



This handout accompanies our free course: [AI Mastery for Executives: A Unique, No-Bullshit Course on AI.](#)

# The Executive's Data Review Toolkit

## Introduction

---

Understanding the nuances of your AI systems isn't just the domain of technicians and engineers—it's a critical responsibility for executives. Your unique perspective can uncover issues that engineers will miss. This toolkit is designed to empower you to review AI data and align your AI's performance with business goals.

---

## Why Hands-On Data Review Matters

---

As an executive, you're in a unique position to bridge the gap between technical capabilities and business objectives. AI systems often interact with users in natural language, making the data more accessible than you might think. By personally reviewing AI interactions, you can:

- **Ensure Alignment with Business Values:** Verify that the AI represents your brand and communicates in a manner consistent with your company's ethos.
  - **Uncover Hidden Issues:** Spot user experience problems, interface issues, or workflow bottlenecks that aggregated reports might miss.
  - **Provide Valuable Feedback:** Your critiques can directly improve the AI's performance, much like coaching a new employee.
-

# The Data Review Process

## Step 1: Set Up a User-Centric Data Viewer

---

### Why This Matters

To effectively review AI interactions, you need to see them as your users do. Technical logs and observability platforms are filled with jargon and irrelevant details that can obscure the real user experience.

### Actions

1. **Collaborate with Your Team:** Discuss the need for a data viewer that mirrors the user interface your customers interact with.
2. **Eliminate Technical Distractions:** Ensure the viewer focuses solely on the AI-user interactions without backend codes, logs, or intermediate AI processing steps.
3. **Simplify Access:** The data viewer should be easily accessible—consider bookmarking it or setting it as your homepage for quick access.

### Pro Tips

- **Start with a Spreadsheet:** For simpler applications, a well-organized spreadsheet can be a quick and effective solution.
- **Prioritize Visual Clarity:** The data viewer should present information clearly, using visuals where appropriate to enhance understanding.
- **Include Key Context:** Make sure each interaction includes relevant details like timestamps, user segments, or any categorization that helps in understanding the context.

## Example of a Data Viewer

---

Here’s a real example of a data viewer used by the CTO of Rechat, that allows him to quickly look at pertinent data. This kind of thing can be built in just a few hours.

LLM Data Review

by Parliance Labs

Tool/Scenario Filter

Tool

All

Scenario

All

Status: Pending  
Record 5 of 186

Stats

Tool	Scenario	Pending	Rejected
Combination	None	182	4

Links

- [Airtable record](#)
- [Langsmith run](#)

Tool: CombinationScenario: NoneAuthor: Synthetic

ChatFuctions

is: MON Jan 08 2024 01:40:19 GMT-0800 (Pacific Standard Time).

USER(2/9)

Prepare an email draft with all the home staging services offered for a listing at 880 Broadway, and schedule the best time to post it on social media via the task manager.

ASSISTANT - Function Call(3/9)

listing-finder({"search":"880 Broadway"})

FUNCTION RESULTS(4/9)

```
[{'address': '880 W BROADWAY',
  'baths': 0,
  'beds': None,
  'close_date': None,
  'close_price': None,
  'id': 'cf9df90f-2fa6-40f5-993b-b4cf6fbab3a1',
  'mls': 'BRIGHT',
  'mls_number': 'PAYK2051636',
  'price': 3000,
  'status': 'Active',
  'type': 'compact_listing'}]
```

ASSISTANT - Function Call(5/9)

email-campaign-creator({"subject":"Home Staging Services for 880 Broadway Listing","recipients":[{"email":"socialmedia@example.com","recipient\_type":"Email"}],"body":"Hello,\n\nWe are excited to offer a range of home staging services for the listing at 880 Broadway. Our services include professional staging, virtual staging, and personalized design consultations. We ensure that every home is presented in the best possible light to attract potential buyers. If you have any questions or would like to schedule a consultation, please feel free to reach out to us. Best regards, Margery Kunisch","status":"draft"})

LLM Output (Editable)

```
"tags": null,
"recipients_count": null,
"text": "Hello, We are excited to offer a range of home staging services for the listing at 880 Broadway. Our services include professional staging, virtual staging, and personalized design consultations. We ensure that every home is presented in the best possible light to attract potential buyers. If you have any questions or would like to schedule a consultation, please feel free to reach out to us. Best regards, Margery Kunisch",
"status": "draft"
}
```

A task has been created to schedule the posting of the email on social media for the home staging services:

AcceptRejectBackResetNext⚡

From my [blog post](#) “Your AI Product Needs Evals”

https://exec-edu.parliance-labs.com/07\_asset\_data\_review\_toolkit.html?ck\_subscriber\_id=2859773944&utm\_source=convertkit&utm\_medium=email&utm\_campaign=Day 8: Our secret weapon for uncovering AI flaw... 3/21

You don't need to build anything as fancy as this, but it shows that it can be done. As an in-between step I often use [Airtable](#) for this task. Remember, do the simplest thing that could possibly work, including a simple spreadsheet.

