Digital Textual Criticism: The New Literacy for Humanities Scholars

Fu-Shih Lin (Distinguished Research Fellow, Institute of History and Philology, Academia Sinica)

Contents

I. Introduction: Beginning from the "Data School"
II. The Pitfalls of "Digitized Data"
    1. The Unfaithful "Copy"
    2. Selective Memory
    3. The Lure of Keywords
III. The Problems of "Born-Digital Data"
    1. Circulating Texts
    2. Plagiaristic Creation
    3. Fake Facts
    4. The Dangers of Quantification
IV. Opportunity or Crisis?
    1. Who Controls the Digital World?
    2. Pay Attention to "Crowdsourcing"
    3. Be Careful of Robots
    4. Consider "Artificial Intelligence"
V. Conclusion: Only Literacy, Not Skill

Abstract

This paper investigates the issue of the "paradigm shift" in digital era humanities research, mainly from the perspective of "textual criticism." All humanities research, whether from the traditional or digital era, must be based on a foundation of "data" and require "textual criticism." But the problems encountered with the forms and applications of data, as well the as the methods and tools of "textual criticism," are constantly changing. Thus, our "model" for digital era humanities research must also necessarily be constantly shifting.

Consequently, in this digital era where "data is king" and "robots" take a dominant role, humanities scholars truly need to study both digital technology and humanities knowledge, combining both humanistic philosophy and "computational thinking." However, this does not guarantee that we can then soar unimpeded through the cosmos of the digital world. After all, these merely describe literacy, not skill. This new type of literacy may be called "digital textual criticism."

I. Introduction: Beginning from the "Data School"

In 1928, the founder of the Institute of History and Philogy—Fu Ssu-Nien — proposed three "whens" and three "missions" in his foreword, *The Aims of Founding the Institute of History and Philology*, for the Bulletin of the Institute of History and Philology, Academia Sinica (IHP). The three "whens" are "improve when one can directly research materials," "improve when a certain discipline can expand on one's research," and "improve when a certain discipline can expand on the tools applied during research" whereas the three "missions" refer to "preserve the *Gu Tinglin Poems* legacy," "expand upon research materials," and "expand upon research tools."

His proposition, simply put, is to "study research materials directly" and expand upon one's research materials and tools. Fu Ssu-nien also concluded that "we are not book readers, but merely those who scour the heavens and earth to look for things." His words were an attempt to break away from traditional disciplines of a book-oriented approach and to seek "things," not just words but also artifacts, images, spoken materials, and sounds.

Revolutionary in early 20th century China, Fu Ssu-nien's proposition gave way for Chinese literature and history researchers to pave a new path of research. As the proposition he set forth was bold and distinct, some began referring to the style of study represented by the Institute of History and Philogy as the "Data School." The label coined was part compliment and part criticism, praising the "Data School" as simple, robust, and broad while criticizing it as dry, boring, and antiquated.

Yet, between the late 20th century to the early 21st, as advanced countries began devoting themselves to the development of "Digital Humanities," we realized that the 90-year-old institute had already started its journey into a new and fashionable field (1984). This may seem bizarre but it is certainly no coincidence.

Fu Ssu-nien's proposition, in terms of digital humanities, was to obtain "Big Data" and utilize various tools and methods for big data analytics. The strong emphasis that the IHP had placed on historical sources was completely in line with the data-oriented approach (digitized data and born-digital data of any form, type, and medium) of digital humanities. "Seeking for things" refers to information retrieval, data warehousing, and data mining. In the world of digital humanities, there is a heavier reliance on machine reading as opposed to "human reading" when searching for and deciphering materials. Books continue to be deconstructed and dissolved into various "texts" and "book readers" are gradually becoming extinct. Juxtaposing the two eras, we can see that Fu's words have inevitably turned into a metaphorical prophecy that has left space for the inclusion and cultivation of digital humanities in the old halls of academia that is the IHP.

As we enter into the digital age, however, digital humanities continue to occupy more spaces within the field of humanities and we seem to be upon the cusp of a paradigm shift

in the field of humanities research. As such, we cannot help but ask, has the paradigm that the IHP represents become antiquated? Will it be replaced? After all, the "*Gu Tinglin Poem* legacy" that Fu Ssu-nien had hoped to preserve was a tradition of evidential scholarship established by Gu Tinglin (1613-1682) and Yan Ruoqu (1636-1704) in the early Qing dynasty over hundreds of years ago, an age where texts and books reigned supreme. The problem is too expansive, however, and requires the test of time. As such, we should perhaps start from "evidence" and consider whether or not "evidencing" is required for humanities research in the digital age or what challenges existing evidential research will face as we ponder the question of the paradigm shift.

