A. JAMES CLARK SCHOOL OF ENGINEERING

THE FISCHELL DEPARTMENT of BIOENGINEERING

UNIVERSITY OF MARYLAND

Team 15: The AI Scribe (VitAI)

Ravidu Hevaganinge, Priya Jain, Matthew Sipper, Aditi Thangavelu, Yaqi Zhao

Advisors: Dr. Yang Tao & Anjana Hevaganinge (University of Maryland Fischell Department of Bioengineering), Dr. Kristen Johnson (Children's National Medical Center)

Introduction & Purpose

- **Electronic Health Records** (**EHRs**) are critical for patient safety, care coordination, and regulatory compliance.
- The **History of Present Illness** (**HPI**) details crucial aspects of a patient's visit such as chief complaint, symptoms, and severity.
- EHR documentation is **time-consuming**, taking 3.5 to 6 hours daily and adding nearly 2 extra hours after hours, contributing to **physician burnout**[1].
- Medical scribing faces issues like **inconsistent training**, **lengthy onboarding** (3-9 months), and **significant costs**[2].
- **Automatic Speech Recognition** (**ASR**) converts spoken language to text.
- **Natural Language Processing** (**NLP**) employs algorithms to understand, interpret, and summarize text

Objective

- **Goal**: Develop an **AI-driven EHR tool** by merging **ASR** and **NLP** methods to rapidly produce clinically-precise **HPIs** from patient-clinician encounters.

Conceptual Approach

Table 1. Overview of Techniques Used in HPI Generation.

Method	Description
Data Collection	The dataset we used is called aci-bench which consists of transcripts and the corresponding HPI. Additionally, we simulated physician-patient medical encounters.
ASR: Whisper	The Whisper model was used to generate raw transcripts from physician-patient encounters.
Transformer: BART	Bidirectional Auto-Regressive Transformer model was used to implement the summarization task (raw transcript → HPI).
Annotation: medspaCy	Medspacy is a medical Named Entity Recognition (NER) package used to annotate transcripts and HPI.

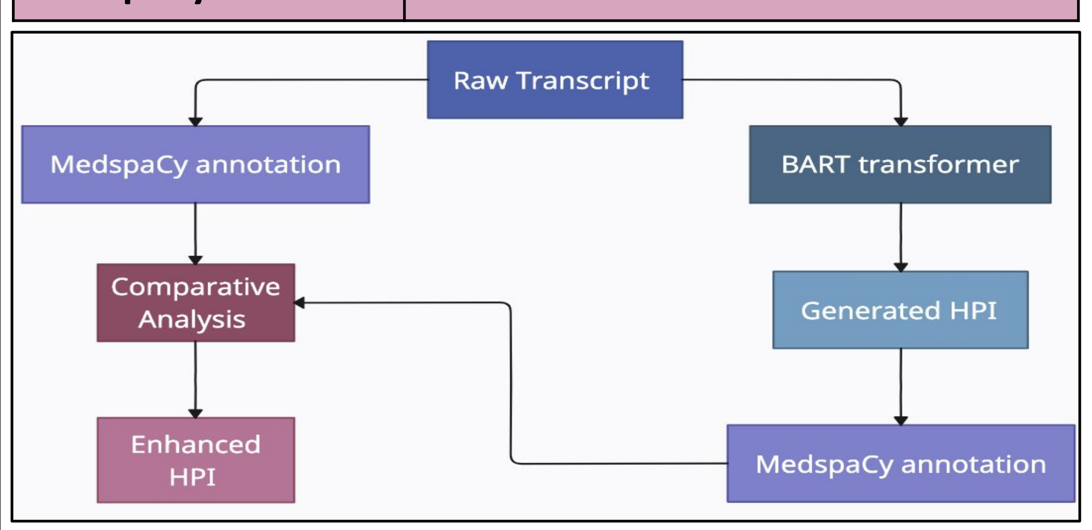


Figure 1. HPI Enhancement Flowchart. MedspaCy annotates raw transcripts and BART-generated HPIs, comparing them to rectify information gaps. The result is an "enhanced HPI" that combines both inputs for greater accuracy, as validated by improved ROUGE-1 scores.

