FHIR-GPT Enhances Health Interoperability with Large Language Models

Yikuan Li, MS[1,2], Hanyin Wang, BMed[1], Halid Z. Yerebakan, PhD[2], Yoshihisa Shinagawa, PhD[2], Yuan Luo, PhD[1]

[1]Northwestern University Feinberg School of Medicine, Chicago, IL, U.S.A

[2]Siemens Medical Solutions, Malvern, PA, U.S.A

Abstract

Advancing health interoperability can significantly benefit health research, including phenotyping, clinical trial support, and public health surveillance. Federal agencies, including ONC, CDC, and CMS, have been collectively collaborating to promote interoperability by adopting Fast Healthcare Interoperability Resources (FHIR). However, the heterogeneous structures and formats of health data present challenges when transforming Electronic Health Record (EHR) data into FHIR resources. This challenge becomes more significant when critical health information is embedded in unstructured data rather than well-organized structured formats. Previous studies relied on multiple separate rule-based or deep learning-based NLP tools to complete the FHIR resource transformation, which demands substantial development costs, extensive training data, and meticulous integration of multiple individual NLP tools. In this study, we assessed the ability of large language models (LLMs) to transform clinical narratives into HL7 FHIR resources. Our experiments on 3,671 snippets of clinical texts demonstrate that the best-performing LLM, namely FHIR-GPT, achieves an exceptional exact match rate of over 90% in the transformation to FHIR medicationstatement resources, when compared to human annotations. FHIR-GPT improved the exact match rates of existing NLP pipelines by 3% for routes, 12% for dose quantities, 35% for reasons, 42% for forms, and over 50% for timing schedules. Our findings provide the foundations for leveraging LLMs to enhance health data interoperability. Future studies will aim to build upon these successes by extending the generation to additional FHIR resources.

Introduction

Interoperability enhances the ability of healthcare providers to deliver safe, effective, and patient-focused care. It also offers novel avenues for individuals and caregivers to access electronic health data for care coordination and management[1]. The promotion of interoperability has become an integral aspect of various healthcare initiatives, spanning from ensuring health equity to responding to public health emergencies [2]. Federal agencies, including the Office of the National Coordinator of Health IT (ONC)[1], the Centers for Disease Control and Prevention (CDC)[3], and the Centers for Medicare & Medicaid Services (CMS)[4], collectively collaborate to promote interoperability through the adoption of Fast Healthcare Interoperability Resources (FHIR), which is a next-generation interoperability standard developed by the Health Level 7 (HL7®) standards development organization[5]. FHIR is specifically designed to facilitate the swift and efficient exchange of health data. FHIR has seen growing adoption in the modeling and integration of both structured and unstructured data for various health research purposes. Its applications range from developing computational phenotyping[6-8] to supporting clinical trials[9-12], building surveillance systems[13,14], and much more. We refer to these two review papers[15,16] for further insights into FHIR applications.

Transforming health data into the FHIR format presents a challenge, as various healthcare organizations have their unique infrastructure, standards, and formats for generating, storing, and organizing health data[17]. This challenge becomes more significant when critical health information is embedded in unstructured data other than well-organized structured formats. There are existing efforts for promoting the transformation of unstructured data into FHIR resources, offered by both academic and commercial sectors. In academic research, Hong et al. [18] integrated clinical NLP tools, including cTAKES[19], MedXN[20], and MedTime[20], to extract clinical entities from corresponding document sections and standardize them into FHIR resources. Wang et al. developed Opioid2FHIR[21], a system that employs multiple deep learning-based natural language

processing (NLP) techniques for opioid information extraction and normalization. In the commercial domain, Google Cloud has released the Healthcare Natural Language API[22], capable of converting medical text input into FHIR resources. Amazon Medical Comprehend[23] can extract and normalize medical concepts into clinical vocabulary, although it lacks the ability to map all extracted information to FHIR resources. Azure Health Data[24] is proficient at converting semi-structured data into FHIR resources but does not handle free-text unstructured input. All the above FHIR transformation tools necessitate sequential collaboration with multiple NLP tools. These include a Named Entity Recognition (NER) tool for extracting medical concepts, a relation extraction tool for identifying relations related to a target concept, a normalization tool for standardizing the extracted concepts into vocabularies, and a reconciliation tool for integrating the normalized concepts into a valid FHIR format. The development and training of each NLP tool is resource-intensive and demands a significant amount of time and data. Creating a pipeline that integrates multiple NLP tools requires substantial computational resources, annotated data, and human effort. Furthermore, as the transformation progresses along the pipeline, the accuracy of the conversion also decreases.