"Evidencing" is essentially "identifying what is true and what is false." You first have to identify truths and falsehoods to then build upon it and conclude whether a fact is true or false. In that sense, empirical research will always require evidence. Regardless of the form or medium, materials will usually include elements that can be examined, such as the creator, the time and location of creation, the storage location, method of transmission, etc. But the materials one must handle in the digital age or in a digital scenario encompasses digitalized data and born-digital data. Whether or not traditional methods and skills of evidencing remain appropriate, therefore, indeed requires reevaluation and rethinking.

II. The Pitfalls of "Digitized Data"

Firstly, we've found that one must be careful of the pitfalls hereinbelow when employing or, rather, reconstructing facts and discovering truths from digitalized data of digital formats converted from traditional literature in their research.

1. The Unfaithful "Copy"

Bibliology and textual criticism have always been an integral part of traditional evidential scholarship, because in the "age of words," manuscripts and printed copies, whether transcribed, block printed, or typographically printed, were often "unfaithful" to the original. Unfaithfulness in the copies is sometimes the result of commercial or technical concerns, interference from political or religious powers, edits or adjustments by later generations because of different understandings or ideologies, and even simple errors or mistakes while transcribing or printing.

Such problems also exist in the conversion of traditional literature into digital data. Take the IHP's Scripta Sinica Database for example, its predecessor is the Automation of Historical Records Project (1984-1990) that began in 1984. The primary purpose of the project was to convert full texts of Chinese historical records into digital texts that can be presented, read, and retrieved on the computer. The first year saw the complete digitalization of pre-Song official history, the *Treatise on Food and Money*, and, subsequently,

entry for the full *Twenty-Five Histories*. At the time (1985), I was a graduate student majoring in history at National Taiwan University and, by chance, worked as a part-time assistant for this project, responsible for proofreading the *Records of the Grand Historian* and *Book of Han* while regularly meeting with IT experts and engineers to discuss how the database would be constructed. At the time, humanities scholars at the IHP insisted that the electronic historical records must be "faithful to the original." The seemingly reasonable and simple request was, now thinking back, a disaster and an impossible mission. The initial phase for Chinese computerization had only been completed around 1980 and there were far from enough interchange codes and digital Chinese characters to faithfully present all the characters used by the historical records. There were, as a result, many "missing words." The tables (*biao*) in the *Records of the Grand Historian* were also dense (ten chapters) which the technology at the time was unable to render. In other words, the electronic version of the *Records of the Grand Historian* deviated greatly from the paperback version.

The issue with missing words and table formats was eventually resolved, but the electronic version was still far from perfect. After the Historical Records Database was built (1990), the IHP immediately began to build a full text database for Confucian classics, historical records, philosophical writings, and miscellaneous works outside of the *Twenty-Five Histories* and in 1993, the database was renamed the Scripta Sinica Database. The toughest challenge remained: how can you create a copy that is "faithful" to the original? For example, the *Daozang* includes many symbols, images, pictures, and tables that are standalone or mixed in with words. How to comprehensively and accurately render them in the computer for retrieval and access is still something we need to improve on.

To save time and labor costs, Optical Character Recognition (OCR) was used to input some of the historical records. Despite OCR's years of history and mature technology, the 90% or even >95% recognition rate, for a database catering to professional research, was not acceptable. Proofreading remains an incredibly valuable industry.

The digitalization of colors, sounds, and audiovisual materials will, I believe, also result in unfaithful copies. But this is not where my expertise lies and I shall not discuss this any further in this paper. In any case, without robust digital preservation strategies, technologies, and management, any form of digital data can be damaged, changed, distorted and even lost or inaccessible which can also be considered as damages or destruction to the original.

2. Selective Memory

Memory is at the core of history, and many have long believed in the existence of what we call "shared memory" or "collective memory" and its intertwinement with ethnic, societal, cultural, and national identity. Many nations have, therefore, at one point and in various forms, attempted to build a collective memory (history). In ancient China, the

written Official History, national history, chorography, genealogy books, Buddhist and Taoist canons, medical prescriptions (e.g. *Taiping Royal Prescriptions*, *General Records of Holy Universal* Relief, and *Prescriptions of Universal Relief*), leishu (category books) (e.g. *Yiwen Leiju*, *Readings of the Taiping Era*, *Cefu Yuangui*, *Imperial Encyclopaedia*), and Collectaneas (e.g. the *Imperial Collection of Four*) can all be considered efforts to build collective memories. The National Digital Archives Plan, Taiwan Digital Archives Program, and Taiwan Cultural Memory Bank initiated by the Taiwan government starting from 2000 are also efforts with similar aims.