Framework and Model Outputs

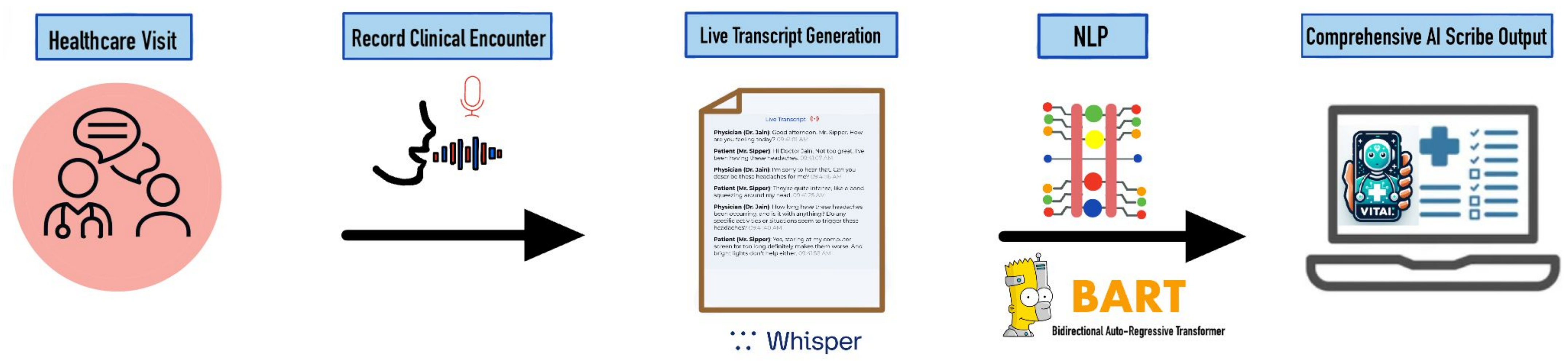


Figure 2 (left). Workflow from Clinical Encounter to AI-Scribed Documentation. The process of capturing and processing a clinical conversation into a comprehensive AI-generated medical document, using Whisper for live transcription and BART for NLP text summarization.

[Doctor] Can you tell me what medications you are currently on?

[Patient] I take 80 mg STRENGTH of lasix DRUG once a day FREQUENCY for my heart LOCATION condition and 20 mg STRENGTH of lisinopril DRUG a day FREQUENCY for my blood pressure.

[Doctor] Alright, we'll keep your prescriptions as they are and continue monitoring. Make sure to track your blood pressure at home and alert me if you notice any increases.

[Patient] Will do.

Figure 3. Raw Transcript. Pertinent targets annotated via medspaCy.

BART Generated HPI:

HISTORY OF PRESENT ILLNESS Brian is a 58-year-old male with a past medical history significant for congestive heart failure and hypertension who presents today for follow-up of his chronic problems. The patient reports that he has been feeling out of sorts lately. He feels tired and fatigued, and he feels kinda lightheaded every once in a while. The symptoms have been going on for about 5 weeks. He denies chest pain, slight cramps, fever, chills, cough, or shortness of breath.

Figure 4. Raw HPI Output. The BART-generated summary is missing pertinent targets and has not been refined.

Enhanced HPI:

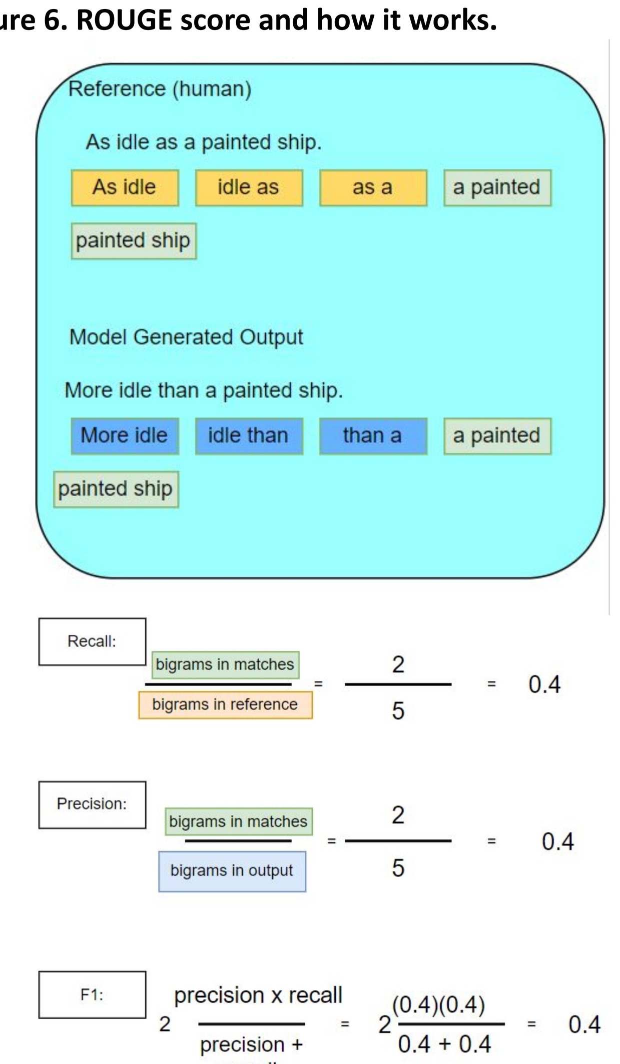
HISTORY OF PRESENT ILLNESS Brian is a 58-year-old male with a past medical history significant for congestive heart failure and hypertension who presents today for follow-up of his chronic problems. The patient reports that he has been feeling out of sorts lately. He feels tired and fatigued, and he feels kinda lightheaded every once in a while. The symptoms have been going on for about 5 weeks. He denies chest pain, slight cramps, fever, chills, cough, or shortness of breath.

