Content Aware and Other Case Studies: Museum of Synthetic History

Egor Kraft
Academy of Fine Arts Vienna
Vienna, Austria
egor@egorkraft.com

Ekaterina Kormilitsyna
Academy of Fine Arts Vienna
Vienna, Austria
kormilitsyna.k@hotmail.com

With the rise of artificial intelligence and related computational tools in everyday dealings with knowledge organisation, production, and distribution, incl. for example archives and history-related applications, we're concerned whether these computational methods 'colonize' and fundamentally change our common approaches to what constitutes studying and knowing a subject matter. We will unpack upon these concerns, looking at phenomena such as a lack of completion and categorisation in biodiversity archives, or new methods of creating artificial fossils as ways of filling gaps within historical datasets and potentially narratives. We also call back into how ontological architectures of computer science have emerged and how they defined ways in which knowledge is accessed. Via the examples of various case studies and thought experiments, the paper tries to examine the initial concern and predict its potential consequences, building upon the question as to what degree machine-learning-based approaches can augment our methods of analysis not just in history but in cultural behaviours.

In other words, how might computational models of ontology be producing an epistemological shift within the quality of knowing by imposing a knowledge system of references, linked nodes, hashtags, and databases that are never entirely complete in representing subjects they are set to define. Thus, asking if we shall hold on to our approaches of comprehension of things and their emergence or instead succumb to the generative, on-demand, a click away, always-at-your-fingertips forms of knowing and comprehending?

Domain Ontology. Computer Science. Meta-Archaeology. AI.

1. INTRODUCTION

> "The whole age of computer has made it where nobody knows exactly what is going on",- President-Elect Donald Trump on the 28th of December 2016.

In a room of the National Archaeological Museum in Athens, people walk alongside stone sculptures showing scars of their excavations; Stones, earth, clay, metals all illuminated by blue light, stored in glass vitrines. At the back of the exhibit, broken into small fragments lie calcified pieces of such metal and stone. A circular structure discoloured by seagrass, eaten to pieces by time and dismantled by humans so eager to preserve it in the name of knowledge. The Antikythera Mechanism, put in place after its discovery in 1901, dragged from the bottom of the sea to be displayed to audiences seeking to understand the past. We lean on this display of objects ripped out of time, as our crutch of engaging with history. Once we find them, they become proof of historical occurrences and on them, we mould our new understandings. The discovery of the Antikythera Mechanism introduced new timelines of the origin of mechanical operation structures, while its concrete purpose still remains unknown. Our 'history' as a notion of understanding the past and the current seems to be object-based on documentational 'proof' which separates 'truth' from 'speculation'. However, what if the creation of these objects were no longer a process of seeking blindly? What if the unknown was to be fulfilled by an analysis of the known?

Figure 1: The Antikythera Mechanism is in the collection of the National Archaeological Museum in Athens,Greece.Tilehamos Efthimiadis, Wikimedia Commons / CC by 2.0

In this text, we are suggesting that there is an epistemological shift in how machine accelerated logistics of information affect the formation of knowledge, archive, and historical record. There are many various ontological theories within the science of being, although not all of them have become so increasingly and forcefully imposed on designing our infrastructures of knowing, learning, and doing, as a set of ontologies upon which computer science was initially set and continues to operate today. Taking this argument further we find it interesting to ask whether the following holds true: a set of ontologies that emerged in computer science colonise other ontologies and render previously established ones obsolete by introducing computationally accelerated ontological structuring and logistics of meaning. Computational ontologies are interdependent with the properties of computational systems, such as scalability, accessibility, interoperability, and others. What are the effects of these properties on how we structure, record, and acquire knowledge? We find it important to look into what issues they may project in relation to studying subjects through their representations in a form of data accessed and delivered via the means of computational networks, i.e outputs derived by machine learning systems or simply web search queries.

Many of the modern machine learning developments were designed as means for automation of information processing while some of them also became means of knowledge production on their own. As opposed to the conventional ideas of knowledge such AI models hold a peculiar quality that can be described with a somewhat capacious word: **programmability**. These precisely programmable systems can generate information upon demand and thus can be seen as forms of programmable knowledge on their own.

What seems ontologically challenging can be inquired via the following question: in which context does the difference hold importance between, let's say, **'a stone image'** generated by a GAN model (General Adversarial Network) capable of generating thousands-a-second of hyper-realistic images of subjects it was trained on, or between a random image delivered via Google search query with the **"stone"** keyword, or a notion of a **stone** within our collective or individual memories. And can any of these derivatives be less ontologically corrupt and thus more "real" than the others? A question similar to the one asked by Joseph Kosuth in his 1965 iconic piece "One and Three Chairs" (Kosuth 1965). He conceptually challenged forms of subject representation. And we believe that a few radically new ones have emerged since the 1960s.