The instigators and primary participants of these cultural works are usually the elite minority or even members of the ruling class. Their jobs are a form of selective memory where they include politically correct materials that fit squarely with their own beliefs, interests, values, and taste into the collective memory. Materials that deviate are ignored, abandoned or deleted. We can refer to the apocrypha and pseudepigraphs of religious classics, some are canonized (included into the Buddhist or Taoist Canon) while others are excluded (left out of the Buddhist or Taoist Canon). The *Imperial Collection of Four* too has Canonized Books (fully canonized), Catalogued Books (where the book catalogs are preserved instead of the full text), Excluded Books (books that are not catalogued or canonized), and Destroyed Books (books that were not canonized but rather destroyed and banned for publication and archiving). Banned books are a collective memory and a collective experience shared by many in both China and Taiwan.

Our collective memory has, therefore, always been compiled by the minority, the only difference is that their tactics were sometimes brutal and sometimes refined. Our history has always been written and broadcasted by those in power, the only difference is that what they said were sometimes moving and sometimes alarming.

How can we recover the forgotten and deleted memories? How can we give voice to those silenced? How can we get a hold of all the materials and reconstruct a truly comprehensive history? These are all rigorous challenges that remain in the digital age. The digitalized data may not be faithful to the original, but as of now, the original is still a highly selective memory.

3. The Lure of Keywords

The advent of digitalized databases and convenient retrieval systems was a source of elation for some humanities scholars, some of whom seemed to believe that as long as they knew a few keywords, they would be able to find any information they needed. They gradually left behind the reference books (i.e. indexes and dictionaries) that they had once relied on and now rarely take reading notes or use index cards. As the quantity of digitalized data has grown, humanities scholars are often left to extensive reading, flipping through, and searching for information, leaving less time for intensive reading and familiarizing

themselves with the literature. The change in habit may seem insignificant but, in a generation or two, will most likely bring about revolutionary changes to humanities research. The key to this change lies in the lure of the keywords.

Keywords are indeed quite useful and it can save scholars a lot of time and energy, delivering researchers a handful of new materials, research topics or even entry into a completely new field in a short amount of time. But the speed is also where the danger lies. Over-reliance on keywords for searches will also result in missing out on important information that results in poor judgement or time wasted from the large amount of junk materials.

Such unfortunate conditions are primarily the result of the abundance of variant and borrowed Chinese characters in traditional Chinese literature or characters with missing strokes and alternative characters that were created due to naming taboos in different generations. Information will be overlooked when searching with simple keywords without sufficient preparation, a synonym chart, or an authority file. It is also very common in Chinese for a character (a phrase) to hold multiple definitions, a thing (a place or person) to have multiple names, and a name to refer to multiple things (places and persons). One would usually have to rely on the text type and context to determine the character's meaning. As such, using the results of keyword searches into word frequency analysis, relation network analysis, and visualized presentations will largely cut into the research's accuracy.

III. The Problems of "Born-Digital Data"

Born-digital data, to a certain extent, also has the aforementioned pitfalls and even more complex issues.

1. Circulating Texts

In ancient China, the spread of information, creation of knowledge, and expression of emotions experienced a long process of transforming from the spoken form to the written form. The earliest texts were written, read, and circulated in the forms of articles, chapters, and scrolls where the person passing on the information could usually add, delete, and change the information and so the difference between the original and the copy could be quite surprising. It was only through the continuous efforts to organize and edit such information between the Spring and Autumn Period (8th Century BCE) and the end of the Han Dynasty (3rd Century BC) that more systematic books and texts were able to form and remain fixed, which is why contemporary scholars often remark that, "this is a work of many men and many ages" when they review pre-Qin historical records. In other words, it is impossible to find the exact author or exact time of creation for ancient books. Even

after the Han dynasty, because of information dissemination, transcriptions, edits, annotations, forgeries, etc., the author and year of creation still poses a large challenge which is why critical analysis of literature, bibliography, and bibliology have always remained such an integral part of humanities research of traditional China.