## Step 2: Establish a Daily Data Review Routine

---

### Why This Matters

Consistent, hands-on review keeps you connected to your AI's performance and user experience. By dedicating a small portion of your day, you can catch issues early and guide your team more effectively.

### Actions

1. **Schedule Time on Your Calendar:** Treat data review as a critical meeting with yourself—block out 15–20 minutes daily.
2. **Focus on Failure Modes:** Prioritize reviewing interactions where the AI may have underperformed or failed to meet user needs.
3. **Review a Representative Sample:** Don't just look at problems; include successful interactions to understand what works well.

### Pro Tips

- **Set Themes for Each Day:** Focus on different features or user segments each day to cover more ground over time.
  - **Use Random Sampling:** To avoid bias, randomly select interactions to review alongside targeted ones.
  - **Keep Notes Handy:** Maintain a journal or digital note where you can jot down immediate thoughts or patterns you notice.
-

# Step 3: Categorize Data for Efficient Review

## Why This Matters

Organizing data into categories helps you spot patterns, understand context, and make your review process more efficient. It allows you to focus on specific areas that are critical to your business objectives. **Your team needs to categorize this data prior to you reviewing it.**

Categorize Before You Review

If the data is not categorized, it will be hard to navigate and analyze. This categorization can be automated with code that inserts tags into the data, with LLMs, or with a combination of both.

## Actions

- 1. **Categorize by Features, Tools, and Skills:** Identify the different functionalities your AI offers, such as email drafting, scheduling, or contact searching.
- 2. **Define Scenarios:** List specific conditions the AI must handle, like "Contact Not Found," "User Provided Invalid Date," or "Multiple Contacts Found."
- 3. **Create a Matrix or Table:** Use a visual aid to map out categories and scenarios for quick reference.

Below is an example of a matrix of features and scenarios that you might use to categorize interactions with an AI.

## Example Features and Scenarios Matrix

Feature	Scenario	Description
Contact Search	Contact Found	Successfully retrieved the correct contact information.
Contact Search	Contact Not Found	No matching contact; AI should suggest creating a new contact or check for typos.

Feature	Scenario	Description
Contact Search	Multiple Contacts Found	Multiple contacts match the query; AI should present options for the user to select.
Email Writing	Successful Send	Email drafted and sent without issues.
Email Writing	Formatting Errors	Issues with email layout, style, or inserted errors by non-native speakers worsening content.
Scheduling	Successful Booking	Meeting scheduled correctly with all details accurate.
Scheduling	Invalid Date Input	User provides an invalid date; AI should recognize and prompt for correction.
Scheduling	Conflicting Appointments	Scheduling conflicts detected; AI should notify user and suggest alternatives.
Data Entry	Error Handling	AI encounters errors during data input; should provide clear guidance to the user.
General Query	Unhandled Questions	AI doesn't understand the query; should ask for clarification or redirect appropriately.

Below is example of applying the above matrix to actual data:

## Example Data With Features and Scenario Grouping Applied

Scenarios Applied to Data

Interaction				
ID	Feature	Scenario	User Query	AI Response

Interaction				
ID	Feature	Scenario	User Query	AI Response
001	Contact Search	Contact Not Found	"Find John's contact info."	"No contact found for 'John'. Would you like to create a new contact?"
002	Scheduling	User Provided Invalid Date	"Schedule a meeting on April 31st."	"April has only 30 days. Please provide a valid date."
003	Scheduling	User Provided Time in the Past	"Set up a call yesterday at 3 PM."	"Cannot schedule meetings in the past. Please choose a future time."
004	Contact Search	Multiple Contacts Found	"Get Alex's email address."	"Multiple contacts named 'Alex' found: Alex Smith, Alex Johnson. Please specify."
005	Scheduling	Conflict with Existing Event	"Book a meeting with Mark at 2 PM tomorrow."	"You already have a meeting scheduled at 2 PM tomorrow. Would you like to reschedule?"
006	Email Writing	Attachment Missing	"Send the report to the team."	"Please attach the report file before sending the email."
007	Data Entry	Invalid Input Format	"Update sales figures to 'ten thousand'."	"Please provide the sales figures in numeric format."
008	Contact Search	User Provided Incomplete Info	"Find contact with phone number 555-1234."	"Multiple contacts have that number. Can you provide a name?"
009	General Query	Unhandled Question	"What's the weather today?"	"I'm sorry, but I cannot provide weather information."

