

Curating Data exam 2020

Introduction

Curating used to be interpreted as a practice of managing an institution's collection (Tyzlik-Carver, 2017). With a more digitally advanced world, shaping the world as we know it and redefining phenomena and ideas, one might challenge the traditional understanding of things. A practice like curation has been redefined as a more abstract phenomena in a digital context. With curation of data, anyone with tools of some sort, like a computer can be a curator of data. One can argue that it is difficult not to be, since it is such an inclusive part of our lives. Whether being collected, digitized, archived, analyzed or presented, data curation is a very current and almost unavoidable discipline used by many.

With topics like big data, AI/machine learning and General Data Protection Regulation (GDPR), curating data has a lot of interesting applications and values, for scholars to work on. With debates about owning data, and letting data get "out of hand", the public's opinion and issues about this matter, is only getting more and more complex. The ever-growing complexity of a digitalized and datafied world, is described as a data revolution, "Second, there is a data revolution underway ... is being radically transformed". (Kitchin, 2014, p. 26). This radical transformation of data and the digital world leaves a demand on people, to have the necessary skills to adapt to this environment. This is the big link from curating data (the course) and this report. Understanding what data curation is, why it is important in our modern lives, the issues it presents and how it influences and shapes our lives.

This report focuses on the theme of exploring the interaction between the physical and digital world. With curating data, this theme can be elaborated to illustrate ideas and thoughts about curating data. This theme corresponds well with some of the literature in the curriculum, and further expands some of the discussions in the classroom.

In the last part of the report, there is in the appendix 4 a link to my R&D area, with further links used in the assignments, like a link to downloading the dataset from assignment 1 (Appendix 4).

This will not be further referenced, but will be an intuitive tool, to go along with reading this report.

Assignment 1. Digital Collection

The digitization, curation and datafication of material objects into a dataset was the object of this assignment. To somehow digitize analogue objects and conceptualize a train of thought through this process, and furthermore to reference this to the curriculum.

Subjective Objects is my attempt to solve this assignment, with a simple idea of collecting household items (material objects) and conceptualize an idea of subjectively valuing collected objects, and define them on my own terms, hence *Subjective Objects*.

The objects have been digitized by taking and uploading pictures, and then collected and sorted in a dataset. The pictures in themselves serve only the purpose to represent the objects' value in a digital visual format, however the variables and categorisations will attempt to add attributes and depth. I base my dataset on 20 different household items, and 6 different categorical values; aesthetic-, practical-, sentimental-, economical- and sharable values, and then the objects possibility of complete digitalisation (POCD). These categories are represented on a scale of 1-6, 6 being the most favourable score of the category, e.g. if an object scores 6 on the economical scale, that means that it is a rather economical object, relative to its field of context/function. The sharable value is scaled on how sharable the object is.

In the dataset I scored object 117 to have a sharable value of 6, meaning I find this particular object to be extraordinarily sharable within its mean of function. Object 117 is a picture of banknotes representing physical money, which I find sharable, since they change hands possibly multiple times a day. The category POCD which might need some explanation, is a measure of how possible I find a completely digital version of an object, to be able to substitute the essence of the material object. POCD is the only category which can get the score; not applicable (N/A). This is not necessarily a low score, however I find the degree of digitizing the object incomprehensible. POCD is sort of the "joker" of the categories, since I find it to be the most debatable.

In the dataset you will see I scored 6 in POCD to object 113. Object 113 is a leather bag which is probably a good example of the debatability of the POCD category. The essence of the leather bag

is to transport objects in a favourable manner, usually books, a computer, and maybe office tools such as paper and pencil. This essence is practically completely digitized in different ways like having a folder on your computer containing subfolders with relevant functions, such as a word document or PDF files. By dragging the folder around different storage applications, or even sending this “bag pack” to a friend, you are substituting the physical function with a digital solution. This does not mean, that the digital version is substitutes the physical object, with it acts as a digital replica, substituting the function.