It appears that born-digital data won't have the same problem because producing an article, a picture, a song or a video automatically generates multiple records, metadata, and basic information (i.e. time, date, spec, etc.) in the system used. Editing records can also be saved automatically. But when born-digital data is being produced through crowdsourcing or when the information is made public on the internet or social media, all or part of the content can now be easily changed, edited, and disseminated, even without malicious intent. The originally fixed text will now duplicate, evolve, and spread like a virus, becoming a circulating text. Unless there is a strong control mechanism and technology, defined user regulations, and self-discipline and compliance from the audience, born-digital data will be, soon after, also considered "a work of many men and many ages."

2. Plagiaristic Creation

Since ancient times, Chinese classic texts have been extensively annotated by scholars across generations, who put layers and layers of notes and commentaries on top of the original texts. Some annotations follow the principle of "the text interprets itself" (using verses in the original texts to explain other verses in the same text), some cite other texts, and others furnish personal commentaries. Most of the time, annotators acknowledged the sources they were citing, but the structured layers consisting of original texts, notes, and commentaries could fall into disorganization after generations of "sorting and compiling," manuscript copying, block printing, and movable type printing, making it difficult for modern scholars to identify the real authors of certain parts of the text. Nevertheless, this is not intentional and thus cannot be considered plagiarism.

The same phenomenon is also present in the classic texts of traditional Chinese medicine. Like the classics of Confucianism and other schools of thought in the pre-Qin era, medical classics (e.g. *Huangdi Neijing*) are also embedded in a tradition of annotation. However, texts about medical prescriptions and herbal medicine are often "copied from copies," sources of citations are often not acknowledged, and personal opinions are often added to copies. In addition, medical texts are often compiled and edited as part of a Leishu. As a result, they often contain works from different authors across different generations. The same phenomenon is also seen in religious texts, especially in Taoist Canons, wherein many descriptions of rituals, recipes for alchemy, and instructions for magics were often passed down secretly from masters to apprentices, sometimes in the form of mnemonic phrases and secret texts, some of which were even claimed to have been passed down by saints and immortals. Editors of such texts were not intentionally

committing plagiarism, but such practices make it difficult for modern scholars to precisely identify by whom and when the texts were written.

In addition, works of two other categories of Chinese classic texts, *Leishu* (similar to encyclopedia in the West) and *Biji* (notebooks), often fail to acknowledge the original sources of citations, including their original authors and the eras in which they were written. Large volumes of government commissioned *Leishu* typically acknowledge the sources clearly and accurately, since their authors are more highly educated and their writing funded by the government. In contrast, many privately written *Leishu* fail to furnish the sources of citations, intentionally or unintentionally plagiarizing works by predecessors, perhaps due to commercial interests, limited knowledge, or the need to show off talent. Such phenomenon is most commonly seen in *Zhiguai* (tales of the strange), fictions, and *Biji* that were written with intention to stress how odd the stories are or to show off how knowledgeable the authors were

Unfortunately, such unintentional plagiarism or "plagiaristic creation" are also present in born-digital data. In the world of Internet and social media, users often take the liberty to "copy" contents of others (in full or in part) by "sharing" posts by others without a adding any notation or commentary. This practice, in fact, essentially creates another text. In some cases, controversial reposts generate more discussions than the original posts per se. If we compile the discussions under the original posts and those under the reposts and consider them to be a volume of text, it is extremely difficult, then, to distinguish the original authors from theplagiarizers. The complexity even goes beyond the characterization of "by different authors across different generations." Furthermore, Wikipedia and other similar texts presents us with another kind of challenge in terms of distinguishing the original from the plagiarized and correct information from wrong information.

3. Fake Facts

From late Qing to early Republic of China, Chinese scholar communities were widely influenced by a trend of thought known as "doubting antiquity." All things ancient were suspected to be possibly fake, leading to a re-examination of every ancient text, ancient artefact, ancient figure, and ancient history. After a series of "cracking down on the fake" movements, some began to believe that many parts of our recorded histories are in fact mythologies or even downright fabrications. Although the "doubting antiquity" trend of thought only remained mainstream for a short period of time, its legacy of "deconstructing" had a lasting effect on historiography, leading to skepticism toward the existence of the Three Sovereigns and Five Emperors and inciting a movement of reconstructing ancient Chinese history that continues to the present day. It became a common sense that historical materials and "historical facts" can in fact be fabricated,

resulting in wariness of historical accounts as people became less gullible.