(Added Sentences)
The patient reports taking 80 mg of lasix once a day. The patient reports taking 20 mg of lisinopril a day.

Figure 5. Finalized HPI post-processing. Pertinent information has been added via enhancement approach in Figure 5.

Performance Metrics

Figure 6. ROUGE score and how it works.



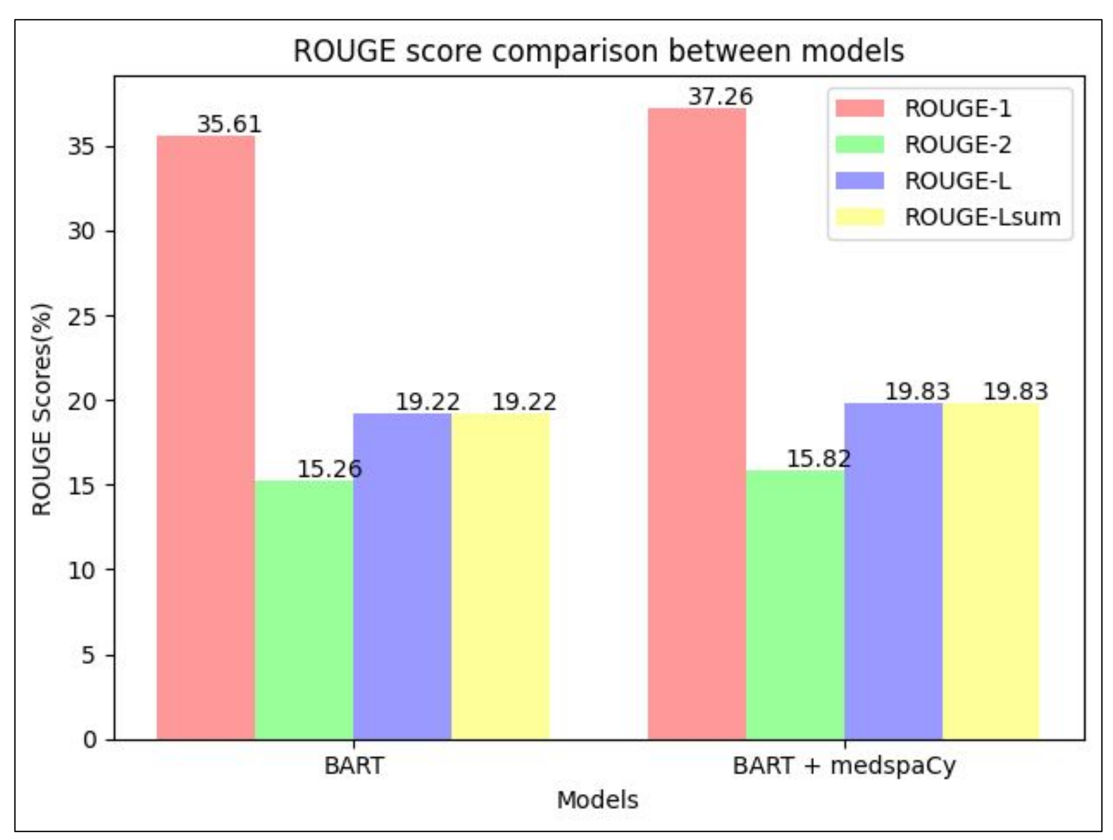


Figure 7. Comparison of ROUGE score between models. BART Model Output Compared to BART Combined with medspaCy. Significant Increase in ROUGE-1 Score Indicates Higher Unigram Overlap Between Combined Summary and Raw Transcript

