UA INNOVATE
HACKATHON AND INNOVATION CHALLENGE

Company	HCA Healthcare
Category	Prototype Innovation
Description of Prompt	HCA Healthcare is the largest for-profit health care company in the United States, serving more than 44 million patients annually across almost 2,800 sites of care. This generates hundreds of millions of units of HL7 data to hundreds of integrated applications daily. Therefore, HCA's Information Technology Group needs a safe and reliable way to deidentify data for rigorous, HIPAA-compliant requirements analysis and testing. This prototype includes the following: 1. Load and manipulate data 2. Redact data 3. De-identify data a. Keep patient identity consistent across messages 4. Demonstrate results in a user interface Learn about HL7 Version 2 and HIPAA Safe Harbor compliance here: • https://www.interfaceware.com/hl7 • https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html • https://docs.webchartnow.com/functions/system-administration/interfaces/sample-hl7-messages/ • https://en.wikipedia.org/wiki/Health_Level_7 Sample Message Repo: https://github.com/hca-foundation/innovate_25
Submission Software	Microsoft Teams and either Google or PowerPoint are acceptable for any presentation. For submitting the three files, you may email them to chase.callahan@hcahealthcare.com and tara.little@hcahealthcare.com
Suggested Languages/Frameworks/Tools	Open to student's discretion
Midpoint Judging/Check-In	The HCA team will do rounds, checking on participants throughout the competition, beginning at least one hour after participants receive the prompt. We will be available in the Hewson lobby for ad-hoc questions
Category Judges	Judges are HCA developers, analysts, and management who will both judge and answer any questions as you work

Innovate 2025 Prototype Innovation Challenge

Background

To de-identify Protected Health Information (PHI) under HIPAA, you must remove or modify the identifiers listed below, ensuring the data cannot be traced back to a specific individual.

- Names
- Address: Including street address, city, county, precinct, and ZIP code
- Birthdate
- Ages over 89: These must be aggregated into a single category of 90 or older.
- Telephone Numbers
- Email Addresses
- Social Security Numbers (SSN)
- Medical Record Numbers (MRN)
- Account Numbers

Rubric

File Submission (100 Points)

Challenge	Points
Sort all messages by Date/Time (`MSH-7`) oldest to newest and write to a file named `messages_sorted.txt`	10
Redact all PHI in HL7 message and write the messages to a file named "messages_redacted.txt". When redacting, use the * character	10
De-identify PHI in messages and write to a file named "messages_deidentified.txt". You must de-identify the items listed in bullets above. Points will be given based on the following:	10
• All messages for a given patient are mapped to a consistent fake identity. For example, if Joe Smith with MRN 123456 appears in five messages within the source file then five deidentified messages should appear in "messages_deidentified.txt" with a consistent fake identity such as Bob Johnson with MRN 5522334411.	15

• All physicians mapped within the PV1 segments should be mapped to consistent fake physicians	5
• PHI that appears in other parts of the message should be redacted with "*". For example, if the patient's date of birth appears within a note segment (NTE segment).	5
• When de-identifying messages, the fake patient's date-of-birth should be within the same year as the real patient's date of birth.	3
• When de-identifying a patient who is 90 years old or older make them 90 years old.	2
• Account numbers may appear in PID-18 or within a patient identifier list in PID-3. An account number in a PID-3 field will have a PID-3.5 (type code) of "AN" to identify that the PID-3.1 of that field is the account number.	5
• When de-identifying messages, the patient's fake State should be in the same State as the real patient. The City, Zip Code, Street, etc., should (obviously) be different.	5
• Mother-to-baby relationships should be maintained with the mother and baby being mapped to fake identities	10
• Special points for catching and mapping fuzzy PHI-leaks of names. For example, names that are often substituted (e.g., "Elizabeth" and "Beth").	5
• Special points for catching and mapping fuzzy PHI-leaks of dates. For example, "August 1, 1984", "8/1/1984", and "1984-08-01" are all the same date.	5
• Special points for redacting "sensitive words" that were identified in other messages for the same patient but are not present in the known PHI-fields of the current message. For example, a SSN that was found in the PID-19 of another message for the patient, but the current message has a blank PID-19, but the SSN appears in the note fields of the current message.	10

User Interface (60 Points)

Challenge	Points
Shows All HL7 messages	6
Implements Pagination/Lazy Loading for message view	8

For each message, delineate between segments	4
For each message, delineate between fields (See Parse Hog for example)	4
For each message, delineate between all remaining delimiters	6
When hovering over a particular field with a message, show a tooltip of the HL7 path of the data. For example, if hovering over the "Last Name" field within a PID segment, the tooltip would read "PID-5.2"	6
Support filtering/searching of messages by Message Control ID	4
Support filtering/searching of messages by MRN	4
Support filtering/searching of messages by Last Name	4
When viewing a particular message, support the option to show the original and de-identified messages side by side	8
Display the MRN as a link that, when clicked, will launch the MRN filter/search using the MRN	6