**Workshop – 3**

**Question 1:** In an earlier module on AI Project Lifecycle & Management, we used a fictional Gen AI project named “AI Data Extractor”. Based on our understanding of the “AI Data Extractor” solution, please fill out the following sections and pointers from the above template:

● ***Model Details:***

**○ What Gen AI API (e.g., from Open AI) or open-source model (e.g., LLama v3) would you use? And Why?**

I will use Propriety Gen AI model “Gemini Pro – 2.5” hosted on Google cloud for the following reasons:

1. AI Data extractor has been identified as Horizon 1 solution which means this is a project which is important and urgent. Proprietary models are readily available, and hence overall delivery is accelerated.
2. HAIL is an insurance company, and we have no information on its technological capabilities. The Propriety Gemini Pro – 2.5 model is easy to use, require low expertise, and Google Cloud offers indemnification for AI model outputs and the data used to train them in specific contexts.
3. The AI Data extractor is a Horizon 1 Solution which means it impacts core business and hence requires great security. By using proprietary models that offer enterprise-grade security features, such as safety filters, citation filters, safety error codes, and many more security features.
4. Gemini Pro -2.5 is Google’s most advanced AI model that is Natively multimodal, understanding input across text, audio, images and video.

By using the Gemini Pro – 2.5 Gen AI model, I believe the AI Avengers team can now focus only on delivering a high-quality solution in the least amount of time.

● ***Intended Use:***

**○ What are the primary use-cases?**

The primary use case is to use the AI Data extractor to automate the data entry for its claim processing. The AI solution extract various required data elements from the claim pdf form and update its claim processing system.

○ **Who are the primary users?**

* The AI data extractor primary users are Business teams who will now verify the elements of the claim form once uploaded.

**○ Are there any out-of-scope use-cases?**

* The solution cannot be used to automate any manual document extractions as it is particularly trained for claim form extraction.

● ***Metrics:***

**○ What are the metrics that should be chosen to evaluate the performance of the solution? Provide the definition or formulae. Use example numbers to show how the metric is calculated.**

1. **Daily/Weekly Error%:**

**Description:** The number of errors that AI Data extractor solution did while performing its task daily/weekly.

**Formula:**

Daily Error%:(# of Errors per claim)/(Total no. of claims processed per day) \* 100

Weekly Error%: (# of Errors per claim)/(Total no. of claims processed per week) \* 100

**Why monitor?:** Plot the data points in a line graph to check if there are error percentages that have crossed acceptable limits and raise incident with support team.

**Example:**

Below table shows the sample calculation for Daily Error%. The same can be applied for weekly as well:

|  |  |  |
| --- | --- | --- |
| Errors per claim | Total claims processed per day | Error% |
| 3 | 210 | = (3/210) \*100 = 1.4% |
| 2 | 190 | = (4/190) \*100 = 2.1% |
| 4 | 180 | = (4/180) \*100 = 2.2% |

2. **Daily/Weekly Avg. Time taken per claim in secs:**

**Description:** Average of the time taken for the AI Data extractor to process each claim in seconds for a 24 hour or a 7 day time period.

**Formula:**

Daily Avg. Time taken:Avg. of Time taken per claim)/ (Total no. of claims processed per day) \* 100

For e.g. in a day lets assume that AI data extractor processed 3 claims then:

|  |  |  |
| --- | --- | --- |
| **Claim no.** | **Time taken in secs** | Daily Avg. taken: (33/3) = 11 secs |
| 121313 | 10 secs |
| 223421 | 15 secs |
| 323212 | 8 secs |
| Total claims: 3 | Total time taken: 33 |

**Why monitor?:** Plot the data points in a line graph to check if the claims processing time was more than acceptable limits. If not, raise the incident with the support team for corrections.

3. **# of Application issues per Week/Month (Reliability):**

Description: Count of issues encountered with the application such as:

1. Application not responding
2. Application has crashed
3. Application not stable

Formula: No. of issues encountered in the application per week/month.