Wireframe

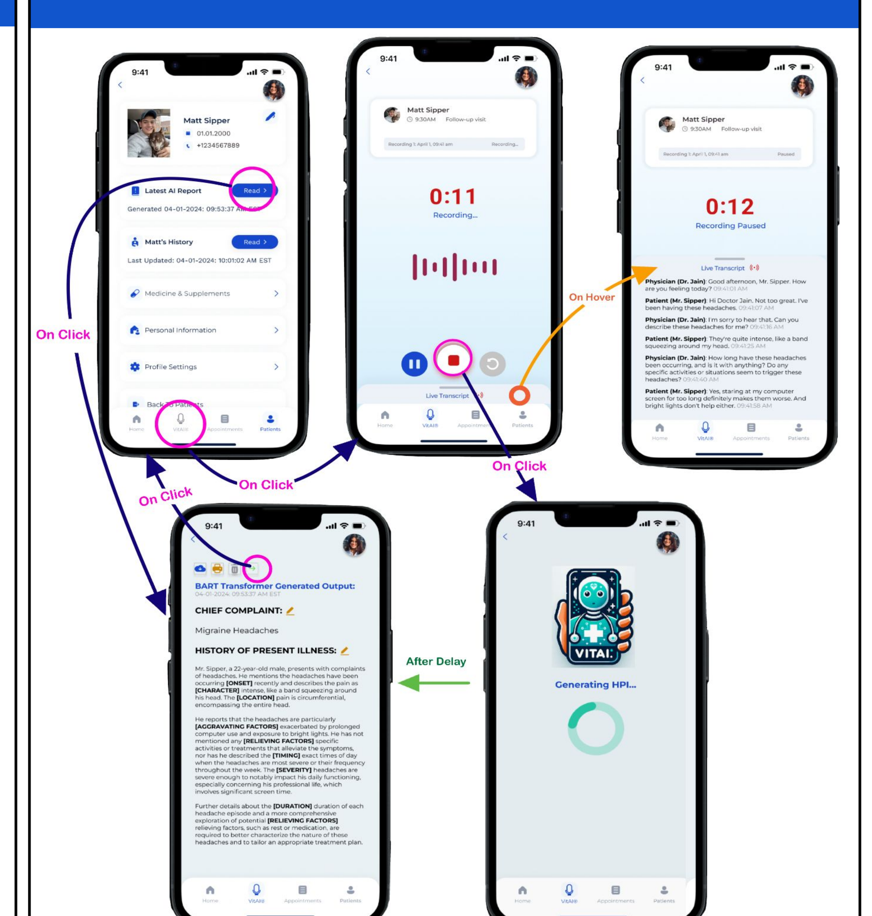


Figure 8. VitAI Interactive Figma Wireframe. VitAI integrated into a clinical data management smartphone interface, highlighting five main screens and navigation pathways. Essential for alignment throughout product development process, it demonstrates functionality and UX design. See demo video.

Conclusions

- Integrated ASR, NLP, and MedSpacy to **efficiently generate comprehensive HPIs**.
- Demonstrated enhanced documentation accuracy and efficiency, **bridging AI with healthcare** needs.
- Validated performance through **improved Rouge scores**, confirming effectiveness.
- Incorporated a **post-processing refinement system** to fine-tune output.
- Developed a **wireframe prototype UI** to anticipate and shape future requirements.

Bioethical Implications:

- Maintain **patient privacy, autonomy, HIPAA** compliance; prevent unauthorized access
- Eliminate **medical biases**, promote **fairness**, and adapt models for multilingual use.
- Survey **public opinion** on AI use in healthcare.

Future Work

- Expand the testing dataset to include **anonymized data** from Children's National Hospital to enhance model accuracy.
- Conduct **comparative analysis** of transformer models to optimize output.
- Develop an **Application Programming Interface** (API) for seamless integration.
- Progress **product development** for a smartphone.
- Utilize a data fabricator to generate **compatible datasets**, crucial for expanding data resources and improve model training.

Acknowledgments

- We would like to thank Stephen Wiafe-Amoako, Aadi Palnitkar, Dr. Bahaa Ghammraoui, and James Ajokubi for their contributions and support.

References

1. How much time is being spent on EHRs? Contemporary Pediatrics. Published April 21, 2022. Accessed April 19, 2024. https://www.contemporarypediatrics.com/view/how-much-time-is-being-spent-on-ehrs

2. Walker KJ, Dunlop W, Liew D, et al. An economic evaluation of the costs of training a medical scribe to work in Emergency Medicine. Emerg Med J. 2016;33(12):865-869. doi:10.1136/emermed-2016-205934

3. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28(9):1773-1784. doi:10.1038/s41591-022-01981-2

4. Lin CY. ROUGE: A Package for Automatic Evaluation of Summaries. In: Text Summarization Branches Out. Association for Computational Linguistics; 2004:74-81. Accessed October 25, 2023. https://aclanthology.org/W04-1013

5. Yim, Ww., Fu, Y., Ben Abacha, A. *et al.* Aci-bench: a Novel Ambient Clinical Intelligence Dataset for Benchmarking Automatic Visit Note Generation. *Sci Data* 10, 586 (2023). https://doi.org/10.1038/s41597-023-02487-3