Figure 2: Larry Aldrich Foundation Fund © 2022 Joseph Kosuth / Artists Rights Society (ARS), New York, Courtesy of the artist and Sean Kelly Gallery, New York

2. ONTOLOGICAL TAKEOVER

Ontology seeks the classification and explanation of entities, as a branch of philosophy, it deals with

questions of origin and existence, but the term has found a modern purpose within the context of AI. Containing the idea of a shared vocabulary, definitions of concepts, and the relationships between them, ontology facilitates an understanding of the architecture of AI systems. Tom Gruber, an American computer scientist recognised for foundational work in ontology engineering in the context of AI, in his 1993 paper "*A Translation Approach to Portable Ontology Specifications*" says, "For knowledge-based systems, what "exists" is exactly that which can be represented." (Gruber, 1993). In other words, in a knowledge-based program, vocabulary **represents** knowledge itself. Computational ontology functions as a database, a structure of information organisation, it is not only the definition for a branch of philosophical discipline but an actual architecture, that largely governs logistics of knowledge and meaning. Computer system ontologies rely on entities such as hypertext, hyperlinks, hashtags, metadata, ascending and descending orders, hierarchies of access, file systems, variables and extensions, executables and more, they are devices and elements of the architecture of knowledge-organisation systems through which they deliver, extract, produce and engage with knowledge itself. Do such developments also bring a change in how we as communities' access and engage with historical knowledge? The answer seems, inevitably, yes.

A reasonable concern to follow, would be how this might change us in return? Some studies suggest an effect of a feedback loop in which the use and implementation of tools create a change in human behaviour. Research at Emory University provides an example of a feedback loop that is intrinsically epistemic: it shows that neural circuits of the brain underwent changes to adapt to Palaeolithic toolmaking, thus playing a key role in primitive forms of communication (Stout 2016, 28-35). Projecting these dynamics onto various forms of computational accelerated forms of engagement with knowledge, we may observe a peculiar relationship in which human interactions with knowledge change to develop structure patterns similar to those of computational ontologies, i.e. hashtags, hypertexts and such. If concerns around the philosophy of language helped us to better learn the correlation between the language, meaning, knowledge, perception, and the world, we may suggest that we will soon need the study to see how computational semantics and generative models affect them too.

Following this thought, the introduction of a network-based knowledge access model can be traced to have brought a database approach to learning behaviours. This could be attributed to both the introduction of the mere accessibility of the vast pool of information and knowledge provided by the internet as well as the methodology through which we have learned to navigate this pool. With knowledge at our fingertips have we been adapting

Machine-derived behaviours, like navigating architectures of knowledge through keywords, hashtags, and reference ontologies rather than internalising it in ways our evolutionary biology suggests? The way information is distributed is defined by the current state of logistics of information technologies. So, what might it mean for attitudes towards information processing and engagement with historical objects when we regard knowledge as something to reference rather than to learn? The focus then lies rather in the development of the quickest architecture for navigation of information rather than infrastructures of passing down knowledge. In this context of a high-speed information highway architecture, we want to look at the process of computational analysis of data and information in particular, historical data, and the growing use of AI investigation tools applied to historical archives, which are described in this text, and these machine-learning outputs of artefacts of the natural sciences as objects of knowledge in this altered state information classification.

So rather than considering how our human interactions with knowledge have adapted in isolation, we consider how our interactions have adapted due to the inclusion of AI mechanisms which in turn rely on ontological knowledge models. AI as an algorithm is a perpetual learning machine, everything else that is gained from it is secondary- its primary function is to learn. Seeking knowledge for seeking's sake.

It must be noted that there are many types of classified AI categories, including Machine learning, Deep learning, Natural language processing, Computer vision, Explainable AI, reactive, limited memory, theory of mind and others. Some of them are ontology-based, while others are self-learning systems. For example, machine learning-based systems use statistical classification of patterns to compare what they have learned from training sets to new data, to determine whether it fits a pattern. Whereas ontological architectures of AI are very different, "Ontology-based AI allows the system to make inferences based on content and relationships, and therefore emulates human performance." (Earley 2020). Considering such ontological dynamics, we must turn a critical eye towards archives and databases, and the biases that are already embedded in them, as well as towards the motivations and intentions behind the applications of computational knowledge production. The 'Museum of Synthetic History' presents a case study of such critical interventions.