● ***Evaluation Dataset:***

**○ Describe the dataset used for evaluating the performance of the solution. How many rows are there? What does each row represent? How many columns are there? What are the column names?**

The dataset to test the model performance contains 10000 real health insurance claims. Each row of the dataset represents summary of the claim that includes a variety of columns that are commonly found in health insurance claims:

**Columns in the Dataset.**

* **Patient\_Id:** A unique identifier for each patient, given by hospital mentioned in the claim form.
* **Patient\_Name**: Full name of the patient (First Name, Middle Name and Last Name).
* **Patient\_Income**: Annual income of the patient in INR.
* **Patient\_Marital\_Status**: Marital status of the patient (Married/Unmarried).
* **Patient\_Age**: Age of the patient in years.
* **Patient\_Gender**: Gender of the patient (M/F).
* **Patient\_DOB**: Date of birth of each patient.
* **Claim\_ID**: A Unique identifier for each claim, system generated.
* **Claim\_Amount**: The amount claimed in INR in .
* **Claim\_Date**: The date when the claim was made.
* **Claim\_Type**: Type of claim (Inpatient, Outpatient, Emergency, Routine).
* **Hospital\_Name**: The name of the hospital where the patient was treated.
* **Hospital\_Specialty**: Specialty of the healthcare provider (e.g., Cardiology, Orthopedics).
* **Hospital\_Location:** Address of the healthcare provider.
* **Health\_Care\_Diagnosis\_Code**: Code representing the diagnosis.
* **Health\_Care\_Procedure\_Code**: Code representing the procedure performed.
* Doctor\_Name: Full Name of the Doctor
* Doctor\_Qualification: Highest medical degree of the doctor.
* Doctor\_Prescription: The prescription of medicines given by doctor.

Apart from this main data set, there will be other data sets that are also useful such as:

Hospital\_Details: Details of all the hospitals such as Hospital name, address, Point\_Of\_Contact, Hospital speciality etc.

Diagnostic\_Details: Details of all the Diagnostic vendors such as Diagnostic name, address, Point\_Of\_Contact, Diagnostic speciality etc.

Doctor\_Details: Details of the doctor such as doctor name, doctor degree, doctor profile, doctor\_treatment\_success\_Rate, doctor\_consulting\_cost etc.

Treatment\_Details: Details of the treatments such as Treatment name, treatment description, treatment process, treatment costs, treatment success rate etc.

Diagnosis\_Details: Details of diagnosis such diagnosis test name, diagnosis\_cost, result\_generation\_time, diagnosis\_description etc.

Medical\_items\_details: Details of commonly used medical items in various treatments such as syringes, cotton rolls, gloves, etc.

● Responsible Usage of AI:

○ Are there any considerations of ethics or fairness or bias that would apply to this solution? If yes, explain why. If not, explain why not.

**Ethics:**  Ethical considerations such as personal information of patients including Date of birth, medical history needs to be secured, encrypted etc.

**Fairness:**  The AI Data extractor solution irrespective of patient gender, claim amount, claim location and doctor should process the claim like it would process any other claim.

**Bias:** The AI data extractor solution irrespective of the handwriting quality, language, spelling mistakes should process the claim like it would process any other claim.

**Regulatory requirements:** Since this is a claims data that contains patient data, therefore the solution should be designed and developed keeping in mind HIPPAA regulatory requirements.

● In Workshop 3 PPT where we discussed AI Strategy, we looked at how having a robust AI/ML Testing Methodology is a key aspect of AI Strategy.

○ In your industry, how has the standard software testing methodology been modified or adapted for AI/ML solutions? Provide a descriptive answer not exceeding half a page in length. Please avoid providing company-specific confidential details.

○ What are the various tools that are used in your industry for testing an AI/ML solution? Please avoid providing company-specific confidential details.

**Question 2**

In Workshop 3 PPT where we discussed AI Strategy, we looked at how having a robust AI/ML Testing Methodology is a key aspect of AI Strategy.

● In your industry, how has the standard software testing methodology been modified or adapted for AI/ML solutions? Provide a descriptive answer not exceeding half a page in length. Please avoid providing company-specific confidential details.

**\*\*\*\* Verify since it is copilot generated\*\*\*\***

In the semiconductor industry, the standard software testing methodology has been significantly adapted to accommodate the unique requirements of AI/ML solutions. Traditional software testing focuses on verifying the correctness of code against predefined requirements, but AI/ML solutions introduce variability and unpredictability due to their data-driven nature.

One major adaptation is the emphasis on data quality and preprocessing. Since AI/ML models rely heavily on data, ensuring the accuracy, completeness, and relevance of training data is crucial. This involves rigorous data validation, cleansing, and augmentation processes to mitigate biases and improve model performance.

Another key modification is the implementation of continuous integration and continuous deployment (CI/CD) pipelines tailored for AI/ML workflows. These pipelines automate the training, testing, and deployment of models, enabling rapid iteration and reducing the time to market. Additionally, specialized testing frameworks and tools are used to evaluate model performance, robustness, and fairness. Techniques such as cross-validation, A/B testing, and adversarial testing are employed to assess model generalization and resilience to adversarial inputs.

Moreover, explainability and interpretability have become critical aspects of AI/ML testing. Understanding how models make decisions is essential for debugging, compliance, and gaining stakeholder trust. Techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are used to provide insights into model behavior and ensure transparency.

