# Generative AI Evaluation Plan for Informed Consent Master Plan

**Date:** 7 Apr 2025

**Sponsor: Acme Corp**

**Prepared by:** John Smith  
**Intended Audience:** Clinical Development, Medical Writing, Regulatory Affairs, AI Governance

## 1. Purpose of the AI Evaluation Plan

This document outlines the **evaluation framework** for assessing artificial intelligence (AI) -generated clinical trial informed consent documents to ensure appropriate **readability, regulatory compliance, clarity, consistency, accuracy, and bias** before stakeholder review and submission.

## 2. Scope of AI-Generated Protocol Evaluation

This document provides guidance for individual studies, applies to all studies involving patients, and where the informed consent documents are **partially or fully generated** using **Generative AI** models. The scope of this document applies to the following approved instances of utilization of AI for the purposes below:

🔹 **Automated Drafting**:

* AI can generate **initial drafts** of informed consent documents based on structured templates, trial protocols, and regulatory guidelines.
* Customization for **specific studies, patient populations, and trial phases**.

🔹 **Personalization & Adaptation**:

* **Demographic-specific tailoring** (e.g., different literacy levels, languages, cultural sensitivities).
* **Dynamic adaptation** for different conditions (e.g., pediatric trials vs. oncology trials).

🔹 **Alternative Consent Modalities**:

* AI can assist in **simplifying consent** for vulnerable populations (e.g., children, elderly, cognitively impaired).
* Generating **multimodal consent** (text-to-speech, video-based consent)

## 3. Quantitative Evaluation Metrics and Review Process

Table 1.0 contains evaluation categories and criteria which may be used in the evaluation of informed consent documents (ICD) where document text has been fully or partially constructed or adapted. Clinical study teams should select the appropriate criteria according to specific trial contexts and documented in the Trial-Level Informed Consent Document Evaluation Summary Report.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Category | Evaluation Criteria | Quantitative Metric | Threshold for Acceptance | Reviewer(s) |
| Clarity | Use of common Words |  | Dale-Chall List | 80%+ Common words |  |
|  | Readability & Comprehension | Ensure patient-friendly language, avoiding jargon | Flesch-Kincaid Grade Level | Grade Level 6-8 | Patient advocate, ethics committees |
|  | Visual clarity |  |  |  |  |
|  | Scientific Accuracy | Alignment of study procedures, risks, and benefits with protocol | Cosine similarity  NLP similarity score with study protocol | >0.85 (high similarity with protocol) | Clinical SME, IRB reviewers |
|  | Regulatory Compliance | Adherence to FDA, EMA, ICH-GCP, and HIPAA | AI-assisted checklist completion rate | 100% compliance criteria met | Regulatory Affairs, Ethics board |
|  | Risk-Benefit Balance | Clarity and neutrality in presenting risks vs. benefits | Sentiment Analysis (neutral tone balance) | <10% sentiment bias | Ethics Committee, Legal review |
|  | Alternative Treatments | Clear disclosure of alternative treatment options | AI-extracted mentions of alternatives | At least on alternative mentioned | Medical writing, ethics |
|  | Bias & Fairness | Avoids coercive language, ensures fair representation | Bias detection model score | <5% deviation from historical trial demographics or diversity standards | ethics |
|  | Free from promotion |  |  |  |  |
|  | Hallucination Rate | Instances where AI generates incorrect/unverifiable claims | % of fabricated content flagged by reviewers | <2% hallucination rate | Medical Writing, Clinical SME |
|  | Informed Decision-Making | Emphasizes voluntary participation and withdrawal rights | AI-check for presence of “voluntary” and “withdraw” | Both terms must be explicitly present | IRB, legal, patient advocate |
| Consistency | Contradiction Detection |  |  |  |  |
|  | Factual consistency |  |  |  |  |

**4. Evaluation Workflow**

The evaluation workflow outlines the necessary steps to evaluate ICDs that are either partially or fully generated by AI.

**Step 1: State AI Model Selection & Versioning**

Complete the Trial-level AI-Generated Informed Consent Document Evaluation Summary Report (T-AICD) Sections: Study & AI Parameters and Section 1a.

**Step 2: Determine Appropriate Evaluation Checks & Execution**

Study teams should determine which measures in Table 1.0 are appropriate for use in the evaluation and document them in the T-AICD.

Study teams should identify and document approved thresholds for each metric prior to evaluation execution.

Following the pre-specification of evaluation metrics and thresholds, the initial execution of evaluation criteria and associated results should be documented in the T-AICD.

**Step 3: Apply Human Expert Review & Scoring**

Each evaluation metric result should be assigned to domain experts (clinical, regulatory, ethics review) for review of quantitative assessments of model performance.

Reviewers acknowledge and approve of the evaluation metrics acceptability following a comparison between AI-generated protocol sections and acceptance thresholds.

In coordination with the appropriate cross functional teams responsible for AI performance, where threshold criteria were exceeded, the AI model and/or pipeline should be modified until the required quality metrics are within acceptable thresholds.

**Step 4: Revision & Iteration**

As needed, conduct AI model and/or pipeline refinements, based on human feedback.

Compare before/after revisions using quantitative quality metrics and iterate until all metrics fall within acceptable thresholds for quality.

**Step 5: Final Validation & Sign-Off**

Approval from key stakeholders (Clinical, Regulatory, Biostats, Medical Writing) should be documented in the T-AICD.

Store version-controlled informed consent storage with an AI evaluation audit trail

AI Model Governance: Ensure explainability & compliance with company AI policies

Regulatory & Sponsor Transparency: Provide AI-assisted sections with documented human review confirmations

**6. Conclusion**

This evaluation plan ensures **scientific rigor, regulatory adherence, and clinical quality** when using Generative AI in protocol development. The combination of **quantitative AI assessment + human expert review** balances innovation with compliance.