Decoding Healthcare's Al Matrix: Connecting Ethics to Patient Reality

SPEAKERS

Caryn Douma, MS, RN, CPPS

Corporate Director of Patient Safety | Penn Medicine

Susan Harkness Regli, PhD

Human Factors Scientist | Penn Medicine



Penn Medicine is...

One of the world's leading academic medical centers, Penn Medicine encompasses the University of Pennsylvania Health System and the Raymond and Ruth Perelman School of Medicine.



50,924 Employees



4,780Perelman School of Medicine staff



2,389
Full-time faculty (including CHOP)

Our Mission

Penn Medicine
is dedicated to
advancing knowledge
and improving
health through...



Patient Care: Exceptional, personalized care to patients with medical conditions of all kinds



Education: Innovative training programs ranging from leadership to global health



Research: Interdisciplinary collaboration improves care and provides the foundation for the next generation of therapies



Community Service: Serving vulnerable populations and conducting outreach initiatives

Our Regional Scope of Clinical Care

Penn Medicine serves patients across the region, in our inpatient and outpatient facilities, and at home.



The Hospital of the University of Pennsylvania



Penn Medicine Princeton Health



Penn Presbyterian Medical Center



Doylestown Health



Pennsylvania Hospital



Penn Medicine at Home



Chester County Hospital



Penn Medicine Medical Group



Lancaster General Health

Patient Safety

STILL NOT HITTING THE MARK



Institute of Medicine Report 1999

To Err is Human

"The single greatest impediment to error prevention in the medical industry is that we punish people for making mistakes." — LUCIAN LEAPE



Crossing the Quality Chasm 2001

Defined quality in terms of six dimensions – safety, effectiveness, patient centeredness, efficiency, timeliness and equity.



NEJM Special Article – The Safety of Inpatient Care 2023

Updated study of adverse events in healthcare



IHI Patient Safety and Artificial Intelligence
Opportunities and Challenges 2025



Achieving Safe Health Care

- Create a fair and just culture staff feel comfortable reporting
- Proactively identify and mitigate preventable harm utilizing technology
- Focus on system improvement through high reliability practices and activities
- Leadership priorities and goals focus on patient safety
- Engage patients and families in care
- Build a learning system for sharing and improvement



Potential Safety Benefits of Al

- Ease of event reporting utilizing prompts for reporters
- Event investigation and analysis summarizing data from reports, EHR and other sources
- Common cause analysis to target interventions and increase impact of improvements
- Prediction of events based on trends and real time data
- Easy, effective querying of policies and related procedures for clinical staff
- Minimize bias (AI can be both a benefit and a concern for bias)
- More effectively engage in collaborative learning across the organization



Problem and Goals

Problem: Difficult to analyze trends and proactively prevent harm to patients

Goal: Integrate patient safety event reporting and automate thematic tagging of incidents to inform ongoing health care operations and initiatives utilizing Natural language processing and AI

Current State

- Safety event reports are self reported after event occurs and quality and quantity is dependent on staff comfort and time
- Investigation and analysis completed retrospectively
- Reports take time to enter in the system and categories are confusing to front line staff
- Only one event type can be chosen
- Lack of governance and oversight over use of data
- Multiple reporting systems within the organization

Future State

- Rapidly answer questions like "What reports are related to workplace violence?"
- Auto suggest multiple event types using NLP/LLMs for analysis and entry
- Remove need for manual manipulation of data
- Utilize AI to capture risks and intervene instead of identifying after harm has occurred
- Organizational process and governance structure for evaluating, granting and restricting access
- One standardized event reporting platform



Patient Safety and Artificial Intelligence

REGULATIONS AND POLICY

Establish clear guidelines for ethical and safe use of Al

- Comprehensive regulations in alignment with state and local laws and reporting requirements
- Data protection and security
- Engage regulators and policymakers in aligning with existing guideline development
- Ensure local engagement and decision making
- Develop standard frameworks for use and protection



Project: NLP and LLMs for Safety Events

Applied Informatics "MadLib"

As a <u>safety executive</u>

Role

When I am <u>deciding where to implement interventions to reduce safety events</u>

Task Context

If I knew all the events that should be tagged with a particular event category

Unknown state

► I would <u>calculate the size of the problem, location(s), and common causes</u>