You could also argue that object nr. 119, which is a computer, has the intrinsic capability of solving the “bag pack function” in a completely digitized way. This is like what Katharina Weinstock writes in *Rearranging the world*, “Everyone with a personal computer has become a de facto archivist, storing and filling thousands of documents, images, and music files” (Weinstock, 2020, p. 232). In this paper she talks about how objects and collections have been rearranged from pre to post internet and digitalisation. In the quote she talks about how everyone equipped with a personal computer has become an archivist, which I find interesting when talking about valuing objects ability to complete digitalization. The computer seems to have been a steppingstone for transferring materialistic objects into a digital world, making the average person a digital curator of their belongings.

The concept of extracting values and ideas behind rather bland objects and digitizing them in a dataset is really the core behind this idea. I decided to make this dataset in the Office Excel work application, since I find the contrast of quantitative “hard” data and strictly organized values and functions which is usually associated with this application, and the more subjective measures of this dataset interesting. Rather than measuring objective values on the objects such as weight and colour, the goal of this dataset is to present values which in the absence of an attached report/explanation (which you are reading right now), would provoke a very different interpretation from person to person.

In Yuk Hui’s “On the Existence of Digital Objects”, Yuk writes, “The idea of the virtual ... Life is living within a virtual” (Hui, 2016, p. 47-48), which describes how the barrier of what is considered real and what is considered virtual, has receded into the background, and how an understanding of

what is real, is understood in the comparison of the virtual. I think this is relevant in my understanding of data curation and everything involved, but also in the context of the theme and my dataset and especially the last category.

An object like 108, the turntable, has been “substituted” and digitalized in many ways with virtual platforms that are able to play digitized recordings of music, like the popular music application/platform Spotify. This sort of technology in the digital world helps define the way we interpret the physical world. The turntable scores 6 on aesthetic, sentimental and POCD, and 1 on Practical and economical. I find this to be an interesting observation (although it is a subjective measure). It is the presence of modern music technology, that defines the subjective values of the turntable, thereby an interaction with the digital and physical world.

If music platforms weren't digitized, the turntable would probably score higher in practical value, since that would be one of the more practical ways of listening to music, and lower on sentimental value since it wouldn't be a “niche” or “antique” item.

Object nr. 117 which I mentioned before, are banknotes. They scored 6 on POCD since I find the ability of completely digitizing money, not just possible but also almost completed. Most people use credit cards and bank transactions as a way of paying, and furthermore take advantage of a digital credit/debt machinery. The central bank of the given currency controls the supply of money, and with the digital world of money, when increasing the supply, they simply “print” by increasing the number on a screen. New ways of digitized monetization like cryptocurrency increases in popularity for its decentralized systems, which increasingly substitutes the physical value/need for banknotes.

This clash between technological innovation and political challenges can be reflected in the book, *The Politics of Mass Digitization* by Nanna Bonde Thylstrup. In this book she writes, “In order to grasp the political stakes of mass digitization, therefore, we need to approach mass digitization projects not as a continuation of the existing politics of cultural memory, or as purely technical endeavours, but rather as emerging sociopolitical and sociotechnical phenomena that introduce new forms of cultural memory politics” (Thylstrup, 2018, p. 4). I find this to be reflective in my thought behind object 117, since the technical endeavours, being the digitalization of banks and decentralized capital, and the political and cultural memory, being the traditional methods of

transaction like cash is in and interesting undergoing digital evolution, and has been for some time. This process raises the question of stakeholders' cooperation and how a new "decreasingly" physical world is formed.

As mentioned, the dataset consists of 20 objects, with a lot of them unmentioned in this report. With the explanation of the basic principles of the categories, and the concept behind the dataset, I hope an appropriate interpretation will be evoked. A big part of this assignment has been the learnings process both practically and creatively, in building and digitizing a dataset. From this process, I have reflected upon the physical and digital interaction, which is a theme throughout my endeavours in this course, and therefore this report.

Assignment 2. Data Selfie

Social media like Facebook, Instagram and Snapchat are used by consumers to interact on a formally envisioned platform to digitize the modern world's social interactions. This digitized form of social interaction gets tracked and stored as data, giving every user on these platforms a sort of data coded history of their activity. On these platforms you can request a constellation of your personal account data and download it. The object of this assignment is to make a data selfie, picking personal data from our social media accounts, and interacting with this data in a certain manner, to represent ourselves. When looking for a platform to download personal data, the first challenge one might run into, is the waiting time, from the request of the data, to the received constellation.