Therefore, we propose harnessing pre-trained large language models (LLMs) to streamline the existing approach which relies on a pipeline of multiple NLP tools, to facilitate the transformation of free-text input into FHIR resources. Our contributions can be summarized as follows:

- We manually annotated a dataset containing 3,671 snippets extracted from discharge summaries, along with their corresponding transformed MedicationStatement resources. To the best of our knowledge, this represents the largest and neatest human-annotated dataset of free-text to FHIR resource transformation pairs.
- We demonstrated that LLMs, especially FHIR-GPT, are able to outperform the existing NLP methods in transforming FHIR resources when evaluated by the exact match rate.

a. Discharge Summary

Discharge Summary:
Discharge Medications:
…
[7. senna 8.6 mg Tablet Sig: One (1) Tablet PO BID P.R.N Constipation]
…
Patient was discharged to long-term care facility.

b. Entity Annotations

Medication	**senna**
Reason	**Constipation**
Route	**PO**
Timing	**BID**
Dose	**One (1)**
Form	**Tablet**
Strength	**8.6 mg**
asNeeded	**P.R.N**

c. FHIR MedicationStatement

```
{'resourceType': 'MedicationStatement',
 'id': '100035T133',
 'subject': {'reference': 'hadm_id100035'},
 'medication': {'reference': {'reference': '#med100035T133'}},
 'reason': [
    {'concept': {'text': 'Constipation',
        'coding': [{'system': 'http://snomed.info/sct',
        'code': '14760008',
        'display': 'Constipation'}]}}],
 'dosage': [
    {'route': {'text': 'PO',
              'coding': [{'system': 'http://snomed.info/sct',
              'code': '26643006',
              'display': 'Oral route'}]},
    'timing': {'repeat': {'frequency': 2, 'period': 1.0, 'periodUnit': 'd'},
               'code': {'coding': [{'system': 'http://terminology.hl7.org/',
                        'code': 'BID',
               'display': 'BID'}]}},
    'asNeeded': True,
    'doseAndRate': [{'doseQuantity': {'value': 1.0}}]}],
 'contained': [
    {'resourceType': 'Medication',
    'id': 'med100035T133',
    'code': {'coding': [
        {'system': 'National Drug Code',
        'code': '00904516561',
        'display': 'sennosides, USP 8.6 MG Oral Tablet'},
        {'system': 'RxNorm',
        'code': '312935',
        'display': 'sennosides, USP 8.6 MG Oral Tablet'}],
            'text': 'senna 8.6 mg Tablet'},
    'doseForm': {'text': 'Tablet',
        'coding': [{'system': 'http://snomed.info/sct',
        'code': '385055001',
        'display': 'Tablet'}]},
    'ingredient': [{'item': {'concept': {'text': 'senna'}},
        'strengthQuantity':
            {'value': 8.6, 'unit': 'milligram',
             'system': 'http://unitsofmeasure.org',
             'code': 'mg'}}]}]}
```

d. Prompts for LLMs

[INSTRUCTIONS]
You are a helpful assistant that can help with medication data extraction.
User will paste a short narrative that describes the administration of a drug.
Please extract the drug route (How drug should enter body), e.g. PO, IV.
< Collapsed for more instructions >

[TEMPLATE]
{"text": "<string>", // the originial text mention of drug route
"coding": [//optional, but MUST lookup from the table below
{"system": "http://snomed.info/sct",
"code": "<code>", # SNOMED code
"display": "<display>" # the display of the code}]}

[EXAMPLES]
For example, the narrative
"Oxycodone-Acetaminophen 5-325 mg Tablet
Sig: 1-2 Tablets PO\nQ4-6H (every 4 to 6 hours) as needed"
You should return a json format:
{'text': 'PO', 'coding': [{'system': 'http://snomed.info/sct', 'code': '26643006','display': 'Oral route'}]}
< Collapsed for 4 more examples >