Lastly, ethical considerations and regulatory compliance play a significant role in AI/ML testing. Ensuring that models adhere to ethical guidelines and industry regulations is paramount. This involves conducting thorough audits, bias detection, and implementing safeguards to prevent unintended consequences.

Overall, the adaptation of software testing methodologies for AI/ML solutions in the semiconductor industry involves a comprehensive approach that addresses data quality, automation, model evaluation, interpretability, and ethical considerations.

● What are the various tools that are used in your industry for testing an AI/ML solution?

**\*\*\*\* Verify since it is copilot generated\*\*\*\***

In the semiconductor industry, testing AI/ML solutions involves a variety of specialized tools designed to ensure the accuracy, robustness, and fairness of models. Here are some commonly used tools:

1. **TensorFlow Extended (TFX)**: This end-to-end platform is used for deploying production-ready machine learning pipelines. It includes components for data validation, model training, and serving, making it ideal for comprehensive testing and deployment of AI/ML models.
2. **Apache Spark MLlib**: Spark's machine learning library is used for large-scale data processing and model training. It supports various algorithms and provides tools for feature extraction, transformation, and model evaluation, which are essential for testing AI/ML solutions.
3. **PyTorch**: Known for its flexibility and ease of use, PyTorch is widely used for developing and testing deep learning models. It offers tools for model debugging, visualization, and performance monitoring, which are crucial for ensuring model reliability.
4. **H2O.ai**: This open-source platform provides tools for building and testing machine learning models. It includes features for automated machine learning (AutoML), model interpretability, and performance evaluation, making it a valuable tool for testing AI/ML solutions.
5. **MLflow**: An open-source platform for managing the end-to-end machine learning lifecycle. MLflow includes components for tracking experiments, packaging code, and deploying models, which are essential for testing and validating AI/ML solutions.
6. **Keras**: A high-level neural networks API, Keras is used for building and testing deep learning models. It provides tools for model training, evaluation, and visualization, which are important for ensuring model accuracy and performance.
7. **Scikit-learn**: This popular machine learning library for Python offers tools for data preprocessing, model training, and evaluation. It supports various algorithms and provides metrics for assessing model performance, making it a key tool for testing AI/ML solutions.

These tools help ensure that AI/ML models in the semiconductor industry are thoroughly tested for accuracy, robustness, and fairness, ultimately leading to reliable and effective solutions.

**Question 3**

In Workshop 3 PPT where we discussed AI Strategy, we also looked at change management at a high level. For the fictional AI Data Extractor solution we looked at in the AI Project Lifecycle and Management module, provide a detailed Change Management plan which answers the following:

● Who are the teams and individuals that should be involved? Provide a RASCI matrix as part of your submission showing who should be: Responsible (R), Accountable (A), Support (S), Consult (C) and Inform (I). As this is a fictional project, feel free to be creative in thinking of the teams and individuals.

A detailed change management plan for an insurance company transitioning from a manual data entry process for claims to an AI data extractor solution is as follows:

**Change Management Plan for Implementing AI Data Extractor Solution**

**1. Introduction**

The insurance company is transitioning from a manual data entry process for claims to an AI data extractor solution. This change aims to improve efficiency, reduce errors, and enhance customer satisfaction. The following plan outlines the steps necessary to manage this change effectively.

**2. Objectives**

* **Improve Efficiency**: Automate data entry to reduce processing time.
* **Reduce Errors**: Minimize human errors in data entry.
* **Enhance Customer Satisfaction**: Provide faster and more accurate claim processing.

**3. Stakeholder Analysis**

Identify and engage key stakeholders who will be affected by the change:

* **Employees**: Data entry staff, claims processors, IT staff.
* **Customers**: Policyholders submitting claims.
* **Management**: Executives and department heads.
* **Vendors**: Providers of the AI data extractor solution.

**4. Communication Plan**

Develop a communication strategy to keep stakeholders informed and engaged:

* **Initial Announcement**: Inform all stakeholders about the upcoming change and its benefits.
* **Regular Updates**: Provide progress updates through emails, meetings, and newsletters.
* **Feedback Channels**: Establish channels for stakeholders to provide feedback and ask questions.

**5. Training and Support**

Ensure employees are adequately trained and supported during the transition:

* **Training Programs**: Conduct training sessions on using the AI data extractor solution.
* **User Manuals**: Provide comprehensive user manuals and guides.
* **Help Desk**: Set up a help desk to assist employees with any issues.