3. MUSEUM OF SYNTHETIC HISTORY

To display the consequences and explore the possibilities of an epistemological shift in machine-lead information architectures, we engage with the 'Museum of Synthetic History', which is an artistic

research project led by Egor Kraft and forms the continuation of the existing project CAS, as a thought experiment, looking at biodiversity and archaeological practices surrounding fossils. We speculate on the role of AI in the analysis and creation of archive data and highlight the concerns which ought to be regarded when using AI for nature-science research. The 'Museum of Synthetic History' as a research project becomes a visual, spatial, and archival output of this investigation, as both: metaphorical imaginations of a museum space filled with synthetic pieces of history created by AI-palaeontologists and real collection and archive of such outputs to further the investigation into the consequences and the biases of the implementation of contemporary problematics in biodiversity and natural science fields.

3.1 On Biodiversity

The Earth's estimated biodiversity is in the order of 10 million species, from which only 10–20% are currently known to science, while the rest still lacks a name, a description, and basic knowledge of its biology. (Krishtalka and Humphrey 2000; Wilson 2003; Costello et al. 2015; Sampaio et al., 2019) The biodiversity extinction crisis is an alarming trend across related fields of science. The rate of biodiversity loss is accelerating, leading to a tendency for "Big Data" production on species observation-based occurrences instead of specimen-based occurrences as a way to map and protect biodiversity (Troudet et al. 2018). During 300 years of biodiversity exploration, many organisms were collected, catalogued, identified, and stored under a systematic order (Sampaio et al., 2019). However, many samples there are, we're barely reaching a quarter of well-documented observable species on Earth, which form the basis of this data-driven research. The consequence of a lack of this knowledge is the loss of irreplaceable sources of high-quality biodiversity data and the proliferation of misidentified records with poor or no corresponding data. All of which, in turn, results in a doubtful source of knowledge for future generations. (Troudet et al. 2018; Sampaio et al., 2019) In other words, in writing the history of life on earth we're currently limiting ourselves to recycling only the existing data in a feedback loop machine of widely available and trending computational methods, such as data-driven and AI-powered research technics. This is an alarming trend in varying fields and within any application of AI since a dataset will never be truly complete.
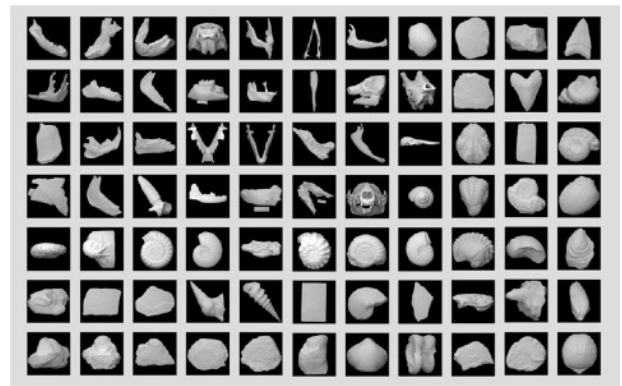

Figure 3: Fragments of a custom-produced dataset of 3D scans of paleontological findings, including fossils, corals & other biogenic items

3.2 Artificial Fossilisation

the University of Bristol, under the supervision of Jakob Vinther, Evan Saitta and their team have been conducting research into artificial fossilisation. The aim of their developed methodology is to aid in the process of finding fossils in order to continue the aim of completing our archives of biodiversity and understanding of paleontological history by reverse engineering fossils.

Their published experimental protocol may indeed change the way fossilisation is studied as they've unlocked methods to manipulate time, not the least force behind the creation of a fossil. Through specially developed technics directed to produce artificial fossils, the research group managed to synthetically compress millions of years of natural processes into a single day in the lab. Those artificial fossils are synthetic by origin, yet visually indistinguishable from the genuine ones, and as material analysis reveals structurally very similar according to the claims in their paper published in 2018 (Saitta, Kaye and Vinther, 2018). 'Artificial maturation', is an approach where high heat and pressure accelerate the chemical degradation reactions that normally occur over millennia when a fossil is buried deep and exposed to geothermal heat and pressure from overlying sediment. Maturation has been a staple of organic geochemists who study the formation of fossil fuels and is similar to the more intense experimental conditions that produce synthetic diamonds.

"The approach we use to simulate fossilisation saves us from having to run a seventy-million-year-long experiment," reported Saitta," We were absolutely thrilled. We kept arguing over who would get to split open the tablets to reveal the specimens. They looked like real fossils - there were dark films of skin and scales, the bones became browned. Even by eye, they looked right." (Starr 2018; Field Museum, 2018)

In their own words they nick-name the procedure 'Easy-Bake fossils', gamifying objects of history as their purpose becomes another one entirely,

specifically that of a tool. They describe the possibilities of their approach as ones of reverse engineering. "Our experimental method is like a cheat sheet. If we use this to find out what kinds of biomolecules can withstand the pressure and heat of fossilisation, then we know what to look for in real fossils."(Field Museum, 2018).