In born-digital data, fabrication is also rampant and even more difficult to tell apart. In online media, what appears to be news stories may in fact be advertisements in disguise. Some media outlets lay down sensational "fake titles" as clickbaits to lure readers into accessing news stories that have nothing to do with their titles. Even worse, some users produce and disseminate fake news out of hatred or for political, religious, and commercial reasons. Social media enabled users to form echo chambers (online communities consisting of people with similar values), where misinformation is easily believed and become widely accepted facts.

While these behaviors may seem to be just people fabricating information, historical facts constructed based on such information are indeed "fake facts." In a future where publishing will only become freer, faster, and more diverse, it is doubtful that the peer review mechanism currently employed by the academia can effectively offer the fact checking needed.

4. The Dangers of Quantification

Chinese governments across dynasties have left us with copious amount of official statistical records. However, scholars have been extremely careful when it comes to using these statistical figures such as population (number of households), land area, commodity price, or losses due to disasters, because the processes employed by ancient governments that produced those statistics, including surveying, recording, compiling, and reporting, were susceptible to errors and inaccuracies due to technical limits, legal limits, or personal interests of the surveyor. Therefore, "quantified historiography" has traditionally not received significant attention by scholars of Chinese history.

In the future, however, we might have more opportunity to employ such statistics, since the digital age is built upon quantification and computing, every data set contains metadata (e.g. time period, resolution, word count, size, area, and volume) that can be accurately measured and described, and the data per se can be divided into smaller units and labeled for counting, analysis, and connecting to other data sets.

In fact, many current research in the field of digital humanities are quantitative in nature, making use of tools such as word frequency, word cloud, social relation network, and visualization. However, inaccuracies and errors may arise if researchers are not scrupulous enough. For example, while it is true that the word "laugh" (*xiao*) is widely associated with positive emotions, and that perhaps the word frequency of "laugh" can be used to identify texts, authors, and eras with more positivity, laughing was oftentimes considered by traditional Chinese medicine a symptom of diseases, or even a disease itself (*xiaoji*, laughing disease). Research that fail to distinguish the two different usages will become a laughing stock.

To cite another example, *dahuang* (rhubarb) is a herb widely used in traditional Chinese medicine, and like all herbs, it went through a process of discovery, recognition, diffusion, gaining popularity, and falling out of popularity. Therefore, one core field of research in pharmaceutical history is the reconstruction of the process that a class of herb went through. The solution seems straightforward: use *dahuang* as a keyword to conduct word frequency count in a large scale Chinese corpus. Complicating the issue, however, is the fact that in addition to being a name of a herbal medicine, *dahuang* also refers to places, banners, and other objects, and the herb *dahuang* also goes by its nickname, *jiangjun* (general). It is therefore difficult to search for the herb *dahuang* accurately in corpora other than those about medicine. By the same token, one should be careful when trying to search for the military rank *jiangjun* (general) as it can be used to refer to the herbal medicine.

Several other examples also help illustrate the challenges faced by researchers. *Zhuyou*, a kind of witchcraft therapy in traditional Chinese medicine, is first mentioned in *Huangdi Neijing*. The origin of *Zhuyou* dates back to ancient times, and it was officially recognized as one of the subjects in medicine during 6th to 15th century (also referred to as *Zhoujin ke, Shujin ke, or Zhuyou ke*). Even after *Zhouyou* was officially abolished as a subject, *Zhuyou* practitioners had still been active in China up until the twentieth century. If we conduct a keyword search with *Zhouyou* in digitized version of *Taishō Tripitaka* and *Daozang*, we will find no mention of the word at all in *Taishō Tripiṭaka* and only six occurrences in *Daozang*, four of which come from *Huangdi Neijing,* one of Daozang's collections. It would be a total mistake, however, to conclude that Buddhists and Taoists in China did not practice or ignored *Zhouyou*. Although Buddhist and Taoist texts rarely, if at all, use the word *Zhouyou*, they still contain a considerable amount of descriptions of witchcraft therapies. Taoism, in particular, valued witchcraft therapies as much as the official medical community did.

These three examples illustrate the problems with digitized data and semantics. However, even more problems may arise as a result of human manipulation and intentional distortion. Today, some of the buzz words in the world of Internet include clicks, views, fans, online popularity, likes, search rank, and best-sellers. Cautious users should be skeptical of those numbers: how were they generated? Are they accurate? Have they been subject to human distortion? However, we still need to rely on these quantified data when studying modern trends of thought, popular culture, consumer behavior, economic cycle, social panic, crowd psychology, online behavior, and the social influence and social image of public figures. Therefore, quantification, albeit dangerous, is a challenge we ought to take on.

