Potential Performance Metrics for Mastercard AI Governance Scorecard

Hallucination Rate

Definition & Relevance

- **What is it?**
  The Hallucination Rate measures the frequency at which a Generative AI system produces responses that are factually incorrect or misleading, deviating from its training data or intended outputs.
- **Why is it important?**
  High hallucination rates in AI-generated content can lead to misinformation, eroding user trust and potentially causing harm, especially in critical sectors like finance and healthcare.
- **Where is it used?**
  This metric is crucial in applications such as AI-powered customer support, financial advisory services, and fraud detection systems, where accuracy and reliability are paramount.

Tracking & Implementation

- **How should this metric be measured?**
  - Implement automated fact-checking tools to validate AI outputs against verified datasets.
  - Utilize confidence scoring models to flag uncertain or potentially erroneous AI-generated responses.
  - Establish human review processes for high-stakes or sensitive AI-generated content.
- **What data sources are needed?**
  AI-generated response logs, confidence scores, user feedback, and benchmark datasets are essential for accurate measurement.
- **How does this metric integrate into Mastercard's AI framework?**
  Monitoring the Hallucination Rate ensures that AI systems align with Mastercard's standards for accuracy, transparency, and regulatory compliance.

Challenges & Considerations

- **What are the limitations of tracking this metric?**
  Fact-checking AI outputs can be complex due to evolving data and contextual nuances. Manual reviews may also be resource-intensive.

- **Are there ethical or regulatory concerns?**
  Inaccurate AI outputs can mislead users, leading to ethical dilemmas and potential regulatory violations, particularly in sectors like finance and healthcare.
- **What are potential solutions or improvements?**
  Enhancing AI models with retrieval-augmented generation (RAG) techniques and continuous fine-tuning using domain-specific knowledge can help mitigate hallucinations.

Use Case Example

- **Healthcare Sector:**
  A transcription tool powered by OpenAI's Whisper model was found to hallucinate in about 1% of transcriptions, sometimes inventing sentences or nonsensical phrases during silences. This highlights the importance of monitoring and mitigating hallucinations to ensure reliable medical documentation.

User Override Rate

Definition & Relevance

- **What is it?**
  The User Override Rate tracks how often human users reject, modify, or override AI-generated recommendations before implementation.
- **Why is it important?**
  A high override rate may indicate issues with the AI system's accuracy, relevance, or user trust, suggesting a need for system improvements.
- **Where is it used?**
  This metric is applicable in areas like risk assessment, fraud detection, customer support automation, and financial decision-making processes.

Tracking & Implementation

- **How should this metric be measured?**
  - Implement user feedback mechanisms allowing analysts to approve, modify, or reject AI recommendations.
  - Monitor the frequency and context of human interventions across various AI applications.
  - Analyze override trends over time to identify patterns and areas for improvement.

- **What data sources are needed?**
 AI recommendation logs, records of manual interventions, user feedback, and real-time decision-tracking data are necessary.
- **How does this metric integrate into Mastercard's AI framework?**
 Understanding the User Override Rate helps Mastercard assess the effectiveness and reliability of AI systems, ensuring they enhance rather than hinder decision-making processes.

Challenges & Considerations

- **What are the limitations of tracking this metric?**
 A high override rate doesn't always signify AI failure; it could reflect conservative decision-making or a lack of user training.
- **Are there ethical or regulatory concerns?**
 AI decisions impacting financial approvals or fraud detection must be explainable to avoid biases and comply with regulations.
- **What are potential solutions or improvements?**
 Incorporating explainability models (e.g., SHAP values) can help users understand AI recommendations, potentially reducing unnecessary overrides.

Use Case Example

- **Healthcare Sector:**
 In clinical settings, clinicians often override AI-generated alerts for potential drug interactions, indicating a need for more accurate and relevant alert systems.

AI-Driven Value Attribution Score

Definition & Relevance

- **What is it?**
 The AI-Driven Value Attribution Score measures the extent to which business outcomes, such as revenue growth, cost savings, or improved customer engagement, can be directly attributed to AI-driven decisions and actions.
- **Why is it important?**
 This metric ensures that investments in AI technologies are delivering measurable benefits, allowing organizations to assess the return on investment (ROI) and make informed decisions about future AI initiatives.

