII. Technical Requirements** 1. As per our understanding, a reporting institution will initially fill up the Reason code associated with the narrative of an STR case. In the event that the narrative does not	**Clarification No. 14** 1. Yes, the AMLC will provide the correct Reason field for a sample of STRs.

adhere to the Reason code (i.e, Reporting institution incorrectly tagged or assigned a reason code for the case/narrative), this is considered a "mismatch" or incorrect. For AMLC's functional requirement to be addressed, will AMLC provide the "correct" Reason Code associated with a narrative? This way, the model will learn from the correct Narrative & Reason code tandem of information. Pls note the following: Correct Reason Code = what AMLC deems is the correct Reason Code Assigned Reason Code = what the Reporting Institution submitted as Reason Code	
2. Are the alerting, information, prioritization, and statistics expected on per STR level or per covered person?	2. Alerting, information, prioritization, and statistics are all expected at the STR-level. As mentioned in the bidding documents, reporting must likewise be available by Covered Person and industry.
3. Once the model detects a mismatch between Narrative and Reason Code, does AMLC want this to be alerted?	3. STRs with mismatched reason and narrative must be included in the monthly report. Daily alert generation is not needed.
4. For monthly reports of STRs with unmatched narrative and reason, do you mean aggregated count by covered person and by industry? After which, is it going to alert a certain user group (e.g, mismatches of 25 STRS from Industry X will alert or be assigned to Personnel A)?	4. Yes to first question. Assigned user group will access/download the monthly report of STRs with unmatched reason and narrative.
Issue No. 17: **Section VII. Technical Requirements** 1. Does AMLC require that the Who, What, Where, When, Why, and How be explicitly identified/itemized as an output of the machine learning model?	**Clarification No. 15** 1. It is not a requirement but would be a helpful functionality to include.
1. If only the Who, Where, When and How, or 4 out of 6 can be detected from the	2. Further to Amendment No. 2, there must at least be a working machine

narrative, will this be considered by AMLC as fulfilling this requirement? If the Machine Learning model is not able to identify completeness of information (cannot identify the 5W and 1H), is there another way to identify if the narrative is complete or not? For example, perhaps a tagging by AMLC? Using this, vendors can just provide the monthly report format as an alternative to fulfilling this requirement.	learning model that attempts to identify completeness of information and learns from new labeled data to improve the accuracy of its predictions. Users must also be able to review the results and manually input correct tags.
Issue No. 18: **Section VII. Technical Requirements** 1. What's the categorization of Predicate Crimes in Training Data? How many of each type? (Detection use-case)	**Clarification No. 16** 1. If this question pertains to STRs, categorization is based on the Reason field in the ARRG. Data collection is still in process.
2. Does AMLC have an existing set of business rules in classifying STRs according to a specific unlawful activity or suspicious circumstance? If yes, can AMLC provide tagging of reason codes with associated unlawful activity? If not, are the unlawful activities or suspicious circumstances all enumerated in the existing STR Form's Reason Codes SI6 (items PC1-PC36?)	2. STRs are labeled according to the reason of suspicion identified by the Covered Persons. Reason of Suspicion is a mandatory field under the 2021 AMLC Registration and Reporting Guidelines.[3]
3. Is this existing prioritization guideline set as a business rule by AMLC internally? Checking the current circular, this priority level is not existing. Kindly clarify.	3. The prioritization was set based on internal guidelines.
4. Can AMLC share the details / relevant factors affecting priority level?	4. The priority level is based primarily on the country's National Risk Assessment on Money Laundering and Terrorist Financing[4].

---

[3] http://www.amlc.gov.ph/images/PDFs/2021%20AMLC%20REGISTRATION%20AND%20REPORTING%20GUIDELINES.pdf

[4] http://www.amlc.gov.ph/images/PDFs/NRAReport20152016.pdf

Issue No. 19:	Clarification No. 17
1. Do you have an existing ETL tool? Please provide product and version.	1. SAS Data Integration Studio (SAS 9.4M5)
2. How does the data get into the Oracle database? Can you describe the process from submission of the CTR / STR reports by institutions to AMLC?	2. The CTR/STR submission and validation process can be gleaned from the ARRG[5]. The data then enters the Oracle database through an internal data collection system.
3. Do you have any existing BI / Dashboarding tools? If yes, what are they and can we use those?	3. No
4. What is the preference of AMLC, the system will be on-premises (of AMLC) or on public cloud?	4. On-premises
5. What are the data retention requirements? How long should data be retained for quick access? How long data should be archived?	5. The vendor's solution must provide ready access to 5 years' worth of output data. Archived data will be stored permanently and, as such, shall not be deleted.
6. Are there any defined Recovery Time Objective (RTO), Recovery Point Objective (RPO), and/or availability requirements?	6. RTO – 4 hours; RPO – 12 hours
7. Kindly provide the required backup retention/size on disk.	7. There is no specific requirement. The primary consideration is that the software operates on the hardware specifications mentioned in the Bidding Documents (pages 43-44).
8. Kindly share a diagram of your current system architecture.	8. This is not currently available.