[TERMINOLOGIES]
Code Display
6064005 Topical route
10547007 Otic route
<Collapsed for 143 more SNOMED CT Codes>

e. Workflow

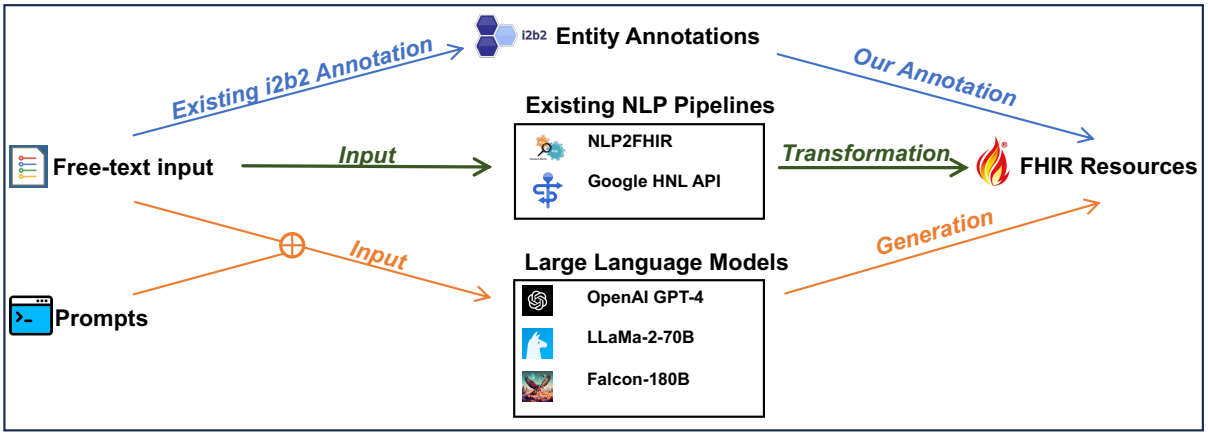

Figure 1. a. A snippet of the discharge summary will be used to generate the FHIR resource. **b**. The i2b2 expert annotates word spans related to medications in the discharge summary. **c**. An example of the transformed FHIR MedicationStatement resource based on our annotations. The same color shading from panel b is used. These results represent the ground truth transformation.**d**. An example of the prompt used to instruct large language models in generating FHIR resources. **e**. The workflow details how we annotate the dataset and compare the performance of Large Language Models and existing NLP pipelines in transforming free-text inputs into associated FHIR resources.

Results

The annotation results are presented in Table 1. In summary, we annotated a total of 3,671 pairs of free-text to FHIR MedicationStatement resources transformations. The free-text input was derived from discharge summaries for 280 admissions. The character lengths of the input data exhibit an average of approximately 66 characters, with a relatively high standard deviation of 65. The annotated resources encompass 625 distinct medications in 26 different forms and are associated with 354 different reasons, as well as 16 administration routes. These elements display varying levels of availability, ranging from approximately 30% for reasons to 65% for timing schedules. SNOMED CT is the most commonly used terminology system, which was applied to medication, form, route, and reason, while HL7's own code set was used for timing schedules. The annotated resources in the .JSON structure have an average number of objects of 58.2 (standard deviation = 16.2) and an average depth of 6.7 (standard deviation = 0.5).

The transformation results are presented in Table 2. In summary, transformation with GPT-4, namely FHIR-GPT, achieved an exceptional exact match rate of over 0.90 for all elements, outperforming both baseline models and all other LLMs. Specifically, when compared to existing NLP pipelines, FHIR-GPT improved the exact match rate by 3% for routes, 12% for dose quantities, 35% for reasons, 42% for forms, and over 50% for timing schedules. Among all LLMs, we observed a trend of increasing accuracy as the parameter size increased. GPT-4, with approximately 1.7 trillion parameters, surpassed the 180 billion parameter Falcon models and further improved upon the 70 billion parameter Llama-2 models. Within all elements, the most challenging ones for both LLMs and existing methods are timing schedules and reasons. Timing schedules, consisting of 10 objects, require calculations and inferences (e.g., inferring the duration based on frequency and distribution), while reasons involve relationship extraction and handling cardinality, as a medication can be taken for more than one reason.

Table 1. Descriptions, examples, and statistics of human annotation for the FHIR *medicationstatement* resource.