For a more detailed explanation of the above table, click the blue button below.

👉 Click To Expand Explanation of Features and Scenarios 👉

#### Tip

The examples above illustrate the AI doing the right thing. The point of this example is to show example categories and scenarios, rather than show you examples of AI failures.

## Pro Tips

- **Use Specific Scenarios:** Clearly define conditions that the AI must handle, which helps in assessing its performance in various situations. By fleshing out these scenarios, you can pinpoint areas for improvement and ensure that the AI responds appropriately to user needs.
- **Reflect Real User Behavior:** Include scenarios that represent common user mistakes or challenges.
- **Use Color Coding:** Assign colors to different scenarios for quicker visual scanning.
- **Leverage Filters:** If using a spreadsheet or data viewer, use filter functions to focus on specific scenarios or features.
- **Update Regularly:** As new features are added or scenarios emerge, update your categories to keep them relevant.

## Step 4: Conduct Binary Evaluations

### Why This Matters

Simplifying your evaluation to a Yes or No question—"Did the AI solve the customer's problem?"—helps you make quick, decisive assessments without getting bogged down in complexity.

### Actions

1. **Ask the Key Question:** For each interaction, determine whether the AI met the user's needs.



2. **Avoid Complex Scoring Systems:** Resist the urge to rate on a scale; keep it simple to maintain consistency.
3. **Document Your Decision:** Clearly mark each interaction as a success or failure.

## Pro Tips

- **Consistency is Key:** Use the same criteria for each evaluation to ensure fairness and reliability.
  - **Trust Your Instincts:** Your business acumen is valuable—don't second-guess your initial judgment.
  - **Note Ambiguities:** If an interaction isn't a clear Yes or No, make a note and consider discussing it with your team for clarity.
- 

## Step 5: Write Constructive Critiques

---

### Why This Matters

Detailed, actionable feedback is essential for improving your AI system. Think of it as coaching an employee—the more specific you are, the better the AI can become.

### Actions

1. **Frame Feedback as Instructions:** Write critiques as if you're guiding a new team member on how to improve.
2. **Be Specific:** Point out exactly what went wrong and suggest how it could be corrected.
3. **Highlight Positive Outcomes:** Occasionally note why certain interactions were successful to reinforce good performance.

### Example Critiques

Below is an example of a data review log with detailed critiques that you might write.

Example Critiques

				Problem Solved (Y/N)	Critique	Metrics Aligned (Y/N)	Notes
2023-10-11	005	Contact Search	Multiple Contacts Found	N	<b>Critique:</b> "When the user searched for 'Alex', the AI found multiple contacts named 'Alex Johnson' and 'Alex Smith' but did not prompt the user to select the correct one. Next time, please display the list of matching contacts and ask the user to choose one."	N	Discuss with team
2023-10-11	006	Scheduling	Invalid Date Input	N	<b>Critique:</b> "The user tried to schedule a meeting on 'February 30th', which is an invalid date. The AI should recognize this and inform the user that the date doesn't exist, then prompt them to provide a valid date for scheduling."	Y	Improvement needed

				Problem Solved (Y/N)	Critique	Metrics Aligned (Y/N)	Notes
2023-10-11	007	Email Writing	Edits Worsening Content	N	<b>Critique:</b> "After drafting an email, the AI accepted user edits that introduced grammatical errors and unclear phrasing. The AI should assist users by suggesting corrections to maintain professionalism and clarity in communications, especially when edits reduce quality."	N	Needs attention
2023-10-11	008	Data Entry	Error Handling	N	<b>Critique:</b> "The AI encountered an error when updating sales figures but only displayed a generic error message. It should provide specific details about the	Y	Error specifics