I decided to go with Facebook, since you can get years of personal data almost instantaneously (depending on the timeframe and activity). Luckily for me, I am not the most active social media user, so the data isn't overwhelming in quantity. Or so I thought. I decided to download my personal data based on categories suggested by the platform, and on the entire account's timespan. The result, was a folder, containing the categories of my choosing, containing JSON-Files within them. Unfortunately, a large number of the folders I collected were trivial not only in the lack of information, but also in lack of visualizable data. One large challenge, when presenting a data selfie from a personal social media account, is to find the balance between relevant and

appropriate data. This reason amongst others, is why I would argue that a self-representation using this data, is challenged by its own limitations.

I have decided to use the data of my reactions on posts and comments. The file containing data about my reactions on posts and comments, contains the data; Actor, Timestamp, Reaction and Title. I have used a digital tool to illustrate the data in a way I find relevant, in portraying myself. I made a “beeswarm plot”, putting timestamp on the X-Axis and color-coding my reactions, with “LIKE” being the red, “ANGER” green and “HAHA” the blue one (Appendix 1). These are the only three different reactions I have used over the span of my Facebook account’s lifetime, which might in itself represent me in a way. Timestamp is sort of scattered over the Axis, since this measure of time requires a converter which I couldn’t bother. The timestamp is however chronically sorted, which is the only necessary part. The reactions are displayed as small dots on the plot sorted chronically from left to right, present day being the latter.

I find the dataset to show an interesting correlation by when I react to stuff, I am more likely to react to more stuff. This is represented in the clump like spread of the data. Only very few reactions are independently represented in the plot, most of them are stacked upon each other, which supports this finding.

It should be clear, that the color-coded representation, shows the lack of diversity from my reactions. I only very rarely use other reactions than “LIKE” (reactions other than like, were first implemented in 2016 (wikipedia)), which I find to also be a representation of me in a way.

Visualizing this data, made me able to more thoroughly analyse it. With the use of a “beeswarm plot” I was able to notice the aforementioned clump like patterns, and the lack of diversity. This helps an in depth understanding of the data, and more so, it helps me to share my thoughts and analysis with reference to the visual information. With reference to the plot, I would assume it might be seemingly easier to understand my thoughts and analysis of the data.

In Rob Kitchin’s *The Data Revolution*, he writes “In rhetorical terms, data are that which exists prior to argument or interpretation that converts them to facts, evidence and information” (Kitchen, 2014, p. 3). Even though I think this is a paraphrase from another source, I find this quote to be reminiscent with my process of this assignment. The data bear only objective value in its core existence, however when manipulated with representation and interpretation, it becomes

information. Much like my representation of my Facebook reactions. I could represent this data in a plethora of ways, however the decisions I have made, help skew the informational qualities of the data.

In the paper *Social media data archives in an API-driven world*, the Authors Amelia acker and Adam Kreisberg writes about data preservation archives. In the paper they write, “information policy, privacy, and law scholars have examined the functional sovereignty that these platforms now exert over society, democracy, and economies of scale by collecting personal data, providing access for third parties, and repurposing it for algorithms, personalization, and advertising technology (Pasquale 2016)” (Kreisberg, 2020, p. 106). This quote talks about how the archiving, extraction and manipulation of data, drives new markets and exerts itself over the social paradigm, affecting a multitude of the social sciences.

I think this is thought-provoking, and very well resonating with my whole experience of collecting personal data from a digitally archived platform. I don't know much about the terms and conditions for Facebook and personal data in terms of public access, however I find it interesting to conspire on a publicly accessible version of all my data, used for “algorithms, personalization, and advertising technology”. Finding data about my interactions on the Facebook platform, and portraying it in a visually pleasing manner, has not only let me to experience the accessibility of my data, but also frightened me of possible exploit.