***Medication Statement* Elements**	**Type**	**Card.**	**Example**	**Description**	**CodeSystem**	**N (%)**	**N, Uniq. Entries**	**N, Uniq. Codes**
identifier	String	1..1	*100035T133*	External identifier	MIMIC+i2b2	3671 (100%)	3,671	3,671
subject	Codeable Reference	1..1	*{'reference': 'hadm_id164366'}*	Who is/was taking the medication	MIMIC	3671 (100%)	280	280
medication		1..1		What medication				
medication-Code	Codeable Concept	0..1	*{'coding': [{'system': 'NDC', 'code': '51079088120', 'display': 'clonazepam 0.5 MG Oral Tablet'}, {'system': 'RxNORM', 'code': '197527','display': 'Clonazepam 500 microgram oral tablet'}, {'system': 'SNOMED', 'code': '322897008', 'display': 'Clonazepam 500 microgram oral tablet'}], 'text': 'clonazepam 0.5 mg Tablet'}*	Codes that identify this medication	NDC / RxNorm / SNOMED CT Medication	3671 (100%)	1762	NDC: 625, RxNorm: 520, SNOMED : 210
doseForm	Codeable Concept	0..1	*{'text': 'Tablet','coding': [{'system': 'SNOMED', 'code': '385055001','display': 'Tablet'}]}*	powder | tablets | capsule +	SNOMED CT Dose Form	1478 (40.3%)	176	26
ingredient. Strength	Quantity	0..1	*{'value': 0.5, 'unit': 'milligram', 'system': 'http://unitsofmeasure.org', 'code': 'mg'}*	Quantity of ingredient presents	unitsofmeasure. org	2383 (64.9%)	188	16
reason	Codeable Concept	0..*	*[{'concept': {'text': 'headache', 'coding': [{'system': 'SNOMED', 'code': '25064002','display': 'Headache'}]}}]*	Reason for why the medication is being/was taken	SNOMED CT Finding	1106 (30.1%)	619	354
dosage		0..*						
asNeeded	Boolean	0..1	*True*	Take "as needed"		3671 (100%)	2	
route	Codeable Concept		*{'text': 'PO', 'coding': [{'system': 'SNOMED', 'code': '26643006', 'display': 'Oral route'}]}*	How medication enters the body	SNOMED CT Route of Admin.	2011 (54.8%)	64	15
timing. repeat	Element	0..1	*{'frequency': 1, 'period': 4.0, 'periodMax': 6.0, 'periodUnit': 'h', 'duration': 3.0, 'durationUnit':'d'}*	Timing schedule	hl7.org/fhir/	2393 (65.2%)	177	6
timing. code	Codeable Concept	0..1	*{'coding': [{'system': 'HL7','code': 'Q4H', 'display': 'Q4H'}]}}*	Code for timing schedule, e.g. 'BID'	hl7.org/fhir/	2287 (62.3%)	17	17
dose-Range	Quantity	0..1	*{"doseQuantity": {"value": 5.0, "unit": "ML"}}*	Amount or range of medication per dose		1378 (37.5%)	53	
dose-Quantity	Range	0..1	*{"doseRange": { "low": { "value": 1.0},"high": { "value": 3.0}}}*			11 (0.30%)	7	

Table 2. Comparison of LLMs and existing NLP pipelines for transforming free-text input into FHIR *MedicationStatement* resources. Performance is evaluated using the exact match rate, which requires that the resources generated by the models precisely match human annotations in all aspects, including structure, codes, and cardinality. Due to version and implementation differences, the existing NLP pipelines cannot generate all the elements included in our annotations. The best-performing model for each element is indicated in bold, while the second-place model is underlined.

Elements of *medicationstatement*	**Large Language Models**			**Existing NLP Pipelines**	
	GPT-4[32]	Falcon-180B[33]	Llama-2-70B[34]	NLP2FHIR[18]	Google Healthcare NL API[22]
medication					
medicationCode	**0.968**	0.899	0.859	0.862	<u>0.963</u>
doseForm	**0.976**	<u>0.890</u>	0.633	0.556	-
ingredient.Strength	**0.980**	<u>0.921</u>	0.792	-	-
reason	**0.902**	0.593	0.169	<u>0.645</u>	-
dosage					
route	**0.902**	0.457	0.516	-	<u>0.871</u>
timing.repeat	**0.947**	0.268	0.221	0.403	-
timing.code	**0.952**	<u>0.818</u>	0.600	0.424	-
doseQuantity/Range	**0.973**	<u>0.864</u>	0.823	0.724	0.854

Methods

In this section, we delve into the technical details employed in data annotation, LLMs usage, and the evaluation process. For an illustrative visual representation of the workflow, please refer to Figure 1.