				Problem Solved (Y/N)	Critique	Metrics Aligned (Y/N)	Notes
					error and guide the user on how to correct the input or whom to contact for support."		
2023-10-11	009	General Query	Unhandled Questions	N	<b>Critique:</b> "The user asked for 'company's quarterly revenue growth', but the AI responded with 'I can't assist with that request.' Instead, the AI should recognize this as a request for financial data and either provide the information or guide the user to the appropriate resource."	N	Expand knowledge
2023-10-11	010	Scheduling	Conflicting Appointments	N	<b>Critique:</b> "When scheduling a meeting at 3 PM on Thursday, the AI didn't alert the user	N	Calendar sync needed

Date	ID	Feature	Scenario	Problem Solved (Y/N)	Critique	Metrics Aligned (Y/N)	Notes
					of an existing appointment at that time. The AI should check the user's calendar for conflicts and suggest alternative times if necessary."		
2023-10-11	011	Contact Search	Contact Not Found	Y	<b>Praise:</b> "Good job informing the user that 'Emma Thompson' was not found and offering to create a new contact. This helps keep contact lists up to date and assists the user efficiently."	Y	Positive example

To see further explanation of the above table, click the blue button below.

 Click To Expand Explanation of Critiques 

**Tip**  
The above table contains a column named "Metrics Aligned (Y/N)". We will discuss metrics in the next step.

Pro Tips

- **Use Clear Language:** Avoid technical jargon; focus on the user experience.
  - **Be Objective:** Focus on the interaction, not the technology behind it.
  - **Prioritize Impact:** Spend more time on critiques that could significantly improve user satisfaction or business outcomes.
- 

## Step 6: Cross-Reference Metrics

---

### Why This Matters

Metrics are only valuable if they align with real-world outcomes. By comparing your evaluations with the associated metrics, you ensure that your team measures what truly matters.

### Actions

1. **Compare Evaluations with Metrics:** For each interaction, look at the metrics your team has recorded.
2. **Sanity-Check the Metrics:** Ask yourself whether the metrics accurately reflect the success or failure of the interaction.
3. **Provide Feedback on Metrics:** If you notice discrepancies, discuss them with your team to refine the measurement approaches.

### Example

- **Interaction ID:** 003
- **Your Evaluation:** Problem Not Solved
- **Metric Reported:** Success (Score of 90%)
- **Discrepancy:** "The AI didn't recognize an invalid date, but the metric indicates a high success rate. This suggests the metric isn't capturing date validation errors."

### Pro Tips

- **Understand Key Metrics:** Familiarize yourself with the metrics your team uses to evaluate AI.

- **Look for Patterns:** If certain metrics misalign with your evaluations, it indicates a need for metric refinement.
  - **Collaborate on Solutions:** Work with your team to adjust metrics to better align with user satisfaction and business goals.
- 

## Step 7: Share Insights and Lead by Example

---

### Why This Matters

Your active involvement signals to your team the importance of data review and continuous improvement. It fosters a culture where everyone is engaged in improving AI performance.

### Actions

1. **Share Your Findings:** Present interesting insights or issues during team meetings or via email updates.
2. **Encourage Open Dialogue:** Invite team members to discuss their observations and suggestions.
3. **Celebrate Successes and Learn from Failures:** Acknowledge both the AI's strengths and areas for improvement.

### Pro Tips

- **Lead by Example:** Your commitment will inspire others to take data review seriously.
  - **Create a Feedback Loop:** Establish regular meetings or channels where data review insights can be shared and acted upon.
  - **Recognize Contributions:** Highlight team members who make significant improvements based on data review.
- 

