

QUALITY ASSURANCE OF APPLICATION OF AI AND MACHINE LEARNING ON ART AND CULTURAL HERITAGE – OPPORTUNITIES AND CHALLENGES FOR FUTURE ART HISTORY

Jan von Bonsdorff

Department of Art History, Uppsala University

AI advancements enable new art-historical research tools, like OpenAI's GPT-4 image-to-text function. Emerging cross-modal translation models can link texts, images, and more. Quality assurance in digitising art sources and training AI models is crucial. Concerns include source origin, content curation, presupposed image models, and potential cultural bias. It is essential for art historians to influence digitisation and AI data processes to ensure quality research when using AI-based tools.

POSITION PAPER

New methods of collecting, classifying, handling, searching, and generating images are rapidly developing through massive digitisation of visual content, new AI algorithms and machine learning. For the discipline of art history, this development brings access to new research tools: OpenAI has recently rolled out the image-to-text function of GPT-4, which enables automatic text annotations of images. Further off, emerging at the research horizon, we find notions of complex cross-modal translation models, enabling intermedial connections between texts, images, sounds, music, and film.

In all cases, such models are fraught with growing pains. Normal source critical questions (källkritik) apply. With my position paper I would like to encourage a general discussion on quality assurance of digitisation and machine learning in connection with art-historical sources. How much and what material is within reach (eg. "fine art" or "stock photos")? Are the digital contents aggregated freely from the internet or contained within a database or a data-driven ontology? And important for the art historian: Is the content curated? Where do metadata and annotations stem from? What historical and cultural framing (old museum, newly acquired private property etc.) has determined the collection in question? What is the cultural bias of the collection? What information is missing or even suppressed?

Let us move from the critique of the archive to the analytical part: In short, any content and any theory or ideology with an art historical bearing can be used for machine learning: style, performativity, Wölfflin's binary concepts – all this is possible to teach an AI (and it has been done). A revolution in the field came in early 2021 with transformer or stable diffusion algorithms based on "ontologies" or

“learned representations” consisting of millions (OpenAI) or even billions (LAION) of paired texts and images (that is, still, a restricted multimodal model).

A “best practice” example of a working and efficient search engine for Fine arts using stable diffusion algorithms is the open Web platform iArt (since 2021; <https://labs.tib.eu/iart>). iArt specializes in differentiated fine art searches, using a complex modular system architecture. The iArt search engine masters iconographically based classification principles that, e.g., examine objects for biblical motifs, but also more general genre themes. Textual prompts can be of any length, within reason. The search engine has the capacity of nearly simultaneously showing a bird’s-eye overview of similar visual results, along with finding a relevant visual answer to a specified research question at a micro scale, either in form of a text or an image, or combinations thereof. The team behind iArt are currently ameliorating and infusing the iArt system with a vast text corpus of thousands of art-historical texts, thus, creating a culturally and visually competent – and curated – backbone to the search engine.

In this way, the iArt team show that it is of utmost importance that art historians can exercise influence of and control the whole process from data gathering to the training of machine learning models and forward to the analysis of the results, thus firmly keeping a provable quality assurance in hand.

But the machine learning field is open for more art-historical input: What about pre-conceived image models? Do we reach an accurate automatic visual description without a proper – and verbalised – fundamental notion of the image? What kind of diverse notions of the image do art historians, semioticians, archivists, museologists, and IT scientists have? What elements and levels are proposed, if there is a structural level in the model? If the term discourse is used, what place does it take? What are the intrinsic and extrinsic features? What are the denotative and connotative features? What about agency and narrative functions? Can we discern linear, centripetal, centrifugal, or rhizomatic models? What about materiality and immateriality? Where is the “/” between text/context?

Throughout this amorphous field, I think it is proper to keep some kind of art-historical compass at hand. Here I add some guide-lines for the treatment of images when facing machine learning:

- take the world views of images seriously
- exploit the discursive conventions and strengths of images, e. g. agency and metaphors
- describe rounded, credible visual counterparts embodying social trends, norms, behaviours, styles, attitudes, moods
- describe pictorial conventions staging the visual arena where something is shown, pointed at and commented – it may even be you (interpellation)
- describe pictorial conventions playing with suggestive ambiguousness, activating curiosity
- exploit the richness of imaginative, flowery and ornate descriptions in natural language

- let the images show a complex and imperfect social world, where poverty, injustice, sexism, xenophobia, racism exist – but also are resisted

Through our discussion, I hope we could compartmentalise the field. Let us use the utility criterion as a tool – what would you like to focus on, if you just had two weeks to teach your undergraduate students the core of art history on an isolated island with no textbooks? Most likely, that is exactly what our emerging tools, the AI models, also should learn.