

AI and Automation in Metadata at NLB: Experiments and Future Directions

Glossary

| | |
|---------------|---|
| BERTScore | A metric that measures the similarity between AI-generated and manually created output. |
| Cataloguing | Process of creating library metadata that describe and index resources to be discoverable in the library's digital catalogue |
| Fixed fields | Fixed-length data in a MARC record where different characters at specific positions describe the record's structure, the intellectual content of the item being catalogued, and the cataloguing standards applied |
| GI-GO | G arbage i n, G arbage o ut ; an idea in computing where the quality of its output is affected by the quality of the input data |
| GPT | G enerative p re-trained t ransformer ; a type of AI model that uses transformer architecture to generate text. |
| Hallucination | False or nonsensical output generated by AI but presented as though factual |
| LLM | L arge l anguage m odel ; a type of artificial intelligence that uses deep learning algorithms to generate text |
| MARC record | M achine r eadable c ataloguing r ecord ; a record using a standardised format of tags, codes and indicators to store and structure bibliographic metadata about resources. |
| Metadata | Data that gives information about other data / resources |
| SGCAT | Custom GPT prototype cataloguing assistant that can draft bibliographic records based on order & selection information and internal cataloguing practices to increase cataloguing efficiency |
| Singheritage | A controlled vocabulary used by NLB to represent different subjects associated with Singapore |
| Turing test | A test to assess quality by assessing whether a human observer can distinguish computer-generated output from human output |
| WARC files | W eb A rchive files ; a file format used to store web content, including web pages, images, and other digital assets, along with associated metadata |
| WAS | W eb A rchive S G ; the archived collection of .sg websites documenting Singaporean life, culture, and history in the 21st century |

Links

1. Paper on metadata generation for WebArchiveSG:

Nair, A., Goh, Z. R., Liu, T., & Huang, A. Y. (2025). Web Archives Metadata Generation with GPT-4o: Challenges and Insights. *Information Technology and Libraries*, 44(2). <https://doi.org/10.5860/ital.v44i2.17305>

2. Link to WebArchiveSG portal: <https://eresources.nlb.gov.sg/webarchives/landing-page>

3. Video demo for SGCAT:

https://drive.google.com/file/d/1js5fYpJwKm9wnEo3QZKICysyCW7E_dnl/view?usp=sharing