CHAI Applied Model Card

Name: Prior Authorization **Developer:** Penguin AI	**Inquires or to report an issue:** fawad@penguinai.co
Release Stage: In-Production **Release Date:** 08/23/2024 **Version:** V1 **Global Availability:** United States of America **Regulatory Approval:** N/A	
Summary: Prior Authorization (PA) Small Language Model (SLM) is designed to assist SMEs while adjudicating/ submitting a PA request. The model is designed to re-imagine the PA decision-making process and make it quicker and more transparent. It incorporates human-in-the-loop feedback mechanisms from SMEs for continuous improvement and accuracy enhancement.	**Uses and Directions:** • **Intended use and workflow:** ◦ The solution is intended to augment payer PA back-office operations teams in getting their work done faster or in reducing PA denials for providers. The SLM automatically analyzes submitted documents and highlights specific sections containing relevant evidence, eliminating the need to read through all pages of documentation. This targeted approach substantially reduces review time and administrative burden. ◦ Additionally, the system provides valuable predictive capabilities at the point of submission, alerting submitting physicians about the probability of request approval ◦ The solution addresses PA requests by analyzing structured and unstructured data and supporting documentation • **Primary intended users:** ◦ Healthcare professionals involved in prior authorization processes ◦ Users should have familiarity with prior authorization procedures, patient documentation and clinical guidelines such as Milliman, InterQual etc. • **How to use:** Our proprietary SLM model comes ready to deploy out-of-the-box, delivering immediate value with minimal setup. For organizations seeking a

personalized touch, users can fine-tune the model on their data while maintaining complete data privacy and ownership. Whether you choose the base model, a custom-tuned version, or the full Digital Worker solution, everything is accessible through secure API endpoints in our HIPAA-compliant Penguin platform.

- **Targeted patient population:** General patient population requiring PA for healthcare services. No specific exclusion criteria based on demographic factors
- **Cautioned out-of-scope settings and use cases:** Not intended for decision-making without human review and should not be used as the sole basis for prior authorization decisions. Also, any changes in medical guidelines and criteria has to be updated on a regular basis

Keywords: None

Warnings

- **Known risks and limitations:**
  - Accuracy of the model improves with usage
  - Places high weightage to human input during RLHF
  - Performance dependent on quality of OCR of input clinical notes
  - May not capture rare or unusual clinical scenarios
- **Known biases or ethical considerations:** Potential data biases inherited from training data
- **Clinical risk level:** Low

Trust Ingredients

AI System Facts:

- **Outcome(s) and output(s):**

  Outcomes:

  - Identify medical evidence according to guidelines to approve or deny a PA request
  - Reduces admin burden by bringing down the time to review documents
  - Highlight missing information at the point of submission to reduce PA denials

  Outputs:

  - Decision such as Approve, Deny & pending state
  - Explainability - Why the request is approved or denied
  - Traceability of the decision back to the submitted documentation
- **Model type:** GPT model; Sentence transformers; Re-ranking models; Semantic chunking algorithms
- **Foundation models used in application:** Llama3
- **Input data source:** Data related to real world Prior Authorizations was used to train the model

- **Output/Input data type:** Clinical notes and documentation, delivered as Unstructured and Structured Data
- **Development data characterization:**
  - Data related to real world Prior Authorizations was used to train the model
  - Supplementary synthetic data was generated and validated by SMEs
- **Bias mitigation approaches:**
  - Input data assessment
  - PII redaction to reduce the impact of certain demographics on model output
  - Data governance by having input data reviewers from business, compliance & tech teams
  - Mitigating input data bias by identifying rows/ annotations that are not congruent (to the Prior Authorization task)
  - Measurement of model bias on validation datasets
  - Medical Director review of data & model to ensure clinical appropriateness
  - Leveling data issues with under-represented demographics
  - Continuous monitoring of model accuracy on different demographics on real-world production data
- **Ongoing Maintenance:**
  - Continuous feedback collection from SMEs to implement RLHF
  - Regular model fine-tuning based on collected feedback
  - Performance monitoring to measure human machine congruence
  - Metrics to measure human engagement within the workflow
  - Training of the scenarios/ broad themes where the model is failing
- **Security and compliance environment practices or accreditations:**
  - **Accreditation**: HIPAA compliance, SOC-2 compliance, Hi-Trust compliance
  - **Security and Compliance Practices**:
    1. Robust Data Protection mechanism including data encryption at rest and in transit using industry-standard protocols.
    2. Multi-tenant architecture is crucial for proper client data segregation, supplemented by strict access controls and identity management.
    3. Network security should implement segmentation, zero-trust architecture, and multiple security layers including VPNs, firewalls, and intrusion detection systems.
- **Transparency, Intelligibility, and Accountability mechanisms:**
  - Human-in-the-loop review process
  - Documented feedback collection and incorporation
  - Model outputs not just the decision but also the reason (supporting evidence within the document) for the decision