Figure 4: *Artificial Fossil. Photo courtesy: Saita et al./ Palaeontology, 62(1), pp.135–150 (2018)*

In this case, archaeological practice becomes a matter of knowing what to look for as opposed to trying to find the undiscovered. From Saitta's statement we understand that in order to combat such problematics like the biodiversity crisis and similar problematics in the paleontological field they intend to attempt to work backwards; To take an organism or marker which is currently in existence, create an artificial fossil of it and review what remains after over a millennium of ageing processes. The remaining markers then become guidelines of what to search for and if found become a new string of the historical narrative of this planet.

Peculiarly we are now faced with a type of 'reversed archaeology', where history is predetermined in a lab and fieldwork becomes a matter of finding the piece which fits the artificially created template. A painting-by-numbers type of paleontological puzzle, which leads to yet another type of recycling of knowledge as opposed to random discovery through seeking; Similar to the problematics which occur when attempting to expand biodiversity data using AI technics on an existing dataset. The consequence of machine-learning knowledge production is that AI approaches questions with the intention of solving them, no matter how much force it must apply to mould the existing data into a solution to the set task. How big is the gap between a DeepDream plate of spaghetti and meatballs morphing into a hellscape of dogs as AI constructs the hallucination with brute force, to archaeologists creating 'easy-bake' versions of fossils and scouring the earth for their counterparts potentially blind to the unknown and undiscovered data around them?

Figure 5: *DEEP DREAM IMAGE; Artist unknown: https://www.stichtingopen.nl/does-android-dream-of-electric-dogs-does-googles-deep-dream-reveal-anything-about-psychedelic-hallucinations/*

3.3 Synthetic Objects

The 'Museum of Synthetic History' builds on these ideas. Preoccupied with the issues of biases in AI-driven research practices today, the 'Museum of Synthetic History' challenges previously established AI-based methodologies, against data from prehistoric and geologic time archives including first stone tools, writing systems, paleontological archives of fossilised plants, organisms, and other biogenic data. How different would an Ai-composed or, ontologically speaking, - synthetic plant fossil seem as opposed to an actual sample from prehistoric floras? Or will AI-manufactured proposals of newly rendered specimens be distinguishable from the remaining millions of actually existing species that never made it to get scientifically catalogued? And, finally, what would that mean to actually produce such objects involving artificial fossilisation technics in terms of philosophical concerns around ontology, agency, and materiality of organic and inorganic subjects? Or what would bone remnants of prehistoric species look like if they were algorithmically composed and then 3d-printed in calcium phosphate? Such engineered bone tissues or artificially maturated stone imprints may come across as indistinguishable from genuine paleontological findings.

What new domains of natural sciences will emerge when the history and ontology of floras, faunas, single-celled organisms, yeasts, moulds, rocks, minerals and those of unearthly origin, is studied by algorithmic forms of knowing? In other words, we can even go so far as to say that the project 'Museum of Synthetic History' is a thought-object experiment into simulating a situation in which the agents of artificial, automated reasoning committed to the conceptualisation of their own emergence and production of their own history and artefacts.

4. GENERATIVE HISTORY

What artificial maturation experiments, biodiversity problematics, and thought experiments into a 'Museum of Synthetic History' have in common is

that they look into the ways data can be produced to fill in what we may define as blank spots in the 'dataset' of our entire history, attempting to create a more 'complete' picture of a subject matter. We are confronted with two radically different ontologies: one archive-based, documental vs. a generative one, the latter intends to produce outputs on demand, whereas the former is rather static and linear. The former is perhaps how history is meant to be, isn't it?

Let's for a second imagine a present that is in constant shift and flux, shaken by earthquakes of change that occur with every newly artificial fossil, created as an algorithm runs a dataset based on the existing documented nature-history archives. A generative historical ontology suggests that its expansion of knowledge of itself occurs through an analysis of itself. Instead of the unknown- we are presented with an ever-shifting 'known', which duplicates into grotesque copies of itself, barely recognisable yet copies all the same. In other words, throughout this text, we have confronted ourselves with structures of how an understanding of paleontological and nature-history sciences operate and how these methodologies are being augmented through the incorporation of computer sciences. There is nothing speculative about the incorporation of these new methodologies. However, they lend themselves to the imaginings described above. Does a generative history position us inside of a generative present, and would such a present look like one described above? We do not know, however imagining such scenarios allows us to visualise the biases and problematics as critical thought when engaging with computational knowledge-producing agents in natural science and historical data analysis.