**6. Implementation Plan**

Outline the steps for implementing the AI data extractor solution:

* **Pilot Phase**: Start with a pilot phase to test the solution with a small group of users.
* **Full Rollout**: Gradually roll out the solution to all users based on the success of the pilot phase.
* **Monitoring and Evaluation**: Continuously monitor the implementation and evaluate its effectiveness.

**7. Risk Management**

Identify and mitigate potential risks associated with the change:

* **Technical Issues**: Ensure robust IT support to address any technical problems.
* **Resistance to Change**: Address employee concerns and resistance through effective communication and training.
* **Data Security**: Implement measures to ensure the security and privacy of data.

**8. Performance Metrics**

Define metrics to measure the success of the change:

* **Efficiency**: Measure the reduction in processing time for claims.
* **Accuracy**: Track the decrease in data entry errors.
* **Customer Satisfaction**: Monitor customer feedback and satisfaction levels.

**9. Review and Continuous Improvement**

Regularly review the change management process and make improvements as needed:

* **Feedback Collection**: Gather feedback from stakeholders to identify areas for improvement.
* **Continuous Improvement**: Implement changes based on feedback to continuously improve the process.

**RASCI Matrix**

A screenshot of a computer

AI-generated content may be incorrect.

● Provide a GANTT chart as part of your submission showing:

○ What are the various milestones that should be a part of the Change Management plan?

○ What are the tasks within each milestone? Who is the owner of the task?

○ What is the timeline for each task? Provide a Start Date and End Date, and also show the duration visually in the GANTT chart.

○ Are there any dependencies between tasks inside a milestone or across milestones? Indicate the same.

Sure, I'd be happy to help you with that! Here's a detailed change management plan for an insurance company transitioning from a manual data entry process for claims to an AI data extractor solution:

**Change Management Plan for Implementing AI Data Extractor Solution**

**1. Introduction**

The insurance company is transitioning from a manual data entry process for claims to an AI data extractor solution. This change aims to improve efficiency, reduce errors, and enhance customer satisfaction. The following plan outlines the steps necessary to manage this change effectively.

**2. Objectives**

* **Improve Efficiency**: Automate data entry to reduce processing time.
* **Reduce Errors**: Minimize human errors in data entry.
* **Enhance Customer Satisfaction**: Provide faster and more accurate claim processing.

**3. Stakeholder Analysis**

Identify and engage key stakeholders who will be affected by the change:

* **Employees**: Data entry staff, claims processors, IT staff.
* **Customers**: Policyholders submitting claims.
* **Management**: Executives and department heads.
* **Vendors**: Providers of the AI data extractor solution.

**4. Communication Plan**

Develop a communication strategy to keep stakeholders informed and engaged:

* **Initial Announcement**: Inform all stakeholders about the upcoming change and its benefits.
* **Regular Updates**: Provide progress updates through emails, meetings, and newsletters.
* **Feedback Channels**: Establish channels for stakeholders to provide feedback and ask questions.

**5. Training and Support**

Ensure employees are adequately trained and supported during the transition:

* **Training Programs**: Conduct training sessions on using the AI data extractor solution.
* **User Manuals**: Provide comprehensive user manuals and guides.
* **Help Desk**: Set up a help desk to assist employees with any issues.

**6. Implementation Plan**

Outline the steps for implementing the AI data extractor solution:

* **Pilot Phase**: Start with a pilot phase to test the solution with a small group of users.
* **Full Rollout**: Gradually roll out the solution to all users based on the success of the pilot phase.
* **Monitoring and Evaluation**: Continuously monitor the implementation and evaluate its effectiveness.

**7. Risk Management**

Identify and mitigate potential risks associated with the change:

* **Technical Issues**: Ensure robust IT support to address any technical problems.
* **Resistance to Change**: Address employee concerns and resistance through effective communication and training.
* **Data Security**: Implement measures to ensure the security and privacy of data.

**8. Performance Metrics**

Define metrics to measure the success of the change:

* **Efficiency**: Measure the reduction in processing time for claims.
* **Accuracy**: Track the decrease in data entry errors.
* **Customer Satisfaction**: Monitor customer feedback and satisfaction levels.

**9. Review and Continuous Improvement**

Regularly review the change management process and make improvements as needed:

* **Feedback Collection**: Gather feedback from stakeholders to identify areas for improvement.
* **Continuous Improvement**: Implement changes based on feedback to continuously improve the process.

I hope this plan helps you manage the transition effectively. If you need any further assistance or have specific questions, feel free to ask! 😊

AI-generated content may be incorrect

**You said:**

Provide a GANTT chart as part of your submission showing: ○ What are the various milestones that should be a part of the Change Management plan? ○ What are the tasks within each milestone? Who is the owner of the task? ○ What is the timeline for each task? Provide a Start Date and End Date, and also show the duration visually in the GANTT chart.

