Automatic for the People: Archives and the Future

Pip Willcox
The National Archives, UK
Pip.Willcox@nationalarchives.gov.uk

Abstract

Artificial intelligence will take half our jobs in the next 20 years; people will be left helpless and subjugated to the will of machines. (We're all doomed.)

Artificial intelligence is a distant dream; our current machine learning methods are overblown hype destined to fail in their promise to improve lives. (We're all doomed.)

Artificial intelligence has already taken over, and our historical biases are baked into it. (We're all doomed.)

Between these extreme positions – exaggerated only slightly – that scream at us from headlines and sidebars, is a world of nuance: of ethical, legal, technological, commercial, environmental, cultural, infrastructural, and practical considerations.

For centuries archives have collected, organised, preserved, and enabled access to materials produced by other activities – administrative, social, cultural, musical, legal, and commercial. What this work has entailed has been transformed by our embrace of digital technologies, and we have seen the term 'archive' borrowed and repurposed by our friends in computer science. Artificial intelligence is one area where the term's two usages approach each other, and where the importance of archives and their function of holding people and systems to account is evident.

In the practice of archiving, there may always have been tension between attention to detail, and descriptions of material 'good enough' to enable their discoverability and re-use for consultation and research. We are seeing this tension through a new lens and at a new scale as significant born-digital archives join digitised, analogue archives in the collections we curate. At The National Archives (UK)[1] we are experimenting with computational approaches to help us in our mission to collect and preserve the record, to connect people with their history, and to support the work of archives.

The National Archives is an essential resource for our democracy, a public good, and an asset for future generations, as we lay out in our vision *Archives for Everyone*.[2] To fulfil this role, we are committed to building trust, and tearing down barriers to access, participation, and understanding. It is frequently noted that while digitising analogue collections enables new knowledge to be created and linked, it creates a further collection that equally requires preservation and curation. As we aim to increase understanding and remain trusted, digitisation also demands contextualisation and explanation, linking across and beyond individual collections to understand, for example, voices and perspectives that are not recorded, reflected and valued. As we look to computational methods to assist us in this, we consider the algorithms we use, their training data, the software we use to understand and create new links, the links themselves, the wider hyperstructure and the uses it enables, the content behind those links, the tools we use to explain the approaches, the people who use our resources, and the standards that underlie all of them.

Drawing on collaborative research at The National Archives, including through the UK Arts and Humanities Research Council's programme Towards a National Collection, this talk explores computational archival science, artificial intelligence, citizen involvement, and post-custodial approaches to challenge doom-laden technological determinism, and how together we might combine 'hand-curated' and 'at-scale' approaches to our shared cultural heritage to ensure automation works 'for the people'.

---

1   https://www.nationalarchives.gov.uk/ (accessed January 12, 2022).

2   https://www.nationalarchives.gov.uk/archives-for-everyone/ (accessed January 12, 2022).