Data Annotation

The HAPI FHIR public test server[25] hosts millions of examples of converted FHIR resources. However, we are unable to retrieve their source data before the conversion. To the best of our knowledge, there is no largely publicly available dataset in the FHIR standard that has been generated from the clinical notes. Therefore, we have decided to annotate a dataset that contains both free-text input and structured output in FHIR resources. The latter will serve as the ground truth against which we can evaluate the performance of our LLMs in FHIR transformation.

We manually annotated the medication-related clinical narratives to adhere to the *MedicationStatement* resource as per FHIR v6.0.0: R6 implementation guide[26]. According to the official FHIR definition, a *MedicationStatement* indicates that a patient may currently be taking a medication, has taken it in the past, or will take it in the future. This transformation holds particular significance because many medication-related details, such as the reasons for administration and dosage instructions, often remain absent in structured data. Clinical notes within the Electronic Health Record (EHR) system frequently represent the sole available source for retrieval and conversion into a standardized format. Clinical notes within the EHR system might be the sole source available for the retrieval and transformation of this information into a standardized format. The MedicationStatement encompasses various contents of medication, including dosage, schedule, reason, form, route, strength, and more. For detailed examples of the elements in the *MedicationStatement* resource, please refer to Table 1.

The clinical text input is obtained from the discharge summaries in the MIMIC-III dataset[27]. The 2018 n2c2 medication extraction challenge[28], essentially a named entity recognition task, provided mentions of medications and the word spans of the medications' associated entities (including drug routes, frequencies, durations, adverse effects, forms, strengths, dosages, and reasons) within the discharge summaries in the MIMIC-III dataset. All entities were manually annotated by clinical experts. We extracted text snippets, each containing mentions of one medication and all its associated entities, from the discharge summaries. We also added some buffer words to ensure that these snippets form complete sentences. These extracted snippets, each related to a specific medication, serve as input for both annotations and transformations.

The human annotation for transformation to the FHIR standard consists of three key steps. The first step involved identifying the elements associated with each medication, and this task was effectively addressed by re-using expert annotations from the n2c2 dataset, which accurately pinpointed the word spans of each element. The second step required standardizing the elements

from free-text into clinical terminology coding systems. The elements were linked to different coding systems, and we have provided a detailed description of which code systems were used in Table 1. Notably, the medication name was encoded in three distinct coding systems. Initially, the medication name was mapped to the patient's prescription table in MIMIC-III, where NDC codes were provided. Input data, for which the medication name couldn't be mapped to the patient's prescription table, were excluded from the dataset. Subsequently, NDC codes were mapped to RxNorm codes and SNOMED CT Medication Codes using the APIs provided by the RxNav toolkit[29]. For all other elements, such as reasons, routes, and forms, the SNOMED CT coding system was primarily used, unless HL7.org provided its own code set. The transformation of these codes relied primarily on manual lookup. We referred to the SNOMED CT Browser, International Edition [30] for display names, codes, and other SNOMED CT terminology details. It's worth noting that we encoded all elements separately (except for reasons) by mapping non-duplicated text inputs to their associated codes. While this coding strategy saved time, it may have resulted in some loss of accuracy, as contextual information wasn't considered. The third step involved assembling the identifiers, codes, texts, extensions, and structures into a complete *MedicationStatement* resource. Throughout the study, we utilized the .json structure format. The converted FHIR medication statements undergo validation by the official FHIR validator[31] to ensure compliance with FHIR standards, including structure, datatypes, cardinalities code sets, display names, etc.

The annotation tasks were primarily conducted by Y. Li, with assistance from H.W., who has clinical expertise and assisted in resolving any ambiguities or uncertainties. We will make the annotated dataset available to the public on PhyisoNet.org for authorized use upon paper acceptance.

Large Language Models

The LLMs we experimented with include OpenAI GPT-4[32], Llama-2-70B[33], and Falcon-180B[34]. OpenAI GPT-4 is the most widely recognized commercial LLM. Llama-2-70B is the largest open-source LLM available for commercial use as of the writing of this manuscript. Falcon-180B is the largest open-source LLM, primarily intended for research purposes. We accessed the GPT-4 APIs through the Azure OpenAI service, as recommended by the responsible use guideline of MIMIC data. The specific model we used is *gpt-4-32k* in its 2023-05-15 version. To enhance efficiency, we made multiple asynchronous API calls. For Llama-2-70B and Falcon-180B, we deployed them on our HIPAA-compliant firewalled local servers with multiple GPU backends. GPTQ [35] was used to accelerate the inference time for Llama-2-70B and Falcon-180B.