In concluding thoughts, at the beginning of this paper, we posed the possibility that the very manner in which one interacts with knowledge has shifted or whether our access to history itself has changed. The case studies and experiments explored above primarily deal with historical archives. When looking at generatively created objects we are looking at a history created in the present. When such methods are being used within a historical investigation, we are rendering archives of the past in real-time. This paradox could be well observed within the case study of the 'Museum of Synthetic History' experiments in which archaeological findings are produced by AI systems which are then artificially matured via physical manipulations. Such objects turn out to be not only illustrations but also case studies of described above problematics. We might also go further and introduce a term - **'post-archaeological object'**, an object which is by all of its measurable material qualities stands for an archival object, but in reality, has been generated only to match these measurable material and symbolic qualities. We can think about the "Ship of Theseus" and whether an object replicated down to its molecular structure it could still be considered the same? Or we might

be tempted to revisit thoughts and ideas on such notions as genuineness, authenticity or even Benjamin's aura, as he argued that the trace of an aura and history of an object may only be brought to light by the chemical analysis (Benjamin, 2003). A criteria which the artificial fossil fulfils, creating the conundrum of this explorations. However, this was not the aim of this text. Instead, here we wanted to highlight the urgency to understand the epidemical nature of somewhat forcefully imposed computational ontologies and their effects on our relationships with the past.

These imaginations are neither dystopian nor utopian; They in no way intend to devalue the richness and incredible applicability of various domains of computer science; however, we suggest that at least some of this criticism and reflections are kept in mind when designing new tools and methods for computational utilities in natural sciences and historical analysis. The 'Museum of Synthetic History' may be seen as a device for critical thinking towards a future in which the past, as in history, is generative, synthetic and merely a-click-away computational output.

3. REFERENCES

Benjamin, W. (1935; 2003). *The work of art in the age of mechanical reproduction.* London: Penguin Books.

Earley, S. (2020). *The Role of Ontology and Information Architecture in AI*. [online] Earley Information Science. Available at: https://www.earley.com/insights/role-ontology-and-information-architecture-ai [Accessed 13 Sep. 2021].

Field Museum (2018). *Press Release: Easy-Bake Fossils*. [online] Field Museum. Available at: https://www.fieldmuseum.org/about/press/easy-bake-fossils [Accessed 7 Sep. 2021].

Krishtalka, L. and Humphrey, P.S. (2000). *Can Natural History Museums Capture the Future?* BioScience, 50(7), p.611.

Goled, S. (2021). *Ontology In AI: A Common Vocabulary To Accelerate Information Sharing* [online] Analytics India Magazine. Available at: https://analyticsindiamag.com/ontology-in-ai-a-common-vocabulary-to-accelerate-information-sharing/ [Accessed 3 Sep. 2021].

Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), pp.199–220.

Gruber, T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies,* 43(5-6), pp.907–928.

Saitta, E.T., Kaye, T.G. and Vinther, J. (2018). Sediment-encased maturation: a novel method

for simulating diagenesis in organic fossil preservation. *Palaeontology,* 62(1), pp.135–150.

Sampaio, Í., Carreiro-Silva, M., Freiwald, A., Menezes, G. and Grasshoff, M. (2019). Natural history collections as a basis for sound biodiversity assessments: Plexauridae (Octocorallia, Holaxonia) of the Naturalis CANCAP and Tyro Mauritania II expeditions. *ZooKeys*, 870, pp.1–32.

Starr, M. (2018). Researchers Have Discovered How to Make Proper Fossils - In a Day. [online] *ScienceAlert*. Available at: https://www.sciencealert.com/fake-fossil-method-baked-in-a-day-artificial-maturation-sediment [Accessed 11 Aug. 2021].

Stout, D. (2016). Tales of a Stone Age Neuroscientist. *Scientific American,* 314(4), pp. 28–35. DOI: 10.1038/scientificamerican0416-28

Thielman, S. (2016). Donald Trump is technology's befuddled (but dangerous) grandfather. [online] The Guardian Available at: https://www.theguardian.com/technology/2016/dec/30/donald-trump-technology-computers-cyber-hacks-surveillance

Troudet, J., Vignes-Lebbe, R., Grandcolas, P. and Legendre, F. (2018). The Increasing Disconnection of Primary Biodiversity Data from Specimens: How Does It Happen and How to Handle It? *Systematic Biology,* 67(6), pp.1110–1119.

Wilson, E.O. (2003). The encyclopedia of life. *Trends in Ecology & Evolution*, 18(2), pp.77–80.