IV. Opportunity or Crisis?

The dangers and problems brought by digital data as mentioned above sometimes

exist independently of one another but at other times affect one another. While some of the problems are remnants of and extensions to those that have already existed before the digital era, others are new problems that emerged in the digital era. Part of these problems can be addressed with traditional approaches of textual criticism, while others have to be tackled with new digital technologies and mindsets. However, using any new tool comes with a price.

1. Who Controls the Digital World?

When the Internet emerged, thereused to be a half serious, half playful saying: "There is no ID (identity) without IP (internet protocol address)." Today, this saying has not received much attention, but with the development of digital technologies and the Internet, we will gradually notice that what does not exist on the Internet, be them individuals, groups, things, or histories, will only fall into oblivion, as if they do not exist at all. Therefore, the easiest way to deprive people of their rights is to deny their access to IP addresses, since losing IP addresses is tantamount to losing identities, messages, works, and friends. Unplugged, we can barely go about our daily lives (engaging in activities like seeing a doctor or shopping).

Similarly, in the future, humanities scholars will not be able to survive outside the digital context, in much the same way fish cannot survive without water. Without the Internet (or other communication systems or cloud systems), we lose access to research data and research tools necessary for our work. However, we should ask certain important questions: Who gets to decide whether certain data can or should be digitalized? Who gets to decide whether certain data can or should exist on the Internet and become accessible? Who gets to decide whether certain data should disappear from the Internet? Who gets to decide whether certain people can have access to the data? Who gets to decide how people should access the data?

In the past, the ownership and use of data were dominated by rulers, social elites, religious authorities, and big merchants. Today, this still seems to be true, but there is now greater emphasis on so-called market mechanisms and wisdom of the crowd. In other words, the tastes, viewpoints, behavioral patterns and actions may play a role in shaping the development of the digital world.

However, is this really the case? The answer may not be as simple as it seems.

2. Pay Attention to "Crowdsourcing"

In the past, humanities scholars tended to work independently (or even in silos), conducting research, collecting data, organizing data, analyzing data, writing paper and publishing paper all on their own. Nowadays, however, globalized Internet makes

crowdsourcing possible, allowing scholars across countries or regions to collaborate on data collection, organization, and analysis and engage in discussions. In addition, collaborative writing is made possible by online collaboration tools, which can also be used for online publishing. Furthermore, data owners (such as libraries, museums, archives, and art museums) sometimes even grant public access to their digital collections or digital spaces, allowing researchers and audience to conduct research, curate exhibitions, and upload their own collections (data). These measures democratize the construction and diffusion of knowledge by enabling the crowd to play a larger role in shaping these processes.

Against this backdrop, the future may seem like brave new world for humanities scholars. Nevertheless, "collaboration" and "crowd" are often synonymous with inconsistent qualities. In the field of humanities, collaborative studies may enjoy advantages such as large sample numbers and higher representativeness, but it is also highly probable that they show compromised quality or offer mediocre viewpoints. Furthermore, as greed is human nature, a collaborative project may see more free riders than real contributors. It is difficult to ensure that every member in a team makes even contribution. Also, in order to accommodate all the participants, a collaborative project may involve more democratic processes, compromise, and the need for standardization and consistency, measures that inevitably undermine the unique perspectives and value judgements of outstanding humanities scholars as well as the diversity of academic communities.

Even more noteworthy is the fact that our team may not even entirely consist of humans!

3. Be Careful of Robots

The defeat of world chess champion by IBM's supercomputer Deep Blue in 1977 and the rise of Google DeepMind's AlphaGo as the best Go player in the world in 2017 after ripping through top human players led to a sober realization that humans had been surpassed—by something they had created. This new creation can be generally referred to as "robots." Although not all robots are smarter than humans, they can now substitute humans in combat, sabotage operations, scams, criminal activities, rescue missions, labor, production, and social occasions. They can also provide services pertaining to medical treatment, entertainment, information, counseling, and learning.