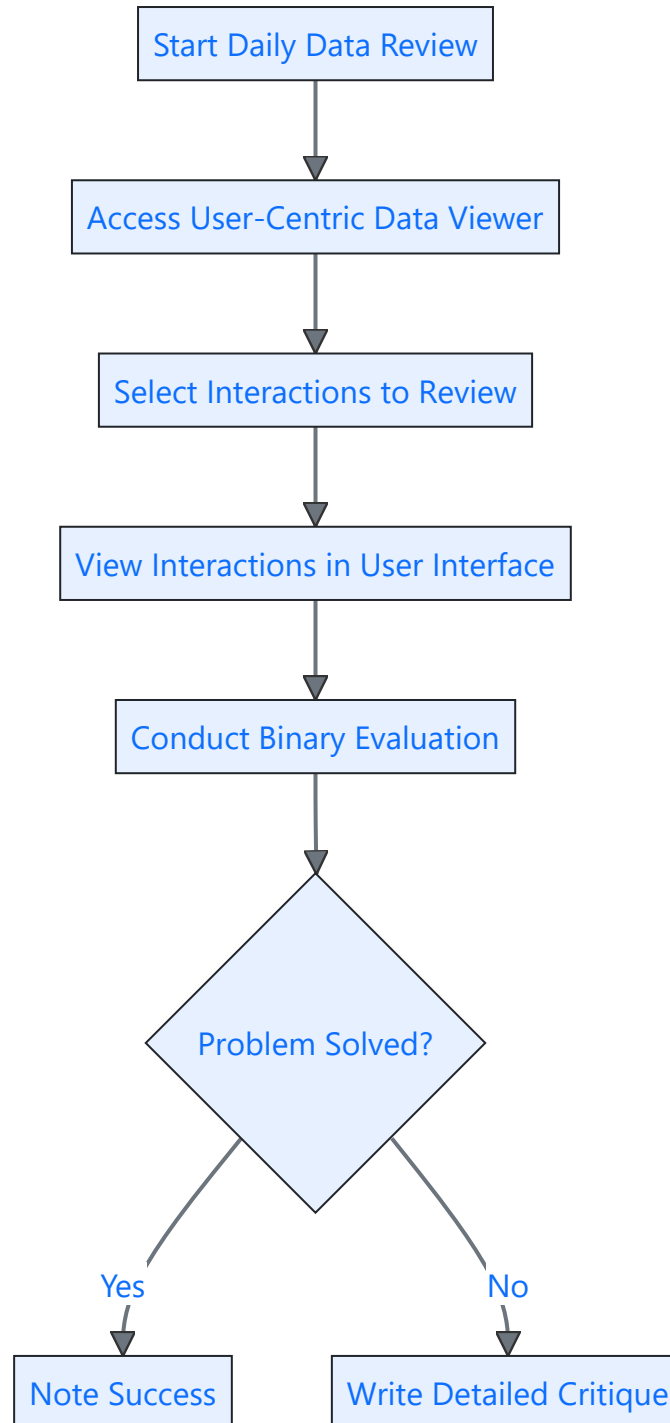
## Recap of the Process

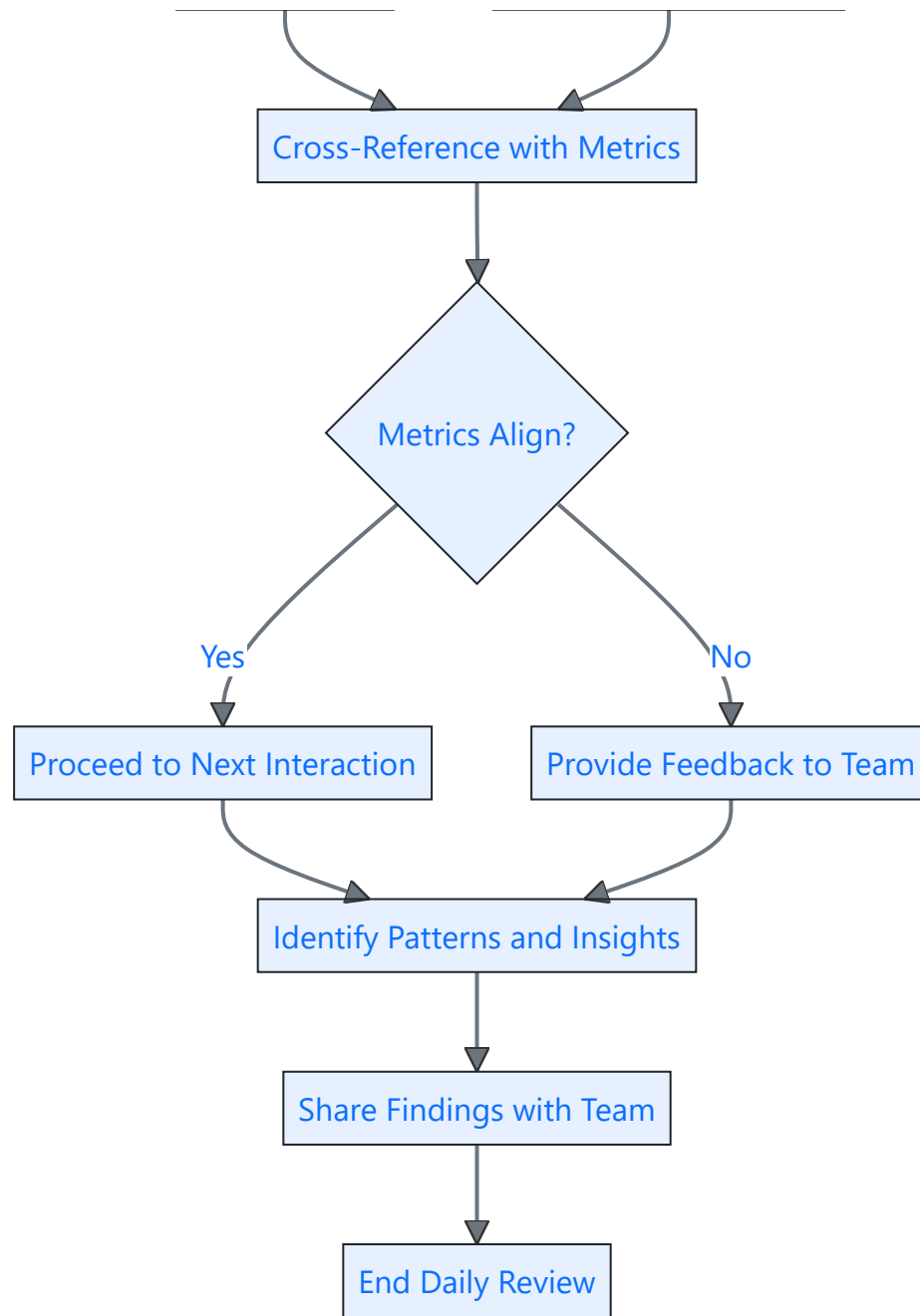
---

Below is a flowchart that summarizes the data review process.









To see further explanation of the above flowchart, click the blue button below.

 Click To Expand Explanation of the Flowchart 

## Common Pitfalls to Avoid

Getting this process right takes time and practice. Here are some common pitfalls to avoid:

Pitfall	Issue	Solution
1. Relying Solely on Technical Tools	Observability platforms and technical logs are designed for engineers and can be overwhelming.	Use a user-centric data viewer that presents interactions as customers experience them.
2. Avoiding Data Review Due to Friction	Complexity and technical barriers can discourage regular data review.	Simplify the process with easy-to-access tools and a streamlined workflow.
3. Trying to Outsource Data Review	Delegating this task entirely to others or relying on AI to review AI can miss critical insights.	Personally engage in data review to leverage your unique perspective.
4. Overcomplicating Evaluations	Using complex scoring systems can create inconsistency and confusion.	Stick to a simple Yes or No evaluation to maintain clarity.
5. Ignoring Failure Modes	Focusing only on successes doesn't address areas needing improvement.	Prioritize reviewing and critiquing interactions where the AI failed to meet expectations.

Pitfall	Issue	Solution
6. Not Providing Specific Feedback	Vague critiques are less actionable and don't guide improvement.	Offer detailed, specific feedback as you would to a human employee.