We required these Language Models (LLMs) to transform the free-text entries into MedicationStatements conforming to the FHIR standard, employing the few-shot prompt settings. Each clinical snippet was individually input into the LLMs to generate the MedicationStatement resource. We used five separate prompts to instruct the LLM to transform the free-text input into the elements of a MedicationStatement resource, including medication details (such as drug name, strength, and form), route, timing, dosage, and reason, respectively. All few-shot prompts adhered to a template with the following order: task instructions, expected output FHIR templates in .JSON format, 4-5 examples of transformations, a comprehensive list of codes from which the model could make selections, and the input text to be transformed. As there was no fine-tuning or domain-specific adaptation in our experiments, we initially had the LLM generate the FHIR format for a small subset of the dataset (N=~100). Then, we manually reviewed the discrepancies between the LLM-generated FHIR output and our human annotations. Common mistakes were identified and used to refine the prompts. There were slight differences in the prompts for each LLM, as different LLMs may be sensitive to different prompts. It's important to note that we did not have access to comprehensive lists of NDC, RxNorm, and SNOMED Medication codes for all medication names, as well as SNOMED Finding codes for reasons. Additionally, even if we had

such comprehensive lists, they would have exceeded the token limits for LLMs. Thus, we did not task LLMs with coding these entities; instead, we instructed them to identify the contexts mentioned in the input text. For other code sets, such as SNOMED CT Form codes, numbering in the hundreds, we allowed LLMs to directly code them. Please see the appendix for prompts.

Evaluation

We compared the transformed resources with the outputs from two existing approaches: NLP2FHIR [18] and Google Healthcare Natural Language (NL) API [22]. The transformation results from both approaches lacked some elements covered by our human annotation and LLMs generation. NLP2FHIR was built based on a previous version of the FHIR implement guide, and the Google Healthcare NL API primarily standardized concepts to UMLS CUIs, rather than SNOMED CT codes, which are more frequently used in our annotations and LLMs' transformations. We made adaptations and adjustments to ensure a fair comparison. We deployed the NLP2FHIR pipeline on our HIPAA-compliant firewalled local servers. We accessed the Google Healthcare NL API through the Google Cloud Healthcare API, which is also compliant with HIPAA regulations.

When evaluating the FHIR resources generated by the LLMs, our initial step was to verify that the output was in valid JSON format. Once the JSON format check was successfully passed, our primary criterion for evaluation was the exact match rate. This criterion required that the resources generated by the LLMs exactly matched the human annotations in all aspects, including structures, codes, and cardinality. Unlike previous studies that reported word scan F1, precision, and recall scores, which considered the transformation as a NER (Named Entity Recognition) task, we did not use these metrics. This decision was made because those metrics may overlook the essential aspects of inferring and standardizing the content based on contexts. Exact identification of the word span does not guarantee the correct corresponding codes can be identified and that the accurate FHIR schema can be derived.

Conclusion

In this study, we provided the foundations of leveraging LLMs to enhance health data interoperability by transforming free-text input into the FHIR resources. The FHIR-GPT model is not only training-free but also improves transformation accuracy. Future studies will aim to build upon these successes by extending the generation to additional FHIR resources and comparing the performance of more LLM models.