I think this little experiment of extracting my personal data speaks volumes to the large scaled datafication of the digital world. Katrina Sluis writes, “The problem with [Facebook's Graph search] is that aggregation says more about us than we consciously know we are making available” (Sluis, 2017, p. 33). This data selfie “experiment” has been eye-opening for the degree of digitally available personal data and helps understand the necessity of current debates like with GDPR. Sluis further writes, “Tracking at all these levels demonstrates the extent to which the social network itself generates a parallel archive of movement recording the interactions of the networked itself, as a simultaneous—but exponentially bigger—living archive. This parallel archive may come to make correlations about ourselves about which we are not yet aware” (IBID). She mentions the exponentially scaling personal data tracking of social media networks, and how this “parallel archive”, which I suppose can be a parallel to the real/physical world, can help us

understand ourselves in ways we otherwise would not. This speaks volumes to the vastness of what data is capable of, and how this digital format interacts with our lives.

In context to the theme of interactions between the physical and digital, I think this self-portrayal from digital social media data, is an interesting addition. It can help expand the idea of understanding the physical world through the digital, and with this assignment, it shows this not with objects, but with myself. Obviously this isn't suggesting that this short illustration can portray a complete set of my personality traits, however I think with this rather "mild" or "bland" example using data from facebook interactions, that one can only imagine the scale and magnitude of this, if done otherwise with other measures. Thereby, likely supporting the argument or exploration of understanding the physical through the digital.

Assignment 3. Folksonomy

To ensure the most effective and fair group work, we met up in person and worked through the whole process together. Working in person creates a more casual environment where you are more likely to bounce ideas off of each other compared to having to write a message first and wait for reply. This way the assignment would be as communal as possible.

The first thing we did was discussing the pros and cons of various software that could be used to put our collection together. Calibre was the obvious first choice, but, as discussed in class, the programme had some issues when it came to more people working in it at the same time and sharing the live changes made in the content of the collection. Therefore, we all ended up agreeing to go with creating a Google Drive for our collection as we all had more experience using that software and felt more comfortable navigating it. It is also favourable that everyone with a link to the drive will have the exact same version of content. To us, Google Drive seems like a more intuitive collections application, which we already are used to using.

An overview of the collection is attached in the appendix, so there won't be any further reference to this appendix, but it might help to understand the whole assignment when reading it (Appendix 2). Furthermore, a link to the collection is also attached in the appendix (Appendix 3).

As the collection had to include sources from the course, we briefly recapped what themes we had discussed in class and ended up agreeing to do a collection on literature that had to do with AI. With AI as our overall category, Jeppe and Anders quickly began browsing additional material while Lasse and Johanne set up the AI Google Drive before joining Anders and Jeppe in their search on the Memory of the World library, Google Scholar, AU Library, and the class syllabus.

As we researched, we discussed what subcategories we could appoint under AI depending on what literature we found most interesting and relatable to class. From that we came up with the category 'machine learning' as in every- and anything about how AI works and learns; 'power' defined as the power that AI can give the user in relation to politics and control; 'creativity' defined as how can AI aid human creativity and what is AI creativity – does it exist?. Originally, we also had 'face recognition' as a subcategory, but we discussed it and concluded that the subject was already contained under 'machine learning'.

To start the collection out, we agreed that each subcategory had to hold at least five literary titles. When we had reached our goal, we collectively sat through each title and decided on what tags we would use and why by skimming through the individual material. We obviously had not read everything (some were whole new books), but we did our best to find the best representative tags by reading summaries, abstracts, and sometimes using the keywords already assigned by the author.

Unfortunately, Google Drive does not have a function to edit metadata or specifically add tags on the files, but you are able to add a description. A quick test proved that the description was a part of the search when using the search engine in the overall library on the drive. That made the description the obvious place for tags. We made the lower limit of at least three tags on each material to ensure most successful use. The upper limit was five tags to ensure focused search results.