Transparency Information:

- **Funding source of the technical implementation:** Penguin AI
- **3rd Party Information:** N/A
- **Stakeholders consulted during design of intervention (e.g. patients, providers):** Medical Director, Nurses

Key Metrics

Usefulness, Usability, and Efficacy	Fairness and Equity	Safety and Reliability
Goal of metric: To evaluate the model's practical effectiveness in Prior Authorization processing	**Goal of metric:** To ensure equitable model performance across diverse demographic groups, medical	**Goal of metric:** To verify model safety, reliability, and compliance with healthcare standards while

through comprehensive assessment of data quality, model performance, and task-specific capabilities.		conditions, and healthcare scenarios while maintaining consistent accuracy and fairness.		ensuring consistent and appropriate outputs across all operational conditions	
Result: Sensitivity of 92.13% and specificity of 91.17% **Data Completeness**: 100% **Unique Record Rate**: 100% **Task Relevancy**: 100% **Data Efficacy**: 80%	**Interpretation:** Following metrics are measured: **Data Completeness**: The proportion of data points with complete values across all fields. **Unique Record Rate**: The proportion of records that are unique in the dataset. **Task Relevancy**: The degree to which the input data aligns with the specific task requirement. **Data Efficacy**: The degree to which output (response) data contains information present in or relevant to the input (prompt) data.	**Result:** **Overall Category Score**: 89% **Data Governance Metrics**: • **Bias Mitigation Score**: 99% • **Criteria Fulfillment Rate**: 99% **Model Evaluation Metrics**: • **Demographic Parity**: 87%	**Interpretation:** Bias Mitigation Score: The degree to which the distribution of demographic attributes in the dataset aligns with the target distribution Criteria Fulfillment Rate: Users would provide the list of criteria against which data should be evaluated to check whether it satisfies all conditions	**Result:** **Overall Category Score**: 99% **Data Governance Metrics**: • **Safety Compliance Rate**: 100% **Model Evaluation Metrics**: • **Model Safety Score**: 99%	**Interpretation:** **Safety Compliance Rate**: The proportion of inputs in the training data that are deemed safe and appropriate

Test Type: • Question accuracy rate • Top-N Context relevance	**Test Type:** Comprehensive fairness assessment including: • Bias detection and measurement • Cross-demographic performance analysis • Multi-stakeholder equity evaluation • Fairness criteria compliance testing	**Test Type:** Safety assessment protocol: **Model safety testing**: • Prompt injection • Document engineering **Reliability testing**: • Reproducibility • Error handling validation • System reliability testing • Performance stability assessment **Security penetration testing**
Testing Data Description: • Historical Prior Authorization requests • Synthetic edge cases and boundary conditions • Multi-payer authorization scenarios • Diverse medical specialty cases • Real-world denial/approval patterns • Out-of-scope query samples	**Testing Data Description:** Specialized dataset incorporating: • Demographically diverse patient populations • Multiple insurance types and plans • Various healthcare provider settings • Different geographic regions • Range of medical conditions	**Testing Data Description:** Comprehensive safety test suite including: • Prompt injection datasets • LLM generated engineered documents • Protected health information scenarios • Rare medical conditions • Complex approval scenarios involving logical/ numeric reasoning • System stress test cases • Security vulnerability tests
Validation Process and Justification: Three-stage validation process: • Automated testing against ground truth • Clinical expert review of outputs • Integration testing in sandbox environments	**Validation Process and Justification:** • Statistical fairness testing • Demographic representation analysis • Performance equity verification • Regular bias monitoring and reporting • Independent fairness audits • Stakeholder feedback integration	**Validation Process and Justification:** Multi-layer safety validation: • Automated safety checks • Compliance audit • Security assessment • Continuous monitoring system • Incident response testing • Regular compliance verification

Resources

- **Evalation References:** Evaluation report or publication, if available: No publication - evaluation metrics are shared above
- **Clinical Trial:** N/A
- **Peer Reviewed Publication(s):** N/A
- **Reimbursement Status:** N/A
- **Patient consent or disclosure required or suggested:** N/A

References

Note: The mention or sharing of any examples, products, organizations, or individuals does not indicate any endorsement of those examples, products, organizations, or individuals by the Coalition for Health AI (CHAI). Any examples provided here are still under review for alignment with existing standards and instructions. We welcome feedback and stress-testing of the tool in draft form.

The information provided in this document is for general informational purposes only and does not constitute legal advice. It is not intended to create, and receipt or review of it does not establish, an attorney-client relationship.

This document should not be relied upon as a substitute for consulting with qualified legal or compliance professionals. Organizations and individuals are encouraged to seek advice specific to their unique circumstances to ensure adherence to applicable laws, regulations, and standards.

For instructions, references, contributors, and disclaimers please refer to the full documentation located at www.chai.org.