

Finding the Conversation in a Haystack:

Leveraging AI to Detect Goals-of-Care Documentation

Deborah Unger, MD FAAHPM, Matthew Gonzales, MD FAAHPM, Melissa Forbin, MBA, Suzanne Engelder, LCSW, Maulin Shah, MD, Guilford Parsons, MD, MS, Azadeh Mobasher, PhD

Background and Context:

It has become well recognized that patients experiencing serious illness and their families benefit from discussions regarding their wishes and values concerning their health, and that these conversations are a vital component of healthcare quality and equity.

At Providence, we have implemented system-wide initiatives to improve documentation of these goals-of-care (GOC) conversations. However, reliably tracking GOC discussions remains difficult, as details are often written narratively within daily documentation. If prompted towards the use of discrete data entry, clinicians frequently find the workflows to be restrictive, with less clinical value.

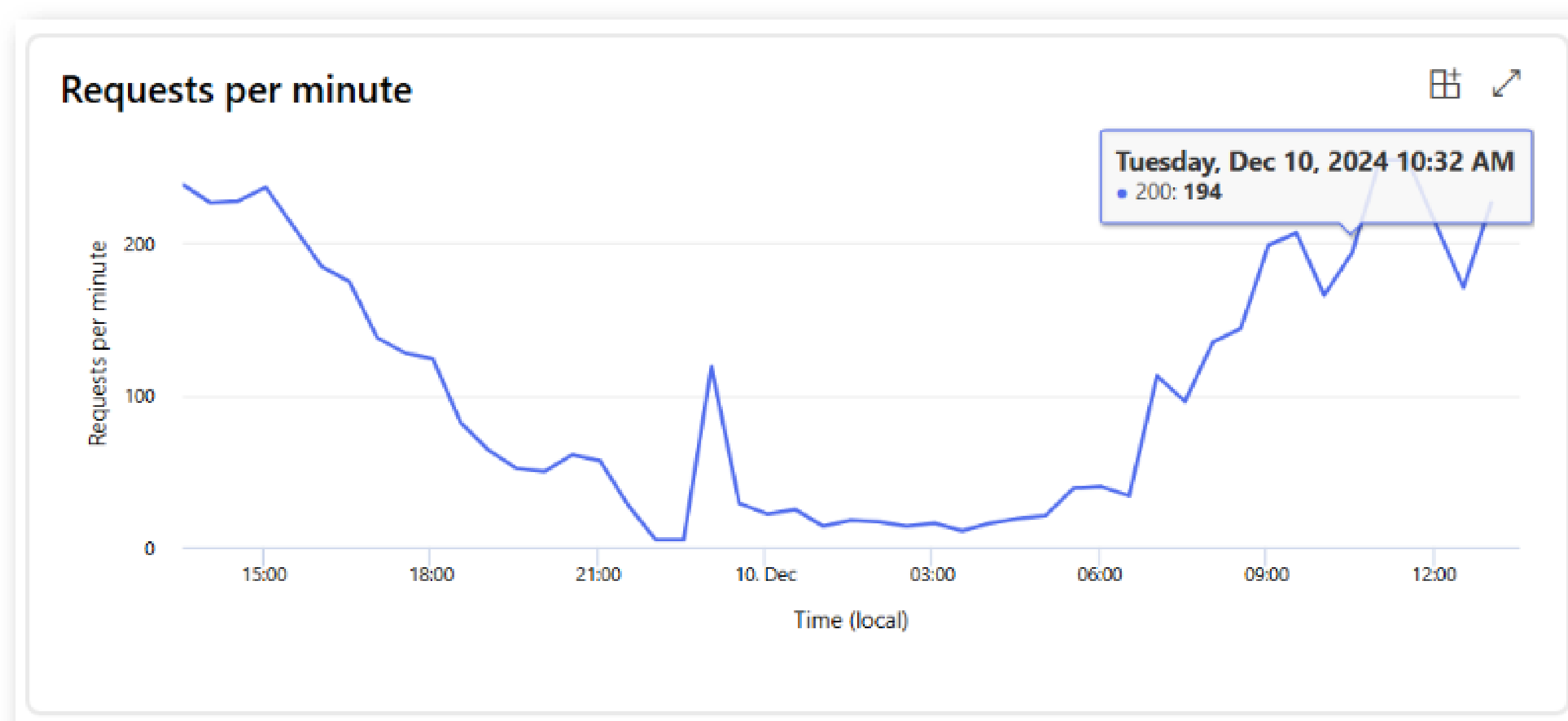
Traditional evaluation methods, such as human assessments and rule-based systems, often face challenges related to cost, accuracy, and scalability.

This led us to develop an AI-powered tool utilizing a general purpose large language model (LLM) to detect GOC conversations written without structured documentation.