We encountered a bit of trouble with the tag 'the dark side of AI' as the search engine would bring up any work that contained any word of the search (like 'the' and 'of') rendering the tag useless. However, Jeppe was quick to discover that if you replaced the spaces with '-', the search would only bring up instances where the five words appear in that exact order – making the tag not useless

after all. We also tried to use some of the same tags where possible, so more material of similar themes would turn up. The most used tags being 'validation', 'pattern recognition', 'politics', 'the dark side of AI', 'BACON', and 'deep learning'.

As the collection will perhaps expand, it will most likely be favourable to add subcategories under subcategories to make it more focused and make it easier to find the exact AI topic you are interested in. As that happens a higher amount of tags will be needed to specify even better.

Assignment 4. Curatorial Statement

Premise

Curating data signifies the curation, manipulation and interpretation of data, and in a world of continually datafied and digitized realities, one can explore numerous aspects and abstractions of this matter.

I've decided to work on the interaction between the physical and the digital world, specifically looking at how data and digitization is playing a larger role in our lives. I've discovered how an interpretation of our world, might just necessarily include a realization of the digital relevance and magnitude for understanding this interpretation. I'm slowly coming to the conclusion that, more and more tangible things might need a digital context, to be fully understood, hence - *Seeing the world, through digital glasses*.

More specifically, what is meant by this is, that we are in a more and more digitized world than ever, and most of our everyday lives are dependent on digital systems. For an example, with a more traditional interpretation of curation and curator, one might think of a collector/inheritor and presenter of materialistic and perhaps artistic objects, likely showcased in a museum. Being a curator was not as inclusive as it arguably is today, with almost everyone having the ability to perform curation, and therefore arguably being a curator (Tyzlik-Carver, 2017). This is illustrated with assignment 1, with the digital collection of materialistic objects. With only few tools and properties anyone can curate. It is partly with the advancement of the digital world, that the

physical world is shaped. Like with the contrast to traditional curators, the digital world has allowed a completely new interpretation and use of curation and curators. Seeing the world through digital glasses, is meant as, looking at things with a digital context or filter in mind, to further understand objects and ideas. This is also an especially interesting phenomena, since it is more and more current. This is the curatorial statement, that summarizes the theme of this report.

Assignment approaches

In our first assignment, we digitized analogue objects, and elaborated on the creative and explorative process of this. I chose to conceptualize an idea of subjectivity, by collecting analogue objects, digitize them with pictures, and addressing subjective values to them. It was an idea of understanding objects, by valuing and comparing them in a digital collection. The collection consists of 20 rather bland household items, with values like sentimental, practical and economical (among others), measured on a scale of 1-6. The dataset was in effect, a tool to observe not necessarily obvious traits of these objects and explore Yuk Hui's idea of using the virtual to understand the real. Through this process I realized the significance of how the digital world shape our interpretation of the physical world and vice versa.

The second assignment had the aim to make a data selfie, to illustrate ourselves through data from social media. In this assignment I used my own personal Facebook data (Facebook reactions on posts and comments) and visualized it in a coordinate like system. In this assignment I discuss the accessibility, function and scalability of a digital world, with the example of me as a digital figure. I reference Amelia Acker and Adam Kreisberg's paper talking about information policy, privacy and law, and the power digital platforms have over the world and all aspects of society. I also reference Katrina Sluis, where she elaborates on the vastness of digital databases and archives on social media. I use these quotes to elaborate the scalability of the data selfie as an illustrative example. I discuss how the use of data in a commercial sense is an ongoing challenge, and how this helps us uncover another part of the digital world.

The third assignment, the folksonomy, was an assignment with the aim to collaboratively build a textual collection, using software of our choosing. We made a collection on Google Drive, and discussed a few subjects in context to the curriculum. We went with a collection of texts within the subject "AI" (Artificial intelligence). The collection consists of several texts with manually added metadata/tags like "politics", and "the dark side of AI". Furthermore, these texts were organized in subcategories like Machine learning and creativity. With further work on this collection, perhaps some inspiration could be provided from Katharina Weinstock, where she writes about implementation of collections, used as forms of art and conceptualization of objects' aesthetics. We have had a rather functional approach to this assignment, but with a twist of artistic swerve and creativity, maybe the assignment might advance to another level.