- **Where is it used?**
  It’s utilized across various domains, including marketing, sales, customer service, and operations, where AI tools are implemented to enhance performance and efficiency.

Challenges & Considerations

- **What are the limitations of tracking this metric?**
  - Attribution Complexity: Determining the exact contribution of AI to specific business outcomes can be challenging due to multiple influencing factors.
  - Data Quality: Accurate attribution requires high-quality, comprehensive data, which may not always be available.
  - Dynamic Environments: Rapid market changes can affect the consistency of attribution models, leading to potential inaccuracies.
- **Are there ethical or regulatory concerns?**
  - Transparency: AI-driven decisions must be explainable to stakeholders to build trust and comply with regulations.
  - Bias and Fairness: There’s a risk that AI models could perpetuate existing biases, leading to unfair outcomes.
  - Privacy: Utilizing customer data for AI analysis must adhere to privacy laws and ethical standards.
- **What are potential solutions or improvements?**
  - Implement Explainability Models: Using tools like SHAP (Shapley Additive explanations) values can help elucidate AI decision-making processes, enhancing transparency.
  - Continuous Monitoring: Regularly updating and validating AI models ensures they adapt to changing environments and maintain accuracy.
  - Cross-Functional Collaboration: Engaging diverse teams in AI development can help identify and mitigate biases, promoting fairness.

Use Case Example

- **Marketing Attribution:**
  Adobe’s Attribution AI enables marketers to understand the impact of each customer interaction across their journey, facilitating executive reporting, budget allocation, and campaign optimization.
- **Sales Optimization:**
  AI-driven lead scoring systems help businesses prioritize prospects, increasing conversion rates and sales efficiency.

Works Cited

Hallucination Rate

1. IBM. (2023). *Understanding AI Hallucinations: Risks & Mitigations.* Retrieved from https://www.ibm.com/think/topics/ai-hallucinations
2. AI Business. (2023). *Combating Generative AI's Hallucination Problem with RAG Models.* Retrieved from https://aibusiness.com/nlp/combating-generative-ai-s-hallucination-problem
3. Stanford AI Research. (2024). *Legal AI Research Report on Hallucinations in LLMs.* Retrieved from https://dho.stanford.edu/wp-content/uploads/Legal_RAG_Hallucinations.pdf
4. The Verge. (2024). *OpenAI's Whisper & Hallucination Issue in Healthcare Transcriptions.* Retrieved from https://www.theverge.com/2024/10/27/24281170/open-ai-whisper-hospitals-transcription-hallucinations-studies

User Override Rate

1. National Library of Medicine (PMC). (2023). *User Override Rates in AI Alert Systems.* Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC10552880/
2. HighRadius. (2023). *Using AI to Forecast Accounts Receivable & User Override Rates.* Retrieved from https://www.highradius.com/resources/Blog/using-ai-to-forecast-account-receivables/

AI-Driven Value Attribution Score

1. PwC. (2023). *Artificial Intelligence ROI: Measuring Business Impact.* Retrieved from https://www.pwc.com/us/en/tech-effect/ai-analytics/artificial-intelligence-roi.html
2. McKinsey. (2023). *The Economic Potential of Generative AI: Measuring Business Value.* Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
3. Forbes. (2023). *How Amazon Uses AI to Boost Revenue Through Personalization.* Retrieved from

https://www.forbes.com/councils/forbescommunicationscouncil/2024/01/05/ai-and-personalization-in-marketing/

4. Adobe Experience Platform. (2023). *Attribution AI Overview.* Retrieved from https://experienceleague.adobe.com/en/docs/experience-platform/intelligent-services/attribution-ai/overview
5. Invoca. (2023). *30 Outstanding Examples of AI in Marketing.* Retrieved from https://www.invoca.com/blog/outstanding-examples-ai-marketing
6. Relevance AI. (2023). *Lead Scoring and Prioritization AI Agents.* Retrieved from https://relevanceai.com/agent-templates-tasks/lead-scoring-and-prioritization-ai-agents