## Best Practices for Effective Data Review

In addition to common pitfalls, here are some best practices to follow:

- **Stay User-Focused:** Always consider interactions from the user's perspective.
- **Be Consistent:** Regular reviews yield better insights over time.
- **Collaborate with Your Team:** Use your findings to guide team efforts and improvements.

## Conclusion

Embracing a hands-on approach to AI data review empowers you to:

- **Align AI Performance with Business Goals:** Ensure your AI systems are meeting the needs of your customers and representing your brand effectively.
- **Enhance AI Capabilities:** Your feedback directly contributes to improving the AI, much like mentoring a high-potential employee.
- **Lead by Example:** Foster a culture of continuous improvement within your team.
- **Inform Strategy:** Use data review insights to shape AI strategy and priorities.

By integrating this data review process into your routine, you're not just overseeing an AI system—you're actively shaping a powerful tool that can significantly enhance your organization's performance. Your hands-on involvement ensures that the AI aligns with your vision and delivers genuine value to your customers.

*Remember, your AI is like a super-employee capable of exponential impact. Investing time in its development and alignment is not just beneficial—it's essential.*

---

By following this comprehensive toolkit, you'll be well-equipped to lead your AI projects to success, ensuring they align with your business values and deliver real value to your customers.

---

© 2024 [Parlance Labs](#). All rights reserved.