Intervention

► In order to <u>determine the nature and location of intervention targets</u>

Measurable Outcome





Legal implications working with safety event data

Challenges



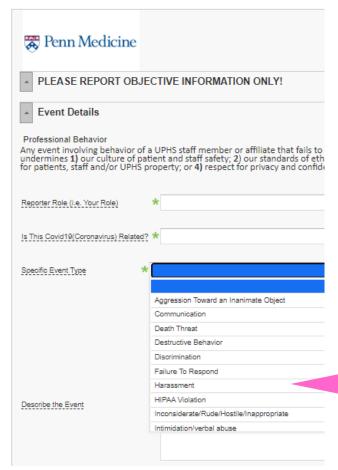
Need for secure storage and governance



Ethical responsibility to patients and staff

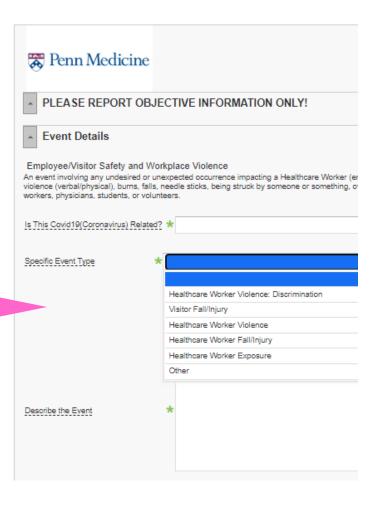


Category Choice not Intuitive



Incident may fit multiple tags







Safety Event Classification using GPT-40



Does this safety event report suggest a physical violence incident occurred?...

Synthetic Example

"Patient became verbally aggressive and started yelling at staff. When asked to calm down, patient threw water pitcher at the wall and attempted to hit the nurse."



...according to the LLM classifier, the safety event report describes a physical violence incident.



On an expert-annotated dataset, the LLM correctly identifies physical violence at 0.84 and verbal abuse/threat of future violence at 0.97



task_prompt = """As a safety leader at a health system, you have been tasked with reviewing narrative reports of safety events, to tag them according to safety event types. For each safety event type, analyze the narrative and determine if the type is explicitly and undeniably present or absent. All data has been de-identified with placeholders like **NAME** **DATE** and **LOCATION**."""

physical_violence_prompt = """Examples of physical violence issues include: Did an act occur at the workplace that can include any of the following: nonverbal, or physical aggression; intimidating, harassing, or humiliating actions; bullying; sabotage; sexual harassment; physical assaults; or other behaviors of concern involving staff, licensed practitioners, patients, or visitors. Can include violence toward an inanimate object (for example, breaking a window)."""



"reasoning": "### Step-by-Step Analysis:

- 1. **Incident Description**: The narrative describes a patient becoming verbally aggressive, yelling at staff, throwing a water pitcher at the wall, and attempting to hit a nurse.
- 2. **Physical Violence**: The patient's actions include throwing a water pitcher and attempting to hit a nurse, which are clear indicators of physical aggression.
- 3. **Aggressor**: The aggressor in this incident is the patient.
- 4. **Victim**: The victim is the nurse, who was the target of the attempted hit.
- 5. **Location**: The specific area or department where the incident occurred is not mentioned in the narrative.
- 6. **Time Context**: The timing details of the incident are not provided in the narrative. ###"



Updates and Future Work

Methodology plans

- Introduce *Automated Prompt Optimization* enable the LLM to refine the initial language given in its prompt to more accurately classify reports without overfitting
- Compare latest models: GPT-4.1, o3, LLaMA 4, Qwen 3
- Expand cohort and dataset to include 2243 new reports (Jul 2024- Dec 2024)

Next validation

- Present results from updated model to two operational coders who do event management
- Ask for correct/incorrect ratings of each tag, as well as tags that may have been missed



By "augmenting human intellect" we mean increasing the capability of man [sic] to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems. Increased capability in this respect is taken to mean mixture of the following: more-rapid comprehension, better comprehension, the possibility of gaining useful degree of comprehension in situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble.

- Douglas Engelbart, Augmenting Human Intellect: A Conceptual Framework, 1962



Conclusion

Paradigm Shift is Needed

- As digital solutions grew to collect more and more discrete data, we realized there are amazing things we can do with that data.
- BUT we are reaching limits of how much we can expect humans (clinicians, operational leaders in risk and safety) to manually enter data
- Al holds a possibility that the collection and summarization of discrete data can be offloaded to a machine, to enable better uses of human capabilities.

What is next?

THE FUTURE OF PATIENT SAFETY

Expand definitions of harm

More robust safety measurement

Patients and families embedded in care

Real time data

Fair and Just Culture

Prioritize safety as a core value

Proactive, predictive -- not reactive

Harness technology responsibly with governance and oversight





Questions?

CONNECTING RISKS THAT MATTER

MATRIX

Thank You!



Connect with us.

Caryn Douma

e: Caryn.Douma@Pennmedicine.upenn.edu

w: pennmedicine.org

in/caryn-douma-0398a021/

Susan Harkness Regli

e: Susan.Regli@pennmedicine.upenn.edu

w: pennmedicine.org

in /in/dr-susan-harkness-regli-7832179/



Backup