Table 1: Comparative performance of language models on GOC identification

| Model | Specificity | Precision | F1 Score | Accuracy |
|--------------------------------|-------------|-------------|-------------|-------------|
| GPT-4o-with-chunking | 0.95 | 0.81 | 0.73 | 0.90 |
| GPT-4o-without-chunking | 0.95 | 0.78 | 0.76 | 0.91 |
| GPT-4o-mini-with-chunking | 0.93 | 0.76 | 0.76 | 0.91 |
| GPT-4o-mini-without-chunking | 0.88 | 0.62 | 0.71 | 0.87 |
| GPT-4 | 0.92 | 0.72 | 0.76 | 0.90 |
| GPT-3.5-Turbo | 0.88 | 0.62 | 0.71 | 0.87 |

Figure 1: Stability of API requests on model launch



Development and Validation:

We built the tool to leverage prompt-engineering powered by OpenAI's ChatGPT to identify the presence of key communication elements of goals of care discussions, leveraging their capabilities in language processing and contextual understanding.

We began with multi-disciplinary consensus validations to create annotation guidelines, with substantial reliability of agreement (Fleiss' kappa, 0.77).

We developed a Python package to test the model through multiple versions of the large language model and selected for the model to maximize specificity and accuracy. We found that GPT 4o(mni) had the highest performance in GOC detection (specificity 0.95, accuracy 0.91, and F1 score of 0.76) along with significant improvements in efficiency and speed.

We also performed subgroup analyses using standard fairness metrics to assess for bias with respect to patients' race and sex (this included parity and equalized odds metrics, and disparate impact scores).

We then developed custom programming and extension code to pass note information to Nebula, Epic's cloud platform, where the model operates, and then returns results through predictive modeling architecture.

Outcomes and Next Steps:

After extensive testing in non-production environments, as well as informatics, clinical, and cybersecurity review, we launched the model in December 2024 as a pilot program at four hospitals. We adjusted the model once for the addition of filters on note type and filing parameters, allowing for optimized efficiency (decreasing the pings to our cloud system per minute).

We have been able to track usage of the model as well as notes identified by the model as containing GOC conversations through the use of SmartData elements.

These notes also display directly to clinicians in the ACP Summary Report page – our current “single source” for information related to advance care planning and goals of care.

Once we perform further validation (using randomized samples for chart review of both positively and negatively tagged results), we have initial governance approval to launch the model across our entire health system.

Figures 2-3: Examples of GOC documentation identified by NotesAI tool

Goals of Care
Palliative care consulted this admission. Patient was determined not to have capacity for complex medical decisions. He assented to the plan of care as discussed with son and daughter. Plan was discussed for discharge to facility with hospice, but as patient's family was working on applying for Medicaid to make this possible; patient medically stabilized. Family has decided to move forward with restorative care with plan to discharge to SNF. Patient is medically stable at this point and he likely will remain in current status long enough to discharge to facility. May still be reasonable to discharge with hospice if goal is to avoid coming back to the hospital.
- CM working on SNF referrals
- If patient were to decompensate, family in agreement they would like to be called and would consider comfort care in the hospital.

Nursing Note
Primary RN asked me to discuss pt's health status with their siblings from Texas as I have taken care of the patient earlier this week and have significant knowledge of overall down trend in health. Hospitalist MD was unavailable to come to bedside to discuss pt status and questions that they had and wanted to discuss at this time.

I started by stating that the patient has slowly began to deteriorate in multiple ways. Arguably most importantly in their unhealing wound. The wound is very complex; per her most recent CT, it had indicated that there was likely spread of infection into the bone. It is also unknown whether the wound is causing the spread of loculated complex fluid collections. Images from the CT were shown to both siblings (with the patient's verbal permission first) to support this information. I also showed them both the recent picture of the wound and several progression photos over the past year. I also discussed the surgical consultation note about the wound and that they deemed that the patient needed surgical intervention to treat the issue of osteomyelitis. However, after the surgical MD consulted neurosurgery as well for input; they both deemed the patient as not a candidate stating "This would be a futile procedure".

Compounding the wound healing is poor nutritional status and frequent emesis. Additionally, the patient is not able to wean off vent, with SBTs discontinued every day this week. The patient's level of lethargy is increasing, sleeping 20-22 hours a day and communicating much less. They are also becoming much more flat and withdrawn. Lastly, the patient remains hemodynamically unstable during dialysis requiring multiple modalities of pressure support with BPs dropping as low as 60s/20s.

With all this information in mind, I discussed the concern for the patient's potential demise considering everything they are fighting against. I revisited her code status as well. The patient's family were thankful for the full update. They both shared their deeply rooted religious background. Their belief is that if the patient were to choose to be DNR that it would be an act of suicide in their God's eyes and that they would be condemned to hell for doing so. They both did understand the moral dilemma that we as health givers are facing when caring for the patient and recognize the situation is very hard for them and for us. Ended the conversation on a good note with thanks from them and other family for all that we are doing for patient.