Significance

The context and magnitude of data, as mentioned in my work, has sparked a new understanding and curiosity of data and the digital world. I've explored the significance in the way we understand the world, by looking at it through "digital glasses". The significance not only in the mere understanding of a physical world, but the mutual interaction with the digital world, and how, in order to understand our physical world, we must realize the digital significance of most or any observations.

Conclusion

This report has illustrated ways of collecting, digitizing, archiving, analyzing and presenting data, and thereby unfolding the concept of curating data.

Inspired by the curriculum, the report references some of the articles and publications, with relevance to the theme and the discussion.

The theme has been the exploration of the interaction between the physical and digital world. This theme is illustrated and discussed throughout the assignments and summed in the curatorial statement "seeing the world through digital glasses", as a way of illustrating thoughts about the

theme and course. Ways of looking into this theme has been illustrated through data curating. Data curation has an intrinsic value to our society, and is integrated on so many levels, that anyone might need to develop a set of data curating skills and knowledge, to better adapt to this ever-going data revolution. Seeing the world through digital glasses, is a way of expressing the necessity of developing these skills, to see and understand the impact and vastness of the digital world.

Bibliography

Acker, Amelia & Kreisberg, Adam. 2020. Social media data archives in an API-driven world

Kitchin, Rob. 2014. "Conceptualising Data." In *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*

Sluis, Katrina. 2017. "Accumulate, Aggregate, Destroy. Database Fever and the Archival Web.

Thylstrup, Nanna Bonde. 2018. *The Politics of Mass Digitization*

Tyzlik-Carver, Magda. 2017. "| Curator | Curating | the Curatorial | Not-Just-Art Curating." Springerin, no. *The Post-Curatorial Turn.*

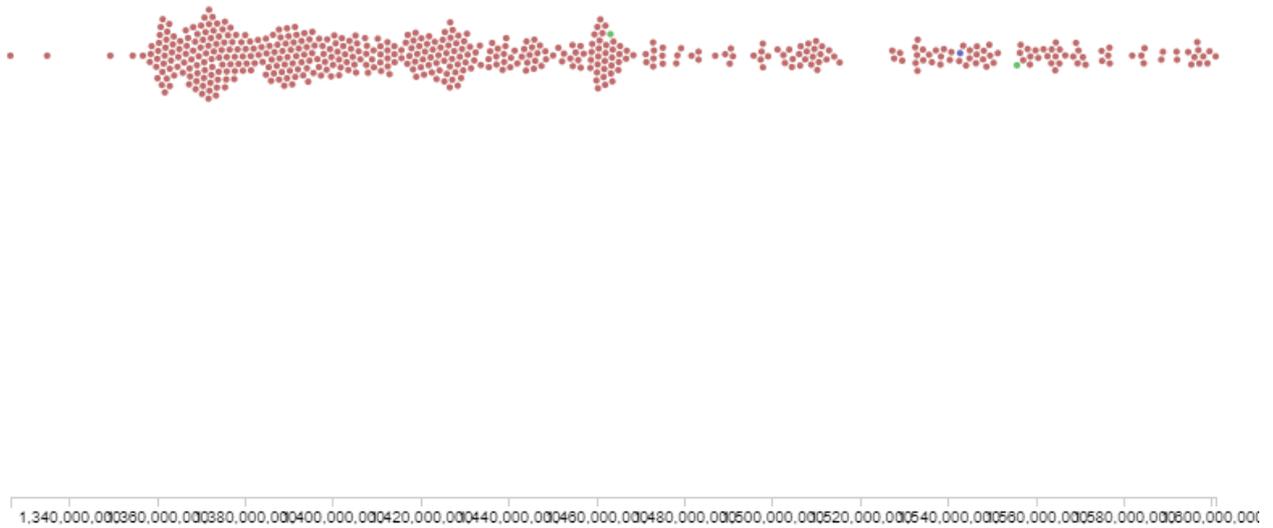
Weinstock, Katharina. 2020. *Rearranging the World*

