Grand Challenges

The Collaborative Computational Project for Arts, Humanities and Culture (CCP-AHC)

Over the next five to ten years, CCP-AHC will become the first network of its kind to support the AHRC community by developing wide-scale access to AI inference for primary source materials

The Community

CCP AHC Collaborative Computational Project for Arts, Humanities and Culture

www.ccpahc.ac.uk

CCP-AHC will be the first network of its kind to support the AHRC research community and have a significant impact both nationally and internationally. The creation of CCP-AHC is timely, given the proliferation of computational tools and techniques—more specifically HPC and AI - into various research agendas of the AHRC. With the AHRC research community increasingly dependent on all aspects of digital research infrastructure (DRI) and with few collaborative efforts focused on developing cohesive visions, initiatives, or DRI assets, the launch of CCP-AHC is very timely.

The Challenge

Access and available services needed to essential services for AHC researchers

Arts, humanities, and culture (AHC) touch every citizen's life. Research in the area contributes to the national wellbeing, the preservation of the natural and built historic environment, and the maintenance of just and strong institutions. It directly addresses phenomena such as creativity and meaning, by turning to archives for evidence of human behaviour which may be preserved as text, image, sound, video, or artefact. By interpreting these traces of human activity, AHC researchers develop a critical understanding of contemporary challenges such as social cohesion, societal resilience, climate change, responsible innovation, and the preservation of diverse languages and intangible practices. The digital transformation of research enables the use of large-scale computational techniques to support interpretations. The recent rapid development of data-intensive techniques suitable for AHC data, driven by public and private investment in artificial intelligence (AI) and machine learning innovation, alongside the consolidation of DRI provision across research areas in the UK offer an opportunity to realise the promise of wider-than-ever availability of the source materials of AHC research.

Galleries, libraries, and museums already hold terabytes of digital collections, countless millions of other holdings, sometimes in personal, local, or community collections, which are yet to be ingested or processed. Existing large-scale born-digital collections, such as the UK Web Archive, are already being curated today at

Grand Challenges

petabyte scale. AHRC-led investments will in the coming decade produce large volumes of digitised holdings in the coming decade. However, born-digital archiving efforts or the production of new digitisations of analogue material do not by themselves produce new results. These data often need to be further transcribed, annotated, or otherwise reformatted before they can be analysed computationally. Researchers need access to services that allow them to ready them for large-scale processing by specific research codes.

The Solution

To accelerate the production of research in AHC by making established techniques available to technical and non-technical AHC researchers

There is no one-size-fits all code or service that will accelerate the production of research uniformly across the varied subfields within the arts, humanities, and culture area. However, there are many currently labour-intensive operations that shared across many disciplines. These tasks have established computational solutions, approach human-level accuracy, including audio transcription, optical character recognition (OCR), and image captioning. These are all powered by accelerated compute and, in many cases, open-source codes and pre-trained models, which are already competitive with state-of-the-art technology. CCP-AHC's vision is to make these techniques available to technical and non-technical AHC researchers working across academia, the non-profit sector, and the culture industries regardless of their affiliation or current access to compute and ability.

This is made complex by the diversity of kinds of data that support AHC research: texts, images, video, audio, and multi-dimensional data derived from physical artefacts are all routinely used. The rapid development of novel approaches also demands new benchmarks to help researchers find appropriate, efficient, sustainable and responsible solutions. In the future, no AHC researcher or innovator will be prevented from exploring the capabilities of AI and machine learning, running on publicly large-scale compute by the uneven distribution of skilled dRTPs or access to DRI within the sector.

The Outcome

Wide-scale access to AI inference for digitised AHC primary source materials

In the future, no AHC researcher or innovator will be prevented from exploring the capabilities of AI and machine learning, running on publicly large-scale compute by the uneven distribution of skilled dRTPs or access to DRI within the sector.

More Information

CoSeC
www.CoSeC.ac.uk
CoSeC@stfc.ac.uk

CCP-AHC
www.ccpahc.ac.uk
ccpahc@durham.ac.uk