Patient Data and Training AI: The OAIC View

Monday 15 September 2025 12:30pm - 1:30pm

Centre for Artificial Intelligence and Digital Ethics (CAIDE)

Patient Data and Training AI: The OAIC View

The Office of the Australian Information Commissioner (OAIC) issued a statement after preliminary inquiries with diagnostic imaging network I-MED Radiology Network Ltd (I-MED), Harrison.ai and Annalise.ai. The inquiries followed media reports that I-MED was allowing Harrison.ai to use its patients' scans to train a diagnostic AI model.[1] Reports raised concerns about I-MEDs compliance with the Privacy Act 1988 (Cth) ('Privacy Act') and the Australian Privacy Principles (APPS).

- APP1 ('open and transparent management of personal information'),
- APP5 ('notification of the collection of personal information') and
- APP6 ('use or disclosure of personal information').

The Privacy Act applies to personal information. This is information or an opinion about an identified individual, or an individual who is reasonably identifiable. Personal information that has been de-identified will no longer be personal information. The report notes:

> Whether a person is 'reasonably identifiable' is an objective test that has practical regard to the context in which the issue arises. Even though it may be technically possible to identify an individual from information, if doing so is so impractical that there is almost no likelihood of it occurring, the information would not generally be regarded as 'personal information'. An individual may not be reasonably identifiable if the steps required to do so are excessively time-consuming or costly in all the circumstances.

What was disclosed?

Patient data disclosed included clinical scans and reports, including X-rays, CT scans and ultrasounds. Before disclosure I-MED processed the data by:

a. segregating the patient data from the underlying dataset,
b. scanning the records with text recognition software,
c. using two hashing techniques (for unique identifiers such as patient ID numbers, and names, addresses and phone numbers),
d. time-shifting dates (to a random date within a specified number of years),
e. aggregating certain fields into large cohorts to avoid identification of outliers, and

[1] https://www.crikey.com.au/2024/09/19/patient-scan-data-train-artificial-intelligence-consent/

f. redacting any text that appears within or within 10% from the boundary of an image scan.

I-MED also imposed contractual obligations on Annalise.ai:

a. prohibiting them from doing any act, or engaging in any practice, that would result in the patient data becoming 'reasonably identifiable',
b. prohibiting them from disclosing or publishing the patient data for any purpose (to prevent wider dissemination of the dataset and accordingly reduce the risk that the patient data may become re-identifiable in the hands of other third parties or the public domain),
c. requiring them to store the patient data in a secure environment, and
d. requiring them to notify I-MED if it inadvertently received any patient personal information.

Furthermore, I-MED developed a Data De-identification Policy and Approach. While patients were not notified of the use or disclosure, and did not provide their consent, I-MED contended this was not required as the patient data had been de-identified.

What did the OAIC say?

The OAIC was satisfied, following the preliminary enquiries, that the patient data shared with Annalise.ai had been de-identified sufficiently that it was no longer personal information for the purposes of the Privacy Act. The OAIC said:

"Our enquiries established that I-MED's de-identification practices reflect many of the practices endorsed by the National Institute of Standards and Technology, including:

- utilising of the 5-Safes Principles,
- ensuring separation of the Annalise.ai and I-MED environments,
- utilising a 'Data Use Agreement Model',
- imposing prescriptive de-identification standards,
- removing or transforming all direct identifiers, and
- utilising top and bottom coding and aggregation of outliers

"[B]etween April 2020 and January 2022, I-MED shared less than 30 million patient studies (a study refers to a complete imaging session for a single patient and may include multiple image types, that together represent a single diagnostic episode), and a similar volume of associated diagnostic reports with Annalise.ai

"During this time, Annalise.ai proactively identified and reported to I-MED a very small number of instances where personal information had been shared with it in error due to failures in the de-identification process. In both cases, the material was subsequently deleted or de-identified.

"Based on the information obtained through the preliminary inquiries, the Commissioner was satisfied that the patient data shared with Annalise.ai had been de-identified sufficiently that it was no longer personal information for the purposes of the Privacy Act.

"Although the steps taken by I-MED could not entirely remove the risk of re-identification, the Commissioner was satisfied that it reduced that risk to a sufficiently low level and was supported by sound data governance practices.

“[T]here were a very small number of occasions where personal information was unintentionally disclosed, but the Commissioner was satisfied that these were relatively minor incidents that were identified and addressed appropriately by I-MED and Annalise.ai.”