Wikipedia, facebook like button

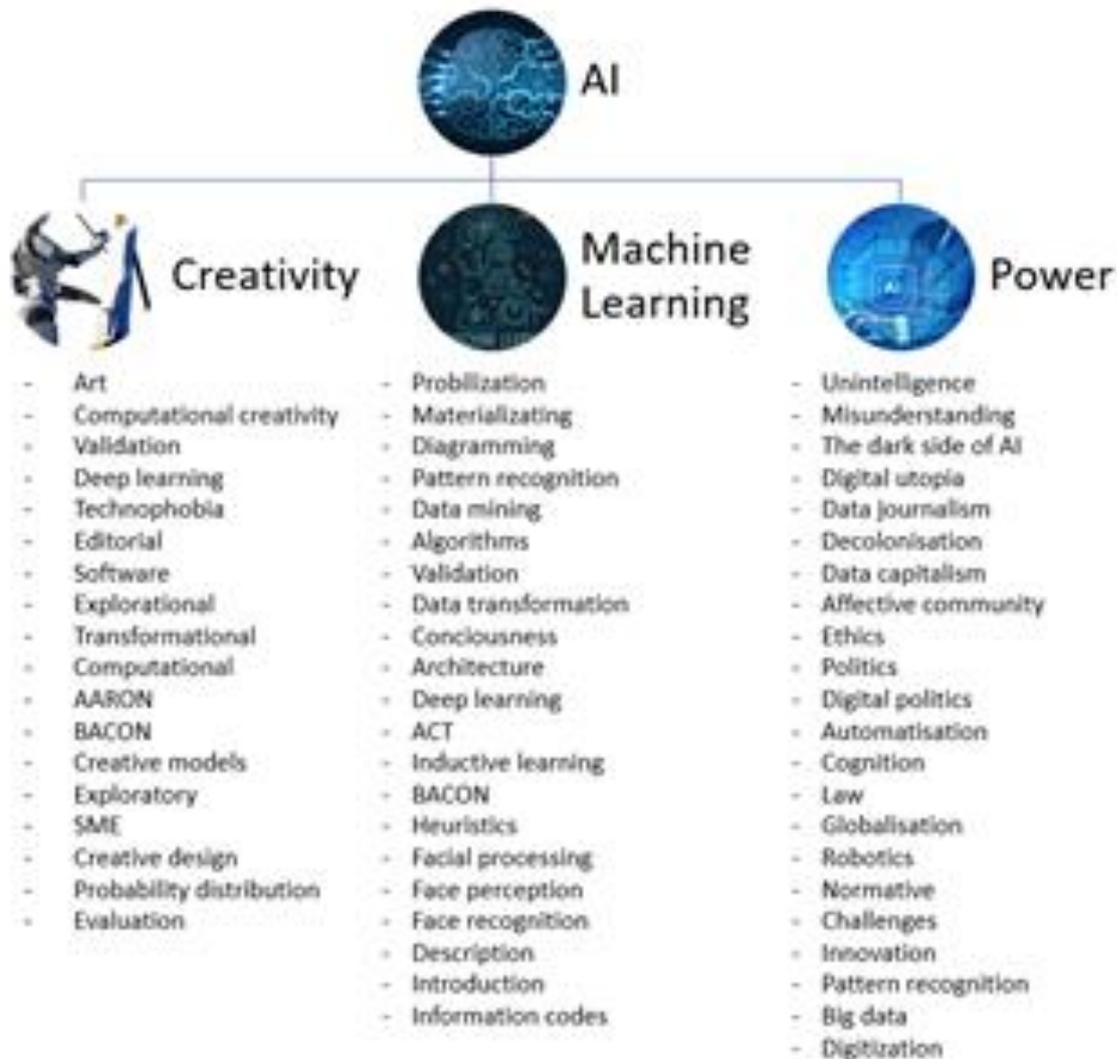
https://en.wikipedia.org/wiki/Facebook_like_button

Appendix

Appendix 1.



Appendix 2.



Appendix 3.

https://drive.google.com/drive/folders/1YmOC_dE_3AwwhrrOGiTljJj_Herqcfcd?usp=sharing

Appendix 4.

<https://jeppekp.hotglue.me/?start/>