In fact, every computer and every smartphone can be considered a robot. What's more, every device, software, and application can also be regarded as a robot. Therefore, robots naturally become a great tool and partner for humanities scholars. In fact, robots currently play almost indispensable roles in the production, storage, management, demonstration, and diffusion of knowledge in humanities. However, robots present us with challenges. We need to learn the languages of robots to give commands and process

information. We need to understand and adapt to their functions and behavioral patterns to become their true masters. Furthermore, we are often times not able to challenge or reject the results produced by robots. If they break down or malfunction, we will be faced with disaster. In other words, in the digital era, humanities scholars and robots are masters and servants/sovereigns and subjects to each other!

Robots can be seen everywhere in our daily life. While some of them are tangible, physical, visible, and even humanoid, some are invisible, hiding or existing only in the cyberspace. In the future, we may be asking what robots can't do rather than what robots can do. A world where human and robots co-exist is gradually taking shape. Consequently, future generations of humanities scholars will have to study behaviors and patterns of both humans and robots. The digital civilization of the future will be co-created by humans and robots. Here, one may be worried about the possibility of robots taking over the world. While such pessimism is unwarranted for now, we may need to consult an AI robot as for what will actually happen in the future.

4. Consider "Artificial Intelligence"

When the term artificial intelligence first appeared, it was sometimes ridiculed as "artisan intelligence." Today, however, AI has become high on the agenda of every industry. Those with and without wisdom, with and without wealth, with and without powers, with and without information access are gradually aware of the fact that artificial intelligence is almost synonymous with wisdom, wealth, power, and information. The aforementioned AlphaGo exemplifies what artificial intelligence is: a robot with the ability to perceive, learn, think, reason, judge, and create.

One day, "things" with artificial intelligence may even become "gods" that humans worship and depend on! At the very least, they will start to dominate and influence the behaviors, emotions, and fates of many. Currently, artificial intelligence is used to analyze big data and to predict the trends and patterns of diseases, weather, production, politics, economy and market. Some base their responses on predictions made by AI, and some even authorize AI to make decisions and take actions on their behalf (such as trading shares).

Then, is it a good thing to have a god-like tool at our disposal ? For better or worse, artificial intelligence and big data will certainly bring about drastic changes to the field of humanities. In the future, we will need to process a colossal amount of data, so it is necessary to rely robots with extraordinary intelligence and computing capabilities. There will be sweeping changes to the methods we use to analyze and process data, the perspectives through which we observe and describe human behaviors and social phenomena, and the ways and tools whereby we present our findings.

We can boldly predict that humanities studies in the future will focus on the correlation, rather than cause and effect relation, between phenomena, that predicting the future will become more important than understanding the past, and that distant reading of data will replace the close reading. The charm of textual narrations will fade away, and works that involve interpreting meanings and explaining facts will give way to simple visualizations—such as numbers, charts, and spatial distribution diagrams—and short expressions similar to poems, idioms, or advertising lines. Moreover, most of the data robots analyze will be produced or fabricated by robots themselves!

Is a world like this a heaven or hell? We should keep thinking about it.

V. Only Literacy, Not Skill

Yi Jing, the Book of Changes, is a classic highly venerated by ancient Chinese. It is said that the title *Yi Jing* possesses three different meanings: *simplicity*, *mutability*, and *persistency*. This concept can be used as a framework to answer the research questions proposed by this paper. Our conclusion is *simple*: first, as a persistent and universal principle, all paradigms, including those in digital humanities, inevitably shift; second, all research in humanities, both traditional and digital, must be grounded in data and require textual criticism; third, as a persistent and universal principle, the form of data, the potential problems that may arise when using data, and methodologies of textual criticism are constantly changing.

Consequently, in this digital era where "data is king" and robots play a dominant role, humanities scholars truly need to study both digital technology and humanities knowledge, combining both humanistic philosophy and "computational thinking." However, this does not guarantee that we can then soar unimpeded through the cosmos of the digital world. After all, these merely describe literacy, not skill. This new type of literacy may be called "digital textual criticism."

Bibliography

Aiden, E. and Michel, Jean-Baptiste. *Uncharted: Big Data as a Lens on Human Culture*. New York: Riverhead Books, 2013.

Bartscherer, T. and Coover, R. eds. *Switching Codes: Thinking through Digital Technology in the Humanities and the Arts*. Chicago and London: University of Chicago Press, 2011.

Bauerlein, M. *The Digital Divide: Arguments for and Against Facebook, Google, Texting, and the Age of Social Networking*. New York: Jeremy P. Tarcher, 2011.

Berry, D. M. *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012. [online]

Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Mass.: The MIT Press, 2015.