**Copilot said:**

I understand that you need a GANTT chart to visually represent the milestones, tasks, owners, and timelines for the Change Management plan. While I can't create the GANTT chart directly in this document, I can provide you with a detailed outline that you can use to create the chart in a tool like Microsoft Project or Excel.

**GANTT Chart Outline for Change Management Plan**

**Milestone 1: Initiation**

* **Task 1.1: Stakeholder Identification**
  + **Owner**: Project Manager
  + **Start Date**: June 10, 2025
  + **End Date**: June 12, 2025
* **Task 1.2: Initial Announcement**
  + **Owner**: Communications Team
  + **Start Date**: June 13, 2025
  + **End Date**: June 14, 2025

**Milestone 2: Planning**

* **Task 2.1: Develop Communication Plan**
  + **Owner**: Communications Team
  + **Start Date**: June 15, 2025
  + **End Date**: June 20, 2025
* **Task 2.2: Develop Training Programs**
  + **Owner**: HR and Training Team
  + **Start Date**: June 21, 2025
  + **End Date**: June 30, 2025

**Milestone 3: Execution**

* **Task 3.1: Conduct Training Sessions**
  + **Owner**: HR and Training Team
  + **Start Date**: July 1, 2025
  + **End Date**: July 15, 2025
* **Task 3.2: Pilot Phase Implementation**
  + **Owner**: IT Team
  + **Start Date**: July 16, 2025
  + **End Date**: July 31, 2025

**Milestone 4: Monitoring and Evaluation**

* **Task 4.1: Monitor Pilot Phase**
  + **Owner**: IT Team
  + **Start Date**: August 1, 2025
  + **End Date**: August 15, 2025
* **Task 4.2: Collect Feedback**
  + **Owner**: Project Manager
  + **Start Date**: August 16, 2025
  + **End Date**: August 20, 2025

**Milestone 5: Full Rollout**

* **Task 5.1: Full Implementation**
  + **Owner**: IT Team
  + **Start Date**: August 21, 2025
  + **End Date**: September 10, 2025
* **Task 5.2: Continuous Monitoring**
  + **Owner**: IT Team
  + **Start Date**: September 11, 2025
  + **End Date**: Ongoing

You can use this outline to create a GANTT chart in your preferred tool. If you need any further assistance or have specific questions, feel free to ask! 😊

Copy

AI-generated content may be incorrect

**You said:**

Are there any dependencies between tasks inside a milestone or across milestones? Indicate the same.

**Copilot said:**

Yes, there are dependencies between tasks both within a milestone and across milestones. Here are the dependencies for the Change Management plan:

**Dependencies Within Milestones**

**Milestone 1: Initiation**

* **Task 1.2: Initial Announcement** is dependent on the completion of **Task 1.1: Stakeholder Identification**. The announcement can only be made once the stakeholders are identified.

**Milestone 2: Planning**

* **Task 2.2: Develop Training Programs** is dependent on the completion of **Task 2.1: Develop Communication Plan**. The training programs should align with the communication strategy.

**Milestone 3: Execution**

* **Task 3.2: Pilot Phase Implementation** is dependent on the completion of **Task 3.1: Conduct Training Sessions**. The pilot phase can only begin once the employees are trained.

**Dependencies Across Milestones**

* **Milestone 3: Execution** is dependent on the completion of **Milestone 2: Planning**. Execution tasks such as training sessions and pilot implementation can only start once the planning tasks are completed.
* **Milestone 4: Monitoring and Evaluation** is dependent on the completion of **Milestone 3: Execution**. Monitoring the pilot phase and collecting feedback can only begin once the pilot phase is implemented.
* **Milestone 5: Full Rollout** is dependent on the completion of **Milestone 4: Monitoring and Evaluation**. The full implementation and continuous monitoring can only start once the pilot phase is monitored and feedback is collected.

These dependencies ensure that each task is completed in a logical sequence, facilitating a smooth transition from one phase to the next. If you need further details or have any other questions, feel free to ask! 😊

**Question 4**

For the AI Data Extractor solution, what data should be collected to measure the business impact of the solution? Focus on the following details:

● What data should be collected at the beginning of the project to create a baseline? Describe this dataset in words. Additionally, provide details on what does each row in the dataset represents and what does each column represent. Hint: This data should allow you to estimate the cost and time involved with the manual data entry.

● What data should be collected after the solution goes live to provide the comparison? Describe this dataset in words. Additionally, provide details on what does each row in the dataset represent and what does each column represent.