---

[5]

http://www.amlc.gov.ph/images/PDFs/2021%20AMLC%20REGISTRATION%20AND%20REPORTING%20GUIDELINES.pdf

9. Will the infrastructure (servers, OS, switches) be part of what will be supplied by the vendor?	9. No, hardware will come from the AMLC.
10. Kindly confirm if data ingestion processes will be executed in parallel with user driven queries.	10. Yes, they will run in parallel.
11. Do you have any concerns / pain points on the current system?	11. This system is the first Artificial Intelligence (AI) solution that the AMLC will procure for this purpose.
12. Do you have existing version control, CI / build server, artifact repository? What are they? Can it be used by the vendor for the to-be system?	12. No, none of the mentioned systems are in place.
13. Will the system be handed over to AMLC administrators for regular operations and maintenance?	13. Yes, the system will be handed over to the AMLC.

This Bid Bulletin shall form an integral part of the Bid Documents.

MA. RHEA M. SANTOS-MENDOZA
Chairperson

ANNEX A

Table 1. Number of reports received yearly

	STRs	Growth Rate	CTRs	Growth Rate
2017	287,265		39,940,922	
2018	491,717	71%	43,114,420	8%
2019	623,201	27%	46,381,173	8%
2020	1,021,605	64%	38,200,000	-18%
2021	4,227,234	314%	42,963,080	12%
Total	**6,651,022**		**210,599,595**	

Table 2. Daily data ingestion

Data Storage and Integration Processing		
Sessions related to ETL/data management *(During overnight BATCH Window)*	***Concurrent Sessions**	***Average Input Data Volume per Session**
Data volumes =< 1 GB	1	700MB
Data volumes =< 10 GB	1	6GB
*Estimated percentage of annual data growth		17%

Table Name	Column Name	Type	Length	SAS Format	SAS Informat	Description
Table 1	sas_id	num	8	21.	21.	SAS system ID
Table 1	sup_agency	char	10	$10.	$10.	Supervising agency abbreviation
Table 1	industry_type	char	100	$100.	$100.	Industry where reporting entity belongs to
Table 1	conf_receipt_no	char	255	$255.	$255.	Reference no. for report submission
Table 1	file_type	char	5	$5.	$5.	Type of report indicating if CTR or STR
Table 1	record_no	num	8	21.	21.	Reference no. for report submission
Table 1	upload_date	num	8	DATETIME25.6	DATETIME25.6	Date/time of uploading to the AMLC database
Table 1	trans_date	num	8	DATETIME25.6	DATETIME25.6	Date/time of transaction
Table 1	trans_date_sas	num	8	21.	21.	Date/time of transaction (sas format)
Table 1	trans_code	char	10	$10.	$10.	Transaction type code (deposit, withdrawal, etc.)
Table 1	trans_code_desc	char	200	$200.	$200.	Transaction type description (deposit, withdrawal, etc.)
Table 1	trans_ref_no	char	50	$50.	$50.	Reference no. for report submission
Table 1	peso_amount	num	8	27.5	27.5	Transaction amount in PHP
Table 1	fx_amount	num	8	27.5	27.5	Transaction amount in foreign currency (if applicable)
Table 1	fx_currency	char	100	$100.	$100.	Transaction currency (if applicable)
Table 1	nature_trans	char	30	$30.	$30.	Nature of transaction
Table 1	corres_bank	char	90	$90.	$90.	Correspondent bank name (if applicable)
Table 1	corresbank_address	char	270	$270.	$270.	Correspondent bank address (if applicable)
Table 1	corresbank_ctrycode	char	100	$100.	$100.	Correspondent bank country code (if applicable)
Table 1	inception_date	num	8	DATETIME25.6	DATETIME25.6	Insurance inception date (if applicable)
Table 1	maturity_date	num	8	DATETIME25.6	DATETIME25.6	Insurance maturity date (if applicable)
Table 1	amount_claim	num	8	27.5	27.5	Insurance amount claim (if applciable)
Table 1	php_annual	num	8	27.5	27.5	PHP Amount of Annual Premium /Excess/Advance Premium/Payment/Top-ups
Table 1	fx_annual	num	8	27.5	27.5	FX Amount of Annual Premium /Excess/Advance Premium/Top-ups
Table 1	reason	char	1000	$1000.	$1000.	Reason for suscpicion (if STR)
Table 1	full_institution_code	char	20	$20.	$20.	Reporting Institution ID
Table 1	institution_branch_name	char	300	$300.	$300.	Reporting Institution branch name
Table 1	submission_type	num	8	21.	21.	Type of submission (A - add, E - edit, D - delete, T - test)
Table 1	narative	char	4000	$4000.	$4000.	Narration of situation/events (if STR)
Table 1	old_account_no	char	40	$40.	$40.	Refers to the previously assigned Account Number of the client or Promissory Note No. for loans, time deposit etc. or Client Stock Ref. No. for securities
Table 1	php_policy_amount	num	8	27.5	27.5	PHP amount of insurance policy
Table 1	fx_policy_amount	num	8	27.5	27.5	FX amount of insurance policy
Table 1	insurance_type	char	30	$30.	$30.	Policy/Insurance Type
Table 1	insurance_terms	num	8	27.5	27.5	Term of insurance (in years)
Table 1	no_shares_units	num	8	27.5	27.5	Number of shares involved in transaction (if applicable)
Table 1	net_asset_value	num	8	27.5	27.5	Net asset value of shares
Table 1	run_date	num	8	DATE9.	DATE9.	System run date (AMLC DB)
Table 1	processed_dttm	num	8	DATETIME25.6	DATETIME25.6	System process dttm (SAS ETL)
Table 2	sev_id	num	8	21.	21.	SAS DQ-generated single entity ID (product of entity resolution)
Table 2	sas_id	num	8	21.	21.	SAS system ID
Table 2	names_id	num	8	21.	21.	Names ID from AMLC DB
Table 2	inst_code	char	20	$20.	$20.	Reporting Institution ID
Table 2	conf_receipt_no	char	255	$255.	$255.	Reference no. for report submission
Table 2	record_no	num	8	21.	21.	Reference no. for report submission