1. Office of the National Coordinator for Health Information Technology (ONC). Interoperability. Accessed Oct 31, 2023, https://www.healthit.gov/topic/interoperability
2. Office of the National Coordinator for Health Information Technology. United States Core Data for Interoperability (USCDI). 2023;
3. Centers for Disease Control and Prevention. Advancing Interoperability for Public Health. Accessed Oct 31, 2023. https://www.cdc.gov/surveillance/policy-standards/interoperability.html
4. CMS Health Informatics and Interoperability Group (HIIG). Federal Interoperability. Accessed Oct 31, 2023. https://www.cms.gov/priorities/key-initiatives/burden-reduction/interoperability/federal-interoperability
5. HL7.org. FHIR Overview. Accessed Oct 31, 2023. https://hl7.org/fhir/overview.html
6. Bauer DC, Metke‐Jimenez A, Maurer‐Stroh S, et al. Interoperable medical data: the missing link for understanding COVID‐19. *Transboundary and emerging diseases*. 2021;68(4):1753-1760.
7. Brandt PS, Pacheco JA, Rasmussen LV. Development of a repository of computable phenotype definitions using the clinical quality language. *JAMIA open*. 2021;4(4):ooab094.
8. Zong N, Sharma DK, Yu Y, et al. Developing a FHIR-based framework for phenome wide association studies: a case study with a pan-cancer cohort. *AMIA Summits on Translational Science Proceedings*. 2020;2020:750.
9. Metke-Jimenez A, Hansen D. FHIRCap: Transforming REDCap forms into FHIR resources. *AMIA Summits on Translational Science Proceedings*. 2019;2019:54.
10. Pfiffner PB, Pinyol I, Natter MD, Mandl KD. C3-PRO: connecting ResearchKit to the health system using i2b2 and FHIR. *PLoS One*. 2016;11(3):e0152722.
11. Reinecke I, Gulden C, Kümmel M, Nassirian A, Blasini R, Sedlmayr M. Design for a modular clinical trial recruitment support system based on FHIR and OMOP. *Digital Personalized Health and Medicine*. IOS Press; 2020:158-162.
12. Zong N, Stone DJ, Sharma DK, et al. Modeling cancer clinical trials using HL7 FHIR to support downstream applications: A case study with colorectal cancer data. *International journal of medical informatics*. 2021;145:104308.
13. Lee H-A, Kung H-H, Lee Y-J, et al. Global infectious disease surveillance and case tracking system for COVID-19: development study. *JMIR Medical Informatics*. 2020;8(12):e20567.
14. Wang X, Lehmann H, Botsis T. Can FHIR support standardization in post-market safety surveillance? *Public Health and Informatics*. IOS Press; 2021:33-37.
15. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. *JMIR medical informatics*. 2021;9(7):e21929.
16. Vorisek CN, Lehne M, Klopfenstein SAI, et al. Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review. *JMIR medical informatics*. 2022;10(7):e35724.
17. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *Journal of big data*. 2019;6(1):1-25.
18. Hong N, Wen A, Shen F, et al. Developing a scalable FHIR-based clinical data normalization pipeline for standardizing and integrating unstructured and structured electronic health record data. *JAMIA open*. 2019;2(4):570-579.
19. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*. 2010;17(5):507-513.
20. Sohn S, Clark C, Halgrim SR, Murphy SP, Chute CG, Liu H. MedXN: an open source medication extraction and normalization tool for clinical text. *Journal of the American Medical Informatics Association*. 2014;21(5):858-865.
21. Wang J, Mathews WC, Pham HA, Xu H, Zhang Y. Opioid2FHIR: A system for extracting FHIR-compatible opioid prescriptions from clinical text. IEEE; 2020:1748-1751.
22. Google Cloud. Use the Healthcare Natural Language API. Accessed Oct 31, 2023. https://cloud.google.com/healthcare-api/docs/how-tos/nlp
23. Bhatia P, Celikkaya B, Khalilia M, Senthivel S. Comprehend medical: a named entity recognition and relationship extraction web service. IEEE; 2019:1844-1851.
24. Microsoft Azure. Azure Health Data Services. Accessed Oct 31, 2023. https://azure.microsoft.com/en-us/products/health-data-services
25. HL7.org. HAPI FHIR. Accessed Oct 31, 2023. https://hapi.fhir.org/

26. HL7.org. Resource MedicationStatement - Content. Accessed Oct 31, 2023. https://build.fhir.org/medicationstatement.html
27. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Scientific data*. 2016;3(1):1-9.
28. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. *Journal of the American Medical Informatics Association*. 2020;27(1):3-12.
29. Zeng K, Bodenreider O, Kilbourne J, Nelson SJ. RxNav: a web service for standard drug information. American Medical Informatics Association; 2006:1156.
30. SNOMED International. SNOMED International SNOMED CT Browser. Accessed Oct 31, 2023. https://browser.ihtsdotools.org/
31. HL7.org. Validate Resources. Accessed Oct 31, 2023. https://validator.fhir.org/
32. OpenAI R. GPT-4 technical report. *arXiv*. 2023:2303.08774.
33. Touvron H, Martin L, Stone K, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:230709288*. 2023;
34. Penedo G, Malartic Q, Hesslow D, et al. The RefinedWeb dataset for Falcon LLM: outperforming curated corpora with web data, and web data only. *arXiv preprint arXiv:230601116*. 2023;
35. Frantar E, Ashkboos S, Hoefler T, Alistarh D. Gptq: Accurate post-training quantization for generative pre-trained transformers. *arXiv preprint arXiv:221017323*. 2022;