Cohen, D. J. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press, 2005.

Davidházi, Péter ed. *New Publication Cultures in the Humanities: Exploring the Paradigm Shift*. Amsterdam: Amsterdam University Press, 2014.

Deegan, M. and McCarty, W. *Collaborative Research in the Digital Humanities*. Farnham: Ashgate, 2012.

Deegan, M. *Text Editing, Print and the Digital World*. Farnham: Ashgate, 2009.

Deyrup, Marta Mestrovic ed. *Digital Scholarship*. New York: Routledge, 2009.

Eve, M. P. *Open Access and the Humanities: Contexts, Controversies and the Future*. Cambridge: Cambridge University Press, 2014.

Gold, M. ed. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. [online]

Hayles, N. K. and Pressman, J. eds. *Comparative Textual Media: Transforming the Humanities in the Postprint Era*. Minneapolis: University of Minnesota Press, 2013.

Inman, J. A., Reed, C. and Sands, P. eds. *Electronic Collaboration in the Humanities: Issues and Options*. Mahwah, N. J.: Lawrence Erlbaum Associates, 2004.

Jockers, M. L. *Macroanalysis: Digital Methods and Literary History*. Champaign: University of Illinois Press, 2013.

Leadbeater, C. *We-Think: Mass Innovation, not Mass Production*. London: Profile Books, 2009.

Lupton, Deborah. *Digital Sociology*. London and New York: Routledge, 2015.

Mandell, L. *Breaking the Book: Print Humanities in the Digital Age*. Chichester, West Sussex, UK: Wiley Blackwell, 2015.

Mayer-Schönberger, V. and Cukier, K. *Big Data: A Revolution That Will Transform how we Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.

McCarty, Willard. *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*. Cambridge: Open Book Publishers, 2010.

McGann, J. *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, Massachusetts: Harvard University Press, 2014.

Meyer, Eric T. and Schroeder, R. *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, Massachusetts: The MIT Press, 2015.

Moretti, F. *Distant Reading*. London: Verso, 2013.

Ragnedda, Massimo and Muschert, Glenn W. eds. *The Digital Divide: The Internet and Social Inequality in International Perspective*. New York: Routledge, 2013.

Ramsay, S. *Reading Machines: Toward an Algorithmic Criticism*. Champaign: University of Illinois Press, 2011.

Ridge, M. *Crowdsourcing our Cultural Heritage*. Farnham: Ashgate, 2014.

Rosenzweig, R. and Grafton, A. *Clio Wired: The Future of the Past in the Digital Age*. New York: Columbia University Press, 2011.

Sabharwal, A. *Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections*. Oxford: Chandos Publishing, 2015.

Schor, N., Weed, E., and Rooney E. eds. *In the Shadows of the Digital Humanities*. Durham, NC: Duke University Press, 2014. [A special issue of *Differences*, 25:1(2014)]

Schreibman, S., Siemens, R., and Unsworth, J. *A Companion to Digital Humanities*. Oxford: Blackwell, 2004. [online]

Schreibman, S., Siemens, R., and Unsworth, J. *A New Companion to Digital Humanities*. Chichester, West Sussex, UK: John Wiley & Sons Inc., 2016. [online]

Shirky, Clay. *Here Comes Everybody: The Power of Organizing without Organizations*. New York: Penguin Press, 2008.

Underberg, Natalie M. and Zorn, E. *Digital Ethnography: Anthropology, Narrative, and New Media*. Austin: University of Texas Press, 2013.

Vandendorpe, C. *From Papyrus to Hypertext: Toward the Universal Digital Library*. Champaign: University of Illinois Press, 2009.

Terras, M., J. Nyhan and E. Vanhoutte, eds., *Defining Digital Humanities: A Reader*. Farnham: Ashgate, 2013.

Wang, Fan-sen. *Fu Ssu-nien: A Life in Chinese History and Politics*. Cambridge; New York: Cambridge University Press, 2000.

Weller, M. *The Digital Scholar: How Technology is Changing Scholarly Practice*. London: Bloomsbury Academic, 2011.

Weller, T. *History in the Digital Age*. Abingdon: Routledge, 2012.

林富士主編，《「數位人文學」白皮書》。臺北市：中央研究院數位文化中心，2017。

曹銘宗撰稿，林富士總策畫，《遠見與承擔：中央研究院數位人文發展史(1984-2015)》。臺北市：中央研究院數位文化中心，2017。