IRM Requirements – Nabla Ambient AI Model Card Mapping

Based on Nabla Ambient AI Model Card, Nov 2025 v2
Review completed March 11, 2026

---

1. Risk Analysis, Potential Risks, and Adverse Impacts

Identify known and potential risks: Nabla identifies transcription accuracy as a key risk, noting the system achieves an average of 95% accuracy but may misinterpret complex medical terminology in challenging audio environments. Clinical risk level is rated as **low**, as no clinical decision-making is performed by the AI.

Potential consequences for users or patients: Clinically significant inaccuracies were noted "occasionally" in a randomized clinical trial (Likert score 2.8 [95% CI, 2.6–3.0]). One Grade 1 (mild) adverse event was reported in the RCT. Consent requirements are clearly documented — if a patient does not consent, the technology must not be used.

Near misses / scenarios considered: Weekly audits flag risk findings across categories including hallucination, translocation of information, omission, incorrect information, and speaker confusion. Audit results average 5% risk findings weekly and are incorporated into the product lifecycle.

---

2. Validity, Reliability, Robustness, Fairness, Intelligibility, Validity & Reliability:

- Transcript and clinical note quality scored **48 out of 50** using the Physician Documentation Quality Instrument (PDQI), designated by CHAI as the top assessment type for Ambient AI.
- Validated across 3,442 unique clinicians and 303,266 encounters across 6 specialties and 3 clinics.
- Peer-reviewed RCT published in NEJM AI demonstrated a 9.5% decrease in time-in-note vs. control group.

Robustness:

- Weekly model re-evaluation and updates address clinician feedback, healthcare taxonomy updates (e.g., ICD-10), client feature requests, and IT/security patching.
- Model versioning is maintained and monitored for drift.

Fairness & Equity:

- Assessed using the **FAVES model** (fair, appropriate, valid, effective, safe).
- Bias mitigation includes generating notes from transcript (not raw audio) to remove pitch/tone bias; diverse training data across national/regional accents, dialects, urban/rural settings, specialty clinics, and age groups; fairness reviews checking word error rates across language variations.

Intelligibility:

- Clinical note summaries can be traced back to original transcribed statements for explainability.
- Natural language explanations are used throughout outputs.
- Audit trails of interactions are provided to clinicians.

---

3. Safety, Security, and Privacy

Patient/User Safety:

- Human-in-the-loop review is built into the core model design and is required for all AI-generated content.
- Consent from the patient (or legal guardian) is required before use and is captured in the clinical note.
- Low clinical risk classification — no autonomous clinical decision-making.

Data Security:

- **SOC 2 Type 2** and **ISO 27001** certified.
- Follows **NIST** and **HITRUST** frameworks for internal security controls.
- Google and Azure serve as sub-processors for core infrastructure.
- Further security validation reports available at trust.nabla.com.

Privacy & Regulatory Compliance:

- **HIPAA** and **GDPR** compliant.
- Client data is **not used to train models by default**.
- Training data consists of synthetic medical encounters, medical terminology databases, and de-identified client feedback.
- Minimal demographic data captured — limited to what is clinically mentioned (language, sex, date of birth per ONC's 45 CFR 170.315).

4. Risk Mitigation and Governance

Policies & Controls:

- Weekly model re-evaluation cycle incorporating clinician feedback, audit findings, taxonomy updates, and security patching.
- Human-in-the-loop is a core design requirement, not optional.
- User-friendly feedback mechanisms are built into the product for ongoing risk identification.

Governance Structures:

- External medical auditing firm provides three layers of verification: (1) medical auditors, (2) operational review by the auditing firm, and (3) review by Nabla's clinical and compliance staff.
- Model versioning is maintained and monitored for drift.
- Accountability mechanisms include error reporting built into the product feedback workflow.
- Audit trails are provided to clinicians for oversight.

5. Data Practices

Data Acquisition:

- Input is audio from the clinical encounter. Contextual patient data from the EHR can optionally be used (not required).
- Input variables include language, gender, date of birth/age, and health status assessment.

Data Storage & Management:

- Hosted on Google and Azure infrastructure with SOC 2 Type 2 and ISO 27001 controls.
- Operational data is real-world clinical setting data; training data is synthetic, from medical databases, and de-identified client feedback.

Data Use:

- Three outputs generated: transcript, clinical note, and patient instructions.
- Client data is not used for model training by default.

- De-identified weekly audit samples are used across diverse clinical settings to verify output quality.
- Bi-directional EHR integration pulls demographic/clinical data and pushes structured notes, patient instructions, vitals, and problem lists.

Summary Assessment

IRM Requirement	Status	Notes
Risk Analysis & Adverse Impacts	Met	Risks documented, adverse event reported in RCT, weekly audits
Validity, Reliability, Fairness, Intelligibility	Met	PDQI scoring, FAVES model, bias mitigation, explainability
Safety, Security & Privacy	Met	SOC2, ISO27001, HIPAA, GDPR, human-in-loop
Risk Mitigation & Governance	Met	Weekly audits, external auditing firm, 3-layer verification
Data Practices	Met	Minimal data capture, no client training data by default, EHR integration documented