Table Name	Column Name	Type	Length	SAS Format	SAS Informat	Description
Table 2	person_type	char	1	$1.	$1.	Type of person relative to transaction (Counterparty, Beneficiary, Subject of suspicion, etc.)
Table 2	birth_date	num	8	DATE9.	DATE9.	Birth date of account holder/s
Table 2	tin	char	30	$30.	$30.	TIN number
Table 2	sssgsis_regno	char	30	$30.	$30.	sss/gsis number (if applicable)
Table 2	telno	char	15	$15.	$15.	Telephone number
Table 2	nature_business	char	35	$35.	$35.	Nature of business
Table 2	person_name_orig	char	450	$450.	$450.	Accountholder name as reported
Table 2	person_name_orig_stnd	char	450	$450.	$450.	Accountholder name after standardization
Table 2	person_address_orig	char	600	$600.	$600.	Accountholder address as reported
Table 2	person_address_orig_stnd	char	600	$600.	$600.	Accountholder address after standardization
Table 2	ind_corp	char	1	$1.	$1.	Corporate entity indicator (Y - corporation, N - individual)
Table 2	cust_ref_no	char	50	$50.	$50.	Accountholder reference number.
Table 2	birth_place	char	90	$90.	$90.	Birth place of account holder/s
Table 2	nationality	char	40	$40.	$40.	Nationality of accountholder
Table 2	id_type	char	100	$100.	$100.	Type of ID provided
Table 2	id_details	char	30	$30.	$30.	Details of ID provided
Table 2	person_benef_relation	char	30	$30.	$30.	Beneficiary relationship (if applicable)
Table 2	person_benef_code	char	1	$1.	$1.	Beneficiary code (if applicable)
Table 2	acct_no	char	50	$50.	$50.	Account number involved
Table 2	cleansed_account_no	char	50	$50.	$50.	DQ standardized account number involved
Table 2	run_date	num	8	DATE9.	DATE9.	System run date (AMLC DB)
Table 2	processed_dttm	num	8	DATETIME25.6	DATETIME25.6	System process dttm (SAS ETL)

ANNEX C

High Risk Predicate Crimes

	UNLAWFUL ACTIVITY	PRIORITY
1.	Terrorism and conspiracy to commit terrorism	High
2.	Financing of terrorism	High
3.	Proliferation financing	High
4.	Drug offenses (violation of specific provisions of the Comprehensive Dangerous Drugs Act of 2002)	High
5.	Graft and Corrupt Practices	High
6.	Plunder	High
7.	Bribery and Corruption of Public Officers	High
8.	Malversation of Public Funds and Property	High
9.	Hijacking and other violations under R.A. 6235; destructive arson and murder	High
10.	Violations of the decree Codifying the Laws on Illegal/Unlawful Possession, Manufacture, Dealing In, Acquisition or Disposition of Firearms, Ammunition or Explosives	High
11.	Violations of the Anti-Trafficking in Persons Act of 2003	High
12.	Violation of the Anti-Photo and Video Voyeurism Act of 2009	High
13.	Violation of the Anti-Child Pornography Act of 2009;	High
14.	Violations of the Special Protection of Children Against Abuse, Exploitation and Discrimination Act	High
15.	Kidnapping for ransom	High